

From Nose to Lung: Using Systems Biology to Fight Pathogens in the Human Respiratory Tract

Dissertation

der Mathematisch-Naturwissenschaftlichen Fakultät

der Eberhard Karls Universität Tübingen

zur Erlangung des Grades eines

Doktors der Naturwissenschaften

(Dr. rer. nat.)

vorgelegt von

Alina Renz

aus Nagold

Tübingen

2021

Gedruckt mit Genehmigung der Mathematisch-Naturwissenschaftlichen Fakultät der
Eberhard Karls Universität Tübingen.

Tag der mündlichen Qualifikation: 13.05.2022

Dekan: Prof. Dr. Thilo Stehle

1. Berichterstatter: Jun.-Prof. Dr. Andreas Dräger

2. Berichterstatter: Prof. Dr.-Ing. Oliver Kohlbacher

3. Berichterstatter: Prof. Dr. Ursula Kummer

For my loved ones

Contents

Abbreviations	vii
Abstract	ix
Kurzfassung	xi
List of Accepted Publications	xiii
1 Introduction	1
2 Objectives	7
3 Results	9
3.1 Assessing the quality and predictive value of genome-scale metabolic models (GEMs)	10
3.2 Curating high-quality GEMs	16
3.3 Identifying potential antiviral targets against SARS-CoV-2	23
4 Discussion and Perspectives	27
Bibliography	31
A How to build high-quality GEMs	47
B Accepted publications	55
B.1 FBA reveals guanylate kinase as potential antiviral target against severe acute respiratory syndrome coronavirus type 2 (SARS-CoV-2)	55
B.2 First GEM of <i>D. pigrum</i> confirms multiple auxotrophies	65
B.3 GEM infected with SARS-CoV-2 mutants confirms guanylate kinase as robust target	87
B.4 Curating and comparing 114 strain-specific GEMs of <i>S. aureus</i>	104
B.5 An updated GEM of <i>P. aeruginosa</i> PA14 to characterize mucin-driven shifts	119

Abbreviations

ATP	adenosine triphosphate
BOF	biomass objective function
CDC	Centers for Disease Control and Prevention
CDM	chemically defined medium
CF	cystic fibrosis
COBRAPy	COntstraint-Based Reconstruction and Analysis for Python
COVID-19	coronavirus disease 2019
E	envelope
EC	Enzyme Commission
ECDC	European Centre for Disease Prevention and Control
ECO	Evidence and Conclusion Ontology
EGC	energy-generating cycles
FAIR	findable, accessible, interoperable, and reusable
FBA	flux balance analysis
FVA	flux variability analysis
GEM	genome-scale metabolic model
GK1	guanylate kinase 1
GPR	gene-protein-reaction
KEGG	Kyoto Encyclopedia of Genes and Genomes
M	membrane
MERS	Middle East respiratory syndrome
MERS-CoV	Middle East respiratory syndrome coronavirus
MIRIAM	minimum information required in annotation of models
MRSA	methicillin-resistant <i>Staphylococcus aureus</i>
N	nucleocapsid
NCBI	National Center for Biotechnology Information
RNA	ribonucleic acid
S	spike
SARS	severe acute respiratory syndrome
SARS-CoV-1	severe acute respiratory syndrome coronavirus type 1
SARS-CoV-2	severe acute respiratory syndrome coronavirus type 2
SBML	Systems Biology Markup Language
SBO	Systems Biology Ontology
SNM	synthetic nasal medium
TCA	tricarboxylic acid
URT	upper respiratory tract
VBOF	viral biomass objective function
VMH	virtual metabolic human

Abbreviations

VOC	variants of concern
VOI	variants of interest
WHO	World Health Organization

Abstract

In December 2019, the world was hit by the global SARS-CoV-2 pandemic. Two years later, the number of infections and deaths is still increasing, affecting everyday's life. Although several vaccines have been successfully approved for SARS-CoV-2, therapeutic strategies are still minimal. As viruses rely on the host's metabolism for replication, analyzing the viral reprogramming of the host cells might reveal potential antiviral targets. Such metabolic alterations can be evaluated and analyzed using genome-scale metabolic models (GEMs). These models represent large metabolic networks that connect metabolites with biochemical reactions facilitated by proteins and encoded by genes. With the help of genomic information, the so-called genotype, we can create metabolic models that can predict the phenotypic behavior of an organism. However, these GEMs can be used to analyze virus-host interactions and predict potential antiviral targets and understand the genotype-phenotype relationship of pathogens and commensals such as *Staphylococcus aureus*. High-quality models with a high predictive value help us to better understand an organism, determine metabolic capabilities in health and disease, identify potential targets for treatment interventions, and analyze the interplay between different cells and organisms. Such models can answer relevant and urgent questions of our time quickly and efficiently and become an indispensable constituent in future research.

In this thesis, I demonstrate (I) how the quality and predictive value of an existing genome-scale metabolic model can be assessed, (II) how high-quality genome-scale metabolic models can be curated, and (III) how high-quality genome-scale metabolic models can be used for model-driven discoveries.

All three points are addressed in the context of pathogens and commensals in the human respiratory tract. To assess the quality and predictive value of GEMs, we collected all currently available models of the pathogen *Staphylococcus aureus*, which colonizes the human nose. We evaluated the models concerning their validity, compliance with the FAIR data principle, quality, simulatability, and predictive value. Using high-quality models with a high predictive value enables model-driven hypotheses and discoveries. However, if no such model is available, one needs to curate a high-quality model. For this purpose, we developed a pipeline that focuses on the model curation of nasal pathogens and commensals. This pipeline is adaptable to incorporate other tools and bacteria, pathogens, or cells while maintaining certain community standards. We demonstrated the applicability of this pipeline by curating the first model of the nasal commensal *Dolosigranulum pigrum*. We showed how to use high-quality GEMs for model-driven discoveries by identifying novel antiviral targets. To do so, we virtually infected human alveolar macrophages in the lung with SARS-CoV-2.

Kurzfassung

Im Dezember 2019 wurde die Welt von der globalen SARS-CoV-2 Pandemie heimgesucht. Zwei Jahre später nimmt die Zahl der Infektionen und Todesfälle immer noch zu und beeinträchtigt das tägliche Leben der Menschen. Obwohl bereits mehrere Impfstoffe gegen SARS-CoV-2 zugelassen wurden, gibt es nur sehr wenige therapeutische Ansätze. Da Viren für ihre Replikation auf den Stoffwechsel des Wirts angewiesen sind, könnten Analysen der viralen Umprogrammierung der menschlichen Zellen mögliche antivirale Ziele aufzeigen. Solche Stoffwechselveränderungen können mithilfe von genom-skaligen Stoffwechselmodellen ausgewertet und analysiert werden. Diese Modelle stellen große Stoffwechselnetzwerke dar, die Metabolite durch biochemische Reaktionen miteinander verbinden, die durch Proteine ermöglicht und von Genen kodiert werden. Mit Hilfe genomischer Information, dem so genannten Genotyp, können wir also Modelle erstellen, die das phänotypische Verhalten eines Organismus vorhersagen können. Diese Stoffwechselmodelle können jedoch nicht nur für die Analyse von Virus-Wirt-Interaktionen und die Vorhersage potentieller antiviraler Ziele verwendet werden, sondern auch zum besseren Verständnis der Genotyp-Phänotyp-Beziehung von Pathogenen und Kommensalen wie *Staphylococcus aureus*. Qualitativ hochwertige Modelle mit einem hohen Vorhersagewert helfen uns, einen Organismus besser zu verstehen, Stoffwechselfähigkeiten in Gesundheit und Krankheit zu bestimmen, potenzielle Ziele für Behandlungsmaßnahmen zu identifizieren und das Zusammenspiel zwischen verschiedenen Zellen und Organismen zu analysieren. Solche Modelle können relevante und dringende Fragen unserer Zeit schnell und effizient beantworten und werden zu einem unverzichtbaren Bestandteil der künftigen Forschung.

In dieser Arbeit zeige ich, (I) wie die Qualität und der Vorhersagewert eines bestehenden genom-skaligen Modells bestimmt werden kann, (II) wie qualitativ hochwertige Stoffwechselmodelle auf Genomebene kuratiert werden können und (III) wie qualitativ hochwertige Stoffwechselmodelle für modellgestützte Entdeckungen genutzt werden können.

Alle drei Punkte werden im Zusammenhang mit Krankheitserregern und Kommensalen im menschlichen Respirationstrakt behandelt. Um die Qualität und den Vorhersagewert von Stoffwechselmodellen zu bewerten, haben wir alle derzeit verfügbaren Modelle des Erregers *Staphylococcus aureus* gesammelt, der in der menschliche Nase vorkommt. Wir bewerteten die Modelle hinsichtlich ihrer Gültigkeit, der Einhaltung des FAIR-Datenprinzips, der Qualität, der Simulierbarkeit und des Vorhersagewerts. Die Verwendung hochwertiger Modelle mit einem großen Vorhersagewert ermöglicht modellgestützte Hypothesen und Entdeckungen. Steht jedoch kein solches Modell zur Verfügung, muss ein hochwertiges Modell selbst erstellt werden. Zu diesem Zweck haben wir eine

Anleitung geschrieben, die sich auf die Modellerstellung von nasalen Pathogenen und Kommensalen konzentriert. Diese Anleitung ist anpassbar, um andere Programme, aber auch andere Bakterien, Krankheitserreger oder Zellen unter Einhaltung gewisser Standards einzubeziehen. Wir haben die Anwendbarkeit dieser Anleitung durch die Erstellung des ersten Modells des nasalen Kommensalen *Dolosigranulum pigrum* demonstriert. Darüber hinaus haben wir gezeigt, wie hochwertige Stoffwechselmodelle für modellgestützte Entdeckungen genutzt werden können, indem wir neue antivirale Targets identifiziert haben. Zu diesem Zweck haben wir menschliche Alveolarmakrophagen in der Lunge virtuell mit SARS-CoV-2 infiziert.

List of Accepted Publications

- 1. FBA reveals guanylate kinase as a potential target for antiviral therapies against SARS-CoV-2.**
Alina Renz, Lina Widerspick, and Andreas Dräger,
Bioinformatics Volume 36, Issue Supplement_2, Pages i813–i821 (2020).
<https://doi.org/10.1093/bioinformatics/btaa813>
- 2. First Genome-Scale Metabolic Model of *Dolosigranulum pigrum* Confirms Multiple Auxotrophies.**
Alina Renz, Lina Widerspick, and Andreas Dräger,
Metabolites 11(4):232 (2021).
<https://doi.org/10.3390/metabo11040232>
- 3. Genome-Scale Metabolic Model of Infection with SARS-CoV-2 Mutants Confirms Guanylate Kinase as Robust Potential Antiviral Target.**
Alina Renz, Lina Widerspick, and Andreas Dräger,
Genes 12(6), 796 (2021).
<https://doi.org/10.3390/genes12060796>
- 4. Curating and comparing 114 strain-specific genome-scale metabolic models of *Staphylococcus aureus*.**
Alina Renz and Andreas Dräger,
npj Systems Biology and Applications 7, 30 (2021).
<https://doi.org/10.1038/s41540-021-00188-4>
Publication in Appendix B on page 104 reproduced with permission from Springer Nature.
- 5. An updated genome-scale metabolic network reconstruction of *Pseudomonas aeruginosa* PA14 to characterize mucin-driven shifts in bacterial metabolism.**
Dawson D. Payne, Alina Renz, Laura J. Dunphy, Taylor Lewis, Andreas Dräger, and Jason A. Papin,
npj Systems Biology and Applications 7, 37 (2021).
<https://doi.org/10.1038/s41540-021-00198-2>
Publication in Appendix B on page 119 reproduced with permission from Springer Nature.

Ask five different astrophysicists to define a black hole, the saying goes, and you'll get five different answers. But ask five biomedical researchers to define systems biology, and you'll get 10 different answers . . . or maybe more.

— Christopher Wanjek

1

Introduction

In December 2019, the world was hit by a global pandemic: the novel SARS-CoV-2 caused the coronavirus disease 2019 (COVID-19). Almost two years later, the virus is still spreading, and the number of infections and deaths increases by the day. As of November 2021, more than 250 million cases and over 5 million deaths due to COVID-19 have been reported [1].

This is, however, by far not the first pandemic of mankind. One of the most famous and fatal pandemics was the Black Death, caused by the bacterium *Yersinia pestis*. It spread in Europe, Asia, and North Africa in the 14th century. During this pandemic, 30-60% of the European population deceased [2]. Also, the Spanish flu, caused by the influenza type A virus H1N1 in the early 20th century, led to approximately 17-50 million deaths worldwide [3, 4]. The Spanish flu was followed by the Asian flu pandemic of influenza A virus subtype H2N2 in 1957 and the Hong Kong flu of influenza A virus subtype H3N2 in 1968 [5].

The currently spreading SARS-CoV-2 belongs to the genus of β -*coronaviruses*. The members of this genus are enveloped, positive-sense, single-stranded RNA viruses. The genome of SARS-CoV-2 was first isolated and sequenced on January 5, 2020 [6], revealing its genome size of approximately 30 kB [7]. The virus has 14 open reading frames that encode for 26 proteins. Included in these 26 proteins are the four structural proteins: spike (S) protein, envelope (E) protein, membrane (M) protein, and nucleocapsid (N) protein. SARS-CoV-2 infects epithelial cells in the upper and lower respiratory tract [8, 9, 10] and immune cells such as alveolar macrophages [11]. Upon infection, most clinical presentations are mild and include fever, cough, malaise, myalgia, headache, as well as taste and smell disturbances. However, in some cases, pneumonia or severe lung, heart, liver, kidney, neurological, or gastrointestinal dysfunction are reported [12, 13].

The severe acute respiratory syndrome coronavirus type 1 (SARS-CoV-1) and the Middle East respiratory syndrome coronavirus (MERS-CoV) are prominent representatives of β -*coronaviruses* besides SARS-CoV-2: SARS-CoV-1 provoked the respiratory

disease severe acute respiratory syndrome (SARS) during an epidemic outbreak from 2003 to 2004, and MERS-CoV caused the Middle East respiratory syndrome (MERS) in an epidemic outbreak in 2012 [14]. After the SARS-CoV-1 epidemic, researchers warned about new genotypes and resulting outbreaks of coronaviruses due to their potential of genetic recombination, the large reservoir in bats, and the culture of eating exotic animals [15]. However, not only coronaviruses hold the potential for hazardous developments: the World Health Organization (WHO) is, for example, closely monitoring the changes in influenza viruses. For that purpose, the global influenza surveillance and response system and the pandemic influenza preparedness framework were initiated as an early warning system for flu outbreaks [16]. Besides influenza and coronaviruses, several other virus families, including α -viruses, filoviruses, henipaviruses, flaviviruses, and bunyaviruses, are considered to have a pandemic potential [17].

Not only viral infections hold the potential for worldwide pandemics. The rise of antibiotic resistance is one of the biggest global threats [18, 19, 20]. After Alexander Fleming identified penicillin in 1928, antibiotics have changed the approaches to treat infectious diseases, revolutionized modern medicine, and saved millions of lives [20, 21]. However, during the last decade, the proportion and the absolute number of multi-drug resistant pathogens have increased dramatically. National and international organizations such as the European Centre for Disease Prevention and Control (ECDC), the US Centers for Disease Control and Prevention (CDC), and the WHO consider infectious diseases by multi-drug resistant bacteria as an emergent global threat [19]. The WHO has published a list of bacteria for which new antibiotics are urgently needed. In this list, twelve so-called ‘priority pathogens’ are cataloged, posing the greatest threat to human health. Based on the urgency of new antibiotics, this list is divided into the three following categories: critical, high, and medium priority (see Table 1.1) [22].

One of the most critical bacterial pathogens of our time is *Pseudomonas aeruginosa*, together with *Acinetobacter baumannii* and *Enterobacteriaceae* species [22]. *A. baumannii* and *P. aeruginosa* are relevant pathogens in the human respiratory tract [23, 24]. *P. aeruginosa* is a Gram-negative opportunistic pathogen. It is one of the most common causes of nosocomial infections [28], causing wound, skin, and urinary tract infections, pneumonia, and sepsis [24]. *P. aeruginosa* is especially prevalent in patients with local or systemic immune deficits, such as burn wounds, cystic fibrosis, acute leukemia, or organ transplants [29]. It is challenging and complicated to treat these patients efficiently due to *P. aeruginosa*’s intrinsic resistance and its ability to acquire additional resistance mechanisms to various antimicrobial agents [28, 30]. It is also listed in the ECDC’s annual epidemiological report about antimicrobial resistance in the European Union. According to their report from 2019, a small but significant decreasing trend in resistances was noted for *P. aeruginosa* compared to 2015. However, high resistance percentages for single and multiple antibiotics persist: more than 30% of the *P. aeruginosa* isolates were resistant to at least one antimicrobial group under surveillance [30].

Besides *P. aeruginosa*, *Staphylococcus aureus* is highly prioritized in the urgency for new antibiotics [22]. Like *Pseudomonas*, *S. aureus* is an opportunistic pathogen. It

Table 1.1: WHO priority pathogens with urgency for new antibiotics [22]. The checkmark (✓) in the column ‘Respiratory relevance’ indicates whether the pathogen has a relevance in the respiratory tract.

Priority	Pathogen	Respiratory relevance
Priority 1: CRITICAL	<i>Acinetobacter baumannii</i>	✓ [23]
	<i>Pseudomonas aeruginosa</i>	✓ [24]
	<i>Enterobacteriaceae</i>	
Priority 2: HIGH	<i>Enterococcus faecium</i>	
	<i>Staphylococcus aureus</i>	✓ [25]
	<i>Helicobacter pylori</i>	
	<i>Campylobacter</i> spp.	
	<i>Salmonellae</i>	
Priority 3: MEDIUM	<i>Neisseria gonorrhoeae</i>	
	<i>Streptococcus pneumoniae</i>	✓ [26]
	<i>Haemophilus influenzae</i>	✓ [27]
	<i>Shigella</i> spp.	

asymptomatically colonizes the nose of approximately one-third of the population [25]. It is a commensal of the human skin and the mucosae but can also be the causative agent of severe bloodstream infections with high morbidity and mortality [31]. The methicillin-resistant form of *S. aureus*, known as MRSA, is one of the most successful modern pathogens as it is the leading cause of severe infections. Besides bacteremia, *S. aureus*, especially MRSA, can cause endocarditis, bone and joint infections, skin and soft tissue infections, and hospital-acquired infections. Almost any item with skin contact, ranging from white coats to pens and mobile phones, can serve as a fomite for MRSA transmission [32]. The ECDC reports a stabilizing or even decreasing percentage of MRSA in the majority of the countries under observation. Nonetheless, it remains an important pathogen in Europe and the US. Besides recommendations for prevention and control of the device- and procedure-associated infections and the prudent use of antimicrobial therapies, the development of new and novel prevention and treatment strategies are required [30, 33].

Despite the recent development of several new antimicrobial and antistaphylococcal drugs, *S. aureus*, and particularly the MRSA strains, remain major pathogens as they evolve and develop new resistance mechanisms [32, 34]. Thus, novel and non-conventional technologies are of high interest to address antimicrobial resistance, not only in *S. aureus* but also in *P. aeruginosa* and other pathogenic bacteria. Approaches range from blocking bacterial virulence factors, monoclonal antibodies to targeting host factors, microbiome interventions, and probiotics [35, 36].

In microbiome interventions, the effect of bacterial interaction and competition is used. For example, after administering broad-spectrum antibiotics, a large portion of

the commensal bacteria in the gut microbiome is eliminated, allowing opportunistic pathogenic bacteria to establish. Introducing a ‘healthy’ gut microbiome with commensal bacteria can help eliminating the pathogens by faster repopulation of the patient [35]. A similar approach is used for interventions with probiotics. Probiotic effects can be mediated by direct competitive exclusion of bacteria or the production of acids or inhibitors [37]. Studies have indicated that regular intake of probiotics reduced the nasal colonization with pathogenic bacteria, including *S. aureus* [38]. One bacterial commensal of interest in the human nasal microbiome is *Dolosigranulum pigrum*. Its presence in the nasopharyngeal passage is associated with a healthy upper respiratory tract. Together with *Corynebacterium pseudodiphtheriticum*, *D. pigrum* is considered a probiotic candidate [39]. Furthermore, the presence of *D. pigrum* is negatively associated with *S. aureus* in the human nasal microbiome, suggesting its inhibiting and antistaphylococcal effect [40]. However, the overall molecular mechanisms behind the bacterial interaction and inhibition remain mostly unclear.

These non-conventional approaches can not only be used to combat antibiotic resistance but also to fight SARS-CoV-2. During the current pandemic, the development of vaccines gained a lot of attention and governmental sponsorship, leading to successful approvals of several vaccines. To this date, however, the number of approved therapeutic strategies against SARS-CoV-2 is minimal. As viruses rely on the host’s metabolism for their replication, the non-conventional approach of targeting host factors seems reasonable. Upon infection, the host cell is reprogrammed to produce viral components, resulting in metabolic changes. These metabolic changes can be analyzed and used to identify potential targets for antiviral therapies. Additionally, targeting host factors holds the benefit of robustness: While targeting capsid proteins of enveloped viruses with small molecules works for some viral infections [41, 42], only a few mutational events might be sufficient for the virus to develop a resistance against the drug. However, viruses are less able to develop resistances against antivirals targeting the host [43].

As the metabolism of human cells is rather complex, the effect of individual changes on the whole metabolic system upon viral infection is challenging to analyze in detail. The same accounts for the metabolism of individual bacteria and the even more complex interaction between commensals and pathogens. One way to understand the complex interplay of genes, proteins, reactions, compounds, and other cellular components is to use systems biology. Systems biology studies biological systems at the system level [44]. It aims to understand the interaction of cellular components, how they form networks, and how they function in the whole cell to generate observable phenotypes [45]. Thus, systems biology does not focus on the components or cellular molecules and their function as in molecular biology; instead, it focuses on the links and relationships that connect these cellular components to form a biochemical reaction network [46]. One of the major modeling approaches for metabolic studies at a systems level is the reconstruction and use of genome-scale metabolic models (GEMs). A GEM mathematically represents the metabolism of an organism. Its reconstruction is facilitated by the advances in high-throughput techniques and whole-genome sequencing in the last 10 to 15 years

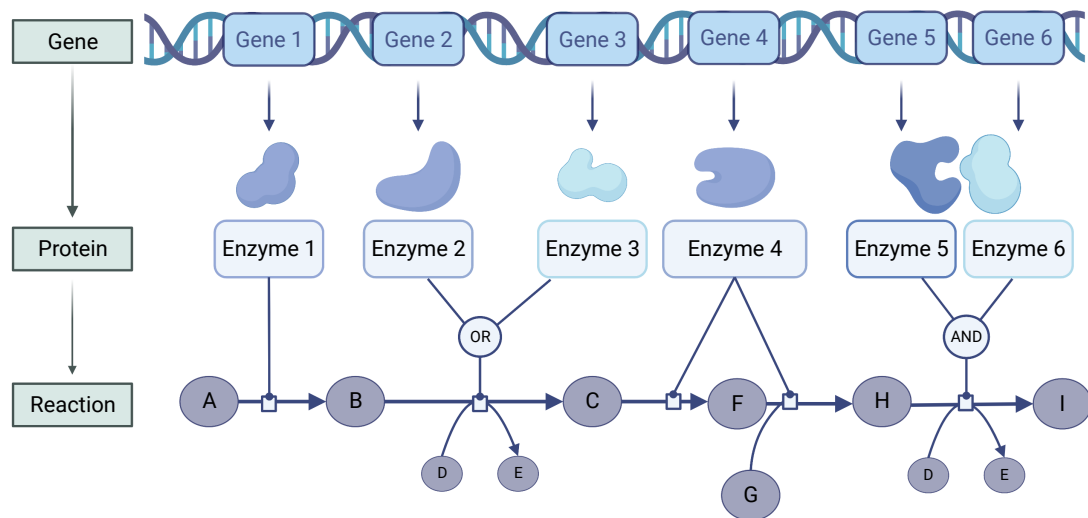


Figure 1.1: The concept of GPRs. In so-called gene-protein-reaction (GPR) associations, the genes on the genome are linked to their gene products, also referred to as proteins. Proteins that act as biocatalysts are also referred to as enzymes. These enzymes facilitate biochemical reactions. An enzyme can catalyze either one reaction (Enzyme 1) or multiple reactions (Enzyme 4). The latter are also known as promiscuous enzymes. Enzymes 2 and 3 are isozymes as both can catalyze the second reaction in the shown network. Enzymes 5 and 6 form an enzyme complex and are both required to catalyze the final reaction. This Figure was created with BioRender.com.

[47, 46]. The information obtained by whole-genome sequencing forms the basis for the GEMs and, thus, understanding the genotype-phenotype relationship. In so-called gene-protein-reaction (GPR) associations, the genes on the genome are linked to their gene products, also referred to as proteins. These proteins, in turn, facilitate biochemical reactions. Various biochemical metabolic reactions are linked via the compounds or metabolites, forming a more extensive network (see Figure 1.1). Eventually, the large metabolic network based on the genotype can simulate the phenotypic behavior of the studied organism [46]. In addition to the genome sequence, experimentally obtained biochemical information, transcriptomics, proteomics, metabolomics, and fluxomics can be incorporated to formulate stoichiometry-based and mass-balanced metabolic reactions.

Looking at the GEM from a network point of view, the nodes represent the compounds, and the reactions represent the edges (see bottom graph of Figure 1.1). Besides the graphical representation of the network in the form of a map, the network can also be represented mathematically in a stoichiometric matrix S . In this matrix, rows list the model's compounds and metabolites while columns represent the model's reactions. The stoichiometric coefficients of this matrix are integer numbers that indicate the number of metabolites participating in the respective reactions. Constraints need to be defined for the metabolic model to simulate the behavior of a GEM *in silico* realistically. As GEMs follow physiochemical laws and rely on environmental conditions, constraints

need to be applied. A physiochemical constraint is the flux balance. As we assume a steady-state for the internal metabolites, all fluxes through the network need to be balanced. In a steady-state, the concentration of the metabolites would not change over time. Thus, the multiplication of the stoichiometric matrix S with a flux vector v needs to equal zero [48, 46]. To follow the basic laws of thermodynamics, another constraint claims that the overall energy E needs to be balanced and conserved [49]. Additional constraints define the bounds of a reaction in the network and regulate the environmental availability of nutrients based on enzyme or transporter capacity and thermodynamics. These constraints can be mathematically formulated as:

- (a) flux balance: $S \cdot v = 0$,
- (b) energy balance: $\Delta E = 0$,
- (c) enzyme or transporter capacity: $v_i \leq v_{\max}$,
- (d) thermodynamics: $0 \leq v_{\min} \leq v_i$.

The constraints span a convex polytope that contains all steady-state flux distributions. An objective function needs to be defined to identify a particular functional state within the solution space. One of the most essential and interesting groups of objectives comprises functions representing expected physiological functions. One such physiological function is the maximization of biomass production and, thus, the growth of an organism. Others include minimizing the adenosine triphosphate (ATP) production or nutrient uptake, or maximizing metabolite production [46, 50]. Using biomass maximization as the objective function and constraints, the model can be optimized using flux balance analysis (FBA) to predict the growth behavior in different environments or conditions.

To summarize, constraint-based reconstruction and analysis of genome-scale metabolic models can be used to predict metabolic alterations. These metabolic changes can, for example, be caused by infections of host cells with a virus. The changes can be analyzed and utilized to identify potential antiviral targets. Moreover, bacterial metabolism can change in different environments. These environments can, for example, be the human nose or lung. The bacterial metabolism is also altered when interacting with other commensals or pathogens. Tracking and understanding the metabolic changes in different conditions will pave the way towards new antiviral and antimicrobial therapeutic strategies. With the help of GEMs and systems biology approaches, we have tools at hand to understand and fight the upcoming challenges in future pandemics and antibiotic resistance.

Nothing in life is to be feared, it is only to be understood. Now is the time to understand more, so that we may fear less.

— Marie Curie

2

Objectives

Systems biology in general and genome-scale metabolic models, in particular, are powerful tools to investigate metabolic changes and potential novel antimicrobial or antiviral targets. As indicated previously, *S. aureus* is a relevant pathogen colonizing the human nose. Due to its high prevalence in the population, antibiotic resistance, and infectious potential, its investigation is crucial. Several *S. aureus* models have already been published, however, with varying quality. To investigate the pathogenicity, resistance mechanisms, and potential antimicrobials, we need a high-quality model with a good predictive value. To identify such a model, we need quality measures for evaluating different aspects of GEMs, such as compliance with standards or the inclusion of laboratory data. High-quality GEMs with high predictive value will be the key to combating antimicrobial resistance and identifying novel treatment strategies using systems biology. One novel treatment strategy is to use the probiotic effect of commensals. *D. pigrum*, for example, is negatively associated with *S. aureus* in the human nose. While several models of *S. aureus* existed but no GEM of *D. pigrum*, we needed to curate a high-quality model of this commensal to enable the investigation of microbial interactions. Having high-quality models at hand facilitates model-driven discoveries for urgent questions of our time, treatment options and antiviral targets in a worldwide pandemic.

Thus, this thesis aims at answering the following three questions:

- (I) How can we assess the quality and predictive value of an existing genome-scale metabolic model?
- (II) How can we curate high-quality genome-scale metabolic models?
- (III) How can we use high-quality genome-scale metabolic models for model-driven discoveries, such as identifying novel antiviral targets against SARS-CoV-2?

All three questions will be answered in the context of pathogens in the human respiratory tract. For questions (I) and (II), we will focus on the upper respiratory tract,

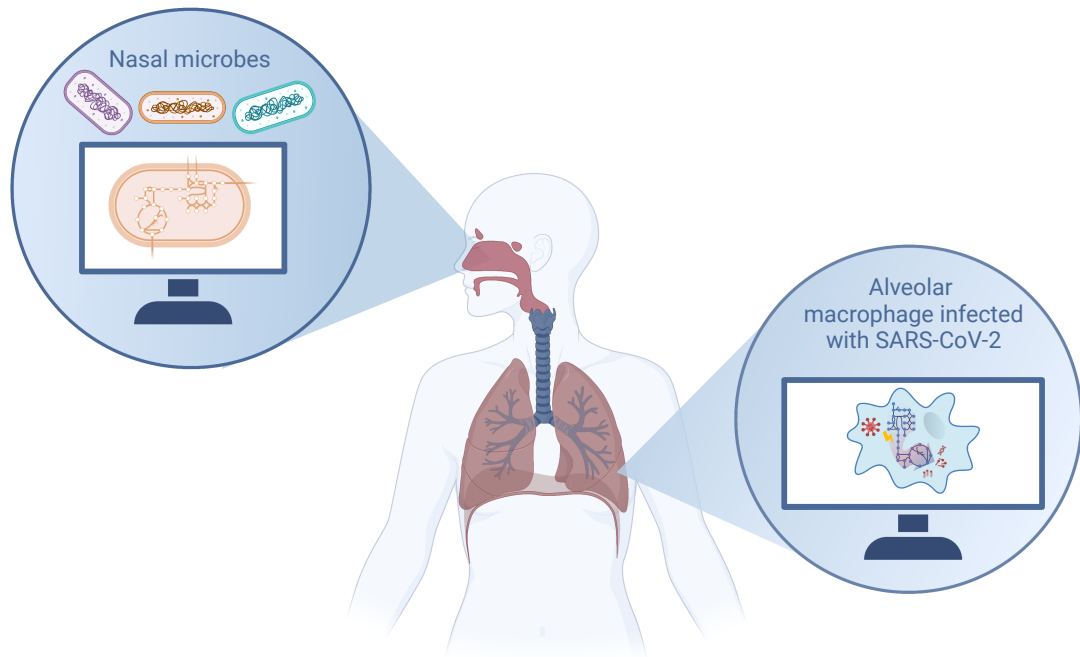


Figure 2.1: Focus of this thesis. In the first and second part of this thesis, we will focus on the upper respiratory tract, namely the nose and human nasal microbes. We will answer the questions (I) how we can assess the quality and predictive value of an existing GEM and (II) how we can curate high-quality GEMs. In the third part of the thesis, we will focus on the lower respiratory tract and (III) investigate potential antiviral targets against SARS-CoV-2. This Figure was created with BioRender.com.

specifically on the human nose and its inhabitants. We will shift the focus from bacteria in the upper respiratory tract to viral infections in the lower respiratory tract to answer question (III). With this transition of the pathogen and infection site, the versatile and adaptable applicability of the methods in systems biology is demonstrated. In Figure 2.1, the human pathogens and the infection or colonization sites are illustrated.

In fact, biology is chaos. Biological systems are the product not of logic but of evolution, an inelegant process. Life does not choose the logically best design to meet a new situation. It adapts what already exists... The result, unlike the clean straight lines of logic, is often irregular, messy.

— John M Barry

3

Results

This chapter consists of three parts. We will focus on each of the three questions stated in Chapter 2. The answers to the questions are part of the following publications:

- (I) How can we assess the quality and predictive value of an existing genome-scale metabolic model?
Curating and comparing 114 strain-specific genome-scale metabolic models of *Staphylococcus aureus*.
- (II) How can we curate high-quality genome-scale metabolic models?
 - a) **First Genome-Scale Metabolic Model of *Dolosigranulum pigrum* Confirms Multiple Auxotrophies.**
 - b) **An updated genome-scale metabolic network reconstruction of *Pseudomonas aeruginosa* PA14 to characterize mucin-driven shifts in bacterial metabolism.**
- (III) How can we use high-quality genome-scale metabolic models for model-driven discoveries, such as identifying novel antiviral targets against SARS-CoV-2?
 - a) **FBA reveals guanylate kinase as a potential target for antiviral therapies against SARS-CoV-2.**
 - b) **Genome-Scale Metabolic Model of Infection with SARS-CoV-2 Mutants Confirms Guanylate Kinase as Robust Potential Antiviral Target.**

3.1 Assessing the quality and predictive value of GEMs

The first genome-scale metabolic model represents the metabolism of *Haemophilus influenzae* published in 1999 [51]. It was followed by one of the best-studied bacteria, *Escherichia coli* in 2000 [52]. The first eukaryotic reconstruction was available three years later for *Saccharomyces cerevisiae* [53]. Since then, the number of modeled organisms and curated reconstructions has increased steadily [46]. The scope of reconstructions is not limited anymore to prokaryotes or single-cell eukaryotes. In 2007, the publication of the first human reconstruction Recon1 [54] paved the way for cell- and tissue-specific metabolic models that can be combined to multi-tissue models. In 2020, the first sex-specific whole-body metabolic reconstructions became available. It comprises 26 organs and six blood cell types [55]. As the number of models increased, the necessity for model databases emerged. Today, various databases exist, such as the BiGG Models database [56], BioModels database [57], including the Path2Models project [58], and the virtual metabolic human (VMH) database [59]. The BiGG Models database currently comprises 108 metabolic reconstructions. The BioModels database contains over 1000 manually curated and non-curated mathematical models of biological and biomedical systems, respectively, and 833 auto-generated models. Of these models, 168 are constraint-based models. Within the Path2Models project, over 140,000 models were automatically generated from pathway resources. The VMH database includes one human and 818 microbe reconstructions. In addition to the mentioned model databases, models can be found and obtained from the supplementary material of relevant scientific publications [60].

The available models are of varying quality. Some comply with the FAIR data principle: they are Findable, Accessible, Interoperable, and Reusable. Especially the upload to a model database and assignment of a specific identifier facilitates the first data principle. By uploading the model to a database and making it publicly available, the model is, in turn, accessible to the research community. Making models available upon request hampers the accessibility, especially when contact information is missing or outdated. The community standard file format for constraint-based reconstructions is Systems Biology Markup Language (SBML) [61]. This format assures the exact representation of a model in different tools and removes translational errors. Thus, SBML files can be imported, read, interpreted, and exported by different software systems, making SBML models interoperable. Using SBML with annotations facilitates the model's reusability as users receive additional information on model instances.

Not all models meet the FAIR data principles. Some models are only available from the supplementary material of scientific publications, others only upon request. Not all models are available in the SBML file format. Instead, publications may also include models in other formats such as Excel spreadsheets, Matlab files, or tabular formats. Parsing these tables into the SBML format is time-consuming, error-prone, and does not ensure a simulatable model. However, even the availability of a model in SBML format does not guarantee a valid and simulatable model.

We will now elaborate on accessing and assessing different models. To demonstrate this process, we collected all available GEMs of the human pathogen *Staphylococcus aureus*. As mentioned before, *S. aureus* colonizes the nose of approximately one-third of the human population, leading to high morbidity and mortality worldwide. Due to the high interest in this human pathogen, many models are available. In total, we collected 114 GEMs of *S. aureus* from seven different resources. The models were downloaded from the BiGG Models database, BioModels database, VMH database, and relevant publications' supplements. The first *S. aureus* model was published in 2005 by Becker *et al.* [62], the latest one in 2019 by Seif *et al.* [63]. Lee *et al.* [64], Bosi *et al.* [65], and the Path2Models project [58] published a collection of different *S. aureus* strains. Figure 3.1 gives an overview of all available GEMs and their features and properties.

After downloading, all SBML models were validated using the built-in validation function in the COBRAPy package [69]. Models that were not available in a valid SBML format underwent debugging and improvement steps. These steps were, for example, conducted for the model collection by Bosi *et al.*, which contained some syntactical errors. The debugging steps included a) adding the missing XML declaration, b) fixing invalid model instance identifiers, c) filling empty compartment lists, d) fixing invalid charge values and characters in chemical formulas, and e) adding missing metabolite compartments. The reader is referred to Figure 3 of the underlying publication [70] in Appendix B.4 on page 104 to see the detailed debugging description of the Bosi models. The models from Lee *et al.* were only available as Excel spreadsheets and converted into the SBML format. Having all models available as valid SBML files facilitates the interoperability of these models.

The valid SBML models were evaluated using the genome-scale metabolic model test suite memote. The standardized testing suite for GEMs provides a library of tests that cover the formal representation of a model, syntactic conventions and annotations, and conceptual and biological integrity for quality control and quality assurance. The score ranges from 0% to 100% and indicates the compliance with community standards and functionality with increasing score [68]. A large part of the score is based on annotation tests, as a lack of standardized annotations hinders the reuse, comparison, and extension of GEMs. Thus, models with a high memote score comply with community standards and include annotations according to the minimum information required in annotation of models (MIRIAM) principle, enabling the reusability of the models.

Another aspect of reusability is the immediate simulatability of models. We simulated the valid SBML models in the default medium they were published to test this aspect. The initial ability to grow depends on a provided objective function that can be optimized, correctly defined constraints, and a model that can produce all constituents of the objective function. Despite the manual effort of translating the Excel spreadsheet into valid SBML files, the models from Lee *et al.* did not show any growth. The same accounts for the models from the Path2Models project. These models do not possess any objective function and instead may rather act as a collection of potential reactions, metabolites, and genes than a functioning model.

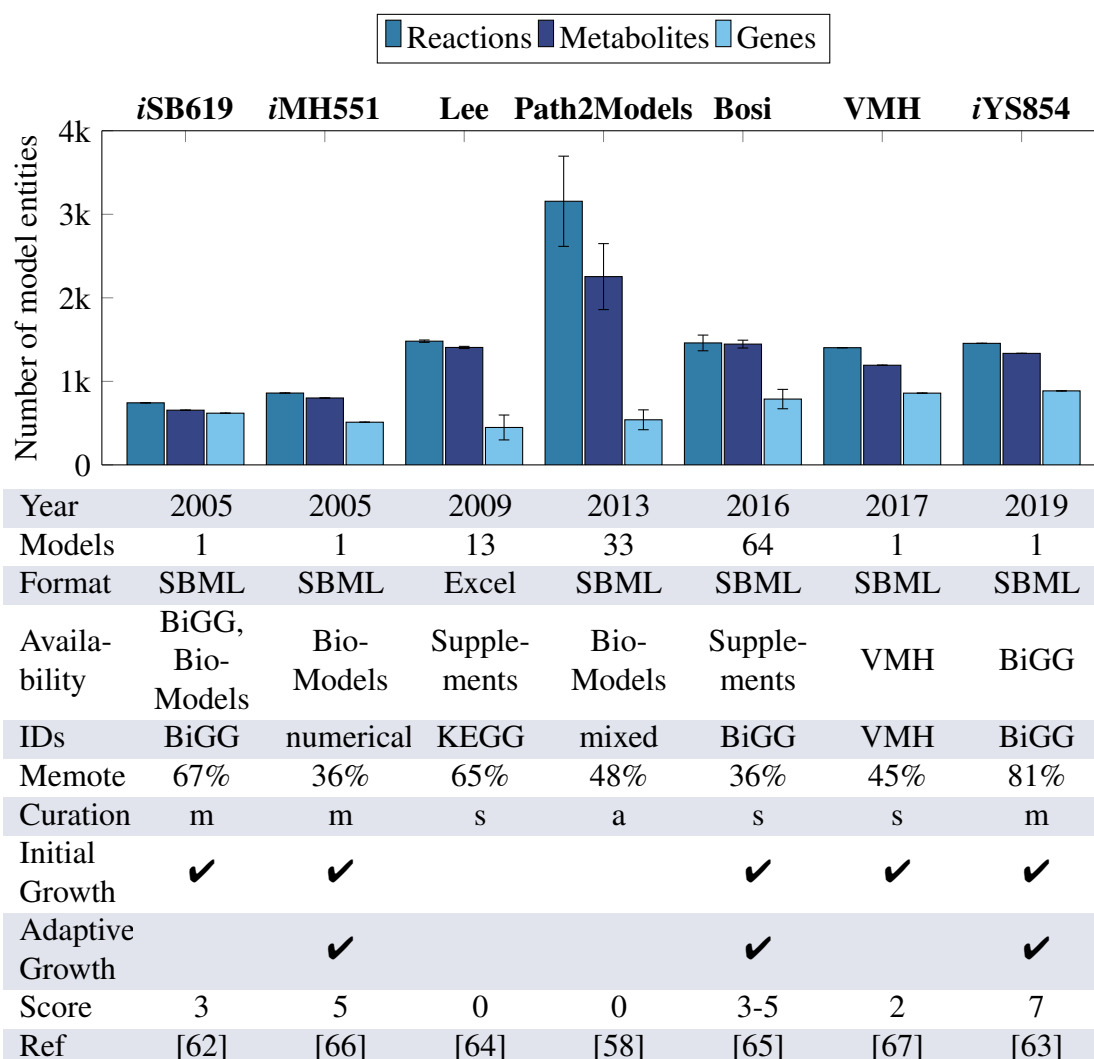


Figure 3.1: Available GEMs of *S. aureus*. All 114 available GEMs of *S. aureus* were downloaded. A collection of different *S. aureus* strains was published by Lee *et al.* [64], Bosi *et al.* [65], and the Path2Models project [58]. The average number of model instances is given together with the standard deviation. The score of the genome-scale metabolic model test suite memote ranges from 0% to 100% and indicates with increasing score the compliance with community standards and functionality [68]. The curation method indicates whether a model was curated manually (m), semi-automatically (s), or automatically (a). The row of initial growth indicates whether the respective models were simulatable directly after downloading and, thus, are directly reusable. The adaptive growth gives information about the *in silico* flexibility of the organism to react to different environmental conditions. We tested the adaptive growth under six different environmental conditions where *S. aureus* grows *in vitro* or *in vivo*, and *in silico* growth should be enabled. Checkmarks (✓) indicate whether initial or adaptive growth was possible. Based on the information from the curation, simulatability, and inclusion of additional laboratory data, a scoring system was developed and applied to all models. The higher the final score, the higher the quality and predictive value of the model.

In addition to the syntactic and initial growth check, we evaluated the predictive value of the models. To this end, we analyzed the growth of the simulatable models in six different environments *S. aureus* is known to grow. Amongst these media was the synthetic nasal medium (SNM), as the primary ecological niche of *S. aureus* is the human nose. Additionally, a chemically defined medium (CDM) was tested with either no carbon source, or glucose or galactose as a carbon source. Furthermore, *S. aureus* occurs in the gut and can colonize the lungs of patients with cystic fibrosis (CF). For these environments, chemical definitions exist that enable *in silico* simulations. The model by Heinemann *et al.* [66] had the most versatile adaptability to varying environments. Some of the Bosi models and the model by Seif *et al.* simulated growth in at least four of the six tested media.

A model's predictive value also increases when laboratory data is included or *in silico* observations are tested in laboratory experiments and vice versa. The model publications were reviewed for information about the inclusion of or verification with experimental data. Amongst others, the growth behavior in different environments, including aerobic and anaerobic conditions, can give valuable insights into the predictive value of a model, as already tested. Additionally, the comparison of predicted gene essentialities, the verification of observed auxotrophies, and the prediction of carbon utilization and catabolism facilitate the assessment of GEMs. However, one always needs to be aware of the strain-specific properties of the organism of interest. Comparing the *in silico* results of one strain with experimental data of another strain has to be conducted only if no strain-specific experimental data is available.

If several strain-specific models of an organism are available, the models can be compared. Identifiers of different databases, however, complicate such a comparison. For *S. aureus*, models with BiGG, Kyoto Encyclopedia of Genes and Genomes (KEGG) [71], VMH, and numerical identifiers were available. Another widely used resource for identifiers is ModelSEED [72]. If the models are well annotated and contain cross-references to various databases, these annotations can facilitate the comparison. However, the cross-references, and following comparisons, are only feasible when the instances are present and up-to-date in both databases. A tool for the complete mapping of identifiers is currently not available. The ModelPolisher, for example, accesses the BiGG Models database and extracts annotations and autocompletion of SBML models [73]. However, the ModelPolisher mainly relies on BiGG identifiers. The broad range of possible reaction and metabolite identifiers complicates the comparability of the models. Comparing the models' gene sets is another alternative. By comparing different strain-specific models, one can identify similarities and differences, search for metabolic modules of interest, or expand and merge the models to obtain the model that fits a particular research question.

After collecting and evaluating all these different model properties and features, we scored the models based on their predictive value. We took model simulatability, growth capabilities in different environments, and experimental verification procedures, including auxotrophies or compliance with physiological data, into account. For this example,

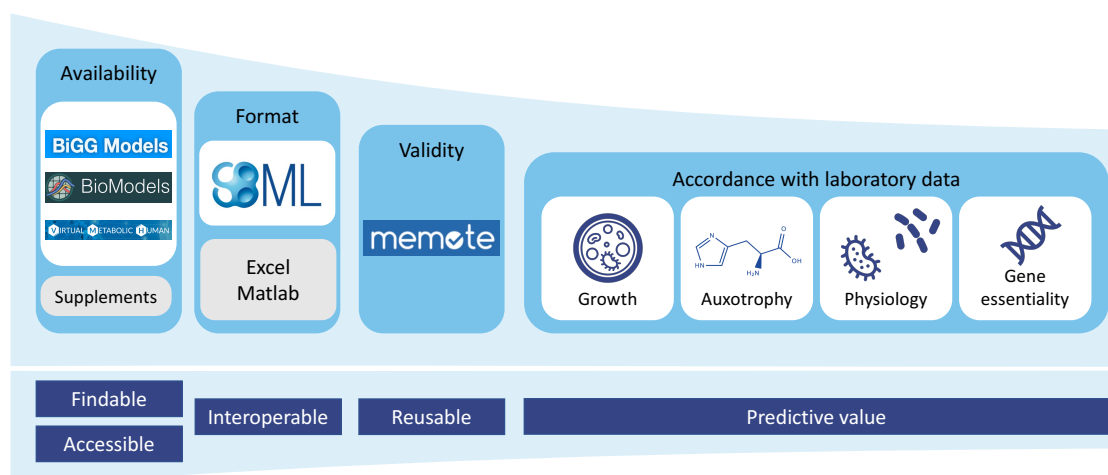


Figure 3.2: Assessment of a GEM's quality and predictive value. Valid models should be stored in a modeling database in SBML format to ensure that models are findable, accessible, interoperable, and reusable (FAIR). Storing models in the supplements of scientific publications or upon request is not recommended. Likewise, storing models as Excel spreadsheets or Matlab files hinders interoperability. GEMs can be validated using the genome-scale metabolic model test suite memote. This test suite also returns a quality score for each model. Models with a high degree of annotations receive a higher memote score than poorly annotated models. A model's predictive value can be determined by comparing model predictions with laboratory data and experiments. These experiments can include growth observations in different environments, identified auxotrophies, different physiological states, or gene essentialities. The larger the overlap between simulation and laboratory observations, the higher the model's predictive value.

the gene essentiality predictions were not included in the score, as experimentally verified essential genes were only available for a few strains. For a detailed description of the scoring of the predictive value, the reader is again referred to the underlying publication starting on page 104 [70].

In conclusion, when assessing the quality and predictive value of available models, one needs to consider several aspects, as summarized in Figure 3.2. Models are easily findable and accessible when they are deposited in modeling databases. These modeling databases are, for instance, the BiGG Models database, BioModels database, including the Path2Models project, and the VMH database. Making models available in the supplements of scientific publications is not recommended, as they are more challenging to find, access, and revise. Models stored in the standardized SBML format are interoperable with various modeling tools and software. The use of Excel spreadsheets or Matlab files often necessitates a translation or transfer to have interoperable models. To make models reusable, they should be valid files with a high degree of annotations. The genome-scale metabolic model test suite memote can validate SBML files and assess the degree of annotations. The higher the memote score, the higher the quality of the model. These aspects cover the FAIR data principles. A model's predictive can be determined by comparing the model simulations to laboratory data and experiments. If an organism grows under certain conditions or in different environments, the model sim-

3.1 Assessing the quality and predictive value of GEMs

ulation should reflect this behavior. However, comparing the growth behavior *in silico* and *in vitro* requires a detailed chemical definition of the growth medium. Auxotrophies or physiological properties observed in laboratory experiments should also be present in the model simulations. If essential genes of an organism are known, they can further be used to assess the predictive value of a GEM. By incorporating genomics, transcriptomics, proteomics, metabolomics, or fluxomics data, the predictive value of a model can increase. Using high-quality models with a high predictive value enables model-driven hypotheses to close knowledge gaps and drive model-driven discoveries. If, however, no such model is available, one needs to curate a high-quality GEM. The following section will focus on the steps necessary to curate high-quality GEMs.

3.2 Curating high-quality GEMs

In 2010, Thiele and Palsson published a protocol for generating high-quality genome-scale metabolic reconstructions [74]. This protocol comprises 96 steps in the following five stages: 1. reconstruction of the draft, 2. refinement of the reconstruction, 3. conversion of the reconstruction into a computable format, 4. evaluation of the network, and 5. prospective use [74]. Today, multiple tools are available to curate GEMs automatically. Amongst these tools are CarveMe [75], AutoKEGGRec [76], AuReMe [77], RAVEN [78], ModelSEED [72], and gapseq [79]. All available tools have their strengths, advantages, but also weaknesses. Mendoza *et al.* systematically assessed available reconstruction tools in 2019 [80]. They concluded that CarveMe and ModelSEED provide ready-to-use GEMs. Furthermore, CarveMe generates networks with high reaction sets, similar to manually curated networks.

Draft reconstruction

We decided to build our workflow upon the reconstruction tool CarveMe to curate the first draft of a strain-specific GEM. CarveMe starts with a universal model from the BiGG Models database [56] and an annotated genome sequence. Applying a top-down approach, CarveMe uses the universal model, specialized templates, and gene annotations to score the certainty that a reaction is present in the organism of interest. The universal model is converted into an organism-specific model by removing all reactions and associated metabolites with low certainty in the model carving step. However, Mendoza *et al.* highlighted in their review on reconstruction tools that users need to be aware of possible false results. For that reason, manual refinement of the initial draft reconstruction is required.

As mentioned in the previous section, memote evaluates the quality of GEMs, not only concerning the content but also the standards [68]. Thus, memote is a valuable tool to track the development of the GEM's curation and refinement process. Another helpful tool is the ModelPolisher [73]. The ModelPolisher accesses the BiGG Models database to autocomplete and annotate models. As CarveMe curates models based on the BiGG Models database, it is possible to use the ModelPolisher to annotate and autocomplete the initial draft reconstructions. Moreover, the ModelPolisher can fix apparent errors in the models and add annotations by using mappings from AnnotateDB [81].

Annotations

Annotations are particularly helpful when comparing various models. Therefore, a high degree of annotations for all model instances is essential for a high-quality GEM. The ModelPolisher adds numerous annotations to the model's metabolites and reactions. The genes, however, are strain-specific, and their annotations need to be included additionally. To curate GEMs with CarveMe, we can use the annotated gene or protein se-

quence files from the National Center for Biotechnology Information (NCBI) database [82]. CarveMe then includes the NCBI gene identifier as GPRs into the model. One possibility to further annotate the genes is to map the NCBI gene or protein identifiers to their respective locus tags. These locus tags are also used in KEGG as gene identifiers of specific species and strains. KEGG has additional cross-references to other databases, including UniProt [83] and NCBI protein identifiers [82].

In addition to cross-references to other databases, the model's instances can be annotated using Systems Biology Ontology (SBO) terms [84]. The SBO comprises a set of controlled, relational vocabularies used in systems biology and computational modeling. SBO terms provide semantic information to a model's standard description to facilitate its efficient reuse. We developed a pipeline that automatically assigns SBO terms to entities in SBML models. Although memote reports missing terms and suggests appropriate ones, we aimed to increase the information content by assigning as specific SBO terms as possible, especially for the reactions. The SBO terms for metabolites and genes are SBO:0000247 for simple chemicals and SBO:0000243 for genes. These terms are already the most specific descriptions of the model instances. For reactions, however, a variety of SBO terms exists. The most general term is SBO:0000176 for biochemical reactions. Our pipeline uses the Enzyme Commission (EC) numbers stored in the annotations of reactions to identify more specific SBO terms. These specific terms range from acetylations, phosphorylations, methylations, and glycosylations, to hydrolysis and various transport reactions. Using SBO terms that describe the biochemical reaction in more detail, we facilitate the analysis and reuse of the curated models and, thus, comply with the FAIR data principle.

Evidence and Conclusion Ontology (ECO) terms can describe the types of scientific evidence for a model's instances. These terms give information about the type of evidence within the biological research domain, such as laboratory experiments, computational methods, or literature curation [85]. The curator can store information about the confidence and evidence of reactions using ECO terms. We developed a workflow to automatically annotate reactions with ECO terms based on their confidence score. Reactions are first checked for their GPRs. If no GPR is available, the reaction is assigned the ECO term with the lowest confidence. Subsequently, each gene is looked up in the UniProt database, and information about the corresponding protein's existence is extracted. ECO terms are then assigned following the reported protein existence level. As for the SBO terms, the addition of ECO terms facilitates the reusability of the model and, thus, complies with the FAIR data principle.

Besides adding SBO and ECO terms, we can further add the pathways to the reactions' annotations, in which the reactions occur. To do so, we can extract the KEGG identifiers from the annotations and looked up associated pathways. These pathways are added to the annotations using the biological qualifier BQB_OCCURS_IN.

Corrections and improvements

As automatically curated reconstructions can contain obscurities, inaccuracies, or errors, several correction and improvement steps are required. The metabolites' chemical formulas and charges are retrieved from the universal BiGG model. However, several chemical formulas and charges are reported for multiple metabolites in the BiGG Models database. The presence and accuracy of the chemical formula and the charge needs to be verified for each metabolite. In cases of uncertainty, other databases, such as ModelSEED [72] or BioCyc [86], can be browsed to guide the decision for choosing the correct chemical formula and charge.

When every metabolite has a chemical formula and charge, reactions can be evaluated for their mass and charge balance. Imbalanced reactions are indicators for inconsistencies or errors in the model. These imbalances can be explained by (i) incorrectly chosen chemical formulas or charges, (ii) missing metabolites in the reaction equation, such as protons, or (iii) a knowledge gap and, consequently, missing information. For the first two reasons, literature research and look-ups in already mentioned databases may dissolve the imbalances. A knowledge gap can only be closed with new insights and additional information. This information can, e.g., be obtained by laboratory experiments. When looking at imbalanced reactions, we must consider that pseudoreactions might not be balanceable. Examples of imbalanced pseudoreactions are exchange, sink, demand, or biomass reactions.

Besides imbalanced reactions, a model can contain orphan and dead-end metabolites. These metabolites are characterized by their low connectivity within the metabolic network. Orphan metabolites are only consumed but not produced by reactions, whereas dead-end metabolites are only produced but not consumed. We can identify additional connecting reactions by browsing different databases, including KEGG [71] and BioCyc [86]. These reactions can then be added to the model to increase the connectivity of the orphan and dead-end metabolites.

GEMs can contain thermodynamically infeasible energy-generating cycles (EGC). Such cycles can produce energy, e.g., in the form of ATP, without consuming any nutrients. Fritzemeier *et al.* summarized 14 energy metabolites that need to be checked for EGCs [87]. A dissipation reaction needs to be added to the model for each energy metabolite. Each dissipation reaction is sequentially maximized after constraining all uptake reactions to zero. Objective values unequal to zero indicate the presence of EGCs, which subsequently need to be eliminated.

CarveMe adds a universal biomass objective function (BOF) to the model. We can improve and specify the BOF of our target organism to increase the predictive value of our model. To do so, we need additional information and data, such as genomics, transcriptomics, proteomics, metabolomics, or lipidomics. The Python package BOFdat can generate and improve a BOF based on organism-specific experimental data [88]. With the provided experimental data, this tool can calculate the stoichiometric coefficients for major macromolecules, inorganic ions and coenzymes, and remaining species-specific

metabolic biomass precursors.

Model extension

As reactions are added to the model to reduce the number of orphan or dead-end metabolites, the model can be further extended to increase its scope. The model can be expanded using the information from additional databases. For example, if the organism and strain of interest have an entry in the KEGG database, we can search for metabolic genes that are so far not included in the model. These genes are linked to reactions and can be incorporated into the model. Another helpful resource is the BioCyc database, where strain-specific information is stored. Depending on the organism, further organism-specific databases might exist. It should be avoided to add orphan or dead-end metabolites or reactions that are not connected to the rest of the network during the model extension step. While adding new reactions, metabolites, and genes, we need to consider the previously mentioned steps and issues concerning annotations, corrections, and improvements. Only reactions associated with strain-specific genes should be added. Including relevant reactions, metabolites, and genes increases the model's scope and, thus, its predictive value.

Model validation

We can compare the *in silico* results to laboratory experiments to evaluate the model's predictive value. As already highlighted in the previous chapter in Figure 3.2, we can evaluate the model's predictive value by testing the growth behavior of our organism in defined environments and under certain growth conditions or confirming described auxotrophies. Physiological features or gene essentialities can additionally be used to evaluate and validate the model. These laboratory results might be obtained from already published literature. If possible, own experiments or collaborations with laboratories can be conducted to verify model predictions. The more accordance we observe between laboratory and simulation results, the higher our curated model's predictive value. Thus, comparing as many physiological features, growth behaviors, auxotrophies, and gene essentialities as possible, is highly recommended. While comparing *in silico* with *in vitro* results, we must consider the strains. If laboratory results are only available for a different strain, we cannot guarantee the accordance between the *in silico* and *in vitro* results. Additionally, we need to consider and simulate the setting in which the laboratory data was collected. For example, the essentiality of genes is dependent on the growth conditions. Thus, simulating the laboratory growth conditions under which the essential genes were determined is essential to make the results comparable.

As we are particularly interested in microbes colonizing the human nose, our model validation should also include evaluating the *in silico* growth behavior in the nasal environment. The synthetic nasal medium (SNM) was developed based on metabolomics data of human nasal secretions to mimic the environmental conditions in the human

nose [89]. As SNM is chemically defined, we can simulate a microbe's growth in this medium. If the organism of interest does not grow in SNM but is reported to occur in the human nose, we have to identify the reason for the missing growth. One reason could be the absence of additionally required reactions in the model. These reactions need to be identified by literature research and model analyses. Subsequently, they need to be added to the model. Another reason could be the isolated examination of the microbe. In the human nose, a magnitude of commensals and pathogens exists and interacts with each other. During these interactions, essential metabolites can be exchanged that are not part of the SNM. We need to identify such metabolites required for growth but not defined in SNM. In following co-culture or community model simulations and in *in vitro* experiments, we can evaluate the role of the identified metabolites.

Model publication

In a final step, the model should be made publicly available to the scientific community to meet the FAIR data principle. For this reason, the model should be uploaded to a model database to make it findable and accessible. Suitable databases are, for instance, the BiGG Models database or the BioModels database. To ensure the interoperability of the model, we need to provide a valid SBML file. As the model includes numerous annotations after the manual refinement steps, the reusability is assured. Thus, a novel high-quality GEM is available to the scientific community. The whole protocol on building high-quality GEMs of (nasal) bacteria can be found in Appendix A on page 53.

Example 1: Novel GEM of *Dolosigranulum pigrum*

Dolosigranulum pigrum is a relatively newly identified opportunistic pathogen [90]. It is a Gram-positive coccus growing in pairs, tetrads, and clusters [91]. *D. pigrum* is sporadically associated with diseases [90, 91]. However, it is also reported to have a potential probiotic effect as its presence is associated with a healthy upper respiratory tract (URT) and a resistance to recurrent ear infections [92]. The URT also includes the human nose, where *D. pigrum* plays a pivotal role in the nasal microbiome [40]: Its presence is negatively associated with *Staphylococcus aureus* and *Streptococcus pneumoniae*, both listed in the WHO's priority pathogens list (see Table 1.1). Brugger *et al.* even identified an inhibiting effect on *S. aureus*, strengthening *D. pigrum*'s role as a potential probiotic. However, the overall mechanism behind this inhibiting effect remains unclear but could be evaluated using GEMs of the organisms of interest. As reported in the previous section, 114 GEMs of *S. aureus* are currently available, but until 2021 none of *D. pigrum*. Due to its increasing importance in microbial communities and interactions with pathogens, we curated the first GEM of *D. pigrum* strain 83VPs-KB5.

We followed the steps described in this section to curate a high-quality GEM. As no model of *D. pigrum* existed in any modeling database, we generated a first draft reconstruction using CarveMe. We added extensive annotations for metabolites, reactions,

and genes, including cross-references to various databases and SBO terms. Reactions were further annotated with ECO terms and associated KEGG pathways. In manual curation steps, we checked the chemical formulas and charges of the metabolites, corrected mass and charge imbalanced reactions, and reduced the amount of orphan and dead-end metabolites. We extended the model using the KEGG database by searching for metabolic genes missing in the model and their associated reactions. The model was then validated using all to date publicly available experimental data. These experimental results included several auxotrophies and the biosynthesis of L-glutamine [40]. We confirmed these auxotrophies and the biosynthesis in our *in silico* simulations. Additionally, the carbohydrate metabolism was investigated. Brugger *et al.* found that *D. pigrum* does not have a tricarboxylic acid (TCA) cycle, which we could confirm with our model. We evaluated the growth behavior in four different environments in the last verification step, including the chemically defined SNM mimicking the human nose [93]. In single culture, three additional metabolites were required for growth in SNM: L-isoleucine and L-methionine, for which auxotrophies were already reported, and meso-2,6-diaminoheptanedioate. For the latter, no biosynthetic pathways could be identified in any database, suggesting either a knowledge gap or an unknown auxotrophy [94].

