

From genome to function: metabolism-wide models of lifecycle stages of parasites

Bruno R. Pinto^{1*}, Gabriela T. Montanaro^{1*}, Mayke B. Alencar¹, Milou Hoving², Jurgen R. Haanstra², Ariel M. Silber¹ and Pranas Grigaitis^{2, #}

¹ Laboratory of Biochemistry of Trypanosomatids – LaBTryps, Department of Parasitology, Institute of Biomedical Science II – ICB II, University of São Paulo – USP. Av. Prof. Lineu Prestes, 1374 – Butantã – São Paulo – SP, Brazil;

² Systems Biology Lab, A-LIFE & AIMMS, Vrije Universiteit Amsterdam, De Boelelaan 1085, 1081 HZ Amsterdam, the Netherlands

* Authors contributed equally to this work.

Correspondence to: p.grigaitis@vu.nl

Abstract

Metabolism is a complex network of biochemical reactions that cells use to obtain free energy transduced from nutrients and synthesize new cellular components. Understanding the wiring of metabolism can aid in metabolic engineering to boost certain metabolic functions, or, alternatively, inspire interventions to block undesired metabolism, e.g. of parasites. The challenge is that interventions to metabolic networks should be studied in the context of the entire, extensive cellular metabolic networks. Incoming and outgoing fluxes may be readily measured but how fluxes are distributed within the metabolic network often remains a black box. Systems biology offers a toolset to shine light inside this black box: constraint-based metabolic models can predict intracellular flux distributions. But how to construct such a model? In this article we will guide you through the steps to make a constraint-based model out of the genome sequence and demonstrate that it may be easier than it sounds. We provide a walkthrough of model construction, curation and validation - with a specific focus on protistan parasites and their lifecycle stages. While genomes can help to infer the entirety of metabolic potential encoded in the organism's genome (a so-called genome-scale metabolic model), cells do not necessarily express all the metabolic enzymes simultaneously. This is particularly relevant for parasites as the different lifecycle stages rely on different sets of metabolic enzymes and host-imposed constraints. We discuss approaches to restrict the genome-wide network to become lifecycle-stage specific and how you can test and use the resulting models.

1. General concepts and background

To grow and function, cells need energy and building blocks for new cells. The precursors of the new cell material can be either taken up from the surroundings, or made by the cell from other molecules. In parallel, the energy is harvested from energy-rich nutrients (e.g. sugars, amino acids, or fats), and stored in the high-energy chemical bonds of adenosine triphosphate (ATP). Metabolic waste that is generated in these processes needs to be excreted from the cell. Thus, the chemistry of life forms large networks of conversion reactions, collectively called metabolism [1].

Molecular (reductionist) approaches in biology typically focus on deciphering the function of a single or small group of genes or proteins, and thus require extensive experimental work. Although this foundational knowledge is essential, it provides an incomplete picture as proteins do not function in isolation, but rather are embedded in large networks. Systems biology is an interdisciplinary field focused on understanding biological systems as a totality: an integrative approach to life. Systems biology considers organisms as complex networks made of interconnected components, with interactions and feedback loops essential to their operation. These interconnections give rise to emergent behaviors and so complicate assessing the behavior of the system as a whole.

The logic of such networks (or significant parts of them) can be encoded into computer models that greatly facilitate understanding of their function. Broadly speaking, modelling approaches to cellular networks fall into three categories (and of course combinations thereof): interaction-, mechanism- and constraint-based models [2, 3]. Interaction-based models catalogue the interactions between proteins in a network without considering network parameters, such as stoichiometry. Such models can, e.g., reveal functional modules within a network. However, when focusing on more dynamic networks, such as metabolism, the other two types of models are more relevant.

Kinetic models are mechanism-based models formulated as a system of ordinary differential equations for metabolite concentrations over time. The individual reactions are described by rate equations that can follow either simple mass action relations or more elaborate kinetics. For enzymes, these rate equations are often Michaelis-Menten type of equations with parameters for maximal velocities of the reaction and affinities of the enzymes for substrates (and products where applicable). Each rate equation often has multiple parameters that need to be experimentally determined or fitted based on pathway-wide flux data; the larger the modeled network, the more parameters values that are needed. Therefore, kinetic models are often made for specific, well-studied pathways rather than metabolism as a whole. Examples for parasites are the glycolysis models of bloodstream-form (BSF) *Trypanosoma brucei* (originally made in 1997 [4]; most recent version in [5]), *Entamoeba histolytica* [6], *Schistosoma mansoni* [7] and *Plasmodium falciparum*-infected red blood cells [8]. These models have been used to predict drug targets and to understand the relevance of allosteric feedforward loops in the regulation of glycolytic flux.

