Universidade do Minho
Escola de Engenharia

Inês Tavares Martins

Predicting the Impact of Psychoactive Compounds on Human Gut Bacteria with Metabolic Modelling

october 2022

Predicting the Impact of Psychoactive Compounds on Human Gut Bacteria with Metabolic Modelling

Inês Tavares Martins

UMinho | 2022

Universidade do Minho
Escola de Engenharia

Inês Tavares Martins

Predicting the Impact of Psychoactive Compounds on Human Gut Bacteria with Metabolic Modelling

Master's Dissertation
Bioinformatics
Information Technologies

Under supervision of
Professor Doctor Miguel Rocha
Doctor Maria Zimmermann-Kogadeeva

october 2022

Statement of integrity

I hereby declare having conducted this academic work with integrity. I confirm that I have not used plagiarism or any form of undue use of information or falsification of results along the process leading to its elaboration.

I further declare that I have fully acknowledged the Code of Ethical Conduct of the University of Minho.

Previsão do Impacto de Compostos Psicoativos nas Bactérias do Intestino Humano com Modelação Metabólica

O intestino humano é composto por muitos micróbios, a microbiota, que vai desde bactérias e vírus até archaea e eukaria, sendo que as bactérias anaeróbicas compõem a maior parte, com cerca de 500-1000 espécies bacterianas. Cada bactéria tem um genoma que abrange milhares de genes e cada pessoa tem um conjunto diferente de bactérias.

A medicação tem surgido como um dos fatores extrínsecos do hospedeiro que tem um forte impacto na variação da composição da microbiota intestinal humana. Uma vez que os recetores dos medicamentos, as enzimas que os metabolizam e os seus transportadores são produtos de genes que apresentam polimorfismos, a genética introduz variabilidade nas respostas dos indivíduos aos medicamentos e pode ser a causa das reações adversas a estes.

O metabolismo celular é um sistema confiável para a análise preditiva dos efeitos secundários dos medicamentos, uma vez que pode ser transformado num modelo preditivo, designado como modelo metabólico à escala genómica.

Este estudo baseia-se na investigação de Maier *et al* 2018, que testou o efeito de 1197 medicamentos que não têm como alvo as bactérias comensais do intestino (como medicamentos para o sistema nervoso) contra bactérias que são representativas do intestino humano, de forma a perceber se estes têm um efeito antibiótico secundário.

Neste estudo, foi analisado o efeito destes medicamentos no crescimento da bactéria *Bacteroides thetaiotaomicron*, *in silico*, através de dois modelos metabólicos (para comparação), um dos quais foi curado manualmente e obtido da literatura e o outro foi reconstruído neste estudo com uma ferramenta automática chamada CarveMe. Este estudo mostra que apenas 4 medicamentos tiveram o mesmo efeito na bactéria *in vitro* e num dos modelos *in silico*, sendo que no outro apenas 2. Algumas hipóteses podem ser formuladas, entre as quais que nenhum destes modelos é representativo do metabolismo da bactéria ou que as condições experimentais não foram exatamente representadas nos modelos *in silico*.

Palavras-chave: inativação de genes, modelos metabólicos, microbioma do intestino

Predicting the Impact of Psychoactive Compounds on Human Gut Bacteria with Metabolic Modelling

The human gut is comprised of many microbes, the microbiota, ranging from bacteria and viruses to archaea and eukarya, being that anaerobic bacteria make up the most part, with about 500-1000 bacterial species. Each bacterium has a genome encompassing thousands of genes and each person has a different set of bacteria.

Medication has been emerging as one of the host extrinsic factors that has a strong impact in the variance of the human gut microbiota composition. Since drug receptors, drug metabolising enzymes and drug transporters are the products of genes that exhibit polymorphisms, genetics introduces variability among the response of individuals and may be the cause of adverse reactions to treatment.

Cellular metabolism is a reliable system for predictive analysis of drugs side effects since it is a genome-wide network that can be turned into a predictive model, designated as genome-scale metabolic model (GSMM).

This study is based on the experimental Maier *et al* 2018 research, that screened 1197 drugs not targeted at human gut commensal bacteria (such as nervous system drugs) against representatives of the human gut, to investigate the antibiotic-like side effect of these drugs.

In the present study, it is analysed the effect of these drugs on the growth of the bacterium *Bacteroides thetaiotaomicron*, *in silico*, through two GSMMs (for comparison purposes), one of which manually curated and obtained from literature and the other reconstructed in this study with the use of an automatic tool named CarveMe. It was found that only 4 drugs had corresponding results *in vitro* and in one of the *in silico* models and only 2 in the other model. Some hypotheses can be made, such as that these GSMMs are not representative of the bacterium metabolism or that the experimental conditions were not exactly represented in the models.

Key words: gene-deletion, metabolic modelling, gut microbiome

Table of Contents

List of Figures

List of Tables

1 Introduction

1.1 Context and Motivation

The human gut harbours countless microbes, the microbiota, which have an active role in functions such as immunological and developmental, and have been linked to brain development, physiology and psychology primarily through neuroactive compound production and degradation [1, 2].

Drugs that act on the nervous system (neuroactive drugs), which include antipsychotics, antidepressants and antianxiety drugs, act on the host central nervous system and their consumption has been associated with modifications in the microbiota composition [1][2][3]. Furthermore, drugs that are not targeted at gut commensals have also been shown to change the microbiome diversity [2].

Therefore, it is crucial to unravel interactions between the microbiome and drugs and take them into account during drug development and selection of medical treatment, in order to control drug responses and side effects [2, 4, 5, 6]. Although numerous studies have reported some interactions, systematic evaluation of these relationships is yet missing [3].

1.2 Aims

The goal of this project is to systematically analyse drug effects on bacterial metabolism and growth, based on data on drug-protein interactions and based on genome-scale metabolic models (GSMM) of gut bacteria. These are further divided into the following aims:

- Reconstruct the GSMM of *Bacteroides thetaiotaomicron* with an automatic tool named CarveMe.
- Examine the GSMM of *B. thetaiotaomicron* for its ability to produce known neuroactive compounds.
- Incorporate drug-protein interaction information from the STITCH and KEGG databases into to the GSMM to predict the effects of drugs, screened *in vitro* in Maier *et al* 2018, on bacterial growth and metabolism.
- Repeat the previous step for a GSMM retrieved from literature that was reconstructed with a different automatic tool named ModelSEED and was manually curated.
- Compare *in silico* results with *in vitro* ones.

1.3 Structure

The dissertation has the following structure:

- Chapter 2: State-of-the-Art
  - — Gut microbiome interindividual variability.
  - — Bi-directional communication between the brain and the gut.
  - — The side effects of drugs in the gut bacteria.
  - — Genome-scale metabolic modelling to predict the effects of a drug on a bacterium.
- Chapter 3: Methods
  - — Automatic reconstruction of *B. thetaiotaomicron*'s genome-scale metabolic model.
  - — *In silico* simulation of drugs' effect on Bacterium's growth and metabolism.
  - — Using a manually curated *B. thetaiotaomicron*'s model to compare and understand drugs' effect results *in silico*.
- Chapter 4: Results
  - — Analyses of automatic model reconstructions.
  - — Investigation of bacterium's ability to metabolise neuroactive compounds.
  - — Results of drugs' effect simulation.
- Chapter 5: Discussion
- Chapter 6: Conclusions

2 State-of-the-Art

2.1 Human Gut Microbiome

The gut flora is comprised of many microbes, the microbiota, ranging from bacteria and viruses to archaea and eukarya, being that anaerobic bacteria make up the most part, hence being the most studied [4][5]. All the genomes belonging to the microbiota constitute the microbiome [5][6].

2.1.1 Composition / Interpersonal Variability

Due to the highly diverse microbial community of the microbiota, present in diverse niches and active at varying rates, numerous interactions among it and with the host exist, such as cross-feeding (metabolic exchange between microorganisms), mutualism, commensalism and competition (e.g., genes producing antimicrobials or competition for the same niche) [7][5][8].

The microbiota composition varies from person to person and throughout life with microbiome's intrinsic and extrinsic factors (such as taxa interactions, in the first case, and diet, genetics, pollutants, cohabitation and use of medication, in the last case) [5][6][2][9][10].

According to a longitudinal study done in a Dutch population, environment and cohabitation were the primary factors that explained microbiome variance (as shown in Fig. 1b) and only 15% of the variability could be explained by extrinsic factors, which means that the microbiome is very unique [5][6][2][9][10].

However, it can be said that the microbiome is fairly stabilised after the age of 3 and, around 65 years old, gradual changes start to occur (characterized, for example, by a depletion in the core abundant genera "Bacteroides" [Fig. 1a] and by changes in the metabolic capacity [11][10][3]) and the microbiome interindividual variation increases [12][11][10].

Gut microbes are responsible for catabolising nutrients that go unmodified to the gut because they are not digested by host enzymes. This results in the production of metabolites, among which are short chain fatty acids (e.g., acetate, propionate and butyrate), branched chain amino acids and gases. This catabolic process might need combinations of microorganisms for it to happen [13][7][5]. Microbes also metabolise other compounds, as for example, bacterial cell wall components. All of these metabolites, that together define the metabolome, can later be absorbed or excreted by the host through metabolic processes and pathways [7][5].

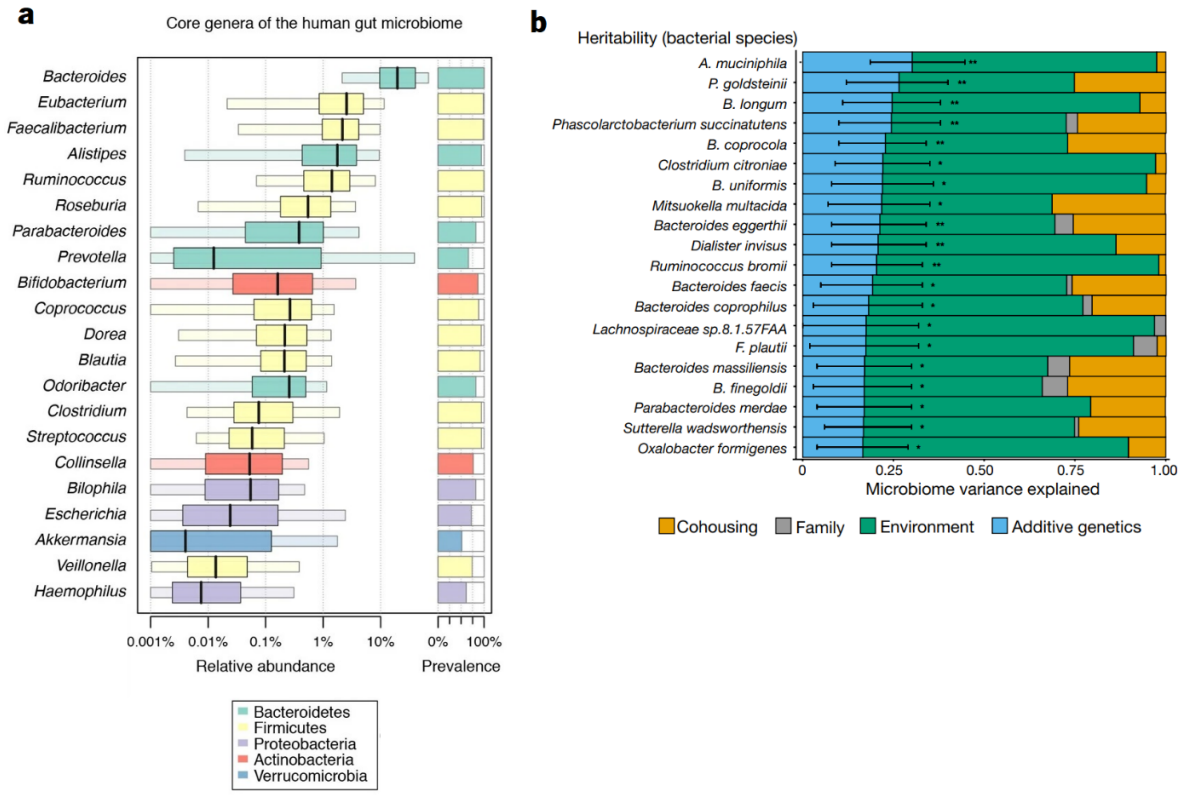


Fig. 1. Gut microbiome composition. **a**, Core of species in the human gut microbiome (from 364 fecal metagenomes of asymptomatic individuals; figure taken from Maier *et al* 2018 [3]). **b**, Heritability and contribution of individual exposures to microbiome variance (figure from [10]).

2.1.2 Gut-Brain Communication Driven by Microbiota-Derived Neuroactive Metabolites

The nervous system allows communication between the brain and all parts of the body, through the transmission of electrical signals along the extension of one neuron (the axon) towards the dendrite of the postsynaptic neuron. This transmission involves the release of neurotransmitters that bind targets in the postsynaptic neuron and can cause its excitation, inhibition or modulation [14].

The nervous system comprises two main parts:

- the central nervous system (CNS), consisting mainly of the spinal cord and the brain and is named so since it integrates information from the rest of the body and controls its activity [15].
- the peripheral nervous system (PNS), that is made up from the nerves that link the skin, limbs and other organs to the CNS [15]. These nerves can be grouped, according to the location of connection to the CNS, into spinal nerves (that connect to the spinal cord) and cranial nerves (that connect directly to the brain). The latter include the vagus nerve, that is the longest cranial nerve since it runs from the brain stem to part of the colon (Fig. 2) [16, 17]. Moreover, the PNS can be divided into:
  — the somatic system, that controls voluntary muscle activity through innervation of skeletal muscles [18][19].

— the autonomic system, that encompasses neurons innervating blood vessels, lymphoid tissue and internal organs, being implicated in the control of physiological processes [18][19]. This system further includes neurons located within the gut, named enteric neurons, which around 500 million of them make up the enteric nervous system (ENS) [20][21].

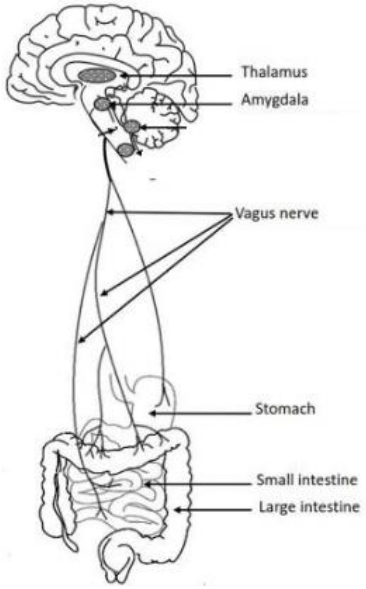


Fig. 2. Gut connection to the brain.

The gut and the central nervous system have been shown in research to have a strong connection, which has been designated as "gut-brain axis" and is modulated by activation of neurons in the gut, by hormones and by immune signals [12][22]. Moreover, metabolites produced by the gut microbiota have gained special interest by research community as they have been implicated in neuronal processes and dysfunction (examples in Table 1) [22]. Evidence regarding some of the metabolites from Table 1 and their involvement in the gut-brain communication will be presented as follows:

- bacteria can metabolise the amino-acid tryptophan and produce precursors of serotonin, a neurotransmitter involved in cognition regulation and mood [23][24]. Tryptophan was present in higher concentrations in plasma of germ-free mice, in contrast to mice with a gut microbiota, which was further supported by reduced metabolism of tryptophan along its dominant metabolic fate, the kynurenine pathway [25][26]. Tryptophan is also degraded into quinolinic acid, which activates the glutamate-gated ion-channel N-methyl-D-aspartate (NMDA) found in neurons and, when the activation is exacerbated, leads to loss of neuronal function and cell damage/death [24].
- neurotransmitter glutamate have been shown to enhance cognitive abilities and memory [22] and are present in much-elevated levels in conventionally colonised mice compared to germ-free mice [26].

- short-chain-fatty-acids (SCFA) are fatty acids produced by gut bacteria during fermentation of complex carbohydrates, of which propionate and acetate are the most abundant metabolites [27][28]. SCFAs are involved in the maturation of immune cells of the CNS and in its homeostasis [22].
- menaquinone/vitamin K2 has shown neuroprotective properties, in a model of neurodegenerative disease, by inhibiting aggregation of neurotoxic proteins inside the neurons [22].

Table 1. Neuroactive metabolites reported as being produced/consumed by bacteria.

Type	Metabolites	Reported observations	CarveMe ID
Neurotransmitters precursors	Tryptophan [29, 30]	Precursor of serotonin (5-HT/5-hidroxitriptamina) and of tryptamine	trp__L_p/e/c
	L-3,4-dihydrophenylalanine (L-DOPA) [26]	Precursor of dopamine	34dhphe_p/e/c
	Tyramine [26]	Precursor of dopamine	tym_p/e
Neurotransmitters	Tryptamine		trypta_c
	Gamma-aminobutyric acid (GABA) [26, 29–31]	Inhibitory neurotransmitter; has been observed in MDD (major depressive disorder) patients [31]	gg4abut_c
	Dopamine [29] [30] [26]	Motivation, reward, hedonistic regulation [30]	dopa_p/e/c
	Glutamate [30] [26]		glu__L_p/e/c
	Nitric oxide [30]		no_p/e/c
	Taurine (amino sulfonic acid) [26]		taur_p/e/c
short-chain-fatty-acids (SCFAs)	Butyrate [29][30]	Regulation of brain-derived neurotrophic factors; Regulation of neuroinflammatory processes [32]; depleted in MDD (major depressive disorder) patients [33]	but_p/e/c
	Propionate [29] [30]		ppa_p/e/c
	Acetate [29] [30]		ac_p/e/c
	Isovaleric acid [30]		3mb_p/e/c
			3mba_e/c
Secondary messenger	Inositol [30]	Regulation of neuronal and glial activity [34]	inost_p/e/c
Vitamins	Menaquinone (vitamin K2) [30]		mqn6/7/8_c
Excitotoxic (damage/death of nerve cells) Neuroinflammatory	Quinolinic acid [30] [31]	Metabolic pathway to quinolinic acid is activated in MDD (major depressive disorder) patients; quinolinic acid impairs neurons inducing depressive symptoms [31]	quln_c
Anti-inflammatory	S-Adenosylmethionine (SAM) [30]		amet_c

2.2 Impact of Drugs on Human Gut Bacteria

Medication has been emerging as one of the host extrinsic factors that has a strong impact in the variance of the human gut microbiota composition [2][35]. Studies have shown that antibiotics, that are supposed to inhibit pathogens, have a side effect on the gut microbiota [2][36]. Furthermore, non-antibiotics have also been linked to changes in microbiome composition [2][35].

Drugs are supposed to have a certain therapeutic outcome while avoiding secondary effects, which are undesirable effects at normal dosages. The study of the genetic influence on the relationship between drug dose and effect is gaining increasing interest since drug receptors, drug metabolising enzymes and drug transporters are the products of genes that exhibit polymorphisms. This influence of genetics introduces variability among response of individuals and may be the cause of adverse reactions to

treatment [37]. The estimated 500-1000 bacterial species that comprise the gut microbiome, with each bacterium having a genome harbouring thousands of genes, together with the fact that each person has a different set of bacteria, adds further variability between individuals and subsequent response to drugs [38].

Cellular metabolism is a reliable system for predictive analysis of drugs side effects since it's a genome-wide network that can be turned into a predictive model, designated as genome-scale metabolic model (GSMM). Using these models together with drug-gene interaction information allows the prediction of linkages between gene and phenotype and the identification of the molecular mechanisms of the side effects of the drugs [39].

Several databases provide information between drugs and their targets, such as Pharos [40], DrugBank [41], BioGRID [42] and Therapeutic Target Database [43], but their data sources can't be easily traced back and neither do they provide links to genome databases, unlike STITCH (search tool for interacting chemicals) which integrates all of this information (Fig. 3) [44][45].

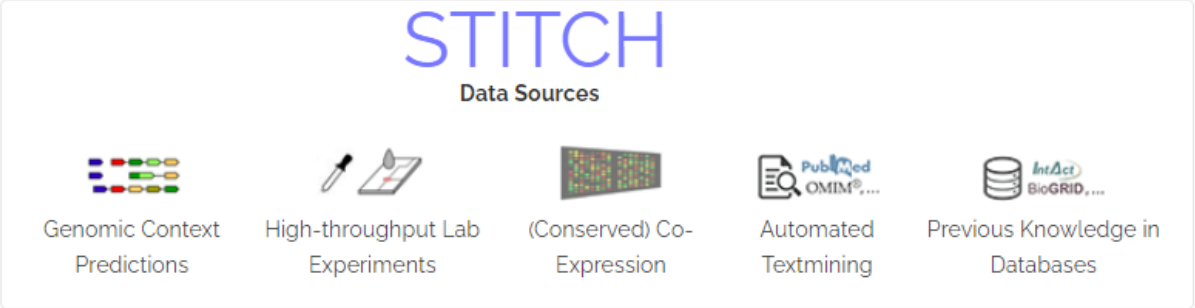


Fig. 3. Information sources of STITCH database (adapted from [46]).

STITCH is a database that collects experimental data about protein-chemical interactions and that predicts protein-chemical associations *de novo* using computational tools, by knowledge transfer between organisms and from other databases, as shown in Fig. 3 [44][47]. These predicted interactions are based on the identification of pairs of genes that are thought to be functionally associated, through genome comparisons [47]. Each predicted interaction is assigned a confidence score, which is representative of how likely the interaction is true. [46]. Experimental interactions are given a uniform confidence score per database. The scores of interactions that are transferred via homology correspond to the probability of finding the associated proteins within the same metabolic pathway. Individual scores are then computed into a combined score. This score ranges from 0 to 1, being that 1 is the highest possible confidence [46][47].

2.3 *In Silico* Simulation of Drug Interaction with Bacteria

2.3.1 Genome-Scale Metabolic Modelling

Profiling microorganisms present in the gut (through metagenomics sequencing and taxonomic assignment), identifying and or quantifying metabolites (metabolomics) and which microbes are responsible their production, understanding metabolites' influence on their host's metabolism (on health and disease states, for example) and evaluating microbe-microbe interactions, is all knowledge that is extremely important for the development of strategies to improve human health. Strategies involving, for instance, the administration of beneficial bacterial (probiotics) or of substrates enhancing their growth (pre-biotics), with the aim of modifying microbiome composition and, consequently, changing its impact on the host [5][7][6].

Approaches namely 16rRNA amplicon sequencing, shotgun metagenomics sequencing, taxonomic assignment and metabolic profiling, mainly of fecal samples (most studied kind of samples, as reviewed in [7]), has generated enormous amounts of data, such as, genomic information, microbial composition and quantification and/or identification of metabolites present in a biological system, respectively [5, 7, 48].

Statistical analysis of this data allows to figure out differences in gut microbial and metabolites' composition under different conditions [49][4]. However, they alone don't allow identification of the contribute of each species to the host's metabolism, nor provide understanding of interactions between microbes [48][50].

Thus, development of descriptive mathematical models is necessary to integrate all of this diverse omics data and enable the computational exploration of complexities of the gut microbiome, consequently, contributing to the knowledge of pathophysiology of diseases and allowing the proposal of personalised interventions [4, 6, 48].

A genome-scale metabolic model (GSMM) of an organism is a mathematical representation of its metabolism's genome-scale network reconstruction (GENRE), which in turn is based on all of the information about the organism, such as genome annotation and biochemical characterisation, assembled from sources like high-throughput data, databases and research papers [51][4].

A GENRE is made up of all organism's metabolic reactions that are linked to its genome. Briefly, the metabolic reconstruction (represented in figure 4) starts with acquiring genome annotations for the organism. Then, biochemical databases (like BiGG, KEGG and MetaCyc) are searched to identify reactions

for the corresponding catalysing enzymes encoded by the genome, the so called gene-protein-reaction (GPR) relationship [4, 51].