As *D. pigrum* is difficult to cultivate in laboratories, we defined a minimal medium. This minimal medium contains 13 amino acids that cannot be synthesized *de novo* by *D. pigrum*, 13 trace minerals, D-glucose as carbon source, tree vitamins, two additional compounds, and oxygen to enable *in silico* aerobic growth. The detailed minimal medium composition can be retrieved from Table 2 of the underlying publication [94] in Appendix B.2 on page 65. This example of the minimal medium definition shows that model-driven hypotheses could help laboratory experiments.

To meet the FAIR data principle described in the previous section, our model is available in the BioModels database as a valid SBML Level 3 version 1 [95] file under the accession number MODEL2012220003. With 86%, the model holds a high memote score, confirming its reusability and a high degree of annotations.

Example 2: Updated GEM of *Pseudomonas aeruginosa*

The described protocol to curate high-quality GEMs can not only be applied to generate new models. It can also be used to update and refine already existing models. We conducted several steps on the *Pseudomonas aeruginosa* PA14 model from Bartell *et al.* [96]. The initial GEM was published in 2007 as SBML Level 2 Version 1 [97] file on the lab's website. It has a memote score of 26%.

We first upgraded the SBML version to the latest SBML edition (Level 3) [61] and enabled the fbc-plugin [98] and the groups-plugin [99]. The low total memote score of the initial model indicates a low degree of annotations. As mentioned previously, ModelPolarizer mainly relies on identifiers from the BiGG Models database. The *P. aeruginosa* PA14 model, however, contains identifiers from the ModelSEED database. We manually and semi-automatically retrieved the respective BiGG Models identifier for all metabo-

lites and reactions, where available. These identifiers were added as annotations. With the help of the ModelPolisher, we added cross-references to several other databases, such as KEGG, MetaNetX [100], or MetaCyc [101]. Using the KEGG annotations, we added associated KEGG pathways to the reactions and annotations to the genes. Finally, we also added SBO terms to the model instances.

After improving the annotations, we corrected mass and charge imbalances. In total, 137 changes were performed to eliminate as many mass and charge imbalances as possible. These changes mainly included adding or removing protons from reactions or changing the chemical formulas or charges of metabolites based on literature or database evidence. During the test for EGC, we saw that the model was able to generate ATP. This EGC was resolved by adding the periplasmic compartment and correcting the reversibility of four reactions in the electron transport chain. The BOF was evaluated and updated to include lipopolysaccharides present in Gram-negative bacteria [102]. It was also re-organized into macromolecular categories to better represent the components required for growth.

The KEGG and MetaCyc database were used to incorporate additional reactions, metabolites, and genes to increase the model's predictive value. Reactions were only added if sufficient evidence was provided in the literature. The updated and corrected model was evaluated using gene essentiality, carbon utilization, and growth predictions. Additionally, transcriptomics data was integrated into the updated *P. aeruginosa* PA14 model.

With these changes and updates, the memote score was increased to 88%. The final updated model is available as SBML Level 3 Version 1 [95] file in the BioModels database under the accession number MODEL2106110001 [103].

3.3 Identifying potential antiviral targets against SARS-CoV-2

After we elaborated on the characteristics of high-quality GEMs and how to build them, we want to focus on working with these models. Having high-quality models with a good predictive value enables us to create model-driven hypotheses. The advantage of such model-driven hypotheses is their targeted and tailored validation in laboratory experiments. Thus, these targeted laboratory experiments may reveal more promising results than conventional experiments. Additionally, positive results may be retrieved faster. Fast and positive results are crucial for questions of high urgency and high importance, like in a worldwide pandemic.

In December 2019, a pneumonia outbreak in Wuhan, China, became the starting point of the worldwide COVID-19 pandemic. COVID-19 is caused by the severe acute respiratory syndrome coronavirus type 2 (SARS-CoV-2). In November 2021, almost two years after the first reported COVID-19 cases, more than 250 million cases and over 5 million deaths were registered [104]. During the first year of the pandemic, the development of vaccines gained a lot of attention and governmental sponsorship. Several vaccines against SARS-CoV-2 are currently available. However, the number of approved therapeutic strategies is still minimal.

We wanted to fill the gap of the urgently needed therapeutic strategies. In 2018, Aller *et al.* published a methodology to curate integrated human-virus metabolic models [105]. As viruses rely on the host cell's metabolism, a relevant human cell type model was required for our analyses to be infected with SARS-CoV-2 *in silico*. Aller *et al.* used a model of human alveolar macrophages as proof of concept for their methodology. This human alveolar macrophage model was published in 2010 by Bordbar *et al.* to give insight into *Mycobacterium tuberculosis* infections [106]. As the model was already used for several studies and human alveolar macrophages are relevant for SARS-CoV-2 infections [107, 11, 10], we chose this cell type as the host cell. For the integration of the virus, we first used the published annotated genome sequence of SARS-CoV-2 to calculate the stoichiometric coefficients of the viral biomass objective function (VBOF). In subsequent analyses, we additionally used sequencing information about the most common mutational variants of the novel coronavirus. In particular, we evaluated the following variants: (i) Alpha, lineage B.1.1.7, (ii) Beta, lineage B.1.351, (iii) Gamma, lineage P.1, (iv) Delta, lineage B.1.617, and (v) Epsilon, lineage B.1.427/429. At the time of evaluation, all variants were either variants of concern (VOC) or variants of interest (VOI) [108, 109]. By November 2021, the Alpha and Epsilon variants have become de-escalated variants, while the Beta, Gamma, and Delta variants are still VOC [110].

We developed a pipeline to automatically calculate the VBOF of the novel coronavirus and its variants. Besides the already mentioned annotated genome sequences, we incorporated information about the virus replication method, the number of genomic

copies per virion, and information about the copy number of the four structural proteins spike (S), envelope (E), membrane (M), and nucleocapsid (N). The nucleotide investment, the amino acid investment, the ATP requirements, and the phosphate liberation were calculated with this information. The stoichiometric coefficients were estimated with the total viral molar mass, and the final VBOF was constructed [111]. In subsequent analyses, we further elaborated on the effect of incorporating lipids into the VBOF. As Aller *et al.* did not consider lipids in their VBOF, we extended their methodology to include relevant fatty acids occurring in the capsid of SARS-CoV-2 into the VBOF [112].

Having the automated pipeline to construct VBOFs at hand, we evaluated the host's metabolic changes upon infections with the SARS-CoV-2 reference genome and its mentioned mutational variants. We wanted to identify potential antiviral targets by knocking out every reaction and evaluating the effect on the host's maintenance and the viral replication capacity. Promising antiviral targets were defined as reactions whose knock-out significantly decreased or completely inhibited viral replication while only minimally affecting the host's maintenance. With this approach, we identified guanylate kinase1 (GK1) as an antiviral target against SARS-CoV-2. GK1 was even robust against the occurring mutational variants.

With this pipeline and high-quality GEMs, the identification of potential antiviral targets can be extended. On the one hand, infections with other viruses can be simulated, as already shown for the mutational variants of SARS-CoV-2. The WHO published a high-priority pathogens list for bacterial infections (see Table 1.1) and for viral infectious diseases. The latter list includes, for example, the nairovirus, the Marburg virus, the Ebola virus, and the family of human coronaviruses [113]. This list could guide further antiviral target identification studies. On the other hand, we can also vary the host cell using our pipeline to study the effect of specific virus infections in various host cell types. We can identify more robust antiviral targets using different host cell types for the integrated host-virus interaction models. One possibility is to use human single-cell sequencing data and available tools to curate cell-specific GEMs. These GEMs can then be used as input for our pipeline and subsequent identification of robust antiviral targets. By combining these two approaches of varying human host cells and the infecting virus, broadly-acting antiviral targets may eventually be identified.

With this example of identifying potential antiviral targets against SARS-CoV-2, we showed that well-curated, high-quality models could be helpful for model-driven discoveries and hypotheses. Potential targets are quickly identified *in silico* and could be validated in targeted laboratory experiments. Especially during a world-threatening pandemic, fast and efficient ways to identify therapeutic approaches are imperative. Besides the high predictive value of the model, the availability and reusability of the models are of high importance to enable the replicability of the results and the extension to further viruses or cell types. With these tools and high-quality models at hand, we might be better prepared for future pandemics and a faster and more efficient development of therapeutic approaches.

3.3 Identifying potential antiviral targets against SARS-CoV-2

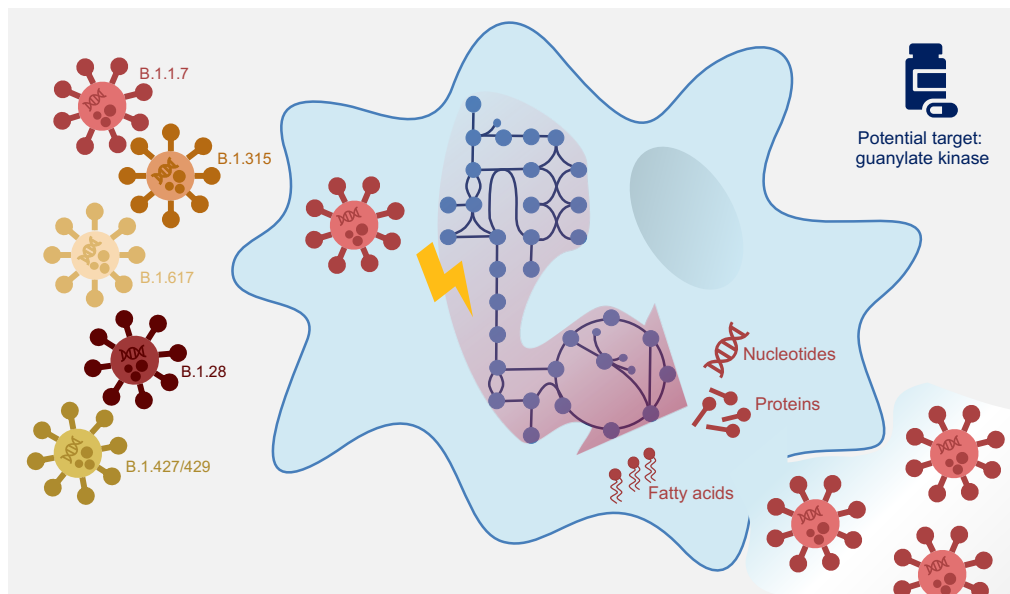


Figure 3.3: Procedure to identify potential antiviral targets against SARS-CoV-2. Integrated host-virus interaction models can be used to examine metabolic shifts in the host's metabolism upon viral infection. We chose human alveolar macrophages as host cells and simulated infections with the mutational variants Alpha (B.1.1.7), Beta (B.1.351), Gamma (P.1 or B.1.28), Delta (B.1.617), and Epsilon (B.1.427/429). *In silico* knock-out experiments were conducted to investigate metabolic necessities for the host cell's maintenance and the viral replication capacity. Knock-outs that significantly inhibited the viral replication capacity while harming the host cell at a minimum level were reported as potential antiviral targets.

The important thing is not to stop questioning. Curiosity has its own reason for existence. One cannot help but be in awe when he contemplates the mysteries of eternity, of life, of the marvelous structure of reality. It is enough if one tries merely to comprehend a little of this mystery each day.

— Albert Einstein

4

Discussion and Perspectives

In this thesis, I demonstrated (I) how the quality and predictive value of existing GEMs can be assessed, (II) how models of high quality can be curated, and (III) how these models can be used for model-driven discoveries.

When we assess a GEM's quality and predictive value, the result might be variable, depending on the research question. If someone is, for example, interested in evaluating a specific metabolic pathway, then one might focus on a model with more detailed information on that specific pathway. If someone, however, is interested in predicting knock-out experiments, then one would choose a model with the highest accordance with gene essentiality data. These two examples demonstrate that there might be no unique answer for the 'best' metabolic model. The answer depends on the research question and the focus of interest. Additionally, not all information or experimental data is available for every organism or strain. Thus, comparing model predictions with *in vitro* experiments is not always possible. This lack might not necessarily imply a model of poorer quality. It rather illustrates the need for further reasonable experiments to fill knowledge gaps. However, one should keep in mind to compare model predictions to experimental data as soon as they become available. Nonetheless, all models should meet the FAIR data principle to be findable, accessible, interoperable, and reusable. These aspects should apply to all GEMs to ensure good scientific practice when working with models. As mentioned before, storing valid models in modeling databases makes them findable and accessible and, thus, reusable. The use of standard formats, such as the SBML file format, is highly recommended to ensure the interoperability of the GEM. To conclude, high-quality models need to meet the FAIR data principle. The decision for a good model, however, might depend on the user's research question.

If no model of an organism is available, or the model is of insufficient quality for a research question, we need to curate a model. In this thesis, we focused on curating a model based on a CarveMe draft. Several other software and tools are currently available for building and curating GEMs – and even more will probably become public.

As mentioned in Chapter 3.2, all these tools have their strengths, advantages, but also weaknesses. Depending on the used tool or software, the pipeline proposed in this thesis might be adapted. The same accounts for newer versions of CarveMe. The standards for the SBML file format, the FAIR data principle, and the MIRIAM principle need to be monitored. Novel or updated standards need to be included in the pipeline to guarantee the high quality of GEMs. Furthermore, we need to keep in mind that models can constantly be improved. Every day, new data is collected that can be incorporated into our models. For example, the first GEM of *Escherichia coli* was published in 2000 [52]. As *E. coli* is one of the best-studied microbes, a wealth of data is available for that organism. Thus, several updates, extensions, and expansions of *E. coli* GEMs are available: 20 *E. coli* GEMs are currently uploaded in the BiGG Models database [56] and 21 in the BioModels database [57]. Besides the adaptation of the pipeline for other tools and novel standards, the pipeline might further be adjusted for personal use and research focus. We are highly interested in the human nasal microbiome and, thus, added the growth evaluation in the synthetic nasal medium to the pipeline. Researchers interested in other environments, phenotypes, or other physiological and metabolic features might want to add additional steps or modify existing ones in the pipeline. However, the model annotation, validation, and publication should always be part of the pipeline to ensure a high-quality model and to meet the FAIR data principle.

We can formulate model-driven hypotheses and make model-driven discoveries with a high-quality GEM. However, we have to keep in mind that we use models. A model is a simplified representation of reality. As it is a simplification, the model does not include all details and information as present in reality. Thus, the simplification can result in the loss of potentially essential information and inaccurate predictions. For that reason, it is crucial to test and validate the model's predictive value with already existing laboratory data. By using GEMs with a high predictive value, we can formulate more precise model-driven hypotheses. These model-driven hypotheses can form an educated guess for targeted experiments or tailor the search area. Furthermore, with GEMs, we look at the systemic level and not the molecular level of individual metabolites or reactions. This systemic view can identify off-target effects that might be missed when focusing on individual reactions or pathways. In the example in Chapter 3.3, we identified potential antiviral targets in human alveolar macrophages against SARS-CoV-2. Aller *et al.* validated this technique for targets against Chikungunya, Dengue, and Zika viruses: Their predictions recovered known targets for existing antiviral drugs [105]. As the technique was already validated for other viruses, we are confident about our identified antiviral targets against SARS-CoV-2. Model-driven hypotheses and discoveries, however, have a much larger field of application than identifying potential antiviral targets. We can predict, for example, growth requirements in a specific environment to facilitate or enable a microbe's cultivation *in vitro*. Other applications include developing or engineering strains to produce specific metabolites, drug targeting in pathogens in general, understanding the metabolism of humans, animals, and pathogens, modeling of interactions between different cells or organisms, and predicting enzyme functions [47]. This wide range of

applications illustrates the usefulness of GEMs: We can rapidly generate experimentally testable hypotheses. By experimentally validating those model-driven hypotheses, we gain novel insights and data, which in turn can be incorporated into the model. This iterative process refines the GEMs and increases their predictive value, by which even better and more accurate model-driven hypotheses can be formulated.

Perspectives

The pipeline proposed in Chapter 3.2 can be used to curate other GEMs. As we are particularly interested in the human nasal microbiome, we can curate, update, or upgrade models of nasal inhabitants. Liu *et al.* have identified bacteria with high prevalence and proportional abundance in seven distinct nasal community state types. Amongst these bacteria are *Moraxella* spp., such as *Moraxella catarrhalis* [114]. *M. catarrhalis* is a human-restricted pathogen of the upper and lower respiratory tract [115]. In the elderly population, however, a nasal microbiome dominated by *Moraxella* species is associated with respiratory health [116]. Due to its high prevalence in one of the nasal community state types, its association with respiratory health in the elderly, and its pathogenic properties, a GEM can help to elucidate the beneficial or harmful interplay with other nasal commensals or pathogens.

Another highly prevalent commensal in the human nose is *Staphylococcus epidermidis* [114, 25], which contributes to a healthy maturation of the nasal microbiome [117]. *S. aureus*'s colonization is negatively associated with *S. epidermidis*' presence [118]. Additionally, *S. epidermidis* mediates antiviral protection against influenza A virus infections [119]. With a GEM of this commensal, not only the role of host-bacterial interactions in antiviral responses might be further investigated. It might also give insight into its negative correlation with *S. aureus*.

A third highly prevalent group of commensals in the human nasal microbiome include *Corynebacterium* species. A low prevalence of *Corynebacterium* is a predictor of high *S. aureus* abundance [114]. Artificial implantation of *Corynebacterium* species into human noses even resulted in eradicating *S. aureus* [120]. One theory for this eradication is that a commensal competition including virulence components may hold the potential for novel antimicrobial compounds [121]. Such competitive interactions could be investigated using high-quality GEMs.

These were just three examples of how this thesis's work could be extended and applied to additional relevant and interesting microbes in the human nose. However, the application is not restricted to nasal bacteria. Almost any bacterium's metabolism can be simulated *in silico* when basic information, like an annotated genome sequence, is available. Additionally, we can expand the scope to viral infections, as shown in Chapter 3.3. The proposed pipeline to identify potential antiviral targets can be applied to other infecting viruses, such as the Dengue, influenza, or measles virus. Furthermore, not only human alveolar macrophages can be virally infected *in silico*; The pipeline can be extended to any other cell type with an existing GEM. These features make the proposed

pipelines and applications highly flexible, adaptable, and reusable.

Conclusion

This thesis provided an overview of quality measures for evaluating different aspects of GEMs. These aspects include compliance with standards, such as the FAIR data principle, the inclusion of laboratory data, and the predictive value. The quality measures were applied to evaluate all currently available models of the human pathogen *S. aureus* responsible for severe infections with high morbidity and mortality. High-quality models with a high predictive value help us better understand an organism, identify metabolic capabilities in health and disease, identify potential targets for treatment interventions, and analyze the interplay between different cells and organisms. If, however, no high-quality model exists, a novel GEM needs to be curated. For that purpose, we developed and adapted a guide to curate high-quality GEMs, especially of nasal bacteria. We can use such models for answering relevant and urgent questions of our time quickly and efficiently. These questions are, however, not restricted to bacteria or the nasal environment: they can be expanded to various pathogens, including viruses, cells, or even tissues and organs in diverse environments. Thus, we evaluated not only the pathogen *S. aureus* and the commensal *D. pigrum* in the nasal environment but also identified potential antiviral targets for human alveolar macrophages infected with SARS-CoV-2. With the transition of the pathogen and infection site, the versatile and adaptable applicability of the methods in systems biology is demonstrated. As we showed how high-quality genome-scale metabolic models can be curated easily and efficiently, they can become an indispensable constituent in future research.

Bibliography

- [1] Ensheng Dong, Hongru Du, and Lauren Gardner. An interactive web-based dashboard to track COVID-19 in real time. *The Lancet Infectious Diseases*, 20(5):533–534, 2020. doi:10.1016/S1473-3099(20)30120-1.
- [2] Lizzie Wade. From black death to fatal flu, past pandemics show why people on the margins suffer most. *Science*, 2020.
- [3] Peter Spreeuwenberg, Madelon Kroneman, and John Paget. Reassessing the Global Mortality Burden of the 1918 Influenza Pandemic. *American Journal of Epidemiology*, 187(12):2561–2567, 2018. doi:10.1093/aje/kwy191.
- [4] Niall P. A. S. Johnson and Juergen Mueller. Updating the accounts: Global mortality of the 1918-1920 ‘spanish’ influenza pandemic. *Bulletin of the History of Medicine*, 76(1):105–115, 2002. URL: <http://www.jstor.org/stable/44446153>.
- [5] Edwin D. Kilbourne. Influenza pandemics of the 20th century. *Emerging Infectious Diseases*, 12(1):9–14, 2006. doi:10.3201/eid1201.051254.
- [6] Nucleotide [Internet]. Bethesda (MD): National Library of Medicine (US), National Center for Biotechnology Information; [1988]. Wuhan seafood market pneumonia virus isolate Wuhan-Hu-1, complete genome - Nucleotide - NCBI – Accession No. NC_045512.1. URL: https://www.ncbi.nlm.nih.gov/nucleotide/NC_045512.1.
- [7] Yinon M. Bar-On, Avi Flamholz, Rob Phillips, and Ron Milo. SARS-CoV-2 (Covid-19) by the numbers. *eLife*, 9, 2020. arXiv:2003.12886, doi:10.7554/eLife.57309.
- [8] Yu Chen, Qianyun Liu, and Deyin Guo. Emerging coronaviruses: Genome structure, replication, and pathogenesis. *Journal of Medical Virology*, 92(4):418–423, 2020. doi:10.1002/jmv.25681.
- [9] Chaolin Huang, Yeming Wang, Xingwang Li, Lili Ren, Jianping Zhao, Yi Hu, Li Zhang, Guohui Fan, Jiuyang Xu, Xiaoying Gu, Zhenshun Cheng, Ting Yu, Jiaan Xia, Yuan Wei, Wenjuan Wu, Xuelei Xie, Wen Yin, Hui Li, Min Liu, Yan Xiao, Hong Gao, Li Guo, Jungang Xie, Guangfa Wang, Rongmeng Jiang, Zhancheng Gao, Qi Jin, Jianwei Wang, and Bin Cao. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *The Lancet*, 395(10223):497–506, 2020. doi:10.1016/S0140-6736(20)30183-5.

- [10] Louise Dalskov, Michelle Møhlenberg, Jacob Thyrssted, Julia Blay-Cadanet, Ebbe Toftgaard Poulsen, Birgitte Holst Folkersen, Søren Helbo Skaarup, David Olagnier, Line Reinert, Jan Johannes Enghild, Hans Jürgen Hoffmann, Christian Kanstrup Holm, and Rune Hartmann. SARS-CoV-2 evades immune detection in alveolar macrophages. *EMBO reports*, 21(12):e51252, 2020. doi: 10.15252/embr.202051252.
- [11] Hin Chu, Jasper Fuk-Woo Chan, Yixin Wang, Terrence Tsz-Tai Yuen, Yue Chai, Yuxin Hou, Huiping Shuai, Dong Yang, Bingjie Hu, Xiner Huang, Xi Zhang, Jian-Piao Cai, Jie Zhou, Shuofeng Yuan, Kin-Hang Kok, Kelvin Kai-Wang To, Ivy Hau-Yee Chan, Anna Jinxia Zhang, Ko-Yung Sit, Wing-Kuk Au, and Kwok-Yung Yuen. Comparative Replication and Immune Activation Profiles of SARS-CoV-2 and SARS-CoV in Human Lungs: An Ex Vivo Study With Implications for the Pathogenesis of COVID-19. *Clinical Infectious Diseases*, 71(6):1400–1409, 2020. doi:10.1093/cid/ciaa410.
- [12] Muge Cevik, Krutika Kuppalli, Jason Kindrachuk, and Malik Peiris. Virology, transmission, and pathogenesis of SARS-CoV-2. *The BMJ*, 371, 2020. doi: 10.1136/bmj.m3862.
- [13] Mona Sadat Mirtaleb, Amir Hossein Mirtaleb, Hassan Nosrati, Jalal Heshmatnia, Reza Falak, and Reza Zolfaghari Emameh. Potential therapeutic agents to COVID-19: An update review on antiviral therapy, immunotherapy, and cell therapy. *Biomedicine and Pharmacotherapy*, 138:111518, 2021. doi:10.1016/j.biopha.2021.111518.
- [14] Zeinab Abdelrahman, Mengyuan Li, and Xiaosheng Wang. Comparative Review of SARS-CoV-2, SARS-CoV, MERS-CoV, and Influenza A Respiratory Viruses. *Frontiers in Immunology*, 11:2309, 2020. doi:10.3389/fimmu.2020.552909.
- [15] Vincent C.C. Cheng, Susanna K.P. Lau, Patrick C.Y. Woo, and Yung Yuen Kwok. Severe acute respiratory syndrome coronavirus as an agent of emerging and reemerging infection. *Clinical Microbiology Reviews*, 20(4):660–694, 2007. doi:10.1128/CMR.00023-07.
- [16] Alan J. Hay and John W. McCauley. The WHO global influenza surveillance and response system (GISRS) — A future perspective. *Influenza and other Respiratory Viruses*, 12(5):551–557, 2018. doi:10.1111/irv.12565.
- [17] Rita M. Meganck and Ralph S. Baric. Developing therapeutic approaches for twenty-first-century emerging infectious viral diseases. *Nature Medicine*, 27(3):401–410, 2021. doi:10.1038/s41591-021-01282-0.

- [18] Ruifang Zhang, Karen Eggleston, Vincent Rotimi, and Richard J. Zeckhauser. Antibiotic resistance as a global threat: Evidence from China, Kuwait and the United States. *Globalization and Health*, 2:6, 2006. doi:10.1186/1744-8603-2-6.
- [19] I. Roca, M. Akova, F. Baquero, J. Carlet, M. Cavaleri, S. Coenen, J. Cohen, D. Findlay, I. Gyssens, O. E. Heure, G. Kahlmeter, H. Kruse, R. Laxminarayan, E. Liébana, L. López-Cerero, A. MacGowan, M. Martins, J. Rodríguez-Baño, J. M. Rolain, C. Segovia, B. Sigauque, E. Taconelli, E. Wellington, and J. Vila. The global threat of antimicrobial resistance: Science for intervention. *New Microbes and New Infections*, 6:22–29, 2015. doi:10.1016/j.nmni.2015.02.007.
- [20] Christopher J Graham. The global threat of antibiotic resistance: what can be done? *Journal of Global Health Reports*, 1:e2017002, 2017. doi:10.29392/joghr.1.e2017002.
- [21] B. Lee Ligon. Sir Alexander Fleming: Scottish researcher who discovered penicillin. *Seminars in pediatric infectious diseases*, 15(1):58–64, 2004. doi:10.1053/j.spid.2004.02.002.
- [22] World Health Organization. WHO publishes list of bacteria for which new antibiotics are urgently needed. *Neurosciences*, 22(2):159, 2017.
- [23] Ming Zen Chen, Po Ren Hsueh, Li Na Lee, Chong Jen Yu, Pan Chyr Yang, and Kwen Tay Luh. Severe community-acquired pneumonia due to *Acinetobacter baumannii*. *Chest*, 120(4):1072–1077, 2001. doi:10.1378/chest.120.4.1072.
- [24] Matteo Bassetti, Antonio Vena, Antony Croxatto, Elda Righi, and Benoit Guery. How to manage *Pseudomonas aeruginosa* infections. *Drugs in Context*, 7, 2018. doi:10.7573/dic.212527.
- [25] Adèle Sakr, Fabienne Brégeon, Jean Louis Mège, Jean Marc Rolain, and Olivier Blin. *Staphylococcus aureus* nasal colonization: An update on mechanisms, epidemiology, risk factors, and subsequent infections. *Frontiers in Microbiology*, 9(OCT):2419, 2018. doi:10.3389/fmicb.2018.02419.
- [26] Aras Kadioglu, Sally Taylor, Francesco Iannelli, Gianni Pozzi, Tim J. Mitchell, and Peter W. Andrew. Upper and lower respiratory tract infection by *Streptococcus pneumoniae* is affected by pneumolysin deficiency and differences in capsule type. *Infection and Immunity*, 70(6):2886–2890, 2002. doi:10.1128/IAI.70.6.2886-2890.2002.
- [27] Paul King. *Haemophilus influenzae* and the lung (Haemophilus and the lung). *Clinical and Translational Medicine*, 1(1):10, 2012. doi:10.1186/2001-1326-1-10.

- [28] Daniel Yordanov and Tanya Strateva. *Pseudomonas aeruginosa* – a phenomenon of bacterial resistance. *Journal of Medical Microbiology*, 58(9):1133–1148, 2009. doi:10.1099/jmm.0.009142-0.
- [29] G. P. Bodey, R. Bolivar, V. Fainstein, and L. Jadeja. Infections Caused by *Pseudomonas aeruginosa*. *Clinical Infectious Diseases*, 5(2):279–313, 1983. doi:10.1093/clinids/5.2.279.
- [30] European Centre for Disease Prevention and Control. Antimicrobial resistance in the EU/EEA (EARS-Net) - Annual Epidemiological Report for 2019, 2020. URL: <https://www.ecdc.europa.eu/sites/default/files/documents/surveillance-antimicrobial-resistance-Europe-2019.pdf>.
- [31] Alessandro Cassini, Liselotte Diaz Högberg, Diamantis Plachouras, Annalisa Quattrocchi, Ana Hoxha, Gunnar Skov Simonsen, Mélanie Colomb-Cotinat, Mirjam E Kretzschmar, Brecht Devleesschauwer, Michele Cecchini, Driss Ait Ouakrim, Tiago Cravo Oliveira, Marc J Struelens, Carl Suetens, Dominique L Monnet, Reinhild Burden of AMR Collaborative Group, Karl Mertens, Thomas Struyf, Boudewijn Catry, Katrien Latour, Ivan N Ivanov, Elina G Dobreva, Arjana Tambic Andrašević, Silvija Soprek, Ana Budimir, Niki Paphitou, Helena Žemlicková, Stefan Schytte Olsen, Ute Wolff Sönksen, Pille Märtin, Marina Ivanova, Outi Lyytikäinen, Jari Jalava, Bruno Coignard, Tim Eckmanns, Muna Abu Sin, Sebastian Haller, George L Daikos, Achilleas Gikas, Sotirios Tsiodras, Flora Kontopidou, Ákos Tóth, Ágnes Hajdu, Ólafur Guðlaugsson, Karl G Kristinsson, Stephen Murchan, Karen Burns, Patrizio Pezzotti, Carlo Gagliotti, Uga Dumpis, Agne Liuimiene, Monique Perrin, Michael A Borg, Sabine C de Greeff, Jos CM Monen, Mayke BG Koek, Petter Elstrøm, Dorota Zabicka, Aleksander Deptula, Waleria Hryniewicz, Manuela Caniça, Paulo Jorge Nogueira, Paulo André Fernandes, Vera Manageiro, Gabriel A Popescu, Roxana I Serban, Eva Schréterová, Slavka Litvová, Mária Štefkovicová, Jana Kolman, Irena Klavs, Aleš Korošec, Belén Aracil, Angel Asensio, María Pérez-Vázquez, Hanna Billström, Sofie Larsson, Jacqui S Reilly, Alan Johnson, and Susan Hopkins. Attributable deaths and disability-adjusted life-years caused by infections with antibiotic-resistant bacteria in the EU and the European Economic Area in 2015: a population-level modelling analysis. *The Lancet. Infectious diseases*, 19(1):56–66, 2019. doi:10.1016/S1473-3099(18)30605-4.
- [32] Nicholas A. Turner, Batu K. Sharma-Kuinkel, Stacey A. Maskarinec, Emily M. Eichenberger, Pratik P. Shah, Manuela Carugati, Thomas L. Holland, and Vance G. Fowler. Methicillin-resistant *Staphylococcus aureus*: an overview of basic and clinical research. *Nature Reviews Microbiology*, 17(4):203–218, 2019. doi:10.1038/s41579-018-0147-4.

- [33] Athena P. Kourtis, Kelly Hatfield, James Baggs, Yi Mu, Isaac See, Erin Epton, Joelle Nadle, Marion A. Kainer, Ghinwa Dumyati, Susan Petit, Susan M. Ray, David Ham, Catherine Capers, Heather Ewing, Nicole Coffin, L. Clifford McDonald, John Jernigan, and Denise Cardo. Vital Signs: Epidemiology and Recent Trends in Methicillin-Resistant and in Methicillin-Susceptible *Staphylococcus aureus* Bloodstream Infections — United States. *MMWR. Morbidity and Mortality Weekly Report*, 68(9):214–219, 2019. doi:10.15585/mmwr.mm6809e1.
- [34] Monica Monaco, Fernanda Pimentel de Araujo, Melania Cruciani, Eliana M. Coccia, and Annalisa Pantosti. Worldwide epidemiology and antibiotic resistance of *Staphylococcus aureus*. In *Current Topics in Microbiology and Immunology*, volume 409, pages 21–56. Springer Verlag, 2017.
- [35] Stephen J. Baker, David J. Payne, Rino Rappuoli, and Ennio De Gregorio. Technologies to address antimicrobial resistance. *Proceedings of the National Academy of Sciences of the United States of America*, 115(51):12887–12895, 2018. doi:10.1073/pnas.1717160115.
- [36] Maria Kechagia, Dimitrios Basoulis, Stavroula Konstantopoulou, Dimitra Dimitriadi, Konstantina Gyftopoulou, Nikoletta Skarmoutsou, and Eleni Maria Fakiri. Health Benefits of Probiotics: A Review. *ISRN Nutrition*, 2013:1–7, 2013. doi:10.5402/2013/481651.
- [37] Hanna Sikorska and Wanda Smoragiewicz. Role of probiotics in the prevention and treatment of methicillin-resistant *Staphylococcus aureus* infections. *International Journal of Antimicrobial Agents*, 42(6):475–481, 2013. doi:10.1016/j.ijantimicag.2013.08.003.
- [38] Ulrich Glück and Jan-Olaf Gebbers. Ingested probiotics reduce nasal colonization with pathogenic bacteria (*Staphylococcus aureus*, *Streptococcus pneumoniae*, and β -hemolytic streptococci). *The American Journal of Clinical Nutrition*, 77(2):517–520, 2003. doi:10.1093/ajcn/77.2.517.
- [39] Rachael Lappan and Christopher S Peacock. *Corynebacterium* and *Dolosigranulum*: future probiotic candidates for upper respiratory tract infections. *Microbiology Australia*, 40(4):172–177, 2019. doi:10.1071/MA19051.
- [40] Silvio D. Brugger, Sara M. Eslami, Melinda M. Pettigrew, Isabel F. Escapa, Matthew T. Henke, Yong Kong, and Katherine P. Lemon. *Dolosigranulum pigrum* Cooperation and Competition in Human Nasal Microbiota. *mSphere*, 5(5), 2020. doi:10.1128/msphere.00852–20.
- [41] Klaus Klumpp and Thibaut Crépin. Capsid proteins of enveloped viruses as antiviral drug targets. *Current Opinion in Virology*, 5:63–71, 2014. doi:https://doi.org/10.1016/j.coviro.2014.02.002.

- [42] Ding Yi Fu, Ya Rong Xue, Xianghui Yu, and Yuqing Wu. Anti-virus reagents targeting the capsid protein assembly. *Journal of Materials Chemistry B*, 7(21):3331–3340, may 2019. doi:10.1039/c8tb02954g.
- [43] Elie Dolgin. The race for antiviral drugs to beat COVID - and the next pandemic. *Nature*, 592(7854):340–343, 2021. doi:10.1038/d41586-021-00958-4.
- [44] Sangdun Choi. *Introduction to systems biology*. Springer, 2007.
- [45] Bernhard Ø Palsson. *Systems biology: Properties of reconstructed networks*. Cambridge University Press, 2006.
- [46] Bernhard Ø. Palsson. *Systems biology: Constraint-based reconstruction and analysis*. Cambridge University Press, 2015.
- [47] Changdai Gu, Gi Bae Kim, Won Jun Kim, Hyun Uk Kim, and Sang Yup Lee. Current status and applications of genome-scale metabolic models. *Genome Biology*, 20(1):1–18, 2019. doi:10.1186/s13059-019-1730-3.
- [48] Liming Liu, Rasmus Agren, Sergio Bordel, and Jens Nielsen. Use of genome-scale metabolic models for understanding microbial physiology. *FEBS Letters*, 584(12):2556–2564, 2010. doi:10.1016/j.febslet.2010.04.052.
- [49] Muhammad Aqeel Ashraf and Muhammad Faheem. Editorial: Energy balances in biological systems. *Nanomaterials and Energy*, 10(1):1–1, 2021. doi:10.1680/jnaen.2021.10.1.1.
- [50] Robert Schuetz, Lars Kuepfer, and Uwe Sauer. Systematic evaluation of objective functions for predicting intracellular fluxes in *Escherichia coli*. *Molecular systems biology*, 3, 2007. doi:10.1038/MSB4100162.
- [51] Jeremy S. Edwards and Bernhard O. Palsson. Systems properties of the *Haemophilus influenzae* Rd metabolic genotype. *Journal of Biological Chemistry*, 274(25):17410–17416, 1999. doi:10.1074/jbc.274.25.17410.
- [52] Jeremy S. Edwards and Bernhard O. Palsson. The *Escherichia coli* MG1655 in silico metabolic genotype: Its definition, characteristics, and capabilities. *Proceedings of the National Academy of Sciences*, 97(10):5528–5533, 2000. doi:10.1073/pnas.97.10.5528.
- [53] Jochen Förster, Iman Famili, Patrick Fu, Bernhard O. Palsson, and Jens Nielsen. Genome-scale reconstruction of the *Saccharomyces cerevisiae* metabolic network. *Genome Research*, 13(2):244–253, 2003. doi:10.1101/gr.234503.

- [54] Natalie C. Duarte, Scott A. Becker, Neema Jamshidi, Ines Thiele, Monica L. Mo, Thuy D. Vo, Rohith Srivas, and Bernhard O. Palsson. Global reconstruction of the human metabolic network based on genomic and bibliomic data. *Proceedings of the National Academy of Sciences of the United States of America*, 104(6):1777–82, 2007. doi:10.1073/pnas.0610772104.
- [55] Ines Thiele, Swagatika Sahoo, Almut Heinken, Johannes Hertel, Laurent Heirendt, Maike K. Aurich, and Ronan M.T. Fleming. Personalized whole-body models integrate metabolism, physiology, and the gut microbiome. *Molecular Systems Biology*, 16(5):e8982, 2020. doi:10.15252/msb.20198982.
- [56] Charles J. Norsigian, Neha Pusarla, John Luke McConn, James T. Yurkovich, Andreas Dräger, Bernhard O. Palsson, and Zachary King. BiGG Models 2020: Multi-strain genome-scale models and expansion across the phylogenetic tree. *Nucleic Acids Research*, 48(D1):D402–D406, 2020. doi:10.1093/nar/gkz1054.
- [57] Rahuman S. Malik-Sheriff, Mihai Glont, Tung V.N. Nguyen, Krishna Tiwari, Matthew G. Roberts, Ashley Xavier, Manh T. Vu, Jinghao Men, Matthieu Maire, Sarubini Kananathan, Emma L. Fairbanks, Johannes P. Meyer, Chinmay Arankalle, Thawfeek M. Varusai, Vincent Knight-Schrijver, Lu Li, Corina Dueñas-Roca, Gaurhari Dass, Sarah M. Keating, Young M. Park, Nicola Buso, Nicolas Rodriguez, Michael Hucka, and Henning Hermjakob. BioModels - 15 years of sharing computational models in life science. *Nucleic Acids Research*, 48(D1):D407–D415, 2020. doi:10.1093/nar/gkz1055.
- [58] Finja Büchel, Nicolas Rodriguez, Neil Swainston, Clemens Wrzodek, Tobias Czuderna, Roland Keller, Florian Mittag, Michael Schubert, Mihai Glont, Martin Golebiewski, Martijn van Iersel, Sarah Keating, Matthias Rall, Michael Wybrow, Henning Hermjakob, Michael Hucka, Douglas B. Kell, Wolfgang Müller, Pedro Mendes, Andreas Zell, Claudine Chaouiya, Julio Saez-Rodriguez, Falk Schreiber, Camille Laibe, Andreas Dräger, and Nicolas Le Novère. Path2Models: large-scale generation of computational models from biochemical pathway maps. *BMC Systems Biology*, 7:116, 2013. doi:10.1186/1752-0509-7-116.
- [59] Alberto Noronha, Jennifer Modamio, Yohan Jarosz, Elisabeth Guerard, Nicolas Sompairac, German Preciat, Anna Dröfn Daníelsdóttir, Max Krecke, Diane Merten, Hulda S. Haraldsdóttir, Almut Heinken, Laurent Heirendt, Stefanía Magnúsdóttir, Dmitry A. Ravcheev, Swagatika Sahoo, Piotr Gawron, Lucia Friscioni, Beatriz Garcia, Mabel Prendergast, Alberto Puente, Mariana Rodrigues, Akansha Roy, Mouss Rouquaya, Luca Wiltgen, Alise Žagare, Elisabeth John, Maren Krueger, Inna Kuperstein, Andrei Zinovyev, Reinhard Schneider, Ronan M.T. Fleming, and Ines Thiele. The Virtual Metabolic Human database: Integrating human and gut microbiome metabolism with nutrition and disease. *Nucleic Acids Research*, 47(D1):D614–D624, 2019. doi:10.1093/nar/gky992.

- [60] Alina Renz, Reihaneh Mostolizadeh, and Andreas Dräger. Clinical Applications of Metabolic Models in SBML Format. In *Systems Medicine*, pages 362–371. Elsevier, 2021. doi:10.1016/b978-0-12-801238-3.11524-7.
- [61] Michael Hucka, Frank T. Bergmann, Claudine Chaouiya, Andreas Dräger, Stefan Hoops, Sarah M. Keating, Matthias König, Nicolas Le Novère, Chris J. Myers, Brett G. Olivier, Sven Sahle, James C. Schaff, Rahuman Sheriff, Lucian P. Smith, Dagmar Waltemath, Darren J. Wilkinson, and Fengkai Zhang. The Systems Biology Markup Language (SBML): Language Specification for Level 3 Version 2 Core Release 2. *Journal of integrative bioinformatics*, 16(2), 2019. doi:10.1515/jib-2019-0021.
- [62] Scott A. Becker and Bernhard O. Palsson. Genome-scale reconstruction of the metabolic network in *Staphylococcus aureus* N315: An initial draft to the two-dimensional annotation. *BMC Microbiology*, 5:8, 2005. doi:10.1186/1471-2180-5-8.
- [63] Yara Seif, Jonathan M. Monk, Nathan Mih, Hannah Tsunemoto, Saugat Poudel, Cristal Zuniga, Jared Broddrick, Karsten Zengler, and Bernhard O. Palsson. A computational knowledge-base elucidates the response of *Staphylococcus aureus* to different media types. *PLOS Computational Biology*, 15(1):e1006644, 2019. doi:10.1371/journal.pcbi.1006644.
- [64] Deok-Sun Lee, Henry Burd, Jiangxia Liu, Eivind Almaas, Olaf Wiest, Albert-László Barabási, Zoltán N Oltvai, and Vinayak Kapatral. Comparative genome-scale metabolic reconstruction and flux balance analysis of multiple *Staphylococcus aureus* genomes identify novel antimicrobial drug targets. *Journal of bacteriology*, 191(12):4015–24, 2009. doi:10.1128/JB.01743-08.
- [65] Emanuele Bosi, Jonathan M. Monk, Ramy K. Aziz, Marco Fondi, Victor Nizet, and Bernhard Ø. Palsson. Comparative genome-scale modelling of *Staphylococcus aureus* strains identifies strain-specific metabolic capabilities linked to pathogenicity. *Proceedings of the National Academy of Sciences of the United States of America*, 113(26):E3801–9, 2016. doi:10.1073/pnas.1523199113.
- [66] Matthias Heinemann, Anne Kümmel, Reto Ruinatscha, and Sven Panke. In silico genome-scale reconstruction and validation of the *Staphylococcus aureus* metabolic network. *Biotechnology and Bioengineering*, 92(7):850–864, 2005. doi:10.1002/bit.20663.
- [67] Stefanía Magnúsdóttir, Almut Heinken, Laura Kutt, Dmitry A. Ravcheev, Eugen Bauer, Alberto Noronha, Kacy Greenhalgh, Christian Jäger, Joanna Baginska, Paul Wilmes, Ronan M. T. Fleming, and Ines Thiele. Generation of genome-scale metabolic reconstructions for 773 members of the human gut microbiota. *Nature Biotechnology*, 35(1):81–89, 2016. doi:10.1038/nbt.3703.

- [68] Christian Lieven, Moritz E. Beber, Brett G. Olivier, Frank T. Bergmann, Meric Ataman, Parizad Babaei, Jennifer A. Bartell, Lars M. Blank, Siddharth Chauhan, Kevin Correia, Christian Diener, Andreas Dräger, Birgitta E. Ebert, Janaka N. Edirisinghe, José P. Faria, Adam M. Feist, Georgios Fengos, Ronan M.T. Fleming, Beatriz García-Jiménez, Vassily Hatzimanikatis, Wout van Helvoirt, Christopher S. Henry, Henning Hermjakob, Markus J. Herrgård, Ali Kaafarani, Hyun Uk Kim, Zachary King, Steffen Klamt, Edda Klipp, Jasper J. Koehorst, Matthias König, Meiyappan Lakshmanan, Dong Yup Lee, Sang Yup Lee, Sunjae Lee, Nathan E. Lewis, Filipe Liu, Hongwu Ma, Daniel Machado, Radhakrishnan Mahadevan, Paulo Maia, Adil Mardinoglu, Gregory L. Medlock, Jonathan M. Monk, Jens Nielsen, Lars Keld Nielsen, Juan Nogales, Intawat Nookaew, Bernhard O. Palsson, Jason A. Papin, Kiran R. Patil, Mark Poolman, Nathan D. Price, Osbaldo Resendis-Antonio, Anne Richelle, Isabel Rocha, Benjamín J. Sánchez, Peter J. Schaap, Rahuman S. Malik Sheriff, Saeed Shoaie, Nikolaus Sonnenschein, Bas Teusink, Paulo Vilaça, Jon Olav Vik, Judith A.H. Wodke, Joana C. Xavier, Qianqian Yuan, Maksim Zakhartsev, and Cheng Zhang. MEMOTE for standardized genome-scale metabolic model testing. *Nature Biotechnology*, 38(3):272–276, 2020. doi:10.1038/s41587-020-0446-y.
- [69] Ali Ebrahim, Joshua A. Lerman, Bernhard O. Palsson, and Daniel R. Hyduke. COBRAPy: CONstraints-Based Reconstruction and Analysis for Python. *BMC Systems Biology*, 7(1):74, 2013. doi:10.1186/1752-0509-7-74.
- [70] Alina Renz and Andreas Dräger. Curating and Comparing 114 Strain-Specific Genome-Scale Metabolic Models of *Staphylococcus aureus*. *npj Systems Biology and Applications*, 7(1):30, June 2021. doi:10.1038/s41540-021-00188-4.
- [71] Minoru Kanehisa, Miho Furumichi, Mao Tanabe, Yoko Sato, and Kanae Morishima. KEGG: new perspectives on genomes, pathways, diseases and drugs. *Nucleic Acids Research*, 45(D1):D353–D361, 2017. doi:10.1093/nar/gkw1092.
- [72] Christopher S. Henry, Matthew DeJongh, Aaron A. Best, Paul M. Frybarger, Ben Linsay, and Rick L. Stevens. High-throughput generation, optimization and analysis of genome-scale metabolic models. *Nature Biotechnology*, 28(9):977–982, 2010. doi:10.1038/nbt.1672.
- [73] Michael Römer, Johannes Eichner, Andreas Dräger, Clemens Wrzodek, Finja Wrzodek, and Andreas Zell. ZBIT Bioinformatics Toolbox: A Web-Platform for Systems Biology and Expression Data Analysis. *PLOS ONE*, 11(2):e0149263, 2016. doi:10.1371/journal.pone.0149263.
- [74] Ines Thiele and Bernhard O. Palsson. A protocol for generating a high-quality genome-scale metabolic reconstruction. *Nature Protocols*, 5(1):93–121, 2010. doi:10.1038/nprot.2009.203.

- [75] Daniel Machado, Sergej Andrejev, Melanie Tramontano, and Kiran Raosaheb Patil. Fast automated reconstruction of genome-scale metabolic models for microbial species and communities. *Nucleic Acids Research*, 46(15):7542–7553, 2018. doi:10.1093/nar/gky537.
- [76] Emil Karlsen, Christian Schulz, and Eivind Almaas. Automated generation of genome-scale metabolic draft reconstructions based on KEGG. *BMC Bioinformatics*, 19(1):467, 2018. doi:10.1186/s12859-018-2472-z.
- [77] Méziane Aite, Marie Chevallier, Clémence Frioux, Camille Trottier, Jeanne Got, María Paz Cortés, Sebastián N. Mendoza, Grégory Carrier, Olivier Dameron, Nicolas Guillaudeux, Mauricio Latorre, Nicolás Loira, Gabriel V. Markov, Alejandro Maass, and Anne Siegel. Traceability, reproducibility and wiki-exploration for “à-la-carte” reconstructions of genome-scale metabolic models. *PLOS Computational Biology*, 14(5):e1006146, 2018. doi:10.1371/journal.pcbi.1006146.
- [78] Hao Wang, Simonas Marcišauskas, Benjamín J. Sánchez, Iván Domenzain, Daniel Hermansson, Rasmus Agren, Jens Nielsen, and Eduard J. Kerkhoven. RAVEN 2.0: A versatile toolbox for metabolic network reconstruction and a case study on *Streptomyces coelicolor*. *PLoS Computational Biology*, 14(10), 2018. doi:10.1371/journal.pcbi.1006541.
- [79] Johannes Zimmermann, Christoph Kaleta, and Silvio Waschina. gapseq: informed prediction of bacterial metabolic pathways and reconstruction of accurate metabolic models. *Genome Biology*, 22(1):1–35, 2021. doi:10.1186/s13059-021-02295-1.
- [80] S. N. Mendoza, B. G. Olivier, D. Molenaar, and B. Teusink. A Systematic Assessment Of Current Genome-Scale Metabolic Reconstruction Tools. *Genome Biology*, 20(158):1–20, 2019. doi:10.1101/558411.
- [81] Matthias König and Jan Grzegorzewski. matthiaskoenig/annotatedb: annotatedb-v0.1.1, 2019. doi:10.5281/ZENODO.3268443.
- [82] Richa Agarwala, Tanya Barrett, Jeff Beck, Dennis A. Benson, Colleen Bollin, Evan Bolton, Devon Bourexis, J. Rodney Brister, Stephen H. Bryant, Kathi Canese, Mark Cavanaugh, Chad Charowhas, Karen Clark, Ilya Dondoshansky, Michael Feolo, Lawrence Fitzpatrick, Kathryn Funk, Lewis Y. Geer, Viatcheslav Gorelenkov, Alan Graeff, Wratko Hlavina, Brad Holmes, Mark Johnson, Brandi Kattman, Viatcheslav Khotomlianski, Avi Kimchi, Michael Kimelman, Masato Kimura, Paul Kitts, William Klimke, Alex Kotliarov, Sergey Krasnov, Anatoliy Kuznetsov, Melissa J. Landrum, David Landsman, Stacy Lathrop, Jennifer M. Lee, Carl Leubsdorf, Zhiyong Lu, Thomas L. Madden, Aron Marchler-Bauer, Adriana Malheiro, Peter Meric, Ilene Karsch-Mizrachi, Anatoly Mnev,

Terence Murphy, Rebecca Orris, James Ostell, Christopher O’Sullivan, Vasuki Palanigobu, Anna R. Panchenko, Lon Phan, Borys Pierov, Kim D. Pruitt, Kurt Roldarmer, Eric W. Sayers, Valerie Schneider, Conrad L. Schoch, Gregory D. Schuler, Stephen T. Sherry, Karanjit Siyan, Alexandra Soboleva, Vladimir Soussov, Grigory Starchenko, Tatiana A. Tatusova, Francoise Thibaud-Nissen, Kamen Todorov, Bart W. Trawick, Denis Vakarov, Minghong Ward, Eugene Yaschenko, Aleksandr Zasypkin, and Kerry Zbicz. Database resources of the National Center for Biotechnology Information. *Nucleic Acids Research*, 46(D1):D8–D13, 2018. doi:10.1093/nar/gkx1095.

- [83] Alex Bateman, Maria Jesus Martin, Claire O’Donovan, Michele Magrane, Rolf Apweiler, Emanuele Alpi, Ricardo Antunes, Joanna Arganiska, Benoit Bely, Mark Bingley, Carlos Bonilla, Ramona Britto, Borisas Bursteinas, Gayatri Chavali, Elena Cibrian-Uhalte, Alan Da Silva, Maurizio De Giorgi, Tunca Dogan, Francesco Fazzini, Paul Gane, Leyla Garcia Castro, Penelope Garmiri, Emma Hatton-Ellis, Reija Hieta, Rachael Huntley, Duncan Legge, Wudong Liu, Jie Luo, Alistair Macdougall, Prudence Mutowo, Andrew Nightingale, Sandra Orchard, Klemens Pichler, Diego Poggioli, Sangya Pundir, Luis Pureza, Guoying Qi, Steven Rosanoff, Rabie Saidi, Tony Sawford, Aleksandra Shypitsyna, Edward Turner, Vladimir Volynkin, Tony Wardell, Xavier Watkins, Hermann Zellner, Andrew Cowley, Luis Figueira, Weizhong Li, Hamish McWilliam, Rodrigo Lopez, Ioannis Xenarios, Lydie Bougueleret, Alan Bridge, Sylvain Poux, Nicole Redaschi, Lucila Aimò, Ghislaine Argoud-Puy, Andrea Auchincloss, Kristian Axelsen, Parit Bansal, Delphine Baratin, Marie Claude Blatter, Brigitte Boeckmann, Jerven Bolleman, Emmanuel Boutet, Lionel Breuza, Cristina Casal-Casas, Edouard De Castro, Elisabeth Coudert, Beatrice Cucho, Mikael Doche, Dolnide Dornevil, Severine Duvaud, Anne Estreicher, Livia Famiglietti, Marc Feuermann, Elisabeth Gasteiger, Sebastien Gehant, Vivienne Gerritsen, Arnaud Gos, Nadine Gruaz-Gumowski, Ursula Hinz, Chantal Hulo, Florence Jungo, Guillaume Keller, Vicente Lara, Philippe Lemercier, Damien Lieberherr, Thierry Lombardot, Xavier Martin, Patrick Masson, Anne Morgat, Teresa Neto, Nevena Nospikel, Salvo Paesano, Ivo Pedruzzi, Sandrine Pilbout, Monica Pozzato, Manuela Pruess, Catherine Rivoire, Bernd Roechert, Michel Schneider, Christian Sigrist, Karin Sonesson, Sylvie Staehli, Andre Stutz, Shyamala Sundaram, Michael Tognolli, Laure Verbregue, Anne Lise Veuthey, Cathy H. Wu, Cecilia N. Arighi, Leslie Arminski, Chuming Chen, Yongxing Chen, John S. Garavelli, Hongzhan Huang, Kati Laiho, Peter McGarvey, Darren A. Natale, Baris E. Suzek, C. R. Vinayaka, Qinghua Wang, Yuqi Wang, Lai Su Yeh, Meher Shruti Yerramalla, and Jian Zhang. UniProt: A hub for protein information. *Nucleic Acids Research*, 43(D1):D204–D212, 2015. arXiv:NIHMS150003, doi:10.1093/nar/gku989.

- [84] M. Courtot, N. Juty, C. Knupfer, D. Waltemath, A. Zhukova, A. Dräger, M. Du-

- montier, A. Finney, M. Golebiewski, J. Hastings, S. Hoops, S. Keating, D. B. Kell, S. Kerrien, J. Lawson, A. Lister, J. Lu, R. Machne, P. Mendes, M. Pocock, N. Rodriguez, A. Villeger, D. J. Wilkinson, S. Wimalaratne, C. Laibe, M. Hucka, and N. Le Novere. Controlled vocabularies and semantics in systems biology. *Molecular Systems Biology*, 7(1):543–543, 2014. doi:10.1038/msb.2011.77.
- [85] Michelle Giglio, Rebecca Tauber, Suvarna Nadendla, James Munro, Dustin Olley, Shoshannah Ball, Elvira Mitranka, Lynn M. Schriml, Pascale Gaudet, Elizabeth T. Hobbs, Ivan Erill, Deborah A. Siegele, James C. Hu, Chris Mungall, and Marcus C. Chibucos. Eco, the evidence & conclusion ontology: Community standard for evidence information. *Nucleic Acids Research*, 47(D1):D1186–D1194, 2019. doi:10.1093/nar/gky1036.
- [86] Peter D. Karp, Richard Billington, Ron Caspi, Carol A. Fulcher, Mario Latendresse, Anamika Kothari, Ingrid M. Keseler, Markus Krummenacker, Peter E. Midford, Quang Ong, Wai Kit Ong, Suzanne M. Paley, and Pallavi Subhraveti. The BioCyc collection of microbial genomes and metabolic pathways. *Briefings in Bioinformatics*, 20(4):1085–1093, jul 2019. doi:10.1093/bib/bbx085.
- [87] Claus Jonathan Fritzscheier, Daniel Hartleb, Balázs Szappanos, Balázs Papp, and Martin J. Lercher. Erroneous energy-generating cycles in published genome scale metabolic networks: Identification and removal. *PLOS Computational Biology*, 13(4):e1005494, 2017. doi:10.1371/journal.pcbi.1005494.
- [88] Jean-Christophe Lachance, Colton J. Lloyd, Jonathan M. Monk, Laurence Yang, Anand V. Sastry, Yara Seif, Bernhard O. Palsson, Sébastien Rodrigue, Adam M. Feist, Zachary A. King, and Pierre-Étienne Jacques. BOFdat: Generating biomass objective functions for genome-scale metabolic models from experimental data. *PLOS Computational Biology*, 15(4):e1006971, 2019. doi:10.1371/journal.pcbi.1006971.
- [89] Bernhard Krismer, Manuel Liebeke, Daniela Janek, Mulugeta Nega, Maren Rautenberg, Gabriele Hornig, Clemens Unger, Christopher Weidenmaier, Michael Lalk, and Andreas Peschel. Nutrient limitation governs *Staphylococcus aureus* metabolism and niche adaptation in the human nose. *PLoS pathogens*, 10(1):e1003862, 2014. doi:10.1371/journal.ppat.1003862.
- [90] Hervé Lécuyer, Juliette Audibert, Astrid Bobigny, Catherine Eckert, Caroline Jannièrre-Nartey, Annie Buu-Hoï, Jean Luc Mainardi, and Isabelle Podglajen. *Dolosigranulum pigrum* causing nosocomial pneumonia and septicemia. *Journal of Clinical Microbiology*, 45(10):3474–3475, 2007. doi:10.1128/JCM.01373-07.

- [91] M. Aguirre, D. Morrison, B.D. Cookson, F.W. Gay, and M.D. Collins. Phenotypic and phylogenetic characterization of some *Gemella*-like organisms from human infections: description of *Dolosigranulum pigrum* gen. nov., sp. nov. *Journal of Applied Bacteriology*, 75(6):608–612, dec 1993. doi:10.1111/j.1365-2672.1993.tb01602.x.
- [92] Rachael Lappan, Kara Imbrogno, Chisha Sikazwe, Denise Anderson, Danny Mok, Harvey Coates, Shyan Vijayasekaran, Paul Bumbak, Christopher C. Blyth, Sarra E. Jamieson, and Christopher S. Peacock. A microbiome case-control study of recurrent acute otitis media identified potentially protective bacterial genera. *BMC Microbiology*, 18(1):1–20, 2018. doi:10.1186/s12866-018-1154-3.
- [93] Bernhard Krismer, Christopher Weidenmaier, Alexander Zipperer, and Andreas Peschel. The commensal lifestyle of *Staphylococcus aureus* and its interactions with the nasal microbiota. *Nature Reviews Microbiology*, 15(11):675–687, 2017. doi:10.1038/nrmicro.2017.104.
- [94] Alina Renz, Lina Widerspick, and Andreas Dräger. First Genome-Scale Metabolic Model of *Dolosigranulum pigrum* Confirms Multiple Auxotrophies. *Metabolites*, 11(4):232, 2021. doi:10.3390/metabo11040232.
- [95] Michael Hucka, Frank T. Bergmann, Stefan Hoops, Sarah M. Keating, Sven Sahle, James C. Schaff, Lucian P. Smith, and Darren J. Wilkinson. The Systems Biology Markup Language (SBML): Language Specification for Level 3 Version 1 Core. *Journal of Integrative Bioinformatics*, 12(2):382–549, 2017. doi:10.1515/jib-2015-266.
- [96] Jennifer A. Bartell, Anna S. Blazier, Phillip Yen, Juliane C. Thøgersen, Lars Jelsbak, Joanna B. Goldberg, and Jason A. Papin. Reconstruction of the metabolic network of *Pseudomonas aeruginosa* to interrogate virulence factor synthesis. *Nature Communications*, 8:14631, 2017. doi:10.1038/ncomms14631.
- [97] Michael Hucka and Andrew Finney. Systems Biology Markup Language (SBML) Level 2: Structures and facilities for model definitions, 2003. Available from COMBINE at <http://identifiers.org/combine.specifications/sbml.level-2.version-1>. URL: <http://identifiers.org/combine.specifications/sbml.level-2.version-1>.
- [98] Brett G. Olivier and Frank T. Bergmann. SBML Level 3 Package: Flux Balance Constraints version 2. *Journal of integrative bioinformatics*, 15(1), 2018. doi:10.1515/jib-2017-0082.
- [99] Michael Hucka and Lucian P. Smith. SBML Level 3 package: Groups, Version 1 Release 1. *Journal of integrative bioinformatics*, 13(3):290, 2016. doi:10.2390/biecoll-jib-2016-290.