Yet in cells, including those of parasites, distinct pathways like glycolysis do not operate in isolation. Constraint-based models offer a broader, metabolic network-wide view. These models aim to include all metabolic reactions in an organism (and are therefore often named

genome-scale metabolic models (GEMs)); this comes at a trade-off of detail: these models generally only include stoichiometry and reaction rate bounds rather than detailed kinetics. Constraint-based models predict network-wide flux distributions, usually referring to optimization of an objective function (e.g. maximal growth rate) dictated by the constraints put on the model (e.g. substrate availability). GEMs are particularly useful for predicting the network-wide effects of perturbations that affect individual metabolic reactions, such as changes in enzyme availability, nutrient limitation, knockout or inhibition of enzyme function. Capitalizing on that, GEMs have been successfully applied to address a wide range of biological questions, such as development of metabolic strategies for synthesizing valuable compounds, discovering gene functions, and essential metabolites, and identifying new drug targets [9].

This article will focus on modeling parasite metabolism, with specific examples from the Kinetoplastea, which are part of the Euglenozoa. Parasite metabolism is an interesting case: parasites often have different hosts and have host-adjusted life-cycle stages with distinct metabolic networks. As in each lifecycle stage only a subset of the genome-encoded metabolic proteins is expressed, the most relevant models are a restriction compared to the genomic information. Constraint-based models have been made for red blood cell and liver stages of *Plasmodium* species (see for review [10]) and *Trypanosoma cruzi* [11]. These models have been used to compare the metabolic networks of different life-stages and have been used to identify drug targets.

In this article, we will describe step-by-step how metabolism-wide constraint-based models can be made for parasites, focusing especially on the lifecycle stage-specific metabolic networks. We will first briefly describe how an initial model is constructed based on information from the genome. The construction can partly be automated but will also require manual curation. This initial, genome-wide model, however, will have limited relevance for parasites with lifecycle stages. We will discuss the types of data that can be used to generate lifecycle-stage specific models. In each lifecycle stage, a different objective might be needed, and we will discuss different options. At the end, we will describe typical strategies to validate the lifecycle stage-specific models and how you can use them.

2. Reconstruction pipeline

The construction of a metabolic model begins with a fully annotated genome (and detailed data on the biochemical reactions carried out by the organism, where applicable) (Figure 1A and 1B). This should also include non-enzymatic reactions. Essential information of reactions, such as substrates and products (stoichiometry), cofactors, reversibility, and cellular compartmentalization must be specified for each reaction. Mathematically, metabolic models are often expressed as a stoichiometric matrix S , with dimensions $m \times n$, where each row corresponds to a compound, in a system with m metabolites, and each column represents a reaction, in a system with n reactions. The elements in this matrix are stoichiometric coefficients that describe the participation of metabolites in each reaction: positive values indicate production, negative values indicate consumption, and zero means that the metabolite is not part of the reaction [12].

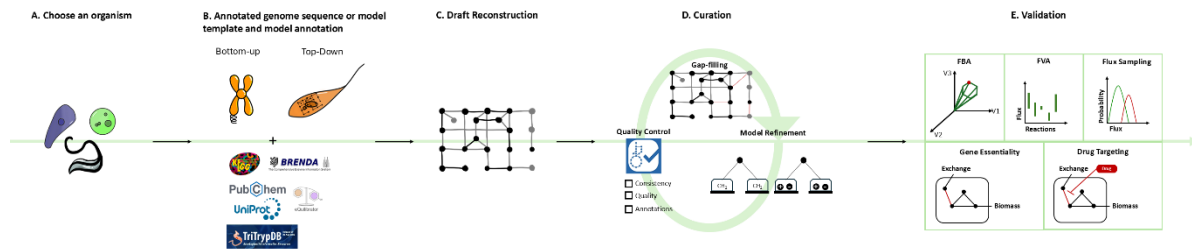


Figure 1. Schematic pipeline of GEM reconstruction. **A.** The first step to reconstruct a genome scale model is to choose an organism of interest. **B.** A draft is made based on the organism of choice, using the annotated genome sequence or an already existing model phylogenetically similar. **C.** The draft reconstruction is done automatically by different software packages. **D.** Once the draft is done, a curation step is necessary, involving gap-filling, quality control and refinement. **E.** Once the curation is over, the model can be validated using different techniques such as flux balance Analysis (FBA), flux variability analysis (FVA), flux sampling, gene essentiality and drug targeting.

Within the network, reactions are categorized as either internal, which take place within the cell, or external (exchange), involving the uptake and excretion of substrates. Internal reactions represent the consumption and/or production of metabolites, creating reaction rates, known as fluxes [13], while exchange reactions represent the movement of metabolites into and out of the system (sources and sinks, respectively). Transport reactions connect the extracellular environment with the cell by representing uptake or excretion of metabolites into and from the simulated cell.