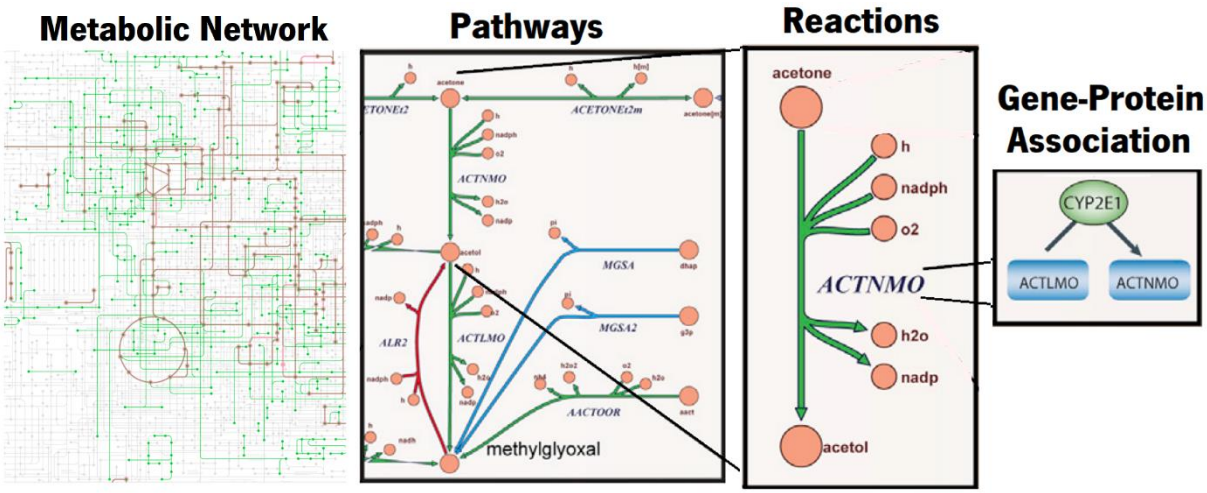


Fig. 4. Genome-scale network reconstruction (adapted from [51]).

The GENRE conversion into a mathematical format initially requires the tabulation of the stoichiometry of each metabolite participating in a reaction, designated as stoichiometric matrix (as shown in figure 5a). Unlike models that necessitate kinetic parameters to be measured, GSMMs rely on the imposition of constraints on the flow patterns of the metabolites to be able to compute a feasible space of flux distributions (or flux map) using constraint-based modelling methods (CBM) such as COBRA (constraint-based reconstruction and analysis) (figure 5b). These constraints, which are equations when referring to balances and are inequalities when imposing bounds, can be of several types [6, 8, 51] :

- Enforcement of mass balance (or mass-balance or flux balance) constraints: the amount of intracellular compounds produced must be equal to the amount of intracellular compounds consumed at the steady state, meaning that the production and consumption cancel out) [6, 8, 51].
- Capacity constraints (or reaction flux bounds): definition of maximum and minimum allowable fluxes through reactions; also define which metabolites can enter from the medium [6, 8, 51].
- Thermodynamic feasibility constraints: thermodynamically feasible directions imposed on reactions [6, 8, 51].
- Substrate uptake rates [6, 8, 51].
- Secretion rates [6, 8, 51].

The computed space of flux distributions represents multiples solutions that satisfy the governing constraints and that are associated with candidate physiological functions that the network produces.

Thus, the next step is defining the biological function that, mathematically, is represented by an objective function (for e.g., if the objective is to predict growth, the objective function is biomass production which is represented by a biomass reaction in the stoichiometric matrix). In order to calculate the reaction fluxes (optimal flux distribution) that minimise or maximise the objective function, it's frequently used a linear programming technique designated as flux balance analysis (FBA), that is based on assumptions of steady-state growth and mass balance. Nevertheless, it presents limitations, one of which is not being able to infer metabolite concentrations [4][51].

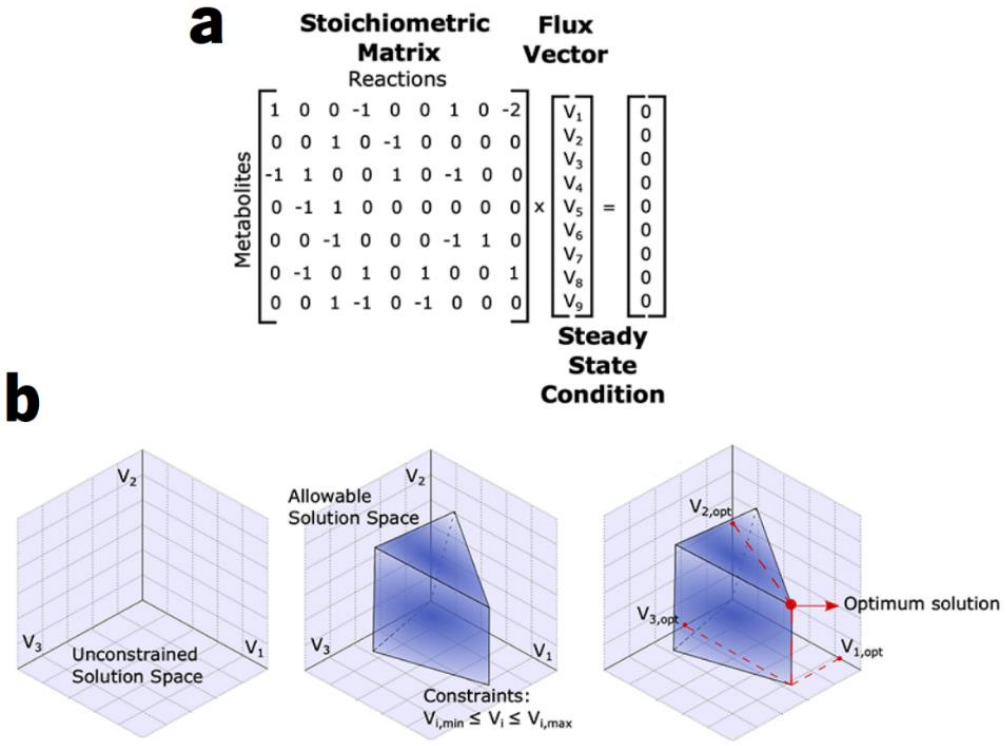


Fig. 5. Mathematical conversion of a genome-scale network reconstruction to a computational genome-scale metabolic model (adapted from [4]). **a,** Stoichiometric matrix. **b,** feasible space of flux distributions (or flux map).

Currently, there are metabolic models for many organisms. In order to provide the scientific community a resource for obtaining these models, repositories have been being created to collect and validate them and to standardise the reaction, the metabolite and the gene identifiers [4, 6]. One example is the Virtual Metabolic Human database [4][52]. However, if the organism doesn't have any GSMM associated or if it has but is not sufficiently descriptive of the organism, a new GSMM can be generated automatically or an existing GSMM can be either manually or automatically curated [4].

2.3.2 Refinement of Model Reconstruction

With the increasing ability to easily determine the complete genome of an organism, through whole genome sequencing, and with the growing number of available software tools to reconstruct GSMMs from genome annotation, the number of metabolic reconstructions is rapidly increasing [53][54][51]. However, problems such as incomplete or incorrect genome annotation (i.e., missing or wrong metabolic function attributed to a gene) and biochemical databases not being organism-specific (meaning that one enzyme might be linked to a certain activity that is not present in the organism), can lead to incorrect models that resulted in wrong predictions of phenotypes when compared with experimental observations. These problems make the refinement/curation of the reconstruction an important part of the process, to ensure high-quality predictions [53][51].

Therefore, after an automatic software tool has been used to create a draft reconstruction, the latter should go through steps of curation that can be performed multiple times and without an order until the final GSMM [53][51]. Some of the most important steps are the following:

- Investigation of the function of a gene through experiments or literature, since it can have a wrong annotation or not have one at all [53][51].
- Removal of reactions that are too generically described, e.g., DNA [53][51].
- Verification of substrates and/or cofactors of enzymes that are associated with multiple reactions in organism-unspecific biochemical databases. This step is necessary because substrate/cofactor specificity of enzymes can be different between organisms and incorrect inclusion of substrates may lead to wrong predictions [53][51].
- Making sure the direction of a reaction is assigned correctly. When no information is available to determine the reaction directionality, the reaction is set by default as reversible. However, a high number of reversible reactions can lead to a free exchange of metabolites between compartments and impact the results. Some textbooks and literature provide rule of thumbs for the directionality of some reactions [53][51].
- Confirming correct compartment location of enzymes (i.e., if they are located in the cytoplasm, in the periplasm or extracellularly). Their incorrect assignment can lead to the addition, without evidence, of reactions that transport metabolites between intracellular compartments and, consequently, result in the misrepresentation of the model [53][51]. Tools such as PSORT [55][56] and Proteome Analyst [57][58] can be used to determine a compartment to an enzyme based on its sequence.
- Verification of the gene-protein-reaction (GPR) rule [53][51]. According to the catalytic mechanism of a metabolic reaction, the GPR rule associated with it can be [54][4, 51, 59]:

— an empty rule, meaning no gene is involved in its catalysis. This is valid since there are reactions that occur spontaneously or that only need small molecules to occur. Reactions with these rules are designated as spontaneous or as non-enzymatic reactions [54][4, 51, 59].

— a single gene rule, i.e., only one gene is required for the reaction catalysis. In this case, one single gene is responsible for a monomeric enzyme, i.e., an enzyme with a single subunit[54] [4, 51, 59].

— a multiple genes rule, when a reaction is catalysed by either an oligomeric enzyme (an enzyme consisting of multiple subunits, all of which necessary for the reaction catalysis) or by isoforms of an enzyme (a highly similar enzyme that can catalyse the same reaction). These rules use the operator AND (Fig. 6) to indicate that all the genes that encode the subunits of an oligomeric enzyme are necessary for the reaction to occur; the operator OR (Fig. 6) to specify the genes that encode the isoforms of the enzyme, which in this case only one of the enzymes is sufficient for the reaction catalysis; or both operators at the same time to describe isoforms and subunits involved [54][4, 51, 59].

Reactions that have associated either of the two previous rules are designated as gene associated reactions or enzymatic reactions [54][4, 51, 59]. Some GPRs retrieved from databases might be wrong due to differences between organisms when it comes to, for instance, subunits composing the enzymes or the reactions they catalyse. Wrong associations affect the results of gene deletion studies [54][4, 51, 59]. The tool GPRuler can reconstruct automatically the GPRs from the name of the organism or from its metabolic model [54][4, 51, 59].

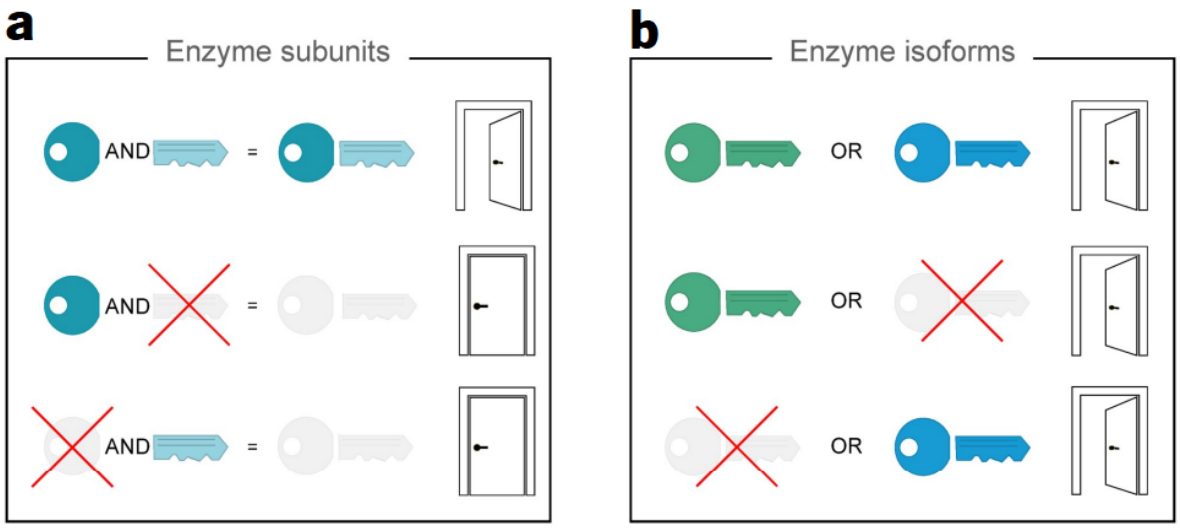


Fig. 6. Representation of the operators AND and OR used in the "multiple genes" GPR rule (taken from [54]). **a,** Scheme representing subunits necessary for an enzyme to catalyse a reaction. **b,** Representation of enzyme isoforms and how they can be used alternatively for reaction catalysis.

- Analysing the biomass reaction. This reaction describes all the intracellular compounds (and their individual contributions) that are essential for the cell to grow. If the biomass composition for the organism is not available, one should determine it experimentally or, at least, estimate it from the genome. Using the genome of other organisms, such as *Escherichia coli*, for this estimation will probably lead to incorrect results since the number of rRNA operons differs between organisms. Furthermore, when a precursor of biomass is not included, its synthesis reaction will not be essential for growth, making the genes that catalyse it not essential as well, and vice-versa. This will consequently lead to wrong results of gene deletion studies [53][4, 51, 59].
- Adding further constraints such as enzyme capacity will result in a smaller set of steady-states flux solutions [53][4, 51, 59].
- Evaluating if the network is able to synthesize biomass precursors or to simulate a physiological function, by testing the capability of the model to carry flux through the reactions. This allows the identification of network gaps which correspond to missing metabolic reactions and functions. Network gaps cause blocked reactions (reactions that cannot carry flux when boundaries are open) and dead-end metabolites (metabolites that are synthetised but not consumed). If the network gaps don't allow the model to be functional, it is necessary to add reactions whilst ensuring network connectivity, the so called gap-filling process [53][4, 51, 59]. The reactions to be added can be selected by multiple ways:
  - analysing through biochemical maps the enzymes that are able to produce the dead-end metabolites, information with which one can search for reactions that consume them [53][4, 51, 59].
  - adding temporary demand and/or sink reactions that add metabolites to the network, in order to make blocked reactions carry flux and, thus, test metabolic functions. This allows the identification of the gap that makes the model not functional and subsequent filling of the gap [53][4, 51, 59].
- Using experimental observations to correct and improve network content [53][4, 51, 59].
- Evaluation of growth rate. A slow growth rate might mean that at least one of the medium components is limiting growth, which can be verified by increasing the uptake rate of each of the components and checking if the growth rate increases. A fast growth rate might be indicative of a non-optimal biomass reaction, reactions that shouldn't be in the model or of missing and/or incorrect constraints [53][4, 51, 59].

Depending on the automation level of the reconstruction tool used, these different steps of curation can be performed manually or automatically and on different stages of the reconstruction process [60].

2.3.3 Model Reconstruction Approaches

In general, genome-scale metabolic reconstruction tools follow one of two main approaches: bottom-up and top-down, as represented in Fig. 7 [60].

The bottom-up approach starts the process with the genome of the organism and generates a draft reconstruction with the respective reactions retrieved from biochemical databases, which will then go through curation steps [60].

The top-down approach distinguishes from the latter since, instead of starting with the genome of the organism, it begins with reconstructing a draft model with all the reactions and metabolites present in a database, that belong to multiple organisms of the same domain. Subsequently, the draft model is turned into an organism-specific model, using genome information to remove the reactions and metabolites that are unlikely to be part of the organism [60].

Several reconstruction tools are available [61], but this study will be focused on the description of ModelSEED [62] and CarveMe [60] which are representing the bottom-up and the top-down approaches, respectively.

ModelSEED is a web tool that automates most of the curation steps mentioned previously in chapter 2.3.2. Since it is a bottom-up reconstruction approach, the first step it performs is annotating a genome sequence using the RAST (Rapid Annotation using Subsystem Technology) fully-automated service [63][62].

After this, the model is constructed with reactions selected from SEED database [62, 64]. This database puts together every information from KEGG database [65] and from 13 available GSMMs [62]. The model generated consists of GPR rules that are constructed from mappings, in SEED database, of gene function to reactions. Furthermore, a biomass reaction is also added based on a template that exists in SEED and that was assembled by curating biomass reactions of 19 existent GSMMs. This reaction includes 39 substrates that are included in every model reconstructed by this tool and 44 substrates that can be additionally added if they satisfy the tool criteria by genomic evidence (e.g., cell wall type being gram-positive or gram-negative) [62]. The relative abundances of the substrates in the biomass reaction are based on measured values in *E. coli* for gram-negative organisms and *Bacillus subtilis* from gram-positive ones.

After this phase, the model has network gaps that don't allow its functionality and goes through a mixed integer linear optimization problem (MILP) where, for example, some reactions are favoured over others (e.g., intracellular biosynthesis pathways are favoured over transport reactions) and penalties are given to reactions that go on an unfavourable thermodynamically way. If the growth medium specific for the

organism is experimentally determined, the optimization also identifies the reactions that must be added to enable growth on the respective medium; if no growth medium information is available, all the metabolites are allowed to be imported from the medium [62]. ModelSEED also performs optimization of the model in order to fit experimental data (optimization steps such as searching inconsistencies between model annotations and gene essentiality data and using the using the GrowMatch [66] algorithm to correct errors that lead to discrepancies between growth *in silico* and *in vivo*.

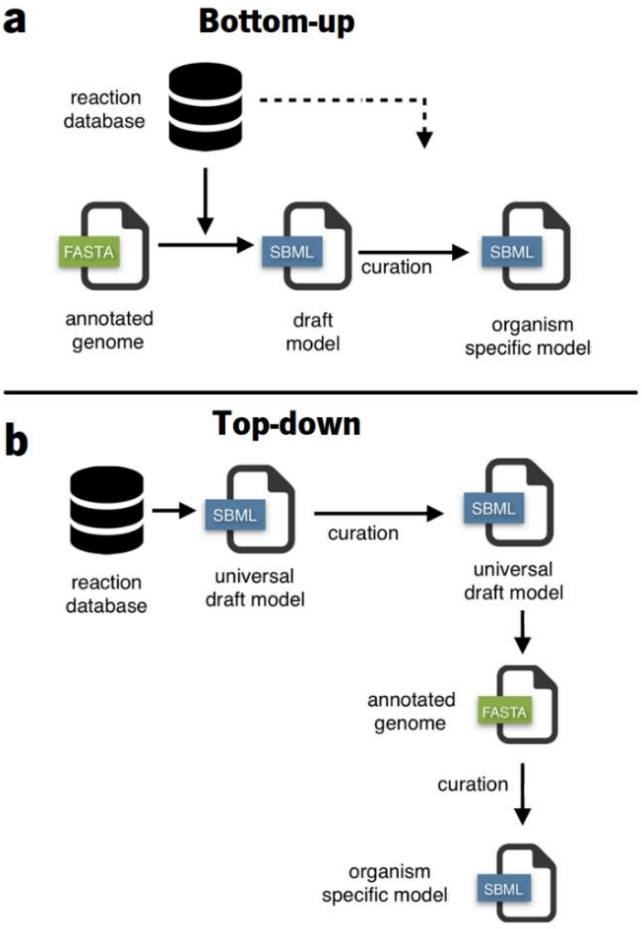


Fig. 7. Metabolic model reconstruction approaches (adapted from [60]). **a,** Represents the bottom-up approach. **b,** Represents the top-down approach.

Regarding the top-down reconstruction approach, CarveMe is a Python-based tool that begins the reconstruction by making a draft model containing all reactions and metabolites that exist in the BiGG Models database (integrates more than 70 published genome-scale metabolic networks) [66] and that are specific for the bacteria domain, thus being designated as the universal draft model of bacterial metabolism [60]. This model then goes through a manual curation process, during which, for instance, an universal biomass equation, adapted from *E.coli*, is added and reversibility of reactions is constrained to make thermodynamically feasible phenotypes. From this universal model, CarveMe generates two

additional models that can be used in the place of the former and that are made specific for gram-positive and gram-negative bacteria by the addition of appropriate cell wall components to the biomass reaction [60].

BiGG database also presents the GPR associations from the original GSMMs, associations which CarveMe uses to makes its own database with the genes and protein sequences. This database is going to be used in the next step, in which the user inputs the genome of the organism and CarveMe aligns gene/protein sequences to the ones in the database and attributes the alignment a score. The score of every isoform of a protein that catalyses a reaction will be summed up, subsequently normalised to a median value of 1 and attributed to the respective reaction. Non-enzymatic reactions (reactions don't need a protein to be catalysed) are given a score of zero and enzymatic reactions that don't have any gene/protein mapped to it are given a -1 score. These scores are then used for the MILP optimization problem that maximizes the number of reactions with higher scores and minimizes the presence of the lower score ones, finally resulting in a model comprising reactions that are more likely to be present in the specific organism. All of this process is schematised in Fig. 8.

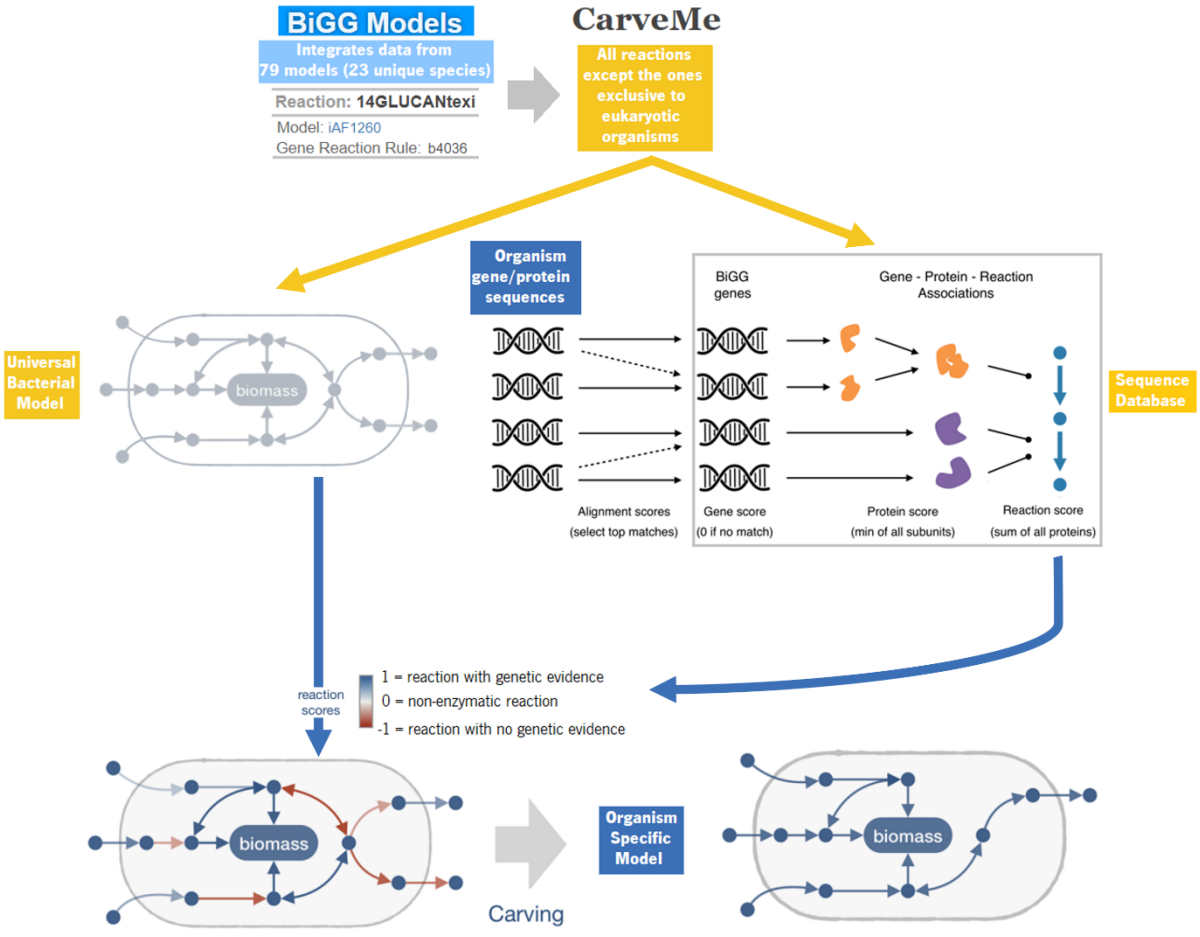


Fig. 8. Scheme about the CarveMe top-down reconstruction process (adapted from [60]).