- [100] Sébastien Moretti, Olivier Martin, T. Van Du Tran, Alan Bridge, Anne Morgat, and Marco Pagni. MetaNetX/MNXref – reconciliation of metabolites and biochemical reactions to bring together genome-scale metabolic networks. *Nucleic Acids Research*, 44(D1):D523–D526, 2016. doi:10.1093/nar/gkv1117.
- [101] Ron Caspi, Richard Billington, Carol A. Fulcher, Ingrid M. Keseler, Anamika Kothari, Markus Krummenacker, Mario Latendresse, Peter E. Midford, Quang Ong, Wai Kit Ong, Suzanne Paley, Pallavi Subhraveti, and Peter D. Karp. The MetaCyc database of metabolic pathways and enzymes. *Nucleic Acids Research*, 46(D1):D633–D639, 2018. doi:10.1093/nar/gkx935.
- [102] R. P. Darveau and R. E.W. Hancock. Procedure for isolation of bacterial lipopolysaccharides from both smooth and rough *Pseudomonas aeruginosa* and *Salmonella typhimurium* strains. *Journal of Bacteriology*, 155(2):831–838, 1983. doi:10.1128/jb.155.2.831-838.1983.
- [103] Dawson D. Payne, Alina Renz, Laura J. Dunphy, Taylor Lewis, Andreas Dräger, and Jason A. Papin. An updated genome-scale metabolic network reconstruction of *Pseudomonas aeruginosa* PA14 to characterize mucin-driven shifts in bacterial metabolism. *npj Systems Biology and Applications*, 7(1):1–8, 2021. doi:10.1038/s41540-021-00198-2.
- [104] John Hopkins University. COVID-19 Map – John Hopkins Coronavirus Resource Center, 2020. URL: <https://coronavirus.jhu.edu/map.html>.
- [105] Sean Aller, Andrew Scott, Mitali Sarkar-Tyson, and Orkun S. Soyer. Integrated human-virus metabolic stoichiometric modelling predicts host-based antiviral targets against Chikungunya, Dengue and Zika viruses. *Journal of the Royal Society Interface*, 15(146):20180125, sep 2018. doi:10.1098/rsif.2018.0125.
- [106] Aarash Bordbar, Nathan E. Lewis, Jan Schellenberger, Bernhard Palsson, and Neema Jamshidi. Insight into human alveolar macrophage and *M. tuberculosis* interactions via metabolic reconstructions. *Molecular Systems Biology*, 6(1):422, 2010. doi:10.1038/msb.2010.68.
- [107] Zaid Abassi, Yara Knaney, Tony Karram, and Samuel N. Heyman. The Lung Macrophage in SARS-CoV-2 Infection: A Friend or a Foe? *Frontiers in Immunology*, 11:1312, 2020. doi:10.3389/fimmu.2020.01312.
- [108] European Centre for Disease Prevention and Control. Risk related to spread of new SARS-CoV-2 variants of concern in the EU/EEA, first update, 2021. URL: <https://www.ecdc.europa.eu/en/publications-data/covid-19-risk-assessment-spread-new-variants-concern-eueea-first-update#no-linkmoz-extension://a7a9e930-91f7-4414-964e-e7d3e863fb8f/>

enhanced-reader.html?openApp&pdf=https%253A%252F%252Fwww.ecdc.europa.eu%252Fsites%252F

- [109] Centers for Disease Control and Prevention (US). SARS-CoV-2 Variants of Concern, 2021. URL: <https://www.cdc.gov/coronavirus/2019-ncov/cases-updates/variant-surveillance/variant-info.html#Concern>.
- [110] European Centre for Disease Prevention and Control. SARS-CoV-2 variants of concern as of 12 November 2021, 2021. URL: <https://www.ecdc.europa.eu/en/covid-19/variants-concern>.
- [111] Alina Renz, Lina Widerspick, and Andreas Dräger. FBA reveals guanylate kinase as a potential target for antiviral therapies against SARS-CoV-2. *Bioinformatics*, 36(Supplement_2):i813–i821, 2020. doi:10.1093/bioinformatics/btaa813.
- [112] Alina Renz, Lina Widerspick, and Andreas Dräger. Genome-Scale Metabolic Model of Infection with SARS-CoV-2 Mutants Confirms Guanylate Kinase as Robust Potential Antiviral Target. *Genes*, 12(6):796, 2021. doi:10.3390/genes12060796.
- [113] Sanjana Kuthya, Casey L. Anthony, Tolulope Fashina, Steven Yeh, and Jessica G. Shantha. World health organization high priority pathogens: Ophthalmic disease findings and vision health perspectives. *Pathogens*, 10(4), 2021. doi:10.3390/pathogens10040442.
- [114] Cindy M. Liu, Lance B. Price, Bruce A. Hungate, Alison G. Abraham, Lisbeth A. Larsen, Kaare Christensen, Marc Stegger, Robert Skov, and Paal Skytt Andersen. *Staphylococcus aureus* and the ecology of the nasal microbiome. *Science Advances*, 1(5), 2015. doi:10.1126/sciadv.1400216.
- [115] Stefan P. W. de Vries, Hester J. Bootsma, John P. Hays, and Peter W. M. Hermans. Molecular Aspects of *Moraxella catarrhalis* Pathogenesis. *Microbiology and Molecular Biology Reviews*, 73(3):389–406, 2009. doi:10.1128/mmb.00007-09.
- [116] Ellen H.A. Van Den Munckhof, Harriet C. Hafkamp, Josephine De Kluijver, Ed J. Kuijper, Maurits N.C. De Koning, Wim G.V. Quint, and Cornelis W. Knetsch. Nasal microbiota dominated by *Moraxella* spp. is associated with respiratory health in the elderly population: A case control study. *Respiratory Research*, 21(1):181, 2020. doi:10.1186/s12931-020-01443-8.
- [117] Qian Liu, Qingyun Liu, Hongwei Meng, Huiying Lv, Yao Liu, Junlan Liu, Hua Wang, Lei He, Juanxiu Qin, Yanan Wang, Yingxin Dai, Michael Otto, and Min Li. *Staphylococcus epidermidis* Contributes to Healthy Maturation of the Nasal

- Microbiome by Stimulating Antimicrobial Peptide Production. *Cell Host and Microbe*, 27(1):68–78.e5, 2020. doi:10.1016/j.chom.2019.11.003.
- [118] Daniel N. Frank, Leah M. Feazel, Mary T. Bessesen, Connie S. Price, Edward N. Janoff, and Norman R. Pace. The Human Nasal Microbiota and *Staphylococcus aureus* Carriage. *PLoS ONE*, 5(5):e10598, 2010. doi:10.1371/journal.pone.0010598.
- [119] Hyun Jik Kim, Ara Jo, Yung Jin Jeon, Sujin An, Kang Mu Lee, Sang Sun Yoon, and Jae Young Choi. Nasal commensal *Staphylococcus epidermidis* enhances interferon- λ -dependent immunity against influenza virus. *Microbiome*, 7(1):80, 2019. doi:10.1186/s40168-019-0691-9.
- [120] Y. Uehara, H. Nakama, K. Agematsu, M. Uchida, Y. Kawakami, A. S.M. Abdul Fattah, and N. Maruchi. Bacterial interference among nasal inhabitants: Eradication of *Staphylococcus aureus* from nasal cavities by artificial implantation of *Corynebacterium* sp. *Journal of Hospital Infection*, 44(2):127–133, 2000. doi:10.1053/jhin.1999.0680.
- [121] Britney L. Hardy, Seth W. Dickey, Roger D. Plaut, Daniel P. Riggins, Scott Stibitz, Michael Otto, and D. Scott Merrell. *Corynebacterium pseudodiphtheriticum* exploits *Staphylococcus aureus* virulence components in a novel polymicrobial defense strategy. *mBio*, 10(1), 2019. doi:10.1128/mBio.02491-18.



How to build high-quality GEMs

Protocol: How to build high-quality GEMs

How to build (nasal) genome-scale metabolic models (GEMs)

1. Decide on an organism of interest

Which organism do you want to work on? _____

Which specific strain? _____

2. Browse different databases

Search in different databases for already existing models of your organism:

BioModels database

Path2Models database

BiGG Models database

Virtual Metabolic Human (VMH) database

Further databases: _____

Draft reconstruction

3. Generate a first draft of your GEM with CarveMe

CarveMe version: _____

Diamond version: _____

CPLEX version: _____

NCBI-accession number of the genome: _____

KEGG-ID (if available): _____

4. Initially analyze the model

Initial properties of the model:

_____ Reactions

_____ Metabolites

_____ Genes

SBML-Version: _____

Memote score: _____

How many genes does the organism have? _____

How many genes code for hypothetical proteins? _____

Protocol: How to build high-quality GEMs

How many genes are still missing in the model (excluding hypothetical)? _____

Annotations

5. Add annotations

Add annotations to the model. Either extract them from the 'notes' field or use ModelPolisher. When you use the ModelPolisher, remember to re-check all bounds, the objective function, and the genes!

ModelPolisher version: _____

- Bounds
- Objective function
- Genes

Memote score: _____

6. Add gene annotations

If your organism also occurs in the KEGG database, add the KEGG gene IDs (locus tags) as gene annotations. The KEGG REST API can subsequently be used to add cross-references to gene IDs in other databases.

Memote score: _____

7. Add SBO terms

The Systems Biology Ontology (SBO) contains a set of controlled vocabularies commonly used in systems biology. Every instance in our GEM can be labeled with an SBO term, including all genes, metabolites, and reactions. SBO terms can be added using libSBML.

- SBO terms for genes
- SBO terms for metabolites
- SBO terms for reactions

Memote score: _____

8. Add ECO terms

The Evidence and Conclusion Ontology (ECO) terms provide information about the curator's confidence about a reaction's inclusion into the model. The use of ECO terms is advised over the use of confidence scores, as confidence scores are not uniquely defined in the literature. ECO terms may be added based on the genes' evidence in a

Protocol: How to build high-quality GEMs

reaction's gene-protein-reaction association (GPR). If no GPR is associated with a reaction, the reaction obtains the ECO term with the lowest confidence score. Based on the evidence level of the genes in the GPR, ECO terms are added to the reactions. A helpful resource is the UniProt database, where the protein existence column gives information about the evidence level of a particular gene (inferred from homology, evidence at protein level, etc.).

9. Add KEGG Pathways

If your organism occurs in the KEGG database, extract the KEGG reaction ID from the annotations of your reactions and identify in which KEGG pathways this reaction occurs. Add all KEGG pathways for a reaction as annotations with the biological qualifier 'OCCURS_IN.'

Corrections and improvements

10. Check the chemical formulas

CarveMe adds the chemical formula of a metabolite to the 'notes' field (depending on the used CarveMe version). Transfer the chemical formula from the 'notes' field to the species description using the libSBML fbc-package. If you used the --fbc2 option during the model generation with CarveMe, check whether the chemical formulas were correctly transferred from the 'notes' field.

Memote score: _____

11. Check the charges

Charges are missing in the description of the model's metabolites. The BiGG Models database can be used to identify the charge for a given metabolite. Keep in mind that sometimes, more than one charge is provided in the database.

Memote score: _____

12. Correct mass and charge imbalances

Mass and charge imbalanced reactions can be reported by COBRApy, but also memote returns a list of mass and charge imbalanced reactions in the report. Evaluate the mass and charge imbalanced reactions and try to fix them.

Memote score: _____

Protocol: How to build high-quality GEMs

13. Reduce or eliminate orphan and dead-end metabolites

Get a list of dead-end and orphan metabolites and try to identify reactions that can be added to connect the metabolites further to the network.

Number of dead-end and orphan metabolites before: _____

Number of dead-end and orphan metabolites after: _____

Memote score: _____

14. Check for energy-generating cycles (EGC)

Models may contain thermodynamically infeasible energy-generating cycles. These models can produce energy without consuming nutrients. Fritzscheier *et al.* (DOI: 10.1371/journal.pcbi.1005494) suggested a pipeline to identify the 14 different energy metabolites ATP, CTP, GTP, UTP, ITP, NADH, NADPH, FAD, FADH, ubiquinol-8, menaquinol-8, 2-demethylmenaquinol 8, acetyl-CoA, and L-glutamate. Add a dissipation reaction for each metabolite, constrain all uptake rates to zero and subsequently maximize each dissipation reaction. If any optimization returns a result unequal zero, you have identified an energy-generating cycle that needs to be eliminated.

Memote score: _____

15. Improve the biomass objective function (BOF)

As CaveMe provides a universal biomass equation, you can further improve and specify the BOF of your organism. One possibility is to use the Python package BOFdat (DOI: 10.1371/journal.pcbi.1006971). The stoichiometric coefficients for (i) major macromolecules, (ii) inorganic ions and coenzymes, and (iii) the remaining species-specific metabolic biomass precursors are calculated and incorporated into the BOF. If genomics, transcriptomics, proteomics, lipidomics, or other experimental data is available, you can use BOFdat to refine your organism's BOF.

Memote score: _____

Model extension

16. Add further reactions

If your organism occurs in the KEGG or BioCyc database, search for metabolic genes that are so far not included in the model. Identify the reactions associated with those genes and add these reactions to the model.

Protocol: How to build high-quality GEMs

Properties of the model:

_____ Reactions
_____ Metabolites
_____ Genes

Memote score: _____

Model validation

17. Compare model predictions to experimental data

To validate your GEM, search the literature for experimental results from your organism, or even strain, of interest. These results could include additional growth media (e.g., M9 or LB), auxotrophies, gene essentialities, or other physiological properties. Simulate the conditions described in the laboratory experiment and compare the *in silico* and *in vitro* results.

Do your *in silico* results agree with the *in vitro* results? Yes No

If your results do not agree with the experimental results, identify where the discrepancy comes from. Check whether you

- need to add reactions to the model
- need to eliminate reactions from the model
- need to change the directionality of reactions
- have identified a knowledge gap

Repeat this procedure for every experimental result.

18. Test growth on SNM

We are highly interested in organisms that grow in the human nose. In Tübingen, a medium that mimics the nasal environment was experimentally developed: the Synthetic Nasal Medium (SNM) (DOI: 10.1371/journal.ppat.1003862).

Use the defined medium to simulate the growth of your organism in the human nose. To do so, change the respective metabolite's uptake rate (lower bound) to 10.0 mmol/(gDW·h). Oxygen should get an uptake rate of 20.0 mmol/(gDW·h). All other exchange reactions for metabolites that are not listed as components of the SNM should have an uptake rate of 0 mmol/(gDW·h).

Does your organism grow on SNM? Yes No

If your organism does not grow on SNM, identify which additional compounds are required to enable growth. For example, you can compare the originally active exchange reactions to the exchange reactions defined in the SNM.

Protocol: How to build high-quality GEMs

If you have identified metabolites required for growth but not defined in SNM, do some literature research to figure out whether the specific metabolite can internally be produced in your organism. If this is the case, add the missing reactions, metabolites, and genes to the model and test the growth on SNM again.

Does your organism grow on SNM now? Yes No

Model publication

19. Make the model available

Make your model available in a modeling database to meet the FAIR data principle (findable, accessible, interoperable, and reusable). Before uploading the model, check one last time whether your model is a valid SBML file.

Is your model a valid SBML file? Yes No

Additionally, check whether your model complies with the minimum standardized content for a metabolic network reconstruction (DOI: 10.15252/msb.20199235, Box 2).

Does your model comply with the standards? Yes No

Final model properties:

_____ Reactions
_____ Metabolites
_____ Genes

Final memote score: _____

Upload the valid SBML file to, e.g., BioModels. You can use a COMBINE archive and the OMEX file format (DOI: 10.1186/s12859-014-0369-z) to share all information of your model, including, e.g., media definitions. For BioModels, you can gain reviewers' access there. Then, the model is initially only visible with the corresponding log-in credentials. After publishing your model in, e.g., a scientific journal, don't forget to make your model publicly available to everyone.

Is your model available to the scientific community? Yes No

B

Accepted publications



Systems

FBA reveals guanylate kinase as a potential target for antiviral therapies against SARS-CoV-2Alina Renz^{1,2,*}, Lina Widerspich¹ and Andreas Dräger^{1,2,3,*}

¹Computational Systems Biology of Infections and Antimicrobial-Resistant Pathogens, Institute for Bioinformatics and Medical Informatics (IBMI) and ²Department of Computer Science, University of Tübingen, Tübingen 72076, Germany and ³German Center for Infection Research (DZIF), partner site Tübingen, Germany

*To whom correspondence should be addressed.

Abstract

Motivation: The novel coronavirus (SARS-CoV-2) currently spreads worldwide, causing the disease COVID-19. The number of infections increases daily, without any approved antiviral therapy. The recently released viral nucleotide sequence enables the identification of therapeutic targets, e.g. by analyzing integrated human-virus metabolic models. Investigations of changed metabolic processes after virus infections and the effect of knock-outs on the host and the virus can reveal new potential targets.

Results: We generated an integrated host–virus genome-scale metabolic model of human alveolar macrophages and SARS-CoV-2. Analyses of stoichiometric and metabolic changes between uninfected and infected host cells using flux balance analysis (FBA) highlighted the different requirements of host and virus. Consequently, alterations in the metabolism can have different effects on host and virus, leading to potential antiviral targets. One of these potential targets is guanylate kinase (GK1). In FBA analyses, the knock-out of the GK1 decreased the growth of the virus to zero, while not affecting the host. As GK1 inhibitors are described in the literature, its potential therapeutic effect for SARS-CoV-2 infections needs to be verified in *in-vitro* experiments.

Availability and implementation: The computational model is accessible at <https://identifiers.org/biomodels/db/MODEL2003020001>.

Contact: renz@informatik.uni-tuebingen.de or draeger@informatik.uni-tuebingen.de

1 Introduction

In December 2019, an outbreak of pneumonia in Wuhan, Hubei province, in China, has aroused the interest of the international community by showing alarming similarities to the outbreaks caused by other β -coronaviruses (β -CoV) like the Severe Acute Respiratory Syndrome (SARS) virus (Huang *et al.*, 2020; Hui *et al.*, 2020).

The febrile respiratory illness caused by the novel coronavirus (SARS-CoV-2) is thought to have spread as a zoonosis from the Huanan Seafood Wholesale Market, which was as a consequence shut down on January 1, 2020 to prevent further transmission events (Hui *et al.*, 2020).

On January 7, first isolation and subsequent deep-sequencing of SARS-CoV-2 from the human lower respiratory tract samples have made the genetic sequence of the virus available to the public by January 12, 2020, thus allowing for the identification of the virus as a Group 2B β -CoV (Huang *et al.*, 2020; Hui *et al.*, 2020). SARS-CoV-2 has 82% sequence similarity with the SARS virus, which has caused an outbreak originating in China in 2002 (Hui *et al.*, 2020; Zhang *et al.*, 2020). The outbreak in 2002 has peaked at a total of 8098 documented cases with a case fatality rate of 9.6% (Huang *et al.*, 2020; Hui *et al.*, 2020). In contrast, the ongoing SARS-CoV-2 has reached, at time of writing, a total of ~ 3 million cases

worldwide and caused over 200 000 deaths (John Hopkins University, 2020; WHO, 2020). The resemblance and the severe global health threat have initiated a swift and determined implementation of public health measures by the Chinese Authorities (Chen *et al.*, 2020; Hui *et al.*, 2020).

While human-to-human transmissions in a nosocomial setting were the primary route of transmission of the SARS virus outbreak, Chinese horseshoe bats have been identified as putative primary reservoir and source of zoonosis for SARS (Huang *et al.*, 2020; Hui *et al.*, 2020; Paraskevis *et al.*, 2020). Moreover, Himalayan palm civets, raccoon dogs and Chinese ferret badgers from the Guangdong wet markets (live animal retail markets) were identified as intermediate reservoirs of zoonosis (Hui *et al.*, 2020; Perlman and Netland, 2009). In the case of SARS, the switch to the human host was allowed by an adaption of the receptor binding domain of the spike (S) protein, conferring to binding capabilities to the human angiotensin-converting enzyme 2 (ACE2) (Huang *et al.*, 2020; Perlman and Netland, 2009). For the SARS-CoV-2, there was initially no proof of efficient human-to-human transmission, however, the rapid increase in cases and distinct clustering of the disease have made it clear that an efficient transmission route in-between humans exists (Chan *et al.*, 2020; Hui *et al.*, 2020). Moreover, Paraskevis *et al.* (2020) have suggested, based on full-genome evolutionary analysis, a zoonosis of SARS-CoV-2 from bats to humans. Accordingly,

Letko *et al.* have recently identified ACE2 as the SARS-CoV-2 entry receptor (Letko and Munster, 2020).

As mentioned above, the SARS-CoV-2 belongs to the subfamily of β -CoV (Huang *et al.*, 2020; Hui *et al.*, 2020). The members of the family are enveloped, single-stranded RNA (ssRNA) viruses with a positive polarity genome of up to 34 kb (Chen *et al.*, 2020; Huang *et al.*, 2020; Zhang *et al.*, 2020). Replication of the RNA genome is performed via an RNA-dependent RNA polymerase (RdRP) in double-membrane vesicles (DMVs), modified to form a reticulovesicular network (Chen *et al.*, 2020; de Wilde *et al.*, 2015; Perlman and Netland, 2009). In the DMVs, the 16 non-structural proteins (NSPs) are directly expressed as polyproteins pp1a and pp1ab from the (+)RNA genome (Chen *et al.*, 2020). Processing of pp1a/1ab by the viral main protease (M^{pro}) is essential to form the replication-transcription complex for the subsequent expression of the viral structural proteins (Chen *et al.*, 2020; Perlman and Netland, 2009).

In total, the NSPs constitute two thirds of the genome's coding capacity (Chen *et al.*, 2020). The rest encodes for structural proteins, such as the spike (S), membrane (M) and envelope (E) proteins, which cover a helical nucleocapsid made up of the nucleocapsid (N) proteins (Chen *et al.*, 2020; Perlman and Netland, 2009). Unlike NSPs, the structural and additional accessory proteins of *coronaviruses* require synthesis as subgenomic messenger RNAs via discontinuous transcription from (-)RNA templates (Chen *et al.*, 2020).

Clinically, the SARS-CoV-2 seems to be milder than the former SARS virus, although they share the symptoms of febrile illness and pneumonia (Chen *et al.*, 2020; Huang *et al.*, 2020; Hui *et al.*, 2020). It was recently shown that the SARS-CoV-2 does not only infect lower respiratory tract cells, but also upper cells in the pharyngeal region (Chen *et al.*, 2020; Huang *et al.*, 2020). Moreover, many *coronaviruses* also infect macrophages and subsequently inhibit an interferon-stimulated genes-mediated antiviral response, increasing immune evasion and pathogenicity (Chen *et al.*, 2020; Cheung *et al.*, 2005; Deng *et al.*, 2017; Huang *et al.*, 2020).

By now, no antiviral treatment for *coronaviruses* has been proven efficacious in a clinical trial (Chen *et al.*, 2020; Huang *et al.*, 2020). Recently, Zhang *et al.* (2020) have found α -ketoamides to be broad-spectrum inhibitors of *coronaviruses* by tissue-dependently inhibiting the M^{pro} of SARS in Vero cells. This inhibition has occurred in a micromolar EC_{50} range, indicating α -ketoamides to be a potential antiviral for the SARS-CoV-2 (Zhang *et al.*, 2020). However, since no therapy or vaccination is available for clinical use, the current standard of care for a SARS-CoV-2 infection is limited to the supportive treatment of symptoms (Chen *et al.*, 2020; Huang *et al.*, 2020).

As no antiviral treatment is currently available for *coronaviruses*, the identification of potential antiviral targets is of great interest. One possibility of identifying new targets is the analysis of metabolic changes in infected cells. Aller *et al.* introduced a procedure for integrated human-virus metabolic models to predict host-based antiviral targets against Chikungunya, Dengue and Zika viruses. The analysis of the integrated human-virus models predicted inhibiting effects of constrained host-reactions on virus production. These predictions included already known targets of existing antiviral drugs, demonstrating the applicability of such analysis methods (Aller *et al.*, 2018).

In our study, we integrated and analyzed a human genome-scale metabolic model (GEM) infected with SARS-CoV-2. As it is shown that *coronaviruses* infect alveolar macrophages (Cheung *et al.*, 2005; Deng *et al.*, 2017; Joel Funk *et al.*, 2012), and Bordbar *et al.* (2010) already published an extensive metabolic model of human alveolar macrophages, this model was used as a host model. For SARS-CoV-2, a virus biomass objective function (VBOF) was generated according to Aller *et al.* (2018). Subsequent analysis of the integrated host-virus model revealed potential targets for antiviral therapies.

2 Materials and methods

The methods used in this article are based on the article by Aller *et al.* (2018). Methods and analyses were adapted and expanded for the coronavirus SARS-CoV-2.

2.1 Generation of VBOF

Analogous to the biomass production or maintenance function in prokaryotic or eukaryotic metabolic models, the VBOF is a pseudo-reaction simulating the production of virus particles. It comprises nucleotides, amino acids and associated energy metabolites required for the production of the virus particles. Due to a lack of knowledge, the stoichiometric information of the virus envelope, and the dynamic information of lipids are not included in the VBOF. Hence, virus entry or exit cannot be modeled. The generation of the VBOF was performed following the seven steps described by Aller *et al.* (2018).

- Step 1: Virus genome and protein information.** The recently published virus genome and protein sequence were obtained from the NCBI database (Geer *et al.*, 2010) with the accession number NC_045512.2 in February 2020. Essential for the calculation of the nucleotide count is the classification of the SARS-CoV-2 in the Baltimore Classification System (Baltimore, 1971) that characterizes viruses based on the replication method. *Coronaviruses* fall into the Group IV classification. These viruses replicate their positive single-stranded RNA (+ssRNA) genomes via a complementary negative ssRNA. The nucleotide counts of the positive strand can be taken from the genome sequence. The nucleotides of the negative strand can be calculated by counting the complementary nucleotides of the positive strand. Both nucleotides are summed up to receive the total nucleotide count. SARS-CoV-2 has a total of 12 proteins, four structural proteins, and 8 NSPs. Structural and NSPs are not expressed equally, and this ratio needs to be considered during the calculation of the VBOF. To this date, no information about the copy number of structural proteins in *Coronaviruses* is reported.
- Step 2: Nucleotide investment.** The nucleotide count of the virus genome and its replication intermediate are summed per nucleotide. The genome copy number (C_g) was assumed to be one. According to Aller *et al.*, the moles of nucleotides were converted into grams of nucleotides per mole of the virus. After similar calculations of the amino acids and the calculation of the total molar weight of the virus based on nucleotide and amino acid content, the stoichiometric coefficients of each nucleotide in the VBOF were calculated.
- Step 3: Amino acid investment.** Analogous to the nucleotide investment, the stoichiometric coefficient of each amino acid was calculated. Instead of the genome copy number, a copy number for structural and NSPs is required. As already mentioned, the copy number of structural proteins is not reported to date. Therefore, we repeated all analyses described in this article for copy numbers of structural proteins $C_{sp}=200, 500, 800$ and 1200. With the information of the total count of each amino acid, weighted by the copy number of structural proteins, and the inclusion of the respective molar mass, the total viral molar mass was calculated (see also Step 6) to eventually calculate the stoichiometric coefficient for every amino acid.
- Steps 4 and 5: Adenosine triphosphate (ATP) requirements, and pyrophosphate liberation.** The calculations of the ATP requirements for the polymerization of amino acids and the calculation of pyrophosphate (PPi) liberation from the nucleotide polymerization were performed as described by Aller *et al.* based on the results of the previous steps. The constants $k_{ATP} = 4$ and $k_{PPi} = 1$ were chosen, as suggested by Aller *et al.* k_{ATP} is the required amount of four ATP molecules for the polymerization of amino acid monomers. The constant k_{PPi} stands for the liberation of one diphosphate molecule per polymerization of nucleotide monomers.

5. Steps 6 and 7: Total viral molar mass and final construction of the VBOF. The total molar mass of SARS-CoV-2 and the final VBOF were calculated and constructed in accordance with Aller *et al.*

2.2 Integration of the SARS-CoV-2 virus into iAB-AMØ-1410

Since Coronaviridae infect alveolar macrophages (Cheung *et al.*, 2005; Deng *et al.*, 2017), the cell-specific GEM of human alveolar macrophages iAB-AMØ-1410 was utilized. This GEM was published in 2010 by Bordbar *et al.* (2010) to give insight into human alveolar macrophages infected with *Mycobacterium tuberculosis*. The cell-specific GEM was constructed based on the human metabolic reconstruction Recon 1 (Duarte *et al.*, 2007). It consists of 3394 reactions and 2572 metabolites. The model was downloaded in SBML Level 2 Version 4 format (Hucka *et al.*, 2008) and upgraded to SBML Level 3 Version 1 format (Hucka *et al.*, 2018) using libSBML (Bornstein *et al.*, 2008). The VBOF was incorporated into the existing macrophage model, using flux bounds of 0 and 1000 as lower and upper bounds. Additionally, aggregated subsystems were added to the reactions as additional information for further analyses (Aller *et al.*, 2018; Zielinski *et al.*, 2015).

2.3 Identification of stoichiometric differences

The human alveolar macrophage model iAB-AMØ-1410 included now two pseudo-reactions for the production or maintenance of the virus and hosts biomass, respectively. The stoichiometric coefficients of shared metabolites within these pseudo-reactions can be compared to identify differences in nucleotide or amino acid requirements. First, the individual amino acid and nucleotide stoichiometric coefficients were normalized against the sum of all metabolites present in the respective biomass objective function (BOF), except for the metabolites for energy requirements. Then, the fold-change (FC) of the normalized amino acid or nucleotide was calculated:

$$FC_i = \log_2 \left(\frac{S_i^V / \sum_k S_k^V}{S_i^H / \sum_k S_k^H} \right), \quad (1)$$

with index i over nucleotides or amino acids and k over all biomass precursors (except for energy requirements). Subscripts H and V represent the use of either the host or virus biomass function. Positive values imply higher usage of nucleotides or amino acids in the virus compared to the host, while negative values imply a lower usage.

2.4 Flux balance analysis and flux variability analysis for the comparison of host- and virus-optimized states

The integration of the VBOF into the iAB-AMØ-1410 model paved the way for the analysis of metabolic changes between host and virus-optimized states. To identify an optimal state, an objective function that is optimized needs to be defined. In general, every reaction in the GEM can serve as objective function. However, biologically meaningful objective functions depend on the research question. For analyses of the growth or survival of the studied organism or cell, biomass production or maintenance reactions can be introduced into the model as pseudo-reactions. In industrial settings, the production of a specific metabolite might be of interest, and hence its production reaction can be set as the objective function. The defined objective function is then optimized with flux balance analysis (FBA). FBA is a mathematical approach using linear optimization to analyze the flow of metabolites through a metabolic network while optimizing for an objective function (Orth *et al.*, 2010). This objective function is optimized under a set of constraints. These constraints are, on the one hand, defined by the stoichiometry of the reactions, and, on the other hand, by limitations of reaction fluxes through upper and lower bounds. Hence, not only the biomass production or maintenance reaction can be adapted to a specific organism or cell type, but also the environment, in which it occurs, can be

adapted accordingly, e.g. by constraining exchange fluxes. In this work, we optimized the host-virus integrated iAB-AMØ-1410 model for either the host biomass maintenance function or the VBOF using FBA.

Since the solutions calculated by FBA are often not unique, the flux variability analysis (FVA) is a method to identify alternate optimal solutions. In FVA, the maximum and minimum possible flux for each reaction in the network is evaluated while constraining the objective value equal or close to the optimal flux (Orth *et al.*, 2010). As for the FBA, the FVA was conducted for both the host- and virus-optimized states. The results of the FVA were used in the subsequent host-derived enforcement to define the upper and lower bounds.

2.4.1 Copy number analysis

As the copy number of structural proteins in SARS-CoV-2 or coronaviruses, in general, is not reported so far, we evaluated the effect of different C_{sp} values on the growth rate in the virus-optimized state. To do so, we varied the copy number C_{sp} between 1 and 1500. The variation of C_{sp} leads to changing stoichiometric coefficients in the VBOF. Since the VBOF is optimized in the virus-optimized state, changing VBOF reactions can lead to changing optimization results.

2.4.2 Optimization analysis

The host-virus integrated model holds both the host biomass maintenance and the VBOF reaction. First, each biomass reaction was optimized individually using FBA. The objective function is defined by setting the objective coefficient to 1. In general, only one reaction (the objective function) has an objective coefficient of 1. However, one can also set the objective coefficient of several reactions to 1 to optimize for both reactions. Hence, the objective coefficient of both biomass functions from host and virus was set to 1 to optimize for both functions. Last, we used COBRAPy's (Ebrahim *et al.*, 2013) tailored objectives option to weight the influence of the two biomass reactions on the objective function. To do so, we created a linear combination of the two biomass reactions weighted by factors varying between 0 and 1, and not exceeding 1 in total. When the host maintenance function weighted 0.2, the VBOF had a weight of 0.8 in the objective function. For each combination, the model was optimized using FBA, and the flux through the two biomass reactions was reported. This analysis was repeated using the results from the host-derived enforcement analysis by adapting the bounds of the reported reactions (Section 2.5.2).

2.4.3 Metabolic analysis

The metabolic analysis was performed for the host- and virus-optimized states. The fluxes of each reaction in the two optimized states were compared by calculating the FC _{r} :

$$FC_r = \log_2 (F_r^V / F_r^H), \quad (2)$$

where the indexation r is over all reactions of the model, and the superscript H indicates flux values from the host and V from the virus-optimized models. The FCs were aggregated into the included subsystems.

2.5 Antiviral target identification

For the identification of potential antiviral targets that preferentially alter the virus growth rate while maintaining the host's biomass maintenance, each reaction in the model was evaluated using two different approaches.

2.5.1 Reaction knock-out

The first approach is a reaction knock-out that is already implemented in COBRAPy. The `single_reaction_deletion` function subsequently sets both bounds of every reaction to zero and optimizes for the chosen objective function. We ran this function twice, once with the host maintenance reaction as the objective function and once with the VBOF. Possible targets were only reported

when the growth of the virus was diminished (<99% of its initial growth rate) and when the growth rate of the virus was below the growth rate of the host-optimized state.

2.5.2 Host-derived enforcement

In the second approach, the effects of changes in flux ranges of the reactions on the VBOF are analyzed, while the metabolic system of the host-optimized state is maintained. To change the flux bounds ensuring maximal biomass maintenance of the host while optimizing for the VBOF, the results of the FVA were utilized. For each reaction, the flux range was defined as described by *Aller et al.* The corresponding minimum (F^-) and maximum (F^+) flux values from the FVA of the host (H) and virus (V) optimization and their relation define the minimum (ε^-) and maximum (ε^+) flux values. Since the cases suggested by *Aller et al.* did not cover all reactions, we added the two more conditional arguments cases 3 and 7:

$$\begin{cases} \text{continue} & \text{if } F_H^+ = F_V^+ \wedge F_H^- = F_V^- & (3) \\ \varepsilon^+ = F_H^+ & & \\ \varepsilon^- = F_H^- - \frac{F_H^+ - F_V^+}{2} & \text{if } F_H^+ > F_V^+ \wedge F_H^- \geq F_V^- & (4) \\ \varepsilon^+ = F_H^- - \frac{F_H^+ - F_V^+}{2} & \text{if } F_H^+ \leq F_V^+ \wedge F_H^- < F_V^- & (5) \\ \varepsilon^- = F_H^- & & \\ \varepsilon^+ = F_H^+ & \text{if } F_H^+ < F_V^+ \wedge F_H^- > F_V^- & (6) \\ \varepsilon^- = F_H^- & & \\ \varepsilon^+ = F_H^+ - \frac{F_H^+ - F_V^+}{2} & & \\ \varepsilon^- = F_H^- - \frac{F_H^+ - F_V^+}{2} & \text{if } F_H^+ \leq F_V^+ \wedge F_H^- \geq F_V^- & (7) \end{cases}$$

The flux values ε^- and ε^+ are set as upper and lower bounds of the corresponding reactions and the model was optimized for the VBOF. The resulting optimization result was compared to the original VBOF growth rate. When the growth rate with adapted bounds was below the threshold of 80% of the initial growth rate, the reaction was reported as potential antiviral target.

3 Results

We developed a GEM of a human macrophage infected with the SARS-CoV-2. To integrate the virus into the human macrophage, a virus biomass reaction representing the virus particle production was added to the model. This virus biomass reaction was generated based on the nucleotide and amino acid sequence.

3.1 Stoichiometric differences

The human alveolar macrophage biomass maintenance function is comprised of several macromolecules, including amino acids, deoxy-ribonucleic acid (DNA) and RNA nucleotides, components for the energy requirements and others, such as fatty acids or phospholipids. In contrast, the VBOF only consists of amino acids, RNA nucleotides and the components for the energy requirements. The stoichiometric coefficients of amino acids and nucleotides are compared by calculating the FC. The result is visualized in *Figure 1*. The stoichiometric coefficients of L-asparagine (N), L-phenylalanine (F), L-threonine (T) and L-tyrosine (Y) are increased. In contrast, the coefficients of L-glutamate (E), L-histidine (H) and L-methionine (M) are reduced. These findings might indicate an up- or down-regulation of the respective metabolic pathways.

3.2 Influence of the copy number of the structural proteins

The calculations for obtaining a virus VBOF include the parameter for the copy number of the structural proteins (C_{sp}). The copy number of structural proteins for some viruses, such as *Alpha-* and *Flaviviruses*, is known and ranges from 180 for *Flaviviruses* (*Mukhopadhyay et al., 2005*) to 240 for *Alphaviruses* (*Strauss and Strauss, 1994*). *Coronaviruses* have four major structural proteins:

the envelope (E) protein, membrane (M) protein, nucleocapsid (N) protein and the spike (S) protein (*Masters, 2006; Mortola and Roy, 2004; Wang et al., 2017*). However, the copy number of those structural proteins is not reported so far. To evaluate the influence of the copy number on the modeling results, we varied the parameter C_{sp} between 1 and 1500. As shown in *Figure 2*, the optimization result of the VBOF depends on the copy number. The growth rate increases for copy numbers ranging from 1 to 58 and decreases for higher copy numbers. For very high copy numbers of structural proteins, the growth rate seems to reach a steady state.

Coronaviruses have a diameter of ~80–120 nm (*Guy et al., 2000*), while *Alpha-* and *Flaviviruses* have only a diameter of ~40–80 nm and 30–50 nm, respectively (*Fraenkel-Conrat and Wagner, 1974*). However, not only the virus size is larger, but also the size of its structural proteins. Since no value for the structure protein copy number in *coronaviruses* was available, we continued our analyses with a C_{sp} value of 500 and verified the results with C_{sp} values of 200, 800 and 1200.

3.3 Optimization of host and virus metabolism

Potential changes in the human alveolar macrophage metabolism during virus infection were first examined by analyzing the following two scenarios: (i) the host cell is not infected and the metabolic system is optimized for the already included maintenance biomass reaction (host optimization). This scenario reflects the normal physiological state of human alveolar macrophages; (ii) the host cell is infected by the virus and is optimized solely for the production of the virus particles (virus optimization). The host optimization results in a biomass maintenance flux of 0.0267 mmol/(gDW·h), while the virus optimization returns a flux of 0.0147 mmol/(gDW·h). When defining both the host maintenance and the virus biomass function as objective functions by assigning both an objective coefficient of one simultaneously, the model only optimizes for the host maintenance reaction while the flux through the VBOF is zero. Last, the optimization result of host and virus metabolism was compared using a new constraint. With varying percentages of the host maintenance reaction and the VBOF on the objective expression, the effect on the respective biomass maintenance or growth function was investigated. The model does not predict an equilibrium state, where both the host maintenance reaction and the VBOF are active. As displayed in *Figure 3*, the switch between the maintenance of the host metabolism and the growth of the virus is at 65% of VBOF contribution to the objective expression. This switch in biomass production is rather insensitive to the structural proteins' copy number. For $C_{sp}=200$, the switch occurs at 63% virus BOF contribution, and for $C_{sp}=800$ and $C_{sp}=1200$ at 66%.

3.4 Metabolic changes in alveolar macrophages after virus infection

The flux distribution for the host- and the virus-optimized states were compared using FBA and FC calculations. For 256 reactions (7.8% of all model reactions), a FC was calculated. As expected from the stoichiometric analysis, reactions related to amino acid and nucleotide metabolism were altered. The main portion of changed reactions consists of transport reactions. However, also several reactions from other subsystems, such as the steroid, fatty acid synthesis and central metabolism, are altered. An overview of altered reactions concerning their FC is given in *Figure 4*. One needs to keep in mind that the FC can only be calculated if neither of the reaction fluxes in the host- and virus-optimized state was zero. Hence, we calculated the absolute change for reactions that have either in the host- or the virus-optimized state a flux of zero. A total of 97 reactions (2.9% of all model reactions) either turned on a previously turned off reaction or vice versa. For example, in the virus-optimized state, the virus turned off 14 reactions related to lipid metabolism.

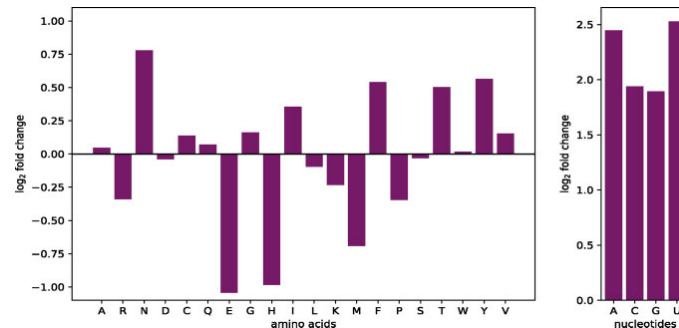


Fig. 1. FC differences in amino acid and nucleotides usage. The stoichiometric coefficients of the alveolar macrophage biomass maintenance function and the VBOF are compared using Equation (1). The left panel displays the FC of all 20 proteinogenic amino acids. The one letter code of the amino acids is used for labeling the x-axis. The right panel displays the FC of the four RNA nucleotides. The one letter code of the RNA nucleotides is used for labeling the x-axis

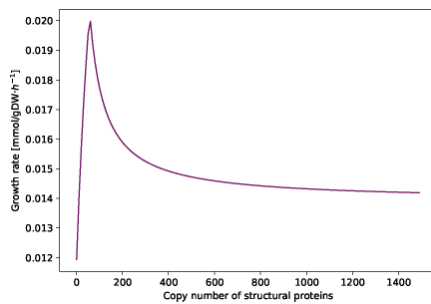


Fig. 2. Influence of C_{sp} on the growth rate. The copy number of structural proteins C_{sp} in *coronaviruses* is not yet reported in literature. Hence, its influence on the growth rate in the virus-optimized state is evaluated by varying C_{sp} between 1 and 1500. For each copy number, the integrated host-virus model was optimized for the VBOF. For C_{sp} values between 58 and 1500, the growth rate decreases

3.5 Identification of metabolic targets for antiviral strategies

FBA and FVA can be used to identify metabolic targets for antiviral strategies by comparing the host- and virus-optimized states after alterations in metabolic pathways.

3.5.1 Knock-out of reactions reveals first promising metabolic target for antiviral strategies

The results from the stoichiometric differences and metabolic changes in alveolar macrophages after virus infections with coronavirus SARS-CoV-2 indicate alternative host- and viral-optimal states. The diverging flux distribution in the two states provides an opportunity for the identification of potential antiviral targets. To identify potential antiviral targets that limit virus production, we implemented two different analysis methods: (i) reaction knock-outs and (ii) host-derived enforcement. The reaction knock-out revealed exactly one reaction (over all tested copy numbers of structural proteins), whose knock-out reduces the flux of the virus biomass to zero, while maintaining the host biomass maintenance at 100%: The guanylate kinase (GK1) reaction that converts ATP and guanosine monophosphate (GMP) to adenosine diphosphate and guanosine diphosphate (GDP):

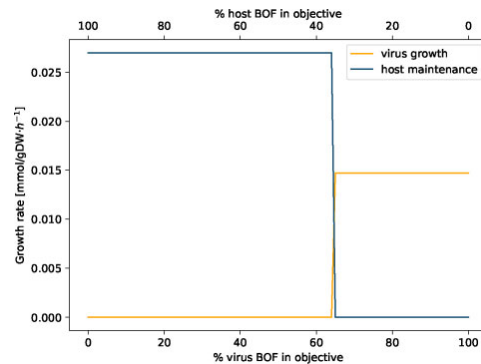


Fig. 3. Linear combination of weighted biomass functions. The objective function was defined as linear combination of the host maintenance and VBOF by weighting the proportion of the respective biomass functions between 0 and 1. The sum of the two weights always sums up to 1. At 65% of virus BOF contribution to the objective expression, a switch between host maintenance and virus production occurs

3.5.2 Host-derived enforcement substantiates the metabolic target and reveals further points of action

This reaction is also observed as a potential target in the second analysis method, the host-derived enforcement. In this approach, the reaction fluxes are constrained to ranges derived from FVA. With flux ranges outside of the optimal state of the virus, the virus production is perturbed while the host maintenance is not affected. For structural protein copy numbers between 500 and 1200, we identified four possible targets, including the GK1 that reduced the virus growth flux to below a threshold of 80% of its initial value. Further potential targets concern the availability of the amino acids L-isoleucine (I), and L-lysine (K), either via the alteration of exchange reactions, or, in case of L-isoleucine, also via the L-isoleucine transporter. In contrast to GK1, where the flux through the reaction is down-regulated (or in case of the reaction knock-outs completely knocked out), the alterations of the mentioned exchange, and transport reactions go into the other direction: the uptake of the amino acids is enabled or even enforced, leading to the host's maintenance while decreasing the growth of the virus to 50% of its initial growth. The enforcement of exchange reactions can be reached by a sufficient supply of the respective amino acid.

B.1 FBA reveals guanylate kinase as potential antiviral target against SARS-CoV-2

i818

A.Renz et al.

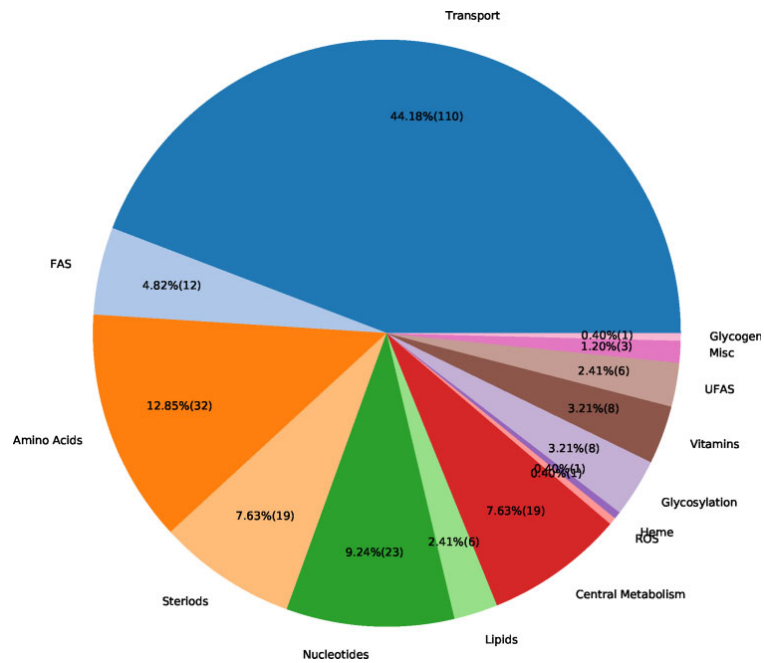


Fig. 4. Altered reactions in virus-optimized state sorted by subsystems. The integrated virus-host model was optimized for the host maintenance and VBOF using FBA. The FC was calculated using Equation (2) and the reactions with FCs were grouped into the aggregated subsystems. Many altered reactions belong to the transport subsystem, or amino acid and nucleotide metabolism

If the host has enough L-lysine in its environment, the maintenance of the host cell is ensured, while the growth of the virus is diminished. We repeated the analysis that evaluates the effect on the growth rates based on the percental involvement of the two biomass functions within the setting of L-lysine excess. Figure 5 illustrates the virus inhibiting effect of L-lysine availability in the environment. Even when the objective function is solely defined by the VBOF (% of virus BOF in objective = 100%), the growth rate of the host still exceeds the growth rate of the virus.

Analogous to L-lysine, the availability of L-isoleucine and L-tyrosine has a positive effect on the growth of the host while decreasing the growth rate of the virus by 50%. The last potential metabolic target for antiviral strategies using the host-derived enforcement approach concerns one transport reaction of L-isoleucine (ILE_{tec}). Enforcing a minimal import of L-isoleucine via this transport has the same effect as the other three targets and results in a diminution of 50% of the virus growth. All targets obtained from the host-derived enforcement for the structural protein copy numbers $C_{sp}=500, 800$ and 1200 are summarized in Table 1.

The analysis of the host-derived enforcement for $C_{sp}=200$ showed among Table 1 listed 29 reactions as potential antiviral targets that decrease the virus growth below the threshold of 80%. While the four already known targets again reduced the growth rate by 50%, the growth reduction of the other targets varies between 70% and 80% of the initial value. Of the 25 new targets, almost half (12 reactions) are associated with nucleotide metabolism. Only four further reactions are affecting the amino acid metabolism and transport mechanisms, respectively. The other reactions affected are part of the central metabolism and miscellaneous reactions.

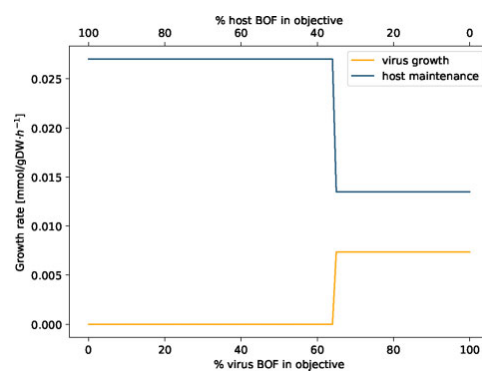


Fig. 5. Linear combination of weighted biomass function with adapted boundaries for L-lysine exchange. The objective function was defined as linear combination of the host maintenance and VBOF by weighting the proportion of the respective biomass functions between 0 and 1. The sum of the two weights always sums up to 1. The upper and lower bounds of the L-lysine exchange were adapted based on the host-derived enforcement results (see Section 2.5.2). Even when the objective function is solely defined by the VBOF (% of virus BOF in objective = 100%), the growth rate of the host still exceeds the growth rate of the virus

3.6 Existing drugs can target the predicted reactions

For the identified potential targets for antiviral therapies, we have searched for existing drugs or compounds.

Table 1. Host-derived enforcement of reactions reduced the growth of the virus

Reaction ID	Growth host (mmol/gDW.h ⁻¹)	Growth virus (mmol/gDW.h ⁻¹)	Regulation
GK1	0.027	0.00736	Down
EX_ile_L(e)	0.027	0.00736	Up
EX_lys_L(e)	0.027	0.00736	Up
ILEtec	0.027	0.00736	Up

Note: The host-derived enforcement analysis (see Section 2.5.2) disclosed reactions reducing the virus growth (initial growth rate 0.0147 mmol/gDW.h⁻¹), while maintaining the host's growth rate at 100%.

3.6.1 Direct inhibition of GK1

As GK1, also known as guanosine monophosphate kinase, was found to be an essential factor for viral growth in this study, the inhibition of the enzyme may be a feasible target in SARS-CoV-2 therapy. The enzyme catalyzes the reversible turnover of GMP or deoxyguanosine monophosphate to GDP or deoxyguanosine diphosphate by binding and transferring a phosphoryl from ATP to GMP (Hible *et al.*, 2006; Khan *et al.*, 2019). GK1 has a highly conserved structure, with a core domain, a GMP-binding domain, a lid domain with catalytic residues, as well as an ATP binding (P)-loop (Hible *et al.*, 2006; Khan *et al.*, 2019). Through activation of ganciclovir in herpes virus treatment and 6-thioguanine or 6-mercaptopurine activation in tumor treatment, GK1 plays an essential role in diverse treatment strategies (Hible *et al.*, 2006; Khan *et al.*, 2019).

Based on the results of this study, we propose a direct inhibitor of GK1 to harbor the potential of SARS-CoV-2 inhibition. Interestingly, such inhibitors have already been described in the context of antiviral therapies (Hible *et al.*, 2006; Jain *et al.*, 2016; Khan *et al.*, 2019; Navé *et al.*, 1992). For instance, Hible *et al.* have demonstrated P1-(5'-adenosyl) P5-(5'-guanosyl) pentaphosphate (Ap5G) to be a potent, bi-substrate inhibitor of GK1 (Hible *et al.*, 2006; Jain *et al.*, 2016). Although Hible *et al.* have only shown the high potency of the inhibition in *Escherichia coli*, Khan *et al.* have demonstrated the binding of Ap5G to introduce a complete closure of the human GK1, indicating inaccessibility of the substrates upon inhibitor binding (Hible *et al.*, 2006; Jain *et al.*, 2016; Khan *et al.*, 2019).

Potent GK1 inhibitors have also been presented by Navé *et al.* (1992). Accordingly, 9-phosphonoalkyl derivatives, such as 9-(6-phosphonoethyl)guanine and 9-(6,6-difluoro-6-phosphonoethyl)guanine, impede GK1 activity via competitive inhibition of GMP and non-competitive inhibition of ATP (Navé *et al.*, 1992).

3.6.2 Acyclonucleotide analogues require GK1 activation

In subsequent work, Navé *et al.* (1995) have tested the antiviral activity of 9-phosphonopentyl derivatives of guanine operating as acyclonucleotide analogues. Acyclonucleotide analogs are pro-drugs, which require the activation by GK1 to form nucleoside triphosphate analogues (Navé *et al.*, 1995). These are known to inhibit viral DNA polymerases via chain termination in diverse herpes virus and retrovirus infections (Navé *et al.*, 1995). In accordance with this, the authors have identified vinyl phosphonates (E)-9-(5-Phosphonopent-4-enyl)guanine and (E)-9-[3-(hydroxymethyl)-5-phosphonopent-4-enyl]guanine to be inhibitors of the human immunodeficiency virus 1 (HIV-1) and human cytomegalovirus (Navé *et al.*, 1995).

As this study has found increased flux through GK1 in SARS-CoV-2 infected cells, a process of pro-drug activation relying on GK1 may allow for increased activation of terminating nucleoside triphosphate analogs in infected cells compared to healthy cells. However, *coronaviruses* are +ssRNA viruses replicated by an RdRP, unlike most viruses targeted by available acyclonucleotide analogs, such as ganciclovir and acyclovir (Chen *et al.*, 2020; Navé *et al.*, 1995; Navé *et al.*, 1992). Moreover, *coronaviruses* lack a viral kinase required for activation of these acyclonucleotide analogs

(Chen *et al.*, 2020; Hible *et al.*, 2006). Despite these limitations, some market-available analogs, such as cidofovir, brincidofovir or favipiravir, do not require a viral kinase activation, and have shown *in vitro* activity against other RNA viruses (Ebola virus) and retroviruses (HIV-1) (Arias *et al.*, 2014; Dunning *et al.*, 2016; Garcia *et al.*, 1998; McCarthy *et al.*, 2016; Navé *et al.*, 1995). Thus, they may be drug candidates worth an exploration in the face of the findings in this study and the current SARS-CoV-2 outbreak.

4 Discussion

As proposed by Aller *et al.* (2018), computational approaches combining FBA and FVA to recover new metabolic antiviral targets are useful, especially in cases of new and emerging viruses, such as the SARS-CoV-2. In this study, we presented a host-virus integrated GEM using the human alveolar macrophage model iAB-AMØ-1410 as host cells and SARS-CoV-2 as virus. We identified potential targets for antiviral therapies using reaction knock-outs and host-derived enforcement approaches and analyzing their metabolic effects on host- and virus-optimized states by optimizing either for the host maintenance or the VBOF.

However, the VBOF constructed in this study only considers amino acids, nucleotides and energy requirements. It does not consider or include virus-host cell recognition, viral entry or the lipid envelope production or release (Timm and Yin, 2012). Especially the metabolic process of lipid envelope production of viruses can give further insight into potential targets for antiviral therapies. First studies with other *coronaviruses*, such as the human coronavirus 229E, suggest elevated and perturbed glycerophospholipids and fatty acids production rates in infected cells (Yan *et al.*, 2019). Yan *et al.* (2019) suggest the lipid metabolism regulation as a potential druggable target for coronavirus infections. Further information about the lipid metabolism of *coronaviruses* can enable the integration of lipids into the VBOF. Analyses with the adapted VBOF could highlight additional potential targets for antiviral therapies.

As their coding capacity is limited, *coronaviruses* strategically regulate host immune response, cell cycle, signaling and metabolism to create a favorable environment for viral replication (de Wilde *et al.*, 2015; Dirmeier *et al.*, 2020; Perlman and Netland, 2009). Accordingly, the viroids depend on cellular enzymes for the formation of progeny, which makes host cell resources a potential target to limit virion production (de Wilde *et al.*, 2015; Dirmeier *et al.*, 2020). The viral hijacking of the cellular metabolic pathways, such as glycolysis, nucleotide and lipid biosynthesis, may shift the environment of the virus to a proliferation promoting environment (Thaker *et al.*, 2019). In this study, we have also demonstrated that SARS-CoV-2 interferes with the host cell organisms, more precisely, the purine biosynthesis pathway to provide for the production of its biomass and, thus, replication.

In our study, we used a GEM. This type of model enables the analysis of metabolic changes under certain constraints. However, only the metabolic changes can be investigated. Further network reconstructions, such as dynamic signaling, regulatory or kinetic network models, can give further insight into changes in signal transduction, regulatory processes or kinetic properties (Bernhard, 2011) of virus infections. Ravindran *et al.* (2019), e.g. analyzed the effect of HIV-1 and hepatitis C virus infections using a large human signaling network. They demonstrated how the infecting virus could bring the dynamically organized host system into its control. Tan *et al.* developed a mathematical model describing the virus-induced interferon (IFN) signaling process. Dynamic analysis and numerical simulations led to the suggestion that a balance between viral replication and IFN-induced regulation is responsible for the dynamic behavior of virus-triggered signaling and also for antiviral responses (Tan *et al.*, 2012). Dynamic modeling of infections with *coronaviruses*, especially with SARS-CoV-2, could broaden the understanding of its effects on the host and give further insight into potential targets for antiviral therapies.

In this study, we used the already published iAB-AMØ-1410 GEM of human alveolar macrophages. This model does currently not include any genes or annotations. It is built upon the first human reconstruction Recon 1 (Duarte *et al.*, 2007). By now, the human

reconstruction Recon3D is available with more than 10 000 reactions, 2000 genes and almost 6000 metabolites (Brunk et al., 2018), including numerous annotations. Following the protocol of Bordbar et al. (2010), a new model of the human alveolar macrophage could be generated based on the newer version of the human reconstruction, Recon3D. The newly curated model could then be used to verify the findings from this study and to identify further targets.

As coronaviruses are reported to infect human alveolar macrophages, and a model for these cells was available, we integrated SARS-CoV-2 into this model. Nevertheless, SARS-CoV-2 is reported also to infect upper and lower respiratory tract cells, including pharyngeal regions (Chen et al., 2020; Huang et al., 2020). GEMs for bronchial epithelial cells and airway epithelial cells (AEC) are already available. Wang et al. reconstructed 126 human tissue-specific GEMs using the metabolic Context-specificity Assessed by Deterministic Reaction Evaluation algorithm. Those models are also built upon Recon 1. Furthermore, the models include fewer numbers of reactions (1242 and 1296, respectively) and lack a biomass maintenance function. Following the protocol for generati9.5 ng high-quality genome-scale metabolic reconstructions by Thiele and Palsson, meaningful models of human bronchial and AEC could be generated (Thiele and Palsson, 2010). However, generating a biomass maintenance function requires much data. Tools, such as BOFdat, can be beneficial for the generation of an appropriate BOF (Lachance et al., 2019). These models can then be used to verify the potential antiviral targets that were found in alveolar macrophages.

The integrated host–virus model suggested the supplementation of L-isoleucine and L-lysine as a potential target for antiviral therapies, as well as the inhibition of the GK1. Since compounds that directly inhibit GK1 are already known, their evaluation and verification in cell culture experiments are required for fast responses to the current spread of SARS-CoV-2 worldwide.

Funding

This work was supported by the Deutsche Forschungsgemeinschaft (German Research Foundation) under Germany's Excellence Strategy—EXC 2124—390838134 and was supported by the German Center for Infection Research.

Conflict of Interest: none declared.

Data availability

The computational model is accessible at <https://identifiers.org/bio-models.db/MODEL2003020001>.

References

Aller, S. et al. (2018) Integrated human-virus metabolic stoichiometric modeling predicts host-based antiviral targets against Chikungunya, Dengue and Zika viruses. *J. R. Soc. Interface*, **15**, 20180125.

Arias, A. et al. (2014) Favipiravir elicits antiviral mutagenesis during virus replication *in vivo*. *Elife*, **3**, e03679.

Baltimore, D. (1971) Expression of animal virus genomes. *Bacteriol. Rev.*, **35**, 235–241.

Bernhard, Ø.P. (2011) *Systems Biology: Simulation of Dynamic Network States*. Cambridge University Press, Cambridge.

Bordbar, A. et al. (2010) Insight into human alveolar macrophage and *M. tuberculosis* interactions via metabolic reconstructions. *Mol. Syst. Biol.*, **6**, 422.

Bornstein, B.J. et al. (2008) LibSBML: an API library for SBML. *Bioinformatics*, **24**, 880–881.

Brunk, E. et al. (2018) Recon3D enables a three-dimensional view of gene variation in human metabolism. *Nat. Biotechnol.*, **36**, 272–281.

Chan, J.F. et al. (2020) A familial cluster of pneumonia associated with the 2019 novel coronavirus indicating person-to-person transmission: a study of a family cluster. *Lancet*, **395**, 514–523.

Chen, Y. et al. (2020) Emerging coronaviruses: genome structure, replication, and pathogenesis. *J. Med. Virol.*, **92**, 418–423.

Cheung, C.Y. et al. (2005) Cytokine responses in severe acute respiratory syndrome coronavirus-infected macrophages *in vitro*: possible relevance to pathogenesis. *J. Virol.*, **79**, 7819–7826.

de Wilde, A.H. et al. (2015) A kinome-wide small interfering RNA screen identifies proviral and antiviral host factors in severe acute respiratory syndrome coronavirus replication, including double-stranded RNA-activated protein kinase and early secretory pathway proteins. *J. Virol.*, **89**, 8318–8333.

Deng, X. et al. (2017) Coronavirus nonstructural protein 15 mediates evasion of dsRNA sensors and limits apoptosis in macrophages. *Proc. Natl. Acad. Sci. USA*, **114**, E4251–E4260.

Dirmeier, S. et al. (2020) Host factor prioritization for pan-viral genetic perturbation screens using random intercept models and network propagation. *PLoS Comput. Biol.*, **16**, e1007587.

Duarte, N.C. et al. (2007) Global reconstruction of the human metabolic network based on genomic and bibliomic data. *Proc. Natl. Acad. Sci. USA*, **104**, 1777–1782.

