Computational Systems Biology and *In Silico* Modeling for Infectious Disease Intervention

Dissertation

der Mathematisch-Naturwissenschaftlichen Fakultät
der Eberhard Karls Universität Tübingen
zur Erlangung des Grades eines
Doktors der Naturwissenschaften
(Dr. rer. nat.)

vorgelegt von
M. Sc. Nantia Leonidou
aus Limassol, Zypern

Tübingen

2024

To my beloved family

∾

Abstract

Infectious diseases, exceeding political and geographical boundaries, persist as significant health threats that burden populations globally. Since ancient times, they have consistently challenged the growth and welfare of all living organisms, including humans. Additionally, the emergence of antimicrobial resistance complicates disease management, diminishing the efficacy of conventional treatments and presenting notable challenges in prevention and control. Recent viral pandemics, such as COVID-19, highlight the urgent need for effective worldwide preparedness and new therapeutic strategies. In response to these challenges, computational systems biology and *in silico* modeling have emerged as powerful assets in combating infectious diseases. Genotype-driven prediction of cellular phenotypes is essential to understand how genetic variations influence disease outcomes, guide personalized treatments, and identify potential drug targets. In this regard, constraint-based modeling methods have evolved to facilitate a mechanistic understanding of metabolic physiology based on imposed constraints and experimental data. This thesis uses systems biology to study the genotype–phenotype relationship of cellular metabolism. More specifically, it explores the applications of constraint-based modeling across multiple projects, aiming to address different infections and antimicrobial-resistant pathogens.

The primary focus of the first project is the antiviral target prediction, leveraging systems biology and constraint-based modeling. This project involves the reconstruction and analysis of cell-specific host-virus metabolic models to detect exploitable inhibitory pathways. The inhibitory efficacy of the identified compounds was also computationally validated across multiple known variants. Nevertheless, the rise in complexity of mathematical models raises problems in quality control and reproducibility. The second project focuses on model standardization and documentation, aiming to enhance the reusability and interoperability of computational models. It presents a Python-based annotation software to automate the assignment of systems biology ontology terms by inferring expert knowledge from model structures. The third project explores the metabolic modeling of human nasal microbiota by focusing on a particular nosocomial pathogenic member named *Acinetobacter baumannii*. The construction of the first collection of metabolic networks for various multidrug-resistant *A. baumannii* strains is presented. Their analysis yields valuable insights into strain-specific metabolic capabilities and vulnerabilities, that could guide the development of novel antimicrobial strategies targeting metabolic pathways. Finally, the last project investigates the metabolic phenotypes of *Rothia mucilaginosa*, an opportunistic pathogen with a multifaceted role in health and disease. The integration of experimental data alongside the construction of the first manually curated metabolic network aim to unravel potential metabolic aspects that could serve as focal points for future research endeavors.

The research presented in this thesis signifies meaningful strides in infectious disease intervention via computational systems biology and metabolic modeling. Through a combination of modeling, data analysis, experiments, and software development, this thesis effectively offers comprehensive insights into previously unexplored areas, thereby advancing the understanding and application of systems biology principles in various fields.

Zusammenfassung

Infektionskrankheiten, die politische und geografische Grenzen überschreiten, stellen nach wie vor eine erhebliche Gesundheitsbedrohung dar und belasten die Bevölkerung weltweit. Seit der Antike stellen sie immer wieder eine Herausforderung für das Wachstum und Wohlergehen aller lebenden Organismen, einschließlich des Menschen, dar. Darüber hinaus erschwert das Auftreten antimikrobieller Resistenzen das Krankheitsmanagement, verringert die Wirksamkeit herkömmlicher Behandlungen und stellt die Prävention und Kontrolle vor große Herausforderungen. Neueste virale Pandemien, wie COVID-19, unterstreichen die dringende Notwendigkeit wirksamer weltweiter Bereitschafts- und neue Therapiestrategien. Als Reaktion auf diese Herausforderungen haben sich die computergestützte Systembiologie und die *In-silico*-Modellierung als leistungsstarke Hilfsmittel zur Bekämpfung von Infektionskrankheiten erwiesen. Die Vorhersage zellulärer Phänotypen auf der Grundlage von Genotypen ist unerlässlich, um zu verstehen, wie genetische Variationen den Krankheitsverlauf beeinflussen, personalisierte Behandlungen anleiten und potenzielle Angriffspunkte für Medikamente identifizieren. In diesem Zusammenhang haben sich einschränkungsbasierte Modellierungsmethoden herausgebildet, die ein mechanistisches Verständnis der Stoffwechselphysiologie auf der Grundlage vorgegebener Einschränkungen und experimenteller Daten ermöglichen. In dieser Arbeit wird die Systembiologie eingesetzt, um die Genotyp-Phänotyp-Beziehung des Zellstoffwechsels zu untersuchen. Insbesondere werden die Anwendungen der einschränkungsbasierten Modellierung in mehreren Projekten zur Bekämpfung verschiedener Infektionen und antimikrobiell resistenter Krankheitserreger untersucht.

Der Hauptschwerpunkt des ersten Projekts ist die Vorhersage antiviraler Ziele, wobei Systembiologie und einschränkungsbasierte Modellierung zum Einsatz kommen. Dieses Projekt umfasst die Rekonstruktion und Analyse von zellspezifischen Wirts-Virus-Stoffwechselmodellen, um ausnutzbare Hemmungspfade zu ermitteln. Die hemmende Wirkung der identifizierten Verbindungen wurde auch rechnerisch für mehrere bekannte Varianten validiert. Die zunehmende Komplexität mathematischer Modelle führt jedoch zu Problemen bei der Qualitätskontrolle und Reproduzierbarkeit. Das zweite Projekt konzentriert sich auf die Modellstandardisierung und -dokumentation, um die Wiederverwendbarkeit und Interoperabilität von Computermodellen zu verbessern. Es stellt ein auf Python basierende Annotationssoftware vor, um die Zuordnung von Begriffen der Systembiologie-Ontologie zu automatisieren, indem Expertenwissen aus Modellstrukturen abgeleitet wird. Das dritte Projekt erforscht die metabolische Modellierung der menschlichen Nasenmikrobiota und konzentriert sich dabei auf einen bestimmten nosokomialen pathogenen Vertreter namens *Acinetobacter baumannii*. Es wird die erste Sammlung von Stoffwechselnetzwerken für verschiedene multiresistente *A. baumannii*-Stämme vorgestellt. Ihre Analyse liefert wertvolle Erkenntnisse über stammspezifische metabolische Fähigkeiten und Schwachstellen, die die Entwicklung neuer antimikrobieller Strategien, die auf Stoffwechselwege abzielen, unterstützen könnten. Das letzte Projekt schließlich untersucht die Stoffwechselphänotypen von *Rothia mucilaginosa*, einem opportunistischen Krankheitserreger mit einer vielschichtigen Rolle in Gesundheit und Krankheit. Die Integration experimenteller Daten sowie die Erstellung des ersten manuell kuratierten metabolischen Netzwerks zielen darauf ab, potenzielle metabolische Aspekte zu entschlüsseln, die als Schwerpunkte für zukünftige Forschungsbemühungen dienen könnten.

Die in dieser Arbeit vorgestellten Forschungsergebnisse stellen einen bedeutenden Fortschritt

bei der Bekämpfung von Infektionskrankheiten durch computergestützte Systembiologie und Stoffwechselmodellierung dar. Durch eine Kombination aus Modellierung, Datenanalyse, Experimenten und Softwareentwicklung bietet diese Arbeit umfassende Einblicke in bisher unerforschte Bereiche und fördert so das Verständnis und die Anwendung systembiologischer Prinzipien in verschiedenen Bereichen.

Acknowledgments

Contents

List of Abbreviations

ADP	adenosine 5'-diphosphate
AGORA	Assembly of Gut Organisms through Reconstruction and Analysis
AIDS	acquired immunodeficiency syndrome
AMP	adenosine 5'-monophosphate
AMR	antimicrobial resistance
API	Application Programming transfer Interface
ATP	adenosine triphosphate
AUC	area under curve
BHI	brain heart infusion
BiGG	Biochemical, Genetical, and Genomical
BLAST	Basic Local Alignment Search Tool
BOF	biomass objective function
CAT	catalase
CBM	constraint-based modeling
CF	cystic fibrosis
CFTR	cystic fibrosis transmembrane conductance regulator
CoA	coenzyme A
COBRA	constraint-based reconstruction and analysis
COBRApy	constraint-based reconstruction and analysis for Python
COVID-19	Coronavirus Disease 2019
CPEC	cyclopentenyl cytosine
CPEU	cyclopentenyl uridine
CTP	cytidine 5'-triphosphate
CTPS1	cytidine 5'-triphosphate synthase 1
DNA	deoxyribonucleic acid
EC	Enzyme Commission
ECO	Evidence and Conclusion Ontology
EDR	energy dissipation reaction
EGC	energy-generating cycle
EPSP	enolpyruvylshikimate phosphate
FAIR	Findable, Accessible, Interoperable, and Reusable
FASTCC	fast consistency check
FBA	flux balance analysis
fbc	flux balance constraints
FC	fold change
FDA	Food and Drug Administration
FVA	flux variability analysis
GEM	genome-scale metabolic model

GISAID	Global Initiative on Sharing All Influenza Data
GK1	guanylate kinase 1
GMP	guanosine 5'-monophosphate
GMPS2	guanosine 5'-monophosphate synthase
GPR	gene-protein-reaction association
HBEC	human bronchial epithelial cell
HDE	host-derived enforcement
HIV	human immunodeficiency virus
HMR	Human Metabolic Reaction
ICU	intensive care unit
ID	identifier
IMP	inosine 5'-monophosphate
IMPD	inosine 5'-monophosphate dehydrogenase
KDPG	2-keto-3-deoxy-6-phosphogluconate
KEGG	Kyoto Encyclopedia of Genes and Genomes
LB	lysogeny broth
LP	linear programming
M9	M9 minimal medium
mAB	monoclonal antibodies
mCADRE	metabolic Context-specificity Assessed by Deterministic Reaction Evaluation
MDR	multidrug-resistant
MEMOTE	metabolic model testing
MIRIAM	Minimal Information Required In the Annotation of Models
MOMA	minimization of metabolic adjustment
mRNA	messenger Ribonucleic Acid
MRSA	methicillin-resistant strains of *Staphylococcus aureus*
NAFLD	non-alcoholic fatty liver disease
NDPK2	nucleoside diphosphate kinase 2
NDPK3	nucleoside diphosphate kinase 3
NSP	non-structural protein
OD	optical density
ODE	ordinary differential equation
OMEX	Open Modeling EXchange format
PATRIC	Pathosystems Resource Integration Center
PEP	phosphoenolpyruvate
pFBA	parsimonious enzyme usage flux balance analysis
PHEIC	public health emergency of international concern
PM	phenotype microarray
PPI	protein-protein interaction
PREDICATE	Prediction of Antiviral Targets
RNA	ribonucleic acid
ROS	reactive oxygen species
RPMI	Roswell Park Memorial Institute
SARS	severe acute respiratory syndrome

SARS-CoV	severe acute respiratory syndrome coronavirus
SARS-CoV-2	severe acute respiratory syndrome coronavirus 2
SBFC	Systems Biology Format Converter
SBGN	Systems Biology Graphical Notation
SBML	Systems Biology Markup Language
SBO	Systems Biology Ontology
SBSCL	Systems Biology Simulation Core Library
SCFM	synthetic cystic fibrosis sputum medium
SNM	synthetic nasal medium
SPODM	superoxide dismutase
SQL	Structured Query Language
SULFCYS	cysteine desulfurase
TSB	tryptic soy broth
UMP	uridine 5'-monophosphate
UMPK5	uridine 5'-monophosphate kinase 5
UTP	uridine 5'-triphosphate
VBOF	viral biomass objective function
VMH	Virtual Metabolic Human
VOC	variants of concern
VZV	varicella zoster virus
WHO	World Health Organization
XML	Extensible Markup Language
XMP	xanthosine 5'-phosphate

List of Publications

1. **New workflow predicts drug targets against SARS-CoV-2 via metabolic changes in infected cells**
   Nantia Leonidou, Alina Renz, Reihaneh Mostolizadeh, and Andreas Dräger
   PLoS Comput Biol Volume 19 (2023).
   https://doi.org/10.1371/journal.pcbi.1010903

2. **SBOannotator: a Python tool for the automated assignment of systems biology ontology terms**
   Nantia Leonidou, Elisabeth Fritze, Alina Renz, and Andreas Dräger
   Bioinformatics Volume 39, Issue 7 (2023).
   https://doi.org/10.1093/bioinformatics/btad437

3. **Exploring the metabolic profiling of *A. baumannii* for antimicrobial development using genome-scale modeling**
   Nantia Leonidou, Yufan Xia, Lea Friedrich, Monika S. Schütz, and Andreas Dräger
   PLoS Pathog Volume 20 (2024).
   https://doi.org/10.1371/journal.ppat.1012528

4. **Genome-scale model of *Rothia mucilaginosa* predicts gene essentialities and reveals metabolic capabilities**
   Nantia Leonidou, Lisa Ostyn, Tom Coenye, Aurélie Crabbé, and Andreas Dräger
   Microbiology Spectrum (2023).
   https://doi.org/10.1128/spectrum.04006-23

5. **Metabolic modeling elucidates phenformin and atpenin A5 as broad-spectrum antiviral drugs**
   Alina Renz[*], Mirjam Hohner[*], Maximilian Breitenbach, Jonathan Josephs-Spaulding, Johanna Dürrwald, Lena Best, Victoria Dulière, Chloé Mialon, Stefanie M. Bader, Raphaël Jami, Georgios Marinos, Nantia Leonidou, Filipe Cabreiro, Marc Pellegrini, Marcel Doerflinger, Manuel Rosa-Calatrava[&], Andrés Pizzorno[&], Andreas Dräger[$], Michael Schindler[$], and Christoph Kaleta[$]
   preprints.org (2023). [*], [&], [$] These authors contributed equally.
   https://doi.org/10.20944/preprints202210.0223.v3

6. **Genome-scale metabolic model of *Staphylococcus epidermidis* ATCC 12228 matches *in vitro* conditions**
   Nantia Leonidou[*], Alina Renz[*], Benjamin Winnerling, Anastasiia Grekova, Fabian Grein, and Andreas Dräger
   bioRxiv (2023). [*] These authors contributed equally.
   https://doi.org/10.1101/2023.12.19.572329

CHAPTER 1

Introduction

The organism is not just a heap of its component parts, but must be regarded as an organized whole with properties which cannot be predicted from those of the separate parts.
– Ludwig von Bertalanffy

The 21st century has introduced a myriad of challenges and crises that span a broad spectrum, including environmental, economic, technological, and social domains. Economic inequalities deepen, contributing to social unrest and unequal resource access, while political instability and resource scarcity compound these complexities. However, amidst this multifaceted landscape, the overarching seriousness lies in the hazards to global health. The Coronavirus Disease 2019 (COVID-19) pandemic clearly illustrated the interconnectedness of these challenges, underscoring how environmental degradation, economic imbalances, and scarce global governance can exacerbate health crises. The rapid spread of infectious diseases, coupled with the potential for their escalation into pandemics, has become a pressing concern. The rise of antimicrobial resistance further complicates disease management, rendering conventional treatments less effective and increasing the risk of prolonged illness and higher mortality. Simultaneously, climate change poses an imminent threat, with its cascading effects on ecosystems, weather patterns, and transmission dynamics of infectious diseases[1]. In the face of a plethora of 21st-century challenges, there is a shared acknowledgment that interdisciplinary approaches and collaborative efforts are imperative for protecting global health and cultivating resilience against prospective threats.

1.1 Chronicles of infectious diseases

Throughout the historical records, diseases and illnesses have persistently afflicted humanity. From ancient times to the present day, humankind has faced a multitude of infectious diseases, each leaving a permanent mark on the trajectory of societal and medical development. The transition from nomadic hunter-gatherer lifestyles to settled agrarian societies (known as the Neolithic Revolution or First Agricultural Revolution) has facilitated the propagation of infectious diseases within human populations[4]. This shift provided secured food supplies, promoted social organization and cultural developments, and initiated trade networks. However, these Neolithic lifeways, combined with the increasing population density, contributed crucially to creating environments that favored the transmission and prevalence of various diseases[5, 6]. New agricultural practices, including animal domestication, supported interactions between humans
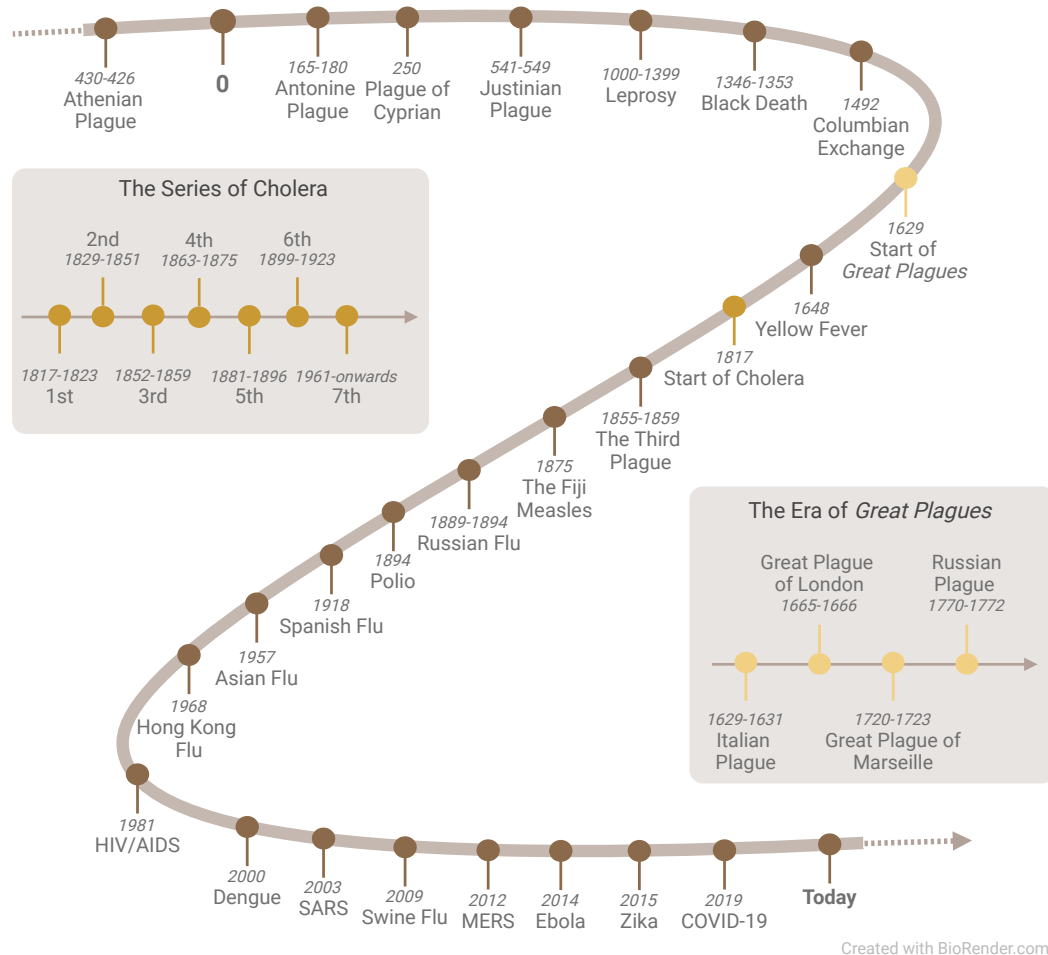
Figure 1.1: Historical timeline of infectious disease outbreaks. A visual odyssey through ancient plagues to modern-world diseases. While certain infectious diseases have historical evidence dating back to earlier times, this figure focuses on their onset and emergence. For instance, leprosy, with osteo-archaeological evidence traced back to 2000 B. C. in India, reached its zenith during the Crusades and European knights' era between the 11th and 14th centuries[2, 3]. In light of the extensive historical record of pandemics and outbreaks, the future of infectious diseases remains unpredictable. The lessons learned from the past emphasize the importance of continuous alertness, international collaboration, and investment in research and preparedness.

and animals, spreading vector-borne and zoonotic diseases[7, 8]. Furthermore, the storage and surplus of food in agrarian societies attracted pests and rodents, serving as additional vectors for disease transmission. These interconnected factors highlight how the agricultural evolution from foraging created conditions conducive to the dissemination of infectious diseases within human populations[9]. An extensive visual exploration of infectious diseases spanning prehistoric to modern epochs is illustrated in Fig. 1.1.

The Athenian Plague of 430 B. C., occurred during the Peloponnesian War and is the earliest historically recorded pandemic. Thucydides, a survivor of this plague, provided a detailed historical narrative of the outbreak, which originated in Ethiopia and rapidly propagated through Egypt and Greece[10]. Centuries later, the Antonine Plague (or the Plague of Galen), weakened the Roman Empire[11], while in the mid-third century, the Plague of Cyprian occurred during Emperor Carus' reign[12]. The causative agents of these historical pandemics remain uncertain, with possibilities including smallpox, measles, and typhus[13–15]. In A. D. 541, the Justinianic Plague, caused by *Yersinia pestis*, hit the Byzantine Empire with extreme weather and trade-aiding transmission[16]. This pandemic, lacking preventive measures, led to widespread mortality and economic decline[17, 18]. The medieval Black Death outbreak marked the onset of an almost five-century-long pandemic known as the second plague pandemic[19]. This event ranks among human history's most consequential infectious disease disasters, leading to substantial depopulation. Known as "bubonic plague" due to the swellings on infected bodies, it caused the death of hundreds of millions and nearly 60 % of the western Eurasian population within eight years[19]. Moreover, the pandemic had a substantial and long-lasting influence on demography and socioeconomic conditions[20]. Between the 17th and 18th centuries, a series of "great plagues" ravaged the European countries (Fig. 1.1). Transmitted by fleas infesting rats, these events left an enduring mark on history, shaping public health strategies[21, 22]. It is proposed that the bacteria might have persisted in rodent reservoirs across Europe, experiencing intermittent re-emergence within the human population over time[23]. Prior to these plagues, during the Columbian Exchange, the inadvertent introduction of Old World diseases to the New World had catastrophic consequences for indigenous populations[24]. Diseases like smallpox, measles, typhus, and cholera caused widespread mortality.

Leprosy, caused by *Mycobacterium leprae*, has been historically documented since 2000 B. C. in India[2], reaching its peak between the 11th and 14th centuries[3]. Anthropologists in 2009 analyzed a 4,000-year-old Indian skeleton, providing the earliest archaeological proof of *M. leprae* infection. Their findings marked a significant milestone in understanding the disease's prehistoric presence in India[2]. Throughout history, leprosy has led to the complete isolation of affected individuals, with sites like Spinalonga in Crete gaining notoriety as isolation centers for leprosy patients in Greece[25]. The WHO recently released alarming data in conjunction with World Leprosy Day. The report indicates that the disease persists in over 120 countries, with an annual report of over 200,000 new cases[26]. Cholera, caused by *Vibrio cholerae*, is an additional bacterial infection and an acute gastrointestinal disease transmitted through contaminated water and food[27, 28]. The seventh cholera pandemic, originating in Indonesia in 1961, has led to major epidemics in countries like Haiti and Zimbabwe[29, 30].

Nowadays, the exploration of ancient infectious diseases has been revolutionized by technological progress, particularly through novel palaeogenetic and palaeoproteomic approaches[31]. A notable and recent example of this advancement is elucidating the origins of the Black Death, whose geographical origin has long puzzled scientists, with potential locations spanning from Western Eurasia to East Asia and China. In 2022, Spyrou et al. identified the Tian Shan region in present-day Kyrgyzstan as the most likely epicenter of the plague[32]. Archaeologists unearthed

numerous graves dating back to 1338–1339, containing inscriptions in Syriac language linking the cause of death to the plague. Genetic analysis of deoxyribonucleic acid (DNA) from the teeth of seven skeletons in two cemeteries in the area, coupled with the detection of *Y. pestis* in three of them, supported this conclusion. Altogether, integrating molecular evidence with conventional methods enables researchers to reassess established assumptions. This approach allows for the analysis of patterns within human genomes, yielding valuable insights into prehistoric populations and opening novel research avenues.

1.2 Outbreaks in the 20th and 21st centuries

From the 19th century onward, following the discovery of the tobacco mosaic virus in 1892, viral pandemics became increasingly prevalent[33]. In recent years, the rising trends in global and regional air travel, trade, and urbanization have intensified the frequent movement of people and goods. This influx not only favors the emergence of problematic pathogens but also enhances the likelihood of novel variants evolving, endangering global health[34].

The "Russian flu" pandemic spanned five years and was the first well-described respiratory pandemic[35]. Two and a half decades later, a series of flu pandemics erupted, with the first being the "Spanish flu", caused by the H1N1 subtype of the influenza A virus (A/H1N1) and claiming over 50 million people worldwide[36]. Further significant flu pandemics that marked the 20th century include the "Asian flu" (A/H2N2) and the "Hong Kong flu" (A/H3N2), both originating from genetic reassortment[37]. A further descendant of the 1918 pandemic virus emerged in 2009 as a triple reassortment of human, avian, and swine influenza viruses[38]. Although the World Health Organization (WHO) reported 18,631 deaths confirmed by laboratory tests, a study from 2013 re-estimated the mortality rates[39]. Their calculations revealed a substantial pandemic mortality burden, attributing between 148,000 and 249,000 respiratory fatalities to influenza in an average non-pandemic period. The HIV/AIDS outbreak, which first gained recognition in the early 1980s, elicited considerable public concern as it progressed to acquired immunodeficiency syndrome (AIDS) and eventually resulted in fatalities. In 2022, approximately 630,000 individuals died from AIDS-associated illnesses, and nearly 40 million people were infected[40]. The response to HIV/AIDS has since involved extensive research, medical advancements, and public health initiatives, making it a pivotal chapter in the ongoing pursuit of understanding and combating infectious diseases[41]. Moreover, the Ebola virus, originated in 1976 within the Democratic Republic of Congo and Sudan, causing severe hemorrhagic fever with high death rates[42]. The virus, transmitted to humans from wildlife (fruit bats), led to sporadic outbreaks in Central Africa and a severe public health crisis in 2014, with its epicenter in West Africa[43].

In late December 2019, the world witnessed the onset of a significant global outbreak characterized by clustered cases of pneumonia of unknown etiology. These cases were reported in Wuhan, the capital of Hubei province in China, marking the beginning of a devastating event with substantial losses. By early 2020, following the exclusion of common respiratory viruses with similar symptomatology, a novel coronavirus, named severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), was found to be responsible for this outbreak[44]. Coronaviruses, part of the *Coronaviridae* family, are enveloped viruses categorized into four genera: alpha-, beta-, gamma-, and delta-coronaviruses[45]. They consist of positive-sense single-stranded ribonucleic acid (RNA) and can infect humans and various animal species. Their genetic makeup is composed of non-structural proteins (NSPs) and four structural proteins, namely the spike (S), nucleocapsid (N), envelope (E), and membrane (M) proteins[45]. The newly emerged *beta-coronavirus*, SARS-CoV-2, exhibits genetic resemblance to severe acute respiratory syndrome

coronavirus (SARS-CoV), with approximately 80 % similarity in their genomes[44]. It belongs to the fourth Baltimore group of viruses, meaning it utilizes a template "-" single RNA antisense strand to synthesize messenger Ribonucleic Acid (mRNA)[46]. Since its discovery, SARS-CoV-2 has spread globally, causing millions of cases and deaths[47]. On March 11, 2020, the WHO declared the pneumonia outbreak a pandemic, terming the resulting disease a COVID-19[48]. This led to extraordinary measures worldwide, such as large-scale quarantines, activity suspensions, and travel restrictions. Notably, the emergence of SARS-CoV-2 has led to re-evaluating historical influenza outbreaks. For years, it was believed that the "Russian flu" was caused by the bacterium, named *Bacillus influenzae* (now *Haemophilus influenzae*)[49], and later by the human *Myxovirus influenzae*[50]. However, scientific investigations from the early 21st century suggested that the pandemic was likely triggered by the human coronavirus OC43 (HCoV–OC43)[51, 52]. Symptoms observed during the "Russian flu", including the distinctive loss of smell and taste, resembled more to COVID-19 than to influenza, while genetic comparisons with OC43 strains supported the observations[51, 52]. This paradigm shift underscores the dynamic nature of our understanding of historical pandemics in the light of contemporary viral insights.

SARS-CoV-2 variants of concerns (VOCs) have been a prominent aspect of the COVID-19 pandemic, introducing complexities to disease dynamics, transmission patterns, and public health strategies[53]. The last VOCs designated by the WHO included the Alpha, Beta, Gamma, Delta, and Omicron variants[54]. Structural protein mutations represent a frequent occurrence, introducing complications in the pathogenesis pathway[55, 56]. Another important aspect of the COVID-19 has been the post-acute sequelae of SARS-CoV-2 infection, or Long COVID. During this condition, individuals experience persistent and long-term symptoms after the acute phase of COVID-19 has resolved[57]. While many people recover within a few weeks, some individuals, including those with mild or asymptomatic initial infections, experience a variety of symptoms persistently over an extended period, lasting for months. The symptoms of Long COVID vary and can afflict multiple organ systems as well as mental health[57–60]. The WHO declared on May 5, 2023, the end of the public health emergency of international concern (PHEIC) status for COVID-19[61]. However, it is crucial to note that this declaration does not signify the elimination of the virus as a considerable threat to global public health. As of April, 2024, the global impact of the pandemic includes over 7 million reported deaths and more than 775 million confirmed cases worldwide[62]. The virus continues to claim lives worldwide, and the potential emergence of new variants remains a constant risk. Ongoing surveillance and genomic sequencing are crucial for promptly identifying emerging variants and assessing their impact on disease dynamics. The dynamic nature of SARS-CoV-2 evolution underscores the importance of adaptability in vaccination strategies and public health interventions to effectively navigate the challenges posed by emerging variants throughout the pandemic.

1.3 The looming threat: navigating the next pandemic stemming from bacterial resistance

The resurgence of infectious diseases once believed eradicated, is a global concern, particularly impacting vulnerable populations and low- to middle-income countries with limited healthcare resources[63]. Additionally, despite being preventable and treatable, numerous bacterial infections remain a substantial threat to public health around the globe. For instance, tuberculosis continues to pose a challenge, with approximately 1.6 million deaths reported in 2022[64]. The infectious diseases spectrum is evolving rapidly in tandem with many shifts influenced by different factors. Prominent determinants acting as driving forces for the (re)emergence of infectious

diseases are social, economic, political, physical, and ecological factors[65, 66]. Climate change represents not only an environmental crisis but also a public health emergency[67]. Another driver of (re)emerging infectious diseases is the constantly increasing population growth, migration, urbanization, and worldwide travel[63, 68, 69].

Most importantly, the excessive and unnecessary use of antimicrobials has fueled the emergence of antimicrobial resistance (AMR) by weakening the efficacy of treatments worldwide. The historical timeline of antibiotic development begins in the early 20th century with the discovery of the arsenic-based pro-drug salvarsan[70], followed by the breakthrough of penicillin[71]. This breakthrough revolutionized bacterial infection treatment, paving the way for new antibiotics. The current AMR crisis has been exacerbated by a gradual decrease in the discovery and development of novel antibiotics, coupled with the evolution of drug resistance among numerous human pathogens. In February 2022, Murray et al. published the first comprehensive and large-scale evaluation of the global AMR burden[72]. For 2019, their predictive statistical models indicated 4.95 million fatalities directly linked to AMR, with a heightened impact observed in low-resource systems. Additionally, pathogen–drug combinations, like the MRSA, have caused over 100,000 deaths attributable to AMR in 2019[72]. Nonetheless, a recent report estimates that AMR is anticipated to result in 10 million deaths annually by 2050[73]. Finally, the increased use of antibiotics as growth-promoting supplements during food-animal production heavily impacts the development of resistance mechanisms and the efficacy of bacterial treatments worldwide[74, 75].

The highly virulent ESKAPE (*Enterococcus faecium*, *Staphylococcus aureus*, *Klebsiella pneumoniae*, *Acinetobacter baumannii*, *Pseudomonas aeruginosa*, and *Enterobacter* spp.) pathogens have been reported to be responsible for 3.57 million AMR-associated deaths[72]. In 2017, the WHO announced a list of bacteria with an urgent need for novel and effective therapeutic strategies[76]. Within this catalog, members of the ESKAPE pathogens were assigned the "critical status", including *Acinetobacter baumannii*. Over the years, extensive research has underscored the threat posed by this Gram-negative pathogen within hospital settings[77, 78]. This concern stems from its inherent resistance to antimicrobial agents, including biocides[79, 80]. *A. baumannii* (from the Greek word *akínētos*, meaning "unmoved") is a non-motile, rod-shaped, and strictly aerobic bacterium. This opportunistic pathogen has a highly adaptable genetic makeup that has led to its endemic presence in intensive care units (ICUs). It primarily affects patients with compromised immune systems, causing various infections such as pneumonia, bacteremia, and endocarditis. The carbapenem-resistant *A. baumannii* is particularly concerning, especially as it poses a serious global threat with high mortality rates[81]. Moreover, it has a propensity to target exposed surfaces and mucous tissues, often colonizing the human nose and showing close associations with infections like those caused by SARS-CoV-2[82–85]. Despite its adaptability, *A. baumannii* has shown susceptibility to conventional drugs, like aminoglycosides, β-lactams, and polymyxins[86].

The looming threat of the next pandemic caused by bacteria is a complex concern going beyond the direct infectious agents themselves. While infectious bacteria undoubtedly constitute a significant risk, the danger is compounded by non-infectious conditions that create an environment conducive to bacterial proliferation. Underlying non-communicable diseases like diabetes, obesity, and cardiovascular conditions compromise the body's immune system, rendering individuals more susceptible to bacterial infections[87, 88]. Additional factors, such as antibiotic overuse and insufficient healthcare infrastructure, can exacerbate the spread of bacterial infections. Cystic fibrosis (CF), caused by mutations in the cystic fibrosis transmembrane conductance regulator (CFTR) gene, is an inherited genetic disorder characterized by autosomal reces-
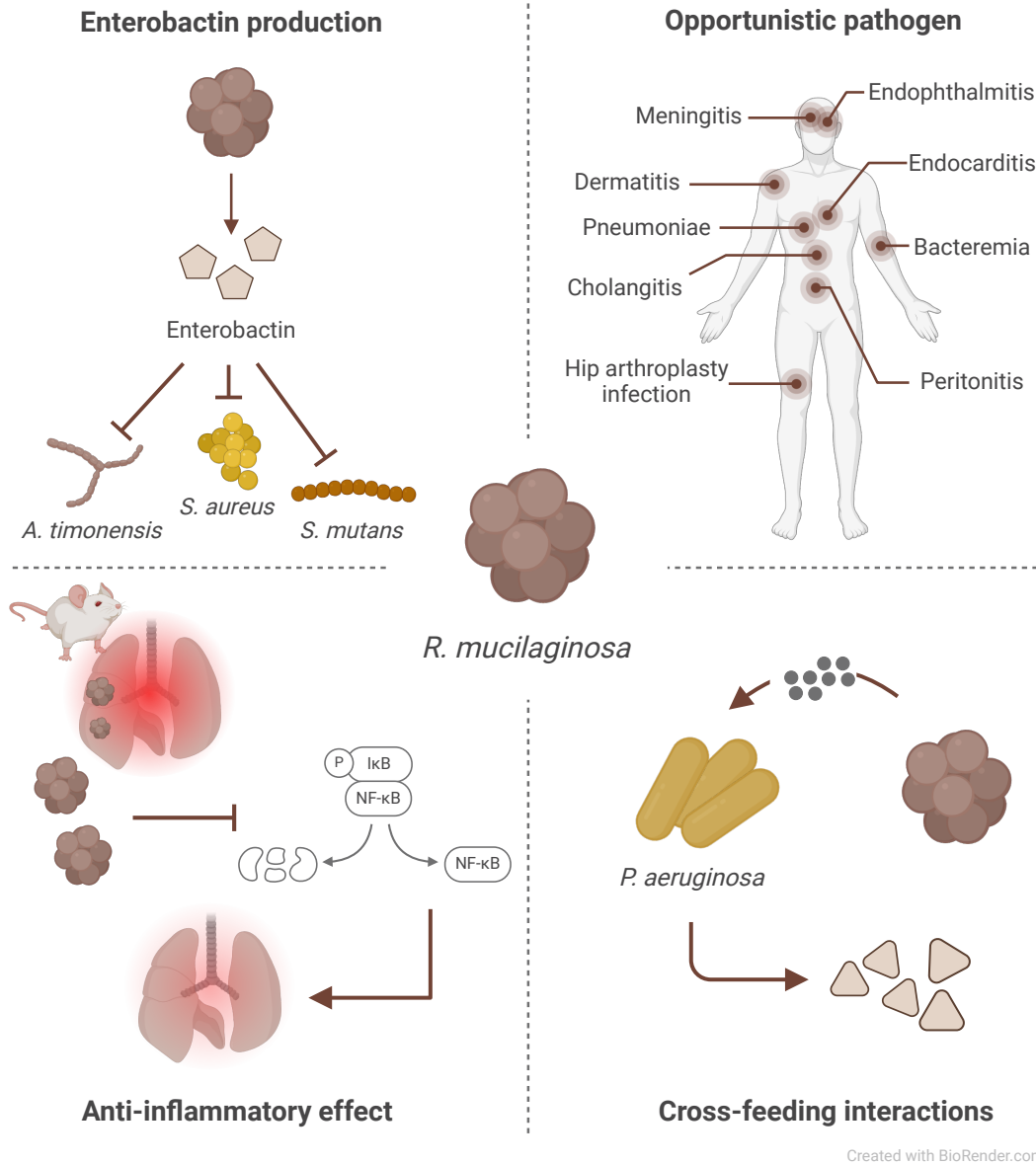
Figure 1.2: The dual role of *R. mucilaginosa* in health and disease. As an opportunistic pathogen, *R. mucilaginosa* acts in individuals with compromised immune functions, leading to severe infections such as endocarditis and sepsis. Concurrently, it produces compounds necessary for *P. aeruginosa* to synthesize essential primary metabolites like glutamate. On the other hand, *R. mucilaginosa* has demonstrated anti-inflammatory properties in the lower respiratory tract, while by producing enterobactin, it exhibited increased virulence against oral microbiota and four methicillin-resistant strains of *Staphylococcus aureus* (MRSA) strains. These findings suggest its potential involvement in both pathogenic and protective mechanisms within the human microbiome.

sive inheritance[89]. This means that the individual inherited two copies of the defective CFTR gene, one copy from each parent. It is a life-limiting disorder that disrupts the transepithelial movement of sodium and chloride ions, resulting in the production of thick and sticky mucus within the airways[90]. It has been estimated that over 160,000 people live with CF worldwide, while 19,516 receive a triple combination of drugs (ivacaftor, tezacaftor, and elexacaftor)[91]. It is characterized as progressive and multisystemic since it leads to ongoing damage and dysfunction of various organ systems, particularly the respiratory and digestive systems[92]. Moreover, the compromised immune clearance creates a hypoxic environment[93], encouraging the colonization of opportunistic microbes, fungi, and viruses[94]. This, leads to the development of persistent and recurrent infections. Chronic respiratory infections are a significant concern for people with cystic fibrosis, and managing them is an essential aspect of CF care.

Several key microbial species play significant roles in CF progression. Early in life, CF environment often involves bacteria such as *S. aureus* and *H. influenzae*[95]. However, chronic infections become more prevalent as the disease progresses and are typically caused by bacteria like *P. aeruginosa* or species within the *Burkholderia cepacia* complex[95]. These microbial species, along with respiratory viruses, are widely recognized as important CF pathogens, driving inflammatory responses and contributing to lung diseases in affected individuals. However, metagenomic sequencing analyses have revealed the high prevalence and increased metabolic activity of a lesser-studied microorganism, called *Rothia mucilaginosa*, in the lungs of CF patients[96, 97]. Guss et al. and Bittar et al. described *R. mucilaginosa* as a "newly" emerged bacterium among CF patients[98, 99]. Additionally, Lim et al. suggested that *R. mucilaginosa* is frequently encountered and metabolically active within the CF airways[100]. *R. mucilaginosa* is a Gram-positive, non-motile, encapsulated, and non-spore-forming bacterium of the *Micrococcaceae* family[101]. It typically grows aerobically but can switch to anaerobic growth through fermentation or alternative pathways, making it an opportunistic microorganism. It is commonly found as a commensal in the oral, lower and upper respiratory tracts, as well as on the skin[101–105].

R. mucilaginosa exhibits a complex role in human health, serving both beneficial and harmful functions, as illustrated in Fig. 1.2. While generally is a harmless commensal, it can become opportunistic, causing infections in immunocompromised individuals. More specifically, it acts as the etiological agent of serious infections such as sepsis, endocarditis, and meningitis[106]. Moreover, Gao et al. demonstrated that *P. aeruginosa* relies on compounds produced by *R. mucilaginosa* to synthesize essential primary metabolites such as glutamate[107]. This symbiotic relationship may contribute to the pathogenesis of *P. aeruginosa* within CF lungs. On the other hand, Rigauts et al. uncovered the anti-inflammatory characteristics of *R. mucilaginosa* within the lower respiratory tract, suggesting a potential impact on the severity of chronic lung diseases[108]. Their work explicitly shows how *R. mucilaginosa* suppresses pro-inflammatory reactions triggered by pathogens or lipopolysaccharides, employing a three-dimensional cell culture model and *in vivo* mouse model. Nonetheless, in 2020, Uranga et al. revealed that *R. mucilaginosa* produces enterobactin, the strongest Fe^{3+}-binding archetypal siderophore known[109]. This contributed to its high virulence against various oral microbiota, including cariogenic organisms like *A. timonensis*, *S. mutans*, and *Streptococcus sp.*, as well as against four MRSA strains. Enterobactin enables scavenging, chelating, and transporting ferric ions from their environment[110]. This ability is crucial for bacterial survival, especially in environments with limited iron availability, as it supports their growth and metabolic processes. Finally, Janek et al. showed that *R. mucilaginosa* is prominently present in the nasal microbiome, underscoring its significant prevalence[111]. Additionally, the researchers observed its susceptibility to specific staphylococcal bacteriocins, suggesting substantial competition with nasal staphylococci and highlighting the influential role of bacteriocins in modulating the nasal microbiota.

1.4 The imperative for pandemic preparedness and global health resilience

The long history of global pandemics has shaped the healthcare systems, medicine, and research, highlighting the importance of timely identification and surveillance of emerging threats. Moreover, COVID-19 demonstrated the remarkable progress possible through collaborative research efforts. This underscores the critical role of sustained investment in scientific research, technological innovation, and expedited vaccine deployment to protect global populations[112]. Vulnerabilities in healthcare systems worldwide have been revealed, calling for the reinforcement of healthcare infrastructure. Additionally, the pandemic illuminated existing health disparities, emphasizing the urgent need for equity in healthcare[113]. Over the last 20 years, medical advancements and improved healthcare have reduced morbidity and mortality from infectious diseases, especially associated with infections in the lower respiratory tract and diarrheal diseases[34]. However, the burden remains high in low- and middle-income countries, particularly for neglected tropical diseases, such as human immunodeficiency virus (HIV), tuberculosis, and malaria[34]. Notably, in 2017, the WHO estimated that at least half of the world still lacks access to essential healthcare[114]. Furthermore, it is crucial to remember that the emergence of COVID-19 was not entirely unforeseen. Scientific studies and warnings dating back to 2007 had identified China as a potential hotspot for a virus outbreak. Factors such as animal-related eating habits, the presence of SARS-CoV-like viruses in horseshoe bats, and the recombination ability of coronaviruses were identified as potential risk factors[115]. Despite these warnings, COVID-19 emerged, underscoring the importance of heeding signs and indicating how ignoring the warnings can lead to devastating consequences.

Looking ahead, the concept of "Disease X" reminds us of the unpredictable nature of infectious diseases. The WHO coined this term to represent a hypothetical, unidentified pathogen capable of triggering a forthcoming epidemic or pandemic[116]. The future threat lies in the emergence and re-emergence of infections. Once again, urbanization, shifting land-use trends, and climate change are driving factors expected to promote disease emergence. These factors are exacerbated by global travel and trade, which facilitate the rapid spread of pathogens. Addressing this evolving landscape necessitates a more forward-looking research within a collaborative global framework dedicated to advancing infectious disease control and research. Measures for pandemic preparedness involve developing and enhancing early warning systems, strengthening international collaboration, investing in research and development, building resilient public health infrastructures, and actively engaging communities. Incorporating the lessons learned from COVID-19 into future public health strategies will better position us to mitigate the impact of emerging health threats and protect the worldwide well-being.

While preventing the emergence of novel pathogens present considerable difficulties, it is possible to effectively confront new diseases or at least decrease the risk of their dissemination. In this regard, the shared global objective is to develop strategies, treatments, and vaccines that can be deployed rapidly in response to unexpected health challenges.

1.5 The crucial role of metabolism in health and disease

As described previously, historically and presently, infectious diseases continually threaten public health, exemplified by recent viral pandemics like COVID-19. Despite advancements in healthcare and disease management, these challenges persist, underscoring the need for preven-

tive measures. Only through a holistic understanding and coordinated efforts can we effectively address the synergistic impact of infectious diseases, viral pandemics, and AMR on a global scale.

The metabolism, intimately linked to a cell's phenotype, stands as the most reliable indicator of its physiological state[117]. It refers to a large and complex network of interconnected biochemical conversions called reactions. Enzymes catalyze the majority of reactions in a process called catalysis. The metabolic network is subdivided into pathways. Each pathway is specialized in converting nutrients into various essential building blocks such as amino acids, nucleotides, and lipids required for cellular development and function. As a result, energy is produced to support various cellular functions, including biomass generation, cell reproduction, maintaining homeostasis, and facilitating biosynthetic processes. The BioCyc database[118] and the Kyoto Encyclopedia of Genes and Genomes (KEGG)[119] are examples of comprehensive repositories of metabolic pathways. Metabolism both impacts and is impacted by cellular processes. Numerous studies have underscored its pivotal role in diverse diseases and conditions, such as diabetes, cancer, and neurodegeneration.

Metabolic alterations in cancer cells influence the development and progression of hematological malignancies, including leukemia, and contribute to immune evasion[120]. Targeting the metabolic vulnerabilities of leukemic cells, such as glycolysis, fatty acid oxidation, and lipid biosynthesis, could provide promising therapeutic opportunities, considering especially non-responsive patients[121]. Furthermore, metabolic reprogramming is crucial in coordinating the immune response, while targeting metabolism has emerged as a potential therapeutic strategy in chronic inflammation[122]. Disruptions in metabolic pathways, such as glycolysis and lipid metabolism, not only contribute to cancer proliferation but also correlate with conditions like diabetes, atherosclerosis, obesity, cardiovascular diseases, and Alzheimer's disease[123–128]. In the context of bacterial metabolism, previous studies indicate that the metabolic state of bacteria not only affects their susceptibility to antibiotics but is also reciprocally influenced by them[129]. Metabolomic studies have shown that antibiotic treatment can alter the metabolite abundance and disrupt bacterial metabolic processes[130]. Consequently, researchers have explored the potential of exogenous metabolite supplementation in laboratory, animal, and computational studies to activate bacterial metabolism, thereby enhancing the effectiveness of antibiotics. Additionally, deciphering bacterial metabolic modulation can also provide insights into the expression of virulence genes[131], the behavior of bacterial communities[132], and the bacterial-host interactions[133].

Finally, viruses, being metabolically inert, rely on host cell metabolism for replication and energy acquisition[134]. They induce cellular changes and rely on host cells for necessary macromolecules, evolving strategies to manipulate host cell metabolism. This genetic manipulation prompts infected cells to produce more virus particles and results in increased metabolic changes[135]. Hence, understanding the metabolic alterations upon viral infections is crucial[136]. Altogether, a system-wide comprehension of the thousands of reactions and metabolites comprising metabolism is paramount for future therapeutic purposes.

1.6   Systems biology and mathematical modeling: decoding biological complexity

The definition and origins of systems biology have been a subject of debate for the past two decades within the research community[137–139]. While it is often perceived as a relatively recent field, its roots can be traced back to earlier scientific endeavors. The conceptual framework of

systems biology emphasizes the study of biological systems as integrated and interconnected networks. This approach is inspired by various disciplines such as cybernetics, mathematical biology, and systems theory[140, 141]. Pioneers in these fields, including Norbert Wiener, Warren Weaver, Paul A. Weiss, and Ludwig von Bertalanffy, laid the groundwork for understanding complex systems in biology as early as the mid-20th century. Their insights into feedback mechanisms, holistic approaches to studying living organisms, and the dynamics of biological systems have established the foundation for the evolution of modern systems biology. O'Malley et al. recognizes two classes of systems biology: "systems-theoretic biology" and "pragmatic systems biology"[137]. The former emphasizes a holistic approach rooted in developmental biology, aiming to understand biological issues through comprehensive system analysis. Conversely, "pragmatic systems biology" focuses on the post-genomic era, leveraging computational models and modern technological tools to investigate molecular interactions.

Radical technological advancements emerging in the mid-1990s and early 2000s marked milestones in systems biology, impacting our understanding of biological systems. The first bacterial genome sequencing[142] and the completion of the Human Genome Project[143] have played a pivotal role in shaping the landscape of systems biology. They providing foundational data and methodologies for studying biological systems at the whole-genome level. More specifically, the sequencing of the *H. influenzae* genome in 1995 marked a remarkable breakthrough, offering insights into the genetic makeup of organisms and catalyzing improvements in sequencing technologies[142]. Similarly, the completion of the Human Genome Project in 2003 provided an extensive catalog of human genes, laying the foundation for studying gene expression, protein interactions, and genetic variation on a genome-wide scale[143]. Additionally, the complexity of biological systems became increasingly apparent, making clear that traditional reductionist approaches alone were insufficient to fully understand biological phenomena[144]. These aspects were the driving forces for the rise of systems biology as a scientific discipline that emphasizes in studying the biological systems as dynamic and interconnected networks rather than focusing solely on individual components[145].

Despite the different opinions on the definition of systems biology, mathematical modeling, quantitative determination of dynamic responses within living cells, and globalism are integral components of a consensus definition[145–148]. They are commonly acceptable denominators that are instrumental in elucidating the complexities of biological systems. Nowadays, it is conventionally endorsed that the systems biology discipline combines mathematical modeling and computational approaches to study biological systems[149]. Defining and understanding the genotype-phenotype relationships has been a prominent challenge in this field. While in straightforward instances of monogenic traits where a single gene represents a phenotype, most phenotypic characteristics result from the interplay of multiple gene products. This complicates the elucidation of the links between genotype and phenotype[150]. The primary goal is to contextualize and interpret various large-scale biological data and unravel the mechanisms of complex systems via an iterative approach. Modeling techniques in systems biology are often distinguished into two categories: top-down and bottom-up. These two classes complement each other and are employed depending on the study's objectives[150, 151]. Top-down systems biology relies on data analysis, extracting biological understanding from high-throughput omics data. Soft models like statistical models or neural networks are often used, with the analysis typically being inductive, leading to hypothesis formation and subsequently to model generation[151]. Conversely, bottom-up systems biology relies on existing knowledge translated into mathematical representations employed to simulate the system's behavior[150, 151]. This approach requires evaluation and validation of the different model structures due to limited knowledge availability.
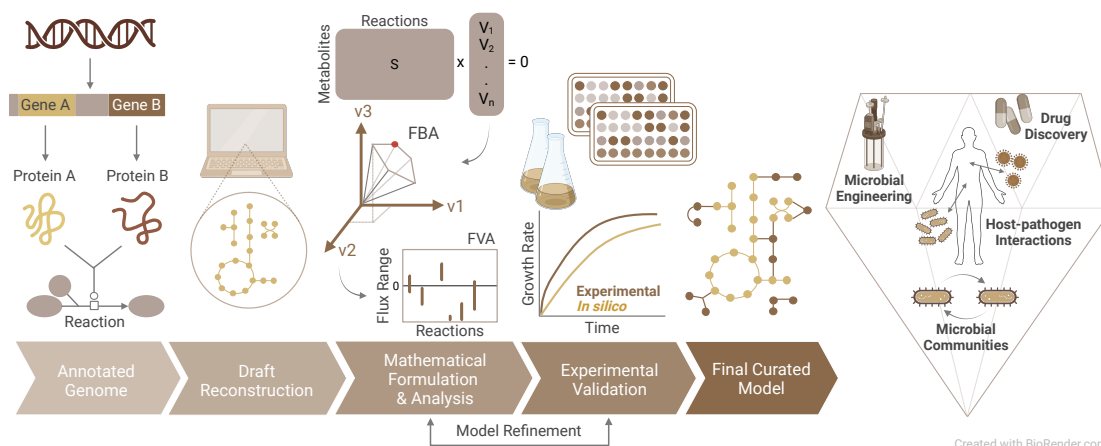
Figure 1.3: Integrated process for the reconstruction and analysis of genome-scale metabolic models. The network reconstruction begins with functional annotation, which involves linking genes with enzymatic functions. This process results in a preliminary draft mathematically represented by the stoichiometric matrix **S**. Constraint-based analysis and linear optimization techniques explore metabolic flux distributions within the solution space, allowing comprehensive exploration of metabolic behaviors. The final model has broad applications in bioengineering, biotechnology, disease understanding, personalized medicine, and drug discovery.

Biological networks are invaluable mathematical tools to represent pairwise connections between various biological entities. They are utilized to understand the interactions among components within a biological system and their collectively functionality is crucial for comprehending an organism's behavior. According to graph theory, biological networks consist of two basic elements: nodes (or vertices) representing individual components and edges depicting their interactions[152]. Depending on the network type, the edges can be directed or undirected. An essential aspect derived from these networks is the degree of a node, which indicates the number of edges connected to a node (i.e., its adjacent nodes). Biological networks enable the visualization and analysis of complex biological phenomena, offering insights into the structure, dynamics, and emergent properties of living systems[153]. They play a pivotal role in bioinformatics, systems biology, and network biology, facilitating the elucidation of cellular signaling pathways, regulatory networks, metabolic pathways, and disease mechanisms. Five main types of biological networks can be categorized based on the type of information they represent: gene interaction, metabolic, protein-protein interaction (PPI), transcriptional (gene) regulatory, and cell signaling networks[154]. Additionally, whole-cell models arose aiming to simulate the behavior of an entire living cell, encompassing its diverse molecular processes and interactions[155–157]. These models integrate data from various omics studies, including genomics, transcriptomics, proteomics, and metabolomics, to capture key cellular functions comprehensively. With this, they simulate the connection between metabolism and other cellular functions associated with cellular growth.

1.6.1   The power of systems biology in deciphering cellular metabolism

Cellular metabolism, being the sum of all biochemical reactions, can be modeled at a genomic scale and is the most extensively described biological network that enables the analysis of genotype-phenotype relationships[158, 159]. As a surrogate for phenotype, it reveals changes resulting from individual or combined x-omics or physiological factors. The well-studied me-

tabolic biochemistry and the increasing availability of annotated genomic sequences expedited the bottom-up assembly of comprehensive biochemical reaction networks. These metabolic networks are called genome-scale metabolic models (GEMs) and consist of interconnected pathways of all known biochemical reactions involving genes, metabolites, enzymes, and other molecules within a cell[160]. Within GEMs, vertices represent chemicals, or metabolites, that are either produced or consumed by other molecules forming multiple reactions[153]. These networks are structured as directed graphs, wherein each edge signifies a biochemical reaction. In irreversible reactions, edges have a single direction, while reversible reactions are represented by edges with double directionality. This organization reflects the flow of metabolites and the directionality of biochemical transformations within the metabolic network. The iterative bottom-up reconstruction and analysis of GEMs involves five main steps, as illustrated in Fig. 1.3. It begins with the functional annotation of the genome and the establishment of links between genes and enzymatic functions. These links are called gene-protein-reaction associations (GPRs) and facilitate the integration of genetic information into constraint-based metabolic networks. There are several types of GPRs, each represented in Boolean logic and depicting different scenarios, including isozymes, protein complexes, multifunctional enzymes, and single enzymes[161]. The simplest form is the "AND" rule, where all genes in a set are required for the associated reaction to occur. Conversely, the "OR" rule dictates that any gene within a set can catalyze the reaction independently. Additionally, more complex rules may involve combinations of "AND" and "OR" relationships, allowing for intricate representations of genetic dependencies within the metabolic network[162]. These rules are crucial for defining the phenotypic outcomes of genetic alterations, as well as elucidating the mechanisms driving a specific phenotype. The annotated genome is used to build a preliminary draft reconstruction that is converted into a mathematical representation called a stoichiometric matrix.

In metabolic modeling, the stoichiometric matrix $\mathbf{S}$ serves as a foundational element, encapsulating the relationships between metabolites and reactions within a biological system[150]. This data matrix entails m rows and n columns, where m is the number of metabolites and n is the number of reactions in the model. Typically, in a biological network, it holds $n > m$[150]. The entries within the stoichiometric matrix, consisting of integers, adhere to chemical principles, thus facilitating the balance of metabolic networks. These integers, known as stoichiometric coefficients, quantitatively express the relative amounts of metabolites involved in reactions, thereby ensuring the preservation of mass and the coherence of reaction equations[150]. Mathematical descriptions of concentration changes over time are done using systems of ordinary differential equations (ODEs) based on reaction rates. Irreversibility or reversibility of reactions is not directly included in $\mathbf{S}$. However, the signs of coefficients within the matrix denote whether the metabolite is produced (positive sign) or consumed (negative sign)[150]. Overall, the $\mathbf{S}$ matrix derived from genomic and bibliomic data embodies diverse attributes (e.g., physicochemical, genetic, biological, mathematical, systemic), ensuring species-specific metabolic networks and enabling precise computational analyses of metabolic dynamics[150]. The dynamic mass balance equation is a fundamental principle in metabolic network modeling that describes how the concentrations of metabolites change over time due to metabolic reactions[150, 163]. Mathematically, it is expressed as

$$\frac{d\vec{x}}{dt} = \mathbf{S} \cdot \vec{v}, \tag{1.1}$$

where $\vec{x}$ represents the vector of metabolite concentrations, $\mathbf{S}$ is the stoichiometric matrix, and $\vec{v}$ denotes the vector of reaction fluxes[150]. The stoichiometric coefficients are denoted by s_{ij}. This equation ensures that mass is conserved in the system, meaning that the total amount of each metabolite remains constant over time. When metabolite concentrations remain constant,

even though reactions may still have fluxes, the system is considered to be in a steady-state, characterized by

$$\mathbf{S} \cdot \vec{v} = \vec{0}. \tag{1.2}$$

Additionally, the dynamic mass balance equation serves as the foundation for simulating and predicting the behavior of metabolic networks under different physiological conditions, aiding in the design and optimization of biotechnological processes.

1.6.2   Constraint-based modeling and flux balance analysis

A frequently utilized powerful approach for analyzing the metabolism is the systems-level constraint-based modeling (CBM) of metabolic networks. Over the past years, an extensive array of constraint-based reconstruction and analysis (COBRA) techniques and algorithms for analyzing GEMs has been developed and encapsulated in computational frameworks[164, 165]. Unlike traditional kinetic models, CBM does not rely on detailed kinetic data but instead employs mass balance principles across the metabolic network. It involves leveraging mathematical constraints to explore and analyze the complex interactions within biological systems. At its core, constraint-based modeling relies on the stoichiometry of biochemical reactions, which is captured in the stoichiometric matrix (stoichiometric modeling). Within the allowable space defined by constraints (e.g., such as mass balance, thermodynamic feasibility, and enzyme capacities), the network can adopt various flux distributions, while any points lying outside it are restricted by the imposed constraints[166, 167]. This approach allows the prediction of metabolic flux distributions and the detection of optimal solutions that maximize or minimize specific objectives, such as growth rate or production yield. Moreover, constraint-based analysis provides valuable insights into metabolic network properties, such as robustness, flexibility, and response to genetic perturbations. When no constraints are imposed, the model simulates the uptake or release of compounds at arbitrary rates that do not reflect real biological systems. Hence, constraints are necessary to approximate the flux solution space with that of living systems, typically representing known growth patterns. Constraints are mathematically depicted as balances (represented by an equation) or bounds (represented by an inequality)[150]. Balance constraints can be the conservation of mass, osmotic pressure, and conservation of energy, while thermodynamic or capacity constraints are examples of bound constraints.

Since the late 1990s, the field of COBRA has evolved to analyze allowable phenotypic states on a genome-scale[150]. COBRA methods heavily rely on optimization techniques, including linear programming (LP), which is used when the problem involves linear constraints and a linear objective function. A widely used computational technique in systems biology to investigate the behavior of metabolic networks is the flux balance analysis (FBA)[167, 168]. FBA relies on the assumption of a steady-state condition and employs linear optimization methods to predict the flux distribution of biochemical reactions within the network. Its objective is to find the best solutions within the solution space delineated by a convex polytope. This space encompasses all possible flux distributions that satisfy the specified constraints, enabling a thorough investigation of metabolic behaviors under different environmental conditions. Linear programming is applied, as the system is under-determined ($n > m$) and is algebraically unsolvable[169]. The underlying maximization problem is mathematically expressed as follows:

$$
\begin{aligned}
\text{maximize} \quad & Z = c^T \vec{v} \\
\text{subject to:} \quad & \mathbf{S} \cdot \vec{v} = \vec{0} \\
& v_{\min} \leq v_r \leq v_{\max} \quad \text{for } r \in \{1, \ldots, n\} \\
& \forall r \in I : 0 \leq v_r,
\end{aligned}
\tag{1.3}
$$

where $\mathbf{S}$ is the stoichiometric matrix, $\vec{v}$ is the vector of fluxes within the network, Z is the linear objective function, I represents an index set containing the indices of all irreversible reactions, and $\vec{c}$ is the vector of coefficients. At its core, FBA relies on optimizing (maximize or minimize) an objective function[168]. The objective function is built by a linear combination of fluxes that quantifies the contribution of each reaction to the phenotype. Therefore, it is vital for computing the optimal network state and flux distribution, addressing the challenge posed by the large solution space of genome-scale networks. Different objective functions are used in metabolic network analysis. These include minimizing adenosine triphosphate (ATP) production to assess energy efficiency and studying nutrient uptake to understand metabolic function in minimal nutrient environments[162]. Additionally, maximizing metabolite production and biomass formation are common objectives, aiming to determine a cell's biochemical production capabilities and maximal growth rate.

When predicting bacterial cellular growth, the focus is on generating biomass, indicating the rate at which metabolic compounds transition into biomass components like lipids, proteins, nucleic acids, and species-specific components[168, 170]. Biomass production is represented *in silico* via incorporating a pseudo-reaction, called biomass objective function (BOF). This is added as an additional column in the stoichiometry matrix, mimicking the consumption of precursor metabolites needed for biomass synthesis. By optimizing the objective function subject to the constraints of the metabolic network, FBA can predict the flux distribution of metabolic reactions that maximize biomass production under specific environmental conditions. The parameters of the biomass reaction are derived from experimental data on biomass constituents, and the reaction is adjusted so that the flux through it is equivalent to the organism's exponential growth rate (μ). In essence, FBA enables the determination of metabolic fluxes that best support cellular growth and viability, providing valuable insights into cellular metabolism and phenotypic behavior. Like the maintenance function in eukaryotic models or biomass production in prokaryotic models, the viral biomass objective function (VBOF) serves as a surrogate reaction that imitates the generation of virus particles[171]. The objective function, together with the demand, exchange, and sink reactions, belong to the class of pseudo-reactions. These reactions are not inferred from the genome and are incorporated within the model purely for modeling purposes[150, 162]. Demand reactions irreversibly consume metabolites stored within the system, while sinks are reversible and supply metabolites to the system that are generated by external reactions. Similarly, exchange reactions transfer metabolites between the different *in silico* cellular compartments. Sink and demand reactions influence intracellular metabolites, whereas exchanges consider extracellular metabolites. Overall, pseudo-reactions enable the modeling of the metabolites' exchange between intracellular space and the external environment. Adding these reactions along with various transporters into the network decreases the amount of orphan (those solely consumed) and dead-end (those solely produced) metabolites and improves network connectivity[162].

Flux variability analysis (FVA) complements FBA by evaluating the range of feasible flux values for each reaction, providing insights into the variability of metabolic fluxes under different conditions or constraints[150, 172]. By systematically perturbing the objective function while maintaining the constraints, FVA calculates the minimum and maximum flux values each reaction can attain while satisfying the metabolic network's constraints. Together, FBA and FVA allow for a comprehensive analysis of metabolic network behavior, from identifying optimal flux distributions to assessing the flexibility and robustness of metabolic pathways. Further approaches utilized for genome-scale optimization include minimization of metabolic adjustment (MOMA), which has been developed to predict changes in metabolic flux distributions when the function of a gene product is lost[173]. In contrast to FBA, MOMA employs quadratic

programming. It aims to identify a new flux distribution within a reduced space of solutions that minimizes the Euclidean distance between the altered solution space caused by the gene knockout and the wild-type state. Moreover, to address the degeneracy issues of FBA solutions, parsimonious enzyme usage flux balance analysis (pFBA) is employed assuming that cells use the minimum amount of energy resources to achieve their metabolic objective[174]. This method involves solving two consecutive linear optimization tasks to identify the optimal flux distribution while reducing the total sum of fluxes. Fong et al. argued that the adaptive evolution of mutated strains could result in phenotypic characteristics closer to those predicted by FBA[175]. Although these methods have been successfully applied in predicting knockouts[173, 176], a debate arises in determining the appropriate wild-type solution for comparison, which has been subject to further investigation[177].

Experimental analysis ensures the accuracy and reliability of predictions generated by the model, aiding in further refinement of the network through gap-filling procedures (Fig. 1.3). Model extension via gap-filling is a non-trivial step, particularly for species that are not closely related to well-studied organisms. Ultimately, the finalized model finds extensive applications across various fields, including bioengineering, biotechnology, disease research, personalized medicine, and drug discovery.

1.6.3   Genome-scale metabolic modeling across domains of life

As mentioned previously, GEMs are fundamental resources in systems biology and metabolic engineering, yielding mechanistic links between genotype and phenotype of the cellular metabolism. Initially reconstructed for individual bacteria, numerous models have been consequently developed for eukaryotes and archaea over time. The reconstruction of GEMs for bacteria has undergone significant development over the past few decades, leading to a rich historical timeline. In the late 1990s, pioneering efforts emerged with the reconstruction of the first metabolic model for *H. influenzae*[169]. As of today, more than 6,000 metabolic models have been created utilizing both automated and semi-automated methodologies, encompassing a wide variety of organisms, including bacteria, archaea, and eukaryotes[178, 179]. Throughout the 2000s, the field witnessed a surge in the reconstruction of metabolic models. It started with model organisms of medical and industrial importance, like *Escherichia coli*[180] and *Saccharomyces cerevisiae*, and continued with multicellular organisms, including plants[181] and humans[182]. The refinement and development of reconstruction techniques, such as manual curation and automated annotation pipelines, led to the generation of additional GEMs. By the 2020s, the field had reached a critical mass, with hundreds of GEMs available for diverse bacteria, encompassing both pathogenic and non-pathogenic organisms[179]. The Assembly of Gut Organisms through Reconstruction and Analysis (AGORA) resource comprises a set of semi-automatically generated GEMs and was designed to represent metabolic activities and interactions within the human gut microbiota[183, 184]. Additionally, multi-species models integrate the metabolic interactions of diverse organisms within communities, offering insights into symbiotic relationships and ecosystem dynamics[185, 186].

With the continuous advancements in computational biology and the availability of high-throughput omics data, human GEMs have become essential tools for studying human metabolism and its implications in health and disease. Over the years, there has been a remarkable rise in the number of human reconstructions, reflecting the growing interest and efforts in understanding the complexities of human metabolism (Fig. 1.4). Recon1, the first generic human metabolic network, was developed through a rigorous reconstruction process, starting
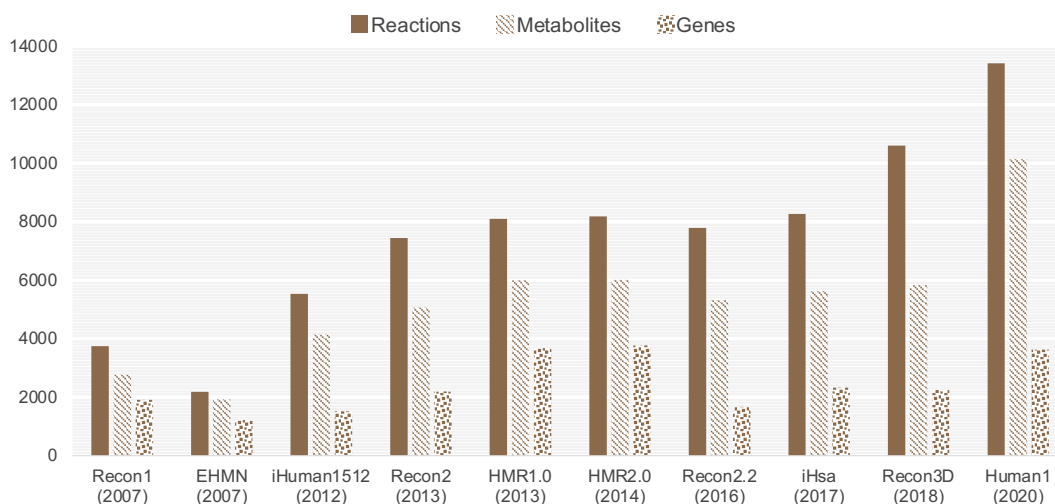
Figure 1.4: Existing mathematical models representing the human metabolism. Collection of GEMs built to represent and simulate the metabolism of human cells. Specific models have undergone iterative revisions, while others originated from the integration of existing reconstructions coupled with omics and other biological datasets.

with the automated extraction of metabolic genes and followed by manual curation of network components[182]. Subsequently, further ReconX models were developed, including Recon2[187], Recon2.2[188], and Recon3D[189]. These iterations expanded upon previous versions, incorporating more reactions, refining gene representation, and enhancing metabolic annotations. The latest addition, Recon3D, represents the most comprehensive and updated metabolic resource. It offers a three-dimensional view of metabolism and enables precise predictions of metabolic responses to various drugs and mutations associated with diseases. Recent curation efforts of the ReconX and Human Metabolic Reaction (HMR) model lineages[190, 191] resulted in the reconstruction of the latest human metabolic network, Human1[192]. These advancements underscore the collaborative efforts within the scientific community to refine and expand our understanding of human metabolism continually. The available human GEMs not only function as a foundational framework but also serve as a versatile platform for integrating and analyzing diverse datasets, including kinetic and omics data. Several methods have been developed supporting the integration of different types of omics data into metabolic reconstructions, enabling the construction of context-specific (tissue- or cell-specific) metabolic models with enhanced biological relevance and predictive power[193, 194]. The computational advances throughout the years, along with the increased availability of omics data and human metabolic models, led to the development of context-specific reconstructions. For instance, the HMR series laid the foundation for the development of several cell-specific GEMs, playing a vital role in studying health conditions like non-alcoholic fatty liver disease (NAFLD), obesity, and diabetes[190, 191, 195]. Furthermore, Thiele et al. employed the generic Recon3D model to reconstruct the first sex-specific and whole-body metabolic network[196]. By combining genomic and clinical data, these models provide valuable understanding of individual health, leading to the advancement of personalized healthcare[197, 198].

Various repositories are accessible for storing metabolic models, granting researchers access to a diverse array of curated models tailored for different organisms and objectives. Among these, the Biochemical, Genetical, and Genomical (BiGG) Models database[199] stands out, hous-

ing multiple genome-scale metabolic network reconstructions, including Recon1 and Recon3D. Similarly, the Virtual Metabolic Human (VMH) database[200] specializes in human metabolic models, catering to specific research needs in this domain. Platforms such as BioModels[201] and ModelSEED[202] provide extensive collections of curated models spanning various organisms. The Metabolic Atlas offers access to the HMR series and Human1 models, alongside reconstructions of metabolic networks for yeast, fly, fruit, rat, mouse, worm, and zebrafish[192]. These repositories are invaluable resources, fostering model sharing, collaboration, and reproducibility across the scientific community.

Currently, the reconstruction of GEMs continues to evolve, driven by advances in omics technologies, computational algorithms, and integrative systems biology approaches. This will pave the way for novel insights into metabolism and its applications in biotechnology, medicine, and environmental science.

1.6.4   Computational model language and standardization

As systems biology gains traction, various biological models are being developed, prompting endeavors to establish a unified language to tackle interoperability and standardization challenges. This initiative spurred extensive international efforts and community dialogues, leading to the development of the Systems Biology Markup Language (SBML)[203]. SBML provides a machine-readable and structured representation format for encoding computational models of biological processes. It facilitates model exchange, dissemination, and reproducibility across different software platforms and research groups[203]. It is based on Extensible Markup Language (XML) and, since its initial release in 2001, multiple editions (levels) have been published. Currently, there are three SBML levels, each building upon its predecessor, considering users' practical experiences. In the latest release introduced in 2019, the SBML Level 3 Version 2[204], a model definition comprises various optional components, including function definitions, species, parameters, compartments, reactions, and rules. The flux balance constraints (fbc) package enhances this level by providing a balanced operation format, supporting the development of interoperable constraint-based model coding formats within a collaborative community[205]. Utilizing the same model representation reduces translation errors and ensures a shared foundation for analysis and simulations. Among others, libSBML[206] and JSBML[207] software libraries have been developed to build, read, write, and manipulate SBML models. The Systems Biology Format Converter (SBFC)[208] and the SBML Toolbox[209] enable importing and exporting models in different formats. Finally, the metabolic model testing (MEMOTE) suite[210] and the SBML Validator from libSBML[206] were developed to assess the model's semantic and syntactic quality.

Central to SBML's effectiveness is the Minimal Information Required In the Annotation of Models (MIRIAM), a set of guidelines ensuring consistency and comprehensiveness in model annotation[211]. MIRIAM-compliant annotations enhance model interpretability and exchangeability. They enable researchers to trace the origins and assumptions of model components, such as reactions and species, fostering transparency and reproducibility in computational studies. Moreover, SBML facilitates the integration of various ontologies and controlled vocabularies to enrich model annotations[212]. By this, SBML ensures that model elements are precisely defined and semantically linked to relevant biological knowledge, enhancing model interpretability, interoperability, and reproducibility[213]. The Systems Biology Ontology (SBO) terms provide a controlled vocabulary for describing the biological meaning and context of model components, such as reactions, species, and compartments[214]. By associating SBO terms with model elements, researchers can convey specific biological semantics, ensuring consistent interpretation

and exchange of models across different computational platforms and research groups. Similarly, Evidence and Conclusion Ontology (ECO) terms enable the annotation of environmental conditions and experimental context associated with model simulations, ensuring transparency and reproducibility in computational studies[215]. Ultimately, integrating SBO and ECO terms within SBML models contributes to model standardization efforts, promoting transparency in computational systems biology research. A further practical approach developed to describe specific parts of cellular metabolism systematically is the Systems Biology Graphical Notation (SBGN)[216, 217]. SBGN is a standardized graphical representation scheme developed to depict biological processes and pathways clearly and intuitively. Existing modeling tools that support the SBGN standard format include the CellDesigner[218, 219] and Newt[220]. By adopting these practices, researchers can communicate complex biological information effectively, facilitating collaboration and comprehension across different domains of systems biology.

1.6.5 Broad utility spectrum of metabolic models

GEMs play a pivotal role in understanding cellular metabolism (mechanistic) and elucidating the complex genotype-phenotype relationships (predictive). As already mentioned, these models provide detailed and flexible frameworks for predicting phenotypes under different conditions. By computationally simulating various metabolic scenarios, GEMs considerably reduce the time and resources typically required for labor-intensive experimental procedures. This empowers researchers to prioritize experiments strategically, focusing efforts on those most likely to yield desired outcomes, and accelerating advancements in fields such as metabolic engineering and biotechnology. Over the past years, GEMs have been extensively utilized to simulate biomass production across different growth media and investigate the utilization patterns of various nutrient sources[221, 222]. Additionally, they facilitate simulation of auxotrophies, fermentation products, minimal growth media compositions, and gene essentialities under different environmental conditions[162, 178]. Nonetheless, constraint-based models have employed in target prediction and drug discovery[223]. They have also been used in studying interspecies metabolic interactions[224–228], cancer metabolism[229] and inborn errors of metabolism[187, 230], as well as host-microbe interactions[231, 232] and astrobiology[233]. Finally, modeling and engineering of microbial secondary metabolism for the development of natural products using GEMs has gained interest within the synthetic biology community[234].

Altogether, GEMs serve as excellent knowledgebases in which extensive information can be collected and updated once more evidence is available. Through these multifaceted applications, they continue to be indispensable tools in advancing our understanding of biological systems and driving innovation in biomedical research. Therefore, their utility is expected to grow remarkably in the future as their applications continue to evolve.

CHAPTER 2

Objectives

The formulation of the problem is often more essential than its solution, which may be merely a matter of mathematical or experimental skill.
– Albert Einstein

The overarching purpose of this thesis is to advance our understanding and control of infectious diseases by applying computational methodologies. The primary objectives include the development of comprehensive computational tools and models representing the dynamics of infectious agents at the molecular and cellular levels. Additionally, integrating biological data aims to enhance model accuracy and generate new hypotheses. This cumulative dissertation systematically examines these objectives, encompassing four distinct research projects and their associated publications.

The main focus of the first project is the development of a computational workflow to rapidly predict effective druggable targets against emerging RNA viruses via metabolic perturbations in infected cells, with a specific focus on COVID-19. For this purpose, a context-specific metabolic model for primary human bronchial epithelial cells related to SARS-CoV-2 infections was reconstructed. Additionally, a computational tool was developed to facilitate the generation and analysis of integrated host-virus metabolic models to identify inhibitory pathways. The identified targets were found to be effective against multiple viral variants, increasing confidence in their effectiveness.

The second project addresses model standardization and reproducibility by improving and automating the assignment of SBO terms to various model entities. The tool, developed for this purpose, enables the integration of expert knowledge-driven classification schemes, ensuring accuracy and reliability in the annotation process. Particular emphasis is placed on biochemical reactions for which both top-level terms and the enzyme functionalities are considered to ensure more accurate annotations.