In order to be able to evaluate if running the MILP problem would lead to different reactions being added to the model and, consequently, give rise to different predictions of a bacterium's capabilities, CarveMe has the option of performing ensemble modelling. This step takes enzymatic reactions without a gene associated to it (have a score of -1) and attributes to them random weighting factors. This means that only these reactions will have a different score every time the MILP problem is solved [25]. This will generate different models, all of which will constitute the ensemble model. Studying these alternative models is important to understand if the reactions that differ are present in most models, hence being able to assume that the predictions of any of the models generated are going to be representative, or not. The more complete a genome is (i.e., the more the DNA sequence is deciphered) and the more genes/proteins mapped to the CarveMe database, the fewer the enzymatic reactions with -1 score. Consequently, the random factor decreases, which means that the models that will be generated each time the optimization problem is solved, should have almost the same reactions present.

CarveMe also allows the user to provide experimental data such as a list of growth medium substrates and perform the necessary curating steps to make the model functional in this medium, such as gap-filling.

2.3.4 Using Metabolic Models to Predict Biological Capabilities

GSMMs can be used to compute perturbations at genetic and environmental (nutritional input) levels, allowing the simulation of many different experimental conditions *in silico* quickly and, consequently, the prediction and analysis of the consequences of these changes [51].

GSMMs have been used for qualitative predictions such as gene essentiality. Since each reaction is linked to the corresponding protein, hence, to the encoding gene(s), removing a reaction from the GSMM and computing growth allows to know if the gene(s) is essential depending if growth can be computed or not without it (figure 9). As genome editing techniques develop, the results of gene knockout studies will be necessary for engineering genomes and, thus, result in a desired phenotype [51].

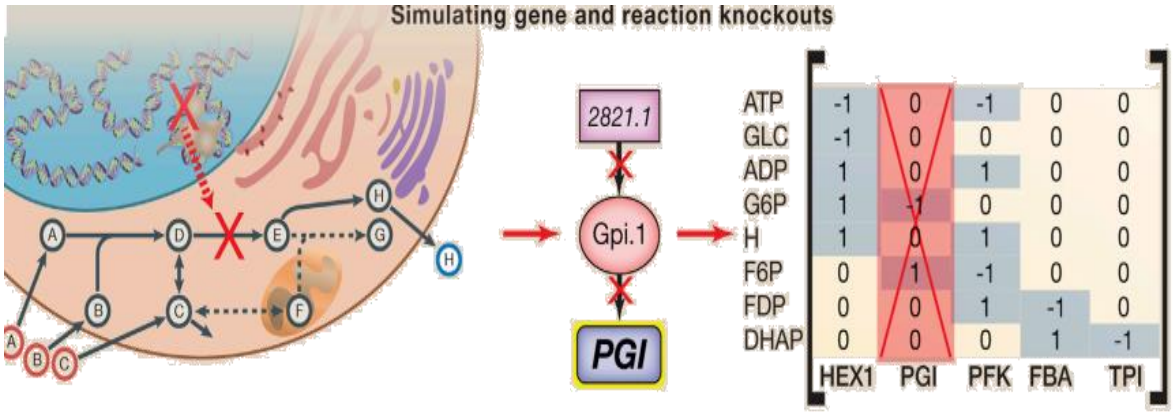


Fig. 9. Gene essentiality prediction (adapted from [51]).

GSMMs have also been used to predict quantitively phenotypic organism functions like (as represented in figure 10) [51]:

- Nutrient utilisation [51].
- Central carbon metabolism fluxes [51].
- Strain abundance and nutrient exchanges [51].

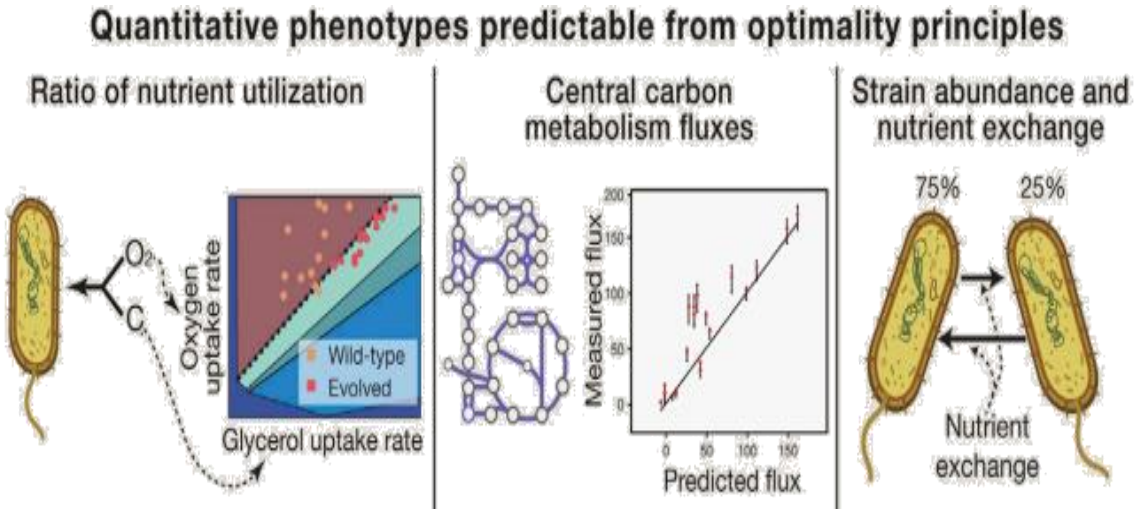


Fig. 10. Quantitative predictions (adapted from [51]).

To address the possible existence of multiple flux solutions that optimize the objective function, flux variability analysis (FVA) can be used to predict the flux range of reactions [51].

Incorrect model predictions are related with incomplete knowledge about an organism that can, in turn, be discovered using algorithms that implement automated approaches designated as “gap-filling”. These have been used to query databases to discover potential metabolic reactions that lead to new hypothesis that can be experimentally tested [51].

Cell-type and condition-specific (e.g., healthy vs disease) GSMMs, resultant of multi-omic data integration through the conversion of experimental data into model constraints, allow evaluation of phenotypic capabilities and consequent identification of molecular differences between different cells or environments [51].

2.3.5 Overview of Available Resources / Tools

Table 2. Overview of available tools.

Analyses	Tools/ Databases	Description
Genome-scale metabolic reconstructions/models' acquisition	AGORA [50]	
	KBase [9]	
	Virtual Metabolic Human database [13]	
Genome-scale metabolic models' refinement	ModelSEED [9]	
	rBioNet [50] [67]	
	metaGEM [68]	Reconstruction of sample-specific metabolic models from metagenomic data
In silico model simulation	COBRA [69]	
	FBA [9] [13]	Linear programming technique to determine the steady-state reaction flux distribution Used for growth prediction/ substrate utilization profiles; maximum production flux
	Distributed FBA [67]	Contribution of each strain to overall production
	FVA [13]	Flux range of reactions Used for production capabilities
	Gurobi [9] [13]	Linear programming solver
	CPLEX [67]	Linear programming solver
	Fast-SL [13]	Synthetic lethality analysis Reaction essentialities
Community modelling	Microbiome Modelling Toolbox [67] (COBRA extension)	Join individual genome-scale metabolic models
	createPanModels.m [70] (function of Microbiome Modelling Toolbox)	Join individual genome-scale metabolic models of the same species (strain-specific)
	MMinte [49]	Pairwise interactions
	MICOM [49]	Community interactions
	mgPipe [70] (function of COBRA)	Integration of microbial abundance
	MetGEMs toolbox [71]	Metabolic functional analysis of microbial communities To assign enzyme functions To associate abundance of enzymes with related diseases (for example)
Large-scale metagenomic data	Human Microbiome Project Consortium [67]	
	MetaHIT [7]	
Statistical tests	SelectKBest [69]	
	Jaccard coefficient [9]	Metabolic dissimilarity test
	Mann-Whitney U test [9]	
	Wilcoxon rank-sum test [9]	
	Spearman's correlation [5]	Connects metabolomics data (metabolites level) with microbiome taxonomic profiles
Databases of reactions	UniProt [69]	Reactions associated with genes
	BLASTP [13]	Reactions associated with enzymes
	KEGG [9]	
	BIGG [9]	
	MetaCyc [6]	

3 Methods

As explained in the Introduction section, according to the *in vitro* results of a 2018 research [3], drugs that are not supposed to target gut commensal bacteria, such as human-targeted drugs, changed the gut microbiome composition. In particular, drugs that targeted the nervous system were overrepresented [3].

To understand the undesirable effects of drugs on the human gut microbiome, it is necessary to study how they interact with bacteria and their impact on bacteria's growth and metabolism.

The study was based on a specific bacterium, the gram-negative *Bacteroides thetaiotaomicron* (strain VPI-5482 and taxonomy identification 226186 [72]), due to its extensive distribution among humans [73] and to being the most commonly isolated obligate anaerobe [74] (thus, having a very complete genome).

The drugs' effect on *B. thetaiotaomicron*'s growth and metabolism was determined by its *in silico* genome-scale metabolic model.

The following sections describe which drugs were studied in this project and how information about drugs' interaction with genes was obtained and integrated for *in silico* simulation of their effects.

3.1 Drug Selection

The first analysis of drug effects on bacteria was focused on the drugs that target the nervous system, screened in the studied paper [3].

As reported Maier et al. [3] and as presented in its "Supplementary Table 1" (shown in Table 3), all the drugs screened in the *in vitro* experiment were categorized into therapeutic classes according to the Anatomical Therapeutic Chemical (ATC) classification presented in Table 4 [75]. Table 3 illustrates two examples of the information extracted from "Supplementary Table 1", where each row corresponds to a drug identified by a unique "Prestwick_ID" and by one or multiple "ATC codes" representing the therapeutic classes they belong to.

Table 3. Partial content of "Supplementary Table 1" of Maier *et al* 2018 [3]. Two examples where chosen for illustration.

prestwick_ID	chemical name	STITCH4 id	ATC codes
Prestw-948	Timolol maleate salt	CID100005478	C07AA06 S01ED01
Prestw-978	Memantine Hydrochloride	CID100004054	N06DX01

Table 4. Anatomical Therapeutic Chemical (ATC) classification. **a**, All therapeutic classes. **b**, Specific classes of nervous system drugs. (Adapted from [75]).

a

Anatomical Therapeutic Chemical (ATC) Classification

A ALIMENTARY TRACT AND METABOLISM

B BLOOD AND BLOOD FORMING ORGANS

C CARDIOVASCULAR SYSTEM

D DERMATOLOGICALS

G GENITO URINARY SYSTEM AND SEX HORMONES

H SYSTEMIC HORMONAL PREPARATIONS, EXCL. SEX HORMONES AND INSULINS

J ANTIINFECTIVES FOR SYSTEMIC USE

L ANTINEOPLASTIC AND IMMUNOMODULATING AGENTS

M MUSCULO-SKELETAL SYSTEM

N NERVOUS SYSTEM

P ANTIPARASITIC PRODUCTS, INSECTICIDES AND REPELLENTS

R RESPIRATORY SYSTEM

S SENSORY ORGANS

V VARIOUS

b

- N NERVOUS SYSTEM
  - N01 ANESTHETICS
  - N02 ANALGESICS
  - N03 ANTIEPILEPTICS
  - N04 ANTI-PARKINSON DRUGS
  - N05 PSYCHOLEPTICS
    - N05A ANTIPSYCHOTICS
    - N05B ANXIOLYTICS
    - N05C HYPNOTICS AND SEDATIVES
  - N06 PSYCHOANALEPTICS
    - N06A ANTIDEPRESSANTS
    - N06B PSYCHOSTIMULANTS, AGENTS USED FOR ADHD AND NOOTROPICS
    - N06C PSYCHOLEPTICS AND PSYCHOANALEPTICS IN COMBINATION
    - N06D ANTI-DEMENTIA DRUGS
  - N07 OTHER NERVOUS SYSTEM DRUGS
    - N07A PARASYMPATHOMIMETICS
    - N07B DRUGS USED IN ADDICTIVE DISORDERS
    - N07C ANTIVERTIGO PREPARATIONS
    - N07X OTHER NERVOUS SYSTEM DRUGS

The drugs that act on the nervous system belong to the ATC class “N”, as shown in Table 4. Therefore, from the “Supplementary Table 1” of the paper [3], all these drugs were gathered by selecting the rows that have at least one of the values in the column “ATC codes” starting with the characters “N0” (as it’s represented in the second row of the Table 3).

This table further presented a column named “STITCH4 id” (Table 3), that attributed an identifier to each drug, with which it is possible to retrieve information from the STITCH database [46] about how a drug interacts with a gene, as it will be explained in the following section.

Lastly, the second analysis of drug effects on bacteria was based on the drugs screened in the same paper [3], that covered all main ATC therapeutic classes and that are not supposed to affect gut commensals. From these, the ones that had information in STITCH were also studied.

3.2 Extracting Interaction Information between Drugs and Bacterial Proteins from the STITCH Database

From the STITCH database [46][44], a file containing known and predicted interactions between drugs and proteins/genes (proteins and genes will be used interchangeably) of *B. thetaiotaomicron* was downloaded (“226186.actions.v5.0.tsv.gz”) (Table 5).

Table 5. STITCH file "226186.actions.v5.0.tsv.gz" partial information. **a**, Table that highlights the 7 different values possible in column "mode". **b**, Table that emphasizes that one drug (e.g. "CIDm00131041") can have an effect on multiple genes.

a

item_id_a	item_id_b	mode	action	a_is_acting	score
226186.BT_1545	CIDm00131041	inhibition	inhibition	f	410
CIDm00131041	226186.BT_1545	inhibition	inhibition	t	410
226186.BT_0638	CIDm06914645	binding	NaN	f	689
CIDm06914645	226186.BT_0638	binding	NaN	f	689
226186.BT_1881	CIDm10400926	pred_bind	NaN	f	182
CIDm10400926	226186.BT_1881	pred_bind	NaN	f	182
226186.BT_3720	CIDm00006410	activation	activation	f	410
CIDm00006410	226186.BT_3720	activation	activation	t	410
226186.BT_1854	CIDs06419702	catalysis	NaN	t	900
CIDs06419702	226186.BT_1854	catalysis	NaN	f	900
226186.BT_0311	CIDm00000783	reaction	NaN	f	777
CIDm00000783	226186.BT_0311	reaction	NaN	t	777
226186.BT_3115	CIDs00063090	expression	NaN	f	152
CIDs00063090	226186.BT_3115	expression	NaN	t	152

b

item_id_a	item_id_b	mode	action	a_is_acting	score
CIDm00131041	226186.BT_1545	inhibition	inhibition	t	410
CIDm00131041	226186.BT_1016	inhibition	inhibition	t	491
CIDm00131041	226186.BT_0303	inhibition	inhibition	t	491
CIDm00131041	226186.BT_1413	inhibition	inhibition	t	410
CIDm00131041	226186.BT_4338	inhibition	inhibition	t	410
CIDm00131041	226186.BT_1234	inhibition	inhibition	t	410

Each row in Table 5a corresponds to an interaction between the values in columns "item_id_a" and "item_id_b" that can either be:

- the gene (e.g. "226186.BT_1545"), in which the characters until the dot represent the NCBI taxonomic ID of the organism ("226186"), followed by the unique gene identifier, the NCBI locus tag ("BT_1545"), that always starts with the organism's initials ("BT").
- the drug ID (e.g. "CIDm00131041"), that is comprised of a prefix "CIDm" or "CIDs" (which corresponds to the "flat" compound, that merges stereo-isomers, or to the stereo-specific compound, respectively) and a suffix that corresponds to the PubChem [76] compound ID [46, 77]; the drug ID will also be henceforth designated as STITCH ID.

Therefore, each row represents a drug-protein interaction. Furthermore, each interaction with a particular drug is repeated in the consecutive row, with "item_id_a" and "item_id_b" swapped. This is to distinguish cases where one of the elements is having an action over the other, but the opposite does not occur (e.g., gene "226186.BT_1545" does not have any effect on the drug "CIDm00131041", but the drug has an effect on the gene, as one can see in the first two rows of Table 5a, where column "a_is_acting" has the respective values "f" [false] and "t" [true]).

Table 5a also shows that the column "mode" can have 7 different values (inhibition, binding, predicted to bind [pred_bind], activation, catalysis, reaction and expression [phenotypic effects or predicted to have the same phenotype] [46]), while the column "action" can only be either inhibition or activation. To study the effect of a drug on a bacterium's growth, the rows containing "inhibition" on both columns "mode"

and “action” were selected, since only this information can be simulated using GSMMs. Subsequently, solely the rows that correspond to a drug inhibiting a gene were kept (i.e., rows that have the drug ID in column “item_id_a” and “t” in column “a_is_acting”).

The “score” column in Table 4 is used to verify if the results have a meaningful interpretation or not, since it is a measure of confidence from STITCH on how likely an interaction is true. Scores rank from 0 to 1, being that 1 is the highest confidence score. In the file, this value is multiplied by 1000.

Depicted in Table 5b is the fact that one drug can interact with multiple genes. All the interactions with a specific drug will be used at the same time to manipulate the *in silico* model and analyse the effect of the drug on the bacterium (as described in the below section 3.5).

3.3 Corresponding Drug-Protein Interaction Information to Selected Drugs

The STITCH IDs of the drugs screened in the paper [3], present in the column “STITCH4 id” of Table 3, were obtained, by the authors of the paper, through a chemical annotation tool called CART [78]. Since a chemical name can correspond to different PubChem compound IDs (Table 6, e.g. “piroxicam” corresponds to IDs “54684470” and “23690938”), CART uses text matching to map the drug’s synonyms to the PubChem ID of the “flat” compound, the suffix of “CIDm” (using the STITCH file “chemical.aliases.v5.0.tsv”).

Table 6. STITCH file “chemical.aliases.v5.0.tsv” partial content.

Flat_chemical	Stereo_chemical	Alias	Source	
CIDm23690938	CIDs23690938	Piroxicam [USAN:BAN:INN:JAN]	845	LeadScope
CIDm23690938	CIDs54684470	Piroxicam-d3	1052	1096

This results in drug IDs with the suffix “CID1” in the paper, because in STITCH files the suffixes “CIDm”/”CID1” (and “CIDs”/CID0”) are interchangeable. However, as the STITCH drug IDs from the paper were obtained without regarding the stereoisomer, the isomers were combined to increase the coverage of information. This means that, in the present work, the suffixes were ignored and the drugs were solely identified by their PubChem ID, not to exclude any information.

3.4 Automatic Reconstruction of Genome-Scale Metabolic Models of *B. thetaiotaomicron*

The drugs’ effect on bacterium’s growth and metabolism was determined for the *in silico B. thetaiotaomicron*’s genome-scale metabolic model.

As presented previously in the state-of-the-art, there are multiple alternative ways of creating a metabolic model. We chose the tool CarveMe in this work since it automates many curation steps and builds a model using only the genome, which is very useful for species that cannot grow in well-defined media [60]. Besides, this tool reconstructs models with BiGG identifiers, allowing the integration of drug-protein information from STITCH, which is not the case of reconstructed models already available in literature such as AGORA models [52], for example. The models generated were analysed using the Python package "COBRApy" [79].

The genome of *B. thetaiotaomicron* was provided to CarveMe as a protein FASTA file (a text-based formatted file that represents amino acid sequences), named as "GCF_000011065.1_ASM1106v1_protein.faa" (downloaded from the National Center for Biotechnology Information [NCBI], using "RefSeq" as the source database [72]).

After this step, CarveMe provides the user various alternatives, some of which were studied in this project under the different analyses performed, to assess their differences and propose the most suitable one. These alternatives are regarding:

- the template used for the "carving" process (i.e. selecting reactions from a model with all bacteria's reactions to generate an organism specific model):
  - universal; models henceforth designated as universal or simply U
  - gram-negative; models identified as gram
- the growth medium:
  - not initialised with any particular medium composition; models henceforward mentioned as NoMedium or complete medium
  - minimal anaerobic M9 medium; models named as M9[-O_2] or only M9
  - lysogeny anaerobic broth; models from now on designated as LB[-O_2] or LB
- the gap-filling process:
  - no gap-filling; models labelled as NoGap
  - during model reconstruction; models identified as gapDuring
  - after model reconstruction; models nominated as gapAfter

This resulted in a total of 14 models, as displayed in Table 7:

Table 7. Resulting models (separated by growth medium).

Models		
U_NoMedium_NoGap	U_LB_gapAfter	U_M9_gapAfter
gram_NoMedium_NoGap	U_LB_gapDuring	U_M9_gapDuring
	U_LB_noGap	U_M9_noGap
	gram_LB_gapAfter	gram_M9_gapAfter
	gram_LB_gapDuring	gram_M9_gapDuring
	gram_LB_noGap	gram_M9_noGap

3.4.1 Exploration of Alternative Models

As mentioned in section 2.3.2, enzymatic reactions that do not have any gene mapped to it have a score of -1 and are given random weighting factors. This means that only these reactions will have a different score every time the MILP problem is solved [60]. Thus, there are alternative solutions that might lead to different reactions being present and, subsequently, give rise to different predictions of a bacterium's capabilities [60].

CarveMe tool allows the generation of an ensemble model, which is an aggregation of n models reconstructed. An ensemble of 100 models was generated using the options of gram-negative template, LB medium and gap-filling after the reconstruction process. The ensemble model was then analysed with the Python library for metabolic model simulation, ReFramed [80, 81].

3.5 *In Silico* Simulation of Drug Effects on Gut Bacteria

To simulate the effect of a drug on the *in silico* model of *B. thetaiotaomicron*, it was necessary to integrate the information gathered from STITCH into the model. This was done using COBRApy tool, as it is described in the following sections.

3.5.1 Performing *in silico* Gene Deletion to Study Growth Inhibiting Drug Effects on *B. thetaiotaomicron*

In order to simulate, *in silico,* the effect of a drug on the growth of a bacterium, the COBRApy function "knock_out_model_genes" was used (Fig. 11). This function takes up a list of genes that are inactivated by the drug and gets the reactions associated to those genes through the GPR rule. If a gene is essential for a reaction to occur, the reaction bounds are set to zero to make the reaction flux null.

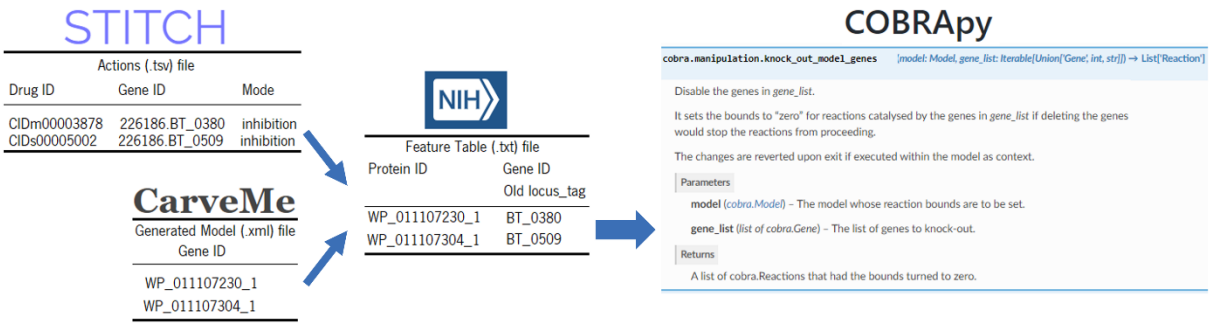


Fig. 11. Scheme that represents the *in silico* process of simulating a drug effect on a bacterium. Initially, all the known and predicted interactions between drugs and the bacterium genes are gathered from STITCH file "226186.actions.v5.0.tsv.gz". Since the gene IDs from STITCH do not match the ones of the generated model with CarveMe, it is necessary to obtain, from NCBI, the feature table file of the bacterium, with which one can match the genes IDs. Finally, for each drug, a list of genes that have interaction information with the former is reunited and then used as input to the "knock_out_model_genes" function from COBRApy.