Dunning, J. et al.; for the RAPIDE-BCV trial team. (2016) Experimental treatment of Ebola virus disease with brincidofovir. *PLoS One*, **11**, e0162199.

Ebrahim, A. et al. (2013) COBRAPy: COntstraints-based reconstruction and analysis for Python. *BMC Syst. Biol.*, **7**, 74.

Fraenkel-Conrat, H. and Wagner, R.R. (1974) *Comprehensive Virology: Descriptive Catalogue of Viruses*. Springer, New York.

Garcia, C.R. et al. (1998) Cidofovir in the treatment of cytomegalovirus (CMV) retinitis. *Ocul. Immunol. Inflamm.*, **6**, 195–203.

Geer, L.Y. et al. (2010) The NCBI BioSystems database. *Nucleic Acids Res.*, **38**, D492–D496.

Guy, J.S. et al. (2000) Characterization of a coronavirus isolated from a diarrheic foal. *J. Clin. Microbiol.*, **38**, 4523–4526.

Hible, G. et al. (2006) Crystal structures of GMP kinase in complex with ganciclovir monophosphate and Ap5G. *Biochimie*, **88**, 1157–1164.

Huang, C. et al. (2020) Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *Lancet*, **395**, 497–506.

Hucka, M. et al. (2008) Systems biology markup language (SBML) Level 2: structures and facilities for model definitions. *Nat. Preced.*, **58**, 10.1038/npre.2008.2715.1.

Hucka, M. et al. (2018) Systems biology markup language (SBML) level 3 version 1 core. *J. Integr. Bioinform.*, **15**, 1.

Hui, D.S. et al. (2020) The continuing 2019-nCoV epidemic threat of novel coronaviruses to global health - the latest 2019 novel coronavirus outbreak in Wuhan, China. *China. Int. J. Infect. Dis.*, **91**, 264–266.

Jain, R. et al. (2016) Insights into open/closed conformations of the catalytically active human guanylate kinase as investigated by small-angle X-ray scattering. *Eur. Biophys. J.*, **45**, 81–89.

Joel Funk, C. et al. (2012) Infection of human alveolar macrophages by human coronavirus strain 229E. *J. Gen. Virol.*, **93**, 494–503.

John Hopkins University. (2020) COVID-19 Map - John Hopkins Coronavirus Resource Center. <https://coronavirus.jhu.edu/map.html> (29 April 2020, date last accessed).

Khan, N. et al. (2019) Solution structure and functional investigation of human guanylate kinase reveals allosteric networking and a crucial role for the enzyme in cancer. *J. Biol. Chem.*, **294**, 11920–11933.

Lachance, J.-C. et al. (2019) BOFdat: generating biomass objective function stoichiometric coefficients from experimental data. *PLoS Comput. Biol.*, **15**, e1006971.

Letko, M. and Munster, V. (2020) Functional assessment of cell entry and receptor usage for SARS-CoV-2 and other lineage B betacoronaviruses. *Nat. Microbiol.*, **5**, 562–569.

Masters, P.S. (2006) The molecular biology of coronaviruses. *Adv. Virus Res.*, **66**, 193–292.

McCarthy, S.D. et al. (2016) A rapid screening assay identifies monotherapy with interferon-γ and combination therapies with nucleoside analogs as effective inhibitors of Ebola virus. *PLoS Negl. Trop. Dis.*, **10**, e0004364.

Mortola, E. and Roy, P. (2004) Efficient assembly and release of SARS coronavirus-like particles by a heterologous expression system. *FEBS Lett.*, **576**, 174–178.

Mukhopadhyay, S. et al. (2005) A structural perspective of the flavivirus life cycle. *Nat. Rev. Microbiol.*, **3**, 13–22.

Navé, J.F. et al. (1992) 9-(Phosphonoalkyl)guanine derivatives as substrates or inhibitors of guanylate kinase. *Arch. Biochem. Biophys.*, **295**, 253–257.

Navé, J.F. et al. (1995) Synthesis, antiviral activity and enzymatic phosphorylation of 9-phosphonopentyl derivatives of guanine. *Antiviral Res.*, **27**, 301–316.

Orth, J.D. et al. (2010) What is flux balance analysis? *Nat. Biotechnol.*, **28**, 245–248.

Paraskevis, D. et al. (2020) Full-genome evolutionary analysis of the novel coronavirus (2019-nCoV) rejects the hypothesis of emergence as a result of a recent recombination event. *Infect. Genet. Evol.*, **79**, 104212.

- Perlman,S. and Netland,J. (2009) Coronaviruses post-SARS: update on replication and pathogenesis. *Nat. Rev. Microbiol.*, **7**, 439–450.
- Ravindran,V. *et al.* (2019) Network controllability analysis of intracellular signalling reveals viruses are actively controlling molecular systems. *Sci. Rep.*, **9**, 1–11.
- Strauss,J.H. and Strauss,E.G. (1994) The alphaviruses: gene expression, replication, and evolution. *Microbiol. Rev.*, **58**, 491–562.
- Tan,J. *et al.* (2012) Modeling and dynamical analysis of virus-triggered innate immune signaling pathways. *PLoS One*, **7**, e48114.
- Thaker,S.K. *et al.* (2019) Viral hijacking of cellular metabolism. *BMC Biol.*, **17**, 59.
- Thiele,I. and Palsson,B. (2010) A protocol for generating a high-quality genome-scale metabolic reconstruction. *Nat. Protoc.*, **5**, 93–121.
- Timm,A. and Yin,J. (2012) Kinetics of virus production from single cells. *Virology*, **424**, 11–17.
- Wang,C. *et al.* (2017) MERS-CoV virus-like particles produced in insect cells induce specific humoral and cellular immunity in rhesus macaques. *Oncotarget*, **8**, 12686–12694.
- WHO. (2020) Coronavirus disease (COVID-19) pandemic. <https://www.who.int/emergencies/diseases/novel-coronavirus-2019> (29 April 2020, date last accessed).
- Yan,B. *et al.* (2019) Characterization of the lipidomic profile of human coronavirus-infected cells: implications for lipid metabolism remodeling upon coronavirus replication. *Viruses*, **11**, 73.
- Zhang,L. *et al.* (2020) Alpha-ketoamides as broad-spectrum inhibitors of coronavirus and enterovirus replication Structure-based design, synthesis, and activity assessment. *J. Med. Chem.*, **63**, 4562–4578.
- Zielinski,D.C. *et al.* (2015) Pharmacogenomic and clinical data link non-pharmacokinetic metabolic dysregulation to drug side effect pathogenesis. *Nat. Commun.*, **6**, 7101.



Article

First Genome-Scale Metabolic Model of *Dolosigranulum pigrum* Confirms Multiple Auxotrophies

Alina Renz ^{1,2,3} , Lina Widerspich ¹ and Andreas Dräger ^{1,2,3,4,*}

- ¹ Computational Systems Biology of Infections and Antimicrobial-Resistant Pathogens, Institute for Bioinformatics and Medical Informatics (IBMI), University of Tübingen, 72076 Tübingen, Germany, renz@informatik.uni-tuebingen.de (A.R.); lina.widerspich@student.uni-tuebingen.de (L.W.)
- ² Department of Computer Science, University of Tübingen, 72076 Tübingen, Germany
- ³ Cluster of Excellence 'Controlling Microbes to Fight Infections', University of Tübingen, 72076 Tübingen, Germany
- ⁴ German Center for Infection Research (DZIF), Partner site Tübingen, 72076 Tübingen, Germany
- * Correspondence: draeger@informatik.uni-tuebingen.de

Abstract: *Dolosigranulum pigrum* is a quite recently discovered Gram-positive coccus. It has gained increasing attention due to its negative correlation with *Staphylococcus aureus*, which is one of the most successful modern pathogens causing severe infections with tremendous morbidity and mortality due to its multiple resistances. As the possible mechanisms behind its inhibition of *S. aureus* remain unclear, a genome-scale metabolic model (GEM) is of enormous interest and high importance to better study its role in this fight. This article presents the first GEM of *D. pigrum*, which was curated using automated reconstruction tools and extensive manual curation steps to yield a high-quality GEM. It was evaluated and validated using all currently available experimental data of *D. pigrum*. With this model, already predicted auxotrophies and biosynthetic pathways could be verified. The model was used to define a minimal medium for further laboratory experiments and to predict various carbon sources' growth capacities. This model will pave the way to better understand *D. pigrum*'s role in the fight against *S. aureus*.

Keywords: *Dolosigranulum pigrum*; genome-scale metabolic model; *Staphylococcus aureus*; interaction; auxotrophy; nose microbiome



Citation: Renz, A.; Widerspich, L.; Dräger, A. First Genome-Scale Metabolic Model of *Dolosigranulum pigrum* Confirms Multiple Auxotrophies. *Metabolites* **2021**, *11*, 232. <https://doi.org/10.3390/metabo11040232>

Academic Editor: Brian Drew

Received: 18 February 2021

Accepted: 6 April 2021

Published: 9 April 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Dolosigranulum pigrum is a rare and rather newly identified opportunistic pathogen [1]. While other microbes, such as *Escherichia coli*, were already detected in the late 19th century [2], *D. pigrum* was first described in 1993 by Aguirre et al. [3]. *D. pigrum* is a Gram-positive, catalase-negative coccus growing in pairs, tetrads, and clusters [3]. In sporadic cases, *D. pigrum* is associated with diseases [1,3–7].

In 2000, the antimicrobial susceptibility and the sources of 27 clinical isolates of *D. pigrum* were determined [8]. The isolation sources ranged from blood and eye cultures from nasopharyngeal swab, sputum, sinus, gastric, and urine specimens to a spinal cord autopsy. The 27 clinical isolates were tested for their susceptibility to 15 antimicrobial agents. *D. pigrum* is a potential pathogen for humans with exceptional resistance to erythromycin but susceptibility to a wide range of other antimicrobial agents [8].

The focus shifted from *D. pigrum* as an opportunistic pathogen to its potential probiotic effect in upper respiratory tract infections in the last years. Together with *Corynebacterium pseudodiphtheriticum*, *D. pigrum* was identified as the nasopharyngeal species associated with a healthy upper respiratory tract (URT) and resistance to recurrent ear infections [9]. Multiple studies strengthen this positive association between a healthy URT and *D. pigrum*, especially in children [10–19]. Several studies indicate a decrease in the abundance of *D. pigrum* after antibiotic treatment [14,18,20], while otopathogenic genera were not affected by

antibiotic treatment [21]. Together with the antimicrobial susceptibility study by LaClaire and Facklam [8] and Lopes et al. [22], this might indicate a high sensitivity of *D. pigrum* to antibiotic agents.

D. pigrum is relevant for the URT and further parts of the respiratory tract. The abundance of *D. pigrum* is decreased in children with cystic fibrosis (CF) compared to healthy children [20]. *D. pigrum* seems to produce significantly less biomass than the conventional CF pathogen *P. aeruginosa* but is crucial for increasing tolerance of the mixed biofilm to most antibiotics [22,23]. However, the role of *D. pigrum* within the microbial communities in patients with CF is currently still not fully understood [24].

As the human nose is part of the upper respiratory tract, *D. pigrum* also plays a pivotal role in the human nasal microbiota [25]. Additionally to the negative association of *D. pigrum* with *Streptococcus pneumoniae*, it is also negatively associated with *Staphylococcus aureus*. Approximately one-third of the human population is permanently colonized by *S. aureus* [26]. It can cause severe infections with high morbidity and mortality [27]. Its methicillin-resistant strains are one of the most successful modern pathogens [28]. Liu et al. identified *D. pigrum* as a predictor of the presence or absence of *S. aureus* [29]. Brugger et al. strengthened the relevance of *D. pigrum* as a potential probiotic due to its inhibiting effects on *S. aureus*. However, the overall mechanisms behind the inhibition remain unclear. Possible mechanisms include nutrient competition or the excretion of primary or secondary metabolites [25].

Such hypotheses could be tested using genome-scale metabolic models (GEMs) of the organisms of interest, e.g., *D. pigrum* and *S. aureus*. Currently, 114 GEMs of *S. aureus* are available [30], but no single GEM of *D. pigrum* exists. Due to its increasing importance in the community with other microbes, such as *S. aureus*, *S. pneumoniae*, or *P. aeruginosa*, the need for a comprehensive and meaningful GEM is of high interest and high significance.

With a community model of *D. pigrum* and other microbes, its interactions and potential probiotic effect could be elucidated. Such interactions are complex and challenging to understand but vital for successful interventions [31]. Especially for the microbial community in the human gut, several studies already investigated the effect of gene knockouts or the absence of a community member [32,33]. The increasing interest and relevance in studying interactions in microbial communities are also highlighted by the increasing number of available tools for modeling bacterial communities, including OptCom, BacArena, or MICOM [32,34,35].

In this work, we introduce the first genome-scale metabolic model of *D. pigrum* strain 83VPs-KB5. This high-quality model comprises multiple annotations and extensive manual curation steps. It was evaluated and validated using all publicly available experimental data to this date. With this model, several auxotrophies were confirmed and additional auxotrophies were identified. To facilitate future laboratory experiments, we developed a chemically defined minimal medium with all the nutritional requirements to cultivate *D. pigrum*. These new findings will pave the way to better understanding *D. pigrum*'s role in the fight against *S. aureus*.

2. Results

The model presented in this article is the first publicly available GEM of *D. pigrum* strain 83VPs-KB5. Based on the latest recommended naming conventions of the community standardization of metabolic models [36], this model is called iDPM21RW. DPM is the species indicator and simultaneously the organism's prefix in KEGG [37]. The curators' names and the year of curation were chosen as iteration identifiers. This GEM of *D. pigrum* comprises 1241 metabolites in 1668 reactions and are encoded by 622 genes. It includes the three compartments cytosol, periplasm, and the extracellular space, which hold 974, 55, and 17 reactions, respectively, excluding transport and exchange reactions. MEMOTE is a metabolic model testing suite that determines for each tested GEM an independent and comparable score within a comprehensive overview. Standardized metabolic model tests and the evaluation of a model's annotations constitute the score [38]. The final

MEMOTE score of *iDPM21RW* amounts to 86%. For comparison, the GEM *iML1515* of *Escherichia coli* [39], for which the first version was published in [40], became steadily updated and improved over the last 20 years by the modeling community and has now reached a MEMOTE score of 91%.

2.1. Properties of the Constructed GEM

The basis for the manual extension was the draft reconstruction automatically curated with CarveMe [41]. It only requires an annotated genome file of the organism of interest. In a simple command line interface, the model can be “carved”. Other tools for the automated reconstruction of GEMs exist besides CarveMe, such as ModelSEED [42], gapseq [43], or KBase [44]. We chose CarveMe as a curation tool as it accesses the BiGG Models database [45] and uses its identifiers. These identifiers are required for subsequent successful use of the ModelPolisher [46] for adding extensive annotations. ModelSEED and gapseq both use ModelSEED identifiers, and thus, applying the ModelPolisher is currently not feasible.

The initial draft reconstruction from CarveMe included only 1499 reactions, 1095 metabolites, and 632 genes. Despite the first impression of a decrease in the number of genes, it needs to be stated that 142 genes were included twice in the initial draft model: once with the prefix G_ and once without this prefix. The duplicated genes were removed, and the 620 genes in the final reconstruction is the number of unique genes. This means that 132 additional genes, 169 reactions, and 146 metabolites were added to the model during the whole manual refinement process. During manual extension based on the KEGG database, 161 reactions, 143 metabolites, and 129 genes were added to the model. An overview of these numbers is given in Figure 1. Metabolic models may contain thermodynamically impossible energy-generating cycles. These models can charge currency metabolites such as adenosine triphosphate (ATP) or reduced nicotinamide adenine dinucleotide phosphate (NADPH) without nutrient consumption [47]. The model *iDPM21RW* was evaluated for the production of 15 energy metabolites while no nutrients were available. None of the tested energy metabolites were produced, and thus, the final model does not contain energy-generating cycles. Of the 1499 reactions, 6.23% are blocked reactions, which means that they cannot carry any flux during flux variability analysis (FVA):x. These blocked reactions might be indicators of knowledge gaps.

2.1.1. Mass and Charge Imbalances

The initial draft model had 858 mass and/or charge imbalanced reactions. After manual refinement of these mass and charge imbalances, more than 82% of the 858 imbalanced reactions were balanced. This increase in balanced reactions is also confirmed by MEMOTE when looking at the mass and charge balance score: The mass balance score increased from 52.7% to 95.6%, and the charge balance score increased from 43.2% to 93.3%. However, 137 reactions were still mass and/or charge imbalanced, none of which were blocked reactions. With novel insights into metabolites’ protonation statuses, the actual participation of metabolites in these reactions and their accurate stoichiometry, and further manual refinement, this number might be reduced even further.

2.1.2. Annotations

The model comprises annotations to various databases. These annotations were added using ModelPolisher [46] and extended manually. For the model reactions, cross-references to the databases MetaNetX [48], Biochemically, Genetically, and Genomically Structured (BiGG) Models [45], UniProt [49], Kyoto Encyclopedia of Genes and Genomes (KEGG) [37], RHEA [50], and BioCyc [51] are included, as are the corresponding EC-numbers, where available. The annotations of the model metabolites contain cross-references to the databases KEGG [37], BiGG [45], BioCyc [51], the Human Metabolome Database (HMDB) [52], MetaNetX [48], and lipidmaps [53]. The gene annotations contain cross-references to KEGG [37] and the NCBI protein database [54].

Additionally, all genes, metabolites, and reactions were further annotated with a term from the Systems Biology Ontology (SBO) [55]. All metabolites were assigned the SBO term SBO:0000247 for “simple chemical”, and all model genes received the SBO term SBO:0000243 coding for “gene”. In total, 22 different SBO terms were assigned to the reactions. The most prominent SBO term with a relative abundance of 31.32% is the SBO:0000176, coding for “biochemical reaction”. All other SBO terms describe more precisely the biochemical reactions, such as the SBO term SBO:0000216 with a relative abundance of 6.5%, coding for a “phosphorylation” reaction. The relative occurrence of all 22 SBO terms is depicted in Figure 1.

The model reactions were further annotated using terms from the Evidence and Conclusion Ontology (ECO) [56]; 38.7% of the model reactions were inferred from background scientific knowledge, 10.1% had similarity evidence, 20.5% held a computational inference, and 30.7% even had sequence similarity evidence. The overall occurrence of the ECO terms is displayed in Figure 1.

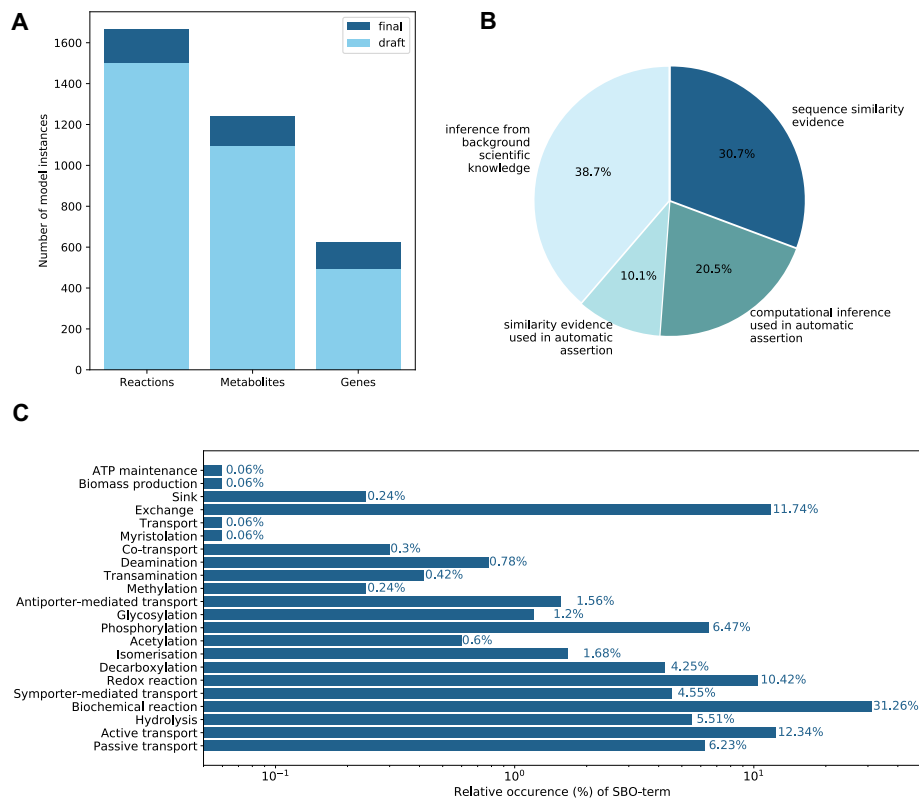


Figure 1. Properties of the genome-scale metabolic model (GEM) *iDPM21RW*. This figure illustrates various model properties. (A) The number of model instances in the draft and the refined final reconstruction is indicated. In total, 132 genes, 169 reactions, and 146 metabolites were added to the final reconstruction. (B) Evidence and Conclusion Ontology (ECO) terms indicate the confidence of inclusion for the model’s reactions. Increasing color intensity corresponds to increasing confidence. (C) Systems Biology Ontology (SBO) terms were used to annotate the models’ reactions further [57]. The axis of the relative occurrence is given as a log scale.

2.1.3. Biomass Objective Function

CarveMe creates a general biomass objective function (BOF) during the curation process [41]. This initial BOF was updated using BOFdat [58]. BOFdat is a Python package to generate and improve a BOF based on organism-specific experimental data. In three steps, the stoichiometric coefficients for major macromolecules, inorganic ions and coenzymes, and other species-specific metabolic biomass precursors were calculated and incorporated into the BOF. With the help of the DNA sequence of *D. pigrum*, five stoichiometric coefficients associated with the macromolecule DNA (deoxyadenosine triphosphate (dATP), deoxythymidine triphosphate (dTTP), deoxycytidine triphosphate (dCTP), deoxyguanosine triphosphate (dGTP), and diphosphate) were updated using the first step of the BOFdat algorithm. In the second step of the BOF dat algorithm, the coefficients of inorganic ions and coenzymes were calculated and updated based on macromolecular weight fractions. Fifteen stoichiometric coefficients associated with coenzymes and inorganic ions were updated, and nine were additionally integrated into the BOF. The coefficients of other macromolecules, such as RNA, proteins, or lipids, could not be updated due to a lack of available experimental data. The same was found for the stoichiometric coefficients of other species-specific metabolic biomass precursors as no required gene essentiality data was available. All metabolites included in the BOF, and their stoichiometric coefficients are listed in the supplementary Table S2.

2.1.4. Subsystems and Groups

The group plugin is available from SBML Level 3 [59]. In total, 82 subsystems were added to the plugin as groups. Reactions associated with these subsystems or pathways were added as members to the respective groups. It needs to be highlighted that the subsystems and pathways were extracted from the KEGG database [37]. Thus, only reactions with annotated KEGG identifiers could be mapped to the respective groups. Among the three groups with the most members and, thus, reactions is the subsystem of metabolic pathways with 411 members, the group of biosynthesis of secondary metabolites with 95 members, and the subsystem of microbial metabolism in diverse environments with 79 associated reactions.

2.2. Evaluating Auxotrophies and Predicted Biosynthesis

After creating and refining a draft reconstruction and its conversion into a mathematical model, the model needs to be verified, evaluated, and validated. In this step, the model-predicted phenotypes are compared with the experimental data [60]. Brugger et al. predicted the biosynthesis, uptake, and degradation of amino acids, carbohydrates, polyamines, and enzyme cofactors in eleven *D. pigrum* strains by evaluating their genetic content [25]. COBRApy [61] was used for all evaluation steps.

2.2.1. Auxotrophies and Biosynthesis

Brugger et al. identified a methionine auxotrophy in all evaluated *D. pigrum* strains. In our model, growth without methionine supplementation was initially possible, indicating the potential for model adaption and refinement. Nineteen reactions were associated with methionine, which were all carefully checked. We identified and removed four reactions without evidence in KEGG [37], BioCyc [51], or a significant hit in a BLAST search [62]. With these alterations, the model is now incapable of producing methionine, as Brugger et al. observed in their study [25]. An ATP-binding cassette (ABC) transporter for the uptake of methionine is present in the model.

D. pigrum has a likely auxotrophy for arginine [25]. We could confirm this observation with our in silico predictions.

Further auxotrophies for the polyamines putrescine and spermidine were predicted [25]. We could also confirm these observations based on our in silico simulations. Additionally, the identified putative ABC-type spermidine transporter and the putative putrescine transporter were already included in the model.

The predicted biotin auxotrophy was initially not observed in the model. For that reason, two biosynthesis reactions were removed from the model, both of which did not have gene–protein reaction (GPR) associations. Instead, the biotin energy-coupling factor (ECF) transporter was added. Brugger et al. identified a biotin-protein ligase in two of the eleven investigated strains. We found the gene for the biotin-protein ligase in the genome of *D. pigrum* strain 83VPs-KB5. Thus, the respective reaction was added to the model.

The last predicted auxotrophy pertained to nicotinic acid (niacin) [25]. This auxotrophy was also observed in the in silico simulations. The identified transporter [25] was already present in the model. The same was found for additional reactions in the conversion of niacin or nicotinamide to NAD⁺ and NADP⁺ with their respective genes. Only one reaction was adapted, as the described enzyme was reclassified into another Enzyme Commission (EC) number with slightly different reactants. The reaction now additionally requires ATP and water instead of a proton and produces adenosine diphosphate (ADP) and a phosphate.

D. pigrum is capable of synthesizing L-glutamine from L-glutamate [25]. All required reactions are included in the model, and in silico simulations verify the production of L-glutamine. All predicted auxotrophies and biosynthesis are summarized in Table 1.

Table 1. Overview of reported auxotrophies and biosynthetic pathways. Brugger et al. investigated auxotrophies and biosynthetic pathways based on functional genomic predictions [25]. Reported auxotrophies and biosynthesis were verified using *iDPM21RW* and in silico predictions. Additionally, reported reactions and transporters were checked for their presence. A black check-mark (✓) indicates a correct prediction or occurrence of the model's instance; a check-mark in gray (✓) indicates a correct prediction or occurrence after model modifications; and a black cross (✗) indicates a discrepancy between the functional genomic predictions and the model. However, we could not find any discrepancy for auxotrophies and biosynthetic pathways.

	Methionine	Arginine	Glutamine	Putrescine	Spermidine	Biotin	Niacin
Auxotrophy	✓	✓		✓	✓	✓	✓
Biosynthesis			✓				
Reported reactions			✓			✓	✓
Transporter	✓	✓		✓	✓	✓	

Several auxotrophies in *D. pigrum* are already reported in the literature [25]. Thus, we investigated further amino acid auxotrophies or de novo biosynthesis capabilities in silico. As seen in Figure 2, only the seven amino acids L-alanine, L-aspartate, L-glutamine, glycine, L-serine, L-asparagine, and L-tyrosine could be synthesized de novo in our simulation. For all other amino acids, *D. pigrum* seems to be dependent on external sources.

2.2.2. Carbohydrate Metabolism

Besides auxotrophies, Brugger et al. also investigated the carbohydrate metabolism of eleven *D. pigrum* strains using functional genomic predictions [25]. They found that there is no tricarboxylic acid (TCA) cycle in *D. pigrum*. Our in silico investigations confirmed this finding: only the two reported reactions catalyzed by fumarate-hydratase (FUM) and the TCA-associated dihydrolipoyl dehydrogenase (AKGDH) are present in the model. Those two reactions are illustrated in Figure 3.

B.2 First GEM of *D. pigrum* confirms multiple auxotrophies

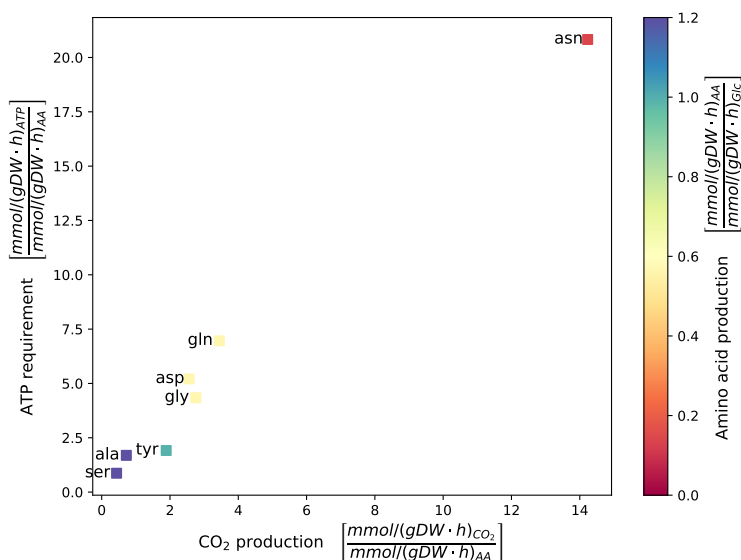


Figure 2. Amino acid production in *D. pigrum*. The exchange reaction of the amino acid of interest was closed to investigate the amino acid production capacity of *D. pigrum* in silico. A sink reaction of the respective amino acid was optimized while maintaining the growth rate at a fixed value of 0.2 mmol/(gDW · h) and maximum growth rate 0.278 mmol/(gDW · h). Only the seven shown amino acids could be synthesized de novo. For every amino acid, the ATP requirement and the CO₂ production were calculated. The color indicates the amino acid production rate concerning the carbon source (glucose). Amino acids are shown with their respective three-letter code.

Brugger et al. identified V-type ATPases in all investigated strains, which can hydrolyze but not synthesize ATP [25]. The model *iDPM21RW* does not currently include any V-type ATPase as there is no corresponding reaction in the BiGG Models database [45].

The authors investigated anaerobic respiratory reductases and did not identify butyryl-CoA-reductases [25]. *iDPM21RW* does not contain the corresponding reaction BTCOARx, confirming the findings by Brugger et al.

Further investigations concerned ten reactions from glycolysis, including glucokinase, phosphoglycerate kinase, and pyruvate kinase. All ten reactions were confirmed with *iDPM21RW*. Additionally, Brugger et al. predicted the presence and absence of various enzymes relevant for homofermentation to lactate. Each enzyme and its corresponding reaction were checked in our model.

Appendix B Accepted publications

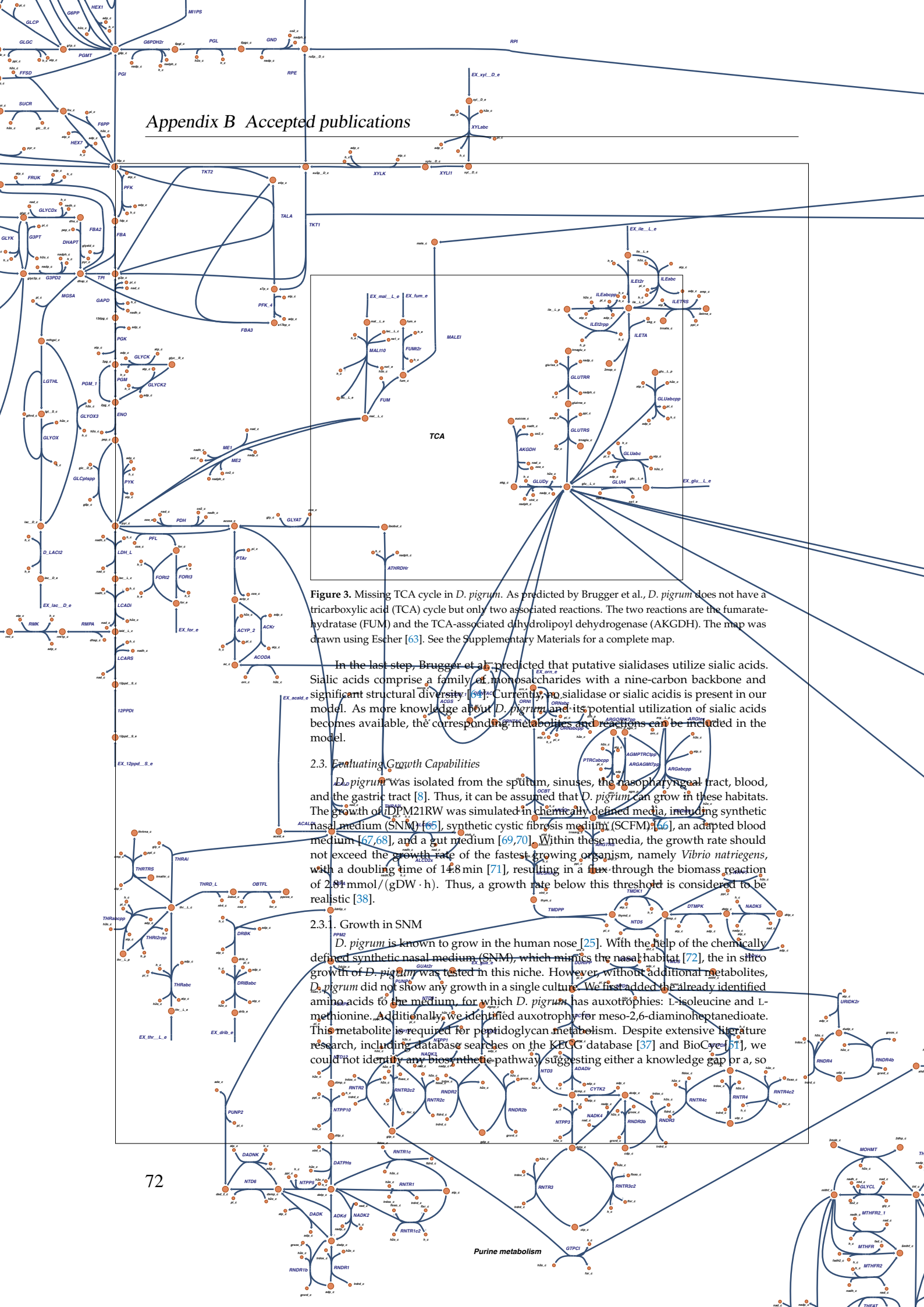


Figure 3. Missing TCA cycle in *D. pigrum*. As predicted by Brugger et al., *D. pigrum* does not have a tricarboxylic acid (TCA) cycle but only two associated reactions. The two reactions are the fumarate-hydratase (FUM) and the TCA-associated dihydrolipoyl dehydrogenase (AKGDH). The map was drawn using Escher [63]. See the Supplementary Materials for a complete map.

In the last step, Brugger et al. predicted that putative sialidases utilize sialic acids. Sialic acids comprise a family of monosaccharides with a nine-carbon backbone and significant structural diversity [64]. Currently, no sialidase or sialic acids is present in our model. As more knowledge about *D. pigrum* and its potential utilization of sialic acids becomes available, the corresponding metabolites and reactions can be included in the model.

2.3. Evaluating Growth Capabilities

D. pigrum was isolated from the sputum, sinuses, the nasopharyngeal tract, blood, and the gastric tract [8]. Thus, it can be assumed that *D. pigrum* can grow in these habitats. The growth of DPM21RW was simulated in chemically defined media, including synthetic nasal medium (SNM) [65], synthetic cystic fibrosis medium (SCFM) [66], an adapted blood medium [67,68], and a gut medium [69,70]. Within these media, the growth rate should not exceed the growth rate of the fastest growing organism, namely *Vibrio natriegens*, with a doubling time of 148 min [71], resulting in a flux through the biomass reaction of 2.04 mmol/(gDW · h). Thus, a growth rate below this threshold is considered to be realistic [38].

2.3.1. Growth in SNM

D. pigrum is known to grow in the human nose [25]. With the help of the chemically defined synthetic nasal medium (SNM), which mimics the nasal habitat [72], the *in silico* growth of *D. pigrum* was tested in this niche. However, without additional metabolites, *D. pigrum* did not show any growth in a single culture. We first added the already identified amino acids to the medium, for which *D. pigrum* has auxotrophies: L-isoleucine and L-methionine. Additionally, we identified auxotrophy for meso-2,6-diaminoheptanedioate. This metabolite is required for peptidoglycan metabolism. Despite extensive literature research, including database searches on the KEGG database [37] and BioCyc [51], we could not identify any biosynthetic pathway suggesting either a knowledge gap or a, so

far, unknown auxotrophy. After those three additions to the medium, the model predicted a realistic growth rate of 0.2824 mmol/(gDW · h) in the SNM.

2.3.2. Growth in SCFM

D. pigrum is also reported to play a role in CF patients, although its role within the microbial community is currently not fully understood [24]. A chemically defined medium is available, mimicking the lungs of patients with CF. The in silico growth capabilities of *D. pigrum* in this synthetic cystic fibrosis medium (SCFM) were evaluated. Similar to the in silico growth of *D. pigrum* in the SNM, the bacterium did not grow in the SCFM without supplementing additional metabolites. Since we expect the trace elements manganese, zinc, copper, cobalt, and nickel to be contained in the medium even without explicit addition, they are not further examined here. The trace elements were simply added to the medium definition. The other required metabolites were riboflavin, thiamine, nicotinate, 4-aminobenzoate, and, as in the SNM, meso-2,6-diaminoheptanedioate. In their preprint from 2019, Brugger et al. stated that all eleven investigated strains of *D. pigrum* lacked genes for the biosynthesis of thiamine and the de novo synthesis of niacin/nicotinate/nicotinamide [73]. For riboflavin, ten of the eleven strains lacked the synthesis cluster of riboflavin. Our model strain *D. pigrum* 83VPs KB was, however, not among the investigated strains. A literature search in several databases, including KEGG [37] and BioCyc [51], and BLAST searches for relevant biosynthetic enzymes did not reveal any hits for the synthesis of thiamine, riboflavin, and nicotinate, confirming the findings of Brugger et al. For the metabolite 4-aminobenzoate, no information was found in the literature. In KEGG [37] and BioCyc [51], the metabolite was reported in *D. pigrum*, but no synthesis pathways were available. No significant BLAST hits were detected for the enzyme aminodeoxychorismate lyase, which catalyzes the synthesis of 4-aminobenzoate. After adding the required metabolites to the medium, the growth rate of *D. pigrum* in SCFM was 0.2824 mmol/(gDW · h).

2.3.3. Growth in the Blood Medium

D. pigrum was isolated from blood samples and even cultivated in aerobic and anaerobic blood culture bottles [1,8]. A chemically defined medium simulating the human blood is available and was used for the in silico simulations [67]. This medium definition was slightly modified and adapted [68]. For the SCFM medium, the trace elements manganese, zinc, copper, cobalt, and nickel are required for growth but are not further investigated here and are only added to the medium definition. Analogously, the compounds 4-aminobenzoate and meso-2,6-diaminoheptanedioate are required for growth, as *D. pigrum* seems to be auxotrophic for those compounds. The in silico simulations predicted a realistic growth rate of 1.908 mmol/(gDW · h) with these metabolites.

As stated above, *D. pigrum* can grow anaerobically in blood cultures. Despite diverse approaches, we could not yet simulate these conditions in our in silico model. There is still much to discover about *D. pigrum*, and with additional information and laboratory experiments, the model could be extended to simulate anaerobic growth in blood cultures.

2.3.4. Growth in the Gastrointestinal Tract

The growth of *D. pigrum* was simulated in the gastrointestinal tract. A defined medium of the European diet from the Virtual Metabolic Human (VMH) database was used for this purpose [69,70]. As in the previously tested media, trace minerals, such as manganese, cobalt, zinc, nickel, and sulfate, were missing in the defined gut medium. The compounds 4-aminobenzoate and meso-2,6-diaminoheptanedioate were again required to enable growth aerobically with a growth rate of 1.088 mmol/(gDW · h).

2.3.5. Definition of a Minimal Medium for *D. pigrum*

The previous analysis of *D. pigrum*'s growth behavior and the investigated auxotrophies indicate specific requirements for its environment and successful colonization. To

obtain a detailed picture of all environmental requirements for successful growth, we defined a minimal medium for the growth of *D. pigrum* with the help of *iDPM21RW* (Supplementary Table S3). This minimal medium contains 33 metabolites, which are all listed in Table 2. It includes the 13 amino acids that cannot be synthesized *de novo* (also see Figure 2) and 13 trace minerals. As a carbon source, D-glucose was chosen. However, in the following section, the growth on different carbon sources is investigated. The three vitamins thiamine (vitamin B₁), riboflavin (vitamin B₂), and niacin (vitamin B₃) are also required to enable growth. 4-aminobenzoate was already mentioned several times to be crucial for *D. pigrum*'s growth. For this reason, it was also included in the minimal medium. The same was found for meso-2,6-diaminoheptanedioate, which was also added to the minimal medium definition. Finally, oxygen is also required for the growth simulations, as anaerobic growth is not yet enabled. Within this minimal medium, the simulated growth rate amounts to 0.2784 mmol/(gDW · h).

Table 2. Definition of a minimal medium for *D. pigrum*. Since *D. pigrum* holds many auxotrophies and several requirements for its environment to grow, we defined a minimal medium containing all relevant metabolites. The minimal medium comprises in total 33 compounds, including the 13 amino acids that cannot be produced, 13 trace minerals, D-glucose as a carbon source, and additional vitamins and required compounds.

Amino Acids	Trace Minerals	Other Molecules
L-leucine	Cl ⁻ (chloride)	D-glucose
L-threonine	K ⁺ (potassium)	4-aminobenzoate
L-arginine	Ca ²⁺ (calcium)	riboflavin
L-lysine	Mg ²⁺ (magnesium)	thiamine
L-proline	Mn ²⁺ (manganese)	niacin
L-glutamate	Co ²⁺ (cobalt)	meso-2,6-diaminoheptanedioate
L-histidine	Zn ²⁺ (zinc)	O ₂ (oxygen)
L-isoleucine	Cu ²⁺ (copper)	
L-methionine	Fe ²⁺ (iron II)	
L-tryptophane	Na ⁺ (sodium)	
L-valine	Ni ²⁺ (nickel)	
L-cysteine	SO ₄ ²⁻ (sulfate)	
L-phenylalanine	HPO ₄ ²⁻ (phosphate)	

2.3.6. Growth on Different Carbon Sources

Little is known about *D. pigrum*. The previous analysis confirmed several auxotrophies and biosynthetic capacities. To further evaluate the metabolic capabilities, the growth on different carbon sources within the previously defined minimal medium was evaluated. The uptake rate of each tested carbon source was set to 10 mmol/(gDW · h). The available mono-, di-, and trisaccharides were tested as sole carbon sources, as seen in Figure 4. As expected, the growth rate increases with increasing amounts of carbon available. Glucose, fructose, and mannose allow the best growth rates for simulations on monosaccharides as sole carbon sources.

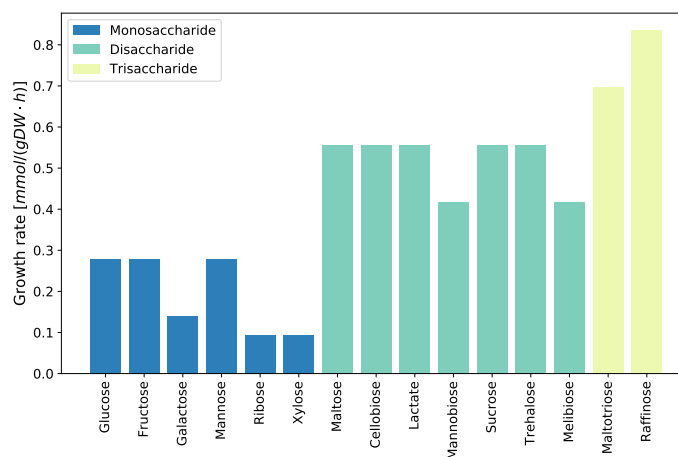


Figure 4. Growth on different carbon sources. *D. pigrum*'s ability to utilize different carbon sources was investigated using the previously defined minimal medium. The available mono-, di-, and trisaccharides were examined concerning the resulting growth rate. As expected, trisaccharides result in a higher growth rate compared to di- and monosaccharides.

2.4. Visualization

A comprehensive map of *D. pigrum*'s metabolism was drawn using Escher [63]. Since a figure would not appropriately capture its large size, the map is included as Supplementary Figure S1 of this publication.

3. Discussion

In this work, we generated *iDPM21RW*: the first genome-scale metabolic model of *Dolosigranulum pigrum*. The basis for the manual extension was the draft reconstruction automatically curated with CarveMe [41].

Models curated by ModelSEED and gapseq could be used to extend the already existing model *iDPM21RW* further. This procedure, however, can be challenging because identifier mapping still holds several difficulties. For this purpose, correct and extensive annotations are indispensable. During curation, we put particular focus on the annotations of reactions, metabolites, and genes. Extensive annotations can hold cross-references to other databases, which facilitates the comparability and interoperability of *iDPM21RW* with models from other databases.

ModelPolisher annotates model instances, such as reactions, metabolites, or compartments but not the genes because they are organism- or even strain-specific. Therefore, the manual addition of gene annotations was required. This was a challenging step because the gene annotations should be strain-specific KEGG identifiers. A direct mapping between the NCBI protein identifiers and the KEGG identifiers was not possible since the NCBI protein identifiers often corresponded to so-called "MULTISPECIES" entries that are not uniquely associated with *D. pigrum*. Reaction or metabolite identifiers are often from different databases, and as already mentioned, mapping is challenging. Strain-specific gene identifiers are, however, sometimes more comfortable to map with the corresponding gene and protein annotation files, and the locus-tag information included. This simplifies model comparisons on gene level.

We added cross-references to several other databases and Systems Biology Ontology (SBO) and Evidence and Conclusion Ontology (ECO) terms. ECO terms [56] provide information about the curator's confidence about a reaction's inclusion into the model. Confidence scores were previously defined by Thiele and Palsson [60] and the Constraint-Based

Reconstruction and Analysis (COBRA) Toolbox. Thiele and Palsson's confidence score 0 indicates the lowest confidence and 4 indicates the highest confidence with biochemical data evidence. To avoid confusion using only numbers, we decided to use ECO terms. These terms are uniquely defined and can directly be accessed via the Minimal Information Requested in the Annotation of Models (MIRIAM) registry initiative at identifiers.org (accessed on 7 April 2021) [74]. Each reaction was assigned one unique ECO term. However, multiple genes can occur within a GPR. We decided to use a conservative approach and to assign the lowest ECO term of all genes to the reaction. One could also think of assigning the highest identified ECO term, but this might require additional manual verification to avoid inducing false confidence.

The biomass objective function (BOF) was improved using the only available omics data, namely genomics. No transcriptomics, proteomics, or lipidomics data are available, which could be used to further improve the BOF and the model itself by adding detected metabolites, reactions, and genes.

Multiple auxotrophies are reported in *D. pigrum*. Brugger et al. predicted that no tricarboxylic acid (TCA) cycle is present [25]. The TCA cycle belongs to the most important central metabolic pathways for energy conservation and biosynthesis of key cellular intermediates, including the amino acid biosynthesis [75]. Thus, it seems not surprising that *D. pigrum* has several auxotrophies, especially for amino acids, resulting from the lacking TCA cycle. *D. pigrum* is not the only microbe missing parts of the TCA cycle. A large number of bacteria are reported to have incomplete or unusual TCA cycles [76,77]. This incompleteness or even absence of the TCA cycle might go back to adaptations to the organism's metabolic lifestyle [76]. However, it might also be the case that apparently "missing" genes are only missing in genome analysis but are revealed in actual biochemical experiments [77]. The observations of Brugger et al. are based on functional genomic prediction, and model curation is based on the genome sequence of *D. pigrum*. Biochemical experiments are required to either confirm the missing TCA cycle or refine the model by adding newly identified reactions.

Further auxotrophies concerned polyamines and vitamins. The polyamines spermidine and putrescine are synthesized from L-arginine and L-methionine in *Escherichia coli* [78], for which *D. pigrum* already harbors auxotrophies. Additionally, *D. pigrum* seems to be auxotrophic for the vitamins thiamine (vitamin B₁), riboflavin (vitamin B₂), and niacin (vitamin B₃). Vitamin B₁ has importance for primary carbohydrate and amino acid metabolism [79]. Our analysis further revealed a 4-aminobenzoate, also called p-Aminobenzoate (PABA), auxotrophy. PABA is a component of folate (vitamin B₉) [80] and, thus, is also associated with the B-vitamins. Rodionov et al. identified transporter proteins for vitamins in various human pathogens, which strictly depend on vitamin uptake [81]. As these transporters are also reported in *D. pigrum*, one could assume that it is also dependent on uptake of the reported B-vitamins. Biochemical experiments are required to confirm all reported auxotrophies.

Having discussed the multiple auxotrophies, it seems apparent that *D. pigrum* has difficulties growing on certain media. The synthetic nasal medium (SNM) and synthetic cystic fibrosis medium (SCFM) mimic two niches, where *D. pigrum* is observed. These habitats, however, are relatively low in nutrient supply. For that reason, metabolites need to be added to the medium definition to enable growth in silico. However, one needs to keep in mind that only single-culture in silico experiments were conducted, combined with in vivo observations. Additional single-culture in vitro growth experiments and coculture in silico experiments might clarify the role of the added metabolites. Sokolovskaya et al. have shown that microbial communities share vitamins. They showed that various mutualisms have evolved between organisms to import and deliver variants of cobamides, including vitamin B₁₂ [82]. It needs to be investigated whether the in silico required nutrients are due to the single culture experiments and are obsolete in multi-culture settings. The same was found for the analysis in the nutrient-rich media simulating the blood and gastrointestinal tract.

Comparing the growth rates between the in silico simulations in the SNM and SCFM to the blood and the gastrointestinal medium, one can observe an increased growth rate for the latter two media. This observation seems reasonable, as the blood and the gastrointestinal medium are rich in nutrients that can be taken up and metabolized compared to the media SNM and SCFM.

With our high-quality model, *iDPM21RW*, we were able to confirm predicted auxotrophies and growth behaviors. Laboratory and biochemical experiments as well as additional omics data can be used to further refine this first-time genome-scale metabolic model of *Dolosigranulum pigrum*. This model will pave the way to better understand its metabolism and its interaction and extrusion of the human pathogen *Staphylococcus aureus*.

4. Materials and Methods

The first draft reconstruction of *Dolosigranulum pigrum* was initially curated using an automated reconstruction tool. Subsequent automated and manual refinement lead to the first genome-scale metabolic model (GEM) of *D. pigrum*.

4.1. Building the Draft Reconstruction

Several tools were used for the draft reconstruction and validation, as explained subsequently.

4.1.1. CarveMe

CarveMe is a fast and automated reconstruction tool for curating genome-scale metabolic models of microbial species and communities [41]. It was used to curate the first draft reconstruction of *D. pigrum* strain 83VPs-KB5. This strain was chosen, as its NCBI assembly level is the only complete genome assembly of *D. pigrum*. Additionally, this strain is the only *D. pigrum* strain in the KEGG database [37]. The coding domain sequence (CDS) of this strain was downloaded from the NCBI assembly database [83], using the accession code ASM19771v1 (RefSeq assembly accession: GCF_007197715.1). With this annotated genome sequence and the default settings of CarveMe version 1.2.2, the initial draft of *D. pigrum* in SBML Level 3 Version 1 format [84] was curated.

4.1.2. ModelPolisher

Subsequently, the ModelPolisher version 2.0.1 was used to annotate the initial draft reconstruction extensively [46]. ModelPolisher matches the identifiers of the model's entities against the BiGG Models database [45]. For each corresponding entry in BiGG, all available information and data about the matched instance are incorporated as annotations into the initial draft reconstruction. ModelPolisher was run within a docker environment using the additional settings `-annotate-with-bigg=true`, `-add-adb-annotations=true`, and `-output-combine=true`. After this annotation step, all gene-protein reaction (GPR) associations, reaction boundaries, and objective coefficients were unreadable by COBRAPy [61] due to inter-conversion difficulties with the SBML flux balance constraints (fbc) package [85]. All unreadable instances were converted to the respective fbc package instances.

4.1.3. MEMOTE

The metabolic model testing suite, MEMOTE determines for each tested GEM an independent and comparable score within a comprehensive overview. Standardized metabolic model tests and the evaluation of a model's annotations constitute the score. Well-annotated and consistent models have a high MEMOTE score [38]. Each improvement step of the *D. pigrum* model was closely monitored by determining the MEMOTE score in each iteration. MEMOTE was used in its command line version.

4.2. Refining the Reconstruction Using Literature Evidence

After the initial draft was curated and annotated, manual refinement steps followed. All manual steps were conducted using COBRAPy [61] and libSBML [86].

4.2.1. Mass and Charge Imbalances

The chemical formula and charge were missing for 65 of the metabolites. They were retrieved from the BiGG Models database [45], added to the respective instance, and used to balance reactions in which they participate.

4.2.2. Add Gene Annotations

The ModelPolisher added annotations for all model instances except for the genes. To this point, only the NCBI protein accession numbers from the CDS file were included in the model. A BLAST [62] search was conducted for every NCBI protein accession number to retrieve the respective GenBank [87] identifiers and to increase the gene annotations' scope. With the help of these GenBank identifiers, the locus tags of the *D. pigrum* genes were identified. These locus tags are also used in the KEGG database [37]. All additionally identified gene annotations were added to the model using libSBML.

4.2.3. Extend Model Manually Using the KEGG Database

Information about *D. pigrum* strain 83VPs KB5 can be found in the KEGG database [37]. The previously retrieved gene annotations were used to compare the already included model genes with the genes listed in the KEGG database to increase the initial reconstruction's scope. Therefore, the identified metabolic reactions, including GPRs, and probable new metabolites, were added to the model. In the next step, dead-end metabolites were identified. Despite an ortholog and homolog search of related nasal microbes available in the BiGG database, the number of dead-end metabolites could not be decreased. Further genes and reactions were added to the model based on these identified metabolites.

4.2.4. Test for Energy-Generating Cycles

GEMs can contain so-called energy-generating cycles. These cycles are thermodynamically impossible since models with such cycles can charge energy metabolites without nutrient consumption [47]. Fritzsche et al. suggested a pipeline to identify 14 different energy metabolites, including adenosine triphosphate (ATP), cytidine triphosphate (CTP), guanosine triphosphate (GTP), uridine triphosphate (UTP), inosine triphosphate (ITP), reduced nicotinamide adenine dinucleotide (NADH), NADPH flavine adenine mononucleotide and dinucleotide, ubiquinol-8, menaquinol-8, 2-demethylmenaquinol 8, acetyl-coA, and L-glutamate as well as the proton exchange between cytosol and periplasm. For each metabolite, a dissipation reaction was defined based on Fritzsche et al. After constraining all uptake reactions to zero, the 15 dissipation reactions were maximized.

4.2.5. Add More Precise SBO Terms

MEMOTE assesses the annotation of model instances with Systems Biology Ontology (SBO) terms [55]. SBO terms provide semantic information about the model instances and allows for explicit and unambiguous understanding of its meaning: the more detailed SBO a term chosen, the more explicit the description given. Metabolites and genes received the general SBO terms for "simple chemical" (SBO:0000247) and "gene" (SBO:0000243), respectively. The reactions' SBO terms were chosen as precisely as possible using an in-house pipeline [57].

4.2.6. Improve Biomass Objective Function

CarveMe adds a universal biomass equation to the carved model. However, this equation was adapted from the biomass composition of *Escherichia coli* [88] to a universal biomass composition [41,89]. To further improve the biomass objective function (BOF) of the *D. pigrum* reconstruction, BOFdat was used [58]. BOFdat is a Python package to generate and improve a BOF based on organism-specific experimental data. In three steps, the stoichiometric coefficients for (i) the major macromolecules, (ii) inorganic ions and coenzymes, and (iii) the remaining species-specific metabolic biomass precursors are generated and incorporated into the BOF. For refinement of the BOF of *D. pigrum*, its genomic DNA

sequence was used as input in the first step. Furthermore, parameters for the dry weight composition are required. Since, at the time of writing, no information about the dry weight composition of *D. pigrum* was available, these parameters were chosen as suggested in the BOFdat documentation. With the DNA sequence and the dry weight composition, the stoichiometric coefficients of the DNA nucleotides deoxyadenosine triphosphate (dATP), deoxythymidine triphosphate (dTTP), deoxyguanosine triphosphate (dGTP), and deoxycytidine triphosphate (dCTP) as well as for diphosphate (ppi) were determined and updated in the BOF. At the time of writing, no transcriptomic, proteomic, or lipidomic data are publicly available. Therefore, the RNA, protein, and lipid macromolecules' coefficients could not be refined within this work.

After determining the stoichiometric coefficients of the macromolecules, the stoichiometric coefficients of the inorganic ions and coenzymes followed. For this step, the BOFdat script was adapted to run in the latest Python version. All inorganic ions or coenzymes were either added to the BOF, or their stoichiometric coefficients were updated.

Experimental gene essentiality data are required for the inclusion and update of additional species-specific metabolic biomass precursors in step (iii). This step aims to identify condition- and species-specific metabolic end goals. As gene essentiality data are also not publicly available at the time of writing, this step was skipped.

4.2.7. Add ECO Terms

The Evidence and Conclusion Ontology (ECO) comprises classes and terms describing different evidence and assertion methods. These terms capture, e.g., the type of evidence that a gene product or a reaction has. ECO terms are helpful for quality control of a model. For every reaction in the model, the GPR association was extracted. All reactions without a GPR were assigned the ECO term ECO:0000001. This term is defined as an inference from background scientific knowledge. For all remaining genes from the GPRs, the UniProt database [49] was consulted. Protein existences were defined as (i) inferred from homology, (ii) predicted, or (iii) evidence at the transcript level. These existences were assigned to their corresponding ECO terms. All assignments are listed in Table 3. If a GPR consists of only one gene, the corresponding ECO term was added to the reaction. If a reaction had a GPR with multiple genes, the gene with the lowest evidence score was added. The ECO terms in Table 3 are sorted from the lowest to the highest evidence scores. Genes that were not found in the UniProt database were assigned the ECO term ECO:0000251 for the similarity evidence used in the automatic assertion. Hence, if one gene in a GPR with multiple genes was not found in UniProt, the reaction was assigned the lowest evidence score, which is the one for genes not found in UniProt.

Table 3. ECO terms and their names and assignments. For every Evidence and Conclusion Ontology (ECO) term, the corresponding name is given together with the assignment. ECO terms are ordered in ascending evidence order.

ECO Term	Term Name	Assignment
ECO:0000001	inference from background scientific knowledge	no GPR
ECO:0000251	similarity evidence used in automatic assertion	GPR but no hit in UniProt
ECO:0000363	computational inference used in automatic assertion	UniProt: 'Predicted'
ECO:0000044	sequence similarity evidence	UniProt: 'Inferred from homology'
ECO:0000009	transcript expression evidence	UniProt: 'Evidence at transcript level'

All ECO terms were added as annotations with the biological qualifier type BQB_IS_DESCRIBED_BY.

4.2.8. Remove Redundant Information

CarveMe stores information about annotations and other databases in the SBML notes field. However, this information is better stored in the annotations field. Since CarveMe and ModelPolisher use the BiGG Models database, the same annotation information is stored twice: once in the notes by CarveMe and once in the annotations field by the ModelPolisher. To avoid this redundancy and to decrease the file size, the annotation information was removed from the notes field.

4.2.9. Add Subsystems and Groups

With the added annotations, the pathways in which a reaction occurs are included in the model. For every reaction that has an annotated KEGG [37] ID, the KEGG representational state transfer (REST) application programming interface (API) was used to retrieve the associated pathways. These pathways were added as further annotations to the reaction with the biological qualifier type BQB_OCCURS_IN. Furthermore, the “groups” plugin [90], available from SBML Level 3 [59,91], was enabled. Every pathway was defined as a group instance, and every reaction occurring in this pathway was added as a member.

4.3. Evaluation and Validation of the Reconstruction

Available knowledge about *D. pigrum* was used and simulated in silico to evaluate and validate iDPM21RW as detailed below.

4.3.1. Evaluating Auxotrophies, Biosynthesis Capabilities, and Carbohydrate Metabolism

We mainly used the results from the functional genomic predictions by Brugger et al. [25] to evaluate the auxotrophies and biosynthetic capabilities. All stated auxotrophies were carefully verified by limiting the respective metabolite’s availability and subsequently optimizing the model. If the in silico simulations revealed no growth after limiting the metabolite’s availability, the auxotrophy was considered confirmed. If growth was possible despite the limitation of its availability, the complete biosynthetic pathway of the respective metabolite was evaluated and carefully checked for every individual reaction. Reactions with limited or insufficient genetic proof were removed from the model. For this evaluation step, we mainly relied on literature research, the two databases KEGG [37] and BioCyc [51], and BLAST searches [62]. For predicted reactions and transporters, the model was checked for the presence of the reported reaction and transporters. Missing reactions or transporters were added to the model with its corresponding genes.

4.3.2. Identification of Additional Auxotrophies

A sink reaction for every amino acid was added to identify additional auxotrophies. This sink reaction was maximized after closing the respective exchange reaction to limit its availability. The growth rate was fixed to 0.2 mmol/(gDW · h). As a medium, the self-defined minimal medium was used (see also Section 4.3.4). If no amino acid production or growth was possible after closing the amino acid’s exchange reaction, *D. pigrum* was considered auxotrophic. If the amino acid could be produced, the amino acid production was set in relation to the sole carbon source (D-glucose). The ATP requirement was calculated by summing up all fluxes of ATP-consuming reactions and by putting them in relation to the amino acid production rate. The CO₂ production rate was computed by setting the CO₂ transport reaction rate in relation to the amino acid production rate.

4.3.3. Evaluating Growth Capabilities in Different Media

The model iDPM21RW was further validated by simulating its growth capabilities in four different environments. The first evaluated habitat was the human nose. For this niche, a chemically defined synthetic nasal medium (SNM) is available [65,72]. As no growth could be simulated with the defined metabolites in the SNM, the identified amino acids for which *D. pigrum* is auxotrophic were added as well. As still no growth was possible,

we further evaluated and identified missing components until a growth in the defined medium could be simulated.

This procedure was repeated for the other three media. The synthetic cystic fibrosis medium (SCFM) mimics the lung of patients with CF and was defined by Palmer et al. [66]. For the blood simulations, an adapted medium initially created for the human reconstruction Recon 2.2 [67] was used. The definition for the European diet was extracted from the Virtual Metabolic Human (VMH) database [69,70]. Each metabolite's exchange reaction and, thus, availability in the analyzed medium was set to 10 mmol/(gDW · h) for determination of the growth rate.

4.3.4. Defining a Minimal Medium for *D. pigrum*

D. pigrum holds many requirements for its environment regarding nutrients due to its multiple auxotrophies. We defined a minimal medium specifically for *D. pigrum* to better cultivate this organism in laboratory settings. For this purpose, we used the SNM medium definition and investigated which metabolites could be removed from the medium while maintaining a realistic growth rate. The uptake rate of each metabolite was set to 10 mmol/(gDW · h). The complete list of minimal medium components is given in Table 2.