Next, known enzymatic reactions in the metabolic model are assigned one or more enzymes inferred from the genome sequence via so-called Gene-Protein-Reaction (GPR) associations. Many enzymes contain multiple subunits that must function together, and some enzymes catalyze more than one reaction. Therefore, GPR associations frequently are not a unique 1:1 match; they may involve multiple genes for a single reaction or one gene for multiple reactions [14, 15].

Once an initial model is generated (Figure 1C), it must undergo multiple refinements using data from various databases, experimental results, and literature to minimize artifacts and biases. This iterative process, more commonly known as model curation (Figure 1D), is an essential part of generating a high-quality reconstruction. The resulting model can then be validated with experimental data (Figure 1E).

Bottom-up or top-down reconstruction?

Bottom-Up Approach

Bottom-up reconstruction is an approach to build GEMs by systematically assembling metabolic networks from the organism's genomic and biochemical data. Starting with genome annotation, this method manually incorporates individual enzymatic reactions, metabolites, and a description of biomass composition to create an organism-specific representation of

metabolism. The process relies heavily on manual, iterative curation and access to experimental data. For these reasons it is highly accurate but labor-intensive [16].

In brief, the bottom-up approach consists of the following steps:

1. Genome annotation: Identify metabolic genes via KEGG, UniProt, ModelSEED or another database.
2. Draft model construction: Convert annotated genes into reactions and pathways. For this it is possible to use freely available tools, for example CarveMe [17].
3. Manual curation and gap-filling: Resolve dead-ends and missing reactions via literature/biochemical evidence [18].
4. Validation: Compare model predictions with experimental data.

Top-Down Approach

The top-down approach is to generate a model from a template GEM, ideally of a biologically related organism. This approach is used to reconstruct metabolic networks when the parts and interactions are not known in advance. Since this type of approach uses homology, the ideal template will be as phylogenetically close as possible to the organism of interest. Several freely available software packages do this [19], for example RAVEN Toolbox [20].

The top-down approach works as follows:

1. Template selection: Select template model of another organism.
2. Homology detection: Use a software to automate homology reconstruction.
3. Validation: Compare model predictions with experimental data.

Comparison between both approaches

The choice between approaches depends on the research objectives and available resources. The bottom-up reconstruction is the preferred method when constructing *de novo* models for well-studied organisms or when maximal accuracy is required. A top-down reconstruction, on the other hand, is more suitable for high-throughput studies, personalized medicine applications, or systems lacking comprehensive biochemical data (Table 1) [21]. A drawback of the top-down approach is that it will never capture the reactions that are present in the organism of interest but absent in the template model.

Table 1: Comparison between GEM reconstruction approaches

	Bottom-up	Top-Down
Input	Genome annotation	Template model
Effort	High (manual curation)	Reduced (automated refinement)
Accuracy	High (organism-specific)	Depends on template model quality

A hybrid methodology that combines aspects of both approaches has become widely used over the years. A typical reconstruction pipeline begins with a template-based draft generation, followed by extensive manual curation (and extensions) to ensure biological accuracy. In

particular, hybrid models help to strike a balance between biological accuracy and investments of time and labor [22].

Constraining the solution space: exchanges and virtual medium

The structure of the metabolic network (stoichiometric ratios, reaction reversibility) defines the so-called solution space, or the set of possible flux distributions in the network that do not violate stoichiometric- or thermodynamic relationships. In an *unbounded* solution space, fluxes in a flux distribution can, theoretically, trend towards infinity. To create a more defined solution space, we use biological reasoning to assign values to upper and lower flux bounds, or to *constrain* the model. The common practice in the field is to set these bounds to 1000 units (*i.e.* [0, 1000] or [-1000, 1000] for irreversible and reversible reactions, respectively).

Cells, cellular compartments, and the extracellular environment are explicitly distinguished in GEMs. Exchange reactions connect the system with its boundary and typically take form of $compound_{extracellular} \leftrightarrow \emptyset$. Manipulating the flux bounds of these reactions will define what nutrients can be taken up and/or excreted. As the first approximation, typically all the exchange reactions are set to [-1000, 1000], *i.e.* uptake and excretion are allowed for every metabolite defined in the extracellular compartment. The resulting scenario is a virtual medium: an environment with high concentrations of ready-to-use nutrients.

The virtual medium can be used to troubleshoot missing transport reactions (*i.e.* whether the intracellular metabolic network can take up all compounds from its environment). Flux bounds of exchange reactions are the primary “knob” to constrain the GEM; typically, a limited set of nutrients is available (*e.g.* defined growth medium) and only a restricted repertoire of molecules is known metabolic by-products. This property also allows another use of GEMs - as tools for growth media design [23]. Setting the flux bounds of exchange reactions to either zero, or non-zero value (as introduced above, ± 1000) will have different consequences as constraints (Table 2).