Using the FBA tool in COBRApy, one can obtain the maximum flux through the objective reaction. In this model, the objective reaction represents the biomass function, which describes the rate of production of the metabolites that make up a bacterium. Hence, with the FBA tool we can predict the bacterium's growth rate.

By running FBA before and after shutting off the flux through the reactions affected by the genes' knock-out (with the previous function), the growth rate of the model without and with the constraints imposed by the drugs was obtained, respectively.

For the purpose of establishing a threshold to distinguish drugs that had effect on growth from those that did not, the percentage difference (also mentioned as "%GrowthChange" henceforth) between the obtained growth rates by running FBA was calculated according to the following equation:

$$\%\text{GrowthChange} = \text{abs}\left(\frac{\text{FBAafter} - \text{FBAbefore}}{\text{FBAbefore}} * 100\right)$$

After this step, a "%GrowthChange" equal or greater than 90% was decided based on histogram's interpretation (Fig. 23, 24, 29 in results section) and was associated with the drug having an inhibitory effect on the bacterium's growth.

These results were then compared with the impact of these drugs *in vitro*, in Maier et al 2018 experiment [3]. The "Supplementary Table 3" of this research paper contains p-values for the impact of each drug on the growth of each gut bacterium used in the experiment. Drugs that led to a significant reduction of growth had a p-value <= 0,01 [3].

Drugs were considered to have matching *in silico* and *in vitro* results when they led to a "%GrowthChange" of equal to or more than 90% and a p-value equal to or below 0,01, as shown in Fig. 12.

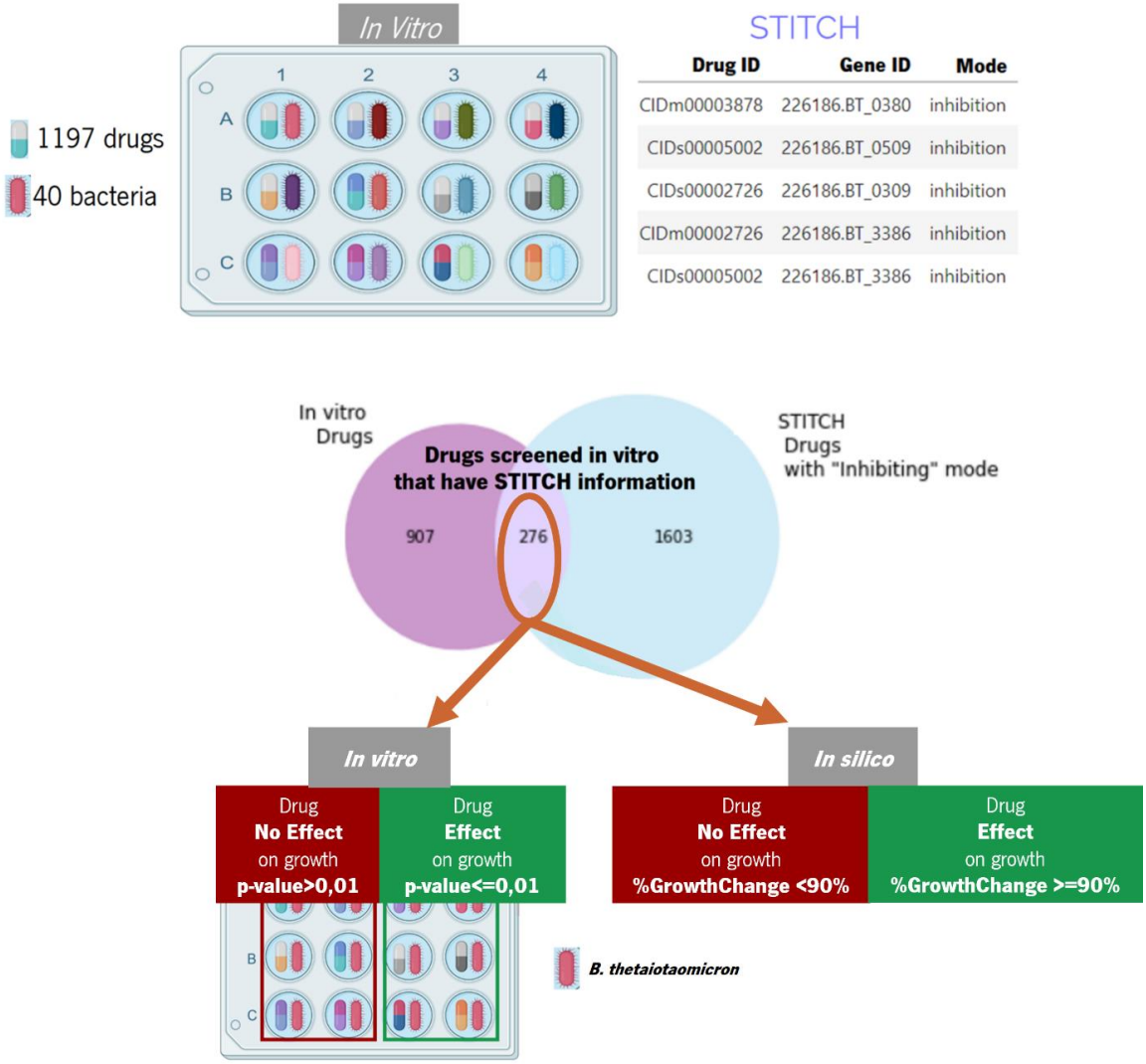


Fig. 12. Process of selecting drugs that have STITCH information, out of the drugs that inhibited *in vitro* the growth of a specific bacterium (in this case, of *B. thetaiotaomicron,* as an example*).* The first step is to find which drugs, from the ones screened *in vitro,* have information in the STITCH file. From these, the drugs that inhibited bacterium's growth *in vitro* are selected. The last step is to compare the effects of these *drugs in silico* with the effects *in vitro*.

3.5.2 Analysis of Drug Effects on the Metabolism of *B. thetaiotaomicron*

With a focus on studying the effect of the drugs on the bacterium's metabolism (i.e., on the reactions that are essential for the organism to be maintained), it was necessary to investigate the number of reactions that were affected.

The COBRApy function "flux_variability_analysis" (FVA) allows to obtain all the fluxes of a reaction that are possible at the optimal growth rate (denominated as fluxes range).

Reactions were considered to be affected when their flux ranges before and after the genes were "knocked-out", were not overlapping 100% (i.e., the minimum or the maximum values were outside the range of fluxes of the reaction that were possible before the constraints imposed on the model by the drug).

3.5.3 Analysis of Drug Effects on the Ability of *B. thetaiotaomicron* to Produce/Consume Neuroactive Metabolites

A list of neuroactive metabolites was gathered from literature, as presented in section 2.1.2. From this list and using COBRApy, it was possible to associate with each metabolite the reactions involved in their production and/or consumption (named as neuroactive reactions in this study).

From the reactions that were affected by a drug (process mentioned in section above 3.5.2), the neuroactive reactions were selected and the neuroactive metabolites were connected to them, to understand if the drug has an influence in the production/consumption of neuroactive metabolites by the bacterium.

3.5.4 Repeating Drug Effects Simulation Analyses using a Manually Curated Model (for Comparison Purposes)

The analyses described in the above-mentioned sections 3.5.1, 3.5.2 and 3.5.3, were repeated for a *B. thetaiotaomicron* GSMM assembled by Heinken et al. 2013 built through a "bottom-up" reconstruction process [82]. This means that a draft model was initially obtained with ModelSEED tool [83], followed by extensive manual curation [82]. The model, designated as "iAH991.xml", was downloaded from the "ThieleLab" website [84] (link "Collection of 11 human gut [...]").

Simulations were also performed in three different media (NoMedium, LB and M9). NoMedium was the default medium of the model and it corresponds to all exchange reactions having (-1000,1000) bounds. In order to simulate LB and M9 media and to be able to compare with the model reconstructed using CarveMe, the bounds of the exchange reactions had to be changed accordingly:

- the exchange reactions that were related with the medium composition had to have their lower bounds modified to -10.
- to remove exchange reactions from the medium, their lower bounds had to be changed to 0.

From this point on, models generated with CarveMe will be further designated by “carveme” models and the ones resulting from “bottom-up” reconstruction approaches, used for comparison analyses, were mentioned as “curated” models.

3.6 Data and Code Availability

https://github.com/inestm28/BioinformaticsDissertation2022

4 Results

This study was based on the *in vitro* results of the Maier *et al* 2018 research [3], in which drugs that are not supposed to target gut commensal bacteria, affected growth of single gut bacteria *in vitro*. In particular, 24% of the human-targeted drugs inhibited the growth of at least one strain *in vitro*, with drugs that act on the nervous system being overrepresented in this group [3].

In order to understand the undesirable secondary effects of drugs on gut bacteria, it is necessary to study how they interact with bacteria and their impact on its growth and metabolism.

This study focused on the impact of all drugs screened in Maier *et al* [3], on one bacterium, *B. thetaiotaomicron*. The drugs' effect on its growth and metabolism was determined with its *in silico* genome-scale metabolic model, through gene deletion studies, i.e., the genes that are predicted to be inhibited by a drug (information present in STITCH database) are "deleted" and the resulting growth is assessed.

First, the GSMM used for the simulation of drug effects was chosen out of 14 different GSMMs automatically reconstructed with CarveMe tool with the aim of assessing robustness of the model reconstruction and compare the models in terms of medium composition, biomass function and included reactions and metabolites [60], as described in the following 4.1 section. Second, we used the models to investigate the information contained in the models on neuroactive metabolism. Third, we simulated the effects of drugs on bacterial growth and metabolism with its *in silico* GSMM, through gene deletion studies, i.e., the genes that are predicted to be inhibited by a drug (information present in STITCH database) are "deleted" and the resulting growth is assessed. Finally, we compared the simulated results to the experimentally observed growth phenotypes. The results of the simulation are presented in sections 4.2 and 4.3.

4.1 Models Automatically Reconstructed based on Combinations of Parameters

In CarveMe, different options for the creation of GSMMs were available. These concerned (as mentioned in methods section 3.4) choices of template (universal or gram), of growth medium (complete medium/NoMedium, LB or M9) and of the gap-filling process (no gap-filling, gap-filling during or gap-filling after the reconstruction process). With the purpose of choosing the alternatives that better represented *B. thetaiotaomicron* and the *in vitro* conditions of the research paper on which this work was based on,

Maier *et al* 2018 [3], 14 different models, comprising combinations of these different options, were generated (models identified in Table 8).

Since it is not possible to find the exact composition of the medium used in the experiments of Maier *et al* 2018 [3] - the modified Gifu Anaerobic Medium (mGAM) - since it contains complex ingredients such as peptone, yeast extracts, and digested serum, which are not chemically defined [85] - it was necessary to reconstruct the models with the available medium options in CarveMe, to verify how they differed between each other and which one would resemble the most with mGAM.

4.1.1 Identification of Medium Composition Differences

In experiments *in vitro*, the growth medium is a solution that contains a variety of nutrients that are essential for the survival and growth of the organism. Likewise, *in silico*, the growth medium is represented by a list of reactions that import the metabolites and co-factors that are available and lead to a specified growth rate that the model has to achieve [79]. These reactions, designated as exchange reactions, are conceptual reactions for modelling influx and efflux across the bacterium boundaries. However, not all the exchange reactions that are part of the GSMM have active exchange fluxes, as represented by the difference between the columns "Total no. of exchange reactions" and "No. of active exchange fluxes" in Table 8: NoMedium models have all exchange reactions with an active flux, LB models have around 20% active and M9 about 7%. This difference is due to their bounds, being that the ones with active exchange fluxes have non-zero flux bounds. When an exchange reaction has a lower bound equal to zero (as in column "Bounds of non-active exchanges" of Table 8), it means a certain metabolite is not being provided to the organism (for instance, if we want to model an anaerobic medium, then we have to set the lower bound of the corresponding exchange reaction "EX_o2_e" to 0, e.g. [0,1000]).

Table 8. Characteristics of the growth media of the models (differently coloured according to the reconstruction's growth medium being "NoMedium", "LB" or "M9").

Models	Total no. of exchange reactions	No. of active exchange fluxes	Bounds of active exchanges	Bounds of non-active exchanges
U_NoMedium	251	251	(-1000.0, 1000.0)	————
gram_NoMedium	253	253	(-1000.0, 1000.0)	————
U_LB	251	50	(-10.0, 1000.0)	(0.0, 1000.0)
gram_LB	253	50	(-10.0, 1000.0)	(0.0, 1000.0)
U_M9	251	17	(-10.0, 1000.0)	(0.0, 1000.0)
gram_M9	253	17	(-10.0, 1000.0)	(0.0, 1000.0)

The composition of LB and M9 growth media is shown in Fig. 13a as well as the partial composition of NoMedium (Fig. 13a,b).

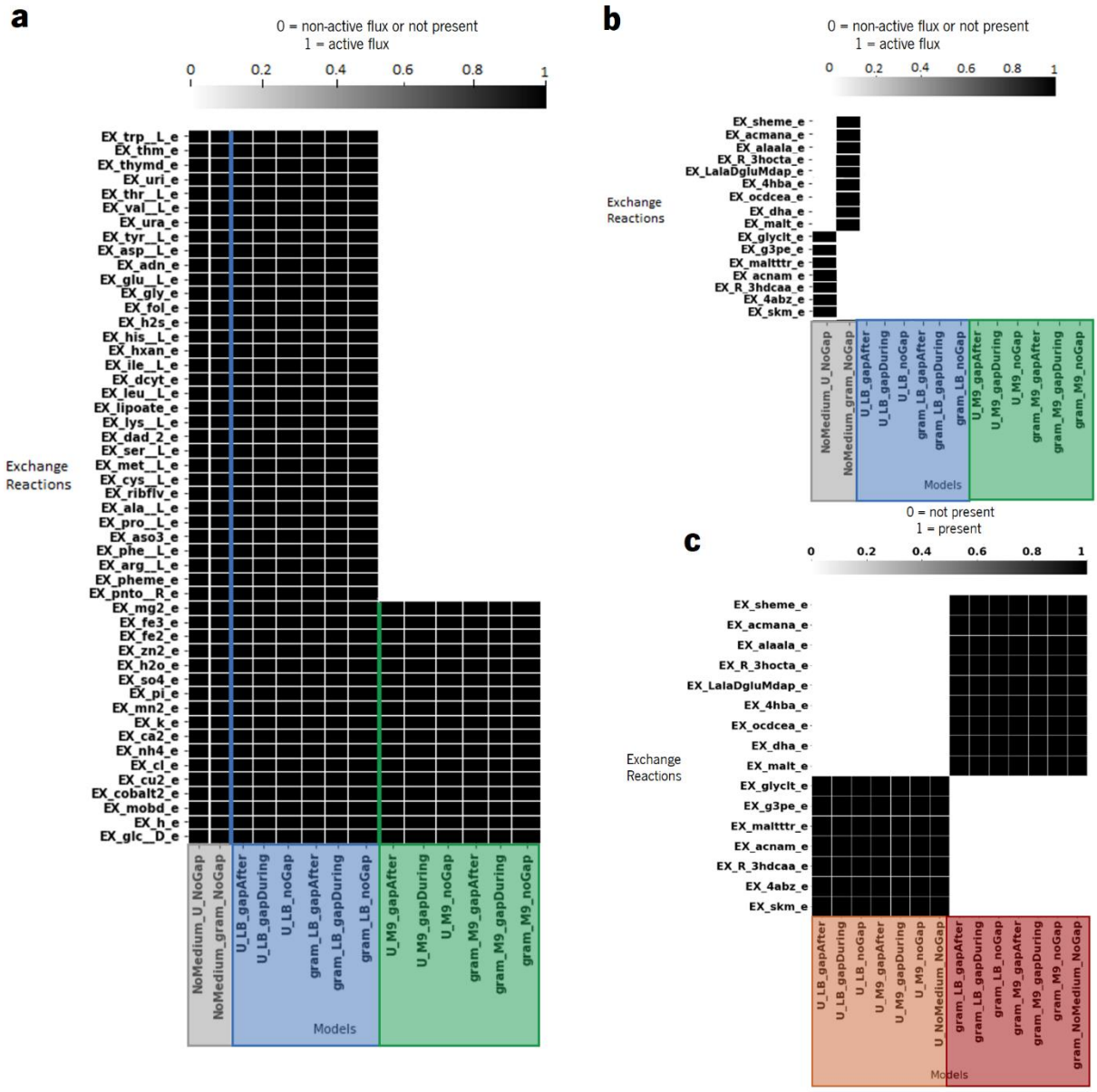


Fig. 13. Models' growth medium composition (which is defined by the exchange reactions that have an active flux) and difference in the exchange reactions that are part of the models. **a**, Heatmap with all exchange reactions that have an active flux for every model, except for the models with no initialised medium ("NoMedium" in x-axis labels) for convenience reasons, since these have around 250 reactions and more than 95% of them are present in both models ("NoMedium_U_NoGap" and "NoMedium_gram_NoGap"); reactions that have an active flux in a specific model are coloured as black and reactions with non-active flux or not present have a white colour; models are also grouped by colours according to their growth medium. **b**, Heatmap with the reactions that have an active exchange flux and that are different between the "NoMedium" models; colour scheme is the same as in the heatmap in **a**. **c**, Heatmap with the exchange reactions that are different between the models based on the universal and on the gram-negative templates; black colour means that the reaction is present and white means the opposite; models are also grouped by colours according to the template used for carving.

From Fig. 13a,b it is possible to gather that 7 exchange reactions only exist in universal models and 9 only in gram. This makes sense since the growth reaction of universal models does not contain membrane and cell wall precursors specific for gram-negative bacteria (as opposed to the growth reaction of gram-negative models) (Fig. 14). These precursors are different for gram-negative bacteria since they have a thinner peptidoglycan layer than gram-positive bacteria and present an outer lipid membrane [86]. These precursors are the following (as shown in Fig. 14):

- “kdo2lipid4_p”, a glycolipid which is an essential component of the outer cell wall, that functions as an hydrophobic anchor of lipopolysaccharides (LPS) [87]
- “murein5px4p_p” (peptidoglycan or murein), a polymer consisting of sugars and amino acids that is part of the peptidoglycan layer [88]
- “pe160/161_c,p”, phosphatidylethanolamines which are the major phospholipids in gram-negative bacteria [89]

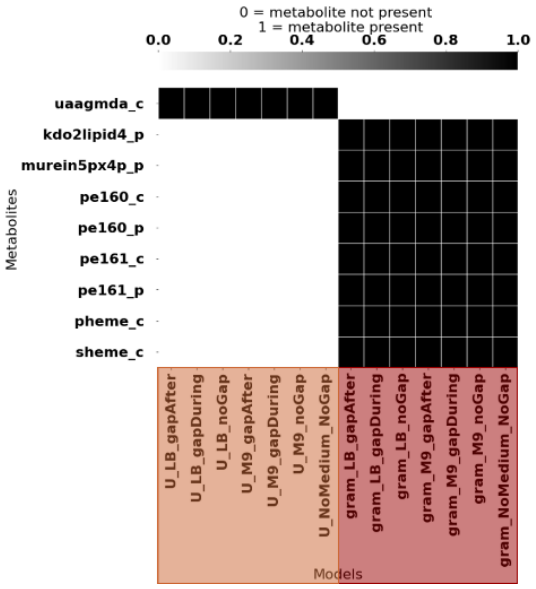


Fig. 14. Metabolites’ difference regarding growth reactions (heatmap where black colour means that the reaction is present and white means the opposite; models are grouped by colours according to the template used for carving being universal or gram).

According to these differences between the growth reactions of universal and gram-negative models, the latter need exchange reactions like “EX_alaala_e”, “EX_LalaDgluMdap_e”, “EX_4hba_e” and “EX_malt_e” (that import murein, amino acids and sugars) (Fig. 13c), which in turn is going to affect the metabolic pathways that are selected during the reconstruction process.

4.1.2 Model from Gram-Negative template, with LB Medium and Gap-Filled After Reconstruction Better Represents *B. thetaiotaomicron* and its *In Vitro* Growth Conditions

To choose the model that will better predict the neuroactive capabilities of the bacterium, an analysis regarding the reactions that produce or consume neuroactive metabolites (henceforward mentioned as neuroactive reactions for simplicity reasons) was performed.

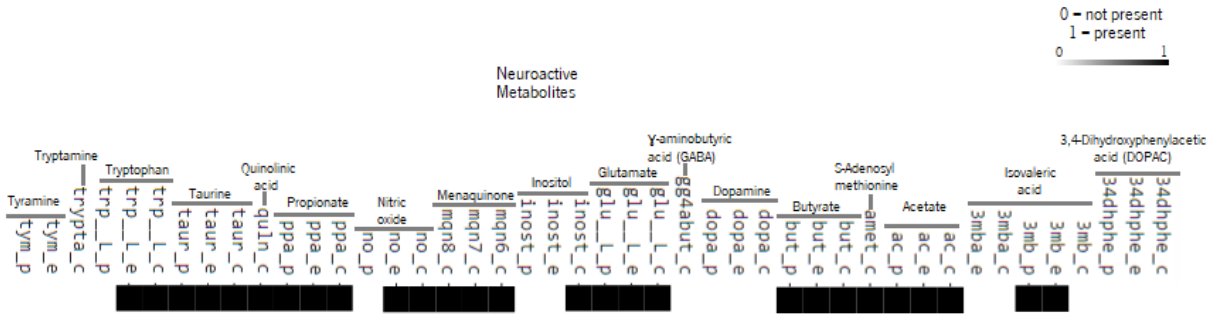


Fig. 15. Neuroactive metabolites (heatmap where black colour means that the metabolite is present in every model and white means the opposite).

From the list of neuroactive metabolites gathered from literature and presented in section 2.1.2, every model presented 12 out of the 17 metabolites (Fig. 15).

As represented in Fig. 16a, only 13 neuroactive reactions (out of 2000) differ between universal and gram models and, out of these, 8 seem to be interchangeable, i.e., reactions that are only present in universal have equivalent functions to reactions that only exist in gram, but some of the precursors are different. These only change because the models use different metabolic pathways. Of these 8, 3 only exist in universal ("AROAT", "GLUtex" and "OHPBAT") and 2 only in gram ("PYDXS" and "UHGADA"). Since the difference was not crucial, it was necessary to analyse how the models differed when it came to all the other reactions.

Regarding all reactions (Fig. 16a), there are around 200 reactions (out of 2000) that differ between universal and gram models, being that 21% are transport reactions (metabolite transport between compartments inside the bacterium or between extracellular/periplasmic space and cytoplasm, as opposed to exchange reactions that correspond to exchanges in the extracellular environment). However, most of the former seem to be interchangeable because of name similarities, which indicate a reaction involved in the same metabolic pathway, such as, for example, reactions "ACt4pp" and "ACt5pp" which are "Na+/Acetate symport (periplasm)" and "Acetate transport in via proton symport", respectively.