4.3.5. Evaluating Growth Capabilities on Different Carbon Sources

With the previously defined minimal medium, the *in silico* growth capabilities of *D. pigrum* on different carbon sources were examined. All available sugar exchange fluxes were extracted from the model and sorted into mono-, di-, and trisaccharides. Each carbon source was tested individually by only enabling the tested carbon source's exchange reaction and by optimizing the model for growth. Growth was also possible for the available polysaccharides, but these were not further investigated.

4.4. Visualization

Escher is a web application for building pathway maps. Reactions, metabolites, and genes can be contextualized within the metabolism of an organism [63]. Besides the web application, an Escher Python package can be run and customized within Jupyter Notebooks [92]. The package can process models using COBRAPy [61]. This Python version of Escher was used to draw parts of *D. pigrum*'s metabolism.

Supplementary Materials: The following are available at <https://www.mdpi.com/2218-1989/11/4/232/s1>, Figure S1: Metabolic map of *D. pigrum*, Table S2: Stoichiometric coefficients of the biomass objective function (BOF), Table S3: Minimal medium for *D. pigrum*.

Author Contributions: Conceptualization, A.R. and A.D.; methodology, A.R. and L.W.; investigation, A.R. and L.W.; software, A.R. and L.W.; visualization, A.R. and L.W.; supervision, A.D.; funding acquisition, A.D.; writing—original draft preparation, A.R.; writing—review and editing, A.R., L.W. and A.D. All authors have read and agreed to the published version of the manuscript.

Funding: This work was funded by the *Deutsche Forschungsgemeinschaft* (DFG, German Research Foundation) under Germany's Excellence Strategy—EXC 2124–390838134—and supported by the German Center for Infection Research (DZIF, doi: 10.13039/100009139) within the *Deutsche Zentren der Gesundheitsforschung* (BMBF-DZG, German Centers for Health Research of the Federal Ministry of Education and Research), grant No. 8020708703. The authors acknowledge support by the Open Access Publishing Fund of the University of Tübingen (<https://uni-tuebingen.de/de/58988>, accessed on 7 April 2021).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The genome-scale metabolic model of *D. pigrum* is available in the BioModels Database [93] as an SBML Level 3 Version 1 file [84] within a COMBINE Archive OMEX file [94] at <https://www.ebi.ac.uk/biomodels/models>, accessed on 7 April 2021 under the accession number MODEL2012220003.

Acknowledgments: We are thankful to Elisabeth Fritze for providing access to the program she designed as part of her bachelor requirements. This program allowed the assignment of hierarchically differentiated SBO terms to our model. We also thank Johannes Zimmermann, Georgios Marinos, and Christoph Kaleta for providing the adapted blood medium.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

ABC	ATP-binding cassette
API	application programming interface
ATP	adenosine triphosphate
BiGG	Biochemically, Genetically, and Genomically structured
BLAST	Basic Local Alignment Search Tool
BOF	biomass objective function
CDS	coding domain sequence
CF	cystic fibrosis
COBRA	Constraint-Based Reconstruction and Analysis
CTP	cytidine triphosphate
dATP	deoxyadenosine triphosphate
dCTP	deoxycytidine triphosphate
dGTP	deoxyguanosine triphosphate
dTTP	deoxythymidine triphosphate
EC	Enzyme Commission
ECF	energy-coupling factor
ECO	Evidence and Conclusion Ontology
FBA	flux balance analysis
fbc	flux balance constraints
FVA	flux variability analysis
GEM	genome-scale metabolic model
GPR	gene–protein reaction
GTP	guanosine triphosphate
HMDB	Human Metabolome Database
ID	identifier
ITP	inosine triphosphate
KEGG	Kyoto Encyclopedia of Genes and Genomes
MIRIAM	Minimal Information Requested in the Annotation of Models
NADH	reduced nicotinamide adenine dinucleotide
NADPH	reduced nicotinamide adenine dinucleotide phosphat
NCBI	National Center for Biotechnology Information
REST	representational state transfer
SBO	Systems Biology Ontology
SCFM	synthetic cystic fibrosis medium
SNM	synthetic nasal medium
URT	upper respiratory tract
UTP	uridine triphosphate
TCA	tricarboxylic acid
VMH	Virtual Metabolic Human

References

1. Lécuyer, H.; Audibert, J.; Bobigny, A.; Eckert, C.; Jannièrre-Nartey, C.; Buu-Hoï, A.; Mainardi, J.L.; Podglajen, I. *Dolosigranulum pigrum* causing nosocomial pneumonia and septicemia. *J. Clin. Microbiol.* **2007**, *45*, 3474–3475. [[CrossRef](#)]
2. Escherich, T. Die Darmbakterien des Neugeborenen und Säuglings. *Fortschr. Med.* **1885**, *3*, 547–554.
3. Aguirre, M.; Morrison, D.; Cookson, B.; Gay, F.; Collins, M. Phenotypic and phylogenetic characterization of some *Gemella*-like organisms from human infections: Description of *Dolosigranulum pigrum* gen. nov., sp. nov. *J. Appl. Bacteriol.* **1993**, *75*, 608–612. [[CrossRef](#)]
4. Hall, G.S.; Gordon, S.; Schroeder, S.; Smith, K.; Anthony, K.; Procop, G.W. Case of synovitis potentially caused by *Dolosigranulum pigrum*. *J. Clin. Microbiol.* **2001**, *39*, 1202–1203. [[CrossRef](#)]

B.2 First GEM of *D. pigrum* confirms multiple auxotrophies

5. Lin, J.C.; Hou, S.J.; Huang, L.U.; Sun, J.R.; Chang, W.K.; Lu, J.J. Acute cholecystitis accompanied by acute pancreatitis potentially caused by *Dolosigranulum pigrum*. *J. Clin. Microbiol.* **2006**, *44*, 2298–2299. [[CrossRef](#)]
6. Hoedemaekers, A.; Schölin, T.; Tonk, B.; Melchers, W.J.; Sturm, P.D. Ventilator-associated pneumonia caused by *Dolosigranulum pigrum*. *J. Clin. Microbiol.* **2006**, *44*, 3461–3462. [[CrossRef](#)]
7. Johnsen, B.O.; Rønning, E.J.; Onken, A.; Figved, W.; Jenum, P.A. *Dolosigranulum pigrum* causing biomaterial-associated arthritis. *APMIS* **2011**, *119*, 85–87. [[CrossRef](#)]
8. LaClaire, L.; Facklam, R. Antimicrobial susceptibility and clinical sources of *Dolosigranulum pigrum* cultures. *Antimicrob. Agents Chemother.* **2000**, *44*, 2001–2003. [[CrossRef](#)]
9. Lappan, R.; S Peacock, C. *Corynebacterium* and *Dolosigranulum*: Future probiotic candidates for upper respiratory tract infections. *Microbiol. Aust.* **2019**, *40*, 172–177. [[CrossRef](#)]
10. Bogaert, D.; Keijsers, B.; Huse, S.; Rossen, J.; Veenhoven, R.; van Gils, E.; Bruin, J.; Montijn, R.; Bonten, M.; Sanders, E. Variability and Diversity of Nasopharyngeal Microbiota in Children: A Metagenomic Analysis. *PLoS ONE* **2011**, *6*, e17035. [[CrossRef](#)]
11. Laufer, A.S.; Metlay, J.P.; Gent, J.F.; Fennie, K.P.; Kong, Y.; Pettigrew, M.M. Microbial communities of the upper respiratory tract and otitis media in children. *mBio* **2011**, *2*, 245–255. [[CrossRef](#)]
12. Bomar, L.; Brugger, S.D.; Yost, B.H.; Davies, S.S.; Lemon, K.P. *Corynebacterium accolens* releases antipneumococcal free fatty acids from human nostril and skin surface triacylglycerols. *mBio* **2016**, *7*. [[CrossRef](#)]
13. Kelly, M.S.; Surette, M.G.; Smieja, M.; Pernica, J.M.; Rossi, L.; Luinstra, K.; Steenhoff, A.P.; Feemster, K.A.; Goldfarb, D.M.; Arscoff-Mills, T.; et al. The Nasopharyngeal Microbiota of Children with Respiratory Infections in Botswana. *Pediatric Infect. Dis. J.* **2017**, *36*, e211–e218. [[CrossRef](#)]
14. Pettigrew, M.M.; Laufer, A.S.; Gent, J.F.; Kong, Y.; Fennie, K.P.; Metlay, J.P. Upper respiratory tract microbial communities, acute otitis media pathogens, and antibiotic use in healthy and sick children. *Appl. Environ. Microbiol.* **2012**, *78*, 6262–6270. [[CrossRef](#)]
15. Lappan, R.; Imbrogno, K.; Sikazwe, C.; Anderson, D.; Mok, D.; Coates, H.; Vijayasekaran, S.; Bumbak, P.; Blyth, C.C.; Jamieson, S.E.; et al. A microbiome case-control study of recurrent acute otitis media identified potentially protective bacterial genera. *BMC Microbiol.* **2018**, *18*, 1–20. [[CrossRef](#)]
16. Biesbroek, G.; Bosch, A.A.; Wang, X.; Keijsers, B.J.; Veenhoven, R.H.; Sanders, E.A.; Bogaert, D. The impact of breastfeeding on nasopharyngeal microbial communities in infants. *Am. J. Respir. Crit. Care Med.* **2014**, *190*, 298–308. [[CrossRef](#)]
17. Biesbroek, G.; Tsvitvadze, E.; Sanders, E.A.; Montijn, R.; Veenhoven, R.H.; Keijsers, B.J.; Bogaert, D. Early respiratory microbiota composition determines bacterial succession patterns and respiratory health in children. *Am. J. Respir. Crit. Care Med.* **2014**, *190*, 1283–1292. [[CrossRef](#)]
18. Bosch, A.A.; De Steenhuijsen Piters, W.A.; Van Houten, M.A.; Chu, M.L.J.; Biesbroek, G.; Kool, J.; Pernet, P.; De Groot, P.K.C.; Eijkemans, M.J.; Keijsers, B.J.; et al. Maturation of the infant respiratory microbiota, environmental drivers, and health consequences. *Am. J. Respir. Crit. Care Med.* **2017**, *196*, 1582–1590. [[CrossRef](#)] [[PubMed](#)]
19. Hasegawa, K.; Linnemann, R.W.; Mansbach, J.M.; Ajami, N.J.; Espinola, J.A.; Petrosino, J.F.; Piedra, P.A.; Stevenson, M.D.; Sullivan, A.F.; Thompson, A.D.; et al. Nasal Airway Microbiota Profile and Severe Bronchiolitis in Infants: A Case-control Study. *Pediatric Infect. Dis. J.* **2017**, *36*, 1044–1051. [[CrossRef](#)] [[PubMed](#)]
20. Prevaes, S.M.; De Winter-De Groot, K.M.; Janssens, H.M.; De Steenhuijsen Piters, W.A.; Tramper-Stranders, G.A.; Wyllie, A.L.; Hasrat, R.; Tiddens, H.A.; Van Westreenen, M.; Van Der Ent, C.K.; et al. Development of the nasopharyngeal microbiota in infants with cystic fibrosis. *Am. J. Respir. Crit. Care Med.* **2016**, *193*, 504–515. [[CrossRef](#)] [[PubMed](#)]
21. Chonmaitree, T.; Jennings, K.; Golovko, G.; Khanipov, K.; Pimenova, M.; Patel, J.A.; McCormick, D.P.; Loeffelholz, M.J.; Fofanov, Y. Nasopharyngeal microbiota in infants and changes during viral upper respiratory tract infection and acute otitis media. *PLoS ONE* **2017**, *12*, e0180630. [[CrossRef](#)]
22. Lopes, S.P.; Ceri, H.; Azevedo, N.F.; Pereira, M.O. Antibiotic resistance of mixed biofilms in cystic fibrosis: Impact of emerging microorganisms on treatment of infection. *Int. J. Antimicrob. Agents* **2012**, *40*, 260–263. [[CrossRef](#)] [[PubMed](#)]
23. Lopes, S.P.; Azevedo, N.F.; Pereira, M.O. Emergent Bacteria in Cystic Fibrosis: In Vitro Biofilm Formation and Resilience under Variable Oxygen Conditions. *BioMed Res. Int.* **2014**, *2014*. [[CrossRef](#)]
24. Lopes, S.P.; Azevedo, N.F.; Pereira, M.O. Developing a model for cystic fibrosis sociomicrobiology based on antibiotic and environmental stress. *Int. J. Med. Microbiol.* **2017**, *307*, 460–470. [[CrossRef](#)]
25. Brugger, S.D.; Eslami, S.M.; Pettigrew, M.M.; Escapa, I.F.; Henke, M.T.; Kong, Y.; Lemon, K.P. *Dolosigranulum pigrum* Cooperation and Competition in Human Nasal Microbiota. *mSphere* **2020**, *5*. [[CrossRef](#)] [[PubMed](#)]
26. Sakr, A.; Brégeon, F.; Mège, J.L.; Rolain, J.M.; Blin, O. *Staphylococcus aureus* nasal colonization: An update on mechanisms, epidemiology, risk factors, and subsequent infections. *Front. Microbiol.* **2018**, *9*, 2419. [[CrossRef](#)] [[PubMed](#)]
27. Schmidt, A.; Bénard, S.; Cyr, S. Hospital Cost of Staphylococcal Infection after Cardiothoracic or Orthopedic Operations in France: A Retrospective Database Analysis. *Surg. Infect.* **2015**, *16*, 428–435. [[CrossRef](#)]
28. Turner, N.A.; Sharma-Kuinkel, B.K.; Maskarinec, S.A.; Eichenberger, E.M.; Shah, P.P.; Carugati, M.; Holland, T.L.; Fowler, V.G. Methicillin-resistant *Staphylococcus aureus*: An overview of basic and clinical research. *Nat. Rev. Microbiol.* **2019**, *17*, 203–218. [[CrossRef](#)] [[PubMed](#)]
29. Liu, C.M.; Price, L.B.; Hungate, B.A.; Abraham, A.G.; Larsen, L.A.; Christensen, K.; Stegger, M.; Skov, R.; Andersen, P.S. *Staphylococcus aureus* and the ecology of the nasal microbiome. *Sci. Adv.* **2015**, *1*. [[CrossRef](#)]



30. Renz, A.; Dräger, A. Curating and Comparing 114 Strain-Specific Genome-Scale Metabolic Models of *Staphylococcus aureus*. *Preprints* **2021**, 2021040244. [[CrossRef](#)]
31. Jansma, J.; El Aidy, S. Understanding the host-microbe interactions using metabolic modeling. *Microbiome* **2021**, *9*, 16. [[CrossRef](#)]
32. Bauer, E.; Thiele, I. From Network Analysis to Functional Metabolic Modeling of the Human Gut Microbiota. *mSystems* **2018**, *3*. [[CrossRef](#)]
33. Diener, C.; Gibbons, S.M.; Resendis-Antonio, O. MICOM: Metagenome-Scale Modeling To Infer Metabolic Interactions in the Gut Microbiota. *mSystems* **2020**, *5*. [[CrossRef](#)]
34. Zomorodi, A.R.; Maranas, C.D. OptCom: A Multi-Level Optimization Framework for the Metabolic Modeling and Analysis of Microbial Communities. *PLoS Comput. Biol.* **2012**, *8*, e1002363. [[CrossRef](#)]
35. Bauer, E.; Zimmermann, J.; Baldini, F.; Thiele, I.; Kaleta, C. BacArena: Individual-based metabolic modeling of heterogeneous microbes in complex communities. *PLoS Comput. Biol.* **2017**, *13*, e1005544. [[CrossRef](#)]
36. Carey, M.A.; Dräger, A.; Beber, M.E.; Papin, J.A.; Yurkovich, J.T. Community standards to facilitate development and address challenges in metabolic modeling. *Mol. Syst. Biol.* **2020**, *16*, e9235. [[CrossRef](#)] [[PubMed](#)]
37. Kanehisa, M.; Sato, Y.; Furumichi, M.; Morishima, K.; Tanabe, M. New approach for understanding genome variations in KEGG. *Nucleic Acids Res.* **2019**, *47*, D590–D595. [[CrossRef](#)]
38. Lieven, C.; Beber, M.E.; Olivier, B.G.; Bergmann, F.T.; Ataman, M.; Babaei, P.; Bartell, J.A.; Blank, L.M.; Chauhan, S.; Correia, K.; et al. MEMOTE for standardized genome-scale metabolic model testing. *Nat. Biotechnol.* **2020**, *38*, 272–276. [[CrossRef](#)]
39. Monk, J.M.; Lloyd, C.J.; Brunk, E.; Mih, N.; Sastry, A.; King, Z.; Takeuchi, R.; Nomura, W.; Zhang, Z.; Mori, H.; et al. iML1515, a knowledgebase that computes *Escherichia coli* traits. *Nat. Biotechnol.* **2017**, *35*, 904–908. [[CrossRef](#)] [[PubMed](#)]
40. Edwards, J.S.; Palsson, B.O. The *Escherichia coli* MG1655 in silico metabolic genotype: Its definition, characteristics, and capabilities. *Proc. Natl. Acad. Sci. USA* **2000**, *97*, 5528–5533. [[CrossRef](#)]
41. Machado, D.; Andrejev, S.; Tramontano, M.; Patil, K.R. Fast automated reconstruction of genome-scale metabolic models for microbial species and communities. *Nucleic Acids Res.* **2018**, *46*, 7542–7553. [[CrossRef](#)]
42. Seaver, S.M.D.; Liu, F.; Zhang, Q.; Jeffryes, J.; Faria, J.P.; Edirisinghe, J.N.; Mundy, M.; Chia, N.; Noor, E.; Beber, M.E.; et al. The ModelSEED Biochemistry Database for the integration of metabolic annotations and the reconstruction, comparison and analysis of metabolic models for plants, fungi and microbes. *Nucleic Acids Res.* **2021**, *49*, D575–D588. [[CrossRef](#)]
43. Zimmermann, J.; Kaleta, C.; Waschina, S. Gapseq: Informed prediction of bacterial metabolic pathways and reconstruction of accurate metabolic models. *Genome Biol.* **2020**. [[CrossRef](#)]
44. Arkin, A.P.; Cottingham, R.W.; Henry, C.S.; Harris, N.L.; Stevens, R.L.; Maslov, S.; Dehal, P.; Ware, D.; Perez, F.; Canon, S.; et al. KBase: The United States department of energy systems biology knowledgebase. *Nat. Biotechnol.* **2018**, *36*, 566–569. [[CrossRef](#)]
45. Norsigian, C.J.; Pusarla, N.; McConn, J.L.; Yurkovich, J.T.; Dräger, A.; Palsson, B.O.; King, Z. BiGG Models 2020: Multi-strain genome-scale models and expansion across the phylogenetic tree. *Nucleic Acids Res.* **2019**. [[CrossRef](#)]
46. Römer, M.; Eichner, J.; Dräger, A.; Wrzodek, C.; Wrzodek, F.; Zell, A. ZBIT Bioinformatics Toolbox: A Web-Platform for Systems Biology and Expression Data Analysis. *PLoS ONE* **2016**, *11*, e0149263. [[CrossRef](#)] [[PubMed](#)]
47. Fritzemeier, C.J.; Hartleb, D.; Szappanos, B.; Papp, B.; Lercher, M.J. Erroneous energy-generating cycles in published genome scale metabolic networks: Identification and removal. *PLoS Comput. Biol.* **2017**, *13*, e1005494. [[CrossRef](#)] [[PubMed](#)]
48. Moretti, S.; Tran, V.D.T.; Mehl, F.; Ibberson, M.; Pagni, M. MetaNetX/MNXref: Unified namespace for metabolites and biochemical reactions in the context of metabolic models. *Nucleic Acids Res.* **2021**, *49*, D570–D574. [[CrossRef](#)]
49. Bateman, A.; Martin, M.J.; O'Donovan, C.; Magrane, M.; Apweiler, R.; Alpi, E.; Antunes, R.; Arganiska, J.; Bely, B.; Bingley, M.; et al. UniProt: A hub for protein information. *Nucleic Acids Res.* **2015**, *43*, D204–D212. [[CrossRef](#)]
50. Morgat, A.; Lombardot, T.; Axelsen, K.B.; Aimo, L.; Niknejad, A.; Hyka-Nouspikel, N.; Coudert, E.; Pozzato, M.; Pagni, M.; Moretti, S.; et al. Updates in Rhea – an expert curated resource of biochemical reactions. *Nucleic Acids Res.* **2017**, *45*, D415–D418. [[CrossRef](#)] [[PubMed](#)]
51. Karp, P.D.; Billington, R.; Caspi, R.; Fulcher, C.A.; Latendresse, M.; Kothari, A.; Keseler, I.M.; Krummenacker, M.; Midford, P.E.; Ong, Q.; et al. The BioCyc collection of microbial genomes and metabolic pathways. *Brief. Bioinform.* **2019**, *20*, 1085–1093. [[CrossRef](#)]
52. Wishart, D.S.; Feunang, Y.D.; Marcu, A.; Guo, A.C.; Liang, K.; Vázquez-Fresno, R.; Sajed, T.; Johnson, D.; Li, C.; Karu, N.; et al. HMDB 4.0: The human metabolome database for 2018. *Nucleic Acids Res.* **2018**, *46*, D608–D617. [[CrossRef](#)] [[PubMed](#)]
53. Sud, M.; Fahy, E.; Cotter, D.; Brown, A.; Dennis, E.A.; Glass, C.K.; Merrill, A.H.; Murphy, R.C.; Raetz, C.R.; Russell, D.W.; et al. LMSD: LIPID MAPS structure database. *Nucleic Acids Res.* **2007**, *35*. [[CrossRef](#)]
54. Geer, L.Y.; Marchler-Bauer, A.; Geer, R.C.; Han, L.; He, J.; He, S.; Liu, C.; Shi, W.; Bryant, S.H. The NCBI BioSystems database. *Nucleic Acids Res.* **2010**, *38*, D492–D496. [[CrossRef](#)] [[PubMed](#)]
55. Courtot, M.; Juty, N.; Knüpfer, C.; Waltemath, D.; Zhukova, A.; Dräger, A.; Dumontier, M.; Finney, A.; Golebiewski, M.; Hastings, J.; et al. Controlled vocabularies and semantics in systems biology. *Mol. Syst. Biol.* **2011**, *7*, 543. [[CrossRef](#)] [[PubMed](#)]
56. Giglio, M.; Tauber, R.; Nadendla, S.; Munro, J.; Olley, D.; Ball, S.; Mitraka, E.; Schriml, L.M.; Gaudet, P.; Hobbs, E.T.; et al. ECO, the Evidence & Conclusion Ontology: Community standard for evidence information. *Nucleic Acids Res.* **2019**, *47*, D1186–D1194. [[CrossRef](#)]
57. Fritze, E. Automating the Assignment of SBO-Terms. Bachelor's Thesis, University of Tübingen, Tübingen, Germany, 2020.

58. Lachance, J.C.; Lloyd, C.J.; Monk, J.M.; Yang, L.; Sastry, A.V.; Seif, Y.; Palsson, B.O.; Rodrigue, S.; Feist, A.M.; King, Z.A.; et al. BOFdat: Generating biomass objective functions for genome-scale metabolic models from experimental data. *PLoS Comput. Biol.* **2019**, *15*, e1006971. [[CrossRef](#)]
59. Keating, S.M.; Waltemath, D.; König, M.; Zhang, F.; Dräger, A.; Chaouiya, C.; Bergmann, F.T.; Finney, A.; Gillespie, C.S.; Helikar, T.; et al. SBML Level 3: An extensible format for the exchange and reuse of biological models. *Mol. Syst. Biol.* **2020**, *16*, e9110. [[CrossRef](#)]
60. Thiele, I.; Palsson, B. A protocol for generating a high-quality genome-scale metabolic reconstruction. *Nat. Protoc.* **2010**, *5*, 93–121. [[CrossRef](#)]
61. Ebrahim, A.; Lerman, J.A.; Palsson, B.O.; Hyduke, D.R. COBRAPy: COnstraints-Based Reconstruction and Analysis for Python. *BMC Syst. Biol.* **2013**, *7*, 74. [[CrossRef](#)]
62. Altschul, S.F.; Gish, W.; Miller, W.; Myers, E.W.; Lipman, D.J. Basic local alignment search tool. *J. Mol. Biol.* **1990**, *215*, 403–410. [[CrossRef](#)]
63. King, Z.A.; Dräger, A.; Ebrahim, A.; Sonnenschein, N.; Lewis, N.E.; Palsson, B.O. Escher: A web application for building, sharing, and embedding data-rich visualizations of biological pathways. *PLoS Comput. Biol.* **2015**, *11*, e1004321. [[CrossRef](#)]
64. Varki, A.; Schauer, R. Chapter 14. Sialic acids. In *Essentials of Glycobiology*, 2nd ed.; Varki, A., Cummings, R.D., Esko, J.D., Freeze, H.H., Stanley, P., Bertozzi, C.R., Hart, G.W., Etzle, M.E., Eds.; Cold Spring Harbor Laboratory Press: Cold Spring Harbor, NY, USA, 2009; Chapter 14.
65. Krismer, B.; Liebecke, M.; Janek, D.; Nega, M.; Rautenberg, M.; Hornig, G.; Unger, C.; Weidenmaier, C.; Lalk, M.; Peschel, A. Nutrient limitation governs *Staphylococcus aureus* metabolism and niche adaptation in the human nose. *PLoS Pathog.* **2014**, *10*, e1003862. [[CrossRef](#)] [[PubMed](#)]
66. Palmer, K.L.; Aye, L.M.; Whiteley, M. Nutritional cues control *Pseudomonas aeruginosa* multicellular behavior in cystic fibrosis sputum. *J. Bacteriol.* **2007**, *189*, 8079–8087. [[CrossRef](#)] [[PubMed](#)]
67. Swainston, N.; Smallbone, K.; Hefzi, H.; Dobson, P.D.; Brewer, J.; Hanscho, M.; Zielinski, D.C.; Ang, K.S.; Gardiner, N.J.; Gutierrez, J.M.; et al. Recon 2.2: From reconstruction to model of human metabolism. *Metabolomics* **2016**, *12*, 109. [[CrossRef](#)] [[PubMed](#)]
68. Bernardes, J.P.; Mishra, N.; Tran, F.; Bahmer, T.; Best, L.; Blase, J.I.; Bordoni, D.; Franzenburg, J.; Geisen, U.; Josephs-Spaulling, J.; et al. Longitudinal Multi-omics Analyses Identify Responses of Megakaryocytes, Erythroid Cells, and Plasmablasts as Hallmarks of Severe COVID-19. *Immunity* **2020**, *53*, 1296–1314.e9. [[CrossRef](#)] [[PubMed](#)]
69. Elmadfa, I. *Österreichischer Ernährungsbericht 2012*; Institut für Ernährungswissenschaften Universität Wien im Auftrag des Bundesministeriums für Gesundheit: Wien, Austria, 2012; p. 424.
70. Magnúsdóttir, S.; Heinken, A.; Kutt, L.; Ravcheev, D.A.; Bauer, E.; Noronha, A.; Greenhalgh, K.; Jäger, C.; Baginska, J.; Wilmes, P.; Fleming, R.M.T.; Thiele, I. Generation of genome-scale metabolic reconstructions for 773 members of the human gut microbiota. *Nat. Biotechnol.* **2016**, *35*, 81–89. [[CrossRef](#)]
71. Lee, H.H.; Ostrov, N.; Wong, B.G.; Gold, M.A.; Khalil, A.S.; Church, G.M. Functional genomics of the rapidly replicating bacterium *Vibrio natriegens* by CRISPRi. *Nat. Microbiol.* **2019**, *4*, 1105–1113. [[CrossRef](#)]
72. Krismer, B.; Weidenmaier, C.; Zipperer, A.; Peschel, A. The commensal lifestyle of *Staphylococcus aureus* and its interactions with the nasal microbiota. *Nat. Rev. Microbiol.* **2017**, *15*, 675–687. [[CrossRef](#)]
73. Brugger, S.D.; Eslami, S.M.; Pettigrew, M.M.; Escapa, I.F.; Henke, M.M.; Kong, Y.; Lemon, K.P. *Dolosigranulum pigrum* cooperation and competition in human nasal microbiota. *bioRxiv* **2019**, 678698. [[CrossRef](#)]
74. Juty, N.; Le Novère, N.; Laibe, C. Identifiers.org and MIRIAM Registry: community resources to provide persistent identification. *Nucleic Acids Res.* **2012**, *40*, D580–D586. [[CrossRef](#)]
75. Kelly, D.J.; Hughes, N.J. The Citric Acid Cycle and Fatty Acid Biosynthesis. In *Helicobacter Pylori*; ASM Press: Washington, DC, USA, 2014; pp. 135–146. [[CrossRef](#)]
76. Huynen, M.A.; Dandekar, T.; Bork, P. Variation and evolution of the citric-acid cycle: A genomic perspective. *Trends Microbiol.* **1999**, *7*, 281–291. [[CrossRef](#)]
77. Cordwell, S.J. Microbial genomes and ‘missing’ enzymes: Redefining biochemical pathways. *Arch. Microbiol.* **1999**, *172*, 269–279. [[CrossRef](#)]
78. Tabor, C.W.; Tabor, H. Polyamines in microorganisms. *Microbiol. Rev.* **1985**, *49*, 81–99. [[CrossRef](#)]
79. Du, Q.; Wang, H.; Xie, J. Thiamin (vitamin B1) biosynthesis and regulation: A rich source of antimicrobial drug targets? *Int. J. Biol. Sci.* **2011**, *7*, 41–52. [[CrossRef](#)]
80. Green, J.M.; Nichols, B.P. p-Aminobenzoate biosynthesis in *Escherichia coli*: Purification of aminodeoxychorismate lyase and cloning of pabC. *J. Biol. Chem.* **1991**, *266*, 12971–12975. [[CrossRef](#)]
81. Rodionov, D.A.; Hebbeln, P.; Eudes, A.; Ter Beek, J.; Rodionova, I.A.; Erkens, G.B.; Slotboom, D.J.; Gelfand, M.S.; Osterman, A.L.; Hanson, A.D.; et al. A novel class of modular transporters for vitamins in prokaryotes. *J. Bacteriol.* **2009**, *91*, 42–51. [[CrossRef](#)] [[PubMed](#)]
82. Sokolovskaya, O.M.; Shelton, A.N.; Taga, M.E. Sharing vitamins: Cobamides unveil microbial interactions. *Science* **2020**, *369*. [[CrossRef](#)]
83. Kitts, P.A.; Church, D.M.; Oise Thibaud-Nissen, F.; Choi, J.; Hem, V.; Sapojnikov, V.; Smith, R.G.; Tatusova, T.; Xiang, C.; Zherikov, A.; et al. Assembly: A resource for assembled genomes at NCBI. *Nucleic Acids Res.* **2015**, *44*, 73–80. [[CrossRef](#)]

84. Hucka, M.; Bergmann, F.T.; Dräger, A.; Hoops, S.; Keating, S.M.; Le Novère, N.; Myers, C.J.; Olivier, B.G.; Sahle, S.; Schaff, J.C.; et al. Systems Biology Markup Language (SBML) Level 3 Version 1 Core. *J. Integr. Bioinform.* **2018**, *15*, 1. [[CrossRef](#)]
85. Olivier, B.G.; Bergmann, F.T. SBML Level 3 Package: Flux Balance Constraints version 2. *J. Integr. Bioinform.* **2018**, *15*, 20170082. [[CrossRef](#)] [[PubMed](#)]
86. Bornstein, B.J.; Keating, S.M.; Jouraku, A.; Hucka, M. LibSBML: An API Library for SBML. *Bioinformatics* **2008**, *24*, 880–881. [[CrossRef](#)]
87. Benson, D.A.; Karsch-Mizrachi, I.; Lipman, D.J.; Ostell, J.; Sayers, E.W. GenBank. *Nucleic Acids Res.* **2011**, *39*, D32–D37. [[CrossRef](#)]
88. Orth, J.D.; Conrad, T.M.; Na, J.; Lerman, J.A.; Nam, H.; Feist, A.M.; Palsson, B. A comprehensive genome-scale reconstruction of *Escherichia coli* metabolism-2011. *Mol. Syst. Biol.* **2011**, *7*, 535. [[CrossRef](#)] [[PubMed](#)]
89. Xavier, J.C.; Patil, K.R.; Rocha, I. Integration of Biomass Formulations of Genome-Scale Metabolic Models with Experimental Data Reveals Universally Essential Cofactors in Prokaryotes. *Metab. Eng.* **2017**, *39*, 200–208. [[CrossRef](#)]
90. Hucka, M.; Smith, L.P. SBML Level 3 package: Groups, Version 1 Release 1. *J. Integr. Bioinform.* **2016**, *13*, 1. [[CrossRef](#)]
91. Renz, A.; Mostolizadeh, R.; Dräger, A. Clinical Applications of Metabolic Models in SBML Format. In *Systems Medicine*; Wolkenhauer, O., Ed.; Academic Press: Oxford, UK, 2020; Volume 3, pp. 362–371. [[CrossRef](#)]
92. Kluyver, T.; Ragan-Kelley, B.; Pérez, F.; Granger, B.; Bussonnier, M.; Frederic, J.; Kelley, K.; Hamrick, J.; Grout, J.; Corlay, S.; et al. Jupyter Notebooks—a publishing format for reproducible computational workflows. In *Positioning and Power in Academic Publishing: Players, Agents and Agendas*; Loizides, F., Schmidt, B., Eds.; IOS Press: Amsterdam, The Netherlands, 2016; pp. 87–90.
93. Malik-Sheriff, R.S.; Glont, M.; Nguyen, T.V.N.; Tiwari, K.; Roberts, M.G.; Xavier, A.; Vu, M.T.; Men, J.; Maire, M.; Kananathan, S.; et al. BioModels—15 years of sharing computational models in life science. *Nucleic Acids Res.* **2020**, *48*, D407–D415. [[CrossRef](#)]
94. Bergmann, F.T.; Adams, R.; Moodie, S.; Cooper, J.; Glont, M.; Golebiewski, M.; Hucka, M.; Laibe, C.; Miller, A.K.; Nickerson, D.P.; et al. COMBINE archive and OMEX format: One file to share all information to reproduce a modeling project. *BMC Bioinform.* **2014**, *15*, 369. [[CrossRef](#)] [[PubMed](#)]

Article

Genome-Scale Metabolic Model of Infection with SARS-CoV-2 Mutants Confirms Guanylate Kinase as Robust Potential Antiviral Target

Alina Renz^{1,2,3} , Lina Widerspich⁴  and Andreas Dräger^{1,2,3,5,*} 

¹ Department of Computer Science, University of Tübingen, 72076 Tübingen, Germany; renz@informatik.uni-tuebingen.de

² Cluster of Excellence 'Controlling Microbes to Fight Infections', University of Tübingen, 72076 Tübingen, Germany

³ Computational Systems Biology of Infections and Antimicrobial-Resistant Pathogens, Institute for Bioinformatics and Medical Informatics (IBMI), University of Tübingen, 72076 Tübingen, Germany

⁴ Bernhard Nocht Institute for Tropical Medicine, Virus Immunology, 20359 Hamburg, Germany; lina.widerspich@bnitm.de

⁵ German Center for Infection Research (DZIF), Partner Site Tübingen, 72076 Tübingen, Germany

* Correspondence: draeger@informatik.uni-tuebingen.de



Citation: Renz, A.; Widerspich, L.; Dräger, A. Genome-Scale Metabolic Model of Infection with SARS-CoV-2 Mutants Confirms Guanylate Kinase as Robust Potential Antiviral Target. *Genes* **2021**, *12*, 796. <https://doi.org/10.3390/genes12060796>

Academic Editors: Giuseppe Novelli and Michela Biancolella

Received: 29 April 2021

Accepted: 21 May 2021

Published: 24 May 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Abstract: The current SARS-CoV-2 pandemic is still threatening humankind. Despite first successes in vaccine development and approval, no antiviral treatment is available for COVID-19 patients. The success is further tarnished by the emergence and spreading of mutation variants of SARS-CoV-2, for which some vaccines have lower efficacy. This highlights the urgent need for antiviral therapies even more. This article describes how the genome-scale metabolic model (GEM) of the host-virus interaction of human alveolar macrophages and SARS-CoV-2 was refined by incorporating the latest information about the virus's structural proteins and the mutant variants B.1.1.7, B.1.351, B.1.28, B.1.427/B.1.429, and B.1.617. We confirmed the initially identified guanylate kinase as a potential antiviral target with this refined model and identified further potential targets from the purine and pyrimidine metabolism. The model was further extended by incorporating the virus' lipid requirements. This opened new perspectives for potential antiviral targets in the altered lipid metabolism. Especially the phosphatidylcholine biosynthesis seems to play a pivotal role in viral replication. The guanylate kinase is even a robust target in all investigated mutation variants currently spreading worldwide. These new insights can guide laboratory experiments for the validation of identified potential antiviral targets. Only the combination of vaccines and antiviral therapies will effectively defeat this ongoing pandemic.

Keywords: SARS-CoV-2; COVID-19; flux balance analysis (FBA); genome-scale metabolic models; target identification; reaction knock-out; structural proteins; purine metabolism; pyrimidine metabolism; B.1.1.7; B.1.351; B.1.617; B.1.28; B.1.427/B.1.429

1. Introduction

Since its emergence in December 2019 [1], individual cases of Severe Acute Respiratory Syndrome (SARS) coronavirus (CoV) type 2 (SARS-CoV-2) infections have evolved into an uncontrolled pandemic. As a result, more than 2.8 million people have lost their lives to or with Coronavirus Disease 19 (COVID-19) by March 2021. COVID-19 symptoms range from pneumonia to severe lung, heart, liver, kidney, neurological or gastrointestinal dysfunction [2]. While great efforts have been employed to provide effective SARS-CoV-2 vaccines [3,4], their success is overshadowed by the emergence of viral escape mutants and the shortcomings in developing targeted antiviral treatments. A meta-analysis by [5] demonstrates that in non-severe cases of COVID-19, there is little to no evidence for effec-

tive use of ribavirin, hydroxychloroquine, umifenovir, lopinavir/ritonavir, or interferon [5]. Even the putative effectiveness of remdesivir is questionable [6,7].

While antiviral medication development was less fruitful, as of March 2021, there are 13 vaccines for SARS-CoV-2 in use, most of them targeting the spike (S) protein [3,8]. Albeit the successes in vaccine development, reports of mutations are increasing. Some of these mutations are even bypassing the immunity provided by several vaccine candidates. Five mutation variants have prevailed, disseminate rapidly, and are classified as variants of concern or variants of interest: (i) B.1.1.7, first detected in the United Kingdom; (ii) P.1 (also called B.1.1.28), first detected in Japan and Brazil; (iii) B.1.351, first detected in South Africa; (iv) B.1.427/B.1.429, first detected in the US [9,10]; and (v) B.1.617, first detected in India [11]. While the consequences of some of these mutations for vaccine efficacy have been reported, the metabolic implications of them remain unclear.

SARS-CoV-2 is a member of the *Betacoronavirus* genus within the *Nidovirales* order [4,12]. The virus has a 27 kb to 32 kb positive sense, single-stranded RNA genome encoding 26 proteins, including the four structural proteins spike (S), envelope (E), membrane (M) and nucleocapsid (N) [4,13]. The S trimers [14,15] scan the host cells surface for the viral entry receptor angiotensin converting enzyme 2 (ACE-2) and therefore initiate the entry process [1,4]. The structural proteins E and M facilitate viral transport, assembly, budding, and release of SARS-CoV-2 virions from infected host cells [1,4]. While N is expressed within the host cytoplasm, the other structural proteins S, E, and M are translated within the endoplasmic reticulum–Golgi intermediate compartment (ERGIC) of the host cell [2,4]. SARS-CoV-2 N supports replicating the viral genome in the cytoplasm and encloses novel viral RNA to form viral ribonucleoprotein complexes (vRNPs) [2]. During the viral replication process's final steps, these cytoplasmic vRNPs are assembled with S, E, and M proteins within the ERGIC [2,4]. The mature virions bud at the ERGIC membrane, forming vesicles which are subsequently released from the host cell via exocytosis [2,4].

Viral lipid envelopes protect the vRNPs and facilitate the particles' entry into host cells [16]. They are usually acquired via budding from the plasma membrane or other cellular organelles [16,17]. Viruses specifically modify host membrane structures, the composition, and the whole host lipid metabolism to favor viral replication [16,18,19]. Many viruses exploit spatiotemporally enriched microdomains or rafts containing different lipid species [19]. To this end, cholesterol, for instance, increases host membrane fluidity for efficient viral entry, replication, and budding, while phosphatidylserine supports viral entry [19]. Altogether, various modifications in viral egress areas determine the differing composition of viral envelopes, thereby influencing their stability and infectivity [19]. As SARS-CoV-2 buds from the ERGIC [2,4], its envelope lipid bilayer resembles this host organelle's composition [2,4]. The viral membrane formation mostly requires cholesterol and phospholipids, while sphingomyelin and cardiolipin are presumably less abundant [20,21].

In our previous work, we have generated an integrated human-virus metabolic model, which combines flux balance analysis (FBA) and flux variability analysis (FVA) to model the metabolic changes within SARS-CoV-2 infected human alveolar macrophages [22]. The GEM is based on the already published and well-developed human alveolar macrophage model iAB-AMØ-1410 by Bordbar et al. [23]. Disabling viral replication in human alveolar macrophages might be an early way of intervention and prevention of the virus's further spread. The model was employed to predict putative antiviral targets such as guanylate kinase 1 (GK1) or the availability of L-isoleucine and L-lysine [22]. Some of these potential targets may be directly targeted by small molecules or antivirals [24–26]. Increasing knowledge of SARS-CoV-2 facilitates the model's improvement by incorporating recent findings of the copy number of the structural proteins [22,27]. The stoichiometric coefficients of the metabolic requirements for amino acids and nucleotides and energy requirements can be refined to predict the viral replication capacity better. Additionally, the lipid requirements were now accounted for in the viral biomass objective function (VBOF). This study presents an updated version of the integrated alveolar macrophage SARS-CoV-2 GEM and the consequences of prominent mutations for predicted metabolic targets.

2. Materials and Methods

2.1. Correcting the Copy Number of Structural Proteins

In the previous version of the VBOF, the copy number of structural proteins was not yet known. We conducted extensive literature research to identify the precise copy number of each structural protein individually. The search was mainly focused on SARS-CoV-2 directly. However, if no information was found for the novel coronavirus, we also searched for information on closely related coronaviruses.

With the identified copy numbers (see Table 2), the stoichiometric coefficients of the nucleotides, amino acids, and energy requirements were re-calculated, as Renz et al. state [22]. However, instead of using a general copy number for all structural proteins, as Aller et al. describe [27], the individual copy numbers of the respective structural proteins were used.

After the VBOF was updated with the corrected stoichiometric coefficients, the knock-out and host-derived enforcement analyses were repeated, as Renz et al. describe [22]. The knock-out experiments were performed by subsequently knocking out each reaction and evaluating its effect on the host's maintenance and viral replication capacity (VBOF). For the host-derived enforcement analyses, the FVA was used to determine flux ranges that allow for 100% maintenance of the host, while decreasing the viral growth by at least 20%. The adapted host-derived enforcement algorithm was used, as Renz et al. describe [22].

2.2. Testing the Targets' Robustness against for Several Mutations

The Global Initiative on Sharing All Influenza Data (GISAID) database has a collection of more than 1.5 million viral sequences of SARS-CoV-2 (May 2021). We set the following filters for the sequences: (i) variant (VUI202012/01 GRY (B.1.1.7) for variant B.1.1.7; GH/501Y.v2 (B.1.351) for variant B.1.351; GR/501Y.V3 (P.1) for variant B.1.1.28; GH/452R.V1 (B.1.429+B.1.427) for variants B.1.429 and B.1.427; and G/452R.V3 (B.1.617+) for variant B.1.617) and (ii) location (Europe/United Kingdom for variant B.1.1.7; Africa for variant B.1.351; South America for variant B.1.1.28; North America/USA for variants B.1.429 and B.1.427; and Asia/India for variant B.1.617). We randomly downloaded ten sequences from each mutation variant with the filters set as described. In addition to the sequences, we downloaded the mutation information given in the metadata. All tested mutations are listed in the Supplementary Table S1. With this information, the stoichiometric coefficients for the VBOF were calculated for every downloaded mutation. As the calculation of the nucleotides' stoichiometric coefficients requires the nucleotide sequence, the downloaded sequences were used directly for this step. For the calculation of the amino acids' coefficients, we used the annotated protein sequence of the SARS-CoV-2 reference sequence (NCBI accession: NC_045512.2) and the mutation information extracted from the metadata files. An algorithm adapted the amino acids from the protein sequence in accordance with the defined mutations, including substitutions, deletions, and introductions of stop codons. With the calculation of the energy requirements and pyrophosphate liberation, all stoichiometric coefficients for the VBOF were available and could be compared. For the first comparison, the mean and standard deviation of all mutations was calculated for each coefficient. These mean values were compared to the wildtype stoichiometric coefficients by calculation the difference. In subsequent analysis, the mean was calculated for the five mutation variants and was then compared to the wildtype. Again, the difference between the coefficients was calculated and visualized. With all generated VBOFs, the reaction knock-out experiments were repeated, as described in the previous section.

2.3. Lipids as Part of the Viral Biomass Objective Function

Literature research was conducted to identify potential fatty acids that occur in the capsid of SARS-CoV-2. As no lipidomics data of SARS-CoV-2 existed at the time of writing, we focused on the five identified lipids phosphatidylcholine, phosphatidylethanolamine, phosphatidylinositol, phosphatidylserine, and cholesterol. The influence of the individual lipids' inclusion into the VBOF on the objective value when optimizing for the VBOF was

evaluated. An overview of the overall procedure for testing the lipids' influence is given in Figure 1.

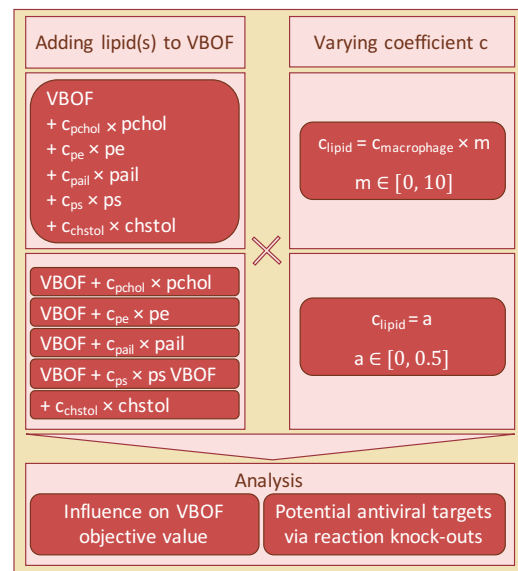


Figure 1. Workflow for the investigation of lipids' influence on the VBOF. The five lipids phosphatidylcholine (pchol), phosphatidylethanolamine (pe), phosphatidylinositol (pail), phosphatidylserine (ps), and cholesterol (chstol) were added together and individually to the VBOF. The stoichiometric coefficients were either an absolute value identical for all lipids, or the initial stoichiometric coefficient from the macrophage biomass function factorized with a multiplication-coefficient. For all scenarios, the influence of the different VBOFs on the objective value was analyzed. Additionally, potential antiviral targets were examined using reaction knock-outs.

As no data were available for the amount of the respective lipids in one virion, we varied the stoichiometric coefficients between 0 and 0.5. The stoichiometric coefficients of the lipids within the macrophage's biomass maintenance function varied from 0.00102 for phosphatidylserine to 0.0315 for phosphatidylcholine (see also Table 1).

Table 1. Stoichiometric coefficients of the five lipids in the macrophage's maintenance function. The stoichiometric coefficients of the five lipids were extracted from the macrophage's maintenance function. Additionally, the BiGG identifiers [28] of the lipids are given. These stoichiometric coefficients formed the starting point for evaluating the lipids' influence on the viral biomass objective function (VBOF).

Lipid	BiGG ID	Coefficient
Phosphatidylcholine	pchol_hs_c	0.03152
Phosphatidylethanolamine	pe_hs_c	0.02110
Phosphatidylinositol	pail_hs_c	0.00374
Phosphatidylserine	ps_hs_c	0.00102
Cholesterol	chsterol_c	0.02093

With the variation of the stoichiometric coefficients between 0 and 0.5, we covered the 14 to 490-fold increase of the stoichiometric coefficients, depending on their initial value. In

the next step, all lipids were added simultaneously to the VBOF. We evaluated the VBOF's objective value using both the lipids' stoichiometric coefficients from the macrophage's maintenance function and their ten-fold value.

To evaluate the effect of the lipids' inclusion on the potential antiviral targets, we again used the stoichiometric coefficients of the macrophage's maintenance function and a multiplication coefficient, ranging from 0 to 10 as the actual coefficient of the lipids is unknown. We conducted the knock-out experiments as Renz et al. describe [22] for each tested coefficient by knocking out each reaction individually and analyzing its effect on both the viral growth and the host's maintenance function. While varying the multiplication coefficient, two additional reactions occurred, whose knock-out decreased the viral growth rate.

To investigate, which lipid influences the knock-out experiments most, we again analyzed the lipids individually. As done for the effect on the VBOF's objective value, we first varied the stoichiometric coefficients between 0 and 0.5. Subsequently, we used a multiplication coefficient ranging from 0 to 10, which was multiplied with the coefficient of the macrophage's maintenance function (see Table 1).

3. Results

3.1. Correcting the Copy Number of Structural Proteins

The single-stranded RNA genome of SARS-CoV-2 has 26 proteins [13], including four structural proteins. These four structural proteins need to be produced by the host in higher amounts than the non-structural proteins. However, the actual number of copies of each structural protein was unknown when the novel coronavirus arose, and the first studies were conducted at the beginning of the year 2020.

After extensive literature research, we collected the latest information about the copy number of the structural proteins of SARS-CoV-2. [15] identify on average 40 copies of the trimeric spike (S) protein on the surface of SARS-CoV-2, resulting in 120 copies of the S protein. [14] estimate the number of S trimers per virion to be 48, resulting in a similar copy number range as [15]. Since [15] use in situ structural analysis and [14] use mathematical estimations, we chose to use a copy number of 120 S proteins for further analysis (see Table 2). The number of the envelope (E) proteins is approximated to 20 copies [29] based on analyses of the OC43 human coronavirus (hCoV) [30] and the transmissible gastroenteritis virus (TGEV) [31]. Exactly like SARS-CoV-2, both viruses belong to the family of *Coronaviridae*, and hCoV also belongs to the same genus *Betacoronavirus* as SARS-CoV-2. Currently, no numbers for the E protein are available for SARS-CoV-2. For that reason, the number is approximated from related coronaviruses. The nucleocapsid (N) packs the viral RNA in so-called vRNPs. [14] observe 38 vRNPs per SARS-CoV-2 virion [14]. Approximately 12 copies of the N protein are located in one vRNP in SARS-CoV-2 [32,33]. Multiplying those two numbers results in 456 copies of the N protein. The amount of membrane proteins is not yet determined for SARS-CoV-2. [13] provide key numbers about SARS-CoV-2, including the copy numbers of the S, M, N, and E protein. However, all copy numbers are derived from SARS-CoV-1 or TGEV. We found precise numbers for the copy number of N proteins in SARS-CoV-2, and [34] determine the estimated ratios of M to N proteins ranging from 3M:1N to 1M:1N with 730 to 2200 N proteins per virion [34]. With this information at hand, we estimated the copy number of M proteins to 1000 by doubling the number of N proteins and rounding them up. The ratio of 2M:1N was chosen based on the article of [13], where the number of N proteins is stated as 1000 copies for SARS-CoV-1 and the number of M proteins as 2000. All used copy numbers are listed in Table 2.

With the updated copy numbers, the stoichiometric coefficients of the nucleotides, amino acids, and energy requirements were re-calculated for the viral biomass objective function (VBOF) of SARS-CoV-2. The subsequent analyses for identifying potential antiviral targets consisted of knock-out and host-derived enforcement experiments, as Renz et al. describe [22]. The guanylate kinase 1 (GK1) remains a promising antiviral target after the adaptations of the copy number of structural proteins based on the knock-out experiments.

Table 2. Copy number of structural proteins. The Copy number of structural proteins (Csp) was determined based on extensive literature research. Besides the reference and the copy number of structural proteins, the investigated organism is given as a source.

Protein	Name	Reference	Source	Csp
S	spike	[15]	SARS-CoV-2	120
E	envelope	[29]	hCOV, TGEV	20
N	nucleocapsid	[14,32]	SARS-CoV-2	456
M	membrane	[34]	SARS-CoV-1	1000

The results of the host-derived enforcement analyses were dependent on the Copy number of structural proteins [22]. As we now identified more precise copy numbers, we can also determine the host-derived enforcement analysis results more precisely. In total, 21 reactions were identified, whose inhibition decreases the viral replication capacity by at least 20% without harming the host's maintenance (100%). These reactions, their inhibition range, and the reduction of the VBOF are visualized in Figure 2. Reactions could be inhibited between 72% and 89%. As seen in the knock-out experiments, the guanylate kinase 1 (GK1) is the only reaction where a complete inhibition (100%) is possible. The ribose-5-phosphate isomerase (RPI) and phosphoribosylpyrophosphate synthetase (PRPPS) are part of the pentose phosphate pathway. glutamine phosphoribosyl-diphosphate amidotransferase (GLUPRT), phosphoribosylglycinamide synthase (PRAGS_r), phosphoribosylglycinamide formyltransferase (GARFT), phosphoribosyl-formylglycinamide synthase (PRFGS), phosphoribosylaminoimidazole synthase (PRAIS), Phosphoribosylaminoimidazole carboxylase (AIRCr), phosphoribosylaminoimidazole-succinocarboxamide synthase (PRASCS), phosphoribosylaminoimidazolecarboxamide formyltransferase (AICART), and inosine monophosphate (IMP) cyclohydrolase (IMPC) are involved in the purines' biosynthetic pathway, more precisely in the biosynthesis of IMP [35]. Reactions associated with the purine adenosine monophosphate (AMP) biosynthesis were also identified as potential targets, namely adenylosuccinate synthase (ADSS), and adenylosuccinate lyase 1 and 2 (ADSL1, ADSL2) [35].

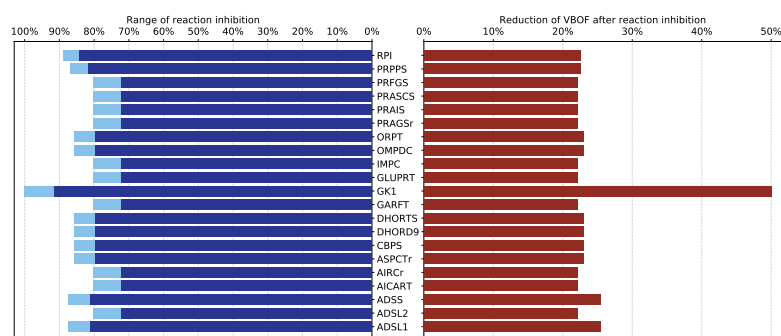


Figure 2. Results of the host-derived enforcement experiments. With the help of the host-derived enforcement, the range and effect of reaction inhibitions on the VBOF can be investigated while keeping the host's maintenance at 100%. The minimum possible reaction inhibition rate to reduce the viral replication capacity (VBOF) is given in dark blue. The maximum inhibition of the reaction does not harm the host's maintenance and is indicated in light blue. The reduction of the VBOF is given in comparison to the un-inhibited state. All reaction identifiers are BiGG identifiers [28]. Table A1 lists all reaction identifiers with their corresponding reaction name and the subsystem they occur in.

Besides the reactions associated with the purine metabolism, the host-derived enforcement analysis also reported reactions from the pyrimidine biosynthesis, such as the carbamoyl-phosphate synthase (CBPS), aspartate carbamoyltransferase (ASPCTr), dihydroorotase (DHORTS), dihydroorotic acid dehydrogenase (DHORD9), orotate phosphoribosyltransferase (ORPT), and orotidine-5'-phosphate decarboxylase (OMPDC) [36].

3.2. Testing the Targets' Robustness for Several Mutations

3.2.1. Analysis of Mutant-Specific Variations in the Viral Biomass

Novel mutations of SARS-CoV-2 emerge on a daily basis. Five mutation variants have prevailed, disseminate rapidly, and are classified as variants of concern or variants of interest: (i) B.1.1.7, (ii) P.1 (also called B.1.1.28), (iii) B.1.351, (iv) B.1.427/B.1.429 [9,10], and (v) B.1.617 [11]. The GISAID was launched in 2008 to promote the international sharing of virus data [37,38]. When the novel coronavirus emerged, GISAID was expanded by a database for sharing sequenced viral genomes of SARS-CoV-2 globally. At the time of writing, more than 1.5 million viral sequences of SARS-CoV-2 are collected in the database. To investigate the mutations' effect on the previously identified potential antiviral targets, sequences of each mutation variant were downloaded from GISAID and analyzed. The stoichiometric coefficients of each variant were calculated as Renz et al. describe [22]: For the calculation of the nucleotides' stoichiometric coefficients, the downloaded RNA sequence was used. The amino acids' stoichiometric coefficients were calculated using the provided information about the identified mutations and the reference (wildtype) protein sequence of the first sequenced SARS-CoV-2. With this information, the abundance of the different amino acids in the different proteins was adapted for each mutation variant. The nucleotide and amino acid counts were subsequently used to calculate the pyrophosphate liberation and the adenosine triphosphates (ATPs) requirements. For each downloaded mutation variant, an individualized VBOF was created with the calculated stoichiometric coefficients.

To assess the mutations' effect on the VBOF's stoichiometric coefficients, we first calculated the mean and standard deviation from all stoichiometric coefficients for all mutations and compared them to the wildtype (WT) coefficients. The mean stoichiometric coefficients of the mutations are very similar to the wildtype's stoichiometric coefficients. The largest difference is observed for the amino acid L-aspartate: The stoichiometric coefficient for L-aspartate is decreased by on average 0.005 in the mutations compared to the wildtype. Figure 3 visualizes the comparison of the mutations' mean stoichiometric coefficients with the wildtype coefficients.

Since we analyzed five distinct mutation variants, the differences in the stoichiometric coefficients were examined further based on these variants. The mean for each stoichiometric coefficient was calculated variant-wise. With this mean, the deviation from the wildtype coefficient was calculated and visualized as a heat-map in Figure 4. This analysis gives further insight into the properties of the individual mutations.

One can observe a pattern for the stoichiometric coefficients of adenosine diphosphate (ADP) and ATP: While the mutation variants B.1.1.7 and B.1.1.28 have decreased stoichiometric coefficients (-0.01) compared to the wildtype, the variants B.1.351 and B.1.427/429 have increased stoichiometric coefficients (0.019 to 0.021). This pattern is most apparent for ADP and ATP, but can also be observed for other stoichiometric coefficients, such as for diphosphate (PPi), L-lysine, L-threonine, or L-valine. To further investigate this pattern, we examined the calculation for the stoichiometric coefficients. Each coefficient is set in relation to the total viral molar mass (M_v), which is the sum of the total molar mass of all nucleotides (G_i) and amino acids (G_j). The mutation variants B.1.1.7 and B.1.1.28 have a higher total viral molar mass compared to the mutation variants B.1.351 and B.1.427/429. This increased total viral molar mass is based on an increased molar mass of both nucleotides (G_i) and amino acids (G_j). As the stoichiometric coefficients for ADP and ATP larger than the other coefficients, this pattern is more apparent.

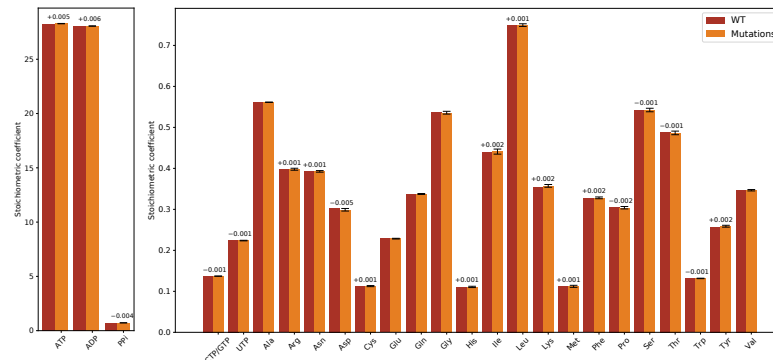


Figure 3. Difference of stoichiometric coefficients between wildtype (WT) and all mutations. The stoichiometric coefficients for all metabolites participating in the viral biomass objective function (VBOF) are compared. WT stoichiometric coefficients are indicated in red, the mean stoichiometric coefficients of all mutation variants are indicated in orange, including standard deviations (black). If the difference of the stoichiometric coefficients between WT and mutation variants was more than 0.001, the difference is indicated above the bars. The stoichiometric coefficients for the metabolites ATP, ADP and PPI, are higher mean compared to the other coefficients. The mutation variants' mean coefficients show little deviation. Additionally, the differences between the stoichiometric coefficients of WT and mutation variants are very small.

However, this pattern does not emerge in all stoichiometric coefficients. There are deviations for, e.g., L-serine. Only the mutation variant B.1.1.7 shows a decreased stoichiometric coefficient compared to the wildtype. We analyzed the documented mutations for this variant and identified two mutations in structural proteins, Spike S982A and N S235F, which only occur in this variant. In both cases, the amino acid L-serine is substituted by another amino acid. As both mutations occur in structural proteins with copy numbers of 120 and 456, respectively, their influence on the amount of amino acid and, thus, the stoichiometric coefficient, is noticeable. Compared to the other mutation variants, variant B.1.1.28 has the highest increase in the stoichiometric coefficient for L-serine. This could be explained by two mutations specific for this variant in the structural spike protein: Spike P26S and Spike R190S. In both cases, other amino acids are replaced by L-serine. As explained for the mutation variant B.1.1.7, the spike protein has a copy number of 120. Changes in these structural proteins can be measurable and influence the stoichiometric coefficient stronger than mutations in non-structural proteins.

The mutation variant B.1.617 does not fit in this pattern. As the variants B.1.315 and B.1.427/429, its stoichiometric coefficients for ADP and ATP are increased, but not as much. Variant B.1.617 has a similar total viral molar mass as B.1.1.7. However, the amount of the nucleotides adenine and uridine is more similar to the variants B.1.315 and B.1.427/429. Having similarities with both pattern groups, variant B.1.617 does fit in neither of the groups. Variant B.1.617 needs less L-isoleucine compared to the wildtype.

