

# Model-driven design of *Mycoplasma* as a vaccine chassis

Model-driven design of *Mycoplasma* as a vaccine chassis

Erika Gaspari

2021



Erika Gaspari

## Propositions

1. A serum-free medium supporting *Mycoplasma* growth is possible.  
(this thesis)
2. Without specific genetic modifications, a live attenuated *Mycoplasma* vaccine will cause an autoimmune reaction in some humans.  
(this thesis)
3. We do not have enough knowledge on microorganisms to assume that they are safe for human applications.
4. The keys of a successful model are spatio-temporal frame definition and integration of decisive biological assumptions.
5. Modellers should focus more on goal-oriented results rather than fanciness of methods.
6. Awards like "best female scientist" are far from a call for gender equality, because they indirectly assume that men and women cannot be evaluated under the same criteria.
7. The way to not realize there is a pandemic out there is to write a doctoral thesis.

Propositions belonging to the thesis, entitled

Model-driven design of *Mycoplasma* as a vaccine chassis

Erika Gaspari

Wageningen, 16 March 2021

# **Model-driven design of *Mycoplasma* as a vaccine chassis**

**Erika Gaspari**

## **Thesis committee**

### **Promotor**

Prof. Dr Vitor A.P. Martins dos Santos  
Professor of Systems and Synthetic Biology  
Wageningen University & Research

### **Co-promotor**

Dr Maria Suarez-Diez  
Associate professor, Laboratory of Systems and Synthetic Biology  
Wageningen University & Research

### **Other members**

Prof. Dr Pier Luigi Martelli, Alma Mater Studiorum Università di Bologna, Italy  
Dr Tjerko Kamminga, MSD Animal Health, Boxmeer  
Prof. Dr Hendriek C. Boshuizen, Wageningen University & Research  
Prof. Dr John Glass, J. Craig Venter Institute, Rockville (MD), USA

This research was conducted under the auspices of the Graduate School VLAG (Advanced studies in Food Technology, Agrobiotechnology, Nutrition and Health Sciences).

# **Model-driven design of *Mycoplasma* as a vaccine chassis**

**Erika Gaspari**

## **Thesis**

Submitted in fulfilment of the requirements for the degree of doctor

at Wageningen University

by the authority of the Rector Magnificus,

Prof. Dr A.P.J. Mol,

in presence of the

Thesis Committee appointed by the Academic Board

to be defended in public

on Tuesday 16 March 2021

at 1.30 p.m. in the Aula

Erika Gaspari

Model-driven design of *Mycoplasma* as a vaccine chassis,

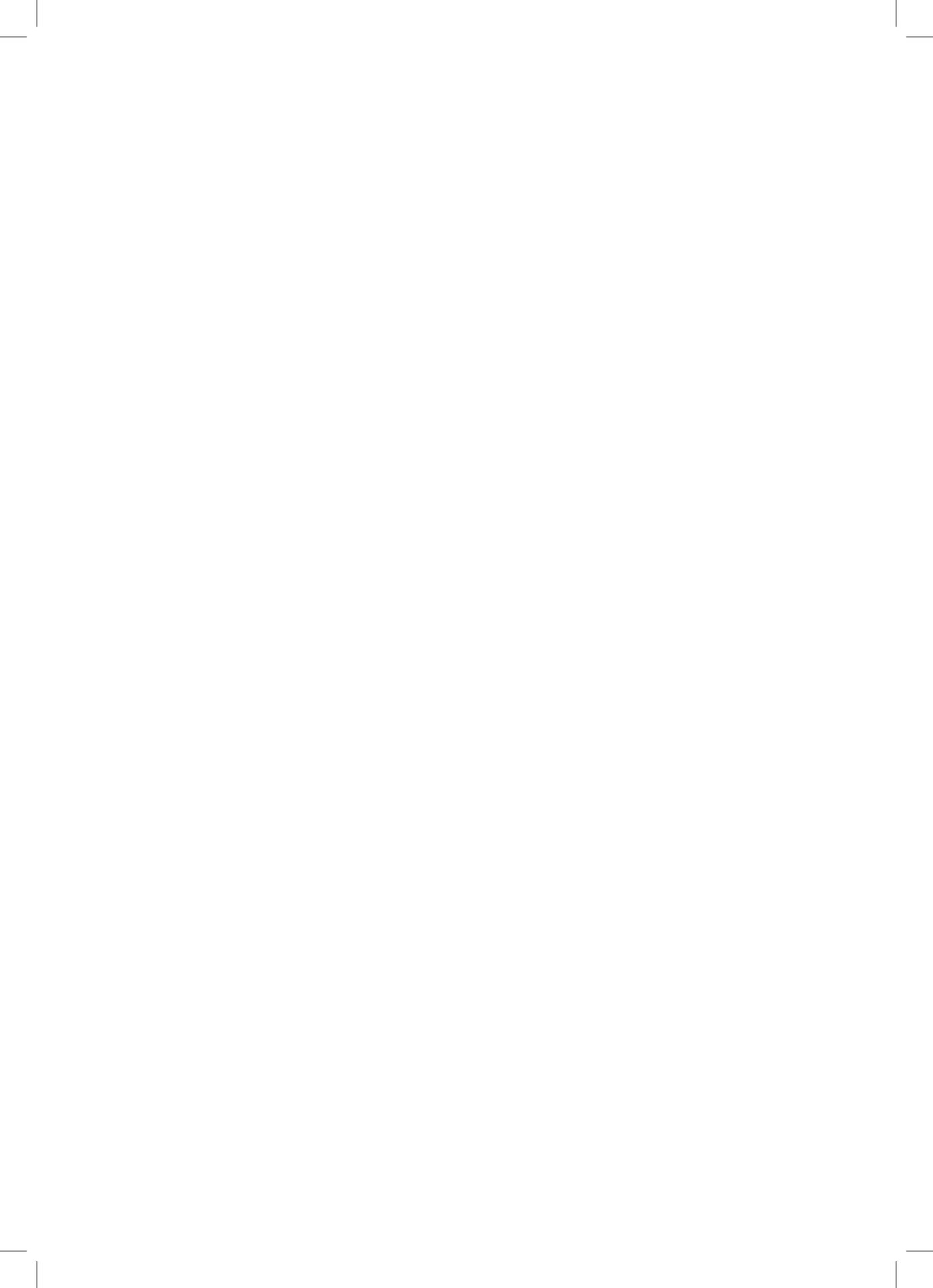
208 pages.

PhD thesis, Wageningen University, Wageningen, the Netherlands (2021)

With references, with summary in English, Italian and Spanish

ISBN: 978-94-6395-686-4

DOI: 10.18174/539593





*Chi ha provato il volo camminerà guardando il cielo,  
perché là è stato e là vuole tornare.*

**Leonardo da Vinci**



# Contents

<b>Chapter 1</b>	Introduction	<b>1</b>
<b>Chapter 2</b>	Model-driven design allows growth of <i>Mycoplasma pneumoniae</i> on serum-free media	<b>37</b>
<b>Chapter 3</b>	Design of a fatty acid-prototrophic <i>Mycoplasma pneumoniae</i> strain through metabolic modelling	<b>65</b>
<b>Chapter 4</b>	Galactocerebroside biosynthesis pathways of <i>Mycoplasma</i> species: an antigen triggering Guillain-Barré-Strohl syndrome	<b>89</b>
<b>Chapter 5</b>	Mathematical models of biosafety circuits designed to limit <i>Mycoplasma</i> growth	<b>119</b>
<b>Chapter 6</b>	Discussion	<b>155</b>
<b>Synopsis</b>	Summary	<b>183</b>
	Riassunto	<b>185</b>
	Resumen	<b>187</b>
<b>About me</b>	Publications & Patents	<b>193</b>
	Training activities	<b>195</b>
	Curriculum Vitae	<b>199</b>
<b>Acknowledgements</b>		<b>203</b>



# Chapter 1

## Introduction

Erika Gaspari



## The world of Mycoplasmas

Mycoplasmas are bacteria of the class *Mollicutes*, the simplest and the smallest free-living prokaryotes. They belong to the *Mycoplasmatales* order and the *Mycoplasmataceae* family. Their name, first used by Albert Bernhard Frank in 1889, comes from the Greek *mykes* (fungus) and *plasma* (formed), because they were identified in consequence of an altered state of plant cell cytoplasm acting as a fungus-like infiltration <sup>1</sup>. They have been later called "pleuropneumonia-like organisms" (PPLo) when isolated from bovines that had contracted a contagious pleuropneumonia <sup>2</sup>.

*Mycoplasma* species can be classified according to the organ where they are isolated from, into respiratory system, genitals or others, or depending on the species they affect. The mycoplasmas that use humans as primary host are *M. amphoriforme*, *M. buccale*, *M. faucium*, *M. lipophilum*, *M. orale*, *M. pneumoniae*, *M. salivarum* (respiratory-tract mycoplasmas), and *M. fermentans*, *M. genitalium*, *M. hominis*, *M. primatum*, *M. spermatophilum* (genitals mycoplasmas). However, mycoplasmas are in total 126 known species that infect not only humans, but a broad range of organisms, comprising farm animals, pets, wild animals and plants <sup>3-10</sup>. Respiratory tract-related mycoplasmas are associated with respiratory distress syndrome, bronchopulmonary dysplasia, intraventricular haemorrhage <sup>11-20</sup> and with atypical (or walking) pneumonia in adults, while the mycoplasmas isolated from the genitals may cause bacterial vaginosis, pelvic inflammatory diseases, cervicitis, infertility, preterm birth, spontaneous abortion and male sterility <sup>21-26</sup>.

Mycoplasmas have some unique characteristics: they lack a peptidoglycan cell wall, so the plasma membrane constitutes the outer boundary of the cell <sup>27-29</sup>, which is trilaminar, lipoprotein-rich and with a uniquely-high percentage of sterols, needed to maintain constant cell volume and to confer resistance to osmotic shocks. In consequence, they cannot be classified into Gram-positive or Gram-negative, although they are more similar to Gram-positive bacteria <sup>30,31</sup>. They can change their shape according to the environmental conditions, from round to oblong, with a diameter in the range of 0.2-0.3  $\mu\text{m}$  <sup>32</sup>. Moreover, they lack nucleus and other membrane-bound organelles. Their genetic material consists of a single DNA duplex and ribosomes are 70S type <sup>33</sup>. Due to their reduced genome, they evolved as unable to synthesize many required growth factors, that they acquire from the host <sup>34</sup>. It is the absence of cell wall that makes mycoplasmas resistant to many antibiotics, such as beta-lactams, which target the cell wall synthesis <sup>28,35</sup>. However, other antibiotics targeting rRNA (macrolides, tetracycline, ketolides and fluoroquinolone) are effective <sup>36,37</sup>. Mycoplasmas catabolize glucose or arginine as primary energetic sources; they live well in presence of animal serum, where they get the sterols from, to use them to synthesize their cell membrane <sup>38-41</sup>.

## Mycoplasmas as model systems to understand and re-programme life

The cell biology of *Mycoplasmas* has for long been interesting not only to mycoplasmologists but also as simple model systems for studying general biological problems, particularly those concerning membrane structure and function <sup>42</sup>. Although mycoplasmas are among the smallest living organisms and own a limited genomic endowment, they are much more complex than one would expect <sup>43,44</sup>.

Mycoplasmas have been often the subjects of studies about minimal genomes, genome transplantation and genome assembly, being in the focus on the development of the most cutting-edge technologies <sup>45,46</sup>. In 1999 Hutchison et al. conducted a comprehensive study on *M. genitalium*, which owns the smallest gene complement of any independently replicating cell so far identified, with a genome of only 580 kb <sup>47</sup>. Their results suggest that 55 to 73% of the protein-coding genes of this bacterium are essential under laboratory growth conditions <sup>48</sup>. This unusually high proportion of essential genes is a consequence of a minimal metabolism and little genomic redundancy. Therefore, *M. genitalium* genome is expected to be close to the minimal set of genes needed to sustain bacterial life. A study by Glass et al. identified 382 essential genes among the 482 protein-coding genes of *M. genitalium* <sup>49</sup>. Two years later, the first synthetic genome was created, consisting of 583 kb and being named "*M. genitalium* JCVI-1.0" <sup>50</sup>. This project presented an impressive technical feat, joining together chemical synthesis and recombination cloning methodology, and definitely represents an important milestone in synthetic biology. Moreover, Lartigue et al. showed that, because mycoplasmas are devoid of cell walls, transfer into the cell of external naked DNA is possible, effectively changing one species into another <sup>51</sup>. About 10 years ago, a fully synthetic reduced genome from *Mycoplasma mycoides* was transplanted into a *Mycoplasma capricolum* recipient <sup>52</sup>. With the aim of building a minimal genome that includes only the genes essential for life, this genome was reduced from 1079 kb to 531 kb (473 genes, of which 149 of unknown function) and named JVC-syn3.0 <sup>53</sup>. These efforts allowed futuristic techniques to become actual applications, which can be used in the future to complete redesign genomes for industrial and medical products <sup>54</sup> and to tailor genomes to specific functions and environments through computational tools for model-driven genome design <sup>55</sup>.

Genome reduction and its consequences have been for long explored not only in *M. genitalium*, but also in *M. pneumoniae*, a bacterial model and one of the best characterized minimal bacteria due to its natural tropism for the human respiratory tract <sup>56,57</sup>. Being one of the simplest organisms able to reproduce outside of host cells, it is indeed used to understand biological organization at a system-level. Among the great number of studies conducted on this organism, three must be highlighted. A main study is the analysis of protein interactions through mass spectrometry conducted by Kuhner et al., which provided insights on the structural information on *M. pneumoniae* proteins and how they interact <sup>58</sup>. Yus et al. mapped the whole metabolic network of *M. pneumoniae* and validated it experimentally <sup>44</sup>. Guell et al. revealed how *M. pneumoniae* uses noncoding RNAs for complex gene regulation <sup>59</sup>.

## ***Mycoplasma pneumoniae***

In this thesis, I focus specially on *M. pneumoniae*, causative agent of atypical pneumonia in humans, whose respiratory-tract epithelium is affected, or occasionally erythrocytes and urogenital cells<sup>60,61</sup>. *M. pneumoniae* is even smaller than most of the *Mycoplasmas*, with a length of 1000-2000 nm and a width of 100-200 nm, being in consequence undetectable by light microscopy. It has a structure related to attachment to the mammal cells it parasitizes<sup>60</sup>. This organelle is made by the proteins HMW1-HMW5, P1, P30, P56, P90 and P116<sup>62,63</sup>. After attaching to the host cells, *M. pneumoniae* evades the immune system by fusion with the host cell, consequent intracellular localization and adjustment of the cell membrane to mimic the composition of a mammal's one<sup>64</sup>. Its genome was sequenced in 1996 and consists of 816394 base pairs, with 687 protein-coding genes<sup>65,66</sup>. *M. pneumoniae* has peculiar features regarding its energy consumption: indeed, only 12-29% of its energy metabolism is directed towards cell growth, while the rest is used in maintenance<sup>67</sup>. This characteristic is indicative of an adaptation to parasitic lifestyle, which prioritizes surviving over population growth.

Although in most cases the symptoms of *M. pneumoniae* infection are mild, they persist for weeks and in some cases require hospitalization, with death occurring especially in elderly or immunocompromised people<sup>68-73</sup>. Extrapulmonary manifestations occur in up to 25% of *M. pneumoniae* cases, with neurological manifestations among the most common ones. Many *M. pneumoniae* outbreaks were documented in the literature since the 60s, with an epidemic cyclic trend of 3-5 years<sup>74-78</sup>. The diagnosis is often complicated: symptoms are delayed and very similar to the ones of other pulmonary conditions<sup>79-81</sup>. Since an implemented vaccine against *S. pneumoniae* is available<sup>82</sup>, *M. pneumoniae* may become the most prevalent bacterial respiratory pathogen and an effective vaccine is desirable, also considering diagnostic difficulties and inefficiency of many antibiotics against its infections.

*M. pneumoniae* can be grafted with selected epitopes from respiratory pathogens to be used as a vaccine. Moreover, whole synthetic pathways can be added to deliver therapeutic molecules directly in the lungs<sup>83</sup>. However, while gene replacement by homologous recombination has been widely used for genome editing of *M. genitalium*, the same approach showed low efficiency in *M. pneumoniae*<sup>84</sup>. Therefore, due to its potentiality in the application of biomedical treatments, efficient genome engineering strategies have been developed, such as gene replacement by an heterologous recombinase from *B. subtilis*<sup>85</sup> or the newest RAGE<sup>86</sup>.

## Mycoplasmas as vaccines

Effective vaccines against mycoplasmas are also needed to avoid massive antibiotic usage in large-scale animal farms and to obviate the diagnostic and treatment difficulties. Indeed, a diagnosis is not always possible due to the big difficulties in isolating the bacterium<sup>87,88</sup> and, when it is possible, only antimicrobials such as tetracyclines and macrolides, which are able to reach high concentrations in the mucus of the respiratory tract, are effective *in vivo*<sup>89-96</sup>. *Mycoplasma* infections result in great economic losses: infected animals show higher mortality rate and grow much slower. This problem affects in particular the cattle, swine and poultry industry, due to the reduced weight gain and poor feed efficiency of these animals<sup>97-102</sup>.

Vaccines are designed to teach the immune system how to fight off certain kinds of germs. There are several types of vaccines, and they can be classified in four main categories: live-attenuated, inactivated, subunit/recombinant/polysaccharide/conjugate, toxoid. Live-attenuated vaccines consist of a weakened (i.e. "attenuated") form of the germ that causes a disease. They create a strong and long-lasting immune response, but because they effectively contain the germ, they might cause adverse effects especially in people with weakened immune systems and health issues. Inactivated vaccines, instead, use a killed version of the germ. The protection they provide is usually weaker than the one of live-attenuated vaccines and several doses need to be injected over time to ensure immunity. The third category of vaccines (subunit/recombinant/polysaccharide/conjugate) comprise those that utilize a specific piece of the germ (proteins, sugars, capsid), allowing a targeted immune response. Adverse effects are a rare problem for this type of vaccines, even in people with health issues, but boosters are needed to ensure protection. Finally, toxoid vaccines use a toxin made by the germ, creating a selective and targeted immunity to that part of the germ that causes the disease, but not to the germ itself. They often need booster shots to allow ongoing protection. Mycoplasmas have been used in several vaccine attempts along the years: several species were inactivated with the aim of developing vaccines for humans, poultry, cattle, swine, sheep and goats, but only some of them reached a clinical trial phase and, in a few cases, were commercialized.

A first experimental vaccine made with *M. pneumoniae* was developed in 1965 and tested on 36 human subjects<sup>103</sup>. Growth-inhibiting antibodies were observed in 10 over 19 volunteers, of which one became ill. Of the non-responding, 7 volunteers developed a respiratory tract disease. Other vaccine studies followed in the 70s, when the mechanisms of immunity to *M. pneumoniae* were investigated by evaluating different vaccination procedures in animal models<sup>104</sup>. The efficacy of an inactivated *M. pneumoniae* vaccine was evaluated in a double-blind study of Marine Corps recruits, obtaining no statistically significant difference between the two groups in protection against bronchitis<sup>105</sup>. The most recent literature review on *M. pneumoniae* vaccine studies<sup>75</sup> reports a total of 67268 subjects were tested in six different clinical trials, and adverse events were observed in 15 studies. Live attenuated vaccines never reached the stage of clinical trials for efficacy evaluation.

*M. gallisepticum* is an avian pathogen, primarily affecting chickens and turkeys<sup>106</sup>, causing a chronic respiratory disease in the first ones and an infectious sinusitis in the second ones<sup>107,108</sup>. Several types of vaccines were developed as avirulent for chicken and non-pathogenic for turkeys: bacterin-based, recombinant proteins, killed or live attenuated. *M. gallisepticum* strains vary a lot in infectivity and virulence. A *M. gallisepticum* S6 strain was made avirulent for both chickens and turkeys and deposited in 1991 in patent US506467A, being named *Mycoplasma* strain Intervet 6/85. The live attenuated vaccine Vaxsafe MG uses a strain called ts-11, then improved to be GapA+ in a strain called ts-304<sup>109–111</sup>. All available vaccines, with their names, types, strains and manufacturers, are summarized by Ishfaq et al.<sup>112</sup>. While these efforts have been successful in reducing the severity of the disease and the horizontal and vertical transmission, some of them have adverse effect, or provide only partial/transient immunity<sup>113</sup>. Moreover, the bacterium evolved in consequence of the vaccine improvements, with the worsening of the disease<sup>114</sup>.

*M. bovis* is a primary cause of calf pneumonia, arthritis, mastitis, keratoconjunctivitis and other conditions<sup>115</sup>. In Europe, *M. bovis* is responsible of 25–33% of cattle pneumonia outbreaks. Moreover, it is becoming resistant to the few antibiotics that are effective on Mycoplasmas, due to the development of antimicrobial resistance<sup>116</sup>. For this reason, several vaccine attempts exist, with the first one reported in 1987<sup>117</sup> consisting of killed *M. bovis* together with other respiratory tract pathogens (PI3 virus and *M. dispar*). This vaccine attempt was abandoned and substituted with formalin-inactivated *M. bovis* and *M. haemolytica*, which reduced the losses. However, formalin-inactivated vaccines have shown to provide only transient protection, and in one case led to increase in disease<sup>118</sup>. Other attempts used saponin as an adjuvant, which unfortunately lyse the cholesterol-rich membrane of *M. bovis* and therefore is not a good candidate<sup>119–122</sup>.

A first inactivated vaccine against *M. agalactiae*, causing contagious caprine pleuropneumoniae, was reported in 1995, when it was tested on goats<sup>123</sup>. A few years later, five sets of vaccines inactivated through different procedures were tested on sheep, but only two of them (phenol- and saponin- inactivated mycoplasmas) turned out to be effective<sup>124</sup>. Instead, Buonavoglia et al. studied the persistence of antibodies in two groups of sheep inoculated with a vaccine in aluminium hydroxide adjuvant and a vaccine in mineral oil adjuvant, finding out this last was giving a much more persistent immunity and full protection<sup>125–127</sup>.

It is worth to mention the vaccines using *M. mycoides* subs. *mycoides* – for contagious bovine pleuropneumoniae (CBPP), and the one using *M. capricolum* subs. *capripneumoniae* – for contagious caprine pleuropneumoniae (CCPP). The first one is a live attenuated vaccine<sup>128</sup>, while the second is a bacterin vaccine<sup>129</sup>, and they were developed respectively in the 60s and in the 80s. The necessity of continuous boosting, together with the gravity of the diseases, would suggest an improvement of these two vaccines is needed<sup>130</sup>.

*M. hyopneumoniae* is the etiological agent of swine enzootic pneumonia and the most sequenced porcine *Mycoplasma*<sup>131–133</sup>; its infection is prevalent in pig production worldwide<sup>30</sup> and damages the cells of pigs from all age groups<sup>134</sup> by predisposing the animal to second invaders or through production of toxic compounds such as hydrogen peroxide and

superoxide radicals<sup>34</sup>. The related disease causes chronic cough, dull hair coat, retarded growth and unthrifty appearance. Although the disease causes little mortality, affected swine are often prone to secondary infections by opportunistic pathogens, resulting in death or stress<sup>87,88,135</sup>. These economic losses were estimated as 200 to 250 million dollars per year in the USA. Vaccination is the most cost-effective strategy for control and prevention of the disease<sup>136,137</sup>.

Several vaccines against *M. hyopneumoniae* were commercialized, with diverse administration schemes depending on type of herd, production system and infection pattern in the farms<sup>138</sup>. All vaccinations are based on inactivated bacterins (killed *Mycoplasma* cultures), consisting of adjuvant whole-cell preparations that induce (partial) protection against development of gross lesions. They improve the daily weight gain and the feed conversion ratio, not only reducing lung lesions and mortality rate, but also shortening the time to reach slaughter weight and, in some cases, improving the meat quality<sup>139</sup>. Until about two decades ago, the protection induced by *M. hyopneumoniae* vaccine attempts was incomplete and the vaccines were not preventing colonisation<sup>138</sup>, although they significantly reduced clinical symptoms and lung lesions. However, improvements were made in the post-genomic era,<sup>140</sup> probably due to the identification of the different strains – J, 232, 7448 and 168 – through sequencing; they have indeed shown a large diversity at genomic, proteomic and virulence levels<sup>135,141,142</sup>. An inactivated *M. hyopneumoniae* bacterin in combination with adjuvant mixture, providing immunity after single administration, was filed in 2007 as swine pneumonia vaccine in patent number US7169394B2. The current commercial vaccines are mostly based on the J-strain and are summarized by Simionatto et al.<sup>140</sup> and Tao et al.<sup>143</sup>.

The disadvantages of existing *Mycoplasma* inactivated vaccines are multiple: they are very expensive, do not protect against infection (i.e. carriers of the virulent *Mycoplasma* remain in the organism) and, in case of animal vaccination, each animal has to be injected individually<sup>144,145</sup>. Live attenuated vaccines consist of a virulence-reduced chassis of the pathogen, which is kept viable but harmless. Although they require fresh media, they have the advantage of activating all the phases of the immune system, provide a more durable and quicker immunity with less frequent boosters, lower costs than the other vaccine types, and beneficial heterologous effects<sup>146–149</sup>. Especially for *Mycoplasma* vaccines, these non-specific effects might be very beneficial: they are expected to increase protection against non-targeted infections, reduce mortality and morbidity rates. Because mycoplasmas have high mutation rates and ability of quick evolution according to the host, off-targets effects might also represent a form of immunomodulation, with the immune system becoming able to handle the changes. The existing live *Mycoplasma* vaccines are not without issues and they need improvement.

All live attenuated *Mycoplasma* vaccines are summarized in [Table 1](#).

**Table 1.** List of live attenuated *Mycoplasma* vaccines. Each organism and strain are reported with the status of the vaccine (old study – abandoned for infeasibility/adverse reaction, in study, in commerce – with name of the vaccine or of the producer) and target organism.

<b>Bacterium</b>	<b>Strain</b>	<b>Status of the vaccine</b>	<b>Target Organism</b>	<b>References</b>
<i>M. gallisepticum</i>	ts-304	In study	Poultry	130
	ts-11	In commerce (Vaxsafe MG)	Poultry	150,151
	F	In commerce (Cevac MG)	Poultry	152
	6/85	In commerce (Nobilis MG)	Poultry	153,154
<i>M. synoviae</i>	MS-H	In commerce (Vaxsafe MS)	Poultry	155
<i>M. hyopneumoniae</i>	168	In commerce (Jiangsu Ac.)	Swine	155–159
	RM48	In commerce (EBVAC)	Swine	143
<i>M. mycoides</i>	Subsp. <i>mycoides</i> LC	In commerce (Pendik Inst.)	Goats	160,161
	Subsp. <i>mycoides</i> SC	In commerce (T1/44)	Cattle	162
<i>M. agalactiae</i>		In commerce (Pendik Inst.)	Sheep	163
<i>M. bovis</i>	<i>No live attenuated vaccine available</i>		Calf	164
<i>M. pneumoniae</i>	Mac-K	Old study	Humans (test on rodents)	165
	FH-P24	Old study	Humans (test on rodents)	166

Vaccination against many mycoplasmas that infects humans, pets and farm animals (e.g. *M. bovis*) are still missing, while those vaccines that are in commerce are outdated or produced by a challenging process. An effective and efficient vaccine against mycoplasmas should be cheap, as their occurrence seems to be prevalent in developing countries<sup>130,167,168</sup>. Moreover, it should grow at a reasonable rate in serum-free medium, to avoid the excessive production costs and the presence of undefined components. Optimally, a *Mycoplasma* vaccine should be flexible, modular and applicable for several mycoplasmas (i.e. “broad-range”). Moreover, it must be safe both in terms of adverse effects for the host and as a genetically-engineered microbe that could potentially be released to the environment.

The fact that mycoplasmas are both pathogenic and genetically of such small size makes them prime candidate for engineering their genome to create a new type of vaccine, i.e. a broad-

spectrum vaccine that can be deployed for a range of animal hosts. Thanks to the extensive biological knowledge on *M. pneumoniae* and on cutting-edge synthetic biology methodologies, such as genome-editing and transplantation tools, a universal *Mycoplasma* chassis that can be used as single- or multi-vaccine in several hosts shows high developmental potentialities ([www.mycosynvac.eu](http://www.mycosynvac.eu) – Figure 1). Such a chassis must be free of virulence determinants from *M. pneumoniae*, must be optimized for fast growth on serum-free medium and must express heterologous antigens from one or more pathogens, and biological adjuvants to create a targeted vector vaccine. Therefore, potentially this chassis could not target only mycoplasmas but also viruses, by expression of a range of epitopes, or could be deployed for other therapeutical applications (cell reprogramming, cell therapy, antibiotics, etc.).

1

Cutting-edge synthetic biology techniques can enormously benefit from the support of computational and mathematical models in the design phase of several aspects: first of all, the large-scale production assumes the model-driven design of a medium that can support the development. Second, gene essentiality must be studied with genome-scale metabolic models to identify pathogenic factors. Computational and mathematical models can also be used for strain design, i.e. the chassis, to ensure it is suitable for vaccine large-scale production, avoiding adverse effects in the hosts. Finally, model simulations can help the design of biocontainment strategies that aim at limiting the growth of the vaccine strain when outside of the desired conditions.

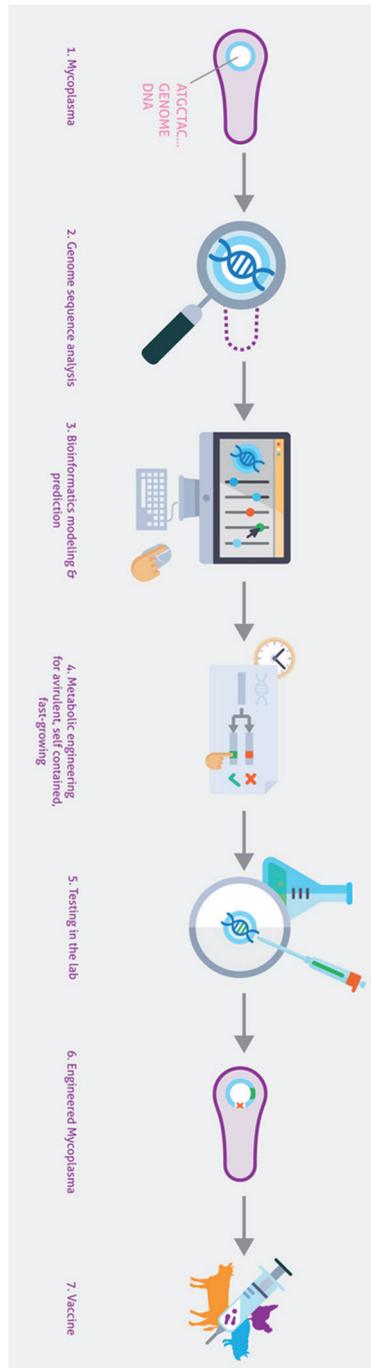


Figure 1. MycoSynVac workflow (from MycoSynVac website: [www.mycosynvac.eu/content/about](http://www.mycosynvac.eu/content/about)).

## Computational models to improve *Mycoplasma* growth

One of the main issues addressed in this thesis is the slow growth of *M. pneumoniae*. Indeed, this bacterium has lost many biosynthesis functions during its evolution, which is reflected in its minimal genome<sup>169</sup>. Yet, *M. pneumoniae* still requires a lot of lipids for the assembly of its cell membrane<sup>170</sup>, which, in lack of cell wall, represents the cell barrier and ensures the surviving of the microorganism. *M. pneumoniae* infects the host, evade its immune system by mimicking its cell membrane<sup>171</sup> and takes from it all the compounds it requires for surviving. However, when cultivated *in vitro*, all these compounds must be supplemented in the growth medium. For this reason, *M. pneumoniae* has always been grown in serum-rich media, which are not suitable for vaccine development and large-scale production both because their high cost and for their non-defined composition (which can vary from batch to batch, thereby resulting in fluctuations in quality). Computational modelling approaches can be used to design serum-free media that allow microbial growth at a rate comparable to the one obtained in rich medium.

Genome-scale models (GEMs) are a computational description of gene-protein-reaction (GPR) associations for an entire organism's metabolism: reactions are stoichiometry-based and mass-balanced, formulated on the base of experimentally obtained gene annotation data. They allow to perform whole-cell analysis of an organism's metabolic capabilities, simulating the conditions of a chemostat. Genome-scale metabolic models allow to give as input external elements such as media components, and, through simulations of the metabolic fluxes, link them to the growth of the bacterium, which is described as biomass yield in the objective function of the model<sup>172-175</sup>.

The first GEM was published in 2000 and reconstructed the metabolism of *Haemophilus influenzae* strain RD<sup>176</sup>; subsequently, GEMs of model organisms such as *Escherichia coli*<sup>177-179</sup> and *Saccharomyces cerevisiae*<sup>180,181</sup> have been developed. GEMs reconstruction techniques evolved in consequence of the rapid improvement of sequencing and omics techniques. As of 2019, 6239 GEMs exist, of which 5897 are of bacterial organisms, 127 of archaea and 215 of eukaryotes<sup>182</sup>. Relevant applications of GEMs include strain development for the production of bio-based chemicals and materials<sup>183-187</sup>, drug targeting in pathogens<sup>188-194</sup>, prediction of enzyme functions<sup>195-200</sup> and interactions among organisms<sup>201-204</sup>.

One of the main computational approaches that are applied to GEMs is constraint-based modelling. Flux Balance Analysis (FBA) is a constraint-based modelling technique that allows the prediction of metabolic fluxes through linear-programming<sup>205-208</sup>. It consists of a mathematical representation of the metabolic reactions, where the stoichiometric coefficients of each reaction are tabulated in a matrix. Metabolic flows are indeed imposed by the stoichiometric coefficients themselves, without requiring kinetic parameters. Constraints consist of lower and upper bounds for reaction fluxes. FBA consists of an optimization of an objective functions when these constraints are applied, assuming steady-state condition, that simulate a chemostat. Optimal value of the selected objective function and a flux distribution compatible with it is then found within the allowable solution space. FBA can be applied to

GEMs for multiple purposes: simulation of knock-outs and knock-ins<sup>209–216</sup>, prediction of cofactor yields<sup>217–220</sup>, completion of genome-scale model reconstructions<sup>221–226</sup>, prediction of gene essentiality<sup>227–232</sup>, maximization of the production of a biomolecule of interest<sup>233–237</sup> or growth prediction and analysis of conditions for maximizing the growth rate<sup>238–245</sup>. It is possible that more than one flux solution leads to the same growth rate; for this reason, it is important to apply as many constraints as possible, especially experimentally-measured uptake rates to constraint exchange reactions, and to combine FBA with Flux Variability Analysis (FVA), that explores the metabolic space obtaining a minimum and maximum possible flux for each reaction that is compatible with the optimal value of the objective function<sup>246</sup>.

Several GEMs have been constructed for *Mycoplasmas*: models of *M. genitalium*, *M. hyopneumoniae* and *M. agalactiae* were built to analyse the bacterial fastidious nutrient requirements and reduced metabolic capabilities due to the reduced genomic content<sup>247–250</sup>. These models were used to design strategies to improve the growth of the bacteria, by addition of critical compounds to the culture media or knock-ins prediction. The model of *M. gallisepticum* was instead built to validate a semi-automated curation methodology for genome-scale metabolic modelling<sup>251</sup>. Finally, model iJW145 of *M. pneumoniae* was built to analyse its metabolic flux distribution and to perform *in silico* knock-out studies<sup>67</sup>.

A customized medium supplementation is not the only approach to improve microbial growth while reducing vaccine production costs: genetic engineering can be used for empowering organisms. Candidate knock-ins can be discovered and investigated through genome-scale modelling and gap-filling algorithms. Gap-filling algorithms are commonly applied to GEMs to “fill their gaps”, i.e. missing reactions that make the metabolic network disconnected or the maximization of the objective function infeasible: although with sequencing technologies our knowledge on the metabolism of many organisms is increasing, gaps in the metabolic networks may exist due to missing reactions, unknown pathways, unannotated or misannotated genes<sup>252–255</sup>. Gap-filling tools have been formulated to improve the network connectivity by i) detecting gaps, ii) change model content (add/remove reactions, change biomass composition, change reaction reversibility), iii) identify responsible genes. Orth & Palsson reviewed all the existing classical gap-filling methods as per 2010, highlighting the rise of a growing number of studies on the usage of these algorithms for different purposes than metabolic network reconstruction, e.g. discovering new metabolic capabilities and gene functions<sup>256</sup>. Since 10 years ago, many more gap-filling algorithms have been released, applied to a broader range of applications<sup>253,257–261</sup>. It is not arduous to understand the potentiality of this algorithmic class and all the possible applications to genome-scale models, for example regarding knock-in discoveries .

## Mathematical models to improve vaccine’s safety

Safety is always a concern when developing a vaccine. In the case of a potential vaccine based on *M. pneumoniae* one such concern is related to finding that this microbe has been linked to a post-infectious autoimmune disease named Guillain-Barré-Strohl syndrome<sup>262,263</sup>, a

neurological disorder. The bacterium triggers the autoimmune response through the exposure of a glycolipid called galactocerebroside.

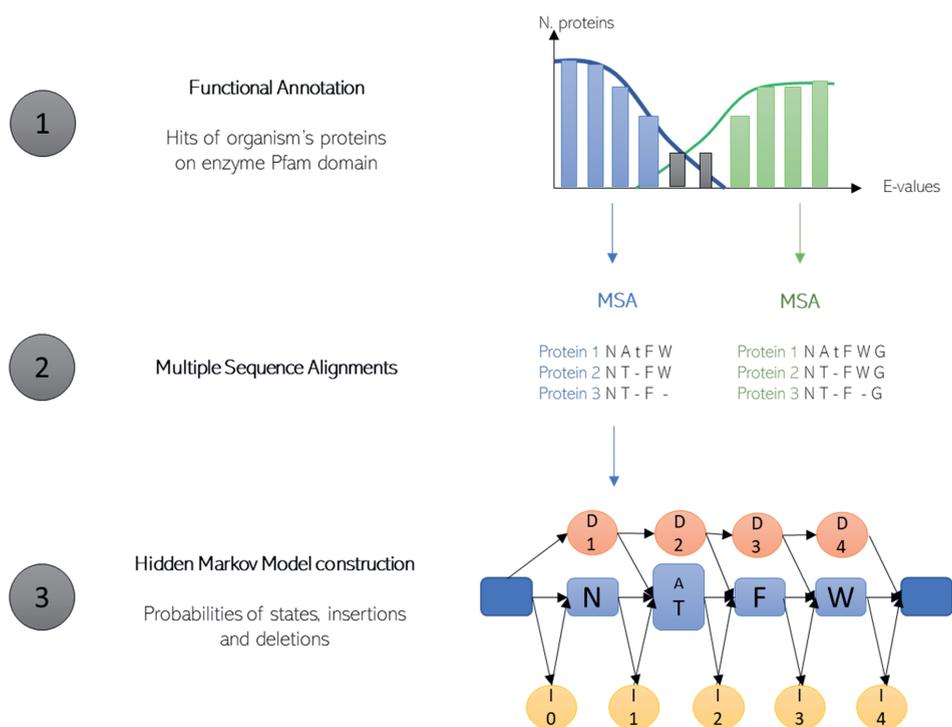
The investigation of this process requires computational protein domain analyses, complemented with mathematical models that allow the identification of the protein motif responsible for the synthesis of the compound. Several methods can be used to build protein motifs relative to a specific function: recently, many machine learning and deep learning methods have been developed for this purpose, mostly with the aim of identifying binding sites<sup>264–267</sup>. For our purpose, i.e. the identification of a conserved protein sequence indicative of a specific enzymatic function, we have chosen Hidden Markov Models (HMMs) as mathematical modelling approach.

HMMs are statistical models that are used to describe the evolution of observable events (“symbols”) that depend on internal factors, which are not directly observable (“states”). The observable symbols depend on the hidden states, which form a Markov chain, governed by transition probabilities. States are chosen ensuring that the next state is determined based on the current state only: for this reason, unobservable quantities must be included. They have been applied to a wide range of issues, comprising speech recognition and digital communication<sup>268</sup>, gene prediction<sup>269</sup>, sequence alignments<sup>270</sup>, modelling DNA sequencing errors<sup>271</sup>, protein structure prediction<sup>272</sup>, ncRNA identification and annotation<sup>273</sup>, RNA folding<sup>274</sup> and many others<sup>275</sup>.

The usage of HMMs is very common for building protein motifs, which can be defined as amino acid sequences shared among different proteins: they can be seen as blocks of conserved sequence within a specific protein family, which is a group of evolutionary-related proteins that will show more or less conserved sequenced among each other. Motifs are recognizable regions that may be defined a unique chemical or biological function. Protein motifs are built as probabilistic models from multiple sequence alignments (MSA), which align three or more biological sequences that are thought to have an evolutionary relationship. MSA can be used for multiple purposes, among which the assessment of sequence conservation of protein domains. HMMs can be used to convert multiple sequence alignments into position-specific scoring systems, which can be then linked to a specific function or utilized as annotation of new functions.

The Pfam database is a large collection and classification of protein families that includes their annotations and MSA, generated using HMMs<sup>276</sup>. Pfam can be used to search specific proteins, to identify new targets for structure determination, to analyse amino acid sequences, to build phylogenetic trees or for functional annotation of genomic data<sup>277–285</sup>. Sometimes, protein families can show differences in the conserved domain regions: van Dam explains how to evaluate predicted occurrences of protein domains and their associated E-values, which might show a bimodal distribution, i.e. two peaks can be associated to two divergent paths<sup>286</sup>. In this case, two new protein profiles may be created. The general process to build new protein profiles from Pfam domains through HMM from MSA is showed in [Figure 2](#). The states of the HMM consists of amino acids. Deletions are insertions are also modelled. The passage from one state to the other is built according to transition probabilities to the next amino acid, to a

deletion or to an insertion. Each state observation depends uniquely on the previous state and on the transition probabilities.



**Figure 2.** Construction of a new HMM protein profile from an existing Pfam domain. 1) Functional annotation of the organism's proteins: enzyme sequences will match the target Pfam domain with different E-values, uncovering hidden differences. 2) Multiple sequence alignment of the sequences from the peaks obtained in step 1. 3) Hidden Markov Model construction on the MSA of interest.

Safety concerns related to live attenuated vaccines might also be related to the nature of Genetically Modified Organisms (GMOs). In general terms, genetic engineering enables scientist to find individual genes that control particular characteristics, separate them from the original source and transfer them into the cells of an organism. As live attenuated vaccine, the chassis must include genetic modifications to render the microorganism non-pathogenic for the host, but only able to stimulate immune response. Genetic engineering has numerous potential applications, as well as many potential risks: when introduced into the environment, GMOs could play more pronounced ecological roles than the wild types, as they are able to reproduce themselves and establish as a persistent population. Moreover, the knowledge of the integrated genes in the organism of provenance does not ensure the same function in the new organism: interferences with other functions and even the entire genome cannot be

predicted and can possibly cause unintended side effects. Potential risks are summarized by Prakash et al.<sup>287</sup> in:

- Genetic contamination and interbreeding: the novel traits might confer selective advantage to the wild types.
- Competition with natural species: if the GMO grows faster than the natural species, it might have competitive advantage over the wild type.
- Increased selection pressure on target and non-target organisms: target and non-target species might adapt to the introduced changes.
- Ecosystem impact: the effect on a single species might extend to the ecosystem.
- Impossibility of follow-up: once a GMO has been introduced into the environment and issues arise, it is impossible to eliminate them.
- Horizontal Gene Transfer (HGT) of recombinant genes to other microorganisms: other organisms might acquire the foreign genes via transformation, transduction or conjugation, which might confer novel traits.

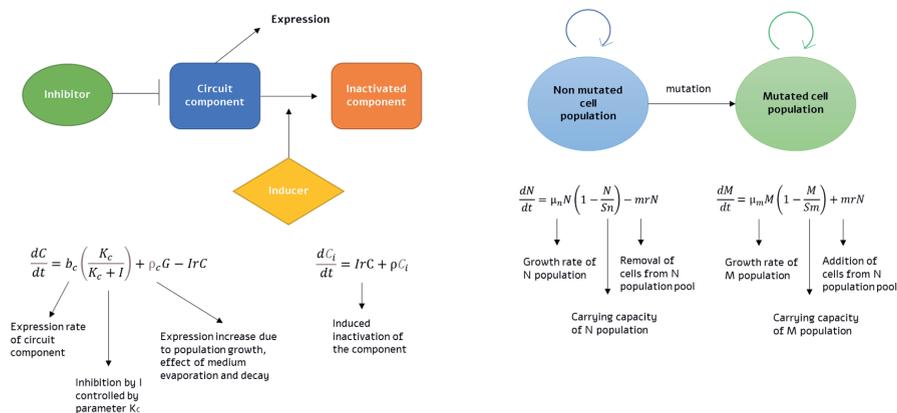
1

Precautions must be taken against these potential risks in the design phase of the GMO, as a live attenuated vaccine is. For this reason, synthetic circuits may be implemented into the organism's genome to impede its growth outside of desired conditions in the long term, thus providing an efficient biocontainment strategy. Therefore, the biosafety circuits may be designed to be induced by an external factor, i.e. "inducer", that activates a cascade response leading to the death of the organism; for this reason, they are also called "killing switch".

The usage of mathematical models to explore biosafety circuits behaviour within the host over time is not very common. However, a few cases are reported in literature: the first toggle switch, an oscillating network called Repressilator, and layered logic gates were constructed in *E. coli* based on mathematical models. The bistable toggle switch, composed of two repressors and two constitutive promoters, was modelled through a system of differential equations describing the evolution over time of the concentrations of the two repressors, as dependent on their synthesis rate, their cooperativity and a dilution/degradation effect<sup>288</sup>. The Repressilator, consisting of three transcriptional repressors, was studied through Fourier analysis to investigate the conditions of convergence towards a stable or unstable steady state and their stochastic fluctuations<sup>289</sup>. Finally, a genetic program that integrates circuits based on biochemical interactions was built as set of layered transcriptional AND gates, i.e. the output promoter of an upstream circuit serves as input promoter for a downstream circuit<sup>290</sup>.

Ordinary Differential Equations (ODEs) models are commonly used in Synthetic Biology to study the dynamics of the concentrations of mRNA and proteins for their suitability to non-linear systems and their dynamics<sup>291</sup>. In the case of biosafety circuits, two main types of simulations are needed: i) dose-responses of the organisms and the circuit's elements to the inducer concentrations; ii) escape mutants that could introduce mutations inactivating the circuit's effects. To predict dose-responses, the ODE system must focus on the characteristics and the evolution of the circuit's components, while to study escape mutants a population-modelling approach is needed. The different ODE modelling approaches are summarized in [Figure 3](#). In the first approach, circuit components (i.e. transcription factors) are modelled through ODEs that follow a Michaelis-Menten or Hill kinetics<sup>292,293</sup>, as dependent on inhibitor

concentration, inducer concentration, rate of expression and degradation/dilution. In the second approach, the ODE system describes an evolution independent on the circuit's characteristics, where the cell population is modelled as composed by two populations (i.e. mutated and non-mutated). This approach is based on the Lotka equation<sup>294</sup> and describes the evolution over time of two populations growing until a certain carrying capacity, where one population slowly substitute the other by removing elements from its pool.



**Figure 3.** Mathematical modelling of biosafety circuits. Left panel). The target of the simulations is an exploration of circuit conformations, with their effect on the bacterial growth, and the estimation of dose-responses (inducer concentration – component expression or inducer concentration – bacterial growth). A system of ODEs describes the process such that the component  $C$  undergoes expression under the parameter  $b_c$ ; the expression is inhibited by inhibitor  $I$  under parameter  $K_c$ . A parameter  $\rho$  controls i) the increase of component expression due to increase of cell population, ii) the effect of medium evaporation – component concentration increases with volume reduction, iii) component's spontaneous degradation. The component undergoes inactivation (increase of  $C_i$  pool) when the inducer  $I$  is added to the medium. Right panel). Mathematical description of the appearance of escape mutants, which will not be affected by the biosafety circuit. The observed cells are divided in two populations: a non-mutated  $N$  and a mutated  $M$  ones. Each population grows with a rate  $\mu$  until a carrying capacity  $S$  is reached. The mutation rate  $mr$  governs the conversion of  $N$  into  $M$  cells: over time, the  $N$  pool is reduced and the  $M$  cells increase in number.

## Objectives of this thesis

The objectives of this thesis were to develop and use modelling frameworks for the design of various aspects of a broad-spectrum live attenuated pneumonia vaccine through modelling frameworks. These aspects comprise the development of a serum-free medium to sustain the large-scale production of the vaccine, chassis modification to improve the microorganism's growth rate, investigation on vaccine's side effects and design of biocontainment strategies for the genetically modified organism. In **chapter 2**, I used metabolic modelling techniques

to design serum-free media to sustain the growth of *M. pneumoniae*. I expanded and improved an existing genome-scale metabolic model of *M. pneumoniae* by i) updating the model on the last genome annotations, ii) introducing a correct non-growth associated energy distribution, and iii) including a membrane-lipids compartment. These improvements enabled me to develop a genome-scale, constraint-based metabolic model [named iEG158\_mpn], which I could deploy to link both metabolism and membrane synthesis to supplementation of the growth medium. Predictions through Flux Balance Analysis allowed the formulation of an efficient serum-free medium, highlighting the importance of the supplementation of two lipids, indeed essential for the membrane formation. The formulated medium supported *M. pneumoniae*'s growth *in vitro*, opening the way to the large-scale production of the vaccine, thanks to the absence of serum (and therefore, of undefined compounds). Crucially, this enables a substantial reduction in costs (and therefore, of undefined compounds) upon the large-scale production of the vaccine. The medium design was not the only attempt to allow a faster growth of *M. pneumoniae*: in **chapter 3**, I have explored the design of a strain capable of synthesizing fatty acids. This strain, called AL, implements a fatty acid biosynthesis pathway from *A. laidlawii*, and resolves the fastidious *Mycoplasma*'s related auxotrophy. While obtaining a successful production of fatty acids by strain AL, I observed a concomitant growth rate reduction respect to the wild-type. For this reason, I used a gap-filling algorithm applied to the metabolic model iEG158\_mpn to identify related growth issues and solutions. My approach suggested that the implementation of a further knock-in would allow a more efficient NADPH regeneration, which is the cofactor most likely used by the pathway, but produced in small amounts by *M. pneumoniae*. In **chapter 4**, I focused on the investigation of possible side effects caused by the vaccine's injection. *M. pneumoniae* has been associated with post-infectious Guillain-Barré-Strohl syndrome, an autoimmune neurodegenerative disease, linked to the synthesis of a glycolipid called galactocerebroside. In this chapter, I performed functional annotation of *M. pneumoniae*'s genome and constructed, through mathematical modelling, a protein motif that might be responsible, when found in the enzyme synthesizing the glycolipid, of the production of galactocerebroside. This allowed me to explore the presence of the motif in all *Mycoplasmas* whose genome sequence was available. Moreover, I reconstructed the pathway leading to the synthesis of the glycolipid, which can be therefore a target for genetic modifications to render the vaccine safe. Again for safety purposes, in **chapter 5** I report on a biocontainment strategy for limiting *Mycoplasma* growth outside of the envisaged production settings. This safety design is needed due to the nature of the vaccine strain, which is a genetically-modified organism. In the chapter, I mathematically explored several killing switches, introduced in *Mycoplasma* genome and induced by external components. These analyses provided information on the optimal concentrations of inducers to be added so to activate the circuits under study and assess the frequency of escape mutants [i.e. *Mycoplasma* cells introducing natural mutations that inactivates the circuit]. I could then predict the efficiency of the killing switches in the long term, and at which **6**, I address the computational approaches used, I recapitulate the status of the vaccine and I discuss its future perspectives. Firstly, I compared the potential of genome-scale metabolic models with that of dynamic metabolic models. Subsequently I address and highlight the importance of serum-free media in growth cultures, with a special focus on their model-driven design. Then I reflect on the safety of live-attenuated vaccines, with examples from various mycoplasmas. Finally, I propose safety-by-design as a strategy to be implemented in the developmental phase of therapeutic and pharmaceutical products.

## Funding

This thesis work has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement n. 634942 (MycoSynVac) and from the Dutch TTW-NWO under agreement n. 15814 (SafeChassis).

## References

1. Krass, C. J. & Gardner, M. W. Etymology of the Term Mycoplasma. *International Journal of Systematic Bacteriology* **23**, 62–64 (1973).
2. Edward, D. G. ff. & Freundt, E. A. The Classification and Nomenclature of Organisms of the Pleuropneumonia Group. *Journal of General Microbiology* **14**, 197–207 (1956).
3. Maggi, R. G. *et al.* Infection with Hemotropic Mycoplasma Species in Patients with or without Extensive Arthropod or Animal Contact. *Journal of Clinical Microbiology* **51**, 3237 (2013).
4. Levisohn, S. & Kleven, S. H. Avian mycoplasmosis (*Mycoplasma gallisepticum*). *Revue scientifique et technique (International Office of Epizootics)* **19**, 425 (2000).
5. Bushnell, R. B. Mycoplasma mastitis. *The Veterinary Clinics of North America. Large Animal Practice* **6**, 301–312 (1984).
6. Maunsell, F. P. *et al.* Mycoplasma bovis infections in cattle. *Journal of Veterinary Internal Medicine* **25**, 772–783 (2011).
7. Chandler, J. C. & Lappin, M. R. Mycoplasmal respiratory infections in small animals: 17 cases (1988–1999). *Journal of the American Animal Hospital Association* **38**, 111–119 (2002).
8. Taylor-Robinson, D. Infections due to species of Mycoplasma and Ureaplasma: an update. *Clinical infectious diseases* 671–682 (1996).
9. Messick, J. B. Hemotrophic mycoplasmas (hemoplasmas): a review and new insights into pathogenic potential. *Veterinary Clinical Pathology* **33**, 2–13 (2004).
10. Neimark, H. & Kirkpatrick, B. C. Isolation and characterization of full-length chromosomes from non-culturable plant-pathogenic Mycoplasma-like organisms. *Molecular microbiology* **7**, 21–28 (1993).
11. Principi, N., Esposito, S., Blasi, F., Allegra, L. & Group, M. S. Role of Mycoplasma pneumoniae and Chlamydia pneumoniae in children with community-acquired lower respiratory tract infections. *Clinical Infectious Diseases* **32**, 1281–1289 (2001).
12. Gardiner, S. J., Gavranich, J. B. & Chang, A. B. Antibiotics for community-acquired lower respiratory tract infections secondary to Mycoplasma pneumoniae in children. *Cochrane Database of Systematic Reviews* (2015).
13. Blanchard, B. *et al.* Electron microscopic observation of the respiratory tract of SPF piglets inoculated with Mycoplasma hyopneumoniae. *Veterinary microbiology* **30**, 329–341 (1992).

14. Saraya, T. The history of *Mycoplasma pneumoniae* pneumonia. *Frontiers in microbiology* **7**, 364 (2016).
15. Kashyap, S. & Sarkar, M. *Mycoplasma pneumoniae*: Clinical features and management. *Lung India: Official Organ of Indian Chest Society* **27**, 75 (2010).
16. Bhandari, V., Hussain, N., Rosenkrantz, T. & Kresch, M. Respiratory tract colonization with mycoplasma species increases the severity of bronchopulmonary dysplasia. *Journal of Perinatal Medicine-Official Journal of the WAPM* **26**, 37–42 (1998).
17. van Waarde, W. M., Brus, F., Okken, A. & Kimpen, J. L. *Ureaplasma urealyticum* colonization, prematurity and bronchopulmonary dysplasia. *European Respiratory Journal* **10**, 886–890 (1997).
18. Kane, J. R., Shenep, J. L., Krance, R. A. & Hurwitz, C. A. Diffuse alveolar hemorrhage associated with *Mycoplasma hominis* respiratory tract infection in a bone marrow transplant recipient. *Chest* **105**, 1891–1892 (1994).
19. Mishra, R., Cano, E., Venkatram, S. & Diaz-Fuentes, G. An interesting case of mycoplasma pneumonia associated multisystem involvement and diffuse alveolar hemorrhage. *Respiratory medicine case reports* **21**, 78–81 (2017).
20. Saliba, E., Sayad, A., Alameddine, L., El-Haddad, K. & Tannous, Z. *Mycoplasma pneumoniae* and atypical acute hemorrhagic edema of infancy. *The American Journal of Emergency Medicine* (2020).
21. Urbanek, C. *et al.* Detection of antibodies directed at M. *hyorhinis* p37 in the serum of men with newly diagnosed prostate cancer. *BMC Cancer* **11**, 233 (2011).
22. Mariotti, E. *et al.* The percentage of CD133+ cells in human colorectal cancer cell lines is influenced by *Mycoplasma hyorhinis* infection. *BMC Cancer* **10**, 120 (2010).
23. Larsen, B. & Hwang, J. *Mycoplasma, Ureaplasma*, and Adverse Pregnancy Outcomes: A Fresh Look. *Infectious Diseases in Obstetrics and Gynecology* **2010**, 521921 (2010).
24. Wiesenfeld, H. C. & Manhart, L. E. *Mycoplasma genitalium* in Women: Current Knowledge and Research Priorities for This Recently Emerged Pathogen. *The Journal of Infectious Diseases* **216**, S389–S395 (2017).
25. Lis, R., Rowhani-Rahbar, A. & Manhart, L. E. *Mycoplasma genitalium* Infection and Female Reproductive Tract Disease: A Meta-analysis. *Clinical Infectious Diseases* **61**, 418–426 (2015).
26. Ljubin-Sternak, S. & Meštrović, T. *Chlamydia trachomatis* and Genital Mycoplasmas: Pathogens with an Impact on Human Reproductive Health. *Journal of Pathogens* **2014**, 183167 (2014).
27. Ryan, K. J. *et al.* *Sherris Medical Microbiology: An Introduction to Infectious Diseases*. (McGraw-Hill, 2004).
28. Bébéar, C., Pereyre, S. & Peuchant, O. *Mycoplasma pneumoniae*: susceptibility and resistance to antibiotics. *Future microbiology* **6**, 423–431 (2011).
29. Wang, X. & Lutkenhaus, J. Characterization of the *ftsZ* gene from *Mycoplasma pulmonis*, an organism lacking a cell wall. *Journal of bacteriology* **178**, 2314–2319 (1996).
30. Thacker, E. L. & Minion, F. C. Mycoplasmosis. *Diseases of swine* **10**, 779–798 (2012).
31. Razin, S. *Methods in Mycoplasmaology V1: Mycoplasma characterization*. vol. 1 (Elsevier, 2012).

32. Hay, R. J., Macy, M. L. & Chen, T. R. Mycoplasma infection of cultured cells. *Nature* **339**, 487–488 (1989).
33. Bashiruddin, J. B. Extraction of DNA from mycoplasmas. in *Mycoplasma protocols* 141–144 (Springer, 1998).
34. Razin, S., Yogev, D. & Naot, Y. Molecular biology and pathogenicity of mycoplasmas. *Microbiology and molecular biology reviews: MMBR* **62**, 1094–156 (1998).
35. Taylor-Robinson, D. & Bebear, C. Antibiotic susceptibilities of mycoplasmas and treatment of mycoplasmal infections. *The Journal of antimicrobial chemotherapy* **40**, 622–630 (1997).
36. Matsuoka, M. *et al.* Characterization and Molecular Analysis of Macrolide-Resistant Mycoplasma pneumoniae. Clinical Isolates Obtained in Japan. *Antimicrobial Agents and Chemotherapy* **48**, 4624 (2004).
37. Pereyre, S. *et al.* In vitro selection and characterization of resistance to macrolides and related antibiotics in Mycoplasma pneumoniae. *Antimicrobial agents and chemotherapy* **48**, 460–465 (2004).
38. Stanbridge, E. Mycoplasmas and cell cultures. *Bacteriological reviews* **35**, 206 (1971).
39. Smith, P. F. & Koostera, W. L. Phospholipids and glycolipids of sterol-requiring Mycoplasma. *Journal of bacteriology* **93**, 1853–1862 (1967).
40. Smith, P. F. The lipids of Mycoplasma. in *Advances in lipid research* vol. 6 69–105 (Elsevier, 1968).
41. Smith, P. F. *The biology of mycoplasmas*. (Elsevier, 2012).
42. Razin, S. The mycoplasmas. *Microbiological reviews* **42**, 414–470 (1978).
43. van Noort, V. *et al.* Cross-talk between phosphorylation and lysine acetylation in a genome-reduced bacterium. *Molecular Systems Biology* **8**, 571 (2012).
44. Yus, E. *et al.* Impact of genome reduction on bacterial metabolism and its regulation. *Science* **326**, 1263–1268 (2009).
45. Leprince, A., van Passel, M. W. J. & dos Santos, V. A. P. M. Streamlining genomes: toward the generation of simplified and stabilized microbial systems. *Current Opinion in Biotechnology* **23**, 651–658 (2012).
46. Labroussaa, F., Baby, V., Rodrigue, S. & Lartigue, C. La transplantation de génomes. *Med Sci (Paris)* **35**, 761–770 (2019).
47. Fraser, C. M. *et al.* The minimal gene complement of Mycoplasma genitalium. *Science* **270**, 397–404 (1995).
48. Hutchison, C. A. *et al.* Global Transposon Mutagenesis and a Minimal Mycoplasma Genome. *Science* **286**, 2165 (1999).
49. Glass, J. I. *et al.* Essential genes of a minimal bacterium. *Proceedings of the National Academy of Sciences of the United States of America* **103**, 425 (2006).
50. Gibson, D. G. *et al.* Complete Chemical Synthesis, Assembly, and Cloning of a Mycoplasma genitalium Genome. *Science* **319**, 1215 (2008).
51. Lartigue, C. *et al.* Genome transplantation in bacteria: changing one species to another. *science* **317**, 632–638 (2007).

52. Gibson, D. G. *et al.* Creation of a Bacterial Cell Controlled by a Chemically Synthesized Genome. *Science* **329**, 52 (2010).
53. Hutchison, C. A. *et al.* Design and synthesis of a minimal bacterial genome. *Science* **351**, aad6253 (2016).
54. Razin, S. & Hayflick, L. Highlights of mycoplasma research—An historical perspective. *Biologicals* **38**, 183–190 (2010).
55. Rees-Garbutt, J. *et al.* Designing minimal genomes using whole-cell models. *Nature Communications* **11**, 836 (2020).
56. Balish, M. F. Mycoplasma pneumoniae, an underutilized model for bacterial cell biology. *Journal of bacteriology* **196**, 3675–3682 (2014).
57. Cozzuto, L. *et al.* MyMpn: a database for the systems biology model organism Mycoplasma pneumoniae. **43**, 618–623 (2015).
58. Kühner, S. *et al.* Proteome Organization in a Genome-Reduced Bacterium. *Science* **326**, 1235 (2009).
59. Güell, M. *et al.* Transcriptome Complexity in a Genome-Reduced Bacterium. *Science* **326**, 1268–1271 (2009).
60. Waites, K. B. & Talkington, D. F. Mycoplasma pneumoniae and its role as a human pathogen. *Clinical Microbiology Reviews* **17**, 697–728 (2004).
61. Sánchez-Vargas, F. M. & Gómez-Duarte, O. G. Mycoplasma pneumoniae—an emerging extra-pulmonary pathogen. *Clinical microbiology and infection* **14**, 105–115 (2008).
62. Baseman, J. B., Cole, R. M., Krause, D. C. & Leith, D. K. Molecular basis for cytoadsorption of Mycoplasma pneumoniae. *Journal of Bacteriology* **151**, 1514 (1982).
63. Drasbek, M., Christiansen, G., Drasbek, K. R., Holm, A. & Birkelund, S. Interaction between the P1 protein of Mycoplasma pneumoniae and receptors on HEp-2 cells. *Microbiology* **153**, 3791–3799 (2007).
64. Dallo, S. F. & Baseman, J. B. Intracellular DNA replication and long-term survival of pathogenic mycoplasmas. *Microbial Pathogenesis* **29**, 301–309 (2000).
65. Weisburg, W. G. *et al.* A phylogenetic analysis of the mycoplasmas: basis for their classification. *Journal of Bacteriology* **171**, 6455 (1989).
66. Himmelreich, R. *et al.* Complete sequence analysis of the genome of the bacterium Mycoplasma pneumoniae. *Nucleic acids research* **24**, 4420–4449 (1996).
67. Wodke, J. A. H. *et al.* Dissecting the energy metabolism in Mycoplasma pneumoniae through genome-scale metabolic modeling. *Molecular Systems Biology* **9**, (2013).
68. Foy, H. M. Infections Caused by Mycoplasma pneumoniae and Possible Carrier State in Different Populations of Patients. *Clinical Infectious Diseases* **17**, S37–S46 (1993).
69. Atkinson, T. P. & Waites, K. B. Mycoplasma pneumoniae infections in childhood. *The Pediatric infectious disease journal* **33**, 92–94 (2014).
70. Kumar, S. Mycoplasma pneumoniae: A significant but underrated pathogen in paediatric community-acquired lower respiratory tract infections. *The Indian journal of medical research* **147**, 23 (2018).