The third objective underlines the utilization of computer modeling techniques to analyze and understand the human nasal microbiota comprehensively. The focus is on developing systems biology models representing various strains within the nasal environment with the ultimate goal of understanding, combating, and preventing primary infections. This thesis addresses this objective by studying a specific human nasal microbiota member, *A. baumannii*. The project includes model reconstruction and validation using experimental data, identification of compounds promoting *A. baumannii*'s cellular biomass, and pinpointing putative essential genes for potential antimicrobial development. Additionally, the project presents the first compilation of
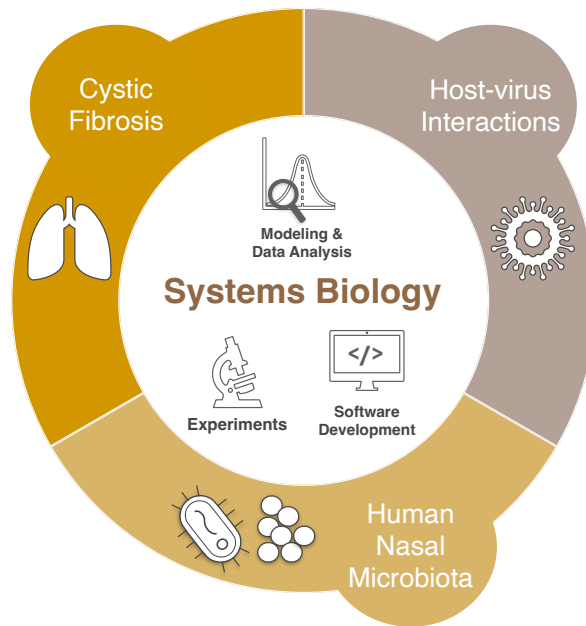
Figure 2.1: Overview of main research areas presented in this thesis. The distinct research fields contribute valuable insights into human health and disease complexities, fostering a holistic understanding of biological processes.

biochemical networks for distinct *A. baumannii* strains in a standardized format, enhancing usability for multi-strain network analysis. These models were analyzed to identify strain-wide growth behavior in different nutrients and essential genes.

The final objective is to unravel the molecular complexities of cystic fibrosis, focusing on studying the metabolic phenotypes of *R. mucilaginosa*. The project describes the reconstruction of the first manually curated genome-scale metabolic model aiming to deepen our understanding of cystic fibrosis and targets capable of altering metabolism. The study explores *R. mucilaginosa*'s metabolic phenotypes through growth experiments and evaluates the model's efficacy in predicting growth behaviors. The model-driven identification of putative essential genes and potential antimicrobial targets showcases its potential in expediting therapeutic interventions.

The demonstrated skills and expertise developed throughout the three-year doctoral studies are summarized in Fig. 2.1, offering a comprehensive overview of the multifaceted research conducted across various aspects of human health.

CHAPTER 3

Results

Research is to see what everybody else has seen and to think what nobody else has thought.
– Albert Szent-Györgyi

This thesis focuses on analyzing complex biochemical networks using computational systems biology and modeling techniques to enhance the understanding, prevention, and intervention strategies against infectious diseases. More specifically, it explores approaches to identify potential drug targets and decipher host-pathogen interactions. The emphasis is on advancing theoretical insights and facilitating the practical application of novel strategies for managing and combating infectious diseases across diverse clinical settings. This chapter introduces four selected projects, while a complete list of publications is available on page xiii.

Firstly, a computational workflow is presented to build context-specific metabolic networks and predict exploitable targets against RNA viruses and their variants. This workflow accelerates the detection of broad-spectrum antiviral agents by analyzing metabolic alterations within infected cells. Subsequently, the SBOannotator is described as the first standalone tool automating the assignment of SBO terms to multiple entities in metabolic models, primarily focusing on biochemical reactions. It considers top-level terms and the functionality of enzymes for precise annotations, aiming to enhance the reproducibility, usability, and analysis of biochemical networks. Additionally, a thorough investigation into bacterial metabolism to address AMR is undertaken. For this purpose, the reconstruction, curation, and analysis of GEM exemplified by a new model for *A. baumannii* is presented. Existing and newly generated experimental data were employed to evaluate the model's performance and enhance the network content. This study goes beyond model creation, curating metabolic networks for multiple *A. baumannii* strains to enhance understanding of metabolic diversity and potential therapeutic targets at a species level. Lastly, the study of CF takes a unique perspective by investigating a less-explored microorganism within the CF lung microbiota, namely *R. mucilaginosa*. The presented work unravels substrate utilization patterns of *R. mucilaginosa*, integrates them into a metabolic network, and identifies nutrient-variant putative essential genes using mathematical modeling. This research contributes the first manually curated GEM for *R. mucilaginosa*, along with high-throughput experimental data. Altogether, these models facilitate the understanding of the metabolic footprint and the advancement of novel therapies. Their high accuracy enables predicting potential antibiotic candidates, guiding future drug discovery efforts.

3.1 Prediction of antiviral drug targets

The global spread of the novel human coronavirus, SARS-CoV-2, has posed substantial challenges to the economy, healthcare, and society. Addressing such a crisis necessitates the development of new vaccines and therapies to efficiently prevent future pandemics. However, vaccines entail limitations such as diminishing immunity over time and reduced efficacy against emerging mutations and variants. Consequently, an essential aspect of pandemic preparedness involves exploring and identifying broadly acting antivirals with robust resistance barriers. Host-directed antiviral approaches are of great interest, as viruses depend on host cell machinery to acquire essential macromolecules, altering cellular metabolism according to their requirements[136]. Here, a new workflow is presented to accelerate the prediction of drug targets against emerging viruses via metabolic perturbations in infected host cells. To this end, a tool was designed to construct integrated host-virus metabolic models, simulate cell infections, and identify host-based pathways that can be targeted to impede SARS-CoV-2 replication.

First of all, cell- or tissue-specific data are collected and undergo a pre-processing procedure. This involves curation tasks, such as computing confidence scores for reactions, binarizing raw transcriptomic data, and determining gene ubiquity scores. Subsequently, the previously developed pymCADRE is employed to construct human cell- and tissue-specific models using gene expression data and topological information from a generic metabolic network. It serves as an open-source Python re-implementation of metabolic Context-specificity Assessed by Deterministic Reaction Evaluation (mCADRE)[235] from the proprietary MATLAB. The tool consists of three main components: (1) ranking, (2) functionality check, and (3) pruning, and provides two optimization methods, FVA and fast consistency check (FASTCC)[236], to ensure internal model consistency. Reactions are ranked based on expression and connectivity evidence, and during pruning, non-core reactions are systematically eliminated while retaining essential metabolic functionalities. Beyond its core functionalities, pymCADRE also includes scripts for data pre-processing, such as assigning confidence scores and binarizing transcriptomic data. Additionally, functional tests ensure metabolic capability, with resultant models provided in SBML[203] format for compatibility with various computational tools.

To evaluate the functionality of pymCADRE and enhance its applicability, particularly in the context of the SARS-CoV-2 outbreak, pymCADRE was utilized here to create a context-specific metabolic network for primary human bronchial epithelial cells (HBECs). The human metabolic network Recon1 [182] served as a generic model to mitigate computational demands.

Table 3.1: Analysis results of the cell-specific reconstructions created using FVA for internal optimizations. The reaction overlap between both models is 99.5 %.

	Pruned Model			Removed Reactions	
	Reactions	Metabolites	Genes	Cores	Non-cores
mCADRE	1,977	1,442	1,905	9	487
pymCADRE	1,973	1,442	1,381	9	489

Remarkably, pymCADRE demonstrated a decreased overall pruning time while preserving the highest possible content similarity compared to the model generated using mCADRE. The final cell-specific model, consisting of 1,973 reactions (1,391 gene-associated; 1,086 metabolic and 305 transport reactions), 1,442 metabolites, and 1,381 genes, exhibited a 99.5 % overlap with the model generated by mCADRE (Table 3.1). This convergence set confidence about the quality of pymCADRE-derived models. Moreover, the final pruned model generated by pym-CADRE includes a BOF representing cellular maintenance requirements. The BOF parameters were extracted from a previously published macrophage model[231]. Furthermore, the model underwent comprehensive validations, including functional, semantic, and syntactic assessments. Benchmarking tools, including Systems Biology Simulation Core Library (SBSCL)[237], MEMOTE[210], and the SBML Validator from libSBML[206], were employed in these validations respectively. In particular, the pymCADRE-derived model exhibited a higher semantic quality score (MEMOTE score of 70 %) compared to previously existing models[235, 238, 239], while it contained neither blocked reactions (reactions that carry zero flux) nor infeasible cycles. Besides that, the model was refined through the inclusion of subsystems for all metabolic reactions. The majority of reactions in the final cell-specific model were transport reactions, with the biosynthesis of other secondary metabolites being the least represented subsystem. This detailed refinement and categorization contribute to more accurately depicting the cellular metabolic landscape.

Nevertheless, inconsistencies in the performance of FASTCC, as implemented in constraint-based reconstruction and analysis for Python (COBRApy)[165], were observed. Multiple iterations of the function revealed differing counts of blocked reactions, resulting in deviations from the ground truth model derived from mCADRE. Since the original algorithm was not modified, discrepancies in the final reactions set suggest variable performance among the built-in functions in Python (COBRApy) and MATLAB (COBRAToolbox[240]). This issue has already been reported and is awaiting resolution. Additionally, the varying performance between the tools may be linked to differences in implementing organic exchange/demand reactions detection. In pymCADRE, this is accomplished in a more robust and fully automated manner based on the chemical formula of the participating metabolites, eliminating the need for hard-coding a specific list of reaction names. This led to the identification of four organic reactions, additionally to those detected by mCADRE. Lastly, another reason for the divergent performance could lie in the MATLAB implementation of the FVA, which is restricted to supporting industrial proprietary CPLEX versions older than V 12.10[241]. Consequently, the latest COBRAToolbox release lacks MATLAB-related binaries for the newest solver release (V 20.1). This limitation is addressed by pymCADRE, which enables the users to choose between the open-source GLPK[242] package and the CPLEX solver for optimization tasks.

Following the main objective to detect potential anti-SARS-CoV-2 targets, the final cell-specific model was enhanced with a VBOF. Similar to the maintenance function in eukaryotic or biomass production in prokaryotic systems, the VBOF serves as a pseudo-reaction that imitates the production of virus particles. This was formulated using established methodologies[171, 243] and consisted of amino acids, nucleotides, lipids, and energy-related metabolites. The VBOF reconstruction considered the viral structure, the genomic reference wildtype sequence, the subsequently encoded proteins and their copy numbers, as well as the energy requirements for nucleotide and peptide bonds. The term "reference wildtype sequence" refers to the first genome that was sequenced and identified as the virus responsible for the COVID-19 pandemic (NCBI accession: NC_045512.2). To represent the lipid requirements, phosphatidylcholine, phosphatidylethanolamine, phosphatidylinositol, phosphatidylserine, cholesterol, and sphingomyelin were added to the viral biomass function. Previous studies examined the impact
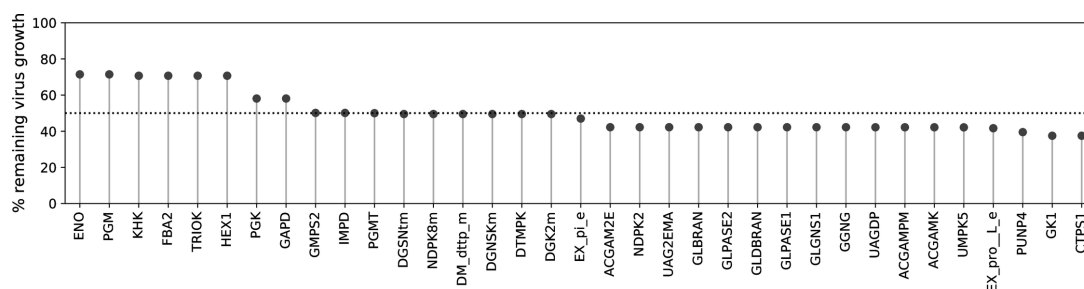
Figure 3.1: Target prediction results upon analysis of the reconstructed cell-specific model using host-derived enforcement. Potential antiviral targets were considered when the viral growth with shifted bounds was below 90 % of its initial growth rate. Enzymes with adjusted bounds involved in purine and pyrimidine metabolism led to a remarkable growth reduction, with additional promising targets identified in carbohydrate metabolism. The dashed line represents the point at which 50 % of the virus remains.

of lipids with varying stoichiometric coefficients on the viral biomass function and antiviral target prediction[243]. However, their final VBOF lacked the incorporation of the lipid composition of a single virion. In this study, stoichiometric coefficients for the lipids were computed based on the virion's surface area, following the method suggested by Nanda et al.[244]. Subsequent FBA optimization of the host-virus model to maximize the host cellular maintenance yielded a flux of $0.2344 \, \text{mmol}/(\text{g}_{\text{DW}} \cdot \text{h})$. Moreover, optimizing the SARS-CoV-2 growth function resulted in a flux of $0.1575 \, \text{mmol}/(\text{g}_{\text{DW}} \cdot \text{h})$. The availability of nutrients considerably impacts metabolic fluxes. Given the primary focus on investigating metabolic changes in infected cells, fluxes strongly influence the simulation outcomes. The human metabolic models lack constraints on the growth medium. This allows all extracellular transport reactions to have a minimum flux value of $-1{,}000.0 \, \text{mmol}/(\text{g}_{\text{DW}} \cdot \text{h})$, enabling unusually high cell maintenance rates. To address this, the chemically defined medium simulating the human blood[245] was incorporated into the final metabolic network as the default medium for further analysis. Cells in the human body are exposed to diverse environments, and the blood medium represents the *in vivo* conditions. However, a minimal medium is required to study specific aspects of cellular metabolism or to understand the minimal requirements for cellular maintenance. Hence, a minimal growth medium for the cell-specific model was defined using COBRApy and considered throughout the analysis. The detailed media compositions are provided in the supplementary files along with the published manuscript.

Stoichiometric modeling was applied on the integrated host-virus model to identify exploitable targets against SARS-CoV-2. The metabolic network was used to detect host-based reactions, which, when constrained, specifically reduce virus growth. Two techniques were applied for this purpose: single-reaction knockouts and host-derived enforcement (HDE)[171]. The first technique involves systematically adjusting individual reaction bounds to zero. The second method constrains all reaction fluxes to ranges obtained from FVA, ensuring the maintenance of the optimal host state while reducing virus propagation[171]. Both approaches confirmed guanylate kinase 1 (GK1) as a promising target to hinder SARS-CoV-2 growth while preserving host cell's maintenance at 100 % (Fig. 3.1). This finding aligns with previous reports of GK1's inhibitory effects in macrophage and lung models[238, 246]. When using HDE, adjusting all reaction fluxes within FVA-determined ranges revealed additional exploitable enzymatic targets. The cytidine 5'-triphosphate synthase 1 (CTPS1) from the *de novo* pyrimi-

dine synthesis pathway emerged as the most promising hit, demonstrating a 62 % inhibition of virus growth and no impact on the host's maintenance. CTPS1 catalyzes the conversion of uridine 5'-triphosphate (UTP) to cytidine 5'-triphosphate (CTP), hence constraining the activity of CTPS1 leads to a decrease in CTP production. It is noteworthy that human cells efficiently restore CTP levels by promptly utilizing alternative routes through the salvage pathway, ensuring the synthesis of the required CTP amounts[247]. Further 33 host-based targets were predicted as potential virus-suppressing candidates by HDE. Specifically, inhibitors involved in carbohydrate metabolism, including the metabolism of amino/nucleotide sugar and sucrose, as well as glycolysis/gluconeogenesis, along with those in purine and pyrimidine biosynthesis, led to a remarkable reduction in viral production. This decrease ranged from 50 % to 58 % of the initial growth rate. Notably, targets from amino/nucleotide sugar and sucrose metabolism exhibited higher antiviral effects than those from glycolysis, fructose, and mannose metabolism. Interestingly, GK1 is closely interconnected with two reported promising targets in the purine metabolism: inosine 5'-monophosphate dehydrogenase (IMPD) and guanosine 5'-monophosphate synthase (GMPS2). IMPD catalyzes the NAD^+-dependent oxidation of inosine 5'-monophosphate (IMP) to xanthosine 5'-phosphate (XMP), followed by GMPS2 generating guanosine 5'-monophosphate (GMP). Using the minimal growth medium not only highlighted nucleoside diphosphate kinase 3 (NDPK3) as a potential target, previously undetected with the blood medium, but also reaffirmed nucleoside diphosphate kinase 2 (NDPK2), uridine 5'-monophosphate kinase 5 (UMPK5), and GK1, all of which are linked within the same pathway. Adjusting the fluxes of NDPK3 resulted in a 44.6 % reduction in virus propagation, which is comparable to the effect of GK1 under the same growth conditions. However, NDPK2 and UMPK5 exhibited the highest viral reduction, emerging as the most promising hits under minimal conditions.

By February, 2022, five SARS-CoV-2 variants, namely Alpha (B.1.1.7), Beta (B.1.351), Gamma (P.1), Delta (B.1.717.2), and Omicron (B.1.1.529), prevailed and spread since the virus's emergence in 2019. The WHO designated them at that time as VOCs[248]. The Global Initiative on Sharing All Influenza Data (GISAID) EpiCoV database[249], launched in February, 2020, hosts over 16.6 million SARS-CoV-2 viral sequences as of April, 2024. It serves as a valuable resource for scientists to investigate the implications of viral variants on metabolic processes and potential therapeutic targets. To examine the effectiveness of predicted targets across all VOCs, multiple VBOFs were created for each mutated sequence retrieved from GISAID employing a methodology analogous to that used for the wildtype sequence. Creating these functions involved the selection of 20 genomic sequences per variant, resulting in the definition of totally 100 individualized biomass functions by calculating the corresponding stoichiometric coefficients. Each function was integrated into the HBEC-specific model, which then was subjected to target prediction. The Prediction of Antiviral Targets (PREDICATE) tool was designed for rapid and automated analysis of multiple sequences associated with a single variant. It expedites the VBOF reconstruction and predicts host-based antiviral targets using integrated metabolic models. Additionally, it provides scripts to modify the reference protein sequences, implementing amino acid mutations (insertions, replacements, duplications, and deletions). Given the analogous composition of RNA viruses, encompassing nucleotides and proteins, this methodology is universally applicable to both single- and double-stranded RNA viruses. Variant-wise comparison of stoichiometric coefficients defining the viral biomass functions for the reference and mutated SARS-CoV-2 sequences revealed differences in the Omicron variant compared to the wildtype. Notably, the stoichiometric coefficients for ATP and adenosine 5'-diphosphate (ADP) in the Omicron variant were remarkably increased. This deviation is ascribed to a lower total viral molar mass and a higher total amino acid count calculated for the Omicron sequences.

Similar differences were observed across additional coefficients. For instance, the Omicron variant exhibited a notable increase in lysine, attributed to the frequent replacement of amino acids by lysine in the spike protein. Conversely, the Delta variant showed a remarkable rise in asparagine due to mutations replacing other amino acids with asparagine. Literature findings confirm these observations, highlighting higher amounts of charged residues in Omicron along with the accumulation of hydrophobic residues in its spike protein[250]. In the next step, PREDICATE integrates each created VBOF into an input cell-specific metabolic model. The subsequent target prediction implemented in PREDICATE considers single-reaction knockouts and HDE, aiming to identify and evaluate the targets' efficacy across the variants of interest. The knockout analysis applied to the HBEC model revealed GK1 as the sole potential inhibitor, while HDE verified the highest inhibitory effect of CTPS1 across all tested variants and the wildtype. GK1 was predicted by HDE as well, though less effective in reducing virus growth. Moreover, the wildtype shared 12 hits with the five SARS-CoV-2 variants, while eight targets were found to be wildtype-specific. Targets associated with carbohydrate metabolism and those involved in purine and pyrimidine metabolism demonstrated promising inhibitory effects across all studied variants. Consequently, nucleotide metabolism, specifically targeting one of the reported enzymes to inhibit SARS-CoV-2, is a promising approach that warrants validation through *in vitro* and cell culture experiments.

The presented workflow and subsequent analyses unveiled enzymatic targets with inhibitory potential in HBECs against SARS-CoV-2. Approved drugs targeting these enzymes were identified by utilizing publicly available resources and databases, including inhibitors for promising candidates such as CTPS1. Virus-infected cells rely on regulated nucleotide metabolism, making it a crucial target for antiviral strategies. This can be achieved through purine and pyrimidine analogs, i.e., modified nucleosides used to stop DNA or RNA polymerase, or by directly inhibiting the enzymes involved in DNA and RNA synthesis. Cyclopentenyl cytosine (CPEC) and acyclovir are known inhibitors CTPS1 and GK1, respectively, acting as analogs that disrupt nucleotide synthesis[251, 252]. Literature findings confirm the antiviral effect of CPEC against a wide range of viruses *in vitro*[253, 254]. Additionally, the therapeutic effect of CPEC has been investigated in cancer studies. In more detail, it has been examined most extensively in leukemic cell lines, and has also been investigated in relation to colorectal carcinoma, brain tumors, and neuroblastoma[255]. Although dose-related hypotension events were observed in patients with colon carcinoma during Phase I trials of CPEC treatment[256], cardiotoxicity could not be replicated in established rat models[257]. Studies have indicated that the deaminated product of CPEC, cyclopentenyl uridine (CPEU), along with cytidine, may serve as potential modulators of CPEC's cytotoxic activity[258, 259]. Therefore, further *in vivo* studies are required to ascertain how it is feasible to establish antiviral effects using CPEC without inducing harmful side effects, whether used alone or in conjunction with other medications. Acyclovir, an approved drug targeting GK1 from the purine synthesis pathway, is used to fight the herpes virus and varicella zoster virus (VZV) infections[260, 261]. In the context of SARS-CoV-2, acyclovir has emerged as a potential antiviral agent against coronaviruses[262], especially in cases where there are indications of VZV reactivation[263]. The authors assumed that this reactivation is linked to the abnormally low lymphocyte count (lymphopenia) observed in COVID-19 patients' blood. Notably, acyclovir's mechanism of action, akin to molnupiravir and remdesivir, involves inhibiting viral replication by mimicking ribonucleosides and causing mutagenic effects. Both molnupiravir and remdesivir belong to the class of carbohydrate-based antiviral drugs and have been studied as therapeutic targets approved by Food and Drug Administration (FDA) to treat SARS-CoV-2[264]. However, acyclovir leads to immediate chain termination, while molnupiravir continues incorporating nucleobases until a mismatch occurs, resulting in an error catastrophe[265]. Remdesivir

functions akin to an ATP analog that causes delayed chain termination[266]. Hence, acyclovir's mode of action suggests a high potential for successful use against SARS-CoV-2 infections. While ritonavir-boosted nirmatrelvir (Paxlovid) and monoclonal antibodiess (mABs) are authorized for COVID-19 and administered intravenously, acyclovir stands out due to its oral administration, offering convenience for self-use. It is essential to note that Paxlovid has been found to have severe implications in patients taking antiarrhythmics, blood thinners, and further medications[267], while mABs (e.g., bamlanivimab-etesevimab, bebtelovimab, and tocilizumab) have undergone revision upon the emergence of the Omicron variant[268]. Under direct enzyme inhibition, fall compounds, like merimepodib and ribavirin, with known pharmacological action that inhibit the predicted host-based target IMPD and have already been tested against emerging viruses[269, 270].

Altogether, the presented workflow, summarized in a four-step process, is designed to contribute to the advancement of effective therapies against emerging viruses and their mutations. It aims to establish the design of broad-spectrum antiviral treatments, serving as a crucial resource for pandemic preparedness. Additionally, the approach focuses on metabolic fluxes within infected cells. This allows for a hypothesis-driven identification of exploitable antivirals, especially considering the time-consuming nature and complex readouts of laboratory tests. The computational workflow introduced herein not only stands to reduce costs significantly but also accelerates the pre-clinical phase, offering applicability to any host cell and RNA virus with outbreak potential. The described software is user-friendly and readily available along with test scripts and a test dataset in a git repository at https://github.com/draeger-lab/pymCADRE. The final host-virus model (*i*HBEC-BOFVBOF-2023) is accessible at https://www.ebi.ac.uk/biomodels/MODEL2202240001 as an SBML Level 3 Version 2[204] file with the fbc package[205] distributed as Open Modeling EXchange format (OMEX) archive[271] including annotations[272].

3.2 Enhancing model standardization and reproducibility

This section presents the article: **Leonidou, N.**, Fritze, E., Renz, A., & Dräger, A. (2023). "SBO-annotator: a Python tool for the automated assignment of systems biology ontology terms." *Bioinformatics*, Volume 39, Issue 7, btad437.

Over the past years, the quantity and size of computational models in biology have increased remarkably, driven by technological advances and increased computational power. Interdisciplinary collaboration among biologists, bioinformaticians, and computational scientists fosters the creation of large-scale models to integrate diverse expertise and comprehend complex biological phenomena. The abundance of biological data, including omics data, provides a rich source for model development, motivating further their widespread inclusion in research. Even though the problem of quality control and quality assurance has already been addressed[210], the increasing prevalence of large-scale metabolic models has drawn attention to challenges related to their reproducibility and clear interpretability[273]. This would necessitate the establishment of a standardized encoding of model-related information, a task frequently hindered by a lack of appropriate tools to assist researchers. Employing precisely defined ontologies facilitates encoding domain-specific expertise and linking disparate data types. In computational modeling, SBO terms facilitate the explicit and unambiguous description of entities, elucidating their roles and characteristics. This section introduces the SBOannotator, the first standalone tool, for the automatic assignment of SBO terms to multiple entities within an SBML[203] model. The tool addresses the laborious step of adding precise descriptions to biochemical reactions
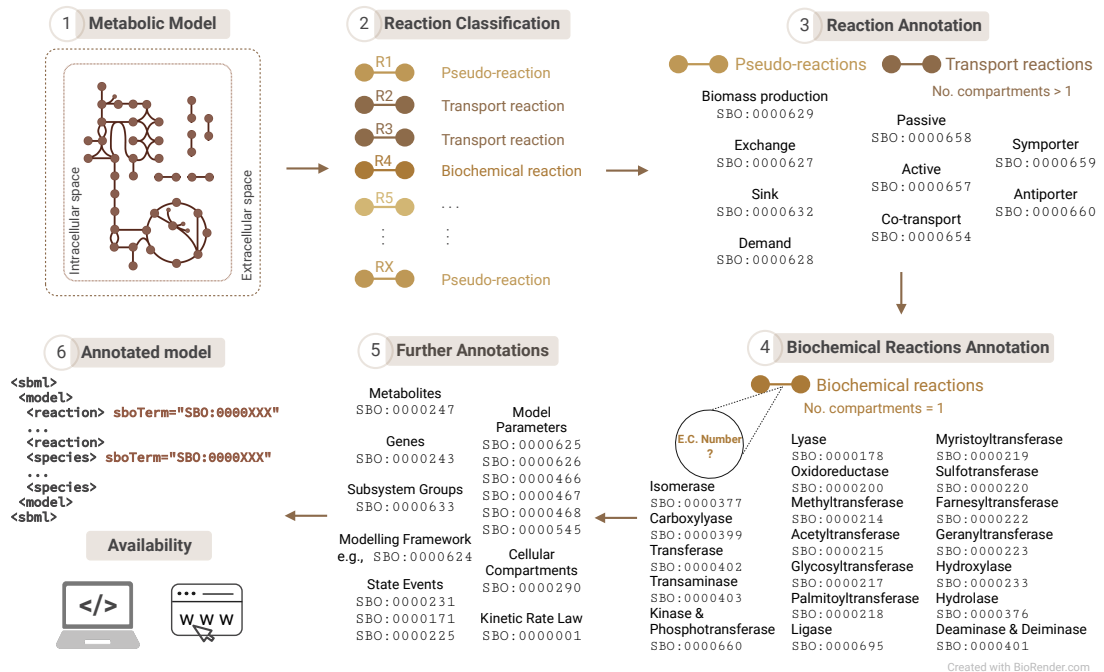
Figure 3.2: Workflow overview for the automated assignment of SBO terms. The SBOannotator enables the automated assignment of SBO terms to multiple model entities within an SBML file. First, the pseudo- and transport reactions are detected and annotated. Then, the biochemical reactions are classified into 19 classes according to their enzymatic function. Finally, further model elements are annotated with their corresponding SBO annotations before saving the final model in SBML format. The SBOannotator is publicly accessible as a web application or a stand-alone tool.

in constraint-based models. The implementation extends beyond top-level terms by considering the functionality of the underlying enzymes to allocate precise and highly specific ontology terms.

The SBOannotator workflow encompasses six main steps, as illustrated in Fig. 3.2. Initially, all reactions within the model are labeled as either (a) transporters, facilitating the movement of molecules across various compartments, (b) simple biochemical reactions, occurring solely in the cytosol, or (c) pseudo-reactions, designed for importing or exporting metabolites to serve modeling purposes. In systems biology modeling, pseudo-reactions do not represent any actual physical process and should not be confused with the pseudo-first-order reactions in chemical kinetics. Pseudo-reactions are classified into exchange (SBO:0000627), sink (SBO:0000632), and demand (SBO:0000628) reactions, while the biomass reactions also belongs to this class. They are incorporated into metabolic models for simulation purposes. The BOF (SBO:0000629) is usually the optimization target in modeling bacterial metabolism and simulates the organism's growth. The SBOannotator further examines the transport reactions and applies advanced classification strategies. In more detail, the categorization mechanism relies on transport properties like the consumption of ATP or phosphoenolpyruvate (PEP) (active transport) or lack thereof (passive transport). Subsequently, the total number of reactants' and products' cellular compartments allows for identifying symporters/antiporters and co-transporters. Symporters transport molecules in the same direction, while antiporters facilitate movement in opposite directions. Co-transporters simultaneously transport two or more substances across a membrane. The remaining biochemical reactions are analyzed in the next step to enable more detailed labeling.

Subsequent steps involve processing remaining biochemical reactions, utilizing an Structured Query Language (SQL) database for efficient mapping between Enzyme Commission (EC) numbers and SBO terms (Fig. 3.2). This pre-established database enhances computational efficiency, particularly when dealing with larger models, streamlining the annotation process.

The SBOannotator assigns the general SBO term of a biochemical reaction (`SBO:0000176`) if the reaction has multiple EC numbers assigned and all belong to different classes resulting in an unambiguous description. Otherwise, the ontologies are accredited based on the respective enzyme's class (e.g., oxidoreductases; `SBO:0000200`, transferases; `SBO:0000402`, hydrolases; `SBO:0000376`). It is important to note that a proper term that describes the ligases (EC class 6) was missing from the SBO graph. This would be necessary to describe reactions involving the formation of DNA, RNA, and protein fragments. After personally contacting the developers of the SBO vocabulary, a new term was introduced (May 16, 2023) for ligases that describes the formation of a covalent bond (`SBO:0000695`). Metabolic reactions falling outside the aforementioned cases and do not have any EC number assigned are given the general SBO term of a biochemical reaction (`SBO:0000176`). The tool can process models with or without pre-existing EC numbers, but it requires the utilization of BiGG[199] identifiers. If the input model lacks EC numbers, the tool initiates an integrated Application Programming transfer Interface (API) call to retrieve the associated data from the BiGG database and subsequently incorporates all missing annotations into the model. Notably, the computational time for this step may vary based on the model's size. Therefore, prior use of an annotation tool, such as ModelPolisher[274], is recommended to optimize performance. Ultimately, the SBOannotator assigns terms to the remaining model entities, including metabolites, genes, cellular compartments, rate laws, and defined parameters. If subsystem groups are declared, the tool allocates the term `SBO:0000633`, while the respective modeling framework (e.g., constraint-based, logical, hybrid) is also assigned an appropriate term (Fig. 3.2).

Evaluation of the SBOannotator's effectiveness in assigning accurate and detailed terms to biochemical reactions involved analyzing 108 metabolic models obtained from the BiGG database. Initially, all models featured five types of SBO terms, representing only top-level terms (biochemical, translocation, demand, sink, and exchange), along with an additional SBO annotation for the biomass reaction. Notably, the SBOannotator assigned 31 specialized terms considering the underlying enzymatic properties. Biochemical reactions constituted the predominant group before and after the application of the SBOannotator. The second most common term in the downloaded models described translocations, while the SBOannotator assigned more specific terms based on transport mechanisms. Consequently, passive transporters were predominantly present across all annotated models, while decarbonylations were the least frequent.

Overall, the SBOannotator stands as a freely available and user-friendly Python tool. Its straightforward usage allows for the rapid annotation of systems biology metabolic networks, focusing on providing accurate and descriptive terms for all chemical reactions. The only prerequisite for utilizing the SBOannotator is a valid SBML[203] format of the input model(s). The SBO terms are directly assigned to SBML model elements, creating a direct one-to-one mapping indicating each element's function. By this, the SBOannotator helps to define more clearly which role individual elements play within a model. Remarkably, the tool takes into account scenarios where a single model entity may have multiple distinct identifiers. SBOannotator is hosted at the TueVis Visualization Server as a web application available at https://sbo-annotator-tuevis.cs.uni-tuebingen.de. Its stand-alone version, all related data, and a demo script to run the code are available in a git repository at https://github.com/draeger-lab/SBOannotator.

3.3   Modeling nasal microbiota for infection prevention

This section presents the article: **Leonidou, N.**, Xia, Y., Friedrich, L., Schütz, M. S., & Dräger, A. (2024). "Exploring the metabolic profiling of *A. baumannii* for antimicrobial development using genome-scale modeling." *bioRxiv*. Manuscript under review in *PLOS Pathogens*.

The nasal cavity serves as a primary reservoir and gateway for microbial colonization, playing a vital role as a link between the interior of the human body and the external environment. Understanding the dynamics of nasal bacteria is particularly crucial in light of the global rise of multidrug-resistant (MDR) bacteria, exemplified by the emergence of carbapenem-resistant *Acinetobacter baumannii*. The urgent need for new antibiotics, emphasized by the WHO, underscores the interconnectedness of nasal microbiota and the broader AMR problem, highlighting the significance of targeted research and preventive therapeutics. *A. baumannii*, known for causing severe infections[79, 81], holds a "critical" status and acquires immense attention[275]. Genome-scale metabolic modeling reveals the complex interaction between metabolites, proteins, and genes and has revolutionized the understanding and analysis of complex microbial metabolism. Therefore, it offers valuable insights into the development of targeted therapeutic interventions and prevention of primary infections[276]. Herein, the exploration of the metabolic profiling of this critical human nasal microbiota member using genome-scale modeling is presented. Specifically, a new metabolic network for the nosocomial *A. baumannii* ATCC 17978 (referred to as *i*ACB23LX) is introduced, adhering to Findable, Accessible, Interoperable, and Reusable (FAIR) data principles and community standards. The model's predictive capability and accuracy were validated through experimental assessments of nutrient utilization and gene essentiality data. Furthermore, existing models for diverse *A. baumannii* strains underwent systematic refinement and evaluation, resulting in the development of the first curated and standardized strain-specific collection of metabolic models.

Over a decade has passed since the initial release of the first mathematical simulation representing the metabolism of *A. baumannii*, followed by subsequent models tailored to specific strains[183, 277–281]. Despite these efforts, the existing metabolic networks exhibit deficiencies, such as a lack of standardization and identifiers, limiting their utility. A literature search revealed one metabolic model of the well-studied strain ATCC 17978, named *i*JS784. By the time of writing, it is exclusively accessible in the form of a dissertation and lacks formal publication in a scientific journal or deposition in a database for mathematical models[282]. Notably, this model has a remarkable limitation as it fails to generate biomass even when all uptake reactions are open and all medium nutrients are available. This deficiency hampers its usability and reproducibility. Therefore, there is a need for a new network that addresses these limitations and provides a comprehensive and well-documented representation of its metabolism. For the construction of a high-quality model the systematic workflow illustrated in Fig. 3.3 was employed, adhering to established community protocols and standards[162, 283]. The newly reconstructed network, named *i*ACB23LX, follows a nomenclature convention where *i* stands for *in silico*, ACB represents the organism- and strain-specific three-letter code from KEGG[119], 23 denotes the reconstruction year, and LX denotes the modelers' initials. The protocol, applicable to all organisms across the tree of life (Eukarya, Bacteria, and Archaea), encompasses eight major stages, ensuring the model's quality and accuracy from the annotated genomic sequence acquisition to validation. CarveMe[284] was utilized for the initial draft model reconstruction, which was subsequently extended and curated manually. Syntactical issues and imbalances in mass and charge were resolved during manual refinement, while missing metabolite charges and chemical
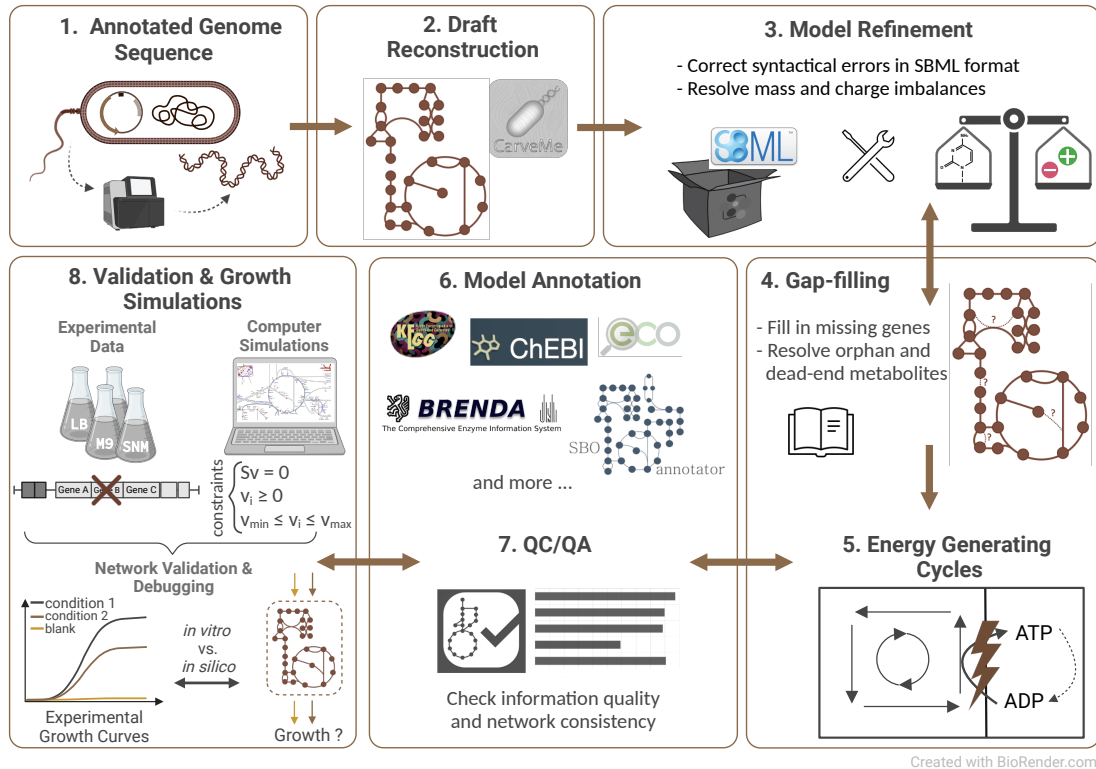
Figure 3.3: Reconstruction workflow of the metabolic network *i*ACB23LX. The workflow consists of eight main steps: extraction of the annotated genome, draft model reconstruction, model refinement, gap-filling, investigation of energy-generating cycles, model annotation, quality control, quality assurance (QC/QA), and model validation using experimental data. The growth simulations consider the examination of growth phenotypes under various conditions.

formulas were accurately defined. The most intensive phase of the workflow involved model extension and gap-filling, encompassing the detection and integration of missing metabolic genes along with their associated reactions and metabolites. Non-biologically occurring dead-end and orphan metabolites were identified and resolved where possible, as their existence implies knowledge gaps in metabolic networks. This step ensured the network's connectivity. Modifications in the model structure, as well as the inclusion of cross-references to multiple functional databases, were implemented using the libSBML[206] library, while all simulations were conducted with COBRApy[165]. With this, the draft model was expanded by 138 reactions, 77 genes, and 110 metabolites distributed across three compartments (cytosol, periplasm, and extracellular space). Overall, *i*ACB23LX comprises 2,321 reactions, 1,660 metabolites, and 1,164 genes (Fig. 3.4). It stands out as the most comprehensive model among previously published ones, maintaining a stoichiometric consistency of 100 %. Over 1,800 reactions have a GPR assigned, while 149 are catalyzed by enzyme complexes.

Additionally, the model was checked for the potential existence of thermodynamically infeasible internal loops, known as energy-generating cycles (EGCs), that could inflate the maximal biomass yields and bias the final predictions[285]. Unlike futile cycles, EGCs have not been experimentally observed. These cycles charge energy metabolites such as ATP and UTP without external nutrient sources, resulting in erroneous energy increases. Therefore, their elimination is
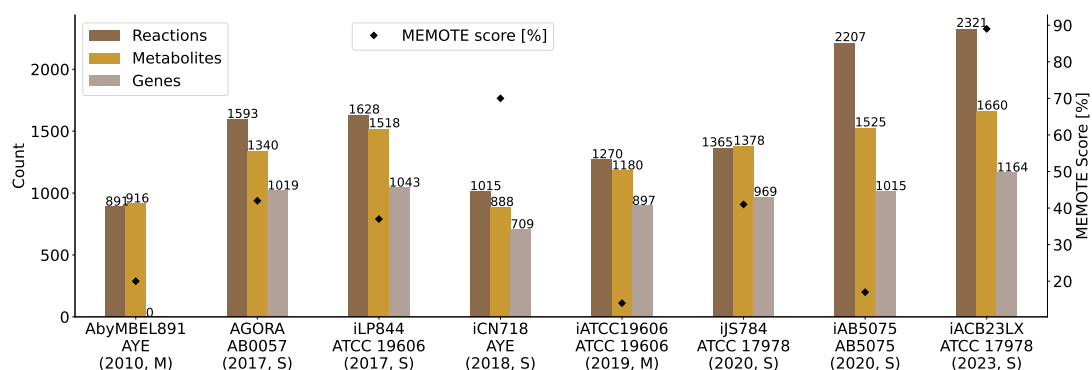
Figure 3.4: Properties of existing strain-specific metabolic networks for *A. baumannii*. The left ordinate shows the counts, the right ordinate represents the MEMOTE scores, while abscissa labels are annotated with the corresponding strains. The reconstruction process is categorized into manual (M, accomplished without the aid of computational tools) and semi-automated (S, draft obtained using an automated reconstruction tool, while subsequent extensions were done manually). Each model is associated with its respective publication year. The *i*ACB23LX exhibits the highest quality score and is more extensive than preceding reconstructions.

crucial in rectifying energy metabolism, as they can render predictions unreliable. For this purpose, energy dissipation reactions (EDRs) were (1) incorporated within the network according to the following scheme:

$$X[c] + H_2O[c] \longrightarrow X^+[c] + H[c] + Pi[c], \qquad (3.1)$$
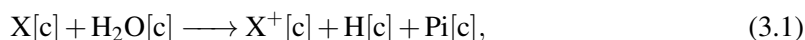
where X is the metabolite of interest and (2) were maximized individually while blocking all influxes[285]. In the final model, none of the tested metabolites could be produced; thus, no EGCs were contained. Thereafter, a multitude of database cross-references was embedded into the model, with SBO terms assigned to every reaction, metabolite, and gene[286]. Additionally, every reaction was linked to an ECO term that conveys the confidence level and the assertion method. The model's semantic and syntactic quality was evaluated using the MEMOTE framework[210] and the SBML Validator from libSBML[206]. The metabolic network, *i*ACB23LX, achieved a MEMOTE score of 89 %, with successful resolution of all syntactical errors. However, the testing algorithm exclusively considers the parent nodes of the SBO directed acyclic graph, disregarding their respective children. Therefore, the assignment of more precise SBO terms wrongly diminishes the quality total score by 2 %. Despite this, the model still outperforms its predecessors, as depicted in Fig. 3.4, securing the highest quality score.

A. baumannii is strictly aerobic and, unlike most *Acinetobacter* species, it is not considered ubiquitous in nature[80, 287]. As a nosocomial pathogen, it is highly prevalent in hospital settings, particularly within intensive care units (ICUs), as well as in the human nasal microbiota[288–290]. Various growth conditions were examined to ensure that *i*ACB23LX accurately recapitulates the growth in well-established growth environments. Initially, the model's capability to simulate strictly aerobic growth was assessed. An accumulation of periplasmic oxygen guided by reactions with high fluxes led to growth when oxygen import was disabled. Individual examination and removal of reactions lacking genetic evidence were performed, including the removal of the periplasmic catalase (BiGG ID: CATpp), a hydrogen peroxide scavenger usually active in the cytosol[291]. This enzyme was either absent in precursor *A. baumannii* GEMs or exclusively present in the cytosol. To enable periplasmic hydrogen peroxide utilization, the model was

Table 3.2: Simulated and empirical growth rates of ATCC 17978 in various growth media. The computationally-defined minimal medium (*i*MinMed), the lysogeny broth (LB) medium, and the synthetic nasal medium (SNM)[294] were tested. Growth rates are given in $mmol/(g_{DW} \cdot h)$, while doubling times are computed in minutes.

	Growth Rates		**Doubling Times**	
	in silico	*in vitro*	*in silico*	*in vitro*
*i*MinMed	0.2097	0.5402	198.32	76.70
LB	0.5926	0.7369	70.18	56.44
SNM	0.5099	0.3592	81.56	115.78

augmented with the inclusion of phenethylamine oxidase (BiGG ID: PEAMNOpp). The minimal amount of metabolites required for *A. baumannii* growth was determined using *i*ACB23LX and M9 minimal medium (M9) as a reference. The minimal medium (*i*MinMed), includes essential carbon ($C_2H_3O_2{}^-$), nitrogen ($NH_4{}^+$), sulfur ($SO_4{}^{2-}$), and phosphorus ($HPO_4{}^{2-}$) sources, in addition to nine transition metals (Ca^{2+}, Cl^-, Cu^{2+}, Fe^{3+}, Co^{2+}, K^+, Mg^{2+}, Mn^{2+}, and Zn^{2+}) crucial for cellular processes. Previous studies have emphasized the significance of nutrient metals for *A. baumannii*'s survival within the host. In particular, the bacterium utilizes these metals as co-factors for vital cellular processes[292]. Notably, manganese and zinc are identified as key elements in the host defense against *A. baumannii*-acquired pneumonia, where calprotectin sequesters them through chelation[293].

The successful bacterial growth utilizing the predicted medium *i*MinMed was experimentally confirmed through a positive empirical growth rate (Table 3.2). To ensure the robustness of the model predictions across diverse environmental conditions, experimental validation and model verification were conducted using the LB and SNM[294]. Computational growth rates below $2.81 \, mmol/(g_{DW} \cdot h)$ were considered realistic based on the doubling time of the fastest organism *Vibrio natriegens*[210]. Predicted and experimentally derived growth rates in various culture media are presented in Table 3.2. The highest rate was observed in LB medium, both *in silico* and *in vitro*, with the model-predicted growth rates in all tested media were experimentally confirmed through positive empirical growth rates. In a nutrient-rich medium, the model demonstrated the highest biomass production flux [$2.1858 \, mmol/(g_{DW} \cdot h)$], although it remained below the growth rate of the fastest-growing organism. Initially, *i*ACB23LX could not simulate realistic growth rates in the examined media. Utilizing the gap-filling function of CarveMe[284], three enzymes were identified (BiGG IDs: PHPYROX, OXADC, and LCYSTAT) whose addition enabled successful growth across all tested media.

The functional validation and refinement of *i*ACB23LX involved high-throughput phenotypic data obtained from two studies published in 2013 and 2014, focusing on the nutrient utilization phenome and gene essentialities of *A. baumannii* ATCC 17978 [295, 296]. These data served as invaluable resources to assess the model's predictive accuracy across a spectrum of metabolic phenotypes. The model's capability to predict carbon and nitrogen utilization was assessed by examining 80 compounds as sole carbon and 48 compounds as sole nitrogen sources. For the remaining metabolites, either no identifier mapping was possible, or they were not part of the metabolic network. Discrepancies between experimental results and model simulations were carefully examined to refine and augment the model. Resolving false negative and false positive predictions entailed the identification of reactions requiring removal or addition to the network. The primary objective was to enhance model accuracy by eliminating reactions lacking genetic evidence, i.e., lacking GPRs. Reactions inaccurately included in the automated draft model

generated by CarveMe[284] were subsequently removed to mitigate false positive predictions. Additionally, missing reactions were identified and incorporated into the network to rectify false negative predictions. The reaction directionality was also reviewed and corrected, based on evidence from the organism-specific BioCyc database[118]. The overall accuracy of iACB23LX for the nutrient utilization tests was 86.3 % and 79.2 % for carbon and nitrogen sources, respectively. In more detail, the final model recapitulates 69 and 38 experimentally-derived phenotypes, respectively. Altogether, improved predictive performance was observed in iACB23LX compared to its predecessors. The model's precision in predicting growth phenotypes of various carbon sources surpassed the accuracy of iATCC19606 (84.3 %) and iLP844 (84 %), while it was comparable to that of iAB5075 (86.3 %).

In the second validation stage, in $silico$ single-gene deletions were conducted to identify essential genes. The results were compared to the strain-specific gene essentiality dataset published by Wang et al. in 2014[296]. These knockouts were simulated using the LB and the rich growth medium (i.e., all import reactions activated). To assess the impact of individual gene deletions on the bacterial biomass production, the fold change (FC) between the model's growth rate before (gr_{WT}) and after (gr_{KO}) a single knockout was computed. This is formulated as follows:

$$FC_{gr} = \frac{gr_{KO}}{gr_{WT}}. \tag{3.2}$$

To this end, when $FC_{gr} = 0$, indicating that its absence hinders the network's ability to generate at least one vital biomass metabolite, resulting in no growth. Correspondingly, if $FC_{gr} = 1$, the gene's removal does not influence the growth phenotype (deemed non-essential), whereas when $0 < FC_{gr} < 1$, the gene's absence partially affects biomass production (referred to as partially essential). Additionally, the impact of gene deletions was examined using two different optimization approaches: FBA[168] and MOMA[173]. In contrast to FBA, MOMA relies on quadratic programming, with the optimization problem centered around minimizing the Euclidean distance within the flux space. Furthermore, MOMA provides an approximation of the metabolic phenotype and relaxes the assumption of optimal growth flux for gene deletions[173]. Utilizing FBA, the model predicted 97, 75, and 991 genes as essential, partially essential, and inessential, respectively, with the LB medium defined. Conversely, optimization with MOMA resulted in 110, 85, and 968 genes with similar classifications. The pathways mainly linked to these genes involved the co-factor and vitamin biosynthesis, amino acid/nucleotide metabolism, energy metabolism, and terpenoid and polyketide metabolism. A detailed examination of the nutrient availability impact on gene essentiality was conducted by performing single-gene knockouts in a rich medium. Both optimization algorithms resulted in more essential genes when the model was subjected to alterations in metabolic behavior due to nutrient absence, i.e., with the LB growth medium, compared to the rich medium. In general, FBA revealed a broader range of genes considered dispensable for growth in both nutritional conditions. In contrast, MOMA categorized more genes as essential or partially essential, with the essential genes identified by FBA constituting a subset of those identified by MOMA. Validation of the in $silico$ simulation results against strain-specific gene essentiality data[296] demonstrated an overall accuracy of 87 % using both optimization methods. This exceeded the performance of all GEMs constructed for $A.$ $baumannii$ (e.g., 80.22 % for iCN718 and 72 % for iLP844), except iAB5075, which performed comparably. Further investigation into false negative genes involved analyzing their proteomes to identify human orthologs. This analysis aimed to eliminate potential cross-linkages with proteins similar to those found in the human host. The absence of certain metabolic pathways or enzymes in the human host is considered a valuable resource for identifying druggable targets against infectious diseases[297]. Among the 37 genes identified as essential by the model contradicting the experi-

Table 3.3: List of existing genome-scale metabolic models for *A. baumannii* that underwent curation. Default growth rates (i.e., model simulated as downloaded), the cellular compartments (C: cytosol, E: extracellular space, P: periplasm, and ER: endoplasmic reticulum), and the database used for the identifiers are listed in the table. MEMOTE scores both pre- and post-manual curation are given in the last column. Brown highlights the newly reconstructed model for the strain ATCC 17978.

	Availability	Used Identifiers	Growth by default $\mathrm{mmol}/(\mathrm{g_{DW}} \cdot \mathrm{h})$	Compartments	MEMOTE
AbyMBEL891[277]	BioModels	Customized	119	Cell	20 % + 17 %
AGORA[183]	VMH	VMH	134	C, E	42 % + 37 %
***i*LP844**[278]	Suppl. Mat.	ModelSeed	15.88	C, E, P	37 % + 21 %
***i*CN718**[279]	BiGG	BiGG	1.31	C, E, P, ER	70 % + 3 %
***i*ATCC1906**[280]	Suppl. Mat.	KEGG	46.34	C, E	14 % + 44 %
***i*JS784**[282]	GitHub	ModelSeed	0.0	C, E, P	41 % + 18 %
***i*AB5075**[281]	Suppl. Mat.	BiGG	1.729	C, E, P	17 % + 50 %
***i*ACB23LX**	BioModels	BiGG	0.1677	C, E, P	89 %

mental data, 17 were found to have no human orthologs, as determined using the NCBI Basic Local Alignment Search Tool (BLAST)p tool[298]. For instance, the enolpyruvylshikimate phosphate (EPSP) synthase (`A1S_2276`) and the chorismate synthase (`A1S_1694`) from the shikimate pathway exhibited no similarity to the human proteome. The shikimate pathway is particularly interesting due to its absence in the metabolome of the human host, while it is important in bacterial metabolism and virulence[299]. Numerous knockout studies have emphasized the importance of enzymes from the shikimate metabolism as promising targets against pathogenic microorganisms such as *Mycobacterium tuberculosis*[300], *Plasmodium falciparum*[301], and *Yersinia enterocolitica*[302]. In an *in vivo* investigation, Umland et al. identified these two gene products as essential using a clinical isolate of *A. baumannii* and a rat abscess infection model[303]. Additional essential genes found to have no human counterparts involved the riboflavin synthase (`A1S_0223`), phosphogluconate dehydratase (`A1S_0483`), and 2-keto-3-deoxy-6-phosphogluconate (KDPG) aldolase (`A1S_0484`). Previous studies have suggested these candidates and their associated metabolic pathways as potential agents against various microorganisms[304–307]. However, their effectiveness against *Acinetobacter* species remains unexplored, presenting an avenue for novel therapeutic strategies. This literature evidence supports the credibility of the results and implies that the genes identified as essential in the computational simulations presented here could be considered potential targets for antimicrobial intervention.

The druggability of the candidate proteins was assessed using the DrugBank database[308]. For instance, the database listed flavin mononucleotide and cobalt hexamine ion as known inhibitors of chorismate synthase with yet unknown functions. Conversely, glyphosate, shikimate-3-phosphate, and formic acid were experimentally determined to interact with EPSP synthase. Additionally, six non-homologous genes were annotated as hypothetical or putative in the KEGG[119] database and/or lacked enzyme-associated information. Aligning their sequences with the DrugBank unveiled two potential drug leads: `A1S_0589` matched the phosphocarrier protein HPr of *Enterococcus faecalis*, and `A1S_0706` resembled sugar phosphatase YbiV of *Escherichia coli*. Despite experimental confirmation of the binding of dexfosfoserine and aspartate beryllium trifluoride to these enzymes, their pharmacological action remain unknown[308]. Overall, *i*ACB23LX reports high agreement with all validation tests, establishing its suitability for systematically elucidating connections between genotypes and phenotypes.

Initiated in 2010 by Kim et al., the first GEM for *A. baumannii* AYE set the foundation for the development of additional metabolic networks (Table 3.3). However, subsequent GEMs

lacked consistency and contained outdated or syntactically invalid information. Here, these issues were systematically addressed, evaluating and debugging seven existing strain-specific models to create a curated, standardized, and updated collection for improved drug development strategies and metabolic engineering. To accomplish this, a workflow was developed with curation steps applicable to all models, focusing on standardizing and enhancing their usability within the community. This workflow closely aligns with the community-driven efforts for reconstructing reusable and translatable models[283]. The curation procedure involves changes in the format, amount, and quality of the included information. Computational determination of minimal growth media addressed inflated growth rates, and growth capabilities were assessed against experimentally-derived media, with LB medium used to identify lethal genes across all examined strains. However, a holistic strain-wise comparison was not possible due to strain-specific identifiers, unsuccessful growth, or absent genes in certain models. Consequently, only the essential genes across all models with identifiers that could be mapped with the Pathosystems Resource Integration Center (PATRIC) ID mapping tool[309] were considered for this analysis. Syntactical correctness and internal consistency were verified using the SBML Validator from libSBML[206]. Additional debugging steps involved addressing warnings, resolving issues related to nomenclature and the fbc extension, declaring missing model attributes, and annotating reactions, metabolites, and genes following MIRIAM guidelines[310] with ModelPolisher[274] and SBOannotator[286]. The final debugging steps included converting all models to the latest SBML[203] format and quality control using MEMOTE[210]. It is essential to highlight that no contextual modifications were made that could impact the models' prediction capabilities. Additionally, common metabolic phenotypes were predicted using the curated models, and the results were compared among the strains. The investigated models represent five different *A. baumannii* strains that are publicly available through databases and online repositories or can be acquired directly from the publications. However, the use of different identifiers hampers the comparison of metabolic networks. More specifically, *i*LP844 and *i*JS784 use ModelSEED[202] identifiers, *i*CN718 and *i*AB5075 use BiGG[199] identifiers, AbyMBEL891 has distinct identifiers not supported by any database, and *i*ATCC1906 utilizes KEGG[119] identifiers. Most models exhibited unrealistic growth rates compared to the doubling time of the fastest-growing organism *V. natriegens*[210], with *i*JS784 showing zero growth even with enabled imports, leading to its exclusion from further analysis (Table 3.3).

To validate the models' reflection of *A. baumannii* metabolic and growth capabilities, the flux through the models' biomass reactions was examined in growth media known to enable *A. baumannii* growth. The majority of models exhibited zero biomass flux in *i*MinMed, while the AGORA model additionally could not simulate growth in LB and SNM. To address these inaccuracies, minimal medium supplementations were identified to enable biomass production in these media. As already mentioned, *i*JS784 was excluded from further examination, together with AbyMBEL891, which hindered the analysis due to its non-standardized identifiers and missing genes. Positive growth rates were achieved for *i*ATCC1906 and *i*AB5075 when *i*MinMed was supplemented D-alanine and D-glucose 6-phosphate as well as GMP, respectively. The AGORA model simulated positive biomass production when meso-2,6-diaminoheptanedioate, menaquinone-8, niacinamide, heme, siroheme, and spermidine were added to LB. Similarly, this analysis revealed compounds that led to positive growth rates when added to SNM and *i*MinMed. Lastly, like with *i*ACB23LX, the LB medium, together with FBA and MOMA, were applied to detect lethal genes in all models. Despite significant efforts, a mapping scheme could not be derived between the strain-specific gene identifiers of *i*LP844 and *i*ATCC1906 to resolve gene identifiers. The following strains were excluded from the analysis due to data limitations, including the absence of strain AB5075 in KEGG[119] and difficulties in mapping gene identifiers

in *i*LP844 and *i*ATCC1906. Moreover, only two models were available for ATCC 17978: the newly created model previously described and presented in this thesis and *i*JS784 that simulated continuously zero growth and was excluded from the analysis. Consequently, an examination was carried out to identify genes necessary for growth among the remaining models across three different strains: AB0057 (AGORA), ATCC 17978 (*i*ACB23LX), and AYE (*i*CN718). More specifically, from totally 392 genes predicted as essential, 34 occurred in all strains. For instance, genes encoding for dephospho-coenzyme A (CoA) kinase, phosphopantetheinyl transferase, shikimate kinase, and chorismate synthase were essential across all strains. As already mentioned, chorismate synthase has no human-like counterpart. This, coupled with its crucial role in supporting growth across three strains, enhances the enzyme's prospects as a potential drug candidate for future therapeutic interventions. These findings highlight potential drug candidates in purine metabolism, transferases, pantothenate and CoA biosynthesis, and amino acid metabolism for future therapies.

The presented metabolic reconstruction and the curated set of strain-specific models facilitate the generation of new model-driven hypotheses concerning *A. baumannii*. These resources are indispensable in exploring the variations in metabolic behavior among different species, elucidating their responses to genetic and environmental changes. The assembled collection of curated and standardized models paves the way for subsequent studies, providing new insights and promoting the development of therapeutic approaches tailored to specific strains and species. This, in turn, facilitates the definition of the entire species and the formulation of new hypotheses. Furthermore, these models advance precision antimicrobial control strategies designed explicitly for *A. baumannii* strains. In summary, the presented workflows and models are multi-purpose tools that can further augment this collection by incorporating additional standardized, strain-specific metabolic reconstructions. This aims to extensively delineate the core and pan metabolic capabilities of *A. baumannii*. The model *i*ACB23LX, along with all curated models, are available at BioModels[201] as an SBML Level 3 Version 1[311] file distributed as OMEX archive[271] including annotation[272].

3.4   Metabolic insights into cystic fibrosis

This section presents the article: **Leonidou, N.**, Ostyn, L., Coenye, T., Crabbé, A., & Dräger, A. (2023). "Genome-scale model of *Rothia mucilaginosa* predicts gene essentialities and reveals metabolic capabilities." *Microbiology spectrum.*

The CF environment promotes the growth of diverse microorganisms, some of which lead to acute and chronic lung infections, while others may positively influence the disease progression.[95]. *R. mucilaginosa* is notably prevalent in the lungs of CF patients, attracting remarkable attention within polymicrobial CF environments[96, 312]. Recent studies have unveiled its anti-inflammatory properties using *in vitro* three-dimensional lung epithelial cell cultures and mouse models mimicking chronic lung diseases *in vivo*[108]. Additionally, it has been linked to severe infections within the human body[106]. However, its metabolic capacity and genotype-phenotype connections in isolated monoculture conditions are poorly known. This study presents the first manually curated and high-quality GEM of *R. mucilaginosa* DSM20746 (referred to as *i*RM23NL), offering a foundation for exploring metabolic phenotypes. The bacterium's substrate utilization behavior was experimentally analyzed, and the results were employed to validate and refine the metabolic reconstruction. Putative essential genes were identified utilizing computational approaches. Their metabolic impact was evaluated under different
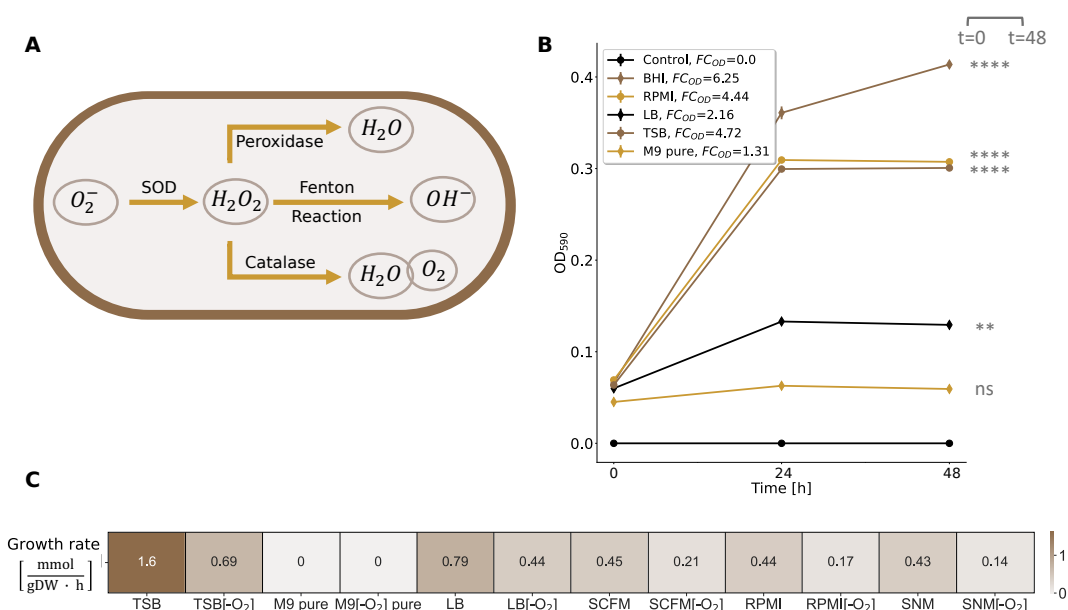
A

B

C

Figure 3.5: Growth behavior of *R. mucilaginosa* DSM20746 in different nutrient environments.
(A) The bacterium's metabolic response under anaerobic stress as represented in *i*RM23NL. **(B)**
Experimentally-derived growth curves in various liquid growth media, along with the corresponding
FCs defined in Eq. (3.3). The data shown here are an average of three biological replicates ($n = 3$).
The correctness of the threshold was verified by performing statistical analysis, whose significance
levels are represented by asterisks. **(C)** *In silico*-simulated growth rates using *i*RM23NL.

nutritional conditions, yielding new hypotheses for future antimicrobial strategies.

The pipeline described in the previous section was employed to construct the metabolic net-
work. Utilizing the bacterium's protein sequence containing more than 1,700 proteins and the
Gram-positive-specific template from CarveMe[284], the preliminary model was reconstructed.
This template provides a more accurate reconstruction by incorporating details about the pep-
tidoglycan layer for the biomass reaction. The initial network comprised 1,015 reactions (141
pseudo-reactions), 788 metabolites, and 265 genes. The reaction list was expanded in the first
gap-filling stage using the annotated genome and growth kinetics data in multiple growth en-
vironments. This resulted in incorporating 50 new GPRs into the model. High-throughput
nutrient screenings and model validation led to the inclusion of further metabolic reactions and
genes. More specifically, 79 reactions, along with 71 novel GPRs, were added, increasing the
genetic coverage. Altogether, over 60 % of transport reactions have GPR, and at least one gene
is assigned to 63 % of all enzymatic reactions. Missing exchange reactions for extracellular me-
tabolites were added, while the strain-specific BioCyc database[118] was used to adjust reaction
reversibility and eliminate duplicates. In cases where organism-specific information was un-
available, data from closely related species was employed. Using the ANIclustermap v.1.1.0[313],
the genome of DSM20746 was compared against 12 *Rothia* species (e.g., *R. aeria*, *R. dento-
cariosa*, and *R. terrae*). The analysis revealed similarity with six tested *Rothia* genomes, par-
ticularly exhibiting a closer relationship with *R. aeria* and *R. dentocariosa*. *R. mucilaginosa*
primarily employs aerobic respiration for ATP generation but shifts to anaerobic metabolism in
oxygen-depleted conditions, adapting to microaerophilic environments like CF lungs[100]. The
initial model could not simulate growth under anoxic conditions, prompting systematic incor-

poration of missing enzymes, including superoxide dismutase (SPODM) and catalase (CAT), responsible for radical reactive oxygen species (ROS) breakdown. Such scavenging enzymes help mitigate the negative effects of ROS during anaerobic respiration[314]. Additional scavenging enzymes, such as glutathione and thioredoxin reductases, necessary for preserving redox equilibrium[315] were already included in the draft model. Altogether, the final iRM23NL comprises 1,162 reactions, 171 exchange and sink reactions, 874 metabolites, and 372 genes. The model entities were enriched with database cross-references[274], assigned precise SBO terms using the SBOannotator[286], and checked for EGCs[285].

Growth kinetics data were employed to validate the performance of the reconstructed metabolic network. The assessment began by evaluating the model's accuracy in simulating biomass production under various environmental conditions and growth media formulations. To achieve this, the growth ability of R. mucilaginosa was experimentally and computationally tested in five different media: three general nutrient media and two defined media. The general media included brain heart infusion (BHI), LB, and tryptic soy broth (TSB), whereas the defined media consisted of M9 pure and the Roswell Park Memorial Institute (RPMI) medium. The BHI was chosen as a baseline for the experimental tests since it is a recognized and well-established environment for the bacterium's proliferation. The FC_{OD} defined as follows:

$$FC_{OD} = \frac{OD_{590\,nm}^{t=48\,h}}{OD_{590\,nm}^{t=0\,h}} \tag{3.3}$$

was used to qualitative compare the *in vitro* to *in silico* results. An $FC_{OD} < 1.4$ represents no growth, while $FC_{OD} > 1.4$ indicates a growth increase over time. Statistical tests were performed to validate the correctness of this threshold and detect potential significant differences between measurements at initial and final time points, indicating growth. The Student's *t*-test, applied to three biological replicates, assessed whether there was a significant difference in optical density (OD) values between initial and final time points. Pre-testing assumptions for data normality and variance homogeneity were conducted using the Shapiro–Wilk and Levene's tests, respectively. Both experimental and computational results demonstrated positive growth across all tested media, excluding the M9 pure (Fig. 3.5). Notably, the highest growth rates were consistently observed in TSB, while M9 pure exhibited the lowest growth, as corroborated by both *in vitro* and *in silico* analyses. The RPMI medium ranked the second-highest in supporting *in vitro* growth, offering a defined medium suitable for R. mucilaginosa's cultivation. Moreover, in contrast to the empirical results, iRM23NL, simulated a lower biomass flux with RPMI compared to LB. Interestingly, six metal ions were supplemented to facilitate a positive simulated growth in RPMI. These included cobalt (Co^{2+}), copper (Cu^{2+}), manganese (Mn^{2+}), zinc (Zn^{2+}), ferric iron (Fe^{3+}), and ferrous iron (Fe^{2+}). These metals have also been confirmed as essential for the *in silico* growth of S. aureus in RPMI[221]. Transition metals, while potentially highly toxic, play a vital role in the survival of all living organisms when present in controlled levels[316]. Notably, these compounds were not listed in the medium formulation provided by the suppliers, raising questions about its accuracy. In all cases, the potential metal co-factor promiscuity in R. mucilaginosa, as suggested by iRM23NL, should be further investigated to determine whether the bacterium can survive without any of the proposed metabolites. Furthermore, iRM23NL was employed to examine the bacterium's growth capabilities within the human nasal environment and CF lungs. For this purpose, the synthetic cystic fibrosis sputum medium (SCFM)[317] and the SNM[294] were used (Fig. 3.5). The computational model successfully demonstrated positive growth in both environments, exhibiting growth rates of $0.43\,mmol/(g_{DW} \cdot h)$ in SNM and $0.45\,mmol/(g_{DW} \cdot h)$ in SCFM. These findings correlate with R. mucilaginosa' established metabolic behavior in CF lungs[96, 312] and its frequent presence in the human nasal cavity[104]. Ad-

ditionally, the model accurately represented *R. mucilaginosa*'s facultative anaerobic respiration capacity, demonstrating successful growth through alternate metabolic routes across all tested nutritional media in the absence of oxygen. Overall, the results emphasize *R. mucilaginosa*'s adaptability to diverse nutritional environments, showcasing optimal growth in nutrient-rich conditions and highlighting specific growth requirements in nutrient-deficient environments.

The second validation step involved defining the complete nutrient utilization phenome of DSM20746. The strain's ability to assimilate and utilize various carbon, nitrogen, phosphorus, and sulfur sources was experimentally determined using high-throughput phenotypic microarray assays, while the results were subsequently compared to the model simulations. Four 96-well Biolog phenotype microarray (PM) microplates were assessed in duplicates, while independent tests confirmed the results. The attained transmittance (T) was converted to absorbance employing the following formula:

$$\text{Absorbance} = 2 - \log_{10}(\%T). \tag{3.4}$$

The area under curve (AUC) was used as a metric to distinguish between growth (AUC $\geq$ 50) and no growth (AUC $<$ 50). The computation of AUCs involved the utilization of the linear trapezoidal rule for interpolating between data points, as expressed by the following equation:

$$\text{AUC}_{(t_{i+1} - t_i)} = \int_{t_i}^{t_{i+1}} f(x)\, dx \approx (t_{i+1} - t_i) \cdot \frac{1}{2}(\text{OD}_{t_{i+1}} + \text{OD}_{t_i}). \tag{3.5}$$

Here, t_i represents the respective measured time point and $i \in \{0, \ldots, e\}$, with e being the final measured time. The trapezoidal rule was iteratively applied to adjacent data points defined along the curve whose summation resulted in the final AUC value. Hence, the final AUC value is defined as follows:

$$\text{AUC}_{t_e} = \sum_{i=0}^{e-1} \text{AUC}_{(t_{i+1} - t_i)}. \tag{3.6}$$

Among the 190 experimentally tested carbon substrates, *R. mucilaginosa* demonstrated the capability to utilize 61, including carboxylates, saccharides, and amino acids, while 10 out of 95 nitrogen sources were found to be viable. Additionally, out of 59 and 35 tested phosphorus and sulfur sources, *R. mucilaginosa* exhibited a loss of metabolic activity for 28 and 10 compounds, respectively. Subsequently, the predictive performance of *i*RM23NL in utilizing various C-, N-, P-, and S-containing substrates was evaluated. Among the 286 Biolog compounds successfully mapped to the BiGG database[199] and then to *i*RM23NL, 126 were identified as extracellular metabolites within the model and considered for further analysis. Discrepancies between the model and phenotypic microarray results were used as a foundation for additional model refinement and were addressed through extensive literature mining and iterative gap analysis. In cases where experimental results suggested the utilization of an undefined compound, missing transporters or enzymatic reactions were added based on strain-specific and gene-related information. If no organism-specific evidence was available, data from genomically identical species were considered. For instance, 3-sulfino-L-alanine was initially missing from the model. To achieve the expected positive utilization phenotype in *i*RM23NL, a BLAST[298] search identified cysteine desulfurase (SULFCYS) and associated transport reactions (proton-mediated; SULFCYSpp, diffusion; SULFCYStex, and ABC transport; SULFCYSabc), showing similarity exceeding 80 % with *R. dentocariosa*.

Furthermore, false negative or false positive predictions often result from missing or inaccurate transporter information. Corrections were made by removing transport reactions lacking gene evidence or adjusting reversibility for export. For example, corresponding irreversible
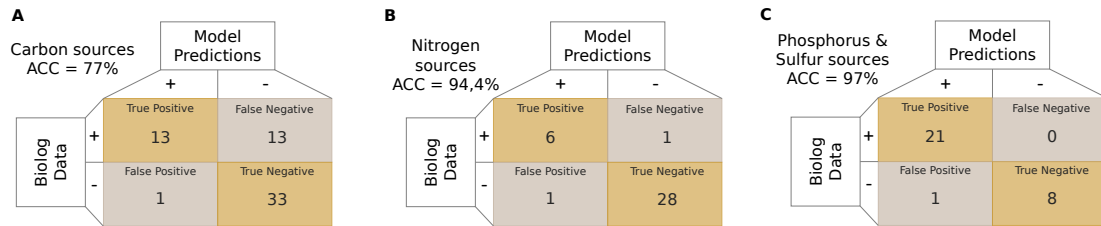
Figure 3.6: Predictive accuracy performance of *i*RM23NL using nutrient utilization data. Only substrates mapped to both BiGG and model identifiers could be analyzed. Multiple registered BiGG identifiers for compounds were also considered during analysis. Positive growth is labeled as "+", while "-" denotes the non-growth phenotype. The overall prediction accuracy of *i*RM23NL was computed using Eq. (3.5).

transporters were added when initial predictions contradicted Biolog assays for L-cysteate and adenosine 5'-monophosphate (AMP). Over 50 transport reactions were added during this refinement step, and 37 erroneous enzymatic functions were removed. However, resolving additional inconsistencies required incorporating metabolic reactions without associated gene evidence, highlighting gaps in the DSM20746 metabolism that necessitate in-depth investigation. Fourteen carbon and nitrogen sources failed to promote the experimentally observed growth in *i*RM23NL, despite the existence of corresponding transport reactions within the network (Fig. 3.6). Further information on their transport or metabolic mechanisms was absent in the genome annotation and the literature. In summary, *i*RM23NL achieved prediction accuracies of 77 % for carbon sources, 94.4 % for nitrogen sources, 100 % for phosphorus, and 92.3 % for sulfur sources (Fig. 3.6). This performance represents a 40 % increase post-comprehensive curation compared to the initial draft model. The refinement notably reduced false positive predictions by 17, leaving only three unresolved mismatches. A remarkable enhancement was observed in predicting nitrogen and phosphorus sources. This high predictive accuracy underscores the accurate reconstruction of core metabolic routes within *i*RM23NL, enabling successful predictions of the catabolism of various compounds like sugars and amino acids.

Given the increased proportion of gene-associated reactions and the high predictive accuracy of the metabolic network, *i*RM23NL was used to predict single gene knockouts. Each metabolic gene was removed from the network, and *i*RM23NL was optimized for biomass production using FBA. The FBA simulation was repeated 100 times to account for the inherent variability of optimization algorithms. Additionally, pFBA was employed, involving two sequential linear optimization problems to determine the optimal flux distribution while minimizing the total sum of flux. Predicted growth rates before and after gene deletions, as defined in Eq. (3.2), served as a measure of gene essentiality. Moreover, different growth conditions (M9 supplemented with glucose, LB, SNM, and SCFM) in both oxic and anoxic environments were considered. Under nutrient-limited conditions, the model consistently predicted a higher number of genes as essential for growth, while these numbers were consistent between oxic and anoxic conditions. In total, four metabolic genes [TrkA family potassium uptake protein (WP_005506372.1), ribulose-phosphate 3-epimerase (WP_005507411.1), glucose-6-phosphate isomerase (WP_005508482.1), and transaldolase (WP_005509117.1)], exhibited partial essentiality across all tested media, contributing to cellular fitness. Most essential genes were involved in nucleotide metabolism, peptidoglycan biosynthesis, and energy metabolism, highlighting their importance in supporting bacterial respiration. Notably, under nutrient-poor conditions, critical genes for survival were associated with the biosynthesis pathways of leucine, valine, and chorismate.

Following that, a protein sequence homology analysis using BLAST[298] against the human proteome was conducted to identify potential targets that could be considered in future therapeutic targets. Only genes deemed essential in both laboratory and synthetically-defined media were considered. In both LB and M9, there were 35 common essential genes, with 20 of them having homologous counterparts in the human genome. Further analysis showed that five genes had over 50 % sequence similarity with homologous proteins, but none exceeded 80 % similarity. Analogously, under aerobic and anaerobic conditions with SCFM and SNM, 45 shared genes were predicted as essential, with 25 having homology to the human genome, and seven showing over 50 % sequence similarity. Examples of these genes include phosphopyruvate hydratase, CTP synthase, and adenylosuccinate synthase, consistently exhibiting human counterparts with over 50 % similarity across all tested conditions. Among the common essential genes in LB and M9, 15 had no homologous hits. Similar observations were made for 20 common essential genes in SCFM and SNM, including orotate phosphoribosyltransferase, type I pantothenate kinas, dihydroneopterin aldolase, and pantetheine-phosphate adenylyltransferase. The main goal in studying *R. mucilaginosa* as a pathogen is to determine essential genes, especially those lacking human counterparts. Identifying these essential genes is crucial and aims to incapacitate the pathogen while ensuring the safety of the host. Simultaneously, *R. mucilaginosa* is being explored as an agent with anti-inflammatory properties. In this context, the focus is on promoting the bacterium's growth by enhancing specific environmental conditions known to support its proliferation. The gene essentiality analysis aims to propose enhancing the production of enzymes encoded by essential genes, as they play a pivotal role in the bacterium's beneficial functions. Nevertheless, knowing the *R. mucilaginosa*-specific essential genes is crucial to avoid inadvertently targeting them during treatments, guaranteeing the bacterium's growth and its anti-inflammatory functions. This dual perspective highlights the therapeutic versatility, encompassing methods to fight *Rothia* when it is harmful, while promoting cell biomass production when its anti-inflammatory effects are advantageous.