3.2.2. Analysis of the Effects of Single Gene Deletions

After highlighting the differences in the stoichiometric coefficients for the different mutation variants, we tested the robustness of our previously identified potential antiviral targets [22]. To do so, we repeated the single-gene-deletion experiments for every mutation variant. Our analysis revealed that in all mutation variants, the guanylate kinase 1 (GK1) is a robust potential antiviral target.

B.3 GEM infected with SARS-CoV-2 mutants confirms guanylate kinase as robust target

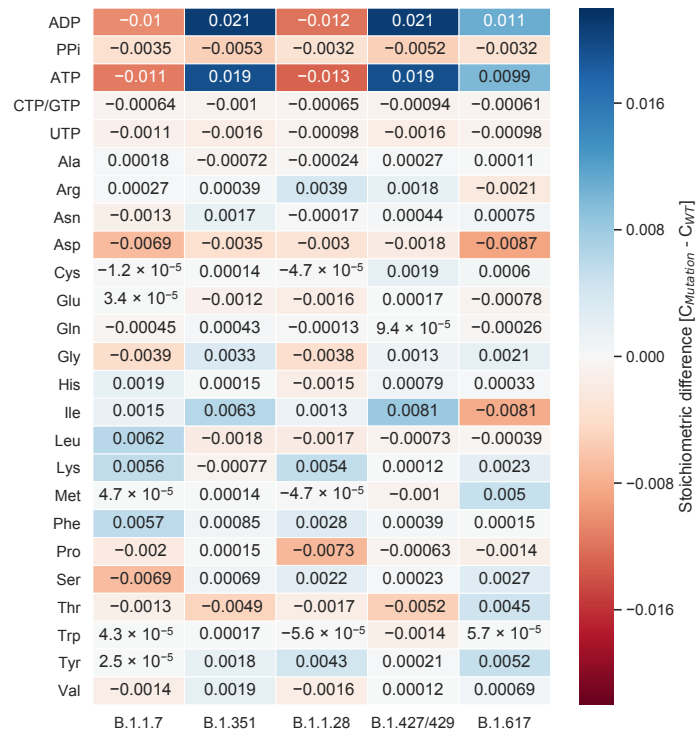


Figure 4. Difference of stoichiometric coefficients between wildtype (WT) and the individual mutations. The deviation between WT and the mean of the individual mutation variants was calculated. Higher stoichiometric coefficients in the mutation compared to the WT are indicated in blue, while lower stoichiometric coefficients are indicated in red. Based on similar sequence length for the mutation variants B.1.1.7 and B.1.1.28 and resulting similar total viral molar masses, a pattern emerges, which is most apparent for the stoichiometric coefficients of ATP and ADP. This pattern, however, is not present for all stoichiometric coefficients. The coefficient for L-serine, for example, is only decreased in the mutation variant B.1.1.7 based on two mutations in two structural proteins. Overall, the deviations from the WT are very small.

3.3. Lipids as Part of the Viral Biomass Objective Function

The transmembrane domain of the envelope (E) protein is located in lipid bilayers mimicking the ERGIC membrane [21]. Ref. [20] described this ERGIC membrane [20] in 1994. The four phospholipids, phosphatidylcholine, phosphatidylethanolamine, phosphatidylinositol, and phosphatidylserine, were observed in the ERGIC while sphingomyelin and cardiolipin were not present [20]. Ref. [21] use an ERGIC-mimetic consisting of the four described phospholipids and cholesterol to investigate the E-protein’s transmembrane domain [21]. The five lipids are also participating in the macrophage’s maintenance function. Thus, their role and influence on the VBOF and antiviral targets were examined.

As the actual amount of lipids in the SARS-CoV-2 virion is not yet determined, we evaluated varying stoichiometric coefficients. In the first experiments, the individual lipids’ effect on the VBOF’s objective value was analyzed. The objective coefficients from the macrophage’s maintenance function varied between 0.001 for phosphatidylserine and 0.031 for phosphatidylcholine. Therefore, we first varied all lipids’ coefficients between 0 and 0.5 and subsequently used a multiplication coefficient between 0 and 10 to multiply

the macrophage's coefficients. Despite an up to 490-fold increase of the stoichiometric coefficient (for phosphatidylserine) compared to its initial value in the macrophage's maintenance function, the VBOF's objective value remained at $0.01886 \text{ mmol}/(\text{gDW} \cdot \text{h})$. This was also the case when all five lipids were added to the VBOF simultaneously.

Knock-out experiments were conducted to identify additional potential antiviral targets. All lipids were included in the VBOF, and the coefficients were varied using a multiplication coefficient. At the five-fold increase of the initial stoichiometric coefficients, two novel reactions emerged as new potential antiviral targets: the methionine synthase (METS) and the 5,10-methylenetetrahydrofolate reductase (FADH_2) (MTHFR). To identify, which lipids are responsible for the emergence of the novel antiviral target, we repeated the described analysis for every lipid individually, once using absolute stoichiometric coefficients ranging from 0 to 0.5 and once using the above-described multiplication coefficient ranging between 0 and 10. By this approach, we identified phosphatidylcholine to be the responsible lipid for the additional antiviral targets. When increasing the initial macrophage's stoichiometric coefficient of phosphatidylcholine by at least 4.76, the two enzymes emerge as potential antiviral targets. At a five-fold increase of phosphatidylcholine and the knock-out of either the methionine synthase or the 5,10-methylenetetrahydrofolate reductase (FADH_2), the viral growth can be inhibited by approximately 1.5%. With increasing amounts of phosphatidylcholine in the VBOF, the knock-out influence of the two reactions on the viral growth increases, as seen in Figure 5: at an eleven-fold increase of phosphatidylcholine, the viral growth rate is decreased by approximately 50%. A twenty-fold increase of phosphatidylcholine inhibits the viral growth even to 30% of its initial growth rate.

It needs to be highlighted that the guanylate kinase 1 (GK1) was a potential antiviral target during all conducted *in silico* experiments evaluating the lipids' effect on potential targets.

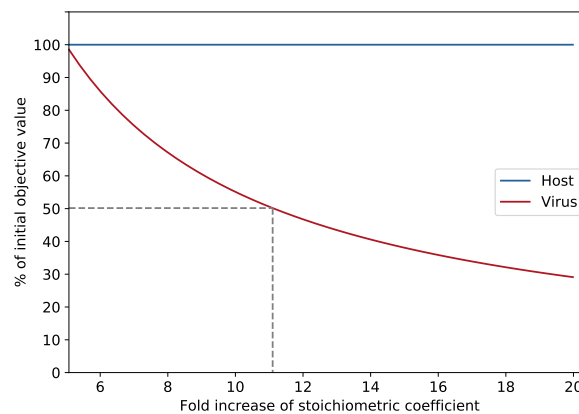


Figure 5. Influence of stoichiometric coefficient on reduction of VBOF during METS knock-out. With increasing factorization of phosphatidylcholine's stoichiometric coefficient, the objective value of the VBOF's optimization decreases during the knock-out of the methionine synthase (METS) reaction. The hosts growth maintenance stays at 100%. At an eleven-fold increase of the initial stoichiometric coefficient extracted from the host's maintenance function results in a 50% decrease of the viral growth rate.

4. Discussion

This study presents an updated viral biomass objective function (VBOF) for the novel coronavirus SARS-CoV-2 based on the latest information of its structural proteins. This VBOF was integrated into an already validated model of human alveolar macrophages [23].

B.3 GEM infected with SARS-CoV-2 mutants confirms guanylate kinase as robust target

The tissue tropism of SARS-CoV-2 comprises most cell types expressing the entry receptor ACE-2, mainly including cell types of the lung, liver, stomach, ileum, kidney, and colon [39,40]. Although SARS-CoV-2 enters the host via the airways, the expression of ACE-2 is comparably low, highlighting the role of possible co-receptors [40]. Nonetheless, human alveolar type 2 cells robustly express ACE-2, while alveolar macrophages possibly express low levels of the entry receptor [40]. It is known that different coronaviruses infect macrophages, such as the human coronavirus strain 229E [41], the Middle East Respiratory Syndrome (MERS) coronavirus [42], and the SARS coronavirus [43]. Also, the novel coronavirus SARS-CoV-2 is reported to infect alveolar macrophages [44]. However, other *in vitro* studies suggest that challenging alveolar macrophages with SARS-CoV-2 does not lead to a productive infection [45]. However, even without productive infection, alveolar macrophages could serve as Trojan horses, which enable viral anchoring within pulmonary parenchyma [39]. Ref. [45] demonstrate that the tissue-resident alveolar macrophages play a crucial role in SARS-CoV-2 immune evasion [44,45] and are hypothesized to support viral pathogenesis [39]. Disabling viral replication in human alveolar macrophages might be an early way of intervention and prevention of the virus's further spread.

We corrected the copy number of structural proteins and the stoichiometric coefficients in the viral biomass objective function (VBOF). The amount of the spike (S) and nucleocapsid (N) proteins were derived from studies on SARS-CoV-2 [14,15,32]. The copy number of the envelope (E) protein is derived from the human coronavirus and the transmissible gastroenteritis virus [29]. Numbers for SARS-CoV-2 are currently not available. Same accounts for the copy number of membrane (M) proteins, where information is only available for SARS-CoV-1 [34]. Especially for the M proteins, a range of potential copy numbers exists, as the ratio of M and N proteins ranges from 3M:1N to 1M:1N [34]. With the N protein's copy number of 456, the M protein's copy number ranges from 456 to 1368. As soon as additional information on the copy numbers of the E and M protein is available for SARS-CoV-2, the stoichiometric coefficients can be refined further.

However, the current refinement still confirmed the guanylate kinase 1 (GK1) as a potential antiviral target. Even for the investigated mutations, the guanylate kinase seems to be a robust target in human alveolar macrophages to interrupt SARS-CoV-2 replication. Ref. [46] conduct a similar study with the human reconstruction RECON2.2 [47] containing a lung biomass objective function and a viral biomass objective function [46]. They also report the guanylate kinase as a potential target for antiviral therapies [46]. In our previous study, we suggested potential drugs that could be repurposed to fight this SARS-CoV-2 pandemic. Amongst these drugs were cidofovir, brincidofovir, and favipiravir [22]. A virtual screening method identified cidofovir as a potentially effective therapeutic against SARS-CoV-2 [48]. A molecular docking study suggests the repurposing of brincidofovir against SARS-CoV-2 [49]. For favipiravir, several clinical trials are listed in the ClinicalTrials database hosted by the U.S. National Library of Medicine [50], running in several countries, including Italy (NCT04336904), Turkey (NCT04474457), and the United States (NCT04358549). However, these therapeutics are only analogs and do not directly inhibit the guanylate kinase. No direct inhibitor of the guanylate kinase is tested for its antiviral effect on SARS-CoV-2 infections at the time of writing. As the guanylate kinase is a robust target for all currently occurring mutation variants, further investigations could be of high interest to fight this pandemic.

Besides the guanylate kinase, additional potential antiviral targets were identified using the host-derived enforcement analysis. These antiviral targets are located in the pentose phosphate pathway, the purine, and the pyrimidine metabolism. It is shown that the pentose phosphate pathway is remarkably deregulated during SARS-CoV-2 replication, which shows potential implications for antiviral therapies [51]. The purine biosynthesis pathway is enhanced upon SARS-CoV-2 infection to support the *de novo* synthesis of purines [52]. First *in vitro* experiments show that the FDA-approved inhibitor of purine biosynthesis methotrexate potentially inhibits viral replication [53,54], protein synthesis, and release [53]. The pyrimidine metabolism is also reported as a potential antiviral target,

especially the dihydroorotate dehydrogenase. Its inhibition by, for example, brequinar or leflunomide is already demonstrated to have antiviral activity against other viruses [55–57], such as rotavirus [58] and Ebola virus [59]. The dihydroorotate dehydrogenase inhibitor PTC299 is shown to arrest SARS-CoV-2 replication *in vitro* [60]. The dihydroorotate dehydrogenase inhibitors S312 and S416 are validated to have high antiviral efficacy *in vivo* [61]. To conclude, our identified antiviral targets are currently under discussion in the scientific community, and for some, the influence and relevance for viral replication are confirmed.

Analyses of the documented mutations revealed that virus variant B.1.617 needs less L-isoleucine compared to the wildtype because of a mutation in the membrane protein, M182T, where L-isoleucine is substituted by L-threonine. As the membrane protein has a copy number of 1000, its replacement could influence the stoichiometric coefficient of the replaced amino acid. Same accounts for the mutation N4402H in the nucleocapsid protein, where L-aspartate is replaced by L-histidine, which might explain the decreased stoichiometric coefficient for L-aspartate. Changes in these structural proteins can be measurable and influence the stoichiometric coefficient stronger than mutations in non-structural proteins.

Alongside the mutation variants that could complicate the fight against SARS-CoV-2 with vaccines, the S protein's glycosylation could impact antibodies' ability to bind to a pathogenic S glycoprotein by shielding its surface [62,63]. Currently, this glycosylation process is not reflected in the VBOF or the model. As soon as more information about the glycosylation is available that can be used to determine a range or precise stoichiometric coefficients, the glycosylation of the spike protein can be incorporated into the model simulations.

The inclusion of lipids in the VBOF opens new perspectives for potential antiviral targets. It is shown that virus infections can dramatically impact on lipid metabolism [64–67]. Upon rhinovirus infection multiple lipid pathways are altered, and changes in phospholipids, lysophospholipids, fatty acids, and inositol phospholipids are observed [66]. For the human coronavirus 229E (hCoV-229E), the host cell lipid response upon infection was comprehensively characterized. Glycerophospholipids and fatty acids were significantly elevated. Lysophosphatidylcholine, which is hydrolyzed from phosphatidylcholine, was significantly elevated and accounted for approximately 60% of all identified lipids with significant elevation [68]. Our study also highlighted phosphatidylcholine as an essential lipid upon SARS-CoV-2 infection, confirming the findings from [68] for hCoV-229E. As metabolic alterations harbor potential antiviral targets, regulating or targeting the lipid metabolism is suggested and discussed [64,66,68]. We identified two novel potential antiviral targets connected with lipid metabolism: the methionine synthase and the 5,10-methylenetetrahydrofolate reductase (FADH₂). S-adenosyl-L-methionine is a pivotal methyl donor in the synthesis of phosphatidylcholine [69,70]. Thus, the synthesis of L-methionine by the 5,10-methylenetetrahydrofolate reductase (FADH₂) and methionine synthase seem to be an antiviral target to disrupt the synthesis of phosphatidylcholine. These novel insights could guide further laboratory experiments for investigating and validating the lipid's role in SARS-CoV-2 infections.

This study confirmed the guanylate kinase 1 (GK1) as a robust antiviral target against SARS-CoV-2 and its arising mutation variants. With the refined copy numbers of structural proteins, the list of further potential antiviral targets was improved, and some targets are already under discussion or even under validation. The inclusion of the lipids into the VBOF opened new perspectives for additional metabolic targets to fight against this pandemic.

Supplementary Materials: The following are available online at <https://www.mdpi.com/2073-4425/12/6/796/s1>, Table S1: List of tested variants. The tested variants are listed together with their GISAID accession number and the observed mutations. The stoichiometric coefficients for the VBOF's compounds in the different virus mutations are listed.

B.3 GEM infected with SARS-CoV-2 mutants confirms guanylate kinase as robust target

Author Contributions: Conceptualization, A.R. and A.D.; methodology, A.R. and L.W.; investigation, A.R. and L.W.; software, A.R.; visualization, A.R.; supervision, A.D.; funding acquisition, A.D.; writing—original draft preparation, A.R. and L.W.; writing—review and editing, A.R., L.W. and A.D.; All authors have read and agreed to the published version of the manuscript.

Funding: This work was funded by the Federal Ministry of Education and Research (BMBF) and the Baden-Württemberg Ministry of Science as part of the Excellence Strategy of the German Federal and State Governments, by the *Deutsche Forschungsgemeinschaft* (DFG, German Research Foundation) under Germany's Excellence Strategy—EXC 2124—390838134, and supported by the German Center for Infection Research (DZIF, doi:10.13039/100009139) within the *Deutsche Zentren der Gesundheitsforschung* (BMBF-DZG, German Centers for Health Research of the Federal Ministry of Education and Research), grant No 8020708703. The authors acknowledge support by the Open Access Publishing Fund of the University of Tübingen (<https://uni-tuebingen.de/en/58988>, accessed on 21 May 2021).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The genome-scale metabolic model of the human alveolar macrophage infected with SARS-CoV-2 is available in the BioModels Database [71] as an SBML Level 3 Version 1 file [72–74] with the flux balance constraints (fbc) extension package [75] within a COMBINE Archive OMEX file [76] under the accession number MODEL2003020001. A supplementary table in Microsoft Excel format is available as Supplementary Materials along with this article.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

ACE-2	angiotensin converting enzyme 2
ADP	adenosine diphosphate
ADSL1	adenylosuccinate lyase 1
ADSL2	adenylosuccinate lyase 2
ADSS	adenylosuccinate synthase
AICART	phosphoribosylaminoimidazolecarboxamide formyltransferase
AIRC	Phosphoribosylaminoimidazole carboxylase
AMP	adenosine monophosphate
ASPCTr	aspartate carbamoyltransferase
ATP	adenosine triphosphate
CBPS	carbamoyl-phosphate synthase
COMBINE	Computational Modeling in Biology Network
COVID-19	Coronavirus Disease 19
Csp	Copy number of structural proteins
DHORD9	dihydroorotic acid dehydrogenase
DHORTS	dihydroorotase
E	envelope
ERGIC	endoplasmic reticulum–Golgi intermediate compartment
FBA	flux balance analysis
fbc	flux balance constraints
FVA	flux variability analysis
GARFT	phosphoribosylglycinamide formyltransferase
GEM	genome-scale metabolic model
GISAID	Global Initiative on Sharing All Influenza Data
GK1	guanylate kinase 1
GLUPRT	glutamine phosphoribosylidiphosphate amidotransferase
hCOV	human coronavirus
hCoV-229E	human coronavirus 229E
ID	identifier
IMP	inosine monophosphate
IMPC	IMP cyclohydrolase

M	membrane
MERS	Middle East Respiratory Syndrome
Mv	viral molar mass
N	nucleocapsid
OMEX	Open Modeling EXchange format
OMPDC	orotidine-5'-phosphate decarboxylase
ORPT	orotate phosphoribosyltransferase
PPi	diphosphate
PRAGSr	phosphoribosylglycinamide synthase
PRAIS	phosphoribosylaminoimidazole synthase
PRASCS	phosphoribosylaminoimidazolesuccinocarboxamide synthase
PRFGS	phosphoribosylformylglycinamide synthase
PRPPS	phosphoribosylpyrophosphate synthetase
RPI	ribose-5-phosphate isomerase
S	spike
SARS	Severe Acute Respiratory Syndrome
SARS-CoV-2	Severe Acute Respiratory Syndrome coronavirus type 2
SBML	Systems Biology Markup Language
TGEV	transmissible gastroenteritis virus
VBOF	viral biomass objective function
vRNP	viral ribonucleoprotein complex
WT	wildtype

Appendix A

Table A1. Reactions from the host-derived enforcement experiments. The reaction identifiers listed in Figure 2 are BiGG identifiers [28]. In this table, the BiGG reaction identifiers are given, together with the reaction name and the subsystem, they occur in.

Reaction-ID	Reaction Name	Subsystem
ADSL1	adenylosuccinate lyase 1	Purine metabolism
ADSL2	adenylosuccinate lyase 2	Purine metabolism
ADSS	adenylosuccinate synthase	Purine metabolism
AICART	phosphoribosylaminoimidazolecarboxamide formyltransferase	Purine metabolism
AIRCr	Phosphoribosylaminoimidazole carboxylase	Purine metabolism
ASPCTr	aspartate carbamoyltransferase	Pyrimidine metabolism
CBPS	carbamoyl-phosphate synthase	Pyrimidine metabolism
DHORD9	dihydroorotic acid dehydrogenase	Pyrimidine metabolism
DHORTS	dihydroorotase	Pyrimidine metabolism
GARFT	phosphoribosylglycinamide formyltransferase	Purine metabolism
GK1	guanylate kinase 1	Purine metabolism
GLUPRT	glutamine phosphoribosylidiphosphate amidotransferase	Purine metabolism
IMPC	IMP cyclohydrolase	Purine metabolism
OMPDC	orotidine-5'-phosphate decarboxylase	Pyrimidine metabolism
ORPT	orotate phosphoribosyltransferase	Pyrimidine metabolism
PRAGSr	phosphoribosylglycinamide synthase	Purine metabolism
PRAIS	phosphoribosylaminoimidazole synthase	Purine metabolism
PRASCS	phosphoribosylaminoimidazolesuccinocarboxamide synthase	Purine metabolism
PRFGS	phosphoribosylformylglycinamide synthase	Purine metabolism
PRPPS	phosphoribosylpyrophosphate synthetase	Pentose phosphate pathway
RPI	ribose-5-phosphate isomerase	Pentose phosphate pathway

B.3 GEM infected with SARS-CoV-2 mutants confirms guanylate kinase as robust target

References

1. Zhou, P.; Yang, X.L.; Wang, X.G.; Hu, B.; Zhang, L.; Zhang, W.; Si, H.R.; Zhu, Y.; Li, B.; Huang, C.L.; et al. A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature* **2020**, *579*, 270–273. [[CrossRef](#)] [[PubMed](#)]
2. Mirtaleb, M.S.; Mirtaleb, A.H.; Nosrati, H.; Heshmatnia, J.; Falak, R.; Zolfaghari Emameh, R. Potential therapeutic agents to COVID-19: An update review on antiviral therapy, immunotherapy, and cell therapy. *Biomed. Pharmacother.* **2021**, *138*, 111518. [[CrossRef](#)] [[PubMed](#)]
3. Li, D.D.; Li, Q.H. SARS-CoV-2: Vaccines in the pandemic era. *Mil. Med. Res.* **2021**, *8*, 1. [[CrossRef](#)] [[PubMed](#)]
4. Zhao, J.; Zhao, S.; Ou, J.; Zhang, J.; Lan, W.; Guan, W.; Wu, X.; Yan, Y.; Zhao, W.; Wu, J.; et al. COVID-19: Coronavirus Vaccine Development Updates. *Front. Immunol.* **2020**, *11*, 602256. [[CrossRef](#)] [[PubMed](#)]
5. Liu, W.; Zhou, P.; Chen, K.; Ye, Z.; Liu, F.; Li, X.; He, N.; Wu, Z.; Zhang, Q.; Gong, X.; et al. Efficacy and safety of antiviral treatment for COVID-19 from evidence in studies of SARS-CoV-2 and other acute viral infections: A systematic review and meta-analysis. *CMAJ* **2020**, *192*, E734–E744. [[CrossRef](#)]
6. Valle, C.; Martin, B.; Touret, F.; Shannon, A.; Canard, B.; Guillemot, J.C.; Coutard, B.; Decroly, E. Drugs against SARS-CoV-2: What do we know about their mode of action? *Rev. Med. Virol.* **2020**, *30*, 1–10. [[CrossRef](#)]
7. Grein, J.; Ohmagari, N.; Shin, D.; Diaz, G.; Asperges, E.; Castagna, A.; Feldt, T.; Green, G.; Green, M.L.; Lescure, F.X.; et al. Compassionate Use of Remdesivir for Patients with Severe Covid-19. *N. Engl. J. Med.* **2020**, *382*, 2327–2336. [[CrossRef](#)]
8. Tregoning, J.S.; Brown, E.S.; Cheeseman, H.M.; Flight, K.E.; Higham, S.L.; Lemm, N.M.; Pierce, B.F.; Stirling, D.C.; Wang, Z.; Pollock, K.M. Vaccines for COVID-19. *Clin. Exp. Immunol.* **2020**, *202*, 162–192. [[CrossRef](#)]
9. European Centre for Disease Prevention and Control. *Risk Related to Spread of New SARS-CoV-2 Variants of Concern in the EU/EEA, First Update*; Technical Report; European Centre for Disease Prevention and Control: Stockholm, Sweden, 2021.
10. Centers for Disease Control and Prevention (US). *SARS-CoV-2 Variants of Concern*; U.S. Department of Health and Human Services, Centers for Disease Control and Prevention: Atlanta, GA, USA, 2021.
11. European Centre for Disease Prevention and Control. *Emergence of SARS-CoV-2 B.1.617 Variants in India and Situation in the EU/EEA—11 May 2021*; Technical Report; European Centre for Disease Prevention and Control: Stockholm, Sweden, 2021.
12. Cui, J.; Li, F.; Shi, Z.L. Origin and evolution of pathogenic coronaviruses. *Nat. Rev. Microbiol.* **2019**, *17*, 181–192. [[CrossRef](#)]
13. Bar-On, Y.M.; Flamholz, A.; Phillips, R.; Milo, R. Sars-cov-2 (Covid-19) by the numbers. *eLife* **2020**, *9*, e57309. [[CrossRef](#)]
14. Klein, S.; Cortese, M.; Winter, S.L.; Wachsmuth-Melm, M.; Neufeldt, C.J.; Cerikan, B.; Stanifer, M.L.; Boulant, S.; Bartenschlager, R.; Chlanda, P. SARS-CoV-2 structure and replication characterized by Situ Cryo-Electron Tomogr. *Nat. Commun.* **2020**, *11*, 5885. [[CrossRef](#)] [[PubMed](#)]
15. Turoňová, B.; Sikora, M.; Schürmann, C.; Hagen, W.J.; Welsch, S.; Blanc, F.E.; von Bülow, S.; Gecht, M.; Bagola, K.; Hörner, C.; et al. Mediat. Three Hinges. *Science* **2020**, *370*, 203–208. [[CrossRef](#)] [[PubMed](#)]
16. Lenard, J. Viral membranes. In *Encyclopedia of Virology*; Springer: Boston, MA, USA, 2008; pp. 308–314. [[CrossRef](#)]
17. de Armas-Rillo, L.; Valera, M.S.; Marrero-Hernández, S.; Valenzuela-Fernández, A. Membrane dynamics associated with viral infection. *Rev. Med. Virol.* **2016**, *26*, 146–160. [[CrossRef](#)] [[PubMed](#)]
18. Miller, S.; Krijnse-Locker, J. Modification of intracellular membrane structures for virus replication. *Nat. Rev. Microbiol.* **2008**, *6*, 363–374. [[CrossRef](#)] [[PubMed](#)]
19. Ketter, E.; Randall, G. Virus Impact on Lipids and Membranes. *Annu. Rev. Virol.* **2019**, *6*, 319–340. [[CrossRef](#)]
20. Schweizer, A.; Clausen, H.; Van Meer, G.; Hauri, H.P. Localization of O-glycan initiation, sphingomyelin synthesis, and glucosylceramide synthesis in Vero cells with respect to the endoplasmic reticulum-Golgi intermediate compartment. *J. Biol. Chem.* **1994**, *269*, 4035–4041. [[CrossRef](#)]
21. Mandala, V.S.; McKay, M.J.; Shcherbakov, A.A.; Dregni, A.J.; Kolocouris, A.; Hong, M. Structure and drug binding of the SARS-CoV-2 envelope protein transmembrane domain in lipid bilayers. *Nat. Struct. Mol. Biol.* **2020**, *27*, 1202–1208. [[CrossRef](#)]
22. Renz, A.; Widerspick, L.; Dräger, A. FBA reveals guanylate kinase as a potential target for antiviral therapies against SARS-CoV-2. *Bioinformatics* **2020**, *36*, i813–i821. [[CrossRef](#)]
23. Bordbar, A.; Lewis, N.E.; Schellenberger, J.; Pálsson, B.; Jamshidi, N. Insight into human alveolar macrophage and M. tuberculosis interactions via metabolic reconstructions. *Mol. Syst. Biol.* **2010**, *6*, 422. [[CrossRef](#)]
24. Hible, G.; Daalova, P.; Gilles, A.M.; Cherfils, J. Crystal structures of GMP kinase in complex with ganciclovir monophosphate and Ap5G. *Biochimie* **2006**, *88*, 1157–1164. [[CrossRef](#)]
25. Navé, J.F.; Eschbach, A.; Halazy, S. 9-(Phosphonoalkyl)guanine derivatives as substrates or inhibitors of guanylate kinase. *Arch. Biochem. Biophys.* **1992**, *295*, 253–257. [[CrossRef](#)]
26. Navé, J.F.; Taylor, D.; Tyms, S.; Kenny, M.; Eggenspillner, A.; Eschbach, A.; Dulworth, J.; Brennan, T.; Piriou, F.; Halazy, S. Synthesis, antiviral activity and enzymatic phosphorylation of 9-phosphonopentenyl derivatives of guanine. *Antivir. Res.* **1995**, *27*, 301–316. [[CrossRef](#)]
27. Aller, S.; Scott, A.; Sarkar-Tyson, M.; Soyer, O.S. Integrated human-virus metabolic stoichiometric modelling predicts host-based antiviral targets against Chikungunya, Dengue and Zika viruses. *J. R. Soc. Interface* **2018**, *15*, 20180125. [[CrossRef](#)] [[PubMed](#)]
28. Norsigian, C.J.; Pusarla, N.; McConn, J.L.; Yurkovich, J.T.; Dräger, A.; Pálsson, B.O.; King, Z. BiGG Models 2020: Multi-strain genome-scale models and expansion across the phylogenetic tree. *Nucleic Acids Res.* **2019**, *48*, D402–D406. [[CrossRef](#)] [[PubMed](#)]

29. Tilocca, B.; Soggiu, A.; Sanguinetti, M.; Babini, G.; De Maio, F.; Britti, D.; Zecconi, A.; Bonizzi, L.; Urbani, A.; Roncada, P. Immunoinformatic analysis of the SARS-CoV-2 envelope protein as a strategy to assess cross-protection against COVID-19. *Microbes Infect.* **2020**, *22*, 182–187. [[CrossRef](#)]
30. Stodola, J.K.; Dubois, G.; Le Coupanec, A.; Desforages, M.; Talbot, P.J. The OC43 human coronavirus envelope protein is critical for infectious virus production and propagation in neuronal cells and is a determinant of neurovirulence and CNS pathology. *Virology* **2018**, *515*, 134–149. [[CrossRef](#)]
31. Godet, M.; L'Haridon, R.; Vautherot, J.F.; Laude, H. TGEV corona virus ORF4 encodes a membrane protein that is incorporated into virions. *Virology* **1992**, *188*, 666–675. [[CrossRef](#)]
32. Yao, H.; Song, Y.; Chen, Y.; Wu, N.; Xu, J.; Sun, C.; Zhang, J.; Weng, T.; Zhang, Z.; Wu, Z.; et al. Molecular Architecture of the SARS-CoV-2 Virus. *Cell* **2020**, *183*, 730–738.e13. [[CrossRef](#)]
33. Lu, S.; Ye, Q.; Singh, D.; Cao, Y.; Diedrich, J.K.; Yates, J.R.; Villa, E.; Cleveland, D.W.; Corbett, K.D. The SARS-CoV-2 nucleocapsid phosphoprotein forms mutually exclusive condensates with RNA and the membrane-associated M protein. *Nat. Commun.* **2021**, *12*, 1–15. [[CrossRef](#)]
34. Neuman, B.W.; Kiss, G.; Kunding, A.H.; Bhella, D.; Baksh, M.F.; Connelly, S.; Droese, B.; Klaus, J.P.; Makino, S.; Sawicki, S.G.; et al. A structural analysis of M protein in coronavirus assembly and morphology. *J. Struct. Biol.* **2011**, *174*, 11–22. [[CrossRef](#)]
35. Pedley, A.M.; Benkovic, S.J. A New View into the Regulation of Purine Metabolism: The Purinosome. *Trends Biochem. Sci.* **2017**, *42*, 141–154. [[CrossRef](#)] [[PubMed](#)]
36. Löffler, M.; Fairbanks, L.D.; Zameitat, E.; Marinaki, A.M.; Simmonds, H.A. Pyrimidine pathways in health and disease. *Trends Mol. Med.* **2005**, *11*, 430–437. [[CrossRef](#)] [[PubMed](#)]
37. Elbe, S.; Buckland-Merrett, G. Data, disease and diplomacy: GISAID's innovative contribution to global health. *Glob. Challenges* **2017**, *1*, 33–46. [[CrossRef](#)]
38. Shu, Y.; McCauley, J. GISAID: Global initiative on sharing all influenza data—From vision to reality. *Eurosurveillance* **2017**, *22*, 30494. [[CrossRef](#)]
39. Abassi, Z.; Knaney, Y.; Karram, T.; Heyman, S.N. The Lung Macrophage in SARS-CoV-2 Infection: A Friend or a Foe? *Front. Immunol.* **2020**, *11*, 1312. [[CrossRef](#)]
40. Qi, F.; Qian, S.; Zhang, S.; Zhang, Z. Single cell RNA sequencing of 13 human tissues identify cell types and receptors of human coronaviruses. *Biochem. Biophys. Res. Commun.* **2020**, *526*, 135–140. [[CrossRef](#)] [[PubMed](#)]
41. Joel Funk, C.; Wang, J.; Ito, Y.; Travanty, E.A.; Voelker, D.R.; Holmes, K.V.; Mason, R.J. Infection of human alveolar macrophages by human coronavirus strain 229E. *J. Gen. Virol.* **2012**, *93*, 494–503. [[CrossRef](#)] [[PubMed](#)]
42. Al-Qahtani, A.A.; Lyroni, K.; Aznaourova, M.; Tseliou, M.; Al-Anazi, M.R.; Al-Ahdal, M.N.; Alkahtani, S.; Sourvinos, G.; Tsatsanis, C. Middle east respiratory syndrome corona virus spike glycoprotein suppresses macrophage responses via DPP4-mediated induction of IRAK-M and PPAR γ . *Oncotarget* **2017**, *8*, 9053–9066. [[CrossRef](#)] [[PubMed](#)]
43. Gu, J.; Gong, E.; Zhang, B.; Zheng, J.; Gao, Z.; Zhong, Y.; Zou, W.; Zhan, J.; Wang, S.; Xie, Z.; et al. Multiple organ infection and the pathogenesis of SARS. *J. Exp. Med.* **2005**, *202*, 415–424. [[CrossRef](#)]
44. Chu, H.; Chan, J.F.W.; Wang, Y.; Yuen, T.T.T.; Chai, Y.; Hou, Y.; Shuai, H.; Yang, D.; Hu, B.; Huang, X.; et al. Comparative Replication and Immune Activation Profiles of SARS-CoV-2 and SARS-CoV in Human Lungs: An Ex Vivo Study with Implications for the Pathogenesis of COVID-19. *Clin. Infect. Dis.* **2020**, *71*, 1400–1409. [[CrossRef](#)]
45. Dalskov, L.; Möhlenberg, M.; Thyrsted, J.; Blay-Cadanet, J.; Poulsen, E.T.; Folkersen, B.H.; Skaarup, S.H.; Olgarnier, D.; Reinert, L.; Enghild, J.J.; et al. SARS-CoV-2 evades immune detection in alveolar macrophages. *EMBO Rep.* **2020**, *21*, e51252. [[CrossRef](#)] [[PubMed](#)]
46. Delattre, H.; Sasidharan, K.; Soyer, O.S. Inhibiting the reproduction of SARS-CoV-2 through perturbations in human lung cell metabolic network. *Life Sci. Alliance* **2021**, *4*, e202000869. [[CrossRef](#)] [[PubMed](#)]
47. Swainston, N.; Smallbone, K.; Hefzi, H.; Dobson, P.D.; Brewer, J.; Hanscho, M.; Zielinski, D.C.; Ang, K.S.; Gardiner, N.J.; Gutierrez, J.M.; et al. Recon 2.2: From reconstruction to model of human metabolism. *Metabolomics* **2016**, *12*, 109. [[CrossRef](#)] [[PubMed](#)]
48. Rahman, M.R.; Banik, A.; Chowdhury, I.M.; Sajib, E.H.; Sarkar, S. Identification of potential antivirals against SARS-CoV-2 using virtual screening method. *Inform. Med. Unlocked* **2021**, *23*, 100531. [[CrossRef](#)] [[PubMed](#)]
49. Hussien, M.A.; Abdelaziz, A.E. Molecular docking suggests repurposing of brincidofovir as a potential drug targeting SARS-CoV-2 ACE2 receptor and main protease. *Netw. Model. Anal. Health Inform. Bioinform.* **2020**, *9*, 1–18. [[CrossRef](#)]
50. National Library of Medicine (U.S.). Available online: <https://clinicaltrials.gov/> (accessed on 29 May 2012).
51. Bojkova, D.; Costa, R.; Bechtel, M.; Ciesek, S.; Michaelis, M.; Cinatl, J. Targeting pentose phosphate pathway for SARS-CoV-2 therapy. *bioRxiv* **2020**. [[CrossRef](#)]
52. Zhang, Y.; Guo, R.; Kim, S.H.; Shah, H.; Zhang, S.; Liang, J.H.; Fang, Y.; Gentili, M.; Leary, C.N.; Elledge, S.J.; et al. SARS-CoV-2 hijacks folate and one-carbon metabolism for viral replication. *Nat. Commun.* **2021**, *12*, 1676. [[CrossRef](#)]
53. Caruso, A.; Caccuri, F.; Bugatti, A.; Zani, A.; Vanoni, M.; Bonfanti, P.; Cazzaniga, M.E.; Perno, C.F.; Messa, C.; Alberghina, L. Methotrexate inhibits SARS-CoV-2 virus replication “in vitro”. *J. Med. Virol.* **2021**, *93*, 1780–1785. [[CrossRef](#)]
54. Stegmann, K.M.; Dickmanns, A.; Gerber, S.; Nikolova, V.; Klemke, L.; Manzini, V.; Schlösser, D.; Bierwirth, C.; Freund, J.; Sitte, M.; et al. The folate antagonist methotrexate diminishes replication of the coronavirus SARS-CoV-2 and enhances the antiviral efficacy of remdesivir in cell culture models. *bioRxiv* **2020**. [[CrossRef](#)]

B.3 GEM infected with SARS-CoV-2 mutants confirms guanylate kinase as robust target

55. Coelho, A.R.; Oliveira, P.J. Dihydroorotate dehydrogenase inhibitors in SARS-CoV-2 infection. *Eur. J. Clin. Investig.* **2020**, *50*, e13366. [[CrossRef](#)]
56. Hoffmann, H.H.; Kunz, A.; Simon, V.A.; Palese, P.; Shaw, M.L. Broad-spectrum antiviral that interferes with *De Novo* Pyrimidine Biosynthesis. *Proc. Natl. Acad. Sci. USA* **2011**, *108*, 5777–5782. [[CrossRef](#)] [[PubMed](#)]
57. Cheung, N.N.; Lai, K.K.; Dai, J.; Kok, K.H.; Chen, H.; Chan, K.H.; Yuen, K.Y.; Tsun Kao, R.Y. Broad-spectrum inhibition of common respiratory RNA viruses by a pyrimidine synthesis inhibitor with involvement of the host antiviral response. *J. Gen. Virol.* **2017**, *98*, 946–954. [[CrossRef](#)] [[PubMed](#)]
58. Chen, S.; Ding, S.; Yin, Y.; Xu, L.; Li, P.; Peppelenbosch, M.P.; Pan, Q.; Wang, W. Suppression of pyrimidine biosynthesis by targeting DHODH enzyme robustly inhibits rotavirus replication. *Antivir. Res.* **2019**, *167*, 35–44. [[CrossRef](#)] [[PubMed](#)]
59. Luthra, P.; Naidoo, J.; Pietzsch, C.A.; De, S.; Khadka, S.; Anantpadma, M.; Williams, C.G.; Edwards, M.R.; Davey, R.A.; Bukreyev, A.; et al. Inhibiting pyrimidine biosynthesis impairs Ebola virus replication through depletion of nucleoside pools and activation of innate immune responses. *Antivir. Res.* **2018**, *158*, 288–302. [[CrossRef](#)] [[PubMed](#)]
60. Luban, J.; Sattler, R.A.; Mühlberger, E.; Graci, J.D.; Cao, L.; Weetall, M.; Trotta, C.; Colacino, J.M.; Bavari, S.; Strambio-De-Castillia, C.; et al. The DHODH inhibitor PTC299 arrests SARS-CoV-2 replication and suppresses induction of inflammatory cytokines. *Virus Res.* **2021**, *292*, 198246. [[CrossRef](#)]
61. Xiong, R.; Zhang, L.; Li, S.; Sun, Y.; Ding, M.; Wang, Y.; Zhao, Y.; Wu, Y.; Shang, W.; Jiang, X.; et al. Novel and potent inhibitors targeting DHODH are broad-spectrum antivirals against RNA viruses including newly-emerged coronavirus SARS-CoV-2. *Protein Cell* **2020**, *11*, 723–739. [[CrossRef](#)] [[PubMed](#)]
62. Grant, O.C.; Montgomery, D.; Ito, K.; Woods, R.J. Analysis of the SARS-CoV-2 spike protein glycan shield reveals implications for immune recognition. *Sci. Rep.* **2020**, *10*, 14991. [[CrossRef](#)]
63. Watanabe, Y.; Allen, J.D.; Wrapp, D.; McLellan, J.S.; Crispin, M. Site-specific glycan analysis of the SARS-CoV-2 spike. *Science* **2020**, *369*, 330–333. [[CrossRef](#)]
64. Abu-Farha, M.; Thanaraj, T.A.; Qaddoumi, M.G.; Hashem, A.; Abubaker, J.; Al-Mulla, F. The Role of Lipid Metabolism in COVID-19 Virus Infection and as a Drug Target. *Int. J. Mol. Sci.* **2020**, *21*, 3544. [[CrossRef](#)]
65. Wu, Q.; Zhou, L.; Sun, X.; Yan, Z.; Hu, C.; Wu, J.; Xu, L.; Li, X.; Liu, H.; Yin, P.; et al. Altered Lipid Metabolism in Recovered SARS Patients Twelve Years after Infection. *Sci. Rep.* **2017**, *7*, 9110. [[CrossRef](#)]
66. Nguyen, A.; Guedán, A.; Mousnier, A.; Swieboda, D.; Zhang, Q.; Horkai, D.; Le Novère, N.; Solari, R.; Wakelam, M.J. Host lipidome analysis during rhinovirus replication in HBEc2s identifies potential therapeutic target. *J. Lipid Res.* **2018**, *59*, 1671–1684. [[CrossRef](#)]
67. Yan, B.; Zou, Z.; Chu, H.; Chan, G.; Tsang, J.O.L.; Lai, P.M.; Yuan, S.; Yip, C.C.Y.; Yin, F.; Kao, R.Y.T.; et al. Lipidomic Profiling Reveals Significant Perturbations of Intracellular Lipid Homeostasis in Enterovirus-Infected Cells. *Int. J. Mol. Sci.* **2019**, *20*, 5952. [[CrossRef](#)] [[PubMed](#)]
68. Yan, B.; Chu, H.; Yang, D.; Sze, K.H.; Lai, P.M.; Yuan, S.; Shuai, H.; Wang, Y.; Kao, R.Y.T.; Chan, J.F.W.; et al. Characterization of the lipidomic profile of human coronavirus-infected cells: Implications for lipid metabolism remodeling upon coronavirus replication. *Viruses* **2019**, *11*, 73. [[CrossRef](#)] [[PubMed](#)]
69. Visioli, F.; Colombo, C.; Monti, S.; Giulidori, P.; Galli, C. S-adenosyl-L-methionine: Role in phosphatidylcholine synthesis and in vitro effects on the ethanol-induced alterations of lipid metabolism. *Pharmacol. Res.* **1998**, *37*, 203–206. [[CrossRef](#)] [[PubMed](#)]
70. Ye, C.; Sutter, B.M.; Wang, Y.; Kuang, Z.; Tu, B.P. A Metabolic Function for Phospholipid and Histone Methylation. *Mol. Cell* **2017**, *66*, 180–193.e8. [[CrossRef](#)] [[PubMed](#)]
71. Malik-Sheriff, R.S.; Glont, M.; Nguyen, T.V.N.; Tiwari, K.; Roberts, M.G.; Xavier, A.; Vu, M.T.; Men, J.; Maire, M.; Kananathan, S.; et al. BioModels—15 years of sharing computational models in life science. *Nucleic Acids Res.* **2020**, *48*, D407–D415. [[CrossRef](#)] [[PubMed](#)]
72. Keating, S.M.; Waltemath, D.; König, M.; Zhang, F.; Dräger, A.; Chaouiya, C.; Bergmann, F.T.; Finney, A.; Gillespie, C.S.; Helikar, T.; et al. SBML Level 3: An extensible format for the exchange and reuse of biological models. *Mol. Syst. Biol.* **2020**, *16*, e9110. [[CrossRef](#)]
73. Renz, A.; Mostolizadeh, R.; Dräger, A. Clinical Applications of Metabolic Models in SBML Format. In *Systems Medicine*; Wolkenhauer, O., Ed.; Academic Press: Oxford, UK, 2020; Volume 3, pp. 362–371. [[CrossRef](#)]
74. Hucka, M.; Bergmann, F.T.; Dräger, A.; Hoops, S.; Keating, S.M.; Le Novère, N.; Myers, C.J.; Olivier, B.G.; Sahle, S.; Schaff, J.C.; et al. Systems Biology Markup Language (SBML) Level 3 Version 1 Core. *J. Integr. Bioinform.* **2018**, *15*, 1. [[CrossRef](#)]
75. Olivier, B.G.; Bergmann, F.T. SBML Level 3 Package: Flux Balance Constraints version 2. *J. Integr. Bioinform.* **2018**, *15*. [[CrossRef](#)]
76. Bergmann, F.T.; Adams, R.; Moodie, S.; Cooper, J.; Glont, M.; Golebiewski, M.; Hucka, M.; Laibe, C.; Miller, A.K.; Nickerson, D.P.; et al. COMBINE archive and OMEX format: One file to share all information to reproduce a modeling project. *BMC Bioinform.* **2014**, *15*, 369. [[CrossRef](#)]

REVIEW ARTICLE OPEN

Curating and comparing 114 strain-specific genome-scale metabolic models of *Staphylococcus aureus*Alina Renz^{1,2,3} and Andreas Dräger^{1,2,3,4}✉

Staphylococcus aureus is a high-priority pathogen causing severe infections with high morbidity and mortality worldwide. Many *S. aureus* strains are methicillin-resistant (MRSA) or even multi-drug resistant. It is one of the most successful and prominent modern pathogens. An effective fight against *S. aureus* infections requires novel targets for antimicrobial and antistaphylococcal therapies. Recent advances in whole-genome sequencing and high-throughput techniques facilitate the generation of genome-scale metabolic models (GEMs). Among the multiple applications of GEMs is drug-targeting in pathogens. Hence, comprehensive and predictive metabolic reconstructions of *S. aureus* could facilitate the identification of novel targets for antimicrobial therapies. This review aims at giving an overview of all available GEMs of multiple *S. aureus* strains. We downloaded all 114 available GEMs of *S. aureus* for further analysis. The scope of each model was evaluated, including the number of reactions, metabolites, and genes. Furthermore, all models were quality-controlled using MEMOTE, an open-source application with standardized metabolic tests. Growth capabilities and model similarities were examined. This review should lead as a guide for choosing the appropriate GEM for a given research question. With the information about the availability, the format, and the strengths and potentials of each model, one can either choose an existing model or combine several models to create models with even higher predictive values. This facilitates model-driven discoveries of novel antimicrobial targets to fight multi-drug resistant *S. aureus* strains.

npj Systems Biology and Applications (2021)7:30; <https://doi.org/10.1038/s41540-021-00188-4>

INTRODUCTION

Staphylococcus aureus is an opportunistic pathogen that asymptotically and permanently colonizes the nose of up to one third of the human population¹. It is a commensal of the mucosae and the human skin, but can also cause severe infections with high morbidity, mortality, and healthcare-associated costs². Methicillin-resistant *S. aureus* (MRSA) is one of the most successful modern pathogens³. In 2017, the WHO published a priority pathogens list for the research and development of new antibiotics. Among the clarithromycin-resistant *Helicobacter pylori* and the vancomycin-resistant *Enterococcus faecium*, *S. aureus*, especially the methicillin-resistant *S. aureus* (MRSA), vancomycin intermediate (VISA), and vancomycin resistant strains (VRSA), are high priority pathogens⁴.

Staphylococcus aureus bacteremia (SAB) is a common infection⁵. The incidence rate ranges from approximately 20 cases per 100,000 persons per year in Canada⁶ to approximately 50 cases per 100,000 persons, inferred from the United States surveillance data⁷. The higher incidence rate might be due to the greater burden of MRSA⁵. SAB can be classified into three categories: (1) Hospital onset of health-care associated infections, e.g., nosocomial; (2) Community onset of health-care associated infections, and (3) community acquired infections⁸. Besides SAB, *S. aureus*, and especially MRSA, is the leading cause of endocarditis, bone and joint infections, skin and soft tissue infections, and further hospital-acquired infections³. A study from 2013 revealed over 80,000 invasive infections and 11,000 deaths per year due to MRSA in the United States. Compared to the previous years, the number of invasive MRSA infections declined slightly⁹. Unfortunately, the rate decline of MRSA infections has recently slowed down according to the "Morbidity and Mortality Weekly Report" of the Centers for Disease Control and Prevention¹⁰,

while the number of methicillin-susceptible *S. aureus* (MSSA) bloodstream infections even slightly increased. In 2017, nearly 120,000 *S. aureus* bloodstream infections and 20,000 associated deaths occurred in the United States¹⁰. Hence, strategies for preventing infections inside and outside acute care settings are required to further reduce the amount of invasive MRSA infections.

The transmission of *S. aureus* in general, and MRSA in particular, is facilitated by the long persistence time of *S. aureus* colonization. Nearly any item with skin contact can serve as fomes. In a hospital setting, this can include coats and clothes from doctors and nursing staff, pens, and mobile devices, such as cell phones³. Studies also suggest that infecting *S. aureus* isolates also persist in households three months after skin infections¹. Even across and within athletic fitness facilities, *S. aureus* is found on different surfaces, including weight plates and treadmill handles¹².

Besides the challenge of controlling *S. aureus* colonization in multiple environments, *S. aureus* strains evolve and adapt to different environments due to variability in diversity, mobile genetic elements (MGEs), and accumulation of mutations¹³⁻¹⁵. Mediators of virulence, immune evasion, and antibiotic resistance are commonly found within the accessory components of the *S. aureus* genomes, consisting of MGEs with pathogenicity islands, chromosomal cassettes, transposons, plasmids, and bacteriophages. Compared to the core genome, the accessory genome is more variable and also often more strain-specific³. MGEs in *S. aureus* can carry antibiotic resistance genes for resistances against penicillin, trimethoprim, erythromycin, clindamycin, and tetracyclines¹⁵. However, strains not only evolve and develop antibiotic resistances, they even replace each other within the same host¹⁴.

¹Computational Systems Biology of Infections and Antimicrobial-Resistant Pathogens, Institute for Bioinformatics and Medical Informatics (IBMI), University of Tübingen, Tübingen, Germany. ²Department of Computer Science, University of Tübingen, Tübingen, Germany. ³Cluster of Excellence 'Controlling Microbes to Fight Infections', University of Tübingen, Tübingen, Germany. ⁴German Center for Infection Research (DZIF), Partner Site Tübingen, Tübingen, Germany. [✉]email: draeger@informatik.uni-tuebingen.de

To fight *S. aureus* infections, several new antimicrobial and antistaphylococcal drugs have been developed recently^{3,13}, including oritavancin and ceftaroline^{16,17}. Despite the development of new antibiotics, *S. aureus* in general, and MRSA in particular, remains a prominent pathogen with persisting high mortality³. Since *S. aureus* will continue to evolve and develop new resistances¹³, the research on *S. aureus* and the development of new antimicrobials is of urgency to fight *S. aureus* infections.

One possibility for the identification of novel targets for antimicrobial therapies is the use of genome-scale metabolic models (GEMs). Advances in high-throughput techniques and whole-genome sequencing facilitate the construction of GEMs^{18,19}. They are reconstructed based on information from genome sequences and experimentally obtained biochemistry^{19,20}. With this information, stoichiometry-based and mass-balanced metabolic reactions can be formulated using gene-protein-reaction associations (GPRs). These stoichiometry-based GEMs can predict metabolic flux values within the constructed network²¹ and optimization techniques. Optimization techniques, such as flux balance analysis (FBA), use linear programming²⁰. Recent advances in the reconstruction of GEMs and the fast analysis and integration of omics data enabled metabolic studies with model-driven hypotheses and context-specific simulations^{22,23}. Among the multiple applications of GEMs is the drug targeting in pathogens and the modeling of interactions among multiple cells or organisms²⁰. These approaches could be used to investigate and develop novel antimicrobials or antistaphylococals. However, depending on the pathogen and strain, various models of *S. aureus* strains might be required to investigate the best antistaphylococcal target for a certain *S. aureus* strain.

In this review, we present all currently available GEMs of *S. aureus* from various databases. The available models were compared regarding their scope, their availability, their format, and their immediate usability. For various reasons, some of the models required revisions, such as converting spreadsheet file formats to SBML²⁴ or ensuring the syntactic validity of SBML files. After having all models available as syntactically valid SBML files, their growth-capabilities, their predictive value, and the similarities between the various models were investigated. This review gives an overview of the available models and their properties to identify the appropriate model for a specific research question.

MODEL OVERVIEW

Introduction of the models

Databases such as BiGG²⁵ or BioModels²⁶ comprise a variety of genome-scale metabolic models. Together with models from other databases and supplementary information from scientific publications, a large number of genome-scale metabolic models of *S. aureus* is available: The BioModels database contains two models of *S. aureus* by Becker et al.²⁷ and Heinemann et al.²⁸, both build in 2005. The BioModels database also harbors the models created within the Path2Models project²⁹. In this project, 33 whole genome metabolism models of *S. aureus* were automatically created and curated between 2012 and 2013²⁹. The BiGG Models Database contains two GEMs of *S. aureus*: the already mentioned model by Becker et al.²⁷ and a recently published model by Seif et al.³⁰ from 2019. Lee et al. published thirteen genome-scale metabolic reconstructions of multiple *Staphylococcus aureus* strains in 2009³¹. In 2016, Bosi et al.³² curated and published 64 genome-scale metabolic models of various *S. aureus* strains. Together with the *S. aureus* model published within the gut microbiota resource of the Virtual Metabolic Human (VMH) Database^{33,34}, a total number of 114 genome-scale metabolic models of *Staphylococcus aureus* exists today.

All models were downloaded, tested, and evaluated using COBRapy³⁵ and MEMOTE³⁶. MEMOTE is an open-source software that contains a standardized and community-maintained set of metabolic model tests³⁶. The overall MEMOTE score comprises information about annotations of metabolites, reactions, and genes, the inclusion of Systems Biology Ontology (SBO) terms, and the model's consistency. Within the annotations sections, the presence and conformity of different database identifiers is evaluated. In the SBO term section, the annotation of model instances with appropriate SBO terms is assessed. The model consistency check comprises tests to evaluate the stoichiometric consistency, mass and charge balances, metabolite connectivity, and unbounded fluxes in default medium³⁶. However, the MEMOTE score currently does not consider information about e.g., realistic growth rates, orphan or dead-end metabolites, stoichiometrically balanced cycles, or duplicated reactions. MEMOTE includes this information in its report but does not incorporate it into the calculated score. The number of model instances and their MEMOTE score are indicated in Fig. 1.

iSB619—GEM by Becker et al. The first, initial draft of an *S. aureus* genome-scale reconstruction was curated by Becker and Palsson in 2005. They reconstructed the *S. aureus* strain N315 with 619 genes, 743 reactions, and 655 metabolites. The GEM was curated based on the key metabolic pathways in the Kyoto Encyclopedia of Genes and Genomes (KEGG) database³⁷. Subsequently, The Institute for Genomic Research (TIGR) website³⁸ was browsed for additional reactions. 91% of all reactions are linked with genes or open reading frames in so-called gene-protein-reaction associations (GPRs). This first-draft GEM is almost completely elementally and charge balanced. The biomass objective function was formulated based on the biomass data from *Bacillus subtilis*³⁹ and substituted where necessary. It contains metabolites, such as amino acids, nucleotides, lipids, and cell wall constituents²⁷. The first *S. aureus* GEM reached a MEMOTE score of 67% and is available as a file in SBML Level 3 Version 1⁴⁰ format with flux balance constraints (fbc) extension⁴¹ and BiGG identifiers.

iMH551—GEM by Heinemann et al. In the same year, the second genome-scale reconstruction of *S. aureus* was published by Heinemann et al. Both research groups curated the *S. aureus* strain N315 and used the KEGG³⁷ and TIGR database³⁸, together with literature for genome regions with limited sequence homology for gene function assignments. A new biomass objective function was specifically defined for *S. aureus* based on integration of literature data from a variety of different *S. aureus* strains. The biomass objective function was built upon the five polymer categories DNA, RNA, proteins, lipids, and cell wall components, and extended by pool solutes. The reconstruction includes 801 metabolites and 860 reactions that are based on 551 genes and simulates aerobic and anaerobic growth²⁸. This *S. aureus* GEM reached a MEMOTE score of 35% and is also available as SBML Level 3 file with fbc extension. The genes are not included in the SBML file.

GEMs by Lee et al. Lee et al. utilized the ERGOTM bioinformatics suite⁴² and the KEGG ligand/reaction database³⁷ to curate metabolic reconstructions of multiple *S. aureus* genomes. The DNA sequence and associated open reading frames (ORFs) or protein sequences were integrated into the ERGO genome database. ORFs were called via a combination of programs and annotated automatically or manually. BLAST was used to compute the protein similarities. Functional assignments, relationship computation, and pathway analyses based on existence of ortholog and protein family clusters led to automated metabolic reconstructions. Manual steps included the review of every gene in the genome, pathway curations, and the consideration and reconciliation of

motif/domain database results for functional assignments. For identified missing steps within a certain pathway, Lee et al. searched for orthologs or published biochemical activities. For all complete, incomplete, or partial EC number annotations, associated reactions were identified via the ERGO pathway collections and KEGG database. Lee et al. used both biomass compositions from Becker et al.²⁷ and Heinemann et al.²⁸ for their analyses. On average, the thirteen *S. aureus* reconstructions included 1476 ± 14 reactions and 1406 ± 11 metabolites. All models are available as Excel spreadsheet files with KEGG identifiers.

GEMs from Path2Models. More than 140,000 freely available and automatically generated mathematical models from pathway representations are available through the Path2Models project. KEGG³⁷, BioCharta⁴³, MetaCyc⁴⁴, and SABIO-RK⁴⁵ served as databases to generate three types of models, including genome-scale metabolic reconstructions. The pipeline for generating GEMs starts with the extraction of pathway data from KEGG³⁷ and MetaCyc⁴⁴. To reconcile the different metabolite and reaction identifiers, MNXref was used⁴⁶. MNXref was further used to define default metabolite formulas and charge states. It allowed the mapping to different databases for a semantical annotation in accordance with the Minimal Information Required In the Annotation of Models (MIRIAM) guidelines⁴⁷. To all GEMs, a default biomass objective function containing all 20 amino acids, RNA and DNA nucleotide precursors, glycogen, and ATP was added. Between 2012 and 2013, 33 *S. aureus* GEMs were curated with the help of this pipeline, including one bovine strain. This strain had 6110 reactions, 4416 metabolites, and 1198 genes. The other *S. aureus* GEMs have on average 3064 ± 103 reactions, 2186 ± 75 metabolites, and 519 ± 12 genes. All models have a MEMOTE score of 48% and are available at the BioModels database as SBML Level 2 files⁴⁸ with mixed nomenclature.

GEMs by Bosi et al. In 2016, Bosi et al. constructed 64 GEMs of different *S. aureus* strains. They started by extending and adding content from KEGG³⁷, Model SEED⁴⁹, and MetaCyc⁴⁴ to the *S. aureus* N315 model iSB619 by Becker et al. This manually curated model was used as reference for other *S. aureus* strains. Shared genes and reactions were identified and subsequently, strain-specific metabolic content available from KEGG³⁷, Model SEED⁴⁹, and BioCyc⁵⁰ was manually added to the strain-specific GEMs. Since an *S. aureus* biomass composition was not available, the biomass objective functions from Becker et al.²⁷ and Heinemann et al.²⁸ were combined and *S. aureus*-specific data regarding the fatty acid composition in the biomass were used to adjust the biomass objective function. A gap-filling step further refined the models. On average, the models have 1460 ± 94 reactions, 1446 ± 47 metabolites, and 788 ± 116 genes with an average MEMOTE score of $36 \pm 1\%$. All models are available as SBML Level 3 files⁵¹ with fbc extension and BiGG nomenclature.

GEM by Magnúsdóttir et al. To elucidate the role of microbial communities in human metabolism and health, Magnúsdóttir et al. semi-automatically generated genome-scale metabolic reconstructions of 773 human gut bacteria, including *S. aureus* USA300-FPR3757³³. By using a comparative metabolic reconstruction method that propagates refinements from one metabolic reconstruction to others, the model quality of all 773 models was improved. The basis for each reconstruction were draft GEMs from Model SEED⁴⁹ and KBase⁵² including gap-filling, refinement via rBioNet⁵³, and quality control and quality assurance testing. Further refinement steps included the verification of reaction directionalities as well as mass and charge imbalances. The reconstructions were extended by gut-microbiota specific

subsystems and central metabolic subsystems, and anaerobic growth was enabled. Leak tests and the removal of infeasible flux loops further refined the model. The *S. aureus* model contains 1403 reactions, 1193 metabolites, and 859 genes, and reached a MEMOTE score of 45%. It is available as SBML Level 3 file with fbc extension and VMH nomenclature.

iYS854—GEM by Seif et al. Seif et al. manually reconstructed a comprehensive genome-scale metabolic model of *S. aureus* USA300 str. JE2 containing 886 genes, 1455 reactions, 1335 metabolites, and 673 three-dimensional protein structures. The GEM was build upon one of the reconstructions of Bosi et al.³². Extensive and detailed manual curation was supported by literature reviews and network evaluations. The initial model was extended by an updated biomass objective function. Model instances, such as genes, reactions, and metabolites, were enriched with cross-references and metadata. More than 50 metabolic sub-modules were examined, curated, and added to the GEM, together with over 200 confidence scores and 300 references. By this manual curation, 569 new metabolic processes, 214 new ORF assignments and 207 new metabolites were added. Experimental validation of the model revealed an 85% agreement with gene essentiality data and 68% agreement with experimental physiological data³⁰. A model evaluation with MEMOTE revealed with 81% the highest MEMOTE score of all tested models. The model is available as SBML Level 3 file with fbc extension and BiGG identifiers.