71. Atkinson, T. P., Balish, M. F. & Waites, K. B. Epidemiology, clinical manifestations, pathogenesis and laboratory detection of *Mycoplasma pneumoniae* infections. *FEMS microbiology reviews* **32**, 956–973 (2008).
72. Mundy, L. M. *et al.* Community-acquired pneumonia: impact of immune status. *American journal of respiratory and critical care medicine* **152**, 1309–1315 (1995).
73. Head, B. M., Trajtman, A., Rueda, Z. v, Vélez, L. & Keynan, Y. Atypical bacterial pneumonia in the HIV-infected population. *Pneumonia* **9**, 1–7 (2017).
74. Klement, E. *et al.* Identification of Risk Factors for Infection in an Outbreak of *Mycoplasma pneumoniae* Respiratory Tract Disease. *Clinical Infectious Diseases* **43**, 1239–1245 (2006).
75. Linchevski, I., Klmenet, E. & Nir-Paz, R. *Mycoplasma pneumoniae* vaccine protective efficacy and adverse reactions—Systematic review and meta-analysis. *Vaccine* **27**, 2437–2446 (2009).
76. Waring, A. L. *et al.* Development of a genomics-based PCR assay for detection of *Mycoplasma pneumoniae* in a large outbreak in New York State. *Journal of Clinical Microbiology* **39**, 1385–1390 (2001).
77. Muldoon, R. L., Raucci, J., Kowalski, J. & Rajashekaraiyah, K. An outbreak of *Mycoplasma pneumoniae* respiratory illness in a semi-closed religious commune. *Annals of Emergency Medicine* **11**, 613–615 (1982).
78. Leibowitz, Z., Schwartzman, P., Epstein, L., Lis, I. & Naot, Y. An outbreak of *Mycoplasma pneumoniae* pneumonia in two kibbutzim: A clinical and epidemiologic study. *Israel Journal of Medical Sciences* **24**, 88–92 (1988).
79. Daxboeck, F., Krause, R. & Wenisch, C. Laboratory diagnosis of *Mycoplasma pneumoniae* infection. *Clinical Microbiology and Infection* **9**, 263–273 (2003).
80. Loens, K., Ursi, D., Goossens, H. & Ieven, M. Molecular diagnosis of *Mycoplasma pneumoniae* respiratory tract infections. *Journal of clinical microbiology* **41**, 4915–4923 (2003).
81. Jacobs, E. Serological diagnosis of *Mycoplasma pneumoniae* infections: a critical review of current procedures. *Clinical infectious diseases* **17**, S79–S82 (1993).
82. Kim, G.-L., Seon, S.-H. & Rhee, D.-K. Pneumonia and *Streptococcus pneumoniae* vaccine. *Archives of Pharmacal Research* **40**, 885–893 (2017).
83. Piñero-Lambea, C. *et al.* *Mycoplasma pneumoniae* Genome Editing Based on Oligo Recombineering and Cas9-Mediated Counterselection. *ACS Synthetic Biology* **9**, 1693–1704 (2020).
84. Krishnakumar, R. *et al.* Targeted Chromosomal Knockouts in *Mycoplasma pneumoniae*. *Applied and Environmental Microbiology* **76**, 5297 (2010).
85. Sun, Z. *et al.* A high-efficiency recombineering system with PCR-based ssDNA in *Bacillus subtilis* mediated by the native phage recombinase GP35. *Applied Microbiology and Biotechnology* **99**, 5151–5162 (2015).
86. Garcia-Morales, L. *et al.* A RAGE Based Strategy for the Genome Engineering of the Human Respiratory Pathogen *Mycoplasma pneumoniae*. *ACS Synthetic Biology* **9**, 2737–2748 (2020).
87. Sibila, M. *et al.* Current perspectives on the diagnosis and epidemiology of *Mycoplasma hyopneumoniae* infection. *The Veterinary Journal* **181**, 221–231 (2009).

88. Friis, N. F. Some recommendations concerning primary isolation of *Mycoplasma suis* pneumoniae and *Mycoplasma flocculare* a survey. *Nordisk veterinærmedicin* **27**, 337 (1975).
89. Krausse, R. & Schubert, S. In-vitro activities of tetracyclines, macrolides, fluoroquinolones and clindamycin against *Mycoplasma hominis* and *Ureaplasma* ssp. isolated in Germany over 20 years. *Clinical microbiology and infection* **16**, 1649–1655 (2010).
90. Guschin, A., Ryzhikh, P., Rumyantseva, T., Gomberg, M. & Unemo, M. Treatment efficacy, treatment failures and selection of macrolide resistance in patients with high load of *Mycoplasma genitalium* during treatment of male urethritis with josamycin. *BMC infectious diseases* **15**, 40 (2015).
91. Kreizinger, Z. *et al.* Antibiotic susceptibility profiles of *Mycoplasma synoviae* strains originating from Central and Eastern Europe. *BMC veterinary research* **13**, 342 (2017).
92. Bradshaw, C. S., Jensen, J. S. & Waites, K. B. New horizons in *Mycoplasma genitalium* treatment. *The Journal of infectious diseases* **216**, S412–S419 (2017).
93. Maes, D., Boyen, F., Haesebrouck, F. & Gautier-Bouchardon, A. v. Antimicrobial treatment of *Mycoplasma hyopneumoniae* infections. *The Veterinary Journal* 105474 (2020).
94. Antunes, N. T. *et al.* In vitro susceptibilities of field isolates of *Mycoplasma mycoides* subsp. *mycoides* large colony type to 15 antimicrobials. *Veterinary microbiology* **119**, 72–75 (2007).
95. Cai, H. Y., McDowall, R., Parker, L., Kaufman, E. I. & Caswell, J. L. Changes in antimicrobial susceptibility profiles of *Mycoplasma bovis* over time. *Canadian Journal of Veterinary Research* **83**, 34–41 (2019).
96. Klein, U. *et al.* New antimicrobial susceptibility data from monitoring of *Mycoplasma bovis* isolated in Europe. *Veterinary Microbiology* **238**, 108432 (2019).
97. Ferraz, M. E. S. *et al.* Lung consolidation caused by *Mycoplasma hyopneumoniae* has a negative effect on productive performance and economic revenue in finishing pigs. *Preventive Veterinary Medicine* **182**, 105091 (2020).
98. Helm, E. T. *et al.* Impact of *Mycoplasma hyopneumoniae* and *Lawsonia intracellularis* on the performance of pigs divergently selected for feed efficiency. *Journal of animal science* **96**, 462–472 (2018).
99. McEwen, S. A. & Fedorka-Cray, P. J. Antimicrobial Use and Resistance in Animals. *Clinical Infectious Diseases* **34**, S93–S106 (2002).
100. Straw, B. E., Shin, S. J. & Yeager, A. E. Effect of pneumonia on growth rate and feed efficiency of minimal disease pigs exposed to *Actinobacillus pleuropneumoniae* and *Mycoplasma hyopneumoniae*. *Preventive Veterinary Medicine* **9**, 287–294 (1990).
101. Maes, D., Verbeke, W., Vicca, J., Verdonck, M. & de Kruif, A. Benefit to cost of vaccination against *Mycoplasma hyopneumoniae* in pig herds under Belgian market conditions from 1996 to 2000. *Livestock Production Science* **83**, 85–93 (2003).
102. Hao, H. *et al.* Benefits and risks of antimicrobial use in food-producing animals. *Frontiers in microbiology* **5**, 288 (2014).
103. Jensen, K. E., Senterfit, L. B., Chanock, R. M., Smith, C. B. & Purcell, R. H. An Inactivated: *Mycoplasma pneumoniae* Vaccine. *JAMA* **194**, 248–252 (1965).

104. Fernald, G. W. & Clyde, W. A. Protective Effect of Vaccines in Experimental Mycoplasma pneumoniae Disease. *Infection and Immunity* **1**, 559 (1970).
105. Wenzel, R. P. *et al.* Field Trial of an Inactivated Mycoplasma pneumoniae Vaccine. I. Vaccine Efficacy. *The Journal of Infectious Diseases* **134**, 571–576 (1976).
106. Ley, D. H. & Yoder Jr, H. W. Mycoplasma gallisepticum infection. *Diseases of poultry* **12**, 807–834 (2008).
107. Wijesurendra, D. S. *et al.* Development of a Mycoplasma gallisepticum infection model in turkeys. *Avian Pathology* **44**, 35–42 (2015).
108. Beaudet, J. *et al.* Transcriptional Profiling of the Chicken Tracheal Response to Virulent Mycoplasma gallisepticum Strain. *Infection and Immunity* **85**, e00343-17 (2017).
109. Kanci, A. *et al.* Evaluation of Mycoplasma gallisepticum (MG) ts-304 vaccine as a live attenuated vaccine in turkeys. *Vaccine* **36**, 2487–2493 (2018).
110. Shil, P. K. *et al.* GapA+ Mycoplasma gallisepticum ts-11 has improved vaccine characteristics. *Microbiology* **157**, 1740–1749 (2011).
111. Kanci, A. *et al.* Is an unstable repeat sequence responsible for the attenuation of the Mycoplasma gallisepticum vaccine strain ts-11. in *15th Congress of the International Organization for Mycoplasmaology* (2004).
112. Ishfaq, M. *et al.* Current status of vaccine research, development, and challenges of vaccines for Mycoplasma gallisepticum. *Poultry Science* **99**, 4195–4202 (2020).
113. Peebles, E. D. *et al.* Effects of different vaccine combinations against Mycoplasma gallisepticum on the digestive and reproductive organ characteristics of commercial egg-laying hens<sup>1,2</sup>. *Poultry Science* **94**, 2898–2904 (2015).
114. Kleven, S. H. Control of avian mycoplasma infections in commercial poultry. *Avian diseases* **52**, 367–374 (2008).
115. Rosengarten, R. & Citti, C. The role of ruminant mycoplasmas in systemic infection. *Mycoplasmas of ruminants: pathogenicity, diagnostics, epidemiology and molecular genetics* **3**, 14–17 (1999).
116. Ayling, R. D., Baker, S. E., Nicholas, R. A. J., Peek, M. L. & Simon, A. J. Comparison of in vitro activity of danofloxacin, florfenicol, oxytetracycline, spectinomycin and tilmicosin against recent field isolates of Mycoplasma bovis. *Veterinary Record* **146**, 745 (2000).
117. Howard, C. J., Stott, E. J., Thomas, L. H., Gourlay, R. N. & Taylor, G. Protection against respiratory disease in calves induced by vaccines containing respiratory syncytial virus, parainfluenza type 3 virus, Mycoplasma bovis and M. dispar. *Veterinary Record* **121**, 372–376 (1987).
118. Nicholas, R. A. J., Ayling, R. D. & Stipkovits, L. P. An experimental vaccine for calf pneumonia caused by Mycoplasma bovis: clinical, cultural, serological and pathological findings. *Vaccine* **20**, 3569–3575 (2002).
119. Dudek, K. *et al.* Analysis of the immune response of calves to various saponin-based adjuvants for an experimental Mycoplasma bovis vaccine. *Acta Veterinaria Hungarica Acta Veterinaria Hungarica* **66**, 226–240 (2018).

120. Dudek, K. *et al.* An experimental vaccine composed of two adjuvants gives protection against *Mycoplasma bovis* in calves. *Vaccine* **34**, 3051–3058 (2016).
121. Dudek, K. & Bednarek, D. T- and B-cell response analysis following calf immunisation with experimental *Mycoplasma bovis* vaccine containing saponin and lysozyme dimer. *Journal of Veterinary Research* **61**, 433–437 (2017).
122. Perez-Casal, J., Prysliak, T., Maina, T., Suleman, M. & Jimbo, S. Status of the development of a vaccine against *Mycoplasma bovis*. *Vaccine* **35**, 2902–2907 (2017).
123. León Vizcaíno, L., Garrido Abellán, F., Cubero Pablo, M. J. & Perales, A. Immunoprophylaxis of caprine contagious agalactia due to *Mycoplasma agalactiae* with an inactivated vaccine. *The Veterinary record* **137**, 266–269 (1995).
124. Tola, S. *et al.* Experimental vaccination against *Mycoplasma agalactiae* using different inactivated vaccines. *Vaccine* **17**, 2764–2768 (1999).
125. Buonavoglia, D. *et al.* Persistence of antibodies to *Mycoplasma agalactiae* in vaccinated sheep. *The new microbiologica* **21**, 209–212 (1998).
126. Buonavoglia, D. *et al.* Long-term immunogenicity and protection against *Mycoplasma agalactiae* induced by an oil adjuvant vaccine in sheep. *Research in Veterinary Science* **88**, 16–19 (2010).
127. Buonavoglia, D. *et al.* An oil-emulsion vaccine induces full-protection against *Mycoplasma agalactiae* infection in sheep. *New Microbiologica* **31** (2008).
128. Thiaucourt F., Aboubakar Y., Wesonga H., Manso-Silvan L. & Blanchard A. Contagious Bovine Pleuropneumoniae Vaccines and Control Strategies: Recent Data. *Dev. Biol.* **119**, 99–111 (2004).
129. Rurangirwa, F. R., McGuire, T. C., Kibor, A. & Chema, S. An inactivated vaccine for contagious caprine pleuropneumonia. *Veterinary Record* **121**, 397 (1987).
130. Jores, J. *et al.* Contagious Bovine and Caprine Pleuropneumonia: a research community's recommendations for the development of better vaccines. *npj Vaccines* **5**, 66 (2020).
131. Liu, W. *et al.* Complete genome sequence of *Mycoplasma hyopneumoniae* strain 168. *Journal of bacteriology* **193**, 1016–1017 (2011).
132. Minion, F. C. *et al.* The genome sequence of *Mycoplasma hyopneumoniae* strain 232, the agent of swine mycoplasmosis. *Journal of bacteriology* **186**, 7123–7133 (2004).
133. Vasconcelos, A. T. R. *et al.* Swine and poultry pathogens: the complete genome sequences of two strains of *Mycoplasma hyopneumoniae* and a strain of *Mycoplasma synoviae*. *Journal of bacteriology* **187**, 5568–5577 (2005).
134. Sibila, M. *et al.* Chronological study of *Mycoplasma hyopneumoniae* infection, seroconversion and associated lung lesions in vaccinated and non-vaccinated pigs. *Veterinary Microbiology* **122**, 97–107 (2007).
135. Vicca, J. *et al.* Evaluation of virulence of *Mycoplasma hyopneumoniae* field isolates. *Veterinary Microbiology* **97**, 177–190 (2003).
136. Maes, D. *et al.* Control of *Mycoplasma hyopneumoniae* infections in pigs. *Veterinary Microbiology* **126**, 297–309 (2008).

137. Vicca, J. *et al.* In vitro susceptibilities of *Mycoplasma hyopneumoniae* field isolates. *Antimicrobial agents and chemotherapy* **48**, 4470–4472 (2004).
138. Haesebrouck, F. *et al.* Efficacy of vaccines against bacterial diseases in swine: what can we expect? *Veterinary Microbiology* **100**, 255–268 (2004).
139. Maes, D. *et al.* Effect of vaccination against *Mycoplasma hyopneumoniae* in pig herds with an all-in/all-out production system. *Vaccine* **17**, 1024–1034 (1999).
140. Simionatto, S., Marchioro, S. B., Maes, D. & Dellagostin, O. A. *Mycoplasma hyopneumoniae*: From disease to vaccine development. *Veterinary Microbiology* **165**, 234–242 (2013).
141. Strait, E. L. *et al.* Real-time PCR assays to address genetic diversity among strains of *Mycoplasma hyopneumoniae*. *Journal of clinical microbiology* **46**, 2491–2498 (2008).
142. Calus, D. *et al.* Protein variability among *Mycoplasma hyopneumoniae* isolates. *Veterinary Microbiology* **120**, 284–291 (2007).
143. Tao, Y., Shu, J., Chen, J., Wu, Y. & He, Y. A concise review of vaccines against *Mycoplasma hyopneumoniae*. *Research in Veterinary Science* **123**, 144–152 (2019).
144. Tzika, E. D. *et al.* Field efficacy study of a novel ready-to-use vaccine against mycoplasma hyopneumoniae and porcine circovirus type 2 in a Greek farm. *Porcine health management* **1**, 15 (2015).
145. Jacob, R., Branton, S. L., Evans, J. D., Leigh, S. A. & Peebles, E. D. Effects of live and killed vaccines against *Mycoplasma gallisepticum* on the performance characteristics of commercial layer chickens. *Poultry Science* **93**, 1403–1409 (2014).
146. Shann, F. Nonspecific Effects of Vaccines and the Reduction of Mortality in Children. *Clinical Therapeutics* **35**, 109–114 (2013).
147. Pasetti, M. F., Simon, J. K., Sztein, M. B. & Levine, M. M. Immunology of gut mucosal vaccines. *Immunological reviews* **239**, 125–148 (2011).
148. Minor, P. D. Live attenuated vaccines: Historical successes and current challenges. *Virology* **479–480**, 379–392 (2015).
149. Benn, C. S., Netea, M. G., Selin, L. K. & Aaby, P. A small jab - a big effect: nonspecific immunomodulation by vaccines. *Trends in Immunology* **34**, 431–439 (2013).
150. Noormohammadi, A. H., Jones, J. F., Underwood, G. & Whithear, K. G. Poor Systemic Antibody Response After Vaccination of Commercial Broiler Breeders with *Mycoplasma gallisepticum* Vaccine ts-11 Not Associated with Susceptibility to Challenge. *Avian Diseases* **46**, 623–628 (2002).
151. Ley, D. H. *et al.* Transmissibility of Live *Mycoplasma gallisepticum* Vaccine Strains ts-11 and 6/85 from Vaccinated Layer Pullets to Sentinel Poultry. *Avian Diseases* **41**, 187–194 (1997).
152. Carpenter, T. E., Mallinson, E. T., Miller, K. F., Gentry, R. F. & Schwartz, L. D. Vaccination with F-Strain *Mycoplasma gallisepticum* to Reduce Production Losses in Layer Chickens. *Avian Diseases* **25**, 404–409 (1981).
153. Evans, J. D. & Leigh, S. A. Differentiation of *Mycoplasma gallisepticum* Vaccine Strains ts-11 and 6/85 from Commonly Used *Mycoplasma gallisepticum* Challenge Strains by PCR. *Avian Diseases* **52**, 491–497 (2008).

154. Branton, S. L. *et al.* The Effects of 6/85 Live *Mycoplasma gallisepticum* Vaccine in Commercial Layer Hens over a 43-Week Laying Cycle on Egg Production, Selected Egg Quality Parameters, and Egg Size Distribution When Challenged Before Beginning of Lay. *Avian Diseases* **46**, 423–428 (2002).
155. Markham, J. F., Scott, P. C. & Whithear, K. G. Field Evaluation of the Safety and Efficacy of a Temperature-Sensitive *Mycoplasma synoviae* Live Vaccine. *Avian Diseases* **42**, 682–689 (1998).
156. Li, Y., Li, P., Wang, X., Yu, Q. & Yang, Q. Co-administration of attenuated *Mycoplasma hyopneumoniae* 168 strain with bacterial DNA enhances the local and systemic immune response after intranasal vaccination in pigs. *Vaccine* **30**, 2153–2158 (2012).
157. Shen, Y., Hu, W., Wei, Y., Feng, Z. & Yang, Q. The immune mechanism of *Mycoplasma hyopneumoniae* 168 vaccine strain through dendritic cells. *BMC Veterinary Research* **13**, 285 (2017).
158. Li, P., Li, Y., Shao, G., Yu, Q. & Yang, Q. Comparison of immune responses to intranasal and intrapulmonary vaccination with the attenuated *Mycoplasma hyopneumoniae* 168 strain in pigs. *Journal of Veterinary Medical Science* **77**, 519–525 (2015).
159. Feng, Z. *et al.* Immune Responses to the Attenuated *Mycoplasma hyopneumoniae* 168 Strain Vaccine by Intrapulmonic Immunization in Piglets. *Agricultural Sciences in China* **9**, 423–431 (2010).
160. de la Fe, C. *et al.* Field trial of two dual vaccines against *Mycoplasma agalactiae* and *Mycoplasma mycoides* subsp. *mycoides* (large colony type) in goats. *Vaccine* **25**, 2340–2345 (2007).
161. Sunder, J., Srivastava, N. C. & Singh, V. P. Preliminary Trials on Development of Vaccine Against *Mycoplasma mycoides* subsp. *mycoides* type LC Infection in Goats. *Journal of Applied Animal Research* **21**, 75–80 (2002).
162. Mbulu, R.-S. *et al.* Contagious bovine pleuropneumonia (CBPP) caused by vaccine strain T1/44 of *Mycoplasma mycoides* subsp. *mycoides* SC. *Veterinary Microbiology* **98**, 229–234 (2004).
163. Foggie, A., Etheridge, J. R., Erdağ, O. & Arisoy, F. Contagious agalactia of sheep and goats studies on live and dead vaccines in lactating sheep. *Journal of Comparative Pathology* **81**, 165–172 (1971).
164. Calcutt, M. J. *et al.* Gap analysis of *Mycoplasma bovis* disease, diagnosis and control: An aid to identify future development requirements. *Transboundary and Emerging Diseases* **65**, 91–109 (2018).
165. Lipman, R. P., Clyde, W. A. & Denny, F. W. Characteristics of Virulent, Attenuated, and Avirulent *Mycoplasma pneumoniae* Strains. *Journal of Bacteriology* **100**, 1037 (1969).
166. Yayoshi, M., Hayatsu, E. & Yoshioka, M. Protective effects of *Mycoplasma pneumoniae* live vaccine or its hyperimmune serum on the experimental infection in mice. *Kansenshogaku zasshi. The Journal of the Japanese Association for Infectious Diseases* **63**, 684–691 (1989).
167. Kumar, A., Verma, A. K., Gangwar, N. K. & Rahal, A. Isolation, characterization and antibiogram of *Mycoplasma bovis* in sheep pneumonia. *Asian Journal of Animal and Veterinary Advances* **7**, 149–157 (2012).
168. Kadir, M. A., Ali, N. H. & Ridha, R. G. M. Prevalence of helminthes, pneumonia and hepatitis in Kirkuk slaughter house, Kirkuk, Iraq. *Iraqi Journal of Veterinary Sciences* **26** (2012).
169. Meseguer, M. A. *et al.* *Mycoplasma pneumoniae*: a reduced-genome intracellular bacterial pathogen. *Infection, Genetics and Evolution* **3**, 47–55 (2003).

170. Gaspari, E. *et al.* Model-driven design allows growth of *Mycoplasma pneumoniae* on serum-free media. *npj Systems Biology and Applications* **6**, 33 (2020).
171. Kusunoki, S., Shiina, M. & Kanazawa, I. Anti-Gal-C antibodies in GBS subsequent to mycoplasma infection: Evidence of molecular mimicry. *Neurology* **57**, 736 (2001).
172. Fang, X., Lloyd, C. J. & Palsson, B. O. Reconstructing organisms in silico: genome-scale models and their emerging applications. *Nature Reviews Microbiology* **18**, 731–743 (2020).
173. Price, N. D., Reed, J. L. & Palsson, B. Ø. Genome-scale models of microbial cells: evaluating the consequences of constraints. *Nature Reviews Microbiology* **2**, 886–897 (2004).
174. O'Brien, E. J., Monk, J. M. & Palsson, B. O. Using Genome-scale Models to Predict Biological Capabilities. *Cell* **161**, 971–987 (2015).
175. Durot, M., Bourguignon, P.-Y. & Schachter, V. Genome-scale models of bacterial metabolism: reconstruction and applications. *FEMS Microbiology Reviews* **33**, 164–190 (2009).
176. Schilling, C. H. & Palsson, B. Ø. Assessment of the Metabolic Capabilities of *Haemophilus influenzae* Rd through a Genome-scale Pathway Analysis. *Journal of Theoretical Biology* **203**, 249–283 (2000).
177. Orth, J. D. *et al.* A comprehensive genome-scale reconstruction of *Escherichia coli* metabolism—2011. *Molecular Systems Biology* **7**, 535 (2011).
178. Feist, A. M. *et al.* A genome-scale metabolic reconstruction for *Escherichia coli* K-12 MG1655 that accounts for 1260 ORFs and thermodynamic information. *Molecular Systems Biology* **3**, 121 (2007).
179. Reed, J. L., Vo, T. D., Schilling, C. H. & Palsson, B. O. An expanded genome-scale model of *Escherichia coli* K-12 (iJR904 GSM/GPR). *Genome Biology* **4**, R54 (2003).
180. Förster, J., Famili, I., Fu, P., Palsson, B. & Nielsen, J. Genome-scale reconstruction of the *Saccharomyces cerevisiae* metabolic network. *Genome Research* **13**, 244–253 (2003).
181. Nookaew, I. *et al.* The genome-scale metabolic model iIN800 of *Saccharomyces cerevisiae* and its validation: a scaffold to query lipid metabolism. *BMC Systems Biology* **2**, 71 (2008).
182. Gu, C., Kim, G. B., Kim, W. J., Kim, H. U. & Lee, S. Y. Current status and applications of genome-scale metabolic models. *Genome Biology* **20**, 121 (2019).
183. Roberts, S. B., Gowen, C. M., Brooks, J. P. & Fong, S. S. Genome-scale metabolic analysis of *Clostridium thermocellum* for bioethanol production. *BMC Systems Biology* **4**, 31 (2010).
184. Feng, J. *et al.* Metabolome- and genome-scale model analyses for engineering of *Aureobasidium pullulans* to enhance polymalic acid and malic acid production from sugarcane molasses. *Biotechnology for Biofuels* **11**, 94 (2018).
185. Pereira, B. *et al.* Reconstruction of a genome-scale metabolic model for *Actinobacillus succinogenes* 130Z. *BMC Systems Biology* **12**, 61 (2018).
186. Becker, J. & Wittmann, C. Bio-based production of chemicals, materials and fuels – *Corynebacterium glutamicum* as versatile cell factory. *Current Opinion in Biotechnology* **23**, 631–640 (2012).
187. Agren, R., Otero, J. M. & Nielsen, J. Genome-scale modeling enables metabolic engineering of *Saccharomyces cerevisiae* for succinic acid production. *Journal of Industrial Microbiology & Biotechnology* **40**, 735–747 (2013).

188. Haanstra, J. R. & Bakker, B. M. Drug target identification through systems biology. *Drug Discovery Today: Technologies* **15**, 17–22 (2015).
189. Dunphy, L. J. & Papin, J. A. Biomedical applications of genome-scale metabolic network reconstructions of human pathogens. *Current Opinion in Biotechnology* **51**, 70–79 (2018).
190. Sertbas, M. & Ulgen, K. O. Genome-Scale Metabolic Modeling for Unraveling Molecular Mechanisms of High Threat Pathogens. *Frontiers in Cell and Developmental Biology* **8**, 566702 (2020).
191. Kim, H. U. *et al.* Integrative genome-scale metabolic analysis of *Vibrio vulnificus* for drug targeting and discovery. *Molecular Systems Biology* **7**, 460 (2011).
192. Mienda, B. S., Salihu, R., Adamu, A. & Idris, S. Genome-scale metabolic models as platforms for identification of novel genes as antimicrobial drug targets. *Future Microbiology* **13**, 455–467 (2018).
193. Raman, K., Yeturu, K. & Chandra, N. targetTB: A target identification pipeline for *Mycobacterium tuberculosis* through an interactome, reactome and genome-scale structural analysis. *BMC Systems Biology* **2**, 109 (2008).
194. Bosi, E. *et al.* Comparative genome-scale modelling of *Staphylococcus aureus* strains identifies strain-specific metabolic capabilities linked to pathogenicity. *Proceedings of the National Academy of Sciences* **113**, E3801 (2016).
195. Vongsangnak, W., Olsen, P., Hansen, K., Krosgaard, S. & Nielsen, J. Improved annotation through genome-scale metabolic modeling of *Aspergillus oryzae*. *BMC Genomics* **9**, 245 (2008).
196. O'Brien, E. J. & Palsson, B. O. Computing the functional proteome: recent progress and future prospects for genome-scale models. *Current Opinion in Biotechnology* **34**, 125–134 (2015).
197. Sánchez, B. J. *et al.* Improving the phenotype predictions of a yeast genome-scale metabolic model by incorporating enzymatic constraints. *Molecular Systems Biology* **13**, 935 (2017).
198. Åkesson, M., Förster, J. & Nielsen, J. Integration of gene expression data into genome-scale metabolic models. *Metabolic Engineering* **6**, 285–293 (2004).
199. Feist, A. M., Scholten, J. C. M., Palsson, B. Ø., Brockman, F. J. & Ideker, T. Modeling methanogenesis with a genome-scale metabolic reconstruction of *Methanosarcina barkeri*. *Molecular Systems Biology* **2**, 2006.0004 (2006).
200. Lerman, J. A. *et al.* In silico method for modelling metabolism and gene product expression at genome scale. *Nature Communications* **3**, 929 (2012).
201. Zhuang, K. *et al.* Genome-scale dynamic modeling of the competition between *Rhodospirillum rubrum* and *Geobacter* in anoxic subsurface environments. *The ISME Journal* **5**, 305–316 (2011).
202. Mahadevan, R., Palsson, B. Ø. & Lovley, D. R. In situ to in silico and back: elucidating the physiology and ecology of *Geobacter* spp. using genome-scale modelling. *Nature Reviews Microbiology* **9**, 39–50 (2011).
203. Rosario, D. *et al.* Understanding the Representative Gut Microbiota Dysbiosis in Metformin-Treated Type 2 Diabetes Patients Using Genome-Scale Metabolic Modeling. *Frontiers in Physiology* **9**, 775 (2018).
204. Islam, M. M., Fernando, S. C. & Saha, R. Metabolic Modeling Elucidates the Transactions in the Rumen Microbiome and the Shifts Upon Virome Interactions. *Frontiers in Microbiology* **10**, 2412 (2019).

205. Orth, J. D., Thiele, I. & Palsson, B. Ø. What is flux balance analysis? *Nature Publishing Group* **28**, 245–248 (2010).
206. Mori, M., Hwa, T., Martin, O. C., de Martino, A. & Marinari, E. Constrained Allocation Flux Balance Analysis. *PLOS Computational Biology* **12**, e1004913 (2016).
207. Kauffman, K. J., Prakash, P. & Edwards, J. S. Advances in flux balance analysis. *Current Opinion in Biotechnology* **14**, 491–496 (2003).
208. Lee, J. M., Gianchandani, E. P. & Papin, J. A. Flux balance analysis in the era of metabolomics. *Briefings in Bioinformatics* **7**, 140–150 (2006).
209. Lee, S. J. *et al.* Metabolic Engineering of Escherichia coli for Enhanced Production of Succinic Acid, Based on Genome Comparison and In Silico Gene Knockout Simulation. *Applied and Environmental Microbiology* **71**, 7880 (2005).
210. Unrean, P., Tee, K. L. & Wong, T. S. Metabolic pathway analysis for in silico design of efficient autotrophic production of advanced biofuels. *Bioresources and Bioprocessing* **6**, 49 (2019).
211. Arora, R., Behera, S., Sharma, N. K. & Kumar, S. Evaluating the Pathway for Co-fermentation of Glucose and Xylose for Enhanced Bioethanol Production Using Flux Balance Analysis. *Biotechnology and Bioprocess Engineering* **24**, 924–933 (2019).
212. Parichehreh, R., Gheshlaghi, R., Mahdavi, M. A. & Elkamel, A. Optimization of lipid production in Chlorella vulgaris for biodiesel production using flux balance analysis. *Biochemical Engineering Journal* **141**, 131–145 (2019).
213. Salleh, A. H. M. *et al.* Gene knockout identification for metabolite production improvement using a hybrid of genetic ant colony optimization and flux balance analysis. *Biotechnology and Bioprocess Engineering* **20**, 685–693 (2015).
214. Chua, P. S. *et al.* Identifying a gene knockout strategy using a hybrid of the bat algorithm and flux balance analysis to enhance the production of succinate and lactate in Escherichia coli. *Biotechnology and Bioprocess Engineering* **20**, 349–357 (2015).
215. Choon, Y. W. *et al.* Gene Knockout Identification Using an Extension of Bees Hill Flux Balance Analysis. *BioMed Research International* **2015**, 124537 (2015).
216. Hon, M. K. *et al.* Identifying a Gene Knockout Strategy Using a Hybrid of Simple Constrained Artificial Bee Colony Algorithm and Flux Balance Analysis to Enhance the Production of Succinate and Lactate in Escherichia Coli. *Interdisciplinary Sciences: Computational Life Sciences* **11**, 33–44 (2019).
217. Liu, J. K. *et al.* Predicting proteome allocation, overflow metabolism, and metal requirements in a model acetogen. *PLOS Computational Biology* **15**, e1006848- (2019).
218. Ghosh, A., Zhao, H. & Price, N. D. Genome-Scale Consequences of Cofactor Balancing in Engineered Pentose Utilization Pathways in Saccharomyces cerevisiae. *PLOS ONE* **6**, e27316- (2011).
219. Grafahrend-Belau, E., Schreiber, F., Koschützki, D. & Junker, B. H. Flux Balance Analysis of Barley Seeds: A Computational Approach to Study Systemic Properties of Central Metabolism. *Plant Physiology* **149**, 585 (2009).
220. Wegrzyn, A. B. *et al.* Cofactors revisited – Predicting the impact of flavoprotein-related diseases on a genome scale. *Biochimica et Biophysica Acta (BBA) - Molecular Basis of Disease* **1865**, 360–370 (2019).

221. Lee, D.-S. *et al.* Comparative Genome-Scale Metabolic Reconstruction and Flux Balance Analysis of Multiple *Staphylococcus aureus* Genomes Identify Novel Antimicrobial Drug Targets. *Journal of Bacteriology* **191**, 4015 (2009).
222. Faria, J. P., Rocha, M., Rocha, I. & Henry, C. S. Methods for automated genome-scale metabolic model reconstruction. *Biochemical Society Transactions* **46**, 931–936 (2018).
223. Rawls, K. D. *et al.* A simplified metabolic network reconstruction to promote understanding and development of flux balance analysis tools. *Computers in Biology and Medicine* **105**, 64–71 (2019).
224. Lee, J., Yun, H., Feist, A. M., Palsson, B. Ø. & Lee, S. Y. Genome-scale reconstruction and in silico analysis of the *Clostridium acetobutylicum* ATCC 824 metabolic network. *Applied Microbiology and Biotechnology* **80**, 849–862 (2008).
225. Kim, T. Y., Sohn, S. B., Kim, Y. bin, Kim, W. J. & Lee, S. Y. Recent advances in reconstruction and applications of genome-scale metabolic models. *Current Opinion in Biotechnology* **23**, 617–623 (2012).
226. Perumal, D., Samal, A., Sakharkar, K. R. & Sakharkar, M. K. Targeting multiple targets in *Pseudomonas aeruginosa* PAO1 using flux balance analysis of a reconstructed genome-scale metabolic network. *Journal of Drug Targeting* **19**, 1–13 (2011).
227. Gatto, F., Miess, H., Schulze, A. & Nielsen, J. Flux balance analysis predicts essential genes in clear cell renal cell carcinoma metabolism. *Scientific Reports* **5**, 10738 (2015).
228. Joyce, A. R. & Palsson, B. Ø. Predicting Gene Essentiality Using Genome-Scale in Silico Models. in *Humana Press* vol. 416 (2008).
229. Toyoshima, M., Toya, Y. & Shimizu, H. Flux balance analysis of cyanobacteria reveals selective use of photosynthetic electron transport components under different spectral light conditions. *Photosynthesis Research* **143**, 31–43 (2020).
230. Peng, C., Lin, Y., Luo, H. & Gao, F. A Comprehensive Overview of Online Resources to Identify and Predict Bacterial Essential Genes. *Frontiers in Microbiology* **8**, 2331 (2017).
231. Edwards, J. S. & Palsson, B. O. Metabolic flux balance analysis and the in silico analysis of *Escherichia coli* K-12 gene deletions. *BMC Bioinformatics* **1**, 1 (2000).
232. Minato, Y. *et al.* Genomewide Assessment of *Mycobacterium tuberculosis* Conditionally Essential Metabolic Pathways. *mSystems* **4**, e00070-19 (2019).
233. Oh, Y.-G., Lee, D.-Y., Lee, S. Y. & Park, S. Multiobjective flux balancing using the NISE method for metabolic network analysis. *Biotechnology Progress* **25**, 999–1008 (2009).
234. Gómez-Cerón, S., Galindo-Betancur, D. & Ramírez-Malule, H. Data set of in silico simulation for the production of clavulanic acid and cephameycin C by *Streptomyces clavuligerus* using a genome scale metabolic model. *Data in Brief* **24**, 103992 (2019).
235. Emenike, V. N., Schulze, M., Schenkendorf, R. & Krewer, U. Model-based optimization of the recombinant protein production in *Pichia pastoris* based on dynamic flux balance analysis and elementary process functions. in *Computer Aided Chemical Engineering* (eds. Espuña, A., Graells, M. & Puigjaner, L.) vol. 40 2815–2820 (Elsevier, 2017).
236. Boyarskiy, S. & Tullman-Ercek, D. Getting pumped: membrane efflux transporters for enhanced biomolecule production. *Current Opinion in Chemical Biology* **28**, 15–19 (2015).

237. Gutierrez, J. M. & Lewis, N. E. Optimizing eukaryotic cell hosts for protein production through systems biotechnology and genome-scale modeling. *Biotechnology Journal* **10**, 939–949 (2015).
238. Mahadevan, R., Edwards, J. S. & Doyle III, F. J. Dynamic Flux Balance Analysis of Diauxic Growth in *Escherichia coli*. *Biophysical Journal* **83**, 1331–1340 (2002).
239. Shameer, S., Vallarino, J. G., Fernie, A. R., Ratcliffe, R. G. & Sweetlove, L. J. Flux balance analysis of metabolism during growth by osmotic cell expansion and its application to tomato fruits. *The Plant Journal* **103**, 68–82 (2020).
240. de Martino, D., MC Andersson, A., Bergmiller, T., Guet, C. C. & Tkačik, G. Statistical mechanics for metabolic networks during steady state growth. *Nature Communications* **9**, 2988 (2018).
241. Emenike, V. N., Schenkendorf, R. & Krewer, U. Model-based optimization of biopharmaceutical manufacturing in *Pichia pastoris* based on dynamic flux balance analysis. *Computers & Chemical Engineering* **118**, 1–13 (2018).
242. Tajparast, M. & Frigon, D. Predicting the accumulation of storage compounds by *Rhodococcus jostii* RHA1 in the feast-famine growth cycles using genome-scale flux balance analysis. *PLOS ONE* **13**, e0191835 (2018).
243. García Sánchez, C. E. & Torres Sáez, R. G. Comparison and analysis of objective functions in flux balance analysis. *Biotechnology Progress* **30**, 985–991 (2014).
244. Choon, Y. W. *et al.* Differential Bees Flux Balance Analysis with OptKnock for In Silico Microbial Strains Optimization. *PLOS ONE* **9**, e102744 (2014).
245. Shastri, A. A. & Morgan, J. A. Flux Balance Analysis of Photoautotrophic Metabolism. *Biotechnology Progress* **21**, 1617–1626 (2005).
246. Gudmundsson, S. & Thiele, I. Computationally efficient flux variability analysis. *BMC Bioinformatics* **11**, 489 (2010).
247. Suthers, P. F. *et al.* A Genome-Scale Metabolic Reconstruction of *Mycoplasma genitalium*, iPS189. *PLOS Computational Biology* **5**, e1000285 (2009).
248. Ferrarini, M. G. *et al.* Insights on the virulence of swine respiratory tract mycoplasmas through genome-scale metabolic modeling. *BMC Genomics* 1–20 (2016) doi:10.1186/s12864-016-2644-z.
249. Kamminga, T. *et al.* Metabolic modeling of energy balances in *Mycoplasma hyopneumoniae* shows that pyruvate addition increases growth rate. *Biotechnology and Bioengineering* **114**, 2339–2347 (2017).
250. Skapski, A. *et al.* Genome-Scale Analysis of *Mycoplasma agalactiae* Loci Involved in Interaction with Host Cells. *PLOS ONE* **6**, e25291 (2011).
251. Bautista, E. J. *et al.* Semi-automated Curation of Metabolic Models via Flux Balance Analysis: A Case Study with *Mycoplasma gallisepticum*. *PLOS Computational Biology* **9**, e1003208- (2013).
252. Pan, S. & Reed, J. L. Advances in gap-filling genome-scale metabolic models and model-driven experiments lead to novel metabolic discoveries. *Current Opinion in Biotechnology* **51**, 103–108 (2018).
253. Ong, W. K., Midford, P. E. & Karp, P. D. Taxonomic weighting improves the accuracy of a gap-filling algorithm for metabolic models. *Bioinformatics* **36**, 1823–1830 (2020).

254. Thiele, I., Vlassis, N. & Fleming, R. M. T. fastGapFill: efficient gap filling in metabolic networks. *Bioinformatics* **30**, 2529–2531 (2014).
255. Orth, J. D. & Palsson, B. Gap-filling analysis of the iJO1366 Escherichia coli metabolic network reconstruction for discovery of metabolic functions. *BMC Systems Biology* **6**, 30 (2012).
256. Orth, J. D. & Palsson, B. Ø. Systematizing the generation of missing metabolic knowledge. *Biotechnology and Bioengineering* **107**, 403–412 (2010).
257. Prigent, S. *et al.* Meneco, a Topology-Based Gap-Filling Tool Applicable to Degraded Genome-Wide Metabolic Networks. *PLOS Computational Biology* **13**, e1005276- (2017).
258. Schroeder, W. L. & Saha, R. OptFill: A Tool for Infeasible Cycle-Free Gapfilling of Stoichiometric Metabolic Models. *iScience* **23**, 100783 (2020).
259. Fouladiha, H. *et al.* Systematically gap-filling the genome-scale model of CHO cells. *bioRxiv* (2020) doi:10.1101/2020.01.27.921296.
260. King, B. *et al.* ProbAnnoWeb and ProbAnnoPy: probabilistic annotation and gap-filling of metabolic reconstructions. *Bioinformatics* **34**, 1594–1596 (2018).
261. Benedict, M. N., Mundy, M. B., Henry, C. S., Chia, N. & Price, N. D. Likelihood-Based Gene Annotations for Gap Filling and Quality Assessment in Genome-Scale Metabolic Models. *PLOS Computational Biology* **10**, e1003882- (2014).
262. Meyer Sauter, P. M. *et al.* Mycoplasma pneumoniae triggering the Guillain-Barré syndrome: A case-control study. *Annals of Neurology* **80**, 566–580 (2016).
263. Meyer, P. M. *et al.* Intrathecal antibody responses to GalC in Guillain-Barré syndrome triggered by Mycoplasma pneumoniae. *Journal of Neuroimmunology* **314**, 13–16 (2018).
264. Paul George, A. A. *et al.* HeMoQuest: A webserver for qualitative prediction of transient heme binding to protein motifs. *BMC Bioinformatics* **21**, (2020).
265. Zhang, F. *et al.* DeepFunc: A Deep Learning Framework for Accurate Prediction of Protein Functions from Protein Sequences and Interactions. *PROTEOMICS* **19**, 1900019 (2019).
266. Bressin, A. *et al.* TriPepSVM: de novo prediction of RNA-binding proteins based on short amino acid motifs. *Nucleic Acids Research* **47**, 4406–4417 (2019).
267. Pan, X. & Shen, H.-B. Predicting RNA–protein binding sites and motifs through combining local and global deep convolutional neural networks. *Bioinformatics* **34**, 3427–3436 (2018).
268. Rabiner, L. R. A tutorial on hidden Markov models and selected applications in speech recognition. *Proceedings of the IEEE* **77**, 257–286 (1989).
269. Munch, K. & Krogh, A. Automatic generation of gene finders for eukaryotic species. *BMC bioinformatics* **7**, 263 (2006).
270. Pachter, L., Alexandersson, M. & Cawley, S. Applications of Generalized Pair Hidden Markov Models to Alignment and Gene Finding Problems. *Journal of Computational Biology* **9**, 389–399 (2002).
271. Lottaz, C., Iseli, C., Jongeneel, C. v & Bucher, P. Modeling sequencing errors by combining Hidden Markov models. *Bioinformatics* **19**, ii103–ii112 (2003).

272. Won, K.-J., Hamelryck, T., Prügél-Bennett, A. & Krogh, A. An evolutionary method for learning HMM structure: prediction of protein secondary structure. *BMC Bioinformatics* **8**, 357 (2007).
273. Weinberg, Z. & Ruzzo, W. L. Sequence-based heuristics for faster annotation of non-coding RNA families. *Bioinformatics* **22**, 35–39 (2006).
274. Harmanci, A. O., Sharma, G. & Mathews, D. H. Efficient pairwise RNA structure prediction using probabilistic alignment constraints in Dynalign. *BMC Bioinformatics* **8**, 130 (2007).
275. Yoon, B.-J. Hidden Markov Models and their Applications in Biological Sequence Analysis. *Current genomics* **10**, 402–415 (2009).
276. El-Gebali, S. *et al.* The Pfam protein families database in 2019. *Nucleic Acids Research* **47**, D427–D432 (2019).
277. Koehorst, J. J. *A fair approach to genomics.* (2019).
278. Adams, M. D. *et al.* The Genome Sequence of *Drosophila melanogaster*. *Science* **287**, 2185 (2000).
279. Lander, E. S. *et al.* Initial sequencing and analysis of the human genome. *Nature* **409**, 860–921 (2001).
280. Waagmeester, A. *et al.* A protocol for adding knowledge to Wikidata, a case report. *bioRxiv* (2020) doi:10.1101/2020.04.05.026336.
281. Bramley, J. C. *et al.* Domain-centric database to uncover structure of minimally characterized viral genomes. *Scientific Data* **7**, 202 (2020).
282. Gavriilidou, A. *et al.* Comparative genomic analysis of Flavobacteriaceae: insights into carbohydrate metabolism, gliding motility and secondary metabolite biosynthesis. *BMC Genomics* **21**, (2020).
283. Gonzalez-Esquer, C. R. *et al.* Demonstration of the potential of *Picochlorum soloeicmus* as a microalgal platform for the production of renewable fuels. *Algal Research* **43**, 101658 (2019).
284. Koehorst, J. J., Saccenti, E., dos Santos, V. M., Suarez-Diez, M. & Schaap, P. J. Expected and observed genotype complexity in prokaryotes: correlation between 16S-rRNA phylogeny and protein domain content. *bioRxiv* 494625 (2018) doi:10.1101/494625.
285. Koehorst, J. J., Saccenti, E., Schaap, P. J., Martins Dos Santos, V. A. P. & Suarez-Diez, M. Protein domain architectures provide a fast, efficient and scalable alternative to sequence-based methods for comparative functional genomics. *F1000Research* **5**, 1987 (2016).
286. van Dam, J. C. J. Semantic systems biology of prokaryotes: heterogeneous data integration to understand bacterial metabolism. *Wageningen: Wageningen University* (2019).
287. Prakash, D., Verma, S., Bhatia, R. & Tiwary, B. N. Risks and Precautions of Genetically Modified Organisms. *ISRN Ecology* **2011**, 369573 (2011).
288. Gardner, T. S., Cantor, C. R. & Collins, J. J. Construction of a genetic toggle switch in *Escherichia coli*. *Nature* **403**, 339–342 (2000).
289. Elowitz, M. B. & Leibler, S. A synthetic oscillatory network of transcriptional regulators. *Nature* **403**, 335–338 (2000).
290. Moon, T. S., Lou, C., Tamsir, A., Stanton, B. C. & Voigt, C. A. Genetic programs constructed from layered logic gates in single cells. *Nature* **491**, 249–253 (2012).

291. Polynikis, A., Hogan, S. J. & di Bernardo, M. Comparing different ODE modelling approaches for gene regulatory networks. *Journal of Theoretical Biology* **261**, 511–530 (2009).
292. Alon, U. *An introduction to systems biology: design principles of biological circuits*. (CRC press, 2019).
293. Hill, A. V. The possible effects of the aggregation of the molecules of haemoglobin on its dissociation curves. *j. physiol.* **40**, 4–7 (1910).
294. Lotka, A. J. The structure of a growing population. *Human Biology* **3**, 459–493 (1931).



# Chapter 2

## Model-driven design allows growth of *Mycoplasma pneumoniae* on serum-free media

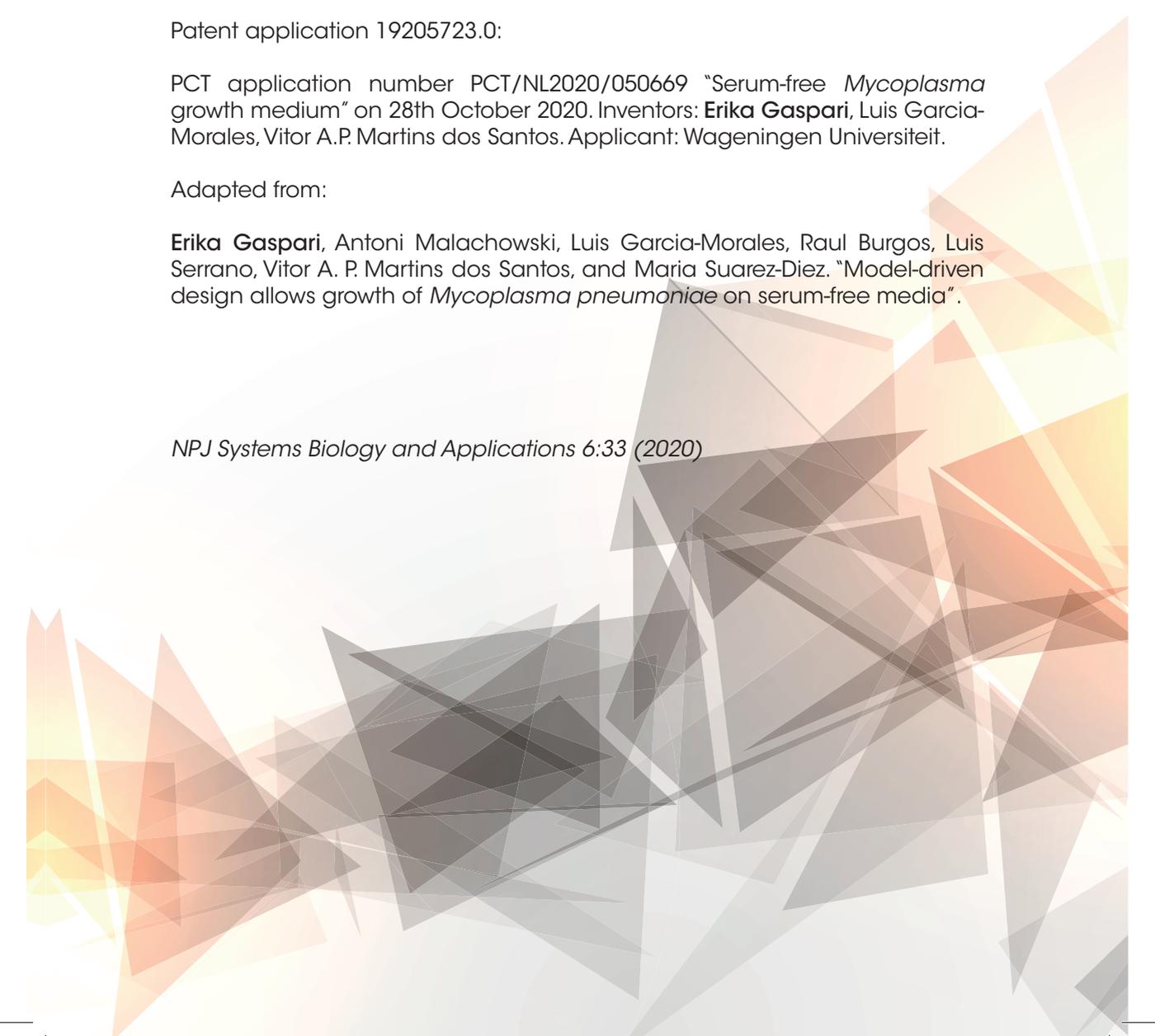
Patent application 19205723.0:

PCT application number PCT/NL2020/050669 "Serum-free *Mycoplasma* growth medium" on 28th October 2020. Inventors: **Erika Gaspari**, Luis Garcia-Morales, Vitor A.P. Martins dos Santos. Applicant: Wageningen Universiteit.

Adapted from:

**Erika Gaspari**, Antoni Malachowski, Luis Garcia-Morales, Raul Burgos, Luis Serrano, Vitor A. P. Martins dos Santos, and Maria Suarez-Diez. "Model-driven design allows growth of *Mycoplasma pneumoniae* on serum-free media".

*NPJ Systems Biology and Applications* 6:33 (2020)



## Abstract

*Mycoplasma pneumoniae* is a slow-growing, human pathogen that causes atypical pneumonia. Because it lacks a cell wall, many antibiotics are ineffective. Due to its reduced genome and dearth of many biosynthetic pathways, this fastidious bacterium depends on rich, undefined medium for growth, which makes large-scale cultivation challenging and expensive.

To understand factors limiting growth, we developed a genome-scale, constraint-based model of *M. pneumoniae* called iEG158\_mpn to describe the metabolic potential of this bacterium. We have put special emphasis on cell membrane formation to identify key lipid components to maximize bacterial growth. We have used this knowledge to predict essential components validated with *in vitro* serum-free media able to sustain growth.

Our findings also show that glycolysis and lipid metabolism are much less efficient under hypoxia; these findings suggest that factors other than metabolism and membrane formation alone affect the growth of *M. pneumoniae*.

Altogether, our modelling approach allowed us to optimize medium composition, enabled growth in defined media and streamlined operational requirements, thereby providing the basis for stable, reproducible and less expensive production.

## Introduction

Mycoplasma are a small genus of bacteria belonging to the Mollicute class,<sup>1</sup> comprising 124 species, 14 of which are human pathogens,<sup>2</sup> while others infect farm animals, herd animals and pets.<sup>3</sup> Most large-scale animal farms make use of antibiotics to fight and prevent infections, which often lead to development of antibiotics-resistant bacteria.<sup>4-6</sup> Moreover, chronic infections caused by several Mycoplasmas are not immediately detectable. Therefore, the use of effective vaccines against various Mycoplasmas in farm animals might be a solution to prevent wide spreading of these chronic infections<sup>7</sup>. It has been suggested that, despite being primarily considered a human pathogen, *Mycoplasma pneumoniae* could potentially be used as a universal chassis to be deployed as single- or multi-vaccine in a range of animal hosts ([www.mycosynvac.eu](http://www.mycosynvac.eu)).

*M. pneumoniae* causes atypical or walking pneumonia.<sup>8</sup> With its highly reduced genome (816 kb), it is among the smallest self-replicating living organisms and has an exclusively parasitic lifestyle. This feature greatly determined its evolution towards minimal functions (e.g. absence of vitamins and lipids synthesis) and inability to detoxify metabolic waste (e.g. hydrogen peroxide), tasks for which this bacterium fully relies on the host.<sup>9</sup> *M. pneumoniae* primarily ferments glucose<sup>10</sup> and uses most of its energy (71 to 88%) for non-growth associated tasks<sup>11</sup>. One of the key features of this pathogen during infections is its ability to evade the immune system: the unique characteristics of its membrane (in particular the high cholesterol contents) allows the mimicking of the host membrane.<sup>12</sup> Moreover, the lack of a cell wall makes it resistant to antibiotics which target this structure.<sup>13</sup>

The development of vaccines based on *M. pneumoniae* requires the large-scale production of this bacterium in a serum-free defined medium. This is due to the high economic cost of rich media, the variable composition of animal-derived compounds<sup>14</sup> and the possible presence of toxins, viruses or antigens that could reduce the sensitivity of immunological assays used on farms to screen for infection.<sup>15</sup> However, the production of *M. pneumoniae* in serum-free media is a major challenge. For decades, the ability to culture *M. pneumoniae* was reported to be possible only in serum-rich media<sup>16</sup> due to a requirement for sterol components.<sup>17</sup> Only in 2009 Yus et al. reported a defined medium based on the metabolic map reconstruction of *M. pneumoniae* that allowed some growth of the bacterium but required daily changes of the medium<sup>18</sup>. This medium is not valid for large-scale growth needed during vaccine production. For these reasons, we focused our study on the model-driven design of a serum-free medium that could improve *M. pneumoniae* growth rate.

We approached the study through metabolic modeling, a method widely used for genome-scale biochemical networks analysis.<sup>19-26</sup> We previously described the development of a genome-scale, constraint-based model (GEM) of *M. pneumoniae* metabolism, iJW145,<sup>11</sup> and its use to carry out predictions related to growth rate and medium design, relying upon a biomass composition that had been formulated from literature,<sup>3</sup> sequencing, proteomics and mass spectrometry data. However, no novelty in the defined medium composition was reported<sup>11</sup>. Several studies have shown that lipids are among the most important component limiting growth, mainly cholesterol, regulating membrane fluidity (i.e. viscosity of the lipid bilayer), and fatty acids.<sup>1,27-30</sup> Determining the lipids composing the membrane of *M. pneumoniae* is challenging, as it is strongly dependent on the composition of the growth medium, growth phase and culture conditions.<sup>31</sup> Some key features remain invariant, such as

the high cholesterol proportion, directly incorporated from the medium.<sup>32</sup> The model iJW145 did not account in detail for lipids in its biomass composition<sup>11</sup>. Therefore, we integrated native lipid pathways of *M. pneumoniae*, and reactions involved in the membrane formation to develop an updated GEM of *M. pneumoniae* metabolism, called iEG158\_mpn.

We aimed to describe the use of the energy available, considering not only cytosolic processes but also transmembrane transport. We used the model iEG158\_mpn to predict the essentiality and synergism of components in serum-free media for *M. pneumoniae* and to study how certain culture conditions may affect the growth of this bacterium, focusing on its intracellular metabolism. These predictions were corroborated with defined-growth media that support robust growth (up to 60% of the biomass obtained with rich medium) and shows the synergism of components. Our study, in combination with the existing literature, suggests membrane formation and adaptability are key aspects to be investigated in relation with *M. pneumoniae* growth, which appears to be limited by several factors.

## Results

### Features of iEG158\_mpn

The model iEG158\_mpn of *M. pneumoniae* metabolism has two compartments, one extracellular and one cytosolic, and consists of 490 reactions, 442 metabolites and 158 genes. Of the total reactions, 314 are gene-associated, 329 are cytosolic conversion reactions, 65 represent transport between the extracellular compartment and the cytosol and vice versa, and 96 are of exchange, meaning they allow uptake/production of compounds. Only 61 of the cytosolic reactions have a zero-flux range, meaning they are not active under the simulated conditions. Excluding the zero-flux reactions, 293 reactions have same minimum and maximum fluxes according to FVA, meaning in total 354 reactions with fixed flux. The high proportion of active reactions and the lack of flexibility in the usage of alternative reactions reflect the evolution of the organism in minimal genome. In the biomass equation, formulated according to Wodke et al.<sup>11</sup>, cofactors and vitamins, known to be essential for growth but with no available measurements, were symbolically included as traces in the biomass equation. Similarly, DNA repair and RNA turnover were considered in the biomass composition by slightly increasing their proportions. Protein turnover is modeled by the addition of degradation reactions,<sup>33</sup> which is reflected in iEG158\_mpn as the decomposition of a general protein compound into amino acids with the associated ATP cost.

The very small size of *M. pneumoniae* contributes to the increase in maintenance costs, as the high surface-to-volume ratio induces higher proton leakage through the membrane. The ATP maintenance reaction was determined in iJW145 by multiplying the number of ATPase complexes on the cell surface by the number of cells in a gram of biomass, so considering only the cytosolic side of the energy requirement. However, the reaction fluxes representation allowed the visualization of a flux of protons out of the cell, required for cytosol deacidification. Moreover, it is stated in Wodke et al.<sup>11</sup> that the proportion of ATP used for cellular tasks not directly related to growth (which corresponds to total non-growth associated maintenance energy in iJW145) is about 76% at 24h after inoculation. We updated the non-growth-associated maintenance reaction (model reaction ID: "ATPM") to specifically include proton leakage component (66% of the total ATP consumption not associated to growth)(model reaction ID: "protonLeak"), as represented in Figure 1. Furthermore, this introduced a link between the non-growth associated maintenance equation and the proton production/consumption reactions.

The maintenance reaction in iEG158\_mpn was further updated according to the results of the sequence alignment for monocarboxylic acids transporters discovery: no lactate or acetate proton symporter were identified to be present in *M. pneumoniae* genome. The respective transport reactions (IDs "LAClt" and "ACtr") were therefore updated in iEG158\_mpn as irreversible, removing the proton symport and the maintenance costs adjusted to the maximum value allowing growth, computed as 10.46 mmol.g<sub>DW</sub><sup>-1</sup>.h<sup>-1</sup> of ATP (g<sub>DW</sub>=grams dry weight) used for non-growth-associated tasks at the quasi-steady state (24h time point).

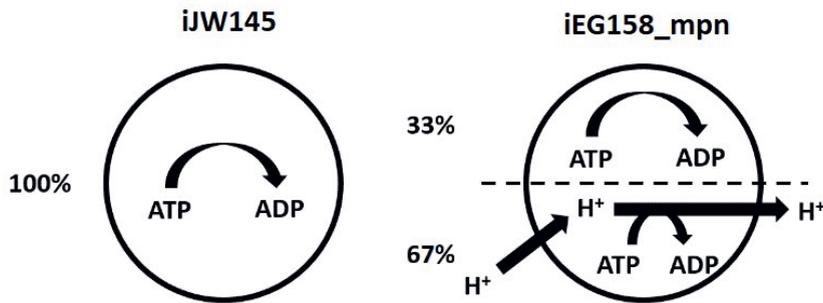
Assuming the total maintenance cost,  $M_t$ , equals to 10.46 mmol.g<sub>DW</sub><sup>-1</sup>.h<sup>-1</sup> of ATP, the energy fraction spent for proton leakage,  $E_p$ , equals to 0.66 and the fraction spent on other maintenance requirements,  $E_L$ , equals 0.33. Therefore, the lower bounds  $P_{LB}$  and  $L_{LB}$  of the correspondent reactions are computed as:

$$P_{LB} = M_t \cdot E_P \cdot 4 \quad [1]$$

$$L_{LB} = M_t \cdot E_L \quad [2]$$

$P_{LB}$  is multiplied by 4 as the ATPase reaction removes four protons at a time.

We simulated the growth rate of *M. pneumoniae* with iEG158\_mpn supplementing *in silico* the medium predicted with model iJW145: the resulting growth rate is 0.053 divisions per hour at quasi-steady state of MPN (24h), correspondent to a doubling time of 13.4 hours, in line with the experimental evidences.



**Figure 1.** Representation of maintenance energy exploitation in two GEMs of *M. pneumoniae*. a) Maintenance energy in model iJW145 and b) Maintenance energy in model iEG158\_mpn. In iEG158\_mpn simulations, the ATPase is observed to run in reverse to allow efflux of protons for cytosol de-acidification. Maintenance reaction in iEG158\_mpn was therefore updated to explicitly include a proton leakage component (66%).

### Reconstruction of membrane lipids in iEG158\_mpn

Because membrane adaptability is one of the key features of *M. pneumoniae*, we assumed that accounting for membrane formation in the model would give important insights into its growth requirements. Focusing on membrane lipid components, we included in iEG158\_mpn the pathways for lipids assembly<sup>34</sup> and reconstructed the membrane lipid composition through an extensive literature search. Articles listed in this section were used to determine lipids percentages in the membrane and their acyl chains composition. Only lipids whose percentage in the membrane has been reported to be higher than 1% of the total lipid biomass were considered. Presence and proportions of the lipids are strongly dependent on the compound's availability in the medium, therefore quantities must be interpreted as ranges.

One of the key features of the membrane of *M. pneumoniae* is the high proportion of cholesterol, from 35% to 50% of the total lipid fraction.<sup>35</sup> *In vitro*, cholesterol is directly incorporated from the medium without being processed by the bacterium. For this reason, in iEG158\_mpn, cholesterol has been included into the biomass composition without being

metabolically processed. Available literature indicates it constitutes about one third of the final membrane and about half of the lipid components of the membrane.