Overall, this study investigates *R. mucilaginosa*' metabolism, a less-explored member of the microbial community and CF patients. The first high-quality and manually curated GEM for *R. mucilaginosa* was reconstructed to enhance the systems-level understanding of its cellular metabolism. Experimental assessments of the bacterium's growth behavior under diverse environmental conditions and the determination of its high-throughput nutrient utilization phenome were determined for the first time. Integrating these *in vitro* findings within the metabolic network led to model's refinement, as well as the inclusion of novel metabolic reactions and genes, suggesting new avenues for research. Ultimately, the study explored the effects of condition-specific single-gene deletions. The *in silico* transposon mutant library created with *i*RM23NL could serve as a foundation for future research and practical applications, offering a robust approach for identifying potential antimicrobial targets and avoiding labor-intensive screening procedures. These applications include the development of rational and condition-specific drug targets and utilization in biotechnology and metabolic engineering. The final metabolic network has a MEMOTE[210] score of 89 % and is available at the BioModels database[201] as an SBML Level 3 Version 1[311] file wrapped in an OMEX archive file[271] together with a metadata file[272]. Finally, a step-by-step tutorial video on how to run simulations with *i*RM23NL is available at https://www.youtube.com/watch?v=Mcgov5H1kq0.

Conclusion

An ounce of prevention is worth a pound of cure.
– Benjamin Franklin

Systems biology is imperative for enhancing our understanding of biological systems and advancing disease intervention. By considering the complex interactions within biological networks, these methodologies offer a holistic perspective crucial for personalized medicine and drug discovery. The predictive capabilities of metabolic modeling facilitate the prediction of cellular responses and behaviors under different conditions. This aids in identifying potential drug targets and optimizing treatment strategies. Additionally, it substantially reduces the time and resources required for traditional experimental processes, particularly during the initial phase of the drug discovery pipeline.

This thesis focuses on computational biology and modeling approaches to analyze biochemical networks at the systems level. It explores methods to identify antiviral drug targets, decode host-pathogen interactions, and optimize future therapeutic methodologies for managing infectious diseases in diverse clinical settings. The conducted research encompasses software development, data analysis, computer modeling, and experimental validation, providing a multifaceted approach that integrates theoretical insights with practical applications.

The first goal of this thesis was to accelerate the prediction of exploitable host-based antiviral targets in the pre-clinical phase. This was accomplished by developing a computational workflow. The workflow is accompanied by software designed to calculate the metabolic stoichiometry of virus production and investigate virus-host interactions. While the workflow application was showcased using SARS-CoV-2 and its variants, its versatility allows for its transfer and utilization to known and emerging RNA viruses. Targeting host cell metabolic pathways offers advantages in robustness and evolutionary stability, leading to the identification of broad-spectrum putative therapeutics. Overall, this approach aims to enhance preparedness for future pandemics involving diverse viruses and host cell types, contributing to effective prevention strategies.

Following the FAIR principles, model standardization and reproducibility are crucial for the transparent and effective dissemination of scientific knowledge. Compliance with these principles enhances the accessibility of models, fosters collaboration, and promotes the reuse of computational tools, contributing to scientific research's overall robustness and reliability. Recognizing this need in systems biology, the tool presented in this thesis, called SBOannotator, successfully enables the automation of the assignment of SBO terms. More specifically, it addresses challenges in accurately annotating biochemical reactions, ensuring adherence to top-

level terms and enzyme functionality. Consequently, the second goal of establishing a systematic and standardized framework for model representation is achieved, thereby contributing to enhanced clarity and consistency in computational biology research.

The human nose serves as an infectious site due to its interface with the external environment, creating a direct link with the internal human body. Moreover, the nasal milieu can act as a reservoir for various bacteria, including those carrying antibiotic resistance determinants, like *A. baumannii*. Tackling the AMR challenge requires a comprehensive understanding of bacterial metabolic capacities. This goal was achieved in the third project by utilizing systems biology and computer modeling techniques to analyze multiple *A. baumannii* strains comprehensively. By creating and analyzing models that simulate the metabolic processes, scientists can acquire knowledge about the nutritional needs and metabolic pathways under various conditions. Exploring the metabolic profiling of *A. baumannii* in diverse growth environments holds clinical importance by providing insights into potential vulnerabilities for targeted interventions. Overall, these insights aim to enhance measures against *A. baumannii*-associated infections and could enable the implementation of personalized decolonization strategies.

Expanding beyond the investigation of nasal microbiota and AMR, this thesis explores the molecular complexities of cystic fibrosis. It focuses on the dual role of *R. mucilaginosa* in health and disease, a microorganism that remains understudied compared to other members of the human microbiota. The presented research endeavors to examine *R. mucilaginosa*'s metabolic processes in depth. This goal is accomplished by constructing the first manually curated GEM to unravel insights into the bacterium's metabolism and identify potential essentialities. The presented study addresses the experimental exploration of *R. mucilaginosa*'s phenotypes through high-throughput experiments, which were subsequently utilized to refine the metabolic network. The model-driven identification of condition-specific essential genes and exploitable antimicrobial targets underscores the bacterium's potential for expediting future therapeutic interventions. Therefore, the metabolic characteristics presented in this research are imperative for both fundamental comprehension and practical applications.

Integrating various data, including genomics and metabolomics, within the systems biology framework holds promise for a broader outlook in personalized interventions and preparation for future pandemics. As we envision the future, these integrated findings open new avenues for research, emphasizing the need for ongoing exploration and collaboration to harness the potential of systems biology in advancing human health. However, it is crucial to recognize that computational-driven hypotheses necessitate rigorous experimental validation and testing in both *in vitro* and *in vivo* settings to ensure their translational impact and clinical applicability. The continuous progress in omics technologies and the ongoing commitment to data integration and standardization efforts will play pivotal roles in shaping the future of systems biology, fostering advancements that hold great potential for improving human health in personalized and precision medicine. Ongoing research proves GEMs as excellent scaffolds for big data analysis, showcasing a distinctive strength in their adaptable nature. These inherently open-ended models allow for continuous expansion with novel biological knowledge, augmenting their predictive capabilities.

The continuing global challenges underscore the critical importance of pandemic preparedness and prevention, emphasizing the need for forward-looking and interdisciplinary approaches to safeguard public health. The research presented in this thesis advances infectious disease intervention through computational systems biology and *in silico* modeling. It offers comprehensive insights into unexplored areas, thereby advancing the application of systems biology principles.

Bibliography

1   C. J. Carlson, G. F. Albery, C. Merow, C. H. Trisos, C. M. Zipfel, E. A. Eskew, K. J. Olival, N. Ross, and S. Bansal. "Climate change increases cross-species viral transmission risk." In: *Nature* 607.7919 (2022), pp. 555–562. DOI: 10.1038/s41586-022-04788-w.

2   G. Robbins, V. M. Tripathy, V. N. Misra, R. K. Mohanty, V. S. Shinde, K. M. Gray, and M. D. Schug. "Ancient skeletal evidence for leprosy in India (2000 BC)." In: *PloS one* 4.5 (2009), e5669. DOI: 10.1371/journal.pone. 0005669.

3   L. Santacroce, R. Del Prete, I. A. Charitos, and L. Bottalico. "*Mycobacterium leprae*: A historical study on the origins of leprosy and its social stigma." In: *Le Infezioni in Medicina* 29.4 (2021), p. 623. DOI: 10.53854/liim-2904-18.

4   A. P. Dobson and E. R. Carper. "Infectious diseases and human population history." In: *Bioscience* 46.2 (1996), pp. 115–126. DOI: 10.2307/1312814.

5   A. R. Omran. "The epidemiologic transition: a theory of the epidemiology of population change." In: *he Milbank Memorial Fund Quarterly* 49.4 (1971), pp. 509–538. DOI: 10.2307/3349375.

6   R. Barrett, C. W. Kuzawa, T. McDade, and G. J. Armelagos. "Emerging and re-emerging infectious diseases: the third epidemiologic transition." In: *Annual review of anthropology* 27.1 (1998), pp. 247–271. DOI: 10. 1146/annurev.anthro.27.1.247.

7   W. H. McNeill. "Plagues and peoples." In: *Anchor Press* 53.1 (1976), pp. 145–147.

8   S. Morand, K. M. McIntyre, and M. Baylis. "Domesticated animals and human infectious diseases of zoonotic origins: domestication time matters." In: *Infection, Genetics and Evolution* 24 (2014), pp. 76–81. DOI: 10. 1016/j.meegid.2014.02.013.

9   N. D. Wolfe, C. P. Dunavan, and J. Diamond. "Origins of major human infectious diseases." In: *Nature* 447.7142 (2007), pp. 279–283. DOI: 10.1038/nature05775.

10   Thucydides. *History of the Peloponnesian War, Book2, Chapter VII*. Trans. by R. Crawley. Digireads.com Publishing, 2017, pp. 89–100. ISBN: 9781420956429.

11   S. Sabbatani and S. Fiorino. "The Antonine Plague and the decline of the Roman Empire." In: *Le Infezioni in medicina* 17.4 (2009), pp. 261–275.

12   K. Harper. "Pandemics and passages to late antiquity: rethinking the plague of c. 249–270 described by Cyprian." In: *Journal of Roman Archaeology* 28 (2015), pp. 223–260. DOI: 10.1017/S1047759415002470.

13   B. A. Cunha. "The cause of the plague of Athens: plague, typhoid, typhus, smallpox, or measles?" In: *Infectious Disease Clinics* 18.1 (2004), pp. 29–43. DOI: 10.1016/S0891-5520(03)00100-4.

14   R. J. Littman. "The plague of Athens: epidemiology and paleopathology." In: *Mount Sinai Journal of Medicine: A Journal of Translational and Personalized Medicine: A Journal of Translational and Personalized Medicine* 76.5 (2009), pp. 456–467. DOI: 10.1002/msj.20137.

15   T. Burki. "Understanding the history of infectious diseases." In: *The Lancet Infectious Diseases* 22.5 (2022), p. 602. DOI: 10.1016/S1473-3099(22)00158-X.

16   D. Stathakopoulos. "The Justinianic plague revisited." In: *Late Antiquity on the Eve of Islam*. Routledge, 2017, pp. 87–107. DOI: 10.1179/byz.2000.24.1.255.

17   Procopius. *History of the Wars*. Trans. by H. B. Dewing. Vol. 1. Loeb Classical Library 48. Cambridge, MA: Harvard University Press, 1914.

18   M. Eisenberg and L. Mordechai. "The Justinianic Plague: an interdisciplinary review." In: *Byzantine and Modern Greek Studies* 43.2 (2019), pp. 156–180. DOI: 10.1017/byz.2019.10.

19   O. J. Benedictow. *The Black Death, 1346-1353: the complete history*. Boydell & Brewer, 2004.

20 D. Herlihy. *The Black Death and the transformation of the West*. Harvard University Press, 1997. ISBN: 9780674076136.

21 B. P. Zietz and H. Dunkelberg. "The history of the plague and the research on the causative agent *Yersinia pestis*." In: *International journal of hygiene and environmental health* 207.2 (2004), pp. 165–178. DOI: 10.1078/1438-4639-00259.

22 J. Aberth. *Plagues in world history*. Rowman & Littlefield Publishers, 2011.

23 L. Seifert, I. Wiechmann, M. Harbeck, A. Thomas, G. Grupe, M. Projahn, H. C. Scholz, and J. M. Riehm. "Genotyping *Yersinia pestis* in historical plague: evidence for long-term persistence of *Y. pestis* in Europe from the 14th to the 17th century." In: *PLoS One* 11.1 (2016), e0145194. DOI: 10.1371/journal.pone.0145194.

24 N. Nunn and N. Qian. "The Columbian exchange: A history of disease, food, and ideas." In: *Journal of Economic Perspectives* 24.2 (2010), pp. 163–188. DOI: 10.1257/jep.24.2.163.

25 C. A. Roberts. *Leprosy: Past and present*. University Press of Florida, 2020.

26 W. H. Organization. *Leprosy - Fact sheet*. Accessed on February 1, 2024. URL: https://www.who.int/news-room/fact-sheets/detail/leprosy.

27 S. M. Faruque, M. J. Albert, and J. J. Mekalanos. "Epidemiology, genetics, and ecology of toxigenic *Vibrio cholerae*." In: *Microbiology and molecular biology reviews* 62.4 (1998), pp. 1301–1314. DOI: 10.1128/MMBR.62.4.1301-1314.1998.

28 D. L. Taylor, T. M. Kahawita, S. Cairncross, and J. H. Ensink. "The impact of water, sanitation and hygiene interventions to control cholera: a systematic review." In: *PLoS one* 10.8 (2015), e0135676. DOI: 10.1371/journal.pone.0135676.

29 A. Mutreja, D. W. Kim, N. R. Thomson, T. R. Connor, J. H. Lee, S. Kariuki, N. J. Croucher, S. Y. Choi, S. R. Harris, M. Lebens, et al. "Evidence for several waves of global transmission in the seventh cholera pandemic." In: *Nature* 477.7365 (2011), pp. 462–465. DOI: 10.1038/nature10392.

30 J. G. Morris Jr. "Cholera—modern pandemic disease of ancient lineage." In: *Emerging infectious diseases* 17.11 (2011), p. 2099. DOI: 10.3201/eid1711.111109.

31 C. J. Houldcroft and S. Underdown. "Infectious disease in the Pleistocene: Old friends or old foes?" In: *American Journal of Biological Anthropology* 182.4 (2023), pp. 513–531. DOI: 10.1002/ajpa.24737.

32 M. A. Spyrou, L. Musralina, G. A. Gnecchi Ruscone, A. Kocher, P.-G. Borbone, V. I. Khartanovich, A. Buzhilova, L. Djansugurova, K. I. Bos, D. Kühnert, et al. "The source of the Black Death in fourteenth-century central Eurasia." In: *Nature* 606.7915 (2022), pp. 718–724. DOI: 10.1038/s41586-022-04800-3.

33 M. Woolhouse, F. Scott, Z. Hudson, R. Howey, and M. Chase-Topping. "Human viruses: discovery and emergence." In: *Philosophical Transactions of the Royal Society B: Biological Sciences* 367.1604 (2012), pp. 2864–2871. DOI: 10.1098/rstb.2011.0354.

34 R. E. Baker, A. S. Mahmud, I. F. Miller, M. Rajeev, F. Rasambainarivo, B. L. Rice, S. Takahashi, A. J. Tatem, C. E. Wagner, L.-F. Wang, et al. "Infectious disease in an era of global change." In: *Nature Reviews Microbiology* 20.4 (2022), pp. 193–205. DOI: 10.1038/s41579-021-00639-z.

35 H. F. Parsons et al. *Report on the influenza epidemic of 1889-90*. Vol. 6387. HM Stationery Office, 1891.

36 D. M. Morens and A. S. Fauci. "The 1918 influenza pandemic: insights for the 21st century." In: *The Journal of infectious diseases* 195.7 (2007), pp. 1018–1028. DOI: 10.1086/511989.

37 Y. Kawaoka, S. Krauss, and R. G. Webster. "Avian-to-human transmission of the PB1 gene of influenza A viruses in the 1957 and 1968 pandemics." In: *Journal of virology* 63.11 (1989), pp. 4603–4608. DOI: 10.1128/jvi.63.11.4603-4608.1989.

38 G. J. Smith, D. Vijaykrishna, J. Bahl, S. J. Lycett, M. Worobey, O. G. Pybus, S. K. Ma, C. L. Cheung, J. Raghwani, S. Bhatt, et al. "Origins and evolutionary genomics of the 2009 swine-origin H1N1 influenza A epidemic." In: *Nature* 459.7250 (2009), pp. 1122–1125. DOI: 10.1038/nature08182.

39 L. Simonsen, P. Spreeuwenberg, R. Lustig, R. J. Taylor, D. M. Fleming, M. Kroneman, M. D. Van Kerkhove, A. W. Mounts, W. J. Paget, and G. C. Teams. "Global mortality estimates for the 2009 Influenza Pandemic from the GLaMOR project: a modeling study." In: *PLoS medicine* 10.11 (2013), e1001558. DOI: 10.1371/journal.pmed.1001558.

40 J. U. N. P. on HIV and A. (UNAIDS). *Global HIV AIDS statistics — Fact sheet*. Accessed on February 5, 2024. URL: https://www.unaids.org/en/resources/fact-sheet.

41 A. Mody, A. H. Sohn, C. Iwuji, R. K. Tan, F. Venter, and E. H. Geng. "HIV epidemiology, prevention, treatment, and implementation strategies for public health." In: *The Lancet* 403.10425 (2024), pp. 471–492. DOI: 10.1016/S0140-6736(23)01381-8.

42 I. Commission et al. "Ebola haemorrhagic fever in Zaire, 1976." In: *Bull. World Health Organ.* 56 (1978), pp. 271–293.

43 J. G. Breman, D. L. Heymann, G. Lloyd, J. B. McCormick, M. Miatudila, F. A. Murphy, J.-J. Muyembé-Tamfun, P. Piot, J.-F. Ruppol, P. Sureau, et al. "Discovery and description of Ebola Zaire virus in 1976 and relevance to the West African epidemic during 2013–2016." In: *The Journal of infectious diseases* 214.suppl_3 (2016), S93–S101. DOI: 10.1093/infdis/jiw207.

44 F. Wu, S. Zhao, B. Yu, Y.-M. Chen, W. Wang, Z.-G. Song, Y. Hu, Z.-W. Tao, J.-H. Tian, Y.-Y. Pei, et al. "A new coronavirus associated with human respiratory disease in China." In: *Nature* 579.7798 (2020), pp. 265–269. DOI: 10.1038/s41586-020-2202-3.

45 P. S. Masters, S. Perlman, et al. "Coronaviridae." In: *Fields virology* 1 (2013), pp. 825–858.

46 D. Baltimore. "Expression of animal virus genomes." In: *Bacteriological reviews* 35.3 (1971), pp. 235–241. DOI: 10.1128/br.35.3.235-241.1971.

47 A. G. Harrison, T. Lin, and P. Wang. "Mechanisms of SARS-CoV-2 transmission and pathogenesis." In: *Trends in immunology* 41.12 (2020), pp. 1100–1115. DOI: 10.1016/j.it.2020.10.004.

48 D. Cucinotta and M. Vanelli. "WHO declares COVID-19 a pandemic." In: *Acta bio medica: Atenei parmensis* 91.1 (2020), p. 157. DOI: 10.23750/abm.v91i1.9397.

49 R. Pfeiffer. "Vorläufige mittheilungen über die erreger der influenza." In: *DMW-Deutsche Medizinische Wochenschrift* 18.02 (1892), pp. 28–28.

50 W. Smith, C. H. Andrewes, P. P. Laidlaw, et al. "A virus obtained from influenza patients." In: *Lancet* (1933), pp. 66–8. DOI: 10.1016/S0140-6736(00)78541-2.

51 H. Brüssow and L. Brüssow. "Clinical evidence that the pandemic from 1889 to 1891 commonly called the Russian flu might have been an earlier coronavirus pandemic." In: *Microbial biotechnology* 14.5 (2021), pp. 1860–1870. DOI: 10.1111/1751-7915.13889.

52 P. Berche. "The enigma of the 1889 Russian flu pandemic: A coronavirus?" In: *La Presse Médicale* 51.3 (2022), p. 104111. DOI: 10.1016/j.lpm.2022.104111.

53 J. Y. Choi and D. M. Smith. "SARS-CoV-2 variants of concern." In: *Yonsei medical journal* 62.11 (2021), p. 961. DOI: 10.3349/ymj.2021.62.11.961.

54 W. H. Organization. *COVID-19 weekly epidemiological update 76– 25 January 2022*. Accessed on February 2, 2024. URL: https://www.who.int/publications/m/item/weekly-epidemiological-update-on-covid-19---25-january-2022.

55 B. Korber, W. M. Fischer, S. Gnanakaran, H. Yoon, J. Theiler, W. Abfalterer, N. Hengartner, E. E. Giorgi, T. Bhattacharya, B. Foley, et al. "Tracking changes in SARS-CoV-2 spike: evidence that D614G increases infectivity of the COVID-19 virus." In: *Cell* 182.4 (2020), pp. 812–827. DOI: 10.1016/j.cell.2020.06.043.

56 S. R. Kannan, A. N. Spratt, K. Sharma, H. S. Chand, S. N. Byrareddy, and K. Singh. "Omicron SARS-CoV-2 variant: Unique features and their impact on pre-existing antibodies." In: *Journal of autoimmunity* 126 (2022), p. 102779. DOI: 10.1016/j.jaut.2021.102779.

57 H. E. Davis, L. McCorkell, J. M. Vogel, and E. J. Topol. "Long COVID: major findings, mechanisms and recommendations." In: *Nature Reviews Microbiology* 21.3 (2023), pp. 133–146. DOI: 10.1038/s41579-022-00846-2.

58 M. S. Woo, J. Malsy, J. Pöttgen, S. Seddiq Zai, F. Ufer, A. Hadjilaou, S. Schmiedel, M. M. Addo, C. Gerloff, C. Heesen, et al. "Frequent neurocognitive deficits after recovery from mild COVID-19." In: *Brain communications* 2.2 (2020), fcaa205. DOI: 10.1093/braincomms/fcaa205.

59 H. Crook, S. Raza, J. Nowell, M. Young, and P. Edison. "Long covid—mechanisms, risk factors, and management." In: *bmj* 374 (2021). DOI: 10.1136/bmj.n1648.

60 H. E. Davis, G. S. Assaf, L. McCorkell, H. Wei, R. J. Low, Y. Re'em, S. Redfield, J. P. Austin, and A. Akrami. "Characterizing long COVID in an international cohort: 7 months of symptoms and their impact." In: *EClinicalMedicine* 38 (2021). DOI: 10.1016/j.eclinm.2021.101019.

61   W. H. Organization. *Statement on the fifteenth meeting of the IHR (2005) Emergency Committee on the COVID-19 pandemic*. Accessed on February 2, 2024. URL: https : / / www . who . int / news / item / 05 - 05 - 2023 - statement - on - the - fifteenth - meeting - of - the - international - health - regulations - (2005) - emergency - committee - regarding - the - coronavirus - disease - (covid - 19) - pandemic ? adgroupsurvey = %7Badgroupsurvey%7D&gclid=EAIaIQobChMI4Ojtsdbe_gIVjQRyCh07igt4EAAYASACEgJ9pfD_BwE& fbclid=IwAR2M8EAyiSrAodhK9p-X582nHkP2AigpSX8pYIsLsPwqYh4SG26RGokGe7E.

62   W. H. Organization. *World Health Organization Coronavirus (COVID-19) Dashboard*. Accessed on February 2, 2024. URL: https://covid19.who.int.

63   P. Venkatesan. "Re-emergence of infectious diseases associated with the past." In: *The Lancet Microbe* 2.4 (2021), e140. DOI: 10.1016/S2666-5247(21)00066-5.

64   World Health Organization. *Tuberculosis*. Accessed on February 21, 2024. URL: https://www.who.int/news-room/fact-sheets/detail/tuberculosis.

65   M. S. Smolinski, M. A. Hamburg, and J. Lederberg. "Microbial threats to health: Emergence, Detection, and Response." In: *Emergence, detection and response* (2003). DOI: 10.17226/10636.

66   R. Khabbaz, B. P. Bell, A. Schuchat, S. M. Ostroff, R. Moseley, A. Levitt, and J. M. Hughes. "Emerging and reemerging infectious disease threats." In: *Mandell, Douglas, and Bennett's principles and practice of infectious diseases* (2015), p. 158. DOI: 10.1016/B978-1-4557-4801-3.00014-X.

67   R. M. San Lio, G. Favara, A. Maugeri, M. Barchitta, and A. Agodi. "How Antimicrobial Resistance Is Linked to Climate Change: An Overview of Two Intertwined Global Challenges." In: *IJERPH* 20.3 (2023), pp. 1–18. DOI: 10.3390/ijerph20031681.

68   S. Allcock, E. H. Young, M. Holmes, D. Gurdasani, G. Dougan, M. S. Sandhu, L. Solomon, and M. Török. "Antimicrobial resistance in human populations: challenges and opportunities." In: *Global health, epidemiology and genomics* 2 (2017), e4. DOI: 10.1017/gheg.2017.4.

69   R. J. Hassing, J. Alsma, M. S. Arcilla, P. J. van Genderen, B. H. Stricker, and A. Verbon. "International travel and acquisition of multidrug-resistant *Enterobacteriaceae*: a systematic review." In: *Eurosurveillance* 20.47 (2015), p. 30074. DOI: 10.2807/1560-7917.ES.2015.20.47.30074.

70   M. I. Hutchings, A. W. Truman, and B. Wilkinson. "Antibiotics: past, present and future." In: *Current opinion in microbiology* 51 (2019), pp. 72–80. DOI: 10.1016/j.mib.2019.10.008.

71   A. Fleming. "On the antibacterial action of cultures of a penicillium, with special reference to their use in the isolation of *B. influenzae*." In: *British journal of experimental pathology* 10.3 (1929), p. 226.

72   C. J. Murray, K. S. Ikuta, F. Sharara, L. Swetschinski, G. R. Aguilar, A. Gray, C. Han, C. Bisignano, P. Rao, E. Wool, et al. "Global burden of bacterial antimicrobial resistance in 2019: a systematic analysis." In: *The Lancet* 399.10325 (2022), pp. 629–655. DOI: 10.1016/S0140-6736(21)02724-0.

73   J. O'neill. *Antimicrobial resistance: tackling a crisis for the health and wealth of nations*. 2014. URL: https://amr-review.org/sites/default/files/AMR%20Review%20Paper%20-%20Tackling%20a%20crisis%20for%20the%20health%20and%20wealth%20of%20nations_1.pdf.

74   B. S. Seal, D. Drider, B. B. Oakley, H. Brüssow, D. Bikard, J. O. Rich, S. Miller, E. Devillard, J. Kwan, G. Bertin, et al. "Microbial-derived products as potential new antimicrobials." In: *Veterinary research* 49 (2018), pp. 1–12. DOI: 10.1186/s13567-018-0563-5.

75   K. Hoelzer, N. Wong, J. Thomas, K. Talkington, E. Jungman, and A. Coukell. "Antimicrobial drug use in food-producing animals and associated human health risks: what, and how strong, is the evidence?" In: *BMC veterinary research* 13 (2017), pp. 1–38. DOI: 10.1186/s12917-017-1131-3.

76   World Health Organization. *WHO publishes list of bacteria for which new antibiotics are urgently needed*. Accessed on February 21, 2024. URL: https://www.who.int/news/item/27-02-2017-who-publishes-list-of-bacteria-for-which-new-antibiotics-are-urgently-needed.

77   L. Dijkshoorn, A. Nemec, and H. Seifert. "An increasing threat in hospitals: multidrug-resistant *Acinetobacter baumannii*." In: *Nature Reviews Microbiology 2007 5:12* 5 (12 Dec. 2007), pp. 939–951. ISSN: 1740-1534. DOI: 10.1038/nrmicro1789.

78   S. Ibrahim, N. Al-Saryi, I. M. Al-Kadmy, and S. N. Aziz. "Multidrug-resistant *Acinetobacter baumannii* as an emerging concern in hospitals." In: *Molecular Biology Reports* 48 (10 Oct. 2021), p. 6987. ISSN: 15734978. DOI: 10.1007/S11033-021-06690-6.

79   P. E. Fournier and H. Richet. "The epidemiology and control of *Acinetobacter baumannii* in health care facilities." In: *Clinical Infectious Diseases* 42 (5 Mar. 2006), pp. 692–699. ISSN: 10584838. DOI: 10.1086/500202/ 2/42-5-692-TBL003.GIF.

80   A. Y. Peleg, H. Seifert, and D. L. Paterson. "*Acinetobacter baumannii*: Emergence of a Successful Pathogen." In: *Clinical Microbiology Reviews* 21 (3 July 2008), p. 538. ISSN: 08938512. DOI: 10.1128/CMR.00058-07.

81   P. Bogaerts, T. Naas, I. Wybo, C. Bauraing, O. Soetens, D. Piérard, P. Nordmann, and Y. Glupczynski. "Outbreak of infection by carbapenem-resistant *Acinetobacter baumannii* producing the carbapenemase OXA-58 in Belgium." In: *Journal of clinical microbiology* 44.11 (2006), pp. 4189–4192. DOI: 10.1128/JCM.00796-06.

82   K. Rangel, T. P. G. Chagas, and S. G. De-Simone. "*Acinetobacter baumannii* infections in times of COVID-19 pandemic." In: *Pathogens* 10.8 (2021), p. 1006. DOI: 10.3390/pathogens10081006.

83   D. Contou, A. Claudinon, O. Pajot, M. Micaëlo, P. Longuet Flandre, M. Dubert, R. Cally, E. Logre, M. Fraissé, H. Mentec, et al. "Bacterial and viral co-infections in patients with severe SARS-CoV-2 pneumonia admitted to a French ICU." In: *Annals of intensive care* 10.1 (2020), pp. 1–9. DOI: 10.1186/s13613-020-00736-x.

84   W. G. Lima, J. C. M. Brito, and W. S. da Cruz Nizer. "Ventilator-associated pneumonia (VAP) caused by carbapenem-resistant *Acinetobacter baumannii* in patients with COVID-19: Two problems, one solution?" In: *Medical hypotheses* 144 (2020), p. 110139. DOI: 10.1016/j.mehy.2020.110139.

85   A. Russo, F. Gavaruzzi, G. Ceccarelli, C. Borrazzo, A. Oliva, F. Alessandri, E. Magnanimi, F. Pugliese, and M. Venditti. "Multidrug-resistant *Acinetobacter baumannii* infections in COVID-19 patients hospitalized in intensive care unit." In: *Infection* 50.1 (2022), pp. 83–92. DOI: 10.1007/s15010-021-01643-4.

86   C. Ayoub Moubareck and D. Hammoudi Halat. "Insights into *Acinetobacter baumannii*: A Review of Microbiological, Virulence, and Resistance Traits in a Threatening Nosocomial Pathogen." In: *Antibiotics* 9.3 (2020), p. 119. DOI: 10.3390/antibiotics9030119.

87   R. M. Carrillo-Larco, C. Anza-Ramírez, G. Saal-Zapata, D. Villarreal-Zegarra, J. H. Zafra-Tanaka, C. Ugarte-Gil, and A. Bernabé-Ortiz. "Type 2 diabetes mellitus and antibiotic-resistant infections: a systematic review and meta-analysis." In: *J Epidemiol Community Health* 76.1 (2022), pp. 75–84. DOI: 10.1136/jech-2020-216029.

88   N. Narayanan, T. Lin, D. Vinarov, T. Bucek, L. Johnson, C. Mathew, S. Chaudhry, and L. Brunetti. "Relationship between multidrug-resistant enterobacterales and obesity in older adults." In: *Infection and Drug Resistance* (2021), pp. 2527–2532. DOI: 10.2147/IDR.S317014.

89   J. L. Bobadilla, M. Macek Jr, J. P. Fine, and P. M. Farrell. "Cystic fibrosis: a worldwide analysis of CFTR mutations—correlation with incidence data and application to screening." In: *Human mutation* 19.6 (2002), pp. 575–606. DOI: 10.1002/humu.10041.

90   P. M. Quinton. "Cystic fibrosis: impaired bicarbonate secretion and mucoviscidosis." In: *The Lancet* 372.9636 (2008), pp. 415–417. DOI: 10.1016/S0140-6736(08)61162-9.

91   J. Guo, A. Garratt, and A. Hill. "Worldwide rates of diagnosis and effective treatment for cystic fibrosis." In: *Journal of Cystic Fibrosis* 21.3 (2022), pp. 456–462. DOI: 10.1016/j.jcf.2022.01.009.

92   G. R. Cutting. "Cystic fibrosis genetics: from molecular understanding to clinical application." In: *Nature Reviews Genetics* 16.1 (2015), pp. 45–56. DOI: 10.1038/nrg3849.

93   D. Worlitzsch, R. Tarran, M. Ulrich, U. Schwab, A. Cekici, K. C. Meyer, P. Birrer, G. Bellon, J. Berger, T. Weiss, et al. "Effects of reduced mucus oxygen concentration in airway *Pseudomonas* infections of cystic fibrosis patients." In: *The Journal of clinical investigation* 109.3 (2002), pp. 317–325. DOI: 10.1172/JCI13870.

94   L. M. Filkins and G. A. O'Toole. "Cystic fibrosis lung infections: polymicrobial, complex, and hard to treat." In: *PLoS pathogens* 11.12 (2015), e1005258. DOI: 10.1371/journal.ppat.1005258.

95   K. E. Price, T. H. Hampton, A. H. Gifford, E. L. Dolben, D. A. Hogan, H. G. Morrison, M. L. Sogin, and G. A. O'Toole. "Unique microbial communities persist in individual cystic fibrosis patients throughout a clinical exacerbation." In: *Microbiome* 1.1 (2013), pp. 1–11. DOI: 10.1186/2049-2618-1-27.

96   M. M. Tunney, T. R. Field, T. F. Moriarty, S. Patrick, G. Doering, M. S. Muhlebach, M. C. Wolfgang, R. Boucher, D. F. Gilpin, A. McDowell, et al. "Detection of anaerobic bacteria in high numbers in sputum from patients with cystic fibrosis." In: *American journal of respiratory and critical care medicine* 177.9 (2008), pp. 995–1001. DOI: 10.1164/rccm.200708-1151OC.

97   M. S. Muhlebach, J. E. Hatch, G. G. Einarsson, S. J. McGrath, D. F. Gilipin, G. Lavelle, B. Mirkovic, M. A. Murray, P. McNally, N. Gotman, et al. "Anaerobic bacteria cultured from cystic fibrosis airways correlate to milder disease: a multisite study." In: *European Respiratory Journal* 52.1 (2018). DOI: 10.1183/13993003.00242-2018.

98   A. M. Guss, G. Roeselers, I. L. Newton, C. R. Young, V. Klepac-Ceraj, S. Lory, and C. M. Cavanaugh. "Phylogenetic and metabolic diversity of bacteria associated with cystic fibrosis." In: *The ISME journal* 5.1 (2011), pp. 20–29. DOI: 10.1038/ismej.2010.88.

99   F. Bittar, H. Richet, J.-C. Dubus, M. Reynaud-Gaubert, N. Stremler, J. Sarles, D. Raoult, and J.-M. Rolain. "Molecular detection of multiple emerging pathogens in sputa from cystic fibrosis patients." In: *PloS one* 3.8 (2008), e2908. DOI: 10.1371/journal.pone.0002908.

100   Y. W. Lim, R. Schmieder, M. Haynes, M. Furlan, T. D. Matthews, K. Whiteson, S. J. Poole, C. S. Hayes, D. A. Low, H. Maughan, et al. "Mechanistic model of *Rothia mucilaginosa* adaptation toward persistence in the CF lung, based on a genome reconstructed from metagenomic data." In: *PloS one* 8.5 (2013), e64285. DOI: 10.1371/journal.pone.0064285.

101   T. Bergan and M. Kocur. "*Stomatococcus mucilaginosus* gen. nov., sp. nov., ep. rev., a member of the family *Micrococcaceae*." In: *International Journal of Systematic and Evolutionary Microbiology* 32.3 (1982), pp. 374–377.

102   I. Olsen, D. Preza, J. A. Aas, and B. J. Paster. "Cultivated and not-yet-cultivated bacteria in oral biofilms." In: *Microbial Ecology in Health and Disease* 21.2 (2009), pp. 65–71. DOI: 10.1080/08910600902907509.

103   S. Guglielmetti, V. Taverniti, M. Minuzzo, S. Arioli, M. Stuknyte, M. Karp, and D. Mora. "Oral bacteria as potential probiotics for the pharyngeal mucosa." In: *Applied and environmental microbiology* 76.12 (2010), pp. 3948–3958. DOI: 10.1128/AEM.00109-10.

104   M. L. Wos-Oxley, I. Plumeier, C. Von Eiff, S. Taudien, M. Platzer, R. Vilchez-Vargas, K. Becker, and D. H. Pieper. "A poke into the diversity and associations within human anterior nare microbial communities." In: *The ISME journal* 4.7 (2010), pp. 839–851. DOI: 10.1038/ismej.2010.15.

105   S. A. Wilbert, J. L. M. Welch, and G. G. Borisy. "Spatial ecology of the human tongue dorsum microbiome." In: *Cell reports* 30.12 (2020), pp. 4003–4015. DOI: 10.1016/j.celrep.2020.02.097.

106   S. G. Dastager, S. Krishnamurthi, N. Rameshkumar, and M. Dharne. *The Family* Micrococcaceae. Springer Berlin, Heidelberg, 2014, pp. 455–498. ISBN: 978-3-642-30138-4. DOI: 10.1007/978-3-642-30138-4_168.

107   B. Gao, T. Gallagher, Y. Zhang, M. Elbadawi-Sidhu, Z. Lai, O. Fiehn, and K. L. Whiteson. "Tracking polymicrobial metabolism in cystic fibrosis airways: *Pseudomonas aeruginosa* metabolism and physiology are influenced by *Rothia mucilaginosa*-derived metabolites." In: *Msphere* 3.2 (2018), pp. 10–1128. DOI: 10.1128/mSphere.00151-18.

108   C. Rigauts, J. Aizawa, S. L. Taylor, G. B. Rogers, M. Govaerts, P. Cos, L. Ostyn, S. Sims, E. Vandeplassche, M. Sze, et al. "*Rothia mucilaginosa* is an anti-inflammatory bacterium in the respiratory tract of patients with chronic lung disease." In: *European Respiratory Journal* 59.5 (2022). DOI: 10.1183/13993003.01293-2021.

109   C. C. Uranga, P. Arroyo Jr, B. M. Duggan, W. H. Gerwick, and A. Edlund. "Commensal oral *Rothia mucilaginosa* produces enterobactin, a metal-chelating siderophore." In: *Msystems* 5.2 (2020), pp. 10–1128. DOI: 10.1128/mSystems.00161-20.

110   K. N. Raymond, E. A. Dertz, and S. S. Kim. "Enterobactin: an archetype for microbial iron transport." In: *Proceedings of the national academy of sciences* 100.7 (2003), pp. 3584–3588. DOI: 10.1073/pnas.0630018100.

111   D. Janek, A. Zipperer, A. Kulik, B. Krismer, and A. Peschel. "High frequency and diversity of antimicrobial activities produced by nasal *Staphylococcus* strains against bacterial competitors." In: *PLoS pathogens* 12.8 (2016), e1005812. DOI: 10.1371/journal.ppat.1005812.

112   H. P. Gurushankara. "Pandemics of the 21st century: lessons and future perspectives." In: *Pandemic Outbreaks in the 21st Century*. Elsevier, 2021, pp. 139–158. DOI: 10.1016/B978-0-323-85662-1.00011-2.

113   K. McGrail, J. Morgan, and A. Siddiqi. "Looking back and moving forward: addressing health inequities after COVID-19." In: *The Lancet Regional Health–Americas* 9 (2022). DOI: 10.1016/j.lana.2022.100232.

114   W. H. Organization. *World Bank and WHO: Half the world lacks access to essential health services, 100 million still pushed into extreme poverty because of health expenses*. Accessed on February 3, 2024. URL: https://www.who.int/about/accountability/results/who-results-report-2022-mtr/who-at-75-years-combatting-21st-century-health-challenges.

115   V. C. Cheng, S. K. Lau, P. C. Woo, and K. Y. Yuen. "Severe acute respiratory syndrome coronavirus as an agent of emerging and reemerging infection." In: *Clinical microbiology reviews* 20.4 (2007), pp. 660–694. DOI: [10.1128/CMR.00023-07](10.1128/CMR.00023-07).

116   W. H. Organization. *Prioritizing diseases for research and development in emergency contexts*. Accessed on February 3, 2024. URL: https://www.who.int/activities/prioritizing-diseases-for-research-and-development-in-emergency-contexts.

117   E. Holmes, I. D. Wilson, and J. K. Nicholson. "Metabolic phenotyping in health and disease." In: *Cell* 134.5 (2008), pp. 714–717. DOI: 10.1016/j.cell.2008.08.026.

118   P. D. Karp, R. Billington, R. Caspi, C. A. Fulcher, M. Latendresse, A. Kothari, I. M. Keseler, M. Krummenacker, P. E. Midford, Q. Ong, et al. "The BioCyc collection of microbial genomes and metabolic pathways." In: *Briefings in bioinformatics* 20.4 (2019), pp. 1085–1093. DOI: [10.1093/bib/bbx085](10.1093/bib/bbx085).

119   M. Kanehisa, M. Furumichi, Y. Sato, M. Ishiguro-Watanabe, and M. Tanabe. "KEGG: integrating viruses and cellular organisms." In: *Nucleic acids research* (Oct. 2020), pp. D545–D551. ISSN: 1362-4962. DOI: 10.1093/nar/gkaa970. aheadofprint.

120   Z. Yu, X. Zhou, and X. Wang. "Metabolic reprogramming in hematologic malignancies: advances and clinical perspectives." In: *Cancer Research* 82.17 (2022), pp. 2955–2963. DOI: [10.1158/0008-5472.CAN-22-0917](10.1158/0008-5472.CAN-22-0917).

121   M. Soltani, Y. Zhao, Z. Xia, M. Ganjalikhani Hakemi, and A. V. Bazhin. "The importance of cellular metabolic pathways in pathogenesis and selective treatments of hematological malignancies." In: *Frontiers in Oncology* 11 (2021), p. 767026. DOI: 10.3389/fonc.2021.767026.

122   M. Certo, H. Elkafrawy, V. Pucino, D. Cucchi, K. C. Cheung, and C. Mauro. "Endothelial cell and T-cell crosstalk: Targeting metabolism as a therapeutic approach in chronic inflammation." In: *British journal of pharmacology* 178.10 (2021), pp. 2041–2059. DOI: [10.1111/bph.15002](10.1111/bph.15002).

123   M. V. Liberti and J. W. Locasale. "The Warburg effect: how does it benefit cancer cells?" In: *Trends in biochemical sciences* 41.3 (2016), pp. 211–218. DOI: 10.1016/j.tibs.2015.12.001.

124   S. Jiang, J. L. Young, K. Wang, Y. Qian, and L. Cai. "Diabetic-induced alterations in hepatic glucose and lipid metabolism: The role of type 1 and type 2 diabetes mellitus." In: *Molecular Medicine Reports* 22.2 (2020), pp. 603–611. DOI: [10.3892/mmr.2020.11175](10.3892/mmr.2020.11175).

125   C. Weber and H. Noels. "Atherosclerosis: current pathogenesis and therapeutic options." In: *Nature medicine* 17.11 (2011), pp. 1410–1422. DOI: [10.1038/nm.2538](10.1038/nm.2538).

126   P. Singla, A. Bardoloi, and A. A. Parkash. "Metabolic effects of obesity: a review." In: *World journal of diabetes* 1.3 (2010), p. 76. DOI: [10.4239/wjd.v1.i3.76](10.4239/wjd.v1.i3.76).

127   S. Bhargava, S. De la Puente-Secades, L. Schurgers, and J. Jankowski. "Lipids and lipoproteins in cardiovascular diseases: a classification." In: *Trends in Endocrinology & Metabolism* 33.6 (2022), pp. 409–423. DOI: [10.1016/j.tem.2022.02.001](10.1016/j.tem.2022.02.001).

128   V. Kumar, S.-H. Kim, and K. Bishayee. "Dysfunctional glucose metabolism in alzheimer's disease onset and potential pharmacological interventions." In: *International Journal of Molecular Sciences* 23.17 (2022), p. 9540. DOI: 10.3390/ijms23179540.

129   J. M. Stokes, A. J. Lopatkin, M. A. Lobritz, and J. J. Collins. "Bacterial metabolism and antibiotic efficacy." In: *Cell metabolism* 30.2 (2019), pp. 251–259. DOI: 0.1016/j.cmet.2019.06.009.

130   W. Y. Chung, Y. Zhu, M. H. Mahamad Maifiah, N. K. Hawala Shivashekaregowda, E. H. Wong, and N. Abdul Rahim. "Exogenous metabolite feeding on altering antibiotic susceptibility in Gram-negative bacteria through metabolic modulation: a review." In: *Metabolomics* 18.7 (2022), p. 47. DOI: [10.1007/s11306-022-01903-w](10.1007/s11306-022-01903-w).

131   W. Eisenreich, T. Dandekar, J. Heesemann, and W. Goebel. "Carbon metabolism of intracellular bacterial pathogens and possible links to virulence." In: *Nature Reviews Microbiology* 8.6 (2010), pp. 401–412. DOI: 10.1128/microbiolspec.VMBF-0027-2015.

132   O. Ponomarova and K. R. Patil. "Metabolic interactions in microbial communities: untangling the Gordian knot." In: *Current opinion in microbiology* 27 (2015), pp. 37–44. DOI: [10.1016/j.mib.2015.06.014](10.1016/j.mib.2015.06.014).

133   J. Nogales and J. Garmendia. "Bacterial metabolism and pathogenesis intimate intertwining: time for metabolic modelling to come into action." In: *Microbial Biotechnology* 15.1 (2022), pp. 95–102. DOI: [10.1111/1751-7915.13942](10.1111/1751-7915.13942).

134   P. Whitaker-Dowling and J. S. Youngner. "Virus-host cell interactions." In: *Encyclopedia of Virology* (1999), p. 1957. DOI: 10.1006/rwvi.1999.0343.

135   N. D. Maynard, M. V. Gutschow, E. W. Birch, and M. W. Covert. "The virus as metabolic engineer." In: *Biotechnology journal* 5.7 (2010), pp. 686–694. DOI: 10.1002/biot.201000080.

136   K. A. Mayer, J. Stöckl, G. J. Zlabinger, and G. A. Gualdoni. "Hijacking the supplies: metabolism as a novel facet of virus-host interaction." In: *Frontiers in immunology* (2019), p. 1533. DOI: 10.3389/fimmu.2019.01533.

137   M. A. O'Malley and J. Dupré. "Fundamental issues in systems biology." In: *BioEssays* 27.12 (2005), pp. 1270–1276. DOI: 10.1002/bies.20323.

138   M. Drack and O. Wolkenhauer. "System approaches of Weiss and Bertalanffy and their relevance for systems biology today." In: *Seminars in Cancer Biology*. Vol. 21. 3. Elsevier. 2011, pp. 150–155. DOI: 10.1016/j.semcancer.2011.05.001.

139   S. Kesić. "Systems biology, emergence and antireductionism." In: *Saudi journal of biological sciences* 23.5 (2016), pp. 584–591. DOI: 10.1016/j.sjbs.2015.06.015.

140   V. Saks, C. Monge, and R. Guzun. "Philosophical basis and some historical aspects of systems biology: from Hegel to Noble-applications for bioenergetic research." In: *International Journal of Molecular Sciences* 10.3 (2009), pp. 1161–1192. DOI: 10.3390/ijms10031161.

141   E. O. Voit. "Perspective: Systems biology beyond biology." In: *Frontiers in Systems Biology* 2 (2022), p. 987135. DOI: 10.3389/fsysb.2022.987135.

142   R. D. Fleischmann, M. D. Adams, O. White, R. A. Clayton, E. F. Kirkness, A. R. Kerlavage, C. J. Bult, J.-F. Tomb, B. A. Dougherty, J. M. Merrick, et al. "Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd." In: *science* 269.5223 (1995), pp. 496–512. DOI: 10.1126/science.7542800.

143   N. H. G. R. Institute. *The Human Genome Project - Fact Sheet*. Accessed on February 7, 2024. URL: https://www.genome.gov/about-genomics/educational-resources/fact-sheets/human-genome-project.

144   B. Palsson. "The challenges of in silico biology." In: *Nature biotechnology* 18.11 (2000), pp. 1147–1150. DOI: 10.1038/81125.

145   H. Kitano. "Systems biology: a brief overview." In: *science* 295.5560 (2002), pp. 1662–1664. DOI: 10.1126/science.1069492.

146   T. Ideker, T. Galitski, and L. Hood. "A new approach to decoding life: systems biology." In: *Annual review of genomics and human genetics* 2.1 (2001), pp. 343–372. DOI: 10.1146/annurev.genom.2.1.343.

147   M. W. Kirschner. "The meaning of systems biology." In: *Cell* 121.4 (2005), pp. 503–504. DOI: 10.1016/j.cell.2005.05.005.

148   F. J. Bruggeman and H. V. Westerhoff. "The nature of systems biology." In: *TRENDS in Microbiology* 15.1 (2007), pp. 45–50. DOI: 10.1016/j.tim.2006.11.003.

149   H. Kitano. "Computational systems biology." In: *Nature* 420.6912 (2002), pp. 206–210. DOI: 10.1038/nature01254.

150   B. Ø. Palsson. *Systems Biology: Constraint-based Reconstruction and Analysis*. Cambridge University Press, 2017. ISBN: 9781139854610.

151   J. Nielsen and M. C. Jewett. "Impact of systems biology on metabolic engineering of *Saccharomyces cerevisiae*." In: *FEMS yeast research* 8.1 (2008), pp. 122–131. DOI: 10.1111/j.1567-1364.2007.00302.x.

152   M. Koutrouli, E. Karatzas, D. Paez-Espino, and G. A. Pavlopoulos. "A guide to conquer the biological network era using graph theory." In: *Frontiers in bioengineering and biotechnology* 8 (2020), p. 34. DOI: 10.3389/fbioe.2020.00034.

153   A.-L. Barabasi and Z. N. Oltvai. "Network biology: understanding the cell's functional organization." In: *Nature reviews genetics* 5.2 (2004), pp. 101–113. DOI: 10.1038/nrg1272.

154   F. Képès. *Biological Networks - Complex Systems and Interdisciplinary Science*. Vol. 3. World Scientific, 2007. DOI: https://doi.org/10.1142/6459.

155   M. Domach, S. Leung, R. Cahn, G. Cocks, and M. Shuler. "Computer model for glucose-limited growth of a single cell of *Escherichia coli* B/r-A." In: *Biotechnology and bioengineering* 26.9 (1984), pp. 1140–1140. DOI: 10.1002/bit.260260303.

156   M. Tomita, K. Hashimoto, K. Takahashi, T. S. Shimizu, Y. Matsuzaki, F. Miyoshi, K. Saito, S. Tanida, K. Yugi, J. C. Venter, et al. "E-CELL: software environment for whole-cell simulation." In: *Bioinformatics (Oxford, England)* 15.1 (1999), pp. 72–84. DOI: 10.1093/bioinformatics/15.1.72.

157   J. R. Karr, J. C. Sanghvi, D. N. Macklin, M. V. Gutschow, J. M. Jacobs, B. Bolival, N. Assad-Garcia, J. I. Glass, and M. W. Covert. "A whole-cell computational model predicts phenotype from genotype." In: *Cell* 150.2 (2012), pp. 389–401. DOI: 10.1016/j.cell.2012.05.044.

158   L. M. Edwards. "Metabolic systems biology: a brief primer." In: *The Journal of physiology* 595.9 (2017), pp. 2849–2855. DOI: 10.1113/JP272275.

159   C. Angione. "Human systems biology and metabolic modelling: a review–from disease metabolism to precision medicine." In: *BioMed research international* 2019 (2019). DOI: 10.1155/2019/8304260.

160   G. Chalancon, K. Kruse, and M. M. Babu. *Metabolic networks, structure and dynamics*. 2013, pp. 1263–1267. ISBN: 978-1-4419-9863-7. DOI: 10.1007/978-1-4419-9863-7_561.

161   K. D. Rawls, B. V. Dougherty, E. M. Blais, E. Stancliffe, G. L. Kolling, K. Vinnakota, V. R. Pannala, A. Wallqvist, and J. A. Papin. "A simplified metabolic network reconstruction to promote understanding and development of flux balance analysis tools." In: *Computers in Biology and Medicine* 105 (2019), pp. 64–71. DOI: 10.1016/j.compbiomed.2018.12.010.

162   I. Thiele and B. Ø. Palsson. "A protocol for generating a high-quality genome-scale metabolic reconstruction." In: *Nature protocols* 5.1 (2010), pp. 93–121. DOI: 10.1038/nprot.2009.203.

163   W. Gottstein, B. G. Olivier, F. J. Bruggeman, and B. Teusink. "Constraint-based stoichiometric modelling from single organisms to microbial communities." In: *Journal of the Royal Society Interface* 13.124 (2016), p. 20160627. DOI: 10.1098/rsif.2016.0627.

164   S. A. Becker, A. M. Feist, M. L. Mo, G. Hannum, B. Ø. Palsson, and M. J. Herrgard. "Quantitative prediction of cellular metabolism with constraint-based models: the COBRA Toolbox." In: *Nature protocols* 2.3 (2007), pp. 727–738. DOI: 10.1038/nprot.2007.99.

165   A. Ebrahim, J. A. Lerman, B. Ø. Palsson, and D. R. Hyduke. "COBRApy: constraints-based reconstruction and analysis for python." In: *BMC systems biology* 7.1 (2013), pp. 1–6. DOI: 10.1186/1752-0509-7-74.

166   N. D. Price, J. L. Reed, and B. Ø. Palsson. "Genome-scale models of microbial cells: evaluating the consequences of constraints." In: *Nature Reviews Microbiology* 2.11 (2004), pp. 886–897. DOI: 10.1038/nrmicro1023.

167   K. Raman and N. Chandra. "Flux balance analysis of biological systems: applications and challenges." In: *Briefings in bioinformatics* 10.4 (2009), pp. 435–449. DOI: 10.1093/bib/bbp011.

168   J. D. Orth, I. Thiele, and B. Ø. Palsson. "What is flux balance analysis?" In: *Nature biotechnology* 28.3 (2010), pp. 245–248. DOI: 10.1038/nbt.1614.

169   J. S. Edwards and B. Ø. Palsson. "Systems properties of the *Haemophilus influenzae* Rd metabolic genotype." In: *Journal of Biological Chemistry* 274.25 (1999), pp. 17410–17416. DOI: 10.1074/jbc.274.25.17410.

170   A. M. Feist and B. Ø. Palsson. "The biomass objective function." In: *Current opinion in microbiology* 13.3 (2010), pp. 344–349. DOI: 10.1016/j.mib.2010.03.003.

171   S. Aller, A. Scott, M. Sarkar-Tyson, and O. S. Soyer. "Integrated human-virus metabolic stoichiometric modelling predicts host-based antiviral targets against Chikungunya, Dengue, and Zika viruses." In: *Journal of The Royal Society Interface* 15.146 (2018), p. 20180125. DOI: 10.1098/rsif.2018.0125.

172   S. Gudmundsson and I. Thiele. "Computationally efficient flux variability analysis." In: *BMC bioinformatics* 11.1 (2010), pp. 1–3. DOI: 10.1186/1471-2105-11-489.

173   D. Segre, D. Vitkup, and G. M. Church. "Analysis of optimality in natural and perturbed metabolic networks." In: *Proceedings of the National Academy of Sciences* 99.23 (2002), pp. 15112–15117. DOI: 10.1073/pnas.232349399.

174   N. E. Lewis, K. K. Hixson, T. M. Conrad, J. A. Lerman, P. Charusanti, A. D. Polpitiya, J. N. Adkins, G. Schramm, S. O. Purvine, D. Lopez-Ferrer, et al. "Omic data from evolved *E. coli* are consistent with computed optimal growth from genome-scale models." In: *Molecular systems biology* 6.1 (2010), p. 390. DOI: 10.1038/msb.2010.47.

175   S. S. Fong, A. P. Burgard, C. D. Herring, E. M. Knight, F. R. Blattner, C. D. Maranas, and B. Ø. Palsson. "*In silico* design and adaptive evolution of *Escherichia coli* for production of lactic acid." In: *Biotechnology and bioengineering* 91.5 (2005), pp. 643–648. DOI: 10.1002/bit.20542.

176   D. Machado and M. Herrgård. "Systematic evaluation of methods for integration of transcriptomic data into constraint-based models of metabolism." In: *PLoS computational biology* 10.4 (2014), e1003580. DOI: 10.1371/journal.pcbi.1003580.

177   R. Mahadevan and C. H. Schilling. "The effects of alternate optimal solutions in constraint-based genome-scale metabolic models." In: *Metabolic engineering* 5.4 (2003), pp. 264–276. DOI: 10.1016/j.ymben.2003.09.002.

178   C. Gu, G. B. Kim, W. J. Kim, H. U. Kim, and S. Y. Lee. "Current status and applications of genome-scale metabolic models." In: *Genome biology* 20 (2019), pp. 1–18. DOI: 10.1186/s13059-019-1730-3.

179   A. Passi, J. D. Tibocha-Bonilla, M. Kumar, D. Tec-Campos, K. Zengler, and C. Zuniga. "Genome-scale metabolic modeling enables in-depth understanding of big data." In: *Metabolites* 12.1 (2021), p. 14. DOI: 10.3390/metabo12010014.

180   J. S. Edwards and B. Ø. Palsson. "The *Escherichia coli* MG1655 in silico metabolic genotype: its definition, characteristics, and capabilities." In: *Proceedings of the National Academy of Sciences* 97.10 (2000), pp. 5528–5533. DOI: 10.1073/pnas.97.10.5528.

181   C. G. de Oliveira Dal'Molin, L.-E. Quek, R. W. Palfreyman, S. M. Brumbley, and L. K. Nielsen. "AraGEM, a genome-scale reconstruction of the primary metabolic network in Arabidopsis." In: *Plant physiology* 152.2 (2010), pp. 579–589. DOI: 10.1104/pp.109.148817.

182   N. C. Duarte, S. A. Becker, N. Jamshidi, I. Thiele, M. L. Mo, T. D. Vo, R. Srivas, and B. Ø. Palsson. "Global reconstruction of the human metabolic network based on genomic and bibliomic data." In: *Proceedings of the National Academy of Sciences* 104.6 (2007), pp. 1777–1782. DOI: 10.1073/pnas.0610772104.

183   S. Magnúsdóttir, A. Heinken, L. Kutt, D. A. Ravcheev, E. Bauer, A. Noronha, K. Greenhalgh, C. Jäger, J. Baginska, P. Wilmes, et al. "Generation of genome-scale metabolic reconstructions for 773 members of the human gut microbiota." In: *Nature biotechnology* 35.1 (2017), pp. 81–89. DOI: 10.1038/nbt.3703.

184   A. Heinken, J. Hertel, G. Acharya, D. A. Ravcheev, M. Nyga, O. E. Okpala, M. Hogan, S. Magnúsdóttir, F. Martinelli, B. Nap, et al. "Genome-scale metabolic reconstruction of 7,302 human microorganisms for personalized medicine." In: *Nature Biotechnology* (2023), pp. 1–12. DOI: 10.1038/s41587-022-01628-0.

185   A. V. Colarusso, I. Goodchild-Michelman, M. Rayle, and A. R. Zomorrodi. "Computational modeling of metabolism in microbial communities on a genome-scale." In: *Current Opinion in Systems Biology* 26 (2021), pp. 46–57. DOI: 10.1016/j.coisb.2021.04.001.

186   R. Mostolizadeh, M. Glöckler, and A. Dräger. "Towards the human nasal microbiome: simulating *D. pigrum* and *S. aureus*." In: *Frontiers in Cellular and Infection Microbiology* 12 (2022), p. 925215. DOI: 10.3389/fcimb.2022.925215.

187   I. Thiele, N. Swainston, R. M. Fleming, A. Hoppe, S. Sahoo, M. K. Aurich, H. Haraldsdottir, M. L. Mo, O. Rolfsson, M. D. Stobbe, et al. "A community-driven global reconstruction of human metabolism." In: *Nature biotechnology* 31.5 (2013), pp. 419–425. DOI: 10.1038/nbt.2488.

188   N. Swainston, K. Smallbone, H. Hefzi, P. D. Dobson, J. Brewer, M. Hanscho, D. C. Zielinski, K. S. Ang, N. J. Gardiner, J. M. Gutierrez, et al. "Recon 2.2: from reconstruction to model of human metabolism." In: *Metabolomics* 12 (2016), pp. 1–7. DOI: 10.1007/s11306-016-1051-4.

189   E. Brunk, S. Sahoo, D. C. Zielinski, A. Altunkaya, A. Dräger, N. Mih, F. Gatto, A. Nilsson, G. A. Preciat Gonzalez, M. K. Aurich, et al. "Recon3D enables a three-dimensional view of gene variation in human metabolism." In: *Nature biotechnology* 36.3 (2018), pp. 272–281. DOI: 10.1038/nbt.4072.

190   A. Mardinoglu, R. Agren, C. Kampf, A. Asplund, I. Nookaew, P. Jacobson, A. J. Walley, P. Froguel, L. M. Carlsson, M. Uhlen, et al. "Integration of clinical data with a genome-scale metabolic model of the human adipocyte." In: *Molecular systems biology* 9.1 (2013), p. 649. DOI: 10.1038/msb.2013.5.

191   A. Mardinoglu, R. Agren, C. Kampf, A. Asplund, M. Uhlen, and J. Nielsen. "Genome-scale metabolic modelling of hepatocytes reveals serine deficiency in patients with non-alcoholic fatty liver disease." In: *Nature communications* 5.1 (2014), p. 3083. DOI: 10.1038/ncomms4083.

192   J. L. Robinson, P. Kocabaş, H. Wang, P.-E. Cholley, D. Cook, A. Nilsson, M. Anton, R. Ferreira, I. Domenzain, V. Billa, et al. "An atlas of human metabolism." In: *Science signaling* 13.624 (2020), eaaz1482. DOI: 10.1126/scisignal.aaz1482.

193   S. Robaina Estévez and Z. Nikoloski. "Generalized framework for context-specific metabolic model extraction methods." In: *Frontiers in plant science* 5 (2014), p. 491. DOI: 10.3389/fpls.2014.00491.

194   S. Opdam, A. Richelle, B. Kellman, S. Li, D. C. Zielinski, and N. E. Lewis. "A systematic evaluation of methods for tailoring genome-scale metabolic models." In: *Cell systems* 4.3 (2017), pp. 318–329. DOI: 10.1016/j.cels.2017.01.010.

195   L. Väremo, C. Scheele, C. Broholm, A. Mardinoglu, C. Kampf, A. Asplund, I. Nookaew, M. Uhlen, B. K. Pedersen, and J. Nielsen. "Proteome-and transcriptome-driven reconstruction of the human myocyte metabolic network and its use for identification of markers for diabetes." In: *Cell reports* 11.6 (2015), pp. 921–933. DOI: 10.1016/j.celrep.2015.04.010.

196   I. Thiele, S. Sahoo, A. Heinken, J. Hertel, L. Heirendt, M. K. Aurich, and R. M. Fleming. "Personalized whole-body models integrate metabolism, physiology, and the gut microbiome." In: *Molecular systems biology* 16.5 (2020), e8982. DOI: 10.15252/msb.20198982.

197   A. Basile, A. Heinken, J. Hertel, L. Smarr, W. Li, L. Treu, G. Valle, S. Campanaro, and I. Thiele. "Longitudinal flux balance analyses of a patient with episodic colonic inflammation reveals microbiome metabolic dynamics." In: *Gut microbes* 15.1 (2023), p. 2226921. DOI: 10.1080/19490976.2023.2226921.

198   I. Thiele and R. M. Fleming. "Whole-body metabolic modelling predicts isoleucine dependency of SARS-CoV-2 replication." In: *Computational and Structural Biotechnology Journal* 20 (2022), pp. 4098–4109. DOI: 10.1016/j.csbj.2022.07.019.

199   C. J. Norsigian, N. Pusarla, J. L. McConn, J. T. Yurkovich, A. Dräger, B. Ø. Palsson, and Z. King. "BiGG Models 2020: multi-strain genome-scale models and expansion across the phylogenetic tree." In: *Nucleic acids research* 48.D1 (2020), pp. D402–D406. DOI: 10.1093/nar/gkz1054.

200   A. Noronha, J. Modamio, Y. Jarosz, E. Guerard, N. Sompairac, G. Preciat, A. D. Daníelsdóttir, M. Krecke, D. Merten, H. S. Haraldsdóttir, et al. "The Virtual Metabolic Human database: integrating human and gut microbiome metabolism with nutrition and disease." In: *Nucleic acids research* 47.D1 (2019), pp. D614–D624. DOI: 10.1093/nar/gky992.

201   R. S. Malik-Sheriff, M. Glont, T. V. N. Nguyen, K. Tiwari, M. G. Roberts, A. Xavier, M. T. Vu, J. Men, M. Maire, S. Kananathan, E. L. Fairbanks, J. P. Meyer, C. Arankalle, T. M. Varusai, V. Knight-Schrijver, L. Li, C. Dueñas-Roca, G. Dass, S. M. Keating, Y. M. Park, N. Buso, N. Rodriguez, M. Hucka, and H. Hermjakob. "BioModels—15 years of sharing computational models in life science." In: *Nucleic Acids Research* 48.D1 (Nov. 2019), pp. D407–D415. ISSN: 0305-1048. DOI: 10.1093/nar/gkz1055.

202   C. S. Henry, M. DeJongh, A. A. Best, P. M. Frybarger, B. Linsay, and R. L. Stevens. "High-throughput generation, optimization and analysis of genome-scale metabolic models." In: *Nature biotechnology* 28.9 (2010), pp. 977–982. DOI: 10.1038/nbt.1672.

203   S. M. Keating, D. Waltemath, M. König, F. Zhang, A. Dräger, C. Chaouiya, F. T. Bergmann, A. Finney, C. S. Gillespie, T. Helikar, et al. "SBML Level 3: an extensible format for the exchange and reuse of biological models." In: *Molecular Systems Biology* 16.8 (Aug. 2020), e9110. ISSN: 1744-4292. DOI: 10.15252/msb.20199110.

204   M. Hucka, F. T. Bergmann, C. Chaouiya, A. Dräger, S. Hoops, S. M. Keating, M. König, N. Le Novère, C. J. Myers, B. G. Olivier, S. Sahle, J. C. Schaff, R. Sheriff, L. P. Smith, D. Waltemath, D. J. Wilkinson, and F. Zhang. "Systems Biology Markup Language (SBML): Language Specification for Level 3 Version 2 Core Release 2." In: *Journal of Integrative Bioinformatics* 16.2 (May 2019), p. 1. ISSN: 1613-4516. DOI: 10.1515/jib-2019-0021.

205   B. G. Olivier and F. T. Bergmann. "SBML Level 3 Package: Flux Balance Constraints version 2." In: *Journal of Integrative Bioinformatics* 15.1 (Mar. 2018), p. 20170082. DOI: 10.1515/jib-2017-0082.

206   B. J. Bornstein, S. M. Keating, A. Jouraku, and M. Hucka. "LibSBML: an API library for SBML." In: *Bioinformatics* 24.6 (2008), pp. 880–881. DOI: 10.1093/bioinformatics/btn051.

207   N. Rodriguez, A. Thomas, L. Watanabe, I. Y. Vazirabad, V. Kofia, H. F. Gómez, F. Mittag, J. Matthes, J. Rudolph, F. Wrzodek, E. Netz, A. Diamantikos, J. Eichner, R. Keller, C. Wrzodek, S. Fröhlich, N. E. Lewis, C. J. Myers, N. Le Novère, B. Ø. Palsson, M. Hucka, and A. Dräger. "JSBML 1.0: providing a smorgasbord of options to encode systems biology models." In: *Bioinformatics* 31.20 (June 2015), pp. 3383–3386. ISSN: 1367-4803. DOI: 10.1093/bioinformatics/btv341.

208   N. Rodriguez, J.-B. Pettit, P. Dalle Pezze, L. Li, A. Henry, M. P. van Iersel, G. Jalowicki, M. Kutmon, K. N. Natarajan, D. Tolnay, et al. "The systems biology format converter." In: *BMC bioinformatics* 17.1 (2016), pp. 1–7. DOI: 10.1186/s12859-016-1000-2.

209   S. M. Keating, B. J. Bornstein, A. Finney, and M. Hucka. "SBMLToolbox: an SBML toolbox for MATLAB users." In: *Bioinformatics* 22.10 (Mar. 2006), pp. 1275–1277. ISSN: 1367-4803. DOI: 10.1093/bioinformatics/btl111.

210 C. Lieven, M. E. Beber, B. G. Olivier, F. T. Bergmann, M. Ataman, P. Babaei, J. A. Bartell, L. M. Blank, S. Chauhan, K. Correia, et al. "MEMOTE for standardized genome-scale metabolic model testing." In: *Nature biotechnology* 38.3 (2020), pp. 272–276. DOI: 10.1038/s41587-020-0446-y.

211 N. Le Novère, A. Finney, M. Hucka, U. S. Bhalla, F. Campagne, J. Collado-Vides, E. J. Crampin, M. Halstead, E. Klipp, P. Mendes, et al. "Minimum information requested in the annotation of biochemical models (MIRIAM)." In: *Nature biotechnology* 23.12 (2005), pp. 1509–1515. DOI: 10.1038/nbt1156.

212 R. Stevens, C. A. Goble, and S. Bechhofer. "Ontology-based knowledge representation for bioinformatics." In: *Briefings in Bioinformatics* 1.4 (Nov. 2000), pp. 398–414. ISSN: 1467-5463. DOI: 10.1093/bib/1.4.398.

213 J. K. Medley, A. P. Goldberg, and J. R. Karr. "Guidelines for Reproducibly Building and Simulating Systems Biology Models." In: *IEEE Transactions on Biomedical Engineering* 63.10 (2016), pp. 2015–2020. DOI: 10.1109/TBME.2016.2591960.

214 M. Courtot, N. Juty, C. Knüpfer, D. Waltemath, A. Zhukova, A. Dräger, M. Dumontier, 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 Novère. "Controlled vocabularies and semantics in systems biology." In: *Molecular Systems Biology* 7.1 (2011), p. 543. DOI: https://doi.org/10.1038/msb.2011.77.

215 M. Giglio, R. Tauber, S. Nadendla, J. Munro, D. Olley, S. Ball, E. Mitraka, L. M. Schriml, P. Gaudet, E. T. Hobbs, I. Erill, D. A. Siegele, J. C. Hu, C. Mungall, and M. C. Chibucos. "ECO, the Evidence amp; Conclusion Ontology: community standard for evidence information." In: *Nucleic Acids Research* 47.D1 (Nov. 2018), pp. D1186–D1194. ISSN: 0305-1048. DOI: 10.1093/nar/gky1036.

216 N. L. Novère, M. Hucka, H. Mi, S. Moodie, F. Schreiber, A. Sorokin, E. Demir, K. Wegner, M. I. Aladjem, S. M. Wimalaratne, et al. "The systems biology graphical notation." In: *Nature biotechnology* 27.8 (2009), pp. 735–741. DOI: https://doi.org/10.1038/nbt.1558.

217 F. T. Bergmann, T. Czauderna, U. Dogrusoz, A. Rougny, A. Dräger, V. Touré, A. Mazein, M. L. Blinov, and A. Luna. "Systems biology graphical notation markup language (SBGNML) version 0.3." In: *Journal of Integrative Bioinformatics* 17.2-3 (2020), p. 20200016. DOI: doi:10.1515/jib-2020-0016.

218 A. Funahashi, Y. Matsuoka, A. Jouraku, M. Morohashi, N. Kikuchi, and H. Kitano. "CellDesigner 3.5: A Versatile Modeling Tool for Biochemical Networks." In: *Proceedings of the IEEE* 96.8 (2008), pp. 1254–1265. DOI: 10.1109/JPROC.2008.925458.

219 I. Balaur, L. Roy, A. Mazein, S. G. Karaca, U. Dogrusoz, E. Barillot, and A. Zinovyev. "cd2sbgnml: bidirectional conversion between CellDesigner and SBGN formats." In: *Bioinformatics* 36.8 (Jan. 2020), pp. 2620–2622. ISSN: 1367-4803. DOI: 10.1093/bioinformatics/btz969.

220 H. Balcı, M. C. Siper, N. Saleh, I. Safarli, L. Roy, M. Kılıçarslan, R. Özaydın, A. Mazein, C. Auffray, Ö. Babur, et al. "Newt: a comprehensive web-based tool for viewing, constructing and analyzing biological maps." In: *Bioinformatics* 37.10 (2021), pp. 1475–1477. DOI: 10.1093/bioinformatics/btaa850.

221 Y. Seif, J. M. Monk, N. Mih, H. Tsunemoto, S. Poudel, C. Zuniga, J. Broddrick, K. Zengler, and B. Ø. Palsson. "A computational knowledge-base elucidates the response of *Staphylococcus aureus* to different media types." In: *PLoS computational biology* 15.1 (2019), e1006644. DOI: 10.1371/journal.pcbi.1006644.

222 C. S. Jensen, C. J. Norsigian, X. Fang, X. C. Nielsen, J. J. Christensen, B. Ø. Palsson, and J. M. Monk. "Reconstruction and validation of a genome-scale metabolic model of *Streptococcus oralis* (iCJ415), a human commensal and opportunistic pathogen." In: *Frontiers in Genetics* 11 (2020), p. 116. DOI: 10.3389/fgene.2020.00116.

223 J. Nielsen. "Systems biology of metabolism: a driver for developing personalized and precision medicine." In: *Cell metabolism* 25.3 (2017), pp. 572–579. DOI: 10.1016/j.cmet.2017.02.002.

224 N. Klitgord and D. Segrè. "Environments that Induce Synthetic Microbial Ecosystems." In: *PLoS Computational Biology* 6 (11 Nov. 2010), p. 1001002. ISSN: 1553734X. DOI: 10.1371/JOURNAL.PCBI.1001002.

225 C. S. Henry, H. C. Bernstein, P. Weisenhorn, R. C. Taylor, J. Y. Lee, J. Zucker, and H. S. Song. "Microbial Community Metabolic Modeling: A Community Data-Driven Network Reconstruction." In: *Journal of Cellular Physiology* 231 (11 Nov. 2016), pp. 2339–2345. ISSN: 10974652. DOI: 10.1002/JCP.25428.

226 C. Zuñiga, L. Zaramela, and K. Zengler. "Elucidation of complexity and prediction of interactions in microbial communities." In: *Microbial Biotechnology* 10 (6 Nov. 2017), pp. 1500–1522. ISSN: 17517915. DOI: 10.1111/1751-7915.12855.

227　A. V. Colarusso, I. Goodchild-Michelman, M. Rayle, and A. R. Zomorrodi. "Computational modeling of metabolism in microbial communities on a genome-scale." In: *Current Opinion in Systems Biology* 26 (June 2021), pp. 46–57. ISSN: 24523100. DOI: 10.1016/J.COISB.2021.04.001.

228　M. Glöckler, A. Dräger, and R. Mostolizadeh. "NCMW: A Python Package to Analyze Metabolic Interactions in the Nasal Microbiome." In: *Frontiers in Bioinformatics* 0 (Feb. 2022), p. 11. ISSN: 2673-7647. DOI: 10.3389/FBINF.2022.827024.

229　N. E. Lewis and A. M. Abdel-Haleem. "The evolution of genome-scale models of cancer metabolism." In: *Frontiers in physiology* 4 (2013), p. 237. DOI: 10.3389/fphys.2013.00237.

230　T. Shlomi, M. N. Cabili, and E. Ruppin. "Predicting metabolic biomarkers of human inborn errors of metabolism." In: *Molecular systems biology* 5.1 (2009), p. 263. DOI: 10.1038/msb.2009.22.

231　A. Bordbar, N. E. Lewis, J. Schellenberger, B. Ø. Palsson, and N. Jamshidi. "Insight into human alveolar macrophage and *M. tuberculosis* interactions via metabolic reconstructions." In: *Molecular systems biology* 6.1 (2010), p. 422. DOI: 10.1038/msb.2010.68.

232　A. Heinken and I. Thiele. "Systematic prediction of health-relevant human-microbial co-metabolism through a computational framework." In: *Gut microbes* 6.2 (2015), pp. 120–130. DOI: 10.1080/19490976.2015.1023494.

233　N. Noirungsee, S. Changkhong, K. Phinyo, C. Suwannajak, N. Tanakul, and S. Inwongwan. "Genome-scale metabolic modelling of extremophiles and its applications in astrobiological environments." In: *Environmental Microbiology Reports* 16.1 (2024), e13231. DOI: 10.1111/1758-2229.13231.

234　S. Qiu, A. Yang, and H. Zeng. "Flux balance analysis-based metabolic modeling of microbial secondary metabolism: Current status and outlook." In: *PLoS Computational Biology* 19.8 (2023), e1011391. DOI: 10.1371/journal.pcbi.1011391.

235　Y. Wang, J. A. Eddy, and N. D. Price. "Reconstruction of genome-scale metabolic models for 126 human tissues using mCADRE." In: *BMC systems biology* 6.1 (2012), pp. 1–16. DOI: 10.1186/1752-0509-6-153.

236　N. Vlassis, M. P. Pacheco, and T. Sauter. "Fast reconstruction of compact context-specific metabolic network models." In: *PLoS computational biology* 10.1 (2014), e1003424. DOI: 10.1371/journal.pcbi.1003424.

237　H. Panchiwala, S. Shah, H. Planatscher, M. Zakharchuk, M. König, and A. Dräger. "The Systems Biology Simulation Core Library." In: *Bioinformatics* 38 (3 Feb. 2022), pp. 864–865. ISSN: 1367-4803. DOI: 10.1093/bioinformatics/btab669.

238　A. Renz, L. Widerspick, and A. Dräger. "FBA reveals guanylate kinase as a potential target for antiviral therapies against SARS-CoV-2." In: *Bioinformatics* 36.Supplement_2 (2020), pp. i813–i821. DOI: 10.1093/bioinformatics/btaa813.

239　B. P. Bannerman, J. Júlvez, A. Oarga, T. L. Blundell, P. Moreno, and R. A. Floto. "Integrated human/SARS-CoV-2 metabolic models present novel treatment strategies against COVID-19." In: *Life science alliance* 4.10 (2021). DOI: 10.26508/lsa.202000954.