Table 2: Interpretation of flux bounds for exchange reactions

Lower bound	Upper bound	Interpretation
0	0	No exchange
0	>0	Only excretion allowed
<0	0	Only uptake allowed
<0	>0	Both uptake and excretion allowed

Objectives and flux balance analysis (FBA)

To obtain flux predictions, GEMs often are used in combination with optimization-based approaches like Flux Balance Analysis (FBA) [24]. In FBA, we assume the metabolic network to operate in a steady-state (*i.e.* the metabolite concentrations do not change), and define an objective to be optimized (minimized/maximized): in most general form, a weighted sum of

fluxes v . For instance, in a system consisting of 3 fluxes (v_1, v_2, v_3), one such objective could be $2 \times v_1 + 1.2 \times v_3$.

The output of such a linear program (optimization problem) is in a set of steady-state fluxes (distribution) in the network with the optimal value of the objective function. Please note that the optimal solutions are usually non-unique, and the optimal solution space (sets of flux distributions resulting in an optimal value of the objective function) should ideally be inspected further. Two principal approaches to this problem exist: optimization- or flux sampling-based methods. Optimization-based methods, namely, Flux Variability Analysis (FVA), is used to infer the interval (span) of the individual flux values corresponding to the optimal solution, and is generally a fast approach to get a feeling of the size of the optimal solution space. On the contrary, sampling of the solution space allows unbiased exploration of the solution space, including suboptimal solutions. However, a brute-force sampling approach is computationally extremely intensive, therefore, many methods have been developed to facilitate this task [25].

To obtain biologically meaningful solutions, a key decision is to choose a representative objective function. As highlighted above, the objective function might be a combination of flux values, yet typically, can be defined as a single “biologically relevant objective” of the cell. What could these objectives be in parasites? There are a few “usual suspects” in the field, however, keep in mind that the inferred objectives might be lifecycle stage-specific: here we will only briefly describe some options (to be elaborated on in Section 3).

Historically, as the first GEMs have been reconstructed for microbes like *Escherichia coli* and *Saccharomyces cerevisiae*, the most frequently used objective function is biomass formation (production of new cells). This objective function represents the hypothesis that cells try to grow and divide as fast as possible: *i.e.*, the metabolic network will ‘convert’ simpler molecules into building blocks needed for new cells in the way that supports the fastest growth of the culture. Other popular objective functions for microorganisms are related to production of high-value metabolites (*e.g.* vitamins [26]) or undesired end-products (*e.g.* off-flavors in food production [27]).

Furthermore, cell growth is not a representative objective function for many cell types (*e.g.* cells from terminally differentiated human tissues or non-replicative stages of parasites); In this case, alternative objectives should be considered, such as maximization of uptake flux for a certain nutrient (*e.g.* glucose). In some cases, maximization of glucose uptake flux corresponds to maximization of biomass production, but they are not interchangeable *stricto sensu*.

Finally, a typical objective function for non-growing cells is maximization of ATP hydrolysis (ATP production in a steady-state). In non-growing cells, the ATP consumption is typically attributed to maintaining the cell livelihood, a.k.a. “ATP maintenance”. In essence, it is a reaction where ATP is hydrolyzed into ADP and inorganic phosphate ($\text{ATP} + \text{H}_2\text{O} \rightarrow \text{ADP} + \text{P}_i + \text{H}^+$). In non-growing cells, the “non-growth-associated ATP maintenance” (NGAM) value (*i.e.* ATP production and hydrolysis flux under no growth) is a value that can be obtained experimentally - we will return to discussing the ATP maintenance parameters in more depth in Section 3.

Quality control of the reconstruction

To produce meaningful predictions, the newly generated GEM should first be checked for inconsistencies (some sort of quality control). Here, we will highlight a few types of potential issues frequently present in draft reconstructions: (i) mass- or charge imbalance, and (ii) energy- or matter-generating cycles. Such checks are likely to be done more than once during the curation process, and it is critical to perform an additional check on the final version of the model to ensure its consistency.

First, the GEM should be checked for violations of mass- and charge conservation, *i.e.* the net sum of masses and charges of reaction substrates and of products should be equal. Most software packages to handle GEMs can perform the consistency check automatically. For these methods to return insightful results, every metabolite in the GEM should ideally be annotated with both the elemental formula and the charge of the molecule (reactions where at least one metabolite is not annotated will be flagged as inconsistent). Note that the charge of (bio-)molecules is typically pH-dependent, and some of them may have a different charge based on their ionization state. It is therefore important to assume a certain pH value and remain adherent to it during the curation process. Typically, an assumption of intracellular pH to be close to neutral (pH=7) is a good starting point.