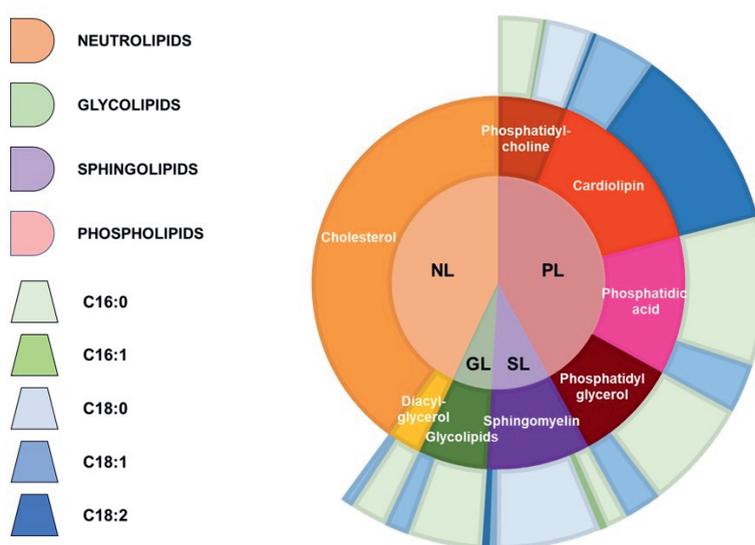
Most of the studies on *M. pneumoniae* utilize Hayflick media, composed for 16% of horse serum, rich in sphingolipids and phosphatidylcholine.<sup>36</sup> Sphingolipids-phosphatidylcholine ratio in *M. pneumoniae* is 2.4,<sup>37</sup> constituting on average 12% and 5% of the membrane lipids, respectively.<sup>38</sup> However, sphingolipids have a good affinity for cholesterol, while phosphatidylcholine repels it, meaning their proportions in the membrane are strongly correlated to the amount of cholesterol directly incorporated from the medium: a higher amount of cholesterol leads to the incorporation of a higher fraction of sphingolipids and a lower fraction of phosphatidylcholine<sup>39</sup>. The interrelation between these two lipids suggests their essentiality in the membrane is strongly related to their fatty acid chains, being the only lipids carrying the unsaturated fatty acid chain C18:2, whose presence might be determinant in *M. pneumoniae* membrane. Moreover, sphingomyelin incorporated into the membrane has a lower proportion of fatty acid chains with 20 carbons or more in respect to the sphingolipids in the serum<sup>40</sup>, suggesting *M. pneumoniae* preferentially incorporates in the membrane sphingomyelin with less than 20 carbons-fatty acid chains. *M. pneumoniae* is a fatty acid and sterol auxotroph, but it can assemble phosphatidic acid, diacylglycerol, phosphatidylglycerol and glycolipids from provided fatty acids in the medium.<sup>34</sup> Glycolipids constitute a range of 5-15% of the total membrane lipids,<sup>41</sup> while the remaining percentage should be constituted by phosphatidylglycerol; phosphatidic acid and diacylglycerol, being lipid synthesis intermediaries, are not found in relevant proportions in the membrane.

The lipids, with their specific molecular weight, were included in the biomass equation to reflect the predicted membrane composition. We assumed a lipid composition is optimal when it confers optimal fluidity, and so permeability, to the membrane of *M. pneumoniae*, where for optimal fluidity it is meant the combination of thickness and viscosity that allows *M. pneumoniae* to perform its membrane-related activity. While the optimal thickness is considered to be given by a range of lipid proportions, i.e. measured when the bacterium is grown on rich-medium with full metabolites and lipids availability, the optimal viscosity is given by the fatty acids distribution that *M. pneumoniae* aims at incorporating in the membrane when lipids availability is limited. Therefore, the acyl chains of the lipids constitute a critical aspect of our study.

With the aim of linking the membrane lipid profile to the availability of fatty acids in the medium, we reconstructed the acyl chains of the various lipids composing the membrane by considering the sum of fatty acid chain fractions previously determined by Wodke et al.<sup>11</sup> and Worliczek et al.<sup>40</sup>; a comparison of the total fatty acid composition between iEG158\_mpn and the ones available in literature is provided in [Supplementary Figure 1](#). *M. pneumoniae* tends to selectively incorporate palmitic acid over stearic acid in the membrane<sup>30</sup>, with a preference for saturated fatty acids, suggesting oleic and linoleic acids are incorporated in a lower proportion with respect to the amount found in the media. This indicates that the proportion of different acyl chains for de-novo assembled phosphatidic acids is conserved during assembly of downstream lipids phosphatidylglycerol, diacylglycerol and glycolipids. We assumed phosphatidylcholine could be degraded *in silico* into glycerol-3-phosphocholine and fatty acids, to represent the re-arrangements of the phospholipid's acyl chains, which could be triggered by membrane lipases *in vivo*<sup>42</sup>. Sphingomyelin, on the other hand, is in our model directly incorporated into the membrane without modifications.

A representative average membrane lipid composition of *M. pneumoniae* grown on serum rich medium is shown in [Figure 2](#), with its respective and total fatty acid chains. In order to predict

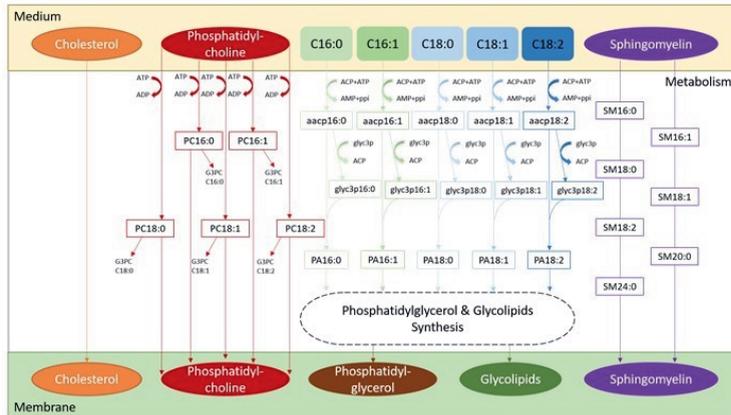
*in silico* the composition of serum-free medium that would maximize the growth of *M. pneumoniae*, we added to the biomass equation the above representative medium composition, excluding those whose supplementation is linked to the addition of serum or would deviate from the established fatty acids proportions (e.g. cardiolipin). We excluded phosphatidic acid from the biomass equation in iEG158\_mpn and consider it solely as precursor of phosphatidylglycerol and glycolipids. Figure 3 consists of a visual representation of lipid pathways and membrane construction as they have been implemented in iEG158\_mpn. The membrane is then composed of cholesterol, phosphatidylcholine, phosphatidylglycerol, glycolipids and sphingomyelin. We reconstructed their pathways within the metabolism of *M. pneumoniae*.



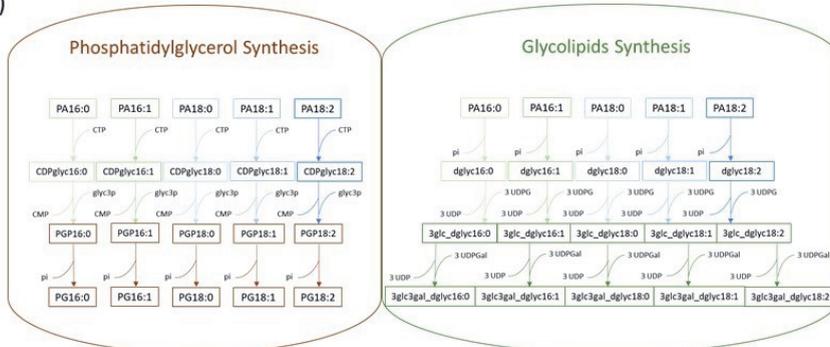
**Figure 2.** Reconstructed average composition of *M. pneumoniae* membrane lipids and their respective fatty acid chains when the bacterium is grown on a serum-rich medium. The inner circle indicates lipid groups: neutrolipids (NL), glycolipids (GL), sphingolipids (SL) and phospholipids (PL). Lipids are represented in the intermediate circle as belonging to the different groups of the inner one; therefore, PL are constituted by phosphatidylcholine, cardiolipin, phosphatidic acid and phosphatidylglycerol, SL by sphingomyelin, GL by glycolipids and NL by cholesterol and diacyl-glycerol. Each lipid is then represented in the outer circle with their average fatty acid chains composition: the different proportions are made of palmitic acyl chains (C16:0), palmitoleic acyl chains (C16:1), stearic acyl chains (C18:0), oleic acyl chains (C18:1) and linoleic acyl chains (C18:2).

The possible ranges of lipids and fatty acids proportions assuming growth in serum-free media are summarized in [Supplementary Table 1](#).

A)



B)



2

**Figure 3.** Implementation scheme of the lipid pathways and membrane formation in iEG158\_mpn, when *M. pneumoniae* is grown on serum-free medium. The wide variety of lipid species was simplified by considering all carry a representative acyl chain distribution instead of a mix with different acyl chain configurations. This simplification considerably reduces the complexity of the model keeping the same amount of quantitative information. A) Cholesterol is directly incorporated in the membrane as well as sphingomyelin (SM), for which all the different fatty acid chains proportions have been considered. Phosphatidylcholine is imported using ATP and either goes to build the membrane or it is degraded into its fatty acid chains and glycerol-phosphocholine (G3PC). All the fatty acids introduced in the medium (C16:0, C16:1, C18:0, C18:1, C18:2) will be linked to an acyl-carrier protein (ACP) at the expense of ATP. The ACP then releases the fatty acid chain to glycerol 3-phosphate (glyc3p) and the product, reacting with the ACP carrying fatty acids, leads to the production of phosphatidic acid (PA). B) PA then undergoes synthesis of phosphatidylglycerol (PG) or glycolipids.

## Model simulations predict the addition of two lipids in growth media greatly increases the growth rate of *M. pneumoniae*

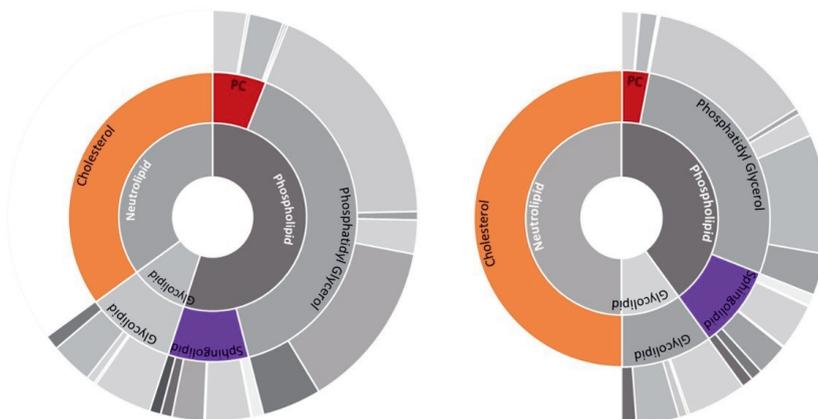
We integrated into iEG158\_mpn reactions involved in fatty acids, lipids and membrane formation pathways to reconstruct the membrane of *M. pneumoniae* and predict essential fatty acids and lipids that, if added to the medium, could allow growth. The lipids in the biomass reaction of iEG158\_mpn have been distributed as reported in [Table 1](#).

**Table 1.** Lipid proportions in terms of percentage of total biomass and percentage of total biomass lipids.

LIPIDS	% OF TOTAL BIOMASS	% OF TOTAL BIOMASS LIPIDS
Glycolipids	2.46	5.82
Cholesterol	18.13	42.92
Phosphatidylcholine	2.58	6.11
Phosphatidylglycerol	12.93	30.62
Sphingomyelin	6.14	14.53

iEG158\_mpn was used to identify media components for increased growth, predicting all the components already highlighted in Yus et al.<sup>18</sup>, with addition of sphingomyelin and phosphatidylcholine. Our analysis pin-points the synergy between the three key lipid components (cholesterol, sphingomyelin and phosphatidylcholine) that were integrated in iEG158\_mpn, in terms of percentages of total lipid mass in the membrane, as follows: the most enriched components must be cholesterol, constituting a range of 35-50% of the total lipids; sphingomyelin and phosphatidylcholine proportions are adjusted accordingly, as sphingomyelin has high affinity for cholesterol, while phosphatidylcholine has a low one.<sup>39</sup> Therefore, they constitute respectively 9-15% and 6-10% of the total lipids. [Figure 4](#) shows the interrelation between cholesterol, phosphatidylcholine and sphingomyelin (or in general sphingolipids) proportions in the membrane of *M. pneumoniae* as simulated by iEG158\_mpn. Sphingomyelin is preferably incorporated carrying an 18-carbon acyl chain and, therefore, its essentiality might be linked to it being the only one carrying the acyl chain C18:2.

The essential fatty acids to be incorporated in a serum-free medium are palmitic (C16:0) and oleic (C18:1) acids, as they are the preferred components of phosphatidic acid in *M. pneumoniae* and its downstream products phosphatidylglycerol and glycolipids. Stearic acid (C18:0) presence is important depending on phosphatidylcholine proportion: once phosphatidylcholine is incorporated, its unsaturated fatty acid chains are replaced with saturated ones, resulting in the membrane di-saturated phosphatidylcholine.<sup>43</sup> Moreover, supplementation of phosphatidylcholine might be important not only as membrane component but also as source of those fatty acids whose presence has been detected in lower percentage with respect to palmitic and oleic acids: linoleic acid (C18:2) and palmitoleic acid (C16:1).

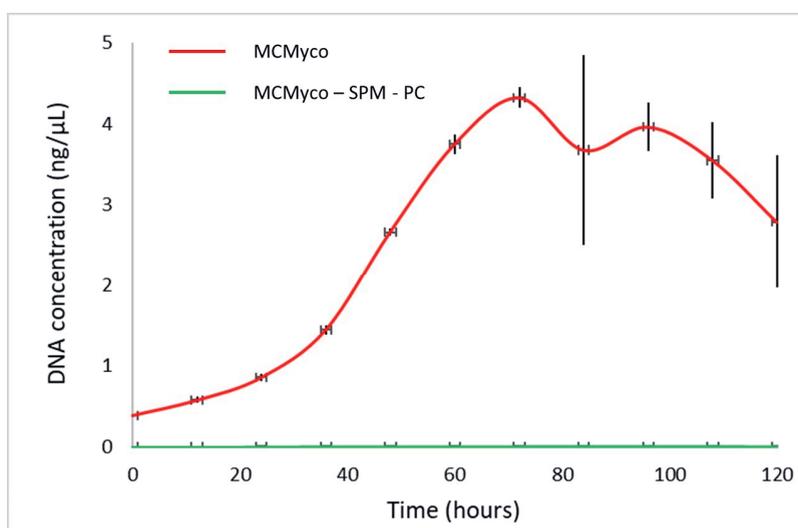


**Figure 4.** Interrelation between cholesterol, sphingomyelin and phosphatidylcholine proportions in the membrane of *M. pneumoniae* as simulated by iEG158\_mpn when grown in silico on the predicted optimal serum-free medium. a) If cholesterol constitutes 35% of the total membrane lipids, phosphatidylcholine (PC) proportion is up to 10%, while sphingomyelin/sphingolipids is reduced at its minimum to 12%. b) If cholesterol constitutes 50% of the membrane lipids, phosphatidylcholine (PC) proportion decreases to 6%, while sphingomyelin increases to 15%. Outer cycle represents the average proportions of acyl chains carried by each lipid, as also showed in Figure 3: the grey colors represent, in order of shading from the lightest to the darkest, C16:0, C16:1, C18:0, C18:1, C18:2.

### The developed serum-free semi-defined medium MCMyco supports robust growth of *M. pneumoniae*

Following the model predictions, we have added sphingomyelin and phosphatidylcholine to the so-called MC medium, previously used for growing *M. florum*.<sup>44</sup> Cholesterol and fatty acids were added as previously reported in Yus et al.<sup>18</sup> to be essential for *M. pneumoniae* growth. The new developed medium has been called MCMyco and its composition consists of Mycoplasma broth base supplemented with 17% yeast extract 15% Solution (Gibco), 0.5% glucose, 500 U/mL of penicillin G and 0.002% phenol red. Delipidated Bovine Serum Albumin was added at a final concentration of 0.5% premixed with the lipid components predicted to be essential by model simulations: fatty acids, cholesterol, sphingomyelin and phosphatidylcholine. We refer to this medium as “semi-defined” due to the presence of the undefined component yeast extract in the MC base. However, yeast extract is reported to contain amino acids, carbohydrates, vitamins and minerals.

MCMyco was used to grow *M. pneumoniae* strain M129, resulting in a doubling time of 17.6 hours after three culture passages, compared to the doubling time of about 8 hours obtained in rich media.<sup>18</sup> The importance of sphingomyelin and phosphatidylcholine is hereby confirmed by the absence of *M. pneumoniae* growth when the two phospholipids are removed from MCMyco (Figure 5).

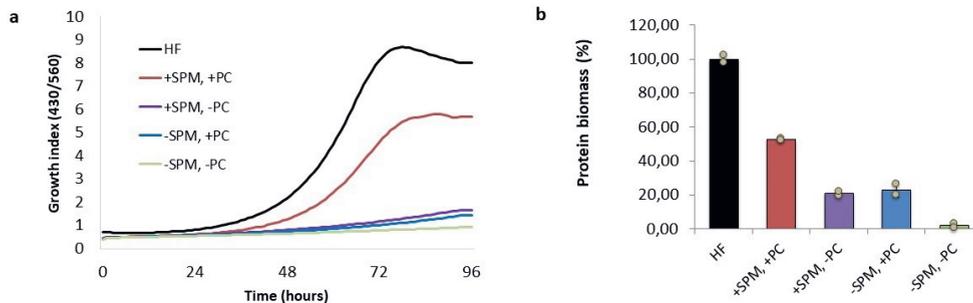


**Figure 5.** Growth of *M. pneumoniae* strain M129 on serum-free semi-defined medium MCMycO measured by genomic DNA quantification after three culture passages. Growth is detected when *M. pneumoniae* is grown on full MCMycO medium (red) and in the same medium after removing both sphingomyelin (SPM) and phosphatidylcholine (PC) (green). In absence of the two phospholipids, *M. pneumoniae* growth is undetectable. Measurements represent an average of 6 replicates and error bars represent the standard deviations of the 6 replicates.

### **Importance of phosphatidylcholine and sphingomyelin for growth of *M. pneumoniae* into the serum-free defined medium vB13**

The defined serum-free medium for *M. pneumoniae* published by Yus et al.<sup>18</sup> only supports very poor growth compared to Hayflick. Therefore, Burgos R., Garcia-Ramallo E., Shaw D., Lluch-Senar M. and Serrano L. improved this medium developing a new serum-free defined medium called vB13, resulting in about 60% of the biomass obtained in Hayflick ([Supplementary Figure 2](#)).

To show the importance of sphingomyelin and phosphatidylcholine in this medium, growth of *M. pneumoniae* was reported in absence of either one of or both the two phospholipids, as shown in [Figure 6](#). The lack of growth in absence of both phospholipids and the very poor growth observed when one or the other is missing confirms the essentiality of the concomitant presence of both phosphatidylcholine and sphingomyelin in a serum-free medium for *M. pneumoniae*.



**Figure 6.** Impact of phosphatidylcholine and sphingomyelin on *M. pneumoniae* strain M129 cell growth using the serum-free vB13 medium. (a) Growth curve analysis determined by metabolic growth index (430/560 absorbance rate) comparing cell growth in a Hayflick rich medium (HF) (black) with cell growth in vB13 (red), vB13 without phosphatidylcholine (PC) (purple), vB13 without sphingomyelin (SPM) (blue) and vB13 without both phospholipids (green). (b) Protein biomass measurement at 96 h, corresponding to the end of the growth curve shown in panel A. Data represent the mean of two replicates and error bars the standard deviation.

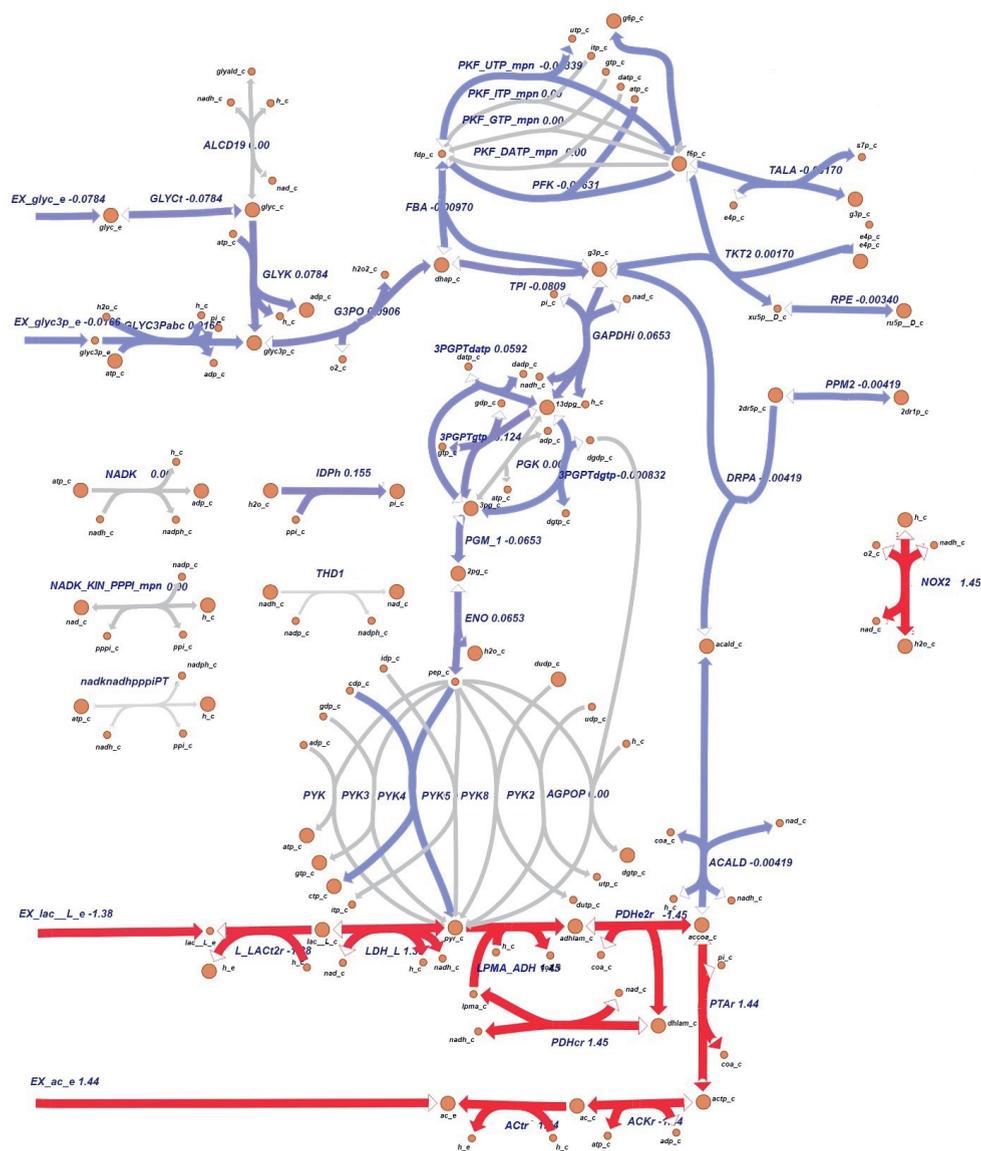
2

### Effect of hypoxia on *M. pneumoniae* metabolism

We used Escher to visualize change of fluxes when simulating different environmental conditions: Figure 7 shows flux changes when oxygen availability is reduced, as no growth occurs in complete absence of oxygen.<sup>45</sup> With no oxygen limitation and glucose uptake rate of  $5.11 \text{ mmol.g}_{\text{DW}}^{-1}.\text{h}^{-1}$ , the model predicts that *M. pneumoniae* takes up  $7.54 \text{ mmol.g}_{\text{DW}}^{-1}.\text{h}^{-1}$  of oxygen and no growth for oxygen uptake rates lower than  $4.81 \text{ mmol.g}_{\text{DW}}^{-1}.\text{h}^{-1}$ . The growth rate when oxygen is limited to  $6 \text{ mmol.g}_{\text{DW}}^{-1}.\text{h}^{-1}$  is limited by about 2.3-fold. FVA results for all the reaction fluxes in baseline and hypoxia conditions are reported in Dataset 1, where the reactions included in Figure 7 are highlighted. Here the fluxes are represented in terms of differences between hypoxia and baseline conditions, given that the solutions for the flux distributions are unique. Despite the overall fluxes' distribution being slightly reduced under oxygen limitation, the number of zero-flux reactions in this condition is 84. Excluding these, 299 reactions have same minimum and maximum allowed fluxes. The uniqueness of these solutions allows direct comparison of both conditions.

Not considering the zero-flux reactions, only 18 show the same range of fluxes comparing the two situations (baseline and reduced oxygen) in iEG158\_mpn, assuming proton leakage and maintenance energy (ID: "ATPM") do not change. All the lipid synthesis reactions show a consistent flux reduction, as well as the reactions related to the carbon degradation pathways and the NAD kinase. Imports of some amino acids and nucleobases (i.e. thymine, ID: "THMt6") show a flux decrease when oxygen is reduced. Contrarily to the general tendency, deoxyuridine kinase and phosphorylase show a higher flux respect to the baseline situation.  $\text{CO}_2$ ,  $\text{H}_2\text{O}_2$ , acetic acid and lactic acid productions increase under hypoxia.

The whole-metabolism visualization of reaction flows under baseline conditions and when oxygen availability is reduced are reported in Dataset 2.



**Figure 7.** Visualization of flux differences in glycolysis reaction when *M. pneumoniae* is under hypoxia (oxygen uptake 6 mmol.gDW<sup>-1</sup>.h<sup>-1</sup>) respect to baseline oxygen availability (7.54mmol.gDW<sup>-1</sup>.h<sup>-1</sup>). Purple is indicative of a small difference (absolute value of flux change <math>< 1</math>) and red of a more considerable one (absolute value of flux change -1.h<sup>-1</sup>.

## Discussion

Owing to its fastidious requirements, *Mycoplasma pneumoniae* has been thus far reported to grow on media containing serum<sup>46,47</sup> and only poorly in defined medium<sup>18</sup>, furthermore requiring daily changes. The absence of serum elements in the culture medium facilitates the large-scale production of *M. pneumoniae* for medical and pharmaceutical purposes, thanks to the reduced costs, greater reproducibility, definite composition and lack of unexpected serum-plate-agglutination (SPA) antigens<sup>15,48</sup> and other unpredictable contaminants.

Hereby, by using *in silico* modeling approaches based on literature data, we propose that sphingomyelin and phosphatidylcholine are essential growth factors allowing *M. pneumoniae* to grow on serum-free media. This prediction was validated by i) adding these two phospholipids, together with cholesterol and fatty acids, to the previously developed MC medium<sup>47</sup>. This led to the development of a medium called MCMyco, supporting robust *M. pneumoniae* growth; and ii) removing sphingomyelin and phosphatidylcholine from the defined serum-free medium vB13 (Burgos R., Garcia-Ramallo E., Shaw D., Lluch-Senar M. and Serrano L.), showing the absence of either both phospholipids or one of the two causes lack of growth. Therefore, our study shows, theoretically and experimentally, that supplementation of sphingomyelin and phosphatidylcholine to serum-free media, together with cholesterol, enables growth of *M. pneumoniae*. Moreover, in agreement with the model, we have shown in vB13 that the effect of the two phospholipids is synergistic and that the different incorporation of these lipids depends on their availability in the medium. Indeed, the lipid composition of the membrane might be strongly variable according to the environmental conditions, which affect its permeability<sup>49</sup> and curvature<sup>50</sup>; for instance, a sphingomyelin-cholesterol packing results in a very tight membrane<sup>51</sup>. Therefore, the key medium lipids herein found confirm that one of the main limitations of *M. pneumoniae* growth resides in its membrane composition, viscosity and permeability. This is, in turn, determined by the proportion of acyl chains in the membrane lipids. While the requirement of cholesterol for growth have been well known for decades,<sup>28,52</sup> phosphatidylcholine supplementation in a medium for *M. pneumoniae* was first reported in 2009<sup>18</sup>. Sphingomyelin, in contrast and to the best our knowledge has not been specifically reported as to be an essential component in media deployed to grow *M. pneumoniae*, although its positive effect on growth was previously suggested in other Mollicutes<sup>53</sup> and observed in Spiroplasma.<sup>54</sup>

Having established that lipids supplementation is a key aspect for *M. pneumoniae* growth, the way they are supplemented might be critical: *M. pneumoniae* cultures where lipids are supplemented as cholesterol-phosphatidylcholine vesicles show a similar growth rate of the ones in serum-rich media.<sup>55</sup>

Medium composition is not the only growth-limiting factor of *M. pneumoniae* growth *in vitro*: batch cultures may result in limitation in oxygen supply as a result of gradients that arise due to biofilm formation. Our analysis shows oxygen level is critical, affecting the whole metabolism, with only a few reaction fluxes that remain intact under hypoxic conditions in respect to baseline conditions. Low oxygen not only pushes the lactate dehydrogenase towards the production of lactate, but also critically affects the whole glycolytic pathway, with a reduction of sugar uptake and ATP production. More NADH is consumed and less is regenerated by the GADPH enzyme, which represents one of the main bottlenecks of the metabolic network. The production and accumulation of compounds such as lactate and

acetate augments, enhancing toxicity and acidifying the medium when exported: it is observed that the doubling time of *M. pneumoniae* in batch culture increases over time due to the decrease in media pH<sup>11</sup> and the accumulation of toxic waste compounds, namely acetic and lactic acid. Cells growing in an acidified medium have a different membrane composition<sup>31</sup>. It is, however, still unclear whether the adaptability of *M. pneumoniae* membrane resides in its genome-coding function, if it is a chemical-physical response of the lipid bilayer to pH changes or if it is a combination of these two aspects.

The cytosol of Mycoplasmas does not maintain a constant pH, with the risk of being unable to maintain the proton gradient required for trans-membrane transport<sup>56</sup> and to alter the internal biochemical state. Our model iEG158\_mpn improves iJW145 by not only integrating the membrane, the specific lipids with their pathways and carried acyl chains, but also rearranging the ATP distribution in the maintenance energy, taking into account the effort made by *M. pneumoniae* while de-acidifying the cytosol: a cytosolic pH decrease is equivalent to an increase in the H<sub>3</sub>O<sup>+</sup> intracellular concentration, which might alter reactions involving H<sup>+</sup> as reactant.<sup>57</sup> This might be a cause of high energy maintenance costs and reduced growth rate. For this reason, a study on the effects of the intracellular pH change on the metabolism of *M. pneumoniae* would be of main interest.

A study of *M. pneumoniae* under different environmental conditions would not only provide insights into the way it reacts to pH changes, but also disclose the strategies bacteria with a reduced genome use to exploit the host's resources to accomplish essential tasks, whose capabilities it has lost during the course of evolution: our analysis highlights a lack of metabolic versatility, reflected by an *in silico* deficiency of flux variability, possibly due to an extremely reduced metabolic network size.

Altogether, our model provides the framework for an integrated approach, incorporating available literature data to systematically discover dependencies that are not of immediate recognition, such as relevant synergies and interconversions, metabolic bottlenecks, cofactor regeneration and redox balancing. Pinpointing knowledge gaps and essential pathways, it demonstrates the usefulness of modeling resources for exploring the metabolic capacity of mycoplasmas, and the use of these insights to rationally influence, modulate and design conditions for optimized performance.

## Methods

### Reconstruction of the genome-scale, constraint-based metabolic model

The metabolic model iJW145<sup>11</sup> was extensively updated, manually curated and expanded to generate the genome-scale metabolic model iEG158\_mpn. Reactions and metabolites have been annotated adding BiGG Models identifiers<sup>58</sup> when possible. Reaction reversibility has been assessed through eQuilibrator,<sup>59</sup> a thermodynamic calculator of the Gibbs energy released by a specific reaction at a set pH and ionic strength (in our case, respectively 7.0 and 0.1M). Reactions where changes in Gibbs energy comprised between -30 and 30 kJ/mol were considered reversible, otherwise irreversible. All model reactions have been manually verified using information from the MyMpn database.<sup>60</sup> Information gathered from literature was used to modify the biomass synthesis reaction adding detailed information on lipid composition. The biomass equation of iEG158\_mpn has been standardized to obtain a molecular weight of 1 g.mmol<sup>-1</sup>. Reactions for fatty acid integration into membrane lipid chains have been added. The genome-scale model iEG158\_mpn is provided in Dataset 3 in SBML format and its syntax and consistency has been checked with SBML Validator<sup>61</sup> and MEMOTE, which reported stoichiometry consistency 99.1%, charge balance 87.8% and metabolite connectivity 97.7%.<sup>62</sup> A list of all reactions and metabolites IDs is given in Dataset 4. The model has been deposited in BioModels<sup>63-65</sup> and assigned identifier MODEL1912060001. A python script to simulate growth with this model, imposing constraints and supplementing *in silico* the medium, is provided as Dataset 5.

2

### Model simulations and medium design

Pathways and whole-metabolism maps for fluxes visualization were constructed with the tool Escher,<sup>66</sup> which graphically depicts a solution for maximum attainable growth rate, as computed by Flux Balance Analysis (FBA), with biomass yield, as a proxy for growth, as the objective function.<sup>67</sup> FBA solves a linear programming problem, which is characterized by having more variables than equations, meaning more than one optimal solution can exist for the same objective function maximization. FBA computes a possible flux distribution leading to the optimal objective solution. The same optimal growth rate might be reached by several different reaction flux combinations. Flux Variability Analysis (FVA) explores the range of each metabolic flux at the optimal solution.<sup>68</sup> Both FBA and FVA were run with biomass synthesis reaction as objective function to maximize, respectively, through the commands *optimize()* and *flux\_variability\_analysis()* of the Python tool *cobrapy*,<sup>69</sup> version 0.5.11 (Python version 3.4.4). Exchange reactions in a genome-scale metabolic model allow metabolites to enter and egress the *in silico* network. Simulations with iEG158\_mpn were performed by conserving the same flux constraints as have been experimentally determined: these assumed unlimited availability of the compounds which are not directly metabolized for energy or carbon and/or cannot be measured experimentally (vitamins, cofactors). Additional fatty acids that could be metabolized or integrated as membrane lipid chains (palmitic acid, palmitelaidic acid, stearic acid, oleic acid, linoleic and linolelaidic acid) have also been assumed to have unlimited

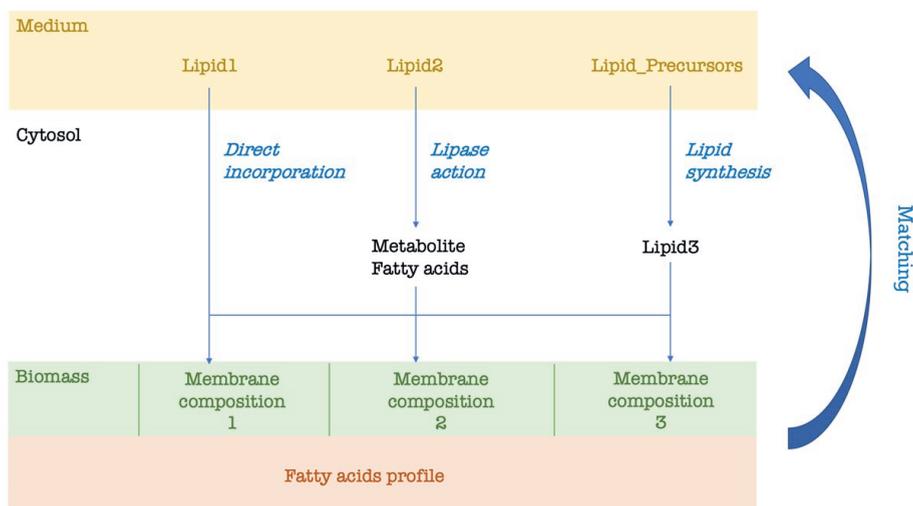
availability as no related uptake rates were experimentally measured. Specifically, exchange reactions corresponding to metabolites whose uptake or secretion rate was not measured were set with bounds of  $\pm 1000$  when production and secretion was possible, with a lower bound of  $-1000$  and an upper bound of  $0$  when only uptake from medium was possible and with a lower bound of  $0$  and an upper bound of  $1000$  when describing secretion.

Growth medium optimization consists of changing the bounds of exchange reactions to enable or limit uptake or secretion in order to find the combination of supplemented compounds that maximizes the flux through the biomass reaction if this corresponds to maximizing growth rate. We tested the possible uptake of additional native compounds not previously considered in *M. pneumoniae* growth media, but for which transporters are present in *M. pneumoniae* genome, by adding the corresponding exchange reactions. To equilibrate the energetics in the model, we assigned a cost of 1 proton for the uptake of uncharged compounds, 1 ATP for the uptake of compounds with a charge of 1 or 2 and 2 ATPs for the one of compounds with charge greater than 2. To design a minimal serum-free chemically defined medium that optimizes *M. pneumoniae* growth, we iterated through all the exchange reactions present to find the combination of bounds that lead to the highest biomass production: we have initially set all the exchange reactions' lower bounds to  $-1000$ , allowing the metabolites of interest into the system. All the lower bounds of the exchange reactions were set to  $0$  one at a time, to be then reprinted to  $-1000$  for the next iteration (next exchange reaction to close). When biomass yield was not changing in consequence, we assumed the component was not required for maximizing growth rate. Each time a component was removed from the list, a new set of iterations was started on all the other components. The minimal medium was obtained once all the remaining components, when removed (i.e. when the correspondent exchange reaction's lower bound was set to  $0$ ), were changing the biomass yield obtained by FBA. The relation between the growth rate and the doubling time has been computed using the exponential function  $e^{\mu T}=2$ , where  $\mu$  is the growth rate in divisions per unit of time and  $T$  is the doubling time.

To simulate hypoxia, we first ascertained, through Flux Balance Analysis, the minimal oxygen uptake rate required for optimal growth by iEG158\_mpn. By allowing *in silico* an infinite supplementation of oxygen ("EX\_O2\_e" reaction lower bound =  $-1000$ ), we ran FBA towards flux maximization of the biomass reaction. Constraining the growth rate to the maximal one, following FBA with objective function "EX\_O2\_e" (minimization of oxygen uptake), we computed the minimal oxygen required. Once this value is established as the minimal oxygen uptake required for optimal growth, we reduced the oxygen supplementation (i.e. subsequent increments of "EX\_O2\_e" reaction lower bound from maximum oxygen uptake rate in baseline conditions towards  $0$ ) until reaching a biomass flux of  $0$ , meaning no growth *in silico*. Following this approach, we extracted the minimum required oxygen to be supplemented *in silico* to allow growth. We then performed simulations of iEG158\_mpn with an intermediate oxygen supplementation resulting from the rounded calculation of the mean of the minimal oxygen uptake required for optimal growth and the minimal oxygen required for allowing growth.

## ***In silico* prediction of medium lipids based on membrane composition**

Lipid composition of the medium was predicted according to the scheme in Figure 8. Each candidate medium lipid was integrated into the cytosol and subsequently into membrane compartment of biomass in three possible alternative ways: through direct integration, after being metabolized or after synthesis from precursors in the medium. Each lipid is considered to carry fatty acids that match experimentally determined fatty acids profile. Several proportions distributions of the membrane composition were considered, according to the ones detected in literature (as in our model: see Supplementary Table 1). Determination of required composition of lipids in the growth medium (see previous section) is then performed balancing both obtained fatty acid profile and variable membrane compositions, in order to predict a consensus medium composition able to allow growth under all the membrane configurations detected during laboratory cultivation.



**Figure 8.** *In silico* strategy for predicting lipid composition of growth medium. All the possible lipids and precursors are provided in the medium for uptake. Each possible membrane composition that is found in the literature can be incorporated in the model as part of the biomass synthesis reaction. In our case, we considered three membrane compositions whose lipid proportions change according to the cholesterol percentage. Fatty acid profile can be found in literature (as in our model) or be experimentally determined. Lipid1 is known to be directly incorporated into the membrane, Lipid2 is metabolized in the cytosol and, whether a lipase acts, degraded into component head group and fatty acids. Lipid3 is synthesized in the cytosol from lipid precursors. Each lipid is integrated into the membrane according to the compositions previously reported in the literature. According to the reported membrane composition and fatty acid profile, a consensus minimal lipid composition of the medium is predicted.

## Sequence alignment

By querying manually curated sequences in the UNIPROT database,<sup>70</sup> we verified the presence of lactate and acetic acid transporters and general monocarboxylic acid transporters in the *M. pneumoniae* genome (Dataset 6). PSI-BLAST algorithm<sup>71</sup> was run with default parameters until convergence was reached to build a sequence profile of the transporters (Position Specific Scoring Matrix provided in Dataset 6), then used to query within *M. pneumoniae* genome for presence of these transporters. In addition, we searched for monocarboxylic transporters in *M. pneumoniae* by running a BLASTp search with default parameters on all its transmembrane proteins, whose list was available on the *MyMpn* portal (August 2018),<sup>60</sup> filtering for proteins containing transmembrane helices.

## Addition of predicted phospholipids into the semi-defined serum-free medium MCMycO

*M. pneumoniae* strain M129 was grown in 75cm<sup>2</sup> Tissue Culture Treated Flasks (Falcon) in MCMycO medium, a serum-free medium based on the MC medium previously used for growing *Mesoplasma florum*.<sup>44</sup> Palmitic acid, oleic acid and cholesterol were added at a final concentration of 16.5 µg mL<sup>-1</sup>, 20 µg mL<sup>-1</sup> and 20 µg mL<sup>-1</sup>, respectively. [Supplementary Table 4](#) reports instructions for MCMycO. Cells were grown and passaged three times in absence or presence of sphingomyelin and phosphatidylcholine at 20 µg mL<sup>-1</sup> before performing any analysis. Each passage was performed by diluting the cells 50 times in fresh serum-free medium after two weeks of culture. Medium was replaced by fresh serum-free medium every 7 days.

Passage 3 cells DNA were measured by qPCR using oligonucleotides MPN628/5 (GCCATTTGGATGGTTATGG) and MPN628/3 (GGTGACCCACTTCCGAGTTA) and the Luna® Universal qPCR Master Mix (New England Biolabs). The DNA quantification was performed using the LightCycler® 480 software (Roche). Cells were diluted in 15 mL to an equivalent of 0.6 pg mL<sup>-1</sup> of DNA and distributed in aliquots of 0.2 mL in 96-well Tissue Culture Treated flat plates (Nunc). Every 12h, the medium from 6 wells were recovered in 1.5 mL tubes and centrifuged 15 min at 12 000g. 0.2 mL of nuclease-free water were added to the empty wells to lyse adherent cells by osmotic shock and then recovered and added to the previous pellet to lyse non-adherent cells. Finally, they were incubated 10 min at 98°C and stored at -20°C. After 120h, the DNA of each sample was measured by qPCR as described above and the growth curve was then plotted and analysed using the R software (version 3.4.4).

## Contribution of predicted phospholipids into the defined serum-free medium vB13

*M. pneumoniae* strain M129 was grown in the defined serum component-free medium vB13, developed by Burgos R., Garcia-Ramallo E., Shaw D., Lluch-Senar M. and Serrano L., based on the medium defined by Yus et al.<sup>18</sup> by experimental screening of different components. vB13 composition is provided in [Supplementary Table 5](#). Growth performance of *M. pneumoniae* in

vB13 was compared to rich medium (Hayflick) by growth curve analysis as follows. *M. pneumoniae* strain M129 was inoculated with frozen stocks (1:200 dilution) in tissue culture flasks of 25cm<sup>2</sup> containing 5ml of medium. Cultures were incubated at 37°C under 5% CO<sub>2</sub>, and protein and DNA biomass were measured at different time points (0h, 24h, 48h, 72, 96h) as follows. For each time point, cells were scraped off from the flasks in the culture medium and 1ml of cell suspension harvested by centrifugation (13100g, 10min). This was performed per duplicate to obtain samples for both, protein and DNA measurements. For protein biomass quantification, the cell pellet was washed twice with PBSx1 and lysed with 100µl of lysis buffer (10mM Tris pH8, 6mM MgCl<sub>2</sub>, 1mM EDTA, 100mM NaCl, 0.1%Triton-X100 plus cocktail of protease inhibitors) prior to duplicate protein measurements using the Pierce™ BCA Protein Assay Kit. For DNA biomass quantification, the cell pellet was directly analysed, and the DNA extracted using the MasterPure DNA purification Kit (Epicentre) following the recommendations of the Kit manufacturer. Finally, extracted DNA for each time point was measured using a fluorometric method (Qubit dsDNA HS assay Kit, Invitrogen).

To assess the contribution of phospholipids in the animal component-free medium, *M. pneumoniae* was grown in a 96-well plate format. Growth curve analysis comparing the performance of rich medium (Hayflick), vB13 and vB13 derivative versions containing only phosphatidylcholine (PC), only sphingomyelin (SPM) or vB13 without PC and SPM were recorded in a Tecan Spark plate reader by determining the growth index value, which is the ratio of absorbance at 430 nm and 560 nm of the culture medium.<sup>18</sup> Each media was tested in duplicate wells containing 200µl of medium inoculated with frozen stocks at 1:100 dilutions. Total cell biomass obtained at the end of the growth curve (96h) was also determined by protein quantification as follows. Briefly, wells of the 96-well plate were washed twice with PBSx1 and attached cells lysed with 100µl of lysis buffer (10mM Tris pH8, 6mM MgCl<sub>2</sub>, 1mM EDTA, 100mM NaCl, 0.1%Triton-X100 plus cocktail of protease inhibitors). Protein concentrations were then determined using the Pierce BCA protein assay kit by measuring absorbance at 562 nm in a Tecan plate reader.<sup>18</sup>

## **Data Availability**

The model has been deposited in BioModels and assigned identifier MODEL1912060001.

## **Acknowledgements**

This work has received funding from European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program under grant agreement n. 634942 (MycoSynVac) and n. 670216 (MYCOCHASSIS).

## **Competing Interests**

Provisional patent has been filed for lipid components of serum-free media for Mycoplasma (Patent application n. EP19205723.0). Patent has been filed for vB13 medium (Patent application n. EP20382261).

## **Author Contributions**

EG, VAPmS and MS-D conceived the project. EG and AM developed the model with suggestions from MS-D and LS. EG and AM performed modeling, simulations, calculations and analysis. EG and LG-M developed the medium MCMycO. LG-M performed MCMycO medium experiments. RB performed vB13 medium experiments. EG wrote the chapter with input of MS-D, AM, VAPmS, LS, LG-M and RB. VAPmS and MS-D supervised the project.

## References

1. Razin, S. The mycoplasmas. *Microbiol. Rev.* **42**, 414–470 (1978).
2. Hayflick, L. & Chanock, R.M. Mycoplasma species of man. *Bacteriol. Rev.* **29**, 185–220 (1965).
3. Razin, S., Yogev, D. & Naot, Y. Molecular biology and pathogenicity of mycoplasmas. *Microbiol. Mol. Biol. Rev.* **62**, 1094–156 (1998).
4. Tilman, D., Cassman, K. G., Matson, P. A., Naylor, R. & Polasky, S. Agricultural sustainability and intensive production practices. *Nature* **418**, 671 (2002).
5. Cabello, F. C. Heavy use of prophylactic antibiotics in aquaculture: a growing problem for human and animal health and for the environment. *Environ. Microbiol.* **8**, 1137–1144 (2006).
6. Van Boeckel, T. P. *et al.* Global trends in antimicrobial use in food animals. *Proc. Natl. Acad. Sci.* **112**, 5649–5654 (2015).
7. Nicholas, R. A. J., Ayling, R. D. & Stipkovits, L. P. An experimental vaccine for calf pneumonia caused by *Mycoplasma bovis*: clinical, cultural, serological and pathological findings. *Vaccine* **20**, 3569–3575 (2002).
8. Grayston, J. T. *et al.* Mycoplasma pneumoniae Infections: Clinical and Epidemiologic Studies. *JAMA* **191**, 369–374 (1965).
9. Collier, A. M. & Clyde, W. A. Relationships Between *Mycoplasma pneumoniae* and Human Respiratory Epithelium. *Infect. Immun.* **3**, 694–701 (1971).
10. Miles, R. J., Taylor, R. R. & Varsani, H. Oxygen uptake and H<sub>2</sub>O<sub>2</sub> production by fermentative *Mycoplasma* spp. *J. Med. Microb.* **34**, 219–223 (1991).
11. Wodke, J. A. H. *et al.* Dissecting the energy metabolism in *Mycoplasma pneumoniae* through genome-scale metabolic modeling. *Mol. Syst. Biol.* **9**, (2013).
12. Razin, S. & Jacobs, E. Mycoplasma adhesion. *J. Gen. Microb.* **138**, 407–422 (1992).
13. Waites, K. B. & Talkington, D. F. Mycoplasma pneumoniae and Its Role as a Human Pathogen. *Clin. Microb. Rev.* **17**, 697–728 (2004).
14. Merten, O. W. Safety issues of animal products used in serum-free media. *Dev. Biol. Stand.* **99**, 167–180 (1999).
15. Ahmad, I., Kleven, A. S. H., Avakian, A. P. & Glisson, J. R. Sensitivity and Specificity of *Mycoplasma gallisepticum* Agglutination Antigens Prepared from Medium with Artificial Liposomes Substituting for Serum. *AAAP* **32**, 519–526 (1988).
16. Laidlaw, P.P. & Elford, W.J. A New Group of Filterable Organisms. *Proceedings of the Royal Society of London* **120**, 292–303 (1936).
17. Edward, B. Y. D. G. F. F. & Fitzgerald, W. A. The Isolation of Organisms of the Pleuropneumonia Group from Dogs. *J. Gen. Microb.* **5**, 566–575 (2018).
18. Yus, E. *et al.* Impact of genome reduction on bacterial metabolism and its regulation. *Science*. **326**, 1263–1268 (2009).
19. Price, N. D., Papin, J. A., Schilling, C. H. & Palsson, B. O. Genome-scale microbial in silico models: the constraints-based approach. *Trends in Microb.* **21**, 162–169 (2003).

20. Westerhoff, H. V & Palsson, B. O. The evolution of molecular biology into systems biology. *Nat. Microb.* **22**, 1249–1252 (2004).
21. Edwards, J. S., Covert, M. & Palsson, B. Metabolic modelling of microbes: the flux-balance approach. *Environ. Microbiol.* **4**, 133–140 (2002).
22. Baloni, P. *et al.* Genome-scale metabolic model of the rat liver predicts effects of diet restriction. *Sci. Rep.* **9**, 9807 (2019).
23. Moreira, T. B. *et al.* A Genome-Scale Metabolic Model of Soybean (*Glycine max*) Highlights Metabolic Fluxes in Seedlings. *Plant Physiol.* **180**, 1912–1929 (2019).
24. Jensen, C. S. *et al.* Reconstruction and Validation of a Genome-Scale Metabolic Model of *Streptococcus oralis* (iCJ415), a Human Commensal and Opportunistic Pathogen. *Front. Genet.* **11**, 116 (2020).
25. Gu, C., Kim, G. B., Kim, W. J., Kim, H. U. & Lee, S. Y. Current status and applications of genome-scale metabolic models. *Genome Biol.* **20**, 121 (2019).
26. Pelicaen, R., Gonze, D., Teusink, B., De Vuyst, L. & Weckx, S. Genome-Scale Metabolic Reconstruction of *Acetobacter pasteurianus* 386B, a Candidate Functional Starter Culture for Cocoa Bean Fermentation. *Front. Microbiol.* **10**, 2801 (2019).
27. Kurzepa, H., Flinton, L. & Vandemark, P. J. Growth of Parasitic *Mycoplasma* Without Serum or Serum Fraction. *J. Bacteriol.* **99**, 908–909 (1969).
28. Razin, S. & Tully, J. G. Cholesterol Requirement of *Mycoplasmas*. *J. Bacteriol.* **102**, 306–310 (1970).
29. Rottem, S. & Kahane, I. *Mycoplasma* cell membranes. *Subcellular chemistry* **20**, 1–336 (2012).
30. Leon, O. & Panos, C. Long-chain fatty acid perturbations in *Mycoplasma pneumoniae*. *J. Bacteriol.* **146**, 1124–1134 (1981).
31. Pollack, J. D., Somerson, N. L. & Senterfit, L. B. Effect of pH on the immunogenicity of *Mycoplasma pneumoniae*. *J. Bacteriol.* **97**, 612–619 (1969).
32. Dahl, J. S. & Dahl, C. E. Effect of Cholesterol on Macromolecular Synthesis and Fatty Acid Uptake by *Mycoplasma capricolum*. *J. Biol. Chem.* **256**, 87–91 (1981).
33. Ku, S., Gavin, A., Aebersold, R. & Serrano, L. Quantification of mRNA and protein and integration with protein turnover in a bacterium. *Mol. Sys. Bio.* **7**:511 (2011). doi:10.1038/msb.2011.38
34. Klement, M. L. R., Öjemyr, L., Tagscherer, K. E., Widmalm, G. & Wieslander, Å. A processive lipid glycosyltransferase in the small human pathogen *Mycoplasma pneumoniae*: Involvement in host immune response. *Mol. Microbiol.* **65**, 1444–1457 (2007).
35. Pollack, J. D., Somerson, N. L. & Senterfit, L. B. Chemical Composition and Serology of *Mycoplasma pneumoniae* Lipids. *J. Infect. Dis.* **127**, S32–S35 (1973).
36. Stoll, L. L. & Spector, A. A. Changes in serum influence the fatty acid composition of established cell lines. *In Vitro* **20**, 732–738 (1984).
37. Rottem, S. Membrane lipids of mycoplasmas. *Biochim. Biophys. Acta* **604**, 65–90 (1980).
38. Rottem, S., Adar, L., Gross, Z. V. I., Eman, N. E. & Davis, P. J. Incorporation and Modification of Exogenous Phosphatidylcholines by *Mycoplasmas*. *J. Bacteriol.* **167**, 299–304 (1986).
39. Doux, J. P. F., Killian, J. A., Nyholm, T. K. M., Slotte, J. P. & Lo, M. Sterols Have Higher Affinity for

- Sphingomyelin than for Phosphatidylcholine Bilayers even at Equal Acyl-Chain Order. *Biophys. J.* **100**, 2633–2641 (2011).
40. Worliczek, H. L., Kämpfer, P., Rosengarten, R., Tindall, B. J. & Busse, H.-J. Polar lipid and fatty acid profiles – Re-vitalizing old approaches as a modern tool for the classification of mycoplasmas? *Syst. Appl. Microbiol.* **30**, 355–370 (2007).
  41. Kornspan, J. D. & Rottem, S. The Phospholipid Profile of Mycoplasmas. *J. Lipids* **2012**, 1–8 (2012).
  42. Salman, M. & Rottem, S. The cell membrane of Mycoplasma penetrans: lipid composition and phospholipase A1 activity. *Biochim. Biophys. Acta - Biomembr.* **1235**, 369–377 (1995).
  43. Rottem, S. & Markowitz, O. Membrane Lipids of Mycoplasma gallisepticum: A Disaturated Phosphatidylcholine and a Phosphatidylglycerol with an Unusual Positional Distribution of Fatty Acids. *Biochemistry* **18**, 2930–2935 (1979). doi:10.1021/bi00581a002
  44. McCoy, R. E. *et al.* Acholeplasma florum, a New Species Isolated from Plants. *Int. J. Syst. Evol. Microbiol.* **34**, 11–15 (1984).
  45. Low, I. E. & Eaton, M. D. Replication of Mycoplasma pneumoniae in Broth Culture. *J. Bacteriol.* **89**, 725 – 728 (1965).
  46. Freundt, E. A. Culture media for classic mycoplasmas. *Methods in Mycoplasmaology* **1**, 127-135 (1983).
  47. Segovia, J. A. *et al.* NLRP3 Is a Critical Regulator of Inflammation and Innate Immune Cell Response during Mycoplasma pneumoniae Infection. *Infect. Immun.* **86** (1) (2018). doi:10.1128/IAI.00548-17
  48. Kenny, G. E., Kaiser, G. G., Cooney, M. K. & Foy, H. M. Diagnosis of Mycoplasma pneumoniae pneumonia: sensitivities and specificities of serology with lipid antigen and isolation of the organism on soy peptone medium for identification of infections. *J. Clin. Microbiol.* **28**, 2087–2093 (1990).
  49. Berglund, A. H., Nilsson, R. & Liljenberg, C. Permeability of large unilamellar digalactosyldiacylglycerol vesicles for protons and glucose – influence of  $\alpha$ -tocopherol, -carotene, zeaxanthin and cholesterol. *Plant Physiol. Biochem.* **37**, 179–186 (1999).
  50. Osterberg, F., Rilfors, L., Wieslander, A., Lindblom, G., Gruner, S.M. Lipid extracts from membranes of Acholeplasma laidlawii A grown with different fatty acids have a nearly constant spontaneous curvature. *Biochim. Biophys. Acta* **1257**, 18–24 (1995).
  51. Finkelstein, A. Water and Nonelectrolyte Permeability of Lipid Bilayer Membranes. *J. Gen. Physiol.* **68**, 127–135 (1976).
  52. Razin, S., Kutner, S., Efrati, H. & Rottem, S. Phospholipid and cholesterol uptake by mycoplasma cells and membranes. *Biochim. Biophys. Acta - Biomembr.* **598**, 628–640 (1980).
  53. Zeiman, E., Tarshis, M. & Rottem, S. Mycoplasma penetrans under nutritional stress: influence on lipid and lipoprotein profiles and on the binding to and invasion of HeLa cells. *FEMS Microb. Lett.* **287**, 243–249 (2008). doi:10.1111/j.1574-6968.2008.01322.x
  54. Hackett, K. J., Ginsberg, A. S., Rottem, S., Henegar, R. B. & Whitcomb, R. F. A defined medium for a fastidious Spiroplasma. *Science*. **237**, 525–527 (1987).
  55. Cluss, R. G., Johnson, J. K. & Somerson, N. L. Liposomes replace serum for cultivation of fermenting mycoplasmas. *Appl. Environ. Microbiol.* **46**, 370-374 (1983).

56. Schummer, U., Schiefer, H.G., Gerhardt, U. The proton gradient across Mycoplasma membranes. *Current Microbiol.* **5**, 371–374 (1981).
57. Bergman, C., Kashiwaya, Y. & Veech, R. L. The Effect of pH and Free Mg<sup>2+</sup> on ATP Linked Enzymes and the Calculation of Gibbs Free Energy of ATP Hydrolysis. *J. Phys. Chem.* **114**, 16137–16146 (2010).
58. King, Z. A. *et al.* BiGG Models: A platform for integrating, standardizing and sharing genome-scale models. *Nucleic Acids Res.* **44**, D515–D522 (2015).
59. Flamholz, A., Noor, E., Bar-even, A. & Milo, R. eQuilibrator — the biochemical thermodynamics calculator. *Nucl. Ac. Res.* **40**, 770–775 (2012).
60. Cozzuto, L. *et al.* MyMpn: a database for the systems biology model organism Mycoplasma pneumoniae. *Nucl. Ac. Res.* **43**, 618–623 (2015).
61. Hucka, M. *et al.* The systems biology markup language ( SBML ): a medium for representation and exchange of biochemical network models. *Bioinformatics* **19**, 524–531 (2003).
62. Lieven, C. *et al.* MEMOTE for standardized genome-scale metabolic model testing. *Nat. Biotechnol.* **38**, 272–276 (2020).
63. Le Novère, N. *et al.* BioModels Database: a free, centralized database of curated, published, quantitative kinetic models of biochemical and cellular systems. *Nucleic Acids Res.* **34**, D689–D691 (2006).
64. Li, C. *et al.* BioModels Database: An enhanced, curated and annotated resource for published quantitative kinetic models. *BMC Syst. Biol.* **4**, 92 (2010).
65. Chelliah, V. *et al.* BioModels: ten-year anniversary. *Nucleic Acids Res.* **43**, D542–D548 (2014).
66. King, Z. A. *et al.* Escher: A Web Application for Building , Sharing , and Embedding Data-Rich Visualizations of Biological Pathways. *PLOS Comp. Biol.* 1–13 (2015). doi:10.1371/journal.pcbi.1004321
67. Orth, J. D., Thiele, I. & Palsson, B. Ø. What is flux balance analysis? *Nat. Publ. Gr.* **28**, 245–248 (2010).
68. Gudmundsson, S. & Thiele, I. Computationally efficient flux variability analysis. *BMC Bioinformatics* **11**, 489 (2010).
69. Ebrahim, A., Lerman, J. A., Palsson, B. O. & Hyduke, D. R. COBRApy: COConstraints-Based Reconstruction and Analysis for Python. *BMC Syst. Biol.* **7**, 74 (2013).
70. Magrane, M. & UniProt Consortium. UniProt Knowledgebase: a hub of integrated protein data. *Database* **2011**, 1–13 (2011).
71. Altschul, S. F. *et al.* Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucl. Ac. Res.* **25**, 3389–3402 (1997).

## Appendices

All additional files can be found at the on-line version of Gaspari et al. "Model-driven design allows growth of *Mycoplasma pneumoniae* on serum-free media". In *NPJ Systems Biology and Applications* 6:33 (2020).

About this manuscript:

"...

**Erika Gaspari** told *ATLA*: "Wageningen University & Research develops a wide range of alternative research methods and models on animal testing. We follow the 3R Rule: try to replace animal testing altogether. If that is not possible, reduce the number of tests that require animals and refine those tests to prevent discomfort as much as possible. In my research, I have focused on the computational design of a pneumonia vaccine for animals, and on the development of a medium to grow the vaccine strain that was animal-serum free. That is important because the blood collection process to obtain animal serum has raised important ethical questions regarding the suffering of the animal caused by the harvesting of the serum. Moreover, a serum-free growth medium enables the large-scale production of an animal vaccine, which represents an alternative to the administration of antibiotics to the animals and, therefore, avoids the rise of antibiotic-resistant bacteria. Finally, my research is an example of drug design performed through computer modelling approaches, whose evolution is opening the way to a future free of animal testing."

..."

Rita Seabra on **Gaspari et al., 2020**.

In: "News and Views", journal *ATLA* (Alternatives to Laboratory Animals) 1-4 (2020)

doi: 10.1177/0261192920979998



# Chapter 3

Design of a fatty acid-prototrophic  
*Mycoplasma pneumoniae* strain  
through metabolic modelling

**Erika Gaspari**, Luis Garcia-Morales, Antoni Malachowski, Ruben G. Van Heck, Maria Lluch-Senar, Raul Burgos, Enrico Orsi, Ruud Weusthuis, Carole Lartigue, Laure Beven, Alain Blanchard, Vitor A.P. Martins dos Santos and Maria Suarez-Diez



## Abstract

*Mycoplasma pneumoniae* is a slow-growing, human parasite causing atypical pneumonia. The lack of cell wall makes many common antibiotics ineffective, and a vaccine is required for prevention of infection. Due to its reduced genome, *Mycoplasma pneumoniae* lacks many biosynthetic pathways, among which, fatty acid biosynthesis. This renders it fastidious to grow, making large-scale vaccine production challenging and expensive.

To overcome this hurdle, we have inserted in *M. pneumoniae* 13 genes from *Acholeplasma laidlawii* coding for fatty acid biosynthesis and tested performance subsequently. We studied the feasibility of the pathway and assessed the energetic bottlenecks with the genome-scale model iEG158\_mpn of *M. pneumoniae*'s metabolism. *In silico* simulations pinpointed the NADPH levels are insufficient for efficient fatty acid production, and suggested a knock-in for NADPH regeneration. The selected strategy would allow a fatty acid-prototrophic strain of *M. pneumoniae* to grow at about the same rate of the wild type strain, in serum-free medium.

Altogether, this synthetic biology study combines the manipulation of bacterial DNA with metabolic modelling, allowing the construction of a growing *M. pneumoniae* strain able to synthesize fatty acids.

## Introduction

*Mycoplasma pneumoniae* is a very small parasite of the class Mollicutes, member of the family Mycoplasmataceae and order Mycoplasmatales. It is characterized by small size (0.1-0.2  $\mu\text{m}$  width, 1-2  $\mu\text{m}$  length) and genome (816394 base pairs), lack of cell wall and periplasmic space, low GC content and unusual nutrition needs, with consequent limited metabolic activity<sup>1-3</sup>. A main consequence of the reduced genome of *M. pneumoniae*, consisting of 687 genes of which 56.6% coding for essential metabolic enzymes, is the loss of function and the lack of a number of pathways: TCA cycle, respiratory electron transport chain, biosynthesis of amino acids, fatty acids and a consistent part of the lipids integrated in the membrane<sup>4,5</sup>. The difficulty in eradicating *M. pneumoniae* infections lies on the ability of this bacterium to persist within the host through membrane-mimicking, as well as on the lack of cell wall, which renders multiple antibiotics such as  $\beta$ -lactams, glycopeptides, sulphonamides, trimethoprim, polymyxins, nalidixic acid and rifampin ineffective<sup>6</sup>. Furthermore, novel therapeutics that target (inhibit) the fatty acid synthase of both Gram-positive and Gram-negative bacteria fail with *M. pneumoniae*, which lacks the multienzyme complex Fatty\_Acid\_Synthase and its related genes<sup>7,8</sup>. Yet fatty acids are indispensable for building the cell membrane, so it uses the membrane-bound acyltransferases FAKA, FakB, ACP, PlsX (Mpn546), PlsY (Mpn350) and PlsC (Mpn299) genes proteins to introduce fatty acids into the phospholipid biosynthetic pathway. To this end, it relies on exogenous host fatty acids for membrane lipid synthesis<sup>9</sup>. The purpose of the work herein described is therefore to re-design *M. pneumoniae* so to render it fatty acid-prototrophic and thus able to synthesise fatty acids and grow in minimal medium at rates comparable to those of the wild type in serum-based medium. To this end, we combine computational-based approaches and targeted genome engineering.