240　L. Heirendt, S. Arreckx, T. Pfau, S. N. Mendoza, A. Richelle, A. Heinken, H. S. Haraldsdóttir, J. Wachowiak, S. M. Keating, V. Vlasov, et al. "Creation and analysis of biochemical constraint-based models using the COBRA Toolbox v. 3.0." In: *Nature protocols* 14.3 (2019), pp. 639–702. DOI: 10.1038/s41596-018-0098-2.

241　I. I. Cplex. "V12. 1: User's Manual for CPLEX." In: *International Business Machines Corporation* 46.53 (2009), p. 157.

242　A. Makhorin. "GLPK (GNU linear programming kit)." In: *http://www. gnu. org/s/glpk/glpk. html* (2008).

243　A. Renz, L. Widerspick, and A. Dräger. "Genome-Scale Metabolic Model of Infection with SARS-CoV-2 Mutants Confirms Guanylate Kinase as Robust Potential Antiviral Target." In: *Genes* 12.6 (2021), p. 796. DOI: 10.3390/genes12060796.

244　P. Nanda and A. Ghosh. "Genome Scale-Differential Flux Analysis reveals deregulation of lung cell metabolism on SARS-CoV-2 infection." In: *PLoS computational biology* 17.4 (2021), e1008860. DOI: 10.1371/journal.pcbi.1008860.

245　J. P. Bernardes, N. Mishra, F. Tran, T. Bahmer, L. Best, J. I. Blase, D. Bordoni, J. Franzenburg, U. Geisen, J. Josephs-Spaulding, et al. "Longitudinal multi-omics analyses identify responses of megakaryocytes, erythroid cells, and plasmablasts as hallmarks of severe COVID-19." In: *Immunity* 53.6 (2020), pp. 1296–1314. DOI: 10.1016/j.immuni.2020.11.017.

246　H. Delattre, K. Sasidharan, and O. S. Soyer. "Inhibiting the reproduction of SARS-CoV-2 through perturbations in human lung cell metabolic network." In: *Life science alliance* 4.1 (2021). DOI: 10.26508/lsa.202000869.

247   D. R. Evans and H. I. Guy. "Mammalian pyrimidine biosynthesis: fresh insights into an ancient pathway." In: *Journal of Biological Chemistry* 279.32 (2004), pp. 33035–33038. DOI: 10.1074/jbc.R400007200.

248   World Health Organization. *Weekly epidemiological update on COVID-19 - 25 January 2022*. Accessed on February 8, 2022. URL: https://www.who.int/publications/m/item/weekly-epidemiological-update-on-covid-19---25-january-2022.

249   S. Khare, C. Gurry, L. Freitas, M. B. Schultz, G. Bach, A. Diallo, N. Akite, J. Ho, R. T. Lee, W. Yeo, et al. "GISAID's Role in Pandemic Response." In: *China CDC Weekly* 3.49 (2021), p. 1049. DOI: 10.46234/ccdcw2021.255.

250   S. Kumar, T. S. Thambiraja, K. Karuppanan, and G. Subramaniam. "Omicron and Delta variant of SARS-CoV-2: A comparative computational study of spike protein." In: *Journal of medical virology* 94.4 (2022), pp. 1641–1649. DOI: 10.1002/jmv.27526.

251   E. D. Clercq. "Antiviral activity spectrum and target of action of different classes of nucleoside analogues." In: *Nucleosides, Nucleotides & Nucleic Acids* 13.6-7 (1994), pp. 1271–1295. DOI: 10.1080/15257779408012151.

252   G. B. Elion. "Mechanism of action and selectivity of acyclovir." In: *The American journal of medicine* 73.1 (1982), pp. 7–13. DOI: 10.1016/0002-9343(82)90055-9.

253   E. De Clercq, J. Murase, and V. E. Marquez. "Broad-spectrum antiviral and cytocidal activity of cyclopentenyl-cytosine, a carbocyclic nucleoside targeted at CTP synthetase." In: *Biochemical pharmacology* 41.12 (1991), pp. 1821–1829. DOI: 10.1016/0006-2952(91)90120-t.

254   V. E. Marquez, M. I. Lim, S. P. Treanor, J. Plowman, M. A. Priest, A. Markovac, M. S. Khan, B. Kaskar, and J. S. Driscoll. "Cyclopentenylcytosine. A carbocyclic nucleoside with antitumor and antiviral properties." In: *Journal of medicinal chemistry* 31.9 (1988), pp. 1687–1694. DOI: 10.1021/jm00117a004.

255   K. J. M. Schimmel, H. Gelderblom, and H.-J. Guchelaar. "Cyclopentenyl cytosine (CPEC): an overview of its *in vitro* and *in vivo* activity." In: *Current cancer drug targets* 7.5 (2007), pp. 504–509. DOI: 10.2174/156800907781386579.

256   P. M. Politi, F. Xie, W. Dahut, H. Ford, J. A. Kelley, A. Bastian, A. Setser, C. J. Allegra, A. P. Chen, J. M. Hamilton, et al. "Phase I clinical trial of continuous infusion cyclopentenyl cytosine." In: *Cancer chemotherapy and pharmacology* 36.6 (1995), pp. 513–523. DOI: 10.1007/BF00685802.

257   K. Schimmel, R. Bennink, K. de Bruin, R. Leen, K. Sand, M. van den Hoff, A. van Kuilenburg, J.-L. Vanderheyden, N. Steinmetz, M. Pfaffendorf, et al. "Absence of cardiotoxicity of the experimental cytotoxic drug cyclopentenyl cytosine (CPEC) in rats." In: *Archives of toxicology* 79.5 (2005), pp. 268–276. DOI: 10.1007/s00204-004-0633-5.

258   S. M. Blaney, F. M. Balis, J. Grem, D. E. Cole, P. C. Adamson, and D. G. Poplack. "Modulation of the cytotoxic effect of cyclopentenylcytosine by its primary metabolite, cyclopentenyluridine." In: *Cancer research* 52.12 (1992), pp. 3503–3505.

259   H. Ford Jr, D. A. Cooney, G. S. Ahluwalia, Z. Hao, M. E. Rommel, L. Hicks, K. A. Dobyns, J. E. Tomaszewski, and D. G. Johns. "Cellular pharmacology of cyclopentenyl cytosine in Molt-4 lymphoblasts." In: *Cancer research* 51.14 (1991), pp. 3733–3740.

260   J. P. Engel, J. A. Englund, C. V. Fletcher, and E. L. Hill. "Treatment of resistant herpes simplex virus with continuous-infusion acyclovir." In: *Jama* 263.12 (1990), pp. 1662–1664.

261   H. H. Balfour Jr, K. A. McMonigal, and B. Bean. "Acyclovir therapy of varicella-zoster virus infections in immunocompromised patients." In: *Journal of Antimicrobial Chemotherapy* 12.suppl_B (1983), pp. 169–179. DOI: 10.1093/jac/12.suppl_b.169.

262   E. L. C. Tan, E. E. Ooi, C.-Y. Lin, H. C. Tan, A. E. Ling, B. Lim, and L. W. Stanton. "Inhibition of SARS coronavirus infection *in vitro* with clinically approved antiviral drugs." In: *Emerging infectious diseases* 10.4 (2004), p. 581. DOI: 10.3201/eid1004.030458.

263   A. Nofal, M. M. Fawzy, S. M. S. E. Deen, and E. E. El-Hawary. "Herpes zoster ophthalmicus in COVID-19 patients." In: *International Journal of Dermatology* 59.12 (2020), p. 1545. DOI: 10.1111/ijd.15240.

264   X. Cao, X. Du, H. Jiao, Q. An, R. Chen, P. Fang, J. Wang, and B. Yu. "Carbohydrate-based drugs launched during 2000- 2021." In: *Acta Pharmaceutica Sinica B* (2022), pp. 3783–3821. DOI: 10.1016/j.apsb.2022.05.020.

265   F. Kabinger, C. Stiller, J. Schmitzová, C. Dienemann, G. Kokic, H. S. Hillen, C. Höbartner, and P. Cramer. "Mechanism of molnupiravir-induced SARS-CoV-2 mutagenesis." In: *Nature structural & molecular biology* 28.9 (2021), pp. 740–746. DOI: 10.1038/s41594-021-00651-0.

266   G. Kokic, H. S. Hillen, D. Tegunov, C. Dienemann, F. Seitz, J. Schmitzova, L. Farnung, A. Siewert, C. Höbartner, and P. Cramer. "Mechanism of SARS-CoV-2 polymerase stalling by remdesivir." In: *Nature communications* 12.1 (2021), p. 279. DOI: 10.1038/s41467-020-20542-0.

267   C. Marzolini, D. R. Kuritzkes, F. Marra, A. Boyle, S. Gibbons, C. Flexner, A. Pozniak, M. Boffito, L. Waters, D. Burger, et al. "Recommendations for the management of drug–drug interactions between the COVID-19 antiviral nirmatrelvir/ritonavir (Paxlovid) and comedications." In: *Clinical Pharmacology & Therapeutics* 112.6 (2022), pp. 1191–1200. DOI: 10.1002/cpt.2646.

268   P. Cavazzoni. "Coronavirus (COVID-19) update: FDA limits use of certain monoclonal antibodies to treat COVID-19 due to the Omicron variant." In: *US Food and Drug Administration* (2022).

269   X. Tong, J. Smith, N. Bukreyeva, T. Koma, J. T. Manning, R. Kalkeri, A. D. Kwong, and S. Paessler. "Merimepodib, an IMPDH inhibitor, suppresses replication of Zika virus and other emerging viral pathogens." In: *Antiviral research* 149 (2018), pp. 34–40. DOI: 10.1016/j.antiviral.2017.11.004.

270   H. S. Te, G. Randall, and D. M. Jensen. "Mechanism of action of ribavirin in the treatment of chronic hepatitis C." In: *Gastroenterology & hepatology* 3.3 (2007), p. 218.

271   F. T. Bergmann, R. Adams, S. Moodie, J. Cooper, M. Glont, M. Golebiewski, M. Hucka, C. Laibe, A. K. Miller, D. P. Nickerson, B. G. Olivier, N. Rodriguez, H. M. Sauro, M. Scharm, S. Soiland-Reyes, D. Waltemath, F. Yvon, and N. Le Novère. "COMBINE archive and OMEX format: one file to share all information to reproduce a modeling project." In: *BMC Bioinformatics* 15 (2014), p. 369. DOI: 10.1186/s12859-014-0369-z.

272   M. L. Neal, M. König, D. Nickerson, G. Mısırlı, R. Kalbasi, A. Dräger, K. Atalag, V. Chelliah, M. T. Cooling, D. L. Cook, S. Crook, M. de Alba, S. H. Friedman, A. Garny, J. H. Gennari, P. Gleeson, M. Golebiewski, M. Hucka, N. Juty, C. Myers, B. G. Olivier, H. M. Sauro, M. Scharm, J. L. Snoep, V. Touré, A. Wipat, O. Wolkenhauer, and D. Waltemath. "Harmonizing semantic annotations for computational models in biology." In: *Briefings in Bioinformatics* 20.2 (Nov. 2018), pp. 540–550. ISSN: 1477-4054. DOI: 10.1093/bib/bby087.

273   A. Ebrahim, E. Almaas, E. Bauer, A. Bordbar, A. P. Burgard, R. L. Chang, A. Dräger, I. Famili, A. M. Feist, R. M. T. Fleming, et al. "Do Genome-scale Models Need Exact Solvers or Clearer Standards?" In: *Molecular Systems Biology* 11.10 (Oct. 2015), p. 831. ISSN: 1744-4292. DOI: 10.15252/msb.20156157.

274   M. Römer, J. Eichner, A. Dräger, C. Wrzodek, F. Wrzodek, and A. Zell. "ZBIT bioinformatics toolbox: a web-platform for systems biology and expression data analysis." In: *PloS one* 11.2 (2016), e0149263. DOI: 10.1371/journal.pone.0149263.

275   E. Tacconelli, E. Carrara, A. Savoldi, S. Harbarth, M. Mendelson, D. L. Monnet, C. Pulcini, G. Kahlmeter, J. Kluytmans, Y. Carmeli, et al. "Discovery, research, and development of new antibiotics: the WHO priority list of antibiotic-resistant bacteria and tuberculosis." In: *The Lancet infectious diseases* 18.3 (2018), pp. 318–327. DOI: 10.1016/S1473-3099(17)30753-3.

276   M. A. Oberhardt, B. Ø. Palsson, and J. A. Papin. "Applications of genome-scale metabolic reconstructions." In: *Molecular systems biology* 5.1 (2009), p. 320. DOI: 10.1038/msb.2009.77.

277   H. U. Kim, T. Y. Kim, and S. Y. Lee. "Genome-scale metabolic network analysis and drug targeting of multidrug resistant pathogen *Acinetobacter baumannii* AYE." In: *Molecular BioSystems* 6.2 (2010), pp. 339–348. DOI: 10.1039/b916446d.

278   L. Presta, E. Bosi, L. Mansouri, L. Dijkshoorn, R. Fani, and M. Fondi. "Constraint-based modeling identifies new putative targets to fight colistin-resistant *A. baumannii* infections." In: *Scientific reports* 7.1 (2017), pp. 1–12. DOI: 10.1038/s41598-017-03416-2.

279   C. J. Norsigian, E. Kavvas, Y. Seif, B. Ø. Palsson, and J. M. Monk. "*i*CN718, an updated and improved genome-scale metabolic network reconstruction of *Acinetobacter baumannii* AYE." In: *Frontiers in genetics* 9 (2018), p. 121. DOI: 10.3389/fgene.2018.00121.

280   Y. Zhu, J. Zhao, M. H. M. Maifiah, T. Velkov, F. Schreiber, and J. Li. "Metabolic responses to polymyxin treatment in *Acinetobacter baumannii* ATCC 19606: integrating transcriptomics and metabolomics with genome-scale metabolic modeling." In: *Msystems* 4.1 (2019), e00157–18. DOI: 10.1128/mSystems.00157-18.

281   J. Zhao, Y. Zhu, J. Han, Y.-W. Lin, M. Aichem, J. Wang, K. Chen, T. Velkov, F. Schreiber, and J. Li. "Genome-scale metabolic modeling reveals metabolic alterations of multidrug-resistant *Acinetobacter Baumannii* in a murine bloodstream infection model." In: *Microorganisms* 8.11 (2020), p. 1793. DOI: 10.3390/microorganisms8111793.

282   J. S. C. Barbosa. "Genome-scale reconstruction of the metabolic network *i*JS784 for *Acinetobacter baumannii* strain ATCC 17978 to address drug target prioritization." Dissertation. National Laboratory for Scientific Computing, Brazil, 2020.

283 M. A. Carey, A. Dräger, M. E. Beber, J. A. Papin, and J. T. Yurkovich. "Community standards to facilitate development and address challenges in metabolic modeling." In: *Molecular Systems Biology* 16.8 (2020), e9235. DOI: 10.15252/msb.20199235.

284 D. Machado, S. Andrejev, M. Tramontano, and K. R. Patil. "Fast automated reconstruction of genome-scale metabolic models for microbial species and communities." In: *Nucleic acids research* 46.15 (2018), pp. 7542–7553. DOI: 10.1093/nar/gky537.

285 C. J. Fritzemeier, D. Hartleb, B. Szappanos, B. Papp, and M. J. Lercher. "Erroneous energy-generating cycles in published genome scale metabolic networks: Identification and removal." In: *PLoS computational biology* 13.4 (2017), e1005494. DOI: 10.1371/journal.pcbi.1005494.

286 N. Leonidou, E. Fritze, A. Renz, and A. Dräger. "SBOannotator: a Python Tool for the Automated Assignment of Systems Biology Ontology Terms." In: *Bioinformatics* (July 2023), btad437. ISSN: 1367-4811. DOI: 10.1093/bioinformatics/btad437.

287 A. Howard, M. O'Donoghue, A. Feeney, and R. D. Sleator. "*Acinetobacter baumannii*: an emerging opportunistic pathogen." In: *Virulence* 3.3 (2012), pp. 243–250. DOI: 10.4161/viru.19700.

288 Q. Damaceno, J. R. Nicoli, and A. Oliveira. "Variability of cutaneous and nasal population levels between patients colonized and infected by multidrug-resistant bacteria in two Brazilian intensive care units." In: *SAGE Open Medicine* 3 (2015), p. 2050312114566668. DOI: 10.1177/2050312114566668.

289 M.-L. Liou, K.-H. Chen, H.-L. Yeh, C.-Y. Lai, and C.-H. Chen. "Persistent nasal carriers of *Acinetobacter baumannii* in long-term-care facilities." In: *American journal of infection control* 45.7 (2017), pp. 723–727. DOI: 10.1016/j.ajic.2017.02.005.

290 C.-H. Chen, M.-L. Liou, C.-Y. Lee, M.-C. Chang, H.-Y. Kuo, and T.-H. Chang. "Diversity of nasal microbiota and its interaction with surface microbiota among residents in healthcare institutes." In: *Scientific reports* 9.1 (2019), pp. 1–10. DOI: 10.1038/s41598-019-42548-5.

291 S. Mishra and J. Imlay. "Why do bacteria use so many enzymes to scavenge hydrogen peroxide?" In: *Archives of biochemistry and biophysics* 525.2 (2012), pp. 145–160. DOI: 10.1016/j.abb.2012.04.014.

292 B. L. Mortensen and E. P. Skaar. "The contribution of nutrient metal acquisition and metabolism to *Acinetobacter baumannii* survival within the host." In: *Frontiers in cellular and infection microbiology* 3 (2013), p. 95. DOI: 10.3389/fcimb.2013.00095.

293 M. I. Hood, B. L. Mortensen, J. L. Moore, Y. Zhang, T. E. Kehl-Fie, N. Sugitani, W. J. Chazin, R. M. Caprioli, and E. P. Skaar. "Identification of an *Acinetobacter baumannii* zinc acquisition system that facilitates resistance to calprotectin-mediated zinc sequestration." In: *PLoS pathogens* 8.12 (2012), e1003068. DOI: 10.1371/journal.ppat.1003068.

294 B. Krismer, M. Liebeke, D. Janek, M. Nega, M. Rautenberg, G. Hornig, C. Unger, C. Weidenmaier, M. Lalk, and A. Peschel. "Nutrient limitation governs *Staphylococcus aureus* metabolism and niche adaptation in the human nose." In: *PLoS pathogens* 10.1 (2014), e1003862. DOI: 10.1371/journal.ppat.1003862.

295 D. N. Farrugia, L. D. Elbourne, K. A. Hassan, B. A. Eijkelkamp, S. G. Tetu, M. H. Brown, B. S. Shah, A. Y. Peleg, B. C. Mabbutt, and I. T. Paulsen. "The complete genome and phenome of a community-acquired *Acinetobacter baumannii*." In: *PLoS one* 8.3 (2013), e58628. DOI: 0.1371/journal.pone.0058628.

296 N. Wang, E. A. Ozer, M. J. Mandel, and A. R. Hauser. "Genome-wide identification of *Acinetobacter baumannii* genes necessary for persistence in the lung." In: *MBio* 5.3 (2014), e01163–14. DOI: 10.1128/mBio.01163-14.

297 D. Barh, S. Tiwari, N. Jain, A. Ali, A. R. Santos, A. N. Misra, V. Azevedo, and A. Kumar. "In silico subtractive genomics for target identification in human bacterial pathogens." In: *Drug Development Research* 72.2 (2011), pp. 162–177. DOI: 10.1002/ddr.20413.

298 S. F. Altschul, W. Gish, W. Miller, E. W. Myers, and D. J. Lipman. "Basic local alignment search tool." In: *Journal of molecular biology* 215.3 (1990), pp. 403–410. DOI: Basiclocalalignmentsearchtool.

299 K. M. Herrmann and L. M. Weaver. "The shikimate pathway." In: *Annual review of plant biology* 50.1 (1999), pp. 473–503. DOI: 10.1146/annurev.arplant.50.1.473.

300 T. Parish and N. G. Stoker. "The common aromatic amino acid biosynthesis pathway is essential in *Mycobacterium tuberculosis*." In: *Microbiology* 148.10 (2002), pp. 3069–3077. DOI: 10.1099/00221287-148-10-3069.

301 S. Tapas, A. Kumar, S. Dhindwal, P. Kumar, et al. "Structural analysis of chorismate synthase from *Plasmodium falciparum*: a novel target for antimalaria drug discovery." In: *International journal of biological macromolecules* 49.4 (2011), pp. 767–777. DOI: 10.1016/j.ijbiomac.2011.07.011.

302 F. Bowe, P. O'Gaora, D. Maskell, M. Cafferkey, and G. Dougan. "Virulence, persistence, and immunogenicity of *Yersinia enterocolitica* O:8 aroA mutants." In: *Infection and immunity* 57.10 (1989), pp. 3234–3236. DOI: 10.1128/iai.57.10.3234-3236.1989.

303 T. C. Umland, L. W. Schultz, U. MacDonald, J. M. Beanan, R. Olson, and T. A. Russo. "*In vivo*-validated essential genes identified in *Acinetobacter baumannii* by using human ascites overlap poorly with essential genes detected on laboratory media." In: *MBio* 3.4 (2012), e00113–12. DOI: 10.1128/mBio.00113-12.

304 N. Farah, V. K. Chin, P. P. Chong, W. F. Lim, C. W. Lim, R. Basir, S. K. Chang, and T. Y. Lee. "Riboflavin as a promising antimicrobial agent? A multi-perspective review." In: *Current Research in Microbial Sciences* (2022), p. 100111. DOI: 10.1016/j.crmicr.2022.100111.

305 B. E. Britigan, Y. Chai, and M. Cohen. "Effects of human serum on the growth and metabolism of *Neisseria gonorrhoeae*: an alternative view of serum." In: *Infection and immunity* 50.3 (1985), pp. 738–744. DOI: 10.1128/iai.50.3.738-744.1985.

306 R. Hommes, P. Postma, D. Tempest, and O. Neijssel. "The influence of the culture pH value on the direct glucose oxidative pathway in *Klebsiella pneumoniae* NCTC 418." In: *Archives of microbiology* 151.3 (1989), pp. 261–267. DOI: 10.1007/BF00413140.

307 T. Lessie and P. Phibbs Jr. "Alternative pathways of carbohydrate utilization in pseudomonads." In: *Annual review of microbiology* 38.1 (1984), pp. 359–388. DOI: 10.1146/annurev.mi.38.100184.002043.

308 D. S. Wishart, Y. D. Feunang, A. C. Guo, E. J. Lo, A. Marcu, J. R. Grant, T. Sajed, D. Johnson, C. Li, Z. Sayeeda, et al. "DrugBank 5.0: a major update to the DrugBank database for 2018." In: *Nucleic acids research* 46.D1 (2018), pp. D1074–D1082. DOI: 10.1093/nar/gkx1037.

309 J. J. Davis, A. R. Wattam, R. K. Aziz, T. Brettin, R. Butler, R. M. Butler, P. Chlenski, N. Conrad, A. Dickerman, E. M. Dietrich, et al. "The PATRIC Bioinformatics Resource Center: expanding data and analysis capabilities." In: *Nucleic acids research* 48.D1 (2020), pp. D606–D612. DOI: 10.1093/nar/gkz943.

310 N. Juty, N. Le Novere, and C. Laibe. "Identifiers. org and MIRIAM Registry: community resources to provide persistent identification." In: *Nucleic acids research* 40.D1 (2012), pp. D580–D586. DOI: 10.1093/nar/gkr1097.

311 M. Hucka, F. T. Bergmann, A. Dräger, S. Hoops, S. M. Keating, N. Le Novère, C. J. Myers, B. G. Olivier, S. Sahle, J. C. Schaff, L. P. Smith, D. Waltemath, and D. J. Wilkinson. "Systems Biology Markup Language (SBML) Level 3 Version 1 Core." In: *Journal of Integrative Bioinformatics* 15.1 (Apr. 2018), p. 1. DOI: 10.1515/jib-2017-0080.

312 Y. W. Lim, R. Schmieder, M. Haynes, D. Willner, M. Furlan, M. Youle, K. Abbott, R. Edwards, J. Evangelista, D. Conrad, et al. "Metagenomics and metatranscriptomics: windows on CF-associated viral and microbial communities." In: *Journal of Cystic Fibrosis* 12.2 (2013), pp. 154–164. DOI: 10.1016/j.jcf.2012.07.009.

313 Y. Shimoyama. *ANIclustermap: A tool for drawing ANI clustermap between all-vs-all microbial genomes.* 2022. URL: https://github.com/moshi4/ANIclustermap.

314 X. Zhao and K. Drlica. "Reactive oxygen species and the bacterial response to lethal stress." In: *Current opinion in microbiology* 21 (2014), pp. 1–6. DOI: 10.1016/j.mib.2014.06.008.

315 P. Korge, G. Calmettes, and J. N. Weiss. "Increased reactive oxygen species production during reductive stress: the roles of mitochondrial glutathione and thioredoxin reductases." In: *Biochimica et Biophysica Acta (BBA)-Bioenergetics* 1847.6-7 (2015), pp. 514–525. DOI: 10.1016/j.bbabio.2015.02.012.

316 M. I. Hood and E. P. Skaar. "Nutritional immunity: transition metals at the pathogen–host interface." In: *Nature Reviews Microbiology* 10.8 (2012), pp. 525–537. DOI: 10.1038/nrmicro2836.

317 K. H. Turner, A. K. Wessel, G. C. Palmer, J. L. Murray, and M. Whiteley. "Essential genome of *Pseudomonas aeruginosa* in cystic fibrosis sputum." In: *Proceedings of the National Academy of Sciences* 112.13 (2015), pp. 4110–4115. DOI: 10.1073/pnas.1419677112.

APPENDIX A

Contributions

All developed ideas, approaches, and research findings presented in this thesis were discussed with my supervisor Prof. Dr. Andreas Dräger (A. D.). Additionally, all visualizations created for the manuscripts and the thesis have been exclusively produced by myself. The following co-workers also contributed to the different projects (ordered alphabetically by their last name):

- Prof. Dr. Tom Coenye   (T. C.)
- Dr. Aurélie Crabbé   (A. C.)
- Lea Friedrich   (L. F.)
- Elisabeth Fritze   (E. F.)
- Dr. Reihaneh Mostolizadeh   (R. M.)
- Lisa Ostyn   (L. O.)
- Dr. rer. nat. Alina Renz   (A. R.)
- PD Dr. rer. nat. Monika S. Schütz   (M. S.)
- Yufan Xia   (Y. X.)

Paper I: New workflow predicts drug targets against SARS-CoV-2 via metabolic changes in infected cells
Conceptualization: N. L., A. R., R. M., and A. D.; data curation, formal analysis, investigation, methodology, software, validation, and visualization: N. L.; funding acquisition, supervision, and project administration: A. D.; manuscript writing: N. L.; manuscript review: N. L., A. R., R. M., and A. D. All authors approved the publishing of the manuscript.

Paper II: SBOannotator: a Python tool for the automated assignment of systems biology ontology terms
Conceptualization: N. L., E. F., A. R., and A. D.; data curation, formal analysis, investigation, and visualization: N. L.; methodology and software: N. L. and E. F.; validation: N. L.; funding acquisition, supervision, and project administration: A. D.; manuscript writing: N. L.; manuscript review: N. L., A. R., and A. D. All authors approved the publishing of the manuscript.

Paper III: Exploring the metabolic profiling of *A. baumannii* for antimicrobial development using genome-scale modeling
Conceptualization: N. L.; model reconstruction: N. L. and Y. X.; formal analysis, investigation, visualization, and software: N.L; curation and analysis of additional models: N. L.; experimental work: L. F.; funding acquisition, supervision, and resources: M. S. and A. D.; manuscript writing: N. L.; manuscript review: N. L., Y. X., L. F., M. S., and A. D. All authors approved the publishing of the manuscript.

Paper IV: Genome-scale model of *Rothia mucilaginosa* predicts gene essentialities and reveals metabolic capabilities
Conceptualization: N. L.; data curation, formal analysis, software, validation, and visualization: N. L.; methodology: N. L. and A. C.; investigation: N. L. and L. O.; funding acquisition and resources: A. D.; supervision: L. O., T. C., A. C., and A. D.; manuscript writing: N. L.; manuscript review: N. L., T. C., A. C., and A. D. All authors approved the publishing of the manuscript.

Paper I: New workflow predicts drug targets against SARS-CoV-2 via metabolic changes in infected cells

RESEARCH ARTICLE

New workflow predicts drug targets against SARS-CoV-2 via metabolic changes in infected cells

Nantia Leonidou [1,2,3] *, Alina Renz [1,2,3], Reihaneh Mostolizadeh [1,2,3,4], Andreas Dräger [1,2,3,4]

1 Computational Systems Biology of Infections and Antimicrobial-Resistant Pathogens, Institute for Bioinformatics and Medical Informatics (IBMI), Eberhard Karls University of Tübingen, Tübingen, Germany, 2 Department of Computer Science, Eberhard Karls University of Tübingen, Tübingen, Germany, 3 Cluster of Excellence 'Controlling Microbes to Fight Infections', Eberhard Karls University of Tübingen, Tübingen, Germany, 4 German Center for Infection Research (DZIF), partner site Tübingen, Germany

* nantia.leonidou@uni-tuebingen.de

Abstract

COVID-19 is one of the deadliest respiratory diseases, and its emergence caught the pharmaceutical industry off guard. While vaccines have been rapidly developed, treatment options for infected people remain scarce, and COVID-19 poses a substantial global threat. This study presents a novel workflow to predict robust druggable targets against emerging RNA viruses using metabolic networks and information of the viral structure and its genome sequence. For this purpose, we implemented pymCADRE and PREDICATE to create tissue-specific metabolic models, construct viral biomass functions and predict host-based antiviral targets from more than one genome. We observed that pymCADRE reduces the computational time of flux variability analysis for internal optimizations. We applied these tools to create a new metabolic network of primary bronchial epithelial cells infected with SARS-CoV-2 and identified enzymatic reactions with inhibitory effects. The most promising reported targets were from the purine metabolism, while targeting the pyrimidine and carbohydrate metabolisms seemed to be promising approaches to enhance viral inhibition. Finally, we computationally tested the robustness of our targets in all known variants of concern, verifying our targets' inhibitory effects. Since laboratory tests are time-consuming and involve complex readouts to track processes, our workflow focuses on metabolic fluxes within infected cells and is applicable for rapid hypothesis-driven identification of potentially exploitable antivirals concerning various viruses and host cell types.

Author summary

The recently emerged human coronavirus SARS-CoV-2 spread worldwide, causing severe challenges in health care, the economy, and society. Developing new vaccines and therapies is essential to prevent the next pandemic efficiently. However, vaccines have the disadvantage of decreased immunity over time, while they lose their efficacy against

Abbreviations: ABC, ATP-binding cassette; **ADP**, Adenosine 5'-diphosphate; **API**, Application Programming transfer Interface; **ATP**, Adenosine 5'-triphosphate; **BiGG**, Biochemical, Genetical, and Genomical; **BMBF**, Federal Ministry of Education and Research (*Bundesministerium für Bildung und Forschung*); **BOF**, Biomass Objective Function; **BRENDA**, Brunswick Enzyme Database; **CDP**, Cytidine 5'-diphosphate; **ChEBI**, Chemicals of Biological Interest; **CMP**, Cytidine 5'-monophosphate; **COBRApy**, Constraints-Based Reconstruction and Analysis for Python; **COVID-19**, Coronavirus Disease 2019; **CPEC**, Cyclopentenyl cytosine; **CPU**, Central Processing Unit; **CTP**, Cytidine 5'-triphosphate; **CTPS1**, CTP synthase 1; **dGTP**, deoxyguanosine 5'-triphosphate; **DNA**, Deoxyribonucleic Acid; **EC**, Enzyme Commission; **EDA**, Explanatory Data Analysis; **ExoN**, Exonuclease; **FBA**, Flux Balance Analysis; **fbc**, flux balance constraints; **FDA**, Food and Drug Administration; **FVA**, Flux Variability Analysis; **GDP**, Guanosine 5'-diphosphate; **GEM**, Genome-scale Metabolic Model; **GEO**, Gene Expression Omnibus; **GISAID**, Global Initiative on Sharing All Influenza Data; **GK1**, Guanylate Kinase 1; **GMP**, Guanosine 5'-monophosphate; **GMPS2**, Guanosine 5'-monophosphate Synthase; **GPR**, Gene-Protein-Reaction associations; **HBEC**, Human Bronchial Epithelial Cell; **HDE**, host-derived enforcement; **HEX1**, Hexokinase; **HIV**, Human Immunodeficiency Virus; **ID**, Identifier; **IMP**, Inosine 5'-monophosphate; **IMPD**, Inosine 5'-monophosphate Dehydrogenase; **KEGG**, Kyoto Encyclopedia of Genes and Genomes; **mCADRE**,

subsequent mutations and variants. Hence, effective pandemic preparedness requires discovering broadly acting antivirals with high resistance barriers. Here, we accelerate the detection of antiviral drug candidates against RNA viruses by analyzing metabolic changes in infected cells. Viruses rely on the host to acquire all macromolecules needed for their replication, re-programming the cellular metabolism according to their needs. For this reason, host-directed approaches are of great importance. We develop software to reconstruct constraint-based models, simulate infections of a cell, and identify host-based metabolic pathways that can be inhibited to suppress viral replication. We identify promising targets with inhibitory effects across multiple variants facilitating further *in vitro* and *in vivo* experiments. Our workflow can be applied to any RNA virus and aims to rapidly identify antiviral targets to better prepare for the next pandemic.

Introduction

In a study published in October, 2007,, scientists studying coronaviruses characterized the situation in China as a ticking "time bomb" for a potential virus outbreak [1]. They had three strong indications to worry: the animal-related eating habits in southern China, the previous appearance of Severe Acute Respiratory Syndrome Coronavirus (SARS-CoV)-like viruses in horseshoe bats, and the ability of coronaviruses to undergo recombination. Since the first major pandemic of the new millennium in 2002, over 4,000 publications on coronaviruses became available, giving insights and leading to the discovery of 36 SARS-related coronaviruses in humans and animals. Eighteen years later, the whole world experiences the realization of this prophecy with the emergence of the Coronavirus Disease 2019 (COVID-19) to be one of the deadliest respiratory disease pandemics since the "Spanish" influenza in 1918 [2]. Scientists globally try to understand the host's immunopathological response, how the novel virus Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) adapts, and how it spreads.

Viruses, being infectious agents, replicate only within the cells of a living organism and re-program them to form other virus particles and accelerate their own reproduction. Their life cycle is divided into four main steps: host cell attachment, penetration, reproduction within the host cell (uncoating, gene expression, replication, and assembly), and release [3]. To increase their mass production, they consume energy from the host cell. This dependency is proved by experimental findings showing considerable metabolic flux alterations in host cells upon infection [4]. To this end, engineering the host metabolism to govern viral infections is of great interest. In fact, one of the largest classes of small-molecule antiviral drugs, the nucleoside and nucleotide analogs, target metabolic enzymes in the nucleotide synthesis resulting in a nucleotide pool imbalance [5]. Examples of such analogs that are already used against RNA viruses are ribavirin [6], acyclovir [7], and remdesivir [8]. Systems-level analysis of gene knock-outs upon bacterial infection with bacteriophage lambda also revealed metabolic genes that, when knocked-out, prevented the phage from replication [9], confirming the engineering of host metabolism as a virus growth regulator.

These laboratory findings highlight the impact of viral biosynthesis on host metabolism and the importance of metabolic alterations in the virus growth minimization. Hence, finding a suitable Virus Biomass Objective Function (VBOF) that reflects the functions of the infected cell is of immense interest. The VBOF is a pseudo-reaction simulating the production of the different virus particles and is analogous to the biomass function used for the metabolic models of prokaryotes and eukaryotes. It consists of energy metabolites, nucleotides, and amino acids, essential for the replication and production of genetic material and proteins. In 2018,

metabolic Context-specificity Assessed by Deterministic Reaction Evaluation; **MEMOTE**, Metabolic Model Testing; **mRNA**, messenger Ribonucleic Acid; **NCBI**, National Centre for Biotechnology Information; **NDPK2**, Nucleoside Diphosphate Kinase 2; **NDPK3**, Nucleoside Diphosphate Kinase 3; **NDPK8m**, Nucleoside Diphosphate Kinase 8; **NSP**, Non-structural Protein; **NSP14**, Non-structural Protein 14; **OMEX**, Open Modeling EXchange format; **PREDICATE**, Prediction of Antiviral Targets; **PUNP4**, Purine-nucleoside phosphorylase; **RAM**, Random-access Memory; **REST**, Representational State Transfer; **RMA**, Robust Multiarray Analysis; **RNA**, Ribonucleic Acid; **SARS**, Severe Acute Respiratory Syndrome; **SARS-CoV**, Severe Acute Respiratory Syndrome Coronavirus; **SARS-CoV-2**, Severe Acute Respiratory Syndrome Coronavirus 2; **SBFC**, Systems Biology Format Converter; **SBML**, Systems Biology Markup Language; **SBO**, Systems Biology Ontology; **SBSCL**, Systems Biology Simulation Core Library; **TCA**, Tricarboxylic Acid Cycle; **UMP**, Uridine 5'-monophosphate; **UMPK5**, Uridine 5'-monophosphate Kinase 5; **UTP**, Uridine 5'-triphosphate; **VBOF**, Virus Biomass Objective Function; **VMH**, Virtual Metabolic Human; **VOC**, Variants of Concern; **VZV**, Varicella-Zoster Virus; **WHO**, World Health Organization; **WT**, wildtype; **XMP**, Xanthosine 5'-phosphate; **ZDV**, *Zentrum für Datenverarbeitung* (Center for Data Processing).

Aller *et al.* present a computational approach to create viral objective functions and predicted critical host reactions of the human macrophages against epidemic viruses, like the Zika virus [10]. The applicability of their method was verified by recovering antecedent antiviral targets and predicting new ones.

Notwithstanding the recent therapeutic advances and the approval of multiple vaccines, COVID-19 remains a substantial global health threat. Currently, great efforts are initiated to detect effective antiviral treatments for this pathogenic agent. Like all viruses, SARS-CoV-2 continuously evolves over time as modifications in its genome occur during replication. Such alterations are typical for viruses that encode their genome in RNA, as enzymes that copy the ribonucleic acid are prone to making errors leading to the presence of copying mistakes during viral replication [11]. It has been reported that SARS-CoV-2, along with all coronaviruses, has relatively low mutation rates ($\sim 10^{-6}$ per site per cycle) compared to other RNA viruses, like the Human Immunodeficiency Virus (HIV)-1 or influenza viruses [12, 13]. This is ascribed to the presence of proofreading and error-correcting enzymes that recognize and repair copying mistakes hindering the development of anti-CoV drugs and vaccines [14]. SARS-CoV-2 encodes an Exonuclease (ExoN) in the Non-structural Protein 14 (NSP14), which participates in the genome proofreading mechanism and results in low mutation rates (or high viral fidelity) [15]. The 5' region of the SARS-CoV-2 genome encodes for two open reading frames (ORF1a/ ORF1ab and ORF1b) which include 16 Non-structural Proteins (NSPs) [16]. These are followed by four structural proteins: nucleocapsid (N), envelope (E), the spike (S) and the membrane (M), and nine accessory proteins (NS) [16].

At the time of writing, five variants of SARS-CoV-2 have been designated as Variants of Concern (VOC) by the World Health Organization (WHO). These are the Alpha (14 December, 2020, United Kingdom (UK), lineage B.1.1.7), Beta (18 December, 2020, South Africa, lineage B.1.351), Gamma (2 January, 2021, Brazil, lineage P.1), Delta (24 March, 2021, India, lineage B.1.617), and Omicron (24 November, 2021, South Africa/Botswana, lineage B.1.1.529) variants [17]. These differ from the conventional virus in terms of their pathogen properties (e.g., transferability, virulence, or susceptibility to the immune response of recovered or vaccinated people). Mutations on the structural proteins occur most frequently and issue complications en route to pathogenesis. The most common mutation of the S protein is the non-synonymous replacement of aspartate by glycine (D614G), which is found to decrease the virus effectivity [18]. Mutations in the E protein have not been reported in any variants, except the Beta and Omicron. These are the substitution of proline by leucine (P71L) [19], and the exchange of the hydrophilic threonine by the hydrophobic isoleucine (T9I) [20].

Identifying potential targets and druggable compounds is of vast concern, and one way to detect them is by analyzing metabolic changes in infected cells. This can be achieved with the help of systems biology and the reconstruction of cell-specific Genome-scale Metabolic Models (GEMs) that recapitulate the metabolism of particular cell types [21]. Targeting the host metabolism has already been suggested as a prospective novel antiviral approach, given the relevance of metabolism in virus infection [22]. Since the emergence of SARS-CoV-2 and within a year several studies have been published trying to identify antiviral targets using constraint-based metabolic modeling and utilizing various approaches and resources [23–28]. For instance, a recent study by Bannerman *et al.* employs a draft model of the airway epithelial cells built from Recon1 [29], refines it using Recon3D [30], and predicts drug targets against SARS-CoV-2 [27]. However, they used pre-existing reconstruction tools and models to obtain a representation of the tissue metabolism.

In 2012 Wang *et al.* publish the metabolic Context-specificity Assessed by Deterministic Reaction Evaluation (mCADRE) algorithm to construct metabolic models based on human gene expression data and network topology information [31]. This tool is implemented in

MATLAB [32], and its functionality is based on the first version of the human model, namely Recon1 [29]. This resulted in its limited usability in the last few years since MATLAB is a commercial and closed-source software.

Here, we present pymCADRE, a re-implementation of mCADRE in Python striving for a more accessible and updated version of the reconstruction tool. Additionally, we implemented scripts for data pre-processing facilitating relevant curation tasks, such as assigning confidence scores to reactions, binarizing raw transcriptomic data, and calculating gene ubiquity scores. Pathological studies already pointed out that SARS-CoV-2 targets the airways and the lungs. The entry and infectivity of enveloped viruses are strongly regulated by proteolytic cleavage of the viral envelope glycoproteins [33]. In the case of SARS-CoV-2, the S protein, when bound in the cell surface, is susceptible to airway protease cleavage, which results in conformational change favoring the entry of the virus into human bronchial epithelial cells [33]. Further single-cell analyses provided insights into the virus replication and the cell tropism, confirming that infection with SARS-CoV-2 is also localized in the bronchial epithelial cells [34, 35]. Hence, we applied pymCADRE to create a novel tissue-specific model of primary Human Bronchial Epithelial Cells (HBECs) based on the already available human metabolic network, Recon1. We updated the model by including a biomass maintenance function that Bordbar *et al.* published in 2010 [36].

We subsequently infected this model *in silico* with the novel SARS-CoV-2 virus by constructing a viral biomass reaction derived from its structural information. Therefore, we created a fully automated computational tool in Python, called Prediction of Antiviral Targets (PREDICATE), which applies the stoichiometric approaches introduced by Aller *et al.* on a metabolic network, constructs a single VBOF, and creates an integrated host-virus model [10]. Subsequently, our tool predicts exploitable cellular metabolic pathways that can be inhibited to suppress virus replication with minimal or no effect on the cell. This is attained using two approaches: the host-derived enforcement (HDE) [10] and single-reaction knock-outs. We applied our automated script to our tissue-specific model Recon1-HBEC and detected potential host-based targets for future COVID-19 therapeutic strategies. We further used PREDICATE and validated the robustness of our predicted targets against all five variants of concern. We underline the identified metabolic reactions as experimentally exploitable drug targets for suppressing SARS-CoV-2 replication in human bronchial epithelial cells. We syntactically validated our model and compared it against the corresponding model reconstructed using mCADRE.

Altogether, our novel workflow can be summarized in a four-step process, as shown in Fig 1, which is fully transferable to any existing RNA virus and any host cell. With this, we aim to support further the development of effective therapies against emerging viruses and their mutations and create a library of drugs to design broad-spectrum antiviral therapies as an essential resource for pandemic preparedness.

Materials and methods

Overview of pymCADRE

The tool can be executed via the command line using:

```
python pymcadre.py
```
or using the provided Jupyter notebook named:
```
main_pymCADRE.ipynb
```
The package can also be found on the Python Package Index [37] (https://pypi.org/project/pymCADRE/) and can be installed using:
```
pip install pymcadre
```
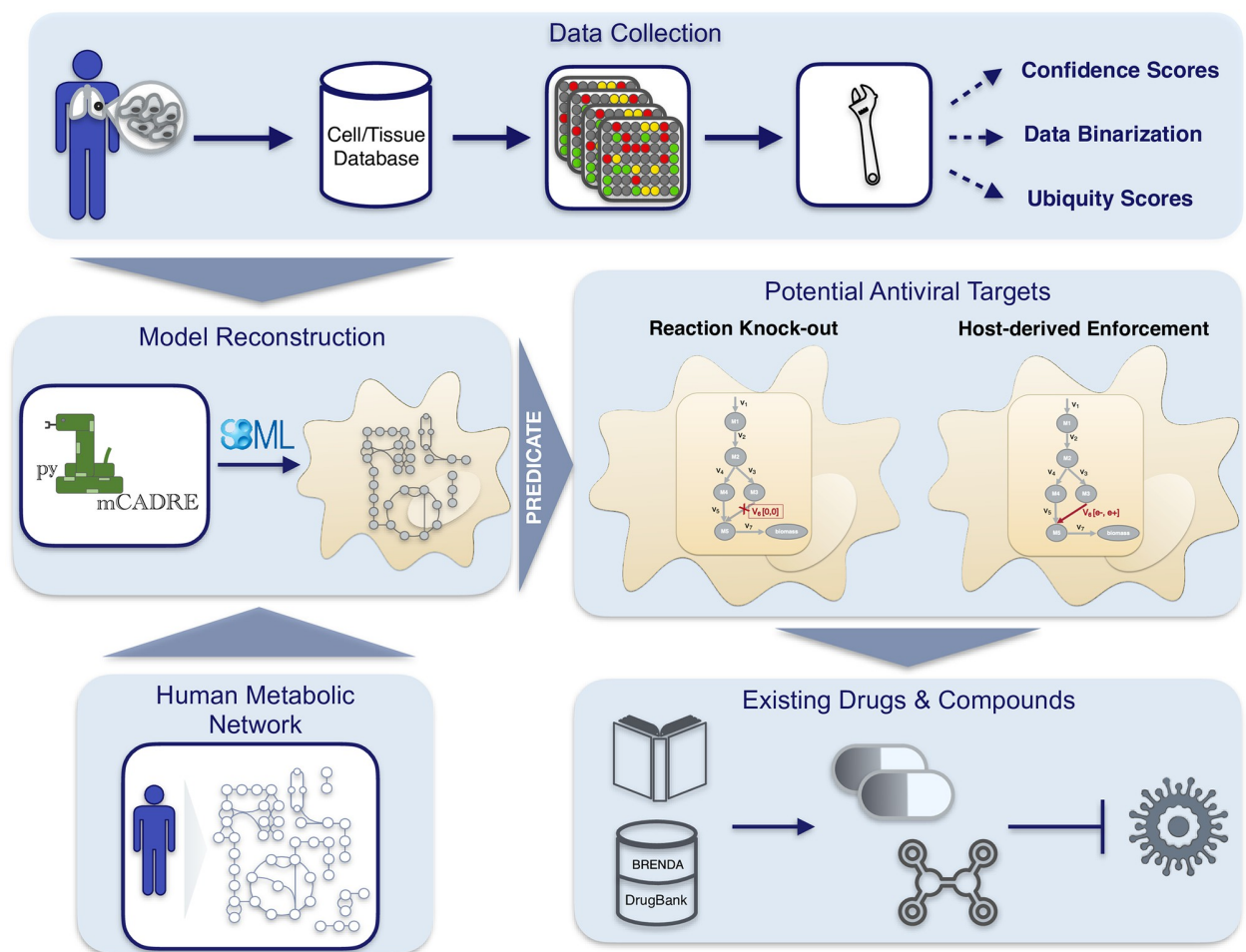
Fig 1. Workflow overview to reconstruct integrated host-virus genome-scale models and detect promising compounds with an antiviral activity.
After collecting and curating the required data (the gene expression data and the human metabolic network), pymCADRE reconstructs a tissue-specific
model using information from the network topology. The reconstructed metabolic network is then infected *in silico* with the virus of interest and is
used to detect promising antiviral targets in an automated process. Detailed description of the process and the respective algorithm, called
PREDICATE, is provided in Materials and Methods. Reaction knock-outs and the host-derived enforcement are used to detect exploitable enzymatic
targets that keep the host maintenance at 100%, while suppressing the virus replication. The resulted top hits are further inspected manually in terms of
already existing drugs and compounds in different databases, such as BRENDA and DrugBank.

Ranking of reactions. The first step in the pymCADRE pipeline is the ranking of all reac-
tions found in the generic model, as Wang *et al.* proposed [31]. The ranking relies on three cri-
teria: expression-based evidence, connectivity of reactions within the model, and confidence-
based evidence. The assignment of evidence scores to reactions aims their division to cores
and non-cores.

After binarizing the gene-expression data, the frequency of a gene's expression across all
experiments of the same tissue is computed; this is the ubiquity score $U(g)$ for each gene g:

$$\forall g \in G : U(g) = \frac{1}{|N|} \sum_{n \in N} X_{g,n} \tag{1}$$

where N is the total number of samples and $X_{g,n} \in \{0, 1\}$ denotes the absence or presence of the
gene g in sample $n \in N$. For instance, if a gene is expressed in three of five samples, its ubiquity

score will be 0.6. These scores are mapped to the corresponding reactions based on Gene-Protein-Reaction associationss (GPRs). That is the expression-based evidence $E_x(r)$ and can be either the minimum or maximum of two ubiquity scores depending on the respective GPRs rule: AND or OR. The expression-based evidence ranges from zero to one, indicating how likely a reaction is present in the selected tissue. More specifically, a score of zero represents a non-active reaction, while reactions with $E_x(r) > 0.9$ define the core set.

Afterwards, non-core reactions are ranked based on the connectivity-based evidence $E_c(r)$, using the network topology information of the generic model. This score defines in which order the reactions should be removed during pruning. The stoichiometric relationships in matrix **S** are applied to determine whether two reactions are connected. A pair of reactions are considered to be linked if they share at least one metabolite. For this purpose, a so-called weighted influence $WI(r)$ calculated as the ratio between $E_x(r)$ and the outgoing influence of each reaction, i.e., the number of reactions connected to it. Then, the actual connectivity-based evidence is determined by the sum of the weighted influences of all reactions adjacent to reaction r. In the supplementary file S2 Fig we graphically illustrate the computation of each score using a toy metabolic network comprising six reactions and four genes coming from four samples. Lastly, the confidence level-based evidence $E_l(r)$ is the third measure of evidence for non-core reactions and indicates the level of biological evidence for the generic model.

Check model function. After classifying the reactions into cores and non-cores, pymCADRE tests the model's ability to produce key metabolites from glucose. These include compounds in the Tricarboxylic Acid Cycle (TCA) and glycolysis, non-essential amino acids, and more. Totally, 38 metabolites were tested based on previously described criteria and used to evaluate similar models by the authors of mCADRE [31]. This list can be expanded and modified utilizing metabolomics data to include tissue-specific metabolites or known abilities of the tissue of interest.

Model pruning. The last step of pymCADRE is to sequentially remove each non-core reaction in a reversed order, i.e., beginning from those with the lowest calculated evidence [31]. The respective reaction will be removed if, and only if, its elimination does not prevent the model from producing key metabolites and the set of core reactions remains consistent. This consistency is tested by determining each reaction's minimum and maximum flux while ensuring that at least one is zero.

More specifically, firstly, the production of precursor metabolites is checked. If this test fails, there is no need to check for model consistency with Flux Variability Analysis (FVA) or FASTCC (time-saving step). If the test leads to successful results, the set of inactive cores and non-cores is determined, and the algorithm moves on with the removal of reactions. Reactions with zero expression will be removed with their corresponding inactive core reactions if sufficiently more non-cores are pruned. On the other hand, if the reaction has expression evidence, pymCADRE only attempts to remove inactive non-cores.

Integration of transcriptomic data in a human genome-scale metabolic model

The functionality of pymCADRE was tested using gene expression data of primary Human Bronchial Epithelial Cell (HBEC) downloaded from the Gene Expression Omnibus (GEO) database (accession number: GSE5264) [38]. All data obtained from GEO underwent manual curation and pre-treatment with scripts that we provided together with the pymCADRE source code. Firstly, the expression data were binarized based on the associated RMA signal intensity values and an absolute call value (i.e., P = present, A = absent, and M = on the border-line detection) was defined. This indicates whether messenger Ribonucleic Acid (mRNA) has

been detected for that specific gene or not, meaning whether it is expressed or not. The second curation step involved collecting confidence scores from the Virtual Metabolic Human (VMH) database, assigned to all reactions in the model. Then, the raw sample transcriptomic data was enhanced with two new information, gene symbol and Entrez identifiers. During binarization, genes present in the sample took the value of one, while marginal and absent calls were assigned to zero. Lastly, the essential ubiquity scores were calculated to represent a single gene's expression frequency across all samples.

The literature-based Recon1 [29] was obtained from the Biochemical, Genetical, and Genomical (BiGG) database [39] and was used as a generic host human model. It consists of 3,741 reactions, 2,766 metabolites, 1,905 transcripts, and 1,497 unique genes. We also incorporated a Biomass Objective Function (BOF) to Recon1 since it does not include one. For this purpose, we used the objective function from the human alveolar macrophage model published by Bordbar *et al.* in 2010 [36]. The biomass reaction with the identifier `biomass_hbec` represents the cellular maintenance requirements such as the ATP maintenance.

In the Recon1 model, there is no constraint growth medium defined; thus, all extracellular transport reactions have a minimum flux value of -1000.0 mmol gDW^{-1} h^{-1}. This means that all exchanges are allowed to carry a flux (rich medium), resulting to unusually high cell growth rates. We have defined here a minimal growth medium using the Constraints-Based Reconstruction and Analysis for Python (COBRApy) built-in function [40], which contains only essential components for growth. Since the availability of nutrients has a major impact on the metabolic fluxes, we re-ran our simulations using the blood medium [41]. The exact compositions of both media are provided in the supplementary file S5 Table.

We manually expanded our model by adding missing exchange reactions to all extracellular metabolites. We also updated all reaction annotations in our tissue-specific model, Recon1-HBEC, by assigning Kyoto Encyclopedia of Genes and Genomes (KEGG) IDs [42] and retrieving the corresponding pathways using the KEGG REST Representational State Transfer (REST) Application Programming transfer Interface (API). These subsystems were incorporated into the model as additional annotations to each reaction with the biological qualifier type `BQB_OCCURS_IN`. The reaction pathways were merged into main classes based on the KEGG classification system (https://www.kegg.jp/kegg/pathway.html). Additionally to the functionality checks incorporated into the mCADRE and consequently into pymCADRE, we examined the presence of futile cycles in our final tissue-specific model. As Fritzemeier *et al.* propose, we tested the production of energy-generating compounds by including energy dissipation reactions and disabling the external uptake of all metabolites [43]. Our final model could not produce any of the tested metabolites, meaning no futile cycles were included. The tested compounds are listed in the supplementary file S1 Table.

The reconstructions were conducted using a 3.3 GHz processor and 16 GB Random-access Memory (RAM), while Metabolic Model Testing (MEMOTE) [44] and the Systems Biology Markup Language (SBML) Validator from the ibSBML [45] were employed to assess the model's quality.

Stoichiometric reconstruction of SARS-CoV-2 biomass objective function

Similar to the biomass production function used for microbial metabolic models, the VBOF is a single pseudo-reaction imitating the production of different virus particles. It consists of nucleotides, amino acids, and components necessary for energy supply. The SARS-CoV-2 virus biomass objective function was created as proposed by Aller *et al.* and as extended by Renz *et al.* The approach considers the viral structure and its genome sequence, the subsequently encoded proteins, and their copy number, as well as the energy requirements for

nucleotide and peptide bonds [10]. The viral genome and protein sequences were downloaded from the National Centre for Biotechnology Information (NCBI) nucleotide database [46] (accession number: `NC_045512.2`, accessed in May, 2020). The genome copy number (G_g) and the number of copies of each of the non-structural proteins (C_{np}) was assumed to be one [10]. Moreover, the copy number of structural proteins was set to 1,000 for membrane proteins (C_m), 456 for nucleocapsid phosphoproteins (C_n), 120 for spike proteins (C_s), and 20 for envelope proteins (C_e) [47].

The SARS-CoV-2 falls into the fourth Baltimore group of viruses (Group IV, positive-sense single-stranded RNA viruses) [48], i.e., it synthesizes mRNA with the help of a template "-" single RNA antisense strand. Thus, the count of nucleotides in the positive strand equals the number of nucleotides in the complementary negative strand. The total moles of each nucleotide in a mole of virus particle were obtained by summing up the nucleotides in the positive and negative strand and multiplying this by the genome copy number. The moles were then converted into grams of nucleotide per mole of the virus by multiplying them with the respective molar mass of the nucleotides [10]. Similar calculations were conducted for the amino acids, as well. Eventually, the stoichiometric coefficients of each nucleotide and amino acid in the VBOF were calculated using the total viral molar mass [10].

For the estimation of the energetic requirements, the ATP requirement per amino acid polymerization and the pyrophosphate liberation during the polymerization of nucleotide monomers were considered. As proposed by Aller *et al.*, four ATP molecules and one pyrophosphate molecule are participating in the formation of nucleotide and amino acid polymers, respectively [10]. Subsequently, the total molar mass of the virus was calculated as the sum of all genome and proteome components.

Finally, to account for the lipid requirements we included phosphatidylcholine (`pchol_hs_c`), phosphatidylethanolamine (`pe_hs_c`), phosphatidylinositol (`pail_hs_c`), phosphatidylserine (`ps_hs_c`), cholesterol (`chsterol_c`), and sphingomyelin (`sphmyln_hs_c`) into the viral biomass function. Renz *et al.* examine the influence of lipids with various stoichiometric coefficients in the viral biomass function and the prediction of antiviral targets. However, they did not incorporated the lipid composition of a single virion into their final viral function [23]. We computed stoichiometric coefficients for these lipids from the surface area of a virion as suggested by Nanda *et al.* [25].

The generated final VBOF was appended into Recon1-HBEC, with a lower bound of zero and an upper bound of 1,000. The individual VBOF components and their stoichiometric coefficients are listed in S1 Table.

Prediction of host-based antiviral targets

Subsequent analysis of Recon1-HBEC allowed us to identify metabolic targets for antiviral therapies. As proposed by Aller *et al.*, Flux Balance Analysis (FBA) and FVA can be used to predict essential host reactions, especially in cases of novel emerging viruses [10]. This can be computationally achieved in two different ways: via single knock-out analysis or via HDE.

The single-reaction knock-out analysis investigates the effect of individual reactions with no flux. Both lower and upper bounds were systematically set to zero once with BOF as the objective function and once with the VBOF. Metabolic targets were reported when the host growth rate was higher than the virus growth rate and when more than 99% of the initial host growth rate was maintained.

A less harmful approach for the cell is the host-derived enforcement. As Aller *et al.* suggest, herein method, the reaction fluxes are constraint to FVA-derived ranges so that the

maintenance of the optimal host state is achieved while reducing the virus propagation [10]. For our analysis, we used an updated version of this method as modified by Renz *et al.* [47]. The re-calculated flux ranges for every reaction were then utilized, and the model was optimized for the VBOF. The resulting optima for the virus production were compared to the original optimal value. Hence, potential antiviral targets were reported when the virus growth rate with altered bounds was beneath the threshold of 90% of the initial growth rate. Additionally, to ensure a reduction of the virus replication, we keep only targets that had a non-zero flux when the VBOF was optimized. Our Recon1-HBEC model was examined for potential antiviral targets using both methods.

Testing targets' robustness against all known variants of concern

To test our targets' robustness, we examined the consequences of concerning SARS-CoV-2 mutations on our predicted metabolic targets. As of February, 2022, five SARS-CoV-2 VOC are known to differ from the conventional virus in terms of their pathogen properties (e.g., transferability, virulence, or susceptibility to the immune response of recovered or vaccinated people). These are the Alpha, Beta, Gamma, Delta, and Omicron variants [17]. Genomic sequences of patients infected with SARS-CoV-2 were retrieved from the Global Initiative on Sharing All Influenza Data (GISAID)'s EpiCoV database [49]. For each variant, we randomly selected 20 sequences adjusting only the location and variants filters as follows: (i) Europe/ United Kingdom for VOC Alpha GRY (B.1.1.7+Q.*), (ii) Africa/South Africa for VOC Beta GH/501Y.V2 (B.1.351+B.1.351.2+B.1.351.3), (iii) South America/Brazil for VOC Gamma GR/ 501Y.V3 (P.1+P.1.*), (iv) Asia/India for VOC Delta GK (B.1.617.2+AY.*), and (v) Africa/ Botswana and Africa/South Africa for VOC Omicron GRA (B.1.1.529) We investigated 100 sample sequences in total. To calculate the amino acid investment per virus, we used the annotated protein sequence of the SARS-CoV-2 reference genome (NCBI accession: NC_045512.2) and the mutation information extracted from GISAID. All used datasets and tested mutations are provided in the supplementary material S3 Table.

We calculated the stoichiometric coefficients of growth-related constituents for each mutated sequence and reconstructed for each one a VBOF as described in the previous sections. To speed up the calculations, we implemented PREDICATE, an automated script, which takes as input one or more genome sequences and computes the metabolic stoichiometry using information from the viral genome, the encoded proteins and their copy numbers, and the energetic requirements. The amino acid coefficients are calculated using the reference protein sequence, which our algorithm mutates by introducing all reported mutations (replacements, insertions, deletions, and duplications) extracted from the metadata. Afterwards, each VBOF is integrated into a given cell-specific metabolic network, in our study Recon1-HBEC, to create a host-virus model. Lastly, PREDICATE applies single-reaction knock-outs and HDE to the integrated model resulting in experimentally testable and robust metabolic virus-suppressing targets. Our script also generates different plots, providing insights into the dataset and a better understanding of the results. To evaluate the mutations' effect on the viral biomass, we computed the mean of all estimated coefficients across all mutated sequences and compared them against the wildtype (WT) coefficients.

PREDICATE can be applied to either one or more nucleotide sequences and all existing RNA viruses. This makes it particularly advantageous and time-saving to simultaneously study multiple viruses and variants.

Results

Tissue-specific reconstruction using pymCADRE

The pymCADRE tool was developed to reconstruct tissue-specific metabolic models based on human gene expression data and topological information from the metabolic network. Like mCADRE, pymCADRE leverages gene expression microarray data, literature-derived evidence, and information from the network topology to build context-specific metabolic models. More accurately, it uses a fully automated way to determine core reactions by setting a threshold to expression-based evidence. Therefore, reactions with scores above this threshold are characterized as core reactions, while the rest constitute the non-core set (more details in the Material and methods section). To test the functionality of pymCADRE and increase its ability to create multiple models of human cells, specifically related to the current outbreak of SARS-CoV-2, we applied pymCADRE to a microarray expression profile dataset of the primary HBEC. Prior to reconstruction, we incorporated a BOF to the first version of the human metabolic network, Recon1 [29], and used it as a generic host human model.

The objective function originates from the human alveolar macrophage model published by Bordbar *et al.* in 2010 (supplementary file S1 Table) [36]. We updated the resulting model by adding subsystems to all the missing metabolic reactions from Recon1. A subsystem-wise classification in Fig 2 indicates that most reactions in the final Recon1-HBEC model belong to the class of transport reactions, while the biosynthesis of other secondary metabolites is the least represented subsystem. Moreover, in Recon1, there is no growth medium defined, and all
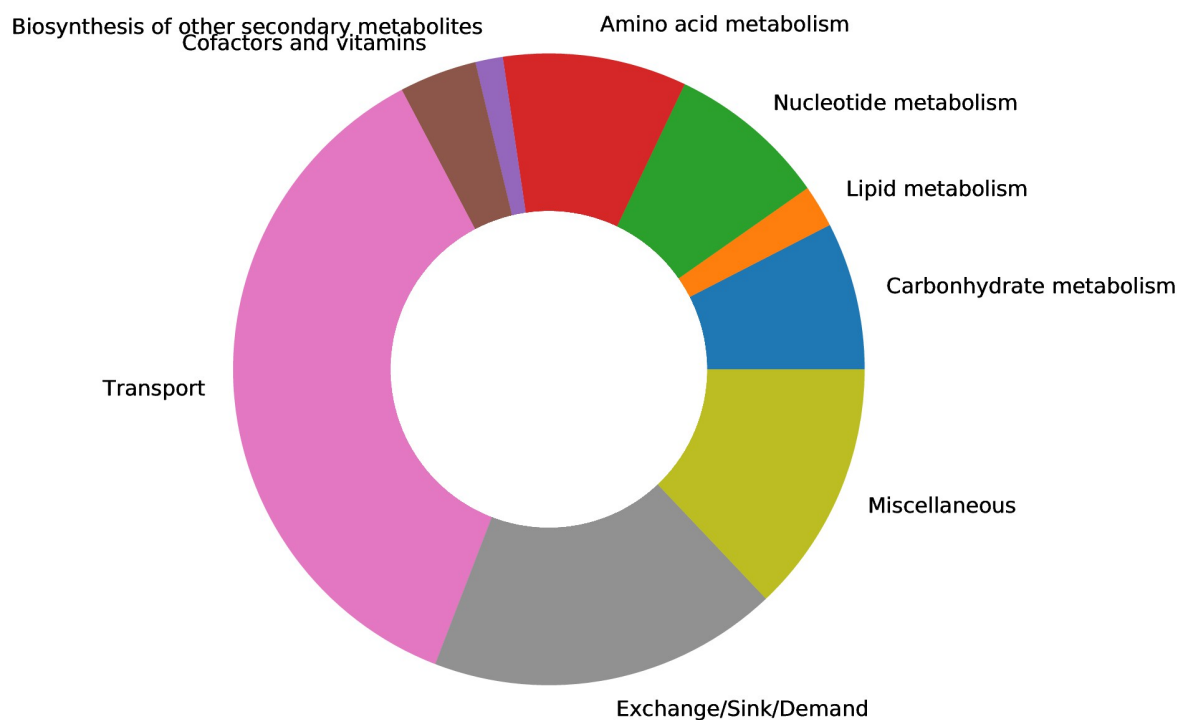
Fig 2. Subsystem-wise classification of all reactions included in Recon1-HBEC. The reaction pathways are merged based on metabolic pathways and according to KEGG. The biomass reaction was assigned to "Miscellaneous." The majority of reactions in the final Recon1-HBEC model are transport reactions, while the least amount of reactions is assigned to the biosynthesis of other secondary metabolites subsystem.

https://doi.org/10.1371/journal.pcbi.1010903.g002

Table 1. Analysis results of the HBEC-specific reconstructions using FVA for internal optimizations. The reaction overlap between both models is over 99.5%.

	Pruned Model			Removed Reactions	
	Reactions	Metabolites	Genes	Cores	Non-cores
mCADRE	1.977	1.442	1.905	9	487
pymCADRE	1.973	1.442	1.381	9	489

https://doi.org/10.1371/journal.pcbi.1010903.t001

extracellular transport reactions are open, i.e., lower fluxes equal -1000.0 mmol gDW^{-1} h^{-1}. Hence, we used the already-defined blood medium [41], and we computationally specified a minimal growth medium using the COBRApy [40] package. SBSCL [50] was used to independently evaluate the FBA problem and confirmed the solution. The exact medium compositions are provided in the supplementary file S5 Table.

The new integrated tissue-specific model Recon1-HBEC contains 1,973 reactions, 1,442 metabolites, and 1,381 genes (Table 1). Almost 70% of all reactions is associated to a gene-protein-reaction rule (1,391; 1,086 metabolic and 305 transport reactions), while 391 metabolic and 264 transport reactions are not related to any gene. To perform internal consistency checks, the user can choose between two COBRApy's [40] tailored optimization functions, FVA [51] and FASTCC [52]. Both methods detect blocked reactions and deliver consistent networks by resolving linear programming problems. Hence, the final pruned model does not contain any blocked reactions. We observed that pymCADRE reduces the pruning time while maintaining the highest possible accuracy compared to the model created with mCADRE (Table 1). With a 3.3 GHz processor and 16 GB RAM on a local computer, mCADRE with FVA demanded $\sim$6 CPU-hours, while pymCADRE $\sim$5 CPU-hours. Totally 1,272 blocked reactions were eliminated from Recon1 during the consistency check. Furthermore, 498 reactions (9 core and 489 non-core) were inactive in the cell type of interest and removed from the generic model during pruning. Inconsistencies were encountered in the performance of FASTCC as implemented in COBRApy. After multiple runs, the function detected a variable number of blocked reactions. This affected the final pruned model, which differed from the ground truth. However, internal optimizations with FASTCC were executed faster compared to FVA. Duplicating the available RAM can reduce the computational time of the pymCADRE twofold.

After the tissue-specific reconstruction, we refined the model using Recon3D [30] and HumanCyc [53]. We further extended the models by adding missing exchange reactions to all extracellular metabolites (71 in the mCADRE and 73 in the pymCADRE model). The final reconstructions shared over 2,040 reactions, meaning an overlap of over 99.5% of all reactions in each model. Hence, we have a considerable convergence between the tools, indicating the high quality of models generated with pymCADRE. Table 2 lists the symmetric difference between both models.

Table 2. Symmetric difference of reactions in the models created by mCADRE and pymCADRE.

mCADRE	
ARTPLM1	R group to palmitate conversion
ARTPLM2	R group to palmitate conversion
PE_HStm	Phosphatidylethanolamine scramblase
RETFA	Retinol acyltransferase
pymCADRE	
Htx	Peroxisomal transport of hydrogen
LRAT	Lecithin retinol acyltransferase

https://doi.org/10.1371/journal.pcbi.1010903.t002
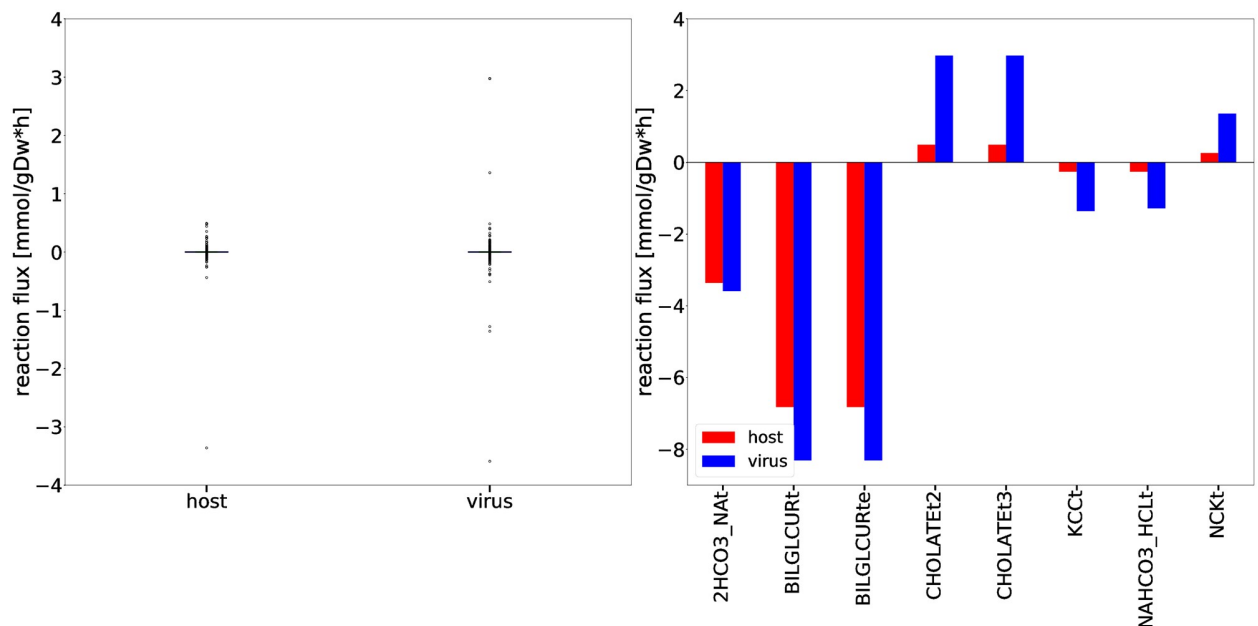
Fig 3. Flux dispersion among host and virus in the Recon1-HBEC model. Distribution of host and virus fluxes as derived from FBA. (a) The flux distributions were computed based on a five-number summary (S7 Table). Remarkable outliers with a flux value greater than 1.0 mmol gDW^{-1} h^{-1} or less than −1.0 mmol gDW^{-1} h^{-1} were investigated separately (b). (b) The fluxes through `2HCO3_NAt`, `BILDGLCURt`, and `BILDGLCURte` are remarkably higher when the model is optimized for both the host and the virus. Overall, all displayed reactions are essential for host maintenance and virus growth.

Additional analysis using FBA allowed us to study the flux dispersion between the host and virus and conclude which reactions are vital for both host maintenance and virus growth. Explanatory Data Analysis (EDA) showed that non-zero fluxes are mostly fluctuating above zero (Fig 3a). Totally 12 numerically distant values (outliers) were observed (Fig 3b). Inspection of the flux distribution vector showed higher viral flux through transporters of essential metabolites, like K$^+$ and Na$^+$, and Adenosine 5'-triphosphate (ATP)-binding cassette (ABC) transporter. Furthermore, the bicarbonate transporter (`2HCO3_NAt`) and the bilirubin beta-diglucuronide transporters (`BILDGLCURt` and `BILDGLCURte`) are used remarkably more by the host and virus to maintain its optimal state, compared to the virus.

Similar to mCADRE, pymCADRE encompasses functionality tests to ensure the fulfillment of the resulting models' basic cellular metabolic capabilities. These tests include the production of various metabolites, such as amino acids and compounds from the TCA when the uptake of glucose is enabled [31]. Additional to this, we tested our model for internal cycles that result in erroneous energy production by testing the production of different energy metabolites when no nutrients are available. [43] Our final model did not include any futile cycles since none of the metabolites could be generated.

The new tissue-specific model created with pymCADRE was converted into SBML Level 3 Version 2 [54] format using the Systems Biology Format Converter (SBFC) [55] and passed the syntactical validation using libSBML [45]. Additionally, the MEMOTE suite Version 0.11.1 was used to assess the GEM quality [44]. MEMOTE reports for a given GEM an independent and comparable score along with a comprehensive overview. This test reported a score of 70% for our integrated model, which indicates a well-annotated model of high quality. Metabolic networks of the same or different tissue possess lower quality scores. For instance, the integrated model of macrophages has a MEMOTE score of 44% [23], while the model of airway

epithelial cells from Bannerman *et al.* has a score of 51% [27]. While these models contain a wide range of reactions and metabolites annotations, the mass and charge imbalances are still high resulting in lower scores. Nevertheless, they contain over 1, 800 blocked reactions, while the integrated macrophage-virus model contains over 140 dead-end and orphan metabolites. The automatically reconstructed models of bronchial and airway epithelial cells from Wang *et al.* have a lower MEMOTE score of 19% [31]. It is mainly attributed to lacking database cross-references and missing Systems Biology Ontology (SBO) [56] terms.

To examine whether pymCADRE functions as expected, we implemented test scripts, which are available at https://github.com/draeger-lab/pymCADRE/.

Since we purposed to use the model to detect possible anti-SARS-CoV-2 targets, we also included a VBOF that imitates the production of virus particles from its different constituents. Following the pipeline developed by Aller *et al.* and extended by Renz *et al.*, we created this pseudo-reaction and used it to infect the new model (Recon1-HBEC) *in silico*. The human bronchial epithelial cell's biomass maintenance function (BOF) encompasses amino acids, DNA and RNA nucleotides, and compounds vital for energy supply, and other macromolecules like fatty acids and phospholipids. Similarly, the VBOF contains amino acids, RNA, lipids, and energy-related compounds (S1 Table, S4 Fig), as well necessary lipids. Analysis of both functions highlights leucine as the most-used amino acid (highest stoichiometric coefficient) in the SARS-CoV-2 growth and the maintenance of the host bronchial cells, while both host and virus utilize only a few tryptophan (Fig 4). Moreover, the same amount of asparagine and phenylalanine is required for the maintenance of the host cell, while the virus needs less phenylalanine. Similar pattern was observed for tyrosine and histidine. Using FBA, optimization of the Recon1-HBEC for the host resulted in a flux for the biomass maintenance function of 0.2344 mmol $gDW^{-1} h^{-1}$, while optimizing the SARS-CoV-2 growth function resulted in a flux of 0.1575 mmol $gDW^{-1} h^{-1}$.