Sometimes, the exact formula is unknown or irrelevant (*e.g.* does not participate in chemical conversion); then, to represent these parts of the molecule, the “remainder” within the formulae can be labeled as “R”. Consider the pair of transfer RNAs, the free tRNA, and its version loaded with the amino acid. Prior to loading, the tRNA could be represented as “RH”, meaning that we are not interested in the molecular composition apart from the site that receives the amino acid residue. The loaded tRNA then would be represented by swapping “H” with the amino acid residue (free amino acid minus the -OH which make up water during the aminoacyl-tRNA ligase reaction). For instance: glycine $C_2H_5NO_2$ will be loaded on the respective tRNA RH to form C_2H_4NOR and H_2O .

Defining the elemental formulae (and charges) right generally greatly helps to detect model inconsistencies leading to further issues, *e.g.* proton imbalance (missing/extra H^+ in the reaction). Such imbalances may give rise to energy- or matter-generating cycles (GCs), *i.e.* sets of reactions that may fuel the *in silico* representation of metabolism without external inflow of nutrients and/or oxygen. Typical GCs are, *e.g.*, sets of transport reactions that generate proton-motive force across membranes (eventually re-phosphorylating ADP to ATP), or intracellular cycles that not only recover their starting products but also additional moieties, typically, hydrocarbon fragments.

To prune some of the most typical GCs, a good sanity check is to create a new reaction representing ATP hydrolysis ($ATP + H_2O \rightarrow ADP + P_i + H^+$), set all exchange flux bounds to [0, 1000] (only excretion allowed; see above for more details), and maximize the flux through the ATP hydrolysis reaction. The optimization result should be 0.0; this would mean no “complete” energy-GCs present. For identifying carbon-GCs, typically it is useful to repeat the same procedure allowing oxygen uptake to the model (*i.e.* oxygen exchange set to [-1000, 1000]). In case the predicted ATP hydrolysis flux does not equal zero, the reactions that carry flux should be inspected in detail to identify the culprit(s) for creating the GC. Imbalances (and

GCs) related to other chemical elements are rare; nonetheless, ideally, all inconsistencies highlighted by the mass- and charge balance check-up should be resolved in a meaningful way.

At the end of the curation process, it might be useful to evaluate the quality (and completeness) of the annotations, together with the consistency of the GEM. To facilitate this process, established software packages like MEMOTE [28] and/or a recent package FROG [29] may be used to assess and benchmark the quality of GEMs. MEMOTE also scores GEMs using a numeric scale (in between 0-100%), which is a combination of assessing the stoichiometric consistency of the model and the coverage of its annotations.

However, the scheme for scoring model annotations in MEMOTE is rather rigid. For instance, completeness of metabolite annotations is scored across 11 naming schema/databases, and it is virtually impossible to acquire close-to-complete mappings for all 11 databases. In our view, the modeler should aim for >90% coverage for 1-2 databases to achieve a reasonable interoperability of model identifiers. Overall, MEMOTE is a diagnostic tool, and the interpretation of a low MEMOTE score is dependent on the actual culprit: stoichiometric inconsistencies are an important issue that can be highlighted by MEMOTE, and should be mended at the best capacity to ensure model consistency. However, a low MEMOTE score coming from the lack of comprehensive annotations should not discourage from using the GEM further, as long as the information is sufficient to ensure a reasonable degree of interoperability.

Naming the genome-scale metabolic model

The historical naming convention for the GEMs focused on three components: (i) flagging the reconstruction of metabolism as an “*in silico*” representation; (ii) acknowledging the main author(s) of the GEM; and (iii) specifying a single metric of the model, typically, number of genes present. This convention has been exercised, among others, for all recent *Escherichia coli* GEMs; as an example, two reconstructions published in 2012 and 2017, respectively, were named *iJO1366* [30] and *iML1515* [22].

Despite the proposed nomenclature, approaches to naming the GEMs are subject to personal taste. Variations on the theme include: instead of genes, modelers choose to include the number of reactions; or use the organism’s binary scientific name instead of the initials of the modelers (e.g. the *Bordetella pertussis* GEM *iBP1780*, [31]). Moving from the convention even further, major iterations of the *S. cerevisiae* GEM were named just “yeast”: the latest version “yeast9” was released in 2024 [32]. Similarly, the latest iteration of the *Homo sapiens* GEM is called Human1, which served as the template for the suite of GEMs representing other metazoan model organisms [33].