3

### Re-designing lipid metabolism in *Mycoplasma pneumoniae*

Lacking cell wall, mycoplasma is highly dependent on its cell membrane, which is reinforced with cholesterol and phospholipids taken from the host<sup>2</sup>. Indeed, one of the key features of *M. pneumoniae*'s membrane is the high proportion of cholesterol, directly incorporated, which can reach up to 50% of the total lipid fraction<sup>10</sup>. Sphingolipids and phosphatidylcholine constitute an average of 12% and 5%, respectively, of the membrane lipids<sup>11</sup>. Although *M. pneumoniae* is a fatty acid and sterol auxotroph, it can assemble phosphatidic acid, diacylglycerol, phosphatidylglycerol and glycolipids from imported fatty acids<sup>12</sup>. While glycolipids and phosphatidylglycerol are found in the membrane, phosphatidic acid and diacylglycerol are biosynthesis intermediaries<sup>13</sup>. Lipids' acyl chains depend on availability of fatty acids in the host. *M. pneumoniae* tends to selectively incorporate palmitic acid over stearic acid in the membrane<sup>14</sup>, with a generic preference for saturated fatty acids over unsaturated ones such as oleic and linoleic acid. For these reasons, the insertion of a pathway for biosynthesis of saturated fatty acid into *M. pneumoniae*'s genome is desirable, as it could improve the flexibility of the bacterium in integrating acyl chains in the membrane's lipids, as well as its growth rate and associated costs for large-scale production. Optimally, the pathway

to knock-in should be well characterized and retrieved from a phylogenetically-close relative of *M. pneumoniae*.

The machinery for biosynthesis of 12-to-18 carbon fatty acids from precursors of 5 carbons or fewer of *Acholeplasma laidlawii* has been extensively studied<sup>15-18</sup> and consists in the concatenation of acetyl chains on short chain-primers operated by an acyl carrier protein (ACP) at the expensive of ATP and NAD(P)H; thirteen genes are involved in the synthesis. The phylogenetic closeness between *M. pneumoniae* and *A. laidlawii* (previously called *Mycoplasma laidlawii*) makes this second particularly suitable for synthetic integration into *M. pneumoniae*'s genome. While not equipped for fatty acid biosynthesis, *M. pneumoniae* is in possession of acyl carrier proteins (ACPs); however, sterols-requiring species have inefficient ACPs<sup>19</sup>. Hence, we set out to insert this pathway from *A. laidlawii* into *M. pneumoniae*'s genome.

To assess the impact of inserting this pathway into *M. pneumoniae* and identify possible further bottlenecks, we deployed computational analyses of the whole-cell metabolism. The recently constructed genome-scale metabolic model iEG158\_mpn<sup>4</sup> of *M. pneumoniae* integrates the membrane formation with the cytosolic reactions, being particularly suitable for *in silico* simulations of the feasibility of the fatty acid biosynthesis knock-in and to elucidate the pathway's energetics, comprising the usage of cofactors. Performance of the pathway was extensively simulated to identify possible bottlenecks. Moreover, predictions on how to improve the growth of the *M. pneumoniae* strain carrying the fatty acid biosynthesis pathway have been conducted with gap filling algorithms. Originally, these algorithms were designed to be used in a metabolic model construction phase, to fill gaps in the model and, therefore, add reactions for which no genome evidence was available<sup>20-24</sup>. To overcome this limitation, we used a gap filling algorithm as a strain-design tool, able to predict knock-ins that would improve the growth rate of *M. pneumoniae*.

## Materials and Methods

### Bacterial strains and growth conditions

*M. pneumoniae* strain M129 and derivative strains were grown at 37°C in 75cm<sup>2</sup> Tissue Culture Treated Flasks (Falcon) in SP4 or the defined serum-free medium vB10 (Supplementary Table S1). Palmitic acid, oleic acid and cholesterol were added to serum-free media at a final concentration of 16.5 µg mL<sup>-1</sup>, 20 µg mL<sup>-1</sup> and 20 µg mL<sup>-1</sup> respectively. Cells were grown in 15 mL SP4 after thawing and passaged three times in vB10 in absence or presence of sphingomyelin and phosphatidylcholine at 20 µg mL<sup>-1</sup> before performing any analysis. Each passage was performed by scrapping the cells off the flask and diluting them 50 times in fresh media until the color shift to orange/yellow.

### Construction of a AL *M. pneumoniae* strain

The first mini-transposon, named L, containing 9 genes *accABCD*, *fabH1*, *fabD*, *fabG1*, *fabB*, *fabZ* and the tetracycline resistance gene (*tetM*) under the *MG\_438* promoter of *Mycoplasma genitalium* (Pich et al. 2006) and the second transposon, named A (4kb), containing genes *acpP*, *acpT*, *fabK* and *acpS* and cholaramphenicol resistance gene, *cat*, under the promoter of the *MG\_438* gene (Calisto et al. 2012), were sequentially transformed into *M. pneumoniae* M129 (Table 1). 7 colonies were successfully isolated containing both transposons.

3

**Table 1.** Thirteen genes from *A. laidlawii* inserted into *M. pneumoniae* M129 genome for reconstituting the fatty acid biosynthesis pathway. Genes are reported with the related reaction's EC number and encoded enzyme name.

Gene	EC number	Enzyme name
<i>accA</i>	6.4.1.2	acetyl-CoA carboxylase, alpha subunit
<i>accB</i>	6.4.1.2	acetyl-CoA carboxylase, biotin carboxyl carrier protein
<i>accC</i>	6.4.1.2	acetyl-CoA carboxylase, biotin carboxylase
<i>accD</i>	6.4.1.2	acetyl-CoA carboxylase, beta subunit
<i>fabD</i>	2.3.1.39	malonyl CoA-acyl carrier protein transacylase
<i>fabH1</i>	2.3.1.180	beta-ketoacyl-(acyl carrier protein) synthase III
<i>fabG1</i>	1.1.1.100	beta-ketoacyl-(acyl carrier protein) reductase
<i>fabZ</i>	4.2.1.59	beta-hydroxyacyl-(acyl-carrier-protein) dehydratase
<i>fabK</i>	1.3.1.9	enoyl-[acyl-carrier-protein] reductase
<i>fabB</i>	2.3.1.41	beta-ketoacyl-(acyl carrier protein) synthetase
<i>acpP</i>		acyl carrier protein
<i>acpT</i>	2.7.8.7	4'-phosphopantetheinyl transferase
<i>acpS</i>	2.7.8.7	Holo-[acyl-carrier-protein] synthase

### Growth conditions and quantification

DNA from passage 3 MPN cells grown in vB10 medium ([Supplementary Table S1](#)) was measured by qPCR using oligonucleotides MPN628/5 (GCCATTTGGATGTTATGG) and MPN628/3 (GGTGACCCACTCCGAGTTA) and the Luna® Universal qPCR Master Mix (New England Biolabs). The quantification was performed using the LightCycler® 480 software (Roche). Cells were diluted in 15 mL to an equivalent of 0.6 pg mL<sup>-1</sup> of DNA and distributed in aliquots of 0.2 mL in 96-well Tissue Culture Treated flat plates (Nunc). Every 12h, the medium from 6 wells was recovered in 1.5 mL tubes and centrifuged 15 min at 12 000g. 0.2 mL of nuclease free water were added to the empty wells to lyse adherent cells by osmotic shock and then recovered and added to the previous pellet to lyse non-adherent cells. Finally, they were incubated 10 min at 98°C and stored at -20°C. After 120h, the DNA of each sample was measured by qPCR and the growth curve was then plotted and analyzed using the R software (version 3.4.4).

### Fatty acid detection

Cells passaged three times in 15 mL of vB10 were scrapped off the flask and centrifuged 15 min at 7500 g. The cell pellet was sent to Bordeaux Metabolome Facility for fatty acid extraction and GC-MS and LC-MS analysis. Given the apparition of non-canonical profile in the GC-MS, LC-MS was used to quantify the percentage of each fatty acid.

### Simulations with genome-scale model iEG158\_mpn

Simulations of the feasibility of the fatty acid biosynthesis pathway inserted into *M. pneumoniae* M129 and of the selected NADPH regeneration system have been conducted on the genome-scale metabolic model iEG158\_mpn<sup>4</sup> using the tool *flux\_balance\_analysis* of package *cobrapy* v. 0.5.11<sup>25</sup>, Python v. 3.4.4 and gurobi solver v. 9.0<sup>26</sup>. Simulations of the predicted knock-ins to integrate were performed with the script "Simulate\_growth.py" ([Supplementary Material](#)), which supplements *in silico* the serum-free medium MCMycO predicted by Gaspari et al. maximizing the growth rate (i.e. the flux through the biomass synthesis reaction) through Flux Balance Analysis (FBA). Although FBA shows one possible flux distribution while maximizing the objective function, the Flux Variability Analysis (FVA) performed by Gaspari et al.<sup>4</sup> on iEG158\_mpn shows the reduced genome of *M. pneumoniae* is reflected in its reduced flux distribution space. This medium is based on MC composition<sup>27</sup>, previously used to grow *A. laidlawii*, and it has been modified with the addition of the lipids required by *M. pneumoniae* to grow (cholesterol, phosphatidylcholine, sphingomyelin).

### Knock-in predictions

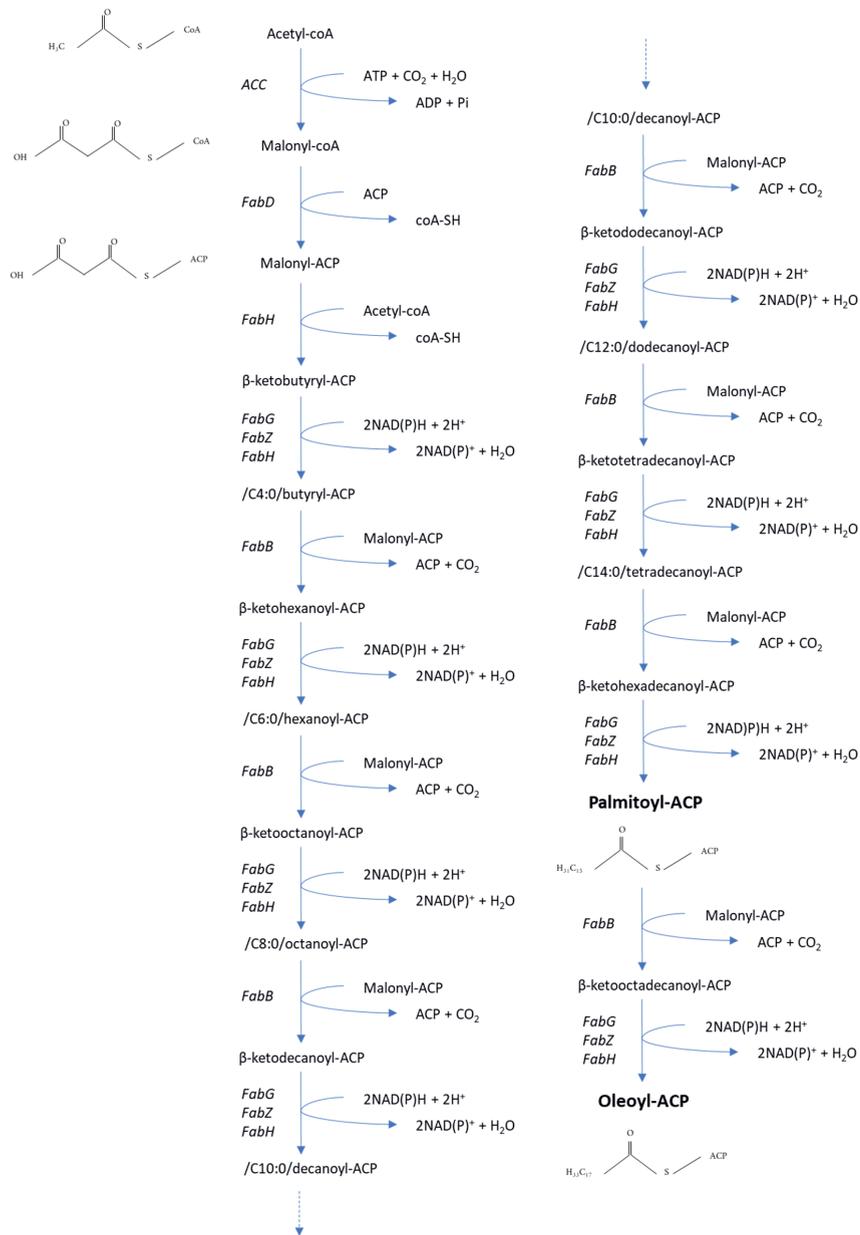
The prediction of knock-ins was conducted through a gap filling algorithm, based on the GapFilling algorithm developed by Reed et al. to refine genome annotation<sup>28</sup>, applied on model iEG158\_mpn, using combination of reactions from the BIGG database (REFBiGG,

provided in XML format in [Supplementary Material](#)), found in bacterial models different from *M. pneumoniae*. REFBiGG consists of 2667 reactions, of which 2182 gene-associated, combined from BiGG models of *E. coli* strain K-12 (e\_coli\_core and iJR904) and strain W (iECW\_1372), *G. metallireducens* (iAF987), *C. ljungdahlii* (iHN637), *Synecosystis* (iJN678), *P. putida* (iJN746), *T. maritima* (iLJ478), *M. tuberculosis* (iNJ661), *S. boydii* (iSBO\_1134), *K. pneumoniae* (iYL1228). The algorithm uses two libraries (Lib\_Phase1.py and Lib\_General.py, in [Supplementary Material](#)) to insert *in silico* combination of one or two reactions into iEG158\_mpn over 50 iterations and simulate the knock-in effect on the growth of *M. pneumoniae* through the script "GapFilling.py" ([Supplementary Material](#)).

## Results

The saturated fatty acids biosynthesis pathway is introduced into *M. pneumoniae* M129 strain by expression of thirteen genes from *A. laidlawii*, which encode for enzymes catalysing reactions that, from starter-molecules of acetyl-coA, lead to the incorporation of acetyl chains from malonyl-coA into acyl carrier proteins and finally into membrane's phospholipids (Figure 1). The pathway consumes 8 acetyl-coA, 7 ATP and 14 NAD(P)H molecules for each palmitoyl chain formed and concatenated to an acyl carrier protein (ACP) and 9 acetyl-coA, 8 ATP and 16 NAD(P)H molecules for each oleoyl-ACP. It is unknown whether the pathway uses NADH or NADPH as cofactors.

The *M. pneumoniae* strain implementing fatty acid biosynthesis was called AL7 and the quantity and entity of the fatty acid produced was measured through GC-MS and LC-MS analysis. Figure 2 shows the comparison of C14, C16 and C18 acyl chains produced by the wild-type strain M129 and the mutant strain AL7 grown on serum-free medium, in presence and absence of phosphatidylcholine and sphingomyelin. Strain AL7 outperforms the growth and the quantity of produced fatty acids of M129 when sphingomyelin and phosphatidylcholine are absent. Moreover, it leads to a much higher proportion of synthesized saturated fatty acids respect to the wild-type strain. While for strain M129 the proportion of saturated and unsaturated fatty acids does not change when sphingomyelin and phosphatidylcholine are added, fractions are consistently different for strain AL7.



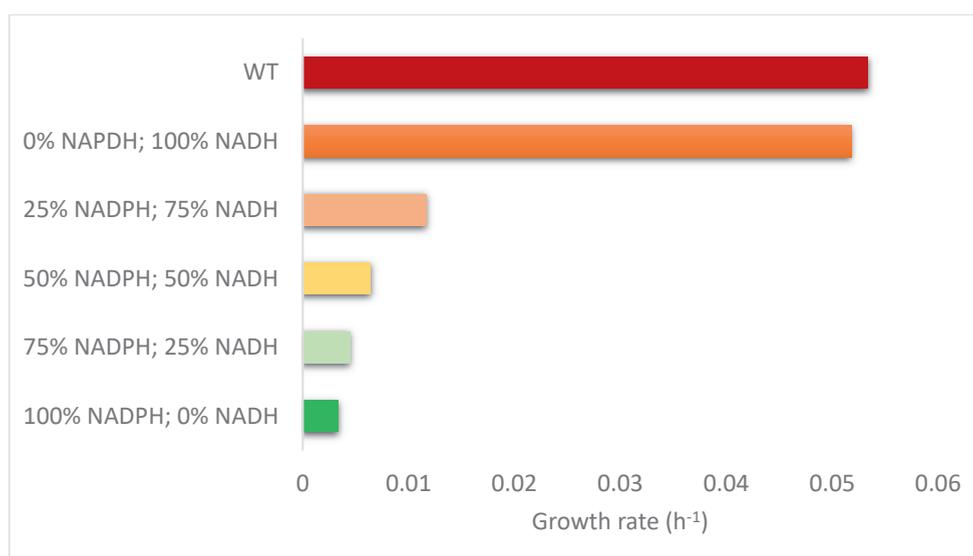
**Figure 1.** Fatty acid biosynthesis pathway from *A. laidlawii* inserted into *Mycoplasma pneumoniae* M129's genome to develop strain AL7. The biosynthesis of palmitoyl chains involves 8 molecules of acetyl-coA (of which 7 are converted to malonyl-coA at the expenditure of 7 ATPs), while the biosynthesis of oleoyl chains consumes 9 molecules of acetyl-coA. The acetyl chains from malonyl-coA are taken by an acyl carrier protein, which will combine subsequent acetyl chains until a length of 16- or 18-carbons is reached. Each time a 2-carbons-chain is concatenated, 2 NAD(P)H are consumed.



**Figure 2.** Fatty acids in the membrane of *M. pneumoniae* strains M129 and AL7, carrying fatty acid biosynthesis pathway. A) Fatty acid quantities (in µg) produced by strains M129 and AL7 grown on serum-free medium, compared to the same growth adding sphingomyelin and phosphatidylcholine (+SP). B) Percentages of saturated (orange) and unsaturated (blue) fatty acids in the membrane of strains M129 and AL7, in absence or presence (+SP) of sphingomyelin and phosphatidylcholine.

The pathway has been inserted *in silico* into model iEG158\_mpn of *M. pneumoniae* M129 metabolism<sup>4</sup> to simulate its feasibility (*in silico* reactions in [Supplementary Table S2](#)). Assuming a reference growth *in silico* of  $0.053\text{h}^{-1}$ , simulated with model iEG158\_mpn grown on the serum-free medium MCMycO developed by Gaspari et al., we simulated the pathway insertion with either NADH or NADPH consumption by the enzymes FabG, FabZ and FabK to elucidate cofactor usage. While the growth of the NADH-dependent fatty acid-prototroph strain results

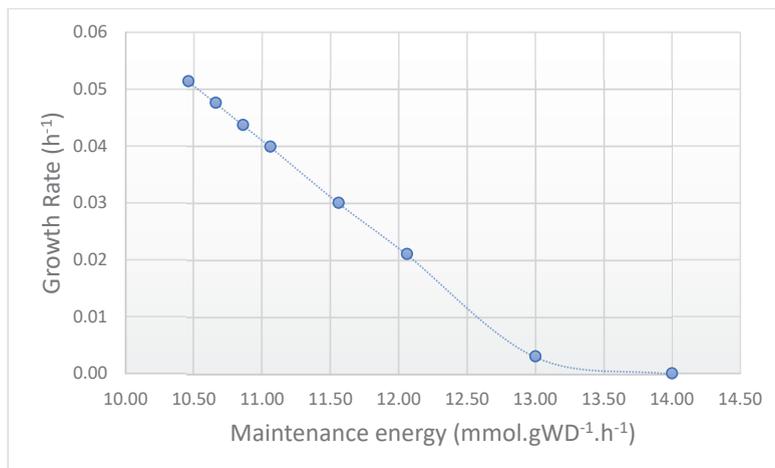
to be 94% of the reference growth, the one with a NADPH-dependent fatty acid biosynthesis pathway results in only 6% of the reference growth ( $0.0033\text{h}^{-1}$ ). A promiscuous cofactor usage, assumed to be 50% NADH and 50% NADPH, would lead to 12% of the reference growth. **Figure 3** shows the effect of the different cofactor usage when simulating the growth of *M. pneumoniae* AL7 with model iEG158\_mpn, compared to the growth of the WT strain M129. This *in silico* results obtained assuming NADPH consumption are in line with the experimental growth on serum-free medium (**Supplementary Table S3**), that shows an AL growth rate of  $0.0012\text{h}^{-1}$ . Therefore, we assumed that the pathway consumes NADPH, and that the difference in the growth rate between the *in vitro* and the *in silico* experiments might be due to metabolic load, enzyme inefficiencies or other factors that cannot be calculated with a genome-scale metabolic model.



**Figure 3.** *In silico* analysis of the effect on growth of different usage of cofactors of the fatty acid synthesis pathway of *M. pneumoniae* strain AL7, compared to the growth rate of the WT strain M129. When 100% of NADH is used in the pathway, the growth would be similar to the WT one, reaching a rate of  $0.0518\text{h}^{-1}$ . If the pathway uses 75% of NADH and 25% of NADPH, the simulated growth rate falls drastically to  $0.0117\text{h}^{-1}$ . If we further increase the percentage of NADPH used in the pathway, simulations lead to growth rates of  $0.0064\text{h}^{-1}$  and  $0.0045\text{h}^{-1}$  when 50% or 25% of NADH are used, respectively. Finally, if we run the simulation assuming the pathway only uses NADPH, the growth rate results to be  $0.0034\text{h}^{-1}$ .

To investigate whether a possible cause of slow growth of strain AL7 was ATP-related, we have run FBA on iEG158\_mpn including the fatty acid biosynthesis pathway for different maintenance energy costs: indeed, the non-growth associated ATP was estimated in Wodke et al. <sup>5</sup> on the wild type strain of *M. pneumoniae*. However, a modified strain might have a different energy consumption. **Figure 4** shows how different maintenance energy costs affect

*in silico* the growth rate of *M. pneumoniae* strain AL7, which decreases proportionally to the ATP used in non-growth associated tasks.



**Figure 4.** Growth rate of *M. pneumoniae* strain AL7 simulated with model iEG158\_mpn integrating the fatty acid biosynthesis pathway for different ATP maintenance costs (non-growth associated).

Moreover, we have looked for strategies to reach a growth rate similar to the one obtained on the supplemented medium, running a modified gap filling algorithm ([Supplementary Material](#)) on model iEG158\_mpn, inserting *in silico* combinations of bacterial reactions from BiGG database into *M. pneumoniae*'s metabolism. The concept is to use an algorithm originally developed to reconstruct genome-scale metabolic models, for strain design: in this way, the result of the computational simulations will lead to genes that when introduced in the genome (knock-ins) increase the growth rate, instead of pathways that make the metabolic model feasible. The results of the simulations suggested slow growth might be due to inefficient NADPH regeneration, providing as output a knock-in of a NADPH-regenerating reaction ([Figure 5](#)), specifically a NADPH transhydrogenase. However, this suggestion is experimentally unfeasible due to the difficulties related to the knock-in of a membrane-bound enzyme in *M. pneumoniae*, so the option of a more efficient NADPH regeneration system was simulated *in silico* by adding a NADPH source in the cytosolic compartment of model iEG158\_mpn. Maximizing the flux through the biomass reaction, a growth rate of 0.051 h<sup>-1</sup> was obtained, corresponding to the 96% of the wild type strain's growth rate in the supplemented medium.

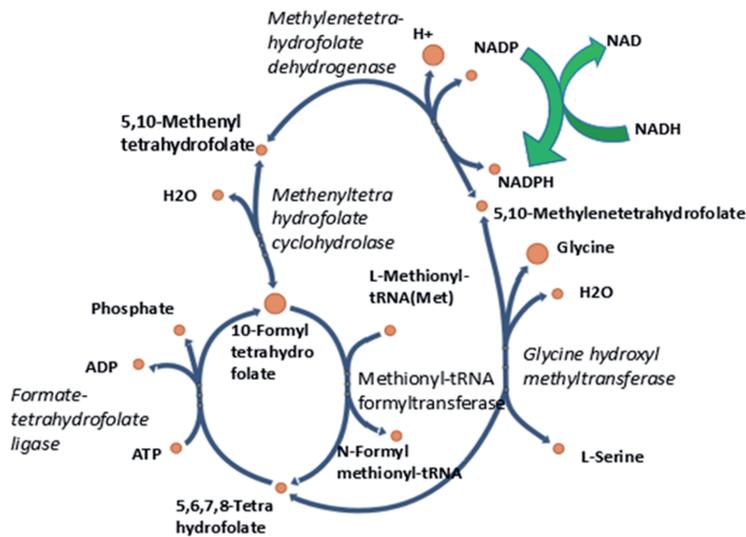
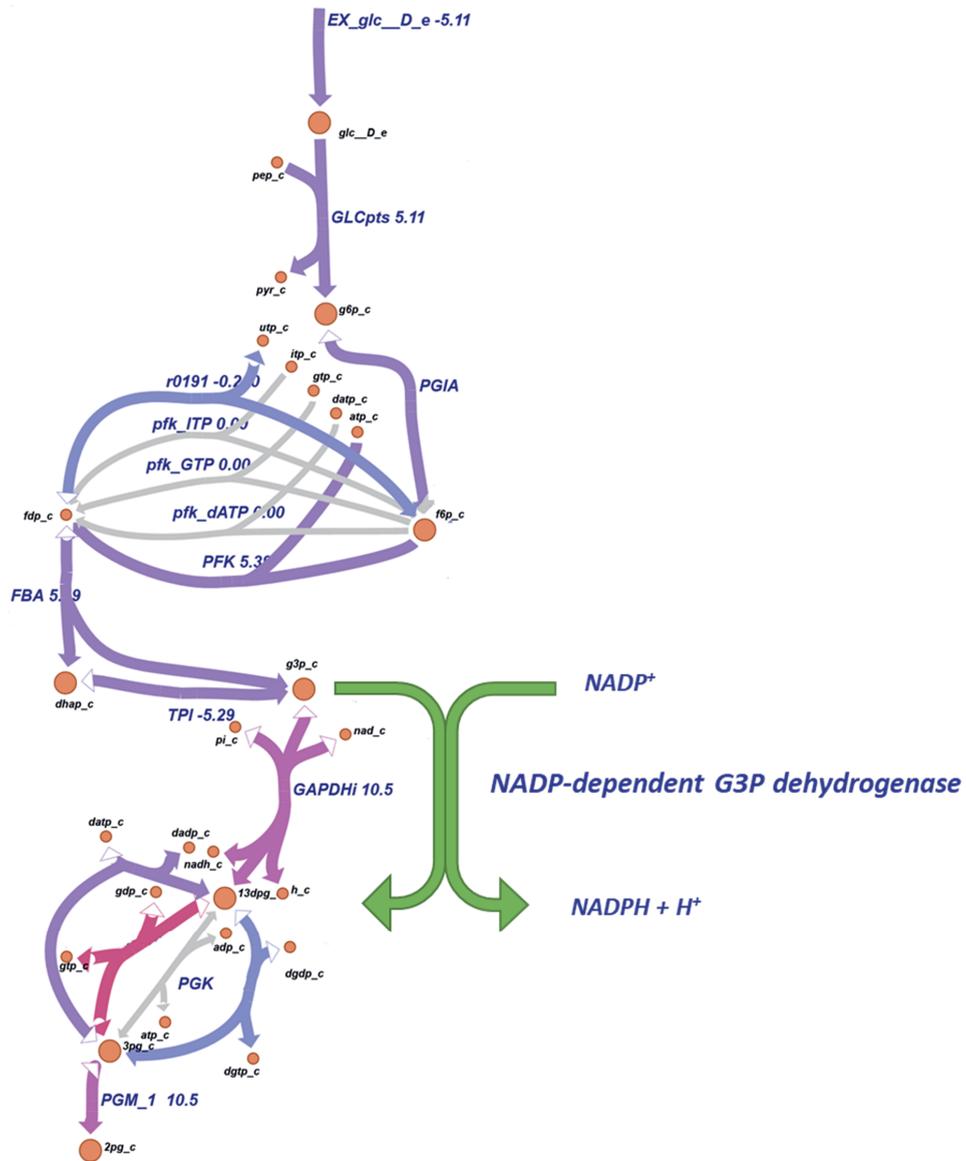


Figure 5. *M. pneumoniae*'s native pathway for NADPH regeneration (blue) compared to the predicted knock-in for more efficient NADPH regeneration (green).

An efficient NADPH regeneration system has been applied in literature on both *E. coli*<sup>29,30</sup> and yeast<sup>31</sup>. The strategy consists in introducing a NADP-dependent glyceraldehyde-3-phosphate dehydrogenase (EC number 1.2.1.13) from *Clostridium acetobutylicum*, converting D-glyceraldehyde 3-phosphate into 3-phospho-D-glyceroyl phosphate reducing a molecule of NADP<sup>+</sup> consuming NADH. It has been shown this system has its maximum efficiency when the native dehydrogenase, regenerating NADH, is replaced by the NADP-dependent one. However, in *M. pneumoniae*'s metabolism, the native glyceraldehyde-3-phosphate dehydrogenase (GAPDH<sub>i</sub> in model iEG158\_mpn, coded by gene *mpn430*) forms one of the main metabolic bottlenecks, being the main NADH regeneration system of the bacterium and carrying a metabolic flux of 10.5 mmol.g<sub>DW</sub><sup>-1</sup>.h<sup>-1</sup> (Figure 6). Therefore, we have simulated *in silico* the replacement of the native GAPDH<sub>i</sub> with the NADP-dependent one and, as an alternative, the knock-in of the second one without replacement. While the growth with the replacement results to be null according to iEG158\_mpn simulations, the addition would lead to 94-96% of the reference growth, i.e. the growth rate of the wild-type strain. Moreover, simulations on iEG158\_mpn (Supplementary Material) suggest the addition of a NADP-dependent dehydrogenase would reduce the NADH regeneration performed by GAPDH<sub>i</sub> by only 1-2%, being the native dehydrogenase favored by the organism, but allowing a sufficient NADPH regeneration to provide enough energy to the fatty acid biosynthesis pathway.



**Figure 6.** Visualization of the initial reactions of the native glycolysis pathway of *M. pneumoniae*, with addition of selected NADPH regeneration system (green). While the native glyceraldehyde-3-phosphate dehydrogenase (GAPDH) converts glyceraldehyde-3-phosphate (g3p\_c) into 3-phospho-D-glyceroyl phosphate (13dpg\_c) regenerating NADH, the in silico knocked-in NADP-dependent glyceraldehyde-3-phosphate dehydrogenase would perform the same conversion regenerating NADPH. Colors are related to flux intensity: the higher is the flux carried by the reaction, the more red-like is the color of the arrow. When the color tends to blue, carried flux is low and can be approximated to null. Numbers indicate the entity of the flux.

## Discussion

We have developed a new *Mycoplasma pneumoniae* strain - called AL7- able to synthesize fatty acids, by inserting thirteen genes from the Mollicute *Acholeplasma laidlawii*. The choice of the synthetic machinery of this bacterium, previously called *Mycoplasma laidlawii*, is due to compatibility of the organism with mycoplasmas, due to their similarity. Indeed, *A. laidlawii* lacks a cell wall; however, unlike mycoplasmas, its cell membrane is mainly made of glycolipids and *Acholeplasma*-specific lipoglycans, while sterols are not present, as this bacterium is able to efficiently synthesize fatty acid precursors<sup>32</sup>. While the cofactor used by *A. laidlawii* for the biosynthesis of fatty acids is not reported, we have used metabolic modelling to investigate cofactor usage in *Mycoplasma* when the machinery for fatty acid biosynthesis is inserted in its genome, concluding that likely consists of NADPH consumption. This was confirmed by the results of *in silico* simulations introducing an artificial NADPH-regeneration route. A gap-filling algorithm was used to investigate the causes of inefficiencies of the pathway and to search for viable alternatives, by knocking-in *in silico* combinations of reactions from other bacteria. Being the NADPH pool limited in *M. pneumoniae*, we have proposed the introduction of a machinery for efficient NADPH regeneration that would leave intact the NADH production, a main metabolic bottleneck of *M. pneumoniae*.

Other causes of growth-related inefficiencies of the fatty acid biosynthesis pathway could be the synthesis and accumulation of toxic compounds and the insufficient ATP level. Gaspari et al. established the ATP level in *M. pneumoniae* metabolism is not a factor limiting growth<sup>4</sup>. However, model iEG158\_mpn does not take into account the cost of expressing the enzymes, which is inversely proportional to the cell volume<sup>33</sup>, and therefore particularly penalizing in mycoplasmas. Our results show that an increase in ATP costs related to non-growth associated tasks might greatly affect the growth rate of *M. pneumoniae*. Inefficient enzyme expression could also affect the functioning of the pathway, as well as the excessive metabolic load caused by the expression of numerous enzymes, a well-known problem when dealing with synthetic biology applications<sup>34-36</sup>. Especially small-size bacteria like mycoplasmas seem to suffer unnatural extra load: cell alterations by heterologous protein production is compensated by an increase in cell size, which cause a decrease in DNA and RNA concentrations<sup>37</sup>.

While most lipids are directly incorporated into the membrane, *M. pneumoniae* might use host phospholipids as a source of fatty acids, as reflected by the presence of lipase on its membrane, as observed in other mycoplasmas<sup>4</sup>. This would allow the rearrangement of acyl chains among the lipids and an efficient mimicking of the host's cell membrane, as well as an adaptation to the environmental conditions by adjustment of thickness and viscosity, which are directly dependent on the fatty acids distribution incorporated into membrane's lipids<sup>4,5,38</sup>. *M. pneumoniae* has indeed an average acyl chain distribution in the membrane that was investigated in literature<sup>4,5,38</sup>. Despite the composition flexibly changes according to the host's availability and environmental conditions, the addition of sphingomyelin and phosphatidylcholine to the serum-free medium used to grow strain AL7 leads to a rearrangements of the saturations in the membrane, suggesting *M. pneumoniae* possess of the machinery for acyl chains transfer. This would imply that a fatty acid source in the medium

might not be needed for growing *M. pneumoniae*, which might be able to use all lipids or only phospholipids as a fatty acid source.

## Conclusion

In this work, we have designed and built a strain of *Mycoplasma pneumoniae* that is fatty acid-prototrophic. We did so by expressing thirteen genes from *Acholeplasma laidlawii* that code for enzymes that form the fatty acid biosynthesis pathway. Thereby, this strain becomes able to synthesize fatty acids and no longer needs them from a rich medium or host. However, the strain still grew poorly in vitro when cultivated on serum-free medium. Through *in silico* simulation of the feasibility of the pathway within the genome-scale metabolic model of *M. pneumoniae*, we identified the remaining growth limitation to be due to inefficient NADPH regeneration. Subsequently, through a gap filling algorithm, we suggested synthetic constructs that would render *M. pneumoniae* able to grow at the same rate as the wild type strain.

## References

1. Lluch-senar, M. *et al.* Defining a minimal cell: essentiality of small ORFs and ncRNAs in a genome-reduced bacterium. *Mol Syst Biol.* **11** (2015).
2. Waites, K. B. & Talkington, D. F. Mycoplasma pneumoniae and its role as a human pathogen. *Clinical Microbiology Reviews* **17**, 697–728 (2004).
3. Weisburg, W. G. *et al.* A phylogenetic analysis of the mycoplasmas: basis for their classification. *Journal of Bacteriology* **171**, 6455 (1989).
4. Gaspari, E. *et al.* Model-driven design allows growth of Mycoplasma pneumoniae on serum-free media. *npj Systems Biology and Applications* **6**, 33 (2020).
5. Wodke, J. A. H. *et al.* Dissecting the energy metabolism in Mycoplasma pneumoniae through genome-scale metabolic modeling. *Molecular Systems Biology* **9**, (2013).
6. Daxboeck, F., Krause, R. & Wensch, C. Laboratory diagnosis of Mycoplasma pneumoniae infection. *Clinical Microbiology and Infection* **9**, 263–273 (2003).
7. Razin, S., Yogev, D. & Naot, Y. Molecular biology and pathogenicity of mycoplasmas. *Microbiology and molecular biology reviews: MMBR* **62**, 1094–156 (1998).
8. Rottem, S. Membrane lipids of mycoplasmas. *Biochimica et Biophysica Acta (BBA) - Reviews on Biomembranes* **604**, 65–90 (1980).
9. Yao, J. & Rock, C. O. How Bacterial Pathogens Eat Host Lipids: Implications for the Development of Fatty Acid Synthesis Therapeutics. *Journal of Biological Chemistry* **290**, 5940–5946 (2015).
10. Pollack, J. D., Somerson, N. L. & Senterfit, L. B. Chemical Composition and Serology of Mycoplasma pneumoniae Lipids. *The Journal of Infectious Diseases* **127**, S32--S35 (1973).
11. Rottem, S., Adar, L., Gross, Z. V. I., Eman, N. E. & Davis, P. J. Incorporation and Modification of Exogenous Phosphatidylcholines by Mycoplasmas. **167**, 299–304 (1986).
12. Klement, M. L. R., Öjemyr, L., Tagscherer, K. E., Widmalm, G. & Wieslander, Å. A processive lipid glycosyltransferase in the small human pathogen Mycoplasma pneumoniae: Involvement in host immune response. *Molecular Microbiology* **65**, 1444–1457 (2007).
13. Kornspan, J. D. & Rottem, S. The Phospholipid Profile of Mycoplasmas. *Journal of Lipids* **2012**, 1–8 (2012).
14. Leon, O. & Panos, C. Long-chain fatty acid perturbations in Mycoplasma pneumoniae. *Journal of Bacteriology* **146**, 1124 LP – 1134 (1981).
15. Saito, Y., Silviu, J. R. & McElhaney, N. Membrane lipid biosynthesis in Acholeplasma laidlawii B: de novo biosynthesis of saturated fatty acids by growing cells. *Journal of bacteriology* **132**, 497–504 (1977).
16. Pollack, J. D. & Tourtellotte, M. E. Synthesis of Saturated Long Chain Fatty Acids from Sodium Acetate-1-C14 by Mycoplasma. *Journal of Bacteriology* **93**, 636 (1967).
17. Rottem, S. & Panos, C. The Synthesis of Long-Chain Fatty Acids by a Cell-Free System from Mycoplasma laidlawii A. *Biochem. Biophys. Res. Commun* **9**, 970 (1970).

18. Smith, P. F. *The biology of mycoplasmas*. (Elsevier, 2012).
19. Rottem, S., Muhsam-Peled, O. & Razin, S. Acyl carrier protein in mycoplasmas. *Journal of Bacteriology* **113**, 586–591 (1973).
20. Orth, J. D. & Palsson, B. Gap-filling analysis of the iJO1366 Escherichia coli metabolic network reconstruction for discovery of metabolic functions. *BMC Systems Biology* **6**, 30 (2012).
21. Faria, J. P., Rocha, M., Rocha, I. & Henry, C. S. Methods for automated genome-scale metabolic model reconstruction. *Biochemical Society Transactions* **46**, 931–936 (2018).
22. Ong, W. K., Midford, P. E. & Karp, P. D. Taxonomic weighting improves the accuracy of a gap-filling algorithm for metabolic models. *Bioinformatics* **36**, 1823–1830 (2020).
23. Pan, S. & Reed, J. L. Advances in gap-filling genome-scale metabolic models and model-driven experiments lead to novel metabolic discoveries. *Current Opinion in Biotechnology* **51**, 103–108 (2018).
24. Thiele, I., Vlassis, N. & Fleming, R. M. T. fastGapFill: efficient gap filling in metabolic networks. *Bioinformatics* **30**, 2529–2531 (2014).
25. Ebrahim, A., Lerman, J. A., Palsson, B. O. & Hyduke, D. R. COBRApy: COstraints-Based Reconstruction and Analysis for Python. *BMC Systems Biology* **7**, 74 (2013).
26. GUROBI OPTIMIZATION INC. Gurobi optimizer reference manual, 2015. URL: <http://www.gurobi.com> 29 (2014).
27. McCoy, R. E. *et al.* Acholeplasma florum, a New Species Isolated from Plantst. *International Journal of Systematic and Evolutionary Microbiology* **34**, 11–15 (1984).
28. Reed, J. L. *et al.* Systems approach to refining genome annotation. *Proceedings of the National Academy of Sciences* **103**, 17480 (2006).
29. Lee, W.-H., Kim, M.-D., Jin, Y.-S. & Seo, J.-H. Engineering of NADPH regenerators in Escherichia coli for enhanced biotransformation. *Applied Microbiology and Biotechnology* **97**, 2761–2772 (2013).
30. Martínez, I., Zhu, J., Lin, H., Bennett, G. N. & San, K.-Y. Replacing Escherichia coli NAD-dependent glyceraldehyde 3-phosphate dehydrogenase (GAPDH) with a NADP-dependent enzyme from Clostridium acetobutylicum facilitates NADPH dependent pathways. *Metabolic Engineering* **10**, 352–359 (2008).
31. Qiao, K., Wasylenko, T. M., Zhou, K., Xu, P. & Stephanopoulos, G. Lipid production in Yarrowia lipolytica is maximized by engineering cytosolic redox metabolism. *Nature Biotechnology* **35**, 173–177 (2017).
32. Lazarev, V. N. *et al.* Complete genome and proteome of Acholeplasma laidlawii. *Journal of bacteriology* **193**, 4943–4953 (2011).
33. Borkowski, O., Ceroni, F., Stan, G. B. & Ellis, T. Overloaded and stressed: whole-cell considerations for bacterial synthetic biology. *Current Opinion in Microbiology* vol. 33 123–130 (2016).
34. Lynch, M. & Marinov, G. K. The bioenergetic costs of a gene. *Proceedings of the National Academy of Sciences* **112**, 15690 (2015).
35. Basan, M. *et al.* Inflating bacterial cells by increased protein synthesis. *Molecular systems biology* **11**, 836 (2015).

36. Shachrai, I., Zaslaver, A., Alon, U. & Dekel, E. Cost of Unneeded Proteins in E. coli Is Reduced after Several Generations in Exponential Growth. *Molecular Cell* **38**, 758–767 (2010).
37. Salman, M. & Rottem, S. The cell membrane of Mycoplasma penetrans: lipid composition and phospholipase A1 activity. *Biochimica et Biophysica Acta (BBA) - Biomembranes* **1235**, 369–377 (1995).
38. Worliczek, H. L., Kämpfer, P., Rosengarten, R., Tindall, B. J. & Busse, H.-J. Polar lipid and fatty acid profiles – Re-vitalizing old approaches as a modern tool for the classification of mycoplasmas? *Systematic and Applied Microbiology* **30**, 355–370 (2007).

## Supplementary Information

**Supplementary Table S1.** Serum-free medium used to grow *Mycoplasma pneumoniae* strains M129 and AL7.

Component	Concentration
CMRL-1066	0.5X
Glucose	0.75%
HEPES	50mM
Ampicillin	100µg/ml
Phenol red pH 7	0.0035%
Glycerol	0.025%
Glutamine	2mM
Spermine	10µg/ml
RNA	1mg/ml
PPLO	15mg/ml
non-lipidate BSA	0.33%
cholesterol	33.3µg/ml
palmitic acid	16.6µg/ml
oleic acid	20µg/ml
thioctic acid	0.2µg/ml
pyridoxamine	0.5µg/ml
nicotinic acid	0.5µg/ml
riboflavin	0.5µg/ml
choline	0.5µg/ml

**Supplementary Table S2.** Fatty acid biosynthesis pathway from *A. laidlawii* as introduced into genome-scale model iEG158\_mpn, with reference to the knocked-in genes, to simulate the synthesis of palmitoyl-ACP / oleoyl-ACP. All reactions were considered irreversible, so they were added to the model with 0 and 1000 lower and upper bounds respectively.

Gene	Enzyme Name	EC number	Reaction ID	Reaction
<i>accA</i>	acetyl-CoA carboxylase, alpha subunit	6.4.1.2	ACCOAhi	accoa_c + atp_c + hco3_c --> adp_c + malcoa_c + pi_c
<i>accB</i>	acetyl-CoA carboxylase, biotin carboxyl carrier protein	6.4.1.2	ACCOAhi	accoa_c + atp_c + hco3_c --> adp_c + malcoa_c + pi_c

<i>accC</i>	acetyl-CoA carboxylase, biotin carboxylase	6.4.1.2	ACCOAhi	accoa_c + atp_c + hco3_c --> adp_c + malcoa_c + pi_c
<i>accD</i>	acetyl-CoA carboxylase, beta subunit	6.4.1.2	ACCOAhi	accoa_c + atp_c + hco3_c --> adp_c + malcoa_c + pi_c
<i>fabD</i>	beta-ketoacyl-(acyl carrier protein) transacylase	2.3.1.39	MCOATA_1	ACP_c + malcoa_c --> coa_c + malACP_c
<i>fabH</i> 1	beta-ketoacyl-(acyl carrier protein) synthase III	2.3.1.180	3OAS40_h	accoa_c + h_c + malACP_c --> actACP_c + co2_c + coa_c
<i>fabG</i> 1	beta-ketoacyl-(acyl carrier protein) reductase	1.1.1.100	16C_SYNTHE SIS/18C_SYN THESIS	actACP_c + 14.0 h_c + 6.0 malcoa_c + 14.0 nadph_c --> ACP_c + 6.0 coa_c + hdca_c + 14.0 nadp_c / actACP_c + 16.0 h_c + 7.0 malcoa_c + 16.0 nadph_c --> ACP_c + 7.0 coa_c + 16.0 nadp_c + ocdca_c
<i>fabZ</i>	beta-hydroxyacyl-(acyl-carrier-protein)dehydratase	4.2.1.59	16C_SYNTHE SIS/18C_SYN THESIS	actACP_c + 14.0 h_c + 6.0 malcoa_c + 14.0 nadph_c --> ACP_c + 6.0 coa_c + hdca_c + 14.0 nadp_c / actACP_c + 16.0 h_c + 7.0 malcoa_c + 16.0 nadph_c --> ACP_c + 7.0 coa_c + 16.0 nadp_c + ocdca_c
<i>fabK</i>	enoyl-[acyl-carrier-protein] reductase	1.3.1.9	16C_SYNTHE SIS/18C_SYN THESIS	actACP_c + 14.0 h_c + 6.0 malcoa_c + 14.0 nadph_c --> ACP_c + 6.0 coa_c + hdca_c + 14.0 nadp_c / actACP_c + 16.0 h_c + 7.0 malcoa_c + 16.0 nadph_c --> ACP_c + 7.0 coa_c + 16.0 nadp_c + ocdca_c
<i>fabB</i>	beta-ketoacyl-(acyl carrier protein) synthetase	2.3.1.41	16C_SYNTHE SIS/18C_SYN THESIS	actACP_c + 14.0 h_c + 6.0 malcoa_c + 14.0 nadph_c --> ACP_c + 6.0 coa_c + hdca_c + 14.0 nadp_c / actACP_c + 16.0 h_c + 7.0 malcoa_c + 16.0 nadph_c --> ACP_c + 7.0 coa_c + 16.0 nadp_c + ocdca_c

**Supplementary Table S3.** Growth curve data of *M. pneumoniae* strains M129 and AL7 as Abs600. Points represent replicates of 3 measurements.

<b>Day n.</b>	<b>M129 Abs600 (x 10<sup>-3</sup>)</b>	<b>Standard Error (x 10<sup>-3</sup>)</b>
1	1.90	0.24
7	1.40	0.51
14	0.76	0.18
21	0.83	0.17
28	0.87	0.14
35	0.71	0.24
<b>Day n.</b>	<b>AL7 Abs600 (x 10<sup>-3</sup>)</b>	<b>Standard Error (x 10<sup>-3</sup>)</b>
1	1.90	0.24
7	2.25	0.18
14	2.25	0.18
21	1.59	0.50
28	2.29	0.55
35	2.45	1.35

Gap-filling algorithm scripts and libraries are provided in the git repository <https://gitlab.com/erika.gaspariwur/phd-thesis-erika-gaspari>.





# Chapter 4

## Galactocerebroside biosynthesis pathways of *Mycoplasma* species: an antigen triggering Guillain-Barré-Strohl syndrome

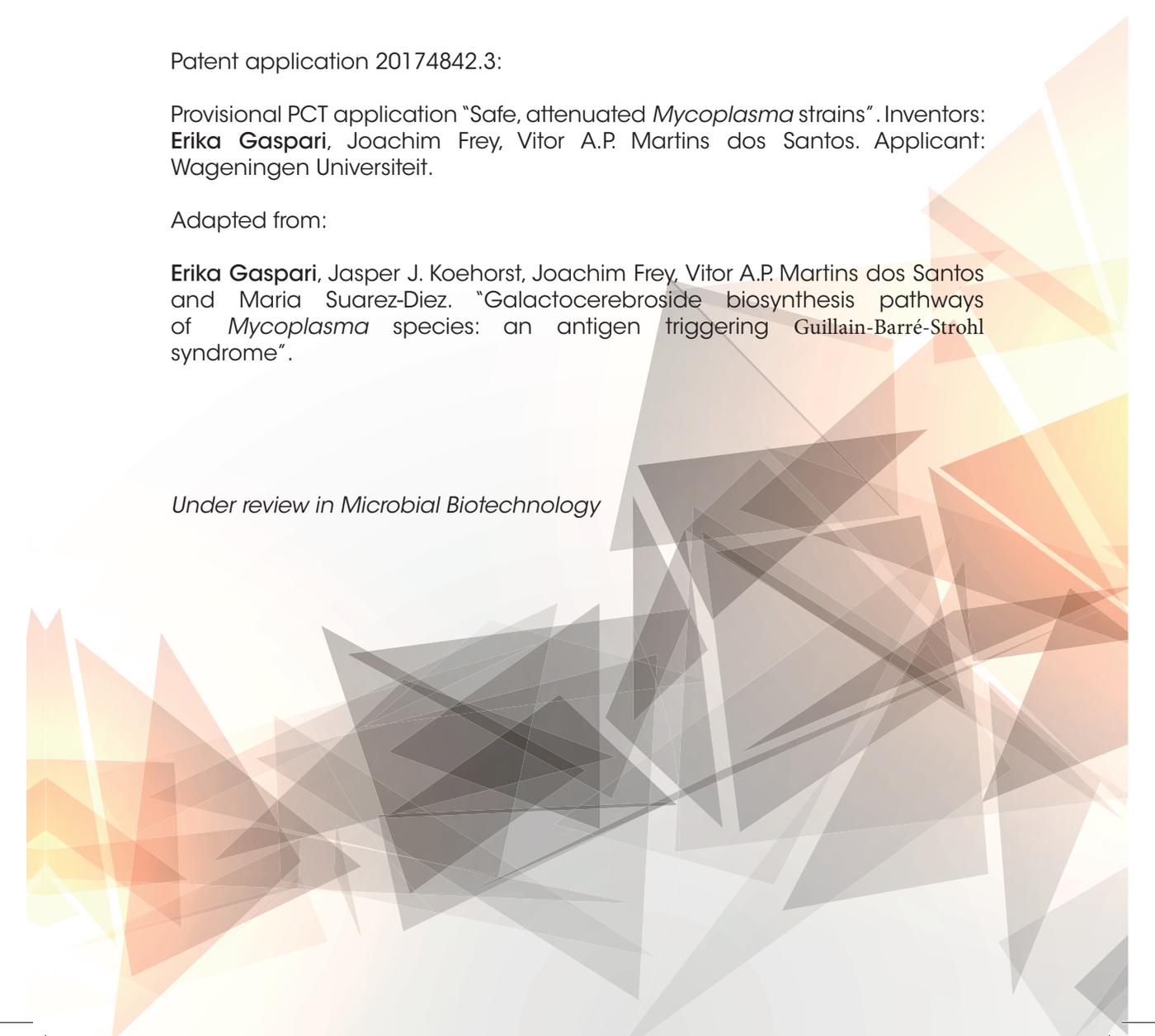
Patent application 20174842.3:

Provisional PCT application "Safe, attenuated *Mycoplasma* strains". Inventors: **Erika Gaspari**, Joachim Frey, Vitor A.P. Martins dos Santos. Applicant: Wageningen Universiteit.

Adapted from:

**Erika Gaspari**, Jasper J. Koehorst, Joachim Frey, Vitor A.P. Martins dos Santos and Maria Suarez-Diez. "Galactocerebroside biosynthesis pathways of *Mycoplasma* species: an antigen triggering Guillain-Barré-Strohl syndrome".

*Under review in Microbial Biotechnology*



## Abstract

Infection by *Mycoplasma pneumoniae* has been identified as a preceding factor of Guillain-Barré-Strohl syndrome. The Guillain-Barré-Strohl syndrome is triggered by an immune reaction against the major glycolipids and it has been postulated that *M. pneumoniae* infection triggers this syndrome due to bacterial production of galactocerebroside. Here, we present an extensive comparison of 224 genome sequences from 104 *Mycoplasma* species to characterize the genetic determinants of galactocerebroside biosynthesis. Hidden Markov models were used to analyze glycosyltransferases, leading to identification of a functional protein domain, termed M2000535, that appears in about a third of the studied genomes. This domain appears to be associated to a potential UDP-glucose epimerase, which converts UDP-glucose into UDP-galactose, a main substrate for the biosynthesis of galactocerebroside. Thus, we hypothesize that strains within this domain can synthesize this compound. These findings clarify the pathogenic mechanisms underlining the triggering of Guillain-Barré-Strohl syndrome by *M. pneumoniae* infections.

## Introduction

*Mycoplasma* species are bacteria representing the smallest free-living organisms on earth. They include several pathogens infecting human, animals and plants. *Mycoplasma* are characterized by the lack of a cell wall, reason for which glycolipids of the membrane are exposed to the host's immune system upon an infection. *Mycoplasma pneumoniae* is a human respiratory pathogen causing atypical (or walking) pneumonia, accounting for approximately 20–30% of all types of pneumonia worldwide (Liu et al., 2009; Varma-Basil et al., 2009; Waites & Talkington, 2004; Waites, Xiao, Liu, Balish, & Atkinson, 2017; Zhang et al., 2016). Diagnosis of *M. pneumoniae* infections is currently performed mostly by PCR tests but remains complicated at an early stage of infection (Miyachi et al., 2009). Since about a decade, research to improve early diagnosis has focused on glycolipid antigens present on the membrane of *M. pneumoniae* (Matsuda, 2015).

For long, *M. pneumoniae* has been suspected as a potential preceding factor of Guillain-Barré-Strohl syndrome (GBS) (Yuki, 2007), which occurs at a frequency of about 5% of the total cases of past *M. pneumoniae* infections (van den Berg et al., 2014) (Meyer Sauteur et al., 2016). The GBS is an autoimmune neurological disorder that is potentially life-threatening. *Campylobacter jejuni* is the first microorganism that was associated to post-infectious outbreak of GBS (Rees, Soudain, Gregson, & Hughes, 1995), and has been found to perform galactocerebroside biosynthesis (Hao et al., 1998). Galactocerebroside has been shown to be immunogenic to a low degree in *M. pneumoniae* infections (Kusunoki et al., 2001; Susuki et al., 2004) and it is postulated that *M. pneumoniae* triggers GBS by inducing anti-galactocerebroside IgG (Meyer Sauteur et al., 2018).

Galactocerebroside, also called galactosylceramide, is a sphingolipid, more specifically a cerebroside, characterized by a galactosyl head group. A similar compound is glucocerebroside, alias glucosylceramide, which instead consists in a cerebroside where the monosaccharide head group is glucose. Glycosphingolipids such as galactocerebroside and glucocerebroside are typically synthesized in bacteria by the enzymes, ceramide galactosyltransferase (or galactosylceramide synthase – reaction EC 2.4.1.47) and ceramide glucosyltransferase (or glucosylceramide synthase – reaction EC 2.4.1.80) of the glycosyltransferase family. The enzyme ceramide galactosyltransferase appears in viruses and cellular organisms; it synthesizes the biosynthesis of galactocerebroside by binding UDP-galactose to a ceramide molecule, releasing UDP.

The glycosyltransferase of *M. pneumoniae*, encoded by gene *mpn483*, has been shown to synthesize galactosyl-ceramide (most likely the beta-variant) using ceramide and UDP-glucose as substrates, both with phosphatidylglycerol or cardiolipin as activators. *M. pneumoniae* has access to ceramide and galactose. It imports ceramide both from *in vitro* growth in axenic medium and from the host *in vivo* during infection (Klement et al., 2007). In addition, it can use the fatty acid chains from incorporated ceramide in other lipids to build up ceramide-based glycolipids. Moreover, *M. pneumoniae* favours the import of glucose *in vivo*, albeit not *in vitro*, in which galactose is preferred (Plackett et al., 1969). Finally, it is postulated but not confirmed that *M. pneumoniae* contains a potential epimerase converting UDP-glucose into UDP-galactose (Dandekar et al., 2000).

Characterization of the galactocerebroside biosynthesis pathway in *M. pneumoniae* will further clarify pathogenic mechanisms and can greatly impact the development of methods early detection and diagnosis of GBS. Moreover, these data are essential for the design of alternative metabolic pathways for *M. pneumoniae* avoiding galactocerebroside formation and to identify alternative *Mycoplasma* species devoid of galactocerebroside for biomedical applications.

We present here a comparative analysis investigation of 9 strains of *M. pneumoniae* and additional 103 currently genome-sequenced *Mycoplasma* species. The goal is to identify the genetic determinants of galactocerebroside biosynthesis and to further characterize the proteins in this pathway.

## Materials and Methods

All available and complete *Mycoplasma* genome sequences were retrieved from the NCBI Genome repository. Overall, 224 genome sequences were obtained, belonging to 104 species that are listed in [Supplementary file 1](#). All genome sequences were re-annotated using the SAPP pipeline (Koehorst et al., 2018). Gene prediction was performed using Prodigal 2.6.3 (Hyatt et al., 2010) and protein sequences were annotated using InterProScan 5.36.75.0 (Jones et al., 2014) to assign functional domains. Annotation data were stored in a triple-store (GraphDB) (Güting, 1994) in a linked data format using the GBOL ontology as schema (van Dam et al., 2019) and systematically queried using SPARQL.

Hidden Markov Models (HMMs) were built and searched for with HMMER v3.3 (Eddy, 1998) - through commands *hmmbuild* and *hmmsearch* - on multiple sequence alignments (MSAs) performed with Clustal Omega 1.2.4 (Sievers et al., 2011; Madeira et al., 2019).

Sequence logos have been generated with WebLogo version 3 (Crooks, Hon, Chandonia, & Brenner, 2004).

## Results

We studied the functionalities associated to the *M. pneumoniae* genome, through an analysis of protein domains, to describe the biosynthesis pathway that leads to the formation of galactocerebroside and compare it to the pathways found in other mycoplasmas. Therefore, we analysed 224 genomes of 104 *Mycoplasma* species (given in [Supplementary file 1](#)), comprising 213 strains. To ensure uniform annotation and a consistent comparison all genomes were re-annotated.

The synthesis of glycosphingolipid such as galactocerebroside, in bacteria, needs a glycosyltransferase enzyme, linking a sugar (galactose) to a phospholipid (ceramide) and building the glycosyl bond. Our functional analysis identified 4 genes containing a glycosyltransferase domain in *M. pneumoniae*. These have been found in the 9 analysed strains and correspond to locus tags *mpn028*, *mpn483*, *mpn075*, and *mpn064* in *M. pneumoniae* M129. While *mpn064* (*deoA*) codes for a thymidine phosphorylase (EC 2.4.2.2) that contains a “glycosyl transferase family 3” domain with InterPro identifier IPR000312, *mpn028*, *mpn483* and *mpn075* contain a “glycosyltransferase 2-like” domain, (IPR001173), associated to proteins that have been linked to glycosphingolipids biosynthesis pathways. Their lengths and E-values of associated glycosyltransferase are shown in [Table 1](#). Interestingly, the glycosyltransferase with more significant E-value is MPN\_028, while the one known to synthesize galactocerebroside is MPN\_483 (Klement et al., 2007).

**Table 1.** The three enzymes MPN\_028, MPN\_075 and MPN\_483 of *Mycoplasma pneumoniae* M129 matching the InterPro domain IPR001173 (“Glycosyltransferase 2-like”), with E-values and gene length.

<b><i>Mycoplasma pneumoniae</i> M129 Enzyme</b>	<b>E-value: glycosyltransferase 2- like domain</b>	<b>Gene length (bp)</b>
<i>MPN_028</i>	$2.7 \times 10^{-26}$	899
<i>MPN_483</i>	$1.2 \times 10^{-25}$	1025
<i>MPN_075</i>	$1.2 \times 10^{-21}$	899

The InterPro IPR001173 domain comprises two Pfam domains: “Glyco\_trans\_2\_3” (PF13632) and “Glyco\_transf\_2” (PF00535); this last is found in two glycosyltransferases of *C. jejuni*, the first microorganism associated to post-infectious outbreak of GBS, then found to perform galactocerebroside biosynthesis, with E-values of  $10^{-29}$  (Putative galactosyltransferase – UniProtKB Q8KWR2) and  $10^{-27}$  (Beta-1,3-galactosyltransferase coded by gene *cgtB* – UniProtKB Q5DT13). Focusing on the *Mycoplasma* species, we therefore assumed the protein sequences containing the Pfam domain PF00535 to be the ones associated to galactocerebroside

synthesis. The distribution of the E-values of this signature, presented in [Figure 1](#), shows a bimodal distribution. The presence of two peaks suggests two similar but distinct domains. Thus, we re-built HMMs on the two separate groups obtaining two new domains, M100535 and M200535, then E-values for these domains on the sequences were re-calculated, as indicated in [Figure 1](#). This approach results in two groups of protein sequences each matching its corresponding motif with much higher significance, i.e. much lower E-values. All the strains of *M. pneumoniae* contain only the second motif M200535. The HMM for M200535 motif is provided in [Supplementary file 2](#). Interestingly, the two glycosyltransferases of *C. jejuni* match M200535 with much higher significance than PF00535: the beta-1,3-galactosyltransferase with E-value  $10^{-43}$  and putative galactosyltransferase with E-value  $10^{-42}$ . An additional glycosyltransferase family 2 protein matching M200535 with E-value  $10^{-42}$  was identified in *C. jejuni* and full results are given in [Supplementary file 3](#).

The two new M100535 and M200535 motifs show some substantial differences: the sequence alignments of the proteins carrying the domains reveal M200535 to be almost a twice as long as M100535. Most differences between the two motifs are present in the first part of the domain, where M200535-containing sequences show a predominance of aspartic acid in positions 344, 348, 428, 430, 436 and 455 ([Figure 2](#)). Consensus sequences for both M100535 and M200535 are provided in [Supplementary file 4](#).

Occurrences of the domains M100535, M200535, here defined, and of domains PF00534, PF13439 or PF13641 associated to alternative glycosyltransferases are summarized in [Table 2](#). In total 73 out of the 104 *Mycoplasma* species analysed match at least one glycosyltransferase domain. It should be noted that any genome containing either M100535 or M200535 also contains PF00535.

**Table 2.** Number of glycosyltransferases found in each *Mycoplasma* specie analysed for domains M100535, M200535, PF00534, PF13439 and PF13641. \* indicates the number of glycosyltransferases among the different strains of the same species, else the same number must be considered for all the strains of the indicated species. † indicates one of the glycosyltransferase domains is overlapping between M100535 and M200535 (has a  $-\text{Log}_{10}\text{E}$ -val of 13.5).