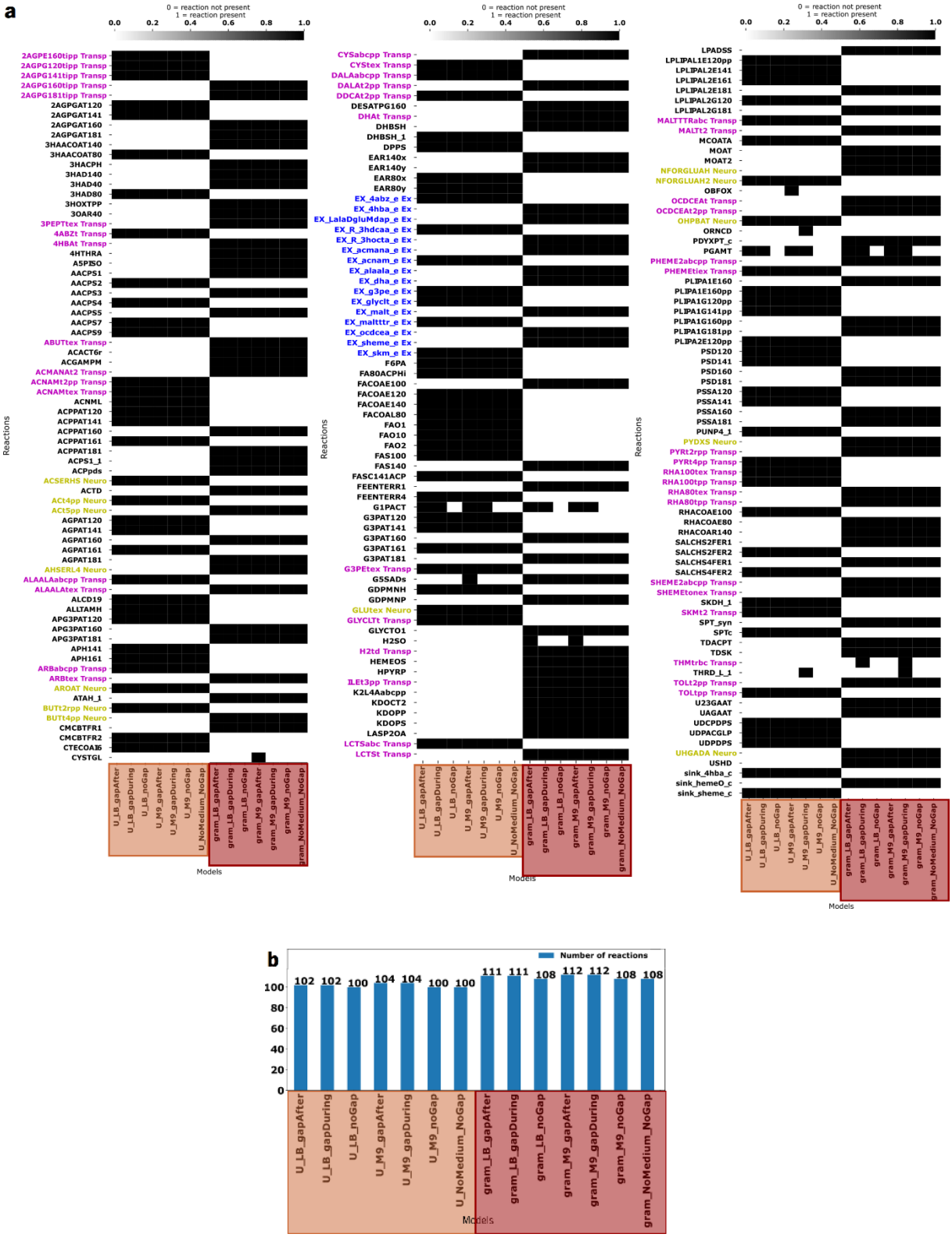


Fig. 16. Reactions that do not exist in at least one model. **a**, Heatmap where the black colour means that the reaction is present in the model and white means the opposite; reactions' names are coloured and named "Transp", "Neuro" and "Ex" to highlight transport, neuroactive and exchange reactions, respectively; models are grouped by colours according to the template used for carving being universal or gram **b**, Bar plot with the number of reactions per model, out of the reactions that don't exist in at least one model (colour scheme is the same for grouping the models).

From this point on, the choice of parameters was restricted to models using the gram-negative template, since these have in average 8 more reactions than universal ones (Fig. 16b) and overall higher FBA predicted values for the growth rate (Fig. 17).

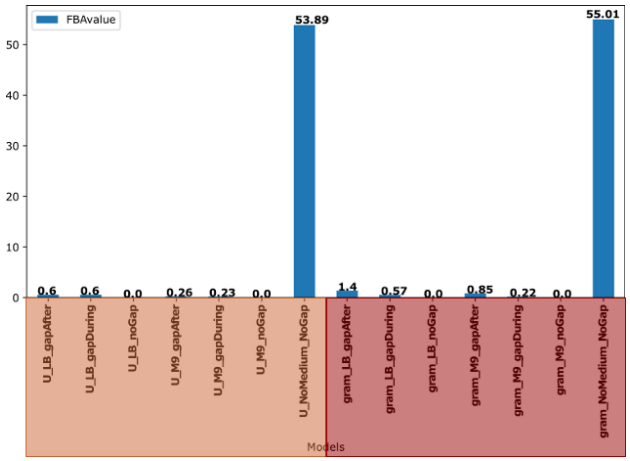


Fig. 17. Models' FBA values/growth rates (models are grouped by colours according to the template used for carving being universal or gram).

Even though the FBA value of the model with complete medium ("gram_NoMedium_NoGap", Fig. 17) is the highest of all models, this medium is not representative of the one in which the experiments of the paper [3] were performed, the modified Gifu Anaerobic Medium (mGAM), as one can observe by its estimated composition in Fig. 18, since the complete medium corresponds to all exchange reactions having non-zero flux bounds which corresponds to a very rich medium that is not the case of mGAM.

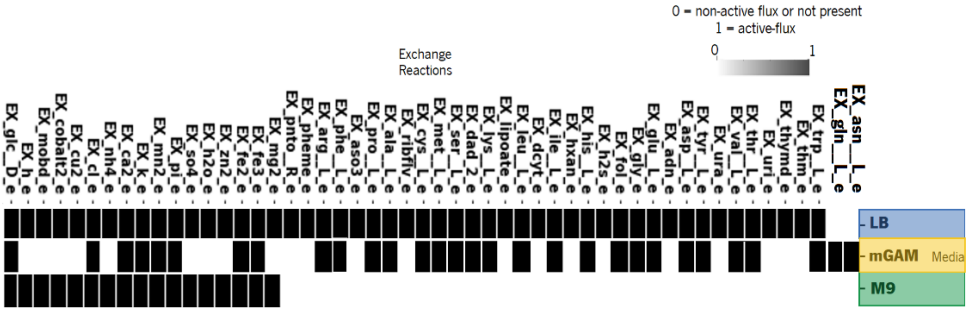


Fig. 18. mGAM deduced composition (heatmap where reactions that have an active flux are coloured as black and reactions with non-active flux or not present have a white colour; media names are distinguished by different colours)

Therefore, models that were initialised with LB medium were chosen for the drug effects simulations that follow in section 4.3, given that this medium resembles more the mGAM medium than M9 (Fig. 18).

Moreover, Fig. 16b shows that the gap-filling process adds 3 to 4 reactions in gram models, hence reducing the reconstructed models to gap-filled models. Between the alternatives for the process of gap-filling, doing it before lead to a very slow growth, which might mean that a substrate might be limiting growth due to something missing in the network. Gap-filling after the reconstruction lead to a higher FBA

value, which makes the model more feasible (Fig. 17). Thereby, the choice of the model "gram_LB_gapAfter" is the most satisfactory for the following analyses.

4.1.3 Alternative models built from the complete genome of *B. thetaiotaomicron* are highly similar

The model above-chosen - "gram_LB_gapAfter" - was generated according to one solution of a specific optimization problem, the MILP problem.

This model reunites all the different models that can be reconstructed with the alternative solutions of the MILP problem. An ensemble model with the same parameters as the "gram_LB_gapAfter" model and comprising the solutions of 100 models was generated.

The heatmap in Fig. 19 highlights the fact that there are only 14 reactions (out of the 2671 unique possible reactions; 0,5%) that are different between the models (out of the 100 models part of the ensemble model), i.e., were selected in at least one of the models but not in all; from these 14, the bottom 10 reactions are present in more than half of the models; from the rest of the 2671 reactions that are present in the database of the tool CarveMe, some were not selected in any model and others were selected in all.

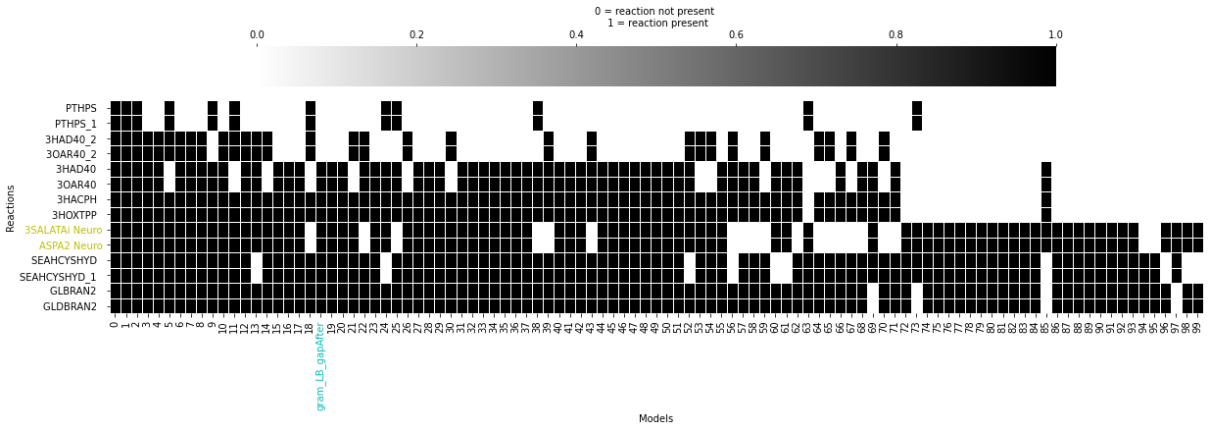


Fig. 19. Reactions of the ensemble model that differ between the 100 models (heatmap first sorted by ascending number of models in which a reaction is present, followed by descendent sort of number of reactions in each model; black colour means that the reaction is present in the model and white means the opposite; the model chosen in the previous section - "gram_LB_gapAfter" – is highlighted to be able to compare its reactions with the alternative models; the neuroactive reactions are also highlighted).

Furthermore, 39% of the models have the exact same reactions as the original "gram_LB_gapAfter", and 6 other only differ in two reactions that seem to be commutable with two other ("3HAD40"/"3HAD40_2" and "3OAR40"/"3OAR40_2", named "3-hydroxyacyl-[acyl-carrier-protein] dehydratase" and "3-oxoacyl-[acyl-carrier-protein] reductase", respectively), since in these models when one doesn't exist, the other does.

These results together demonstrate that the alternative models only have a very small percentage of reactions that change between themselves, not being substantially different from the model generated originally ("gram_LB_gapAfter"). Hence, it is possible to assume that using this model for the analyses will lead to results that can have a meaningful interpretation.

4.2 Automatically Reconstructed Model has the Potential to Produce/ Consume Neuroactive Metabolites

The "gram_LB_gapAfter" model has a total of 12 out of the selected 17 neuroactive metabolites and a total of 138 reactions (out of 2154; 6%), some of which are involved in their production and some in their consumption (Fig. 20).

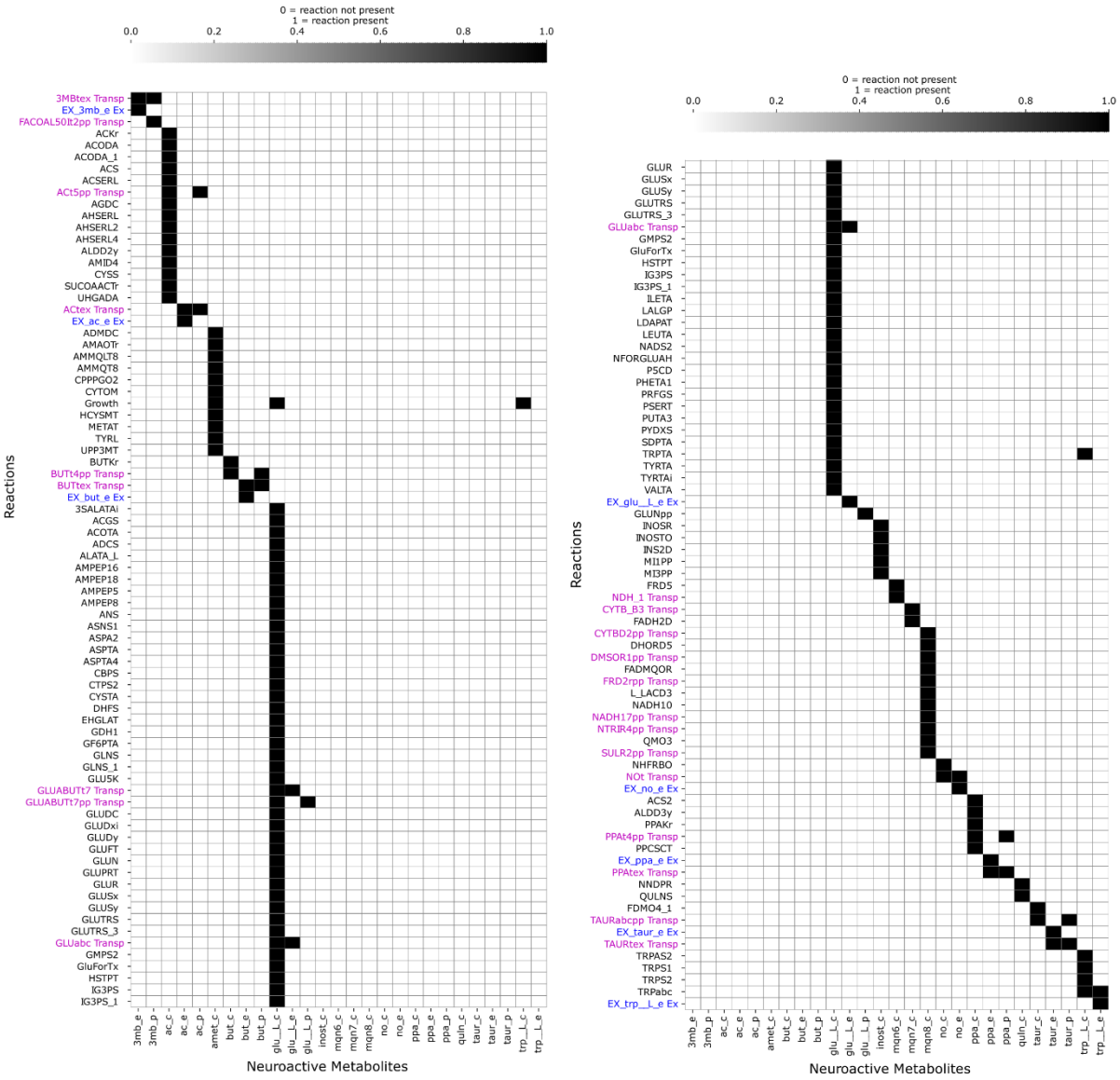


Fig. 20. Reactions that produce or consume the selected neuroactive metabolites (heatmap where the black colour means that the reaction is present in the model and white means the opposite; reactions' names are coloured and named "Transp" and "Ex" to highlight transport and exchange reactions, respectively; heatmap is divided into two parts, by the y-axis, due to size reasons).

4.3 *In Silico* Simulations Capture a Small Number of Experimentally Observed Drug Effects

According to Maier *et al.* 2018 [3], drugs that act on the nervous system (henceforth mentioned, interchangeably, as neuroactive drugs) inhibited, *in vitro*, gut bacteria more than other medications. To study this, they screened 198 different neuroactive drugs, being that 193 are human-targeted, 1 is targeted at protozoa, 1 at metazoan parasites and 3 are veterinary drugs (Fig. 21).

The research also comprised other drugs that cover all main ATC therapeutic classes, most of which are human-targeted and the rest are supposed to only inhibit pathogens but also affected gut commensals (Fig. 21).

In total, the paper screened 1200 compounds, of which 89 are not drugs (i.e., are compounds with biological roles (e.g. vitamins) or compounds that are being investigated, in clinical trials, as possible treatments) and 3 (Prestw-1105/385/425) did not have a measured impact on bacterial growth (no p-value).

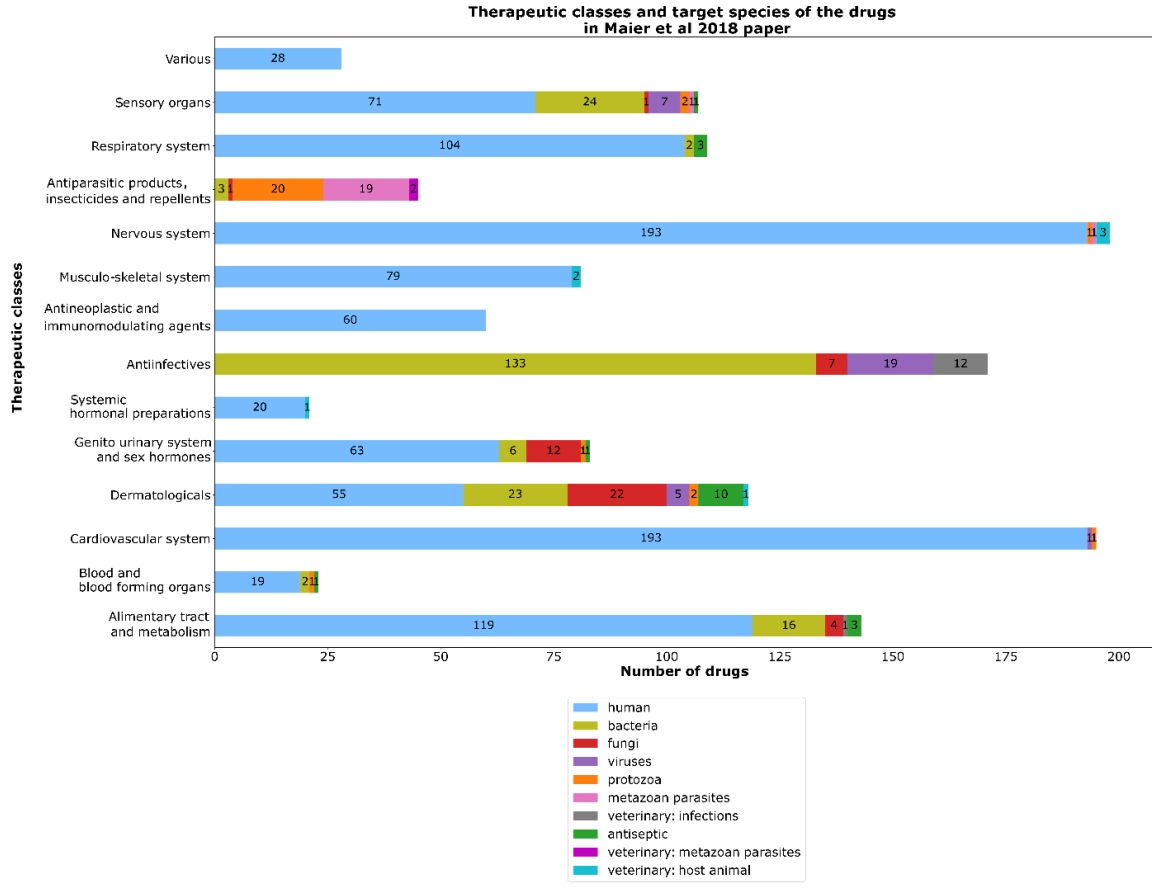


Fig. 21. Representation of therapeutic classes and target species of the drugs screened in Maier *et al* 2018 (13% of the drugs belong to more than one therapeutic class, hence adding up these values doesn't make up the total of 1111 drugs). Each bar corresponds to drugs that belong to one specific therapeutic class and is subdivided by the target species of the drugs.

In the next sections, we aimed at simulating the effects of these drugs on bacterial growth with genome-scale models and flux balance analysis.

4.3.1 Neuroactive Drugs have Divergent *In Vitro* and *In Silico* Effects on Bacterium's Growth

The experimental results showing bacterial inhibition by neuroactive drugs are intriguing, since they are supposed to target receptors that are absent in bacteria (like dopamine and serotonin receptors). The idea then was to investigate, *in silico*, which proteins are affected in *B. thetaiotaomicron* that restrict the flux of reactions in a way that lead to its growth inhibition *in vitro*, using drug-protein interactions from STITCH database.

Each one of the 198 neuroactive drugs is identified by a unique Prestwick ID. Out of these, 4 Prestwick IDs correspond to 2 STITCH IDs, being that each pair of Prestwick IDs corresponds to 1 STITCH ID (Table 9).

Table 9. Neuroactive drugs from Maier *et al* 2018 that have different Prestwick IDs but same STITCH ID.

Prestwick ID	STITCH ID
Prestw-109, Prestw-935	CIDs00001207
Prestw-1271, Prestw-692	CIDs00002771

Of the 196 unique STITCH IDs, only 41 have reported inhibiting interactions with genes of *B. thetaiotaomicron* (Fig. 22a). These interactions comprise 36 genes in the "gram_LB_gapAfter" model (Fig. 22b).

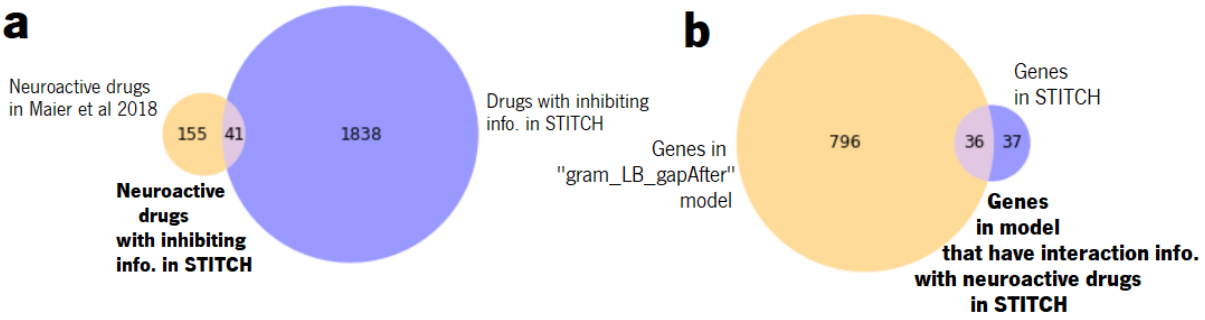


Fig. 22. Inhibiting interactions, from STITCH database, involving neuroactive drugs (from Maier *et al* 2018) and *B. thetaiotaomicron* genes. **a**, 41 Neuroactive drugs that have inhibiting information with *B. thetaiotaomicron* in STITCH. **b**, 36 *B. thetaiotaomicron* proteins from "gram_LB_gapAfter" model that have, in STITCH, inhibiting interaction information with the 41 neuroactive drugs (37 genes in STITCH don't have a correspondent gene in carveme, because STITCH uses "old_locus_tag" but some genes in the FASTA file [used for "gram_LB_gapAfter"] only have the "new_locus_tag"). Bold font in figure is used to highlight the intersection information.

From these 41 drugs, none of them inhibited *B. thetaiotaomicron*'s growth *in vitro*.

In silico and using the chosen model "gram_LB_gapAfter", 3 drugs – acetaminophen, gabapentin, vigabatrin – inhibited growth (absolute percentage change in growth rate of 100%, Fig. 23a,b) when the genes, with inhibiting interactions with the corresponding drug, were "knocked-out" at the same time.

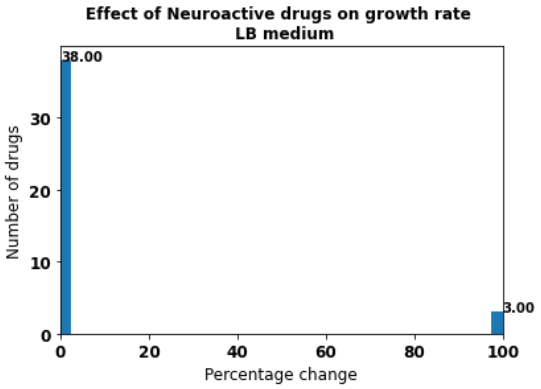


Fig. 23. Effect of drugs on growth rate, in "gram_LB_gapAfter" model (drugs that lead to 100% change are the ones that had an effect; in this case, 3 drugs).

These *in silico* results happened because each of these 3 drugs has interaction information in STITCH with one essential gene in "gram_LB_gapAfter" model (acetaminophen and gabapentin with "BT_1806" and vigabatrin with "BT_3935") (Table 10). A gene is essential when restricting the flux of the reactions that depend on it lead to a null growth rate. However, these interactions have a score of around 0.4 out of 1, which is a low confidence value, and, according to STITCH database, they have no experimental evidence and were transferred via orthology from another organism. This might mean that these interactions are incorrect and, if so, these drugs would have an inhibitory effect *in silico*, like they didn't *in vitro*.