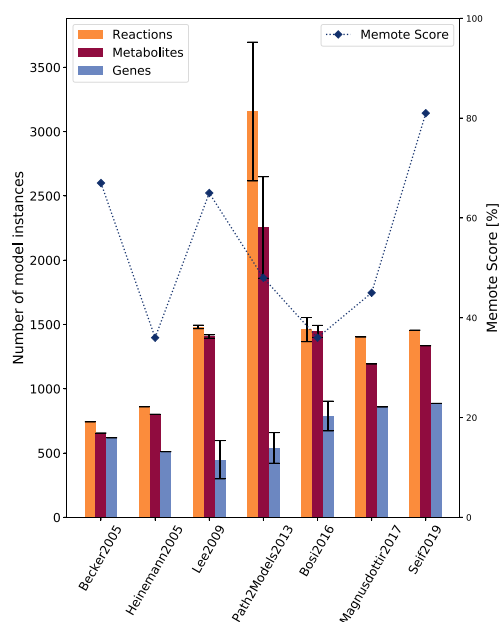


Fig. 1 Properties of all available *S. aureus* models and their scopes. For all models, the number of reactions, metabolites, and genes in the model is illustrated. MEMOTE conducts standardized and community-maintained metabolic tests for quality control and quality assurance of genome-scale metabolic models (GEMs) and assigns the tested model a score ranging from 0 to 100%. Lee et al.³¹, Bosi et al.³², and the Path2Models Project²⁹ published a collection of different *S. aureus* models. For the collections, the mean number of model instances is shown and the error bar indicates the standard deviation (s.d.).

B.4 Curating and comparing 114 strain-specific GEMs of *S. aureus*

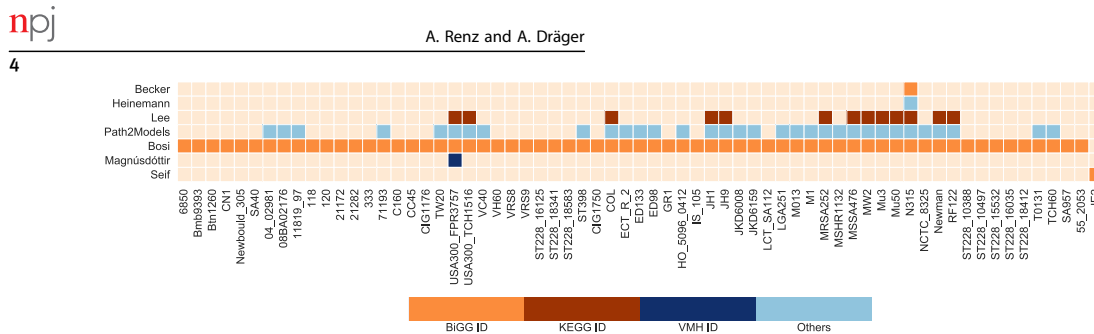


Fig. 2 Occurrence of models for *S. aureus* strains. Sixty-five different strains of *S. aureus* are available over the seven publications. Some *S. aureus* strains, such as the *S. aureus* strain USA300-FPR3757, occur in several publications, for other strains, only one publication is available. The colors indicate the utilized metabolite and reaction identifiers in the respective models. Models with similar or same identifiers can be compared more easily.

Presence of strains

The 114 currently available GEMs divide into 65 different *S. aureus* strains. In Fig. 2, the diverse *S. aureus* strains and their occurrence in the different publications is illustrated. Some strains, such as USA300-FPR3757 or N315 occur several times in different databases. Others, like the GEM for *S. aureus* strain JE2 occur only once in literature so far. The colors indicate the metabolite and reaction identifier in the respective model. Among the five models of the strain N315, two models exist that both carry BiGG identifiers. Models with same identifiers can be compared more easily than models with discriminating identifiers. Thirteen *S. aureus* strains occur at least in three different databases or publications with varying identifiers.

Due to the vast amount of different *S. aureus* strains, we elucidate only the strains that are shared over multiple databases. As already mentioned, the GEMs of the *S. aureus* strain N315 are the most prevalent. This strain was isolated from the pharyngeal smear of a Japanese patient in 1982⁵⁴. It is a methicillin-resistant *S. aureus* (MRSA). The only effective antibiotic against it was vancomycin. However, in 1997, a vancomycin-resistant MRSA strain, Mu50, was discovered in a Japanese infant with a surgical wound infection⁵⁴. The closely related strain Mu3 is a hetero vancomycin-intermediate MRSA strain. Strains with heterogeneous vancomycin resistance can spontaneously produce cells with increasing resistance against vancomycin^{55,56}.

The isolates JH1 and JH9 stem from a series of MRSA isolates obtained from a patient receiving extensive therapy. These strains are also vancomycin-intermediate *S. aureus*. The first isolate, JH1, was taken before the chemotherapy and was fully susceptible to vancomycin. The last isolate, JH9, from the end of the therapy showed decreased susceptibility to vancomycin⁵⁷.

The *S. aureus* strains of type USA300 are clones of the community-acquired MRSA^{58,59}. It causes invasive infections in children and adults in the USA⁵⁸, but also in Canada and Europe⁵⁹. It is suggested that USA300 is more virulent than other community-acquired MRSA strains⁵⁸. FPR3757 is a multidrug-resistant USA300 strain with acquired mobile genetic elements (MGEs) encoding resistance and virulence determinant that probably lead to enhanced pathogenicity⁵⁹. The other USA300 isolate, TCH1516, also named USA300-HOU-MR, was isolated at the Texas Children's Hospital in 2007. Significant differences to other MRSA strains lie within the plasmid content and the antibiotic susceptibility profiles⁵⁸.

MW2 is another community-acquired MRSA isolate. It carries a wide range of virulence and resistance genes⁶⁰. At the moment, more than fifteen different pathogenicity islands are identified in *S. aureus*. Interestingly, MW2 contains almost the same complement of pathogenicity islands as USA300-TCH1516⁵⁸. In contrast, the *S. aureus* strain COL contains six pathogenicity islands, such as Mu50, but in different combinations⁵⁸. COL is one of the first MRSA

isolates from the early 1960s. It is a penicillinase-negative strain^{61,62}. In contrast to the highly virulent MW2 strain, where virulence factors are found outside of prophages, fewer virulence factors are found outside of prophages in *S. aureus* strain Newman. This strain carries four integrated prophages and two large pathogenicity islands with important contributions for the pathogenesis. This *S. aureus* strain is susceptible to methicillin⁶³.

As the Newman strain, the *S. aureus* isolate MSSA476 is a methicillin-susceptible clone. It is a community-acquired strain, such as MW2. It was isolated in 1998 and susceptible to most commonly used antibiotics, excluding penicillin and fusidic acid⁶⁴. In contrast MRSA252 is a clinically important hospital-acquired MRSA lineage. It is genetically diverse to other *S. aureus* strains⁶⁴.

S. aureus does not only infect humans, it is also the cause of a mastitis in cattle. Strain RF122 contains genomic features that distinguish the human and the bovine pathogens⁶⁵.

Eight different *S. aureus* isolates belong to the South German clone lineage ST228. This clone spread over 10 years in a hospital in Switzerland. The isolates were collected between 2001 and 2008. The eight isolates represent the evolutionary history of the clone. As many others, it is an MRSA⁶⁶.

MODEL IMPROVEMENTS

A variety of different *S. aureus* models from various strains is available. However, not all 114 downloaded *S. aureus* models were of the same quality: Some SBML files were syntactically invalid, others utilized an older SBML format, or were not available as SBML file at all. To provide a collection of usable and updated SBML models, we performed debugging and/or improvement steps on some of the models. Models with valid SBML files of the latest level were not improved. All debugging and improvement steps served the purpose of standardizing and annotating the models. No content changes were performed that affect model calculations.

GEMs by Bosi et al.

The 64 *S. aureus* models by Bosi et al.³² were downloaded and evaluated using COBRAPy³⁵. The built-in validity check for SBML files returned a number of errors. In a first step, a pipeline for debugging the errors was created. All files lacked the XML declaration, which was added together with the XML version number and the encoding attribute. According to the SBML language specifications, metabolite, reaction, and model identifier need to fulfill certain properties⁶⁷, e.g., model identifiers cannot start with a number. The identifiers were adapted according to the guidelines. The downloaded SBML file contained an empty compartment list, which was filled with the compartments during the debugging. As the compartment list comprises all cellular compartments in which metabolites and reactions occur, the

Table 1. Overview over the available *S. aureus* models.

Model	Year	Model count	Availability	Format	MEMOTE score	Initial growth	Curation
iSB619 ²⁷	2005	1	BiGG & BioModels Database	SBML L3V1 with fbc	67%	✓	m
iMH551 ²⁸	2005	1	BioModels Database	SBML L3V1 with fbc	36%	✓	m
Lee ³¹	2009	13	Supplements	Excel spreadsheet file	65%		s
Path2Models ²⁹	2013	33	BioModels Database	SBML L2V4	48%		a
Bosi ³²	2016	64	Supplements	SBML L3V1 with fbc	36%	✓	s
Magnúsdóttir ³³	2017	1	Virtual Metabolic Human (VMH) Database	SBML L3V1 with fbc	45%	✓	s
iYS854 ³⁰	2019	1	BiGG Models Database	SBML L3V1 with fbc	81%	✓	m

All GEMs were downloaded from the respective database or from the supplements of the publication. Their format and SBML version were determined. The initial growth was tested (indicated with the symbol ✓ in case of success) and the MEMOTE³⁶ score was calculated for each model. The information for the models by Bosi et al. was determined after the debugging steps (see Fig. 3 steps 1–7). These debugging steps only served the purpose of generating valid SBML files. No additional improvements, which could increase the MEMOTE score, were performed at this point. The curation column indicates, whether the model was curated manually (m), automatically (a), or semi-automatically (s).

different compartments were extracted from the metabolites' information and subsequently incorporated into the compartment list. The charges, chemical formulas, and compartments of the models' metabolites were adapted or added, where necessary. After these debugging steps, the models were exported as valid SBML files and evaluated with MEMOTE. The MEMOTE score of 36% in Table 1 is the score after these debugging steps, since MEMOTE requires a syntactically valid SBML file as input.

Since a pipeline for altering all 64 *S. aureus* GEMs already existed, we added further steps to the pipeline to extend the models with respect to their annotations. With the use of the Systems Biology Ontology (SBO), semantic information about model components can be provided. This information allows an explicit and unambiguous understanding of the components' meaning⁶⁸. For the model genes and metabolites, appropriate SBO terms were defined. Reactions were divided into metabolic and transport reactions, each receiving different SBO terms. Transport reactions were even further refined to active, passive, or co-transport with antiporters or symporters. After the assignment of appropriate SBO terms, further annotations were added using ModelPolisher⁶⁹. ModelPolisher accesses the BiGG Models Database for the annotation and autocompletion of SBML models⁶⁹. With the help of the ModelPolisher, additional metadata was incorporated for the different model instances. After those extensions, the MEMOTE score of the 64 GEMs increased on average to 83 ± 1%, which is an average improvement of 47%. The complete pipeline for debugging and extending all 64 models and saving them as valid SBML files is summarized in Fig. 3.

All debugging and extension steps served the purpose of making the models simulatable. Since reaction-sets, metabolite-sets, or gene-sets were not altered, the models' simulation behavior is not affected. However, the models can now directly be used, as they are now all available as valid SBML files.

GEMs by Lee et al.

The thirteen GEMs by Lee et al. were available as Excel spreadsheet. For all reactions and metabolites in the list, the respective information, such as reaction or metabolite name, or chemical formula was extracted from the KEGG database³⁷, where available. Based on the information from the KEGG database and the Excel spreadsheet, a consensus model including all reactions was created. Both biomass objective functions from Becker et al.²⁷ and Heinemann et al.²⁸ were added to the consensus model, as well as exchange reactions for all extracellular metabolites. ModelPolisher⁶⁹ was used for annotating the model. Based on this consensus model, the individual models of the thirteen *S. aureus* strains were curated: The strain-specific reactions listed in the Excel spreadsheet were added to the respective models, and

the biomass objective function from Becker et al. was adapted strain-specifically. The KEGG database was browsed for the strain-specific gene identifiers. The models now include on average 491 ± 8 genes, except for *S. aureus* strain RF122, where no strain-specific KEGG gene identifier was available. Further annotations, such as KEGG annotations and EC-codes were added to the models. Despite manual effort, all thirteen models do not show growth for neither of the biomass objective functions. The MEMOTE score for all models excluding the model for the *S. aureus* strain RF122 reached 66%. Since the GEM for the RF122 strain does not contain any genes, its MEMOTE score only adds up to 57%. Comparing the originally published models concerning model simulations and growth predictions is not possible, because only Excel spreadsheets with reactions and metabolites were available.

GEMs from Path2Models

The 33 models from the Path2Models project are the only models of *S. aureus* that are still SBML Level 2 Version 4⁷⁰. Since the fbc package is officially only available from Level 3, it is not yet integrated in the files. We updated all models to SBML Level 3 Version 1⁴⁰ with the fbc package enabled using libSBML⁷¹. However, the original chemical formulas did not match the scheme that the official fbc package⁷² requires. In order to avoid creating syntactically invalid SBML files, all chemical formulas needed to be adapted according to the fbc specification⁷². The original chemical formulas can still be found in the `notes` field. This `notes` field further contained a variety of annotations from different databases, including BRENDA⁷³, KEGG³⁷, MetaCyc⁴⁴, MetaNetX⁴⁶, Rhea⁷⁴, BiGG²⁵, Reactome⁷⁵, Model SEED⁴⁹, Unipathway⁷⁶, the Human Metabolome Database (HMDB)⁷⁷, ChEBI⁷⁸, and InChI⁷⁹. All database annotations that can be found in the identifiers.org⁴⁷ registry were transferred to the annotations, using identifiers.org uniform Resource Identifiers (URIs). The service identifiers.org provides directly resolvable identifiers from a multitude of different databases. The final and valid SBML files were evaluated using MEMOTE. The total score for the GEMs from the Path2Models project increased from 48 to 59% and all models are now available as SBML Level 3 files. Again, no changes on the reaction, metabolite, or gene content were performed, which would affect the model simulations.

MODEL ANALYSIS

In the following section, we examined the available models for their predictive value and their similarity. As the growth behavior of *S. aureus* is reported in various defined media, the models' capability of reflecting growth under these conditions indicates

B.4 Curating and comparing 114 stain-specific GEMs of *S. aureus*

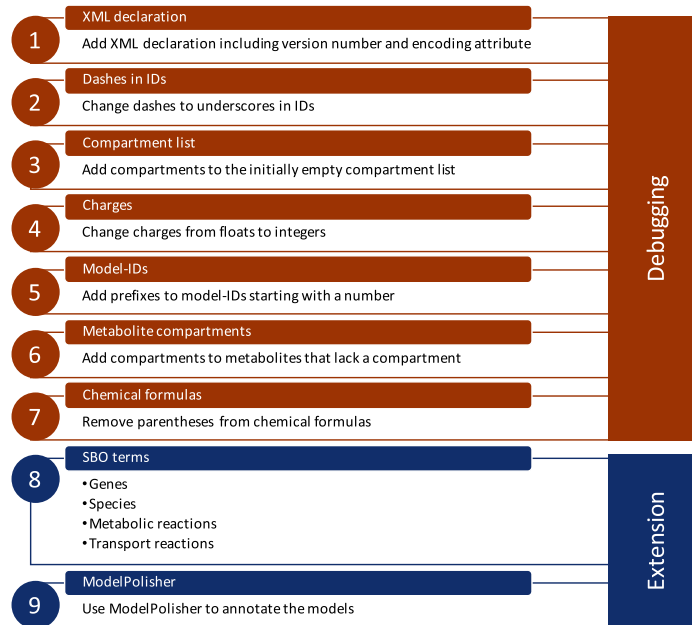


Fig. 3 Debugging and extension steps in GEMs by Bosi et al. Not all the 64 SBML files downloaded from the supplement of Bosi et al. did directly pass the syntactic validation. In seven steps, the errors reported in the validity check were solved to receive valid SBML files. The valid files were then further extended with appropriate SBO terms for genes, species, and reactions. In a final step, annotations were added to the model using ModelPolisher⁶⁹.

the predictive value of the model. Subsequently, the publications were checked for the inclusion of experimental data in the models or the verification of model-driven hypothesis. Additionally, the predictions of gene essentialities using different models are compared. In the last step, the models' similarities were examined concerning their reaction and gene content.

Growth capabilities

The growth of genome-scale metabolic models on different media is an important characteristic of a model's capabilities and flexibility to reflect the organisms behavior in different environments. Since *S. aureus* is known to grow in a variety of different environments, its growth was simulated in chemically defined environments to investigate the model's capabilities.

Chemically defined medium (CDM). The CDM is a complete defined medium with 18 amino acids, two purines, and six vitamins and initially developed to study the slime production by coagulase-negative staphylococci⁸⁰. It was used by Halsey et al. to study the amino acid catabolism in *S. aureus*⁸¹. Either no carbon source was added (CDM), or glucose (CDM_glc) or galactose (CDM_gal) was added to the medium. The growth of *S. aureus* strain JE2 is already computationally and experimentally validated and verified on CDM and its variants³⁰.

Synthetic nasal medium (SNM). The primary ecological niche of *S. aureus* is the human nose^{82,83}. Krismer et al. developed a defined synthetic nasal medium (SNM) based on the composition of nasal fluid components determined by metabolomics^{84,85}. This medium was initially developed to monitor the growth of *S. aureus* under similar physiological conditions as in the nose. Growth in

this medium is experimentally verified for the *S. aureus* strains USA300 LAC and Newman. Since the medium is chemically defined, it can also be established in growth simulations in systems biology.

Gut medium. Already in the 1950s and 1960s, the intestinal colonization of *S. aureus* was reported⁸². Recent interest in the gut microbiome revealed and enlightened the relevant role and influence of *S. aureus* on the intestinal microbial ecology and diversity^{83,86–89}. Intestinal colonization by *S. aureus* is, e.g., assumed to induce pseudo-membranous colitis and to change the gut microbial ecology⁸⁹. Alterations in the composition of the gut microbiota can result in the development of chronic diseases, such as type 2 diabetes, colorectal cancer, and obesity⁹⁰. Hence, studying the role of *S. aureus* in the context of the gut microbiome is of high relevance. Magnúsdóttir et al. generated 773 genome-scale metabolic reconstructions for 773 members of the human gut microbiome, including *S. aureus*. To simulate the growth in the gut, they chemically defined a medium according to experimental data. The medium definition was extracted from the *S. aureus* model created by Magnúsdóttir et al.³³. Magnúsdóttir et al. validated two of the 773 genome-scale reconstructions experimentally, where *S. aureus* was not included. However, as their model grew in the defined medium, and *S. aureus* is reported to colonize the intestine, we inferred that growth should be possible.

SCFM. *S. aureus* does not only occur on the human skin, in the human nose^{82,83}, or the nasopharyngeal tract^{91,92}. It is furthermore observed in patients, especially in children, with cystic fibrosis (CF)⁹³, an autosomal recessive disease. As one of the earliest and also most prevalent pathogens, *S. aureus* causes chronic airway infections in patients with CF⁹⁴. To investigate the role of *S. aureus*

and other associated pathogens, such as *Pseudomonas aeruginosa*⁹³, Palmer et al. developed a synthetic cystic fibrosis medium (SCFM), mimicking the nutritional composition of the sputum of patients with CF by chromatographic and enzymatic analyses of the CF sputum. This medium was initially created to analyze the nutritional behavior of *Pseudomonas aeruginosa* in CF sputum⁹⁵. Clinical isolates of *S. aureus* are reported to grow in SCFM⁹⁶.

Since the thirteen models by Lee et al. and the 33 models from the path2models project did not exhibit any growth in full medium, these models were not included in the analysis of growth capabilities. During the analysis, three of the models by Bosi et al. reported a low growth rate of 0.00186 mmol/(gDW × h) without any active exchange reactions (models *sa_118*, *sa_gxl*, and *sa_lct*). A positive growth rate without active exchange reactions can be an indicator for futile cycles and a necessity for manual verification and refinements.

Not all models by Bosi et al. were capable of growing on any of the tested media. In total, 33 out of the 61 remaining models were not able to grow on any of the tested media. This might be explained by the auxotrophies for amino acids and vitamins in several *S. aureus* strains observed by Bosi et al.³². None of the models by Bosi et al. grew on the SCFM or the gut medium. For SNM and the CDM compositions, different patterns emerged: ten strains, including N315, only grew on SNM, while six strains only grew on the CDM with galactose. Seven strains grew on all three variants of the CDM and the remaining five strain models grew on both the SNM and all CDM. The model *iSB619* by Becker et al. only grew on the gut medium, while the model *iMH551* by Heinemann et al. returned a positive growth rate for all tested media types. The model *iYS854* exhibits growth on almost all tested media, except for the SCFM and the gut medium. It is comparable to the models by Bosi et al., with the difference of a higher growth rate. The model by Magnúsdóttir only grew on its own gut medium. In Fig. 4, the growth capabilities of the various *S. aureus* GEMs under different environmental conditions is illustrated.

Presence of experimental data

Besides the correct prediction of growth in a defined environment, a model's predictive value also increases when laboratory data is included or in silico observations are verified in laboratory experiments.

Automatically curated GEMs. The models from the Path2Models project were automatically constructed. Within automated reconstruction processes, the inclusion of experimental data for individual models is complicated. For this reason, the GEMs from the Path2Models project do not contain experimental data²⁹. Moreover, the models are not simulatable and, thus, can also not predict any growth. Verification of model predictions is hence not possible.

Semi-automatically curated GEMs. Curating a collection of multiple GEMs is time and labor intense. Manual reconstruction would take a significant amount of time. Thus, the models from Lee et al., Bosi et al., and Magnúsdóttir et al. were constructed semi-automatically.

Lee et al. verified their models using gene essentiality analysis and growth experiments of two models. They found literature evidence and experimental verification for six of the 44 identified genes that were essential in all strains in silico. The growth experiments supported their minimal-medium predictions³¹.

The models from Bosi et al. were examined for the correct simulation of already known auxotrophies. Furthermore, the predictions of the growth capability in the presence of spermidine, and the growth on chemically defined media were verified in laboratory experiments for several strains³².

The model from Magnúsdóttir et al. was curated based on literature-derived experimental data. However, it is not specified which experimental data is used exactly. Metabolic predictions of two of the 773 reconstructions were validated against experimental data³³.

Manually curated GEMs. Becker et al., Heinemann et al., and Seif et al. manually curated their strain-specific GEMs. The in silico growth predictions of the model *iSB619* in a minimal medium were compared to laboratory experiments. Becker et al. additionally predicted essential genes. As this was the first available GEM of *S. aureus*, no experimental data was available to compare the predicted essential genes with²⁷. The model *iMH551* was compared to available knowledge about auxotrophies in *S. aureus*. The model's growth predictions under aerobic and anaerobic conditions were validated against available experimental evidence²⁸.

The model *iYS854* underwent the most experimental verifications compared to all other models. Its predictions are in 85%

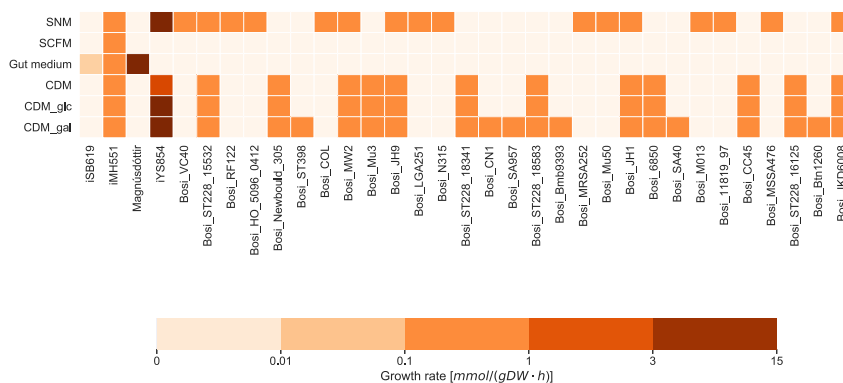


Fig. 4 Growth rate of *S. aureus* GEMs in different media. All models with initial growth (see Table 1) were tested on the different media. For the GEMs by Bosi et al.³², the prefix “Bosi” was added to the model strain name. The other models are named according to their published model ID or, in case of the model from the VMH database, by the author. Media types are the synthetic nasal medium (SNM), synthetic cystic fibrosis medium (SCFM), gut medium, chemically defined medium (CDM), CDM with glucose (CDM_glc), and CDM with galactose (CDM_gal). Models that did not show growth in any of the tested media were excluded. The color-bar indicates the growth rate: the darker the color, the higher the growth rate of the model organism on the given medium.

agreement with gene essentiality experiments. The in silico predictions of the catabolism of carbon sources are in 68% agreement with experimental physiological data. They compared the models' growth predictions on various media with laboratory experiments, and performed extensive condition-specific GEM validation and evaluation in the presence and absence of glucose.

Prediction of gene essentialities

Another indicator for the predictive value of a model is the correctness of predicted gene essentialities. The essentiality of a gene depends on the environment and the availability of nutrients. To identify essential genes in silico, each gene is individually knocked out in a so-called single gene deletion analysis and its effect on the growth rate is evaluated. This analysis, however, requires a model's capacity to simulate growth in the investigated environment. As the models from the Path2Models project and Lee et al. did not show any initial growth (see Table 1), these models were excluded from the single gene deletion analysis. Additionally, this review aims to compare models from different sources. Since the models from the Path2Models project and Lee et al. were already excluded from this analysis, only two strains remain with more than one model: *S. aureus* USA300-FPR3757 and *S. aureus* N315. Two models from Bosi et al. and Magnúsdóttir et al. are available for the strain USA300-FPR3757, which can simulate growth. The model from Magnúsdóttir et al. contains gene identifiers that cannot be resolved within the PATRIC database⁹⁷, leading to its exclusion from this analysis. With only one remaining model from Bosi et al., a comparison of predicted gene essentialities for the strain USA300-FPR3757 is not possible anymore.

Becker et al., Heinemann et al., and Bosi et al. curated models for the strain N315 simulating growth. The model from Heinemann et al., however, had to be excluded from the single-gene-deletion analysis as the model did not contain any GPRs and, thus, no genes. We downloaded the list of 302 essential genes for N315 from the Database of Essential Genes (DEG)⁹⁸ and mapped all genes to the respective KEGG gene identifier. The medium is indicated as a rich medium in the DEG, but no further description of the chemical definition is given. Therefore, all exchange reactions were opened for the single gene deletion analysis.

The model from Bosi et al. predicted 117 essential genes, while the model from Becker et al. predicted 80. Of the 302 essential genes from the DEG, only 176 and 107 genes were present in the models from Bosi et al. and Becker et al., respectively. From the 117 predicted essential genes by Bosi et al., 27 (23.1%) were predicted correctly, while 90 (76.9%) of the predicted essential genes are not in accordance with the experimentally derived essential genes. Similarly, from the 80 predicted essential genes by Becker et al., 18 (22.5%) were predicted correctly, while 62 (77.5%) of the predicted essential genes are not listed in the DEG. One possible explanation for the similar predictions of essential genes is that the models from Bosi et al. are based on the model from Becker et al. The low number of true positive predicted essential genes could indicate further refinement potential of the two models.

Similarities between models

The analysis of the growth capabilities implied a clustering of models with similar growth behavior, especially for the models by Bosi et al. To identify further similarities between the models, the reaction sets were compared. Mapping identifiers between different databases induces a bias, since a complete mapping is currently not feasible. Tools, such as ModelPolisher³⁹ can be helpful for annotating and comparing models. However, these tools rely on cross-references in various databases, which holds some challenges: The tools can only search with the correct

identifier; if a model, however, has identifiers not included in the database, the tools will not find any annotations for that model instance. One other challenge lies within the administration and topicality of the databases. Changes in one database might not be reported or updated in the cross-references of other databases, leading to erroneous allocations that would bias the result of the comparison.

Heat maps of reaction similarity. Since the models have diverging identifiers, we divided them into three different groups. The first group comprises the 33 models from the Path2Models project with consistently mixed identifiers. The second group includes all thirteen models by Lee et al. with KEGG IDs. The third group includes all models with BiGG identifiers, namely all models by Bosi et al., as well as the models *iSB619* and *iYS854*. Furthermore, this third group contains the model created by Magnúsdóttir et al. This model possesses VMH identifiers, however, those identifiers can easily be converted to BiGG identifiers since they bear a resemblance to the BiGG IDs. Within these groups, all reactions were listed and checked for their occurrence in the models. With this table of reaction occurrences, the Jaccard distance was calculated between all pairwise combinations of the models.

With this distance matrix, the heat-map in Fig. 5 was created. The models *iSB619*, Magnúsdóttir, and *iYS854* vary widely between each other and the models by Bosi et al. Within the Bosi models, clusters of more and less similar models can be identified (Fig. 5c). Such clusters are expected, as we assumed that genetically similar strains also lead to more similar GEMs, due to the gene–protein–reaction associations (GPRs). For example, the two closely related USA300 strains TCH1516 and FPR3757 have a distance value of 0.015, while the distance to one of the isolates of the ST228 lineage (ST228-16035) is 0.160. Strain MRS252 is reported to be genetically diverse compared to other *S. aureus* strains. Its distance, however, to the USA300-TCH1516 strain is smaller (0.06) than the distance to the isolates of the ST228 lineage. Hence, the genetic differences between the different strains are not necessarily reflected in their respective GEMs so far.

The distances between the models with BiGG IDs (group three) ranged from 0 to 0.8, with the maximal distances between the models *iSB619*, Magnúsdóttir, and *iYS854*. The models by Lee et al., however, are more similar, indicated by the scaling of the color-bar that ranges from 0 to 0.05. The model of the *S. aureus* strain TCH1516 differs the most from all other models (Fig. 5b). Unlike the models from Bosi et al., the two USA300 strains (TCH1516 and USA300) do not cluster. They have a distance of 0.037. In contrast to the models of Lee et al., the strain TCH1516 does not stand out in the groups with BiGG IDs and the Path2Models models.

Most distances between the models from the Path2Models project (group one), ranged from 0.25 to 0.35. However, the model of strain RF122 protrudes with a mean distance of 0.62. This trend can also be observed in the heat-map of the models by Lee et al., but not as prominent as in Fig. 5a. One possible explanation is given in the taxonomy for the *S. aureus* strain RF122, which is an bovine mastitis-associated isolate with notable differences to human clones of *S. aureus*⁹⁹. This difference is, however, not as obvious in the Models of Bosi et al. compared to the models of Lee et al. and the Path2Models project.

Venn diagrams of gene similarity. Despite significant effort to standardize and consistently annotate all models using different annotating tools, such as the ModelPolisher, or database requests for aliases from databases like BiGG or ModelSEED, a satisfying comparison of the reaction sets between different identifiers is still not possible. For example, for the models with KEGG identifiers from Lee et al., we could not use the ModelPolisher, as this annotation tool currently requires BiGG identifiers. For that reason, we browsed the BiGG Models Database locally for cross-references

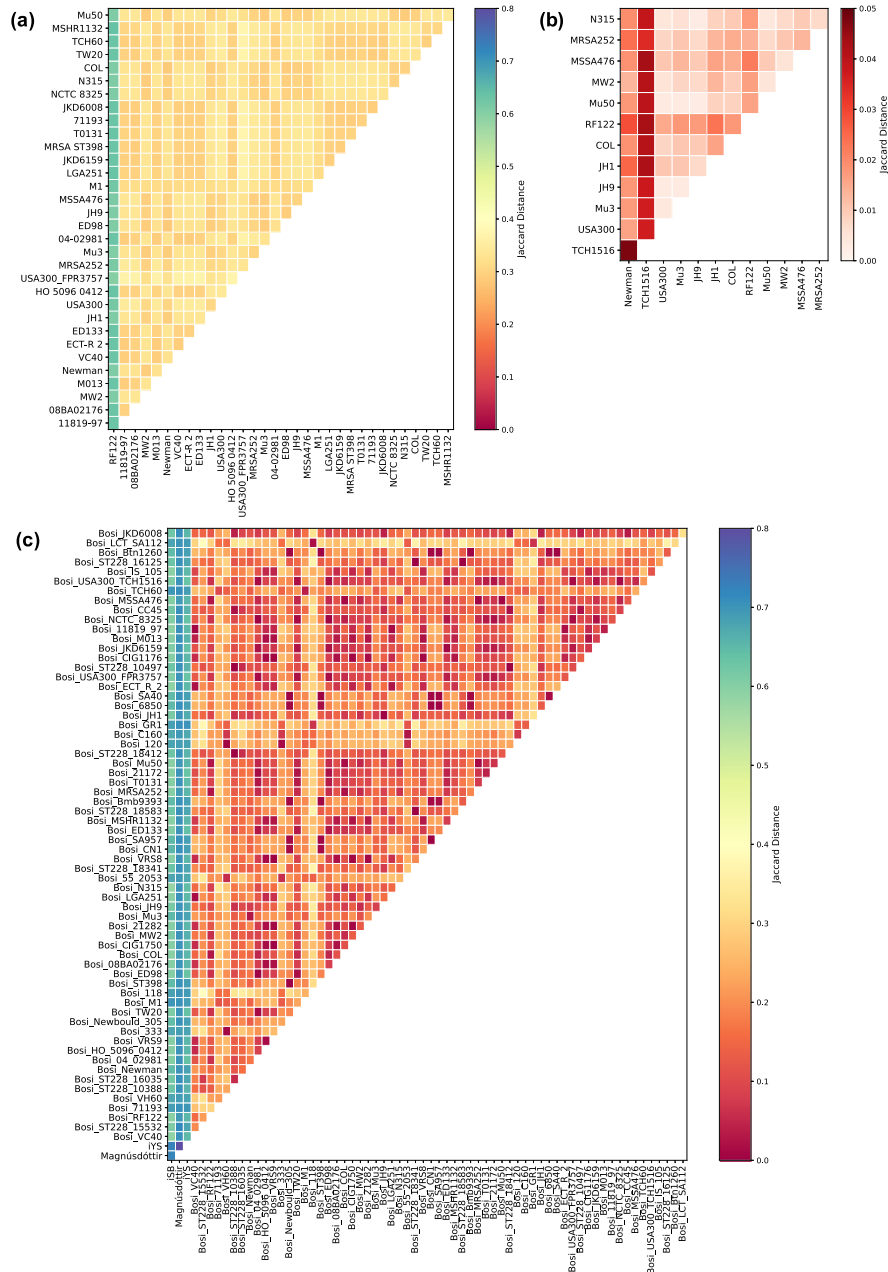


Fig. 5 Model comparison based on Jaccard distance between reaction sets. The models were divided into three groups based on their metabolite and reaction identifiers: (a) has all models of the Path2Models project with consistently mixed identifiers, (b) has all models with KEGG identifiers (hence, all GEMs by Lee et al.), and (c) contains all models with BiGG identifiers. Within the three groups, all pairwise Jaccard distances were calculated based on the models' reaction sets. The distances are displayed in the heat map. The color bar range is equal for (a) and (c) for better comparison. As the distances in (b) are much smaller, the color bar's range was adapted.

to KEGG identifiers. Unfortunately, 842 out of 1486 KEGG reaction identifier were not referenced at all in BiGG, 359 KEGG identifiers were not uniquely mapped to a BiGG identifier, and only 285 identifiers were uniquely mapped. We checked some of the non-referenced KEGG identifiers in the ModelSEED database for aliases but could not determine the respective identifiers.

For that reason, we looked at the gene content of the models. Most models used KEGG gene identifiers, regardless of the identifier database of the reactions and metabolites. As the different strains have strain-specific gene identifiers, the following analysis was conducted strain-wise. Strains with at least three models from various resources were taken into account (see also Fig. 2): For eleven strains, three models are available, for the strain USA300-FPR3757, four models are present in this collection, and for the strain N315, five models are available. However, the SBML file of the N315 model by Heinemann et al. does not include any genes. Thus, the model was excluded from the comparison. Same accounts for the RF122 strain-specific model by Lee et al., which also does not contain any genes. For this reason, the model was also excluded from the analysis. By that, the strain RF122 did no longer fulfill the criterion of at least three available models.

The gene sets from the remaining models were compared. As indicated, most models used KEGG gene identifiers, but not all. The model by Magnúsdóttir et al. included strain-specific and unspecific PATRIC identifiers³⁷. With the help of the PATRIC ID mapping service, the respective KEGG gene identifiers were extracted. However, this was only feasible for the strain-specific identifiers. Despite significant effort, the unspecific identifiers could not be resolved, as no mapping scheme could be identified. Thus, from the 859 genes included in the Magnúsdóttir model, only 192 could be resolved to KEGG identifiers.

Model iSB619 contained new locus tags, whereas the KEGG identifiers correspond to the old locus tags. With the GenBank flat file (gbff)¹⁰⁰ of *S. aureus* strain N315, the locus tags were mapped. For the 619 new locus tags 611 respective old locus tags, and thus KEGG identifiers, were extracted.

The models by Bosi et al. included mostly KEGG gene identifiers. Within the strains JH1 and JH9, the gene identifiers were truncated by the included word "DRAFT" to make them consistent with the actual KEGG identifiers. For example, the initial identifier *SaurJH1DRAFT_0595* was truncated to the correct KEGG identifier *SaurJH1_0595*.

After these mapping and adapting steps, the gene sets within the different strains from the different resources were compared, and Venn diagrams were created as shown in Fig. 6. Across all twelve comparisons, the models by Bosi et al. have the largest portion of genes that are solely reflected in these models. This number varies between 20.1% in the N315 strain and 59% in the Newman strain. As these models have the highest gene content on average with approximately 788 ± 116 genes per model, this seems apparent. The models from the Path2Models project have an average gene content of 519 ± 12 genes per model, and the models by Lee et al. contain 488 ± 149 genes on average. It was already mentioned that the gene identifiers from the JH1 and JH9 models by Bosi needed to be adapted. Despite this adaptation, only half of the gene content is present in the other models as well. For the Newman, MW2, and Mu3 strains, we further analyzed the gene identifiers after these observed discrepancies between the gene contents with the models from the other two databases. These three strain-specific models from Bosi include non-strain-specific gene identifiers, which could not be mapped to the corresponding strain-specific gene identifier.

The models from Lee et al. and the Path2Models project are relatively similar concerning their gene content. Since both models are curated based on the KEGG database, this similarity is evident. The four models of the *S. aureus* USA300-FPR3757 strain have a gene content overlap of 15.7%. The model by

Magnúsdóttir et al. has only 0.3% gene content that is not reflected in the other three models. However, one needs to keep in mind that many genes in the model are not strain-specific and could not be mapped and compared.

With these twelve gene content comparisons, we again calculated the Jaccard distance between the models from Bosi et al., Lee et al., and the Path2Models project. As already visible from the Venn diagrams, the models from Lee and the Path2Models project are most similar with respect to their gene content. They have a mean Jaccard distance of 0.288 ± 0.004 . However, one might have speculated that the models are more similar based on the Venn diagrams. It needs to be highlighted that the Venn diagrams are calculated based on the gene content of all compared models. In contrast, the Jaccard distance calculates pairwise distances and, thus, only considers two models at once. For that reason, the models from Lee et al. and the Path2Models project are still the most similar ones, but their identity might not be as large as first expected when looking at the Venn diagrams. The Bosi models have a mean distance to the Lee models of 0.666 ± 0.179 and to the Path2Models project models a mean distance of 0.616 ± 0.203 .

Although the different models from the various databases reflect the same strain, the models have distinct diversities. This can be explained by the differences in the reconstruction process. How the model is curated seems to play a pivotal role for the final model and its model instances. Thus, the reconstruction method needs to be chosen carefully, and manual or semi-automated additions might be required.

Decision guidance

With the vast amount of different strain-specific *S. aureus* models, the identification of the suitable GEM for a specific research question or purpose might become difficult. Table 2 gives an overview about the main features of the *S. aureus* GEMs. The features were assigned based on the strengths of the different models or model collections after the model improvement steps. If one is interested in simulatable models, the table guides the reader to the corresponding models. By combining different required features, the selection can be tailored. If one needs, e.g., a model with BiGG IDs that grows on different media, the models by Bosi et al. or the model iYS854 are suggested, depending on the desired strain. High MEMOTE scores indicate a high degree of annotations, which facilitates the re-usability and comparability of a model.

A predictive value score was calculated based on the model analysis regarding their growth capabilities and the presence of experimental data. If a model was not simulatable, it received a predictive value score of 0. Otherwise, a score of 1 was added. For growth capabilities in one environment, a score of 1 was added; for growth in multiple environments, 2 was added. For every experimental verification procedure, such as growth verifications, auxotrophies, compliance with physiological data, or other experiments, a score of 1 was added. The prediction of essential genes was not included in this score, as this analysis was only conducted for two models. By this scheme, the model iYS854 had the highest predictive value score of 7, followed by iMH551 and some models by Bosi et al. The models by Bosi et al. received a score between 3 and 5, as some models do not predict growth in any tested environment, while others do. As the models from Lee et al. and the Path2Models project are not simulatable, they received a predictive value score of 0. Models with high predictive value score and high MEMOTE score are recommended for further use, while models with low predictive value score might need further refinement and experimental verification before usage.

This table does not contain strain-specific information. Including the information from Figs. 2 and 4 will further guide the decision for a suitable model.

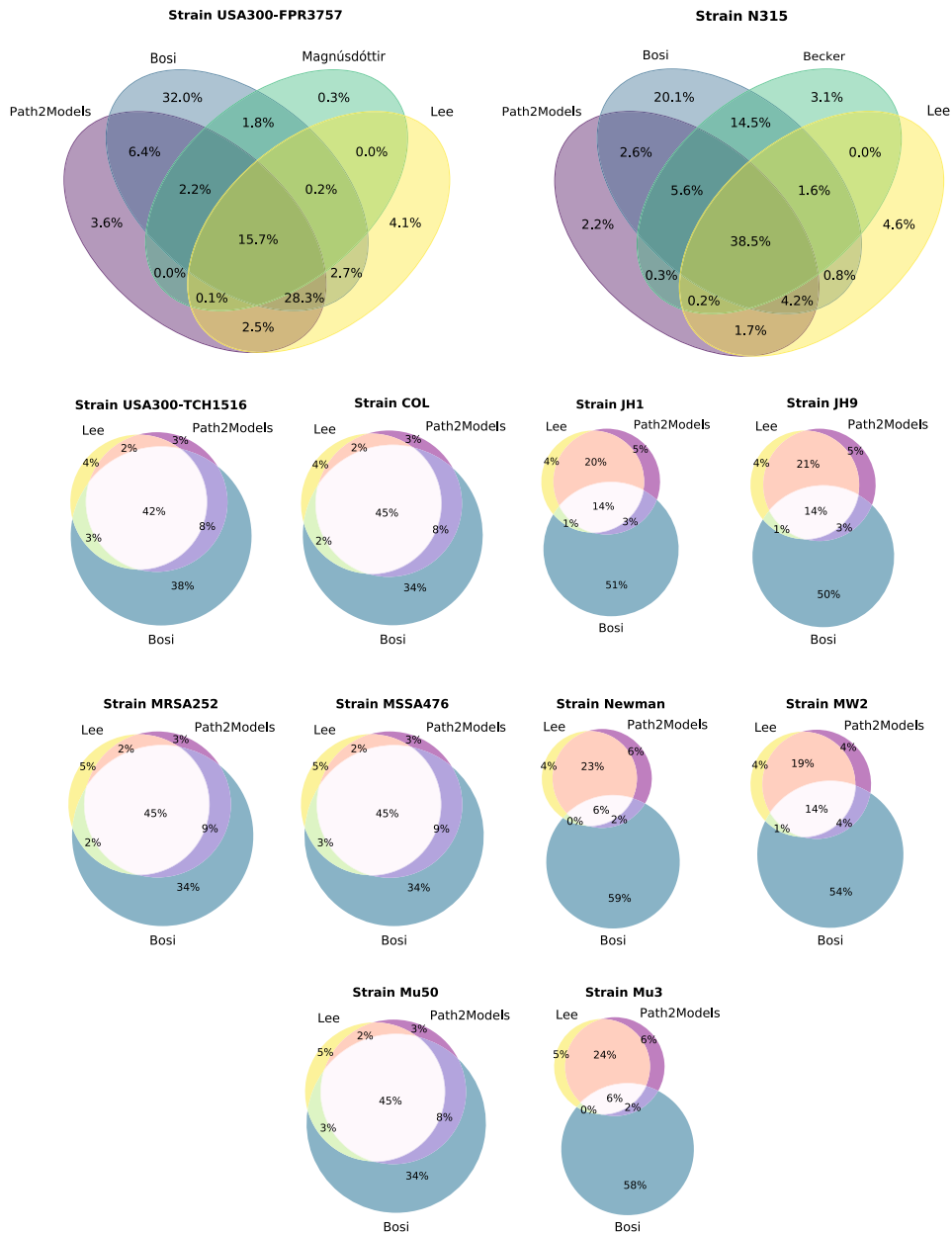


Fig. 6 Strain-specific model comparison based on gene sets. For all models occurring in at least three different resources, the gene content was compared strain-specifically. After unifying the gene identifiers to KEGG IDs, Venn diagrams were created comparing the gene content. The models from Bosi et al. have, on average, the highest gene content, explaining the large fraction of genes occurring only in these models. The models by Lee et al. and the Path2Models project seem more similar, which could be explained by the fact that both are curated based on the KEGG database. Although all models in one Venn diagram (and thus, one comparison) represent the same strain, the models have differences, indicating the influence of the reconstruction method on the final model content.

B.4 Curating and comparing 114 strain-specific GEMs of *S. aureus*

Table 2. Feature-based decision guidance.

Feature	iSB619	iMH551	Lee	Path2Models	Bosi	Magnúsdóttir	iYS854
Database				✓			
Simulatable models	✓	✓			✓	✓	✓
BiGG IDs	✓				✓		✓
Kyoto Encyclopedia of Genes and Genomes (KEGG) IDs			✓				
Growth on different media		✓			✓		✓
High MEMOTE score	✓		✓		✓		✓
Predictive value score	3	5	0	0	3–5	2	7

The main features of the *S. aureus* GEMs are listed and indicated with the symbol ✓ when present. The models are assigned to the features based on their strengths after the model improvement steps. A predictive value score was calculated as described in the section “Decision guidance”. With the help of the features and the predictive value score, one can identify the best suited model for the research question of interest.

DISCUSSION

The analyses show that despite genomic and genetic similarities, GEMs of related strains are not necessarily similar to each other. This accounts for both models of the same strain curated by different research groups and to related strains curated by the same group. One example is the model from Magnúsdóttir et al. with the *S. aureus* strain USA300-FPR3757 and the corresponding model from Bosi et al. Despite it is the same strain, the GEMs are quite different in their reaction content. In contrast, the two strain-specific models of the strains MRSA252 and USA300-TCH1516 by Bosi et al. are quite similar despite the genetic diversity of the strain MRSA252. This observation might have several reasons. The first, and probably most striking, reason is the incompleteness of the models. As high-quality genome-scale metabolic reconstructions require manual curation and evaluation¹⁰¹, and many models introduced in this review were created automatically or semi-automatically, some models might lack general or strain-specific reactions. This lack of required reactions is also visible when optimizing the flux distributions of the models. For multiple models, no growth could be simulated in FBA, not even in full medium. This was especially the case for the automatically curated models from the Path2Models project and the semi-automatically curated models from Lee et al. But also some of the semi-automatically curated models from Bosi et al. did not show any growth. Thus, a connection between automated or semi-automated curation and the functionality of the models seems to exist. However, automated or semi-automated curation does not necessarily result in poor growth prediction, especially when the basis for the (semi-) automated processes underwent significant manual curation. The other models from Bosi et al. showed growth on up to four different media. The semi-automatically constructed model by Magnúsdóttir et al. could be simulated on one medium, which is also the case for the manually curated model iSB619. Furthermore, some of the *S. aureus* strains have plasmids carrying additional genes. For a strain-specific model, these additional genes need to be incorporated into the GEM as well. Especially the metabolic and transporter genes are relevant for the strain-specific model. The plasmid of the *S. aureus* strain N315, e.g., carries a gene for the cadmium resistance transporter CadD, which facilitates the export of cadmium ions and other cationic compounds¹⁰². Besides further proteinogenic genes, the plasmid of strain N315 also carries a gene for the penicillin-hydrolyzing class A β -lactamase enzyme. These two genes are, e.g., also present on the plasmid of the *S. aureus* strain USA300-TCH1516.

As explained previously, the challenge lies within the different reaction and metabolite identifiers. In this review, we additionally tried to annotate the GEMs further to simplify the comparison of models with differing identifiers. However, only approximately

one third of all reactions and metabolites are annotated with identifiers of external databases. It is still challenging to find all cross-references for a particular metabolite or reaction in a specific database. For that reason, we additionally evaluated the gene content of the strain-specific models, as most models contained identifiers from the KEGG database. The gene identifiers from other databases were mapped to the KEGG identifiers. Again, a bias is introduced when identifiers are mapped between databases: On the one hand, not all identifiers can be resolved in the other database. On the other hand, some identifiers do not comply with the databases' identifiers scheme and do not have annotations. This makes an automated mapping of several hundred identifiers infeasible. Extensive manual labor would be necessary to map these identifiers. The usage of consistent identifiers that comply with the database scheme and additional annotations is highly recommended and would simplify the re-usability, translatability, and comparability of models¹⁰³. The comparison of the strain-specific models' gene content confirmed that GEMs from different resources could vary, despite their genetic equality, highlighting the relevance of the curation process on the resulting GEM. This observation is even more explicit when comparing the models by Lee et al. and from the Path2Models project: both rely on the KEGG database. However, the models are not equal, as the two groups used different approaches for the curation of the models.

Missing reactions and strain-specific genes might also affect the growth behavior of a strain-specific model on a given medium. Only the model iMH551 showed growth on all tested media. Additional growth experiments for specific *S. aureus* strains can help to identify the missing growth capabilities of the model. The model's ability to adapt to different environmental conditions is crucial to simulate an organism in silico. This is also reflected in the predictive value score, which was assigned to the models. Especially for models with a low predictive value score, additional experiments would help determine and also increase the predictive value of the model.

The models from Lee et al., the Path2Models project, Bosi et al., and Magnúsdóttir et al. are curated automatically or semi-automatically. Except for the models from Bosi et al., all models have a comparatively low predictive value score than the manually curated models. The models from the Path2Models project and Lee et al. have a score of 0. The low score from the Path2Models projects' models might go back to the lack of experimental data in both the curation and verification process, thus highlighting its importance for predictive genome-scale metabolic reconstructions. The low score for the models from Lee et al. accentuates the importance of standardized GEMs, which allow re-usability. Although the models from Bosi et al. are curated semi-automatically, their predictive value scores are comparable

high. They based their pipeline on a manually refined model and verified their predictions with experimental data. More experimental data accompany more knowledge. The latest model, *YS854* has the highest predictive value score, was manually curated, and extensively experimentally validated. The result of such a time- and labor-intensive work is a GEM with a high predictive value and a strong recommendation for future usage.

CONCLUSION AND OUTLOOK

In this review, all 114 currently available genome-scale metabolic models (GEMs) of *Staphylococcus aureus* were presented and evaluated. It serves as guide for the different available reconstructions in various databases, using differing metabolite and reaction identifiers. Some models originally comprise a large number of reactions, metabolites, and genes, after undergoing several manual curation steps and extensive annotating. Such models have a high MEMOTE score. The model with the highest MEMOTE score is the *YS854* model by Seif et al. Other models have a vast amount of reactions and metabolites, such as the reconstructions of the Path2Models project. Such models could, e.g., serve as information sources for the reconstruction or refinement of already existing strain-specific models. Based on the information regarding availability, model format, MEMOTE score, growth behavior, used database identifiers, predictive value, and similarities between models, together with a previously defined research question, the appropriate genome-scale reconstruction can be chosen from the vast amount of available GEMs. Another approach would be to use the strengths of every reconstruction and incorporate it into merged or combined models, which increase the correctness and the predictive value of a strain-specific model. Despite the vast amount of presented models in this review, there is no suitable model for every *S. aureus* strain available. Furthermore, missing annotations or identifiers that do not comply with the database identifier scheme impede the models' re-usability and comparability. Standardization of all models would be desirable but is currently not feasible with the available tools without extensive manual labor for hundreds of identifiers. No omics data was incorporated into many of the published GEMs so far. Information about transcription profiles, for example, can help to refine metabolic reconstructions to better reflect the metabolic state of an organism in a defined environment. The incorporation of omics data can thus increase the predictive value of genome-based metabolic reconstructions¹⁰⁴.

However, with the help of the already available reconstructions and further information, strain-specific models could be created or extended. Information from literature, merging of strain-specific models, and manual curation steps could further improve the predictive value of simulations and analyses of metabolic features of *S. aureus*. Having predictive GEMs can eventually lead to the identification of novel targets for antimicrobial therapies in the fight against antibiotic resistant strains of *S. aureus*.

DATA AVAILABILITY

The availability of all models, including the improved models, is listed in the supplementary Table S1. The model collection was deposited in BioModels¹⁰⁵ within COMBINE archive files (in OMEX format)¹⁰⁶ and assigned the identifiers (1) [MODEL2007110001](https://doi.org/10.1093/nar/gkz1054/5614178), (2) [MODEL2007150001](https://doi.org/10.1093/nar/gkz1054/5614178), and (3) [MODEL2007150002](https://doi.org/10.1093/nar/gkz1054/5614178).

CODE AVAILABILITY

All the necessary scripts and resources for model modifications and improvements are available in a git repository at github.com/draeger-lab/S_aureus_GEMs_Collection.

Received: 16 October 2020; Accepted: 25 May 2021;
Published online: 29 June 2021

REFERENCES

- Sakr, A., Brégeon, F., Mège, J. L., Rolain, J. M. & Blin, O. *Staphylococcus aureus* nasal colonization: an update on mechanisms, epidemiology, risk factors, and subsequent infections. *Front. Microbiol.* **9**, 2419 (2018).
- Schmidt, A., Bénard, S. & Cyr, S. Hospital cost of staphylococcal infection after cardiothoracic or orthopedic operations in France: a retrospective database analysis. *Surg. Infect.* **16**, 428–435 (2015).
- Turner, N. A. et al. Methicillin-resistant *Staphylococcus aureus*: an overview of basic and clinical research. *Nat. Rev. Microbiol.* **17**, 203–218 (2019).
- Tacconelli, E. et al. Discovery, research, and development of new antibiotics: the WHO priority list of antibiotic-resistant bacteria and tuberculosis. *Lancet Infect. Dis.* **18**, 318–327 (2018).
- van Hal, S. J. et al. Predictors of mortality in *Staphylococcus aureus* bacteremia. *Clin. Microbiol. Rev.* **25**, 362–386 (2012).
- Laupland, K., Ross, T. & Gregson, D. *Staphylococcus aureus* bloodstream infections: risk factors, outcomes, and the influence of methicillin resistance in Calgary, Canada, 2000–2006. *J. Infect. Dis.* **198**, 336–343 (2008).
- Klevens, R. M. et al. Invasive methicillin-resistant *Staphylococcus aureus* infections in the United States. *J. Am. Med. Assoc.* **298**, 1763–1771 (2007).
- Friedman, N. D. et al. Health care-associated bloodstream infections in adults: a reason to change the accepted definition of community-acquired infections. *Ann. Intern. Med.* **137**, 791–797 (2002).
- Dantes, R. et al. National burden of invasive methicillin-resistant *Staphylococcus aureus* infections, United States, 2011. *JAMA Intern. Med.* **173**, 1970–1979 (2013).
- Kourtis, A. P. et al. Vital signs: epidemiology and recent trends in methicillin-resistant and in methicillin-susceptible *Staphylococcus aureus* bloodstream infections—United States. *MMWR* **68**, 214–219 (2019).
- Eells, S. J. et al. Persistent environmental contamination with USA300 methicillin-resistant *Staphylococcus aureus* and other pathogenic strain types in households with *S. aureus* skin infections. *Infect. Control Hosp. Epidemiol.* **35**, 1373–1382 (2014).
- Dalman, M. et al. Characterizing the molecular epidemiology of *Staphylococcus aureus* across and within fitness facility types. *BMC Infect. Dis.* **19**, 69 (2019).
- Monaco, M., Pimentel de Araujo, F., Cruciani, M., Coccia, E. M. & Pantosti, A. In *Current Topics in Microbiology and Immunology*, vol. 409, 21–56 (Springer, 2017).
- Azarian, T. et al. Intrahost evolution of methicillin-resistant *Staphylococcus aureus* USA300 among individuals with reoccurring skin and soft-tissue infections. *J. Infect. Dis.* **214**, 895–905 (2016).
- Malachowa, N. & Deleo, F. R. Mobile genetic elements of *Staphylococcus aureus*. *Cell. Mol. Life Sci.* **67**, 3057–3071 (2010).
- Corey, G. R. et al. Pooled analysis of single-dose oritavancin in the treatment of acute bacterial skin and skin-structure infections caused by Gram-positive pathogens, including a large patient subset with methicillin-resistant *Staphylococcus aureus*. *Int. J. Antimicrob. Agents* **48**, 528–534 (2016).
- Arshad, S. et al. Ceftaroline fosamil monotherapy for methicillin-resistant *Staphylococcus aureus* bacteremia: a comparative clinical outcomes study. *Int. J. Infect. Dis.* **57**, 27–31 (2017).
- Mienda, B. S., Sallihu, R., Adamu, A. & Idris, S. Genome-scale metabolic models as platforms for identification of novel genes as antimicrobial drug targets. *Future Microbiol.* **13**, 455–467 (2018).
- Liu, L., Agren, R., Bordel, S. & Nielsen, J. Use of genome-scale metabolic models for understanding microbial physiology. *FEBS Lett.* **584**, 2556–2564 (2010).
- 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**, 1–18 (2019).
- Dräger, A. & Planatscher, H. in *Encyclopedia of Systems Biology* (eds. Dubitzky, W. et al.) 1249–1251 (Springer, 2013).
- Lewis, N. E., Nagarajan, H. & Palsson, B. O. Constraining the metabolic genotype-phenotype relationship using a phylogeny of in silico methods. *Nat. Rev. Microbiol.* **10**, 291–305 (2012).
- O'Brien, E., Monk, J. & Palsson, B. Using genome-scale models to predict biological capabilities. *Cell* **161**, 971–987 (2015).
- Renz, A., Mostolizadeh, R. & Dräger, A. In *Systems Medicine* (ed. Wolkenhauer, O.) vol. 3, 362–371 (Academic Press, 2020). <https://www.sciencedirect.com/science/article/pii/B9780128012383115247>.
- Norsigian, C. J. et al. BiGG Models 2020: multi-strain genome-scale models and expansion across the phylogenetic tree. *Nucleic Acids Res.* <https://doi.org/10.1093/nar/gkz1054/5614178> (2020).
- Glont, M. et al. BioModels: expanding horizons to include more modelling approaches and formats. *Nucleic Acids Res.* **46**, D1248–D1253 (2018).
- Becker, S. A. & Palsson, B. O. Genome-scale reconstruction of the metabolic network in *Staphylococcus aureus* N315: an initial draft to the two-dimensional annotation. *BMC Microbiol.* **5**, 8 (2005).