SPECIES	N. GLYCOSYLTRANSFERASES FOUND WITH DOMAIN:				
	M100535	M200535	PF00534	PF13439	PF13641
<i>M. agalactie</i> *	1	1/2	0	0	1
<i>M. agassizii</i>	1	0	1	0	0
<i>M. alligatoris</i>	1	2	0	0	1
<i>M. alvi</i>	5	3	5	0	2
<i>M. amphoriforme</i>	0	5	1	0	0
<i>M. anatis</i>	1	1	0	0	0
<i>M. anseris</i>	1	0	0	0	0
<i>M. arginini</i>	1	0	0	0	0
<i>M. arthritis</i>	1	0	0	0	0
<i>M. bovigenitalium</i>	2	2	1	1	2
<i>M. bovirhinis</i>	1	0	0	0	0
<i>M. bovis</i>	1	1	0	0	1

<i>M. bovoculi</i>	0	2	0	0	0
<i>M. buteonis</i>	1	0	0	0	0
<i>M. californicum*</i>	1/2	1	1	1	2
<i>M. canadense</i>	1	0	0	0	0
<i>M. canis</i>	1	0	0	0	0
<i>M. capricolum</i>	1	0	0	0	1
<i>M. cloacale</i>	1	0	0	0	0
<i>M. collis</i>	1	1	1	0	0
<i>M. colombinum</i>	1	0	0	0	0
<i>M. columborale</i>	1	0	0	0	0
<i>M. conjunctivae</i>	0	1	1	1	0
<i>M. cricetuli</i>	1	0	0	0	0
<i>M. crocodyli</i>	1	2	0	0	1
<i>M. cynos</i>	1	0	0	0	0
<i>M. elephantis†</i>	2/3	2/3	0	0	0
<i>M. falifaucium</i>	1	0	1	1	0
<i>M. felis</i>	1	0	0	0	0
<i>M. fermentans</i>	1	1	1	1	1
<i>M. gallinarum</i>	1	1	0	0	1
<i>M. gallisepticum</i>	1	1	0	0	0
<i>M. gallopavonis</i>	1	0	0	0	0
<i>M. genitalium*</i>	1	2/3/4	0	0	0
<i>M. girerdii</i>	2	0	0	0	0
<i>M. glycophyllum</i>	1	0	0	0	0
<i>M. hyorhinis</i>	1	0	0	0	0
<i>M. imitans</i>	0	1	0	0	0
<i>M. iners</i>	1	0	0	0	0
<i>M. iowae</i>	2	0	0	0	1
<i>M. leachii</i>	1	0	0	0	0
<i>M. leonicaptivi</i>	1	0	0	0	0
<i>M. lipofaciens</i>	1	1	0	0	0
<i>M. melagridis</i>	1	0	0	0	0
<i>M. moatsii</i>	2	6	2	0	1
<i>M. mobile</i>	1	2	1	0	0
<i>M. molare</i>	1	0	0	0	0
<i>M. mustelae</i>	1	0	0	0	0
<i>M. mycoides</i>	1	2	0	0	3
<i>M. penetrans</i>	0	3	0	0	1
<i>M. phocidae</i>	1	0	0	0	0
<i>M. phocirhinis</i>	1	0	1	1	0
<i>M. pirum</i>	2	9	4	1	1
<i>M. pneumoniae</i>	0	3	0	0	0
<i>M. primatum</i>	0	3	0	0	0
<i>M. pullorum</i>	1	0	0	0	0
<i>M. pulmonis</i>	1	0	0	0	0
<i>M. putrefaciens</i>	1	0	0	0	0
<i>M. simbae</i>	1	0	1	1	0
<i>M. sp. 5H</i>	0	3	1	0	0

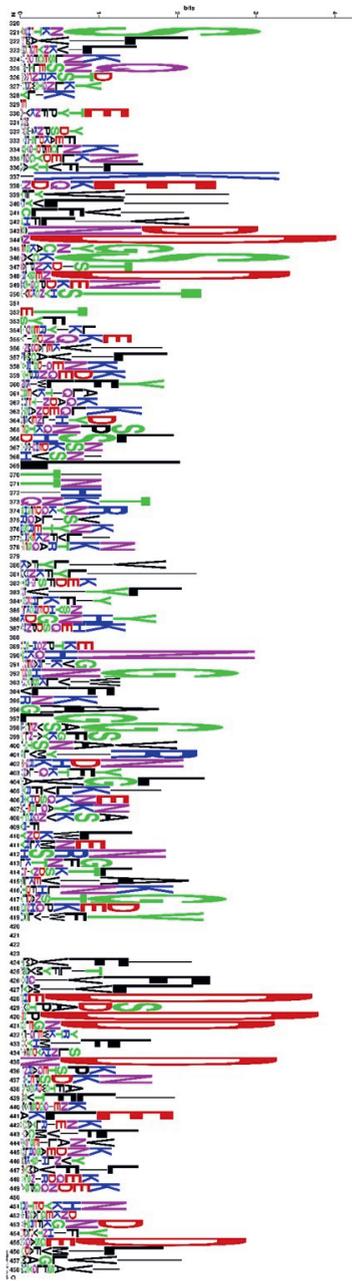
<i>M. sp. Bg1</i>	0	0	0	0	0
<i>M. sp. CAG:472</i>	1	1	2	3	0
<i>M. sp. CAG:611*</i>	0	2/3	3	2	1
<i>M. sp. CAG:776</i>	1	9	3	2	2
<i>M. sp. CAG:877</i>	1	6	5	4	0
<i>M. sp. CAG:956</i>	3	7	2	1	1
<i>M. sp. G5847</i>	3	1	0	0	1
<i>M. sp. PE</i>	1	2	1	0	0
<i>M. sp. UBA710</i>	2	4	0	0	0
<i>M. sturni</i>	1	0	0	0	0
<i>M. subdolum</i>	1	0	0	0	0
<i>M. synoviae</i>	1	0	0	0	0
<i>M. testudinis</i>	2	7	4	0	0
<i>M. yeatsii</i>	1	0	0	0	0

Functional analysis on *M. pneumoniae* reveals that *mpn257* contains Pfam domain “GDP-mannose 4,6 dehydratase” (PF16363) comprised in the InterPro domain IPR016040 (E-value:  $2.4 \cdot 10^{-46}$ ), found in sequences annotated as UDP-6-glucose 4-epimerases, that interconvert UDP-glucose into UDP-galactose. Therefore, we assume *M. pneumoniae* has the ability of converting UDP-glucose into UDP-galactose.

We can conclude the pathway for galactocerebroside synthesis in *Mycoplasma pneumoniae* *in vivo* is most likely as represented in [Figure 3](#).

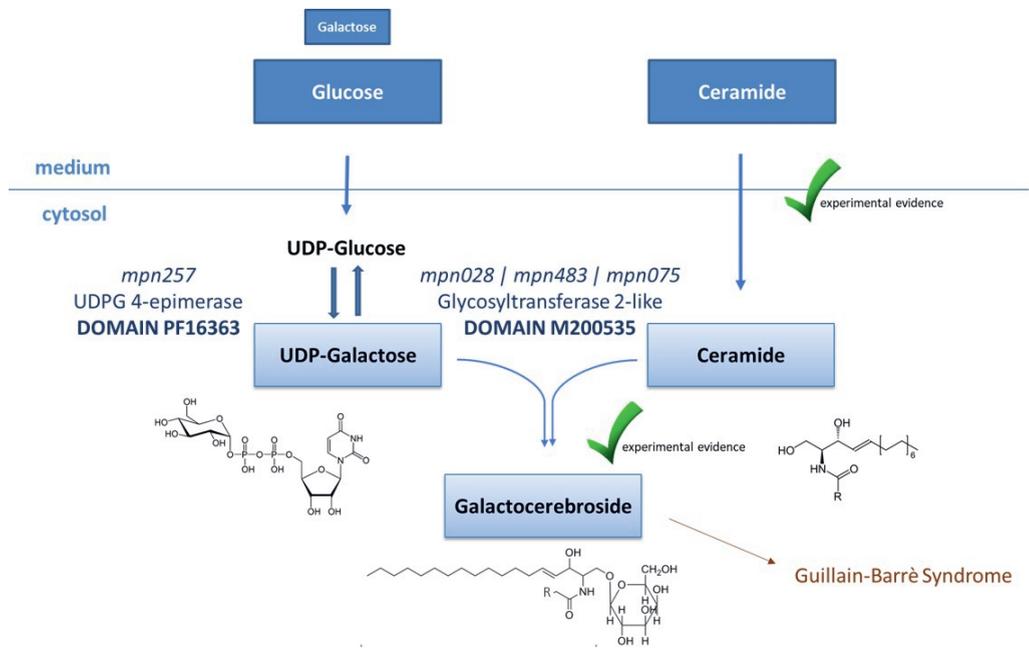
Domain Pfam PF01370 is associated with a functionally equivalent epimerase. Interestingly, analysis of all *Mycoplasma* species, shows that almost only the species containing at least one domain, M200535 matched the UDP-glucose-epimerase domains Pfam PF16363 and/or PF01370 ([Figure 4](#)). Exceptions are *M. sp. Bg1* (not containing any glycosyltransferase domain but matching an epimerase domain) and *Mycoplasma iowae* (containing a glycosyltransferase domain different from M200535).



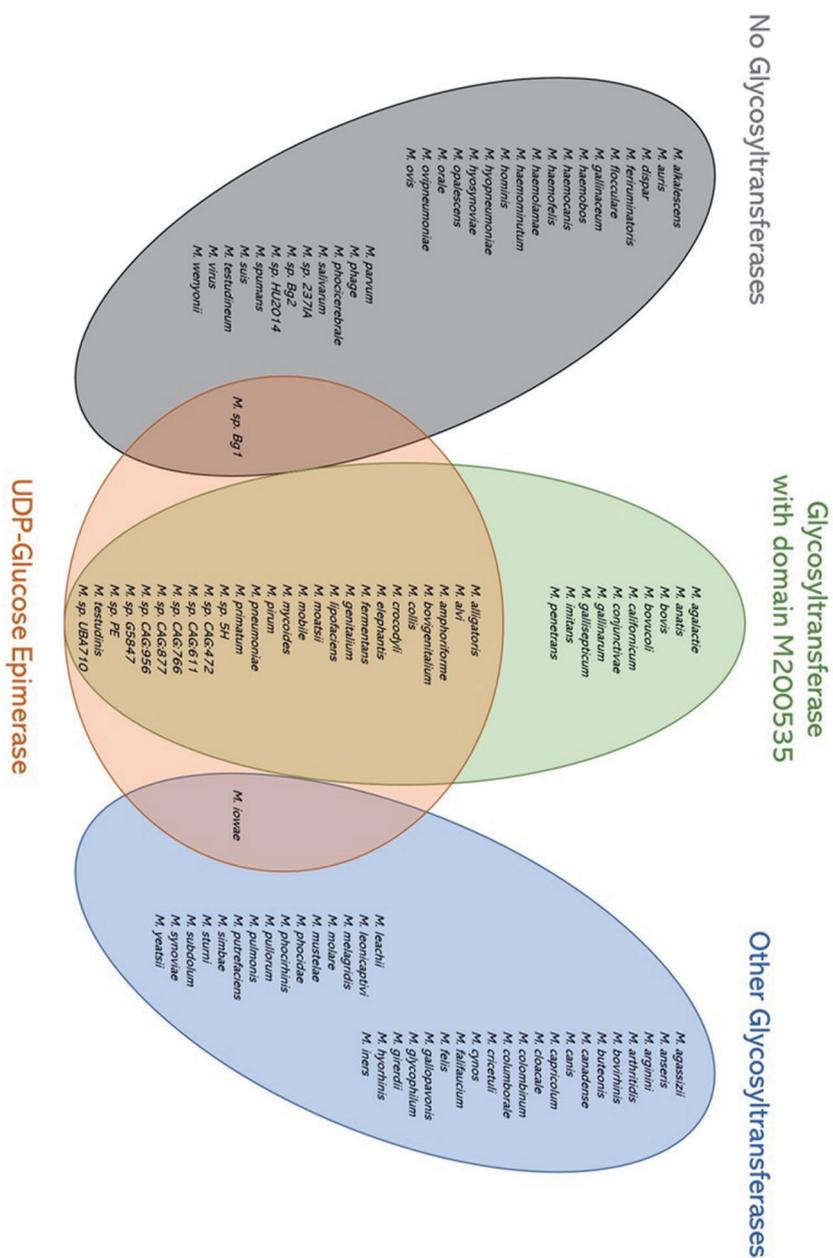


4

**Figure 2.** Sequence logo of protein domains M200535 from sequence position 320 to 458. M200535 is abundant in aspartic acid (D) in positions 344, 348, 428, 430, 436 and 455.



**Figure 3.** *In vivo* galactocerebroside biosynthesis pathway in *Mycoplasma pneumoniae* according to functional analysis. *M. pneumoniae* favors the import on glucose over galactose in vivo. Once in the cytosol, UDP-glucose is converted into UDP-galactose by the epimerase containing the functional domain PF16363, encoded by gene *mpn257*. UDP-galactose is then used by one of the glycosyltransferases encoded by *mpn028*, *mpn483* or *mpn075*, all containing the functional domain M200535. The import of ceramide and the synthesis of galactocerebroside are supported by experimental evidences.



**Figure 4.** Venn diagram of the three glycosyltransferase-representing groups of mycoplasmas (no glycosyltransferase – 31 species - in grey, at least one glycosyltransferase with domain M200535 – 36 species - in green, other glycosyltransferases with domains different from M200535 – 37 species - in blue) intersecting the group of mycoplasmas containing at least one UDP-glucose epimerase (28 species - in red). Except for *M. sp. Bg1* and *M. iowae*, only species with M200535 show presence of UDP-glucose epimerase (26 species).

## Discussion

Our results show that all strains of *M. pneumoniae* analysed have at least three glycosyltransferases encoded in the genome (in *M. pneumoniae* strain M129 these are MPN\_075, MPN\_028 and MPN\_483) that can potentially perform the synthesis of galactocerebroside. The highly significant match of M200535 with *C. jejuni* galactosyltransferases indicates that the sequences responsible of galactocerebroside synthesis contain the functional domain M200535, which, in this study, was found to be a distinctive motif with specific feature, while previously included in the Pfam domain PF00535. The *Mycoplasma* species and strains containing at least a protein with the functional domain M200535 are the only *Mycoplasma* (with the two exceptions *M. iowae* and the taxonomically not yet defined *M. sp. Bg1*) showing concomitant presence of a UDP-glucose epimerase domain, which converts UDP-glucose into UDP-galactose. UDP-galactose is used in the galactocerebroside synthesis and favoured, over UDP-glucose, as a substrate by the glycosyltransferase operating the linkage. In fact, M200535 contains conserved aspartic acid residues, which in other microorganisms such as *E. coli* have been found to be catalytically essential for glycosyltransferase to exploit their  $\beta$ -transferase activity on UDP-sugars (Griffiths et al., 1998).

Our analysis pinpoints that almost all the *Mycoplasma* species infecting humans contain the domain M200535 (Figure 4). Other *Mycoplasma* species without the functional domain M200535 show an absence of UDP-glucose epimerase. About half of them lack any domain associated to glycosyltransferase function; therefore, this set of species would be most likely unable to synthesize galactocerebroside that triggers GBS. Among the human-related *Mycoplasma* species, those with no glycosyltransferase domain are *M. hominis* and *M. orale*, while among respiratory tract-related species, present in organisms other than humans, are the calves' *M. dispar*, the swine's *M. flocculare* and *M. hyopneumoniae*, the caprine's *M. ovipneumoniae*, dog's *M. spumans* and turtle's *M. testudineum*. The rest of the species without functional domain M200535, and therefore lacking UDP-glucose epimerase, comprise *Mycoplasma* species with one or more glycosyltransferase enzymes containing domains. Glycosyltransferases with domains such as M100535 do not show conserved presence of aspartic acid residues, indicating they might be unable to use UDP-sugars as substrate and, in consequence, to perform the synthesis of galactocerebroside. Human *Mycoplasma* species with glycosyltransferases lacking the domain M200535 have not been detected, while the following non-human respiratory tract-related species have been identified with glycosyltransferase and absence of M200535: tortoise's *M. agassizii*, domestic animal's *M. arginine*, bovine *M. bovirhinis* and *M. canadense*, wild bird's *M. buteonis*, goat's *M. capricolum*, pigeon's *M. columborale*, canine *M. cynos*, pig's *M. hyorhinis*, chicken's *M. iners* and *M. synoviae*, turkey's *M. melagridis*, dog's *M. molare*, mink's *M. mustelae*, seal's *M. phocidae* and *M. phocirhinis*, avian *M. pullorum* and lion's *M. simbae*.

Our work leads to the suggestion of genetic modifications that would validate the hypotheses formulated by computational analysis. The characterization of the pathway indeed suggests which genes should be a primary target for genetic modifications to avoid biosynthesis of galactocerebroside in *M. pneumoniae*. The most intuitive strategy would be the knock-out of the genes encoding for the glycosyltransferases blocking the transfer of galactose to ceramide.

However, this seems to be non-trivial: in a global transposon mutagenesis inactivation experiment of *M. genitalium* the gene encoding for the glycosyltransferase MG\_517, homologous of MPN\_483 in *M. pneumoniae* (Klement et al., 2007), remained untouched suggesting the gene encoding for this enzyme is essential (Glass et al., 2006). Although it is proven that *mg517* and *mpn483* share 77% of gene sequence similarity, the enzymatic activities slightly differ in terms of specificity (Andres et al., 2011). The essentiality of *mpn483*, not excluded by gene transposon analysis (Lluch-senar et al., 2015), is expected as *M. pneumoniae* uses this enzyme to perform synthesis of many other lipids in the membrane (Klement, 2007), which we know are crucial for its survival (Gaspari et al., 2019). In the same way, the gene transposon analysis conducted by Lluch-Senar et al. suggests *mpn075* and *mpn257*, respectively coding for the glycosyltransferase MPN\_075 and the epimerase MPN\_257, might be essential, at the contrary of *mpn028*, coding for the third glycosyltransferase, which is reported to be non-essential. A validation of our computational analysis would consist in knocking out *mpn028* and replacing *mpn483* and *mpn075* with genes coding for glycosyltransferases that do not possess the motif M200535. To facilitate the genetic modification, genes coding for glycosyltransferases of other mycoplasmas should be chosen, among the ones not containing the motif M200535 (Figure 4). However, the essentiality of genes *mpn075* and *mpn257* remains to be clarified, due to the uncertainty of the gene transposon analysis method in establishing gene essentiality. Moreover, the essentiality of the gene arises only after several strain passages (Lluch-Senar et al., 2015). The knock-out of *mpn257*, coding for the UDP-glucose epimerase, would indeed constitute an additional layer of safety: the limited import of galactose *in vivo* is not a sufficient condition for assuming limiting quantities for biosynthesis of galactocerebroside, as the bacterium could take the needed amount of galactose from glucose conversion.

*M. pneumoniae* is the *Mycoplasma* species which metabolome, transcriptome and proteome are among the best-studied, hence giving access to profound basic research opportunities and development of novel approaches in biotechnology and biomedicine (Chen et al., 2016; Maier et al., 2013; Trussart et al., 2017; Yus, Lluch-Senar, & Serrano, 2019; Yus et al., 2009). Therefore, an experimental validation of our computational analysis would represent a suitable approach to render this bacterium safe in its numerous applications. In fact, other attempts to avoid biosynthesis of galactocerebroside seem to be not trivial: it would not be possible to knock-out genes involved in sphingolipid transporter since this transporter would not only import ceramide but also very important lipids such as sphingomyelin that were shown to be essential to the survival of *M. pneumoniae* (Gaspari et al., 2019). Instead, the introduction of a ceramidase, disassembling ceramide into its sphingosine backbone and fatty acids chains, is another potential strategy to integrate. In bacteria, this enzyme has been found in *Pseudomonas aeruginosa* (Kita, Okino, & Ito, 2000; Okino, Tani, Imayama, & Ito, 1998) and *Mycobacterium tuberculosis* (Okino, Ikeda, & Ito, 2014) to be neutral (Tani, Okino, Sueyoshi, & Ito, 2004) and reversible (Ito, Okino, & Tani, 2014), so not always favouring degradation of ceramide but also its synthesis, according to the environmental and cytosolic conditions. However, *M. pneumoniae* affinity for sphingomyelin is unique, therefore the degradation of ceramide into sphingosine and fatty acids could be of advantage for *M. pneumoniae* to build up sphingomyelin, which it typically imports unchanged from the medium (Worliczek et al., 2007). The *in silico* characterization of the protein domain that might be responsible of the galactocerebroside biosynthesis has potential impact as a target for drugs related to post-infectious GBS: the elucidation of the pathway and its analysis on mycoplasmas could be used

to assess the risk of post-infectious GBS development, helping in the development of therapeutical strategies for early diagnose and/or control of GBS.

Moreover, in this manuscript we report a group of mycoplasmas that lacks both UDP-glucose epimerase, converting UDP-glucose into UDP-galactose, and any glycosyltransferase domain, therefore unable to complex lipids with sugars. This group consists of *M. alkalescens*, *M. auris*, *M. dispar*, *M. feriruminatoris*, *M. flocculare*, *M. gallinaceum*, *M. haemobos*, *M. haemocanis*, *M. haemofelis*, *M. haemolamae*, *M. haemominutum*, *M. hominis*, *M. hyopneumoniae*, *M. hyosynoviae*, *M. opalescens*, *M. orale*, *M. ovipneumoniae*, *M. ovis*, *M. parvum*, *M. phage*, *M. phocicerebrale*, *M. salivarum*, *M. spumans*, *M. suis*, *M. testudineum*, *M. virus*, *M. wenyonii* and the uncharacterized *Mycoplasma* species named 237IA, Bg2 and HU2014. Lacking the whole machinery for galactocerebroside synthesis, this group of *Mycoplasma* species may be considered suitable for biomedical applications in terms of reducing the risk of post-infectious GBS.

## Conclusion

All *Mycoplasma pneumoniae* strains have genes encoding for glycosyltransferases, of which at least two are essential and at least one has been proved to encode for an enzyme (MPN\_483 in M129) that can synthesize glycosphingolipids such as galactocerebroside. Most likely, MPN\_028 and MPN\_075 perform the same synthesis, as they show high significant match for motif M200535. This motif was found as well in *C. jejuni*, the first microorganism linked to GBS through galactocerebroside biosynthesis. While the access of *M. pneumoniae* to galactose *in vivo*, when the import of glucose is at higher rate, remains unclear, the presence of a UDP-glucose epimerase MPN\_257, converting UDP-glucose into UDP-galactose, could make the synthesis of galactocerebroside possible. We can conclude that all wild type strains of *M. pneumoniae* are potentially capable of synthesizing galactocerebroside and will most likely be able to do so even with substitution or knock-out of the galactosyltransferase MPN483 or the epimerase MPN\_257, which will be problematic for certain medical applications. Similarly, this will be the case for *Mycoplasma* species with the functional domain M200535, presenting also a UDP-glucose epimerase domain. Our data show a set of *Mycoplasma* species that could serve as alternatives for such biomedical applications or that could provide pathway genes to modify *M. pneumoniae* accordingly.

## **Conflict of Interest**

Patent application n. EP20174842.3.

All authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## **Author Contributions**

JF suggested the study. EG, JJK, JF, MS-D and VAPMdS conceived the project. JJK performed the functional domain analysis. EG developed the motif model. EG wrote the chapter with input of MS-D, JF, VAPMdS and JJK, JF, MS-D and VAPMdS supervised the project.

## **Funding**

This work has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement n. 634942 (MycoSynVac) and from the Dutch TTW-NWO under agreement n. 15814 (SafeChassis).

## References

- Beckman, E. M., Porcelli, S. A., Morita, C. T., Behar, S. M., Furlong, S. T., & Brenner, M. B. (1994). Recognition of a lipid antigen by CD1-restricted  $\alpha\beta$ + T cells. *Nature*, *372*(6507), 691–694. <https://doi.org/10.1038/372691a0>
- Braun, G. S., Wagner, K. S., Huttner, B. D., & Schmid, H. (2006). Mycoplasma pneumoniae: Usual suspect and unsecured Diagnosis in the acute setting. *Journal of Emergency Medicine*, *30*(4), 371–375. <https://doi.org/10.1016/j.jemermed.2005.07.015>
- Caporale, C. M., Papola, F., Fioroni, M. A., Aureli, A., Giovannini, A., Notturmo, F., ... Uncini, A. (2006). Susceptibility to Guillain-Barré syndrome is associated to polymorphisms of CD1 genes. *Journal of Neuroimmunology*, *177*(1), 112–118. <https://doi.org/https://doi.org/10.1016/j.jneuroim.2006.05.018>
- Chen, Z., Shao, X., Dou, X., Zhang, X., Wang, Y., Zhu, C., ... Yan, Y. (2016). Role of the Mycoplasma pneumoniae/Interleukin-8/Neutrophil Axis in the Pathogenesis of Pneumonia. *PLoS One*, *11*(1), e0146377–e0146377. <https://doi.org/10.1371/journal.pone.0146377>
- Crooks, G. E., Hon, G., Chandonia, J., & Brenner, S. E. (2004). WebLogo: A Sequence Logo Generator, 1188–1190. <https://doi.org/10.1101/gr.849004.1>
- Dandekar, T., Huynen, M., Regula, J. T., Ueberle, B., Zimmermann, C. U., Andrade, M. A., ... Bork, P. (2000). Re-annotating the Mycoplasma pneumoniae genome sequence: adding value, function and reading frames. *Nucleic Acids Research*, *28*(17), 3278–3288. <https://doi.org/10.1093/nar/28.17.3278>
- Eddy, S. R. (1998). Profile hidden Markov models. *Bioinformatics*, *14*(9), 755–763. <https://doi.org/10.1093/bioinformatics/14.9.755>
- Foy, H. M. (1993). Infections Caused by Mycoplasma pneumoniae and Possible Carrier State in Different Populations of Patients. *Clinical Infectious Diseases*, *17*(Supplement\_1), S37–S46. [https://doi.org/10.1093/clinids/17.Supplement\\_1.S37](https://doi.org/10.1093/clinids/17.Supplement_1.S37)
- Gaspari, E., Malachowski, A., Garcia-morales, L., Burgos, R., Serrano, L., & Vitor, A. P. (2019). Model-driven design allows growth of Mycoplasma pneumoniae on serum-free media, 1–19.
- Glass, J. I., Assad-Garcia, N., Alperovich, N., Yooseph, S., Lewis, M. R., Maruf, M., ... Venter, J. C. (2006). Essential genes of a minimal bacterium. *Proceedings of the National Academy of Sciences*, *103*(2), 425–430. <https://doi.org/10.1073/pnas.0510013103>
- Griffiths, G., Cook, N. J., Gottfridson, E., Lind, T., Lidholt, K., & Roberts, I. S. (1998). Characterization of the Glycosyltransferase Enzyme from the Escherichia coli K5 Capsule Gene Cluster and Identification and Characterization of the Glucuronyl Active Site \*, *273*(19), 11752–11757.
- Güting, R.H. (1994). GraphDB: A Data Model and Query Language for Graphs in Databases.
- Hao, Q., Saida, T., Kuroki, S., Nishimura, M., Nukina, M., Obayashi, H., & Saida, K. (1998). Antibodies to gangliosides and galactocerebroside in patients with Guillain-Barré syndrome with preceding Campylobacter jejuni and other identical infections. *Journal of Neuroimmunology*, *81*(1), 116–126. [https://doi.org/https://doi.org/10.1016/S0165-5728\(97\)00166-5](https://doi.org/https://doi.org/10.1016/S0165-5728(97)00166-5)
- Hata, A., Honda, Y., Asada, K., Sasaki, Y., Kenri, T., & Hata, D. (2008). Mycoplasma hominis meningitis in a neonate: Case report and review. *Journal of Infection*, *57*(4), 338–343. <https://doi.org/https://doi.org/10.1016/j.jinf.2008.08.002>
- Hyatt, D., Chen, G.-L., LoCascio, P. F., Land, M. L., Larimer, F. W., & Hauser, L. J. (2010). Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics*,

11(1), 119. <https://doi.org/10.1186/1471-2105-11-119>

- Ito, M., Okino, N., & Tani, M. (2014). New insight into the structure, reaction mechanism, and biological functions of neutral ceramidase. *Biochimica et Biophysica Acta - Molecular and Cell Biology of Lipids*, 1841(5), 682–691. <https://doi.org/10.1016/j.bbalip.2013.09.008>
- Jones, P., Binns, D., Chang, H.-Y., Fraser, M., Li, W., McAnulla, C., ... Hunter, S. (2014). InterProScan 5: genome-scale protein function classification. *Bioinformatics*, 30(9), 1236–1240. <https://doi.org/10.1093/bioinformatics/btu031>
- Kita, K., Okino, N., & Ito, M. (2000). Reverse hydrolysis reaction of a recombinant alkaline ceramidase of *Pseudomonas aeruginosa*. *Biochimica et Biophysica Acta - Molecular and Cell Biology of Lipids*, 1485(2–3), 111–120. [https://doi.org/10.1016/S1388-1981\(00\)00029-9](https://doi.org/10.1016/S1388-1981(00)00029-9)
- Klement, M. L. R., Öjemyr, L., Tagscherer, K. E., Widmalm, G., & Wieslander, Å. (2007). A processive lipid glycosyltransferase in the small human pathogen *Mycoplasma pneumoniae*: Involvement in host immune response. *Molecular Microbiology*, 65(6), 1444–1457. <https://doi.org/10.1111/j.1365-2958.2007.05865.x>
- Koehorst, J. J., van Dam, J. C. J., Saccenti, E., Martins dos Santos, V. A. P., Suarez-Diez, M., & Schaap, P. J. (2018). SAPP: functional genome annotation and analysis through a semantic framework using FAIR principles. *Bioinformatics*, 34(8), 1401–1403. <https://doi.org/10.1093/bioinformatics/btx767>
- Kusunoki, S., Chiba, A., Hitoshi, S., Takizawa, H., & Kanazawa, I. (1995). Anti-gal-C antibody in autoimmune neuropathies subsequent to mycoplasma infection. *Muscle & Nerve*, 18(4), 409–413. <https://doi.org/10.1002/mus.880180407>
- Kusunoki, S., Shiina, M., & Kanazawa, I. (2001). Anti-Gal-C antibodies in GBS subsequent to mycoplasma infection: Evidence of molecular mimicry. *Neurology*, 57(4), 736–738. <https://doi.org/10.1212/WNL.57.4.736>
- Liu, Y., Ye, X., Zhang, H., Xu, X., Li, W., Zhu, D., & Wang, M. (2009). Antimicrobial Susceptibility of *Mycoplasma pneumoniae* Isolates and Molecular Analysis of Macrolide-Resistant Strains from Shanghai, China. *Antimicrobial Agents and Chemotherapy*, 53(5), 2160–2162. <https://doi.org/10.1128/AAC.01684-08>
- Lluch-senar, M., Delgado, J., Chen, W., Lloréns-rico, V., Reilly, F. J. O., Wodke, J. A. H., ... Gavin, A. (2015). Defining a minimal cell: essentiality of small ORFs and ncRNAs in a genome-reduced bacterium, 1–7.
- Madeira, F., Park, Y. M., Lee, J., Buso, N., Gur, T., Madhusoodanan, N., ... Lopez, R. (2019). The EMBL-EBI search and sequence analysis tools APIs in 2019. *Nucleic Acids Research*, 47(W1), W636–W641. <https://doi.org/10.1093/nar/gkz268>
- Maier, T., Marcos, J., Wodke, J. A. H., Paetzold, B., Liebeke, M., Gutiérrez-Gallego, R., & Serrano, L. (2013). Large-scale metabolome analysis and quantitative integration with genomics and proteomics data in *Mycoplasma pneumoniae*. *Mol. BioSyst.*, 9(7), 1743–1755. <https://doi.org/10.1039/C3MB70113A>
- Mason, C. (2019). *Mycoplasma bovis* infections in dairy calves. *Livestock*, 24(5), 214–220. <https://doi.org/10.12968/live.2019.24.5.214>
- Matsuda, K. (2015). A novel therapeutic strategy for mycoplasma infectious diseases. *Personalized Medicine Universe*, 4, 32–39. <https://doi.org/10.1016/j.pmu.2015.04.005>
- Meyer Sauter, P. M., Huizinga, R., Tio-Gillen, A. P., Drenthen, J., Unger, W. W. J., Jacobs, E., van Rossum A and Jacobs, B. C. (2018). Intrathecal antibody responses to GalC in Guillain-Barré syndrome triggered by *Mycoplasma pneumoniae*. *Journal of Neuroimmunology*, 314(November 2017), 13–16. <https://doi.org/10.1016/j.jneuroim.2017.11.011>

- Meyer Sauteur, P. M., Huizinga, R., Tio-Gillen, A. P., Roodbol, J., Hoogenboezem, T., Jacobs, E., ... Jacobs, B. C. (2016). Mycoplasma pneumoniae triggering the Guillain-Barré syndrome: A case-control study. *Annals of Neurology*, *80*(4), 566–580. <https://doi.org/10.1002/ana.24755>
- Miyachi, A., Miyazaki, A., Shingu, Y., Matsuda, K., Dohi, H., & Nishida, Y. (2009). Synthesis and absolute structures of Mycoplasma pneumoniae  $\beta$ -glyceroglycolipid antigens. *Carbohydrate Research*, *344*(1), 36–43. <https://doi.org/10.1016/j.carres.2008.09.028>
- Miyashita, N., Akaike, H., Teranishi, H., Ouchi, K., & Okimoto, N. (2013). Macrolide-Resistant Mycoplasma pneumoniae Pneumonia in Adolescents and Adults: Clinical Findings, Drug Susceptibility, and Therapeutic Efficacy. *Antimicrobial Agents and Chemotherapy*, *57*(10), 5181–5185. <https://doi.org/10.1128/AAC.00737-13>
- Mulvey, J. M., Padowitz, A., Lindley-Jones, M., & Nickels, R. (2007). Mycoplasma pneumoniae associated with Stevens Johnson syndrome. *Anaesthesia and Intensive Care*, *35*(3), 414–417. <https://doi.org/10.1177/0310057x0703500317>
- Okino, N., Ikeda, R., & Ito, M. (2014). Expression , Purification , and Characterization of a Recombinant Neutral Ceramidase from Mycobacterium tuberculosis, *8451*(May). <https://doi.org/10.1271/bbb.90645>
- Okino, N., Tani, M., Imayama, S., & Ito, M. (1998). Purification and characterization of a novel ceramidase from pseudomonas aeruginosa. *Journal of Biological Chemistry*, *273*(23), 14368–14373. <https://doi.org/10.1074/jbc.273.23.14368>
- Paulsen, I. T., Sliwinski, M. K., & Saier, M. H. (1998). Microbial genome analyses: global comparisons of transport capabilities based on phylogenies, bioenergetics and substrate specificities11Edited by G. Von Heijne. *Journal of Molecular Biology*, *277*(3), 573–592. <https://doi.org/https://doi.org/10.1006/jmbi.1998.1609>
- Plackett, P., Marmion, B. P., Shaw, E. J., & Lemcke, R. M. (1969). IMMUNOCHEMICAL ANALYSIS OF MYCOPLASMA PNEUMONIAE. *Australian Journal of Experimental Biology and Medical Science*, *47*(2), 171–195. <https://doi.org/10.1038/icb.1969.19>
- Rees, J. H., Soudain, S. E., Gregson, N. A., & Hughes, R. A. C. (1995). Campylobacter jejuni Infection and Guillain-Barré Syndrome. *New England Journal of Medicine*, *333*(21), 1374–1379. <https://doi.org/10.1056/NEJM199511233332102>
- Klement, M. (2007). *The advantages of being small: Glycosyltransferases in many dimensions and glycolipid synthesis in Mycoplasma Pneumoniae*. Institutionen för biokemi och biofysik.
- Schalock, P. C., & Dinulos, J. G. H. (2005). Mycoplasma pneumoniae-induced Stevens-Johnson syndrome without skin lesions: fact or fiction? *Journal of the American Academy of Dermatology*, *52*(2), 312–315. <https://doi.org/10.1016/j.jaad.2004.07.044>
- Sieling, P. A., Chatterjee, D., Porcelli, S. A., Prigozy, T. I., Mazzaccaro, R. J., Soriano, T., ... et, al. (1995). CD1-restricted T cell recognition of microbial lipoglycan antigens. *Science*, *269*(5221), 227–230. <https://doi.org/10.1126/science.7542404>
- Sievers, F., Wilm, A., Dineen, D., Gibson, T. J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., Soding, J., Thompson, J. D., Higgins, D. G. (2011). Fast, scalable generation of high-quality protein multiple sequence alignment using Clustal Omega. *Molecular Systems Biology*, *7*(1), 539. <https://doi.org/10.1038/msb.2011.75>
- Susuki, K., Odaka, M., Mori, M., Hirata, K., & Yuki, N. (2004). Acute motor axonal neuropathy after Mycoplasma infection. *Neurology*, *62*(6), 949–956. <https://doi.org/10.1212/01.WNL.0000115123.42929.FD>
- Tani, M., Okino, N., Sueyoshi, N., & Ito, M. (2004). Conserved amino acid residues in the COOH-terminal tail are indispensable for the correct folding and localization and enzyme activity of

- neutral ceramidase. *Journal of Biological Chemistry*, 279(28), 29351–29358. <https://doi.org/10.1074/jbc.M404012200>
- Andrés E. (2011). Expression and Characterization of a *Mycoplasma genitalium* Glycosyltransferase in Membrane Glycolipid Biosynthesis, 286(41), 35367–35379. <https://doi.org/10.1074/jbc.M110.214148>
- Trussart, M., Yus, E., Martinez, S., Baù, D., Tahara, Y. O., Pengo, T., ... Serrano, L. (2017). Defined chromosome structure in the genome-reduced bacterium *Mycoplasma pneumoniae*. *Nature Communications*, 8(1), 14665. <https://doi.org/10.1038/ncomms14665>
- van Dam, J. C. J., Koehorst, J. J., Vik, J. O., Martins dos Santos, V. A. P., Schaap, P. J., & Suarez-Diez, M. (2019). The Empusa code generator and its application to GBOL, an extendable ontology for genome annotation. *Scientific Data*, 6(1), 254. <https://doi.org/10.1038/s41597-019-0263-7>
- van den Berg, B., van der Eijk, A. A., Pas, S. D., Hunter, J. G., Madden, R. G., Tio-Gillen, A. P., ... Jacobs, B. C. (2014). Guillain-Barré syndrome associated with preceding hepatitis E virus infection. *Neurology*, 82(6), 491 LP – 497. <https://doi.org/10.1212/WNL.0000000000000111>
- Varma-Basil, M., Dwivedi, S. K. D., Kumar, K., Pathak, R., Rastogi, R., Thukral, S. S., ... Chaudhary, R. (2009). Role of *Mycoplasma pneumoniae* infection in acute exacerbations of chronic obstructive pulmonary disease. *Journal of Medical Microbiology*, 58(3), 322–326. <https://doi.org/https://doi.org/10.1099/jmm.0.003335-0>
- Waites, K. B., & Talkington, D. F. (2004). *Mycoplasma pneumoniae* and its role as a human pathogen. *Clinical Microbiology Reviews*, 17(4), 697–728. <https://doi.org/10.1128/CMR.17.4.697-728.2004>
- Waites, K. B., Xiao, L., Liu, Y., Balish, M. F., & Atkinson, T. P. (2017). *Mycoplasma pneumoniae* from the Respiratory Tract and Beyond. *Clinical Microbiology Reviews*, 30(3), 747–809. <https://doi.org/10.1128/CMR.00114-16>
- Worliczek, H. L., Kämpfer, P., Rosengarten, R., Tindall, B. J., & Busse, H.-J. (2007). Polar lipid and fatty acid profiles – Re-vitalizing old approaches as a modern tool for the classification of mycoplasmas? *Systematic and Applied Microbiology*, 30(5), 355–370. <https://doi.org/https://doi.org/10.1016/j.syapm.2007.03.004>
- Yuki, N. (2007). Ganglioside mimicry and peripheral nerve disease. *Muscle and Nerve*, 35(6), 691–711. <https://doi.org/10.1002/mus.20762>
- Yus, E., Lluch-senar, M., & Serrano, L. (2019). Determination of the Gene Regulatory Network of a Genome-Reduced Bacterium Highlights Alternative Regulation Independent of Transcription Factors Article Determination of the Gene Regulatory Network of a Genome-Reduced Bacterium Highlights Alternative Regulation Independent of Transcription Factors, 143–158. <https://doi.org/10.1016/j.cels.2019.07.001>
- Yus, E., Maier, T., Michalodimitrakis, K., Van Noort, V., Yamada, T., Chen, W. H., ... Serrano, L. (2009). Impact of genome reduction on bacterial metabolism and its regulation. *Science*, 326(5957), 1263–1268. <https://doi.org/10.1126/science.1177263>
- Zhang, Y., Zhou, Y., Li, S., Yang, D., Wu, X., & Chen, Z. (2016). The Clinical Characteristics and Predictors of Refractory *Mycoplasma pneumoniae* Pneumonia in Children. *PLoS One*, 11(5), e0156465–e0156465. <https://doi.org/10.1371/journal.pone.0156465>

## Supplementary Information

### Supplementary File 1

List of *Mycoplasma* genome sequences used in the study, obtained from NCBI repository. The 224 genome sequences belong to 104 *Mycoplasma* species and 213 *Mycoplasma* strains.

GCA_000008205.1	<i>Mycoplasma hyopneumoniae</i> J
GCA_000008225.1	<i>Mycoplasma hyopneumoniae</i> 7448
GCA_000008245.1	<i>Mycoplasma synoviae</i> 53
GCA_000008365.1	<i>Mycoplasma mobile</i> 163K
GCA_000008405.1	<i>Mycoplasma hyopneumoniae</i> 232
GCA_000011225.1	<i>Mycoplasma penetrans</i> HF-2
GCA_000011445.1	<i>Mycoplasma mycoides</i> subsp. <i>mycoides</i> SC str. PG1
GCA_000012765.1	<i>Mycoplasma capricolum</i> subsp. <i>capricolum</i> ATCC 27343
GCA_000020065.1	<i>Mycoplasma arthritis</i> 158L3-1
GCA_000023165.1	<i>Mycoplasma mycoides</i> subsp. <i>capri</i> str. GM12
GCA_000023685.1	<i>Mycoplasma mycoides</i> subsp. <i>capri</i> str. GM12
GCA_000025365.1	<i>Mycoplasma gallisepticum</i> str. R(high)
GCA_000025385.1	<i>Mycoplasma gallisepticum</i> str. F
GCA_000025845.1	<i>Mycoplasma crocodyli</i> MP145
GCA_000026765.1	<i>Mycoplasma conjunctivae</i>
GCA_000027325.1	<i>Mycoplasma genitalium</i> G37
GCA_000027345.1	<i>Mycoplasma pneumoniae</i> M129
GCA_000063605.1	<i>Mycoplasma agalactiae</i>
GCA_000085865.1	<i>Mycoplasma hominis</i> ATCC 23114
GCA_000089865.1	<i>Mycoplasma agalactiae</i>
GCA_000092585.1	<i>Mycoplasma gallisepticum</i> str. R(low)
GCA_000143865.1	<i>Mycoplasma mycoides</i> subsp. <i>mycoides</i> SC str. Gladysdale
GCA_000143945.1	<i>Mycoplasma pneumoniae</i> FH
GCA_000145705.1	<i>Mycoplasma hyorhinis</i> HUB-1
GCA_000148625.1	<i>Mycoplasma fermentans</i> JER
GCA_000167595.1	<i>Mycoplasma genitalium</i> G37
GCA_000178375.1	<i>Mycoplasma alligatoris</i> A21JP2
GCA_000179035.2	<i>Mycoplasma suis</i> str. Illinois
GCA_000183185.1	<i>Mycoplasma hyopneumoniae</i> 168
GCA_000183365.1	<i>Mycoplasma leachii</i> PG50
GCA_000183385.1	<i>Mycoplasma bovis</i> PG45
GCA_000186005.1	<i>Mycoplasma fermentans</i> M64
GCA_000186985.3	<i>Mycoplasma haemofelis</i> Ohio2
GCA_000192395.2	<i>Mycoplasma capricolum</i> subsp. <i>capripneumoniae</i> M1601
GCA_000195875.1	<i>Mycoplasma pulmonis</i> UAB CTIP
GCA_000200735.1	<i>Mycoplasma haemofelis</i> str. Langford 1
GCA_000203215.1	<i>Mycoplasma suis</i> KI3806
GCA_000209735.1	<i>Mycoplasma fermentans</i> PG18
GCA_000211295.1	<i>Mycoplasma hyorhinis</i> MCLD
GCA_000211545.6	<i>Mycoplasma gallisepticum</i> S6
GCA_000218525.2	<i>Mycoplasma ovipneumoniae</i> SC01
GCA_000219375.1	<i>Mycoplasma bovis</i> Hubei-1
GCA_000221305.2	<i>Mycoplasma anatis</i> 1340

GCA\_000222995.2 *Mycoplasma columbinum* SF7  
 GCA\_000224105.1 *Mycoplasma putrefaciens* KS1  
 GCA\_000227355.2 *Mycoplasma iowae* 695  
 GCA\_000238995.1 *Mycoplasma haemocanis* str. Illinois  
 GCA\_000241125.1 *Mycoplasma hyorhinitis* GDL-1  
 GCA\_000258925.1 *Mycoplasma canis* PG 14  
 GCA\_000258945.1 *Mycoplasma canis* UF31  
 GCA\_000258965.1 *Mycoplasma canis* UF33  
 GCA\_000258985.1 *Mycoplasma canis* UFG1  
 GCA\_000259005.1 *Mycoplasma canis* UFG4  
 GCA\_000266865.1 *Mycoplasma agalactiae* 14628  
 GCA\_000270525.1 *Mycoplasma bovis* HB0801  
 GCA\_000277795.1 *Mycoplasma wenyonii* str. Massachusetts  
 GCA\_000281235.1 Candidatus *Mycoplasma haemolamae* str. Purdue  
 GCA\_000283755.1 *Mycoplasma pneumoniae* 309  
 GCA\_000286675.1 *Mycoplasma gallisepticum* VA94\_7994-1-7P  
 GCA\_000286695.1 *Mycoplasma gallisepticum* NC95\_13295-2-2P  
 GCA\_000286715.1 *Mycoplasma gallisepticum* NC96\_1596-4-2P  
 GCA\_000286735.1 *Mycoplasma gallisepticum* NY01\_2001.047-5-1P  
 GCA\_000286755.1 *Mycoplasma gallisepticum* WI01\_2001.043-13-2P  
 GCA\_000292405.1 *Mycoplasma genitalium* M2321  
 GCA\_000292445.1 *Mycoplasma genitalium* M6282  
 GCA\_000292485.1 *Mycoplasma genitalium* M6320  
 GCA\_000292505.1 *Mycoplasma genitalium* M2288  
 GCA\_000313635.1 *Mycoplasma hyorhinitis* SK76  
 GCA\_000319365.1 Candidatus *Mycoplasma haemominutum* 'Birmingham 1'  
 GCA\_000319655.2 *Mycoplasma pneumoniae* PO1  
 GCA\_000319675.2 *Mycoplasma pneumoniae* PI 1428  
 GCA\_000327395.1 *Mycoplasma feriruminatoris*  
 GCA\_000328725.1 *Mycoplasma cynos* C142  
 GCA\_000331085.2 *Mycoplasma pneumoniae* M129-B7  
 GCA\_000367185.1 *Mycoplasma flocculare* ATCC 27716  
 GCA\_000367445.1 *Mycoplasma alkalescens* 14918  
 GCA\_000367765.1 *Mycoplasma auris* 15026  
 GCA\_000367785.1 *Mycoplasma arginini* 7264  
 GCA\_000367805.1 *Mycoplasma bovigenitalium* 51080  
 GCA\_000376625.1 *Mycoplasma putrefaciens* Mput9231  
 GCA\_000380285.1 *Mycoplasma yeatsii* 13926  
 GCA\_000383515.1 *Mycoplasma hyorhinitis* ATCC 17981  
 GCA\_000385095.1 *Mycoplasma synoviae* ATCC 25204  
 GCA\_000387745.2 *Mycoplasma pneumoniae* 19294  
 GCA\_000400855.1 *Mycoplasma hyopneumoniae* 168-L  
 GCA\_000420105.1 *Mycoplasma orale* ATCC 23714  
 GCA\_000420225.1 *Mycoplasma moatsii* ATCC 27625  
 GCA\_000427215.1 *Mycoplasma hyopneumoniae* 7422  
 GCA\_000428625.1 *Mycoplasma arginini* ATCC 23838  
 GCA\_000433455.1 *Mycoplasma* sp. CAG:877  
 GCA\_000433475.1 *Mycoplasma* sp. CAG:956  
 GCA\_000434175.1 *Mycoplasma* sp. CAG:611  
 GCA\_000437315.1 *Mycoplasma* sp. CAG:472  
 GCA\_000438195.1 *Mycoplasma* sp. CAG:776  
 GCA\_000477415.1 *Mycoplasma parvum* str. Indiana

GCA_000485555.1	Mycoplasma salivarium ATCC 23064
GCA_000496815.1	Mycoplasma hyorhinis DBS 1050
GCA_000508245.1	Mycoplasma ovnis str. Michigan
GCA_000518305.1	Mycoplasma imitans ATCC 51306
GCA_000524555.1	Mycoplasma bovoculi M165/69
GCA_000526955.1	Mycoplasma cricetuli ATCC 35279
GCA_000620005.1	Mycoplasma spumans ATCC 19526
GCA_000621085.1	Mycoplasma gallinarum DSM 19816
GCA_000622165.1	Mycoplasma molare ATCC 27746
GCA_000622205.1	Mycoplasma leonicaptivi ATCC 49890
GCA_000685905.1	Mycoplasma pirum ATCC 25960
GCA_000686245.1	Mycoplasma pirum MPI25960
GCA_000686585.1	Mycoplasma lipofaciens ATCC 35015
GCA_000687775.1	Mycoplasma felifaucium ATCC 43428
GCA_000687795.1	Mycoplasma testudinis ATCC 43263
GCA_000687815.1	Mycoplasma elephantis ATCC 51980
GCA_000687855.1	Mycoplasma glycyphilum ATCC 35277
GCA_000691265.1	Mycoplasma hyosynoviae
GCA_000691285.1	Mycoplasma hyosynoviae
GCA_000691305.1	Mycoplasma hyosynoviae
GCA_000691325.1	Mycoplasma hyosynoviae
GCA_000691345.1	Mycoplasma hyosynoviae
GCA_000691365.1	Mycoplasma hyosynoviae
GCA_000691385.1	Mycoplasma hyosynoviae
GCA_000695835.1	Mycoplasma californicum
GCA_000696015.1	Mycoplasma bovis CQ-W70
GCA_000701485.1	Mycoplasma sturni DSM 22021
GCA_000701785.1	Mycoplasma alvi ATCC 29626
GCA_000701805.1	Mycoplasma iners ATCC 19705
GCA_000701825.1	Mycoplasma collis ATCC 35278
GCA_000701845.1	Mycoplasma columborale ATCC 29258
GCA_000701865.1	Mycoplasma felis ATCC 23391
GCA_000702705.1	Mycoplasma simbae ATCC 49888
GCA_000702745.1	Mycoplasma cloacale ATCC 35276
GCA_000702785.1	Mycoplasma primatum ATCC 25948
GCA_000702805.1	Mycoplasma anseris ATCC 49234
GCA_000712175.1	Mycoplasma columbinum ATCC 29257
GCA_000712185.1	Mycoplasma opalescens ATCC 27921
GCA_000723365.1	Mycoplasma amphoriforme A39
GCA_000733685.1	Mycoplasma iowae DK-CPA
GCA_000733865.1	Mycoplasma buteonis
GCA_000733995.1	Mycoplasma pneumoniae M29
GCA_000753815.1	Mycoplasma ovipneumoniae NM2010
GCA_000759375.2	Mycoplasma hominis
GCA_000767725.1	Mycoplasma hominis ATCC 27545
GCA_000770195.1	Candidatus Mycoplasma gireddii
GCA_000815065.1	Mycoplasma flocculare ATCC 27399
GCA_000828855.1	Mycoplasma canadense
GCA_000829335.1	Mycoplasma californicum HAZ160_1
GCA_000838165.1	Mycoplasma virus P1
GCA_000843645.2	Mycoplasma phage phiMFV1
GCA_000875755.1	Mycoplasma yeatsii GM274B

GCA_000941075.1	<i>Mycoplasma dispar</i>
GCA_000947915.1	<i>Mycoplasma hominis</i>
GCA_000965765.1	<i>Mycoplasma gallinaceum</i>
GCA_000969625.1	<i>Mycoplasma meleagridis</i> ATCC 25294
GCA_000969765.1	<i>Mycoplasma synoviae</i> ATCC 25204
GCA_000988065.1	<i>Mycoplasma canis</i>
GCA_001017595.1	<i>Mycoplasma hominis</i>
GCA_001043135.1	<i>Mycoplasma bovis</i>
GCA_001063305.1	<i>Mycoplasma hominis</i>
GCA_001199495.1	<i>Mycoplasma</i> sp. HU2014
GCA_001272715.1	<i>Mycoplasma pneumoniae</i> 39443
GCA_001484825.1	<i>Mycoplasma meleagridis</i>
GCA_001547975.1	<i>Mycoplasma arginini</i>
GCA_001553195.1	<i>Mycoplasma canis</i> PG 14
GCA_001611865.1	<i>Mycoplasma bovis</i>
GCA_001611895.1	<i>Mycoplasma bovis</i>
GCA_001611935.1	<i>Mycoplasma bovis</i>
GCA_001637235.1	<i>Mycoplasma gallinarum</i>
GCA_001639665.1	<i>Mycoplasma meleagridis</i>
GCA_001641205.1	<i>Mycoplasma</i> sp. Bg1
GCA_001641225.1	<i>Mycoplasma</i> sp. Bg2
GCA_001645765.1	Candidatus <i>Mycoplasma haemobos</i>
GCA_001705605.1	<i>Mycoplasma hyorhinis</i>
GCA_001715025.1	<i>Mycoplasma ovipneumoniae</i>
GCA_001715035.1	<i>Mycoplasma ovipneumoniae</i>
GCA_001715045.1	<i>Mycoplasma ovipneumoniae</i>
GCA_001715085.1	<i>Mycoplasma ovipneumoniae</i> ATCC 29419
GCA_001715095.1	<i>Mycoplasma ovipneumoniae</i>
GCA_001715105.1	<i>Mycoplasma ovipneumoniae</i>
GCA_001715115.1	<i>Mycoplasma ovipneumoniae</i>
GCA_001900245.1	<i>Mycoplasma pullorum</i>
GCA_001906845.1	<i>Mycoplasma hominis</i>
GCA_001917275.1	<i>Mycoplasma</i> sp. CAG:611_25_7
GCA_002009275.1	<i>Mycoplasma bovis</i>
GCA_002193015.1	<i>Mycoplasma hyopneumoniae</i>
GCA_002213485.1	<i>Mycoplasma hyopneumoniae</i>
GCA_002214445.1	<i>Mycoplasma hyosynoviae</i>
GCA_002215425.1	Candidatus <i>Mycoplasma girerdii</i>
GCA_002245785.1	<i>Mycoplasma testudineum</i>
GCA_002257505.1	<i>Mycoplasma hyopneumoniae</i>
GCA_002272945.1	<i>Mycoplasma agassizii</i>
GCA_002276225.1	<i>Mycoplasma agassizii</i>
GCA_002298905.1	<i>Mycoplasma</i> sp. UBA710
GCA_002356075.1	<i>Mycoplasma bovis genitalium</i>
GCA_002688685.1	<i>Mycoplasma bovirhinis</i>
GCA_002736285.1	<i>Mycoplasma dispar</i>
GCA_002952835.1	<i>Mycoplasma hominis</i>
GCA_003115795.1	<i>Mycoplasma arginini</i>
GCA_003147565.1	<i>Mycoplasma synoviae</i>
GCA_003208575.1	<i>Mycoplasma alkalescens</i>
GCA_003253435.1	<i>Mycoplasma auris</i>
GCA_003265155.1	<i>Mycoplasma wenyonii</i>

GCA_003269445.1	Mycoplasma cloacale
GCA_003285045.1	Mycoplasma anseris
GCA_003285065.1	Mycoplasma anatis
GCA_003332325.1	Mycoplasma phocidae
GCA_003383595.3	Mycoplasma phocicerebrale
GCA_003583385.1	Mycoplasma gallopavonis
GCA_003586085.1	Mycoplasma synoviae
GCA_003609735.1	Mycoplasma bovirhinis
GCA_003663725.1	Mycoplasma hominis
GCA_003688445.1	Mycoplasma subdolum
GCA_003704055.1	Mycoplasma fermentans MF-I1
GCA_003704075.1	Mycoplasma fermentans MF-I2
GCA_003855455.1	Mycoplasma sp. 237IA
GCA_003925935.1	Mycoplasma californicum
GCA_003925955.1	Mycoplasma californicum
GCA_003925975.1	Mycoplasma californicum
GCA_003967215.2	Mycoplasma californicum
GCA_003967235.2	Mycoplasma californicum
GCA_004127945.1	Mycoplasma penetrans
GCA_004216495.1	Mycoplasma phocirhinis
GCA_004335975.1	Mycoplasma sp. PE
GCA_004335995.1	Mycoplasma sp. 5H
GCA_004362335.1	Mycoplasma testudineum
GCA_004365095.1	Mycoplasma mustelae
GCA_004365165.1	Mycoplasma hyosynoviae

### Supplementary File 2

Motif M200535, suggested to be responsible for Galactocerebroside biosynthesis, as Hidden Markov Model. The model of motif M200535 in HMM format is available in the git repository <https://gitlab.com/erika.gaspariwur/phd-thesis-erika-gaspari>.

### Supplementary File 3

Result of HMM search of motif M200535 in *Campylobacter jejuni*. Available in the git repository <https://gitlab.com/erika.gaspariwur/phd-thesis-erika-gaspari>.

### Supplementary File 4

Consensus sequences for motifs M100535 and M200535.

*M200535*

```
XXXXXXXXXX XXXXXXXXXXXX XXXXXXXXXXXX XXXXXXXXXXXX XXXXXXXXXXXX XXXXXXXXXXXX XXXXXXXXXXXX XXXXXXXXXXXX
XXXXXXXXXX XXXXXXXXXXXX XXXXXXXXXXXX XXXXXXXXXXXX XXXXXXXXXXXX XXXXXXXXXXXX XXXXXXXXXXXX XXXXXXXXXXXX
```



### **Patent application EP20174842.3**

Title: Safe, attenuated *Mycoplasma* strains

Field: The invention relates to the field of infectious diseases and the prevention thereto. The invention relates to methods for the generation of a live, attenuated *Mycoplasma* strain that does not induce Guillain-Barré syndrome, and to the use of such strain, inter alia to protect against *Mycoplasma* infections.

Abstract: The invention relates to a *Mycoplasma* strain lacking a M200535 domain (SEQ ID NO:1)-containing galactosyl-ceramide synthase with EC 2.4.1.47 and lacking an UDP-glucose 4-epimerases with EC 5.1.3.2. The invention further relates to an immunogenic composition comprising said *Mycoplasma* strain and a pharmaceutically or veterinary acceptable excipient; and to methods of preventing a disease caused by a *Mycoplasma* species in a mammal, comprising administering an immunizing amount of the *Mycoplasma* strain.





# Chapter 5

Mathematical models of  
biosafety circuits designed to limit  
*Mycoplasma* growth

**Erika Gaspari**, Alicia Broto-Hernandez, Robert W. Smith, Mark Isalan, Vitor A.P. Martins dos Santos and Maria Suarez-Diez



## Abstract

Biocontainment and biosafety strategies are becoming more and more common as genetically modified organisms (GMOs) are produced at large-scale and deployed in open environments. Undesired proliferation of GMOs can be prevented through the application of these strategies, for example by inserting biosafety circuits into the organism's genome. Here, we mathematically and computationally explore several candidate biosafety circuits, of different entities, sizes and scopes, for limiting *Mycoplasma pneumoniae* growth, a bacterium deployed as GMO in vaccines and other pharmaceutical, medical and biotechnological applications. Ordinary differential equations models were used to investigate on circuit's efficacy, feasibility and possible appearance of escape mutants, and to predict its long-term effect on the bacterium. Our analysis unveils key design principles that need to be implemented for successful biocontainment of *M. pneumoniae* growth.

## Introduction

Since the 1970s, when the first recombinant-DNA technology was developed, Genetically-Modified Organisms (GMOs) have been more and more deployed as cell factories in large-scale production of biotherapeutic, health and food products, as well as in agriculture and other bioengineered systems aiming at producing biofuels, chemicals, biomaterials etc.<sup>1-3</sup>. The broad usage synthetic biologists make of GMOs can be extremely beneficial, but risk management strategies (such as biocontainment) are required to prevent accidental proliferation in the environment and damage to natural ecosystems<sup>4</sup>. Synthetic biology is a discipline that involves the redesigning of these organisms to create new useful functions, but also ensures their safety through proper risk management. Risks on humans related to Synthetic Biology have been identified and categorized into allergies, antibiotic resistance, carcinogens and pathogenicity<sup>5</sup>. For this reason, several biosafety strategies have been developed over the years, mostly consisting of suicide systems<sup>6,7</sup>, artificial introduction of auxotrophies for essential natural compounds<sup>8</sup> or non-natural ones<sup>9</sup>, or a combination of the two<sup>10,11</sup>. The existing microbial biosafety devices with their strengths and weaknesses, have been categorized by Wright et al. into five categories: toxin-antitoxin systems, other toxin systems (e.g. using restriction enzymes), essential gene complementation, miscellaneous systems and conditional origins of plasmid replication<sup>12</sup>. The aim of biocontainment is to protect against escape mechanisms such as mutagenic drift<sup>13</sup>. Most of the existing implemented strategies fail in providing total protection<sup>14</sup>: killing switches might undergo deactivation of the toxic product or of the circuit itself through the raise of escape mutants, and metabolic auxotrophies can be circumvented by hunting of the essential metabolites from the environment<sup>15,16</sup>.

Mycoplasmas are a group of bacteria belonging to the Mollicutes class. Their main characteristic, besides the small genome, is the absence of cell wall, for which they are considered biological study models. Mycoplasmas have raised interest in recent years for their suitability in biotechnological and pharmaceutical applications, such as vaccines, as indicated by the numerous publications reporting mycoplasmas attenuation and genome-editing<sup>17-23</sup>. Moreover, in 2008, Gibson et al. successfully assembled the first synthetic bacterial genome by inserting a minimal set of genes into DNA-depleted *M. genitalium* cells, obtaining the so-called *Mycoplasma laboratorium*<sup>24</sup>. The possibility of horizontal gene transfer between mycoplasmas has been long investigated and evidences have been reported in species located in the same ecological site<sup>23</sup> and in mycoplasmas of the same species<sup>25</sup>.

The growing synthetic biology applications on mycoplasmas<sup>26-29</sup>, especially regarding genome-engineering and genome-transplantation<sup>30,31</sup>, raise the need of effective and efficient biocontainment strategies. Here, we present the mathematical exploration through dynamic and regression modelling approaches of several candidate single- or multi-layer biosafety circuits inserted, into *M. pneumoniae* genome, with the aim of limiting its growth by i) targeting key survival genes to inhibit cell division and ii) inducing suicide of the bacterium by killing switches.

All the circuits herein described act at the transcriptional level. Players of these circuits involve the Cas9 endonuclease, a DNA-binding protein with remarkable modularity whose usage in cell death programming has already been reported<sup>32</sup>, the LacI repressor, a transcription factor that regulates genes in response to lactose or lactose analogues, and the TetR repressor<sup>33</sup>, regulating the expression of the tetracycline pump, a key determinant of bacterial resistance against tetracycline antibiotics<sup>34</sup>. LacI and TetR have been widely used in orthogonal circuits for safety applications in Synthetic Biology because they can be modulated by external elements having limited inference with the chassis: these factors are indeed expressed independently from the native cell transcription machinery and are, consequently, uncoupled from evolutionary constraints. Moreover, the conversion of DNA information into proteins is executed entirely by the orthogonal circuit's components, functionally isolated from the endogenous expression machinery<sup>35,36</sup>. In *E. coli*, LacI expression was controlled by Isopropyl  $\beta$ -D-1-thiogalactopyranoside (IPTG) and TetR by anhydrotetracycline (aTc)<sup>37-40</sup>. Therefore, the external factors act as inducers of the biosafety circuits, enforcing biocontainment. However, biosafety circuits are continuously subjected to mutations each time the microbe where they are inserted divides. This means that a microbial population starts including so-called "escape mutants", which are not killed by the activation of the biosafety circuit. After a certain number of divisions, escape mutants will take over, creating a biosafety circuit-resistant population. Single-component kill switches generally fail in the long term due to point mutations inactivating the killing mechanism, which generally occur at frequencies of  $10^{-3}$  to  $10^{-7}$  per cell per hour<sup>41</sup>. For this reason, multiple copies of a circuit must be taken into consideration.

All the highlighted risks suggest the need for an integral approach to safety, or "Safety-by-Design"<sup>42-44</sup>, which requires modelling and scenario simulations. This is because it is not enough to correct safety issues in an ongoing process: they must be anticipated, predicted and inserted into the design of the product itself at a very early stage.