Table 10. Neuroactive drugs with *in silico* effect in carveme model.

Drug	Acetaminophen (paracetamol)	Gabapentin	Vigabatrin
ATC code	N02BE01 (analgesic)	N03AX12 (antiepileptic)	N03AG04 (antiepileptic)
STITCH ID	CID000001983	CID000003446	CID000005665
Target species	Human	Human	Human
Target species' proteins	PTGS1 (COX1) PTGS2 (COX2)	CACNA2D SLC6A1 (GAT1)	ABAT

Target proteins' classification	Oxidoreductases	Ion channels Solute carrier family	Aminotransferases
B. thetaiotaomicron orthologous genes in KEGG database	———————	———————	———————
B. thetaiotaomicron essential genes in "gram_LB_gapAfter" model	BT_1806 (oxidoreductase: acyl-CoA dehydrogenase family protein)	BT_3935 (transferase: aminotransferase class I/II-fold pyridoxal phosphate-dependent enzyme)	
STITCH score	0,471 (not experimental)	0,410 (not experimental)	

In order to understand if the differences between *in vitro* and *in silico* results were due to models lacking information, the analyses were repeated for a *B. thetaiotaomicron*'s GSMM, assembled and gone through extensive manual curation by Heinken *et al.* 2013 (model designated, from this point onwards, as curated and, for simplicity reasons, "gram_LB_gapAfter" model is going to be named as carveme) [82]. For comparison purposes, the medium was modified in order to resemble LB medium from carveme model.

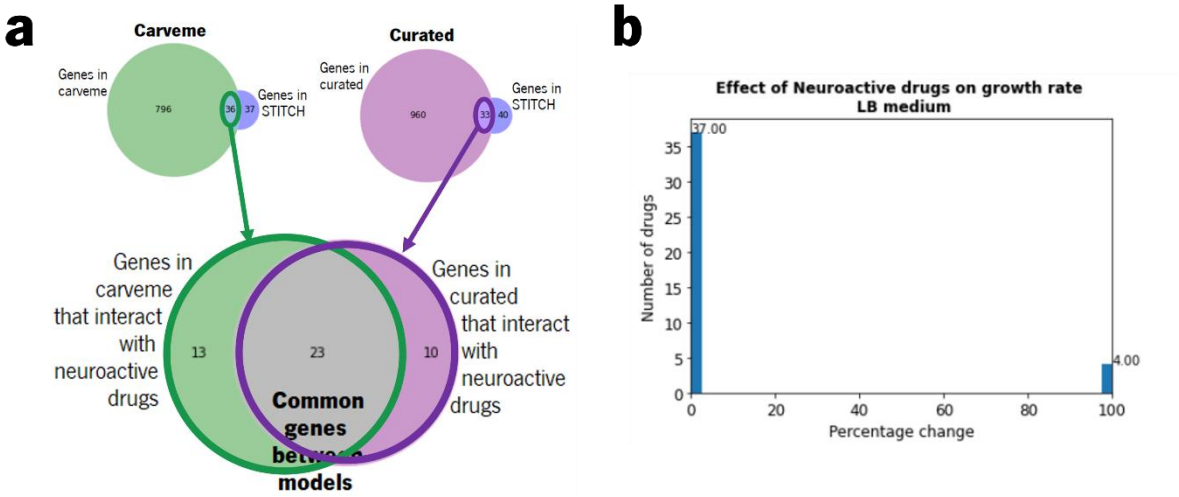


Fig. 24. Neuroactive drugs and curated model. **a,** 33 *B. thetaiotaomicron* genes from curated model (out of 73 in STITCH) have, in STITCH, inhibiting interaction information with the 41 neuroactive drugs and 23 genes are common between carveme and curated models. **b,** Effect of drugs on growth rate (drugs that lead to 100% change are the ones that had an effect; in this case, 4 drugs).

Fig. 24c shows that 23 genes are common between carveme and curated models. And in Fig. 24a, we can see that the curated model has 3 genes less than carveme, with interaction information from STITCH.

From Fig. 24b and from Table 11, we can see that, instead of the 3 above-mentioned drugs that had an *in silico* effect on growth in the carveme model, in the curated model 4 different drugs had an effect,

namely Carbamazepine, Valproic acid, Lamotrigine and Topiramate. Each of these drugs affects 2 essential genes in curated, "BT_1225" and "BT_0382", that do not exist in carveme (Table 11).

Table 11. Neuroactive drugs with *in silico* effect in curated model.

Drug	Carbamazepine	Valproic acid	Lamotrigine	Topiramate
ATC code	N03AF01 (antiepileptic)	N03AG01 (antiepileptic)	N03AX09 (antiepileptic)	N03AX11 (antiepileptic)
STITCH ID	CID000002554	CID000003121	CID000003878	CID000005514
Target species	Human	Human	Human	Human
Target species' proteins	SCN1-5/8-11A	ABAT SSADH GAD CACNA1-T	SCN1-5/8-11A	SCN1-5/8-11A GRI[A/K] CACNA1-L GABR CA2/4
Target proteins' classification	Ion channels	Ion channels Oxidoreductases Transferases Lyases	Ion channels	Ion channels (ligand-gated: GABA, glutamate) Lyases
B. thetaiotaomicron orthologous genes in KEGG database	————————	BT_2570 (lyase: glutamate decarboxylase)	————————	————————
B. thetaiotaomicron essential genes in curated model	BT_1225 (oxidoreductase: GDP-L-fucose synthase) BT_0382 (oxidoreductase: NAD-dependent epimerase/dehydratase family protein)			
STITCH score	0,388 (not experimental)			

From the essential genes affected in carveme, "BT_1806" is not present in the curated model, while "BT_3935" is present but is not essential.

The drug "valproic acid" affects the human gene "GAD" (glutamate decarboxylase 1) that corresponds, according to KEGG database [90], to the *B. thetaiotaomicron*'s orthologous gene "BT_2570" (glutamate decarboxylase) (Table 11). This gene is present in the curated model, but it is not essential (meaning that even if it was present, this gene alone wouldn't inhibit growth) and it does not have inhibiting interaction in STITCH (so even though the drug interacts with this protein, it can interact with it in a different way from inhibition, such as, for instance, activation or binding).

4.3.2 Neuroactive Drugs Influence Neuroactive Metabolism

With STITCH information, it is also possible to investigate if the neuroactive drugs can have any effect on *B. thetaiotaomicron*'s neuroactive reactions and, consequently, on the production and/or consumption

of neuroactive metabolites by the bacterium. This information can then be verified experimentally by mass spectrometry-based metabolomics approaches.

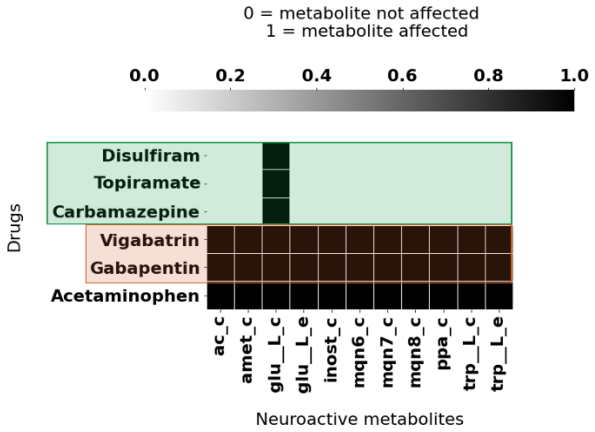


Fig. 25. Neuroactive metabolites that might be possibly affected by neuroactive drugs in carveme model (heatmap with black colour representing a metabolite that was affected by a drug and with white colour representing the opposite; the drugs that are coloured are the ones that also affect neuroactive reactions in curated model).

Looking at Fig. 25, and connecting the information there with the one from Table 1 and from Fig. 20, it is noticeable that, out of the 41 neuroactive drugs, 6 have an effect in neuroactive reactions in the carveme model. The neuroactive metabolites that might be affected by the neuroactive drugs that had *in silico* effect on growth - vigabatrin, gabapentin and acetaminophen - are acetate (ac_c), S-adenosylmethionine (amet_c), glutamate (glu__L_c/e), inositol (inost_c), menaquinone (mqn6/7/8_c), propionate (ppa_c) and tryptophan (trp__L_c/e). Additionally, the drugs topiramate, disulfiram and carbamazepine might also affect glutamate.

Regarding the curated model, a total of 8 neuroactive drugs had an effect on neuroactive reactions (Fig. 26).

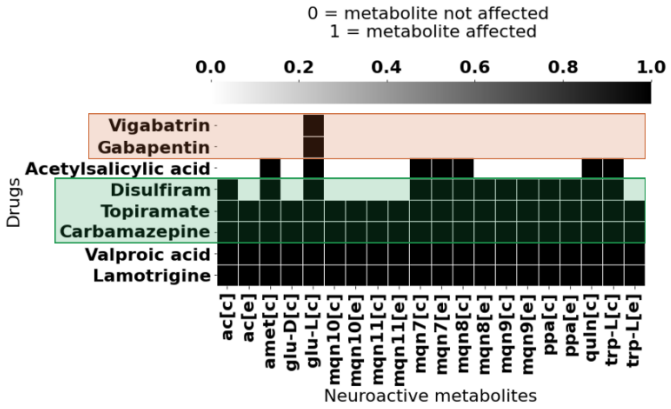


Fig. 26. Neuroactive metabolites that are affected by neuroactive drugs in curated model (heatmap with black colour representing a metabolite which had its production and/or consumption affected by a drug and with white colour representing the opposite; the drugs that are coloured are the ones that also affect neuroactive reactions in carveme model).

5 out of the 8 drugs (vigabatrin, gabapentin, disulfiram, topiramate and carbamazepine) also had an effect in the carveme model. However, carveme and curated models have basically inverted results:

- disulfiram, topiramate and carbamazepine only affect glutamate in carveme, but affect acetate, S-adenosylmethionine, glucose, menaquinone, propionate and tryptophan in curated; plus quinolinic acid (quln[c]), that is not affected in carveme.
- vigabatrin and gabapentin only affect glutamate in curated, but affect acetate, S-adenosylmethionine, glucose, menaquinone, propionate and tryptophan in carveme; plus inositol, that is not affected in curated.

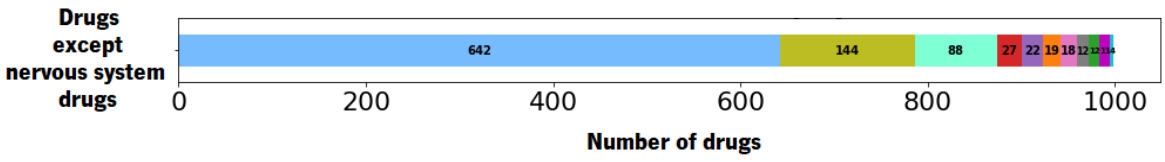
These results are likely due to the fact that the neuroactive reactions indirectly affected by some drugs are not the same when comparing carveme with curated.

4.3.3 *In Silico* Results Match *In Vitro* Growth Inhibition Effect of Two Percent of the Non-Commensal Targeting Drugs

Besides the drugs that act upon the human nervous system, Maier *et al* 2018 [3] also screened drugs from other ATC therapeutic classes (Fig. 21). However, it focused on drugs that are not supposed to affect gut commensals, to generate a systematic resource of the effect of drugs in altering gut microbiome composition (i.e., changing the species that are present in the gut) [3]. These drugs comprised, for example, the ones which have the human cells as targets (642 drugs) and antibiotics that are supposed to only inhibit pathogens (and not commensals, 144 drugs), as shown in Fig. 27 [3].

Even though some of the compounds studied are not considered drugs (88 “not a drug” , Fig. 27), (i.e., are compounds with biological roles (e.g. vitamins) or compounds that are being investigated, in clinical trials, as possible treatments), they were designated, both in the present work and in the research paper, as drugs (for simplicity reasons, making the words “compound” and “drug” interchangeable).

Removing the 198 nervous system drugs, investigated in the previous section 4.3.1, from the 1197 compounds from Maier *et al* [3], a total of 999 compounds were selected for the investigation of their effect in *B. thetaiotaomicron*’s GSMM (Fig. 27).



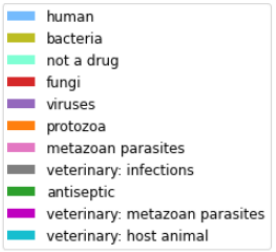


Fig. 27. Target species of the 999 compounds from all ATC classes, except from the nervous system one.

Every one of the 999 compounds is identified by a unique Prestwick ID. However, not all of them have a unique STITCH ID. 29 Prestwick IDs comprise 14 unique STITCH IDs, being that 13 pairs match to one unique STITCH ID and 3 match to a last one (Table 12). This gives a total of 984 unique STITCH IDs. The drugs with the same STITCH ID will have the same genes inhibited and, subsequently, the same *in silico* results.

Table 12. Drugs from Maier *et al* 2018 that have different Prestwick IDs but same STITCH ID (coloured rows correspond to the drugs in STITCH database that have inhibiting interactions with *B. thetaiotaomicron*).

Prestwick ID	STITCH ID
Prestw-1075/1081	CIDs00004946
Prestw-1086/1089	CIDs00000401
Prestw-1516/198	CIDs00002083
Prestw-847/857	CIDs00003914
Prestw-860/864	CIDs00069216
Prestw-1097/911	CIDs00003779
Prestw-135/373	CIDs00004112
Prestw-182/545	CIDs00003913
Prestw-233/565	CIDs00003661
Prestw-256/257	CIDs00005538
Prestw-285/985	CIDs00005645
Prestw-45/791	CIDs00001301
Prestw-536/953	CIDs00002249
Prestw-411/440/697	CIDs00000225

Out of these 984 STITCH IDs, only 232 (Fig. 28a) have inhibiting interactions with *B. thetaiotaomicron*'s genes according to STITCH, which encompass 96 genes in the carveme model and 128 in the curated one (Fig. 28b). Due to some STITCH IDs having more than 1 Prestwick ID (Table 12), as mentioned above, these 232 STITCH IDs correspond to 242 Prestwick IDs. Out of the latter, 41 inhibited the growth of *B. thetaiotaomicron in vitro* (Fig. 29a) and, *in silico*, 10 had an effect in carveme and 20 in curated (Fig. 29b), being that 7 had effect in both models.

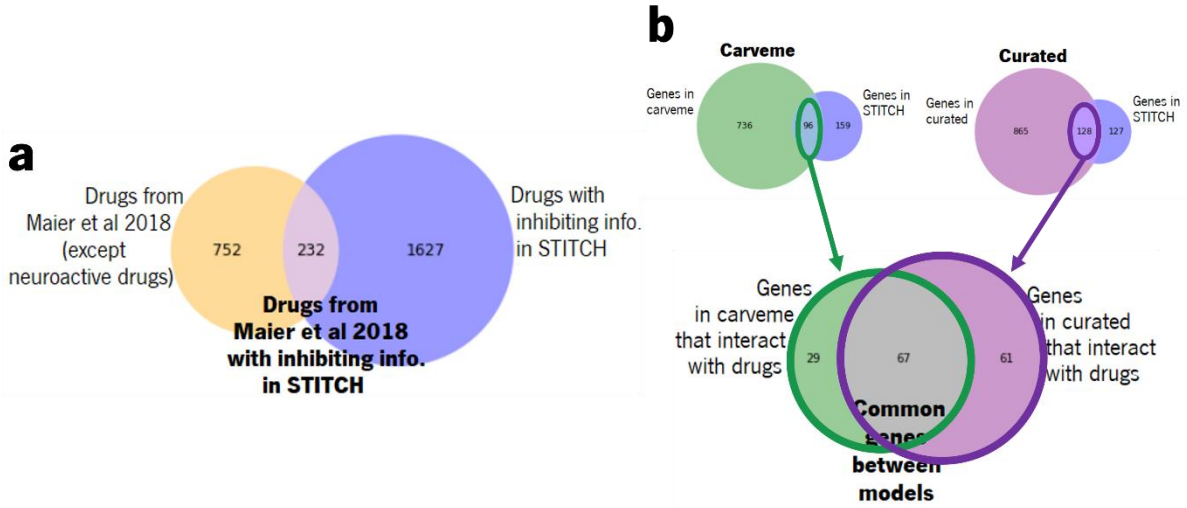


Fig. 28. Inhibiting interactions, from STITCH database, involving all drugs from Maier *et al* 2018 (except neuroactive drugs) and *B. thetaiotaomicron* genes. **a**, 232 Drugs that have inhibiting interaction information with *B. thetaiotaomicron*, in STITCH. **b**, 96 *B. thetaiotaomicron* genes from carveme and 128 from curated (out of 255 proteins in STITCH) have, in STITCH, inhibiting interaction information with the 232 drugs' STITCH IDs and 67 genes are common between carveme and curated models.

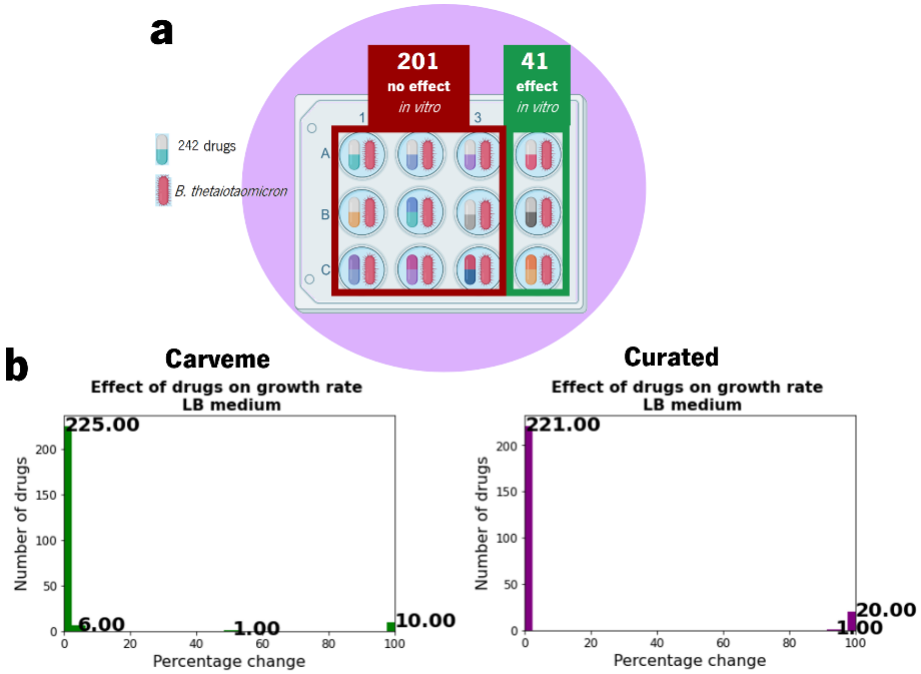


Fig. 29. Effect of the 242 drugs on *B. thetaiotaomicron*'s growth. **a,** Drugs' effect *in vitro* (191 no effect/41effect). **b,** Drugs' effect *in silico*, in both carveme and curated models, out of 242 unique Prestwick IDs.

From the 242 drugs, only 2 – Prestw-808 (Furazolidone) and Prestw-732 (Streptozotocin) – had an effect *in vitro* and in both models (carveme and curated) and 2 more *in vitro* and in the curated model only – Prestw-237 (Ofloxacin) and Prestw-1479 (Triclosan) – as Table 13 highlights.

Table 13. 60 Drugs (out of the 242 that had, in STITCH database, inhibiting interactions with *B. thetaiotaomicron*'s proteins) that had effect either *in vitro* (first 41 rows with p-value <=0.01 [pink rows]), *in silico* (in carveme [green rows] and curated [purple rows] models) or in both (yellow rows) (corresponding STITCH IDs and drugs names are in Supplementary Table 1).

Prestwick ID	Proteins in STITCH	Proteins in carveme	Proteins in curated	Affected reactions carveme	Affected reactions curated	% Growth change carveme	% Growth change curated	Effect *in vitro* (p-value)	Target species
1109	3	1	0	0	0	0	0	0.00	bacteria
1233	3	3	3	16	4	0	0	0.00	human
525	3	1	0	0	0	0	0	0.00	bacteria
151	3	1	0	0	0	0	0	0.00	bacteria
808	1	1	1	1156	780	100	100	0.00	bacteria
113	7	3	2	0	20	0	0	0.00	bacteria
208	1	0	0	0	0	0	0	0.00	bacteria
1415	1	1	1	0	6	0	0	0.00	human
1157	3	1	0	0	0	0	0	0.00	bacteria
1446	4	0	0	0	0	0	0	0.00	bacteria
1265	4	0	0	0	0	0	0	0.00	bacteria
238	4	0	0	0	0	0	0	0.00	bacteria
237	9	2	2	0	802	0	100	0.00	bacteria
1343	4	0	0	0	0	0	0	0.00	bacteria
766	3	1	0	0	0	0	0	0.00	bacteria
756	7	4	0	0	0	0	0	0.00	human
1401	4	0	0	0	0	0	0	0.00	bacteria
37	1	0	0	0	0	0	0	0.00	protozoa
1056	1	1	1	0	6	0	0	0.00	viruses
1194	4	0	0	0	0	0	0	0.00	antiseptic
333	1	1	1	2	95	0	0	0.00	viruses
168	7	5	4	16	4	0	0	0.00	bacteria
1479	20	5	10	0	752	0	100	0.00	antiseptic
1378	3	1	0	0	0	0	0	0.00	bacteria
699	1	1	0	0	0	0	0	0.00	human
1	1	1	1	2	0	0	0	0.00	human
1303	4	0	0	0	0	0	0	0.00	bacteria
370	13	5	0	0	0	0	0	0.00	human
390	3	1	0	0	0	0	0	0.00	bacteria
732	34	22	29	1151	759	100	100	0.00	human
708	1	1	0	0	0	0	0	0.00	antiseptic
267	3	1	0	0	0	0	0	0.00	fungi
376	3	1	0	0	0	0	0	0.00	bacteria
740	1	1	0	0	0	0	0	0.00	not a drug
1203	14	10	7	0	27	0	0	0.00	human
487	4	0	0	0	0	0	0	0.00	human
205	1	1	0	0	0	0	0	0.00	human
368	1	1	0	0	0	0	0	0.01	human
126	3	1	0	0	0	0	0	0.01	protozoa
478	3	1	0	0	0	0	0	0.01	human
1114	3	1	0	0	0	0	0	0.01	viruses
736	2	1	2	0	802	0	100	0.02	bacteria
1467	20	8	12	0	752	0	100	0.11	human
1314	15	4	10	0	752	0	100	0.24	human
94	1	1	1	0	783	0	100	0.52	human
1337	21	7	14	0	752	0	100	0.74	human
105	31	2	13	0	794	0	100	0.80	human
275	22	2	3	0	803	0	100	0.81	human
1097/911	35	9	32	1186	4	100	0	0.93/1.00	human
1134	2	2	2	161	760	51	94	1.00	human
14	2	2	1	12	752	0	100	1.00	bacteria
257/256	9	7	4	1108	803	100	100	1.00/1.00	human
1210	2	2	2	1165	804	100	100	1.00	human
741	26	19	17	1167	784	100	100	1.00	not a drug
1198	28	1	13	0	794	0	100	1.00	human
1285	2	2	2	1165	804	100	100	1.00	human
1118	2	1	2	0	802	0	100	1.00	bacteria

489	2	1	2	0	802	0	100	1.00	bacteria
441	20	12	8	1152	7	100	0	1.00	human

Regarding the 2 drugs that had an effect *in vitro* and in both *in silico* models:

- the drug "Prestw-808" affected one gene "BT_0347" that is essential in both models (Table 14). This drug-protein interaction was transferred via orthology (STITCH score 0,6) from another species ("Gallus gallus").
- the drug "Prestw-732" did not affect any essential gene in either of the models (Table 14). However, it had many reported protein interactions in STITCH (34) and more than 60% of these proteins were present in carveme and in curated (22 and 29, respectively, as shown in Table 14). Consequently, their knock-out lead to many reactions having their fluxes changed (around 50% of the total number of reactions in both models) (Table 13). Even though the drug is targeted at humans, according to Maier *et al* 2018 [3], it has been previously reported as having antibacterial activity.