Recently, there was a call to standardize the meta-data collection for GEMs to improve the FAIRness¹ of GEMs: the “standard-GEM” [34] specification. There, the naming convention relates more to the “yeast” or “human” GEM, suggesting that either a trivial name, or the species identifier (e.g. human-GEM, or sapiens-GEM) is used. However, we discourage from

¹ FAIR stands for Findable, Accesible, Interoperable and Reusable

using the naming convention used in the “standard-GEM” specification: the “canonical” system generates GEM identifiers with a high likelihood to be unique, an important consideration for FAIRness. The historical naming conventions allow for a wide range of alternatives if a name is reserved and do not suggest a historical sequence of GEMs.

3. From a genome-scale reconstruction to stage-specific models

A genome-scale reconstruction of metabolism is generally a start - not the endpoint. While the GEM represents the overall metabolic *capacity* of your organism of interest, at every given moment, the *factual capacity* will always be a subset of the reactions in the network. Here we have to acknowledge the elephant in the room: the chemistry of life requires catalysts, namely, enzymes - which are a great expenditure of resources for the cell [35], including taking up physical volume of the cell [36]. Therefore, even the unicellular, free-living microbes are highly unlikely to express the entire (metabolic) genome at once. In multicellular organisms, such expectation is, in fact, the norm: differentiated tissues execute highly specific gene expression programs. Somewhere in between these two extremes between free-living bacteria and well-differentiated cells, the unicellular parasites typically exhibit multi-stage lifecycles: it is well-known for many parasites that the metabolic map usually changes drastically upon transition to another lifecycle stage [37].

To obtain a good representation of lifestage-specific metabolism, it is critical to extensively curate the species-specific GEM (which we will call the “pan-GEM”, referring to the theoretical metabolic capacity as a union [“pan”] of capacity at different lifecycle stages) prior to crafting lifestage-specific GEMs (Figure 2A). The model reduction process shall, in most cases, generate a lot of gaps in the pan-GEM. These gaps will be a mix of true- and false-negatives (e.g. enzyme-encoding gene not expressed during the lifecycle stage vs. true expression, but below detection limit) due to incomplete datasets. Acknowledging these limitations, the pan-GEM should guide the critical assessment of the impact of losing certain metabolic functions to the lifestage-specific metabolic network.

To make the lifestage-specific metabolic network, three parts (Figure 2B) are needed:

1. Information to reduce the pan-GEM to the network of that lifecycle-stage.
2. Definition of the environment in which the lifestage dwells. The availability of specific nutrients is a determining factor for metabolic reprogramming and, consequently, for metabolic flux simulations.
3. An objective function that can steer the algorithm in making a functional model. The selected objective functions (e.g. biomass production or ATP hydrolysis) should be included into the pan-GEM to represent the variety of metabolic objectives the organism of interest might adhere to. In different lifecycle-stage GEMs, different reactions should be flagged as objectives.

We will first discuss the data to reduce the network, setting lifecycle stage-specific context, and the choice of the objective function. Then, we will introduce the general idea for a model reduction algorithm(s) (Figure 2C), and the software for creating lifecycle stage-specific GEMs

(Figure 2D). Note that even though these lifecycle stage specific models are by definition not genome-wide anymore, we will keep referring to them as GEMs.

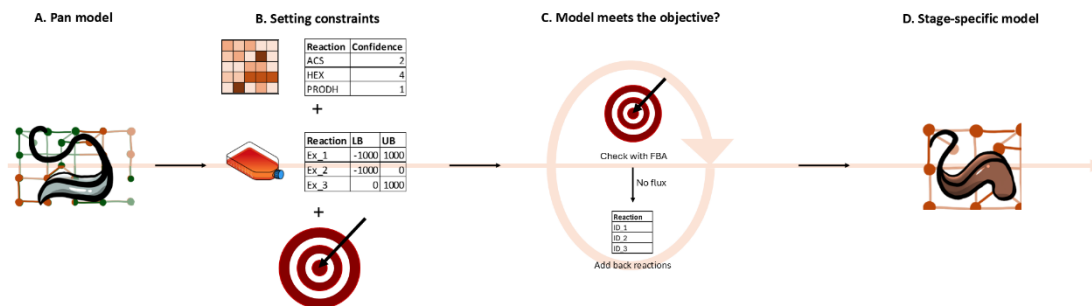


Figure 2. Schematic pipeline of stage-specific GEM. **A.** To build a pan model stage-specific, it is necessary to include a set of constraints. **B.** It is possible to obtain a list of reactions present in the life stage of interest, by using OMICs datasets or information already published in known databases. Other constraints, such as an objective function and medium components are necessary to match with the biological situations that the life stage of interest actually faces. Once the information is gathered, it is necessary to attribute to the reactions high, medium and low confidence scores, depending on the algorithm of use. **C.** In the end, the algorithm has to check if the objective function has flux to maintain or remove reactions. If the objective function does not have flux, there is a need to add back reactions and do the FBA again. **D.** If the objective function has flux, the stage-specific model is complete. The network is smaller than that of the pan-model.