### Mathematical models of biosafety circuits

Quantitative epidemiological models have been used in the past for evaluating biosecurity strategies, extrapolating formulas to describe how infectious disease parameters interact in populations<sup>45-47</sup>. Mathematical models such as Ordinary Differential Equations (ODEs) models are determinant when studying biocontainment strategies to predict the long term effect of circuits, with a special focus on the arising of escape mutants, i.e. cells developing mutations that make them resistant to the circuit, the time it takes before they take over the non-resistant cell population and the consequence on the biosafety of the vaccine. These mutants will indeed show no death when the biosafety circuit inducers are inserted into the medium.

Here, we present the mathematical modelling of selected biosafety circuits inserted into *M. pneumoniae*. The models were calibrated and constructed on available experimental data taken by partners (*Imperial College London*). Several circuits were inserted into *M. pneumoniae* genome through mini-transposon vectors, with the aim of controlling bacterial growth by inhibiting cell division and induce death when *Mycoplasma* was growing outside the desired conditions. According to the issue to address, different types of data were used in the

parametrization of the models. When studying conformation of biosafety circuits, characteristics and dose-responses of inducer concentration–protein expression or inducer concentration–bacterial growth, two types of data were used: the monitoring of the expression of circuit components through the mCherry reporter fluorescence and the concomitant bacterial growth curves, which are expressed in terms of rate of absorbances ABS430/ABS560 over time. These absorbances measure pH change: while the cell population grows, the medium acidifies and the ABS430 increases. Data were taken for different circuit configurations and inducer concentrations. The study of the arising of escape mutants within the cell population over time, with consequent inactivation of the circuits by the bacterium through spontaneous mutations, was performed with a mathematical model calibrated on bacterial growth curves when the killing switch was inserted in the genome, under the assumption that only mutated cells would be viable in presence of the circuit’s inducers.

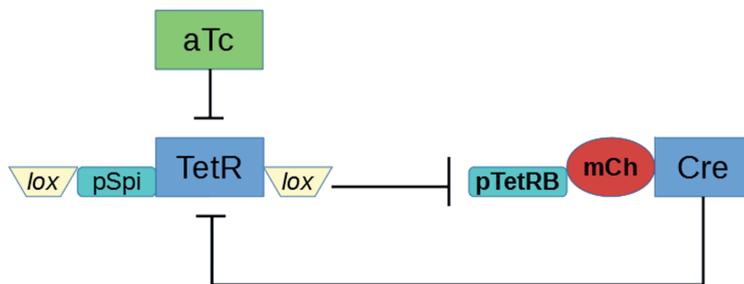
We used these models to investigate their potential, to properly account for the dynamics of the circuits described, and to identify key design principles for a biocontainment strategy for the industrial application of this pathogen.

## Results

Mathematical models of different entities have been developed for several biosafety circuits/killing switches to: a) Investigate on the interaction between the elements of a TetR-Cre killing switch implementing different promoters; b) study the arising and the impact of escape mutants on clonal and non-clonal populations carrying a LacI-Cas9 killing switch; c) simulate growth curves over clonal strain passages to highlight the importance of circuit's redundancy; d) provide an exploration on the functioning of a full biosafety network, implementing three different concatenated and interdependent circuits. According to the issue to address, different experimental data were used to calibrate the models (Materials and methods).

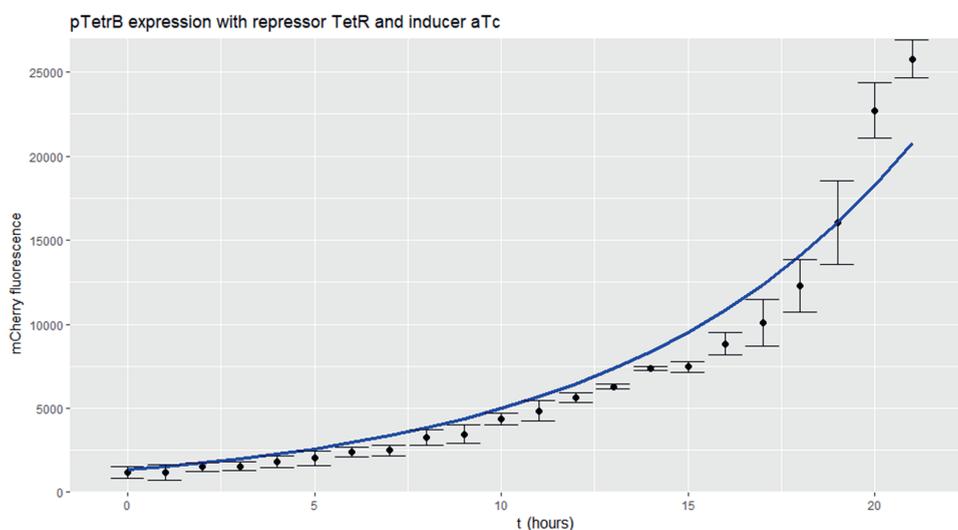
### Dynamic modelling of the TetR-Cre killing switch for different pTetR promoters

A TetR-Cre killing switch for leading *M. pneumoniae* to death when growing outside the desired conditions was investigated (Figure 1). The circuit consists of: TetR (Tet Repressor Protein), constitutively transcribed by *M. pneumoniae* and controlled by promoter pSpi, functions as a homodimer, with each monomer having a DNA-binding domain. It inhibits the Cre recombinase which carries out site-specific recombination between two DNA recognition sites (lox sites). TetR is inserted between lox sites and, therefore, inhibited by Cre, which acts through excision, cutting the DNA at the position of the lox sites. When aTc is added to the medium, TetR is inactivated: this leaves Cre and the reporter mCherry free to express, as the inhibition by TetR is not occurring in presence of the aTc inducer. Cre and mCherry expression are controlled by promoter pTetRB. The killing entity of this circuit is given by the substitution of mCherry with a toxic element, leading the cells to death.



**Figure 1.** Scheme of a killing switch candidate to be inserted in *M. pneumoniae*'s genome. TetR is inserted between lox sites and its expression is governed by the promoter pSpi. It is inactivated by the addition of aTc to the medium. TetR inhibits Cre and mCherry reporter expression, which are under the control of promoter pTetRB. Cre inhibits TetR expression through excision in the lox sites. When the mCherry reporter is substituted by a toxic element, the addition of aTc to the medium will lead to cell death.

The study of the TetR-Cre killing switch was approached through dynamic modelling, where simulations and step-by-step parameter estimation allowed the characterization of the interactions between the elements. Two promoters controlling mCherry and Cre expression were investigated: pTetRA and pTetRB. The comparison between the simulations of the circuit with the one or the other promoter ([Supplementary Figure S1](#)) shows pTetRB is more suitable for implementation, due to its stability and absence of leakiness. [Figure 2](#) reports the simulation of mCherry expression with the TetR-Cre model compared to experimental data over time (hours).

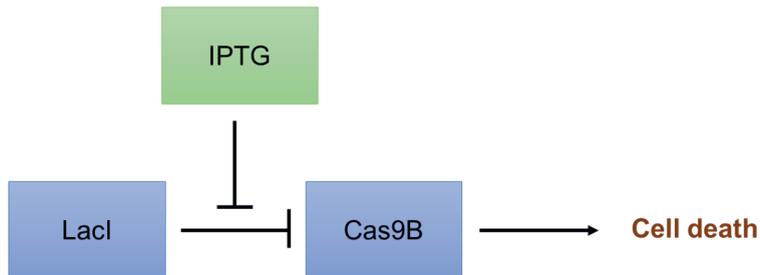


**Figure 2.** mCherry expression simulated with model (blue) of TetR-Cre killing switch implementing promoter pTetRB when 50 ng/ml aTc is added to the medium, repressing TetR expression and, therefore, activating the Cre recombinase. Model is compared to experimental data (black). Error bars represent standard deviation of 3 replicates .

### Investigation of escape mutations in *M. pneumoniae* clonal and non-clonal strains

A LacI-Cas9 killing switch was inserted into *M. pneumoniae*'s genome. The killing switch ([Figure 3](#)) consists of LacI (Lac repressor), constitutively expressed by the bacterium and inactivated IPTG, an analogous of allolactose. LacI inhibits the Cas9B nuclease. Expression of Cas9B nuclease and its complex with the guiding RNA leads to DNA scission and cell death. In this case the guiding RNA was designed to target 8 sites on DNA. When IPTG is added to the medium, Cas9B is not repressed and functions as a nuclease. When *M. pneumoniae* cells are pre-treated with IPTG (killing switch-inducer) in the previous passage, the next passage will show no difference in terms of growth curves in presence or absence of IPTG, indicating that all the cells growing have been mutated in the circuit's insertion site ([Supplementary Figure](#)

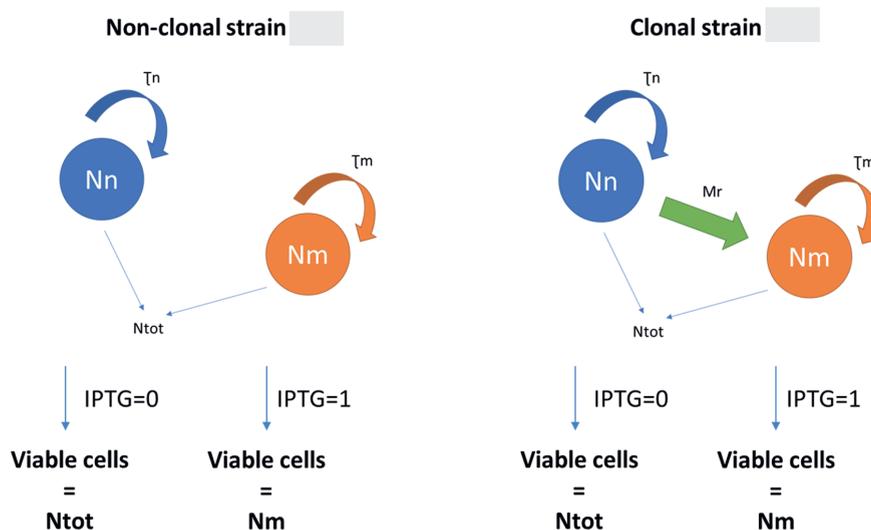
S2). To investigate this process, we modelled the two strain populations while growing at the passage when they are treated for the first time with IPTG.



**Figure 3.** LacI-Cas9 killing switch used to study the escape mutants present at the strain passage after IPTG pretreatment. LacI, constitutively expressed, inhibits the nuclease Cas9B, that cuts the DNA leading to cell death. The inhibition of LacI, and the consequent activation of Cas9, happens when IPTG is added to the medium.

Two strains were considered in the study: a non-clonal strain (nCS) and a clonal strain CS. In non-clonal strains, the circuit is inserted at unknown random position and positions will vary in daughter cells. On the contrary, clonal strains are generated from a selected cell, where the circuit is inserted at a specific position. The clonal strain population consists of cells that have inserted the circuit in the same position.

The escape mutant problem requires a population-level framework, largely independent of the circuit's specifications. In the model approach we assumed that, for each strain, two cell populations are monitored in the growth curve: a non-mutated cell population  $Nn$ , whose growth will be affected by IPTG addition, and a mutated cell population  $Nm$ , whose growth is not sensitive to IPTG. When IPTG is absent, the sum of the two populations  $Ntot$  is visible in the growth curve, while, when IPTG is present, only the mutated population will be visible (Figure 4).

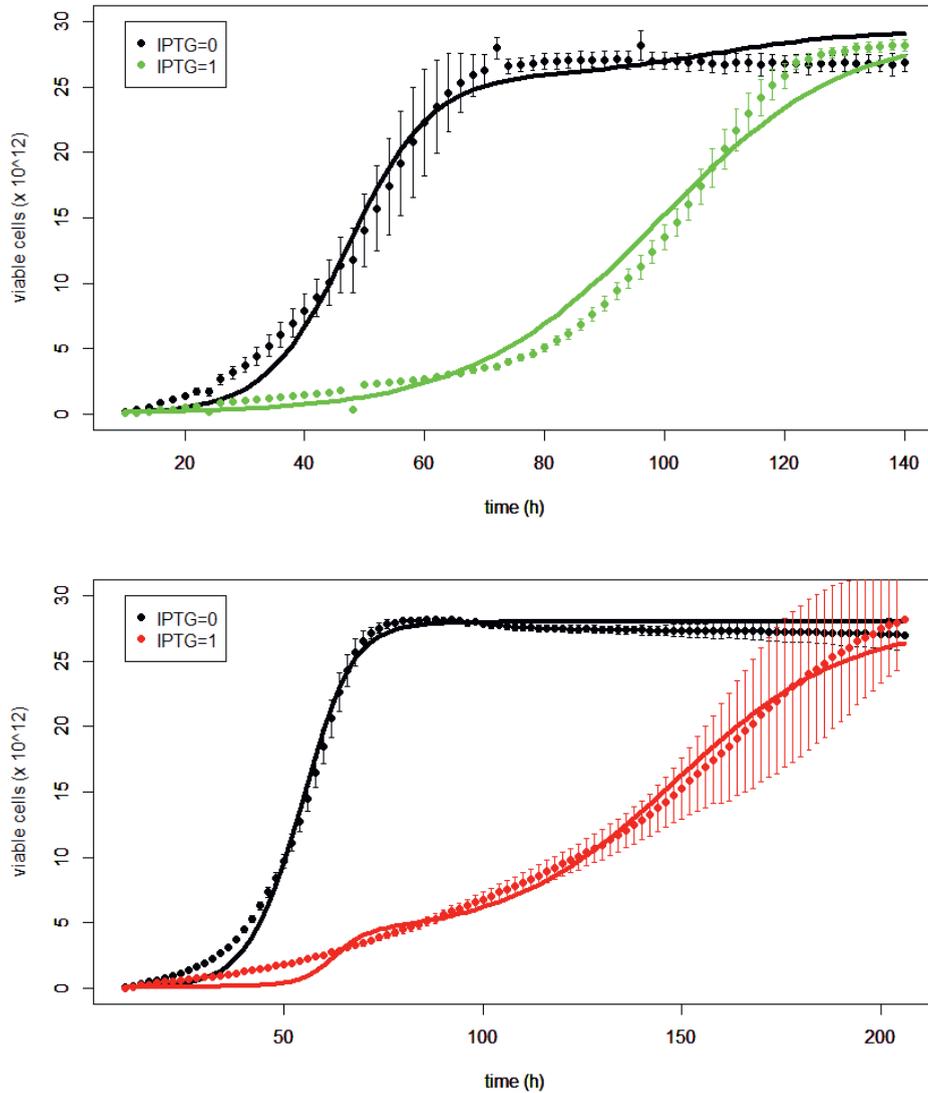


**Figure 4.** Schemes of the models for strains nCS and CS used to study the effect of mutations on the circuit, where  $N_n$ = non-mutated cell population;  $N_m$  = mutated cell population;  $N_{tot}$  = total cell population;  $M_r$  = mutation rate.  $N_n$  and  $N_m$  populations grow with division times  $\tau_n$  and  $\tau_m$ .

To explain the difference in growth between strain nCS and strain CS when IPTG is present, we assumed that for nCS growing in presence of inducer, all the mutated cells are already present in the beginning of the growth curve, co-existing with the non-mutated population. On the contrary, the clonal strain CS will accumulate mutations along the curve when IPTG is present, with a mutation rate  $M_r$ . These two different assumptions are due to the observation that, while for strain CS the growth rate is obviously different when IPTG is present or not, for nCS the growth rate does not change in presence or absence of the inducer and corresponds to  $10^{12}$  cells/h.

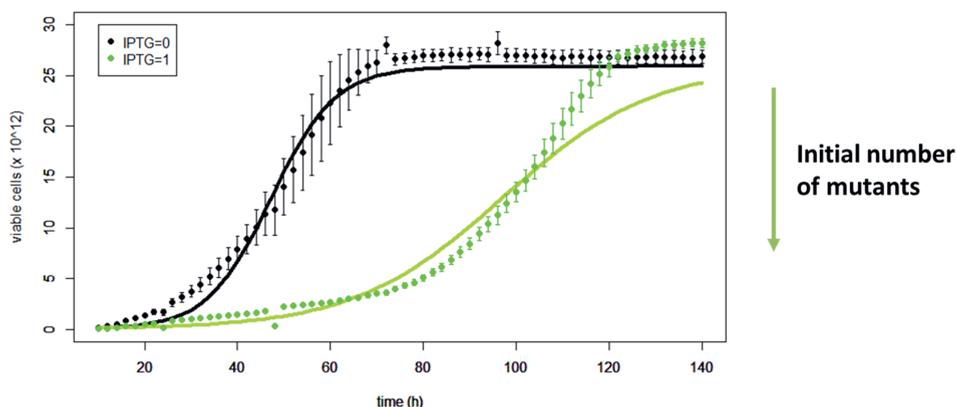
$M_r$  has been estimated as a parameter to correspond to a value of  $3.8 \cdot 10^{-11}$ . Comparison of measured data and model simulations for both nCS and CS, in presence or absence of the inducer IPTG, are shown in [Figure 5](#).

Initial values greatly change the time course, and small variations in the number of detected cells in the presence of IPTG greatly impact the initial hours, as shown in [Supplementary Figure S3](#), where increasing the initial mutant cell number to  $10^5$  represent a better fit to the experimental data for the first 72h of the mutant growth curve.



**Figure 5.** Model simulations for strains nCS (upper panel) and CS (bottom panel) compared to growth curves in presence or absence of IPTG, where dots represent average of experimental data (triplicates) with standard deviations and lines constitute the ODE models trained on the experimental data. In absence of IPTG, the total viable cells  $N_{tot}$  (black) are the sum of the non-mutated cell population and the mutated cell population. In presence of IPTG, viable cells are represented by the mutated cell population only (green for nCS, red for CS). While the growth rate varies for CS when IPTG is added or not, the growth rate does not vary within nCS curves in presence or absence of IPTG, confirming the correctness of the model assumption stating mutated cells must accumulate with a rate  $Mr$  along the CS curve, while the initial mutated cells are sufficient to explain the curve for nCS.

Simulations were done to simulate the effect of a potential strengthening of the circuit, assuming to start from a lower number of mutant cells. Figure 6 shows that a more efficient killing switch in the previous strain passage (i.e. leading to less starting mutant cells in the next passage) would lead anyway to a high number of escape mutants in the next passage, suggesting multiple copies of Cas9 or multiple switches acting on cell growth are needed to avoid *M. pneumoniae* to acquire resistance to the circuit at the short-term. This can be done by introducing multiple copies of the circuit or by implementing a biosafety network where multiple interdependent switches are present.



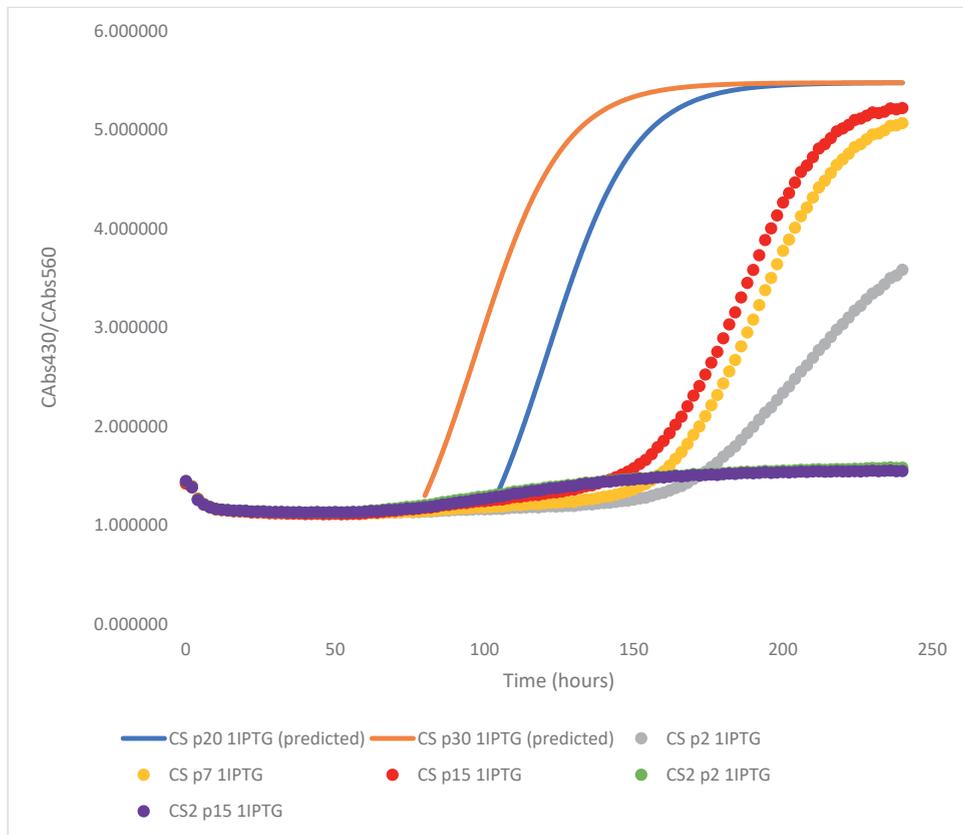
**Figure 6.** Simulation of the growth of *M. pneumoniae* non-clonal strain nCS in or not in presence of IPTG. The mutated cell population is visible when IPTG is added to the medium, due to lack of death in presence of the inducer. The model (line) was simulated assuming a lower number of accumulated mutants in the beginning of the growth rate. This would indicate an increased strength of the circuit when exploiting its function of killing switch. The model was compared to the data (dots) with the original growth rate. The study suggests multiple copies of Cas9 might be needed to avoid the escape mutants to nullify the effect of the killing switch.

5

### Modelling the impact on population dynamics of variations on circuit copy number

Growth curves for two clonal strains, CS and CS2, have been studied for three different strain passages in CS case (passages 2, 7 and 15) and for two passages in CS2 case (passages 2 and 15). The CS strain carries one copy of the LacI-Cas9 killing switch, while the CS2 strain carries two copies of the same circuit. The aim of the study is to predict the long-term effect of the circuit on *M. pneumoniae* growth, monitoring the arising of escape mutants at latest passages. The study was performed through regression models, applied to background-corrected data, where pH change-growth rate (ABS430/ABS560) was corrected for the effect of the medium evaporation; we refer to the background-corrected data as CAbs43/Cabs560. The choice of regression models is linked to the problem complexity as well as to the importance of the curve parameters along the strain passages. Models were built predicting the curve parameters over strain passages; the parameters considered were a) time to reach growth capacity, b) growth rate, c) maximum growth level and d) death rate. We define as time to reach growth

capacity the time required for the different strain passages to reach the necessary capacity that allows to enter in the exponential phase. For CS data, we assume the correspondent  $C_{abs430}/C_{abs560}$  (rate of absorbances after background correction) is 1.30, as shown in [Supplementary Figure S4](#). The duration of the lag phase reduces over and over with subsequent strain passages; for example, in passage 2 the strain with IPTG will start the exponential phase 116 hours after the strain without IPTG. However, along the passages the strain accumulates mutations: the same separation in passage 15 corresponds to 61 hours. The exponential phase of the strain with or without IPTG is predicted to start at the same time (80 hours) after 30 passages. [Supplementary Figure S5](#) shows the evolution of the time to reach growth capacity for strain CS. The unitless parameter describing the growth rate  $r$  remains constant over strain passages, with a value of 0.1366 when IPTG is absent and 0.0684 when IPTG is present. The carrying capacity  $K$  is the same for all CS data and equal to 4.1765. In the same way, the death rate  $dr$  is equal to 0.0041789. [Figure 7](#) shows the evolution curves over strain passages for strain CS in presence of IPTG, compared to data for CS2 (carrying two copies of the circuit), showing no accumulation of mutants along strain passages, confirming an increase in circuit copy number is effective against the inactivation of the biosafety strategy by escape mutants.

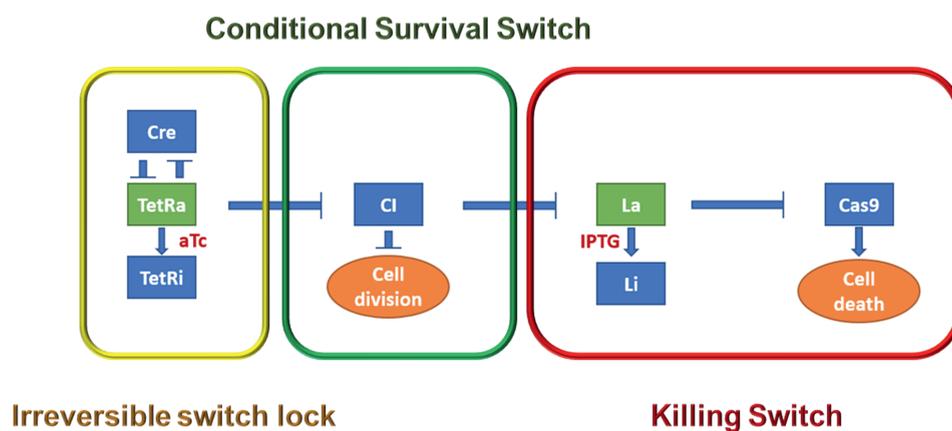


**Figure 7.** Simulation of the growth of *M. pneumoniae* clonal strain CS in presence of 1mM IPTG for strain passages 20 (green) and 30 (orange), compared to experimental data for CS strain passages 2 (grey), 7 (yellow) and 15 (blue), and CS2 strain passages 2 (purple) and 15 (red), overlapping between each other. Growth data are shown in terms of background-corrected rate of absorbance CAbs430/CAbs560 over time.

### General exploration of a multi-circuit biosafety network

A biosafety network (Figure 8), based on inducible promoter systems that modulate gene expression at a transcriptional level, was designed for limiting *Mycoplasma* growth *in vitro*. This biosafety network constitutes of three different but linked and interdependent synthetic submodules: a) an Irreversible Switch Lock, involving TetR, constitutively transcribed by *M. pneumoniae*. The Cre recombinase is and TetR mutually inhibit each other with the mechanisms previously explained for the TetR-Cre killing switch. When (anhdyrotetracycline) aTc is added to the medium, TetR will be inactivated, inducing Cre recombinase expression. Addition of aTc will irreversibly remove TetR cassette, causing cell death; b) Conditional Survival Switch, where the CI repressor, inserted in *M. pneumoniae*'s genome through a vector, represses key survival genes, inhibiting cell division by turning-off conditional survival

functions; c) Killing Switch, comprising Lacl, constitutively expressed, that silences Cas9, introduced through vector. Lacl inhibits Cas9, a nuclease that, when not repressed, will cut *M. pneumoniae*'s double-strand break leading to cell death. The addition in the medium of IPTG inactivates Lacl, triggering Cas9 activity, with the mechanisms previously explained for the Lacl-Cas9 killing switch.



**Figure 8.** The candidate configuration studied of the biosafety network, composed by three circuits: an irreversible switch lock (yellow box), where Cre inhibits TetR (constitutively transcribed), that mutually inhibits Cre. Addition of aTc to the medium, inactivates TetR, leading downstream to cell death. In a second circuit, called conditional survival switch (green box), Cl represses key survival genes, inhibiting cell division. Finally, the third circuit, or killing switch (red box), consists of Cas9 activation that leads to cell death. Cas9 is repressed by Lacl (L), constitutively expressed in its active form. (La), but when IPTG is added, Lacl is inactivated (Li), leading to the expression of Cas9.

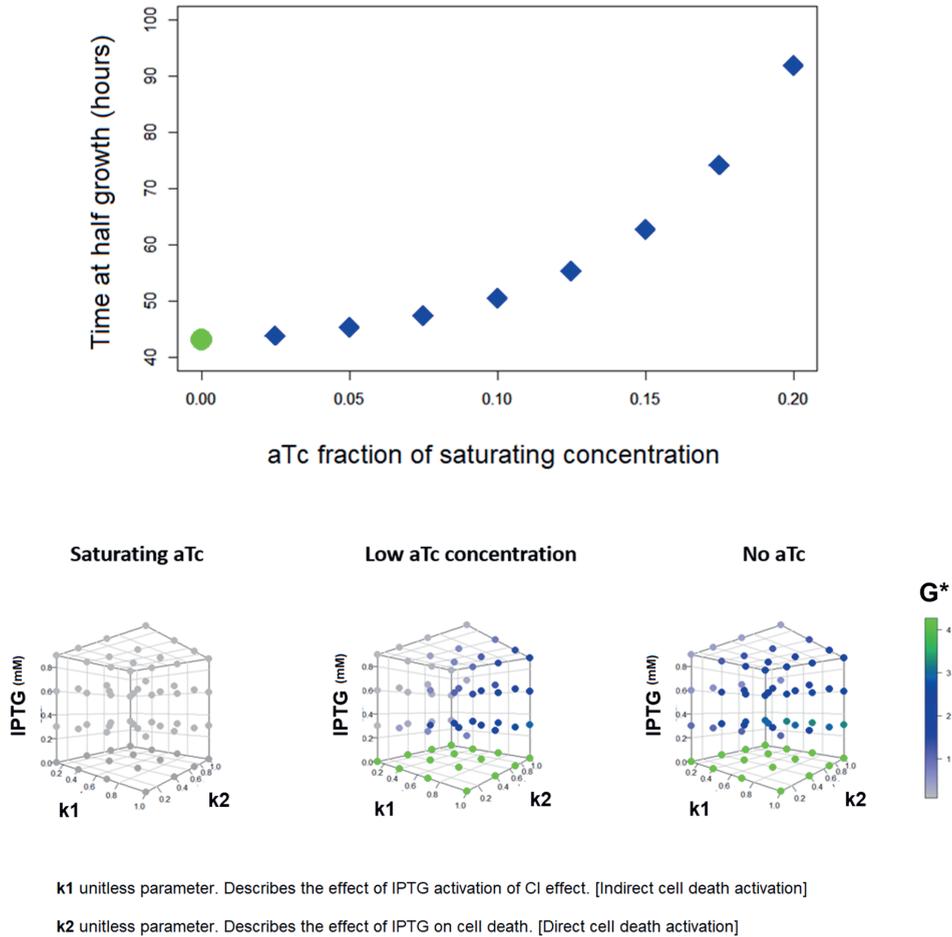
The configuration of the network provides multiple layers of biosafety, which allows to prevent growth of the strain outside the defined *in vitro* conditions, and interdependency between elements of different circuits, to achieve 100% killing and apply genetic barriers *in vivo*.

A simplification of this biosafety network where Lacl was pruned ([Supplementary Figure S6](#)) has been converted into an Ordinary Differential Equations system to simulate its dynamics, compute inducers dose-responses curves and investigate on the effects of the different promoter strengths. When designing the model, we assumed the bacterial cell growth  $G$ , at the population level, is only affected by growth delay and cell death. The model describing the experimental data trend is shown in [Supplementary Figure S7](#).

Simulations of the circuit when induced by aTc suggest that the  $t_{1/2}$  growth exponentially increases with the aTc concentrations. aTc concentrations are indicated as fractions of saturating aTc concentration, in a range 0-1. When adding only aTc to the medium, the circuit

shows an effect on growth delay but not on maximum growth level (curve's steady state); indeed, the horizontal asymptote for stationary phase remains the same of the wild type, as shown in [Supplementary Figure S8](#). Further simulations for three aTc concentrations (saturating, low and zero) show the addition of both inducers (aTc and IPTG) to the medium. Model predictions indicate which growth level was reached for different inducers' concentrations, where  $G^*$  is the maximum reachable growth level, i.e. the carrying capacity (horizontal asymptote in stationary phase) of *M. pneumoniae*'s growth when the circuit is not inserted in the genome. Other factors considered in the simulations are the affinity for promoters: experimental data are insufficient for estimating the related parameters; therefore, simulations were done considering parameters' affinity strength in a range 0-1. Examples of simulations the biosafety network are reported in [Figure 9](#).

Results show that a) the growth delay increases non-linearly with the relative aTc concentration, b) aTc does not affect the curve steady state, c) growth levels are maximized when  $k1$  and  $k2$  are high (weak promoters), d) IPTG buffers the effect of aTc on cell death.



**Figure 9.** Simulations of the effects that the insertion of a biosafety circuit into *M. pneumoniae* genome would have on its maximum growth ( $G^*$ ). Top panel: effect of the different concentration of aTc added to the medium on the time at half maximal growth ( $t_{1/2}$ ). Concentrations are indicated as fractions of saturating concentrations. The growth delay is predicted to be dependent on the aTc concentration with an exponential trend. Bottom panel: overview of the combination of effects of different aTc concentrations (saturating, low – 10% of saturating, and zero), IPTG concentration and promoter affinities ( $k_1$ ,  $k_2$ ) have on the growth of *M. pneumoniae*, comparing to the maximum reachable growth ( $G^*$ , i.e. when the circuit is not inserted).

## Discussion

Over the last two decades, mycoplasmas have been providing access to enormous basic research opportunities and development of novel approaches in biotechnology and biomedicine<sup>48-52</sup>. Their growing applications, for example as live-attenuated vaccines, lead to the need of efficient and safe biocontainment strategies. In this study, we have investigated numerous biosafety circuits, each providing useful leads on design principles to adopt for successfully limiting the growth of mycoplasmas outside desired conditions. The complexity of the designs and the laborious cultivation of *M. pneumoniae*, owing to its slow growth rate, make mathematical modelling approaches particularly suitable to study these issues, allowing a computational exploration of the predicted dose-responses and long-term effects of the circuits.

On studying the TetR-Cre killing switch, we have investigated the interaction between the circuit's elements and the impact of biological noise, here represented by promoter leakiness. Model and data analysis led us to select the promoter allowing less leakage. The model developed for studying the arising of escape mutants in clonal and non-clonal strains carrying the LacI-Cas9 killing switch focused on two cell populations: mutated and nonmutated. This model is largely independent of the details of the circuit. We have investigated the arising of escape mutants, induced *in vitro* by pre-treatment with the inducer, concluding a solution for obtaining an efficient biocontainment strategy over time is the duplication of the number of endonucleases, for example by introducing multiple copies of a killing switch in the bacterial genome, or to build a biosafety network made of different interdependent circuits, leading to both cell death and inhibition of cell division. Therefore, we have studied the long-term effect of the one or two copies of a killing switch on *Mycoplasma* clonal strains' growth over strain passages, predicting the passage at which the circuit is completely inactivated. The importance of multiple copies of the endonuclease is visible in the non-clonal strain nCS, carrying one copy of the LacI-Cas9 switch, which is easily inactivated by the bacterium after one inducer-pre-treatment passage. Although the clonal strain CS is more robust, our results clearly show at least two copies of the circuit are needed, as implemented in strain CS2. Finally, we have performed a mathematical exploration of a multi-circuit biosafety network, consisting of three linked synthetic circuits triggering cell death and inhibition of the cell division. From this study we extracted dose-responses curves for circuit's inducers concentrations, suggesting the inducer of the killing switch-component of the network (IPTG) seems to buffer the effect of the inducer activating the conditional survival switch (aTc).

Promoter leakiness can also be a problem in biocontainment devices because it can give selective growth advantage to cells with mutated suicide functions<sup>15</sup>. Additionally, it has been reported that with cell density, promoter tightness can vary<sup>53</sup>. One of the circuits tested (TetR-Cre switch) shows the importance of this issue when designing a biosafety circuit, at the contrary of the LacI-Cas9 switch, which seems more robust in terms of promoter leakiness. In addition to promoter leakiness, the practical usefulness of kill switches can be influenced by induction capability and the ability of the inducer to reach all cells and activate gene expression. All the circuits we explored in this work rely on the usage of inducers, which is

common in killing/suicide switches<sup>54-56</sup>. Although the strain carrying two copies of the LacI-Cas9 killing switch gave robust results at the long-term (> 15 strain passages), at higher cell densities, reaching all cells may be a problem that can result in inconsistent induction. Furthermore, less inducer may be available per cell, also due to possible cell aggregation and biofilm formation.

Although the construction of effective biosafety mechanism is underway, the escape of even a single, errant bacterium could have widespread consequences, so it is crucial that these safety mechanisms be well explored. In this work, our attention has focused not only on the construction of the switches, but also on understanding their potential failure modes, through novel mathematical approaches. This concept follows the idea of Safety-by-Design, that supports the identification of risks as early as possible in the developmental phase.

## Conclusion

In this work, we have investigated in depth several candidate synthetic circuits to be inserted into *M. pneumoniae* genome for its efficient biocontainment. The combination of mathematical modelling with through experiments allows to extract information from the data to provide reliable predictions on inducers dose-responses and long-term effects of selected biosafety circuits on *Mycoplasma* growth. Our analysis allowed to make predictions on the circuit components and the whole biosafety device over time, reaching two main conclusions: an efficient biocontainment strategy to limit *Mycoplasma* growth *in vitro* must feature redundancy, multiple layers or a combination of the two. As for the moment, the best results in growth limitation are reached by a double killing switch carried by a clonal strain of *M. pneumoniae*.

## Materials and Methods

The TetR-Cre killing switch, the LacI-Cas9 killing switch with the effect of pre-treatments on clonal and non-clonal strains carrying the circuit, and the full biosafety network, made of three submodules, have been investigated through dynamic modelling, where the circuits were described through Ordinary Differential Equations (ODEs) systems, whose parameters have been estimated from available experimental data. All the below steps for model construction, parameter estimation, model validation, model simulation and visualization have been performed with tools of R 4.0.0<sup>57</sup> in RStudio suite, version 1.2.5042<sup>58</sup>. The ODE systems have been solved with the R package *deSolve*<sup>59</sup>. The equations' parameters have been estimated through Maximum Likelihood optimization with the command *mle2* of the R package *bbmle* v. 1.0.23.1<sup>60</sup> from experimental data, reported below for each model. Bacterial growth data are reported over time in terms of rate of absorbances ABS430/ABS560. These absorbances measure pH change: while the cell population grows, the medium acidifies and the ABS430 increases, describing the S-shaped curve that is typical of bacterial growth.

### TetR-Cre killing switch: mathematical model

The TetR-Cre killing switch is described through the following 3 ODEs:

$$\begin{aligned}\frac{dCh}{dt} &= b_{cre} \left( \frac{K_{cre}}{K_{cre} + TetR} \right) + esp_{cre} G, & [1] \\ \frac{dTetR}{dt} &= b_{tet} \left( \frac{K_{tet}}{K_{tet} + aTc} \right) + esp_{tet} G, & [2] \\ \frac{daTc}{dt} &= 0. & [3]\end{aligned}$$

Where *Ch* refers to the mCherry concentration, *TetR* to the concentration of TetR, *G* to the culture growth and *aTc* to aTc concentration. Being under the control of the same promoter, we assumed mCherry fluorescence describes the Cre recombinase's activity.

The first equation describes growth-dependent expression of mCherry that is repressed by TetR, following a Hill function with coefficients  $K_{cre}$   $b_{cre}$ , where  $b_{cre}$  describe mCherry's fluorescence increase rate and  $K_{cre}$  the inhibition of *TetR* respect to mCherry (and *Cre*) expression. The  $esp_{cre}$  parameter describes the effect of medium evaporation (leading to increase in fluorescence) and the element's decay. Although these processes should be described by two separate parameters, the entity of the training data does not make them identifiable; therefore, they were grouped under the same parameter. The second equation describes the evolution of *TetR* concentration, increasing with a rate described by  $b_{tet}$ , and inhibited by *aTc* under parameter  $K_{tet}$ .  $esp_{tet}$  describes medium evaporation's effect and element's decay. The last equation represents the assumption that *aTc* levels remain constant, therefore its change over time equals 0.

For the TetR-Cre killing switch, only the exponential phase of the growth curve was modelled and used in the parametrization. The dynamic model was built parametrizing circuit's

submodules one at a time. First, parameters for mCherry expression under promoter pTetRB were estimated ( $b_{cre}$ ,  $esp_{cre}$ ); equation [1], in this first estimation, lacks the part related to TetR inhibition ( $\frac{K_{cre}}{K_{cre}+TetR}$ ), as TetR has not yet been introduced to the circuit. Then, TetR inhibition was inserted *in silico* into the circuit to estimate the parameter  $K_{cre}$  and TetR initial concentration; in this step, equation [1] is complete, while equation [2] lacks aTc inhibition ( $\frac{K_{tet}}{K_{tet}+aTc}$ ). Finally, aTc-induction was added to the circuit, obtaining the complete ODE system and the remaining parameter values to be estimated:  $b_{tet}$ ,  $K_{tet}$ ,  $esp_{tet}$ .

### TetR-Cre killing switch: experimental data

The model was trained step-by-step, estimating the parameters of the ODE system one equation at a time from the experimental data, as reported in [Supplementary File S1](#). The data used for model training are provided in [Supplementary Table S1](#), where the circuit is implemented step-by-step, adding an element at a time: the first implementation of the circuit consists of aTc addition inhibiting TetR, which inhibits the mCherry expression. The second implementation step adds the Cre recombinase to the circuit, whose expression is controlled by the same promoter that controls mCherry. The third implementation inserts TetR between lox sites, excision points of the Cre recombinase; this last step completes the killing switch implementation, that only misses the substitution of mCherry with a toxic element to induce cell death. For each implementation step, data of ABS430/ABS560 (pH change-bacterial growth curve) and mCherry fluorescence  $Ch$  are reported using different promoters: p438 (constitutive, not inducible, used as control), pTetRA and pTetRB. For each promoter used, 0 ng/ml aTc or 50 ng/ml aTc have been added to the medium; data are reported in triplicates for each condition. pTetRB will be selected as promoter to implement in the killing switch. In this data, the reporter mCherry's fluorescence is indicative of the promoter activity. Being under the control of the same promoter, we assumed mCherry fluorescence describes the Cre recombinase's activity.

### Investigation on escape mutants with LacI-Cas9 killing switch: mathematical model

The LacI-Cas9 killing switch, induced by the addition of IPTG to the medium, was investigated to study the arise of escape mutants on a non-clonal strain nCS and a clonal strain CS, where non-clonal strains consist of cells where the biosafety circuit was randomly inserted at unknown position, while clonal strains have originated by the same cell, with a specific circuit position.

Models for nCS and CS have been built as ODEs systems. nCS is described by 3 ODEs with 5 parameters to be estimated from the experimental data:

$$\frac{dNn}{dt} = \mu_n Nn \left(1 - \frac{Nn}{S_n}\right) + \frac{ki}{ki+IPTG}, \quad [4]$$

$$\frac{dNm}{dt} = \mu_m Nm \left(1 - \frac{Nm}{S_m}\right), \quad [5]$$

$$\frac{dIPTG}{dt} = 0. \quad [6]$$

CS models is a system of 3 ODEs with 6 parameters to estimate from the experimental data:

$$\frac{dNn}{dt} = \mu_n Nn \left(1 - \frac{Nn}{S_n}\right) + \frac{ki}{ki+IPTG} - mrNn, \quad [7]$$

$$\frac{dNm}{dt} = \mu_m Nm \left(1 - \frac{Nm}{S_m}\right) + mrNn, \quad [8]$$

$$\frac{dIPTG}{dt} = 0. \quad [9]$$

In both models,  $Nn$  and  $Nm$  describe the evolution over time of, respectively, the non-mutated cell population and the mutated cell population. The sum of the two ( $Ntot$ ) is visible when IPTG is absent, while only the second population is visible when IPTG is present. For both strains, the initial inoculum  $Ntot$  was detected as  $5 \cdot 10^5$  cells. Initial value for  $Nm$  was experimentally measured as 3000 cells for nCS and as 4 cells for CS. All the cell population curves follow a logistic growth with carrying capacity  $S_n$  or  $S_m$  and growth rates  $\mu_n$  and  $\mu_m$ . The  $Nn$  population is inhibited by IPTG, while the  $Nm$  population is not. IPTG concentration does not change over time and it is added at initial value of 0 when absent and value of 1 when present.

### Investigation on escape mutants with LacI-Cas9 killing switch: experimental data

Experimental data used to estimate parameters for both CS and nCS consists of rate of absorbances (ABS430/ABS560), measuring the growth of the two *M. pneumoniae* strains over time. Data have been transformed into their natural logarithm and number of cells have been computed through equation [10]:

$$N_{cells}^F = e^{\frac{D \ln(2)}{Td}} + \ln(N_{cells}^I). \quad [10]$$

Where  $N_{cells}^F$  and  $N_{cells}^I$  are, respectively, the final and the initial number of cells,  $D$  is the duration of the growth in hours (206h for CS, 141h for nCS),  $Td$  the doubling time (8h). All the ABS430/ABS560 growth data have been then rescaled into the interval  $[N_{cells}^I; N_{cells}^F]$  with the R command *rescale* of package *scales*<sup>61</sup>. [Supplementary Figure S9](#) show experimental data in terms of ABS430/ABS560, provided in [Supplementary Table S2](#). This data reports the growth of both CS and nCS strains in terms of rate of absorbances ABS430/ABS560 over time when the LacI-Cas9 killing switch is inserted in *M. pneumoniae* genome. Data consists of rate of absorbances of the strains in absence or presence of the inducer; in this last condition, IPTG is added at a concentration of 1mM to the medium. Data are provided with mean and standard deviation, calculated on the basis of three replicates.

The growth rates of strain nCS in presence or absence of the inducer IPTG have been verified through regression with 5-parameters log-logistic model [11]:

$$f(t) = c + \frac{(d-c)}{(1+e^{b(\ln t - \ln h)})^f}. \quad [11]$$

With the growth rate  $\mu$  of the curve being the first derivative at the curve's inflection point  $h$ :

$$\mu = \left(\frac{df(t)}{dt}\right)_h. \quad [12]$$

Log-logistic regression and inflection point of the two curves (nCS with no IPTG, nCS with 1mM IPTG) are shown in [Supplementary Figure S10](#).

### **Background correction of ABS430/ABS560 experimental data**

An effect of the medium volume reduction in the culture well is observed in the detection of the rate of absorbances ABS430/ABS560 for the clonal strains data types. Specifically, the volume reduction, due to evaporation, affects mainly ABS560 detection, while the effect on ABS430 is minimal. This is visible in data consisting of the detection of ABS430/ABS560 in absence of the bacterium ([Supplementary Figure S11](#)). Linear regression was performed on ABS430/ABS560 data in absence of the bacterium through the R command *lm* for both ABS430 and ABS560, to capture the different entity of the effect of the volume reduction on the detection. Assuming both ABS430 and ABS560, in absence of the microorganism, should not change over time, the slope of the line obtained by linear regression was subtracted from the point-by-point mean of the original data, provided in triplicates. Slopes of the lines correspond to values of  $1^{-8}$  for ABS430 and  $-2^{-7}$  for ABS560. The rates of absorbances Cabs430/Cabs560 were recomputed with the background-corrected absorbances and standard deviations were recalculated.

The difference between ABS430/ABS560 and Cabs430/Cabs560 was used to correct the data for all the strain types, for each condition and time point. All the background corrected curves, compared to the original data, are provided in [Supplementary File S2](#).

### **Prediction of growth curves over strain passages for clonal strains CS and CS2: regression model**

The study was performed through regression models, applied to background-corrected data, where pH change-growth rate (ABS430/ABS560) was corrected for the effect of the medium evaporation; we refer to the background-corrected data as CAbs43/Cabs560. Models were built predicting the curve parameters over strain passages; the parameters considered were a) time to reach growth capacity, b) growth rate, c) maximum growth level and d) death rate. The R package *Growthcurver*<sup>62</sup> was used to extract  $N_0$  (population size at the beginning of the exponential growth),  $K$  (carrying capacity) and  $r$  (growth rate) parameters for each strain passage. The population growth  $N_t$  over time  $t$  for each strain passage is described by equation [13]:

$$N_t = \frac{K}{1 + \left(\frac{K - N_0}{N_0}\right)e^{-rt}} \quad [13]$$

The death rate was estimated through linear regression with the R command *lm*.

## Prediction of growth curves over strain passages for clonal strains CS and CS2: experimental data

All the growth curve's parameters were estimated for the clonal strain CS, carrying one copy of the circuit. Data used for model's calibration are provided in [Supplementary Table S3](#), where growth curves (background-corrected rate of absorbances A430/A560) are measured for strain passages 2, 7 and 15, each in absence/presence of IPTG (1mM when present) and 274 hours. Data for clonal strain CS2, carrying two copies of the killing switch, are reported in the same file. However, the limited number of passages in which the growth curves were measured (passage 2 and 15) made it impossible to estimate the curve's parameters.

## Multi-circuits biosafety network: mathematical model

Each component's development over time was described with an ODE and is assumed to follow Hill kinetics. We assumed transcription time is negligible respect to translation time and that the mRNA levels correlate with the protein's ones, therefore the model describes the protein concentration and translational evolution of each component. An additional ODE was added to the system to describe the bacterial growth over time. The system consists of six ODEs with 16 parameters, of which 15 must be estimated from the data, while one is known (the horizontal asymptote  $G_{max}$ , the maximum growth level reached by the wild type strain):

$$\frac{dCre}{dt} = b_{Cre} \left( \frac{K_{Cre\_Tet}^4}{K_{Cre\_Tet}^4 + TetRa^4} \right) - d_r Cre, \quad [14]$$

$$\frac{dTetRa}{dt} = b_{TetRa} \left( \frac{K_{Tet\_Cre}}{K_{Tet\_Cre} + Cre} \right) - atc TetRa - d_r TetRa, \quad [15]$$

$$\frac{dTetRi}{dt} = atc TetRa - d_r TetRi, \quad [16]$$

$$\frac{dCI}{dt} = b_{CI} \left( \frac{K_{CI\_Tet}^4}{K_{CI\_Tet}^4 + TetRa^4} \right) - d_r CI, \quad [17]$$

$$\frac{dIPTG}{dt} = 0, \quad [18]$$

$$\frac{dG}{dt} = \mu_{max} \left\{ \left[ \left( \frac{k_1(k_3 + IPTG)}{k_1(k_3 + IPTG) + IPTGCI} \right) \left( \frac{k_2}{k_2 + IPTG} \right) \right] - \left( \frac{G}{G_{max}} \right)^m \right\} \left( \frac{K_{G\_CI}}{K_{G\_CI} + CI} \right)^n G. \quad [19]$$

5

The decay rate  $d_r$  was estimated starting from a value determined from the repressilator's protein half-lives as  $d_r = \ln(2) / t_{1/2}$ <sup>38</sup>, to extract the correct order of magnitude of this parameter, which otherwise would be hardly identifiable. We assumed in the time of observation all the circuit's elements decay with the same rate. In this ODE system, the components' concentrations of the simplified circuit are indicated as Cre, TetRa (active TetR), TetRi (inactive TetR), CI and IPTG. The parameter  $atc$  represents the strength of aTc effect, equal to the fraction of saturating aTc concentration added;  $b_{Cre}$ ,  $b_{TetRa}$ ,  $b_{CI}$  are, respectively, the maximum production rates of Cre, TetR and CI;  $K_{Cre\_Tet}$ ,  $K_{Tet\_Cre}$  and  $K_{CI\_Tet}$  are the Michaelis-Menten constants for respectively Cre, TetR and CI kinetics. According to the simplification, the assumption that the growth is only dependent on growth delay and cell death leads to only two circuit components directly affecting the growth, G, through a coherent feed-forward

loop: IPTG and CI. The growth equation (eq. 19) is a modification of the Gompertz model.<sup>63,64</sup> In the equations,  $\mu_{max}$  indicates the maximum growth rate,  $G_{max}$  the maximum carrying capacity,  $m$  and  $n$  are the Hill or curvature parameters. Hill coefficients are fixed to 4 for TetR repression as TetR acts couple of homodimers. While the cell death is affected by both IPTG and CI under the parameters  $k_1$ ,  $k_2$  and  $k_3$ , the growth delay is only directly influenced by CI under the parameter  $K_{G,CI}$ .

### **Multi-circuits biosafety network: experimental data**

Parameters of the ODE system were estimated from experimental data consisting of rate of absorbances ABS430/ABS560 over time. *M. pneumoniae* growth data points used to train the model are reported in [Supplementary Table S4](#) as mean of 3 replicates. The modelling procedure must be considered as semi-quantitative as the ODE system has been trained on data representing the growth curve of the bacterium with the inserted biosafety network, but without addition of the inducers. In consequence, many parameters are expected to be non-identifiable.

### **Funding**

This work has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement n. 634942 (MycoSynVac) and from the Dutch TTW-NWO under agreement n. 15814 (SafeChassis).

## References

1. Moe-Behrens, G., Davis, R. & Haynes, K. Preparing synthetic biology for the world. *Frontiers in Microbiology* **4**, 5 (2013).
2. Menozzi, D. *et al.* A stakeholder engagement approach for identifying future research directions in the evaluation of current and emerging applications of GMOs. *Bio-based and Applied Economics* **6**, 57–79 (2017).
3. Halecký, M. & Kozliak, E. Modern Bioremediation Approaches: Use of Biosurfactants, Emulsifiers, Enzymes, Biopesticides, GMOs. in *Advanced Nano-Bio Technologies for Water and Soil Treatment* (eds. Filip, J., Cajthaml, T., Najmanová, P., Černík, M. & Zbořil, R.) 495–526 (Springer International Publishing, 2020). doi:10.1007/978-3-030-29840-1\_24.
4. Nordmann, B. D. Issues in biosecurity and biosafety. *International Journal of Antimicrobial Agents* **36**, S66–S69 (2010).
5. Hewett, J. P., Wolfe, A. K., Bergmann, R. A., Stelling, S. C. & Davis, K. L. Human health and environmental risks posed by synthetic biology R&D for energy applications: A literature analysis. *Applied Biosafety* **21**, 177–184 (2016).
6. Molin, S. *et al.* Conditional Suicide System for Containment of Bacteria and Plasmids. *Bio/Technology* **5**, 1315–1318 (1987).
7. Li, Q. & Wu, Y.-J. A fluorescent, genetically engineered microorganism that degrades organophosphates and commits suicide when required. *Applied Microbiology and Biotechnology* **82**, 749–756 (2009).
8. Curtiss III, R. Biological Containment and Cloning Vector Transmissibility. *The Journal of Infectious Diseases* **137**, 668–675 (1978).
9. Mandell, D. J. *et al.* Biocontainment of genetically modified organisms by synthetic protein design. *Nature* **518**, 55–60 (2015).
10. Ronchel, M. C. & Ramos, J. L. Dual System To Reinforce Biological Containment of Recombinant Bacteria Designed for Rhizoremediation. *Applied and Environmental Microbiology* **67**, 2649 (2001).
11. Wright, O., Delmans, M., Stan, G.-B. & Ellis, T. GeneGuard: A Modular Plasmid System Designed for Biosafety. *ACS Synthetic Biology* **4**, 307–316 (2015).
12. Wright, O., Stan, G.-B. & Ellis, T. Building-in biosafety for synthetic biology. *Microbiology* **159**, 1221–1235 (2013).
13. Kelle, A. Synthetic biology and biosecurity. *EMBO reports* **10**, S23–S27 (2009).
14. Wikmark, O.-G. *et al.* *Synthetic biology: biosafety and contribution to addressing societal challenges*. (2016). doi:10.13140/RG.2.2.29987.25121.
15. Knudsen, S. *et al.* Development and testing of improved suicide functions for biological containment of bacteria. *Applied and environmental microbiology* **61**, 985–991 (1995).

16. Pasotti, L., Zucca, S., Lupotto, M., Cusella De Angelis, M. G. & Magni, P. Characterization of a synthetic bacterial self-destruction device for programmed cell death and for recombinant proteins release. *Journal of Biological Engineering* **5**, 8 (2011).
17. Ishag, H. Z. A., Xiong, Q., Liu, M., Feng, Z. & Shao, G. Development of oriC-plasmids for use in *Mycoplasma hyorhinis*. *Scientific Reports* **7**, 10596 (2017).
18. Rideau, F. *et al.* Cloning, Stability, and Modification of *Mycoplasma hominis* Genome in Yeast. *ACS Synthetic Biology* **6**, 891–901 (2017).
19. Blötz, C. *et al.* Development of a replicating plasmid based on the native oriC in *Mycoplasma pneumoniae*. *Microbiology* **164**, 1372–1382 (2018).
20. Karas, B. J. *et al.* Direct Transfer of a *Mycoplasma mycoides* Genome to Yeast Is Enhanced by Removal of the *Mycoides* Glycerol Uptake Factor Gene *glpF*. *ACS Synthetic Biology* **8**, 239–244 (2019).
21. Lartigue, C. *et al.* Attenuation of a Pathogenic *Mycoplasma* Strain by Modification of the *obg* Gene by Using Synthetic Biology Approaches. *mSphere* **4**, e00030-19 (2019).
22. Piñero Lambea, C. *et al.* *Mycoplasma pneumoniae* Genome Editing Based on Oligo Recombineering and Cas9-Mediated Counterselection. *ACS synthetic biology* (2020) doi:10.1021/acssynbio.0c00022.
23. Sirand-Pugnet, P. *et al.* Being Pathogenic, Plastic, and Sexual while Living with a Nearly Minimal Bacterial Genome. *PLOS Genetics* **3**, e75- (2007).
24. Gibson, D. G. *et al.* *Complete Chemical Synthesis, Assembly, and Cloning of a Mycoplasma genitalium Genome*. <http://science.sciencemag.org/>.
25. Teachman, A. M., French, C. T., Yu, H., Simmons, W. L. & Dybvig, K. Gene transfer in *Mycoplasma pulmonis*. *Journal of bacteriology* **184**, 947–951 (2002).
26. Perez-Casal, J., Prysliak, T., Maina, T., Suleman, M. & Jimbo, S. Status of the development of a vaccine against *Mycoplasma bovis*. *Vaccine* **35**, 2902–2907 (2017).
27. Ishfaq, M. *et al.* Current status of vaccine research, development, and challenges of vaccines for *Mycoplasma gallisepticum*. *Poultry Science* **99**, 4195–4202 (2020).
28. Yattoo, M. *et al.* Novel Candidates for Vaccine Development Against *Mycoplasma Capricolum* Subspecies *Capripneumoniae* (Mccp)—Current Knowledge and Future Prospects. *Vaccines* **7**, 71 (2019).
29. Akbarzadeh-Niaki, M., Derakhshandeh, A., Kazemipour, N., Eraghi, V. & Hemmatzadeh, F. A novel chimeric recombinant protein PDHB-P80 of *Mycoplasma agalactiae* as a potential diagnostic tool. *Molecular Biology Research Communications* **9**, 123–128 (2020).
30. Garcia-Morales, L. *et al.* A RAGE Based Strategy for the Genome Engineering of the Human Respiratory Pathogen *Mycoplasma pneumoniae*. *ACS Synthetic Biology* **9**, 2737–2748 (2020).
31. Paetzold, B. 1981-. Synthetic biology in *Mycoplasma pneumoniae*. *TDX (Tesis Doctorals en Xarxa)* (2013).
32. Qi, L. S. *et al.* Repurposing CRISPR as an RNA-Guided Platform for Sequence-Specific Control of Gene Expression. *Cell* **152**, 1173–1183 (2013).
33. Lewis, M. Allosteric and the lac Operon. *Journal of Molecular Biology* **425**, 2309–2316 (2013).

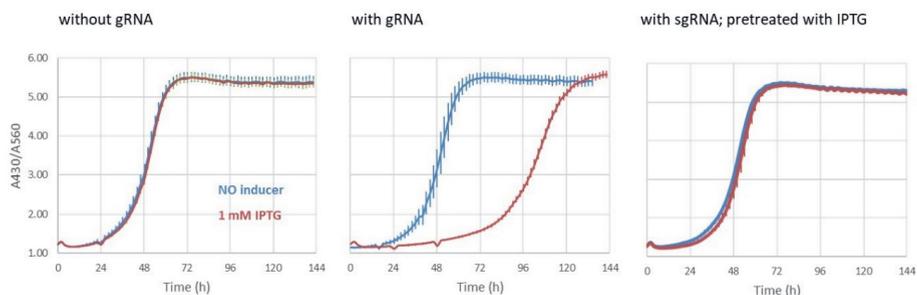
34. Ramos, J. L. *et al.* The TetR Family of Transcriptional Repressors. *Microbiology and Molecular Biology Reviews* **69**, 326 (2005).
35. An, W. & Chin, J. W. Synthesis of orthogonal transcription-translation networks. *Proceedings of the National Academy of Sciences* **106**, 8477 (2009).
36. Liu, C. C., Jewett, M. C., Chin, J. W. & Voigt, C. A. Toward an orthogonal central dogma. *Nature chemical biology* **14**, 103–106 (2018).
37. Lutz, R. & Bujard, H. Independent and Tight Regulation of Transcriptional Units in Escherichia Coli Via the LacR/O, the TetR/O and AraC/I1-I2 Regulatory Elements. *Nucleic Acids Research* **25**, 1203–1210 (1997).
38. Elowitz, M. B. & Leibler, S. A synthetic oscillatory network of transcriptional regulators. *Nature* **403**, 335–338 (2000).
39. Gardner, T. S., Cantor, C. R. & Collins, J. J. Construction of a genetic toggle switch in Escherichia coli. *Nature* **403**, 339–342 (2000).
40. Kogenaru, M. & Tans, S. J. An improved Escherichia colistrain to host gene regulatory networks involving both the AraC and LacI inducible transcription factors. *Journal of Biological Engineering* **8**, 2 (2014).
41. Molin, S. *et al.* Suicidal genetic elements and their use in biological containment of bacteria. *Annual Review of Microbiology* **47**, 139–166 (1993).
42. Kelty, C. Beyond implications and applications: The story of “safety by design.” *NanoEthics* **3**, 79–96 (2009).
43. van de Poel, I. & Robaey, Z. Safe-by-design: from safety to responsibility. *Nanoethics* **11**, 297–306 (2017).
44. Asin-Garcia, E., Kallergi, A., Landeweerd, L. & dos Santos, V. A. P. M. Genetic Safeguards for Safety-by-design: So Close Yet So Far. *Trends in Biotechnology* (2020).
45. Hewlett, A. L., Whitney, S. E., Gibbs, S. G., Smith, P. W. & Viljoen, H. J. Mathematical Modeling of Pathogen Trajectory in a Patient Care Environment. *Infection Control & Hospital Epidemiology* **34**, 1181–1188 (2013).
46. Rock, K. S. *et al.* Assessing Strategies Against Gambiense Sleeping Sickness Through Mathematical Modeling. *Clinical Infectious Diseases* **66**, S286–S292 (2018).
47. Smith, D. R. Epidemiologic tools for biosecurity and biocontainment. *Veterinary Clinics: Food Animal Practice* **18**, 157–175 (2002).
48. Chen, Z. *et al.* Role of the Mycoplasma pneumoniae/Interleukin-8/Neutrophil Axis in the Pathogenesis of Pneumonia. *PLoS one* **11**, e0146377–e0146377 (2016).
49. Maier, T. *et al.* Large-scale metabolome analysis and quantitative integration with genomics and proteomics data in Mycoplasma pneumoniae. *Mol. BioSyst.* **9**, 1743–1755 (2013).
50. Trussart, M. *et al.* Defined chromosome structure in the genome-reduced bacterium Mycoplasma pneumoniae. *Nature Communications* **8**, 14665 (2017).

51. Yus, E., Lluch-senar, M. & Serrano, L. Determination of the Gene Regulatory Network of a Genome-Reduced Bacterium Highlights Alternative Regulation Independent of Transcription Factors. *Cell Syst.* **9**, 143–158 (2019).
52. Yus, E. *et al.* Impact of genome reduction on bacterial metabolism and its regulation. *Science* **326**, 1263–1268 (2009).
53. Guerrero, F., Carbonell, V., Cossu, M., Correddu, D. & Jones, P. R. Ethylene Synthesis and Regulated Expression of Recombinant Protein in *Synechocystis* sp. PCC 6803. *PLOS ONE* **7**, e50470- (2012).
54. Elešnik, H. *et al.* Biosafety of biotechnologically important microalgae: Intrinsic suicide switch implementation in cyanobacterium *Synechocystis* sp. PCC 6803. *Biology Open* **5**, 519–528 (2016).
55. Contreras, A., Molin, S. & Ramos, J.-L. Conditional-Suicide Containment System for Bacteria Which Mineralize Aromatics. *Applied and Environmental Microbiology* **57**, 1504 (1991).
56. Simon, A. J. & Ellington, A. D. Recent advances in synthetic biosafety. *F1000Research* **5**, F1000 Faculty Rev-2118 (2016).
57. Bunn, A. G. A dendrochronology program library in R (dplR). *Dendrochronologia* **26**, 115–124 (2008).
58. The R User Conference, useR! 2011. Book of contributed abstracts.
59. Soetaert, K., Petzoldt, T. & Setzer, R. W. *Solving Differential Equations in R: Package deSolve*. *JSS Journal of Statistical Software* vol. 33 <http://www.jstatsoft.org/> (2010).
60. Bolker, B., R Development Core Team & Giné-Vázquez, I. *Package "bbmle" Title Tools for General Maximum Likelihood Estimation Description Methods and functions for fitting maximum likelihood models in R*. <https://github.com/bbolker/bbmle> (2020).
61. Package 'scales'. Scale functions for visualization. <https://scales.r-lib.org>, <https://github.com/r-lib/scales>
62. Sprouffske, K. & Wagner, A. Growthcurver: an R package for obtaining interpretable metrics from microbial growth curves. *BMC Bioinformatics* **17**, 172 (2016).
63. Zwietering, M. H., Jongenburger, I., Rombouts, F. M. & van 't Riet, K. Modeling of the bacterial growth curve. *Applied and environmental microbiology* **56**, 1875–1881 (1990).
64. Gompertz, B. *On the Nature of the Function Expressive of the Law of Human Mortality, and on a New Mode of Determining the Value of Life Contingencies*. *Philosophical Transactions of the Royal Society of London* vol. 115 (1825).