Stoichiometric modeling of the integrated host-virus model predicts targets against SARS-CoV-2

To analyze the host-virus interactions from a metabolic point of view, we created an integrated stoichiometric model of human bronchial epithelial cells infected with SARS-CoV-2. We then used our model to detect host-based reactions, which, when constrained, reduce the virus production the most. According to Aller *et al.*, this analysis can be computationally implemented through systematically setting individual lower and upper bounds to zero (i.e., reaction knock-outs). Applying this approach, we identified a single target enzyme, which if knocked-out, completely inhibits the virus while keeping the host maintenance at 100% of its initial growth rate. This enzyme is called Guanylate Kinase 1 (`GK1`, EC-Number: 2.7.4.8, KEGG Reaction ID: `R00332`) and catalyzes the conversion of ATP and Guanosine 5'-monophosphate (GMP) to Adenosine 5'-diphosphate (ADP) and Guanosine 5'-diphosphate (GDP) (KEGG Reaction ID: `R00332`):

$$ATP + GMP \rightleftharpoons ADP + GDP.$$
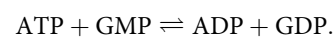
To ensure the maintenance of the metabolic network in a host-optimized state while suppressing the viral growth, we applied the HDE (see Materials and methods) [10, 47]. We constrained all reaction fluxes to ranges obtained from FVA, allowing the attainment of host-optimal state and suppressing the virus production at most. This approach verified the enzymatic target `GK1` and revealed further possible compounds that could inhibit the viral production without harming the host cell. The most promising novel hit was the CTP synthase 1 (`CTPS1`) from the *de novo* pyrimidine synthesis pathway that, when constrained, inhibited the
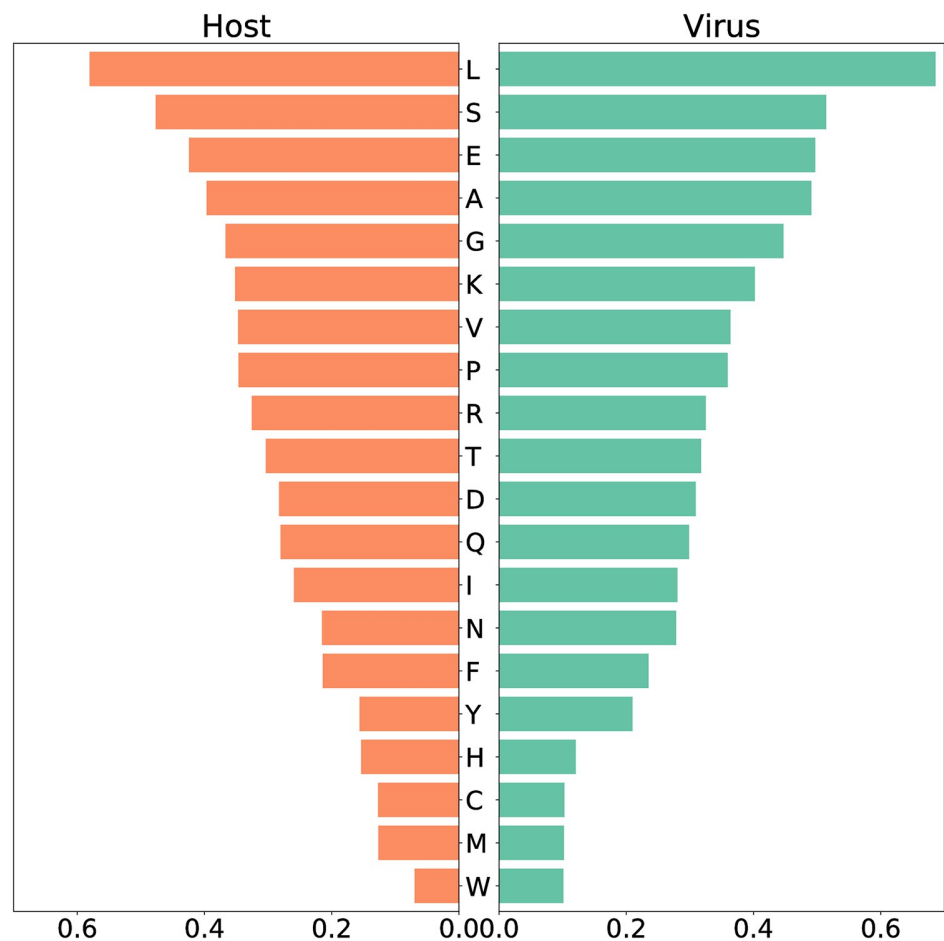
Fig 4. Amino acid usage between host and virus based on the stoichiometric coefficients. The two panels show the amino acid composition of the host maintenance function (left) and the virus biomass (right). The amino acids are annotated using the one-letter code (S6 Table). Both host and virus use mostly leucine (L) for their maintenance/growth, while tryptophan (W) is needed at least. The same amount of asparagine (N) and phenylalanine (F) is required for the maintenance of the host cell, while the virus needs less phenylalanine. Similar pattern can be observed for tyrosine (Y) and histidine (H).

https://doi.org/10.1371/journal.pcbi.1010903.g004

virus growth by 62% with no effect on the host's maintenance (100% of the initial rate). CTPS1 catalyzes the formation of Cytidine 5'-triphosphate (CTP) from Uridine 5'-triphosphate (UTP). It is important to note here that when the activity of CTPS1 is constrained, and therefore the formation of CTP, host cells can use alternative routes through the salvage pathway using Cytidine 5'-diphosphate (CDP) and/or Cytidine 5'-monophosphate (CMP) to restore the CTP levels directly. Similar results were observed for GK1 with adapted bounds. Further 33 enzymatic targets with inhibitory effects on the virus production were reported using the HDE approach. These concerned, for instance, restricting the extracellular exchange of L-proline and phosphate (EX_pro__L_e and EX_pi_e) and constraining the enzymatic activity in the metabolism of purines (PUNP4, IMPD, GMPS2, NDPK8m, and DGNSKm) and pyrimidines (UMPK5, NDPK2, and DTMPK). Moreover, inhibiting the functionality of enzymes in carbohydrate metabolism, more specifically in the amino/nucleotide sugar metabolism and sucrose metabolism (e.g., ACGAMK, UAGDP, and PGMT) as well as in glycolysis/
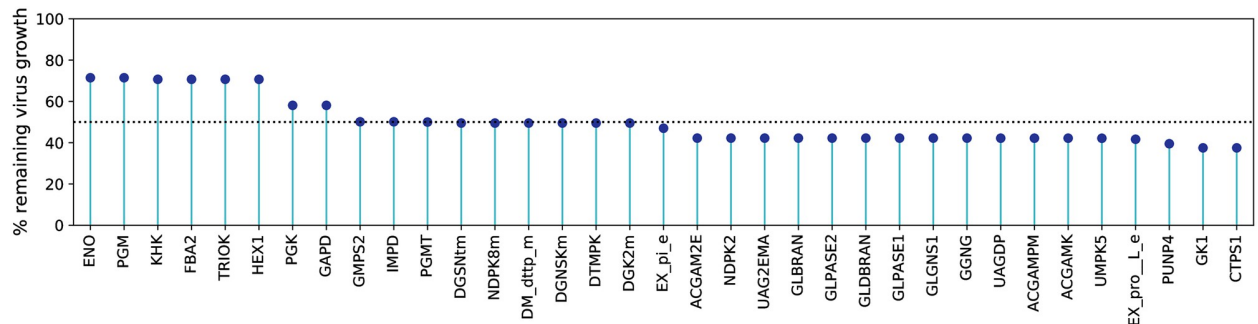
Fig 5. Enzymatic targets of SARS-CoV-2 from the HDE experiments applied to the Recon1-HBEC model. Potential antiviral targets were reported when the virus rate of growth with shifted bounds was beneath the threshold of 90% of its initial growth rate. Enzymes with adapted bounds from the purine and pyrimidine metabolism led to a remarkable virus decrease, while further promising targets were reported from the carbohydrate metabolism. The dashed line represents the 50% of the virus remaining.

https://doi.org/10.1371/journal.pcbi.1010903.g005

gluconeogenesis (GAPD, PGK, HEX1, FBA2, PGM, and ENO) led to a remarkable decrease in the viral production by 50% to 58% of the initial growth. Fig 5 illustrates all antiviral targets predicted using HDE against the percentage of the remaining virus growth after constraining the reaction bounds. Detailed information about all reactions is included in the supplementary material S2 Table.

The GK1 has been recently identified as an essential component for viral propagation and a potential target for antiviral therapies against SARS-CoV-2 in the human alveolar macrophage model [23] and further cell lines [24, 28]. Renz et al. showed that GK1 could decrease the virus production up to 50% without damaging the macrophages' maintenance (100%) [23]. Our host-derived enforcement on the bronchial epithelial cells also reported GK1 as a potential anti-SARS-CoV-2 target, however, with a similar impact on the virus production compared to CTPS1. Similarly, ENO, PGK, and PGM have been predicted as targets to inhibit the production of SARS-CoV-2 by Delattre et al..

Interestingly, the already identified robust target GK1 is closely interconnected with two reported promising targets in the purine metabolism (Fig 6, created with Newt [56] using the using the Systems Biology Graphical Notation (SBGN) [57]) using the using the Systems Biology Graphical Notation (SBGN) [58]). IMPD catalyses the NAD$^+$-dependent oxidation of Inosine 5'-monophosphate (IMP) to Xanthosine 5'-phosphate (XMP) that is subsequently used by GMPS2 to generate GMP. From this, we suggest that focusing on the purine metabolism, and more specifically on the action of one of these enzymes to inhibit SARS-CoV-2 is a well-established approach that needs to be validated in vitro and in cell culture experiments.

Metabolic fluxes are highly affected by the nutrients' availability. Since our approaches mainly focus on studying the metabolic changes in infected cells, fluxes play a major role in the simulation outcomes. So far, we have focused on a chemically defined medium simulating the human blood [41]. Additionally, we examined the virus inhibition that our targets could reach using a minimal growth medium computed with linear optimization [40]. A novel enzymatic target was reported from the pyrimidine metabolism with the minimal medium defined, named NDPK3. Like NDPK2, UMPK5, and GK1, it belongs to the class of phosphotransferases with a phosphate group as an acceptor and was highlighted as a hit target from the single-reaction deletions and HDE. NDPK3 constrained resulted in a decrease of 44.6% of the virus growth, which is comparable to the effect of GK1 using the minimal medium. However, NDPK2 and UMPK5 with adapted upper and lower fluxes led to higher viral reduction and
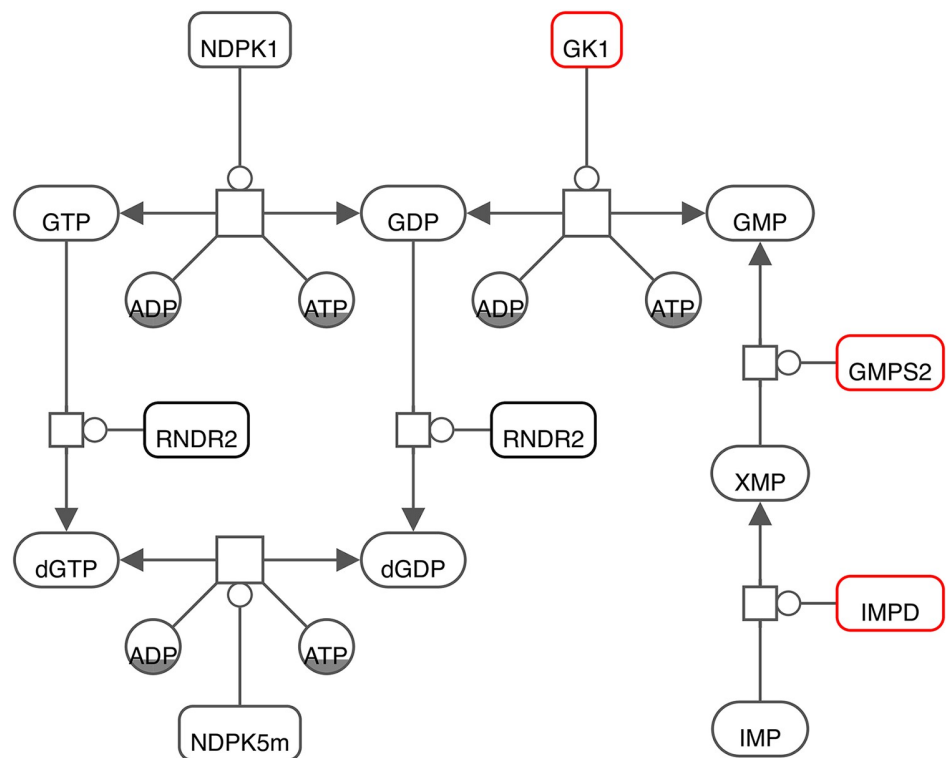
Fig 6. Graphical illustration of the interconnection between promising targets reported from the purine metabolism (red colored). To annotate reactions and metabolites, BiGG Identifiers were utilized.

https://doi.org/10.1371/journal.pcbi.1010903.g006

appeared as the best hits. The HDE-derived metabolic targets using the minimal medium are shown in S1 Fig and the medium definition is provided in the supplementary file S5 Table.

Altogether, we created a fully automated computer tool, which simulates the virus growth in target cells with the help of metabolic networks. Subsequently, our tool applies the above-mentioned host-dependent approaches, HDE, and reaction knock-outs, and predicts enzymatic targets with high inhibitory potency against the virus. The SBML model [59] of the SARS-CoV-2-infected bronchial epithelial cell (iHBEC-BOFVBOF-2023) is available at the BioModels Database [60] as an SBML Level 3 Version 2 file [54] with flux balance constraints (fbc) package [61] distributed as Open Modeling EXchange format (OMEX) archive [62] including annotation [63].

Predicted targets are robust against all known variants of concern

Novel mutations of RNA viruses emerge daily, and as of February, 2020, five SARS-CoV-2 variants have prevailed and spread since its emergence in 2019. These are the Alpha (B.1.1.7), Beta (B.1.351), Gamma (P.1), Delta (B.1.717.2), and Omicron (B.1.1.529) variants [17] and have been marked as VOC. Since the beginning of the COVID-19 pandemic, there has been an exponential growth in the number of stored genome sequences within large databases. The WHO asked all scientists around the world to upload their data on the GISAID database and help accelerate the response against health threats to humankind [49]. In January, 2020, the GISAID's EpiCoV database launched, becoming the most popular repository for SARS-CoV-2 as it gathers over eight million viral sequences by February, 2022. To examine the variants'

effect on the predicted metabolic targets, sequences for all VOC were downloaded from GISAID and investigated further.

We reconstructed a SARS-CoV-2 VBOF using the same approaches as with the reference (wildtype) sequence for each retrieved mutated sequence. We reconstructed 100 individualized biomass functions and tested each to detect enzymes that inhibit the virus's growth while keeping the host maintenance at maximum. To speed up the reconstruction and analysis processes, we developed an automated script to analyze more than one sequence simultaneously (Fig 7). Additionally, we implemented an algorithm to modify reference protein sequences and introduce amino acid mutations (replacements, insertions, deletions, and duplications) and named this tool Prediction of Antiviral Targets. Since RNA viruses are composed of similar building blocks, nucleotides, and proteins, our pipeline can be applied to any single- or double-stranded RNA virus that could infect any cell or tissue type.

To evaluate the mutations' effect on the viral biomass, we calculated the mean of all estimated coefficients across all mutated sequences and compared them against the WT stoichiometries. We did this by looking at the variant-wise differences to the WT. Fig 8 shows how much the variant-wise calculated mean of coefficients deviates from the stoichiometries calculated for the reference sequence. We observed a remarkable increase in the stoichiometric coefficients of ATP and ADP between the Omicron variant and the WT. This pattern is mainly distinct to the ATP and ADP but is observed for the majority of the stoichiometric coefficients. We analyzed the mathematical calculations that led to the stoichiometric coefficients to explain this further. All coefficients depend on the total viral molar mass M_v, which is derived from the sum of the mass of the genome (G_i) and proteome (G_j) [10]. The randomly downloaded genomic sequences of the Omicron variant contained a higher amount of NNN stretches (i.e., nucleotides that could not be determined via sequencing) compared to the other variants. Consequently, the Omicron variant has a decreased count of nucleotides (G_i) and amino acids (G_j), thus a lower total molar mass M_v. Moreover, the overall moles of energy (A^{TOT}) needed to assemble the structural and non-structural proteins strongly influences the stoichiometric
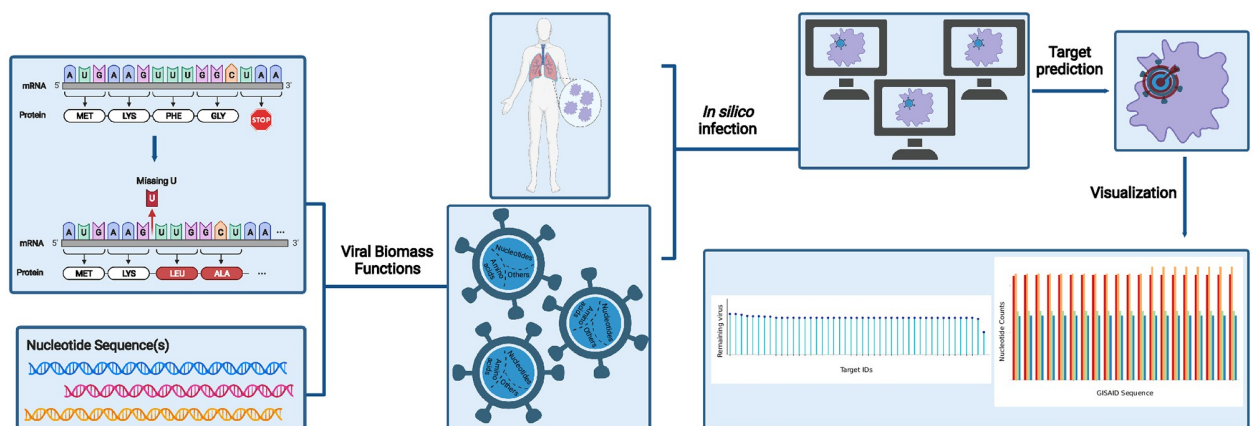
Fig 7. Overview of PREDICATE developed to create viral biomass reactions and predict host-based antiviral targets using host-virus models. First of all, our algorithm, PREDICATE, modifies the amino acids in the protein sequence according to the defined mutations. The mutated protein sequence and the nucleotide sequences are further employed to calculate the stoichiometric coefficients for the virus biomass functions. Reaction knock-outs and the host-derived enforcement are applied to reveal enzymatic reactions that suppress the viral replication. The final step generates various plots, providing insights into the dataset and a better understanding of the results. This pipeline can be applied to either one or more nucleotide sequences and all existing RNA viruses. This makes it particularly advantageous and time-saving when studying multiple variants of a single virus. The number of genomic input sequences equals the number of the calculated VBOF. The Materials and Methods section describes the implemented approaches to predict antiviral targets. Figure created with BioRender (BioRender.com).

https://doi.org/10.1371/journal.pcbi.1010903.g007

	Alpha	Beta	Gamma	Delta	Omicron
ADP	-2e-05	0.045	0.0037	0.035	0.14
ATP	-0.0017	0.041	0.0015	0.033	0.13
CTP/GTP	-0.001	-0.0024	-0.0015	-0.0019	-0.0082
PPi	-0.0027	-0.0058	-0.0038	-0.0044	-0.021
UTP	-0.0017	-0.0035	-0.0023	-0.0025	-0.012
Ala	-0.0013	-0.00012	-0.00037	0.00032	-0.0077
Arg	-0.00023	0.00069	0.0039	-0.0015	0.0023
Asn	-0.001	0.00075	0.00019	0.0025	-0.0027
Asp	-0.0062	-0.0029	-0.0027	-0.011	-0.011
Cys	7.5e-05	0.00023	1.7e-05	0.0053	0.00068
Gln	0.00011	0.00068	6e-05	0.00025	-0.013
Glu	-8.1e-06	-0.00066	-0.00036	-0.0018	0.0087
Gly	-0.0035	0.0034	-0.0044	0.003	0.0066
His	0.0015	-0.00025	-0.00078	0.0013	0.0033
Ile	0.0017	0.0064	0.00085	-0.0089	0.0046
Leu	0.0052	-0.0006	-0.00044	0.00046	0.0082
Lys	0.0053	0.00077	0.0051	0.0019	0.015
Met	6.1e-05	0.00023	1.7e-05	0.0053	0.00067
Phe	0.0052	0.0011	0.0026	-0.00047	0.0051
Pro	-0.00099	0.00017	-0.0071	-0.0011	-0.0022
Ser	-0.0044	0.0012	0.0022	0.00073	-0.0019
Thr	-0.0011	-0.0043	-0.00078	0.0078	0.0096
Trp	7.3e-05	0.00027	1.9e-05	0.00021	0.0008
Tyr	0.00013	0.0022	0.0038	0.005	0.0042
Val	-0.00063	0.002	-0.00089	2e-05	0.0033

Scale bar: $C_{variant} - C_{wildtype}$, ranging from -0.04 (dark red) to 0.04 (dark blue).

Fig 8. Variant-wise comparison of stoichiometric coefficients derived directly mutated sequences and the wildtype. The difference between the average stoichiometric coefficients of the individual variants and the reference sequence was computed. Red color highlights decreased stoichiometric coefficients in the variants, while increased coefficients are colored by blue. A remarkable increase can be observed in the stoichiometric coefficients of ATP and ADP between the Omicron variant and the wildtype. The stoichiometries of charged and hydrophobic amino acids were higher for the Omicron variant. All in all, the variations between mutants and wildtype are very small.

https://doi.org/10.1371/journal.pcbi.1010903.g008

coefficient of ATP [10]. The A^{TOT} is related to the total amino acids counts (X_j). Overall, the Alpha variant included more deletions in the spike (`H69del`, `Y144del`, and `V70del`) and NSP6 (`F108del`, `G107del`, and `S106del`) proteins compared to the Omicron variant. The same holds for the Beta, Gamma, and Delta variants. This affected the X_j, which was higher for

the Omicron variant. Altogether, a decreased total viral molar mass and a higher total amino acids count resulted in the apparent rise of the ATP and ADP stoichiometric coefficients for the Omicron variant. Accordingly, the absolute differences between the WT and the Omicron variant are higher than the rest.

When looking at the differences between the amino acids and the WT stoichiometric coefficients, a noticeable increase in the Omicron Variant can be observed for lysine. For this reason, we inspected the respective amino acid mutations. In more than half of the Omicron-related genomic sequences, other amino acids were more often replaced by lysine (Spike N440K, Spike N764K, Spike N856K, Spike N969K, Spike T478K, Spike N679K, Spike T547K, and N R203K). In contrast, the substitution of lysine by other amino acids is rarely occurred (NSP3 K38R and Spike K417N). This also affected the stoichiometric coefficient of asparagine. As most of these mutations emerge in the spike protein, which has the highest copy number, their impact on the amino acid count, and consequently, the stoichiometric coefficient, is considerable. Among the variants for which a higher coefficient was computed for asparagine than the WT, the greatest increase was observed for the Delta variant. This could be justified by the presence of mutations, in which mostly an amino acid is being replaced by asparagine (Spike D950N, M S197N, NSP16 H186N, NSP3 K1693N, NSP8 K37N, and NSP3 K902N). A substitution of asparagine by another amino acid occurs only in three mutation types. Thus, overall there is an increase in the total amount of asparagine, and therefore, in the stoichiometric coefficient for the Delta variant. Lastly, the Omicron variant needs the least glutamine (−0.013) and the most lysine (0.015) compared to the WT.

To verify the validity of our calculations, we searched in the literature to find evidence about the amino acid composition of the different variants. For instance, we observed higher stoichiometric coefficients of charged and hydrophobic residues in the Omicron variant compared to the Delta. Recently, computational analyses indicated in the Omicron variant an increased amount of arginine, lysine, aspartate, and glutamate that contribute to the formation of salt bridges [64]. The same study pointed out the accumulation of the hydrophobic residues, phenylalanine and isoleucine, in the spike protein of same variant.

After investigating the mutations' impact on the viral stoichiometric coefficients, we tested the effectiveness of the previously identified targets against the SARS-CoV-2 variants repeating the single-reaction deletions and HDE experiments. Our single-reaction knock-outs indicated GK1 to be the only potent antiviral inhibitor. All host-based targets detected from the HDE analysis to have an inhibitory effect on SARS-CoV-2 for all variants are shown in Fig 9. Targets were reported as potentially effective when the virus growth rate with altered bounds was lower than the threshold of 90% of its initial growth rate. The CTPS1 was reported to have the highest virus inhibitory effect across all Variants of Concern. After its inhibition, the virus growth dropped to 24.4–37.5% of its initial growth in the host cell. Further possible compounds were found to inhibit the viral production while keeping the host at maximum. Eight targets in total were detected to be WT-specific: ACGAM2E, DGK2m, DGNSKm, DGSNtm, HEX1, NDPK8m, PUNP4, and UAG2EMA. Except for CTPS1, GK1 was a common target, which constraint led to a reduced virus growth, however not as effective as CTPS1. Moreover, the five SARS-CoV-2 variants shared twelve additional hits with the wildtype (WT) that reported inhibitory effects (S5 Fig). Our integrated host-virus model suggested the supplementation of L-proline and phosphate in the host's environment as potential targets ensuring the cell's maintenance. Moreover, four targets from the carbohydrate metabolism (UAGDP, ACGAMPM, ACGAMK, and PGMT) showed a remarkable inhibitory effect in all studied variants, while once more targeting the metabolism of purines and pyrimidines seemed promising for all virus variants.
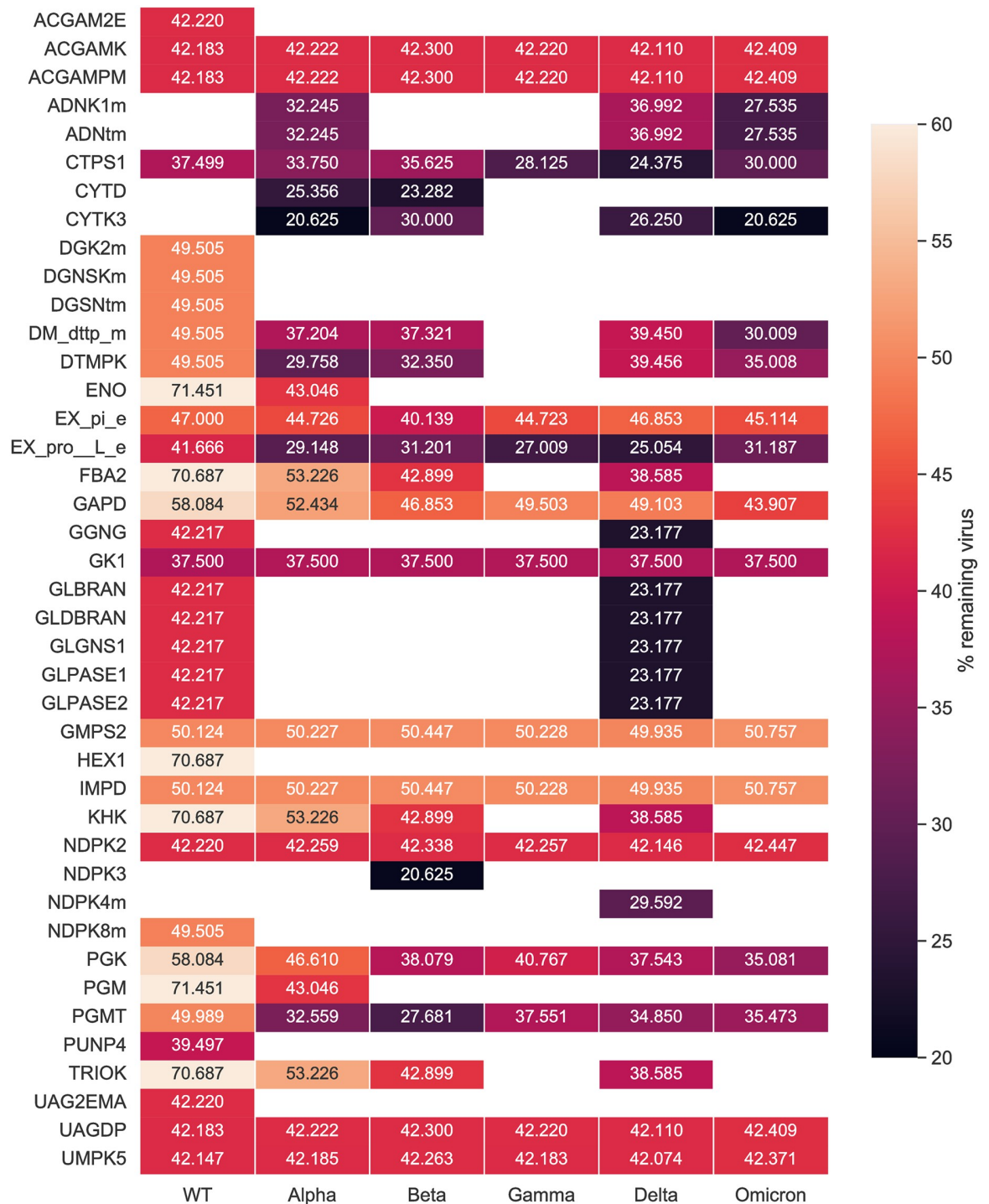
Fig 9. Results of the host-derived enforcement applied to all known variants of concern. The range and effect of reaction inhibitions on the VBOF were calculated while keeping the host's maintenance at 100%. Only targets predicted across all retrieved sequences for a single variant were considered robust and were examined further. Empty cells in the heatmap represent targets that were not predicted as potential inhibitors for the corresponding variant. CTPS1 showed the highest inhibitory effect against the virus at all studied variants, followed by GK1. Targeting the amino sugar and nucleotide carbohydrate metabolism highlighted to be a robust hit against all studied variants.

https://doi.org/10.1371/journal.pcbi.1010903.g009

Existing drugs and inhibitors could target predicted enzymes to hinder the growth of SARS-CoV-2

The computational approaches used here allowed the prediction of diverse enzymatic targets that could inhibit the SARS-CoV-2 replication in primary human bronchial epithelial cells. For these targets, we evaluated their corresponded enzymes considering already existing approved drugs using the BRENDA [65] and DrugBank [66] databases. We found various hitherto approved drugs and compounds that target some of the predicted reactions and could inhibit them, including those targeting the very promising enzymes CTPS1 and GK1. Table 3 lists examples of already existing drugs that inhibit our predicted anti-SARS-CoV-2 target reactions. These compounds and drugs could be used as an indication to validate the computational predictions made here experimentally.

Like all living cells, virus-infected cells require nucleotides to synthesize deoxyribonucleic and ribonucleic acid to strengthen their proliferation. Hence, nucleotide metabolism is regulated to establish constant pools of pyrimidines and purines. Various drugs targeting the nucleotide metabolism in viral infections represent a therapeutic approach to limit viral replication. There are two main strategies to rewire the nucleotide metabolism: via purine and pyrimidine analogs (i.e., modified nucleosides used to stop DNA or RNA polymerase) or directly inhibiting the enzymes involved in DNA and RNA synthesis. The majority of our predicted targets are involved in the purine and pyrimidine metabolism.

We conducted extensive literature research and highlighted the cytidine analog Cyclopentenyl cytosine (CPEC) as an already known competitive inhibitor of the CTP synthetase CTPS1 (Table 3). In CPEC, the ribose is substituted by an unsaturated carbocyclic ring and must undergo three times phosphorylation to form CPEC-TP that finally acts as an inhibitor of CTPS1 [71]. A nucleoside analog named acyclovir is an approved drug that acts against GK1 (Table 3). In acyclovir, the sugar in the deoxyguanosine is substituted by an acyclic side chain, a (2-hydroxyethoxy)methyl substituent, at position nine. The viral DNA polymerase is competitively inhibited by acyclovir which acts as an analog to deoxyguanosine 5'-triphosphate (dGTP). This results in chain termination since the adherence of further nucleosides is prevented by the absence of the 3'-hydroxyl group [72].

The second approach is the direct inhibition of enzymes related to nucleotide synthesis. In the past few years, diverse enzyme inhibitors have been known to treat viral infections. One such antiviral, merimepodib, targets the action of inosine-5'-monophosphate dehydrogenase (IMPD) and has already been tested against emerging RNA viruses (e.g., Zika, Ebola, Lassa, Junin, and Chikungunya viruses) [73]. Merimepodib has also been examined in the context of SARS-CoV-2 and has demonstrated *in vitro* suppression of viral inhibition [74]. Our methods reported the IMPD as a promising hit for therapies against all SARS-CoV-2 variants with 49.9% virus reduction. Together with merimepodib, DrugBank and BRENDA list ribavirin as an inhibitor with known pharmacological action. Several studies have postulated that ribavirin's mechanism of action lies on various not mutually exclusive pathways [75]. Lastly, with

Table 3. Exemplary selection of already approved drugs and compounds that act against proteins associated with our predicted anti-SARS-CoV-2 target reactions and could possibly used for antiviral therapies. All listed drugs have known pharmacological action and are sorted based on the predicted percentage of virus reduction in the wildtype sequence.

Reaction	EC-Number	Approved drug	Reference (PubMed ID)	Predicted % virus reduction
CTPS1	6.3.4.2	CPEC	10930994 [67]	62.5
GK1	2.7.4.8	Acyclovir	1316735 [68]	62.5
PUNP4	2.4.2.1	Ganciclovir	24107682 [69]	60.5
IMPD	1.1.1.205	Ribavirin	4197928 [70]	49.9

https://doi.org/10.1371/journal.pcbi.1010903.t003

our methods, we identified the purine-nucleoside phosphorylase (PUNP4) for which ganciclovir has known inhibitory effects (Table 3).

Discussion

Studying human metabolism guides the understanding of diverse diseases by determining the cells' health. The existence of high-quality genome-scale reconstructions facilitates systems-based insights into metabolism. As complex organisms, humans embody multiple cell and tissue types, each with different functions and metabolisms, leading to the essential use of cell- or tissue-specific metabolic networks to enable the accurate prediction of the cells' metabolic behavior. Here, we presented pymCADRE, a re-implementation of mCADRE [31] in Python that allows the reconstruction of tissue-specific human models based on human gene expression data and network topology information. Similar to the original mCADRE algorithm, pymCADRE consists of three parts: (1) ranking, (2) consistency check, and (3) pruning, enabling the user to choose between two optimization methods, FVA and FASTCC, to check for model consistency (S3 Fig). We enriched our implementations with data pre-processing scripts that simplify multiple data curation tasks.

We used our tool to create a tissue-specific model of primary Human Bronchial Epithelial Cell (HBEC) to investigate SARS-CoV-2 infections. We used the human metabolic network Recon1 as a generic model to test our tool to avoid high computational costs. When FVA was used, pymCADRE proceeded faster than mCADRE, maintaining the highest possible similarity to the ground truth, i.e., the mCADRE-derived model. The two models showed a reaction overlap of almost 100%, suggesting a substantial similarity between both implementations and demonstrating confidence about the quality of the pymCADRE models. Since we did not modify the initial mCADRE algorithm, the varying amount of reactions in the final tissue-specific models suggests variable performance among built-in functions in COBRA Toolbox [76] and COBRApy [40]. More specifically, we observed divergent results among the two programming languages when FASTCC was employed. In both cases, the function is implemented as described by Vlassis *et al.* [52]; however the pythonic version detected a varying number of blocked reactions after multiple runs. The bug has already been reported and awaits resolve. Additionally, the detected inactive reactions were dissimilar compared to the reactions in the mCADRE model. This was not the case when the COBRApy methods,

```
flux_variability_analysis()
```
or
```
find_blocked_reactions()
```

from the package 🗀 cobra.flux_analysis were employed. Moreover, the current version of FVA in MATLAB only supports the industrial proprietary CPLEX versions older than V 12.10 [77]. The latest solver release, V 20.1 (released in December, 2020, does not yet include MATLAB-related binaries, and hence, FVA from the COBRAToolbox is of restricted use. This problem is resolved by pymCADRE, as the latest version of COBRApy enables the users to choose among the open-source GLPK package and the CPLEX solver from IBM to perform optimization tasks. Another reason for the divergent performance among both tools could be the implementation of organic exchange/demand reactions detection. We achieved this in a more powerful and fully automated script. Thus, pymCADRE detected four additional organic exchange/demand reactions in Recon1, affecting the result of consistency checks. The utilized human generic model, Recon1, does not include a BOF. We updated the generic human model by including a BOF extracted from the macrophage model published by Bordbar *et al.* [36].

Furthermore, we used our model and simulated an infection with the SARS-CoV-2 to better understand the host's impact on the virus and vice versa. For this purpose, we generated a SARS-CoV-2 VBOF based on the protocol of Aller *et al.* to create an integrated metabolic model aiming the analysis of host-virus interactions and the identification of effective targets for antiviral therapeutic strategies [10]. They recovered already known antiviral targets for the Chikungunya, Dengue, and Zika viruses within the human macrophage cell, verifying their approach's robustness. As Aller *et al.* suggested, FBA and FVA can be employed to predict essential host reactions, especially in cases of novel viruses. Two different computational experiments achieved this: single-reaction knock-outs and host-derived enforcement. Both approaches verified GK1 as a target to restrict the SARS-CoV-2 growth without harming the host. GK1 has already been reported to show inhibitory effects in the macrophage and the lung models [24, 47]. Similarly, our results confirmed enzymatic hits from the glycolysis (ENO, PGM, and PGK) that have been previously described in the literature [24].

However, our methods revealed further novel targets with promising results. CTPS1 constrained, resulted in a remarkable suppression of the viral replication in the cell, similar to GK1. The pre-print from Rao *et al.* experimentally describes how the SARS-CoV-2 activates CTPS1 deploying several proteins to induce the synthesis of CTP, suppressing thus the interferon production and downstream immune response [78]. The authors suggest targeting the inhibition of the host enzyme CTPS1 as a potential approach to restore the interferon induction and, therefore, to hinder the SARS-CoV-2 replication. Notwithstanding, CPEC has been previously described to exhibit antiviral and anti-tumor effects. More specifically, it is known to be active against a wide range of RNA and DNA viruses (e.g., influenza, herpes simplex, and yellow fever) *in vitro* [79, 80]. Similarly, modulating the pyrimidine ribonucleotide levels has been a widely studied approach in treating cancer. As of today, it has been examined most extensively in leukemic cell lines, but also in the context of colorectal carcinoma, brain tumor, and neuroblastoma [81]. Although dosis-related hypotension events occurred in patients with colon carcinoma treated with CPEC in Phase I trials [82], the cardiotoxicity could not be reproduced in established rat models [83]. Studies have reported the deaminated product of CPEC, CPEU, as well as cytidine as potential modulators of the cytotoxic activity of CPEC [84, 85]. It remains to be investigated *in vivo* to which extent it is possible to establish antiviral activities with CPEC without toxic side-effects, but also in combination with other drugs.

Most of our hits fall into the purine and pyrimidine metabolism and are tightly coupled. This implies and verifies that drugs targeting the nucleotide metabolism exemplify a common therapeutic strategy to restrict SARS-CoV-2 replication. We conducted extensive literature and database search and found acyclovir that targets GK1 from the purine synthesis pathway. So far, acyclovir is the standard gold treatment of infections with the herpes virus and the Varicella-Zoster Virus (VZV) [7, 86]. In the context of SARS-CoV-2, acyclovir has been proposed in studies as a drug with an antiviral potential against coronaviruses [87], more specifically SARS-CoV-2 concurrently with signs of reactivation of VZV [88]. The authors assumed that this reactivation is coupled to the unusually low count of lymphocytes (lymphopenia) in the COVID-19 patients' blood. Its mechanism of action resembles molnupiravir, which has been granted the Food and Drug Administration (FDA)-Emergency Use Authorization against SARS-CoV-2 infections [89]. Both drugs target the viral replication by mimicking ribonucleosides and causing mutagenic effects. Compared to acyclovir, which leads to immediate chain termination, molnupiravir continues incorporating of nucleobases until a mismatch occurs, resulting in an error catastrophe. The only FDA-approved drug called remdesivir acts similarly and is an ATP analog and causes delayed chain termination. Hence, acyclovir's mechanism of action indicates a high potential for successful use against SARS-CoV-2 infections. Intravenous ritonavir-boosted nirmatrelvir (Paxlovid) has also received the Emergency Use

Authorizations by FDA medication used in COVID-19 patients. Antiviral effects occur in an earlier stage as it prevents viral replication by inhibiting protein synthesis. Its disadvantage is that it may cause adverse effects upon drug-drug interactions since ritonavir can be dangerous for patients taking antiarrhythmics, blood thinners, and further medications [90]. Nevertheless, several monoclonal Antibodies (mABs), such as bebtelovimab, bamlanivimab-etesevimab, and tocilizumab, have been authorized for intravenous administration and subsequently revised with the emergence of the Omicron variant [91]. On the contrary, acyclovir can be administered orally, making it easier for self-use.

Besides that, we predicted enzymatic candidate targets from the carbohydrate metabolism. In more detail, reactions from the amino/nucleotide sugar and sucrose metabolism demonstrated higher antiviral effects than targets from the glycolysis or the fructose and mannose metabolism. Among others, carbohydrates are essential components of viral particles, with some playing a crucial role in their attachment and penetration into host cells [92]. They have been extensively studied as therapeutic targets against viral infections, while two of the already FDA-approved drugs to treat SARS-CoV-2 [93], remdesivir and molnupiravir, belong into the class of carbohydrate-based antiviral drugs. Our results demonstrated that targeting the path leading to the production of the sialic acid N-Acetyl-D-mannosamine (ManNAc), could result in up to 57.9% SARS-CoV-2 inhibition.

Moreover, we tested two different growth media to validate the robustness of our predicted targets. GK1 was shown to be more effective against the virus with the blood medium defined, compared to the minimal defined medium. Using both media, NDPK2 demonstrated the same inhibitory effect as UMPK5, while nucleoside diphosphate kinase 3 (NDPK3) constrained with the minimal medium showed a higher effect on virus replication.

We further validated the robustness of our host-based targets against all five variants of concern (Alpha, Beta, Gamma, Delta, and Omicron). To accelerate the VBOF reconstruction, we developed PREDICATE to analyze multiple sequences for a single variant rapidly and in an automated way. Within this tool, we also implemented an algorithm to modify reference protein sequences and introduce amino acid mutations. Our implementations are transferable to all RNA viruses, as they are composed of the same building blocks. Firstly, we evaluated the mutations' effect on the computed stoichiometric coefficients variant-wise for the corresponding mutations. The high stoichiometric coefficients for ATP and ADP are consequences of decreased total viral molar masses and increased total amino acid counts. We observed increased use of lysine in the Omicron variant because most mutations replace amino acids with lysine. The opposite effect was observed in Omicron for asparagine. All single-reaction deletions across all variants highlighted NDPK1 as a potential robust antiviral inhibitor. The NDPK1 also proved by HDE to have the highest inhibitory effect against SARS-CoV-2, without harming the host cell. Besides that, supplementation of L-histidine, L-threonine, L-lysine, L-proline, and L-tryptophan in the host's environment shown to interrupt the virus's growth in all five SARS-CoV-2 variants.

Future improvements need to be done to make pymCADRE computationally feasible with more complex and more comprehensive models, including Recon2.2 [94] and Recon3D [30]. Currently, pymCADRE and mCADRE need a large amount of computational time to complete the ranking of reactions when a more complex generic model, like Recon3D, is used. Both tools are automatically killed during pruning as there is no sufficient memory for them to process further reactions. However, we used Recon3D to fill missing knowledge in our model Recon1-HBEC. Our targets' effectiveness needs to be verified in more updated networks that better represent the human metabolism. So far, we tested the results of our pipeline using gene expression data from cell lines originating from primary cells that are easier to handle and analyze. With these, we verified targets already described for SARS-CoV-2 and ensured prediction

accuracy. Further future target validation step would be to employ RNA-Seq data of primary cells that retain more traits of living cells and capture the entire transcriptome, consequently, the gene and transcript abundance. This will enable the detection of further unknown enzymatic targets guiding novel antiviral therapies.

Our integrated bronchial-specific metabolic model could be further expanded and investigated regarding the consequences of any upcoming mutation in the predicting antiviral targets. Models created by pymCADRE could be utilized to simulate the interaction of bacterial pathogens or symbionts and detect potential antiviral targets for drugs against emerging viruses on different host cells quickly. This new software provides the basis for systematic studies of a wide range of integrated computer models for host-pathogen interaction. It reduces the time for creating such models maintaining the highest possible similarity compared to the ground truth model. Our methods are based on the metabolic fluxes of infected cells and the interactions between the host cell and the virus. The latter remain unaffected by evolutionary changes. This, together with the fact that virus replication generally depends on conserved cellular pathways, drastically increases the likelihood of identifying druggable targets with broad antiviral activity. In addition, our predicted host-based targets are derived based on human patient data increasing thus their clinical relevance and their potential to achieve higher efficacy in COVID-19 therapies. Our database-derived drug compounds are experimentally supported and have already been suggested for other single-stranded RNA viruses, opening up the potential of experimentally verifying their safety, toxicity, and efficacy in cell culture experiments and in *in vitro* assays. Moreover, their optimum dosage and route of administration at different infection stages must be determined, since metabolic approaches do not consider that.

Altogether, we propose a complete workflow to create constraint-specific models and use them to predict host-based antiviral targets based on metabolic changes in infected cells. Targeting the host cell metabolic pathways has the benefit of robustness and evolutionary stability while it enables the re-purposing of already available drugs and leads to broad-spectrum putative therapeutics. For some viral infections, such as those caused by enveloped viruses, e.g., HIV, hepatitis B, or the human papillomavirus, it can be effective to target viral proteins with enzymatic activity (e.g., the protease or viral polymerase). However, focusing on viral proteins enhances the evolution of resistance, mainly when used in monotherapy, while new variants carry underlying resistances. Additionally, these direct-acting antivirals are highly virus-specific, preventing from pan-viral efficacy and hindering pandemic preparedness. With that, targeting the host's metabolism using our approaches restrains the emergence of resistance. It reveals host pathways and enzymes essential for viral replication but dispensable for cellular maintenance and survival. Our pipeline has the advantage that applies to all RNA viruses that infect host cells, remarkably reduces the duration of target identification and compound selection, and accelerates the pre-clinical phase. Focusing on the metabolic changes of infected cells, we aim to apply our methods for rapid identification of potential antiviral targets to efficiently prevent future pandemics concerning various viruses and host cell types, promoting pandemic preparedness.

Supporting information

S1 Fig. Results of the host-derived enforcement after defining the minimal growth medium. After constraining the fluxes of `NDPK2` and `UMPK5`, 49.8% of the initial virus remained in the host. Compared to the blood medium, these targets proved to have a greater impact on the virus growth leading to a higher decrease than `GK1`.
(TIF)

S2 Fig. Overview of the evidence-based ranking of reactions in pymCADRE. The evidence-based ranking of reactions in pymCADRE is conducted similarly to mCADRE and consists of three main parts: (A) After binarizing tissue-specific data, the frequency of a gene's expression across all experiments of the same tissue is computed; this is the ubiquity score $U(g)$ for each gene g. The expression-based evidence $E_x(r)$ is computed for each gene-associated reaction r from ubiquity scores. Reactions with a sufficiently high $E_x(r)$ value are denoted as core reactions. Non-active reactions have zero expression-based evidence. (B) Non-core reactions are ranked based on the connectivity-based evidence $E_c(r)$, using the generic models' network topology and the weighted influence $WI(r)$. Figure re-created from Wang *et al* [31].
(TIF)

S3 Fig. Hierarchical organization of the pymCADRE code and its dependencies. The three main scripts are colored with purple, while intermediate scripts are orange-colored. First of all, the 🗀 `rank_reactions.py` module is executed, followed by 🗀 `prune_model.py`. The module 🗀 `check_model_function.py` is connected to main and intermediate scripts and is used multiple times within a single run. Figure created with yEd [95].
(TIF)

S4 Fig. Categorization of the compounds needed for the growth of SARS-CoV-2. The VBOF includes totally four nucleotides, five energy-related metabolites, 20 proteinogenic amino acids, and six fatty acids.
(TIF)

S5 Fig. Hits from the host-derived enforcement with inhibitory effect across all examined variants of concern. Only hits shared by all virus variants are displayed. The range and effect of reaction inhibitions on the VBOF were calculated while keeping the host's maintenance at 100%.
(TIF)

S1 Table. Overview of compounds and their stoichiometric coefficients in the host and viral biomass functions together with the energy-generating compounds. From the listed metabolites, `adp_c`, `h_c`, `pi_c` and `ppi_c` are the reaction products, while the rest the reactants.
(XLSX)

S2 Table. Detailed information of all antiviral targets predicted using the host-derived enforcement (HDE).
(XLSX)

S3 Table. Summary of datasets used for all variants of concern. All variants are listed along with their GISAID accession number, the associated mutations and submission details.
(XLSX)

S4 Table. The stoichiometric coefficients all the molecules included in the VBOFs created for all five examined variants of concern.
(XLSX)

S5 Table. Growth media definitions.
(XLSX)

S6 Table. Amino acids and their three-letter and one-letter codes, and their molecular weight used to construct the SARS-CoV-2 VBOF. The molecular weights were derived from the Chemicals of Biological Interest (ChEBI) database [96].
(XLSX)

S7 Table. Five-number summary of reaction fluxes in host and virus. The summary consists of five values: minimum, first quartile (25th percentile), median (50th percentile), third quartile (75th percentile), and maximum.
(XLSX)

S1 File. Python script of the PREDICATE tool written in a Jupyter Notebook format.
(IPYNB)

S2 File. The SBML model of the integrated host-SARS-CoV-2 bronchial epithelial cell.
(XML)

Acknowledgments

Author Contributions

Conceptualization: Nantia Leonidou, Alina Renz, Reihaneh Mostolizadeh, Andreas Dräger.

Data curation: Nantia Leonidou.

Formal analysis: Nantia Leonidou.

Funding acquisition: Andreas Dräger.

Investigation: Nantia Leonidou.

Methodology: Nantia Leonidou.

Project administration: Andreas Dräger.

Resources: Nantia Leonidou.

Software: Nantia Leonidou.

Supervision: Andreas Dräger.

Validation: Nantia Leonidou.

Visualization: Nantia Leonidou.

Writing – original draft: Nantia Leonidou.

Writing – review & editing: Nantia Leonidou, Alina Renz, Reihaneh Mostolizadeh, Andreas Dräger.

References

1. Cheng VC, Lau SK, Woo PC, Yuen KY. Severe acute respiratory syndrome coronavirus as an agent of emerging and reemerging infection. Clinical microbiology reviews. 2007; 20(4):660–694. https://doi.org/10.1128/CMR.00023-07 PMID: 17934078

2. Taubenberger JK, Morens DM. 1918 Influenza: the mother of all pandemics. Revista Biomedica. 2006; 17(1):69–79. https://doi.org/10.3201/eid1201.050979 PMID: 16494711

3. Ryu WS. Part III. RNA Viruses. In: Ryu WS, editor. Molecular Virology of Human Pathogenic Viruses. Boston: Academic Press; 2017. p. 149–150.

4. Maynard ND, Gutschow MV, Birch EW, Covert MW. The virus as metabolic engineer. Biotechnology journal. 2010; 5(7):686–694. https://doi.org/10.1002/biot.201000080 PMID: 20665642

5.   Leyssen P, De Clercq E, Neyts J. Molecular strategies to inhibit the replication of RNA viruses. Antiviral research. 2008; 78(1):9–25. https://doi.org/10.1016/j.antiviral.2008.01.004 PMID: 18313769

6.   Feld JJ, Hoofnagle JH. Mechanism of action of interferon and ribavirin in treatment of hepatitis C. Nature. 2005; 436(7053):967–972. https://doi.org/10.1038/nature04082 PMID: 16107837

7.   Engel JP, Englund JA, Fletcher CV, Hill EL. Treatment of resistant herpes simplex virus with continuous-infusion acyclovir. Jama. 1990; 263(12):1662–1664. https://doi.org/10.1001/jama.1990.03440120084042 PMID: 2308204

8.   Warren TK, Jordan R, Lo MK, Ray AS, Mackman RL, Soloveva V, et al. Therapeutic efficacy of the small molecule GS-5734 against Ebola virus in rhesus monkeys. Nature. 2016; 531(7594):381–385. https://doi.org/10.1038/nature17180 PMID: 26934220

9.   Maynard ND, Birch EW, Sanghvi JC, Chen L, Gutschow MV, Covert MW. A forward-genetic screen and dynamic analysis of lambda phage host-dependencies reveals an extensive interaction network and a new anti-viral strategy. PLoS genetics. 2010; 6(7):e1001017. https://doi.org/10.1371/journal.pgen.1001017 PMID: 20628568

10. Aller S, Scott A, Sarkar-Tyson M, Soyer OS. Integrated human-virus metabolic stoichiometric modelling predicts host-based antiviral targets against Chikungunya, Dengue and Zika viruses. Journal of The Royal Society Interface. 2018; 15(146):20180125. https://doi.org/10.1098/rsif.2018.0125 PMID: 30209043

11. Smith EC. The not-so-infinite malleability of RNA viruses: Viral and cellular determinants of RNA virus mutation rates. PLoS pathogens. 2017; 13(4):e1006254. https://doi.org/10.1371/journal.ppat.1006254 PMID: 28448634

12. Drake JW. Rates of spontaneous mutation among RNA viruses. Proceedings of the National Academy of Sciences. 1993; 90(9):4171–4175. https://doi.org/10.1073/pnas.90.9.4171 PMID: 8387212

13. Bar-On YM, Flamholz A, Phillips R, Milo R. Science Forum: SARS-CoV-2 (COVID-19) by the numbers. elife. 2020; 9:e57309. https://doi.org/10.7554/eLife.57309 PMID: 32228860

14. Domingo E, Holland J. RNA virus mutations and fitness for survival. Annual review of microbiology. 1997; 51(1):151–178. https://doi.org/10.1146/annurev.micro.51.1.151 PMID: 9343347

15. Robson F, Khan KS, Le TK, Paris C, Demirbag S, Barfuss P, et al. Coronavirus RNA Proofreading: Molecular Basis and Therapeutic Targeting. Molecular cell. 2020; 79(5):710–727. https://doi.org/10.1016/j.molcel.2020.07.027 PMID: 32853546

16. Kim D, Lee JY, Yang JS, Kim JW, Kim VN, Chang H. The architecture of SARS-CoV-2 transcriptome. Cell. 2020; 181(4):914–921. https://doi.org/10.1016/j.cell.2020.04.011 PMID: 32330414

17. World Health Organization. COVID-19 weekly epidemiological update 76– 25 January 2022;. Available from: https://www.who.int/publications/m/item/weekly-epidemiological-update-on-covid-19---25-january-2022.

18. Korber B, Fischer WM, Gnanakaran S, Yoon H, Theiler J, Abfalterer W, et al. Tracking changes in SARS-CoV-2 spike: evidence that D614G increases infectivity of the COVID-19 virus. Cell. 2020; 182 (4):812–827. https://doi.org/10.1016/j.cell.2020.06.043 PMID: 32697968

19. Timmers LFSM, Peixoto JV, Ducati RG, Bachega JFR, de Mattos Pereira L, Caceres RA, et al. SARS-CoV-2 mutations in Brazil: from genomics to putative clinical conditions. Scientific reports. 2021; 11 (1):1–14. https://doi.org/10.1038/s41598-021-91585-6 PMID: 34099808

20. Kannan SR, Spratt AN, Sharma K, Chand HS, Byrareddy SN, Singh K. Omicron SARS-CoV-2 variant: Unique features and their impact on pre-existing antibodies. Journal of autoimmunity. 2022; 126:102779. https://doi.org/10.1016/j.jaut.2021.102779 PMID: 34915422

21. Becker SA, Palsson BO. Context-specific metabolic networks are consistent with experiments. PLoS computational biology. 2008; 4(5):e1000082. https://doi.org/10.1371/journal.pcbi.1000082 PMID: 18483554

22. Mayer KA, Stöckl J, Zlabinger GJ, Gualdoni GA. Hijacking the supplies: metabolism as a novel facet of virus-host interaction. Frontiers in immunology. 2019; p. 1533. https://doi.org/10.3389/fimmu.2019.01533 PMID: 31333664

23. Renz A, Widerspick L, Dräger A. FBA reveals guanylate kinase as a potential target for antiviral therapies against SARS-CoV-2. Bioinformatics. 2020; 36(Supplement_2):i813–i821. https://doi.org/10.1093/bioinformatics/btaa813 PMID: 33381848

24. Delattre H, Sasidharan K, Soyer OS. Inhibiting the reproduction of SARS-CoV-2 through perturbations in human lung cell metabolic network. Life science alliance. 2021; 4(1). https://doi.org/10.26508/lsa.202000869 PMID: 33234678

25. Nanda P, Ghosh A. Genome Scale-Differential Flux Analysis reveals deregulation of lung cell metabolism on SARS-CoV-2 infection. PLoS computational biology. 2021; 17(4):e1008860. https://doi.org/10.1371/journal.pcbi.1008860 PMID: 33835998

26. Cheng K, Martin-Sancho L, Pal LR, Pu Y, Riva L, Yin X, et al. Genome-scale metabolic modeling reveals SARS-CoV-2-induced metabolic changes and antiviral targets. Molecular systems biology. 2021; 17(11):e10260. https://doi.org/10.15252/msb.202110260 PMID: 34709707

27. Bannerman BP, Júlvez J, Oarga A, Blundell TL, Moreno P, Floto RA. Integrated human/SARS-CoV-2 metabolic models present novel treatment strategies against COVID-19. Life science alliance. 2021; 4 (10). https://doi.org/10.26508/lsa.202000954 PMID: 34353886

28. Kishk A, Pacheco MP, Sauter T. DCcov: Repositioning of drugs and drug combinations for SARS-CoV-2 infected lung through constraint-based modeling. Iscience. 2021; 24(11):103331. https://doi.org/10.1016/j.isci.2021.103331 PMID: 34723158

29. Duarte NC, Becker SA, Jamshidi N, Thiele I, Mo ML, Vo TD, et al. Global reconstruction of the human metabolic network based on genomic and bibliomic data. Proceedings of the National Academy of Sciences. 2007; 104(6):1777–1782. https://doi.org/10.1073/pnas.0610772104 PMID: 17267599

30. Brunk E, Sahoo S, Zielinski DC, Dräger A, Mih N, Gatto F, et al. Recon3D enables a three-dimensional view of gene variation in human metabolism. Nature Biotechnology. 2018; 36:272–281. https://doi.org/10.1038/nbt.4072 PMID: 29457794

31. Wang Y, Eddy JA, Price ND. Reconstruction of genome-scale metabolic models for 126 human tissues using mCADRE. BMC systems biology. 2012; 6(1):1–16. https://doi.org/10.1186/1752-0509-6-153 PMID: 23234303

32. MATLAB. version R2020a. Natick, Massachusetts: The MathWorks Inc.; 2020.

33. Kam YW, Okumura Y, Kido H, Ng LFP, Bruzzone R, Altmeyer R. Cleavage of the SARS coronavirus spike glycoprotein by airway proteases enhances virus entry into human bronchial epithelial cells *in vitro*. PloS one. 2009; 4(11):e7870. https://doi.org/10.1371/journal.pone.0007870 PMID: 19924243

34. Ravindra NG, Alfajaro MM, Gasque V, Huston NC, Wan H, Szigeti-Buck K, et al. Single-cell longitudinal analysis of SARS-CoV-2 infection in human airway epithelium identifies target cells, alterations in gene expression, and cell state changes. PLoS biology. 2021; 19(3):e3001143. https://doi.org/10.1371/journal.pbio.3001143 PMID: 33730024

35. Ryu G, Shin HW. SARS-CoV-2 infection of airway epithelial cells. Immune network. 2021; 21(1). https://doi.org/10.4110/in.2021.21.e3 PMID: 33728096

36. Bordbar A, Lewis NE, Schellenberger J, Palsson BØ, Jamshidi N. Insight into human alveolar macro-phage and M. tuberculosis interactions via metabolic reconstructions. Molecular systems biology. 2010; 6(1):422. https://doi.org/10.1038/msb.2010.68 PMID: 20959820

37. Python Package Index—PyPI;. https://pypi.org/.

38. Ross AJ, Dailey LA, Brighton LE, Devlin RB. Transcriptional profiling of mucociliary differentiation in human airway epithelial cells. American journal of respiratory cell and molecular biology. 2007; 37 (2):169–185. https://doi.org/10.1165/rcmb.2006-0466OC PMID: 17413031

39. Norsigian CJ, Pusarla N, McConn JL, Yurkovich JT, Dräger A, Palsson BO, et al. BiGG Models 2020: multi-strain genome-scale models and expansion across the phylogenetic tree. Nucleic Acids Research. 2019; 48(D1).

40. Ebrahim A, Lerman JA, Palsson BO, Hyduke DR. COBRApy: constraints-based reconstruction and analysis for python. BMC systems biology. 2013; 7(1):1–6. https://doi.org/10.1186/1752-0509-7-74 PMID: 23927696

41. Bernardes JP, Mishra N, Tran F, Bahmer T, Best L, Blase JI, et al. Longitudinal multi-omics analyses identify responses of megakaryocytes, erythroid cells, and plasmablasts as hallmarks of severe COVID-19. Immunity. 2020; 53(6):1296–1314. https://doi.org/10.1016/j.immuni.2020.11.017 PMID: 33296687

42. Kanehisa M, Sato Y, Kawashima M. KEGG mapping tools for uncovering hidden features in biological data. Protein Science. 2021. https://doi.org/10.1002/pro.4172 PMID: 34423492

43. Fritzemeier CJ, Hartleb D, Szappanos B, Papp B, Lercher MJ. Erroneous energy-generating cycles in published genome scale metabolic networks: Identification and removal. PLoS computational biology. 2017; 13(4):e1005494. https://doi.org/10.1371/journal.pcbi.1005494 PMID: 28419089

44. Lieven C, Beber ME, Olivier BG, Bergmann FT, Ataman M, Babaei P, et al. MEMOTE for standardized genome-scale metabolic model testing. Nature biotechnology. 2020; 38(3):272–276. https://doi.org/10.1038/s41587-020-0446-y PMID: 32123384

45. Bornstein BJ, Keating SM, Jouraku A, Hucka M. LibSBML: an API library for SBML. Bioinformatics. 2008; 24(6):880–881. https://doi.org/10.1093/bioinformatics/btn051 PMID: 18252737

46. Geer LY, Marchler-Bauer A, Geer RC, Han L, He J, He S, et al. The NCBI BioSystems database. Nucleic acids research. 2010; 38(suppl_1):D492–D496. https://doi.org/10.1093/nar/gkp858 PMID: 19854944

47.   Renz A, Widerspick 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(6). https://doi.org/10.3390/genes12060796 PMID: 34073716

48.   Baltimore D. Expression of animal virus genomes. Bacteriological reviews. 1971; 35(3):235–241. https://doi.org/10.1128/br.35.3.235-241.1971 PMID: 4329869

49.   Khare S, Gurry C, Freitas L, Schultz MB, Bach G, Diallo A, et al. GISAID's Role in Pandemic Response. China CDC Weekly. 2021; 3(49):1049. https://doi.org/10.46234/ccdcw2021.255 PMID: 34934514

50.   Panchiwala H, Shah S, Planatscher H, Zakharchuk M, König M, Dräger A. The Systems Biology Simulation Core Library. Bioinformatics. 2022; 38:864–865. https://doi.org/10.1093/bioinformatics/btab669 PMID: 34554191

51.   Gudmundsson S, Thiele I. Computationally efficient flux variability analysis. BMC bioinformatics. 2010; 11(1):1–3. https://doi.org/10.1186/1471-2105-11-489 PMID: 20920235

52.   Vlassis N, Pacheco MP, Sauter T. Fast reconstruction of compact context-specific metabolic network models. PLoS computational biology. 2014; 10(1):e1003424. https://doi.org/10.1371/journal.pcbi.1003424 PMID: 24453953

53.   Romero P, Wagg J, Green ML, Kaiser D, Krummenacker M, Karp PD. Computational prediction of human metabolic pathways from the complete human genome. Genome biology. 2005; 6(1):1–17. https://doi.org/10.1186/gb-2004-6-1-r2 PMID: 15642094

54.   Hucka M, Bergmann FT, Chaouiya C, Dräger A, Hoops S, Keating SM, et al. Systems Biology Markup Language (SBML): Language Specification for Level 3 Version 2 Core Release 2. Journal of Integrative Bioinformatics. 2019; 16(2):1. https://doi.org/10.1515/jib-2019-0021 PMID: 31219795

55.   Rodriguez N, Pettit JB, Dalle Pezze P, Li L, Henry A, van Iersel MP, et al. The systems biology format converter. BMC bioinformatics. 2016; 17(1):1–7. https://doi.org/10.1186/s12859-016-1000-2 PMID: 27044654

56.   Courtot M, Juty N, Knüpfer C, Waltemath D, Zhukova A, Dräger A, et al. Controlled vocabularies and semantics in systems biology. Molecular Systems Biology. 2011; 7(1):543. https://doi.org/10.1038/msb.2011.77 PMID: 22027554

57.   Balcı H, Siper MC, Saleh N, Safarli I, Roy L, Kılıçarslan M, et al. Newt: a comprehensive web-based tool for viewing, constructing and analyzing biological maps. Bioinformatics. 2021; 37(10):1475–1477. https://doi.org/10.1093/bioinformatics/btaa850 PMID: 33010165

58.   Touré V, Dräger A, Luna A, Dogrusoz U, Rougny A. The Systems Biology Graphical Notation: Current Status and Applications in Systems Medicine. In: Wolkenhauer O, editor. Systems Medicine. vol. 3. Oxford: Academic Press; 2020. p. 372–381.

59.   Keating SM, Waltemath D, König M, Zhang F, Dräger A, Chaouiya C, et al. SBML Level 3: an extensible format for the exchange and reuse of biological models. Molecular Systems Biology. 2020; 16(8): e9110. https://doi.org/10.15252/msb.20199110 PMID: 32845085

60.   Malik-Sheriff RS, Glont M, Nguyen TVN, Tiwari K, Roberts MG, Xavier A, et al. BioModels—15 years of sharing computational models in life science. Nucleic Acids Research. 2020; 48:D407–D415. https://doi.org/10.1093/nar/gkz1055 PMID: 31701150

61.   Olivier BG, Bergmann FT. SBML Level 3 Package: Flux Balance Constraints version 2. Journal of Integrative Bioinformatics. 2018; 15(1):20170082. https://doi.org/10.1515/jib-2017-0082 PMID: 29522419

62.   Bergmann FT, Adams R, Moodie S, Cooper J, Glont M, Golebiewski M, et al. COMBINE archive and OMEX format: one file to share all information to reproduce a modeling project. BMC Bioinformatics. 2014; 15:369. https://doi.org/10.1186/s12859-014-0369-z PMID: 25494900

63.   Neal ML, König M, Nickerson D, Mısırlı G, Kalbasi R, Dräger A, et al. Harmonizing semantic annotations for computational models in biology. Briefings in Bioinformatics. 2018; 20(2):540–550. https://doi.org/10.1093/bib/bby087

64.   Kumar S, Thambiraja TS, Karuppanan K, Subramaniam G. Omicron and Delta variant of SARS-CoV-2: A comparative computational study of spike protein. Journal of medical virology. 2021. PMID: 34914115

65.   Jeske L, Placzek S, Schomburg I, Chang A, Schomburg D. BRENDA in 2019: a European ELIXIR core data resource. Nucleic acids research. 2019; 47(D1):D542–D549. https://doi.org/10.1093/nar/gky1048 PMID: 30395242

66.   Wishart DS, Knox C, Guo AC, Shrivastava S, Hassanali M, Stothard P, et al. DrugBank: a comprehensive resource for in silico drug discovery and exploration. Nucleic acids research. 2006; 34(suppl_1): D668–D672. https://doi.org/10.1093/nar/gkj067 PMID: 16381955

67.   Verschuur AC, Van Gennip AH, Leen R, Meinsma R, Voute P, Van Kuilenburg AB. *In vitro* inhibition of cytidine triphosphate synthetase activity by cyclopentenyl cytosine in paediatric acute lymphocytic leukaemia. British journal of haematology. 2000; 110(1):161–169. https://doi.org/10.1046/j.1365-2141.2000.02136.x PMID: 10930994

68. O'Brien JJ, Campoli-Richards DM. Acyclovir. Drugs. 1989; 37(3):233–309. https://doi.org/10.2165/00003495-198937030-00002

69. Furihata T, Kishida S, Sugiura H, Kamiichi A, Iikura M, Chiba K. Functional analysis of purine nucleoside phosphorylase as a key enzyme in ribavirin metabolism. Drug Metabolism and Pharmacokinetics. 2014; 29(2):211–214. https://doi.org/10.2133/dmpk.DMPK-13-NT-065 PMID: 24107682

70. Streeter DG, Witkowski J, Khare GP, Sidwell RW, Bauer RJ, Robins RK, et al. Mechanism of action of 1-*β*-D-ribofuranosyl-1, 2, 4-triazole-3-carboxamide (Virazole), a new broad-spectrum antiviral agent. Proceedings of the National Academy of Sciences. 1973; 70(4):1174–1178. https://doi.org/10.1073/pnas.70.4.1174 PMID: 4197928

71. Clercq ED. Antiviral activity spectrum and target of action of different classes of nucleoside analogues. Nucleosides, Nucleotides & Nucleic Acids. 1994; 13(6-7):1271–1295. https://doi.org/10.1080/15257779408012151

72. Elion GB. Mechanism of action and selectivity of acyclovir. The American journal of medicine. 1982; 73 (1):7–13. https://doi.org/10.1016/0002-9343(82)90055-9 PMID: 6285736

73. Tong X, Smith J, Bukreyeva N, Koma T, Manning JT, Kalkeri R, et al. Merimepodib, an IMPDH inhibitor, suppresses replication of Zika virus and other emerging viral pathogens. Antiviral research. 2018; 149:34–40. https://doi.org/10.1016/j.antiviral.2017.11.004 PMID: 29126899

74. Bukreyeva N, Mantlo EK, Sattler RA, Huang C, Paessler S, Zeldis J. The IMPDH inhibitor merimepodib suppresses SARS-CoV-2 replication *in vitro*. BioRxiv. 2020; https://doi.org/10.1101/2020.04.07.028589

75. Te HS, Randall G, Jensen DM. Mechanism of action of ribavirin in the treatment of chronic hepatitis C. Gastroenterology & hepatology. 2007; 3(3):218. PMID: 21960835

76. Heirendt L, Arreckx S, Pfau T, Mendoza SN, Richelle A, Heinken A, et al. Creation and analysis of bio-chemical constraint-based models using the COBRA Toolbox v. 3.0. Nature protocols. 2019; 14 (3):639–702. https://doi.org/10.1038/s41596-018-0098-2 PMID: 30787451

77. Cplex II. V12. 1: User's Manual for CPLEX. International Business Machines Corporation. 2009; 46 (53):157.

78. Rao Y, Wang TY, Qin C, Espinosa B, Liu Q, Ekanayake A, et al. Targeting CTP synthetase 1 to restore interferon induction and impede nucleotide synthesis in SARS-CoV-2 infection. bioRxiv. 2021. https://doi.org/10.1101/2021.02.05.429959 PMID: 33564769

79. De Clercq E, Murase J, Marquez VE. Broad-spectrum antiviral and cytocidal activity of cyclopentenylcy-tosine, a carbocyclic nucleoside targeted at CTP synthetase. Biochemical pharmacology. 1991; 41 (12):1821–1829. https://doi.org/10.1016/0006-2952(91)90120-T PMID: 1710119

80. Marquez VE, Lim MI, Treanor SP, Plowman J, Priest MA, Markovac A, et al. Cyclopentenylcytosine. A carbocyclic nucleoside with antitumor and antiviral properties. Journal of medicinal chemistry. 1988; 31 (9):1687–1694. https://doi.org/10.1021/jm00117a004 PMID: 3411597

81. Schimmel KJM, Gelderblom H, Guchelaar HJ. Cyclopentenyl cytosine (CPEC): an overview of its *in vitro* and *in vivo* activity. Current cancer drug targets. 2007; 7(5):504–509. https://doi.org/10.2174/156800907781386579 PMID: 17691910

82. Politi PM, Xie F, Dahut W, Ford H, Kelley JA, Bastian A, et al. Phase I clinical trial of continuous infusion cyclopentenyl cytosine. Cancer chemotherapy and pharmacology. 1995; 36(6):513–523. https://doi.org/10.1007/BF00685802 PMID: 7554044

83. Schimmel K, Bennink R, de Bruin K, Leen R, Sand K, van den Hoff M, et al. Absence of cardiotoxicity of the experimental cytotoxic drug cyclopentenyl cytosine (CPEC) in rats. Archives of toxicology. 2005; 79 (5):268–276. https://doi.org/10.1007/s00204-004-0633-5 PMID: 15902424

84. Blaney SM, Balis FM, Grem J, Cole DE, Adamson PC, Poplack DG. Modulation of the cytotoxic effect of cyclopentenylcytosine by its primary metabolite, cyclopentenyluridine. Cancer research. 1992; 52 (12):3503–3505. PMID: 1596909

85. Ford H Jr, Cooney DA, Ahluwalia GS, Hao Z, Rommel ME, Hicks L, et al. Cellular pharmacology of cyclopentenyl cytosine in Molt-4 lymphoblasts. Cancer research. 1991; 51(14):3733–3740. PMID: 1712247

86. Balfour HH Jr, McMonigal KA, Bean B. Acyclovir therapy of varicella-zoster virus infections in immuno-compromised patients. Journal of Antimicrobial Chemotherapy. 1983; 12(suppl_B):169–179. https://doi.org/10.1093/jac/12.suppl_B.169 PMID: 6313596

87. Tan ELC, Ooi EE, Lin CY, Tan HC, Ling AE, Lim B, et al. Inhibition of SARS coronavirus infection *in vitro* with clinically approved antiviral drugs. Emerging infectious diseases. 2004; 10(4):581. https://doi.org/10.3201/eid1004.030458 PMID: 15200845

88. Nofal A, Fawzy MM, Deen SMSE, El-Hawary EE. Herpes zoster ophthalmicus in COVID-19 patients. International Journal of Dermatology. 2020; 59(12):1545. https://doi.org/10.1111/ijd.15240 PMID: 33040343

89. Kabinger F, Stiller C, Schmitzová J, Dienemann C, Kokic G, Hillen HS, et al. Mechanism of molnupiravir-induced SARS-CoV-2 mutagenesis. Nature structural & molecular biology. 2021; 28(9):740–746. https://doi.org/10.1038/s41594-021-00651-0 PMID: 34381216

90. Marzolini C, Kuritzkes DR, Marra F, Boyle A, Gibbons S, Flexner C, et al. Recommendations for the management of drug-drug interactions between the COVID-19 antiviral nirmatrelvir/ritonavir (Paxlovid) and comedications. Clinical Pharmacology & Therapeutics. 2022. https://doi.org/10.1002/cpt.2646 PMID: 35567754

91. Cavazzoni P. Coronavirus (COVID-19) update: FDA limits use of certain monoclonal antibodies to treat COVID-19 due to the Omicron variant. US Food and Drug Administration. 2022.

92. Mathez G, Cagno V. Viruses like sugars: how to assess glycan involvement in viral attachment. Microorganisms. 2021; 9(6):1238. https://doi.org/10.3390/microorganisms9061238 PMID: 34200288

93. Cao X, Du X, Jiao H, An Q, Chen R, Fang P, et al. Carbohydrate-based drugs launched during 2000-2021. Acta Pharmaceutica Sinica B. 2022. https://doi.org/10.1016/j.apsb.2022.05.020 PMID: 36213536

94. Swainston N, Smallbone K, Hefzi H, Dobson PD, Brewer J, Hanscho M, et al. Recon 2.2: from reconstruction to model of human metabolism. Metabolomics. 2016; 12(7):1–7. https://doi.org/10.1007/s11306-016-1051-4 PMID: 27358602

95. yWorks GmbH. yEd. 2019.

96. Hastings J, Owen G, Dekker A, Ennis M, Kale N, Muthukrishnan V, et al. ChEBI in 2016: Improved services and an expanding collection of metabolites. Nucleic acids research. 2016; 44(D1):D1214–D1219. https://doi.org/10.1093/nar/gkv1031 PMID: 26467479

APPENDIX C

Paper II: SBOannotator: a Python tool for the automated assignment of systems biology ontology terms

Systems biology

SBOannotator: a Python tool for the automated assignment of systems biology ontology terms

Nantia Leonidou [iD] [1,2,3,4], Elisabeth Fritze[2], Alina Renz [iD] [1,2,4], Andreas Dräger [iD] [1,2,3,4,*]

[1]Computational Systems Biology of Infections and Antimicrobial-Resistant Pathogens, Institute for Bioinformatics and Medical Informatics (IBMI), Eberhard Karl University of Tübingen, 72076 Tübingen, Germany
[2]Department of Computer Science, Eberhard Karl University of Tübingen, 72076 Tübingen, Germany
[3]German Center for Infection Research (DZIF), partner site Tübingen, Germany
[4]Cluster of Excellence 'Controlling Microbes to Fight Infections,' Eberhard Karl University of Tübingen, 72076 Tübingen, Germany

*Corresponding author. Computational Systems Biology of Infections and Antimicrobial-Resistant Pathogens, Institute for Bioinformatics and Medical Informatics (IBMI), Eberhard Karl University of Tübingen, 72076 Tübingen, Germany. E-mail: andreas.draeger@uni-tuebingen.de

Associate Editor: Pier Luigi Martelli

Abstract

Motivation: The number and size of computational models in biology have drastically increased over the past years and continue to grow. Modeled networks are becoming more complex, and reconstructing them from the beginning in an exchangeable and reproducible manner is challenging. Using precisely defined ontologies enables the encoding of field-specific knowledge and the association of disparate data types. In computational modeling, the medium for representing domain knowledge is the set of orthogonal structured controlled vocabularies named Systems Biology Ontology (SBO). The SBO terms enable modelers to explicitly define and describe model entities, including their roles and characteristics.

Results: Here, we present the first standalone tool that automatically assigns SBO terms to multiple entities of a given SBML model, named the SBOannotator. The main focus lies on the reactions, as the correct assignment of precise SBO annotations requires their extensive classification. Our implementation does not consider only top-level terms but examines the functionality of the underlying enzymes to allocate precise and highly specific ontology terms to biochemical reactions. Transport reactions are examined separately and are classified based on the mechanism of molecule transport. Pseudo-reactions that serve modeling purposes are given reasonable terms to distinguish between biomass production and the import or export of metabolites. Finally, other model entities, such as metabolites and genes, are annotated with appropriate terms. Including SBO annotations in the models will enhance the reproducibility, usability, and analysis of biochemical networks.

Availability and implementation: SBOannotator is freely available from https://github.com/draeger-lab/SBOannotator/.

1 Introduction

Ontologies are used to share common knowledge and its application across communities (Stevens *et al.* 2000). While concepts in biology are adequately covered by appropriate ontologies, model-related semantics are encoded by standardized Systems Biology Ontology (SBO) terms (Courtot *et al.* 2011). The SBO is a set of orthogonal controlled vocabulary terms used to explicitly and unambiguously describe the semantics of model instances. They are divided into eight orthogonal vocabularies and can be employed to annotate a model and describe various entities. For instance, they may represent the type or role of a single component in a model streamlining the understanding and meaning of this entity. The more specific the SBO term is, the more precise the description. As of January 2023, they consist of 694 terms, with 24 newly added in the last 3 years. Generally, such terms ensure model reproducibility and exchangeability as they record and categorize the semantics of model components. From the release of SBML Level 2 Version 2 in the fall of 2006 to the current edition (SBML Level 3 Version 2 Release 2, Hucka *et al.* 2019), the SBML format has supported annotating its components using SBO terms to unambiguously mark their semantics

and extend their scope. At this point, adding general, top-level SBO terms to a model can be done automatically. However, adding precise descriptions for biochemical reactions in constraint-based models, e.g. glycosylation or hydrolysis, remains a laborious and complicated step. After precise categorization, all terms must be determined and added individually to each occurrence. A higher-level SBO term specificity in reactions can enable new model analysis methods. For instance, similar to the gene set enrichment analysis, by counting the occurrence of SBO terms, one could easily deduce the types of over-catalyzed reactions, either for complete models or selected pathways. Hence, their automated assignment is of great importance. Here, we implemented an expert knowledge-driven classification scheme implemented in Python called SBOannotator, which can be easily used to assign SBO terms in a given SBML model (Keating *et al.* 2020) automatically.