B.4 Curating and comparing 114 stain-specific GEMs of *S. aureus*

28. Heinemann, M., Kümmel, A., Ruinatscha, R. & Panke, S. In silico genome-scale reconstruction and validation of the *Staphylococcus aureus* metabolic network. *Biotechnol. Bioeng.* **92**, 850–864 (2005).
29. Büchel, F. et al. Path2Models: large-scale generation of computational models from biochemical pathway maps. *BMC Syst. Biol.* **7**, 116 (2013).
30. Seif, Y. et al. A computational knowledge-base elucidates the response of *Staphylococcus aureus* to different media types. *PLoS Comput. Biol.* **15**, e1006644 (2019).
31. Lee, D.-S. et al. Comparative genome-scale metabolic reconstruction and flux balance analysis of multiple *Staphylococcus aureus* genomes identify novel antimicrobial drug targets. *J. Bacteriol.* **191**, 4015–4024 (2009).
32. Bosi, E. et al. Comparative genome-scale modelling of *Staphylococcus aureus* strains identifies strain-specific metabolic capabilities linked to pathogenicity. *Proc. Natl Acad. Sci. USA* **113**, E3801–9 (2016).
33. Magnúsdóttir, S. et al. Generation of genome-scale metabolic reconstructions for 773 members of the human gut microbiota. *Nat. Biotechnol.* **35**, 81–89 (2016).
34. Noronha, A. et al. The virtual metabolic human database: integrating human and gut microbiome metabolism with nutrition and disease. *Nucleic Acids Res.* **47**, D614–D624 (2019).
35. Ebrahim, A., Lerman, J. A., Palsson, B. O. & Hyduke, D. R. COBRApy: COstraints-based reconstruction and analysis for python. *BMC Syst. Biol.* **7**, 74 (2013).
36. Lieven, C. et al. MEMOTE for standardized genome-scale metabolic model testing. *Nat. Biotechnol.* **38**, 272–276 (2020).
37. Kanehisa, M., Sato, Y., Furumichi, M., Morishima, K. & Tanabe, M. New approach for understanding genome variations in KEGG. *Nucleic Acids Res.* **47**, D590–D595 (2019).
38. Peterson, J. D. The comprehensive microbial resource. *Nucleic Acids Res.* **29**, 123–125 (2001).
39. Dauner M, S. U. Stoichiometric growth model for riboflavin-producing *Bacillus subtilis*. *Biotechnol. Bioeng.* **76**, 132–143 (2001).
40. Hucka, M. et al. Systems biology markup language (SBML) level 3 version 1 core. *J. Integr. Bioinform.* **15**, 1 (2018).
41. Olivier, B. G. & Bergmann, F. T. SBML level 3 package: flux balance constraints version 2. *J. Integr. Bioinform.* **15**, 20170082 (2018).
42. Overbeek, R. et al. The ErgoTM genome analysis and discovery system. *Nucleic Acids Res.* **31**, 164–171 (2003).
43. Schaefer, C. F. et al. PID: The pathway interaction database. *Nucleic Acids Res.* **37**, D674–D679 (2009).
44. Caspi, R. et al. The MetaCyc database of metabolic pathways and enzymes—a 2019 update. *Nucleic Acids Res.* **48**, D445–D453 (2019).
45. Wittig, U., Rey, M., Weidemann, A., Kania, R. & Müller, W. SABIO-RK: an updated resource for manually curated biochemical reaction kinetics. *Nucleic Acids Res.* **46**, D656–D660 (2018).
46. Moretti, S., Tran, V. D. T., Mehl, F., Ibberson, M. & Pagni, M. MetaNetX/MNXref: unified namespace for metabolites and biochemical reactions in the context of metabolic models. *Nucleic Acids Res.* **49**, D570–D574 (2021).
47. Juty, N., Le Novère, N. & Laibe, C. Identifiers.org and MIRIAM Registry: community resources to provide persistent identification. *Nucleic Acids Res.* **40**, D580–D586 (2012).
48. Finney, A. & Hucka, M. Systems biology markup language: level 2 and beyond. *Biochem. Soc. Trans.* **31**, 1472–1473 (2003).
49. Henry, C. S. et al. High-throughput generation, optimization and analysis of genome-scale metabolic models. *Nat. Biotechnol.* **28**, 977–982 (2010).
50. Karp, P. D. et al. The BioCyc collection of microbial genomes and metabolic pathways. *Brief. Bioinform.* **20**, 1085–1093 (2019).
51. Keating, S. M. et al. SBML level 3: an extensible format for the exchange and reuse of biological models. *Mol. Syst. Biol.* **16**, e9110 (2020).
52. Arkin, A. P. et al. KBase: The United States Department of Energy systems biology knowledgebase. *Nat. Biotechnol.* **36**, 566–569 (2018).
53. Thorleifsson, S. G. & Thiele, I. rBioNet: a COBRA toolbox extension for reconstructing high-quality biochemical networks. *Bioinformatics* **27**, 2009–2010 (2011).
54. Kuroda, M. et al. Whole genome sequencing of methicillin-resistant *Staphylococcus aureus*. *Lancet* **357**, 1225–1240 (2001).
55. Hiramatsu, K. et al. Dissemination in Japanese hospitals of strains of *Staphylococcus aureus* heterogeneously resistant to vancomycin. *Lancet* **350**, 1670–1673 (1997).
56. Neoh, H. M. et al. Mutated response regulator graR is responsible for phenotypic conversion of *Staphylococcus aureus* from heterogeneous vancomycin-intermediate resistance to vancomycin-intermediate resistance. *Antimicrob. Agents Chemother.* **52**, 45–53 (2008).
57. Mwangi, M. M. et al. Tracking the in vivo evolution of multidrug resistance in *Staphylococcus aureus* by whole-genome sequencing. *Proc. Natl Acad. Sci. USA* **104**, 9451–9456 (2007).
58. Highlander, S. K. et al. Subtle genetic changes enhance virulence of methicillin resistant and sensitive *Staphylococcus aureus*. *BMC Microbiol.* **7**, 99 (2007).
59. Diep, B. A. et al. Complete genome sequence of USA300, an epidemic clone of community-acquired methicillin-resistant *Staphylococcus aureus*. *Lancet* **367**, 731–739 (2006).
60. Baba, T. et al. Genome and virulence determinants of high virulence community-acquired MRSA. *Lancet* **359**, 1819–1827 (2002).
61. Gill, S. R. et al. Insights on evolution of virulence and resistance from the complete genome analysis of an early methicillin-resistant *Staphylococcus aureus* strain and a biofilm-producing methicillin-resistant *Staphylococcus epidermidis* strain. *J. Bacteriol.* **187**, 2426–2438 (2005).
62. Dyke, K. G., Jevons, M. P. & Parker, M. T. Penicillinase production and intrinsic resistance to penicillins in *Staphylococcus aureus*. *Lancet* **1**, 835–838 (1966).
63. Baba, T., Bae, T., Schneewind, O., Takeuchi, F. & Hiramatsu, K. Genome sequence of *Staphylococcus aureus* strain Newman and comparative analysis of staphylococcal genomes: polymorphism and evolution of two major pathogenicity islands. *J. Bacteriol.* **190**, 300–310 (2008).
64. Holden, M. T. et al. Complete genomes of two clinical *Staphylococcus aureus* strains: evidence for the evolution of virulence and drug resistance. *Proc. Natl Acad. Sci. USA* **101**, 9786–9791 (2004).
65. Herron-Olson, L., Fitzgerald, J. R., Musser, J. M. & Kapur, V. Molecular correlates of host specialization in *Staphylococcus aureus*. *PLoS ONE* **2**, e1120 (2007).
66. Vogel, V., Falquet, L., Calderon-Copete, S. P., Basset, P. & Blanc, D. S. Short term evolution of a highly transmissible methicillin-resistant *Staphylococcus aureus* clone (ST228) in a Tertiary care hospital. *PLoS ONE* **7**, e38969 (2012).
67. Hucka, M. et al. Systems Biology Markup Language (SBML) level 3 version 2 core release 2. *J. Integr. Bioinform.* **16**, 1 (2019).
68. Courtot, M. et al. Controlled vocabularies and semantics in systems biology. *Mol. Syst. Biol.* **7**, 543–543 (2014).
69. Römer, M. et al. ZBIT Bioinformatics Toolbox: a web-platform for systems biology and expression data analysis. *PLoS ONE* **11**, e0149263 (2016).
70. Hucka, M., Finney, A., Hoops, S., Keating, S. M. & Le Novère, N. Systems biology markup language (SBML) Level 2: structures and facilities for model definitions. *Tech. Rep. Nat. Preced.* <https://doi.org/10.1038/npre.2008.2715.1> (2008).
71. Bornstein, B. J., Keating, S. M., Jouraku, A. & Hucka, M. LibSBML: an API Library for SBML. *Bioinformatics* **24**, 880–881 (2008).
72. Olivier, B. G. & Bergmann, F. T. The systems biology markup language (SBML) level 3 package: flux balance constraints. *J. Integr. Bioinform.* **2**, 660–690 (2015).
73. Chang, A. et al. BRENDA, the ELIXIR core data resource in 2021: new developments and updates. *Nucleic Acids Res.* **49**, D498–D508 (2021).
74. Lombardot, T. et al. Updates in Rhea: SPARQLing biochemical reaction data. *Nucleic Acids Res.* **47**, D596–D600 (2019).
75. Jassal, B. et al. The reactome pathway knowledgebase. *Nucleic Acids Res.* **48**, D498–D503 (2020).
76. Morgat, A. et al. UniPathway: a resource for the exploration and annotation of metabolic pathways. *Nucleic Acids Res.* **40**, D761–D769 (2012).
77. Wishart, D. S. et al. HMDB 4.0: the human metabolome database for 2018. *Nucleic Acids Res.* **46**, D608–D617 (2018).
78. Hastings, J. et al. ChEBI in 2016: improved services and an expanding collection of metabolites. *Nucleic Acids Res.* **44**, D1214–D1219 (2016).
79. Heller, S. R., McNaught, A., Pletnev, I., Stein, S. & Tchekhovskoi, D. InChI, the IUPAC International Chemical Identifier. *J. Cheminform.* **7**, 23 (2015).
80. Hussain, M., Hastings, J. G. M. & White, P. J. A chemically defined medium for slime production by coagulase-negative staphylococci. *J. Med. Microbiol.* **34**, 143–147 (1991).
81. Halsey, C. R. et al. Amino acid catabolism in *Staphylococcus aureus* and the function of carbon catabolite repression. *mBio* **8**, e01434–16 (2017).
82. Williams, R. E. Healthy carriage of *Staphylococcus aureus*: its prevalence and importance. *Bacteriol. Rev.* **27**, 56–71 (1963).
83. Acton, D. S., Tempelmann Plat-Sinnige, M. J., Van Wamel, W., De Groot, N. & Van Belkum, A. Intestinal carriage of *Staphylococcus aureus*: How does its frequency compare with that of nasal carriage and what is its clinical impact? *Eur. J. Clin. Microbiol. Infect. Dis.* **28**, 115–127 (2009).
84. Krismer, B. et al. Nutrient limitation governs *Staphylococcus aureus* metabolism and niche adaptation in the human nose. *PLoS Pathog.* **10**, e1003862 (2014).
85. Krismer, B., Weidenmaier, C., Zipperer, A. & Peschel, A. The commensal lifestyle of *Staphylococcus aureus* and its interactions with the nasal microbiota. *Nat. Rev. Microbiol.* **15**, 675–687 (2017).
86. Piewngam, P. et al. Pathogen elimination by probiotic *Bacillus* via signalling interference. *Nature* **562**, 532–537 (2018).
87. Ren, D. et al. Effects of mixed lactic acid bacteria on intestinal microbiota of mice infected with *Staphylococcus aureus*. *BMC Microbiol.* **18**, 109 (2018).

88. Dong, D. et al. Effects of intestinal colonization by *Clostridium difficile* and *Staphylococcus aureus* on microbiota diversity in healthy individuals in China. *BMC Infect. Dis.* **18**, 1–8 (2018).
89. Sannasiddappa, T. H., Costabile, A., Gibson, G. R. & Clarke, S. R. The influence of *Staphylococcus aureus* on gut microbial ecology in an in vitro continuous culture human colonic model system. *PLoS ONE* **6**, e23227 (2011).
90. Clemente, J. C., Ursell, L. K., Parfrey, L. W. & Knight, R. The impact of the gut microbiota on human health: an integrative view. *Cell* **148**, 1258–1270 (2012).
91. Bogaert, D. et al. Colonisation by *Streptococcus pneumoniae* and *Staphylococcus aureus* in healthy children. *Lancet* **363**, 1871–1872 (2004).
92. Rosenfeld, M. et al. Prevalence of cystic fibrosis pathogens in the oropharynx of healthy children and implications for cystic fibrosis care. *J. Cyst. Fibros.* **11**, 456–457 (2012).
93. Hurley, M. N. *Staphylococcus aureus* in cystic fibrosis: Problem bug or an innocent bystander? *Breathe* **14**, 87–90 (2018).
94. Schwerdt, M. et al. *Staphylococcus aureus* in the airways of cystic fibrosis patients—a retrospective long-term study. *Int. J. Med. Microbiol.* **308**, 631–639 (2018).
95. Palmer, K. L., Aye, L. M. & Whiteley, M. Nutritional cues control *Pseudomonas aeruginosa* multicellular behavior in cystic fibrosis sputum. *J. Bacteriol.* **189**, 8079–8087 (2007).
96. Sweeney, E. et al. An ex vivo cystic fibrosis model recapitulates key clinical aspects of chronic *Staphylococcus aureus* infection. *Microbiology* **167**, 000987 (2021).
97. Davis, J. J. et al. The PATRIC Bioinformatics Resource Center: expanding data and analysis capabilities. *Nucleic Acids Res.* **48**, D606–D612 (2020).
98. Luo, H., Lin, Y., Gao, F., Zhang, C. T. & Zhang, R. DEG 10, an update of the database of essential genes that includes both protein-coding genes and noncoding genomic elements. *Nucleic Acids Res.* **42**, D574 (2014).
99. Herron, L. L. et al. Genome sequence survey identifies unique sequences and key virulence genes with unusual rates of amino acid substitution in bovine *Staphylococcus aureus*. *Infect. Immunity* **70**, 3978–3981 (2002).
100. Clark, K., Karsch-Mizrachi, I., Lipman, D. J., Ostell, J. & Sayers, E. W. GenBank. *Nucleic Acids Res.* **44**, D67–D72 (2016).
101. Thiele, I. & Palsson, B. A protocol for generating a high-quality genome-scale metabolic reconstruction. *Nat. Protoc.* **5**, 93–121 (2010).
102. Crupper, S. S., Worrell, V., Stewart, G. C. & Landolo, J. J. Cloning and expression of cadD, a new cadmium resistance gene of *Staphylococcus aureus*. *J. Bacteriol.* **181**, 4071–4075 (1999).
103. Carey, M. A., Dräger, A., Beber, M. E., Papin, J. A. & Yurkovich, J. T. Community standards to facilitate development and address challenges in metabolic modeling. *Mol. Syst. Biol.* **16**, e9235 (2020).
104. Ramon, C., Gollub, M. G. & Stelling, J. Integrating -omics data into genome-scale metabolic network models: principles and challenges. *Essays Biochem.* **62**, 563–574 (2018).
105. Chelliah, V. et al. BioModels: ten-year anniversary. *Nucleic Acids Res.* **43**, D542–D548 (2015).
106. Bergmann, F. T. et al. COMBINE archive and OMEX format: one file to share all information to reproduce a modeling project. *BMC Bioinform.* **15**, 369 (2014).

ACKNOWLEDGEMENTS

This work was funded by the *Deutsche Forschungsgemeinschaft* (DFG, German Research Foundation) under Germany's Excellence Strategy—EXC 2124—390838134 and supported by the Cluster of Excellence “Controlling Microbes to Fight Infections” (CMFI). A.D. is supported by the German Center for Infection Research (DZIF, doi: 10.13039/100009139) within the *Deutsche Zentren der Gesundheitsforschung* (BMBF-DZG, German Centers for Health Research of the Federal Ministry of Education and Research), grant № 8020708703.

AUTHOR CONTRIBUTIONS

A.R. conceived the study and improved the models. A.D. supervised the study. A.R. and A.D. wrote the manuscript. Both authors reviewed and approved the final manuscript.

COMPETING INTERESTS

The authors declare no competing interests.

ADDITIONAL INFORMATION

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41540-021-00188-4>.

Correspondence and requests for materials should be addressed to A.D.

Reprints and permission information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021

ARTICLE OPEN



An updated genome-scale metabolic network reconstruction of *Pseudomonas aeruginosa* PA14 to characterize mucin-driven shifts in bacterial metabolism

Dawson D. Payne¹, Alina Renz^{2,3,4,7}, Laura J. Dunphy^{1,7}, Taylor Lewis⁵, Andreas Dräger^{2,3,4,6} and Jason A. Papin¹✉

Mucins are present in mucosal membranes throughout the body and play a key role in the microbe clearance and infection prevention. Understanding the metabolic responses of pathogens to mucins will further enable the development of protective approaches against infections. We update the genome-scale metabolic network reconstruction (GENRE) of one such pathogen, *Pseudomonas aeruginosa* PA14, through metabolic coverage expansion, format update, extensive annotation addition, and literature-based curation to produce iPau21. We then validate iPau21 through MEMOTE, growth rate, carbon source utilization, and gene essentiality testing to demonstrate its improved quality and predictive capabilities. We then integrate the GENRE with transcriptomic data in order to generate context-specific models of *P. aeruginosa* metabolism. The contextualized models recapitulated known phenotypes of unaltered growth and a differential utilization of fumarate metabolism, while also revealing an increased utilization of propionate metabolism upon MUC5B exposure. This work serves to validate iPau21 and demonstrate its utility for providing biological insights.

npj Systems Biology and Applications (2021)7:37; <https://doi.org/10.1038/s41540-021-00198-2>

INTRODUCTION

The mucosal barrier is a hydrated mucus gel that lines wet epithelial cells throughout the body, including eyes, mouth, lungs, and the gastrointestinal and urogenital tracts^{1,2}. It serves as a key mechanism of protection against pathogens. The component responsible for the gel-like properties of the mucosal layer is the glycoprotein mucin³. The dysregulation of mucins underlies diseases like cystic fibrosis⁴ and chronic obstructive pulmonary disorders². As mucins are involved in the clearance of microbes⁵, a dysregulation of mucins can result in pathogen overgrowth and severe infections⁶. While some bacterial species, including pathogenic strains of *Pseudomonas*³, are capable of residing within the mucosal layer, mucins typically impair the formation of biofilms and surface attachment⁷. Furthermore, mucins are reported to downregulate virulence genes involved in siderophore biosynthesis, quorum sensing, and toxin secretion¹. By disturbing these key mechanisms of infection, mucins attenuate the virulence and infective potential of *P. aeruginosa*.

Elucidating the metabolic responses of *P. aeruginosa* to mucins can enable the development of protective approaches against infection⁸. Genome-scale metabolic network reconstructions (GENREs) and associated genome-scale metabolic models (GEMs) are well suited for this purpose as they can enable the prediction of cellular behavior under different biological conditions such as the absence or presence of different mucins in an environment⁹. A GENRE can also be used to contextualize high-throughput data, such as transcriptomics or proteomics data¹⁰. Gene expression data can, for example, be used to constrain specific predicted metabolic fluxes¹¹ and thereby increase the predictive value of the model. Metabolically active pathways under different conditions

can be identified by integrating high-throughput data with a metabolic network¹².

P. aeruginosa is a critical bacterial species in the 'Priority Pathogens List' for research and development of new antibiotics published by the World Health Organization (WHO)¹³. However, the lack of novel antibiotics^{14,15} emphasizes the need for the development of innovative and protective therapeutic approaches. This pressing need for protective strategies coupled with new insights from recent research present an opportunity to further refine the GENRE of the highly virulent strain UCBPP-PA14 by Bartell et al.¹⁶. An updated GENRE can be used to better understand the metabolic underpinnings of *P. aeruginosa* infections and ultimately develop new therapeutic strategies from those insights.

Here, we present iPau21, an updated GENRE of *P. aeruginosa* strain UCBPP-PA14 metabolism. We improve predictions of carbon source utilization and growth yields in order to better recapitulate the behavior of the pathogen. Metabolic network coverage is expanded through the addition of genes, reactions, and metabolites supported by literature evidence. The quality of the reconstruction was improved through an update of standardized formatting, improved annotation, and the addition of binning metabolites representing macromolecular categories to assist with analysis. The metabolic network model was validated by comparing phenotypic predictions to experimental datasets^{16–21} and the quality of the reconstruction was assessed with the MEMOTE benchmarking software²². This updated reconstruction was further contextualized with recently published transcriptomic data¹ in order to demonstrate its utility in elucidating the metabolic shifts of *P. aeruginosa* after exposure to mucins. The validated reconstructions will serve as a key resource for the

¹Department of Biomedical Engineering, University of Virginia, Charlottesville, VA, USA. ²Institute for Bioinformatics and Medical Informatics (IBMI), University of Tübingen, Tübingen, Germany. ³Department of Computer Science, University of Tübingen, Tübingen, Germany. ⁴Cluster of Excellence 'Controlling Microbes to Fight Infections', University of Tübingen, Tübingen, Germany. ⁵Department of Chemical Engineering, University of Massachusetts Amherst, Amherst, MA, USA. ⁶German Center for Infection Research (DZIF) partner site, Tübingen, Germany. ⁷These authors contributed equally: Alina Renz, Laura J. Dunphy. ✉email: papin@virginia.edu

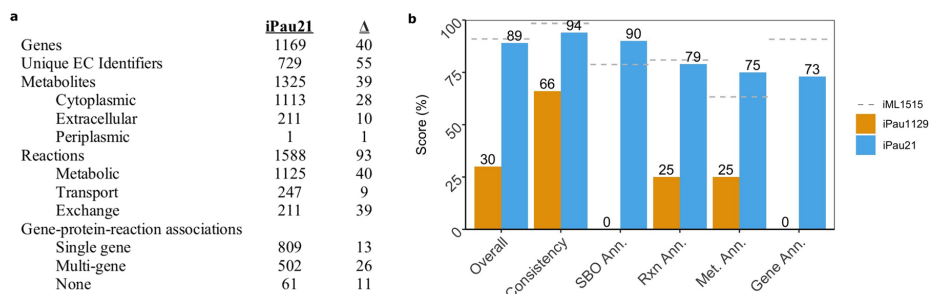


Fig. 1 Characteristics and MEMOTE benchmarking of iPau21. **a** Properties of iPau21 as compared to iPau1129. **b** MEMOTE scores of iPau21, iPau1129, and iML1515, a high-quality reconstruction of *E. coli*.

Pseudomonas and microbial metabolic modeling communities and the insights into mucin-driven metabolic shifts in *P. aeruginosa* may serve to inform the future development of therapeutic strategies.

RESULTS

An updated network reconstruction of *Pseudomonas aeruginosa* metabolism

A metabolic network reconstruction of *P. aeruginosa* PA14 (iPau1129) was previously published¹⁶ and served as a starting point for an updated reconstruction (iPau21). The metabolic coverage of the reconstruction was expanded, the format and annotations were updated, and an ATP-generating loop was resolved in order to produce a refined model with improved accuracy and extensive annotation.

We expanded iPau1129 by 40 genes, 24 metabolites, and 76 reactions (Fig. 1a) through manual curation based on literature evidence (Supplementary Data 1). Many of these additions served to increase the utility of the reconstruction for simulation (such as the addition of 33 exchange reactions), while others expanded metabolic pathways for amino acid metabolism and glycerophospholipid metabolism. A periplasmic compartment containing hydrogen was added to the reconstruction to better represent the electron transport chain and ATP synthase, which eliminated all ATP-generating loops in the metabolic network. The format was updated from SBML Level 2²³ to Level 3²⁴, which enables additional functionality such as the utilization of several extension packages and the transfer of information content to dedicated new data structures. Annotations from various databases were added to metabolites, reactions, and genes where possible.

The overall quality of the updated reconstruction was assessed using MEMOTE²², a recently developed GENRE test suite. The MEMOTE score of iPau21 improved in all subcategories when compared to iPau1129 resulting in an increase of the overall score from 30% to 89% (Fig. 1b and Supplementary Materials). The scores in annotation subcategories were increased by adding annotations and SBO terms to metabolites, reactions, and genes in the updated GENRE. The consistency of the metabolic network was improved through the correction of imbalanced reactions and the resolution of energy generating cycles that were present in iPau1129.

The biomass objective function (BOF) was updated to better reflect the macromolecular components found experimentally in *P. aeruginosa* including the inclusion of lipopolysaccharide^{25–27}. BOF substrates were organized into corresponding macromolecular categories (i.e., DNA, RNA, protein, lipid) to better represent the categories of components that are required for growth.

Model validation

Validation of iPau21 was performed by comparing in silico predictions of biomass flux, carbon source utilization, and gene essentiality to experimental data. Biomass flux and subsequent doubling time predictions in simulated lysogeny broth (LB), synthetic cystic fibrosis media (SCFM), and glucose minimal media were compared to experimental values found in literature (Fig. 2a, Supplementary Data 2)^{17–19}. Doubling time predictions of iPau21 were 25%, 19%, and 22% more accurate than those of iPau1129 in simulated LB, SCFM, and glucose minimal media, respectively. Compared to the original model, iPau21 doubling times are higher, which reflects the resolution of the ATP-generating loop that previously allowed the model to costlessly convert ADP to ATP. The iPau21 doubling time prediction on glucose minimal media of 40.2 min showed agreement with experimental data, falling within the range of experimentally determined values¹⁹. Model doubling time predictions on LB and SCFM were faster than observed experimentally, which is consistent with metabolic network models that are structured to predict the optimal growth of an organism.

Carbon source utilization predictions were compared to previously collected experimental results across 91 carbon sources¹⁶. Utilization was predicted by iPau21 with an accuracy of 89% and Matthews correlation coefficient (MCC) of 0.78, while iPau1129 demonstrated an accuracy of 80% and MCC of 0.62 (Fig. 2b and Supplementary Data 3). This increase in accuracy was achieved through the completion of pathways that allow for the utilization of more carbon sources and the removal of an unsupported reaction that previously allowed for the utilization of D-malate. Carbon source predictions of iPau21 remain incorrect for 10 carbon sources. Five of the incorrect predictions are due to the absence of metabolic pathways required for growth on certain carbon sources. When addressing these predictions, our literature survey was unable to provide sufficient evidence for these pathways so the predictions remain incorrect and we opted to not gapfill without that additional evidence. To correct these predictions, summary reactions could be added to the reconstruction, but these reactions would lack the mechanistic granularity of associated genes and could have negative impacts on other aspects of the reconstruction. The other five incorrect predictions were caused by the presence of metabolic pathways that allow for the erroneous growth on the associated carbon sources. In each of these cases, the pathway was investigated and the corresponding genes were verified through the KEGG²⁸ and ModelSEED²⁹ databases, but there was not strong enough evidence to warrant changes in the reconstruction^{28,30}. Some of these discrepancies may be due to considerations that are outside of the scope of the network, such as transcriptional processes. For example, in the case of D-serine, PA14 has the ability to metabolize this carbon source but expression of this gene is not triggered by the

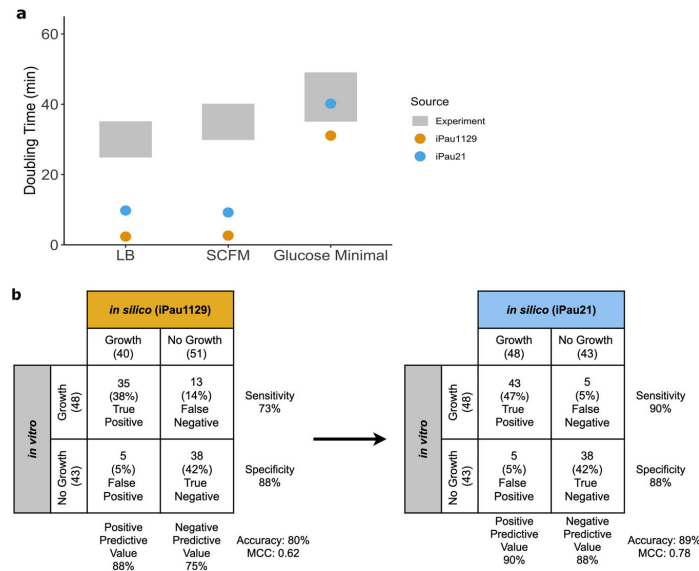


Fig. 2 Updated reconstruction of *P. aeruginosa* enables accurate growth rate, gene essentiality, and carbon source utilization predictions. **a** Model doubling time predictions compared to experimental results gathered from literature. Gray bar represents the experimental range. **b** Model carbon source utilization predictions compared to results gathered from literature.

presence of D-serine so it is unable to grow on this single carbon source in vitro³¹. These inaccurate predictions could be improved by modifying constraints in the metabolic network model. However, since the gene-protein-reaction (GPR) rules were found to be valid and the prediction error could be due to unaccounted for regulatory control, we opted to leave the pathways intact. Overall, we were able to increase carbon source utilization prediction accuracy by 9% in comparison to the previously published model.

Gene essentiality predictions were compared to a published dataset comprised of the overlap of essential genes identified through the growth of strains PAO1 and PA14 transposon insertion mutants in LB media^{20,21}. The number of genes accounted for by iPau21 was expanded to 1169 and the gene essentiality prediction accuracy was maintained at 91%, which is equivalent to iPau1129 (Supplementary Data 4). Gene essentiality was predicted by iPau21 with a MCC of 0.50, compared to a value of 0.44 by iPau1129. Three genes labeled as “SPONTANEOUS,” “unassigned,” and “Unassigned” were removed from the reconstruction given that these labels did not correspond to genes belonging to *P. aeruginosa*. Gene essentiality data was not used for curation of the metabolic network given the variability in gene essentiality screens and the resultant challenges with data interpretation³². Instead, model predictions were compared to gene essentiality data as one facet of validation. As a reference, iPau21 has a gene essentiality prediction accuracy of 91%, which is near the 93% accuracy of iML1515, a well-curated reconstruction of *Escherichia coli*³³.

Transcriptome-guided modeling of *P. aeruginosa* metabolism in the presence of human mucins

Mucins are the primary macromolecules in mucosal layers known to modulate microbial phenotypes². In order to investigate how the metabolism of *P. aeruginosa* shifts when it comes into contact with mucins, in vitro transcriptomic data was integrated with iPau21 to generate contextualized models that offer more

biologically accurate representations of associated metabolic phenotypes. Analysis of the structure and pathway utilization in these transcriptome-guided models offers insights into the metabolic shifts that arise when *P. aeruginosa* is exposed to mucins.

Transcriptomic profiles of *P. aeruginosa* PAO1 grown in agrobacterium minimal medium with thiamine, glucose, and casamino acids (ABTGC) medium supplemented with either MUC5AC, MUC5B, or mucin-glycans were collected from literature^{1,34}. MUC5AC and MUC5B are mucin types found both individually and together at different sites of the human body that *P. aeruginosa* is known to infect⁸. The mucin-glycans used in the published experiments were isolated from the backbone of MUC5AC. The experiments were performed with strain PAO1, which has a highly similar genome to strain PA14³⁵. The main difference between the strains is the presence of additional gene clusters in PA14 (most linked to virulence) that we would not expect to have a large effect on overall metabolism. PAO1 genes in the transcriptomic dataset were mapped to PA14 orthologs and then the data was integrated with the iPau21 using the RIPTiDe algorithm³⁶. RIPTiDe uses transcriptomic evidence to create context-specific metabolic models representative of a parsimonious metabolism consistent with the transcriptional investments of an organism. This analysis resulted in four contextualized models that more accurately represent the metabolism of *P. aeruginosa* when grown without mucin exposure (ABTGC) and when exposed to MUC5AC, MUC5B, and glycans.

Flux samples were generated for each model and BOF flux did not vary significantly among the contextualized models (less than five percent change), recapitulating the phenotype that was observed experimentally¹. The flux distributions underlying the BOF values were compared across models using non-metric multidimensional scaling (NMDS) in order to compare the metabolic mechanisms of growth utilized by the condition-specific metabolisms (Fig. 3a). The fluxes from the 378 consensus reactions (shared across all models) were used for this analysis.

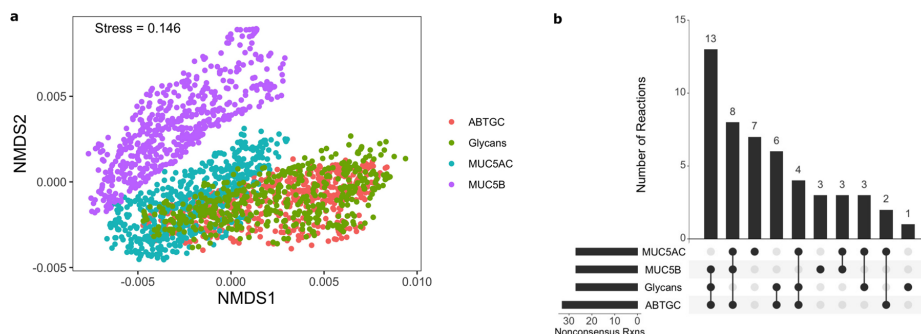


Fig. 3 Contextualization of updated reconstruction shows shifts in *P. aeruginosa* metabolism in response to mucins and mucin components. **a** NMDS analysis of flux samples ($n = 500$) from each contextualized model. **b** Comparison of non-consensus reactions present within models displays subsets of reactions that are shared by groups of contextualized models.

NMDS analysis revealed that among the tested conditions, the sampled flux distributions from the MUC5B model clustered the furthest from the ABTGC condition. This result indicates that although there was not a significant difference in the BOF value, exposure to MUC5B caused the largest shift in the metabolic pathways utilized for growth. MUC5AC clustered the second furthest away, while Glycans clustered most closely to the ABTGC model, showing that there was a variable metabolic response to different mucins and glycans by *P. aeruginosa*. Mucin-glycans do not contain the same level of structural and biochemical complexity as MUC5AC and MUC5B, which may account for the slight metabolic shift observed in the Glycans model relative to the MUC5AC and MUC5B models. MUC5AC and MUC5B are known to differ from each other in terms of charge, shape, and glycosylation³⁷. These differences could explain the variable metabolic response they elicit in *P. aeruginosa*. Additionally, of the two only MUC5B has been shown to be critical for murine mucociliary transport and antibacterial defense³⁸. One mechanism of MUC5B antibacterial effects could be through modulation of pathogen metabolism, which would explain the larger shift in metabolism observed when *P. aeruginosa* was exposed to MUC5B. The conserved BOF flux values and separation observed between clusters of flux samples suggest that while *P. aeruginosa* metabolism is modulated by the presence of mucins, its versatility allows for the utilization of alternative metabolic pathways in order to avoid a growth defect.

The differences between networks were further investigated through the metabolites that were produced and consumed by models in silico. This analysis offers a snapshot of the substrates used and byproducts of particular metabolic states, which can be informative of the metabolism underlying that state. All models were found to consume the same metabolites with some small differences in specific flux values however, there were key differences in the metabolites that models produced (Supplementary Data 5). The graded differences between models seen in NMDS were highlighted by the production of formate by the models. The ABTGC and Glycans models produced substantially higher amounts of formate than the MUC5AC model, while the MUC5B model did not contain the formate exchange reaction. Therefore, with our model, we are able to predict subtle shifts in *P. aeruginosa* metabolism in response to different environmental mucins.

Human mucins shift *P. aeruginosa* metabolism

Further analysis was conducted on the contextualized models to better understand the shifts in metabolism that resulted in the observed dissimilarities in the NMDS analysis. Reactions not

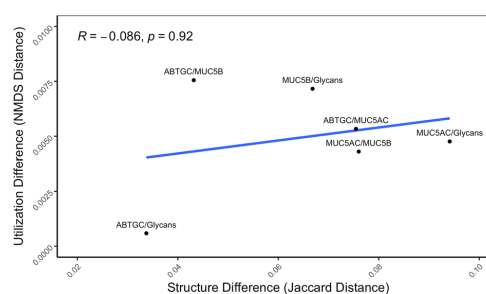


Fig. 4 Network utilization does not correlate with network structure. The distance between median NMDS coordinates for each pair of networks was calculated as a metric of difference in network utilization, while the Jaccard distance of network reactions for each pair of networks was calculated as a metric of structural difference. Spearman's correlation shows an insignificant relation between the two metrics ($p = 0.92$).

shared across all models (non-consensus reactions) were identified and compared to investigate how network structure varies across models (Fig. 3b). This analysis revealed a set of 13 reactions shared by the ABTGC, MUC5B, and Glycans models but absent from the MUC5AC network. This result suggests that while MUC5B displayed the largest functional differences in metabolism, MUC5AC is the most structurally unique of our models. Additionally, we found that there was no correlation between network structure and utilization among our contextualized models (Fig. 4, p -value = 0.92). Since the NMDS analysis revealed that the ABTGC and MUC5B models had the largest difference in functional metabolism, these two models were further investigated to find key attributes that underlie these large differences. Random forest analysis was conducted on the flux samples from consensus reactions of the ABTGC and MUC5B models to find which reactions were most differentially utilized between the two cases (Fig. 5). Two reactions corresponding to fumarate transport were in the top seven most discriminating reactions between models, suggesting that there was a differential utilization of reactions involved in fumarate metabolism. The MUC5B model utilized the fumarate reactions more highly than the ABTGC model and contained a fumarase reaction that was not present in the ABTGC model, which further suggests that fumarate metabolism is a key point of difference between the models. This observation recapitulates what was noted in the original paper that produced the transcript data used for contextualization¹. Of the top six most

Reaction Name	Pathway	MUC5B Shift
Phosphate acetyltransferase	Propionate metabolism	↑
Succinyl-CoA synthetase (ADP-forming)	Propionate metabolism	↓
Propionyl adenylateCoA ligase (AMP-forming)	Propionate metabolism	↑
Fumarate transport via proton symport (3 H)	Fumarate metabolism	↑
PropanoateCoA ligase (AMP-forming)	Propionate metabolism	↑
Propionate kinase	Propionate metabolism	↑
Fumarate reversible transport via symport	Fumarate metabolism	↑

Fig. 5 Random forest analysis between ABTGC and MUC5B shows the networks differ most in terms of propionate and fumarate metabolism utilization. The top seven most discriminating reactions between the two models belong to propionate and fumarate metabolism. MUC5B utilizes these two types of metabolism more highly than the ABTGC model.

discriminating reactions, five corresponded to propionate metabolism and were more highly utilized by the MUC5B model than the ABTGC model. While there was no propionate in the simulated (or in vitro) media, it is a known byproduct of mucin fermenters and has anti-lipogenic and anti-inflammatory properties in humans^{39,40}. This analysis revealed that the exposure to MUC5B elicited the largest shift in metabolism compared to MUC5AC and Glycans. Further, an increased utilization of fumarate and propionate metabolism during simulated growth was responsible for this shift.

DISCUSSION

We generated an updated network reconstruction of *P. aeruginosa* PA14 metabolism with considerable improvements in model annotation and accuracy of growth rate and carbon source utilization predictions. The metabolic reaction coverage of the reconstruction was expanded, the format and annotations were updated to be consistent with current best practices, and an ATP-generating loop was resolved. Model improvements were quantified through various metrics such as accuracy of growth yield and carbon source utilization predictions as well as MEMOTE benchmarking²².

The updated network reconstruction was contextualized using transcriptomic data in order to investigate the shifts in metabolism that occur when *P. aeruginosa* is exposed to mucins present in the human body. This analysis recapitulated an unaltered growth rate and differential fumarate metabolism that has been reported in literature and also revealed an increased utilization of propionate metabolism in the presence of mucins. Propionate is a short chain fatty acid with beneficial effects to human health such as anti-lipogenic, anti-inflammatory, and anti-carcinogenic action^{39,41}. While propionate is not present in the ABTGC medium, it is known to be produced by bacteria such as *Akkermansia muciniphila* when they come into contact with and catabolize mucins⁴⁰. This shift of *P. aeruginosa* metabolism towards propionate metabolism may indicate a cross-feeding mechanism where MUC5B mucins signal to *Pseudomonas* to prepare to metabolize the propionate produced by other microbes as they break down the mucins. Once validated, this insight could be used to develop therapeutic strategies for *P. aeruginosa* infections of body sites containing MUC5B such as the lung, oral cavity, and middle ear¹. Antibiotics could be designed to target proteins for propionate metabolism in order to combat drug-resistant strains that cannot be treated with traditional antibiotics.

While the updates made to the model broadly improved the model accuracy, there were incorrect predictions about carbon source utilization, gene essentiality, and growth rate that were not able to be addressed. Some incorrect predictions are due to a lack of literature evidence, such as incorrect carbon source utilization predictions that are due to the absence of metabolic

pathways in the model. Other incorrect predictions are due to factors that are outside of the scope of the model, such as the incorrect prediction of growth on D-serine that is caused by the transcriptional regulation of *dsdA*³¹. There are other opportunities for further curation that would result in additional improvements to the MEMOTE score, which can be further interrogated by uploading the iPau21 reconstruction to the MEMOTE website (memote.io).

The transcriptomic data that was used to contextualize the model was collected through experiments with *P. aeruginosa* strain PAO1. Therefore, the genes in the transcriptomic dataset were mapped to their PA14 orthologs before being integrated with the network reconstruction. While the genomes of PAO1 and PA14 are highly similar, the PA14 (6.5 Mb) genome is slightly larger than PAO1 (6.3 Mb) and contains gene clusters that are not present in PAO1³⁵. The genes absent in PAO1 therefore would not be accounted for in the transcriptomic dataset. However, since most of these genes are linked to virulence, they should not have large effects on whole metabolism as simulated here. Therefore, we expect that this application of the model would allow the identification of broad shifts in metabolism due to exposure to mucins irrespective of the specific strain simulated.

The improvements in the *P. aeruginosa* metabolic network reconstruction were made to reconcile key disagreements between in silico predictions and in vitro results, ultimately producing a higher quality metabolic network reconstruction. Through the update process, we identified key predictions that remain incorrect and offer targets for further curation. The application of the model to investigate metabolic shifts that occur upon exposure to mucins recapitulated phenotypes observed in literature and offered mechanistic insights that would be difficult to delineate experimentally. This application of the reconstruction serves as an example of how the reconstruction and associated models can provide insights into context-specific metabolism. Ultimately, this reconstruction can serve as a resource for investigating the metabolism of *P. aeruginosa* in a variety of settings and conditions.

METHODS

Genome-scale metabolic reconstructions and models (GENREs and GEMs)

GENREs are network reconstructions that represent the metabolic capabilities of an organism and can be analyzed for various applications. An organism's genes are connected to the proteins they code for and the reactions that those proteins catalyze. These associations are stored as gene-protein-reaction (GPR) relationships with the reactants and products of each reaction cataloged in a stoichiometric matrix. Metabolites in the reconstruction are assigned to compartments that mirror biologically discrete spaces such as the cytosol and the extracellular space. Exchange and transport reactions allow metabolites to flow between the compartments in the reconstruction. A GENRE is turned into a GEM (Genome-Scale Metabolic Model) by adding reaction bounds that capture the flux constraints and the reversibility of reactions. The flux bounds dictate the amount and direction of flux that a reaction can carry. Objective functions (OFs) that represent metabolic goals are added to the model to simulate biological processes. GEMs can be analyzed using flux balance analysis (FBA)-based methods to investigate and gain insights into the metabolic state of a network⁴². The updated GENRE was named iPau21 according to the community standard naming convention⁴³.

Adding annotations

Initially, the PA14 reconstruction did not contain extensive annotations for metabolites, reactions, or genes. ModelPolisher⁴⁴ can be used to annotate metabolites and reactions of a metabolic model. To do so, identifiers of the BiGG database⁴⁵ (BiGG-IDs) are required as metabolite or reaction identifiers, respectively. Since the identifiers of the model were obtained from the ModelSEED database³⁰, BiGG-IDs needed to be determined. For each metabolite, the BiGG-IDs were assessed manually. Since this is a very

time-consuming procedure, the BiGG-IDs for the reactions were resolved in a semi-automated way: The cross-references of the ModelSEED database to other databases, such as BiGG or KEGG²⁸, were used to automatically obtain the BiGG-IDs for the respective ModelSEED reaction identifier. If more than one BiGG-ID was returned, the correct identifier was determined by manual inspection of the respective reaction. The BiGG-IDs of the metabolites and reactions were added as biological qualifier ('BQB_IS') annotations to the model using libSBML Version 5.17.0⁴⁶. The annotations were added in accordance with the MIRIAM guidelines⁴⁷. After adding the BiGG-IDs to the model, ModelPolisher was used for further annotations of the model's reactions and metabolites for references to other databases, such as KEGG, MetaNetX⁴⁸, or MetaCyc⁴⁴.

For the reactions, the obtained KEGG annotations were used to further add all pathways that are associated with the respective reaction to the model. The pathways were obtained using the KEGG-ID and KEGG API to request all associated pathways. The pathways were then added to the respective reactions using the biological qualifier "BQB_OCCURS_IN" in libSBML.

The identifiers of the model genes are from the KEGG database. With the help of libSBML, the KEGG gene annotation was added to the model. For further gene annotations, the KEGG API was used to request NCBI⁴⁹ Protein IDs and Uniprot⁵⁰ IDs, which were subsequently added as respective annotations to the model. Additionally, the ID mapper from PATRIC⁵¹ was used to request RefSeq and NCBI⁴⁹ gene identifiers, as well as identifiers of the ASAP database.

Systems Biology Ontology (SBO)⁵² terms can give semantic information or be used for annotation purposes. In our network reconstruction, all genes were labeled as genes with the SBO-term "SBO:0000243". All metabolites without a valid SBO-term were labeled as simple chemicals with the SBO-term "SBO:0000247". Transport reactions were divided into (1) active transport if ATP is required for the respective transport reaction (SBO:0000657), (2) passive transport if no external energy is required (SBO:0000658), (3) symporter-mediated transport if two or more molecules are transported into the same relative direction across a membrane (SBO:0000659), or (4) antiporter-mediated transport if two or more molecules are transported in relative opposite directions across a membrane (SBO:0000660). All metabolic reactions were labeled as biochemical reactions with the SBO-term "SBO:0000176".

Upgrading SBML version

The initial PA14 reconstruction was represented in SBML Level 2 Version 1⁵³. The current reconstruction was updated to the latest SBML edition (Level 3)⁵⁴. With the help of libSBML, both the fbc-plugin⁵⁵ and the groups-plugin⁵⁶ were enabled. Initially, the chemical formulas and charges of the metabolites were stored in the notes field. With the fbc-plugin, the charges were added as features of the metabolites to the reconstruction. The fbc-plugin also enables the addition of gene products to the reconstruction. In the initial reconstruction, the subsystems of the reactions were saved in the notes field. With libSBML and the groups-plugin, the subsystems were extracted from the notes field and added as groups to the reconstruction. For each subsystem, a list of reactions associated with that pathway according to the notes was created and added to the subsystem as members.

Correcting charge and mass imbalances

A list of all mass- and charge-imbalanced reactions was extracted from the reconstruction. From this list, all exchange, sink, demand and biomass reactions were excluded. Each remaining reaction was manually checked by looking up the reaction-ID in ModelSEED²⁹: (1) If the reaction status in ModelSEED was balanced ("OK"), but differed from the reaction equation in the reconstruction, the reaction was adapted according to ModelSEED and again checked for imbalances. (2) If the reaction in ModelSEED also had an imbalanced reaction status, other databases like MetaCyc⁵⁷, BiGG⁴⁵, or KEGG²⁸ were explored and the reactions were adapted according to the respective reactions in the external databases. Where required, chemical formulas, charges, and coefficients were corrected, or chemical compounds were added or subtracted from the reactions according to the respective database reaction. All changed reactions are listed in Supplementary Data 1.

Assessing the quality of the reconstruction

MEMOTE is an open-source software that provides a measure for model quality²². Every change and improvement of the model was continuously

documented and quality-assessed using MEMOTE Version 0.9.11. Full MEMOTE reports are provided for iPau1129, iPau21, and iML1515 (Supplementary Materials). Gene essentiality predictions were compared to a published dataset that was originally used to validate iPau1129¹⁶. This dataset comprises the overlap of essential genes identified through the growth of PAO1 and PA14 transposon insertion mutants in LB media^{20,21}. Carbon source utilization predictions were compared to previously collected experimental results¹⁶. Prediction accuracy was calculated as the number of correction predictions divided by the number of total predictions. Matthews correlation coefficient (MCC) was calculated in order to assess the quality of predictions⁵⁸. Biomass flux and subsequent doubling time predictions in lysogeny broth (LB), synthetic cystic fibrosis media (SCFM), and glucose minimal media were compared to experimental values found in literature (Fig. 1c)¹⁷⁻¹⁹.

Literature-based updates

Previous work identified multiple areas where the original reconstruction (iPau1129) was unable to accurately recapitulate experimental data. This assessment included 18 incorrect carbon source predictions¹⁶ and several incorrect gene essentiality predictions⁵⁹. Pathways and gene-protein-reaction rules related to each incorrect prediction were manually curated to reflect the most recent evidence from literature, KEGG, and MetaCyc. In the absence of sufficient evidence, no changes were made, even if this absence of a change meant a prediction would remain uncorrected.

Evaluating and updating the BOF

Macromolecular categories represented in the dry weight of *P. aeruginosa* were identified through a literature survey. Metabolites in the biomass objective function (BOF) were organized into these macromolecular categories in order to better represent the components required for growth. During organization, no additional metabolites were added and the ratios of metabolites in the BOF were kept the same.

The BOF was also updated to include lipopolysaccharide (cpd17065) to reflect its presence in Gram-negative bacteria⁶⁰. A metabolite representing biomass was also added to the products of the BOF to represent the accumulation of biomass.

Addition of exchange reactions

A list of all extracellular metabolites in the reconstruction was compiled and compared to a list of all exchange reactions in the reconstruction. Exchange reactions were added for 33 extracellular metabolites that previously did not have one.

Removal of energy generating cycles

Exchange reactions were closed and the objective function was set to energy dissipation reactions for electron carriers (ATP, NADH, NADPH, FADH₂, and H⁺). The model was able to generate flux for only the ATP energy dissipation objective function, which indicated that an energy generating cycle existed. The cycle was resolved through the addition of a periplasm compartment to contain hydrogen involved in the electron transport chain and correcting the reversibility of four participating reactions.

RIPTiDe contextualization & analysis

Published transcriptomic data was integrated with the model using RIPTiDe³⁶. The transcriptomic data was normalized then translated from PAO1 genes to the orthologous PA14 genes prior to integration⁶¹. ABTGC medium was simulated in silico and applied to the model (Supplementary Data 2). Then, RIPTiDe was used to produce the contextualized models for in vitro media conditions.

NMDS analysis was conducted on flux samples from each contextualized model ($n = 500$ samples per model) using the Vegan package in R⁶². Only consensus reactions across all four contextualized models were included in the flux sample dataset and a constant was added to each flux value in the dataset to make all data points positive to facilitate comparison. Median fluxes for every reaction in each model are provided in Supplementary Data 6.

Random forest analysis was conducted on flux sampling data ($n = 500$ samples per model) from the consensus reactions of the ABTGC and MUC5B models using the randomForest package in R⁶³. Reactions that

were differentially present in contextualized models were identified and connected to their corresponding metabolic pathways manually.

The Jaccard distance of network structures was calculated by comparing the reactions contained in pairs of networks⁶⁴. The NMDS distance was calculated as the distance between the median NMDS coordinates of network pairs. Spearman's correlation was used to calculate a *p*-value for the relationship between network structure and network utilization across all pairs of networks.

DATA AVAILABILITY

All data for this project is available on GitHub (<https://github.com/dawsonpayne/iPau21>). The genome-scale metabolic model iPau21 is available in the BioModels Database⁶⁵ as an SBML Level 3 Version 1 file within a COMBINE Archive OMEX file⁶⁶ including the contextualized models and metadata⁶⁷ at <https://identifiers.org/biomodels.db/MODEL2106110001>.

CODE AVAILABILITY

Code used for the GENRE update and analyses have been archived in a GitHub repository (<https://github.com/dawsonpayne/iPau21>).

Received: 10 March 2021; Accepted: 17 September 2021;

Published online: 08 October 2021

REFERENCES

- Wheeler, K. M. et al. Mucin glycans attenuate the virulence of *Pseudomonas aeruginosa* in infection. *Nat. Microbiol.* **4**, 2146–2154 (2019).
- Co, J. Y. et al. Mucins trigger dispersal of *Pseudomonas aeruginosa* biofilms. *NPJ Biofilms Microbiomes* **4**, 1–8 (2018).
- Bansil, R. & Turner, B. S. Mucin structure, aggregation, physiological functions and biomedical applications. *Curr. Opin. Colloid Interface Sci.* **11**, 164–170 (2006).
- Henke, M. O., John, G., Germann, M., Lindemann, H. & Rubin, B. K. MUC5AC and MUC5B mucins increase in cystic fibrosis airway secretions during pulmonary exacerbation. *Am. J. Respiratory Crit. Care Med.* **175**, 816–821 (2007).
- Tilley, A. E., Walters, M. S., Shaykhiev, R. & Crystal, R. G. Cilia dysfunction in lung disease. *Annu. Rev. Physiol.* **77**, 379–406 (2015).
- Rose, M. C. & Voynow, J. A. Respiratory tract mucin genes and mucin glycoproteins in health and disease. *Physiological Rev.* **86**, 245–278 (2006).
- Caldara, M. et al. Mucin biopolymers prevent bacterial aggregation by retaining cells in the free-swimming state. *Curr. Biol.* **22**, 2325–2330 (2012).
- Hoffman, C. L., Lalsiamthara, J., & Aballay, A. Host mucin is exploited by *Pseudomonas aeruginosa* to provide monosaccharides required for a successful infection. *MBio*, **11** (2020).
- Fan, J., Qiu, J. & Wei, Q. Extremely rare case of intravascular solitary fibrous tumour in the inferior vena cava with review of the literature. *Diagnostic Pathol.* **14**, 1–6 (2019).
- Oberhardt, M. A., Palsson, B. Ø. & Papin, J. A. Applications of genome-scale metabolic reconstructions. *Mol. Syst. Biol.* **5**, 320 (2009).
- Shlomi, T., Cabili, M. N., Herrgård, M. J., Palsson, B. Ø. & Ruppin, E. Network-based prediction of human tissue-specific metabolism. *Nat. Biotechnol.* **26**, 1003–1010 (2008).
- Usalte, R., Patil, K. R., Grotkjær, T., Nielsen, J. & Regenberg, B. Global transcriptional and physiological responses of *Saccharomyces cerevisiae* to ammonium, L-alanine, or L-glutamine limitation. *Appl. Environ. Microbiol.* **72**, 6194–6203 (2006).
- World Health Organization *WHO Publishes List of Bacteria for Which New Antibiotics Are Urgently Needed* (WHO, 2017).
- Kmieciewicz, Z. Few novel antibiotics in the pipeline, WHO warns. *Br. Med. J. (Online)* **358** (2017).
- Carlet, J. et al. Society's failure to protect a precious resource: antibiotics. *Lancet* **378**, 369–371 (2011).
- Bartell et al. Reconstruction of the metabolic network of *Pseudomonas aeruginosa* to interrogate virulence factor synthesis. *Nat. Commun.* **8**, 1–13 (2017).
- Kwon, T., Huse, H. K., Vogel, C., Whiteley, M. & Marcotte, E. M. Protein-to-mRNA ratios are conserved between *Pseudomonas aeruginosa* strains. *J. Proteome Res.* **13**, 2370–2380 (2014).
- LaBauve, A. E. & Wargo, M. J. Growth and laboratory maintenance of *Pseudomonas aeruginosa*. *Curr. Protoc. Microbiol.* **25**, 6E–1 (2012).
- Oberhardt, M. A., Puchalka, J., Fryer, K. E., Dos Santos, V. A. M. & Papin, J. A. Genome-scale metabolic network analysis of the opportunistic pathogen *Pseudomonas aeruginosa* PAO1. *J. Bacteriol.* **190**, 2790–2803 (2008).
- Liberati, N. T. et al. An ordered, nonredundant library of *Pseudomonas aeruginosa* strain PA14 transposon insertion mutants. *Proc. Natl Acad. Sci. USA* **103**, 2833–2838 (2006).
- Jacobs, M. A. et al. Comprehensive transposon mutant library of *Pseudomonas aeruginosa*. *Proc. Natl Acad. Sci. USA* **100**, 14339–14344 (2003).
- Lieven, C. et al. MEMOTE for standardized genome-scale metabolic model testing. *Nat. Biotechnol.* **38**, 272–276 (2020).
- Finney, A. & Hucka, M. Systems biology markup language: Level 2 and beyond. *Biochem. Soc. Trans.* **31**, 1472–1473 (2003).
- Keating, S. M. et al. SBML Level 3: an extensible format for the exchange and reuse of biological models. *Mol. Syst. Biol.* **16**, e9110 (2020).
- Falla, J. A., Bauda, P. & Block, J. C. Isolation of cell enveloped layers of *Pseudomonas fluorescens*. *J. Microbiological Methods* **7**, 285–294 (1988).
- Yeager, C. M., Bottomley, P. J. & Arp, D. J. Requirement of DNA repair mechanisms for survival of *Burkholderia cepacia* G4 upon degradation of trichloroethylene. *Appl. Environ. Microbiol.* **67**, 5384–5391 (2001).
- Huszczynski, S. M., Lam, J. S. & Khursigara, C. M. The role of *Pseudomonas aeruginosa* lipopolysaccharide in bacterial pathogenesis and physiology. *Pathogens* **9**, 6 (2020).
- Kanehisa, M., Furumichi, M., Sato, Y., Ishiguro-Watanabe, M. & Tanabe, M. KEGG: integrating viruses and cellular organisms. *Nucleic Acids Res.* **49**, D545–D551 (2021).
- Seaver, S. M. D. et al. The ModelSEED Biochemistry Database for the integration of metabolic annotations and the reconstruction, comparison and analysis of metabolic models for plants, fungi and microbes. *Nucleic Acids Res.* **49**, D575–D588 (2021).
- Henry, C. S. et al. High-throughput generation, optimization and analysis of genome-scale metabolic models. *Nat. Biotechnol.* **28**, 977–982 (2010).
- Li, G. & Lu, C. D. The cryptic *dsdA* gene encodes a functional D-Serine dehydratase in *Pseudomonas aeruginosa* PAO1. *Curr. Microbiol.* **72**, 788–794 (2016).
- Blazier, A. S. & Papin, J. A. Reconciling high-throughput gene essentiality data with metabolic network reconstructions. *PLoS Computational Biol.* **15**, e1006507 (2019).
- Monk, J. M. et al. iML1515, a knowledgebase that computes *Escherichia coli* traits. *Nat. Biotechnol.* **35**, 904–908 (2017).
- Tan, S. Y. et al. Identification of five structurally unrelated quorum-sensing inhibitors of *Pseudomonas aeruginosa* from a natural-derivative database. *Antimicrobial Agents Chemother.* **57**, 5629–5641 (2013).
- Lee et al. Genomic analysis reveals that *Pseudomonas aeruginosa* virulence is combinatorial. *Genome Biol.* **7**, 1–14. (2006).
- Jenior, M. L., Moutinho, T. J. Jr, Dougherty, B. V. & Papin, J. A. Transcriptome-guided parsimonious flux analysis improves predictions with metabolic networks in complex environments. *PLoS Computational Biol.* **16**, e1007099 (2020).
- Bonser, L. R. & Erle, D. J. Airway mucus and asthma: the role of MUC5AC and MUC5B. *J. Clin. Med.* **6**, 112 (2017).
- Roy et al. Muc5B is required for airway defence. *Nature* **505**, 412–416 (2014).
- Hosseini, E., Grootaert, C., Verstraete, W. & Van de Wiele, T. Propionate as a health-promoting microbial metabolite in the human gut. *Nutr. Rev.* **69**, 245–258 (2011).
- El Hage, R., Hernandez-Sanabria, E., Calatayud Arroyo, M., Props, R. & Van de Wiele, T. Propionate-producing consortium restores antibiotic-induced dysbiosis in a dynamic in vitro model of the human intestinal microbial ecosystem. *Front. Microbiol.* **10**, 1206 (2019).
- Vinolo, M. A., Rodrigues, H. G., Nachbar, R. T. & Curi, R. Regulation of inflammation by short chain fatty acids. *Nutrients* **3**, 858–876 (2011).
- Raman, K. & Chandra, N. Flux balance analysis of biological systems: applications and challenges. *Brief. Bioinforma.* **10**, 435–449 (2009).
- Carey, M. A., Dräger, A., Beber, M. E., Papin, J. A. & Yurkovich, J. T. Community standards to facilitate development and address challenges in metabolic modeling. *Mol. Syst. Biol.* **16**, e9235 (2020).
- Römer et al. ZBIT bioinformatics toolbox: a web-platform for systems biology and expression data analysis. *PLoS ONE* **11**, e0149263 (2016).
- Norsigian et al. BiGG Models 2020: multi-strain genome-scale models and expansion across the phylogenetic tree. *Nucleic Acids Res.* **48**, D402–D406 (2020).
- Bornstein, B. J., Keating, S. M., Jouraku, A. & Hucka, M. LibSBML: an API library for SBML. *Bioinformatics* **24**, 880–881 (2008).
- Juty, N., Le Novère, N. & Laibe, C. Identifiers.org and MIRIAM Registry: community resources to provide persistent identification. *Nucleic Acids Res.* **40**, D580–D586 (2012).
- Moretti, S., Tran, V. D. T., Mehl, F., Ibberson, M. & Pagni, M. MetaNetX/MNXref: unified namespace for metabolites and biochemical reactions in the context of metabolic models. *Nucleic Acids Res.* **49**, D570–D574 (2021).
- Sayers et al. Database resources of the national center for biotechnology information. *Nucleic Acids Res.* **47**, D23 (2019). Database issue.
- The UniProt Consortium. UniProt: The universal protein knowledgebase in 2021. *Nucleic Acids Res.* **49**, D480–D489 (2021).

51. Wattam et al. Improvements to PATRIC, the all-bacterial bioinformatics database and analysis resource center. *Nucleic Acids Res.* **45**, D535–D542 (2017).
52. Courtot et al. Controlled vocabularies and semantics in systems biology. *Mol. Syst. Biol.* **7**, 543 (2011).
53. Finney, A., Hucka, M., & Le Novère, N. *Systems Biology Markup Language (SBML) Level 2: Structures and Facilities for Model Definitions*. Retrieved from COMBINE (2006), <http://identifiers.org/combine.specifications/sbml.level-2.version-2>.
54. Hucka et al. (2018). The systems biology markup language (SBML): Language specification for level 3 version 1 core. *J. Integr. Bioinform.* **15** (2018).
55. Olivier, B. G. & Bergmann, F. T. SBML level 3 package: flux balance constraints version 2. *J. Integr. Bioinform.* **15** (2018).
56. Hucka, M. & Smith, L. P. SBML level 3 package: groups, version 1 release 1. *J. Integr. Bioinform.* **13**, 290 (2016).
57. Caspi et al. The MetaCyc database of metabolic pathways and enzymes—a 2019 update. *Nucleic Acids Res.* **48**, D445–D453 (2020).
58. Matthews, B. W. Comparison of the predicted and observed secondary structure of T4 phage lysozyme. *Biochim. Biophys. Acta (BBA)-Protein Struct.* **405**, 442–451 (1975).
59. Dunphy, L. J., Yen, P. & Papin, J. A. Integrated experimental and computational analyses reveal differential metabolic functionality in antibiotic-resistant *Pseudomonas aeruginosa*. *Cell Syst.* **8**, 3–14 (2019).
60. Darveau, R. P. & Hancock, R. E. Procedure for isolation of bacterial lipopolysaccharides from both smooth and rough *Pseudomonas aeruginosa* and *Salmonella typhimurium* strains. *J. Bacteriol.* **155**, 831 (1983).
61. Winsor et al. Enhanced annotations and features for comparing thousands of *Pseudomonas* genomes in the *Pseudomonas* genome database. *Nucleic Acids Res.* **44**, D646–D653 (2016).
62. Oksanen et al. Community ecology package. *R package version, 2* (2013).
63. Liaw, A. & Wiener, M. Classification and regression by randomForest. *R. N.* **2**, 18–22 (2002).
64. Hardle, W. K. & Simar, L. In *Applied Multivariate Statistical Analysis* (eds. Hardle, W. K. & Simar, L.) 3rd edn. (Springer, 2012).
65. Malik-Sheriff, R. S. et al. "BioModels—15 years of sharing computational models in life science". *Nucleic Acids Res.* **48**, D407–D415 (2020).
66. Bergmann, F. T. et al. "COMBINE archive and OMEX format: One file to share all information to reproduce a modeling project". *BMC Bioinform.* **15**, 369 (2014).
67. Neal, M. L. et al. "Harmonizing semantic annotations for computational models in biology". *Brief. Bioinform.* **ume 20**, 540–550 (2019).

ACKNOWLEDGEMENTS

The authors would like to thank our colleagues Jennifer Bartell and Anna Blazier for their previous work on iPau1129 and their helpful suggestions for our curation of iPau21. This work is supported with U.S. federal funds from the National Institutes of Health (R01 AI154242) and funded by the *Deutsche Forschungsgemeinschaft* (DFG, German Research Foundation) under Germany's Excellence Strategy—EXC 2124–390838134 within the Cluster of Excellence CMFI (Controlling Microbes to

Fight Infections). A.D. is funded by the Germany Center for Infection Research (DZIF) within the *Deutsche Zentren der Gesundheitsforschung* (BMBF-DZG, Germany Centers for Health Research of the Federal Ministry of Education and Research), grant no. 8020708703.

AUTHOR CONTRIBUTIONS

L.D. and J.P. conceived the project. D.P., A.R., L.D. and T.L. researched for GENRE updates. D.P., A.R. and L.D. prepared the GENRE. D.P. ran simulations with the GENRE and produced/analyzed contextualized models. D.P., L.D. and J.P. interpreted the results. D.P., A.R. and L.D. wrote the manuscript. All authors critically revised the manuscript and approved the final version.

COMPETING INTERESTS

The authors declare no competing interests.

ADDITIONAL INFORMATION

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41540-021-00198-2>.

Correspondence and requests for materials should be addressed to Jason A. Papin.

Reprints and permission information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021