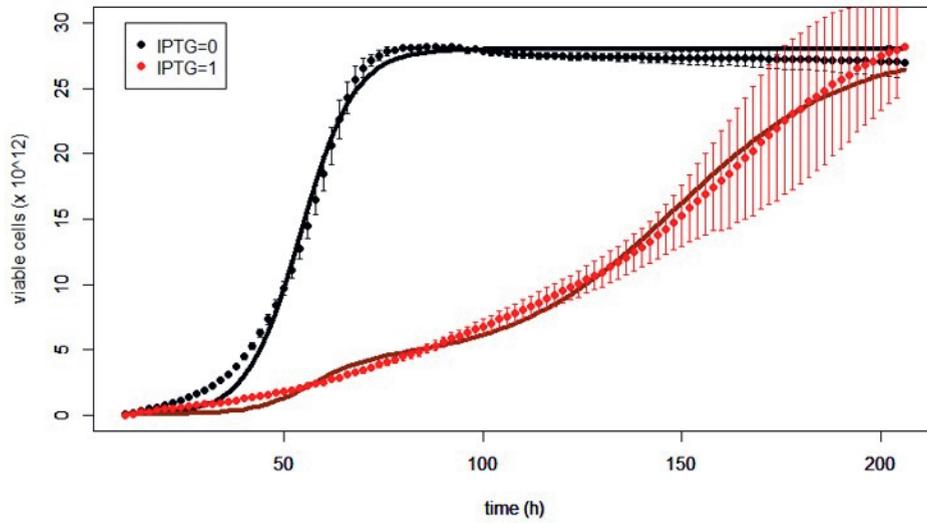
## Supplementary Information

### Supplementary Figures

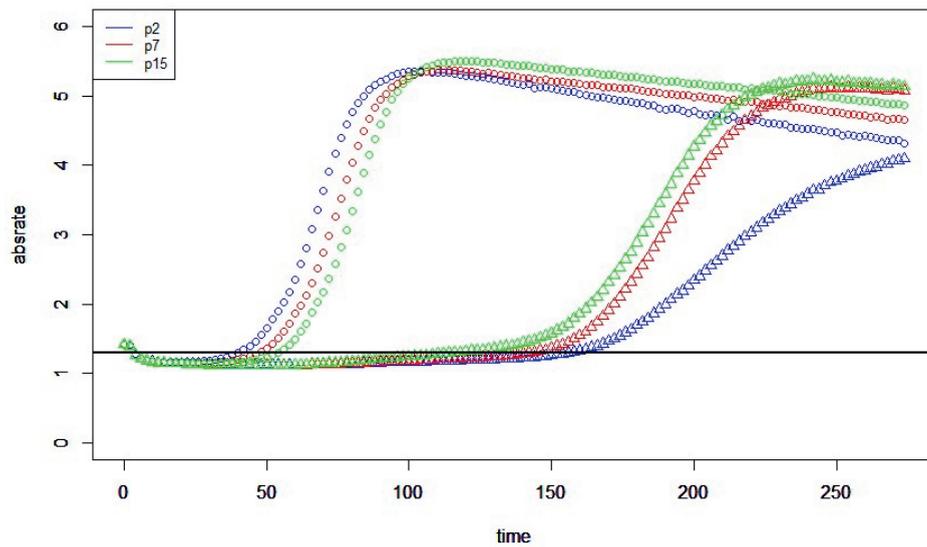
**Supplementary Figure S1 (git repository).** ODE model for TetR-Cre switch simulated with pTetRB (green) and pTetRA (blue). Model simulations are compared to experimental data (black). Measurements and simulations are in terms of mCherry fluorescence over time (hours). The three plots for each parameter represent the step-by-step building of the circuit.



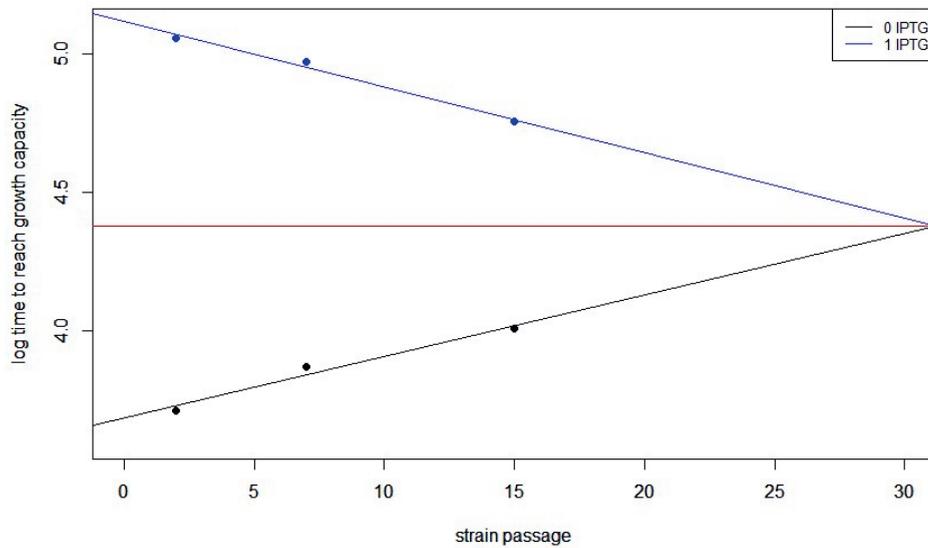
**Supplementary Figure S2.** Effect of IPTG-pretreatment on non-clonal strain nCS monitored through growth curves (pH change-based, rate of absorbances A430/A560). First panel) control. When guided RNA is not present, IPTG addition to the medium has no effect on the bacterial growth because the Cas9 nuclease is not cleaving the DNA sites. Second panel) Effect of IPTG addition on bacterial growth when the circuit is fully implemented. In presence of IPTG (red), less cells are present respect to when IPTG is absent (blue). Third panel) Effect of IPTG addition on bacterial growth after pre-treatment with IPTG on previous strain passage. Only the cells that have mutated the circuit are surviving in the pre-treatment passage. Therefore, at the next passage, IPTG will have no effect on bacterial growth.



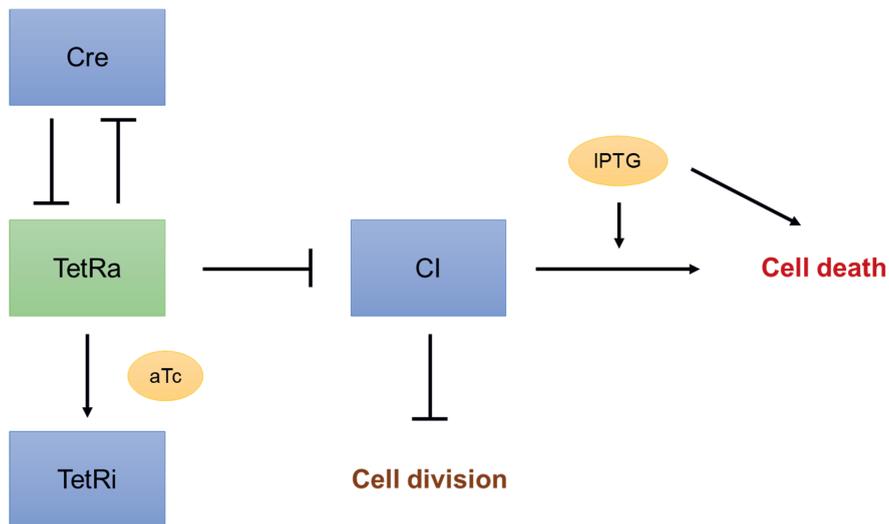
**Supplementary Figure S3.** Model simulation of clonal strain CS when a variation in the initial number of detected cells is introduced, impacting the growth curve at the first hours. Initial cell mutant cell number have been increased to 10<sup>5</sup>, showing the initial cell number might have been underestimated in the experimental detection.



**Supplementary Figure S4.** Estimation of rate of absorbances (after background correction) necessary to reach capacity for growth (black line) for strain CS.

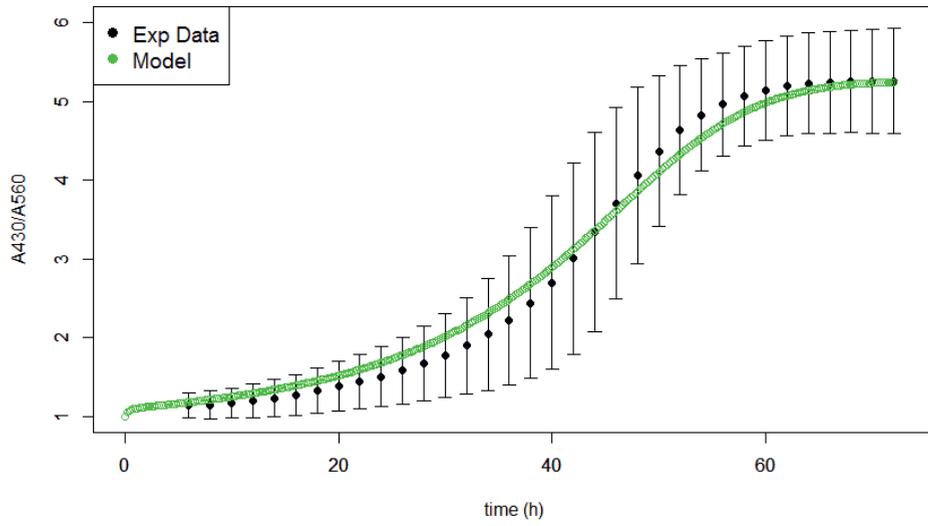


**Supplementary Figure S5.** Evolution of the time to reach growth capacity over strain passages for clonal strain CS. Cells will start growing at the same time, whether IPTG is present or not, once reached strain passage 30.

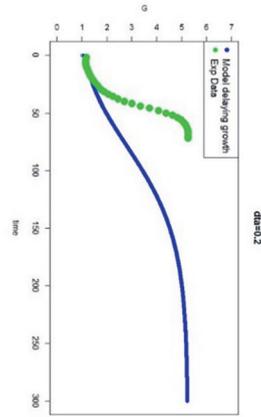
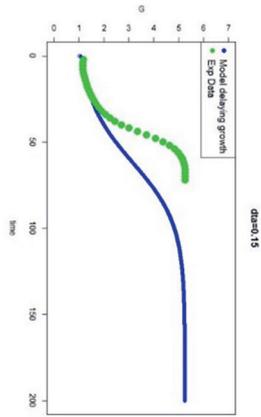
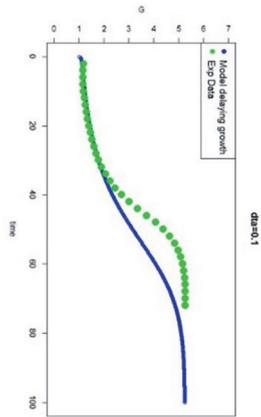


5

**Supplementary Figure S6.** Simplification of the multi-circuits biosafety network, used to facilitate the modelling procedure.

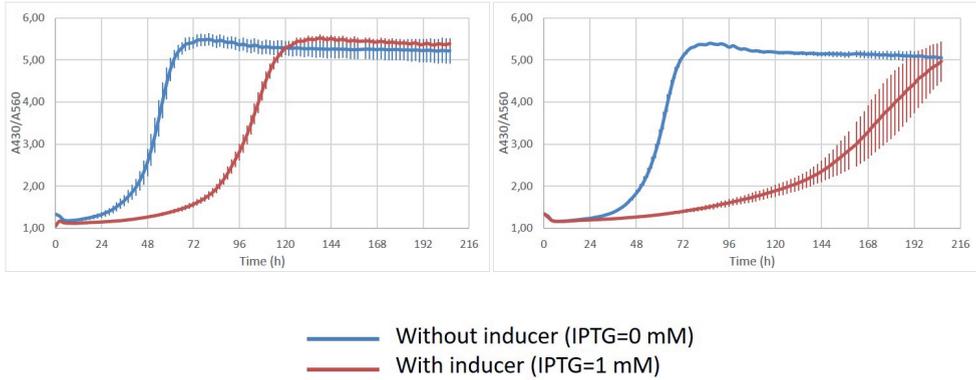


**Supplementary Figure S7.** Model of the full biosafety network (green) describing experimental data in terms of ABS430/ABS560 over time (black).

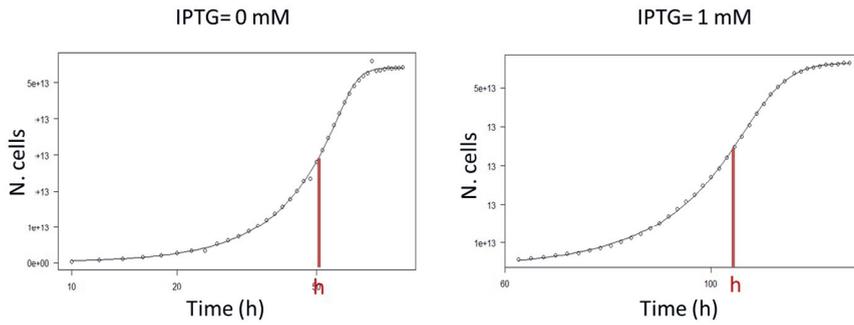


5

**Supplementary Figure S8.** Effect on growth delay of different aTc concentration added to the medium, modeled in terms of fraction of saturating aTc concentration (“dta”). Growth (G) is represented in terms of ABS430/ABS560 over time (hours).



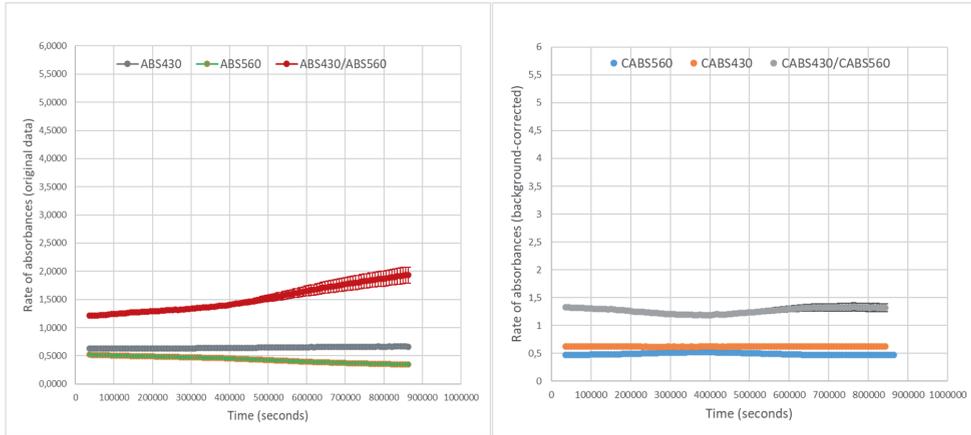
**Supplementary Figure S9.** Training data for model LacI-Cas9 for clonal and non-clonal strains in terms of rate of absorbances (ABS430/ABS560) over time (hours), with and without IPTG, for right panel) strain nCS and left panel) strain CS.



Regression fit with 5-Parameters Log-logistic: 
$$f(t) = c + \frac{(d - c)}{(1 + e^{b(\ln t - \ln h)})^f}$$

Growth rate is the first derivative of  $f(t)$  at the inflection point  $h$ : 
$$\left(\frac{df(t)}{dt}\right)_h$$

**Supplementary Figure S10.** Log-logistic regression to estimate inflection point  $h$  and growth rate of the curves for non-clonal strain nCS in presence or absence of IPTG.



**Supplementary Figure S11.** Background correction of data taken in absence of the bacterium for ABS430, ABS560 and their rate. The two absorbances should be constant over time. Left panel) before background correction. Right panel) after background correction.

### Supplementary Tables

Supplementary Tables S1, S2, S3, S4 are available in the git repository <https://gitlab.com/erika.gaspariwur/phd-thesis-erika-gaspari>.

5

### Supplementary Files

Supplementary Files S1, S2 are available in the git repository <https://gitlab.com/erika.gaspariwur/phd-thesis-erika-gaspari>.



# Chapter 6

## Discussion

Erika Gaspari



In this thesis, I have developed and used model-driven frameworks to support the development of a pneumonia vaccine. This work is part of the MycoSynVac project ([www.mycosynvac.eu](http://www.mycosynvac.eu)), aiming at using cutting-edge synthetic biology methods to engineer *Mycoplasma pneumoniae* as a universal chassis for vaccination, so that can be deployed as single or multi-vaccine in a range of animal hosts. Among the main specific objectives underlying that goal, I focused on making designs of serum-free growth media and on the strains themselves to improve the large-scale production of the vaccine, as well as on designs to enhance the product's safety. Thereby, I made a step towards bringing the product to the market.

One of the primary challenges I tackled was the design of strategies to allow the reproducible, high-quality and large-scale production of the vaccine chassis. The fact that *Mycoplasma* has very fastidious growth requirements and various auxotrophies made this objective a main part of MycoSynVac, as it is impossible to think of producing the vaccine at a large-scale and commercialize the product without being able to grow the bacterium at a reasonable rate, in defined conditions and at reduced costs. That is why in **chapters 2 and 3** I designed a modelling framework that could be deployed for this purpose, while at the same time being flexible with regards to the experimental applications. I wanted to build a model that could be deployed for medium development and empowered-strain design (i.e. being able to efficiently simulate knock-ins and knock-outs), but also that could be easily complemented with new pathways and compartments, representing a versatile approach to perform *in silico* studies on *Mycoplasma*. To be successful and to represent accurately the bacterium, the model had to integrate all the biological knowledge available at the moment on this microbe. Therefore, I implemented the biological assumptions in the model itself, primarily linking the bacterial membrane formation to the metabolism of *Mycoplasma*. This strategy was successful and allowed me to fulfil the aims for which the model was developed namely, the design of a serum-free medium enabling *Mycoplasma* growth and with predictive power to simulate genetic modifications. I think an important aspect of this framework is that a very intuitive methodology was used: the right biological assumptions and their implementation in the model made the existing standard metabolic modelling approaches more than enough to provide powerful simulations, making the model easy to use and reproducible. Furthermore, when developing a vaccine, especially if this has to be designed as a universal chassis, antigens and epitopes are surely one of the main aspects. Although I have not taken part of the peptide-expression production (conducted by the MycoSynVac partner *ATG biosynthetics*, [www.mycosynvac.eu](http://www.mycosynvac.eu)), I have designed the model such that it could be deployed for maximization of protein expression and for a *Mycoplasma pneumoniae* chassis displaying any epitope. Moreover, the model can be used to simulate the strain without pathogenic factors and for the prediction of essential genes.

To be suitable for the market, a vaccine chassis must be *safe*. Safety aspects to be considered are various and derive by diverse foreseen consequences. First of all, one must analyse the potential of the vaccine strain to cause harm, although the strain has been made non-pathogenic. The fact the bacterium is no longer able to cause an infection does not mean it cannot have adverse effects when injected into the host. Indeed, *Mycoplasma pneumoniae* is known to trigger an autoimmune disorder in humans. In **chapter 4** I therefore studied the pathway that leads to the synthesis of the compound triggering this disorder and identified the factors that should be substituted to avoid an autoimmune response by the host. By analysing the ability of all the genome-sequenced *Mycoplasmas* to produce such a compound, I asked myself whether *Mycoplasma pneumoniae* was actually the right candidate among the *Mycoplasmas* to develop a universal vaccine chassis. Safety does not pertain only adverse reactions in the host, but also the identity of the vaccine strain as a genetically-engineered organism. Synthetic microbes' impact on the environment where they are deployed must be taken into consideration when developing a live attenuated vaccine. For this reason, safeguards and biocontainment strategies have been implemented into the chassis' genome,

in order to prevent its growth outside the desired conditions. In **chapter 5**, I have supported the design of these biosafety strategies (implemented by the MycoSynVac partner *Imperial College of London*) with mathematical frameworks, predicting the effect and efficiency of the circuits over time.

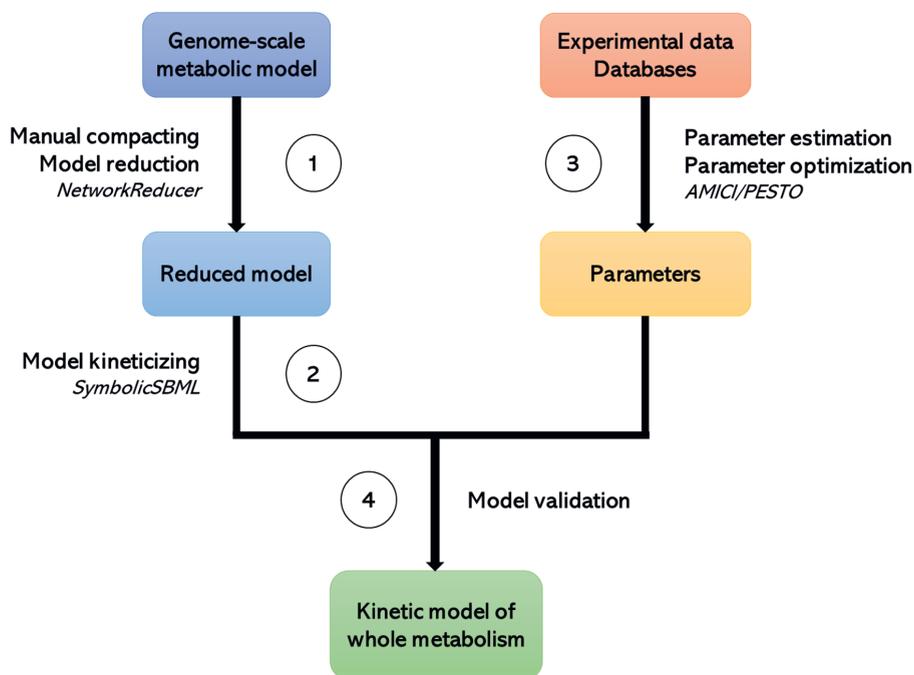
Altogether, the four primary research chapters in this thesis address all the issues for which model frameworks have been used in the design phase of MycoSynVac. Jointly, there are some methodological aspects that should be discussed further, which I do in the following subsections. Firstly, I relied on a metabolic model for medium and chassis design; this type of models assumed steady state conditions, thus simulating accurately what happens in, for instance, chemostats. However, these models do not take into account the dynamics that take place in between steady states. Secondly, although model-driven design has been reported several times in literature, I believe this approach should become a standard, especially considering companies that still make massive usage of animal serum in growth media. Serum-rich media are largely used due to their versatility, but ethical concerns have been raised regarding the serum's harvesting. Moreover, they are expensive and non-defined. Serum-free media can be designed to be efficient and versatile at the same time. A third main point of discussion regarding this thesis is the deployment of *Mycoplasma pneumoniae* as a live attenuated vaccine, and the deployment of live attenuated vaccines in general. Thus, below, I discuss the *known unknowns* of Mycoplasmas and of microbial organisms used in biomedical applications, as well as the *unknown unknowns*, which we are not aware of by definition, but we should take into consideration. Finally, I lead the reader to a reflection on how to tackle the unknowns, proposing *Safety-by-Design* as a standard to be implemented in the developmental phases of pharmaceutical and biomedical products.

### **Steady-state vs dynamic metabolic models.**

In two chapters of this thesis, I developed and deployed genome-scale, constraint-based metabolic models (GEMs) for the purpose designing defined growth media and optimizing *Mycoplasma* strains. The main mathematical method applied to genome-scale models, Flux Balance Analysis (FBA), simulates metabolic fluxes at steady state. The deployment of FBA to GEMs has two main limitations: i) GEMs only describe the metabolic (and transport) fluxes of an organism, without accounting for non-metabolic functions, and ii) FBA can only simulate the metabolic fluxes at the steady state. Numerous attempts of improving GEMs exist, which have led to the development of "next-generation genome-scale models". For instance, ME-Models combine metabolic information with gene expression, allowing to simulate metabolic fluxes considering enzyme costs and the prediction of gene expression levels, maximum feasible growth, nutrient-limited phenotypes<sup>1</sup>. These models have been further improved by integrating membrane protein translocation, which is useful for predicting membrane protein composition<sup>2</sup>. Metabolic models with protein structural properties also exist that account for the effect of these characteristics on the cell's metabolism and predict drug binding sites<sup>3-5</sup>.

Regulatory interactions have been integrated into GEMs by creating probabilistic transcriptional and metabolic networks, which can be used to assess the effects of genetic perturbations <sup>6</sup>. This probabilistic regulation of metabolism, as a result of genetic and environmental perturbations affecting the interactions between genes and transcription factors, has been considered in the so-called PROM models. More recently, GECKO models account for enzyme kinetics, abundances and turnover numbers, to constraint the metabolic fluxes such that they are not able to exceed the maximum capacity <sup>7</sup>. The simulation of these model types relies on the assumption of the steady state, meaning that they are solved by setting the product of the stoichiometry matrix and the vector of unknown fluxes to 0. For this reason, dynamic Flux Balance Analysis (dFBA) was developed to incorporate a rate of change of flux constraints, so kinetic expressions. With respect to FBA, which solves a linear programming problem, dFBA transforms it to a nonlinear programming one. dFBA then consists of solving of system of equations dividing the total time period into several intervals, where in each time interval the optimization problem is reduced to a linear programming one <sup>8</sup>. The same concept was used by Karr et al., who remarkably developed a whole-cell model integrating 28 cellular process submodules, among which metabolism, for *M. genitalium* <sup>9</sup>. The original objective is to integrate the various layers of cellular processes and thereby predict full phenotype from genotype. This last model is simulated in a modular fashion, with each cellular process being considered both as a singular entity and as a part of a larger system, will all the calculations for each submodule being carried out within a given time step. Accordingly, submodules are simulated separately from the other cellular processes over a single time step of 1 second, and then synchronized to all the other submodules before the next time step. However, a numerical convergence criterion is clearly missing, as well as clear results of the study and reliable predictions by the model. Altogether, this makes it a *bello che non balla* (a beautiful one, but which cannot dance) among models.

To overcome this limitation I tried to provide numerical convergence and time continuity by developing a kinetic model for the whole metabolism of *M. pneumoniae*. Such a dynamic model has the advantage of simulating metabolic species (even internal metabolites) with no steady state assumption, and allowing for parameter optimization and sensitivity analysis. The construction was based on a Python pipeline called DMPY <sup>10</sup>, automatedly constructing Ordinary Differential Equations (ODEs) models from GEMs and, therefore, able to simulate the metabolite concentrations over time. Although the metabolic network of *M. pneumoniae* was consistently reduced, kinetic models of this size are considered large (despite the *M. pneumoniae* network being very small compared to other microorganisms). This means that a very large amount of parameters has to be retrieved from experimental data and databases such as BRENDA <sup>11</sup>, SABIO-RK <sup>12</sup> and eQuilibrator <sup>13</sup> [this last is not a direct retrieval, but a computation based on chemical constraints]. Regrettably, in practice, generally only a minimal number of parameters can be retrieved, whereas the majority must be estimated/optimized. An overview of all the steps needed to build a kinetic model of *M. pneumoniae* metabolism is shown in [Figure 1](#).

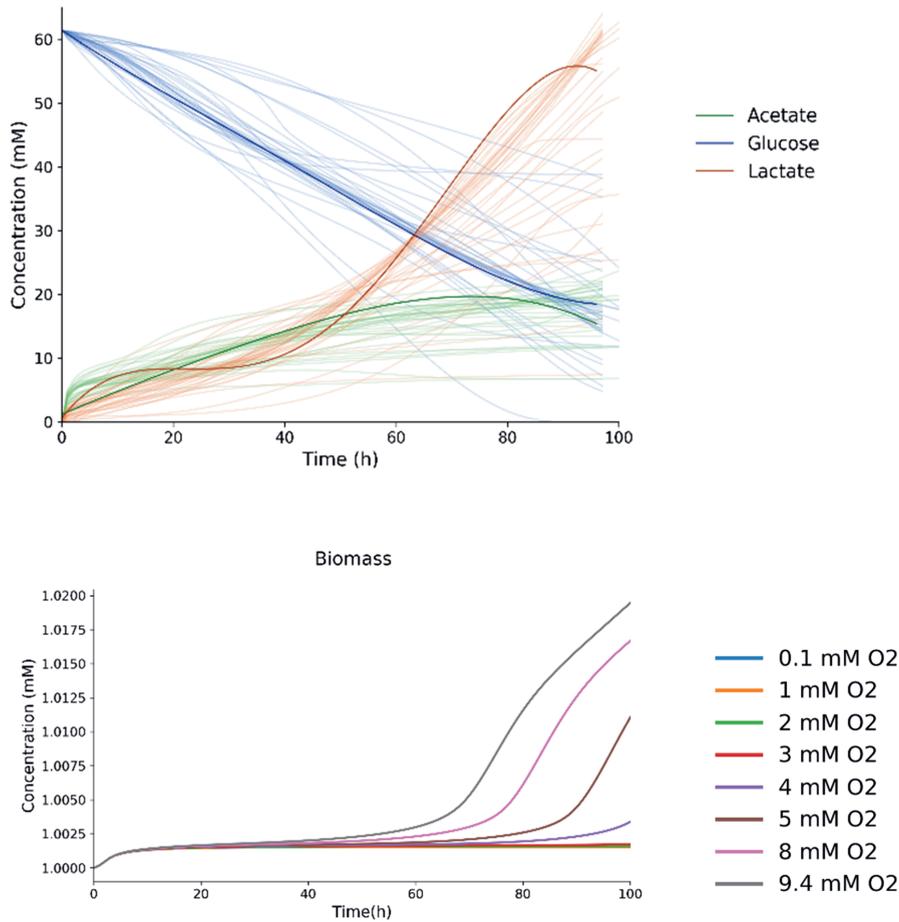


**Figure 1.** Overview of the process for the construction of a whole-metabolism kinetic model for an organism from a genome-scale, constraint-based metabolic model. 1) The metabolic network is reduced, both manually and through the usage of a tool, i.e. NetworkReducer<sup>14</sup>. The result is a reduced model. A model that is “reduced-enough” for this purpose is considered to be around 100 reactions. 2) The biochemical reactions are kineticized with SymbolicSBML (van Rosmalen, unpublished), which generate a system of ODEs. 3) Available parameters are retrieved from the databases and the remaining ones estimated/optimized from available experimental data through the tools PESTO and AMICI<sup>15,16</sup>. 4) Parameters are inserted in the kinetic equations and the model is simulated and compared to available experimental data for validation.

Despite the computationally-intensive and laborious process to implement the workflow, the model was able to correctly capture glucose consumption, acetate and lactate production (Figure 2a) when compared to available experimental data, but not, for example, the biomass yield at different oxygen concentrations (Figure 2b), whose prediction was unrealistic. The same simulation was instead reliable when performed with the GEM (chapter 2)<sup>17</sup>, providing realistic oxygen concentrations and predicting the expected effects on the metabolic pathways. Furthermore, I could only validate the model on the same data on which the parameters were estimated, due to the lack of available time-course experimental data. Even a cross-validation procedure would have made the estimation impossible for the insufficient amount of data. Indeed, parameter optimization can be considered both an advantage and a disadvantage of kinetic modelling: while it allows to model internal metabolites with time continuity, it also requires a consistent amount of experimental data over time to estimate the

parameters from. In my case, the only available experimental data of this type were provided for glucose consumption, acetate production and lactate production. This data consisted only of two replicates over four time points, taken every 24 hours. Therefore, the simulation of lactate, acetate and glucose over time alone cannot be considered as a satisfactory validation of the model.

Furthermore, I was unable to perform any other prediction with the model that could be considered reliable, as further experimental data for the comparison to model simulations were missing. From the bulk of my analysis, I believe the network reduction had an impact on the quality of the model, which unfortunately lost important information when entire pathways were lumped in one reaction. However, this step is needed to avoid a parameter space that would be too big to allow the estimation and optimization. A second main problem was the lack of parameters in the databases, and of experimental data to estimate them from. The optimal type of data would comprise replicates over time of both extracellular (e.g. glucose, acetate and lactate) and intracellular metabolites; these last are difficult to measure in a time-course, but strongly required for the parameter estimation. One of the main advantages of constraint-based modelling is that it does not require parameter estimation, and the related parameter search in databases, complex parameter estimation tools and laborious manual curation to make the model feasible and reliable. Moreover, what is important in modelling is a clear definition of the spatio-temporal frame where the focus of the simulations lie. Our attempt to develop a kinetic model for *M. pneumoniae* metabolism might have encountered the same issue of Karr et al.<sup>9</sup> when simulating processes that take place at different time frames and compartments. To combine different kinetic processes at different time frames and with very different parameter ranges for a high number of equations did not, for now and at my best knowledge, produce reliable results, although a model feasibility was reached both in my and in Karr's et al. cases. Nevertheless, even when reaching the same predictive power of the related GEM, would it be worth it? I believe an advantage that dynamic models provide over GEMs is the potential of introducing transcriptional, translation regulation and feedback regulatory mechanisms into the model of a biological system, but effective methods for the integration of these compartments with the metabolic one has still a long way to go.



**Figure 2.** Simulations performed with the kinetic model of *M. pneumoniae* metabolism. Top panel) Predictions of glucose consumption, acetate production and lactate production over time (thin lines) compared to experimental data (thick line). Bottom panel) Simulations of biomass production for different oxygen concentrations.

### The importance of model-driven design and serum-free media.

Prediction of minimal serum-free growth medium is necessary in particular when the microbe in question has difficulties in growing and it is used in high-added value processes in pharmaceutical or biotechnological applications, such as large-scale productions of vaccines or biomolecules of interests (e.g. antibiotics, biofuels etc.)<sup>18</sup>. An example is *Pseudomonas putida* KT2440, widely used in biotechnological applications and which is able to grow on a wide range of compounds. In this case, a medium analysis was needed to understand the usage of complex mixtures of sources<sup>19</sup>, which might lead to challenges when using the

bacterium in bioremediation and biotransformation processes. However, in this work no specific medium prediction was performed. In the ambit of biodiesel production, maximization of growth rate through FBA on *Chlorella vulgaris*' metabolic network<sup>20</sup> was used to predict conditions for maximizing lipid production, leading to the definition of optimal nitrogen, oxygen and light conditions, but again there was no media design done. Unrelated to the bacterial world, the medium for cultivating Chinese Hamster Ovary (CHO) cells, used for the production of the majority of biopharmaceutical drugs, was computationally predicted through model-driven designs attempts<sup>21,22</sup>. The approaches comprised kinetic models<sup>23,24</sup>, machine learning<sup>25</sup>, small-scale metabolic flux analysis<sup>26-29</sup> and FBA applied to a genome-scale model<sup>30-33</sup>. Similarly, Huguera et al. gave insights on the amino acid metabolism of stem cells (specifically, human mesenchymal stem cells), contributing to the design of animal serum-free medium<sup>34</sup>. A chemically defined minimal medium supporting *Staphylococcus aureus* growth was predicted through model-driven design by Machado et al.<sup>35</sup> on the basis of a previously developed medium unable to support growth of two strains. Quite common are works that regards phenotype predictions and carbon-fluxes characterization according to genome-scale model simulations on different media composition<sup>36-41</sup>. These model-driven design works led to the choice of optimal existing media to grow the organisms of interest, but not to the formulation of the media themselves. Instead, one of the first applications of metabolic modelling for medium design purposes is in 2007, when Baart et al. constructed a genome-scale model of *Neisseria meningitis*, and used it to predict substrate preferences and design a minimal growth medium, afterwards tested in batch and chemostat cultures<sup>42</sup>. Similarly to what I described in **chapter 2**, the key of the successful medium prediction was the integration of membrane-transport functions. Tejera et al. used genome-scale metabolic modelling to identify auxotrophies in *Campylobacter jejuni* and design a defined medium containing, among the others, the auxotrophic substrates, and therefore support the growth of the microorganism<sup>43</sup>. Model-driven medium designed was also applied to *Akkermansia muciniphila*, after the identification of metabolic bottlenecks and the need of N-acetylglucosamine in the culture medium<sup>44</sup>. In addition to genome-scale modelling, other efficient methodologies have been applied for the discovery of culture media *in silico*, such as genetic algorithms to optimize the amino acid composition of cell culture media for the marine sponge *Dysidea etheria*<sup>45</sup>. Furthermore, Teusink et al. studied the relations between biomass yield on different glucose concentrations and the maintenance energy of *Lactobacillus plantarum*<sup>46</sup>. Finally, an alternative approach for media design is the one described by Oberhardt et al., who developed the Known Media Database (KOMODO) consisting of more than 18000 strain-media combinations. This database can be used to predict defined culture media for new organisms based on transitivity properties and phylogenetics analyses<sup>47</sup>.

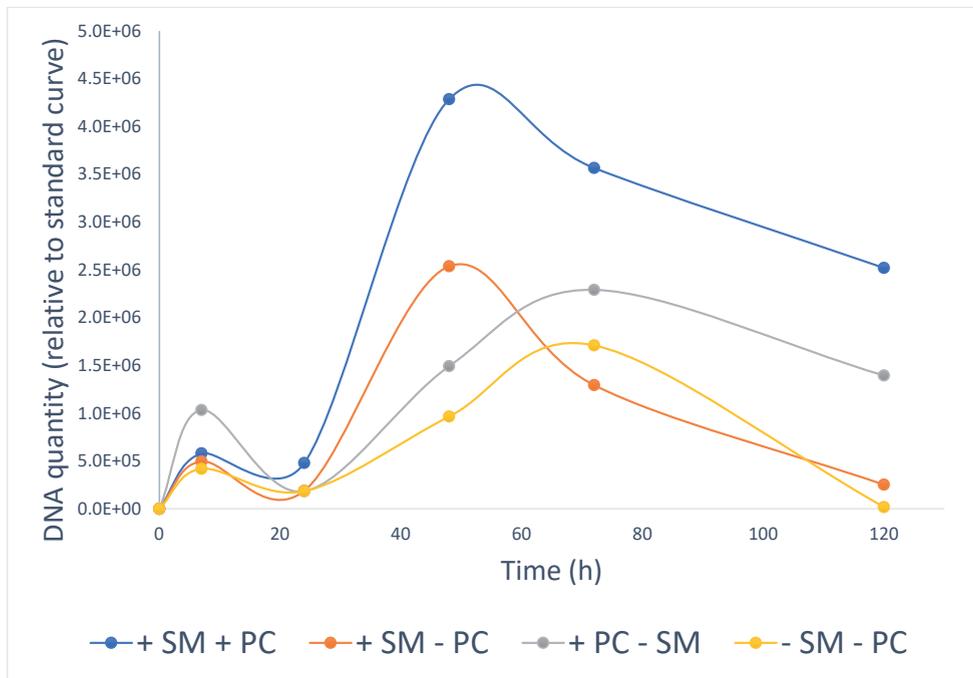
The procedure for model-driven design that I implemented is intuitive for a computational biologist and does not involve any complex approach, as it pertains analyses that are commonly performed with genome-scale metabolic models. I believe this model-driven approach should be applied for medium design whenever the known medium for growing the microorganism contains animal serum. Indeed, many ethical concerns have been raised regarding the harvesting of the animal blood, which might cause suffering<sup>48-50</sup>. The usage of animal serum in growth media is due to its richness in growth factors and versatility (can be used in many

culture applications); however, serum composition is not fully defined and may vary between batches. Many biomedical and biotechnological applications require the knowledge of the exact components used to feed the microorganism; therefore, the usage of serum-free media should be enhanced not only to protect the animals and reduce production costs, but also to have full knowledge on the growth medium composition. This is essential to guarantee reproducibility and consistent quality in production processes, in particular in pharmaceutical industries. Moreover, although versatility is considered one of the main reasons why serum-rich media are so broadly used, serum-free media can also be versatile and a fraction of the cost and applicable to a range of organisms. As an example, based on the robust results obtained on *M. pneumoniae*, we did preliminary tests on additional mycoplasmas grown on the MCMycO medium I have developed. Table 1 shows other mycoplasmas can grow on MCMycO. Although tested only on first strain passage (meaning stable growth over strain passages cannot be assumed), these results are promising and show how modelling is not just a solid approach for medium design, but can also reveal media versatilities, especially regarding group of organisms with similar genetic makeup and, therefore, metabolic requirements.

**Table 1.** Growth of mycoplasmas on MCMycO shows alternatives to serum-rich media are possible. Three mycoplasmas plus *M. pneumoniae* have been grown at first strain passage on MCMycO with various lipid (SM=sphingomyelin, PC=phosphatidylcholine) concentrations. The "+" indicates presence of growth, i.e. observed colour change of the sample in the tube with phenol red.

	MCMycO 20µg/mL SM, 20µg/mL PC	MCMycO 50µg/mL SM, 50µg/mL PC	MCMycO 20µg/mL SM 0µg/mL PC	MCMycO 0µg/mL SM 20µg/mL PC	MCMycO 0µg/mL SM 80µg/mL PC	Serum- rich medium
<i>M. bovis</i>					+	+
<i>M. cynos</i>		+			+	+
<i>M. gallisepticum</i>	+		+	+	+	+
<i>M. pneumoniae</i>	+		+	+	+	+

To be taken into consideration is anyway that growth peaks are reached at different times. For example, *M. pneumoniae* shows growth in almost all the tested lipid mixtures, but grows consistently only in presence of both SM and PC (Figure 3).



**Figure 3.** Growth over time of *M. pneumoniae* on MCMycro with various lipid (SM=sphingomyelin, PC=phosphatidylcholine) presence. Growth curves show only the “full” MCMycro (i.e. with both lipids) allows consistent growth of the bacterium. Curves’ initial increase at 8h might be due to growth in rich medium in the previous strain passage. After subsequent decrease around 24h, the bacterial population shows a sharp growth when sphingomyelin is present (blue and orange curves). Despite all conditions allow growth, only in presence of both PC and SPM the curve has both a sharp increase and remains stable over time, until 120h.

Some *Mycoplasmas* are even more difficult to grow: *M. hyopneumoniae*, the swine pathogen, is one of them and, although several vaccines exist, a fully serum-free medium is not yet available. The available genome-scale metabolic model of this bacterium is recent and well-characterized <sup>51</sup>. In this work, the authors sequenced the genome of strain 11 of *M. hyopneumoniae* and used it to reconstruct the genome-scale model of its metabolism, finding out the addition of pyruvate to FRIIS medium increased the growth rate of the bacterium. However, FRIIS medium is supplemented with horse serum and swine serum <sup>52</sup>; therefore, a serum-free medium was not developed. A possible solution could be reached by applying the same approach we designed for *M. pneumoniae*: integrating a membrane compartment, which can be built based on available membrane lipids data, into the genome-scale metabolic model, introduce in the model all the singular lipid pathways and link them to the membrane and the medium supplementation. While for *M. pneumoniae* available literature data were used, lipidomics platform might be used to avoid incomplete and inconsistent information on membrane lipid entities and proportions. The collection of data on relevant metabolites, followed by a model-driven approach, might be able to represent in the future a solid framework for medium design.

## Safety of live attenuated vaccines

I would like to propose the reader a reflection on the usage and development of live attenuated vaccines. The organisms used in this type of vaccines are to be "non-pathogenic". However, can we always ensure the organism is totally non-pathogenic, or is it only non-pathogenic for the disease we are aware of? Taking the case of *M. pneumoniae*, the bacterium is known for causing atypical pneumoniae in humans. Furthermore, there are numerous report cases of *M. pneumoniae* being correlated to a series of other disorders, with the Guillain-Barré-Strohl syndrome being only one of those. Muller et al. observed that *M. pneumoniae* worsened the clinical symptoms of Tourette-Syndrome (TS) patients<sup>53</sup>; therefore, Müller et al. performed a subsequent study to determine antibody titers against *M. pneumoniae* in TS patients. They found out the percentage of TS patients with positive antibody titers against *M. pneumoniae* was much higher than the percentage of the control group (59% vs 3%), suggesting the involvement of IgA antibodies<sup>54</sup>. *M. pneumoniae* has been also linked in several report cases to post-infectious cold antibody haemolytic anaemia (CAHA). Although it has been suggested that the lipid-rich *Mycoplasma* surface might play a role as adjuvant in the formation of complexes with receptors on human cells, the causes are not fully understood<sup>55,56</sup>. Other cases found links of *M. pneumoniae* infections with Obsessive-Compulsive Disease<sup>57</sup>, myocarditis and perimyocarditis<sup>58-62</sup>, Kawasaki disease<sup>63,64</sup>, Stevens-Johnson syndrome<sup>65-67</sup>, hepatitis<sup>68-71</sup>, hemophagocytic syndrome<sup>72-74</sup>, arthritis<sup>75-78</sup>, rhabdomyolysis<sup>79-81</sup>, otitis<sup>82,83</sup>, ocular manifestations<sup>84,85</sup>, glomerulonephritis<sup>86,87</sup> and other extrapulmonary manifestations, even in absence of pneumonia, including almost all organs in the human body<sup>88</sup>. Although correlation does not mean causation and most of the cases must be considered rare or linked to pre-existing pathologies, these numerous reports should make us reflect on how little we know on *M. pneumoniae*, its interaction with the human body and, therefore, its possible adverse effects. The US Secretary of Defense Donald Rumsfeld said in 2002 that "*there are known knowns. There are things we know we know. We also know there are known unknowns. That is to say, we know there are some things we do not know. But there are also unknown unknowns, the ones we don't know we don't know*" ([www.scientificamerican.com/article/rumsfelds-wisdom](http://www.scientificamerican.com/article/rumsfelds-wisdom)). While this phrase originally referred to the limitation of intelligence reports, it applies to scientific research too, whose role is to investigate the known unknowns. However, scientists are called to consider the unknown unknowns as well<sup>89</sup>. As it is for now impossible for scientist to fully understand one of the smallest living organisms, it is not difficult to think that this affirmation could be valid for any other living organism. Vaccine clinical trials have, among the others, the specific function of assessing the vaccine's safety, revealing all side effects. However, millions of euros are first spent to fund pre-clinical development of live attenuated vaccines, whose usage should be discouraged due to our lack of knowledge on living organisms and their effects on the host. Indeed, most live attenuated vaccines for human diseases in commerce have been abandoned over the years as a consequence of the many side effects. The bacterial live attenuated vaccines still in commerce are against anthrax, cholera (*V. cholerae*), plague (*Y. pestis*), tuberculosis (*M. bovis* BCG) and typhoid (*S. typhi* Ty21a)<sup>90-96</sup>. Viral live attenuated vaccines in commerce are for influenza, measles, mumps, rubella, polio, rotavirus, smallpox, chickenpox (also called varicella), yellow fever and Japanese encephalitis<sup>97-107</sup>.

An example of potential unknown side effects of the usage of live microorganisms in biomedical applications can be overviewed by exploring the presence of the motif responsible for biosynthesis of galactocerebroside, inducing autoimmune response and Guillain-Barré-Strohl syndrome in humans. Table 2 provides a list of organisms where the protein motif was found in glycosyltransferases or similar enzymes, following the same approach described in chapter 4.

**Table 2.** Number of M200535 motifs (linked to galactocerebroside synthesis in chapter 4) found in enzymes of *V. cholerae*, *L. lactis*, *L. monocytogenes*, *L. casei*, *A. baumannii*.

<b>NAME OF THE MICROORGANISM</b>	<b>TYPE OF ENZYME</b>	<b>N. OF MOTIFS</b>
<b><i>VIBRIO CHOLERAE</i></b>	Glycosyltransferase family 2	1
<b><i>VIBRIO CHOLERAE</i></b>	Sugar transferase	1
<b><i>LACTOCOCCUS LACTIS</i></b>	Glycosyltransferase family 2	2
<b><i>LACTOCOCCUS LACTIS</i></b>	Putative glycosyltransferase	1
<b><i>LISTERIA MONOCYTOGENES</i></b>	Glycosyltransferase	2
<b><i>LISTERIA MONOCYTOGENES</i></b>	Hyaluronan synthase	2
<b><i>LISTERIA MONOCYTOGENES</i></b>	Glycosyltransferase family 2	1
<b><i>LACTOBACILLUS CASEI</i></b>	Glycosyltransferase family 2	1
<b><i>LACTOBACILLUS CASEI</i></b>	Uncharacterized protein	1
<b><i>ACINETOBACTER BAUMANNII</i></b>	Glycosyltransferase (cell wall)	1
<b><i>ACINETOBACTER BAUMANNII</i></b>	Glycosyltransferase 200	1
<b><i>ACINETOBACTER BAUMANNII</i></b>	Glycosyltransferase 67	1

This list is the result of a search on a restricted list of organisms of one motif related to one specific disease. Therefore, it is easy to imagine the potential microorganisms causing side effects on humans when applied to biomedical applications. Moreover, especially in the case of live attenuated vaccine, I believe such a search should be performed long prior to the clinical trial phase and so in the design phase, to avoid the waste of important economic resources, at least for what concerns known links with disorders. Projects that regard the development of a new vaccine, therapy or product should comprise, in their pre-clinical phases, not only the exploration of possible side effects in, e.g., risk-related work packages, but also the active design of strategies to implement in these terms. Although the scientific community cannot be aware of all the unknown unknowns, a Safety-by-Design approach would help in addressing those issues that are not immediately obvious and that might not necessarily constitute a problem along the project development, allowing the scientists to be ready to tackle them.

## Safety-by-design

In recent years, the European Union has funded several research efforts on the environmental impact of a deliberate release of genetically engineered microbes<sup>108-117</sup>. De Lorenzo concluded that the environmental impact was approximately the same between genetically engineered microbes and native ones. The studies have revealed that synthetic microbes can attain a transient advantage in a population, but it is difficult for them to survive in the long run because they can be rapidly destroyed by competitors or predators, and strong constraints for their proliferation are also imposed by ecological conditions involved; this notion is consistent with the fact that most attempts to genetically engineer microbes for environmental applications have had little success to date<sup>118,119</sup>. However, the emergence of CRISPR/Cas9, a new genome-editing technology, has had tremendous effects on the synthetic-biology field. This technology not only improves the accuracy and efficiency of editing of pathogens', animals', plants', and human genomes but also yields traceless modification of genomes in a short period. Therefore, the technology can be utilized to enhance the pathogenicity, virulence, or transmission of toxins or bacteria or to disrupt the essential genes in humans, animals, and plants. In addition, several recent studies indicate that CRISPR/Cas9 has "off-target" effects, which could result in undefined health consequences<sup>120,121</sup>. Furthermore, easy and low-cost operation increases the risks of intentional abuse<sup>122</sup>. In a recent report submitted to the U.S. Senate by the U.S. Intelligence Agency in 2016 and 2017, the CRISPR/Cas9 genome-editing technology was suggested to have a potential as a weapon of mass destruction (<https://www.technologyreview.com/2016/02/09/71575/top-us-intelligence-official-calls-gene-editing-a-wmd-threat>).

The concerns explained above suggest the need of a way to investigate and take into account safety concerns at early stage and throughout the whole developmental phase of biotechnological and biomedical products. While only recently explored in the synthetic biology field, this concept is not new and was introduced in totally unrelated fields, for example the construction world<sup>123</sup>, nanotechnology applications<sup>124</sup> and others<sup>125,126</sup>, being currently described as Safety-by-Design (SbD).

In addition to incorporating risk assessment from the onset and for the entire process development cycle, SbD deals with risk management in synthetic biology through the implementation of genetic safeguards, biocontainment and biosafety strategies. Such implementations allow the risks to be anticipated and assessed at a R&D phase or during the product design, leading to the development of measures to minimize or, optimally, avoid potential risks. Asin-Garcia et al. have drawn a library for SbD genetic safeguards<sup>127</sup>, classifying them into four categories, acting through biological isolation (restricting the viability to defined environments) or genetic isolation (preventing horizontal and vertical gene transfer):

- Auxotrophy-related: the microorganism is made unable to synthesize an essential molecule, which in consequence must be provided, e.g. in the growth medium.
- Genetic circuits and switches: biosafety circuits are introduced into the microorganism's genome to limit its growth or survival.

- Semantic containment: alteration of the microorganism's genome, e.g. introducing noncanonical DNA that encode for noncanonical amino acids, to prevent horizontal gene transfer.
- Combinatorial systems: combination of multiple strategies.

The strategy implemented in **chapter 5** belongs to the group of genetic circuits, being killing switches a type of circuit that, when introduced into the genome of the organism, leads it to death when induced by an exogenous molecule. The underlying modelling approaches I applied enabled to provide optimized designs of the circuits and to take into account their burden to the cell. Importantly, it allowed me to design possible scenarios over a range of conditions and foresee the effects of the experimental implementations conducted in the laboratory by cooperation partners (Boto-Hernandez et al., unpublished). Although consistent results are reached by the strain with two copies of the killing switch, a further way to improve the safety could be the construction of a multi-level device. For example, the double kill switch could be complemented with riboswitches, riboregulators or other RNA tools. These RNA-based tools provided promising results in the biocontrol of cyanobacteria<sup>128-131</sup>. Riboswitches are regulatory RNA elements that, in response to binding of a specific ligand, undergo a conformational change, which finally results in an ON/OFF switch of gene expression; most of the ones in use are induced by theophylline<sup>131-134</sup>.

While the potential risks related to synthetic biology are thoroughly discussed, SbD is very rarely implemented at an R&D phase, and the strategy outcomes are not yet widely used commercially. Although the Scientific Committees of the European Commission warned on the risks related to synthetic biology and recommended the development of predictive tools and SbD in the design phase, there is still a regulatory gap and the market does not seem to follow these concerns<sup>135,136</sup>.

We are living a modern age where people get educated on the internet, often with no discernment between facts and unfounded opinions and where scientific knowledge is continuously discredited by *no-vax*, *no-mask* and a series of *no-X* movements. It is our best as scientist to show our knowledge is solid and that no matter the age where we live: the scientific method is always as exact as the limits of knowledge allows and up-to-date. Early-stage stakeholder engagement, rigorous scientific method and the (model-driven) deployment of rigorous risk assessment and risk management strategies provides a solid framework for the development of new biotechnology products and processes that contribute to meet the global challenges of our times.

## Concluding remarks

This thesis considers a set of modelling frameworks used to design, simulate and predict various aspects of the development of *Mycoplasma pneumoniae* as a universal vaccine chassis. I resolved the main objectives that the model approaches were called to address, but I believe more has to come. First of all, I think the model-driven medium design approach I have used should be a standard methodology for media development, to eliminate the need of serum-rich media. Secondly, I am confident that the scientific community will move more and more towards a predictive approach, where products' safety issues are not addressed when they show up, but foreseen and avoided in the developmental phase, through the implementation of biosafety strategies or, simply, by abandoning the product and look for a safer, improved version of it. Although there is still a lot of scepticism towards computational approaches, they might constitute the future for what concerns the design of biomedical and pharmaceutical products. Will one day laboratory testing be just a distant memory?

## Funding

This work has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement n. 634942 (MycoSynVac) and from the TTW-NWO under grant agreement n. 15814 (SafeChassis).

## References

1. Lerman, J. A. *et al.* In silico method for modelling metabolism and gene product expression at genome scale. *Nature Communications* **3**, 929 (2012).
2. Liu, J. K. *et al.* Reconstruction and modeling protein translocation and compartmentalization in *Escherichia coli* at the genome-scale. *BMC Systems Biology* **8**, 110 (2014).
3. Zhang, Y. *et al.* Three-Dimensional Structural View of the Central Metabolic Network of *Thermotoga maritima*. *Science* **325**, 1544 (2009).
4. Chang, R. L. *et al.* Structural Systems Biology Evaluation of Metabolic Thermotolerance in *Escherichia coli*. *Science* **340**, 1220 (2013).
5. Chang, R. L., Xie, L., Bourne, P. E. & Palsson, B. O. Antibacterial mechanisms identified through structural systems pharmacology. *BMC Systems Biology* **7**, 102 (2013).
6. Chandrasekaran, S. & Price, N. D. Probabilistic integrative modeling of genome-scale metabolic and regulatory networks in *Escherichia coli* and *Mycobacterium tuberculosis*. *Proceedings of the National Academy of Sciences of the United States of America* **107**, 17845–50 (2010).
7. Sánchez, B. J. *et al.* Improving the phenotype predictions of a yeast genome-scale metabolic model by incorporating enzymatic constraints. *Molecular Systems Biology* **13**, 935 (2017).
8. Mahadevan, R., Edwards, J. S. & Doyle III, F. J. Dynamic Flux Balance Analysis of Diauxic Growth in *Escherichia coli*. *Biophysical Journal* **83**, 1331–1340 (2002).
9. Karr, J. R. *et al.* A Whole-Cell Computational Model Predicts Phenotype from Genotype. *Cell* **150**, 389–401 (2012).
10. Smith, R. W., van Rosmalen, R. P., Martins dos Santos, V. A. P. & Fleck, C. DMPy: A Python package for automated mathematical model construction of large-scale metabolic systems. *BMC Systems Biology* **12**, (2018).
11. Jeske, L., Placzek, S., Schomburg, I., Chang, A. & Schomburg, D. BRENDA in 2019: a European ELIXIR core data resource. *Nucleic Acids Research* **47**, D542–D549 (2018).
12. Wittig, U., Rey, M., Weidemann, A., Kania, R. & Müller, W. SABIO-RK: an updated resource for manually curated biochemical reaction kinetics. *Nucleic Acids Research* 1–5 (2017) doi:10.1093/nar/gkx1065.
13. Flamholz, A., Noor, E., Bar-Even, A. & Milo, R. EQUilibrator - The biochemical thermodynamics calculator. *Nucleic Acids Research* **40**, 770–775 (2012).
14. Erdrich, P., Steuer, R. & Klamt, S. An algorithm for the reduction of genome-scale metabolic network models to meaningful core models. *BMC systems biology* **9**, 48 (2015).
15. Stapor, P. *et al.* PESTO: Parameter ESTimation TOolbox. *Bioinformatics* **34**, 705–707 (2017).
16. Fröhlich, F., Kaltenbacher, B., Theis, F. J. & Hasenauer, J. Scalable Parameter Estimation for Genome-Scale Biochemical Reaction Networks. *PLOS Computational Biology* **13**, e1005331 (2017).
17. Gaspari, E. *et al.* Model-driven design allows growth of *Mycoplasma pneumoniae* on serum-free media. *npj Systems Biology and Applications* **6**, 33 (2020).

18. Kamminga, T., Slagman, S.-J., Martins dos Santos, V. A. P., Bijlsma, J. J. E. & Schaap, P. J. Risk-Based Bioengineering Strategies for Reliable Bacterial Vaccine Production. *Trends in Biotechnology* **37**, 805–816 (2019).
19. Molina, L., Rosa, R. la, Nogales, J. & Rojo, F. *Pseudomonas putida* KT2440 metabolism undergoes sequential modifications during exponential growth in a complete medium as compounds are gradually consumed. *Environmental Microbiology* **21**, 2375–2390 (2019).
20. Parichehreh, R., Gheshlaghi, R., Mahdavi, M. A. & Elkamel, A. Optimization of lipid production in *Chlorella vulgaris* for biodiesel production using flux balance analysis. *Biochemical Engineering Journal* **141**, 131–145 (2019).
21. Pan, X., Streefland, M., Dalm, C., Wijffels, R. H. & Martens, D. E. Selection of chemically defined media for CHO cell fed-batch culture processes. *Cytotechnology* **69**, 39–56 (2017).
22. Traustason, B., Cheeks, M. & Dikicioglu, D. Computer-Aided Strategies for Determining the Amino Acid Composition of Medium for Chinese Hamster Ovary Cell-Based Biomanufacturing Platforms. *International Journal of Molecular Sciences* **20**, 5464 (2019).
23. Robitaille, J., Chen, J. & Jolicoeur, M. A Single Dynamic Metabolic Model Can Describe mAb Producing CHO Cell Batch and Fed-Batch Cultures on Different Culture Media. *PLOS ONE* **10**, e0136815- (2015).
24. Simon, L. & Karim, M. N. Modeling and control of amino acid starvation-induced apoptosis in CHO cell cultures. in *Proceedings of the 2002 American Control Conference (IEEE Cat. No.CH37301)* vol. 2 1579–1584 vol.2 (2002).
25. Zampieri, G., Coggins, M., Valle, G. & Angione, C. A poly-omics machine learning method to predict metabolite production in CHO cells. *Sciforum Electronic Conference Series* vol. 2.
26. Sengupta, N., Rose, S. T. & Morgan, J. A. Metabolic flux analysis of CHO cell metabolism in the late non-growth phase. *Biotechnology and Bioengineering* **108**, 82–92 (2011).
27. Quek, L.-E., Dietmair, S., Krömer, J. O. & Nielsen, L. K. Metabolic flux analysis in mammalian cell culture. *Metabolic Engineering* **12**, 161–171 (2010).
28. Niu, H., Amribt, Z., Fickers, P., Tan, W. & Bogaerts, P. Metabolic pathway analysis and reduction for mammalian cell cultures—Towards macroscopic modeling. *Chemical Engineering Science* **102**, 461–473 (2013).
29. Ahn, W. S. & Antoniewicz, M. R. Metabolic flux analysis of CHO cells at growth and non-growth phases using isotopic tracers and mass spectrometry. *Metabolic Engineering* **13**, 598–609 (2011).
30. Selvarasu, S. *et al.* Combined in silico modeling and metabolomics analysis to characterize fed-batch CHO cell culture. *Biotechnology and Bioengineering* **109**, 1415–1429 (2012).
31. Hefzi, H. *et al.* A Consensus Genome-scale Reconstruction of Chinese Hamster Ovary Cell Metabolism. *Cell Systems* **3**, 434–443.e8 (2016).
32. Calmels, C., McCann, A., Malphettes, L. & Andersen, M. R. Application of a curated genome-scale metabolic model of CHO DG44 to an industrial fed-batch process. *Metabolic Engineering* **51**, 9–19 (2019).
33. Fouladiha, H. *et al.* A metabolic network-based approach for developing feeding strategies for CHO cells to increase monoclonal antibody production. *bioRxiv* 751347 (2020) doi:10.1101/751347.

34. Higuera, G. A. *et al.* Patterns of Amino Acid Metabolism by Proliferating Human Mesenchymal Stem Cells. *Tissue Engineering Part A* **18**, 654–664 (2011).
35. Machado, H. *et al.* Strain-Specific Metabolic Requirements Revealed by a Defined Minimal Medium for Systems Analyses of *Staphylococcus aureus*. *Applied and Environmental Microbiology* **85**, e01773-19 (2019).
36. O'Brien, E. J., Lerman, J. A., Chang, R. L., Hyduke, D. R. & Palsson, B. Genome-scale models of metabolism and gene expression extend and refine growth phenotype prediction. *Molecular Systems Biology* **9**, (2013).
37. Nookaew, I. *et al.* The genome-scale metabolic model iIN800 of *Saccharomyces cerevisiae* and its validation: a scaffold to query lipid metabolism. *BMC Systems Biology* **2**, 71 (2008).
38. Fang, X., Lloyd, C. J. & Palsson, B. O. Reconstructing organisms in silico: genome-scale models and their emerging applications. *Nature Reviews Microbiology* **18**, 731–743 (2020).
39. O'Brien, E. J., Monk, J. M. & Palsson, B. O. Using genome-scale models to predict biological capabilities. *Cell* **161**, 971–987 (2015).
40. Massaiu, I. *et al.* Integration of enzymatic data in *Bacillus subtilis* genome-scale metabolic model improves phenotype predictions and enables in silico design of poly- $\gamma$ -glutamic acid production strains. *Microbial Cell Factories* **18**, 1–20 (2019).
41. Vu, T. T. *et al.* Genome-scale modeling of light-driven reductant partitioning and carbon fluxes in diazotrophic unicellular cyanobacterium *Cyanospora* sp. ATCC 51142. *PLoS Comput Biol* **8**, e1002460 (2012).
42. Baart, G. J. E. *et al.* Modeling *Neisseria meningitidis* metabolism: from genome to metabolic fluxes. *Genome Biology* **8**, R136 (2007).
43. Tejera, N. *et al.* Genome-Scale Metabolic Model Driven Design of a Defined Medium for *Campylobacter jejuni* M1cam. *Frontiers in Microbiology* **11**, 1072 (2020).
44. van der Ark, K. C. H. *et al.* Model-driven design of a minimal medium for *Akkermansia muciniphila* confirms mucus adaptation. *Microbial biotechnology* **11**, 476–485 (2018).
45. Munroe, S., Sandoval, K., Martens, D. E., Sipkema, D. & Pomponi, S. A. Genetic algorithm as an optimization tool for the development of sponge cell culture media. *In Vitro Cellular & Developmental Biology - Animal* **55**, 149–158 (2019).
46. Teusink, B. *et al.* Analysis of Growth of *Lactobacillus plantarum* WCFS1 on a Complex Medium Using a Genome-scale Metabolic Model. *Journal of Biological Chemistry* **281**, 40041–40048 (2006).
47. Oberhardt, M. A. *et al.* Harnessing the landscape of microbial culture media to predict new organism-media pairings. *Nature Communications* **6**, 8493 (2015).
48. van der Valk, J. *et al.* Optimization of chemically defined cell culture media – Replacing fetal bovine serum in mammalian in vitro methods. *Toxicology in Vitro* **24**, 1053–1063 (2010).
49. Jochems, C. E. A., van der Valk, J. B. F., Stafleu, F. R. & Baumans, V. The Use of Fetal Bovine Serum: Ethical or Scientific Problem? *Alternatives to Laboratory Animals* **30**, 219–227 (2002).