2 Results

The SBOannotator workflow comprises six main steps (Fig. 1). At first, all reactions found within the model are labeled as either (i) *transporters* that move molecules across
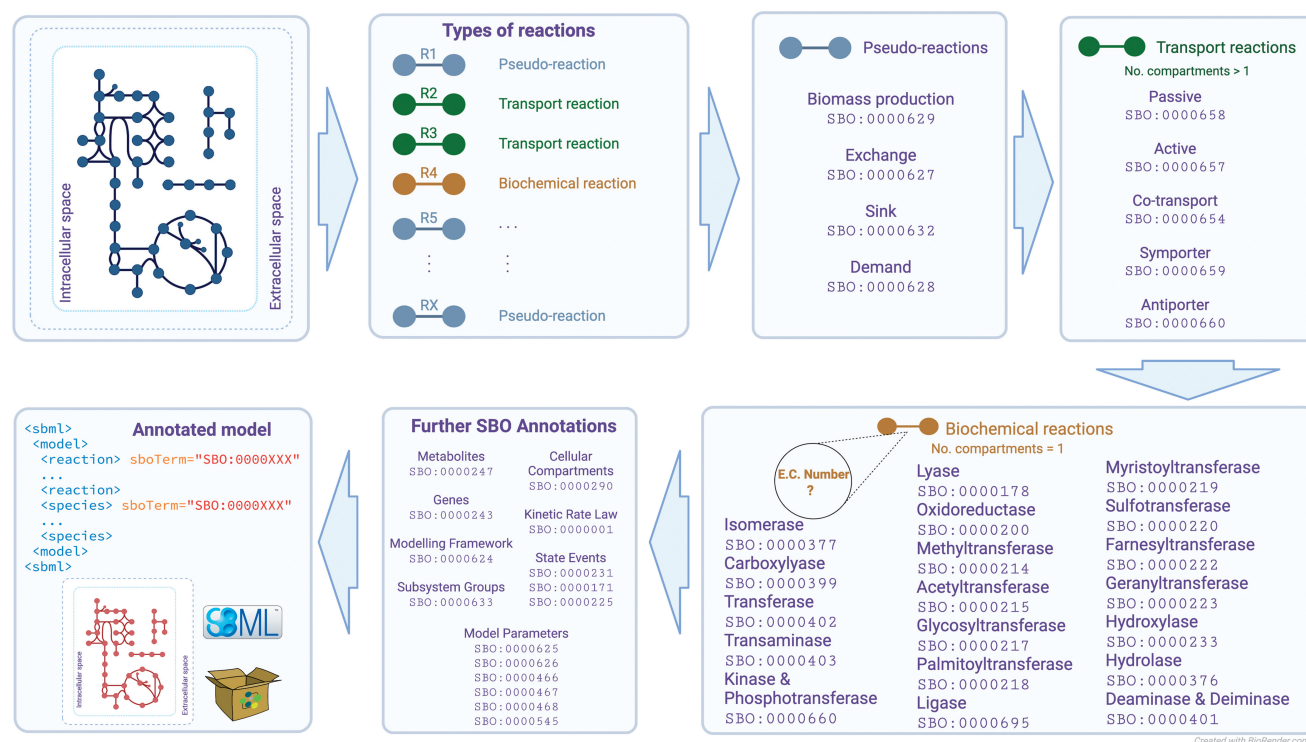
Figure 1. Overview of SBOannotator's pipeline. The software enables the automated assignment of SBO terms to multiple model entities in a given SBML file. First, the pseudo- and transport reactions are characterized and classified. Then, the biochemical reactions are subdivided into 18 classes based on the underlying enzymatic function. Finally, further model elements are annotated by their respective SBO annotations, and the final model is stored in SBML format. The main advantage is the detailed categorization of biochemical reactions and the allocation of specialized terms that precisely capture related and exchangeable information.

different compartments, (ii) *simple biochemical reactions* that only take place in the cytosol, and (iii) *pseudo-reactions* that import or export metabolites and serve modeling purposes. Pseudo-reactions in systems biology modeling do not correspond to any actual physical process and should not be confused with the pseudo-first-order reactions from the field of chemical kinetics. They are subdivided into demand, exchange, and sink reactions. The biomass objective function also belongs to this class. SBOannotator processes further by examining the transport reactions and assigning appropriate SBO terms. The classification mechanism in this step is comparably advanced since several types of transporters exist. The decision relies on the main characteristics of the different classes, such as the presence of one (passive transport) or more reaction participants, and the consumption of adenosine triphosphate (ATP) or phosphoenolpyruvate (PEP) (active transport). The remaining biochemical reactions are processed in the next step to enable more detailed labeling. For this purpose, the SBOannotator employs an Structured Query Language (SQL) database that contains mappings between Enzyme Commission (EC) numbers and the respective SBO terms. As the model's size increases, using an already-defined database accelerates the computational time needed for their annotation. Our mappings could be divided into three main categories: (i) one-to-one mapping; one SBO term represents EC numbers from a single sub-subclass (e.g. transamination); (ii) one-to-few mapping; one SBO term maps only a subset of EC numbers belonging in a single sub-subclass (e.g. myristoylation); and (iii) one-to-many mapping; one SBO term covers a large subset of EC numbers within one sub-subclass (e.g.

acetylation). Supplementary Table S1 lists all mappings in detail. It is important to note that a proper term that describes the ligases (EC class 6) was missing from the SBO graph. This would be necessary to describe, for instance, reactions involving the formation of deoxyribonucleic acid (DNA), ribonucleic acid (RNA), and protein fragments. After contacting the developers of the SBO vocabulary, a new term was introduced (16 May 2023) for ligases that describes the formation of a covalent bonds (SBO:0000695). The SBOannotator is designed to handle models with or without EC numbers assigned. However, they should utilize the Biochemical, Genetical, and Genomical (BiGG, Norsigian *et al.* 2020) identifiers. If the input model provides no EC numbers, an integrated Application Programming transfer Interface (API) call requests the necessary information from the BiGG database and adds all missing annotations into the model. Depending on the model's size, this step may increase the computational time. Hence, we recommend the prior use of an annotation tool, such as ModelPolisher (Römer *et al.* 2016). We have tested the performance of SBOannotator in assigning descriptive and more precise terms to biochemical reactions using 108 metabolic models from the BiGG database. All downloaded models contained only five types of SBO annotation representing only top-level terms. The biochemical reactions made up the largest group before and after the SBOannotator (see Supplementary Tables S2 and S3). However, their coverage was reduced from 57.9% to 18.9%, meaning a large percentage of the initial reactions got a more specific term (see Supplementary Figs S1 and S2). Finally, SBOannotator assigns SBO terms to the remaining model entities, such as

metabolites and genes. The final annotated SBML model is stored in the current directory with the tag `_SBOannotated`.

3 Conclusion

Unlike other annotations, like the EC numbers, the SBO terms are directly assigned in SBML (Keating *et al.* 2020) as individual attributes to any model element. This results in a direct one-to-one mapping that indicates the function of that element inside the model. In contrast, numerous EC numbers and further references may be provided to the same element to show the link between an annotation and a model element, each with different qualifiers. For instance, a reaction in SBML may involve a so-called modifier in addition to reactants and products. This could be an inhibitor, an (enzymatic) catalyst, or some other stimulator. If multiple EC numbers link a modifier to a specific enzyme, the SBOannotator can interpret this information and add a precise SBO term for "enzymatic catalyst". Without specifying the exact mechanism of this catalysis, the role of the modifier is now defined through an "is a"-relationship: this modifier is an enzymatic catalyst. Once assigned, other software, like the SBMLsqueezer 2 (Dräger *et al.* 2015), may interpret this information to automatically derive suitable reaction rate laws, where it is of great importance if the modifier is an inhibitor or an enzyme. With this, the information from the EC numbers becomes directly accessible and interpretable: instead of a potentially extensive list of various annotations, there is a single attribute with a defined value. By this, SBOannotator helps to define more clearly which role individual elements play within a model.

Overall, the SBOannotator is a freely available and user-friendly Python tool. It can be easily employed to rapidly annotate systems biology metabolic networks in SBML format with appropriate SBO terms, with particular emphasis on allocating precise and descriptive terms to all chemical reactions. The minimal requirement for the tool is a valid SBML format of the input model(s). SBOannotator will then proceed with the labeling of model components with terms based on the defined model entities and attributes. The assignment of enzyme-based SBO terms to reactions is hinged upon the existence of standardized BiGG (Norsigian *et al.* 2020) identifiers. However, expanding the usability of SBOannotator by enabling the utilization of further database identifiers to extract enzymatic information would be of great importance. So far, SBOannotator is a standalone application. Its integration into existing software, such as ModelPolisher (Römer *et al.* 2016), could be worthwhile, as long as an abstract use of the tool by the users is possible. Lastly, since the main emphasis of the tool is the precise annotation of biochemical reactions, it has been developed on the basis of genome-scale metabolic models. However, SBO terms may also be useful in other modeling frameworks, such as, the automated assignment of rate laws for dynamic simulations. Hence, the SBOannotator could be extended to assign additional terms specific to different model types, including dynamic, stochastic, and population models.

Supplementary data

Supplementary data are available at *Bioinformatics* online.

Funding

Conflict of interest

None declared.

Data availability

The SBOannotator tool, all related data, and a demo script to run the code are available in a git repository at https://github.com/draeger-lab/SBOannotator/. Along with this article, a supplementary table in Comma-separated Values (CSV) format is available.

References

Courtot M, Juty N, Knüpfer C *et al.* Controlled vocabularies and semantics in systems biology. *Mol Syst Biol* 2011;7:543.

Dräger A, Zielinski DC, Keller R *et al.* Sbmlsqueezer 2: context-sensitive creation of kinetic equations in biochemical networks. *BMC Syst Biol* 2015;9:68.

Hucka M, Bergmann FT, Chaouiya C *et al.* The systems biology markup language (SBML): language specification for level 3 version 2 core release 2. *J Integr Bioinform* 2019;16:1. https://doi.org/10.1515/jib-2019-0021.

Keating SM, Waltemath D, König M *et al.*; SBML Level 3 Community members. SBML level 3: an extensible format for the exchange and reuse of biological models. *Mol Syst Biol* 2020;16:e9110. https://doi.org/10.15252/msb.20199110.

Norsigian CJ, Pusarla N, McConn JL *et al.* BiGG Models 2020: multi-strain genome-scale models and expansion across the phylogenetic tree. *Nucleic Acids Res* 2020;48:D402–6. https://doi.org/10.1093/nar/gkz1054.

Römer M, Eichner J, Dräger A *et al.* ZBIT bioinformatics toolbox: a web-platform for systems biology and expression data analysis. *PLoS One* 2016;11:e0149263. https://doi.org/10.1371/journal.pone.0149263.

Stevens R, Goble CA, Bechhofer S. Ontology-based knowledge representation for bioinformatics. *Brief Bioinform* 2000;1:398–414. https://doi.org/10.1093/bib/1.4.398.
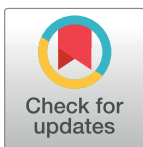
APPENDIX D

Paper III: Exploring the metabolic profiling of *A. baumannii* for antimicrobial development using genome-scale modeling

PLOS PATHOGENS

Exploring the metabolic profile of *A. baumannii* for antimicrobial development using genome-scale modeling

Nantia Leonidou [1,2,3,4,5] *, Yufan Xia [2], Lea Friedrich [6], Monika S. Schütz [4,6], Andreas Dräger [1,4,5,7]

1 Computational Systems Biology of Infections and Antimicrobial-Resistant Pathogens, Institute for Bioinformatics and Medical Informatics (IBMI), Eberhard Karl University of Tübingen, Tübingen, Germany, **2** Department of Computer Science, Eberhard Karl University of Tübingen, Tübingen, Germany, **3** Cluster of Excellence 'Controlling Microbes to Fight Infections', Eberhard Karl University of Tübingen, Tübingen, Germany, **4** German Center for Infection Research (DZIF), partner site Tübingen, Germany, **5** Quantitative Biology Center (QBiC), Eberhard Karl University of Tübingen, Tübingen, Germany, **6** Interfaculty Institute for Microbiology and Infection Medicine, Institute for Medical Microbiology and Hygiene, University Hospital Tübingen, Tübingen, Germany, **7** Data Analytics and Bioinformatics, Institute of Computer Science, Martin Luther University Halle-Wittenberg, Halle (Saale), Germany

* nantia.leonidou@uni-tuebingen.de

Abstract

With the emergence of multidrug-resistant bacteria, the World Health Organization published a catalog of microorganisms urgently needing new antibiotics, with the carbapenem-resistant *Acinetobacter baumannii* designated as "critical". Such isolates, frequently detected in healthcare settings, pose a global pandemic threat. One way to facilitate a systemic view of bacterial metabolism and allow the development of new therapeutics is to apply constraint-based modeling. Here, we developed a versatile workflow to build high-quality and simulation-ready genome-scale metabolic models. We applied our workflow to create a metabolic model for *A. baumannii* and validated its predictive capabilities using experimental nutrient utilization and gene essentiality data. Our analysis showed that our model *i*ACB23LX could recapitulate cellular metabolic phenotypes observed during *in vitro* experiments, while positive biomass production rates were observed and experimentally validated in various growth media. We further defined a minimal set of compounds that increase *A. baumannii*'s cellular biomass and identified putative essential genes with no human counterparts, offering new candidates for future antimicrobial development. Finally, we assembled and curated the first collection of metabolic reconstructions for distinct *A. baumannii* strains and analyzed their growth characteristics. The presented models are in a standardized and well-curated format, enhancing their usability for multi-strain network reconstruction.

Author summary

The emergence of multidrug-resistant bacteria, particularly carbapenem-resistant *Acinetobacter baumannii*, has become a severe global health threat. This pressing issue

necessitated the development of new antibiotics, as highlighted by the World Health Organization. To address this need, we aimed to create comprehensive metabolic models to better understand bacterial metabolism and aid in developing novel therapeutic strategies. In this study, we developed a versatile workflow to construct high-quality, simulation-ready genome-scale metabolic models for bacterial pathogens. Applying this workflow, we constructed a metabolic model for *A. baumannii* and validated its accuracy using experimental data. The model successfully replicated observed metabolic phenotypes and identified essential genes without human counterparts, suggesting potential targets for new antibiotics. Additionally, we assembled and curated the first collection of metabolic reconstructions for distinct *A. baumannii* strains, analyzing their growth characteristics. These standardized and well-curated models enhance usability, facilitating multi-strain network reconstruction and further research. These findings provide a robust tool for understanding *A. baumannii*'s metabolism, guiding the development of new antimicrobial therapies.

Introduction

In the 21st century, treating common bacterial infections has become a global health concern. The rapid emergence of pathogens with newly developed resistance mechanisms led to the ineffectiveness of hitherto used antimicrobial drugs. According to their resistance patterns, bacteria are classified into three main categories: multidrug-resistant (MDR, resistant to at least one agent in more than three antibiotic categories), extensively drug-resistant (XDR, non-susceptible to one or two categories), and pandrug-resistant (PDR, non-susceptible to all drugs in all categories) [1]. Pathogens from the last two classes are called "superbugs". In February 2022, Murray *et al.* developed predictive statistical models within a large-scale global study and estimated 1.27 million deaths directly associated with antimicrobial resistance (AMR) [2]. The same study underlines the highly virulent ESKAPE pathogens (*Enterococcus faecium*, *Staphylococcus aureus*, *Klebsiella pneumoniae*, *Acinetobacter baumannii*, *Pseudomonas aeruginosa*, and *Enterobacter* spp.) as the primary cause of AMR-related deaths, while the World Health Organization (WHO) announced in 2017 the urgent need for novel and effective therapeutic strategies against these microorganisms, assigning them the "critical status".

Over the years, numerous studies highlighted the Gram-negative human pathogen *Acinetobacter baumannii* of substantial concern in hospital environments attributable to its high intrinsic resistance against antimicrobial agents, including biocides [3–6]. *A. baumannii* (from the Greek word *akínētos*, meaning "unmoved") is a rod-shaped, non-motile, and strictly aerobic bacterium. It is an opportunistic pathogen whose adaptable genetic apparatus has caused it to become endemic in intensive care units (ICUs), affecting immunocompromised patients, causing pneumonia, bacteremia, endocarditis, and more. Especially the carbapenem-resistant *A. baumannii* poses a serious global threat with high mortality rates [7–9]. It targets exposed surfaces and mucous tissues, colonizes the human nose [10–12], and is frequently associated with Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) infections [13–16]. The skin has shown to be a community reservoir for *A. baumannii* in a very small percentage of samples [17, 18], while its prevalence in the soil is a frequent misconception as species from the genus *Acinetobacter* are ubiquitous in nature [5, 19]. Finally, it shows susceptibility to commonly used drugs, like *β*-lactams, aminoglycosides, and polymyxins. The strain ATCC 17978 is a widely studied nosocomial strain notable for its fully sequenced genome that provides a comprehensive genetic framework for researchers. Historically, it was first isolated in 1951

from a 4-month-old infant with fatal meningitis [20]. Smith *et al.* sequenced its complete genome using high-density pyrosequencing [20]. They reported the presence of putative pathogenic islands with virulence genes, including a large island with transposons and integrases, as well as elements resembling the Legionella/Coxiella Type IV secretion system, crucial for pathogenesis. Moreover, it is a human-adapted isolate and was the first to be discovered with an active type VI secretion system (T6SS) encoded on its chromosome [21]. Another remarkable characteristic is its utilization of microcin Mcc17978 to combat other bacteria such as closely related *Acinetobacter* species and *E. coli* through contact and T6SS-independent competition activity [22]. Previous studies have noted ATCC 17978 for its susceptibility to clinically important antibiotics, serving as a valuable model for studying the basic biology of *A. baumannii*, antibiotic susceptibility, and the mechanisms by which resistance can develop [23–26]. Finally, its lower level of antibiotic resistance makes it valuable for studying the basic mechanisms of antibiotic action and resistance development, while it is easier to genetically manipulate [27].

Systems biology, and especially the field of genome-scale metabolic network analysis, is the key to exploring genotype-phenotype relationships, better understanding mechanisms of action of such threatening pathogens, and ultimately developing novel therapeutic strategies. Genome-scale metabolic models (GEMs) combined with constraint-based modeling (CBM) provide a well-established, fast, and inexpensive *in silico* framework to systematically assess an organism's cellular metabolic capabilities under varying conditions having only its annotated genomic sequence [28]. As of today, GEMs have numerous applications in metabolic engineering, contributing to the formulation of novel hypotheses towards the detection of new potential pharmacological targets [29].

It has been more than a decade since the release of the first mathematical model representing *A. baumannii*'s metabolism. Kim *et al.* integrated biological and literature data to manually build AbyMBEL891, representing the strain AYE [30]. This model was further employed as an essential foundation for future reconstructions; however, its non-standardized and missing identifiers limited its use. Following a tremendous increase in the volume of literature and experimental data on *A. baumannii* (over 5,670 articles published between 2010 and 2017 according to PubMed), two new strain-specific metabolic networks emerged: the *i*LP844 [31] and the AGORA (Assembly of Gut Organisms through Reconstruction and Analysis) model [32]. Both networks were reconstructed through a semi-automated process and simulated the metabolism of two distinct strains: ATCC 19606 and AB0057, respectively. With the help of transcriptomic data of sampled colistin responses and *i*LP844, it was observed that the type strain ATCC 19606 underwent metabolic reprogramming, demonstrating a stress condition as a resistance mechanism against colistin exposure. Alterations in gene essentiality phenotypes between treated and untreated conditions enabled the discovery of putative antimicrobial targets and biomarkers. Moreover, the model for AB0057 was part of an extensive resource of GEMs built to elucidate the impact of microbial communities on host metabolism. The amount of mass- and charge-balanced reactions in these models is very high; however, they carry few to no database references. Norsigian *et al.* improved and expanded AbyMBEL891 to finally create the high-quality model *i*CN718 with a prediction accuracy of over 80% in experimental data [33], while Zhu *et al.* built a GEM for ATCC 19606 (*i*ATCC19606) integrating multi-omics data [34]. Compared to *i*LP844, *i*ATCC19606 incorporates metabolomics data together with transcriptomic data enabling the deciphering of bactericidal activity upon polymyxin treatment and the interplay of various metabolic pathways. Last but not least, in 2020, the first *in vivo* study on *A. baumannii* infection was published utilizing constraint-based modeling [35]. This time, the collection of strain-specific models was enriched with the first GEM for the hyper-virulent strain AB5075 (*i*AB5075). The model was validated using various

experimental data, while transcriptomics data was leveraged to identify critical fluxes leading to mouse bloodstream infections. Our literature search revealed one additional metabolic model of *A. baumannii* ATCC 17978, named *i*JS784 [36]. As of the time of writing, this model has not been officially published in a scientific journal or been deposited in a mathematical models' database. Nonetheless, the model cannot produce biomass even when all uptake reactions are open and all medium nutrients are available to the cell, making it unusable and hampering reproducibility.

We expanded the collection of *A. baumannii* GEMs by building a high-quality model for the nosocomial strain ATCC 17978, named *i*ACB23LX. The presented model follows the FAIR data principles and community standards and recapitulates experimentally-derived phenotypes with high predictive capability and accuracy scores. We enriched the model with numerous database cross-references and inferred the minimal nutritional requirements computationally. Moreover, we used this model to investigate the organism's growth ability in defined media and a medium simulating human nasal secretions while we assessed its ability to predict essential genes using different optimization approaches. Among the examined strains, ATCC 17978 is one of the most well-studied, with a substantial amount of experimental data available that can be used to direct model refinement and validation. Besides that, we systematically refined and evaluated all pre-existing reconstructions' performance to finally create the first compendium of curated and standardized models for *A. baumannii*. With this, we aim to support further studies to give new insights into this pathogen and promote strain- and species-specific therapeutic approaches.

Materials and methods

Growth curves of *A. baumannii*

Growth curves for *A. baumannii* strains AB5075, ATCC 17978, ATCC 19606, and AYE were recorded in Luria-Bertani (LB), *i*MinMed supplemented with acetate as sole carbon source (0.2% weight per volume), and synthetic nasal medium (SNM) [97]. Overnight cultures of the strains grown in LB were harvested by centrifugation, and the cell pellet was washed once with 5 mL of phosphate buffered saline (PBS). Cells were then re-suspended in the medium used for the growth curves. The starting optical density $(OD)_{600nm}$ was adjusted to 0.1 and growth curves were recorded in 2 mL of medium for 12 h in triplicates using a Tecan infinite M200 PRO plate reader and 12-well plates covered with a plastic lid. Plates were incubated with linear shaking at 37°C and the OD_{600nm} was measured every 15 min. Growth rates were determined as the slope of the linear part of the curves plotting the natural logarithm of OD_{600nm} against time.

Metabolic network reconstruction workflow

Fig 1 illustrates the workflow developed to create the high-quality genome-scale metabolic network *i*ACB23LX, following the state-of-the-art protocol by Thiele *et al.* [28]. Our workflow consists of eight major steps starting from the extraction of an annotated genome until the model validation using experimental data. Modifications in the model structure, as well as the inclusion of cross-references to multiple functional databases, were done using the libSBML [46] library, while all simulations were conducted via the constraints-based reconstruction and analysis for Python (COBRApy)-0.22.1 suite [58] that includes functions commonly used for simulations.

The individual steps are described below in more detail with respect to the reconstruction of *i*ACB23LX.
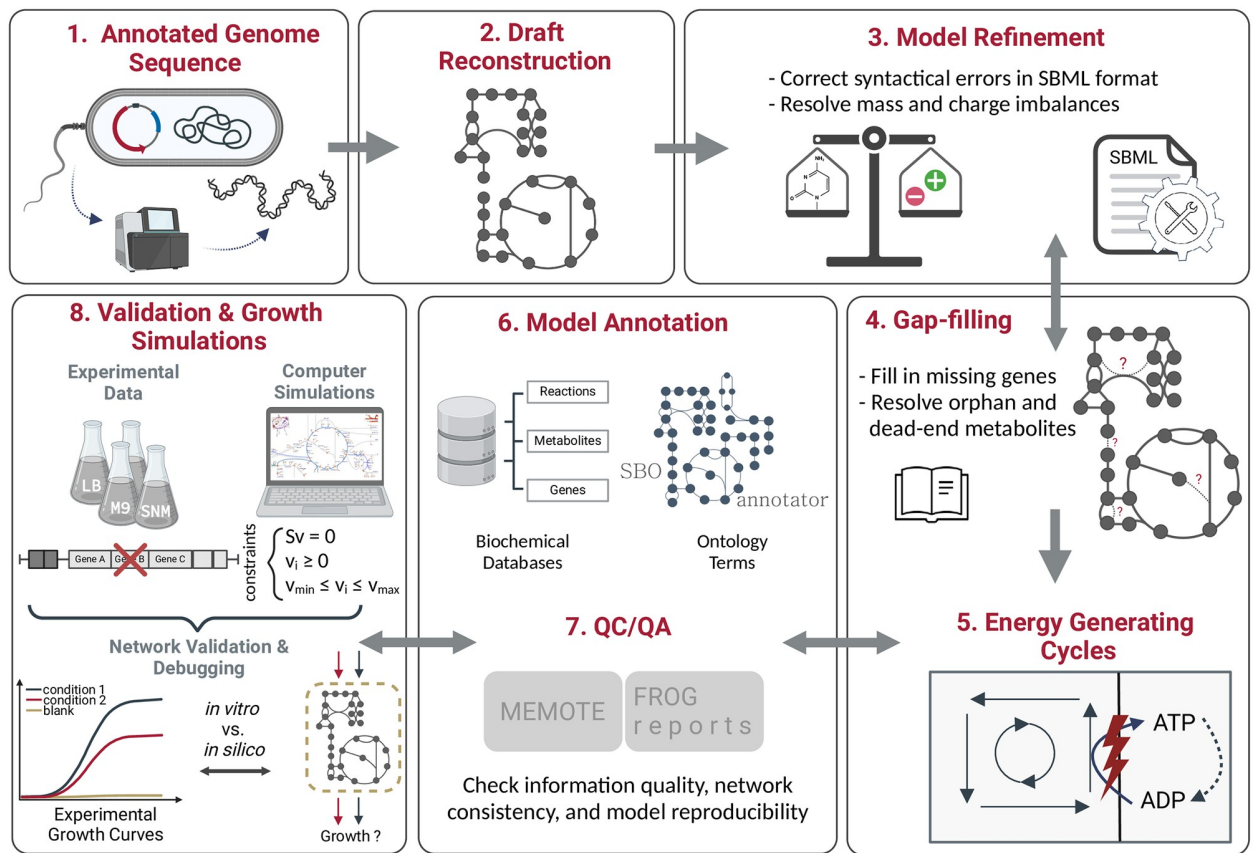
Fig 1. Workflow developed for the metabolic network reconstruction of *i*ACB23LX. The created workflow consists of eight main steps: extraction of the annotated genome, draft model reconstruction, model refinement, gap-filling, investigation of energy-generating cycles, model annotation, quality control and quality assurance (QC/QA), and model validation using experimental data. Growth simulations include the examination of growth requirements and the definition of a minimal growth medium. The last six processes are continuously iterated until the model reaches a satisfied quality and can recapitulate known phenotypes. Figure created with BioRender.com.

https://doi.org/10.1371/journal.ppat.1012528.g001

Draft reconstruction. The draft model was built with CarveMe version 1.5.1 using the annotated genome sequence of ATCC 17978 [39]. This was downloaded from the National Centre for Biotechnology Information (NCBI) at https://www.ncbi.nlm.nih.gov and has the assembly accession number ASM1542v1 [20]. Seven strain-specific assemblies are registered in NCBI; however, the chosen entry is also registered in Kyoto Encyclopedia of Genes and Genomes (KEGG) [38], which supports the model extension. The genome is 3.9 Mbp long and has two plasmids (pAB1 and pAB2). We set the SBML flavor to activate the extension for flux balance constraints (fbc) version 2 [98]. This extension enables semantic descriptions for domain-specific elements, such as metabolite chemical formulas and charges, along with reaction boundaries and gene-protein-reaction associations (GPRs). Moreover, the optional parameter gramneg provided CarveMe was selected to employ the specialized template for the Gram-negative bacteria. Compared to the Gram-positive template, the Gram-negative template comes with phosphatidylethanolamines, murein, and a lipopolysaccharide unit. Its biomass reaction involves membrane and cell wall components resulting in more accurate gene essentiality predictions in the lipid biosynthesis pathways.

Manual refinement and extension. We started the manual refinement of the draft model by resolving syntactical errors within the Systems Biology Markup Language (SBML) [40]

model file using the SBML Validator from the libSBML [46]. These errors involve improper file structure, incorrect or missing tags, missing or improperly formatted attributes, and invalid values. Missing metabolite charges and chemical formulas were retrieved from the Biochemical, Genetical, and Genomical (BiGG) [70] and ChEBI [99] databases, while mass- and charge-imbalanced reactions were corrected. The most intense part of the workflow is the manual network extension and gap-filling. This was done using the organism-specific databases KEGG [38] and BioCyc [100], together with ModelSEED [71]. We mapped the new gene locus tags to the old ones using the GenBank General Feature Format (GFF) [101] and added missing metabolic genes along with the respective reactions and metabolites into our model. The network's connectivity was ensured by resolving as many dead-end (only produced but not consumed) and orphan (only consumed but not produced) metabolites as possible. Also, reactions with no connectivity were not included in the model, while reactions with no organism-specific gene evidence were removed from the model.

Erroneous energy generating cycles. Energy-generating cycles (EGCs) are thermodynamically infeasible loops found in metabolic networks and have not been experimentally observed, unlike futile cycles. EGCs charge energy metabolites such as adenosine triphosphate (ATP) and uridine triphosphate (UTP) without any external source of nutrients, which may lead to incorrect and unrealistic energy increases. Their elimination is crucial for correcting the energy metabolism, as they can inflate the maximal biomass yields and render unreliable predictions. We checked their existence in *i*ACB23LX applying an algorithm developed by Fritzemeier *et al.* [41].

We created a Python script that (1) defines and adds energy dissipation reactions (EDRs) in the network:

$$X[c] + H_2O[c] \longrightarrow X^+[c] + H[c] + Pi[c],$$
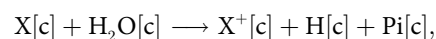
where X is the metabolite of interest and (2) maximizes each EDR while blocking all influxes. This can be formulated as follows:

$$\max(v_{edr})$$

subject to

$$\begin{aligned} S \cdot \vec{v} &= 0 \\ \forall i \notin E : v_i^{\min} &\leq v_i \leq v_i^{\max} \\ \forall i \in E : v_i &= 0 \end{aligned} \tag{1}$$

where *edr* is the index of the current dissipation reaction, S is the stoichiometric matrix, $\vec{v}$ the flux vector, E the set of all exchange reactions, and $v_i^{\min}$ and $v_i^{\max}$ the upper and lower bounds. The existence of EGCs is indicated by a positive optimal value of v_{edr}.

Totally we examined 14 energy metabolites: ATP, cytidine triphosphate (CTP), guanosine triphosphate (GTP), UTP, inosine triphosphate (ITP), nicotinamide adenine dinucleotide (NADH), nicotinamide adenine dinucleotide phosphate (NADPH), flavin adenine dinucleotide (FADH2), flavin mononucleotide (FMNH2), ubiquinol-8, menaquinol-8, demethylmenaquinol-8, acetyl-coenzyme A (CoA), and L-glutamate. Moreover, we tested the proton exchange between cytosol and periplasm.

In the case of existing EGCs, we examined the directionality and the gene evidence of all participated reactions using organism-specific information from BioCyc as reference [100].

Database annotations. In this stage, the model was enriched with cross-linkings to various functional databases. Reactions and metabolites were annotated with relevant databases (e.g., KEGG [38], BRENDA [102], and UniProt [103]). These were included in the model as

controlled vocabulary (CV) Terms following the Minimal Information Required In the Annotation of Models (MIRIAM) guidelines [104] and the resolution service at https://identifiers.org/. We used ModelPolisher [105] to complete the missing available metadata for all metabolites and genes. Similarly, metabolic genes were annotated with their KEGG [38], NCBI Protein, and RefSeq identifiers using the GFF file [101]. Systems Biology Ontology (SBO) terms were assigned to different model entities using a freely accessible standalone tool called SBOannotator [43]. The SBO terms are structured controlled vocabularies used in computational modeling to define and describe model entities unambiguously [106]. The SBOannotator was developed to automatically assign precise SBO terms to SBML models, mainly focusing on biochemical reactions, thereby enhancing the reproducibility and usability of biochemical networks [43]. In addition, Evidence and Conclusion Ontology (ECO) terms were added to every reaction to capture the type of evidence of biological assertions with `BQB_IS_DESCRIBED_BY` as a biological qualifier. They are useful during quality control and mirror the curator's confidence about the inclusion of a reaction. When multiple genes encode a single reaction, an ECO term was added for every participant gene. Both terms were incorporated into the final model.

Finally, reactions were annotated with the associated subsystems in which they participate using the KEGG [38] database and the biological qualifier `BQB_OCCURS_IN`. Moreover, the "groups" plugin was activated [48]. Every reaction that appeared in a given pathway was added as a `groups:member`, while each pathway was created as a group instance with `sboTerm="SBO:0000633"` and `groups:kind="partonomy"`.

Quality control and quality assurance. The metabolic model testing (MEMOTE) [45] version 0.13.0 was used to assess and track the quality of our model after each modification, providing us with information regarding the model improvement. The Flux variability, Reaction deletion, Objective function, and Gene deletion (FROG) analysis framework was utilized to assess model reproducibility, ensuring reusability and results verification [47]. The resulting reports include comprehensive analyses of flux variability, reaction deletion, objective function, and gene deletion, providing a thorough evaluation of model performance. The final model was converted into the latest SBML Level 3 Version 2 [48] format using the libSBML package, while the SBML Validator tracked syntactical errors and ensured a valid format of the final model [46].

Constraint-based analysis

The most frequently used constraint-based modeling approach is the flux balance analysis (FBA) that determines a flux distribution via optimization of the objective function and linear programming [59]. Prior to this, the metabolic network is mathematically encoded using the stoichiometric matrix *S* formalism. This structure delineates the connectivity of the network, and it is formed by the stoichiometric coefficients of all participating biochemical reactions. The rows and columns are represented by the metabolites and the mass- and charge-balanced reactions respectively. At steady state, the system of linear equations derived from the network is defined as follows:

$$\mathbf{S} \cdot \vec{v} = 0 \tag{2}$$

with $\mathbf{S}$ being the stoichiometric matrix and $\vec{v}$ the flux vector. With no defined constraints, the flux distribution may be determined at any point within the solution space. This space must be further restricted since the system is under-determined and algebraically insoluble. An allowable solution space is defined by a series of imposed constraints that are followed by cellular functions. Altogether the FBA optimization problem, with mass balance, thermodynamic, and

capacity constraints, is defined as:

$$\begin{aligned}
\text{max or min} \quad & Z = c^T \vec{v} \\
\text{subject to :} \quad & \mathbf{S} \cdot \vec{v} = 0 \\
& 0 \leq v_i & \forall \text{ irreversible reactions } i \\
& v_{\min} \leq v_i \leq v_{\max} & \text{for } i = 1, \ldots, n.
\end{aligned} \tag{3}$$

Here, n is the amount of reactions, Z represents the linear objective function, and $\vec{c}$ is a vector of coefficients on the fluxes $\vec{v}$ used to define the objective function.

Growth simulations

Strict aerobic growth check. At the time of writing, CarveMe does not include reconstruction templates to differentiate between aerobic and anaerobic species. The directionality of reactions that produce or consume oxygen may affect the model's ability to grow anaerobically. *A. baumannii* is defined to be a strictly aerobic species. Hence, we tested whether our model could grow with no oxygen supplementation. For this purpose, we examined all active oxygen-producing reactions under anaerobic conditions. We corrected their directionality based on the organism-specific information found in BioCyc [100] and kept only those with associated gene evidence.

Defining a minimal growth medium. To determine the minimal number of nutrients needed for the bacterium to grow, we defined a minimal medium using *i*ACB23LX. We determined the minimal amount of metabolites needed for growth using the M9 minimal medium (M9) (S1 Table) as a reference (*i*MinMed). We modeled growth on *i*MinMed by enabling the uptake of all metabolites that constitute the medium. The lower bound for the rest of the exchanges was set to 0 mmol/($g_{DW} \cdot$ h). The final minimal medium is listed in Table 1 and in S1 Table. It consists of nine transition metals, a carbon source, a nitrogen source, a sulfur source, and a phosphorus source. The aerobic environment was simulated by setting the lower bound for the oxygen exchange to −10 mmol/($g_{DW} \cdot$ h).

Table 1. Composition of the computationally defined minimal growth medium, *i*MinMed. It consists of nine transition metals, a carbon source, a nitrogen source, a sulfur source, and a phosphorus source. Oxygen was used to represent aerobic conditions.

	Molecular Formula	Name
Carbon source	$C_2H_3O_2{}^-$	Acetate
Nitrogen source	$NH_4{}^+$	Ammonium
Sulfur source	$SO_4{}^{2-}$	Sulfate
Phosphorus source	$HPO_4{}^{2-}$	Phosphate
	Ca^{2+}	Calcium
	Cl^-	Chloride
	Cu^{2+}	Copper
	Fe^{3+}	Ferric iron
Transition metals	Co^{2+}	Cobalt
	K^+	Potassium
	Mg^{2+}	Magnesium
	Mn^{2+}	Manganese
	Zn^{2+}	Zinc
Oxygen source	O_2	Oxygen

Growth in chemically defined media. We utilized experimentally verified growth media to examine the growth capabilities of *i*ACB23LX. The LB medium serves as a common medium for the cultivation of *A. baumannii*. Consequently, we conducted an assessment of our model's capacity to accurately simulate growth in this particular medium. Furthermore, we examined the growth of our model in the human nasal niche, considering that *A. baumannii* has been isolated from nasal samples within ICUs [10–12]. For this purpose, we utilized the SNM that imitates the human nasal habitat [97]. In all cases, if macromolecules or mixtures were present, we considered the constitutive molecular components for the medium definition. As our model was initially unable to reproduce growth on the applied media, we deployed the gap-filling option from CarveMe to detect missing reactions and gaps in the network [39]. All growth media formulations are available in S1 Table.

Rich medium definition. To investigate our model's growth rate when all nutrients are available to the bacterial cell, we defined the rich medium. For this purpose, we enabled the uptake of all extracellular metabolites by the model setting the lower bound of their exchange reactions to −10 mmol/($g_{DW} \cdot h$).

Model validation

Evaluation of carbon and nitrogen utilization. We employed the previously published Biolog Phenotypic Array data by Farrugia *et al.* for *A. baumannii* ATCC 17978 to validate the functionality of our model [26]. According to the experimental guidelines provided by Farrugia *et al.*, we utilized M9 for all simulations [26]. The medium was then supplemented with D-xylose as a carbon source for the nitrogen testings, while ammonium served as the only nitrogen source for the carbon tests. As D-xylose was initially not part of the model, we conducted an extensive search in the organism-specific databases KEGG [38] and BioCyc [100] to include missing reactions.

The phenotypes were grouped by their maximal kinetic curve height. A trait was considered positive ("growth") if the height exceeded the 115 and 101 OmniLog units for a nitrogen and carbon source, respectively. The prediction accuracy was evaluated by comparing the *in silico*-derived phenotypes to the Biolog results. More specifically, the overall model's accuracy (ACC) was calculated by the overall agreement:

$$ACC = \frac{TP + TN}{TP + TN + FP + FN} \tag{4}$$

where true positive (TP) and true negative (TN) are correct predictions, while false positive (FP) and false negative (FN) are inconsistent predictions. Discrepancies were resolved via iterative manual model curation.

Gene perturbation analysis. We performed *in silico* single-gene deletions on *i*ACB23LX to detect essential genes. For this purpose, we utilized the `single_gene_deletion` function from the COBRApy [58] package. A gene is considered to be essential if a flux of 0.0 mmol/($g_{DW} \cdot h$) is predicted through the biomass reaction after setting the lower and upper bounds of the associated reaction(s) to 0.0 mmol/($g_{DW} \cdot h$).

Additionally, we examined the effect of gene deletions using two different optimization approaches: FBA [59] and minimization of metabolic adjustment (MOMA) [60]. Contrary to FBA, MOMA is based on quadratic programming, and the involved optimization problem is the Euclidean distance minimization in flux space. Moreover, it approximates the metabolic phenotype and relaxes the assumption of optimal growth flux for gene deletions [60].

The results were compared to gene essentiality data [57]. Wang *et al.* generated a random mutagenesis dataset including 15,000 unique transposon mutants using insertion sequencing

(INSeq) [57]. Four additional transposon sequencing (Tn-seq) libraries of multiple *A. baumannii* isolates were employed to validate the model and increase confidence [57, 68, 77–79]. In this case only genes with an ortholog in `ASM1542v1` were considered. Analogously to the experimental settings, the nutrient uptake constraints were set to define the LB medium. From the 453 genes experimentally identified as essential, 191 could be compared to our predictions. The rest were not part of *i*ACB23LX due to their non-metabolic functions. To measure the effect of a single deletion, we calculated the fold change (FC) between the model's growth rate after (gr_{KO}) and before (gr_{WT}) a single knockout. This is formulated as follows:

$$FC_{gr} = \frac{gr_{KO}}{gr_{WT}} \tag{5}$$

To this end, if $FC_{gr} = 0$, the deleted gene is classified as essential, meaning its removal prevented the network from producing at least one key biomass metabolite predicting no growth. Similarly, if $FC_{gr} = 1$, the deletion of the gene from the network did not affect the growth phenotype (labeled as inessential), while when $0 < FC_{gr} < 1$, the removal of this gene affected partially the biomass production (labeled as partially essential). The complete lists of the gene essentiality results are available in S2 and S3 Tables.

To explore the potential of *in silico* identified essential genes as new drug candidates against *A. baumannii* infections, we probed the queries of predicted false negative candidates against the human proteome using Basic Local Alignment Search Tool (BLAST) [107]. The protein sequences were aligned to the human protein sequences using the default settings of the NCBI BLASTp tool (word size: 6, matrix: BLOSUM62, gap costs: 11 for existence and 1 for extension). To eliminate adverse effects and ensure no interference with human-like proteins, queries with any non-zero alignment score with the human proteome were not considered. Lastly, we searched the DrugBank database version 5.1.9 to find inhibitors or ligands known to act with the enzymes encoded by the non-homologous genes [66].

Curation of existing metabolic networks

Previously reconstructed models of *A. baumannii* for multiple strains were collected and curated following community standards and guidelines. For this, we created a workflow that comprises four main steps and utilizes model validation and annotation tools. This can be applied to any metabolic network in SBML [40] format and follows the community "gold standards" strictly, as proposed by Carey *et al.* [37]. The curation steps involved changes in the format, amount, and quality of the included information. The context has not been altered in any way that could impact the models' prediction capabilities. We employed a combination of already existing tools to analyze, simulate, and quality-control the models (COBRApy [58], `MEMOTE` [45], and the SBML Validator [46]). Different database cross-references were incorporated in the models using ModelPolisher [105] and following the MIRIAM guidelines [104], while the libSBML library [46] was used to manipulate the file format and convert to the latest version. To resolve inflated growth rates, we determined computationally-defined minimal growth media. The growth capabilities were examined with respect to various experimentally-derived growth media, while the LB medium was applied to identify lethal genes. A strain-wise comparison was not feasible due to strain-specific identifiers, no successful growth, or missing genes. Hence, we investigated the essential genes across all models with identifiers that could be mapped with the Pathosystems Resource Integration Center (PATRIC) ID mapping tool [108].

To begin with the debugging, we examined the syntactical correctness and internal consistency of the downloaded files using the SBML Validator from the libSBML library [46]. Two models (*i*CN718 and *i*JS784) could not pass the validator check and reported errors since they

were not in a valid SBML [40] format right after their attainment. We made *i*CN718 valid by deleting the reaction `DNADRAIN` for which neither a reactant nor a product was assigned since the associated metabolite was not part of the model. Similarly, the empty `groups` attribute was removed from *i*JS784, converting the file into a valid format. Warnings were detected for *i*ATCC19606, and *i*AB5075 due to missing definition of the fbc extension (available at the latest Level 3 release [98]) and the non-alphanumeric chemical formulas. We resolved these issues by defining the fbc list `listOfGeneProducts` and the species attribute `chemicalFormula`. In more detail, we extracted the given GPRs from the notes field and defined individual `geneProduct` classes with `id`, `name`, and `label`. The attribute `chemicalFormula` was set equal to the species chemical formulas extracted from the notes and is particularly essential in reaction's validation and balancing. Following the SBML [40] specifications regarding its constitution, in case of ambiguous formulas separated by a semicolon (`;`), the first molecular representation was chosen. With this, the genes and metabolites' chemical formulas became part of the file's main structure. Since *i*ATCC19606 carried KEGG [38] identifiers, we could extract the metabolites' chemical formulas from the database and add them to the model. Moving on with the file extension, we declared the remaining missing attributes from reactions, metabolites, and genes that are required according to the SBML [40] language guidelines. More specifically, we defined the `metaid` attribute when missing, while we fixed any errors regarding the identifiers nomenclature. Further extension involved the annotation of reactions, metabolites, and genes with a plethora of database cross-references following the MIRIAM guidelines [104]. For this, we employed ModelPolisher that complements and annotates SBML [40] models with additional metadata using the BiGG Models knowledgebase as reference [105]. We also defined precise SBO terms with the `sboTerm` attribute using the SBOannotator [43]. The final step of debugging involved the conversion of all models to the newest available format SBML Level 3 Version 2 [48], as well as the quality control using `MEMOTE` [45].

Results

Reconstruction process of the metabolic network *i*ACB23LX

To build a high-quality model for *A. baumannii* ATCC 17978, we developed a workflow, as depicted in Fig 1, adhering closely to the community standards [37] (see Materials and methods).

We named the newly reconstructed network *i*ACB23LX, where *i* stands for *in silico*, ACB is the organism- and strain-specific three-letter code from the KEGG [38] database, 23 the year of reconstruction, and LX the modellers' initials. Our protocol involves eight major stages starting from the attainment of the annotated genomic sequence until the model validation, applies to any organism from the tree of life (Archaea, Bacteria, and Eukarya), and ensures the good quality and correctness of the final model. CarveMe [39] was used to build a preliminary model, which was subsequently extended and curated manually. We resolved SBML [40] syntactical issues and mass and charge imbalances during manual refinement while we defined missing metabolite charges and chemical formulas. Our final model contains no mass-imbalanced reactions and only two charge-imbalanced reactions. After extensive efforts, resolving all charge imbalances was impossible since all participated metabolites are interconnected to multiple reactions within the network. Hence, any modification in their charge resulted in newly introduced imbalances. The model extension process involved incorporating missing metabolic genes considering the network's connectivity. Dead-end and orphan metabolites do not exist biologically in the species, implying knowledge gaps in metabolic networks. Moreover, reactions including such metabolites are not evaluated in FBA. Therefore, reactions with
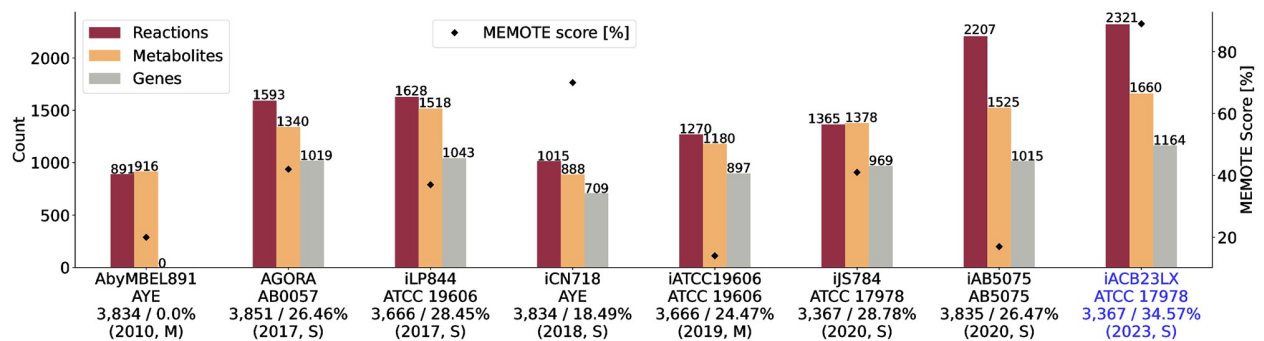
Fig 2. Properties of all metabolic networks for *A. baumannii*. Blue highlights the metabolic network for ATCC 17978 presented in this publication. The left ordinate shows the counts, while the right ordinate represents the MEMOTE scores. The abscissa labels are annotated with the respective strains, each accompanied by the count of open reading frames (ORFs) and the percentage of model gene coverage. The reconstruction process is divided into manual (M, nocomputational tool was used to reconstruct and refine the model) and semi-automated (S, draft obtained via an automated reconstruction tool, while further extension was done manually) and is written together with the publication year. The new model presented in this work exhibits the highest quality score and is more comprehensive and complete than the preceding reconstructions.

https://doi.org/10.1371/journal.ppat.1012528.g002

zero connectivity and no organism-specific gene evidence were omitted from the gap-filling. We extended the draft model by 138 reactions, 77 genes, and 110 metabolites in three compartments (cytosol, periplasm, and extracellular space). All in all, *i*ACB23LX comprises 2,321 reactions, 1,660 metabolites, and 1,164 genes (Fig 2). It is the most comprehensive model, while its stoichiometric consistency lies at 100% and contains no unconserved metabolites. Over 1,800 reactions have a GPR assigned, while 149 are catalysed by enzyme complexes (GPR contains at least two genes connected via a logical AND).

Furthermore, we tested our model for EGCs to prevent having thermodynamically infeasible internal loops that bias the final predictions [41]. We defined energy dissipation reactions (EDRs) for 15 energy metabolites, enabling transmission of cellular energy. Each reaction was individually added to the model and set as the objective function, while all uptakes were constrained to zero (see Materials and methods). A non-zero optimization result indicated an energy-generating cycle, which was then removed. With this, our final model, *i*ACB23LX, contains no EGCs. As shown in Fig 1, a plethora of database cross-references was embedded in the model, while SBO terms were defined for every reaction, metabolite, and gene [43]. Additionally, each reaction was mapped to an ECO term representing the confidence level and the assertion method (Fig 3).

To assess the model's quality, we utilized the Metabolic Model Testing tool (MEMOTE) [45] and the SBML [40] Validator from the libSBML library [46]. Our metabolic network, *i*ACB23LX, achieved a MEMOTE score of 89% with all syntactical errors resolved. Our model undoubtedly exhibits the highest quality score among its predecessors (Fig 2). Notably, the MEMOTE testing algorithm considers only the parent nodes of the SBO directed acyclic graph and not their respective children. Assigning more representative SBO terms does not increase the final score but reduces it by 2%. Finally, we assessed the model's reproducibility using FROG analyses [47] and submitted the reports along with our model to enable verification of results. The final model is available in SBML Level 3 Version 2 [48] and JavaScript Object Notation (JSON) formats with the fbc and groups plugins available.

*i*ACB23LX is of high quality and exhibits an increased predictive accuracy

Prediction and experimental validation of bacterial growth on various nutritional environments. Constraint-based modeling approaches, such as FBA, estimate flux rates
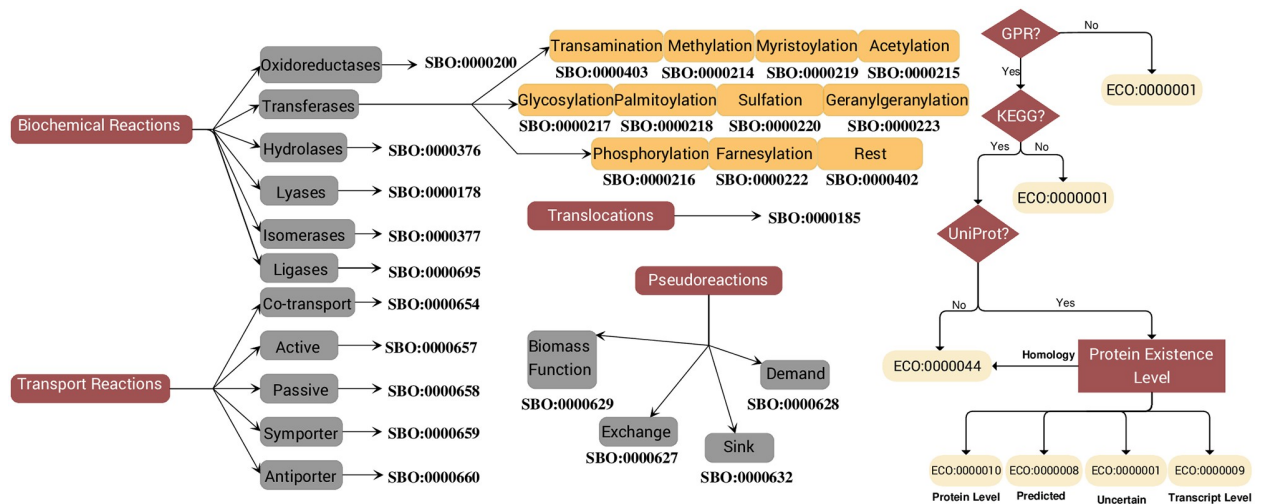
Fig 3. Schematic representation of the SBO and ECO terms mapping. It follows the graphs defined in the repository for biomedical ontologies Ontology Lookup Service (OLS) [42]. The SBO terms were added using the SBOannotator tool [43]. The ECO terms annotated metabolic reactions and were declared based on the presence of GPR along with KEGG and UniProt annotations. Providing UniProt identifiers, the Protein Existence Level guides the mapping to appropriate ECO terms. Figure created with yEd [44].

indicating how metabolites flow through the metabolic network and predict cellular phenotypes for various growth scenarios. *A. baumannii* is known to be strictly aerobic and, compared to the majority of *Acinetobacter* species, it is not considered ubiquitous in nature. As a nosocomial pathogen, it has been mostly detected in hospital environments, particularly in the ICUs, and within the human nasal microbiota [10–12]. We examined various growth conditions to ensure that *i*ACB23LX recapitulates these already known and fundamental phenotypes.

First, we tested our model's capability to simulate a strictly aerobic growth. For this purpose, we examined the directionality of all active oxygen-producing and -consuming reactions when the oxygen uptake was disabled (see S1 Fig). We observed an accumulation of periplasmic oxygen by reactions that carried remarkably high fluxes, leading to growth even when oxygen import was turned off. We examined each reaction individually and removed those without gene evidence to correct this. More specifically, we removed the periplasmic catalase (CATpp), one of bacteria's main hydrogen peroxide scavengers. This enzyme is typically active in the cytosol [49] and was not part of any precursor *A. baumannii* GEM or was found only in cytosol (*i*LP844 [31]). To fill the gap and enable the usage of the periplasmic hydrogen peroxide, we added the (PEAMNOpp) in the model. Eventually, *i*ACB23LX demonstrated growth only in the presence of oxygen using a rich medium (all exchange reactions are open).

Furthermore, we determined the minimal number of metabolites necessary for growth using *i*ACB23LX and the M9 as a reference. Minimal growth media typically consist of carbon, nitrogen, phosphorus, and sulfur sources, as well multiple inorganic salts and transition metals. These metals are crucial for the growth and survival of all three domains of life; however, they can be transformed into toxic compounds in hyper-availability [50]. The exact composition of our minimal medium (*i*MinMed) is shown in Table 1. It comprises nine transition metals, acetate as the carbon source, ammonium as a nitrogen source, sulfate as a sulfur source, and phosphate as a phosphorus source. Previous studies have highlighted the importance of nutrient metals for *A. baumannii* to survive within the host. More specifically, the bacterium utilizes these metals as co-factors for vital cellular processes [51]. Manganese and zinc have

Table 2. Simulated and empirical growth rates of ATCC 17978 in various growth media. The tested media are the computationally-defined minimal medium (*i*MinMed), the LB, and the SNM. Computational growth rates are given in mmol/($g_{DW} \cdot$ h), while *in vitro* rates in h^{-1}. Doubling times are calculated in minutes. The media formulations are available in S1 Table.

	Growth Rates		Doubling Times	
	in silico	*in vitro*	*in silico*	*in vitro*
*i*MinMed	0.5503	0.5402	75.57	76.70
LB	0.6065	0.7369	68.57	56.44
SNM	0.2914	0.3592	142.72	115.78

https://doi.org/10.1371/journal.ppat.1012528.t002

also been studied as essential determinants of host defense against *A. baumannii*-acquired pneumonia through their sequestering by calprotectin via a type of bonding called chelation [52]. In the computational simulations, growth rates below 2.81 mmol/($g_{DW} \cdot$ h) were considered realistic. This threshold corresponds to the doubling time of the fastest-growing organism, *Vibrio natriegens*, which is 14.8 minutes [45]. Table 2 displays the predicted growth rates of *i*ACB23LX in the respective culture media. In LB, our model simulated with the highest growth rate; 0.6065 mmol/($g_{DW} \cdot$ h). With our self-defined minimal medium, *i*MinMed, our model exhibited the lowest rate; 0.5503 mmol/($g_{DW} \cdot$ h). Notably, our experimental validation revealed similarity between the growth rates obtained from the *in vitro* respiratory curves (Table 2 and S2 Fig) and those predicted by our *in silico* simulations. Additionally, we examined the growth rate of our model in a rich medium, in which all nutrients were available to the model. With this, the flux through the biomass production was the highest, 2.1858 mmol/($g_{DW} \cdot$ h), as expected. This is still less than the growth rate of the fastest organism, increasing the confidence in our model's consistency and simulation capabilities. Initially, *i*ACB23LX could not predict any realistic growth rate for the simulated media. Using the gap-filling function of CarveMe [39], we detected three enzymes (PHPYROX, OXADC, and LCYSTAT) whose addition into the metabolic network resulted in successful growth in all tested media.

Functional validation of *i*ACB23LX using nutrient utilization data. Multiple *in silico* approaches have hitherto been employed to predict lethal genes and to assess growth metrics on different carbon/nitrogen sources for severe pathogenic organisms including *Mycobacterium tuberculosis* [53, 54] and *Staphylococcus aureus* [55, 56]. In 2013 and 2014, two studies were published that examined the catabolic phenome and gene essentialities of the strain ATCC 17978 [26, 57]. We used these datasets to evaluate the overall performance (functionality and accuracy) of *i*ACB23LX.

Our first validation experiment assessed the accuracy of our model's carbon and nitrogen catabolism potentials using the large-scale phenotypic data provided by Farrugia *et al.* [26]. While the authors tested a larger number of compounds overall, we were only able to examine 80 carbon sources and 48 nitrogen sources. For the remaining molecules, either no BiGG identifier existed, or they were not part of the metabolic network. Following the experimental protocol by Farrugia *et al.*, we applied the M9 medium and enabled D-xylose as the sole carbon source for the nitrogen testings. As D-xylose was initially not part of the reconstructed network, we conducted extensive literature and database search to include associated missing reactions. This improved the prediction accuracy, especially for the carbon sources, where an amelioration of 19% was achieved. In more detail, despite the comprehensive manual curation, the first draft model was reconstructed using the automated tool CarveMe [39]. This resulted in the incorrect inclusion of transport reactions, which were consequently removed to reduce the number of false positive predictions. In both cases, our main objective was to improve the accuracy while keeping the number of orphan and dead-end metabolites low and removing

only reactions with no gene evidence (lack of assigned GPR). Similarly, missing reactions were identified and included in the network to eliminate the false negative predictions. For instance, in accordance with the phenotypic data, ATCC 17978 should not be able to grow when utilizing D-trehalose as the sole carbon source. Our model initially predicted a growth phenotype for this carbon source. To overcome this conflict, we deleted the reaction `TREP` with no organism-specific gene evidence, i.e., no assigned GPR. However, it was not feasible to resolve all inconsistencies since adding transport reactions to resolve false positives or false negatives in the nitrogen testings led to more false predictions in the carbon sources. More specifically, when adenosine, inosine, L-homoserine, and uridine are utilized as sole carbon sources, the model should not predict growth, while as sole nitrogen sources they should result in a non-zero objective value. In this case, adding transporters would resolve false predictions in the nitrogen tests, while it would have induced more false predictions in the carbon sources tests. Altogether, *i*ACB23LX exhibited an overall accuracy of 86.3% for the carbon and 79.2% for the nitrogen sources test (Fig 4c). By adding their corresponding transport reactions, we resolved discrepancies regarding uridine, inosine, adenosine, and L-homoserine. Our model was able to catabolize 49 sole carbon and 40 sole nitrogen sources (see Fig 4a and 4b), recapitulating totally 69 and 38 experimentally-derived phenotypes, respectively.

We further assessed the ability of *i*ACB23LX to predict known gene essentialities. First, 1,164 *in silico* single gene deletions were conducted on both LB and rich growth media, respectively to identify all lethal gene deletions. Subsequently, the ratio between the growth rate after and before the respective knockouts (FC_{gr}) was calculated, and the genes were classified accordingly (see Materials and methods). For the optimization, two mathematics-based approaches from the COBRApy [58] package were deployed: the FBA [59] and the MOMA [60]. Between the two methods, a similar distribution of the FC_{gr} values was observed (Fig 5a and 5b). Using FBA, 97, 75, and 991 genes were predicted to be essential, partially essential, and inessential on LB, respectively. Similarly, optimization with MOMA resulted in 110, 85, and 968 genes (Fig 5c and S3 and S2 and S3 Tables). These genes were primarily associated with the biosynthesis of cofactors and vitamins, the amino acid/nucleotide metabolism, the energy metabolism, and the metabolism of terpenoids and polyketides. Additionally, we examined in more detail how nutrition availability impacts the gene essentiality by conducting single-gene knockouts in the rich medium. Both optimization methods yielded a higher number of essential genes when the model had to adapt its metabolic behavior due to lacking nutrients, i.e., with LB, compared to the rich medium (Fig 5c and S2 and S3 Tables). In general, FBA detected more genes to be dispensable for growth in both nutritional environments. On the other hand, MOMA classified more genes as essential or partially essential (Fig 5c and S2 and S3 Tables), while genes from FBA build a subset of the essential genes derived by MOMA. Furthermore, we validated the prediction accuracy of *i*ACB23LX using gene essentiality data. First, we analyzed the transposon mutant library developed by Wang *et al.* as it examines the same *A. baumannii* genome that we used to build our model [57].

Using this dataset and the LB medium, our model demonstrated a prediciton accuracy of 87% with both optimization methods (Fig 5d). To enhance the robustness and generalizability of our model across diverse datasets, we compared our model's predictions to four additional Tn-seq datasets. A total of 43 genes were labeled as essential in all studied Tn-seq data and were predicted as essential by our model. The derived predictive accuracies ranged between 86.6% and 88.9% (Fig 5e), underscoring the efficacy of our model across diverse high-throughput gene essentiality datasets. We further analyzed the predicted false negative genes and probed their proteomes to investigate the existence of human homologs (see Materials and methods and S4 Table). Our aim was to remove cross-linkings to human-similar proteins, as pathways or enzymes absent in humans are valuable sources of druggable targets against
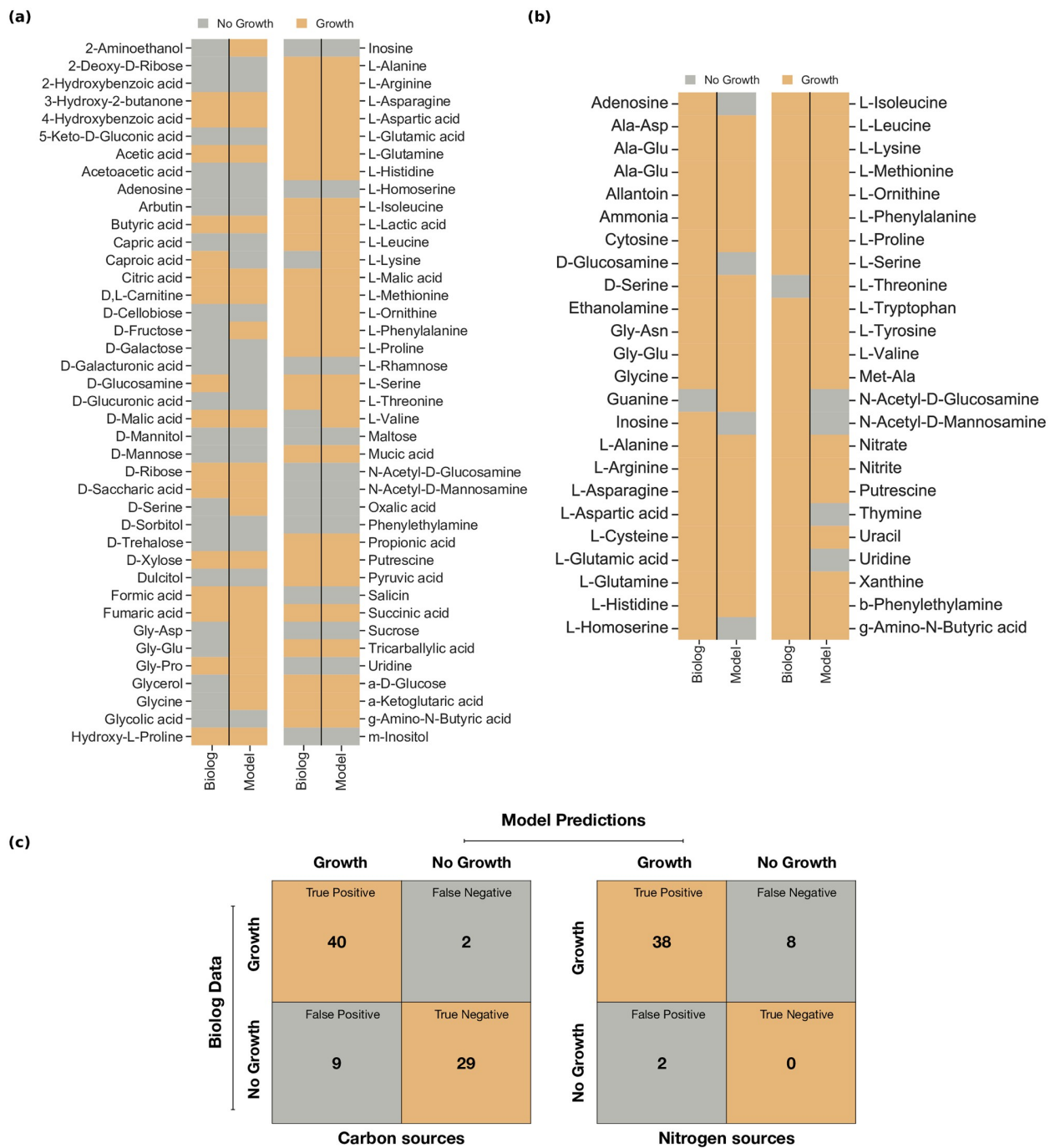
Fig 4. Model predictions compared to the Biolog experimental measurements for various carbon and nitrogen sources. From the Biolog data, only substances mappable to model metabolites were included, while the M9 medium was applied. **(a)** and **(b)** The model's ability to catabolize various carbon and nitrogen sources was assessed using the strain-specific phenotypic data by Farrugia *et al.* [26]. Grey indicates no growth, and orange indicates growth. Totally, 80 and 48 compounds were tested as sole carbon and nitrogen sources, respectively. Out of these, 69 and 38 phenotypes were recapitulated successfully by *i*ACB23LX. **(c)** Confusion matrices of model predictions and Biolog experimental measurements. The overall accuracy of *i*ACB23LX is 86.3% for the carbon (left matrix) and 79.2% for the nitrogen (right matrix) testings. Orange represents correct predictions, and grey represents wrong predictions.

https://doi.org/10.1371/journal.ppat.1012528.g004

(a)

(b)

(c)

(d)

(e)

FBA

	TP	FN	FP	TN	ACC
Gallagher *et al.* (2015)	960	26	107	71	88.57%
Geisinger *et al.* (2020)	960	39	107	58	87.46%
Bai *et al.* (2021)	952	38	115	59	86.86%
Bai *et al.* (2023)	972	34	95	63	88.92%

MOMA

	TP	FN	FP	TN	ACC
Gallagher *et al.* (2015)	952	34	102	76	88.32%
Geisinger *et al.* (2020)	952	47	102	63	87.20%
Bai *et al.* (2021)	944	46	110	64	86.60%
Bai *et al.* (2023)	964	42	90	68	88.66%
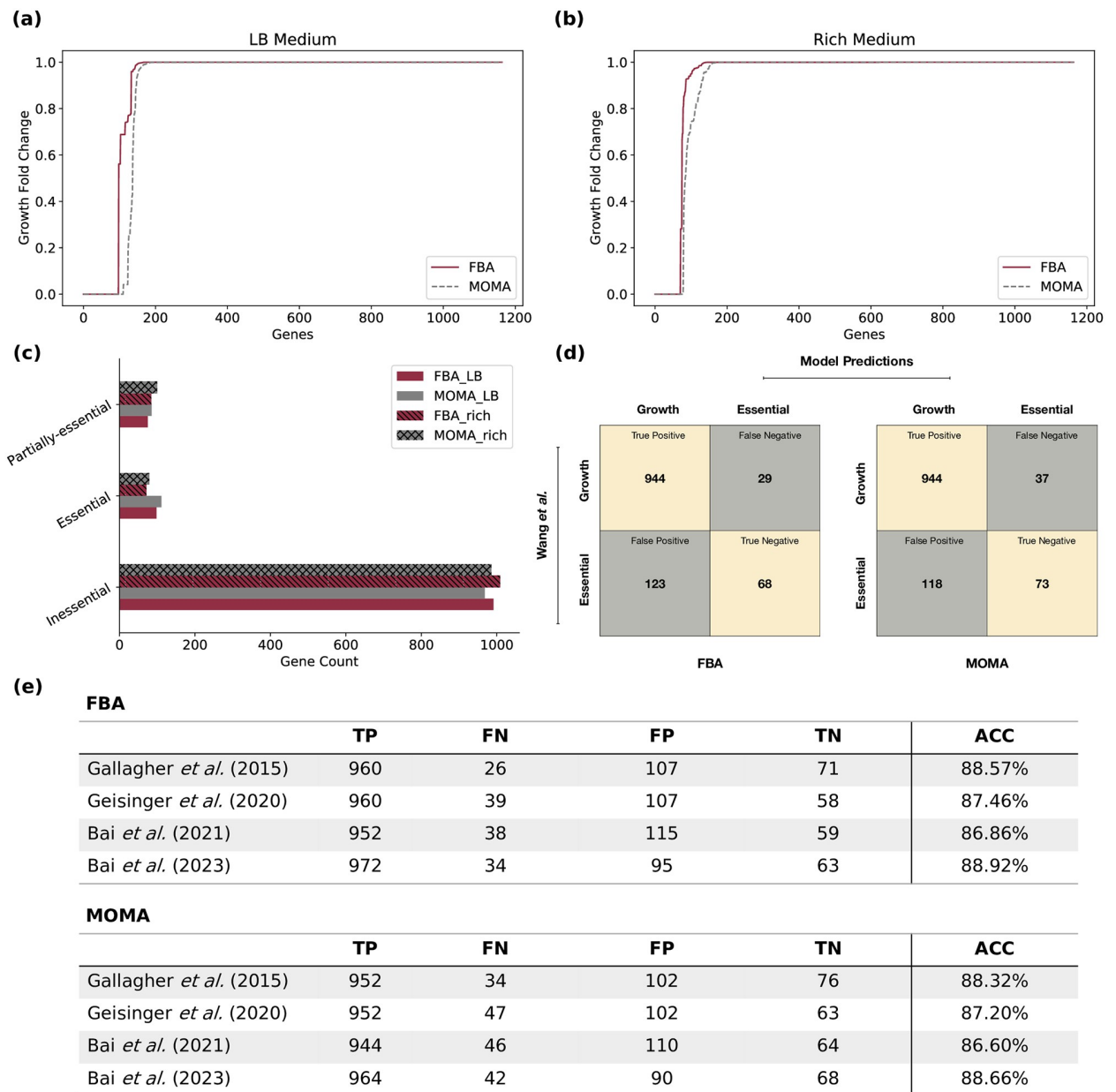
Fig 5. Gene essentiality analysis using *i*ACB23LX. (a) and **(b)** Distribution of the FC_{gr} values calculated for all genes included in *i*ACB23LX. Red lines represent FBA predictions and grey are ratios derived with MOMA. Totally 1,164 knockouts were conducted using each method in LB and rich media. (**c**) Classification of gene essentialities in essential, inessential, and partially essential based on their FC_{gr} values. (**d**) Accuracy of gene essentiality predictions based on empirical data. The *in silico* results were compared to the Wang *et al.* transposon library [57]. The LB medium was applied to mirror the experimental settings. The metabolic network exhibited 87% accuracy with FBA (left) and MOMA (right). Beige indicates correct predictions; grey indicates incorrect predictions. (**e**) Comparative analysis of essential genes predicted by *i*ACB23LX versus those identified in multiple Tn-seq studies.

infectious diseases [61]. Out of the 37 genes that our model predicted to be essential using FBA and MOMA, which contradicted the experimental results, 17 were found to be non-homologous to human (S4 Table). Genes with a zero similarity score were defined as non-homologous, while those with a non-zero similarity score were excluded to avoid targeting human-

like proteins. Some examples are the genes encoding the enolpyruvylshikimate phosphate (EPSP) synthase (`A1S_2276`), chorismate synthase (`A1S_1694`), riboflavin synthase (`A1S_0223`), phosphogluconate dehydratase (`A1S_0483`), dihydrofolate reductase (DHFR) (`A1S_0457`), and 2-keto-3-deoxy-6-phosphogluconate (KDPG) aldolase (A1S 0484). The EPSP synthase converts the shikimate-3-phosphate together with phosphoenolpyruvate to 5-o-(1-carboxyvinyl)-3-phosphoshikimic acid. Subsequently, the chorismate synthase catalyses the conversion of the 5-o-(1-carboxyvinyl)-3-phosphoshikimic acid to chorismate, the seventh and last step within the shikimate pathway [62]. Chorismate is the common precursor in the production of the aromatic compounds tryptophan, phenylalanine, and tyrosine, as well as folate and menaquinones during the bacterial life cycle. The shikimate pathway is of particular interest due to its absence from the human host metabolome and its vital role in bacterial metabolism and virulence. Moreover, the enzyme riboflavin synthase catalyses the final step of riboflavin (vitamin B2) biosynthesis with no participating cofactors. Riboflavin can be produced by most microorganisms compared to humans, who have to externally uptake them via food supplements. Also, it plays an important role in the growth of different microbes, especially due to its photosynthesizing property that marks it as a non-invasive and safe therapeutic strategy against bacterial infections [63]. Lastly, the phosphogluconate dehydratase catalyses the dehydration of 6-phospho-D-gluconate to KDPG, the precursor of pyruvate and 3-phospho-D-glycerate [64]. This enzyme is part of the Entner–Doudoroff pathway that catabolizes glucose to pyruvate, similarly to glycolysis, but using a different set of enzymes [65].

We further assessed the druggability of our essential non-homologous proteins and investigated the existence of inhibitors or compounds known to interact with the enzymes. For this, we used the online DrugBank database that contains detailed information on various drugs and drug targets [66]. In all cases, the listed drugs are of unknown pharmacological action, and there is still no evidence indicating the enzymes' association with the molecule's mechanism of action. For instance, the flavin mononucleotide and the cobalt hexamine ion were listed as known inhibitors of yet unknown function against the chorismate synthase, while glyphosate, shikimate-3-phosphate, and formic acid have been experimentally found to act with EPSP synthase. Six non-homologous genes were marked as hypothetical or putative in the KEGG [38] database and/or lacked enzyme-associated information. We searched for drug leads by aligning the query sequences against the DrugBank's database to find homologous proteins. Two out of six were found to have a protein hit. More specifically, the protein encoded by `A1S_0589` was found to have high sequence identity with the phosphocarrier protein HPr of *Enterococcus faecalis* (Bit-score: 48.5), while the translation product of `A1S_0706` resembles the sugar phosphatase YbiV of *Escherichia coli* (Bit-score: 225.3). According to DrugBank, dexfosfoserine and aspartate beryllium trifluoride have been experimentally determined to bind to these enzymes; however, their pharmacological action is still unknown. The S4 Table lists all non-homologous essential genes reported for *i*ACB23LX.

Overall, *i*ACB23LX exhibits high agreement to all validation tests and can, therefore, be used to systematically derive associations between genotypes and phenotypes.

A curated collection of *A. baumannii* metabolic models

In 2010, Kim *et al.* published the first GEM for the multidrug-resistant strain *A. baumannii* AYE [30]. After that, multiple studies provided new data and genomic analyses were published, paving new ways towards its update and refinement [26, 57, 67, 68]. Since then, a variety of GEMs was developed aiming at the empowering of drug development strategies and the enforcement of metabolic engineering by formulating new and reliable hypotheses (Table 3). However, the amount and format of information contained are inconsistent, with some being

Table 3. List of genome-scale metabolic models curated for *A. baumannii*, along with information relevant to the manual refinement. Default growth rates (i.e., model simulated as downloaded), the cellular compartments (C: cytosol, E: extracellular space, P: periplasm, and ER: endoplasmic reticulum), and the reactions and metabolites identifiers are listed in the table. MEMOTE scores before and after manual curation are given in the last column. Dark red highlights our reconstruction for the strain ATCC 17978. After manual curation, our model developed following our workflow in Fig 1 has the highest quality score and comes along with a minimal medium defined.

	Availability	Used Identifiers	Growth by default mmol/($g_{DW} \cdot$ h)	Compartments	MEMOTE	BioModels ID
AbyMBEL891 [30]	BioModels	Customized	119.0	Cell	20% + 17%	MODEL2406250010
AGORA [32]	VMH	VMH	134.0	C, E	42% + 37%	MODEL2406250011
*i***LP844** [31]	Suppl. Mat	ModelSeed	15.88	C, E, P	37% + 21%	MODEL2406250005
*i***CN718** [33]	BiGG	BiGG	1.31	C, E, P, ER	70% + 3%	MODEL2406250007
*i***ATCC19606** [34]	Suppl. Mat	KEGG	46.34	C, E	14% + 44%	MODEL2406250008
*i***JS784** [36]	GitHub	ModelSeed	0.0	C, E, P	41% + 18%	MODEL2406250006
*i***AB5075** [35]	Suppl. Mat	BiGG	1. 729	C, E, P	17% + 50%	MODEL2406250009
*i***ACB23LX**	BioModels	BiGG	0.5503	C, E, P	89%	MODEL2309120001

https://doi.org/10.1371/journal.ppat.1012528.t003

syntactically invalid or of older formats. Here, we systematically analyzed the quality of all seven currently existing GEMs, reporting their strengths and weaknesses and debugging them to finally build a curated, standardized, and updated collection. To do so, we developed a workflow with curation steps applicable to all models aiming at the standardization and usability of published GEMs by the community (Fig 6a). This closely follows the community-driven workflow published by Carey *et al.* for the reconstruction of reusable and translatable models [37]. The curation procedure includes a series of stages aiming at modifying data format, data amount, and information quality. It is important to note that no contextual modifications were conducted that could affect the model's prediction capabilities (see Materials and methods).