The data for network reduction

-Omics datasets

The process of making lifecycle-stage specific models begins with compiling the information on the makeup of the lifestage of interest, usually in the form of large biological datasets (-omics). Two kinds of -omics data are typically used, transcriptomics and proteomics (and combination of both, if available). Both types of data come with trade-offs: currently, transcriptomics techniques, like RNA-sequencing (RNA-seq), are often still more sensitive than proteomics (e.g., mass spectrometry-based proteomics), generating more data-rich datasets. However, detection of protein makes a stronger case for the enzymatic reaction to take place: mRNA levels frequently are not a good indicator of the protein level [38] due to post-transcriptional regulation; in the end, frequently there is no protein translated at all.

Furthermore, protein expression alone does not necessarily imply enzymatic activity in cases when post-translational modifications or compartmentalization are required. In such cases, profiling of post-translational modifications (e.g. phosphoproteomics) or use of compartment-specific proteomics datasets (e.g. mitochondrial proteome [39]) may help to infer the enzymatic activity. Furthermore, most of the computational methods to reduce GEMs from -omics data rely on thresholding, so it is critical to place the cut-off right. A threshold too low

may preserve metabolic functions that should not be present in the stage-specific model, while the opposite may remove pathways that are actually operational.

When using a single type of data (and/or single dataset), some data exploration (e.g. plotting the values, computing percentiles etc.) might help to get a feeling for numbers and select the threshold value. This decision becomes non-trivial when blending multiple datasets coming from different techniques (e.g. transcriptomics and proteomics, or protein expression vs. phosphorylation degree), since different values have to be harmonized to form a weighing scheme.

Beyond -omics datasets: parsing organism-specific databases

For some (usually biomedically- or biotechnologically relevant) organisms, decades of research have yielded a vast amount of biochemical knowledge from experiments. Keeping in mind that -omics datasets will likely be incomplete, other streams of knowledge should be harnessed if available. Databases which systemize evidence from multiple data sources (e.g. evidence for the presence of activity of specific enzymes) are available for many model organisms such as *E. coli*, or budding yeast *S. cerevisiae*. The [TrypanoCyc](#) [40] and [LeishCyc](#) [41] databases store information for *T. brucei* and *Leishmania major*. For instance, TrypanoCyc contains a massive amount of manually curated entries regarding known- and predicted enzyme activities (the latter means that the reactions take place in cell lysates, yet the responsible enzyme(s) have not been identified), evidence-scored (from -2 to 4, increasing support from data) inspired by the schema introduced by Thiele and Palsson [16].

Maintaining organism-specific databases is a laborious task, and many databases have been unfortunately discontinued and/or are not updated anymore. In such cases, one might resort to more generic, well-established resources, like [KEGG](#) or [BioCyc](#), to aggregate information from, e.g., closely-related organisms.

Constraints for the stage-specific context

Next to a different network topology, also the environment of the lifecycle stage should be taken into account. In many cases, the “environment” could be interpreted synonymously as the growth medium. Cells can only take up what is provided in their growth medium and time-resolved measurements can further specify the values of exchange fluxes.

Using measured fluxes as quantitative constraints

From experiments of concentration profile over time, specific metabolic fluxes can be calculated if the biomass is also quantified. These fluxes can be used as quantitative constraints on the exchange flux in the model. If metabolite measurements are done during growth, it is important to also follow biomass changes.

Flux calculation for lifecycle stages that grow

It is important to realize that in growing conditions, the cell number is increasing over time. The change in the metabolite over time is then a combination of a specific flux (e.g. per unit

biomass) and the increase of cell numbers: flux calculations should take this into account. Thus, the specific flux q for a metabolite can be calculated with the formula:

$$q_{\text{metabolite}} = \frac{\mu}{Y} \text{ [42] (Eq. 1)}$$

in which the μ is the specific growth rate in (cells/cells)/time and Y is the biomass yield on the metabolite in cells/molar metabolite. The μ can be obtained from an exponential fit to biomass data over time. The Y can be obtained by plotting the biomass (e.g. number of cells over the metabolite amount (from time-resolved measurements)). If there is a constant flux, all points will be on a line. The slope of the linear fit through the points will give the Y on the metabolite. This method has been used to calculate the glucose consumption flux in batch cultures of *T. brucei* [43] and that paper shows the plots and the calculations for μ , Y and q_{glucose} .