Table 14. Drugs with both *in vitro* and *in silico* effect.

Drug	Prestw-808	Prestw-732	Prestw-237	Prestw-1479
Name	Furazolidone	Streptozotocin	Ofloxacin	Triclosan
ATC code	G01AX06 (Gynecological antiinfectives and antiseptics - nitrofurans [inhibit glucose mechanism])	L01AD04 (Antineoplastic/ alkylating agents)	J01MA01/12 (Antibacterial for systemic use) S01AX11/19 S02AA16 (Antiinfective)	D08AE04 D09AA06 (Antiseptics and disinfectants)
STITCH ID	CID000003435	CID000005300	CID000004583	CID000005564
Target species	Bacteria	Human	Bacteria	Antiseptic
Target species' proteins	———	———	DNA gyrase DNA Topoisomerase	Enoyl-acyl carrier protein reductase enzyme
Target proteins' classification	———	Binds to phosphate, amino, sulfhydryl, hydroxyl, and imidazole groups, commonly found in nucleic acids and other macromolecules	Isomerases	Oxidoreductases
B. thetaiotaomicron orthologous genes in KEGG database	———	———	BT_0899 (DNA gyrase) BT_3579 (DNA topoisomerase)	BT_4188 (oxidoreductase)
B. thetaiotaomicron essential genes in carveme model	BT_0347 (transketolase; involved in pentose phosphate pathway)	———	———	———
B. thetaiotaomicron essential genes in curated model	BT_0347 (transketolase)	———	———	BT_1225 (oxidoreductase: GDP-L-fucose synthase) BT_0382

				(oxidoreductase: NAD-dependent epimerase/dehydratase family protein)
STITCH score	0,604 (not experimental)	———————	———————	0,382 (not experimental)

As for the drugs that inhibited growth *in vitro* but only had an effect in the curated model:

- the drug "Prestw-237" only affected two genes – "BT_2048" and "BT_0106" - (Table 13), but none of them were essential (Table 14). However, their "knock-out" restricted the flux of two reactions, "FOLR" and "DHFR", to zero, which, consequently, lead to a high number of affected reactions (50% of total reactions).
  - In carveme, the drug also targeted "BT_2048", but "BT_0106" was not present and instead targeted "BT_3386". These only affected the reaction "FOLR2", but its fluxes' range did not change and, subsequently, didn't affect any other reactions.
  - Moreover, according to KEGG and to STITCH database, this drug targets *B. thetaiotaomicron*'s genes "BT_0899" and "BT_3579", but these genes are not present in carveme nor in curated models [90].
- The drug "Prestw-1479" targeted two essential genes in curated, namely "BT_1225" and "BT_0382" (Table 13).
  - These genes don't exist in carveme.
  - Furthermore, this drug targets "BT_4188" according to KEGG but, even though it is present in both models and it is essential in curated, it does not have inhibiting interactions in STITCH.

Regarding the 201 drugs that did not inhibit growth *in vitro*, 20 did inhibit in at least one of the models (17 in curated, 8 in carveme and 5 in both, as Table 13 shows). From these:

- the 8 drugs (6 unique STITCH IDs) that inhibited carveme affected 6 essential genes altogether, 1 of which is also essential in curated (BT_2123), 3 exist but are not essential there and 2 don't exist (Table 15).
- the 17 drugs (16 unique STITCH IDs) that inhibited curated, only 8 targeted essential genes. A total of 12 essential genes were affected, being that 1 is also essential in carveme, 5 exist in carveme but are not essential there and 6 don't exist (Table 15). The drugs that did not affect essential genes lead, in average, to a higher number of affected reactions than the ones that targeted essential genes (796 vs 768, respectively).

Table 15. Essential genes affected by drugs that didn't inhibit growth *in vitro* but inhibited *in silico*; (*) means that the drug inhibited growth but it didn't affect any essential gene.

Prestwick ID	Essential gene(s) carveme	Essential gene(s) curated	Model's presence
736	————	Inhibited curated (*)	
1467	————	BT_1225	Not in carveme
		BT_0382	Not in carveme
		BT_0674	Exists in carveme but not essential
1314	————	BT_1225	Not in carveme
		BT_0382	Not in carveme
94	————	BT_3845	Exists in carveme but not essential
1337	————	BT_1225	Not in carveme
		BT_0382	Not in carveme
105	————	Inhibited curated (*)	
275	————	Inhibited curated (*)	
1097/911	BT_1806	————	Not in curated
1134	————	Inhibited curated (*)	
14	————	BT_0373	Exists in carveme but not essential
257/256	BT_2123	BT_2123	Exists and it's essential in both
1210	BT_3261	Inhibited curated (*)	Exists in curated but not essential
741	BT_4503		Not in curated
		BT_3845	Exists in carveme but not essential
		BT_2797	Exists in carveme but not essential
1198	————	Inhibited curated (*)	
1285	BT_3261	Inhibited curated (*)	Exists in curated but not essential
1118	————	Inhibited curated (*)	
489	————	Inhibited curated (*)	
441	BT_3935	————	Exists in curated but not essential

In order to understand the differences between the models, regarding the effect of the drugs, it was necessary to study how the number of proteins that have information in STITCH, the number of affected proteins, the number of affected reactions and the effect on the bacterium's growth, changed together (Fig. 30). However, in this analysis, the drugs that affected essential genes (15 drugs) were removed from the 242 drugs, in order to be able to generally understand the relationship between the affected proteins and respective affected reactions.

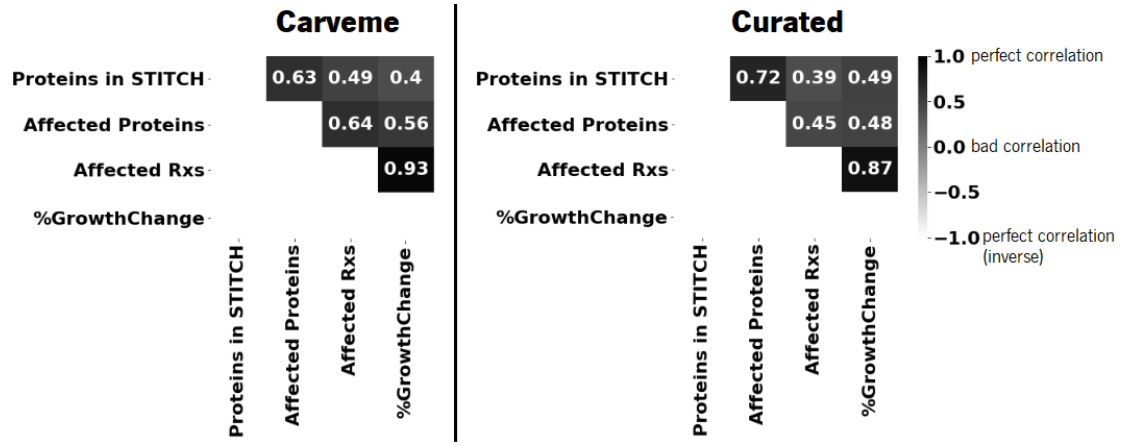


Fig. 30. Relationship between the number of proteins that have interactions in STITCH ("Proteins in STITCH"), the number of affected proteins in the *in silico* models ("Affected Proteins"), the number of affected reactions (i.e., reactions for which their fluxes' ranges before and after the genes were "knocked-out", were not 100% overlapping) ("Affected Rxs") and the percentage of change in growth rate ("%GrowthChange"); this applies to all the 242 drugs mentioned before, except the drugs that affected essential genes in order to be able to understand the correlation between the number of affected genes/proteins and the number of affected reactions; left panel is the correlation heatmap for carveme and the right panel is for curated, where a value of 1 or -1 means a perfect correlation, a value above 0.6 was considered a good correlation and the absolute values below 0.6 are considered a bad correlation. The correlation values correspond to the Pearson correlation coefficient.

Looking at Fig. 30 and at Fig. 31 (that shows the distribution of the number of affected proteins and the percentage of affected reactions by the 242 drugs), overall, one can make some observations about these results, such as:

- There is a higher correlation between the number of proteins that have inhibiting information in STITCH and the number of affected proteins in curated than in carveme (0.72 vs 0.63, respectively).
  - This means that curated has more genes that interact with the drugs than carveme. This matches the information from Fig. 28b, which shows that curated has more genes interacting with drugs than carveme. Hence, when there is more information from STITCH, it is more likely that more proteins will be affected in curated than in carveme.
- Regarding the majority of the drugs (the ones that did not have an effect either *in vitro* or *in silico*), the number of proteins affected by them was overall equal to or less than 5, in both models(Fig. 31a). However, in curated, more drugs led to a higher percentage of affected reactions (around 7% of the drugs affected the flux of about 20% of the reactions) than in carveme (each drug affected less than 10% of the reactions) (Fig. 31b). This is corroborated by the better correlation between the number of affected proteins and the number of affected reactions in carveme than in curated (0.64 vs 0.45, respectively) (Fig. 30).
  - Nevertheless, it is a weak relationship in both, meaning that an increase in the number of affected reactions does not mean it was caused by a higher number of affected proteins.
  - This difference between carveme and curated might happen because the percentage of gene associated reactions in curated is higher than in carveme (71% vs 69%, respectively; Table 16), which means that one gene will probably be involved in a higher number of reactions in curated than in carveme, thus, once the gene is "knocked-out", it will affect a higher number of reactions.
- The correlation between affected reactions and percentage growth change is high in both models 0.93 and 0.87, in carveme and curated, respectively), meaning that when the number of affected reactions is elevated, it's very likely that the change in growth will be high as well, leading more likely to inhibition of growth.

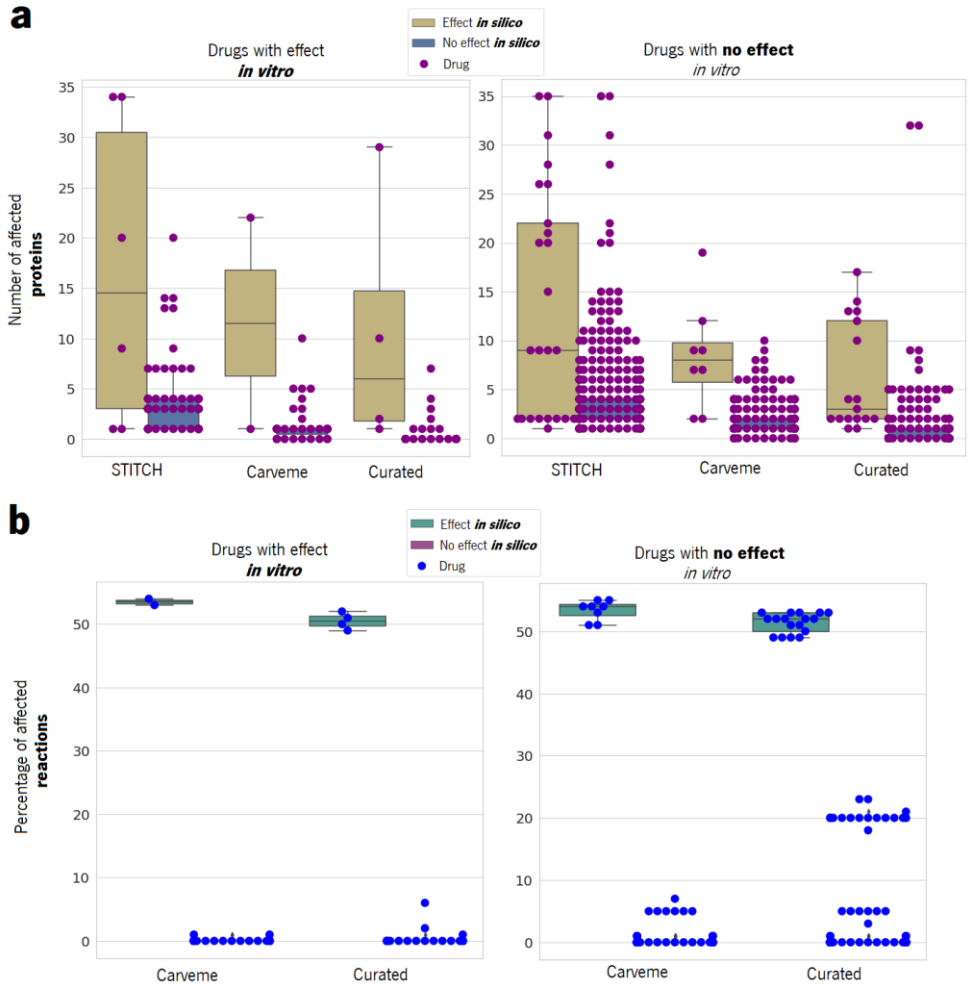


Fig. 31. Distribution of the proteins and reactions affected by the 242 drugs that had STITCH inhibiting information with *B. thetaiotaomicron*'s proteins. **a,** Number of proteins that were affected/"knocked-out" in the carveme and curated models (grouped by drugs that had effect *in silico* [left boxplots] and by the ones that didn't [right boxplots]); figure divided into two panels, in which the left panel regards drugs that had effect *in vitro* and the right one regards drugs that didn't have effect *in vitro*. **b,** Percentage of affected reactions out of each models' total number of reactions (i.e., reactions for which their fluxes' ranges before and after the genes were "knocked-out", were not 100% overlapping); grouping of drugs and scheme of figures is the same as in **a**.

4.3.4 Non-Commensal Targeting Drugs Influence Neuroactive Metabolism

Taking a look at the drug effects on the bacterial neuroactive metabolism, and restricting to the drugs that had an effect *in vitro* and in at least one of the models (streptozotocin, furazolidone, ofloxacin and triclosan):

- these drugs affected the production/consumption of the same neuroactive metabolites (acetate, S-adenosylmethionine, glutamate, menaquinone, propionate and tryptophan), in both models, except that in carveme additionally inositol was affected and in curated quinolinic acid (Fig. 32).

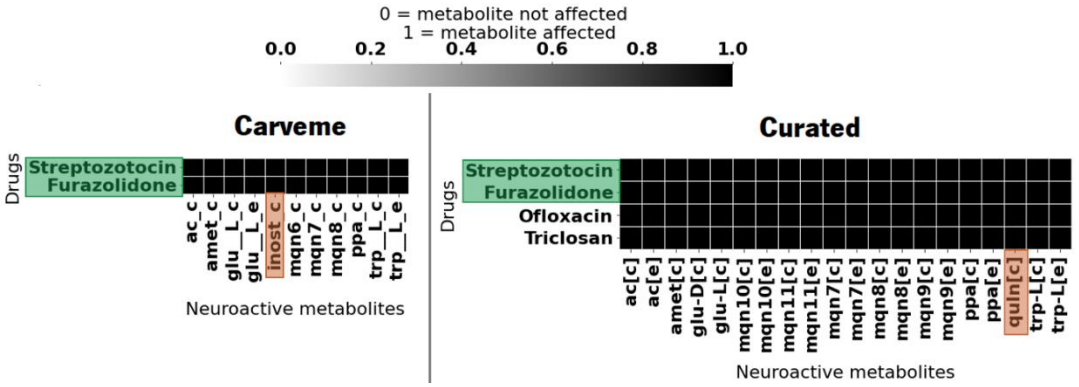


Fig. 32. Neuroactive metabolites that might be possibly affected by drugs, out of the 242, that inhibited *in vitro* and in at least one of the *in silico* models (heatmap with black colour representing a metabolite that was affected by a drug and with white colour representing the opposite; the drugs that are coloured are the ones that affect in both models, carveme and curated; the coloured metabolites are the ones that differ between the models).

5 Discussion

The *in vitro* results of the Maier *et al* 2018 research [3], showed undesirable side effects of drugs that aren't supposed to target gut commensal bacteria, on the gut microbiome composition. In particular, there was an overrepresentation of neuroactive drugs among the human-targeted drugs that had effects of bacterial growth [3]. However, the mechanisms of the drug inhibition of bacterial growth remain largely unknown. In this study, we aimed at identifying potential mechanisms of drug-bacteria interactions in silico using GSMM and flux balance analysis combined with the information on drug-protein interactions from the STITCH database.

In summary, the *in silico* results were the following:

- out of the 198 neuroactive drugs screened *in vitro*:
  - 41 had inhibiting interaction information, in STITCH, with proteins of *B. thetaiotaomicron* and, out of these:
    - none had effect on the growth of *B. thetaiotaomicron in vitro.*
    - 3 inhibited growth in carveme model (acetaminophen, gabapentin and vigabatrin); affected essential genes "BT_1806", that is not present in curated, and "BT_3935", that is present but is not essential in curated.
    - 4 different drugs inhibited in curated (carbamazepine, valproic acid, lamotrigine and topiramate); affected essential genes "BT_1225" and "BT_0382", that do not exist in carveme.
- out of the 999 non-commensal targeted drugs screened *in vitro*:
  - 242 had inhibiting interaction information with proteins of *B. thetaiotaomicron*:
    - 41 inhibited growth *in vitro* and of these*:*
      - 2 also had an effect in both *in silico* models; one of them affected the gene "BT_0347" that is essential in both models.
      - 2 had effect in the curated model only; affected the non-essential genes "BT_2048" and "BT_0106". The latter was not present in carveme.
    - 201 didn't inhibit growth *in vitro* but:
      - 8 in carveme; all affected essential genes, 6 in total, 1 essential in curated as well, 3 exist in curated but are not essential and 2 don't exist.
      - 17 had an effect on growth in curated; 8 drugs targeted essential genes, 12 in total, 5 of which exist in carveme but are not essential.
      - 5 drugs inhibited growth in both models.

Overall, combining experimental results with *in silico* ones, there are some hypotheses one can make:

- the drug interactions with the affected essential genes, in both models, were erroneous [90].
- LB medium might not be completely representative of mGAM medium, which can lead to different essential genes for *B. thetaiotaomicron*'s growth (as one can see in Table 16) . Thus, the drug-protein interactions from STITCH might occur, but not lead to inhibition of growth.
- the GPR rules might be incorrect or incomplete, since they are generated during the model reconstruction process, hence possibly leading to incorrect predictions of essential genes.
- these genes can be essential, but the dose used in the experiment might have been too small to have had an effect on bacterium's growth.

Table 16. Summary of carveme and of curated models, not only for LB medium, but also for complete and M9 media (fluxes have the unit mmol / [gDW h]).

	CarveMe model			Curated model		
Reconstruction approach	"top-down"			"bottom-up"		
Biomass equation - species	*Escherichia coli* [60]					
Medium composition - species	Complete *Mycoplasma genitalium* [60]	LB *Bacillus subtilis* and *Shewanella oneidensis* [60]	M9 *E. coli* [60]	Complete (default)	LB (from LB carveme models)	M9 (from M9 carveme models)
Growth rate	55,0	1,4	0,9	86,8	5,7	1,7
No. of proteins in FASTA	4636			4636		
No. of genes/proteins in model	832			993		
Essential genes	40	67 (8%)	120	61	124 (12%)	200
Total no. of reactions	2151	2154	2155	1528		
Universally blocked reactions	55			89		
Exchange reactions	253			280		
Exchange reactions with active flux	253	50	17	280	49	21
Exchange reactions with active flux – absolute fluxes' mean	23,1	2,4	5,2	73,0	22,2	44,8
Exchange reactions bounds	(-1000,1000)	active (-10,1000) | non-active (0,1000)	active (-10,1000) | non-active (0,1000)	(-1000,1000)	active (-10,1000) | non-active (0,1000)	active (-10,1000) | non-active (0,1000)
Medium specific blocked reactions	55	685 (32%)	760	89	530 (35%)	540
Zero flux reactions	1825	1833	1774	945	843	920
Transport reactions	637			87		
Purely metabolic reactions	1255	1258 (58%)	1259	1154 (75%)		

Purely metabolic reactions with non-zero flux	180	212	311	465	582	554
Purely metabolic reactions with non-zero flux - absolute fluxes' mean	68,5	2,4	1,0	109,8	17,5	22,9
Gene associated reactions (enzymatic reactions)	1489 (69%)			1084 (71%)		
Gene associated reactions that are purely metabolic	1047 (49%)			1034 (68%)		

Altogether, the results show that:

- carveme and curated models do not have the same essential genes in LB medium (67 vs 124, respectively; Table 16), which means that, to be able to rely on the *in silico* results, there is still a need to do more experimental research on *B. thetaiotaomicron*'s, in order to find the most accurate biomass composition, that, consequently, is going to change gene essentiality.
- both *in silico* models are missing genes that have interactions in STITCH (37 and 40, respectively; Fig. 24c), which means that some reactions in the models are not being affected. These reactions could lead to a change in bacterium's growth, thus leading to even more drugs inhibiting growth.
- carveme and curated models differ in the genes present (as shown in Fig. 24a and Fig. 28b and Table 16 – 832 vs 993), in the number of blocked reactions (685 vs 530) and in the number of gene associated reactions (1489 vs 1084, Table 16). Therefore, when one drug affects one gene in one model, it affects certain reactions that might not be affected in the other model (which is further supported by the difference in the drugs' effect on the neuroactive metabolism, since it affects the production/consumption of different neuroactive metabolites, as shown in Fig. 25, 26 and 32).
- It is necessary to do more upstream perturbations experiments in order to discover more drug-protein inhibiting interactions and, subsequently, be able to perform more accurate *in silico* predictions.

6 Conclusions

In the present work, an automatic genome-scale metabolic reconstruction of *B. thetaiotaomicron* was generated and subsequently used to systematically investigate drugs' effect on its growth and its metabolism and, moreover, its results were compared with the ones from a manually curated model reconstruction. The drugs have been previously screened *in vitro* in the Maier *et al* 2018 research paper. These drugs were selected for two different reasons:

- to understand why neuroactive drugs had an effect on bacteria, given that they are supposed to target receptors that don't exist in bacteria.
- to examine why medication that is not targeted at gut commensals, affects them.

The goal was to evaluate if the drugs had corresponding results *in vitro* and *in silico* and, if so, be able to examine the proteins that are being affected.

Neither of the *in silico* models predicted well the drugs' effect on the bacterium's growth. These results suggest that:

- the models are inaccurate, particularly due to probable incomplete and/or incorrect biomass composition and possibly leading to wrong gene essentiality predictions.
- some drug-protein interaction information retrieved from STITCH database was erroneous because they were transferred via orthology from other organisms.
- the medium used in the *in silico* models was not representative of the one used in the *in vitro* screen experiment.

Since Maier *et al.* 2018 paper [3] monitored the growth of 40 bacterial strains. In the future work, it would be important to extend the analysis performed in this study for *B. thetaiotaomicron* to the other 39 bacteria. In this way, we can test whether drug effects can be predicted better for other species and compare *in silico* predictions for drug effects between different species.

References

1. Flowers, S.A., Evans, S.J., Ward, K.M., McInnis, M.G., Ellingrod, V.L.: Interaction Between Atypical Antipsychotics and the Gut Microbiome in a Bipolar Disease Cohort. Pharmacotherapy. 37, 261–267 (2017).

2. Schmidt, T.S.B., Raes, J., Bork, P.: The Human Gut Microbiome: From Association to Modulation. Cell. 172, 1198–1215 (2018).

3. Maier, L., Pruteanu, M., Kuhn, M., Zeller, G., Telzerow, A., Anderson, E.E., Brochado, A.R., Fernandez, K.C., Dose, H., Mori, H., Patil, K.R., Bork, P., Typas, A.: Extensive impact of non-antibiotic drugs on human gut bacteria. Nature. 555, 623–628 (2018).

4. Esvap, E., Ulgen, K.O.: Advances in Genome-Scale Metabolic Modeling toward Microbial Community Analysis of the Human Microbiome. ACS Synth. Biol. 10, 2121–2137 (2021).

5. Lamichhane, S., Sen, P., Dickens, A.M., Orešič, M., Bertram, H.C.: Gut metabolome meets microbiome: A methodological perspective to understand the relationship between host and microbe. Methods. 149, 3–12 (2018).