Five *A. baumannii* strains have been reconstructed throughout the years, with AYE and ATCC 19606 having two reconstructions each. All models are publicly stored and can be downloaded either from a database/repository [BioModels, Virtual Metabolic Human (VMH) [69], BiGG [70], and GitHub] or directly from the publication's additional material. The use of
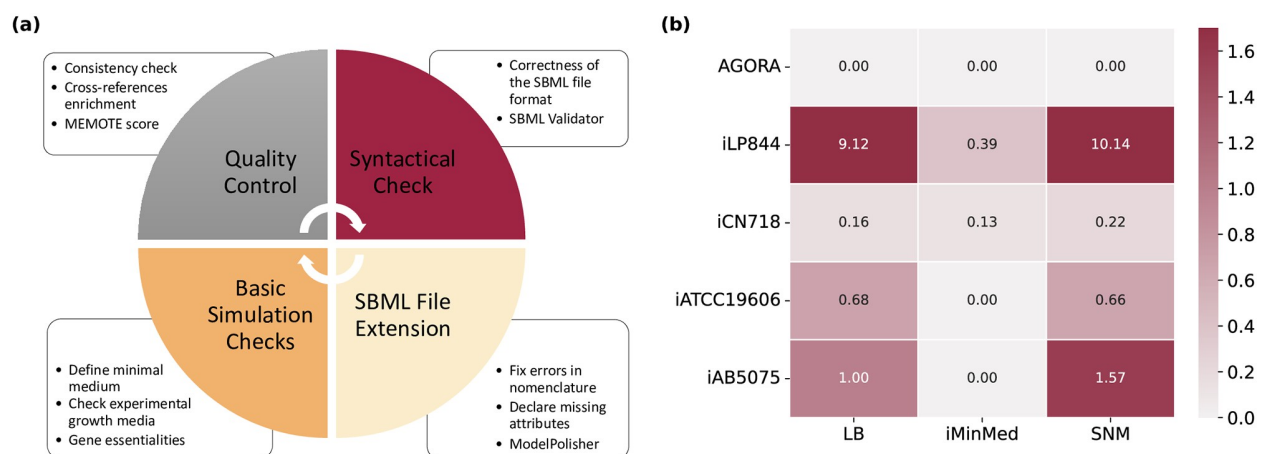
Fig 6. Collection of strain-specific *A. baumannii* metabolic models. (a) Debugging workflow to curate and evaluate already published models. Following the community standards, the existing *A. baumannii* models were curated and transformed into re-usable, simulatable, and translatable models. Quality controls and metabolic standardized tests were conducted using `MEMOTE`, while the validity of the file format and syntax were examined with the SBML Validator [46]. ModelPolisher enhanced the models with missing metadata. **(b)** *In silico*-derived growth rates in various media. The empirical and predicted growth rates of *i*ACB23LX are listed in Table 2.

https://doi.org/10.1371/journal.ppat.1012528.g006

distinct identifiers prevents the metabolic networks from being compared to each other. More specifically, *i*LP844 and *i*JS784 carry ModelSEED [71] identifiers for reactions and metabolites, while *i*CN718 and *i*AB5075 BiGG [70] identifiers. AbyMBEL891 uses distinct identifiers not supported by any database, and *i*ATCC19606 includes identifiers derived from KEGG [38]. Most of the models resulted in an unrealistic and inflated growth rate (reference: doubling time of the fastest growing organism *V. natriegens*) in their defined medium, while *i*JS784 showed a zero growth even when all imports were enabled (Table 3). Hence, this model was excluded from further analysis. For each of the remaining GEM, we defined the minimum growth requirements that result in a non-zero and realistic objective value. For instance, the AGORA model required at least 21 compounds (mostly metal ions), while oxygen was sufficient for AbyMBEL891 to simulate a non-zero growth (S5 Table).

Since these models should successfully reflect the bacterium's metabolic and growth capabilities (S2 Fig), we examined the flux through their biomass reaction in various growth media known to induce *A. baumannii* growth (Fig 6b). The majority resulted in a biomass flux of 0.0 mmol/($g_{DW} \cdot$ h) in the *i*MinMed, while the AGORA model could not simulate growth in LB and SNM as well. Thus, we investigated and identified minimal medium supplementations needed to enable cellular biomass production. As already mentioned, *i*JS784 was excluded from further examination (Table 3), together with AbyMBEL891 that debilitated the analysis due to its non-standardized identifiers and its missing genes and GPRs. When the *i*MinMed for *i*ATCC19606 and *i*AB5075 was supplemented with D-alanine and D-glucose 6-phospahte as well as guanosine 5'-phosphate (GMP), respectively, their biomass reactions carried a positive flux rate of 0.5279 mmol/($g_{DW} \cdot$ h) and 0.6477 mmol/($g_{DW} \cdot$ h). Supplementation of meso-2,6-diaminoheptanedioate, menaquinone-8, niacinamide, heme, siroheme, and spermidine into the medium of the AGORA model resulted in a positive growth rate of 1.9430 mmol/($g_{DW} \cdot$ h). Similarly, when supplementing the SNM with glycyl-L-asparagine, the derived growth rate was 1.5020 mmol/($g_{DW} \cdot$ h), while the *i*MinMed needed to be extended with 12 additional components (resulted growth rate: 1.2789 mmol/($g_{DW} \cdot$ h)). Lastly, like with *i*ACB23LX, the LB medium, together with FBA and MOMA, were applied to detect lethal genes in all models (S6 and S7 Tables). Despite remarkable efforts, we could not derive a mapping scheme between the strain-specific gene identifiers of *i*LP844 and *i*ATCC19606 to resolve PROKKA or HMPREF identifiers. Similar issues arose with *i*AB5075. Thus, a strain-wise comparison of essential genes would be feasible only for the strain ATCC 17978. Subsequently, we examined which genes were necessary for growth among the remaining models across three different strains: AYE (*i*CN718), ATCC 17978 (*i*ACB23LX), and AB0057 (AGORA). Totally, 392 genes were identified as essential, while 34 occurred in all three strains. For instance, when the genes encoding for dephospho-CoA kinase, phosphopantetheinyl transferase, shikimate kinase, or chorismate synthase were deleted from the three strains, no growth could be simulated in the LB medium. As already mentioned, the gene encoding the chorismate synthase has no human-like counterpart. This, together with the fact that it was detected to be vital for growth across three distinct strains, increases its potential to be a drug candidate for future therapies. Generally, most essential genes are members of the purine metabolism and encode various transferases. Besides this, the pantothenate and CoA biosynthesis and the amino acid metabolism were found to be a prominent target pathways for further drug development.

Discussion

The historical timeline of past pandemics shows the imposed threat of bacteria in causing repetitive outbreaks with the highest death tolls [72], such as cholera and plague. By 2050, antimicrobial-resistant pathogens are expected to kill 10 million people annually [73], while the

antibiotics misuse accompanied by the ongoing Coronavirus Disease 2019 (COVID-19) crisis exacerbated this global threat. It is noteworthy that elevated morbidity rates were ascribed to bacterial co/secondary infections during previous viral disease outbreaks [74–76]. Hence, developing effective antibiotic regimens is of urgent importance. Here, we present the most recent and comprehensive ready-to-use blueprint GEM for the Gram-negative pathogen *A. baumannii*. For this, we created a workflow that applies to any living organism and ensures the reconstruction of high-quality models following the community standards. Our model, *i*ACB23LX, was able to simulate growth in SNM that mimics the human nasal niche, the experimentally defined medium LB, and the model-derived *i*MinMed. With *i*MinMed we denoted the minimal number of compounds needed to achieve non-zero growth. This medium contains totally 14 compounds, including transition metals and energy sources. Transition metals have been shown to participate in important biological processes and are vital for the survival of living organisms [50]. We confirmed the computationally predicted growth rates by comparing them to our empirically determined growth kinetics data. With this, we ensured that our model recapitulates growth phenotypes in media that reflect *Acinetobacter*-associated environments.

Furthermore, we validated *i*ACB23LX quantitatively and qualitatively using existing experimental data and observed remarkable improvements compared to precursory models. More specifically, our model predicted experimental Biolog growth phenotypes on various carbon sources with an overall agreement of 86.3%. This agreement is higher than the prediction capability of *i*ATCC19606 (84.3%) and *i*LP844 (84%), and comparable to that of *i*AB5075 (86.3%). Similarly, *i*ACB23LX exhibited 79.2% predictive accuracy on nitrogen sources tests, while this increases to 87.5% after further refinement. Improving and re-defining the biomass objective function (BOF) based on accurate strain-specific experimental data would be the next step to diminish the number of inconsistent predictions and to further improve the network and its predictive potential. During gene lethality analysis in LB medium, our model predicted 110 genes with MOMA to be essential, while 97 of them were also reported by FBA to impair the growth. Generally, after enriching the nutritional input with all available compounds (rich medium), less lethal genes resulted, meaning that *A. baumannii* undergoes metabolic alterations when nutrients are lacking. Our *in silico* results, when compared to five different strain-specific gene essentiality data [57, 68, 77–79], achieved accuracies between 88.60% and 88.92%. The predictive accuracies are remarkably higher than all GEMs built for *A. baumannii* (e.g., 80.22% for *i*CN718 and 72% for *i*LP844), except *i*AB5075 which performed comparably. This comprehensive analysis ensures that our model is well-validated and highly reliable, making it a valuable tool for predicting gene essentiality in various biological and experimental settings.

Subsequently, we examined more carefully our false negative predictions and searched for putative drug targets that could be employed for future therapeutics. More specifically, we focused on genes found to be essential for growth and encode proteins with no human counterparts (S4 Table). Our study highlighted EPSP and chorismate synthases from the shikimate pathway as prominent target candidates with no correlation to the human proteome. Several knockout studies have highlighted the importance of enzymes from the shikimate metabolism as potential targets against infections caused by threatening microorganisms, e.g., *Mycobacterium tuberculosis* [80], *Plasmodium falciparum* [81], and *Yersinia enterocolitica* [82]. Umland *et al.* identified these two gene products as essential in an *in vivo* study using a clinical isolate of *A. baumannii* (AB307-0294) and a rat abscess infection model [83]. This increases the confidence of our results and indicates that genes found to be essential *in silico* should be considered as potential antimicrobial targets. Moreover, our predicted target DHFR has been

extensively studied as a primary target for antibacterial and anticancer drug development, given its pivotal role in nucleotide biosynthesis [84–88].

Trimethoprim (TMP), an antifolate antibacterial agent, selectively inhibits the bacterial DHFR, a crucial enzyme catalyzing the tetrahydrofolic acid (THF) formation. Sulfamethoxazole (SMX), a bacteriostatic sulfonamide antibiotic, competitively inhibits dihydropteroate synthase, responsible for the formation of dihydrofolate (DHF). Together as cotrimoxazole or TMP-SMX, both compounds effectively treat bacterial infections by collectively inhibiting folic acid synthesis, essential for bacterial growth and replication [89, 90]. Further studies have shown that antifolates effectively inhibit *A. baumannii* DHFR, demonstrating potent antibacterial activity against multidrug-resistant strains and highlighting their potential for further antibiotic development [91, 92]. Finally, targeting folate biosynthesis is a well-established strategy against infectious diseases due to its absence in higher eukaryotic organisms. Altogether, selectively targeting bacterial proteins vital for key cellular functions, such as cell wall biosynthesis, translation, and deoxyribonucleic acid (DNA) replication, is a well-established strategy in antibiotic development. Similarly, numerous studies have suggested one of our further candidates, riboflavin, as a potential antimicrobial agent for further investigation [63]. Additionally, the Entner–Doudoroff pathway, in which our candidate targets phosphogluconate dehydratase and KDPG aldolase act to produce pyruvate, is similar to glycolysis but involves different enzymes. This pathway has been firstly discovered in *P. saccharophila* [65] and later in *E. coli* [93]. Meanwhile, it is vital for the survival of further pathogenic microorganisms, like *Neisseria gonorrhoeae*, *K. pneumoniae*, and *P. aeruginosa* [94–96]. However, these targets have not yet been examined in *Acinetobacter* species and could be a source of antimicrobial therapeutic strategies. Hence, these biosynthetic routes could be a valuable resource for targets to fight bacterial infectious diseases. Finally, we investigated the druggability of our essential non-homologous genes. We searched the DrugBank database to find compounds known to inhibit these genes and that are already approved by the Food and Drug Administration (FDA). Our analysis resulted in drugs that have been found to interact with the gene product of interest; however their pharmacological action is yet unknown. We further probed the hypothetical and putative non-homologous genes against the DrugBank's sequence database to find homologous proteins and determine their activity. Also in this case, the resulted drugs were listed with still undetermined pharmacological action. These putative and yet unexplored targets with inhibitory potential are of great interest in the context of developing new classes of antibiotics. Overall, our model reached a MEMOTE score of 89%, which is the highest score reported for this organism.

Moreover, we improved and assessed all previously published models and created the first curated strain-specific collection of metabolic networks for *A. baumannii*. We created a debugging workflow consisting of four major steps to systematically analyze and curate constraint-based models focusing on their standardization and the FAIR data principles. We applied this workflow and curated a total of seven metabolic models for *A. baumannii*. In addition, most of the models simulated growth rates by default that were unrealistic when compared to the fastest growing organism [45]. Therefore, we determined the minimal number of components needed for these models to result in non-inflated biomass production rates. The defined minimal media were mostly composed of metal ions (e.g., cobalt, iron, magnesium) that are essential for bacterial growth. For the model *i*JS784, the minimization process was infeasible; thus, the model was not considered for further analysis. We also examined the growth ability of these models in three media (SNM, LB, and *i*MinMed) and compared them to our model, *i*ACB23LX. When the models simulated a zero flux through the biomass reaction, we continued by detecting the minimal amount of metabolites supplemented in the medium that resulted in a non-zero growth rate. These would enable the detection of gaps and assist in future improvement of the models. It is important to note here that with this curation, we

opted for a systematical assessment of the previously reconstructed models and the detection of their assets and liabilities. Consequently, we did not undertake any contextual modification that could alter the models' predictive capabilities. Finally, we predicted lethal genes among comparable and simulatable models of *A. baumannii*. Our analysis incorporated three strains (AYE, AB0057, and ATCC 17978), and we examined the effect of genetic variation across strains in the gene essentiality. We highlighted once again the shikimate pathway, as well as the purine metabolism, the pantothenate, and CoA biosynthesis, and the amino acid metabolism as candidate routes to consider for future new classes of antibacterial drugs with potential effect across multiple *A. baumannii* strains. The curated models, together with our model, would benefit the future prediction of candidate lethal genes by reducing the considerable resources needed for classical whole-genome essentiality screenings. All in all, this collection of simulation-ready models will forward the selection of a suitable metabolic network based on individual research questions and help define the entire species and new hypothesis.

Our new metabolic reconstruction and the curated collection of further strain-specific models will guide the formulation of ground-breaking and reliable model-driven hypotheses about this pathogen and help examine the diversity in the metabolic behavior of different *A. baumannii* species in response to genetic and environmental alterations. Additionally, they can be utilized as knowledge bases to detect critical pathways related to responses against multiple antibiotic treatments. This will ultimately strengthen the development of advanced precision antimicrobial control strategies against multidrug-resistant (MDR) *A. baumannii* strains.

Taken together, our workflows and models can be employed to expand this collection further with additional standardized strain-specific metabolic reconstructions to finally define the core and pan metabolic capabilities of *A. baumannii*.

Supporting information

S1 Fig. Oxygen-producing and -consuming reactions found in *i*ACB23LX together with their anaerobic fluxes. All flux rates are written in orange and are given in mmol/($g_{DW} \cdot$ h). The reaction abbreviations are as follows: `O2tpp`, O_2 transport via diffusion between periplasm and cytosol; `CATpp`, periplasmatic catalase; `H2O2tex`, hydrogen peroxide transport via diffusion; `CAT`, catalase; `O2tex`, O_2 transport via diffusion between periplasm and extracellular space; `EX_h2o2_e`, hydrogen peroxide exchange and `EX_o2_e`, O_2 exchange. Figure generated with Escher [109].
(TIF)

S2 Fig. Experimentally-derived growth curves of *A. baumannii*. The growth curves for *A. baumannii* strains AB5075, ATCC 73217978, ATCC 19606, and AYE were measured in LB and SNM. Additionally, the *in silico*-defined minimal medium (*i*MinMed) was tested for all strains.
(TIF)

S3 Fig. Comparative analysis of essential genes predicted by *i*ACB23LX versus those identified in Tn-seq libraries [57, 68, 77–79]. (a) Metabolic subsystems distribution of all essential genes reported in various Tn-seq studies and predicted using *i*ACB23LX (true negatives). **(b)** Venn diagrams of essential genes from five examined Tn-seq datasets compared to essential genes predicted by the model developed in this study.
(TIF)

S1 Table. *In silico* formulations of examined media compositions. Metabolites are described by BiGG [70] identifiers.
(XLSX)

S2 Table. *In silico* gene knockout results using FBA. The ratio column describes the growth rate change before and after the respective knockout.
(XLSX)

S3 Table. *In silico* gene knockout results using MOMA. The ratio column describes the growth rate change before and after the respective knockout.
(XLSX)

S4 Table. Metabolic genes found to be essential for growth in *i*ACB23LX and encode proteins with no human counterparts.
(XLSX)

S5 Table. Computationally-defined minimal growth media for previously published models. Due to inflated growth rates in most published *A. baumannii* GEMs, we established minimal media supporting non-zero biomass flux.
(XLSX)

S6 Table. Gene lethality predictions using previously published *A. baumannii* models and FBA.
(CSV)

S7 Table. Gene lethality predictions using previously published *A. baumannii* models and MOMA. This offers a complementary perspective on the essential genes in the organism's metabolism.
(CSV)

S8 Table. Summary of essential genes predicted by *i*ACB23LX and confirmed in Tn-seq libraries [57, 68, 77–79]. Five independent Tn-seq datasets were utilized for the comparison. Both FBA and MOMA were used with the LB medium defined to predict essential genes. The table also includes the intersection of essential genes identified by both methods, along with their associated orthologs in the case of different studied genomes. The computed accuracies verified the high predictive performance of our model.
(XLSX)

Acknowledgments

Author Contributions

Conceptualization: Nantia Leonidou, Andreas Dräger.

Data curation: Nantia Leonidou.

Formal analysis: Nantia Leonidou, Lea Friedrich.

Funding acquisition: Monika S. Schütz, Andreas Dräger.

Investigation: Nantia Leonidou, Lea Friedrich.

Methodology: Nantia Leonidou, Yufan Xia.

Project administration: Monika S. Schütz, Andreas Dräger.

Resources: Monika S. Schütz, Andreas Dräger.

Software: Nantia Leonidou, Yufan Xia.

Supervision: Monika S. Schütz, Andreas Dräger.

Validation: Nantia Leonidou.

Visualization: Nantia Leonidou.

Writing – original draft: Nantia Leonidou.

Writing – review & editing: Nantia Leonidou, Monika S. Schütz, Andreas Dräger.

References

1. Magiorakos AP, Srinivasan A, Carey RB, Carmeli Y, Falagas M, Giske C, et al. Multidrug-resistant, extensively drug-resistant and pandrug-resistant bacteria: an international expert proposal for interim standard definitions for acquired resistance. Clinical microbiology and infection. 2012; 18(3):268–281. https://doi.org/10.1111/j.1469-0691.2011.03570.x PMID: 21793988

2. Murray CJ, Ikuta KS, Sharara F, Swetschinski L, Aguilar GR, Gray A, et al. Global burden of bacterial antimicrobial resistance in 2019: a systematic analysis. The Lancet. 2022; 399(10325):629–655. https://doi.org/10.1016/S0140-6736(21)02724-0 PMID: 35065702

3. Fournier PE, Richet H. The epidemiology and control of *Acinetobacter baumannii* in health care facilities. Clinical Infectious Diseases. 2006; 42:692–699. https://doi.org/10.1086/500202 PMID: 16447117

4. Dijkshoorn L, Nemec A, Seifert H. An increasing threat in hospitals: multidrug-resistant *Acinetobacter baumannii*. Nature Reviews Microbiology 2007 5:12. 2007; 5:939–951. https://doi.org/10.1038/nrmicro1789 PMID: 18007677

5. Peleg AY, Seifert H, Paterson DL. *Acinetobacter baumannii*: Emergence of a Successful Pathogen. Clinical Microbiology Reviews. 2008; 21:538. https://doi.org/10.1128/CMR.00058-07 PMID: 18625687

6. Ibrahim S, Al-Saryi N, Al-Kadmy IMS, Aziz SN. Multidrug-resistant *Acinetobacter baumannii* as an emerging concern in hospitals. Molecular Biology Reports. 2021; 48:6987. https://doi.org/10.1007/S11033-021-06690-6 PMID: 34460060

7. Bogaerts P, Naas T, Wybo I, Bauraing C, Soetens O, Piérard D, et al. Outbreak of infection by carbapenem-resistant *Acinetobacter baumannii* producing the carbapenemase OXA-58 in Belgium. Journal of clinical microbiology. 2006; 44(11):4189–4192. https://doi.org/10.1128/JCM.00796-06 PMID: 16957031

8. Cai B, Echols R, Magee G, Arjona Ferreira JC, Morgan G, Ariyasu M, et al. Prevalence of carbapenem-resistant Gram-negative infections in the United States predominated by *Acinetobacter baumannii* and *Pseudomonas aeruginosa*. In: Open forum infectious diseases. vol. 4. Oxford University Press; 2017.

9. Iovleva A, Mustapha MM, Griffith MP, Komarow L, Luterbach C, Evans DR, et al. Carbapenem-Resistant *Acinetobacter baumannii* in US hospitals: diversification of circulating lineages and antimicrobial resistance. Mbio. 2022; 13(2):e02759–21. https://doi.org/10.1128/mbio.02759-21 PMID: 35311529

10. Damaceno Q, Nicoli JR, Oliveira A. Variability of cutaneous and nasal population levels between patients colonized and infected by multidrug-resistant bacteria in two Brazilian intensive care units. SAGE Open Medicine. 2015; 3:2050312114566668. https://doi.org/10.1177/2050312114566668 PMID: 26770762

11. Liou ML, Chen KH, Yeh HL, Lai CY, Chen CH. Persistent nasal carriers of *Acinetobacter baumannii* in long-term-care facilities. American journal of infection control. 2017; 45(7):723–727. https://doi.org/10.1016/j.ajic.2017.02.005 PMID: 28284750

12. Chen CH, Liou ML, Lee CY, Chang MC, Kuo HY, Chang TH. Diversity of nasal microbiota and its interaction with surface microbiota among residents in healthcare institutes. Scientific reports. 2019; 9(1):1–10. https://doi.org/10.1038/s41598-019-42548-5 PMID: 30992494

13. Rangel K, Chagas TPG, De-Simone SG. *Acinetobacter baumannii* infections in times of COVID-19 pandemic. Pathogens. 2021; 10(8):1006. https://doi.org/10.3390/pathogens10081006 PMID: 34451470

14. Contou D, Claudinon A, Pajot O, Micaëlo M, Longuet Flandre P, Dubert M, et al. Bacterial and viral co-infections in patients with severe SARS-CoV-2 pneumonia admitted to a French ICU. Annals of intensive care. 2020; 10(1):1–9. https://doi.org/10.1186/s13613-020-00736-x PMID: 32894364

15. Lima WG, Brito JCM, da Cruz Nizer WS. Ventilator-associated pneumonia (VAP) caused by carbapenem-resistant *Acinetobacter baumannii* in patients with COVID-19: Two problems, one solution? Medical hypotheses. 2020; 144:110139. https://doi.org/10.1016/j.mehy.2020.110139 PMID: 32758905

16. Russo A, Gavaruzzi F, Ceccarelli G, Borrazzo C, Oliva A, Alessandri F, et al. Multidrug-resistant *Acinetobacter baumannii* infections in COVID-19 patients hospitalized in intensive care unit. Infection. 2022; 50(1):83–92. https://doi.org/10.1007/s15010-021-01643-4 PMID: 34176088

17. Seifert H, Dijkshoorn L, Gerner-Smidt P, Pelzer N, Tjernberg I, Vaneechoutte M. Distribution of *Acinetobacter* species on human skin: comparison of phenotypic and genotypic identification methods. Journal of clinical microbiology. 1997; 35(11):2819–2825.

18. Dijkshoorn L, Van Aken E, Shunburne L, Van Der Reijden T, Bernards A, Nemec A, et al. Prevalence of *Acinetobacter baumannii* and other *Acinetobacter* spp. in faecal samples from non-hospitalised individuals. Clinical microbiology and infection. 2005; 11(4):329–332. https://doi.org/10.1111/j.1469-0691.2005.01093.x PMID: 15760432

19. Howard A, O'Donoghue M, Feeney A, Sleator RD. *Acinetobacter baumannii*: an emerging opportunistic pathogen. Virulence. 2012; 3(3):243–250. https://doi.org/10.4161/viru.19700 PMID: 22546906

20. Smith MG, Gianoulis TA, Pukatzki S, Mekalanos JJ, Ornston LN, Gerstein M, et al. New insights into *Acinetobacter baumannii* pathogenesis revealed by high-density pyrosequencing and transposon mutagenesis. Genes & development. 2007; 21(5):601–614. https://doi.org/10.1101/gad.1510307 PMID: 17344419

21. Weber BS, Miyata ST, Iwashkiw JA, Mortensen BL, Skaar EP, Pukatzki S, et al. Genomic and functional analysis of the type VI secretion system in *Acinetobacter*. PloS one. 2013; 8(1):e55142. https://doi.org/10.1371/journal.pone.0055142 PMID: 23365692

22. Bisaro F, Shuman HA, Feldman MF, Gebhardt MJ, Pukatzki S. *Acinetobacter baumannii* ATCC 17978 encodes a microcin system with antimicrobial properties for contact-independent competition. Microbiology. 2023; 169(6):001346. https://doi.org/10.1099/mic.0.001346 PMID: 37289493

23. Adams MD, Goglin K, Molyneaux N, Hujer KM, Lavender H, Jamison JJ, et al. Comparative genome sequence analysis of multidrug-resistant *Acinetobacter baumannii*. Journal of bacteriology. 2008; 190 (24):8053–8064. https://doi.org/10.1128/JB.00834-08 PMID: 18931120

24. Valentine SC, Contreras D, Tan S, Real LJ, Chu S, Xu HH. Phenotypic and molecular characterization of *Acinetobacter baumannii* clinical isolates from nosocomial outbreaks in Los Angeles County, California. Journal of clinical microbiology. 2008; 46(8):2499–2507. https://doi.org/10.1128/JCM.00367-08 PMID: 18524965

25. Antunes LC, Imperi F, Carattoli A, Visca P. Deciphering the multifactorial nature of *Acinetobacter baumannii* pathogenicity. PloS one. 2011; 6(8):e22674. https://doi.org/10.1371/journal.pone.0022674 PMID: 21829642

26. Farrugia DN, Elbourne LD, Hassan KA, Eijkelkamp BA, Tetu SG, Brown MH, et al. The complete genome and phenome of a community-acquired *Acinetobacter baumannii*. PloS one. 2013; 8(3): e58628. https://doi.org/10.1371/journal.pone.0058628 PMID: 23527001

27. Jacobs AC, Thompson MG, Black CC, Kessler JL, Clark LP, McQueary CN, et al. AB5075, a highly virulent isolate of *Acinetobacter baumannii*, as a model strain for the evaluation of pathogenesis and antimicrobial treatments. MBio. 2014; 5(3):10–1128. https://doi.org/10.1128/mBio.01076-14 PMID: 24865555

28. Thiele I, Palsson BØ. A protocol for generating a high-quality genome-scale metabolic reconstruction. Nature protocols. 2010; 5(1):93–121. https://doi.org/10.1038/nprot.2009.203 PMID: 20057383

29. Oberhardt MA, Palsson BØ, Papin JA. Applications of genome-scale metabolic reconstructions. Molecular systems biology. 2009; 5(1):320. https://doi.org/10.1038/msb.2009.77 PMID: 19888215

30. Kim HU, Kim TY, Lee SY. Genome-scale metabolic network analysis and drug targeting of multi-drug resistant pathogen *Acinetobacter baumannii* AYE. Molecular BioSystems. 2010; 6(2):339–348. https://doi.org/10.1039/b916446d PMID: 20094653

31. Presta L, Bosi E, Mansouri L, Dijkshoorn L, Fani R, Fondi M. Constraint-based modeling identifies new putative targets to fight colistin-resistant *A. baumannii* infections. Scientific reports. 2017; 7(1):1–12. https://doi.org/10.1038/s41598-017-03416-2 PMID: 28623298

32. Magnúsdóttir S, Heinken A, Kutt L, Ravcheev DA, Bauer E, Noronha A, et al. Generation of genome-scale metabolic reconstructions for 773 members of the human gut microbiota. Nature biotechnology. 2017; 35(1):81–89. https://doi.org/10.1038/nbt.3703 PMID: 27893703

33. Norsigian CJ, Kavvas E, Seif Y, Palsson BO, Monk JM. iCN718, an updated and improved genome-scale metabolic network reconstruction of *Acinetobacter baumannii* AYE. Frontiers in genetics. 2018; 9:121. https://doi.org/10.3389/fgene.2018.00121 PMID: 29692801

34. Zhu Y, Zhao J, Maifiah MHM, Velkov T, Schreiber F, Li J. Metabolic responses to polymyxin treatment in *Acinetobacter baumannii* ATCC 19606: integrating transcriptomics and metabolomics with genome-scale metabolic modeling. Msystems. 2019; 4(1):e00157–18. https://doi.org/10.1128/mSystems.00157-18 PMID: 30746493

35. Zhao J, Zhu Y, Han J, Lin YW, Aichem M, Wang J, et al. Genome-scale metabolic modeling reveals metabolic alterations of multidrug-resistant *Acinetobacter Baumannii* in a murine bloodstream infection model. Microorganisms. 2020; 8(11):1793. https://doi.org/10.3390/microorganisms8111793 PMID: 33207684

36. Barbosa JSC. Genome-scale reconstruction of the metabolic network iJS784 for *Acinetobacter baumannii* strain ATCC 17978 to address drug target prioritization [dissertation]. National Laboratory for Scientific Computing, Brazil; 2020. Available from: http://bdtd.ibict.br/vufind/Record/LNCC_66db4adf05fede6c0e4bf5b7e6cd9069.

37. Carey MA, Dräger A, Beber ME, Papin JA, Yurkovich JT. Community standards to facilitate development and address challenges in metabolic modeling. Molecular Systems Biology. 2020; 16(8):e9235. https://doi.org/10.15252/msb.20199235 PMID: 32845080

38. Kanehisa M, Furumichi M, Sato Y, Ishiguro-Watanabe M, Tanabe M. KEGG: integrating viruses and cellular organisms. Nucleic acids research. 2020; https://doi.org/10.1093/nar/gkaa970 PMID: 33125081

39. Machado D, Andrejev S, Tramontano M, Patil KR. Fast automated reconstruction of genome-scale metabolic models for microbial species and communities. Nucleic acids research. 2018; 46(15):7542–7553. https://doi.org/10.1093/nar/gky537 PMID: 30192979

40. Keating SM, Waltemath D, König M, Zhang F, Dräger A, Chaouiya C, et al. SBML Level 3: an extensible format for the exchange and reuse of biological models. Molecular Systems Biology. 2020; 16(8): e9110. https://doi.org/10.15252/msb.20199110 PMID: 32845085

41. Fritzemeier CJ, Hartleb D, Szappanos B, Papp B, Lercher MJ. Erroneous energy-generating cycles in published genome scale metabolic networks: Identification and removal. PLoS computational biology. 2017; 13(4):e1005494. https://doi.org/10.1371/journal.pcbi.1005494 PMID: 28419089

42. Jupp S, Burdett T, Leroy C, Parkinson HE. A new Ontology Lookup Service at EMBL-EBI. SWAT4LS. 2015; 2:118–119.

43. Leonidou N, Fritze E, Renz A, Dräger A. SBOannotator: a Python Tool for the Automated Assignment of Systems Biology Ontology Terms. Bioinformatics. 2023; p. btad437. https://doi.org/10.1093/bioinformatics/btad437 PMID: 37449910

44. yWorks GmbH. yEd. 2019;.

45. Lieven C, Beber ME, Olivier BG, Bergmann FT, Ataman M, Babaei P, et al. MEMOTE for standardized genome-scale metabolic model testing. Nature biotechnology. 2020; 38(3):272–276. https://doi.org/10.1038/s41587-020-0446-y PMID: 32123384

46. Bornstein BJ, Keating SM, Jouraku A, Hucka M. LibSBML: an API library for SBML. Bioinformatics. 2008; 24(6):880–881. https://doi.org/10.1093/bioinformatics/btn051 PMID: 18252737

47. König M. matthiaskoenig/fbc curation: fbc curation-v0.0.6; 2020. Available from: https://doi.org/10.5281/zenodo.3711541.

48. Hucka M, Bergmann FT, Chaouiya C, Dräger A, Hoops S, Keating SM, et al. Systems Biology Markup Language (SBML): Language Specification for Level 3 Version 2 Core Release 2. Journal of Integrative Bioinformatics. 2019; 16(2):1. https://doi.org/10.1515/jib-2019-0021 PMID: 31219795

49. Mishra S, Imlay J. Why do bacteria use so many enzymes to scavenge hydrogen peroxide? Archives of biochemistry and biophysics. 2012; 525(2):145–160. https://doi.org/10.1016/j.abb.2012.04.014 PMID: 22609271

50. Hood MI, Skaar EP. Nutritional immunity: transition metals at the pathogen–host interface. Nature Reviews Microbiology. 2012; 10(8):525–537. https://doi.org/10.1038/nrmicro2836 PMID: 22796883

51. Mortensen BL, Skaar EP. The contribution of nutrient metal acquisition and metabolism to *Acinetobacter baumannii* survival within the host. Frontiers in cellular and infection microbiology. 2013; 3:95. https://doi.org/10.3389/fcimb.2013.00095 PMID: 24377089

52. Hood MI, Mortensen BL, Moore JL, Zhang Y, Kehl-Fie TE, Sugitani N, et al. Identification of an *Acinetobacter baumannii* zinc acquisition system that facilitates resistance to calprotectin-mediated zinc sequestration. PLoS pathogens. 2012; 8(12):e1003068. https://doi.org/10.1371/journal.ppat.1003068 PMID: 23236280

53. Raman K, Rajagopalan P, Chandra N. Flux balance analysis of mycolic acid pathway: targets for anti-tubercular drugs. PLoS computational biology. 2005; 1(5):e46. https://doi.org/10.1371/journal.pcbi.0010046 PMID: 16261191

54. Ma S, Minch KJ, Rustad TR, Hobbs S, Zhou SL, Sherman DR, et al. Integrated modeling of gene regulatory and metabolic networks in *Mycobacterium tuberculosis*. PLoS computational biology. 2015; 11 (11):e1004543. https://doi.org/10.1371/journal.pcbi.1004543 PMID: 26618656

55. Lee DS, Burd H, Liu J, Almaas E, Wiest O, Barabási AL, et al. Comparative genome-scale metabolic reconstruction and flux balance analysis of multiple *Staphylococcus aureus* genomes identify novel

antimicrobial drug targets. Journal of bacteriology. 2009; 191(12):4015–4024. https://doi.org/10.1128/JB.01743-08 PMID: 19376871

56. Seif Y, Monk JM, Mih N, Tsunemoto H, Poudel S, Zuniga C, et al. A computational knowledge-base elucidates the response of *Staphylococcus aureus* to different media types. PLoS computational biology. 2019; 15(1):e1006644. https://doi.org/10.1371/journal.pcbi.1006644 PMID: 30625152

57. Wang N, Ozer EA, Mandel MJ, Hauser AR. Genome-wide identification of *Acinetobacter baumannii* genes necessary for persistence in the lung. MBio. 2014; 5(3):e01163–14. https://doi.org/10.1128/mBio.01163-14 PMID: 24895306

58. Ebrahim A, Lerman JA, Palsson BO, Hyduke DR. COBRApy: constraints-based reconstruction and analysis for python. BMC systems biology. 2013; 7(1):1–6. https://doi.org/10.1186/1752-0509-7-74 PMID: 23927696

59. Orth JD, Thiele I, Palsson BØ. What is flux balance analysis? Nature biotechnology. 2010; 28(3):245–248. https://doi.org/10.1038/nbt.1614 PMID: 20212490

60. Segre D, Vitkup D, Church GM. Analysis of optimality in natural and perturbed metabolic networks. Proceedings of the National Academy of Sciences. 2002; 99(23):15112–15117. https://doi.org/10.1073/pnas.232349399 PMID: 12415116

61. Barh D, Tiwari S, Jain N, Ali A, Santos AR, Misra AN, et al. In silico subtractive genomics for target identification in human bacterial pathogens. Drug Development Research. 2011; 72(2):162–177. https://doi.org/10.1002/ddr.20413

62. Bentley R, Haslam E. The shikimate pathway—a metabolic tree with many branches. Critical reviews in biochemistry and molecular biology. 1990; 25(5):307–384. https://doi.org/10.3109/10409239009090615 PMID: 2279393

63. Farah N, Chin VK, Chong PP, Lim WF, Lim CW, Basir R, et al. Riboflavin as a promising antimicrobial agent? A multi-perspective review. Current Research in Microbial Sciences. 2022; p. 100111. https://doi.org/10.1016/j.crmicr.2022.100111 PMID: 35199072

64. Meloche H, Wood W. The mechanism of 6-phosphogluconic dehydrase. Journal of Biological Chemistry. 1964; 239(10):3505–3510. https://doi.org/10.1016/S0021-9258(18)97751-3 PMID: 14245409

65. Entner N, Doudoroff M. Glucose and gluconic acid oxidation of *Pseudomonas saccharophila*. J Biol Chem. 1952; 196(2):853–862. https://doi.org/10.1016/S0021-9258(19)52415-2

66. Wishart DS, Feunang YD, Guo AC, Lo EJ, Marcu A, Grant JR, et al. DrugBank 5.0: a major update to the DrugBank database for 2018. Nucleic acids research. 2018; 46(D1):D1074–D1082. https://doi.org/10.1093/nar/gkx1037 PMID: 29126136

67. Dorsey CW, Tomaras AP, Actis LA. Genetic and phenotypic analysis of *Acinetobacter baumannii* insertion derivatives generated with a transposome system. Applied and Environmental Microbiology. 2002; 68(12):6353–6360.

68. Gallagher LA, Ramage E, Weiss EJ, Radey M, Hayden HS, Held KG, et al. Resources for genetic and genomic analysis of emerging pathogen *Acinetobacter baumannii*. Journal of bacteriology. 2015; 197(12):2027–2035. https://doi.org/10.1128/JB.00131-15 PMID: 25845845

69. Noronha A, Modamio J, Jarosz Y, Guerard E, Sompairac N, Preciat G, et al. The Virtual Metabolic Human database: integrating human and gut microbiome metabolism with nutrition and disease. Nucleic acids research. 2019; 47(D1):D614–D624. https://doi.org/10.1093/nar/gky992 PMID: 30371894

70. Norsigian CJ, Pusarla N, McConn JL, Yurkovich JT, Dräger A, Palsson BO, et al. BiGG Models 2020: multi-strain genome-scale models and expansion across the phylogenetic tree. Nucleic Acids Research. 2019; 48(D1). https://doi.org/10.1093/nar/gkz1054 PMID: 31696234

71. Henry CS, DeJongh M, Best AA, Frybarger PM, Linsay B, Stevens RL. High-throughput generation, optimization and analysis of genome-scale metabolic models. Nature biotechnology. 2010; 28(9):977–982. https://doi.org/10.1038/nbt.1672 PMID: 20802497

72. Piret J, Boivin G. Pandemics throughout history. Frontiers in microbiology. 2021; 11:631736. https://doi.org/10.3389/fmicb.2020.631736 PMID: 33584597

73. O'Neill J, et al. Review on antimicrobial resistance. Antimicrobial resistance: tackling a crisis for the health and wealth of nations. 2014;(4).

74. Chien YW, Klugman KP, Morens DM. Bacterial pathogens and death during the 1918 influenza pandemic. New England Journal of Medicine. 2009; 361(26):2582–2583. https://doi.org/10.1056/NEJMc0908216 PMID: 20032332

75. Sheng ZM, Chertow DS, Ambroggio X, McCall S, Przygodzki RM, Cunningham RE, et al. Autopsy series of 68 cases dying before and during the 1918 influenza pandemic peak. Proceedings of the National Academy of Sciences. 2011; 108(39):16416–16421. https://doi.org/10.1073/pnas.1111179108 PMID: 21930918

76. Morris DE, Cleary DW, Clarke SC. Secondary bacterial infections associated with influenza pandemics. Frontiers in microbiology. 2017; 8:1041. https://doi.org/10.3389/fmicb.2017.01041 PMID: 28690590

77. Geisinger E, Mortman NJ, Dai Y, Cokol M, Syal S, Farinha A, et al. Antibiotic susceptibility signatures identify potential antimicrobial targets in the *Acinetobacter baumannii* cell envelope. Nature communications. 2020; 11(1):4522. https://doi.org/10.1038/s41467-020-18301-2 PMID: 32908144

78. Bai J, Dai Y, Farinha A, Tang AY, Syal S, Vargas-Cuebas G, et al. Essential gene analysis in *Acinetobacter baumannii* by high-density transposon mutagenesis and CRISPR interference. Journal of bacteriology. 2021; 203(12):10–1128. https://doi.org/10.1128/jb.00565-20 PMID: 33782056

79. Bai J, Raustad N, Denoncourt J, van Opijnen T, Geisinger E. Genome-wide phage susceptibility analysis in *Acinetobacter baumannii* reveals capsule modulation strategies that determine phage infectivity. PLoS Pathogens. 2023; 19(6):e1010928. https://doi.org/10.1371/journal.ppat.1010928 PMID: 37289824

80. Parish T, Stoker NG. The common aromatic amino acid biosynthesis pathway is essential in *Mycobacterium tuberculosis*. Microbiology. 2002; 148(10):3069–3077. https://doi.org/10.1099/00221287-148-10-3069 PMID: 12368440

81. Tapas S, Kumar A, Dhindwal S, Kumar P, et al. Structural analysis of chorismate synthase from *Plasmodium falciparum*: a novel target for antimalaria drug discovery. International journal of biological macromolecules. 2011; 49(4):767–777. https://doi.org/10.1016/j.ijbiomac.2011.07.011 PMID: 21801743

82. Bowe F, O'Gaora P, Maskell D, Cafferkey M, Dougan G. Virulence, persistence, and immunogenicity of *Yersinia enterocolitica* O:8 aroA mutants. Infection and immunity. 1989; 57(10):3234–3236.

83. Umland TC, Schultz LW, MacDonald U, Beanan JM, Olson R, Russo TA. In vivo-validated essential genes identified in *Acinetobacter baumannii* by using human ascites overlap poorly with essential genes detected on laboratory media. MBio. 2012; 3(4):e00113–12. https://doi.org/10.1128/mBio.00113-12 PMID: 22911967

84. Zhang Y, Chowdhury S, Rodrigues JV, Shakhnovich E. Development of antibacterial compounds that constrain evolutionary pathways to resistance. Elife. 2021; 10:e64518. https://doi.org/10.7554/eLife.64518 PMID: 34279221

85. Estrada A, Wright DL, Anderson AC. Antibacterial antifolates: from development through resistance to the next generation. Cold Spring Harbor perspectives in medicine. 2016; 6(8):a028324. https://doi.org/10.1101/cshperspect.a028324 PMID: 27352799

86. Sadaka C, Ellsworth E, Hansen PR, Ewin R, Damborg P, Watts JL. Review on abyssomicins: Inhibitors of the chorismate pathway and folate biosynthesis. Molecules. 2018; 23(6):1371. https://doi.org/10.3390/molecules23061371 PMID: 29882815

87. Frey KM, Viswanathan K, Wright DL, Anderson AC. Prospective screening of novel antibacterial inhibitors of dihydrofolate reductase for mutational resistance. Antimicrobial agents and chemotherapy. 2012; 56(7):3556–3562. https://doi.org/10.1128/AAC.06263-11 PMID: 22491688

88. Wróbel A, Drozdowska D. Recent design and structure-activity relationship studies on the modifications of DHFR inhibitors as anticancer agents. Current Medicinal Chemistry. 2021; 28(5):910–939. https://doi.org/10.2174/0929867326666191016151018 PMID: 31622199

89. Close SJ, McBurney CR, Garvin CG, Chen DC, Martin SJ. Trimethoprim-sulfamethoxazole activity and pharmacodynamics against glycopeptide-intermediate *Staphylococcus aureus*. Pharmacotherapy: The Journal of Human Pharmacology and Drug Therapy. 2002; 22(8):983–989. https://doi.org/10.1592/phco.22.12.983.33599 PMID: 12173801

90. Li J, Bi W, Dong G, Zhang Y, Wu Q, Dong T, et al. The new perspective of old antibiotic: in vitro antibacterial activity of TMP-SMZ against *Klebsiella pneumoniae*. Journal of Microbiology, Immunology and Infection. 2020; 53(5):757–765. https://doi.org/10.1016/j.jmii.2018.12.013 PMID: 30857922

91. Songsungthong W, Yongkiettrakul S, Bohan LE, Nicholson ES, Prasopporn S, Chaiyen P, et al. Diaminoquinazoline MMV675968 from Pathogen Box inhibits *Acinetobacter baumannii* growth through targeting of dihydrofolate reductase. Scientific Reports. 2019; 9(1):15625. https://doi.org/10.1038/s41598-019-52176-8 PMID: 31666629

92. Wu H, Chen H, Zhang J, Hu X, Xie C, Cao W, et al. The anti-multidrug-resistant *Acinetobacter baumannii* study on 1, 3-diamino-7H-pyrrolo [3, 2-f] quinazoline compounds. Molecules. 2022; 27 (23):8609. https://doi.org/10.3390/molecules27238609 PMID: 36500701

93. Eisenberg RC, Dobrogosz WJ. Gluconate metabolism in *Escherichia coli*. Journal of Bacteriology. 1967; 93(3):941–949.

94. Britigan BE, Chai Y, Cohen M. Effects of human serum on the growth and metabolism of *Neisseria gonorrhoeae*: an alternative view of serum. Infection and immunity. 1985; 50(3):738–744.

95. Hommes R, Postma P, Tempest D, Neijssel O. The influence of the culture pH value on the direct glucose oxidative pathway in *Klebsiella pneumoniae* NCTC 418. Archives of microbiology. 1989; 151 (3):261–267. https://doi.org/10.1007/BF00413140 PMID: 2650650

96. Lessie T, Phibbs P Jr. Alternative pathways of carbohydrate utilization in pseudomonads. Annual review of microbiology. 1984; 38(1):359–388. https://doi.org/10.1146/annurev.mi.38.100184.002043 PMID: 6388497

97. Krismer B, Liebeke M, Janek D, Nega M, Rautenberg M, Hornig G, et al. Nutrient limitation governs *Staphylococcus aureus* metabolism and niche adaptation in the human nose. PLoS pathogens. 2014; 10(1):e1003862. https://doi.org/10.1371/journal.ppat.1003862 PMID: 24453967

98. Olivier BG, Bergmann FT. SBML level 3 package: flux balance constraints version 2. Journal of integrative bioinformatics. 2018; 15(1). https://doi.org/10.1515/jib-2017-0082 PMID: 29522419

99. Hastings J, Owen G, Dekker A, Ennis M, Kale N, Muthukrishnan V, et al. ChEBI in 2016: Improved services and an expanding collection of metabolites. Nucleic acids research. 2016; 44(D1):D1214–D1219. https://doi.org/10.1093/nar/gkv1031 PMID: 26467479

100. Karp PD, Billington R, Caspi R, Fulcher CA, Latendresse M, Kothari A, et al. The BioCyc collection of microbial genomes and metabolic pathways. Briefings in bioinformatics. 2019; 20(4):1085–1093. https://doi.org/10.1093/bib/bbx085 PMID: 29447345

101. Clark K, Karsch-Mizrachi I, Lipman DJ, Ostell J, Sayers EW. GenBank. Nucleic acids research. 2016; 44(D1):D67–D72. https://doi.org/10.1093/nar/gkv1276 PMID: 26590407

102. Chang A, Jeske L, Ulbrich S, Hofmann J, Koblitz J, Schomburg I, et al. BRENDA, the ELIXIR core data resource in 2021: new developments and updates. Nucleic Acids Research. 2021; 49(D1):D498–D508. https://doi.org/10.1093/nar/gkaa1025 PMID: 33211880

103. UniProt: the universal protein knowledgebase in 2021. Nucleic acids research. 2021; 49(D1):D480–D489. https://doi.org/10.1093/nar/gkaa1100 PMID: 33237286

104. Juty N, Le Novere N, Laibe C. Identifiers. org and MIRIAM Registry: community resources to provide persistent identification. Nucleic acids research. 2012; 40(D1):D580–D586. https://doi.org/10.1093/nar/gkr1097 PMID: 22140103

105. 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(2):e0149263. https://doi.org/10.1371/journal.pone.0149263 PMID: 26882475

106. Courtot M, Juty N, Knüpfer C, Waltemath D, Zhukova A, Dräger A, et al. Controlled vocabularies and semantics in systems biology. Molecular systems biology. 2011; 7(1):543. https://doi.org/10.1038/msb.2011.77 PMID: 22027554

107. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. Journal of molecular biology. 1990; 215(3):403–410. Basic local alignment search tool. https://doi.org/10.1016/S0022-2836(05)80360-2 PMID: 2231712

108. Davis JJ, Wattam AR, Aziz RK, Brettin T, Butler R, Butler RM, et al. The PATRIC Bioinformatics Resource Center: expanding data and analysis capabilities. Nucleic acids research. 2020; 48(D1):D606–D612.

109. King ZA, Dräger A, Ebrahim A, Sonnenschein N, Lewis NE, Palsson BO. Escher: a web application for building, sharing, and embedding data-rich visualizations of biological pathways. PLoS computational biology. 2015; 11(8):e1004321. https://doi.org/10.1371/journal.pcbi.1004321 PMID: 26313928

Paper IV: Genome-scale model of *R. mucilaginosa* predicts gene
essentialities and reveals metabolic capabilities

Genome-scale model of *Rothia mucilaginosa* predicts gene essentialities and reveals metabolic capabilities

Nantia Leonidou,[1,2,3,4,5] Lisa Ostyn,[6] Tom Coenye,[6] Aurélie Crabbé,[6] Andreas Dräger[1,4,7]

AUTHOR AFFILIATIONS See affiliation list on p. 21.

ABSTRACT Cystic fibrosis (CF), an inherited genetic disorder caused by mutations in the cystic fibrosis transmembrane conductance regulator gene, results in sticky and thick mucosal fluids. This environment facilitates the colonization of various microorganisms, some of which can cause acute and chronic lung infections, while others may positively impact the disease. *Rothia mucilaginosa*, an oral commensal, is relatively abundant in the lungs of CF patients. Recent studies have unveiled its anti-inflammatory properties using *in vitro* three-dimensional lung epithelial cell cultures and *in vivo* mouse models relevant to chronic lung diseases. Apart from this, *R. mucilaginosa* has been associated with severe infections. However, its metabolic capabilities and genotype-phenotype relationships remain largely unknown. To gain insights into its cellular metabolism and genetic content, we developed the first manually curated genome-scale metabolic model, *i*RM23NL. Through growth kinetics and high-throughput phenotypic microarray testings, we defined its complete catabolic phenome. Subsequently, we assessed the model's effectiveness in accurately predicting growth behaviors and utilizing multiple substrates. We used constraint-based modeling techniques to formulate novel hypotheses that could expedite the development of antimicrobial strategies. More specifically, we detected putative essential genes and assessed their effect on metabolism under varying nutritional conditions. These predictions could offer novel potential antimicrobial targets without laborious large-scale screening of knockouts and mutant transposon libraries. Overall, *i*RM23NL demonstrates a solid capability to predict cellular phenotypes and holds immense potential as a valuable resource for accurate predictions in advancing antimicrobial therapies. Moreover, it can guide metabolic engineering to tailor *R. mucilaginosa*'s metabolism for desired performance.

IMPORTANCE Cystic fibrosis (CF) is a genetic disorder characterized by thick mucosal secretions, leading to chronic lung infections. *Rothia mucilaginosa* is a common bacterium found in various parts of the human body, acting as a normal part of the flora. In people with weakened immune systems, it can become an opportunistic pathogen, while it is prevalent and active in CF airways. Recent studies have highlighted its anti-inflammatory properties in the lower pulmonary system, indicating the intricate relationship between microbes and human health. Herein, we have developed the first manually curated metabolic model of *R. mucilaginosa*. Our study examined the previously unknown relationships between the bacterium's genotype and phenotype and identified essential genes that impact the metabolism under various conditions. With this, we opt for paving the way for developing new strategies in antimicrobial therapy and metabolic engineering, leading to enhanced therapeutic outcomes in cystic fibrosis and related conditions.

KEYWORDS *i*RM23NL, *Rothia mucilaginosa* DSM20746, ATCC 25296, constraint-based modeling, flux balance analysis, flux variability analysis, mathematical network, genome-

scale metabolic model, metabolic engineering, pathway analysis, SBML, Gram-positive, nasal microbiome, lung infections, cystic fibrosis, antimicrobial strategies

Rothia mucilaginosa is a Gram-positive, encapsulated, non-motile, and non-spore-forming bacterium of the *Micrococcaceae* family (1, 2). While it is mainly aerobic, it may also grow anaerobically as it can switch to fermentation or other non-oxygen-involving pathways. *R. mucilaginosa* is a common commensal of the normal oral, upper and lower respiratory tract, and part of the skin florae in humans (1, 3–6). This means it coexists harmlessly within the host and may even provide benefits. Nonetheless, it can also act as an opportunistic pathogen, particularly in individuals with weakened immune systems, as an etiological agent of serious infections such as endocarditis, sepsis, and meningitis (7). Janek et al. highlighted the high prevalence of *R. mucilaginosa* within the nasal microbiome (8). Moreover, they report its susceptibility to certain staphylococcal bacteriocins, indicating its major competition with the nasal staphylococci and the substantial impact of bacteriocins in shaping the nasal microbiota. In 2020, Uranga et al. revealed that *R. mucilaginosa* produces the strongest Fe^{3+}-binding archetypal siderophore known, called enterobactin (9). This attribute contributes to its competition with oral microbiota (the cariogenic *S. mutans*, *A. timonensis*, and *Streptococcus* sp.) as well as four methicillin-resistant strains of *S. aureus* (MRSA). Enterobactin is a type of siderophore produced by bacteria to scavenge, chelate, and transport ferric irons from their surrounding environment. These are essential for bacteria when iron is scarce as they facilitate their acquisition necessary for their growth and metabolic processes.

Prior metagenomic sequencing analyses have unveiled the prevalence of *R. mucilaginosa* at high abundances and its enhanced metabolic activity in the lungs of cystic fibrosis (CF) patients (10, 11). CF is caused by the hereditary mutation of the cystic fibrosis transmembrane conductance regulator (CFTR) gene that disrupts the transepithelial movement of ions, leading to an aberrant accumulation of thick and sticky mucus within the airways. The impaired immune clearance creates a hypoxic environment (12) promoting the polymicrobial colonization of opportunistic microbes together with fungi and viruses, ultimately resulting in persistent and recurring infections (13). Guss et al. and Bittar et al. declared *R. mucilaginosa* as an emerging CF bacterium (14, 15), while Lim et al. provided evidence supporting that *R. mucilaginosa* is a frequently encountered and metabolically active inhabitant of CF airways (16). Additionally, a study from 2018 shows that the opportunistic pathogen *Pseudomonas aeruginosa*, which frequently causes infections in CF patients, builds essential primary metabolites, like glutamate, by utilizing compounds produced by *R. mucilaginosa* (17). This symbiotic interaction implies that *P. aeruginosa* benefits from its neighboring microbes, which contributes to its pathogenesis in the CF lungs. On the other hand, Rigauts et al. revealed the anti-inflammatory activity of *R. mucilaginosa* in the lower respiratory tract, which could impact the seriousness of chronic lung diseases (18).

In systems biology, genome-scale metabolic models (GEMs) represent comprehensive reconstructions of organisms' metabolic networks. They are built using genomic sequences and comprise all known biochemical reactions and associated genes. These models provide systems-level insights into cellular metabolism, allowing researchers to simulate and analyze the flow of metabolites through these networks (19). The interactions among reactions and metabolites in a metabolic model are mathematically represented with a stoichiometric matrix (20). In the past years, an array of *in silico* methods has been developed to analyze GEMs and derive valuable hypotheses. Flux balance analysis (FBA) is such a powerful computational technique that operates on the principle of achieving a steady state by optimizing the flux (rate) of metabolites through reactions while accounting for various constraints such as stoichiometry, thermodynamics, and uptake/secretion boundaries (21). Applying flux balance analysis on a GEM provides insights into the intricate biological system interactions. This analytical approach facilitates the prediction of cellular phenotypes and identification of promising drug targets and contributes to optimizing biotechnological processes (22). Moreover,

such models can guide genetic engineering by suggesting genetic modifications that could enhance desired product formation or cellular behavior. Further applications include ameliorating culture media by incorporating components that increase bacterial growth rates. So far, GEMs have been an invaluable resource in the systems biology field that helped untangle the metabolism of various organisms and especially of high-threat pathogens (23, 24). As described above, *R. mucilaginosa* has gained great interest in the context of polymicrobial CF environments. However, its metabolic capabilities and genotype-phenotype relationships in isolated monoculture settings remain largely unexplored.

Here, we present the first manually curated and high-quality GEM of *R. mucilaginosa*, *i*RM23NL, striving to understand its metabolism and unique phenotypes under diverse conditions. Our simulation-ready network accounts for thousands of reactions and is available in a standardized format following the community guidelines (25). Through growth kinetic experiments and high-throughput phenotypic microarray assays, we validated *i*RM23NL's accuracy in predicting growth and substrate utilization patterns. We refined the reconstruction by comparing the *in vitro* results to *in silico* simulations, resulting in novel metabolic reactions and genes. To our knowledge, this is the first study presenting high-throughput nutrient utilization and comprehensive growth data for *R. mucilaginosa*. Finally, we employed FBA to formulate novel gene essentiality hypotheses that could expedite the development of antimicrobial strategies. Figure 1 summarizes the experimental and computational work presented here.

RESULTS

Reconstruction of a high-quality metabolic model for *R. mucilaginosa* DSM20746

The pipeline we previously developed (26) was used to build the first high-quality and manually curated GEM of *R. mucilaginosa* DSM20746 (ATCC 25296). An initial draft metabolic model was derived with CarveMe (27) and is based on the Biochemical, Genetical, and Genomical (BiGG) identifiers (28). The translated sequence with over 1,700 proteins and the Gram-positive-specific template were employed. This enabled us to build a more precise reconstruction considering information on the peptidoglycan layer for the biomass objective function (BOF). The draft network contained 1,015 reactions (141 pseudo-reactions), 788 metabolites, and 265 genes (Fig. 2). In the first gap-filling stage (Draft_2), we expanded the list of reactions based on the annotated genome and growth kinetics data in diverse growth environments. For this, we extensively indexed organism-specific literature and databases and included additional enzymatic reactions together with 50 new gene-protein-reaction associations (GPRs). Subsequently, high-throughput nutrient utilization assays and model validation incorporated further reactions and their associated metabolic genes. Non-metabolic genes, which take part in other cellular processes, e.g., signaling pathways or transcription, were not considered. In total, 95 reactions, together with their associated metabolites, were newly added into the model, along with 121 novel GPRs, increasing the genetic coverage. Over 60% of the transport reactions have a GPR assigned, while 63% of the total enzymatic reactions have at least one gene assigned. Moreover, missing exchange reactions were added to all extracellular metabolites to represent the exchange of substrates between the extracellular environment and the model. The strain-specific BioCyc31 database was further employed to correct the reversibility of biochemical reactions, while duplicated reactions and metabolites were eliminated. In all cases, when no organism-specific information was available, we leveraged data from closely related species based on our phylogenomic analysis (Fig. 3). According to the calculated average nucleotide identity (ANI) matrix, *R. mucilaginosa* exhibits a similarity to six out of the 12 tested *Rothia* genomes. More specifically, it shares a greater resemblance with *R. aeria* and *R. dentocariosa* underscoring a closer evolutionary relationship between these species.

R. mucilaginosa is primarily aerobic, efficiently generating ATP through oxic respiration; however, in low-oxygen or oxygen-absent conditions, it shifts to anaerobic
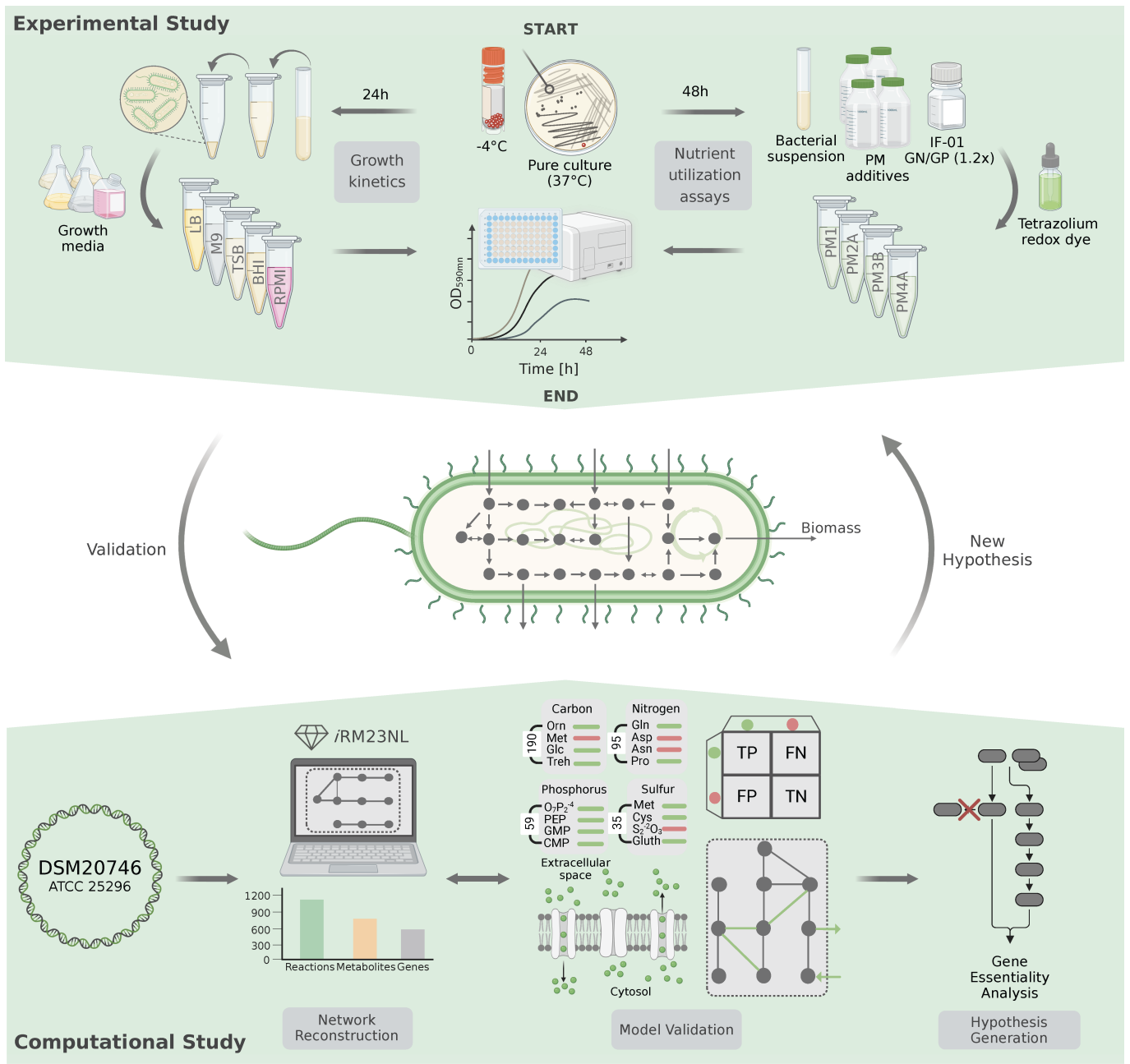
FIG 1 Construction and validation flowchart of the metabolic network for *R. mucilaginosa*, *i*RM23NL. The study is divided into the experimental and computational phases. The proteome-derived metabolic reconstruction and curation was done based on the workflow we described elsewhere (26).

metabolism to produce energy. This metabolic adaptability enables *R. mucilaginosa* to adapt in microaerophilic environments like the oxygen-restricted conditions in CF lungs (16). Our draft model lacked the ability to demonstrate anaerobic growth. Therefore, we investigated the metabolic cascade and systematically incorporated missing enzymes to ensure that the model can simulate growth even in the absence of oxygen by identifying and integrating alternative pathways. This refinement included the incorporation of enzymatic reactions, such as the superoxide dismutase (SPODM) and catalase (CAT) that are responsible for the breakdown of radical reactive oxygen species (ROS) and shielding the cell against oxidative damage (Fig. 4 Panel A). Such scavenging enzymes play an integral role in counteracting the harmful effects of ROS during anaerobic respiration (31). However, during this process, we found no associated GPRs for CAT within the
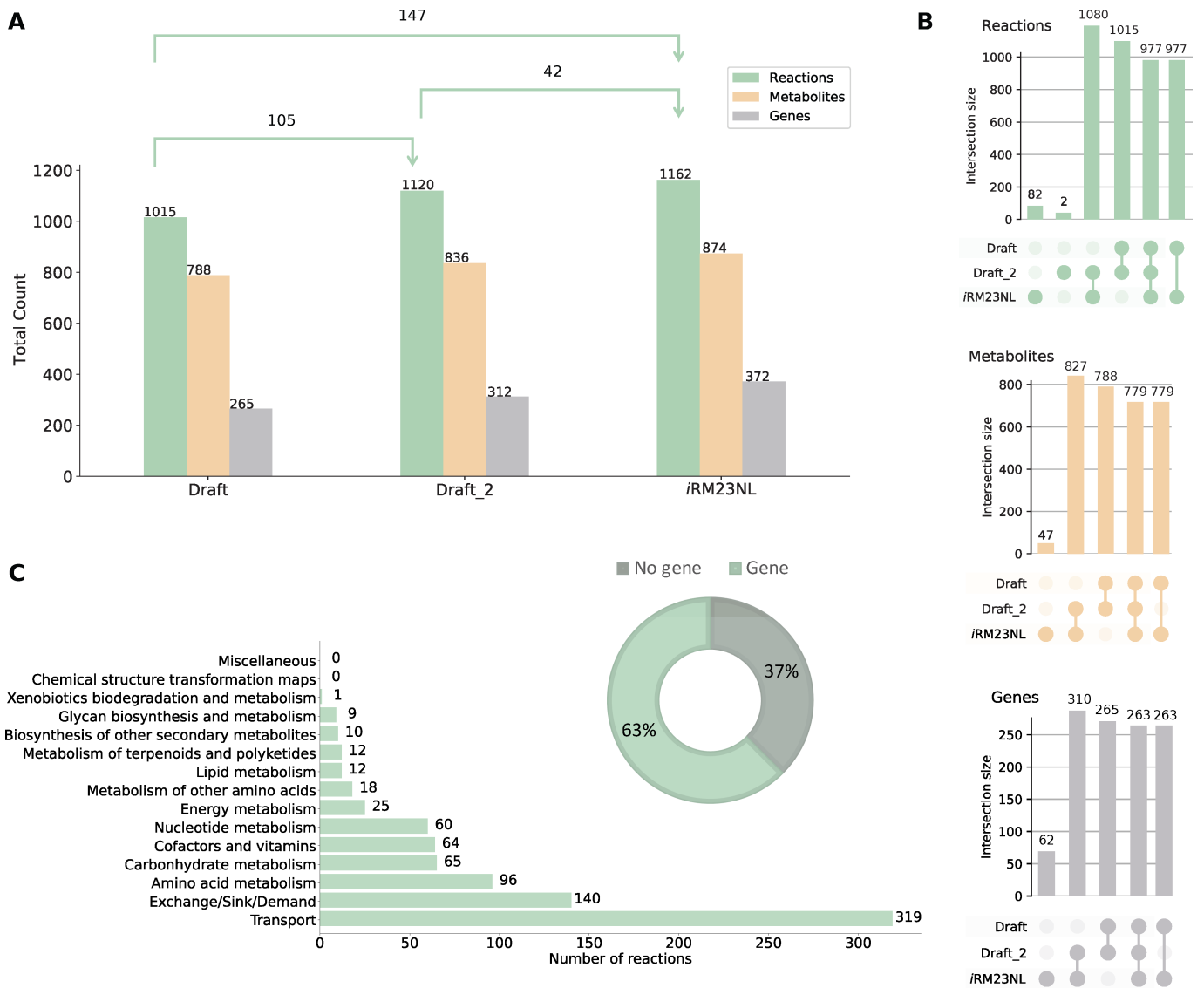
FIG 2 Properties of the genome-scale metabolic model for *R. mucilaginosa* DSM20746, *i*RM23NL. (A) Evolution of metabolic network content from its initial draft to the final stage of extensive manual gap-filling. The shifts in the sets' sizes are also displayed in each stage. The first stage of gap-filling is denoted by Draft_2, while the final stage is upon validation with experimental data. (B) UpSet plots comparing sets between three model versions created using the UpSetPlot package (29). The numbers indicate the cardinality of the respective set. (C) Subsystem-level statistics within pathways along with the distribution of gene- and non-gene-associated reactions. The pathway analysis was limited to reaction identifiers that could be successfully mapped to Kyoto Encyclopedia of Genes and Genomes (KEGG) (30) reactions.

organism-specific BioCyc database. Additional scavenging enzymes like glutathione and thioredoxin reductases essential for maintaining the redox balance (32) were already present in the initial draft model (GTHOr, GTHRDabc2pp, and TRDR). Altogether, the final model, *i*RM23NL, contains 1,162 reactions (619 gene-associated; 65 catalyzed by enzyme complexes, 70 catalyzed by isozymes, and 484 by simple gene association), 171 exchange and sink reactions, 874 metabolites (558 in cytoplasm, 148 in periplasm, and 168 in the extracellular space), and 372 genes (Fig. 2). The model's metabolic coverage is at 3.12%, which indicates a high level of modeling detail regarding reactions, enzymes, and their associated genes (33). Additionally, we enriched the model elements with numerous database cross-references (34), while appropriate and precise Systems Biology Ontology (SBO) terms were assigned to each model entity using the SBOannotator package (35). The presence of no energy-generating cycles (EGCs) was
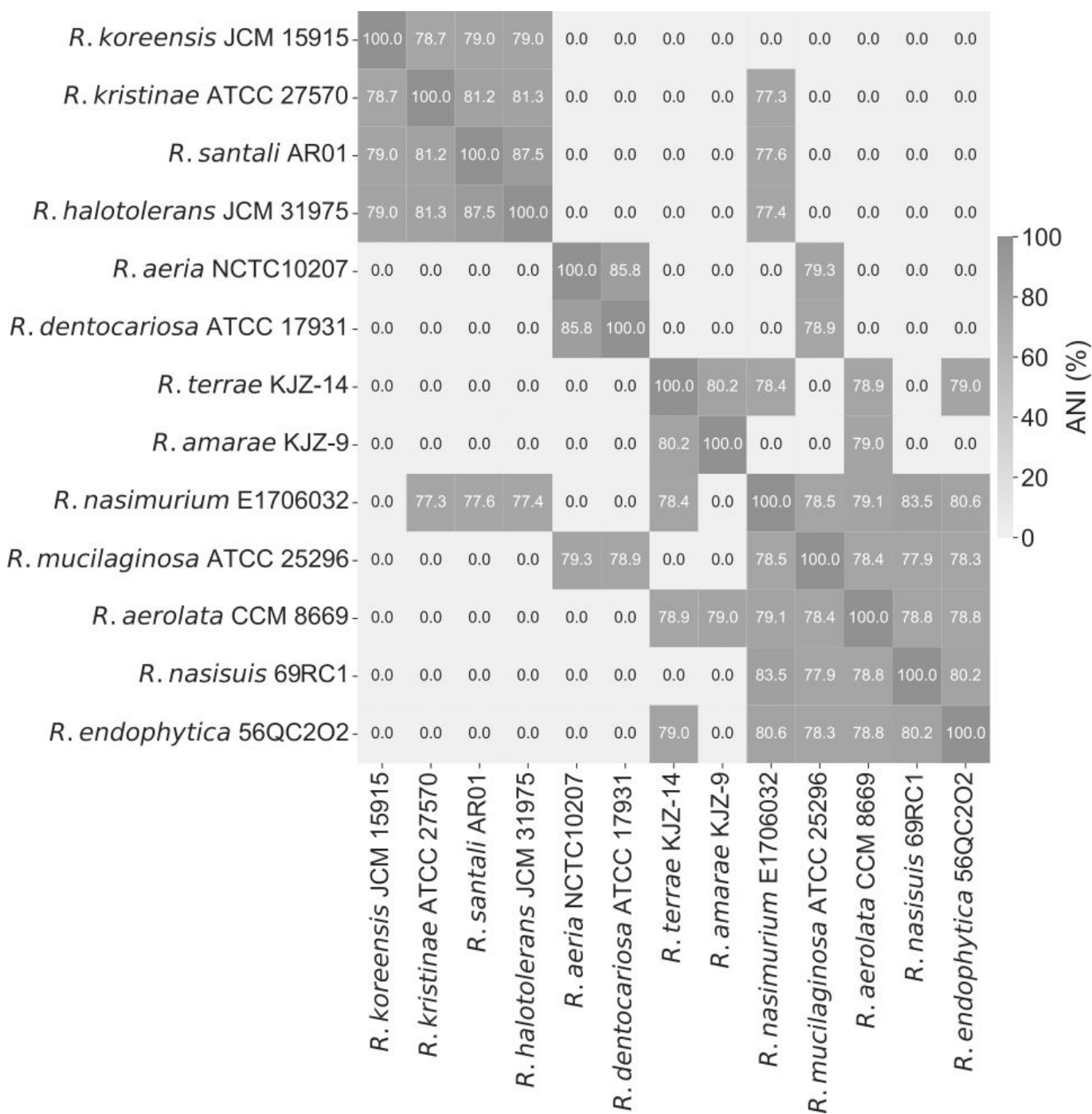
FIG 3 Phylogenomic all-vs-all analysis between 13 *Rothia* species. Based on the calculated ANI matrix, *R. mucilaginosa* is mostly similar to six out of 13 genomes, with higher similarity to *R. aeria* and *R. dentocariosa*.

ensured and controlled after each curation stage, and the mass- and charge-imbalances were corrected. With this, the final Metabolic Model Testing (MEMOTE) (36) score of *i*RM23NL is 89%, while with highly specific SBO terms, the score drops by 2%. The final curated model, *i*RM23NL, is available as a supplementary file in Systems Biology Markup Language (SBML) Level 3 Version 1 (37) and JavaScript Object Notation (JSON) formats with the flux balance constraints (fbc) and groups plugins available.