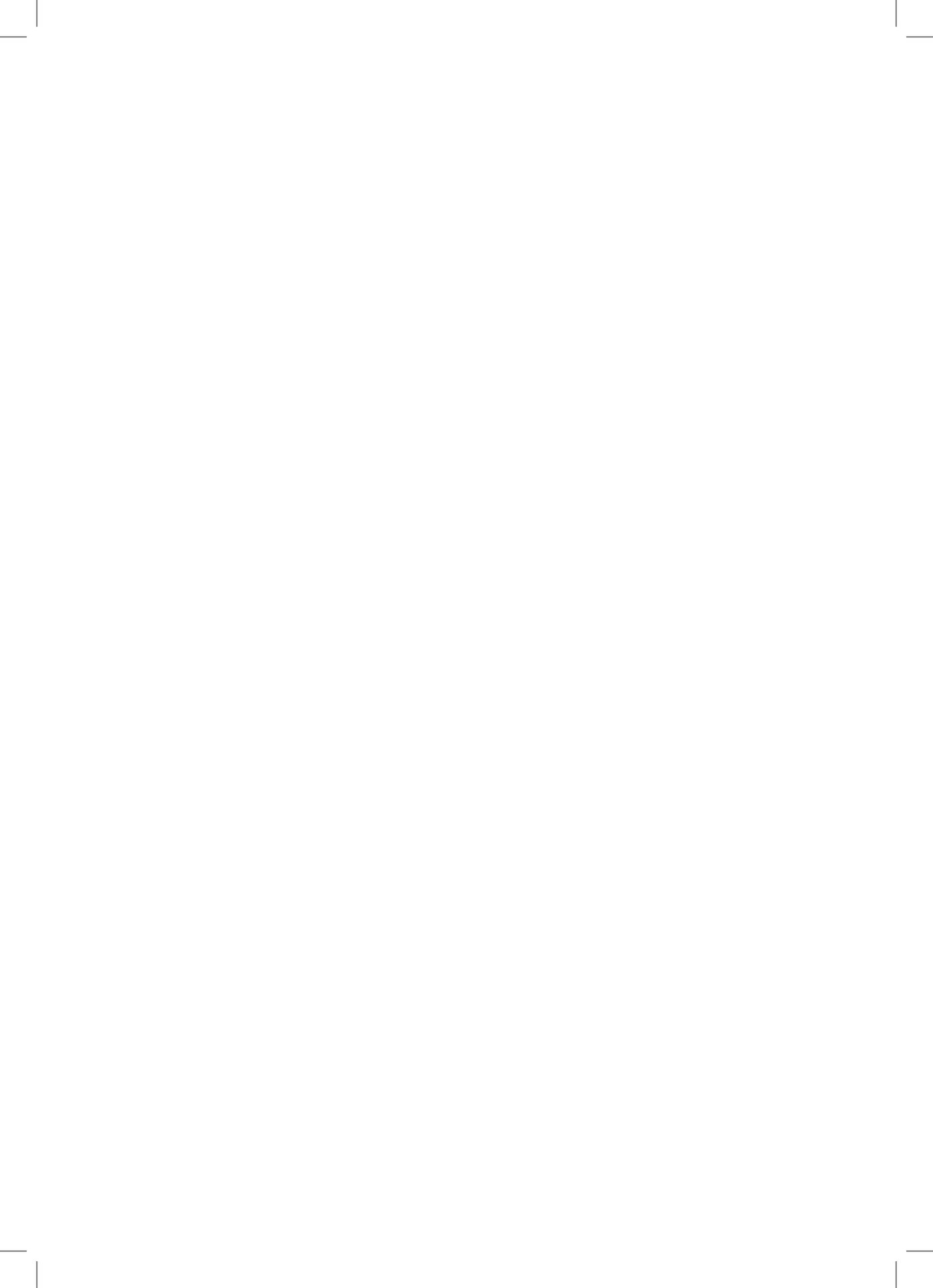
50. Mellor, D. J. & Gregory, N. G. Responsiveness, behavioural arousal and awareness in fetal and newborn lambs: experimental, practical and therapeutic implications. *New Zealand Veterinary Journal* **51**, 2–13 (2003).
51. Kamminga, T. *et al.* Metabolic modeling of energy balances in *Mycoplasma hyopneumoniae* shows that pyruvate addition increases growth rate. *Biotechnology and Bioengineering* **114**, 2339–2347 (2017).
52. Friis, N. F. Some recommendations concerning primary isolation of *Mycoplasma suis* pneumoniae and *Mycoplasma flocculare* a survey. *Nordisk veterinaermedicin* **27**, 337 (1975).
53. Müller, N., Riedel, M., Foderreuther, S., Blendinger, C. & Abele-Horn, M. Tourette's syndrome and mycoplasma pneumoniae infection. *American Journal of Psychiatry* **157**, 481-a (2000).
54. Müller, N. *et al.* *Mycoplasma pneumoniae* infection and Tourette's syndrome. *Psychiatry Research* **129**, 119–125 (2004).
55. Khan, F. Y. & A.yassin, M. *Mycoplasma pneumoniae* associated with severe autoimmune hemolytic anemia: case report and literature review. *Brazilian Journal of Infectious Diseases* **13**, 77–79 (2009).
56. Daxböck, F., Zedtwitz-Liebenstein, K., Burgmann, H. & Graninger, W. Severe hemolytic anemia and excessive leukocytosis masking mycoplasma pneumonia. *Annals of Hematology* **80**, 180–182 (2001).
57. Ercan, T. E., Ercan, G., Severge, B., Arpaozu, M. & Karasu, G. *Mycoplasma pneumoniae* infection and obsessive-compulsive disease: a case report. *Journal of child neurology* **23**, 338–340 (2008).
58. Agarwala, B. N. & Ruschhaupt, D. G. Complete heart block from mycoplasma pneumoniae infection. *Pediatric Cardiology* **12**, 233–236 (1991).
59. Scapini, J. P., Flynn, L. P., Sciacaluga, S., Morales, L. & Cadario, M. E. Confirmed *Mycoplasma pneumoniae* endocarditis. *Emerging infectious diseases* **14**, 1664–1665 (2008).
60. Kenney, R. T. *et al.* Mycoplasmal Pericarditis: Evidence of Invasive Disease. *Clinical Infectious Diseases* **17**, S58–S62 (1993).
61. Szymanski, M., Petric, M., Saunders, F. E. & Tellier, R. *Mycoplasma pneumoniae* Pericarditis Demonstrated by Polymerase Chain Reaction and Electron Microscopy. *Clinical Infectious Diseases* **34**, e16–e17 (2002).
62. Meseguer, M. A. *et al.* *Mycoplasma Pneumoniae* Pericarditis and Cardiac Tamponade in a ten-year-old girl. *The Pediatric Infectious Disease Journal* **15**, (1996).
63. Tang, Y. *et al.* Kawasaki disease associated with *Mycoplasma pneumoniae*. *Italian journal of pediatrics* **42**, 83 (2016).
64. Lee, M. N. *et al.* *Mycoplasma pneumoniae* infection in patients with Kawasaki disease. *Korean Journal of Pediatrics* **54**, 123 (2011).
65. Ravin, K. A. *et al.* *Mycoplasma pneumoniae* and atypical Stevens-Johnson syndrome: a case series. *Pediatrics* **119**, e1002–e1005 (2007).
66. Olson, D. *et al.* Outbreak of *Mycoplasma pneumoniae*–associated Stevens-Johnson syndrome. *Pediatrics* **136**, e386–e394 (2015).
67. Mulvey, J. M., Padowitz, A., Lindley-Jones, M. & Nickels, R. *Mycoplasma pneumoniae* associated with Stevens Johnson syndrome. *Anaesthesia and intensive care* **35**, 414–417 (2007).

68. Quioc, J.-J. *et al.* Acute cholestatic hepatitis revealing *Mycoplasma pneumoniae* infection without lung involvement in an adult patient. *European Journal of Gastroenterology & Hepatology* **21**, (2009).
69. Grüllich, C., Baumert, T. F. & Blum, H. E. Acute *Mycoplasma pneumoniae* infection presenting as cholestatic hepatitis. *Journal of clinical microbiology* **41**, 514–515 (2003).
70. Daxboeck, F., Gattringer, R., Mustafa, S., Bauer, C. & Assadian, O. Elevated serum alanine aminotransferase (ALT) levels in patients with serologically verified *Mycoplasma pneumoniae* pneumonia. *Clinical Microbiology and Infection* **11**, 507–510 (2005).
71. Romero-Gómez, M. *et al.* Acute hepatitis due to *Mycoplasma pneumoniae* infection without lung involvement in adult patients. *Journal of Hepatology* **44**, 827–828 (2006).
72. Ishida, Y. *et al.* Hemophagocytic lymphohistiocytosis secondary to *Mycoplasma pneumoniae* infection. *Pediatrics International* **46**, 174–177 (2004).
73. Mizukane, R. *et al.* An Elderly Patient with Hemophagocytic Syndrome due to Severe *Mycoplasma* Pneumonia with Marked Hypercytokinemia. *Respiration* **69**, 87–91 (2002).
74. Bruch, L. A., Jefferson, R. J., Pike, M. G., Gould, S. J. & Squier, W. *Mycoplasma pneumoniae* infection, meningoencephalitis, and hemophagocytosis. *Pediatric Neurology* **25**, 67–70 (2001).
75. Pönkä, A. Arthritis associated with *Mycoplasma pneumoniae* infection. *Scandinavian journal of rheumatology* **8**, 27–32 (1979).
76. Ramirez, A. S. *et al.* Relationship between rheumatoid arthritis and *Mycoplasma pneumoniae*: a case–control study. *Rheumatology* **44**, 912–914 (2005).
77. Johnson, S. M., Bruckner, F. & Collins, D. Distribution of *Mycoplasma pneumoniae* and *Mycoplasma salivarium* in the synovial fluid of arthritis patients. *Journal of Clinical Microbiology* **45**, 953–957 (2007).
78. Ataee, R. A. *et al.* Simultaneous detection of *Mycoplasma pneumoniae*, *Mycoplasma hominis* and *Mycoplasma arthritidis* in synovial fluid of patients with rheumatoid arthritis by multiplex PCR. *Archives of Iranian Medicine* **18**, 0 (2015).
79. Khan, F. Y. & Sayed, H. Rhabdomyolysis associated with *Mycoplasma pneumoniae* pneumonia. *Hong Kong Med J* **18**, 247–249 (2012).
80. Weng, W.-C., Peng, S. S.-F., Wang, S.-B., Chou, Y.-T. & Lee, W.-T. *Mycoplasma pneumoniae*–associated transverse myelitis and rhabdomyolysis. *Pediatric neurology* **40**, 128–130 (2009).
81. Oishi, T. *et al.* Rhabdomyolysis Associated with Antimicrobial Drug–Resistant *Mycoplasma pneumoniae*. *Emerging infectious diseases* **18**, 849 (2012).
82. Yatsyshina, S. *et al.* Detection of respiratory pathogens in pediatric acute otitis media by PCR and comparison of findings in the middle ear and nasopharynx. *Diagnostic microbiology and infectious disease* **85**, 125–130 (2016).
83. Soběslavský, O., Syrůček, L., Brůčková, M. & Abrahamovič, M. THE ETIOLOGICAL ROLE OF *MYCOPLASMA PNEUMONIAE* IN OTITIS MEDIA IN CHILDREN. *Pediatrics* **35**, 652 (1965).
84. Weinstein, O., Shneck, M., Levy, J. & Lifshitz, T. Bilateral acute anterior uveitis as a presenting symptom of *Mycoplasma pneumoniae* infection. *Canadian Journal of Ophthalmology* **41**, 594–595 (2006).

85. Salzman, M. B., Sood, S. K., Slavin, M. L. & Rubin, L. G. Ocular Manifestations of Mycoplasma pneumoniae Infection. *Clinical Infectious Diseases* **14**, 1137–1139 (1992).
86. Saïd, M.-H. *et al.* Mycoplasma pneumoniae-associated nephritis in children. *Pediatric Nephrology* **13**, 39–44 (1999).
87. Siomou, E., Kollios, K. D., Papadimitriou, P., Kostoula, A. & Papadopoulou, Z. L. Acute nephritis and respiratory tract infection caused by Mycoplasma pneumoniae: case report and review of the literature. *The Pediatric Infectious Disease Journal* **22**, (2003).
88. Narita, M. Pathogenesis of extrapulmonary manifestations of Mycoplasma pneumoniae infection with special reference to pneumonia. *Journal of Infection and Chemotherapy* **16**, 162–169 (2010).
89. Logan, D. C. Known knowns, known unknowns, unknown unknowns and the propagation of scientific enquiry. *Journal of Experimental Botany* **60**, 712–714 (2009).
90. Harris, J. B. Cholera: Immunity and Prospects in Vaccine Development. *The Journal of Infectious Diseases* **218**, S141–S146 (2018).
91. Verma, S. K. & Tuteja, U. Plague Vaccine Development: Current Research and Future Trends. *Frontiers in Immunology* **7**, 602 (2016).
92. Schragar, L. K., Harris, R. C. & Vekemans, J. Research and development of new tuberculosis vaccines: a review. *F1000Research* **7**, 1732 (2018).
93. Meiring, J. E., Giubilini, A., Savulescu, J., Pitzer, V. E. & Pollard, A. J. Generating the Evidence for Typhoid Vaccine Introduction: Considerations for Global Disease Burden Estimates and Vaccine Testing Through Human Challenge. *Clinical Infectious Diseases* **69**, S402–S407 (2019).
94. Odey, F., Okomo, U. & Oyo-Ita, A. Vaccines for preventing invasive salmonella infections in people with sickle cell disease. *Cochrane Database of Systematic Reviews* (2018) doi:10.1002/14651858.CD006975.pub4.
95. Donegan, S., Bellamy, R. & Gamble, C. L. Vaccines for preventing anthrax. *Cochrane Database of Systematic Reviews* (2009) doi:10.1002/14651858.CD006403.pub2.
96. Wang, X., Zhang, X., Zhou, D. & Yang, R. Live-attenuated Yersinia pestis vaccines. *Expert Review of Vaccines* **12**, 677–686 (2013).
97. Yun, S.-I. & Lee, Y.-M. Japanese encephalitis. *Human Vaccines & Immunotherapeutics* **10**, 263–279 (2014).
98. Griffin, D. E. Measles Vaccine. *Viral Immunology* **31**, 86–95 (2017).
99. Bandyopadhyay, A. S., Garon, J., Seib, K. & Orenstein, W. A. Polio vaccination: past, present and future. *Future Microbiology* **10**, 791–808 (2015).
100. Bruijning-Verhagen, P. & Groome, M. Rotavirus Vaccine: Current Use and Future Considerations. *The Pediatric Infectious Disease Journal* **36**, (2017).
101. Lambert, N., Strelbel, P., Orenstein, W., Icenogle, J. & Poland, G. A. Rubella. *The Lancet* **385**, 2297–2307 (2015).
102. Voigt, E. A., Kennedy, R. B. & Poland, G. A. Defending against smallpox: a focus on vaccines. *Expert Review of Vaccines* **15**, 1197–1211 (2016).

103. Marin, M., Marti, M., Kambhampati, A., Jeram, S. M. & Seward, J. F. Global Varicella Vaccine Effectiveness: A Meta-analysis. *Pediatrics* **137**, e20153741 (2016).
104. Monath, T. P. & Vasconcelos, P. F. C. Yellow fever. *Journal of Clinical Virology* **64**, 160–173 (2015).
105. Schmader, K. Herpes Zoster. *Annals of Internal Medicine* **169**, ITC17–ITC32 (2018).
106. Su, S.-B., Chang, H.-L. & Chen, K.-T. Current Status of Mumps Virus Infection: Epidemiology, Pathogenesis, and Vaccine. *International Journal of Environmental Research and Public Health* **17**, 1686 (2020).
107. Jefferson, T., Rivetti, A., di Pietrantonj, C. & Demicheli, V. Vaccines for preventing influenza in healthy children. *Cochrane Database of Systematic Reviews* (2018) doi:10.1002/14651858.CD004879.pub5.
108. de Lorenzo, V. Environmental biosafety in the age of Synthetic Biology: Do we really need a radical new approach? *BioEssays* **32**, (2010).
109. Gelinsky, E. & Hilbeck, A. European Court of Justice ruling regarding new genetic engineering methods scientifically justified: a commentary on the biased reporting about the recent ruling. *Environmental sciences Europe* **30**, 52 (2018).
110. Brodmann, T. *et al.* Safety of Novel Microbes for Human Consumption: Practical Examples of Assessment in the European Union. *Frontiers in Microbiology* **8**, 1725 (2017).
111. Hilbeck, A., Meyer, H., Wynne, B. & Millstone, E. GMO regulations and their interpretation: how EFSA's guidance on risk assessments of GMOs is bound to fail. *Environmental Sciences Europe* vol. 32 (2020).
112. Parker, M. T. & Kunjapur, A. M. Deployment of Engineered Microbes: Contributions to the Bioeconomy and Considerations for Biosecurity. *Health Security* **18**, 278–296 (2020).
113. Pei, L. & Schmidt, M. Fast-Growing Engineered Microbes: New Concerns for Gain-of-Function Research? *Frontiers in Genetics* **9**, 207 (2018).
114. Cichocka, D. *et al.* European Union research and innovation perspectives on biotechnology. *Journal of Biotechnology* **156**, 382–391 (2011).
115. Na, K. I. *A decade of EU-funded GMO research*. <http://ec.europa.eu/research/research-eu> (2001).
116. *European Commission-Daily News*.
117. van den Akker, H. C. M. & Wassenaar, | A L M. *Potential introduction of unapproved GM animals and GM products in the Netherlands*. [www.rivm.com](http://www.rivm.com) (2012).
118. Bouchez *et al.* Ecological study of a bioaugmentation failure. *Environmental Microbiology* **2**, 179–190 (2000).
119. Cases, I. & de Lorenzo, V. Genetically modified organisms for the environment: stories of success and failure and what we have learned from them. *International microbiology: the official journal of the Spanish Society for Microbiology* **8 3**, 213–22 (2005).
120. Yin, H. *et al.* Partial DNA-guided Cas9 enables genome editing with reduced off-target activity. *Nature Chemical Biology* **14**, 311–316 (2018).
121. Zhang, Q. *et al.* Potential high-frequency off-target mutagenesis induced by CRISPR/Cas9 in Arabidopsis and its prevention. *Plant molecular biology* **96**, 445–456 (2018).

122. Wang, F. & Zhang, W. Synthetic biology: Recent progress, biosafety and biosecurity concerns, and possible solutions. *Journal of Biosafety and Biosecurity* **1**, 22–30 (2019).
123. Marc, W., John, G. & Steven, H. Can Design Improve Construction Safety?: Assessing the Impact of a Collaborative Safety-in-Design Process. *Journal of Construction Engineering and Management* **131**, 1125–1134 (2005).
124. Donaldson, K., Murphy, F., Schinwald, A., Duffin, R. & Poland, C. A. Identifying the pulmonary hazard of high aspect ratio nanoparticles to enable their safety-by-design. *Nanomedicine* **6**, 143–156 (2010).
125. van de Poel, I. & Robaey, Z. Safe-by-Design: from Safety to Responsibility. *NanoEthics* **11**, 297–306 (2017).
126. Czornyj, E., Newcomer, D., Schroeder, I., Wayne, N. L. & Merlic, C. A. Proceedings of the 2016 Workshop Safety By Design – Improving safety in research laboratories. *Journal of Chemical Health & Safety* **25**, 36–49 (2018).
127. Asin-Garcia, E., Kallergi, A., Landeweerd, L. & Martins dos Santos, V. A. P. Genetic Safeguards for Safety-by-design: So Close Yet So Far. *Trends in Biotechnology* (2020)  
doi:<https://doi.org/10.1016/j.tibtech.2020.04.005>.
128. Higo, A., Isu, A., Fukaya, Y., Ehira, S. & Hisabori, T. Application of CRISPR Interference for Metabolic Engineering of the Heterocyst-Forming Multicellular Cyanobacterium *Anabaena* sp. PCC 7120. *Plant and Cell Physiology* **59**, 119–127 (2018).
129. Taton, A., Ma, A. T., Ota, M., Golden, S. S. & Golden, J. W. NOT Gate Genetic Circuits to Control Gene Expression in Cyanobacteria. *ACS Synthetic Biology* **6**, 2175–2182 (2017).
130. Ueno, K., Tsukakoshi, K. & Ikebukuro, K. Riboregulator elements as tools to engineer gene expression in cyanobacteria. *Applied Microbiology and Biotechnology* **102**, 7717–7723 (2018).
131. Ma, A. T., Schmidt, C. M. & Golden, J. W. Regulation of Gene Expression in Diverse Cyanobacterial Species by Using Theophylline-Responsive Riboswitches. *Applied and Environmental Microbiology* **80**, 6704 (2014).
132. Topp, S. *et al.* Synthetic Riboswitches That Induce Gene Expression in Diverse Bacterial Species. *Applied and Environmental Microbiology* **76**, 7881 (2010).
133. Nakahira, Y., Ogawa, A., Asano, H., Oyama, T. & Tozawa, Y. Theophylline-Dependent Riboswitch as a Novel Genetic Tool for Strict Regulation of Protein Expression in Cyanobacterium *Synechococcus elongatus* PCC 7942. *Plant and Cell Physiology* **54**, 1724–1735 (2013).
134. Till, P., Toepel, J., Bühler, B., Mach, R. L. & Mach-Aigner, A. R. Regulatory systems for gene expression control in cyanobacteria. *Applied Microbiology and Biotechnology* vol. 104 1977–1991 (2020).
135. Epstein, M. M. & Vermeire, T. Scientific Opinion on Risk Assessment of Synthetic Biology. *Trends in Biotechnology* **34**, 601–603 (2016).
136. EFSA Scientific Committee, More, S, Bampidis, V, Benford, D, Bragard, C, Halldorsson, *et al.*, 2020. Scientific Opinion on the evaluation of existing guidelines for their adequacy for the microbial characterisation and environmental risk assessment of microorganisms obtained through synthetic biology. *EFSA Journal* 2020; **18**(10):6263, 50 pp. <https://doi.org/10.2903/j.efsa.2020.6263>





# Synopsis





## Summary

The work I carried out in this thesis is part of a multinational effort that aimed to develop a universal chassis for the vaccination of a range of animals, including humans, pets and farm animals ([www.mycosynvac.eu](http://www.mycosynvac.eu)). With this thesis, I worked towards providing an engineering foundation for the design and production of *Mycoplasma pneumoniae* as a vaccine chassis. Since this and other mycoplasmas have fastidious growth requirements and must be grown in media containing animal serum, the focus of **chapter 2** and **chapter 3** is on exploring ways to enable their growth at a rate comparable to that obtained on serum-rich media, through the development of a serum-free medium and a re-design of the strain itself. Achieving this would greatly facilitate (and reduce the cost of) large-scale production of the vaccine. In **chapter 4** I focus on ascertaining possible adverse effects caused by the vaccine into the host, whereas in **chapter 5** I propose various biocontainment strategies to prevent the proliferation of the vaccine strain, outside the desired settings.

In **chapter 2**, I used genome-scale, constraint-based modelling of metabolism to design the first serum-free medium that supports the growth of *Mycoplasma pneumoniae* with a rate comparable to the one obtained in rich medium. Specifically, I updated an existing genome-scale metabolic model of *Mycoplasma pneumoniae* by introducing genes and reactions from recent annotations, fine-tuning the non-growth associated energy distribution, and expanding it by integrating a membrane assembly pathway with the detailed composition of its lipids. In this way, the medium supplementation was linked to the membrane formation, which is a key aspect of the surviving of *Mycoplasma pneumoniae*. Therefore, the assumptions related the medium requirements reflected the importance of the membrane lipids. Through model simulations, a serum-free medium MCMyco was experimentally developed and supported robust growth of *Mycoplasma pneumoniae in vitro*. This work was published and formed the basis for a submitted PCT patent application.

To enable its growth on serum-free medium was not the only strategy to increase *Mycoplasma pneumoniae's* growth rate. This bacterium is a fatty-acid auxotroph, despite its need for fatty acids to build up its cell membrane. In **chapter 3**, I designed a strain of *Mycoplasma pneumoniae* able to synthesize mid-length fatty acids, and then integrated them (computationally) as acyl chains in the membrane lipids. For the computational design of this chassis, I combined the genome-scale modelling approach described above with a gap-filling algorithm, modified to find genetic modifications that would have increased the growth rate of *Mycoplasma pneumoniae*. Because of the inability of this strain to grow at a rate comparable with the one of the wild-type strain, I introduced *in silico* exogenous reactions from phylogenetically-close bacteria. Simulations revealed a bottleneck related to NADPH, one the cofactor mostly used in the fatty acid biosynthesis pathway, and suggested genetic modifications to be applied for a more efficient NADPH regeneration. The experimental implementation of these suggestions is currently being finalised by collaborators and the complete work is expected to result in a scientific publication.

*Mycoplasma pneumoniae* has been linked to post-infectious Guillain-Barré-Strohl syndrome (an autoimmune neurodegenerative disorder) through the synthesis of a glycolipid called galactocerebroside. In **chapter 4**, I reconstructed the biosynthesis pathway of this component and I carried out an investigation aimed at finding enzymes to target on to prevent its synthesis. I extended this analysis to all mycoplasmas that could synthesize this molecule and found that several of them do not have this potential. Therefore, such group of mycoplasmas could be considered safe for application as far as this condition is concerned. A protein motif that, when found in the enzyme synthesizing the glycolipid, could be an indication of synthesis capability in *Mycoplasmas* was built. This motif can be used as a signature to investigate on the potentiality of galactocerebroside synthesis in microorganisms. This work was the basis for a patent application and a scientific article.

In **chapter 5**, I developed mathematical models to analyse and sustain the experimental design of biosafety circuits to insert into the *Mycoplasma pneumoniae*'s genome, so to limit its growth when outside of the desired operating conditions. Different Ordinary Differential Equations modelling techniques were followed according to the desired predictions: when it was the circuit conformation to study, or the dose-responses of the organism and the circuit's elements respect to the inducer concentrations, the related equations system focused on the characteristics and evolution of the circuit components. To study escape mutants that could introduce mutations inactivating the circuit's effects, I used a population modelling approach. Finally, to predict growth curves' evolution over strain passages, I used regression models on the curve's parameters. The model yielded robust results with a killing switch inserted in the bacterial genome in double copy, which showed to be resistant to inactivating mutations over several strain passages. These promising results have been tested experimentally by collaborators and the joint outputs of the work are to be published in a scientific journal.

Finally, in **chapter 6**, I addressed the computational approaches used, I recapitulated the status of the vaccine and I discussed its future perspectives. Firstly, I compared the potential of genome-scale metabolic models with that of dynamic metabolic models. Subsequently I addressed and highlighted the importance of serum-free media in growth cultures, with a special focus on their model-driven design. Then I reflected on the safety of live-attenuated vaccines, with examples from various mycoplasmas. Finally, I proposed safety-by-design as a strategy to be implemented in the developmental phase of therapeutic and pharmaceutical products.

## Riassunto

Il lavoro svolto in questa tesi è frutto di una cooperazione internazionale, la quale ha unito gli sforzi per sviluppare un "telaio" (*chassis*) universale per la vaccinazione dell'uomo, di animali domestici e da fattoria contro la polmonite atipica ([www.mycosynvac.eu](http://www.mycosynvac.eu)). Con questa tesi, ho mirato a fornire fondamenti ingegneristici per il disegno e la produzione del batterio *Mycoplasma pneumoniae* come vaccino. Poiché questo batterio e altri micoplasmi hanno requisiti di crescita complicati e devono essere coltivati in medium ricchi di siero animale, lo scopo del **capitolo 2** e del **capitolo 3** è l'esplorazione di strategie che permettano la crescita dei micoplasmi su di un medium non contenente siero animale, oltre a ridisegnare geneticamente il ceppo vaccinale. Il raggiungimento di questo traguardo non solo facilita la produzione in larga scala del vaccino, ma ne riduce notevolmente i costi. Nel **capitolo 4**, ho focalizzato l'attenzione sulla predizione di possibili reazioni avverse causate dal vaccino nell'uomo. Nel **capitolo 5**, ho proposto misure di biocontenimento per prevenire la proliferazione del ceppo vaccinale al di fuori delle condizioni desiderate.

Nel **capitolo 2**, ho utilizzato una tecnica di modellizzazione del metabolismo batterico su scala genomica per sviluppare un medium senza siero animale che permetta la coltivazione di *Mycoplasma pneumoniae* con un tasso di crescita comparabile a quello ottenuto in medium contenenti siero animale. Nello specifico, ho aggiornato ed ampliato un modello metabolico esistente, introducendovi geni e reazioni metaboliche ottenute da annotazioni più recenti, ritoccandone la distribuzione energetica and espandendolo tramite l'integrazione dei pathway responsabili dell'assemblaggio della membrana batterica, con composizione dettagliata dei lipidi. In questo modo, ho creato la connessione tra il supplemento del medium e la formazione della membrana batterica, aspetto che si è rivelato un elemento chiave per permettere la sopravvivenza e la crescita di *Mycoplasma pneumoniae*. Tramite simulazioni modellistiche, è stato quindi sviluppato il medium "MCMyc", non contenente siero animale, il quale permette una crescita robusta dei micoplasmi. Questo lavoro è stato pubblicato su Nature e la proprietà intellettuale è protetta da brevetto.

Lo sviluppo del medium di coltivazione non è stata l'unica strategia adottata per incrementare il tasso di crescita di *Mycoplasma pneumoniae*. Questo batterio è auxotrofo per gli acidi grassi (ovvero, non è in grado di sintetizzarli), nonostante la necessità di questi nel processo di formazione della membrana. Pertanto, nel **capitolo 3**, ho disegnato, tramite simulazioni di modificazioni genetiche, un ceppo di *Mycoplasma* che abbia le capacità di sintetizzare acidi grassi, per poi integrarli come catene aciliche nei lipidi della membrana. Come modelli computazionali, ho combinato un approccio di modellizzazione su scala genomica con un algoritmo di *gap-filling* (letteralmente: "riempimento degli spazi vuoti"), il quale è stato sviluppato in modo da fare emergere modificazioni genetiche che permettano al batterio di crescere più velocemente. Ho quindi simulato l'inserzione di reazioni esogene nel genoma micoplasmatico, ovvero reazioni proprie di altri batteri filogeneticamente vicini ai micoplasmi. Le simulazioni hanno rivelato le difficoltà dei micoplasmi nel rigenerare il cofattore NADPH,

utilizzato principalmente dagli enzimi che sintetizzano acidi grassi, suggerendone la relativa modificazione genetica per ottenere una rigenerazione più efficiente.

*Mycoplasma pneumoniae* è stato correlato all'insorgenza della sindrome di Guillain-Barré-Strohl, una malattia autoimmune e neurodegenerativa, tramite la sintesi di un glicolipide chiamato galactocerebroside. Nel **capitolo 4**, ho ricostruito il percorso di biosintesi di questo componente in *Mycoplasma pneumoniae*; inoltre, ho individuato gli enzimi da modificare per prevenire la formazione del galactocerebroside. Ho esteso questa analisi a tutti i micoplasmi, rilevando che molti di essi non hanno il potenziale per sintetizzare tale glicolipide. Il gruppo di micoplasmi così identificati può quindi considerarsi sicuro per applicazioni biomediche e farmaceutiche sull'uomo. Inoltre, ho ricostruito una sequenza proteica che possa servire da "etichetta", indicatrice del fatto che l'organismo in possesso di enzimi contenenti tale sequenza sarebbe in grado di produrre galactocerebroside, con conseguente problema di sicurezza in caso di applicazioni biomediche e farmaceutiche sull'uomo. Questo lavoro è protetto da brevetto.

Nel **capitolo 5**, ho sviluppato modelli matematici che avessero lo scopo di analizzare e sviluppare circuiti di biosicurezza da inserire nel genoma di *Mycoplasma pneumoniae* per limitare la sua crescita qualora questa avvenga fuori dalle condizioni desiderate. Diverse tecniche di modellizzazione basate su equazioni differenziali sono state utilizzate con gli obiettivi di: studiare la configurazione ottimale del circuito e la sua risposta a diverse dosi di induttori; predire la formazione di ceppi batterici che abbiano la capacità di introdurre mutazioni che vanificherebbero gli effetti del circuito di sicurezza; simulare la crescita batterica dei ceppi contenenti il circuito di sicurezza. Per raggiungere il primo obiettivo, ho utilizzato dei modelli matematici che si focalizzano sulle caratteristiche dei circuiti di biosicurezza; per il secondo, ho invece modellizzato la popolazione batterica; infine, per il terzo, ho utilizzato delle tecniche di regressione sui parametri delle curve di crescita batterica. Tramite questi modelli, ho ottenuto la conformazione ottimale della strategia di biosicurezza: un circuito introdotto nel genoma batterico in doppia copia, che ha dimostrato resistenza alle mutazioni per un tempo sufficientemente lungo.

Infine, nel **capitolo 6**, ho esaminato tutti gli approcci computazionali usati, ricapitolato lo stato corrente del vaccino e discusso le prospettive future. Ho quindi confrontato il potenziale dei modelli su scala genomica con quello dei modelli metabolici dinamici; successivamente, ho evidenziato l'importanza del medium sviluppato. Ho poi introdotto una riflessione sulla sicurezza dei vaccini vivi attenuati, fornendo esempi da vari micoplasmi. In conclusione, ho proposto l'utilizzo di strategie di *safety-by-design* ("sicurezza in base alla progettazione") nelle fasi di sviluppo di prodotti farmaceutici e terapeutici.

## Resumen

El trabajo realizado en esta tesis es el resultado de la cooperación internacional, que ha combinado esfuerzos para desarrollar un "chasis" universal para la vacunación de una serie de animales, incluidos los seres humanos, las mascotas y los animales de granja ([www.mycosynvac.eu](http://www.mycosynvac.eu)). Con esta tesis, pretendía proporcionar bases de ingeniería para el diseño y producción de la bacteria *Mycoplasma pneumoniae* como vacuna. Dado que esta bacteria y otros micoplasmas tienen complicados requisitos de crecimiento, y deben cultivarse en medios ricos en suero animal, el propósito del **capítulo 2** y del **capítulo 3** es explorar estrategias que permitan el crecimiento de micoplasmas en un medio que no contenga suero animal, así como rediseñar genéticamente la cepa de vacunación. Lograr este objetivo no sólo facilita la producción a gran escala de vacunas, sino que también reduce significativamente sus costos. En el **capítulo 4**, me centro en predecir posibles reacciones adversas causadas por la vacuna en humanos. En el **capítulo 5**, propongo medidas de biocontención para evitar la proliferación de la cepa de vacuna fuera de las condiciones deseadas.

En el **capítulo 2**, utilicé una técnica de modelado del metabolismo bacteriano a escala genómica para desarrollar un medio libre de suero animal, que permitiera el cultivo de *Mycoplasma pneumoniae* con una tasa de crecimiento comparable a la obtenida en medios que contienen suero animal. Específicamente, actualicé y amplifiqué un modelo metabólico existente, introduciendo genes y reacciones metabólicas obtenidas de anotaciones más recientes, retocando su distribución de energía y expandiéndola a través de la integración de vías responsables de ensamblar la membrana bacteriana, con su detallada composición de lípidos. De esta manera, creé la conexión entre el suplemento del medio y la formación de la membrana bacteriana, que resultó ser un elemento clave en la supervivencia y el crecimiento de *Mycoplasma pneumoniae*. A través de simulaciones de modelos, se desarrolló el medio "MCMyc", que no contiene suero animal, lo que permite un crecimiento robusto in vitro. Este trabajo ha sido publicado en la revista científica Nature y su propiedad intelectual protegida por patente.

El desarrollo del medio de cultivo no fue la única estrategia adoptada para aumentar la tasa de crecimiento de *Mycoplasma pneumoniae*. Esta bacteria es auxotrófica a los ácidos grasos (es decir, no es capaz de sintetizarlos), a pesar de la necesidad de estos en el proceso de formación de sus membranas. Así, en el **capítulo 3**, diseñé, a través de simulaciones de modificación genética, una cepa de *Mycoplasma* que tenía la capacidad de sintetizar ácidos grasos, y luego integrarlos como cadenas acíclicas en lípidos de membrana. Para el diseño computacional de este chasis, combiné un enfoque de modelado a escala genómica con un algoritmo de *gap-filling* (que fue adaptado de tal manera que predecía las modificaciones genéticas responsables de un mayor crecimiento de *Mycoplasma pneumoniae*). Entonces simulé la inserción de reacciones exógenas en el genoma micoplasmático, que se producen en bacterias filogenéticamente cercanas a la *Mycoplasma*. Las simulaciones revelaron las dificultades de los micoplasmas en la regeneración del cofactor NADPH, utilizado

principalmente por enzimas que sintetizan ácidos grasos, sugiriendo su modificación genética para lograr una regeneración más eficiente.

*Mycoplasma pneumoniae* se ha relacionado con la aparición del síndrome de Guillain-Barré-Strohl, una enfermedad autoinmune y neurodegenerativa, a través de la síntesis de un glicolípido llamado galactocerebrosido. En el **capítulo 4**, reconstruí la ruta de la biosíntesis de este componente; además, investigué para encontrar enzimas que modificar para evitar la formación de galactocerebrosido. Extendí este análisis a todos los micoplasma, encontrando que muchos de ellos no tienen el potencial de sintetizar este glicolípido. Por lo tanto, este grupo de micoplasmas puede considerarse seguro para aplicaciones humanas. Además, reconstruí una secuencia de proteínas que podría servir como una "etiqueta", una indicación de que el organismo que contiene enzimas con esa secuencia sería capaz de producir galactocerebrosido, resultando en un problema de seguridad en caso de aplicaciones humanas. Este trabajo está protegido por patente.

En el **capítulo 5**, desarrollé modelos matemáticos que analizarían y desarrollarían circuitos de bioseguridad para insertarlos en el genoma de *Mycoplasma pneumoniae* para limitar su crecimiento si se produce fuera de las condiciones deseadas. Se han utilizado varias técnicas de modelado basadas en ecuaciones diferenciales con los objetivos de: estudiar la configuración óptima del circuito y su respuesta a diferentes dosis de inductores; predecir la formación de cepas bacterianas que tienen la capacidad de introducir mutaciones que anularían los efectos del circuito de seguridad; simular el crecimiento bacteriano de cepas que contienen el circuito de seguridad. Para el primero, utilicé modelos matemáticos que se centraban en las características de los circuitos de bioseguridad; para el segundo, modelé la población bacteriana en su lugar; finalmente, para la tercera utilicé técnicas de regresión en los parámetros de las curvas de crecimiento bacteriana. A través de estos modelos, obtuve la conformación óptima de la estrategia de bioseguridad: un circuito introducido en el genoma bacteriano por duplicado, que demostró resistencia a mutaciones durante un tiempo suficientemente largo.

Por último, en el **capítulo 6** examiné todos los enfoques computacionales utilizados, resumí el estado actual de la vacuna y analicé las perspectivas futuras. Entonces comparé el potencial de los modelos a escala genómica con el de los modelos metabólicos dinámicos; más tarde, destacué la importancia del medio desarrollado. A continuación, presenté una reflexión sobre la seguridad de las vacunas vivas atenuadas, proporcionando ejemplos de varios micoplasmas. En conclusión, propuse el uso de estrategias de seguridad basada en el diseño en las fases de desarrollo de productos farmacéuticos y terapéuticos.





## About the author





## Publications & Patents

**Erika Gaspari**, Annika Franke, Diana Robles-Diaz, Robert Zweigerdt, Ingo Roeder, Thomas Zerjatke and Henning Kempf. "Paracrine mechanisms in early differentiation of human pluripotent stem cells: insight from a mathematical model". *Stem Cell Research*, 32, 1-7 (2018).

**Erika Gaspari**, Antoni Malachowski, Luis Garcia-Morales, Raul Burgos, Luis Serrano, Vitor A.P. Martins dos Santos and Maria Suarez-Diez. "Model-driven design allows growth of *Mycoplasma pneumoniae* on serum-free media". *NPJ Syst Biol Appl*, 6, 33 (2020).

**Erika Gaspari**, Jasper J. Koehorst, Joachim Frey, Vitor A.P. Martins dos Santos and Maria Suarez-Diez. "Galactocerebroside biosynthesis pathways of *Mycoplasma* species: an antigen triggering Guillain-Barré-Strohl syndrome". *Submitted for publication in Microbial Biotech*.

Alicia Broto-Hernandez, **Erika Gaspari**, Robert W. Smith, Maria Suarez-Diez, Vitor A.P. Martins dos Santos and Mark Isalan. "A biosafety circuit designed to limit *Mycoplasma* growth". *Prepared for submission*.

Rita Seabra on **Gaspari et al., 2020**: "Modelling helps to design serum-free medium for optimal *in vitro* culture of *Mycoplasma pneumoniae*" (2020). *News & Views, Alternatives to Laboratory Animals*, 1-4.

Patent Application EU19205723.0 (28<sup>th</sup> October 2020). "Serum-free *Mycoplasma* growth medium". Inventors: **Erika Gaspari**, Luis Garcia-Morales, Vitor A.P. Martins dos Santos. Applicant: Wageningen Universiteit.

Provisional Patent Application EU20174842.3. "Safe, attenuated *Mycoplasma* strains". Inventors: **Erika Gaspari**, Joachim Frey, Vitor A.P. Martins dos Santos. Applicant: Wageningen Universiteit.



## Overview of completed training activities

### Discipline-specific activities

NAME OF THE COURSE/MEETING	ORGANIZING INSTITUTE(S)	CITY, COUNTRY	YEAR
<b>MYCOSYNVAC 1<sup>ST</sup> GENERAL MEETING</b>	INRAE	Bordeaux, France	2016
<b>MYCOSYNVAC 2<sup>ND</sup> GENERAL MEETING</b>	MSD	Boxmeer, the Netherlands	2017
<b>MYCOSYNVAC 3<sup>RD</sup> REVIEW MEETING</b>	CRG	Barcelona, Spain	2019
<b>MYCOSYNVAC 1<sup>ST</sup> REVIEW MEETING</b>	European Commission	Brussels, Belgium	2018
<b>MYCOSYNVAC 2<sup>ND</sup> REVIEW MEETING</b>	European Commission	Brussels, Belgium	2020
<b>BIOSB 2018</b>	BioSB	Lunteren, the Netherlands	2018
<b>SSBSS 2018</b>	SSBSS group directors	Siena, Italy	2018
<b>ICSB 2019</b>	OIST	Okinawa, Japan	2019
<b>WHOLE-CELL MODELLING</b>	CRG	Barcelona, Spain	2017
<b>BIG DATA IN THE LIFE SCIENCE</b>	VLAG	Wageningen, the Netherlands	2017

### General courses

NAME OF THE COURSE/MEETING	ORGANIZING INSTITUTE(S)	CITY, COUNTRY	YEAR
<b>VLAG PHD WEEK</b>	VLAG	Baarlo, the Netherlands	2017
<b>MACHINE LEARNING A-Z: HANDS-ON PYTHON AND R IN DATA SCIENCE</b>	Udemy	Online	2020
<b>CAREER ORIENTATION</b>	Hertz / VLAG	Wageningen, the Netherlands	2020

## Optionals

NAME OF THE COURSE/MEETING	ORGANIZING INSTITUTE(S)	CITY, COUNTRY	YEAR
<b>PREPARATION OF RESEARCH PROPOSAL</b>	WUR, SSB	Wageningen, the Netherlands	2016-2020
<b>PHD STUDY TOUR (ORGANIZER)</b>	WUR, MIB/SSB	North Europe (DE-DK-SE)	2017
<b>WEEKLY GROUP MEETINGS (CSB)</b>	WUR, SSB	Wageningen, the Netherlands	2016-2020

## Teaching Obligations

NAME OF THE COURSE	YEAR
<b>MOLECULAR SYSTEMS BIOLOGY</b>	2017
<b>MOLECULAR SYSTEMS BIOLOGY</b>	2018
<b>MOLECULAR SYSTEMS BIOLOGY</b>	2019

## Student supervision

NAME OF THE STUDENT	PHD YEAR	MSC/BSC
<b>ANTONI MALACHOWSKI</b>	2017-2018	MSc
<b>HIDDE BLOKZIJL</b>	2017-2018	BSc
<b>IMIE NIEUWLAND</b>	2018-2019	MSc





## Curriculum Vitae

Erika Gaspari was born on the 21<sup>st</sup> of August 1990 in Varzi, a small town in the province of Pavia, near Milan, Italy. Three are the things she loves the most: science, adventures and spending time with friends and family. She lived in Bressana Bottarone, in the province of Pavia, until 2009, when she finished her scientific high-school studies. In the same year, she moved to Rome, until 2011 when she started her bachelor studies in Biotechnology, at the University of Pavia. After graduating in 2014, she moved to Bologna to enrol in the international master in Bioinformatics. In Bologna, she spent the happiest years of her life, and she considers this city her favourite place in the world. During her studies, she has worked as waitress in several restaurants, as high-school and university student's helper – mainly in mathematics, and as tourist entertainer in a resort in Salento, in the South of Italy, during the summer. In 2016, she worked as an intern on mathematical models of stem cell treatments in Dresden, Germany, for 6 months, before going back to Bologna for her master graduation in July 2016. In October 2016, she started as PhD candidate at Wageningen University & Research, working on the projects MycoSynVac and SafeChassis. Since November 2020, she has a position as Project Officer at the European & Developing Countries Clinical Trials Partnership (EDCTP) in The Hague.





# Acknowledgements





## Thank you...

I suppose this is the most read section of the thesis, so I tried to make it nice and complete, and above all not to forget anyone.

Those of the PhD have not been four simple years for me, both from a personal and a work point of view, but if I managed to complete them I must especially thank two people: **Vitor**, your contribution was really important, if not indispensable. It took you a while to get to know me, but once you knew who I was, you believed in me more than anyone else. You stimulated my desire to keep working just at the moment I thought I had lost it, and you put me in many exciting projects, leaving me complete freedom to decide on the technical modalities. You have invested time, energy and money to understand and promote my work. For that, I can never thank you enough.

The other person without whom I would not have been able to complete my doctorate is **Luis**. If I hadn't started working with you at MycoSynVac during my second year of PhD, I would probably have quit. You are the other person who believed in me, and also the collaborator with whom I had the most pleasure to work. We understood each other immediately, and in fact over time you also became a very good friend. When they gave me the first draft of the cover of my thesis, there was a man in a lab coat. On first impact I thought "but I am a woman, and I do not wear a lab coat", but then I decided to leave it there, because the guy in the lab coat without whom this thesis would have never been completed is you! I'm going to miss working with you, you're a terrific scientist, as well as a great friend. Thank you Luis.

And now my girls: **Niru** and **Nhung**. I couldn't have asked for better office mates. How much we laughed, how many beautiful and bad moments we shared. For me, our friendship is really special: we are three women with a very different cultural background, born in very different places, and in different stages of life, but as if by magic we never noticed. We just noticed what we had in common, and the rest didn't matter anymore. You're among those people I want next to me all my life. **Niru**, can I say you're one of the most special persons I know? You are kind, sweet, caring, selfless, humble, funny, intelligent,... I can't find you a flaw. And that's why you have a wonderful man, **Rajaram**, by your side. Thank you guys for your friendship. **Nhung**, how much you've been through in these four years. Whoever, in your place, would have quit, but you're incredibly strong. I hope that all the bad times have already passed, and that from now on there is only room for happiness for you, because that is what you deserve. There is also a third girl of course: **Melanie**, how could you abandon us! We missed you so much... you are a special girl, and definitely one of the funniest people I know. Please don't ever change.

Oh and **Peter**, sorry for all the noise of our laughing, although perhaps one day you will admit we were your favorite office neighbours ever. Thank you for putting up with us.

**Maria**, thank you for being so patient with me. You taught me so much and much more than you think, both scientifically and personally. Thank you for always being punctual, organized and for having an answer to every problem. You're a fantastic supervisor.

I also have to say a huge thank you to four people who introduced me to the world of Systems Biology. **Prof. Peverali**, you were the first to see my potential in this field. Thank you for supporting and valuing me. **Prof. Martelli**, you are the one that most of all made me passionate about this field. I can say with certainty that your teachings have been decisive in my career choice. Thank you for all your help. **Walter**, it's your fault that I ended up in the Netherlands: after meeting you, I thought all the Dutch were like you, in fact I would say I was very much hoping for it. I miss you so much, and I really miss working with you. **Thomas**, the man who hates social networks and without WhatsApp, I have been sending you emails for months. I associate you and Walter with the best memories of my months in Dresden. Thank you for everything you taught me, and thank you for your patience as a supervisor. I wish you every success, I hope that one day we will meet again.

There is a group of friends that I was really lucky to meet at SSB, and that I have been missing a lot in the last part of my PhD. **Basti**, I absolutely have to start with you. You were my first friend at SSB, the first person who came up to me and immediately tried to make me feel at home, as part of the group. I was very lucky to meet you. You are a wonderful friend, always ready to help others. I hope I've been as good friend to you as you have been to me. I always think of you, I wonder if you're okay, and if you're happy. I miss you very much. **Maarten**, in the acknowledgments usually people write funny things about you, and they are right, because your funny part is the one that you notice right away. But today I want to go against the grain, and do some romance (first and last time, so remember it well). You're one of the people I missed the most at SSB, and also one of those I've been happiest to know. I consider you a good friend, and I hope you will never change. Well, then I remember when I completely covered your screen with post-it and when I circulated the picture of you sleeping at the SSB meeting, but this is another matter... **Benoit**, cazzo, you too have abandoned me. For the last year I had no one to complain with about the crap wine and food of the Netherlands. But don't worry, I went on complaining alone. I wish you three guys the best, and so much success in life.

**Ruben**, another of the people I was really lucky to meet, and who I've missed a lot. I have to thank you too if I finished the PhD: maybe I never told you, but when after a year I thought about giving up and taking a position elsewhere, what you told me was decisive to make me change my mind, to continue and do my utmost. You're a really good friend, thank you.

**Dorett**, maybe we interacted more after you left than when you were still at SSB. And once I got to know you a little bit more, I regretted not getting to know you better before. I remember when you told me you were cold. Maybe you're shy, but you're not cold at all, you know how to give a lot to people. You're a beautiful person, never underestimate yourself.

Among my colleagues, there are some with whom I have had the pleasure of collaborating. First of all **Rik**, colleague, but also friend. Smart ass, and great cook. The pragmatism personified. Sounds like the perfect combination. **Enrique**, it's nice to have someone around who always smiles, and who is always ready to have a laugh. You are a person who brings joy, and also a great scientist, as well as super creative (in fact, I copied you a couple of templates, but I never admitted it out of envy). I really enjoyed working with you. **Niels**, I actually took your place in MycoSynVac when you finished your PhD contract, but we still had time to work

together. I wonder if one day we will find ourselves discussing science again in front of a good Bordeaux... I wish you the best with your new business! **Jasper**, you're a funny guy and a hard worker. It was nice to work with you, especially when we pulled out a paper from a few hours of work. But let's not tell anyone. Oh wait...

My first student, **Antoni**! Well, you already know, my favorite one also. I don't know if it was you learning from me or the other way around... You're a guy with monstrous abilities, incredibly intelligent and sensitive. I wish you the best, and so much success!

I also want to thank my colleagues at SSB: **Will** (Weiqi), the first and only guy who managed to share the office with me, Niru, Nhung and Melanie without despairing. Worthy of praise for sure! You left before I could learn to pronounce your name correctly. Thank you for putting up with us in the office! **Bart**, they're going to make you a statue for your patience. **Cristina F.**, what a pity you joined us just before the covid crisis, I would have liked to spend more time together. **Sara M.**, you're another one of those people I was sorry to know little about. It would have been nice to share the office for longer. **Rob**, thank you for your contribution to this thesis on the work on biosafety circuits, and for supervising me in the first year of PhD. **Wasin**, it is always nice to bother you. **Edo**, the only Italian besides me in SSB before Cristina's arrival. Let's keep the flag up! **Lyon**, mortacci tua Lyon, they didn't fire you yet? Then you could possibly devote yourself to cooking (just joking of course, please don't do that). Ah, you still have to pay for my "mafiosi cousins" at the Christmas lunch... **Willemijn**, of course I didn't forget about you! Our super secretary, very organized and always available to help us. Thank you!

Thanks also to **Emma, Anna D., Anna M., Marta, Sanjee, Sabine, Sara B., Christos, Maria M., Jesse, Linde, Nikolas, Kal** and **Tom**. With some more, with others less, but we shared time, chatter, scientific discussions and events. I wish you success in life, during and after SSB.

I thank all the collaborators of MycoSynVac. It is impossible to list everyone, but I want to mention a few that, for me and for my work, made a difference. **Joachim**, if all the scientists were like you, the world would be a better place. You are for me an example of intelligence, acuity, humility and professionalism. It was really a honor to work with you, and a great pleasure to chat during meetings. **Tjerko**, first colleague, then collaborator, and finally opponent. In fact, I couldn't think of anyone better suited to understand this thesis in its entirety. Do you remember when Peter said we should take your doctoral thesis as an example? I actually did! I don't know if the thesis is as good as yours, but you will be the one to tell me. Also from MSD: **Jetta** and **Paul**, thanks also to you, for your professionalism and sympathy. And also for the life teachings - I'm sure Paul knows what I'm talking about. **Carole**, what a pity you couldn't be my opponent, I would have really loved it. It was a pleasure to meet you and work with you and **Alain. Alicia**, we've been working together all the years of my PhD. Yours is the first project I started working on, and the last chapter I wrote for the thesis. You're a phenomenal scientist, it remains a mystery to me how you did all that work on your own. You're great! Last but not least, the splendid Biofaction team (**Markus, Camillo, Sandra, Lei**). Surely the most original and creative company I know, you're really great, as well as an extremely nice group of people. Meeting you guys was one of the reasons I was happy to go to the MycoSynVac meetings.

**Frits**, working with you was really a great pleasure. You are so good and passionate about your work that you have made me passionate too, so much that I have been in doubt until the very last moment to pursue a career as a patent attorney. But you are not only an example of professionalism and a hard worker, you are also a very pleasant person, with whom you can talk for hours without ever being bored. Thank you!

The shoooots group! **Sven**, you're simply great. A great everything: a great friend, a great worker, a great father, a great stubborn, a great deal of fun. Even though we now live far away, there is always a way to see each other and spend an evening together as in the old days, and I hope that continues to be the case. **Cristina G.**, you went up there in the Dutch colonia and we see each other very little, but I hope to see you soon and that you continue to rock your job. **Alex**, my favorite Russian. Also the only one I know actually, but that's not the point. A brute with the heart of a teddy bear. A big hug to you! **Irene**, sometimes I regret not being part of MIB just because I can't work with you. You're great! **Juanan**, we didn't meet when you were working in Wageningen, but during one of your many visits. What a pity we didn't meet before! Thanks also to you, MIB or ex-MIB: **Monika, Lara, Martijn, Jorrit, Jeroen, Gerben, Ran, Caifang, Catalina**. Thanks for making my stay in Wageningen more enjoyable. And thanks to **Hugo**, definitely one of the funniest Dutch I know! **Emmy**, you are such a kind and helpful person, thank you for being you. **Wen**, you are a great deal of fun, and a bit crazy: I love it. I miss having lunch with you and all the girls. I wish we could see each other more.

Grazie al gruppo di amici italiani a Wageningen: **Camilla, Francio, Enrico, Fabian, Lavinia, Marco, Nicolò, Riccardo, Stefano, Giulia** ed **Edo**. Ormai siamo tutti sparsi qua e là, ma non mi dimenticherò mai di voi. Quanto mi mancheranno le serate alla casa del popolo!

Ringrazio anche il gruppo di amici italiani a Utrecht: **Chiara, Enrico, Ambra, Carlo, Dario, Francesco, Loredana, Laura, Gabriela, Manuele, Mario, Soraya, Giulia, Matteo, Giampy, Gessica**. Diciamocelo, l'Olanda non ci piace, ma quando si ha un bel gruppo di amici tutti i posti diventano belli. Mi mancate.

I miei anni ad Utrecht non sarebbero stati gli stessi senza il fantastico team delle COI.ONE! Grazie ragazze: **Alessandra, Bea, Laura, Lucia, Monica, Stefania, Alessia, Arianna, Chiara, Claudia, Vinella, Marghe, Marta, Mylene, Rosa, Sonia F., Sonia C., Vania**, il presidente **Roberto** e il coach **Mario**. Nessuno aveva mai avuto il potere di farmi alzare la domenica mattina prima di voi. E infatti è successo solo un paio di volte. Urge una rimpatriata!

**Ioana e Ginevra**, le mie ex coinquiline ormai lontane. Anche voi mi mancate molto. È sempre bello rincontrarsi e raccontarsi le nuove avventure. Vi auguro tanta felicità nella vita, siete due amiche che mi porterò sempre nel cuore.

E meno male che ci siete voi, gli amici di lunga data, e quelli di sempre. **Angelica**, posso dire che l'unica gioia del 2020 sia stata quella di ritrovarti. E proprio il ritrovarti ha reso il momento di estremo dolore che ho passato in quei mesi più lieve. Passare di nuovo dei momenti con te, il **Carlo** e la **Barbara** mi ha riempito il cuore di gioia. Ti prego, non cambiare mai, ti voglio bene. **Mary**, ma alla fine, com'è che ci siamo conosciute? Secondo me c'è di mezzo quella cresta da galletta che portavi anni or sono...Devo aver pensato: "ma perché questa ancora non è amica mia?". E infatti non ci è voluto molto. Grazie amica, per esserci sempre. **Dani**, perché

non ci siamo conosciuti prima?! Ma che bello sarebbe stato? In un paio d'anni, sei diventato uno dei miei amici più stretti. Sono fortunata ad averti come amico, perché sei fantastico. Sei più forte di quello che pensi, e un giorno sarai felice come vuoi tu. Ti voglio bene.

**Martina**, da quanto tempo non ci vediamo. Sono così fiera di te amica mia: trasferirti negli Stati Uniti, uscire da una relazione che non ti rendeva felice, intraprendere una carriera faticosa ma soddisfacente. Mi sarebbe piaciuto averti accanto oggi mentre difendo il PhD, perché noi in queste occasioni ci siamo sempre l'una per l'altra. Maledetto covid! Ti voglio bene, spero di poterti riabbracciare presto.

Sono molto orgogliosa anche di te, **Eugenio**. Hai preso e te ne sei andato da un posto che non ti rendeva felice e non ti dava possibilità. Prima o poi riuscirò a venire a trovarti (però pure tu torna qualche volta eccheccazz ti sei barricato in UK?). Ti auguro tanta, tanta felicità.

**Markus**, my "little" Markus, we now know each other for many years, and back to Malta, I am very happy I met such a special friend as you are. Thank you.

**Michael**, ed eccoci qua, dalla tua vicina di banco in Università a Pavia al dottorato in Olanda. Questo sì che è essere vecchi. Tu chiamami pure Gargamella, io continuerò a circolare le tue foto mentre dormi a lezione, mentre una voce riecheggerà per sempre nelle aule dell'università: Negri ma io cosa devo dire? Banane!

**Michele**, grazie per esserci sempre per un bicchiere ed una chiacchierata quelle rare volte che torno a casa. Hai addirittura attraversato il Po per venire a trovarmi, immergendoti nel fantastico mondo dei paesanotti. Questa sì che è amicizia.

Il mondo è veramente piccolo, ma vi pare normale che ci siamo ritrovati qui in Olanda? **Giulia**, abbooonaaaa, il fatto che tu e **Nicco** abitate così vicini mi ha fatto passare un po' di nostalgia di Bologna. No non è vero, la nostalgia è rimasta, ma quando mi avete detto che vi sareste trasferiti qui ero felicissima. Chissà se faremo anche il prossimo trasferimento nello stesso paese, tocca mettersi d'accordo! Ma ovunque finiremo, spero che rimarremo amici per sempre. **Valeria, Ronci, Mattia**, certo che è strano rivedersi dopo 10 anni in Olanda. Quello che è ancora più strano è che quando ci vediamo è come se questi 10 anni non siano passati, se fossimo ancora a Roma al bar dopo delle lezioni. Mi siete mancati moltissimo, è stato bello ritrovarvi qui.

Se devo pensare agli anni più belli della mia vita, sicuramente penso a Bologna, specialmente per un gruppo di persone che l'hanno resa il mio posto preferito sulla faccia della Terra: i miei ex coinquilini **Simone, Nicola, Deborah, Vince, Giovannina, Carla** e **Antonio**. Vi penso sempre ragazzi, mi manca da morire vivere con voi. E poi gli ex compagni di corso: **Giulia, Niccolò, Alessandro, Andrea, Claudio, Francesco, Laura, Paolo, Stefano, Grant, Shadab, Atena** e anche **Giuseppe** (eri il nostro tutor, ma alla fine per noi sei sempre stato un amico, parte del gruppo). Quando penso a voi, la nostalgia è grande, ma i ricordi dei momenti felici di più.

If there is one person that I really have to thank more than anyone else is the love of my life, **Fran**. We met right at the beginning of my PhD. We never admitted it, but we fell in love right away. And then all the difficulties: you in Sweden, then in Spain, then in the United Arab

Emirates, and I always here in the Netherlands. But I couldn't stay that far away from you, and as in the most beautiful love stories, last year you left everything to come to me. You put your career aside for me to do mine. How can I ever thank you enough? Thank you for everything you've done for me, thank you for supporting me in times when the pain was huge, thank you for all the travels, adventures and experiences we've had together, and thank you for sharing the everyday life with me. Ti amo.

**Marisol y Paco**, hay el dicho de que nunca te llevas bien con tus suegros. Aquellos que afirman esto es porque nunca os han conocido, porque es imposible no amaros. Gracias por ser siempre tan serviciales y amorosos. **Andrea y Antonio**, gracias a vosotros también. El nacimiento del pequeño **Hernán** ha sido una inmensa alegría para mí, y no puedo esperar a ejercer mi papel de tía. Vais a ser unos grandes padres.

Ho voluto lasciare per ultima la mia famiglia, come per ritagliarmi un angolino e metterci le persone che amo di più. **Mamma, papà, nonna, zio**, se oggi divento dottoressa è anche grazie a voi, perché tutti gli sforzi che ho fatto, tutte le giornate che ho passato distrutta dalla stanchezza, tutti i libri che ho studiato, sono stati soprattutto per voi, perché foste orgogliosi di me. Siete sempre stati voi il motore che muoveva la mia ambizione, la ragione per cui ho puntato sempre più in alto. Mi avete trasmesso quei valori di impegno e di serietà verso il proprio lavoro che sono stati indispensabili per non mollare, anche quando pensavo di non farcela. **Nonno**, lo so che c'è anche il tuo zampino, soprattutto in questi ultimi mesi, quando ho dovuto terminare la tesi mentre ero sopraffatta dal dolore. Grazie per avermi dato la forza. Spero che oggi siate orgogliosi di me, come io lo sono di voi.

*E infine grazie a te, l'amore e il dolore più forti che abbia mai provato. Non avrei mai pensato che sarei stata capace di finire una tesi di dottorato, iniziare un nuovo lavoro, cambiare casa ed organizzare un matrimonio nello stesso momento. L'ho fatto per te, perché volevo che mi vedessi come una super mamma. Volevo essere la tua eroina, il tuo esempio. Un giorno ci rincontreremo bambino mio, non ti dimenticherò mai.*



The research described in this thesis was financially supported by the Laboratory of Systems and Synthetic Biology of Wageningen University and by the European Union's Horizon 2020 research and innovation program, under grant agreement No. 634942.

Financial support from Wageningen University for printing this thesis is gratefully acknowledged.

Cover design by Stefanie van der Herik, HerikMedia.

Thesis layout by Proefschriftmaken.nl

Printed by Proefschriftmaken.nl on FSC-certified recycled paper.