6. van der Ark, K.C.H., van Heck, R.G.A., Martins Dos Santos, V.A.P., Belzer, C., de Vos, W.M.: More than just a gut feeling: constraint-based genome-scale metabolic models for predicting functions of human intestinal microbes. Microbiome. 5, 78 (2017).

7. Hornung, B., Martins dos Santos, V.A.P., Smidt, H., Schaap, P.J.: Studying microbial functionality within the gut ecosystem by systems biology. Genes Nutr. 13, 1–19 (2018).

8. Saa, P., Urrutia, A., Silva-Andrade, C., Martín, A.J., Garrido, D.: Modeling approaches for probing cross-feeding interactions in the human gut microbiome. Comput. Struct. Biotechnol. J. 20, 79–89 (2022).

9. Kumar, M., Ji, B., Babaei, P., Das, P., Lappa, D., Ramakrishnan, G., Fox, T.E., Haque, R., Petri, W.A., Bäckhed, F., Nielsen, J.: Gut microbiota dysbiosis is associated with malnutrition and reduced plasma amino acid levels: Lessons from genome-scale metabolic modeling. Metab. Eng. 49, 128–142 (2018).

10. Gacesa, R., Kurilshikov, A., Vich Vila, A., Sinha, T., Klaassen, M.A.Y., Bolte, L.A., Andreu-Sánchez, S., Chen, L., Collij, V., Hu, S., Dekens, J.A.M., Lenters, V.C., Björk, J.R., Swarte, J.C., Swertz, M.A., Jansen, B.H., Gelderloos-Arends, J., Jankipersadsing, S., Hofker, M., Vermeulen, R.C.H., Sanna, S., Harmsen, H.J.M., Wijmenga, C., Fu, J., Zhernakova, A., Weersma, R.K.: Environmental factors shaping the gut microbiome in a Dutch population. Nature. 604, 732–739 (2022).

11. Wilmanski, T., Diener, C., Rappaport, N., Patwardhan, S., Wiedrick, J., Lapidus, J., Earls, J.C., Zimmer, A., Glusman, G., Robinson, M., Yurkovich, J.T., Kado, D.M., Cauley, J.A., Zmuda, J., Lane, N.E., Magis, A.T., Lovejoy, J.C., Hood, L., Gibbons, S.M., Orwoll, E.S., Price, N.D.: Gut microbiome pattern reflects healthy ageing and predicts survival in humans. Nat. Metab. 3, 274–286 (2021).

12. Chakrabarti, A., Geurts, L., Hoyles, L., Iozzo, P., Kraneveld, A.D., La Fata, G., Miani, M., Patterson, E., Pot, B., Shortt, C., Vauzour, D.: The microbiota–gut–brain axis: pathways to better brain health.

Perspectives on what we know, what we need to investigate and how to put knowledge into practice. Cell. Mol. Life Sci. 79, 1–15 (2022).

13. Devika, N.T., Raman, K.: Deciphering the metabolic capabilities of Bifidobacteria using genome-scale metabolic models. Sci. Rep. 9, 1–9 (2019).

14. What are the parts of the nervous system?, https://www.nichd.nih.gov/health/topics/neuro/conditioninfo/parts.

15. Hagan, C.E., Bolon, B., Keene, C.D.: Nervous System. Elsevier Inc. (2012).

16. Vagus Nerve: Function, Stimulation, and More, https://www.healthline.com/human-body-maps/vagus-nerve#anatomy-and-function, last accessed 2022/10/25.

17. The Vagus Nerve, https://teachmeanatomy.info/head/cranial-nerves/vagus-nerve-cn-x/, last accessed 2022/10/25.

18. Pavlov, V.A., Chavan, S.S., Tracey, K.J.: Molecular and Functional Neuroscience in Immunity. Annu. Rev. Immunol. 36, 783 (2018).

19. Bradley, P.B.: The somatic motor system. In: Introduction to Neuropharmacology. pp. 35–42. Butterworth-Heinemann (1989).

20. Fleming, M.A., Ehsan, L., Moore, S.R., Levin, D.E.: The Enteric Nervous System and Its Emerging Role as a Therapeutic Target. Gastroenterol. Res. Pract. 2020, (2020).

21. Cook, T.M., Mansuy-Aubert, V.: Communication between the gut microbiota and peripheral nervous system in health and chronic disease. Gut Microbes. 14, 1–20 (2022).

22. Ahmed, H., Leyrolle, Q., Koistinen, V., Kärkkäinen, O., Layé, S., Delzenne, N., Hanhineva, K.: Microbiota-derived metabolites as drivers of gut–brain communication. Gut Microbes. 14, (2022).

23. Luan, H., Wang, X., Cai, Z.: Mass spectrometry-based metabolomics: Targeting the crosstalk between gut microbiota and brain in neurodegenerative disorders. Mass Spectrom. Rev. 38, 22–33 (2019).

24. Spichak, S., Bastiaanssen, T.F.S., Berding, K., Vlckova, K., Clarke, G., Dinan, T.G., Cryan, J.F.: Mining microbes for mental health: Determining the role of microbial metabolic pathways in human brain health and disease. Neurosci. Biobehav. Rev. 125, 698–761 (2021).

25. Clarke, G., Grenham, S., Scully, P., Fitzgerald, P., Moloney, R.D., Shanahan, F., Dinan, T.G., Cryan, J.F.: The microbiome-gut-brain axis during early life regulates the hippocampal serotonergic system in a sex-dependent manner. Mol. Psychiatry. 18, 666–673 (2013).

26. Lai, Y., Liu, C.W., Yang, Y., Hsiao, Y.C., Ru, H., Lu, K.: High-coverage metabolomics uncovers microbiota-driven biochemical landscape of interorgan transport and gut-brain communication in mice. Nat. Commun. 12, 1–16 (2021).

27. Tan, J., McKenzie, C., Potamitis, M., Thorburn, A.N., Mackay, C.R., Macia, L.: The Role of Short-Chain Fatty Acids in Health and Disease. Elsevier Inc. (2014).

28. Silva, Y.P., Bernardi, A., Frozza, R.L.: The Role of Short-Chain Fatty Acids From Gut Microbiota in Gut-Brain Communication. Front. Endocrinol. (Lausanne). 11, 25 (2020).

29. Dinan, T.G., Cryan, J.F.: Brain-Gut-Microbiota Axis and Mental Health. Psychosom. Med. 79, 920–926 (2017).

30. Valles-Colomer, M., Falony, G., Darzi, Y., Tigchelaar, E.F., Wang, J., Tito, R.Y., Schiweck, C., Kurilshikov, A., Joossens, M., Wijmenga, C., Claes, S., Van Oudenhove, L., Zhernakova, A., Vieira-Silva, S., Raes, J.: The neuroactive potential of the human gut microbiota in quality of life and depression. Nat. Microbiol. 4, 623–632 (2019).

31. Suda, K., Matsuda, K.: How Microbes Affect Depression: Underlying Mechanisms via the Gut–Brain Axis and the Modulating Role of Probiotics. Int. J. Mol. Sci. 23, 1–17 (2022).

32. Tran, S.M.S., Hasan Mohajeri, M.: The role of gut bacterial metabolites in brain development, aging and disease. Nutrients. 13, 1–41 (2021).

33. Caspani, G., Kennedy, S., Foster, J.A., Swann, J.: Gut microbial metabolites in depression: Understanding the biochemical mechanisms. Microb. Cell. 6, 454–481 (2019).

34. López-Gambero, A.J., Sanjuan, C., Serrano-Castro, P.J., Suárez, J., Fonseca, F.R. De: The Biomedical Uses of Inositols: A Nutraceutical Approach to Metabolic Dysfunction in Aging and Neurodegenerative Diseases. Biomedicines. 8, (2020).

35. Maier, L., Typas, A.: Systematically investigating the impact of medication on the gut microbiome. Curr. Opin. Microbiol. 39, 128–135 (2017).

36. Becattini, S., Taur, Y., Pamer, E.G.: Antibiotic-Induced Changes in the Intestinal Microbiota and Disease. Trends Mol. Med. 22, 458–478 (2016).

37. Westervelt, P., Cho, K., Bright, D.R., Kisor, D.F.: Drug–gene interactions: Inherent variability in drug maintenance dose requirements. P T. 39, 630–637 (2014).

38. Gilbert, J.A., Blaser, M.J., Caporaso, J.G., Jansson, J.K., Lynch, S. V., Knight, R.: Current understanding of the human microbiome. Nat. Med. 24, 392 (2018).

39. Shaked, I., Oberhardt, M.A., Atias, N., Sharan, R., Ruppin, E.: Metabolic Network Prediction of Drug Side Effects. Cell Syst. 2, 209–213 (2016).

40. Pharos : Illuminating the Druggable Genome, https://pharos.nih.gov/, last accessed 2022/10/30.

41. Wishart, D.S., Feunang, Y.D., Guo, A.C., Lo, E.J., Marcu, A., Grant, J.R., Sajed, T., Johnson, D., Li, C., Sayeeda, Z., Assempour, N., Iynkkaran, I., Liu, Y., Maclejewski, A., Gale, N., Wilson, A., Chin, L., Cummings, R., Le, Di., Pon, A., Knox, C., Wilson, M.: DrugBank 5.0: A major update to the DrugBank database for 2018. Nucleic Acids Res. 46, D1074–D1082 (2018).

42. BioGRID | Database of Protein, Chemical, and Genetic Interactions, https://thebiogrid.org/, last accessed 2022/10/30.

43. Zhou, Y., Zhang, Y., Lian, X., Li, F., Wang, C., Zhu, F., Qiu, Y., Chen, Y.: Therapeutic target database update 2022: Facilitating drug discovery with enriched comparative data of targeted agents. Nucleic Acids Res. 50, D1398–D1407 (2022).

44. Szklarczyk, D., Santos, A., Von Mering, C., Jensen, L.J., Bork, P., Kuhn, M.: STITCH 5: augmenting protein-chemical interaction networks with tissue and affinity data. Nucleic Acids Res. 44, D380–

D384 (2016).

45. Mottini, C., Napolitano, F., Li, Z., Gao, X., Cardone, L.: Computer-aided drug repurposing for cancer therapy: Approaches and opportunities to challenge anticancer targets. Semin. Cancer Biol. 68, 59–74 (2021).

46. STITCH: chemical association networks, http://stitch.embl.de/cgi/input.pl?UserId=zaZX8Q1vNvhF&sessionId=VuqqdWbWsMj9.

47. von Mering, C., Jensen, L.J., Snel, B., Hooper, S.D., Krupp, M., Foglierini, M., Jouffre, N., Huynen, M.A., Bork, P.: STRING: known and predicted protein-protein associations, integrated and transferred across organisms. Nucleic Acids Res. 33, (2005).

48. Sen, P., Orešič, M.: Metabolic modeling of human gut microbiota on a genome scale: An overview. Metabolites. 9, (2019).

49. Hale, V.L., Jeraldo, P., Mundy, M., Yao, J., Keeney, G., Scott, N., Cheek, E.H., Davidson, J., Green, M., Martinez, C., Lehman, J., Pettry, C., Reed, E., Lyke, K., White, B.A., Diener, C., Resendis-Antonio, O., Gransee, J., Dutta, T., Petterson, X.M., Boardman, L., Larson, D., Nelson, H., Chia, N.: Synthesis of multi-omic data and community metabolic models reveals insights into the role of hydrogen sulfide in colon cancer. Methods. 149, 59–68 (2018).

50. Magnúsdóttir, S., Heinken, A., Kutt, L., Ravcheev, D.A., Bauer, E., Noronha, A., Greenhalgh, K., Jäger, C., Baginska, J., Wilmes, P., Fleming, R.M.T., Thiele, I.: Generation of genome-scale metabolic reconstructions for 773 members of the human gut microbiota. Nat. Biotechnol. 35, 81–89 (2017).

51. O'Brien, E.J., Monk, J.M., Palsson, B.O.: Using genome-scale models to predict biological capabilities. Cell. 161, 971–987 (2015).

52. Virtual Metabolic Human, https://www.vmh.life/.

53. Thiele, I., Palsson, B.: A protocol for generating a high-quality genome-scale metabolic reconstruction. Nat. Protoc. 5, 93–121 (2010).

54. Filippoid, M. Di, Damianiid, C., Pesciniid, D.: GPRuler: Metabolic gene-protein-reaction rules automatic reconstruction. (2021).

55. PSORT, https://www.psort.org/.

56. Yu, N.Y., Wagner, J.R., Laird, M.R., Melli, G., Rey, S., Lo, R., Dao, P., Cenk Sahinalp, S., Ester, M., Foster, L.J., Brinkman, F.S.L.: PSORTb 3.0: improved protein subcellular localization prediction with refined localization subcategories and predictive capabilities for all prokaryotes. Bioinformatics. 26, 1608–1615 (2010).

57. Proteome Analyst, http://pa.wishartlab.com/pa/pa/index.html.

58. Szafron, D., Lu, P., Greiner, R., Wishart, D.S., Poulin, B., Eisner, R., Lu, Z., Anvik, J., Macdonell, C., Fyshe, A., Meeuwis, D.: Proteome Analyst: custom predictions with explanations in a web-based tool for high-throughput proteome annotations. Nucleic Acids Res. 32, W365 (2004).

59. Bernstein, D.B., Sulheim, S., Almaas, E., Segrè, D.: Addressing uncertainty in genome-scale metabolic model reconstruction and analysis. Genome Biol. 22, 1–22 (2021).

60. Machado, D., Andrejev, S., Tramontano, M., Patil, K.R.: Fast automated reconstruction of genome-scale metabolic models for microbial species and communities. Nucleic Acids Res. 46, 7542–7553 (2018).

61. Passi, A., Tibocha-Bonilla, J.D., Kumar, M., Tec-Campos, D., Zengler, K., Zuniga, C.: Genome-scale metabolic modeling enables in-depth understanding of big data. Metabolites. 12, (2022).

62. Henry, C.S., Dejongh, M., Best, A.A., Frybarger, P.M., Linsay, B., Stevens, R.L.: High-throughput generation, optimization and analysis of genome-scale metabolic models. Nat. Biotechnol. 28, 977–982 (2010).

63. Aziz, R.K., Bartels, D., Best, A., DeJongh, M., Disz, T., Edwards, R.A., Formsma, K., Gerdes, S., Glass, E.M., Kubal, M., Meyer, F., Olsen, G.J., Olson, R., Osterman, A.L., Overbeek, R.A., McNeil, L.K., Paarmann, D., Paczian, T., Parrello, B., Pusch, G.D., Reich, C., Stevens, R., Vassieva, O., Vonstein, V., Wilke, A., Zagnitko, O.: The RAST Server: rapid annotations using subsystems technology. BMC Genomics. 9, (2008).

64. Overbeek, R., Begley, T., Butler, R.M., Choudhuri, J. V., Chuang, H.Y., Cohoon, M., de Crécy-Lagard, V., Diaz, N., Disz, T., Edwards, R., Fonstein, M., Frank, E.D., Gerdes, S., Glass, E.M., Goesmann, A., Hanson, A., Iwata-Reuyl, D., Jensen, R., Jamshidi, N., Krause, L., Kubal, M., Larsen, N., Linke, B., McHardy, A.C., Meyer, F., Neuweger, H., Olsen, G., Olson, R., Osterman, A., Portnoy, V., Pusch, G.D., Rodionov, D.A., Rückert, C., Steiner, J., Stevens, R., Thiele, I., Vassieva, O., Ye, Y., Zagnitko, O., Vonstein, V.: The subsystems approach to genome annotation and its use in the project to annotate 1000 genomes. Nucleic Acids Res. 33, 5691–5702 (2005).

65. Kanehisa, M., Furumichi, M., Sato, Y., Kawashima, M., Ishiguro-Watanabe, M.: KEGG for taxonomy-based analysis of pathways and genomes. Nucleic Acids Res. 1, 13–14 (2013).

66. Kumar, V.S., Maranas, C.D.: GrowMatch: An Automated Method for Reconciling In Silico/In Vivo Growth Predictions. PLOS Comput. Biol. 5, e1000308 (2009).

67. Heinken, A., Ravcheev, D.A., Baldini, F., Heirendt, L., Fleming, R.M.T., Thiele, I.: Systematic assessment of secondary bile acid metabolism in gut microbes reveals distinct metabolic capabilities in inflammatory bowel disease. Microbiome. 7, 1–18 (2019).

68. Zorrilla, F., Buric, F., Patil, K.R., Zelezniak, A.: MetaGEM: Reconstruction of genome scale metabolic models directly from metagenomes. Nucleic Acids Res. 49, (2021).

69. Fang, X., Monk, J.M., Mih, N., Du, B., Sastry, A. V., Kavvas, E., Seif, Y., Smarr, L., Palsson, B.O.: Escherichia coli B2 strains prevalent in inflammatory bowel disease patients have distinct metabolic capabilities that enable colonization of intestinal mucosa. BMC Syst. Biol. 12, 1–10 (2018).

70. Baldini, F., Hertel, J., Sandt, E., Thinnes, C.C., Neuberger-Castillo, L., Pavelka, L., Betsou, F., Krüger, R., Thiele, I.: Parkinson's disease-associated alterations of the gut microbiome predict disease-relevant changes in metabolic functions. BMC Biol. 18, 1–21 (2020).

71. Patumcharoenpol, P., Nakphaichit, M., Panagiotou, G., Senavonge, A., Suratannon, N., Vongsangnak, W.: MetGEMs Toolbox: Metagenome-scale models as integrative toolbox for uncovering metabolic functions and routes of human gut microbiome. PLoS Comput. Biol. 17, 1–18 (2021).

72. Btheta - Genome - Assembly - NCBI, https://www.ncbi.nlm.nih.gov/assembly/GCF_000011065.1#/qa.

73. Ryan, D., Jenniches, L., Reichardt, S., Barquist, L., Westermann, A.J.: A high-resolution transcriptome map identifies small RNA regulation of metabolism in the gut microbe Bacteroides thetaiotaomicron. Nat. Commun. 11, (2020).

74. Mohsin, M., Tanaka, K., Kawahara, R., Kondo, S., Noguchi, H., Motooka, D., Nakamura, S., Khong, D.T., Nguyen, T.N., Hoang, T.N., Yamamoto, Y.: Whole-genome sequencing and comparative analysis of the genomes of Bacteroides thetaiotaomicron and Escherichia coli isolated from a healthy resident in Vietnam. J. Glob. Antimicrob. Resist. 21, 65–67 (2020).

75. KEGG BRITE: Anatomical Therapeutic Chemical (ATC) Classification, https://www.genome.jp/brite/br08303.

76. PubChem, https://pubchem.ncbi.nlm.nih.gov/.

77. Kuhn, M., Szklarczyk, D., Franceschini, A., Von Mering, C., Jensen, L.J., Bork, P.: STITCH 3: zooming in on protein-chemical interactions. Nucleic Acids Res. 40, (2012).

78. Deghou, S., Zeller, G., Iskar, M., Driessen, M., Castillo, M., Van Noort, V., Bork, P.: CART—a chemical annotation retrieval toolkit. Bioinformatics. 32, 2869–2871 (2016).

79. Ebrahim, A., Lerman, J.A., Palsson, B.O., Hyduke, D.R.: COBRApy: COnstraints-Based Reconstruction and Analysis for Python. BMC Syst. Biol. 7, 1–6 (2013).

80. GitHub - cdanielmachado/reframed, https://github.com/cdanielmachado/reframed.

81. ReFramed, https://reframed.readthedocs.io/en/latest/.

82. Heinken, A., Sahoo, S., Fleming, R.M.T., Thiele, I.: Systems-level characterization of a host-microbe metabolic symbiosis in the mammalian gut. Gut Microbes. 4, 28–40 (2013).

83. Henry, C.S., Dejongh, M., Best, A.A., Frybarger, P.M., Linsay, B., Stevens, R.L.: High-throughput generation, optimization and analysis of genome-scale metabolic models. Nat. Biotechnol. 2010 289. 28, 977–982 (2010).

84. In silico Reconstructions | ThieleLab, https://www.thielelab.eu/in-silico-models.

85. Gifu Anaerobic Broth, Modified (GAM) Composition, https://himedialabs.com/TD/M2079.pdf.

86. Silhavy, T.J., Kahne, D., Walker, S.: The Bacterial Cell Envelope. Cold Spring Harb. Perspect. Biol. 2, (2010).

87. Wang, X., Quinn, P.J., Yan, A.: Kdo2-lipid A: structural diversity and impact on immunopharmacology. Biol. Rev. Camb. Philos. Soc. 90, 408 (2015).

88. Vollmer, W., Bertsche, U.: Murein (peptidoglycan) structure, architecture and biosynthesis in Escherichia coli. Biochim. Biophys. Acta - Biomembr. 1778, 1714–1734 (2008).

89. Bogdanov, M., Pyrshev, K., Yesylevskyy, S., Ryabichko, S., Boiko, V., Ivanchenko, P., Kiyamova, R., Guan, Z., Ramseyer, C., Dowhan, W.: Phospholipid distribution in the cytoplasmic membrane of Gram-negative bacteria is highly asymmetric, dynamic, and cell shape-dependent. Sci. Adv. 6,

(2020).

90. KEGG: Kyoto Encyclopedia of Genes and Genomes, https://www.genome.jp/kegg/.

Supplementary Information

Supplementary Table 1. Prestwick ID/STITCH ID/Drug name matching of the drugs (out of the 242 that had, in STITCH database, inhibiting interactions with B. thetaiotaomicron's proteins) that had effect either in vitro, in silico (in carveme and curated models) or in both.

PrestwickID	STITCHID	Drugname	PrestwickID	STITCHID	Drugname
1109	05361912	Rifabutin	708	00002335	Benzethonium chloride
1233	06333887	Auranofin	267	00002812	Clotrimazole
525	05381226	Rifampicin	376	00002794	Clofazimine
151	00003255	Erythromycin	740	00003194	Ebselen
808	00003435	Furazolidone	1203	00003385	5-fluorouracil
113	00002764	Ciprofloxacin hydrochloride monohydrate	487	00002958	Daunorubicin hydrochloride
208	00005578	Trimethoprim	205	00610479	Tolfenamic acid
1415	00003363	Floxuridine	368	00002351	Bepridil hydrochloride
1157	06323497	Rifapentine	126	00004046	Mefloquine hydrochloride
1446	00004259	Moxifloxacin	478	00003333	Felodipine
1265	00005379	Gatifloxacin	1114	00060787	Saquinavir mesylate
238	00003948	Lomefloxacin hydrochloride	736	00002617	Cefazolin sodium salt
237	00004583	Ofloxacin	1467	00005591	Troglitazone
1343	00005257	Sparfloxacin	1314	00004829	Pioglitazone
766	00005479	Tinidazole	94	00002265	Azathioprine
756	00003054	Diethylstilbestrol	1337	00077998	Rosiglitazone Hydrochloride
1401	00003229	Enoxacin	105	00002949	Danazol
37	00004993	Pyrimethamine	275	00003339	Fenofibrate
1056	00006256	Trifluridine	1097/911	00003779	(-)-Isoproterenol hydrochloride / (+)-Isoproterenol (+)-bitartrate salt
1194	00005441	Thimerosal	1134	00000596	Cytarabine
333	00005726	Zidovudine, AZT	14	00005320	Sulfacetamide sodic hydrate
168	00004509	Nitrofurantoin	257/256	00005538	Retinoic acid / Isotretinoin
1479	00005564	Triclosan	1210	00002088	Alendronate sodium
1378	00054688	Clarithromycin	741	00000925	Nadide
699	00003606	Hexestrol	1198	00005527	Tranilast
1	00008646	Azaguanine-8	1285	00060852	Ibandronate sodium
1303	00051081	Pefloxacine	1118	00002622	Cefepime hydrochloride
370	00002333	Benzbromarone	489	00002650	Ceftazidime pentahydrate
390	00003443	Fusidic acid sodium salt	441	00000450	Estradiol-17 beta
732	00005300	Streptozotocin			