The first validation step of *i*RM23NL aimed to evaluate its ability to correctly simulate biomass production across diverse environmental conditions and growth media

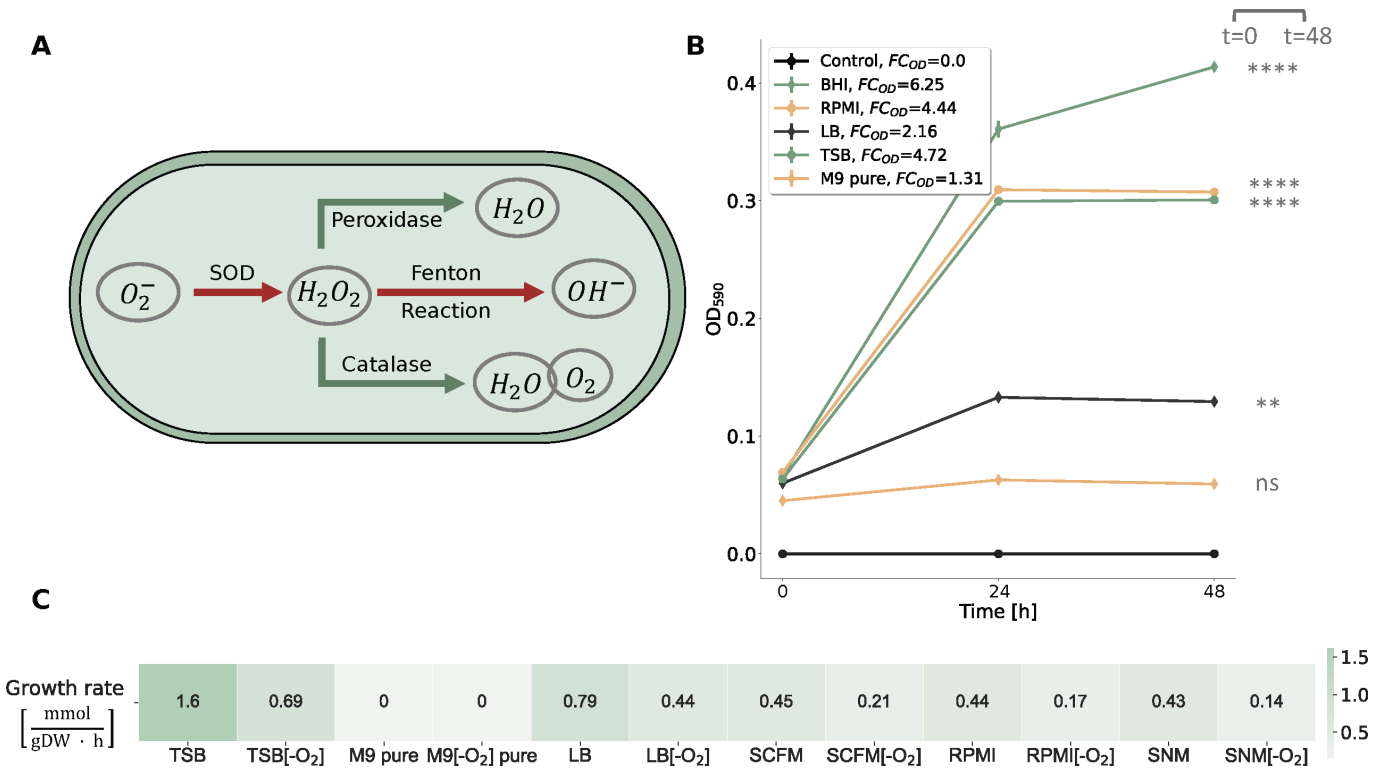**FIG 4** Investigation of *R. mucilaginosa*'s growth behavior in different nutrient media. (A) Metabolic response of *R. mucilaginosa* under anaerobic stress as represented in *i*RM23NL. Reduction process of oxygen (O_2^-) generating ROS is indicated by red arrows, while pathways highlighted in green arrows represent reactions governed by ROS scavenging enzymes leading to bacterial cell detoxification. (B) Experimentally-derived growth curves for *R. mucilaginosa* DSM20746 in multiple liquid growth media along with the respective fold changes (FCs) of the acquired optical densities (ODs) at 590 nm, as defined in Equation 1. The data shown here are an average of three biological replicates ($n = 3$). Based on the experimental results, a threshold of $FC_{OD} = 1.4$ was established to qualitatively describe bacterial growth. We verified the correctness of the threshold by performing statistical analysis as described in Materials and Methods. All data are normally distributed, while there is no significant difference between their variances. The asterisks flag the significance levels. The BHI medium was used as a baseline, while the Control line represents blank measurements of pure media. Bacterial growth was aerobically measured by the OD at 590 nm (ordinate) at three distinct time points ranging from 0 h to 48 h (abscissa). (C) *In silico*-simulated growth rates using *i*RM23NL. Detailed *in silico* media formulations are provided in Table S2.

formulations. To elucidate the bacterium's optimal conditions and metabolic preferences, we experimentally tested five commonly used media, including three general nutrient media; brain heart infusion (BHI) and Luria-Bertani (LB), and tryptic soy broth (TSB), and two defined media; M9 minimal medium (M9) pure and Roswell Park Memorial Institute (RPMI) (Fig. 4 Panel B). The BHI medium was used as a baseline for the *in vitro* experiments since it is a known and well-established environment for the growth of *R. mucilaginosa* and enabled us to compare the bacterium's growth characteristics to the newly tested media. For the *in silico* simulations, we applied FBA and added additional constraints to the linear programming problem defined in Equation 5. In more detail, we specified the flux constraints such that only extracellular metabolites defined in the medium of interest could flow freely through the system (unconstrained, finite fluxes), while the remaining fluxes were constrained to zero. We compared the *in vitro* to the *in silico* observed growth using the FC_{OD} as a qualitative measure of growth (see Materials and Methods). Furthermore, we compared the OD at the start and the end of the experiment, considering a statistically significant difference between these measurements as an indication of growth. Our metabolic network, *i*RM23NL, simulated positive fluxes through the biomass reaction for all tested media except for the M9 pure medium, where a zero flux was observed. These findings align with the experimentally observed data. More specifically, there is no statistically significant difference in OD between the initial and final time-points in M9 pure medium (*P*-value = 0.1202 and

FC_{OD} < 1.4) indicating no significant growth. Conversely, in the remaining examined media, statistically significant growth was observed (*P*-value = 0.00006–0.00142 and FC_{OD} > 1.4) indicating significant growth in these settings. The highest aerobic growth rate was predicted in TSB [1.6 mmol/(g_{DW} · h)], while the lowest biomass production flux was recorded for the M9 pure medium containing only essential salts. However, the RPMI medium followed as the second-highest in supporting bacterial *in vitro* cellular growth, offering a defined medium suitable for *R. mucilaginosa*'s cultivation. Although *R. mucilaginosa* increased its biomass after 24 h, it slightly declined after 48 h. On the other hand, the simulated network resulted in a contrary outcome compared to the expected experimental effect. More specifically, *i*RM23NL simulated a lower flux through biomass [0.44 mmol/(g_{DW} · h)] with RPMI when compared to LB. It is important to note here that in order to simulate growth in RPMI medium, six metal ions [cobalt (Co^{2+}), copper (Cu^{2+}), manganese (Mn^{2+}), zinc (Zn^{2+}), ferric iron (Fe^{3+}), and ferrous iron (Fe^{2+})] were supplemented. These compounds were missing from the providers' medium formulation. Our findings underscored *R. mucilaginosa*'s adaptability to various nutritional environments, growing best in nutrient-rich conditions while revealing specific growth requirements beyond minimal settings.

We further employed *i*RM23NL to examine whether it could generate biomass within the human nasal environment and the CF lungs. For this purpose, we performed *in silico* simulations using the synthetic cystic fibrosis sputum medium (SCFM) (38) and synthetic nasal medium (SNM) (39) (Fig. 4 Panel C). Our computational model successfully simulated positive growth in both media, with a growth rate of 0.43 mmol/(g_{DW} · h) in SNM and 0.45 mmol/(g_{DW} · h) in SCFM. These results align with the documented metabolic activity of *R. mucilaginosa* in CF lungs and its frequent isolation from the human nasal cavity. Notably, the observed growth rates closely resembled the flux rate predicted for biomass production in RPMI medium. Additionally, we confirmed that *i*RM23NL accurately represented *R. mucilaginosa*'s capacity for facultative anaerobic respiration. In more detail, when the oxygen uptake was turned off *i*RM23NL could successfully exhibit growth using alternative metabolic pathways across all tested nutritional media. When the oxygen level was decreased, the model predicted up to 68% reduction in biomass yield compared to aerobic conditions. Consequently, the remarkably lower anaerobic rates in all tested media mimic *R. mucilaginosa*'s inherent facultative anaerobic capabilities.

Nutrient utilization profile of *R. mucilaginosa* and predictive performance of *i*RM23NL

We experimentally characterized the metabolic phenotype of *R. mucilaginosa* DSM20746 using four 96-well Biolog PM microplates (Fig. 5). These high-throughput assays serve as proxies for bacterial growth by measuring cellular respiration across several conditions. Active respiration in the minimal medium is detected by the reduction of tetrazolium dye over time, indicating the utilization of the provided sole source (40). We cultivated our strain in a minimal medium supplemented with various sources, and growth was monitored over 48 h to identify suitable nutrients for the bacterium (as described in Materials and Methods). The derived OD measurements were normalized according to the average growth over replicates per plate and converted to qualitative data representing non-growth (NG) or growth (G). In total, we tested the uptake and utilization of 379 distinct carbon, nitrogen, phosphorus, and sulfur substrates. *R. mucilaginosa* demonstrated the ability to utilize 61 of 190 tested carbon substrates, including carboxylates, saccharides, and amino acids, while 10 of 95 were found to be viable nitrogen sources (Fig. 5 Panel B). Furthermore, out of 59 tested phosphorus sources, *R. mucilaginosa* exhibited a loss of metabolic activity for 28 compounds, resulting in a non-viable phenotype, while only 71.4% of all analyzed sulfur substrates supported positive growth. More specifically, 6 inorganic phosphorus (IP), 14 organic phosphorus (OP), 2 cyclic nucleoside monophosphates (cNMPs), and 9 nucleoside monophosphates (NMPs) were successfully utilized as sole phosphorus sources (Fig. 5, Panel C). The experimentally
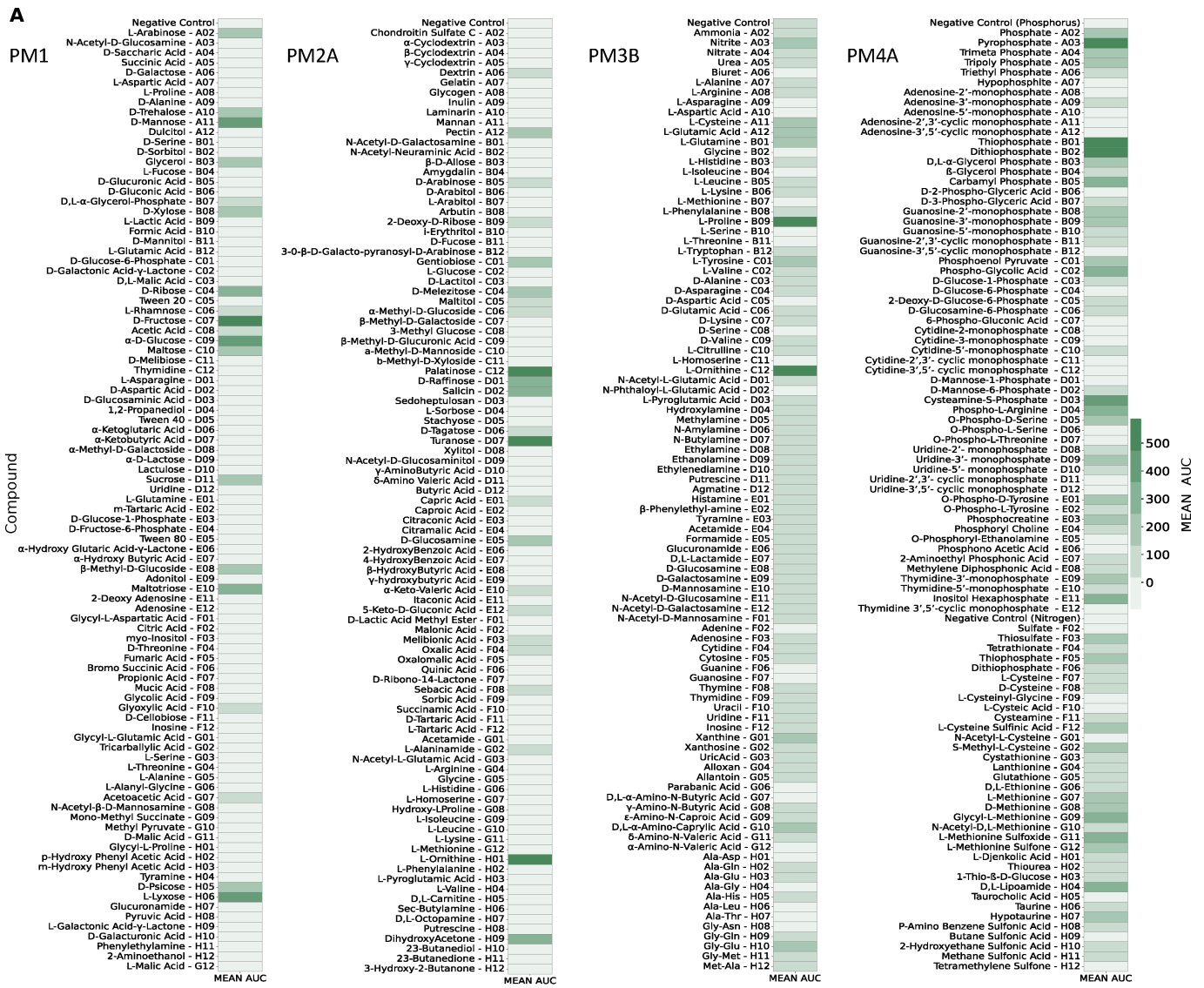
FIG 5 Complete experimentally-derived nutrient utilization phenome of *R. mucilaginosa* DSM20746. (A) Utilization of individual nutrients by the bacterium across four Biolog phenotypic microarrays. Bacterial growth was measured by OD at 590 nm. (B) Numerical summary nutrient sources experimentally tested in each Biolog phenotype microarray (PM), classified into those resulting in bacterial growth and those that *R. mucilaginosa* could not utilize. (C) Categorization of all tested phosphorous sources during the high-throughput Biolog assay. Utilization of totally 31 phosphorus sources resulted in positive phenotype (green chart), while the cell exhibited an inability to utilize the remaining 28 (orange chart).

defined nutrient utilization phenome of *R. mucilaginosa* can be found in Fig. S1. An overview of all experimentally tested substrates, along with the assay results, can be

found in Table S3. We independently confirmed the Biolog nutrient utilization data by testing the ability of DSM20746 to grow on minimal media in the presence of 10 compounds (see Materials and Methods, Fig. S2).

Additionally, we evaluated the predictive performance of our metabolic model by using various C-, N-, P-, and S-containing substrates. All compounds from the high-throughput phenotypic data were mapped to BiGG (28) identifiers and subsequently to *i*RM23NL. In total, 286 could be successfully mapped to the BiGG database. From these, 126 existed as extracellular metabolites in *i*RM23NL and were considered for further analysis. Model simulations were performed under aerobic conditions with the minimal medium defined in Table S2 and FBA (see Materials and Methods). An extracellular reaction was enabled for each tested substrate to force the model to use its transporters. Discrepancies between the Biolog data and the model simulations were utilized as basis for hypotheses to further improve and refine the network reconstruction. We resolved most inconsistencies via extensive literature mining and iterative gap analysis. For this, we used the organism- and strain-specific BioCyc (33) database. Throughout this process, we encountered different scenarios regarding incorrect model predictions. These included compounds present in all compartments, including the extracellular space, as well as substrates defined within the intracellular space and periplasm, with no transporter defined toward the extracellular space. If the experimental results indicated utilization of an undefined compound, we searched BioCyc (33) to find strain-specific and gene-based missing transporters or enzymatic reactions. When no organism-specific evidence was available, we sought supporting data from genomically identical species (Fig. 3). For instance, the compound 3-sulfino-L-alanine (3sala) was initially absent from any compartment in the preliminary draft model. Since no strain-specific information was available, we conducted a homology-based search using Basic Local Alignment Search Tool (BLAST) (41) to find genes with high similarity (similarity threshold: > 80%) in related species. Subsequently, we identified cysteine desulfurase (SULFCYS) along with three associated transport reactions (proton-mediated; SULFCYSpp, diffusion; SULFCYStex, and ABC transport; SULFCYSabc) that displayed a similarity over 80% with *R. dentocariosa*. These components were consequently incorporated into *i*RM23NL, resulting in the expected positive utilization phenotype. Generally, false negative or false positive predictions arise from missing or erroneous involvement of transporters, respectively. We resolved false positives by removing transport reactions lacking supporting gene evidence or adjusting their reversibility to facilitate export solely. More specifically, initial model predictions indicated that *i*RM23NL could not sustain growth when supplied with either L-cysteate (Lcyst) or AMP (amp) as sole sources, while Biolog assays indicated the opposite. To rectify this, we introduced the corresponding irreversible transporters (LCYStex and AMPt) and enabled their *in silico* utilization of these compounds. Moreover, several metabolites (e.g., phosphoenolpyruvate; pep, trimetaphosphate; tmp, hypotaurine; hyptaur, and inorganic triphosphate; pppi) which were absent from the initial draft model but exhibited positive growth in utilization assays were subsequently incorporated into the final network, leading to additional true positives predictions. All in all, over 50 transport reactions were added into the network, while 37 wrongly added enzymatic functions were removed. We also incorporated novel GPRs encoding over 60 biochemical reactions. Nevertheless, we identified approximately 20 instances where the resolution of inconsistencies necessitated the inclusion of metabolic reactions lacking associated gene evidence. For instance, to enable the utilization of L-aspartate, we introduced a transporter via diffusion from extracellular to periplasm (ASPtex), for which no associated GPR was available. Similar scenarios arose for other compounds, e.g., D-galactose, D-glucuronate, and acetate. These instances underscore knowledge gaps in the metabolism of DSM20746 that require in-depth investigation. In total, 14 carbon and nitrogen sources failed to promote growth in *i*RM23NL. Surprisingly, all of these sources had corresponding transport reactions *i*RM23NL but still remained ineffective (e.g., L-fucose, L-arabinose, and L-rhamnose) and

nitrogen (L-tyrosin). We could not find further information on their transport or metabolic mechanism either in the genome annotation or the literature.

In summary, the final prediction accuracy of nutrient assimilation and utilization achieved by *i*RM23NL was 77% for carbon sources (MCC for PM1 = 0.52 and PM2A = 0.58), 94.4% for nitrogen sources (MCC = 0.82), 97% for phosphorus and sulfur sources (ACC = 100%; MCC = 1.0 and ACC = 92.3%; MCC = 0.82, respectively) (Fig. 6). Our model's performance was notably increased by 40% post-comprehensive curation compared to the initial draft model. Our refinement reduced false positive predictions by 17, leaving only 3 unresolved mismatches. The most remarkable improvement was in nitrogen and phosphorus sources predictions. The high predictive accuracy indicates that core metabolic pathways and multiple catabolic routes of DSM20746 have been accurately reconstructed within *i*RM23NL. Consequently, the network can predict the catabolism of numerous common compounds, such as sugars and amino acids.

Gene essentiality predictions using *i*RM23NL

Given the increased percentage of gene-associated reactions (Fig. 2, Panel C) and the high predictive accuracy of the metabolic reconstruction, we employed *i*RM23NL further to predict exploitable single gene knockouts. For this purpose, we systematically removed each biochemical reaction from the network and optimized *i*RM23NL to produce biomass using FBA. To mitigate the inherent variability of the optimization algorithms, we repeated our FBA simulation 100 times. Additionally, we employed parsimonious enzyme usage flux balance analysis (pFBA), which involves solving two sequential linear optimization problems to determine the flux distribution of the optimal solution while minimizing the total sum of flux. Then, we compared the predicted growth rates before and after introducing the simulated gene deletion. The FC_{gr} between the knocked-out and wild-type growth rates was employed as a proxy for the gene's essentiality. We proceeded with condition-specific *in silico* single gene deletions. For this purpose, we utilized a minimal and a nutrient-rich medium (M9 supplemented with glucose and LB) as well as two growth media that mimic the intra-human nasal passages and the lungs of CF patients [SNM (39) and SCFM (38)]. Generally, when subjected to nutrient-limited conditions, the model predicted a higher number of genes as essential for growth, while the count of essential genes remained consistent among oxic and anoxic conditions (Fig. 7, Panel A). In total, 4 metabolic genes exhibited a partially essential effect across all tested media. This indicates that these genes promote cellular fitness, and their deletion partially impairs the bacterium's capacity to generate biomass. These genes are the Trka family potassium uptake protein (WP_005506372.1), ribulose-phosphate 3-epimerase (WP_005507411.1), glucose-6-phosphate isomerase (WP_005508482.1), and transaldolase (WP_005509117.1). The majority of essential genes are involved in nucleotide metabolism, peptidoglycan biosynthesis, or the energy metabolism. These over-represented subsystems among the identified essential genes suggest their importance in supporting the bacterium's respiration (Fig. S4). Nevertheless, in nutrient-poor conditions (M9 medium) genes from the biosynthesis of leucine (2-isopropylmalate synthase; WP_005508679.1 and 3-isopropylmalate dehydratase; WP_005507445.1), valine (ketol-acid reductoisomerase; WP_005508646.1 and dihydroxy-acid dehydratase; WP_005509229.1), and chorismate (shikimate kinase; WP_005508729.1 and 3-dehydroquinate dehydratase; WP_005504658.1) were found to be critical for the organism's survival. Tables S4 and S5 list in detail the predicted essential genes, each corresponding to specific approaches employed in this study.

Subsequently, we conducted a protein sequence homology analysis with BLAST (42) against the human proteome to identify potential targets that could be exploited in future therapeutic strategies. For this, only genes highlighted as essential in both laboratory and synthetically defined media were considered (Fig. 7, Panel B). Overall, 35 essential genes were common in LB and M9, of which 20 common genes reported homologous counterparts in the human genome. Further analysis revealed that among these genes, five genes exhibited over 50% sequence similarity with homologous
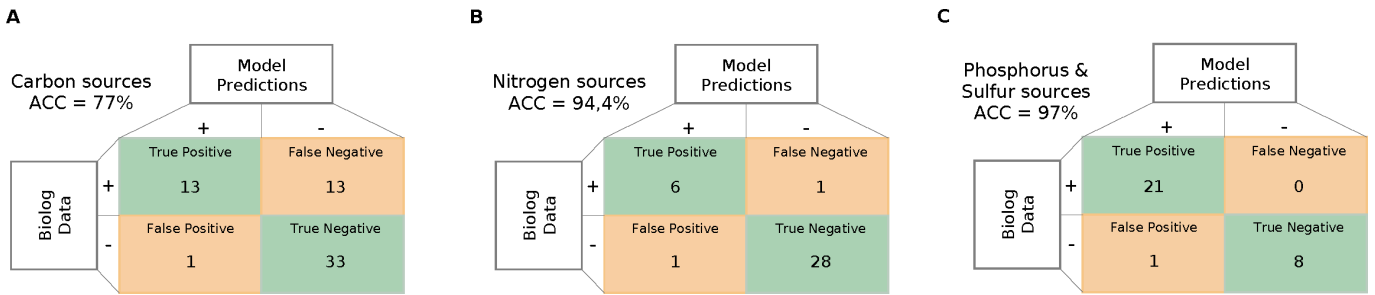
FIG 6 Predictive accuracy performance of *i*RM23NL using nutrient utilization data. Only substrates that exhibited complete mapping to both BiGG and model identifiers could be analyzed. Green represents correct predictions, and orange represents inconsistent predictions. The overall prediction accuracy of *i*RM23NL was computed using Equation 6.

proteins although none resulted in over 80% similarity. Similarly, when *i*RM23NL was simulated with SCFM and SNM in both aerobic and anaerobic conditions, 45 shared genes were predicted to be essential. Homology analysis against the human genome yielded 25 genes with exhibited homology in the human genomes, with 7 demonstrating over 50% sequence similarity. For instance, genes encoding proteins such as phosphopyruvate hydratase (WP_005506838.1), CTP synthase (WP_044141843.1), and adenylosuccinate synthase (WP_005509175.1) consistently exhibited human counterparts with similarity exceeding 50% across all tested growth media and oxygen levels. Among the essential genes shared between both LB and M9, 15 of them did not have any homologous hits. The same was observed for 20 common essential genes
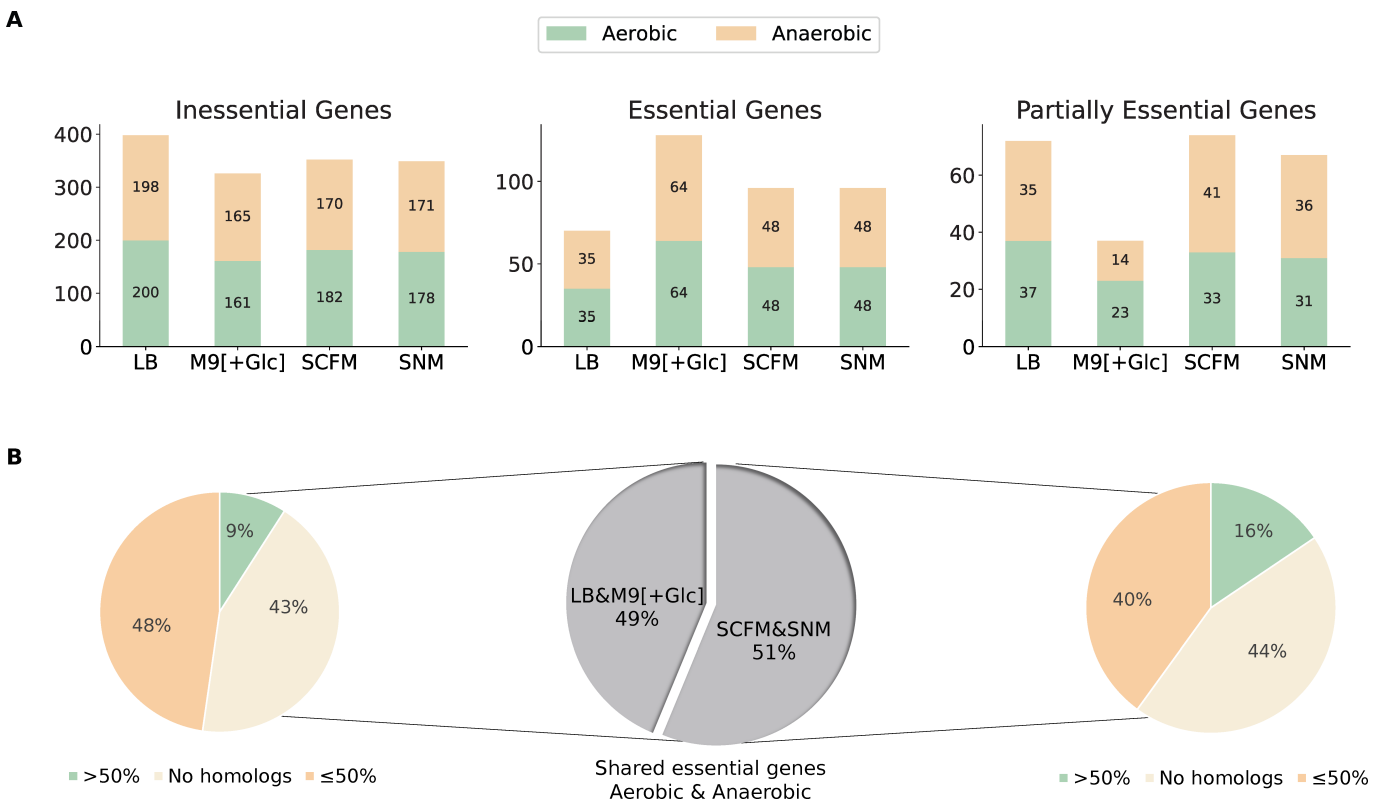
FIG 7 Comparative analysis of novel gene essentialities in *i*RM23NL across four distinct growth media. (A) Classification of network-derived single gene deletions within *i*RM23NL, classified into essential, inessential, and partially essential genes, when subjected to aerobic (green) and anaerobic (orange) environments. Details regarding the classification schema can be found in Materials and Methods. (B) Protein sequence homology analysis of genes predicted to be essential in the laboratory media (LB and M9 pure supplemented with glucose) and the synthetically defined SNM and SCFM in both oxygen-rich and oxygen-limited conditions. The percentage identity threshold was set to 50% similarity to the human proteome.

in SCFM and SNM. Some examples of these genes include orotate phosphoribosyl-transferase (WP_005507935.1), type I pantothenate kinase (WP_005505041.1), dihydro-neopterin aldolase (WP_005507619.1), and pantetheine-phosphate adenylyltransferase (WP_005508106.1). A more detailed comparison can be found in Table S6.

Our *in silico* transposon mutant analysis using *i*RM23NL could serve as a basis for several research and practical applications from rational and condition-specific drug target development to biotechnological applications and metabolic engineering.

DISCUSSION

The metabolic phenome of *R. mucilaginosa*, a bacterium with both beneficial and pathogenic behavior, remains still largely unexplored. Investigating its metabolic traits is of great importance as it holds the potential to unveil unique capabilities, including substrate utilization, byproduct production, and contributions to host-microbe interactions. *R. mucilaginosa* is a versatile microbe found in humans' oral, respiratory, and skin flora, where it coexists harmoniously. However, in immunocompromised individuals, *R. mucilaginosa* can act as an opportunistic pathogen, causing severe infections. Our study focuses on the metabolic aspects of *R. mucilaginosa*, particularly its behavior in isolated cultures. In 2019, a 17-species bacterial community model was reconstructed to simulate the polymicrobial community of the CF airways (43). This model accurately predicted the abundance of specific bacteria within patients' CF lung communities by linking metabolomics and 16S rRNA gene sequencing data. However, studying a bacterium's metabolism and genotype-phenotype relationships in monoculture provides a more controlled knowledge base. This allows for the precise manipulation of variables, enhancing our understanding of its individual traits, genetic makeup, metabolic pathways, and responses to stimuli (22, 23). Moreover, one can elucidate the bacterium's unique contributions to nutrient uptake, substrate production, and growth dynamics, crucial for understanding its role in a broader ecosystem. Monoculture studies identify key genes and pathways, revealing how the bacterium functions autonomously. Such analysis serves as a valuable reference, differentiating inherent characteristics from those influenced by external interactions. To this end, we empirically analyzed the metabolic phenome of *R. mucilaginosa* DSM20746 and developed the first high-quality strain-specific GEM of *R. mucilaginosa*, called *i*RM23NL. We considered literature and database organism-specific evidence to manually gap-fill the model and include highly relevant biochemical reactions. Phylogenetic analysis of further *Rothia* species provided insights into the relationship and genetic diversity between these species and was utilized to extend the metabolic network's completeness. Our model is simulation-ready, follows strictly community standards (25), and exhibits a high content quality MEMOTE score.

R. mucilaginosa is primarily aerobic and can perform oxic respiration by efficiently generating energy in the form of adenosine triphosphate (ATP) (1). However, when oxygen is limited or absent, *R. mucilaginosa* switches to anaerobic metabolism, which may involve fermentation or other alternative pathways to generate energy. As already mentioned, *R. mucilaginosa* has been previously found to be metabolically active in CF lungs where the oxygen levels are notably restricted (16). This indicates that the bacterium undergoes metabolic shift and can survive in microaerophilic environments. Various ROS products emerge as byproducts in the bacterial response to the fluctuating oxygen levels (34). In more detail, the cascade of ROS is initiated by the formation of O_2^- upon univalent oxygen reduction within the electron transport chain (ETC). Extreme oxygen fluctuations may be lethal and can ultimately damage cellular structure. The detoxifying pathway includes the enzymes superoxide dismutase (SOD), catalase, and peroxidase that break down lethal radicals to water and oxygen enabling the cell to neutralize the oxidative stress (44) (see Fig. 4). However, the exact anaerobic respiration mechanism of *R. mucilaginosa* must be thoroughly examined in experimental settings.

Since *R. mucilaginosa*'s metabolic behavior and adaptability are mainly yet unknown, we started by testing its growth behavior in various nutrient media. Exploring how bacteria react to various growth conditions within the human body is pivotal

for understanding diseases and developing effective treatments. Moreover, they are essential for evaluating their evolution and adaptation to different environmental conditions, leading to new ecological niches in which the bacterium could be metabolically active. We ultimately validated *i*RM23NL using our growth kinetics data in various growth media. Overall, *i*RM23NL's predictions were in line with the experimental observations. *R. mucilaginosa* demonstrated higher experimental growth in nutrient-rich media. The model successfully simulated growth for most media, while no biomass production was achieved in the M9 pure medium. When comparing LB to RPMI, the simulated growth rate was higher in LB, while the empirical growth in RPMI was twice as high as that in LB. This can be attributed to the fact that computer models cannot mimic the entire experimental settings and lack kinetic parameters. As of September 2023, bacteria like *S. aureus*, *B. subtilis*, and *E. coli* have been extensively researched for decades, with hundreds of thousands of PubMed (45) entries since the early 1990s. In contrast, *R. mucilaginosa*'s scientific prominence only began in the 21st century, with only 423 publications to date, indicating significant knowledge gaps crucial for metabolic reconstructions. More specialized BOF would enhance the predictive power and would reflect a more organism-specific metabolism. Therefore, this scarcity underscores the urgent need for further research efforts to uncover the hidden facets of *R. mucilaginosa*'s metabolism and its significance. Notably, to simulate *in silico* growth in RPMI and SCFM media, six metal ions needed to be supplemented. These metals have also been confirmed as essential for the *in silico* growth of *S. aureus* in RPMI (41). According to the model's predictions RPMI, supplementation with manganese, zinc, and molybdate was required. Transition metals could be highly toxic; however, in controlled levels are important in the survival of all living organisms (46). For instance, they are involved in redox catalysis, needed for energy production through respiration, and in non-redox catalysis, necessary for many biosynthetic and metabolic processes. Additionally, transition metals are required for the activity of many enzymes, including those involved in genomic replication and repair and nitrogen fixation. However, since these compounds were absent from the providers' medium formulation for RPMI, we speculate that the provided medium definition may not be exact. In all cases, the suggested metal co-factor promiscuity in *R. mucilaginosa* by *i*RM23NL needs to be examined to shed light on whether the bacterium could survive in the absence of one of the suggested metals.

Moreover, we experimentally characterized the strain's ability to assimilate and utilize substrates using high-throughput phenotypic microarray assays. The utilization of various nitrogen sources did not result in active respiration, indicating that the bacterial genome lacks genes encoding for respective transporters. We used the phenotypic results to validate and extend our metabolic reconstruction, *i*RM23NL. Inconsistencies between the model and the phenotypic microarray results served as a basis for further model refinement. We enriched the model with missing transport reactions and their respective GPRs by referring to the organism- and strain-specific BioCyc (47) registry and the General Feature Format (GFF) annotation file. All in all, characterizing and determining the repertoire of nutrient sources a strain can use or assimilate is a critical factor of pathogenesis. It provides valuable insights into how pathogens adapt to host environments and evade host defenses. Our transporter-augmented model reflects a high accuracy degree with the experimental data regarding using carbon, nitrogen, phosphorus, and sulfur sources. Discrepancies between computational and empirical results highlight areas of current uncertainty knowledge regarding the metabolism of *R. mucilaginosa*. They could be attributed to non-metabolic factors that fall beyond the metabolic models' scope, including regulatory processes, gene expression, and signaling pathways. However, targeted experiments are needed to fill the remaining network gaps and reveal novel enzymatic processes.

Considering the predictive precision of our metabolic reconstruction, we utilized *i*RM23NL to derive novel hypotheses. We examined the effects of condition-specific single gene knockouts on the bacterial capacity to produce biomass. Gene essentiality

analysis is inherently contingent upon specific conditions. In the context of constraint-based metabolic modeling, a plethora of constraints are established, with the availability of nutrients, often the growth medium, being the most prevalent. By altering the availability of these nutrients, the environmental conditions are modified, consequently exerting a profound influence on the metabolic state and growth of an organism (48, 49). However, the true strength and versatility of GEMs lie in their ability to rapidly generate condition-specific hypotheses on a large scale, circumventing the need for labor-intensive and expensive screenings that may not always yield direct success. Various models, spanning organisms like *A. baumannii*, *E. coli*, *S. cerevisiae*, *P. falciparum*, and *P. aeruginosa*, demonstrated predictive accuracies ranging from 72% to 93% (50–60). Additionally, gene essentiality analysis has been instrumental in identifying potential drug targets for diseases such as cancer and viral infections, aligning well with both *in vitro* (61–63) and *in vivo* (64) data. Therefore, we utlized our GEM and created a high-throughput *in silico*-derived transposon mutant library considering two standard growth media, LB and M9, along with two growth media formulated to mimic the environment within the human body, SNM and SCFM. In this regard, we identified putative essential and partially essential genes and assessed their potential vulnerability under varying nutritional environments. With this, we opted for detecting candidate genes that could be considered in future antimicrobial and anti-inflammatory strategies in immunocompromised and CF patients. With this, we opted for identifying candidate genes for future research that hold promise for experimental validation. Determining which essential genes have human counterparts is of great importance for antibiotic drug development, as it helps assess potential side effects and cross-species effects on human genes targeted by antibiotics. Moreover, it provides insights into the molecular mechanisms of host-pathogen interactions, explaining how pathogens manipulate host cells and evade the immune system. Utilizing our GEM offers promising venues for future targeted engineering strategies without the need for laborious large-scale screening of knockouts and mutant libraries. This methodology would facilitate the rapid design of metabolic gene knockout strains by eliminating the associated reaction(s) from the model. Finally, CF lungs represent a highly dynamic environment (65, 66). However, GEMs are adaptable and can be tailored to reflect the metabolic capabilities of bacteria across diverse environmental conditions.

The main objective in our endeavor to combat *R. mucilaginosa* as an opportunistic pathogen causing infections (7) is identifying essential genes, particularly those without human counterparts. Determining these essential genes is crucial as we aim to neutralize the pathogen without harming the host. Simultaneously, we are exploring *R. mucilaginosa* as an agent with anti-inflammatory properties (18). In this context, we opt for promoting *Rothia*'s growth, focusing on modulating the environmental conditions that have been reported to do so. Once the key pathways involved in the beneficial functions of *R. mucilaginosa* are known, our gene essentiality predictions can be exploited to boost activation of these pathways. Nonetheless, being aware of *R. mucilaginosa*-specific essential genes is crucial to avoid inadvertently targeting them during therapeutic treatments, ensuring both the bacterium's growth and its anti-inflammatory activities. With this dual perspective, we indicate *R. mucilaginosa*'s therapeutic variety, including developing strategies to combat the bacterium, when it is detrimental while increasing cell biomass production when its anti-inflammatory properties are beneficial. The latter could benefit human health in the context of cystic fibrosis. However, these model-driven hypotheses need to be extensively validated via *in vitro* and *in vivo* experiments.

Altogether, creating a genome-scale metabolic network for *R. mucilaginosa* reveals insights that would have been resource-intensive to acquire using traditional wet-lab means. Understanding the metabolic complexities of *R. mucilaginosa* is essential for expanding our basic understanding of bacterium's microbiology and would benefit various practical applications. In medicine, it could facilitate the development of strategies to deal with caused infections, while in biotechnology, it would allow us to use its metabolic abilities for bioprocessing and bioengineering purposes. Hence,

our high-quality metabolic network, *i*RM23NL, could provide a systematic and detailed framework for analyzing *R. mucilaginosa*'s metabolism, yielding valuable insights with broad-reaching impacts.

MATERIALS AND METHODS

Experimental settings

Bacterial strain and growth conditions

The *R. mucilaginosa* DSM20746 (ATCC 25296) used for the experimental work in this study is a type strain, and it was purchased from the American Type Culture Collection (ATCC, US). To create an inoculum, the bacterium was streaked onto nutrient agar (NA, Neogen, Heywood, UK) plates from a cryopreserved glycerol stock stored at −80°C using a sterile loop. Subsequently, the plates were incubated at 37°C for 48 h to form colonies (pure cultures). It is important to note that each biological replicate was conducted using pure cultures derived from the initial frozen stock (no sub-culturing). This ensures maintaining the genetic and phenotypic characteristics of the strain without introducing any potential mutations or adaptations.

Growth kinetics protocol

R. mucilaginosa overnight liquid cultures were prepared by adding bacterial colonies from pure cultures to 5 mL BHI (Neogen, Heywood, UK) and were put at 37 mL in a shaking incubator for 24 h. The initial OD was assessed and, if necessary, adjusted via up-concentration or dilution to achieve $OD_{590nm} = 0.25$. Then, the bacterial suspension was subjected to centrifugation at 10,000 RPM for 5 min, and the resulting pellet was re-suspended in the medium of interest at a dilution of 1:10. Ultimately, the inoculated growth media were transferred to a sterile 96 well-plate, including three technical replicates for each tested condition together with their corresponding control conditions (sterile growth media). The outer wells were filled with milliQ water (MQ) to prevent evaporation. The respective OD_{590nm} was measured aerobically at three distinct time points (0, 24, and 48 h) using an EnVision microplate reader (Perkin Elmer, Waltham, MA, USA). The microplates were incubated at 37°C during the interim periods between measurements. The final growth curves were generated for three biological replicates (n = 3) for the following growth media: BHI (baseline medium), LB (Neogen, Heywood, UK), M9 pure, RPMI medium (RPMI-1640 Sigma-Aldrich), and TSB (Neogen, Heywood, UK). In the M9 pure medium, only salts were considered. For detailed information regarding the constitution of M9, see Table S1. The rest of the media were prepared according to the providers' instructions.

The raw data were normalized by subtracting the blank values from the measured ODs and were summarized by calculating the arithmetic mean across all replicates. To interpret the growth of bacterial cells in all tested media and compare their growth characteristics, we employed the FC_{OD} ratio, which is defined as follows:

$$FC_{OD} = \frac{OD_{590nm}^{t\ =\ 48h}}{OD_{590nm}^{t\ =\ 0h}} \tag{1}$$

In this context, we define FC_{OD} below 1.4 as no growth, while FC_{OD} ratios greater than 1.4 indicate a growth increase over time. This FC_{OD} threshold was chosen by analyzing experimental data and growth curves. A value of 1.4 was selected, considering the range of calculated growth rates. Statistical tests, as described below, validated the threshold's reliability in accurately discerning growth from no growth in the bacterial cultures.

Phenotypic microarray screenings

DSM20746 was tested for utilizing multiple carbon, nitrogen, phosphorus, and sulfur sources. Biolog Phenotype Microarrays (PM, Hayward, CA, USA) were employed to test the utilization of 190 carbon (PM1 and PM2A), 95 nitrogen (PM3B), 59 phosphorus (PM4A), and 35 sulfur sources (PM4A). These assays use a tetrazolium redox dye to enable a colorimetric detection of active cell respiration across different nutrient sources (40). Normal cell respiration is indicated by the formation of a purple color as a result of the reduction of the colorless dye during incubation.

The PM plates were prepared following the manufacturer's protocol for Gram-positive bacteria. Table 1 lists the assay set up for of PM plates. However, modifications were made during the cell suspension preparation. The strain was grown on nutrient agar plates without undergoing sub-culturing. Using an inoculation loop, individual colonies were picked and suspended in an inoculating fluid (IF-0) at an absorbance of 0.0915 at 590 nm. Per the established protocol, 81% of transmittance (T) should be achieved. Given our measurement of OD, the subsequent conversion of transmittance to absorbance was carried out employing the following formula:

$$\text{Absorbance} = 2 - \log_{10}(\%T) \tag{2}$$

In each well of a 96 well-plate, we introduced 100 µL of cell suspension, followed by a 48 h incubation period at 37°C. Bacterial growth was measured by the OD at 590 nm using a VICTOR Nivo Multimode microplate reader. Each PM plate was tested in duplicate.

The subsequent analysis of the acquired data included calculating the arithmetic mean across all technical and biological replicates for all measured n time points. Background noise was also removed, and the data were normalized by subtracting the blank values from the actual measurements. The area under curve (AUC) was used to distinguish between growth (AUC ≥ 50) and no growth (AUC < 50). The computation of the AUCs was carried out by leveraging the linear trapezoidal rule that expresses the interpolation between data points:

$$\text{AUC}_{(t_{i+1} - t_i)} = \int_{t_i}^{t_{i+1}} f(x)\,dx \approx (t_{i+1} - t_i) \cdot \frac{1}{2}(\text{OD}_{t_{i+1}} + \text{OD}_{t_i}) \tag{3}$$

where t_i is the respective measured time point and $i \in \{0, ..., e\}$. More specifically, the trapezoidal rule is iteratively applied to adjacent data points defined along the curve whose summation resulted in the final AUC value. Hence for n measured data points, the final AUC value is defined as follows:

$$\text{AUC}_{t_e} = \sum_{i=0}^{e-1} \text{AUC}_{(t_{i+1} - t_i)} \tag{4}$$

Finally, we repeated this across the spectrum of tested compounds within the microarray plates.

TABLE 1 Assay configuration for diverse Biolog PM microplates combinations[a]

	For 1× PM
IF-0a GN/GP (1.2×)	10.0
Dye mix (100×)	0.12
PM additive (12x)	1.0
81%T cell suspension	0.88
Total volume	12.0

[a]Volumes are expressed in µL. The provided volume quantities are adequate for inoculating the specified number of plates in this study, using 100 µL/well with an additional excess.

Independent confirmatory testings of Biolog data

To independently confirm the Biolog data, we applied the growth kinetics protocol described above to 10 compounds. Although the base inoculating fluid (IF) used for the metabolic PM plates is proprietary, it is considered to reflect a minimal medium composed mainly of salts and buffers (40, 67). Hence, we used the M9 pure medium supplemented with different substrates to perform the independent tests (Fig. S2). The following compounds were examined: α-D-glucose, D-mannose, adonitol, L-ornithine, L-methionine, salicin, succinate, L-alanine, L-malate, and L-histidine. We also included negative controls of substrates with the Biolog inoculation fluid zero (IF-0). To ensure accuracy, triplicates were carried out for each tested compound. The M9 pure medium and the exact concentrations of added substrates are described in Table S1. All bacterial cell suspensions were prepared in 1:10 dilutions, and the OD_{590nm} were measured at 0, 24, and 48 h using an EnVision microplate reader (Perkin Elmer, Waltham, MA, US) and the associated software package.

We computed the arithmetic mean across the three replicates from the collected data set for each measured time point. Additionally, we performed a background correction to mitigate the influence of background noise or unwanted signal interference present in the measured ODs.

Statistical hypothesis analysis

We conducted statistical tests to evaluate the chosen threshold and potential statistically significant differences between measurements at the initial and final time-points, thereby indicating the significant growth or no growth. Specifically, we employed the Student's *t*-test for each experimental condition, taking into account the data from the three biological replicates. The null hypothesis is formulated as following: there is no significant difference between the measured OD values in starting and end time-points. Prior to hypothesis testing, we checked the correctness of associated assumptions. More specifically, we assessed data normality through the Shapiro–Wilk test and verified the homogeneity of variances using the Levene's test.

Computational framework and modeling methodology

Phylogenomic analysis

We supported the gap-filling process using evidence of closely related species within the *Rothia* genus. Employing ANIclustermap v.1.1.0 (68), we conducted a comprehensive genomic comparison involving *R. mucilaginosa* DSM20746 and 12 distinct *Rothia* species: *R. koreensis*, *R. kristinae*, *R. santali*, *R. halotolerans*, *R. aeria*, *R. dentocariosa*, *R. terrae*, *R. amarae*, *R. nasimurium*, *R. mucilaginosa*, *R. aerolata*, *R. nasisuis*, and *R. endophytica* (see Fig. 3). In brief, ANIclustermap creates an all-vs-all genome ANI clustermap and groups microbial genomes based on their genetic similarity. ANI is a pairwise measure to classify bacterial genomes according to their genetic similarity. It is defined as the genetic similarity across all orthologous genes shared between any two genomes (69, 70). It serves as a powerful tool for distinguishing strains of the same species or closely related species.

Draft model reconstruction and curation

The proteome of *R. mucilaginosa* DSM20746 (GCF_000175615.1) served as the basis for reconstructing a draft metabolic network. The DSM20746 (ATCC 25296) represents a type strain obtained from the throat, and its genetic and proteomic sequences were retrieved from National Centre for Biotechnology Information (NCBI) (https://www.ncbi.nlm.nih.gov). The genome sequence was annotated using the NCBI Prokaryotic Genome Annotation Pipeline (PGAP) (71). An initial draft model was built using CarveMe 1.5.1 (27). CarveMe uses mixed-integer linear programming (MILP) to convert a universal model into an organism-specific one by deleting metabolites and reactions

with low occurrence scores within the specific organism of interest. The universal BOF might yield incorrect gene essentiality predictions for biosynthesis pathways that rely on precursors unique to Gram-positive bacteria due to the absence of specific membrane and cell wall information. Hence, we chose the specialized Gram-positive template instead of the universal one to build our model more accurately. This Gram-positive template incorporates cell wall and membrane components specific to Gram-positive bacteria in contrast to the universal biomass reaction defined in CarveMe (27). Specifically, the BOF developed for *i*RM23NL includes essential macromolecules such as nucleotides and amino acids, co-enzymes, and inorganic ions. Moreover, it encompasses cell wall components like lipoteichoic acids, a peptidoglycan unit, and glycerol teichoic acids. The growth-associated (GAM) energy requirements are integrated into the biomass reaction (labeled as Growth), while non-growth associated maintenance (NGAM) is explicitly considered in the model, expressed by the reaction ATPM.

We conducted an extensive two-staged iterative gap-filling to address incomplete or missing information in the metabolic model. Gaps or missing reactions can arise for various reasons, such as incomplete genome annotations or undiscovered enzymatic activities. For this purpose, we leveraged information from both the bibliome and biochemical databases, including BioCyc (47). Thus, we ensured that the model could support the growth and viability of the organism under specific conditions.

We applied our previously published pipeline (26) to curate further the model based on community standards. The pipeline consists of eight steps, from which five (step 3–step 4) are related to model curation and ensure a high quality of the final model. In Summary, ModelPolisher (34) and SBOannotator (35) were employed to enrich the model with multiple cross-references, while the mass- and charge-unbalanced reactions were fixed. Further annotations integrated into the model encompassed: Evidence and Conclusion Ontology (ECO) terms representing the confidence level and the assertion method (biological qualifier: BQB_IS_DESCRIBED_BY), KEGG (30) subsystems as *groups:member* (biological qualifier: BQB_OCCURS_IN), and gene annotations. The latter was done by mapping the gene locus tags to the old tags using the GenBank GFF (72). Finally, we checked the presence of potential EGCs that could bias the final predictions (73). To manipulate the model structure, we employed the libSBML library (74).

The SBML Validator from libSBML (74) was used to assure a correct syntax of the model, while the quality control was carried out employing MEMOTE (36). However, it is worth noting that, as we discussed in our previous publication, MEMOTE considers only the parent nodes of the SBO directed acyclic graph excluding their respective children (26). Hence, MEMOTE was used carefully and not as an absolute quality indicator.

Linear programming: formulation of assumptions and constraints

FBA is used to determine the flux distribution through optimization of the objective function, typically the maximization of biomass production rate, under steady-state conditions (21). To address the under-determined nature of the system, constraints are imposed to define an allowable solution space that aligns with cellular functions. These constraints, encompassing mass balance, thermodynamics, and capacity, contribute to the FBA maximization problem. The linear programming problem used to obtain growth rates is described as follows:

$$
\begin{aligned}
\text{maximize} \quad & Z = c^T \vec{v} \\
\text{subject to:} \quad & \mathbf{S} \cdot \vec{v} = 0 \\
& v_{\min} \leq v_r \leq v_{\max} \qquad \text{for } r \in \{1, ..., n\} \\
& \forall r \in I : 0 \leq v_r
\end{aligned}
\tag{5}
$$

where $\vec{v}$ is the vector of fluxes within the network, $\mathbf{S}$ is the stoichiometric matrix, Z is the linear objective function, $\vec{c}$ is the vector of coefficients, and I represents an index set containing the indices of all irreversible reactions. The dimensionality of vector $\vec{v}$

matches the number of reactions, denoted as n in the system, and is consistent with the n columns in the matrix **S**.

The unit mmol/($g_{DW} \cdot$ h) is utilized to denote the predicted growth rates since the biomass consistency (rate at which biomass is produced per unit of dry weight in the model) of iRM23NL approaches 1 mmol/($g_{DW} \cdot$ h) (36, 75, 76). Consequently, direct comparisons can be made between experimentally observed and predicted growth rates. These conversions maintain their validity under the assumptions of constant volume and steady-state inherent in constraint-based modeling (75).

Bacterial growth analysis and nutrient utilization assays

Bacterial cell growth within various media and multiple substrate utilization evaluations were determined by solving Equation 5. The medium and the nutrient source of interest defined additional constraints. To achieve this objective, adjustments were made to the upper and lower limits of exchange reactions, as appropriate. We set specific uptake rates for key components within the growth medium as follows: the uptake rate of transition metals was set at 5.0 mmol/($g_{DW} \cdot$ h), the uptake rate of oxygen under aerobic conditions was established at 20.0 mmol/($g_{DW} \cdot$ h), and the rest media components equal to 10.0 mmol/($g_{DW} \cdot$ h). As previously mentioned, the M9 pure medium was used for the substrate utilization assays. Only substrates present in the metabolic network as intra- or extracellular metabolites were considered for the *in silico* validation. The results from the experimental and the *in silico* growth tests were categorized into "growth" (G) or "non-growth" (NG). Here, "growth" indicates the network's ability to generate biomass and, therefore, a positive growth rate. The model's overall prediction performance was assessed using the following statistical parameters: overall agreement (ACC):

$$ACC = \frac{TP + TN}{TP + TN + FP + FN} \tag{6}$$

and Matthews Correlation Coefficient (MCC):

$$MCC = \frac{(TP \cdot TN - FP \cdot FN)}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}} \tag{7}$$

where true negative (TN) and true positive (TP) represent accurate predictions, and false negative (FN) and false positive (FP) indicate incorrect predictions. Inconsistencies were resolved via iterative manual network gap-filling. For all FBA simulations, we employed the Constraints-Based Reconstruction and Analysis for Python (COBRApy) (77) package. All growth media definitions are listed in Table S2.

Gene lethality analysis

The *in silico* single-gene knockouts were performed as described in our previous study using FBA (26). To address the degeneracy issue of optimization, we additionally ran our FBA simulations in a total of 100 independent runs. Furthermore, we utilized pFBA, a method that allows us to ascertain the flux distribution of the optimal solution while concurrently minimizing the overall flux sum (78). The results were categorized as either essential $FC_{gr} = 0$, inessential ($FC_{gr} = 1$), or partially essential ($0 < FC_{gr} < 1$), where FC_{gr} denotes the FC bacterial growth rate before and after deletion (27). Shared essential genes between FBA and pFBA, as well as all tested conditions, were further aligned against the human genome using BLAST (42).

AUTHOR AFFILIATIONS

[1]Computational Systems Biology of Infections and Antimicrobial-Resistant Pathogens, Institute for Bioinformatics and Medical Informatics (IBMI), Eberhard Karl University of Tübingen, Tübingen, Germany

[2]Department of Computer Science, Eberhard Karl University of Tübingen, Tübingen, Germany

[3]Cluster of Excellence 'Controlling Microbes to Fight Infections', Eberhard Karl University of Tübingen, Tübingen, Germany

[4]German Center for Infection Research (DZIF), partner site Tübingen, Tübingen, Germany

[5]Quantitative Biology Center (QBiC), Eberhard Karl University of Tübingen, Tübingen, Germany

[6]Laboratory of Pharmaceutical Microbiology (LPM), Ghent University, Ghent, Belgium

[7]Data Analytics and Bioinformatics, Institute of Computer Science, Martin Luther University Halle-Wittenberg, Halle (Saale), Germany

AUTHOR ORCIDs

Nantia Leonidou http://orcid.org/0000-0002-0248-6679
Tom Coenye http://orcid.org/0000-0002-6407-0601
Aurélie Crabbé http://orcid.org/0000-0003-3084-4418
Andreas Dräger http://orcid.org/0000-0002-1240-5553

AUTHOR CONTRIBUTIONS

Nantia Leonidou, Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Project administration, Software, Validation, Visualization, Writing – original draft, Writing – review and editing | Lisa Ostyn, Investigation, Supervision | Tom Coenye, Supervision, Writing – review and editing | Aurélie Crabbé, Methodology, Supervision, Writing – review and editing | Andreas Dräger, Funding acquisition, Resources, Supervision, Writing – review and editing

DATA AVAILABILITY

Supplementary data are available along with this article. Additionally, iRM23NL is available at the BioModels Database (79) as an SBML Level 3 Version 1 (37) file wrapped in an Open Modelling EXchange format (OMEX) (80) archive file together with a metadata file (81). Access the model at https://identifiers.org/biomodels.db/MODEL2310240001.

ADDITIONAL FILES

The following material is available online.

Supplemental Material

Figure S1 (Spectrum04006-23-s0001.pdf). Experimentally-derived catabolic phenome of *R. mucilaginosa* DSM20746.

Figure S2 (Spectrum04006-23-s0002.pdf). Growth curves of the independent confirmatory tests for validating the Biolog PM results.

Figure S3 (Spectrum04006-23-s0003.pdf). Detailed comparative analysis of gene essentiality in silico predictions using iRM23NL.

Figure S4 (Spectrum04006-23-s0004.pdf). Distribution of essential genes in metabolic subsystems.

Table S1 (Spectrum04006-23-s0005.xlsx). Reconstitution of 10X M9 medium and amount of substrates supplemented.

Table S2 (Spectrum04006-23-s0006.xlsx). Detailed definition of growth media used for in silico simulations.

Table S3 (Spectrum04006-23-s0007.xlsx). Summary of high-throughput Biolog PM results.

Table S4 (Spectrum04006-23-s0008.xlsx). List of essential genes predicted by FBA in different nutritional media.

Table S5 (Spectrum04006-23-s0009.xlsx). List of essential genes predicted by pFBA in different nutritional media.

Table S6 (Spectrum04006-23-s0010.xlsx). Comparative analysis of predicted essential genes for *R. mucilaginosa* in different nutrient environments.

Open Peer Review

PEER REVIEW HISTORY (review-history.pdf). An accounting of the reviewer comments and feedback.

REFERENCES

1. Bergan T, Kocur M. 1982. *Stomatococcus mucilaginosus* gen.nov., sp.nov., ep. rev., a member of the family *Micrococcaceae*. Int J Syst Bacteriol 32:374–377. https://doi.org/10.1099/00207713-32-3-374
2. Collins MD, Hutson RA, Båverud V, Falsen E. 2000. Characterization of a *Rothia*-like organism from a mouse: description of *Rothia nasimurium* sp. nov. and reclassification of *Stomatococcus mucilaginosus* as *Rothia mucilaginosa* comb. nov. Int J Syst Evol Microbiol 50:1247–1251. https://doi.org/10.1099/00207713-50-3-1247
3. Olsen I, Preza D, Aas JA, Paster BJ. 2009. Cultivated and not-yet-cultivated bacteria in oral biofilms. Microb Ecol Health Dis 21:65–71. https://doi.org/10.1080/08910600902907509
4. Guglielmetti S, Taverniti V, Minuzzo M, Arioli S, Stuknyte M, Karp M, Mora D. 2010. Oral bacteria as potential probiotics for the pharyngeal mucosa. Appl Environ Microbiol 76:3948–3958. https://doi.org/10.1128/AEM.00109-10
5. Wos-Oxley ML, Plumeier I, von Eiff C, Taudien S, Platzer M, Vilchez-Vargas R, Becker K, Pieper DH. 2010. A poke into the diversity and associations within human anterior nare microbial communities. ISME J 4:839–851. https://doi.org/10.1038/ismej.2010.15
6. Wilbert SA, Mark Welch JL, Borisy GG. 2020. Spatial ecology of the human tongue dorsum microbiome. Cell Rep 30:4003–4015. https://doi.org/10.1016/j.celrep.2020.02.097
7. Dastager SG, Srinivasan Krishnamurthi NR, Dharne M. 2014. The family *Micrococcaceae*. In The prokaryotes
8. Janek D, Zipperer A, Kulik A, Krismer B, Peschel A. 2016. High frequency and diversity of antimicrobial activities produced by nasal *Staphylococcus* strains against bacterial competitors. PLoS Pathog 12:e1005812. https://doi.org/10.1371/journal.ppat.1005812
9. Uranga CC, Arroyo P, Duggan BM, Gerwick WH, Edlund A. 2020. Commensal oral *Rothia mucilaginosa* produces enterobactin, a metal-chelating siderophore. mSystems 5:e00161-20. https://doi.org/10.1128/mSystems.00161-20
10. Tunney MM, Field TR, Moriarty TF, Patrick S, Doering G, Muhlebach MS, Wolfgang MC, Boucher R, Gilpin DF, McDowell A, Elborn JS. 2008. Detection of anaerobic bacteria in high numbers in sputum from patients with cystic fibrosis. Am J Respir Crit Care Med 177:995–1001. https://doi.org/10.1164/rccm.200708-1151OC
11. Lim YW, Schmieder R, Haynes M, Willner D, Furlan M, Youle M, Abbott K, Edwards R, Evangelista J, Conrad D, Rohwer F. 2013. Metagenomics and metatranscriptomics: windows on CF-associated viral and microbial communities. J Cyst Fibros 12:154–164. https://doi.org/10.1016/j.jcf.2012.07.009
12. Worlitzsch D, Tarran R, Ulrich M, Schwab U, Cekici A, Meyer KC, Birrer P, Bellon G, Berger J, Weiss T, Botzenhart K, Yankaskas JR, Randell S, Boucher RC, Döring G. 2002. Effects of reduced mucus oxygen concentration in airway *Pseudomonas* infections of cystic fibrosis patients. J Clin Invest 109:317–325. https://doi.org/10.1172/JCI13870
13. Filkins LM, O'Toole GA. 2015. Cystic fibrosis lung infections: polymicrobial, complex, and hard to treat. PLoS Pathog 11:e1005258. https://doi.org/10.1371/journal.ppat.1005258
14. Guss AM, Roeselers G, Newton ILG, Young CR, Klepac-Ceraj V, Lory S, Cavanaugh CM. 2011. Phylogenetic and metabolic diversity of bacteria associated with cystic fibrosis. ISME J 5:20–29. https://doi.org/10.1038/ismej.2010.88
15. Bittar F, Richet H, Dubus J-C, Reynaud-Gaubert M, Stremler N, Sarles J, Raoult D, Rolain J-M. 2008. Molecular detection of multiple emerging pathogens in sputa from cystic fibrosis patients. PLoS One 3:e2908. https://doi.org/10.1371/journal.pone.0002908
16. Lim YW, Schmieder R, Haynes M, Furlan M, Matthews TD, Whiteson K, Poole SJ, Hayes CS, Low DA, Maughan H, Edwards R, Conrad D, Rohwer F. 2013. Mechanistic model of *Rothia mucilaginosa* adaptation toward persistence in the CF lung, based on a genome reconstructed from metagenomic data. PLoS One 8:e64285. https://doi.org/10.1371/journal.pone.0064285
17. Gao B, Gallagher T, Zhang Y, Elbadawi-Sidhu M, Lai Z, Fiehn O, Whiteson KL. 2018. Tracking polymicrobial metabolism in cystic fibrosis airways: *Pseudomonas aeruginosa* metabolism and physiology are influenced by

Rothia mucilaginosa-derived metabolites. mSphere 3:10–1128. https://doi.org/10.1128/mSphere.00151-18

18. Rigauts C, Aizawa J, Taylor SL, Rogers GB, Govaerts M, Cos P, Ostyn L, Sims S, Vandeplassche E, Sze M, Dondelinger Y, Vereecke L, Van Acker H, Simpson JL, Burr L, Willems A, Tunney MM, Cigana C, Bragonzi A, Coenye T, Crabbé A. 2022. *Rothia mucilaginosa* is an anti-inflammatory bacterium in the respiratory tract of patients with chronic lung disease. Eur Respir J 59:2101293. https://doi.org/10.1183/13993003.01293-2021

19. Passi A, Tibocha-Bonilla JD, Kumar M, Tec-Campos D, Zengler K, Zuniga C. 2021. Genome-scale metabolic modeling enables in-depth understanding of big data. Metabolites 12:14. https://doi.org/10.3390/metabo12010014

20. Palsson BØ. 2006. Systems biology: properties of reconstructed networks. Cambridge university press.

21. Orth JD, Thiele I, Palsson BØ. 2010. What is flux balance analysis? Nat Biotechnol 28:245–248. https://doi.org/10.1038/nbt.1614

22. Oberhardt MA, Palsson BØ, Papin JA. 2009. Applications of genome-scale metabolic reconstructions. Mol Syst Biol 5:320. https://doi.org/10.1038/msb.2009.77

23. Gu C, Kim GB, Kim WJ, Kim HU, Lee SY. 2019. Current status and applications of genome-scale metabolic models. Genome Biol 20:121. https://doi.org/10.1186/s13059-019-1730-3

24. Sertbas M, Ulgen KO. 2020. Genome-scale metabolic modeling for unraveling molecular mechanisms of high threat pathogens. Front Cell Dev Biol 8:566702. https://doi.org/10.3389/fcell.2020.566702

25. Carey MA, Dräger A, Beber ME, Papin JA, Yurkovich JT. 2020. Community standards to facilitate development and address challenges in metabolic modeling. Mol Syst Biol 16:e9235. https://doi.org/10.15252/msb.20199235

26. Leonidou N, Xia Y, Friedrich L, Schütz MS, Dräger A. 2023. Exploring the metabolic profiling of *A. baumannii* for antimicrobial development using genome-scale modeling. bioRxiv. https://doi.org/10.1101/2023.09.13.557502

27. Machado D, Andrejev S, Tramontano M, Patil KR. 2018. Fast automated reconstruction of genome-scale metabolic models for microbial species and communities. Nucleic Acids Res 46:7542–7553. https://doi.org/10.1093/nar/gky537

28. Norsigian CJ, Pusarla N, McConn JL, Yurkovich JT, Dräger A, Palsson BO, King Z. 2020. BiGG models 2020: multi-strain genome-scale models and expansion across the phylogenetic tree. Nucleic Acids Res 48:D402–D406. https://doi.org/10.1093/nar/gkz1054

29. Lex A, Gehlenborg N, Strobelt H, Vuillemot R, Pfister H. 2014. UpSet: visualization of intersecting sets. IEEE Trans Vis Comput Graph 20:1983–1992. https://doi.org/10.1109/TVCG.2014.2346248

30. Kanehisa M, Furumichi M, Sato Y, Ishiguro-Watanabe M, Tanabe M. 2021. KEGG: integrating viruses and cellular organisms. Nucleic Acids Res 49:D545–D551. https://doi.org/10.1093/nar/gkaa970

31. Zhao X, Drlica K. 2014. Reactive oxygen species and the bacterial response to lethal stress. Curr Opin Microbiol 21:1–6. https://doi.org/10.1016/j.mib.2014.06.008

32. Korge P, Calmettes G, Weiss JN. 2015. Increased reactive oxygen species production during reductive stress: the roles of mitochondrial glutathione and thioredoxin reductases. Biochim Biophys Acta 1847:514–525. https://doi.org/10.1016/j.bbabio.2015.02.012

33. Monk J, Nogales J, Palsson BO. 2014. Optimizing genome-scale network reconstructions. Nat Biotechnol 32:447–452. https://doi.org/10.1038/nbt.2870

34. Römer M, Eichner J, Dräger A, Wrzodek C, Wrzodek F, Zell A. 2016. ZBIT bioinformatics toolbox: a web-platform for systems biology and expression data analysis. PLoS One 11:e0149263. https://doi.org/10.1371/journal.pone.0149263

35. Leonidou N, Fritze E, Renz A, Dräger A. 2023. SBOannotator: a python tool for the automated assignment of systems biology ontology terms. Bioinformatics 39:btad437. https://doi.org/10.1093/bioinformatics/btad437

36. Lieven C, Beber ME, Olivier BG, Bergmann FT, Ataman M, Babaei P, Bartell JA, Blank LM, Chauhan S, Correia K, et al. 2020. MEMOTE for standardized genome-scale metabolic model testing. Nat Biotechnol 38:272–276. https://doi.org/10.1038/s41587-020-0446-y

37. Hucka M, Bergmann FT, Dräger A, Hoops S, Keating SM, Le Novère N, Myers CJ, Olivier BG, Sahle S, Schaff JC, Smith LP, Waltemath D,

Wilkinson DJ. 2018. The systems biology markup language (SBML): language specification for level 3 version 1 core. J Integr Bioinform 15:20170080. https://doi.org/10.1515/jib-2017-0080

38. Turner KH, Wessel AK, Palmer GC, Murray JL, Whiteley M. 2015. Essential genome of *Pseudomonas aeruginosa* in cystic fibrosis sputum. Proc Natl Acad Sci U S A 112:4110–4115. https://doi.org/10.1073/pnas.1419677112

39. Krismer B, Liebeke M, Janek D, Nega M, Rautenberg M, Hornig G, Unger C, Weidenmaier C, Lalk M, Peschel A. 2014. Nutrient limitation governs *Staphylococcus aureus* metabolism and niche adaptation in the human nose. PLoS Pathog 10:e1003862. https://doi.org/10.1371/journal.ppat.1003862

40. Bochner BR, Gadzinski P, Panomitros E. 2001. Phenotype microarrays for high-throughput phenotypic testing and assay of gene function. Genome Res 11:1246–1255. https://doi.org/10.1101/gr.186501

41. Seif Y, Monk JM, Mih N, Tsunemoto H, Poudel S, Zuniga C, Broddrick J, Zengler K, Palsson BO. 2019. A computational knowledge-base elucidates the response of *Staphylococcus aureus* to different media types. PLoS Comput Biol 15:e1006644. https://doi.org/10.1371/journal.pcbi.1006644

42. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. J Mol Biol 215:403–410. https://doi.org/10.1016/S0022-2836(05)80360-2

43. Henson MA, Orazi G, Phalak P, O'Toole GA. 2019. Metabolic modeling of cystic fibrosis airway communities predicts mechanisms of pathogen dominance. mSystems 4:e00026-19. https://doi.org/10.1128/mSystems.00026-19

44. Baron S. 1996. Medical microbiology

45. Wheeler DL, Chappey C, Lash AE, Leipe DD, Madden TL, Schuler GD, Tatusova TA, Rapp BA. 2000. Database resources of the national center for biotechnology information. Nucleic Acids Res 28:10–14. https://doi.org/10.1093/nar/28.1.10

46. Hood MI, Skaar EP. 2012. Nutritional immunity: transition metals at the pathogen-host interface. Nat Rev Microbiol 10:525–537. https://doi.org/10.1038/nrmicro2836

47. Karp PD, Billington R, Caspi R, Fulcher CA, Latendresse M, Kothari A, Keseler IM, Krummenacker M, Midford PE, Ong Q, Ong WK, Paley SM, Subhraveti P. 2019. The BioCyc collection of microbial genomes and metabolic pathways. Brief Bioinform 20:1085–1093. https://doi.org/10.1093/bib/bbx085

48. Gasperotti A, Brameyer S, Fabiani F, Jung K. 2020. Phenotypic heterogeneity of microbial populations under nutrient limitation. Curr Opin Biotechnol 62:160–167. https://doi.org/10.1016/j.copbio.2019.09.016

49. Camacho A, Rochera C, Picazo A. 2022. Effect of experimentally increased nutrient availability on the structure, metabolic activities, and potential microbial functions of a maritime Antarctic microbial mat. Front Microbiol 13:900158. https://doi.org/10.3389/fmicb.2022.900158

50. Presta L, Bosi E, Mansouri L, Dijkshoorn L, Fani R, Fondi M. 2017. Constraint-based modeling identifies new putative targets to fight colistin-resistant *A. baumannii* infections. Sci Rep 7:3706. https://doi.org/10.1038/s41598-017-03416-2

51. Norsigian CJ, Kavvas E, Seif Y, Palsson BO, Monk JM. 2018. iCN718, an updated and improved genome-scale metabolic network reconstruction of *Acinetobacter baumannii* AYE. Front Genet 9:121. https://doi.org/10.3389/fgene.2018.00121

52. Zhao J, Zhu Y, Han J, Lin Y-W, Aichem M, Wang J, Chen K, Velkov T, Schreiber F, Li J. 2020. Genome-scale metabolic modeling reveals metabolic alterations of multidrug-resistant *Acinetobacter baumannii* in a murine bloodstream infection model. Microorganisms 8:11. https://doi.org/10.3390/microorganisms8111793

53. Zhu Y, Zhao J, Maifiah MHM, Velkov T, Schreiber F, Li J. 2019. Metabolic responses to polymyxin treatment in *Acinetobacter baumannii* ATCC 19606: integrating transcriptomics and metabolomics with genome-scale metabolic modeling. mSystems 4:e00157-18. https://doi.org/10.1128/mSystems.00157-18

54. Monk JM, Lloyd CJ, Brunk E, Mih N, Sastry A, King Z, Takeuchi R, Nomura W, Zhang Z, Mori H, Feist AM, Palsson BO. 2017. *i*ML1515, a knowledge-base that computes *Escherichia coli* traits. Nat Biotechnol 35:904–908. https://doi.org/10.1038/nbt.3956

55. Joyce AR, Reed JL, White A, Edwards R, Osterman A, Baba T, Mori H, Lesely SA, Palsson BØ, Agarwalla S. 2006. Experimental and

computational assessment of conditionally essential genes in *Escherichia coli*. J Bacteriol 188:8259–8271. https://doi.org/10.1128/JB.00740-06

56. Abdel-Haleem AM, Hefzi H, Mineta K, Gao X, Gojobori T, Palsson BO, Lewis NE, Jamshidi N. 2018. Functional interrogation of *Plasmodium* genus metabolism identifies species- and stage-specific differences in nutrient essentiality and drug targeting. PLoS Comput Biol 14:e1005895. https://doi.org/10.1371/journal.pcbi.1005895

57. Oh Y-K, Palsson BO, Park SM, Schilling CH, Mahadevan R. 2007. Genome-scale reconstruction of metabolic network in *Bacillus subtilis* based on high-throughput phenotyping and gene essentiality data. J Biol Chem 282:28791–28799. https://doi.org/10.1074/jbc.M703759200

58. Nanda P, Patra P, Das M, Ghosh A. 2020. Reconstruction and analysis of genome-scale metabolic model of weak Crabtree positive yeast *Lachancea kluyveri*. Sci Rep 10:16314. https://doi.org/10.1038/s41598-020-73253-3

59. Oberhardt MA, Puchałka J, Fryer KE, Martins dos Santos VAP, Papin JA. 2008. Genome-scale metabolic network analysis of the opportunistic pathogen *Pseudomonas aeruginosa* PAO1. J Bacteriol 190:2790–2803. https://doi.org/10.1128/JB.01583-07

60. Harrison R, Papp B, Pál C, Oliver SG, Delneri D. 2007. Plasticity of genetic interactions in metabolic networks of yeast. Proc Natl Acad Sci U S A 104:2307–2312. https://doi.org/10.1073/pnas.0607153104

61. Aller S, Scott A, Sarkar-Tyson M, Soyer OS. 2018. Integrated human-virus metabolic stoichiometric modelling predicts host-based antiviral targets against Chikungunya, Dengue and Zika viruses. J R Soc Interface 15:146. https://doi.org/10.1098/rsif.2018.0125

62. Bidkhori G, Benfeitas R, Elmas E, Kararoudi MN, Arif M, Uhlen M, Nielsen J, Mardinoglu A. 2018. Metabolic network-based identification and prioritization of anticancer targets based on expression data in hepatocellular carcinoma. Front Physiol 9:916. https://doi.org/10.3389/fphys.2018.00916

63. Agren R, Mardinoglu A, Asplund A, Kampf C, Uhlen M, Nielsen J. 2014. Identification of anticancer drugs for hepatocellular carcinoma through personalized genome-scale metabolic modeling. Mol Syst Biol 10:721. https://doi.org/10.1002/msb.145122

64. Renz A, Hohner M, Breitenbach M, Josephs-Spaulding J, Dürrwald J, Best L, Jami R, Marinos G, Cabreiro F, Dräger A, Schindler M, Kaleta C. 2023. Metabolic modeling elucidates phenformin and atpenin A5 as broad-spectrum antiviral drugs. Biol Life sci. https://doi.org/10.20944/preprints202210.0223.v2

65. Zhao J, Schloss PD, Kalikin LM, Carmody LA, Foster BK, Petrosino JF, Cavalcoli JD, VanDevanter DR, Murray S, Li JZ, Young VB, LiPuma JJ. 2012. Decade-long bacterial community dynamics in cystic fibrosis airways. Proc Natl Acad Sci U S A 109:5809–5814. https://doi.org/10.1073/pnas.1120577109

66. Greenwald MA, Wolfgang MC. 2022. The changing landscape of the cystic fibrosis lung environment: from the perspective of *Pseudomonas aeruginosa*. Curr Opin Pharmacol 65:102262. https://doi.org/10.1016/j.coph.2022.102262

67. Zhou L, Lei X-H, Bochner BR, Wanner BL. 2003. Phenotype microarray analysis of *Escherichia coli* K-12 mutants with deletions of all two-component systems. J Bacteriol 185:4956–4972. https://doi.org/10.1128/JB.185.16.4956-4972.2003

68. Shimoyama Y. 2022. ANIclustermap: a tool for drawing ANI clustermap between all-vs-all microbial genomes

69. Konstantinidis KT, Tiedje JM. 2005. Genomic insights that advance the species definition for prokaryotes. Proc Natl Acad Sci U S A 102:2567–2572. https://doi.org/10.1073/pnas.0409727102

70. Goris J, Konstantinidis KT, Klappenbach JA, Coenye T, Vandamme P, Tiedje JM. 2007. DNA-DNA hybridization values and their relationship to whole-genome sequence similarities. Int J Syst Evol Microbiol 57:81–91. https://doi.org/10.1099/ijs.0.64483-0

71. Tatusova T, DiCuccio M, Badretdin A, Chetvernin V, Nawrocki EP, Zaslavsky L, Lomsadze A, Pruitt KD, Borodovsky M, Ostell J. 2016. NCBI prokaryotic genome annotation pipeline. Nucleic Acids Res 44:6614–6624. https://doi.org/10.1093/nar/gkw569

72. Clark K, Karsch-Mizrachi I, Lipman DJ, Ostell J, Sayers EW. 2016. GenBank. Nucleic Acids Res 44:D67–D72. https://doi.org/10.1093/nar/gkv1276

73. Fritzemeier CJ, Hartleb D, Szappanos B, Papp B, Lercher MJ. 2017. Erroneous energy-generating cycles in published genome scale metabolic networks: identification and removal. PLoS Comput Biol 13:e1005494. https://doi.org/10.1371/journal.pcbi.1005494

74. Bornstein BJ, Keating SM, Jouraku A, Hucka M. 2008. LibSBML: an API library for SBML. Bioinformatics 24:880–881. https://doi.org/10.1093/bioinformatics/btn051

75. Gottstein W, Olivier BG, Bruggeman FJ, Teusink B. 2016. Constraint-based stoichiometric modelling from single organisms to microbial communities. J R Soc Interface 13:124. https://doi.org/10.1098/rsif.2016.0627

76. Chan SHJ, Cai J, Wang L, Simons-Senftle MN, Maranas CD. 2017. Standardizing biomass reactions and ensuring complete mass balance in genome-scale metabolic models. Bioinformatics 33:3603–3609. https://doi.org/10.1093/bioinformatics/btx453

77. Ebrahim A, Lerman JA, Palsson BO, Hyduke DR. 2013. COBRApy: constraints-based reconstruction and analysis for python. BMC Syst Biol 7:1–6. https://doi.org/10.1186/1752-0509-7-74

78. Lewis NE, Hixson KK, Conrad TM, Lerman JA, Charusanti P, Polpitiya AD, Adkins JN, Schramm G, Purvine SO, Lopez-Ferrer D, Weitz KK, Eils R, König R, Smith RD, Palsson BØ. 2010. Omic data from evolved *E. coli* are consistent with computed optimal growth from genome-scale models. Mol Syst Biol 6:390. https://doi.org/10.1038/msb.2010.47

79. Malik-Sheriff RS, Glont M, Nguyen TVN, Tiwari K, Roberts MG, Xavier A, Vu MT, Men J, Maire M, Kananathan S, Fairbanks EL, Meyer JP, Arankalle C, Varusai TM, Knight-Schrijver V, Li L, Dueñas-Roca C, Dass G, Keating SM, Park YM, Buso N, Rodriguez N, Hucka M, Hermjakob H. 2020. BioModels—15 years of sharing computational models in life science. Nucleic Acids Res 48:D407–D415. https://doi.org/10.1093/nar/gkz1055

80. Bergmann FT, Adams R, Moodie S, Cooper J, Glont M, Golebiewski M, Hucka M, Laibe C, Miller AK, Nickerson DP, Olivier BG, Rodriguez N, Sauro HM, Scharm M, Soiland-Reyes S, Waltemath D, Yvon F, Le Novère N. 2014. COMBINE Archive and OMEX format: One file to share all information to reproduce a modeling project. BMC Bioinformatics 15:369. https://doi.org/10.1186/s12859-014-0369-z

81. Neal ML, König M, Nickerson D, Mısırlı G, Kalbasi R, Dräger A, Atalag K, Chelliah V, Cooling MT, Cook DL, et al. 2019. Harmonizing semantic annotations for computational models in biology. Brief Bioinform 20:540–550. https://doi.org/10.1093/bib/bby087