Flux calculation when biomass is constant

In cases where biomass is constant, flux calculations are relatively simple. This clearly applies to lifecycle stages that are stationary (e.g. stumpy form *T. brucei*). The simpler method might also apply for proliferative lifecycle stages, if the measurements are done in a timeframe where growth can be neglected. Because there is no increase in biomass, if there is a constant flux in these experiments, a plot of the metabolite concentration over time will show a straight line. The slope of a linear fit through the points will yield concentration/time. Correction for incubation volume and division by the (average) amount of biomass in the incubation will give the specific flux q in molar amount/time/cell number. An example for glucose and glycerol fluxes from a 12-hour incubation of *T. brucei* can be found in [44]².

Qualitative or semi-quantitative constraints based on growth medium composition

Sometimes, and especially working with organisms growing on rich media (such as parasites or mammalian cells), it is complicated to quantify the consumption of all media components. Even without flux quantification, we can still use information from the medium composition to constrain exchange fluxes. For instance, we could approach this in a qualitative manner (binary decision whether to allow uptake of the metabolites or not) based on the media composition. If the component is absent from the media, the flux bound is set to [0, 1000] (see previous section). Qualitative constraints could be used to test whether cells could grow on different media; identify nutritional dependencies (auxotrophies); test whether cells can grow anaerobically (without oxygen).

Furthermore, in the absence of further information, we could resort to some simplification to derive reasonable exchange flux bounds. For such cases, we could follow the assumption that after a certain time period τ (e.g. 24 or 48 hours) the cells will halt growth because one (or more) nutrients has been depleted in the medium. Then we could infer the maximal possible

² It remains a challenge to keep biomass constant while in relevant conditions. The example of this 12-hour study was done in a growth medium. Therefore, growth can be expected, but they calculated fluxes as if the cell numbers stayed constant. Close inspection of the data shows that not all the points are perfectly on the line. This may be because with a doubling time of BSF *T. brucei* of approximately 6 hours there may have been some growth in their experiments. The authors only report no cell death, but if they would have monitored cell numbers they may have been able to use the flux calculation for growing cells.

uptake flux over τ from the concentration of the component X (denoted as x in [mmol/L]) of the starting media, cell dry mass (m in [gDW/cell]) and initial cell seeding density (D in [cell/L]):

$$LB_x = -1.0 \times \frac{x \text{ [mmol/L]}}{D \text{ [cell/L]} \times m \text{ [gDW/cell]}} \times \frac{1}{\tau} [h^{-1}] \text{ (Eq. 2)}$$

This represents the maximal theoretical uptake rate, which likely will be considerably lower in real-life, and is highly dependent on the growth rate of the cells. However, this can be a useful first approximation to, e.g., narrow down media components of interest to be quantified in more precision.

Objective functions for different life-stages

As we have seen above ([Objectives and flux balance analysis \(FBA\)](#)), to calculate flux distributions, models, and therefore also the stage-specific models, will need an objective function that is maximized or minimized. Biomass production is a logical objective for the growing lifecycle stages of parasites. However, non-growing lifecycle stages should be represented by a different objective function. We will first discuss the objective functions for growing lifecycle stages of parasites, with the focus on the least trivial to determine: biomass formation. Then we will discuss alternatives for objective functions for non-growing life-stages.

For growing life-stages: biomass objective function (BOF)

The build-up of new cells is not explicitly encoded in the genome: a virtual reaction needs to be included that produces “cell biomass”. The biomass objective function (BOF), or, colloquially, the biomass equation, represents the fractions of (macro-)molecules that are parts of the cell in the ratios they are present in dry cell biomass. Generally, a biomass equation will consist of DNA, RNA, protein, lipids, and carbohydrates. If quantities are known, minor components like cofactors and metal ions may be included as well. Although producing the biomass precursors (and assembling them to macromolecules) requires a lot of energy, usually there is a substantial fraction of energy investment that is not explicitly accounted for. We name the additional ATP requirement the growth-associated ATP maintenance (GAM): it is represented in the BOF as hydrolysis of ATP to ADP and phosphate. We will describe the process of computing this parameter, as well as non-growth-associated ATP maintenance (NGAM), introduced in Section 2. NGAM will be used to set the flux bounds for an ATP hydrolysis reaction, **separate** from the BOF.

As mentioned before, the major components of biomass are largely the same among different cells, but their fractional contribution to biomass may differ. Therefore, the stoichiometric coefficients (quantities of components per dry weight) need to be tailored to the organism of interest. Proteins, nucleotides, and lipids take up most of the biomass by mass, so it is important to get a good estimate or measurements for these.

To date, there is no standardized protocol for experimentally determining *all* cellular components. These values either need to be experimentally measured or inferred from available data - we will briefly discuss combining both approaches. For the latter, a good starting point is to check whether there is a related organism for which a BOF already exists. Then, consider which fractions may differ in (the life-cycle stages of) your organism - variations

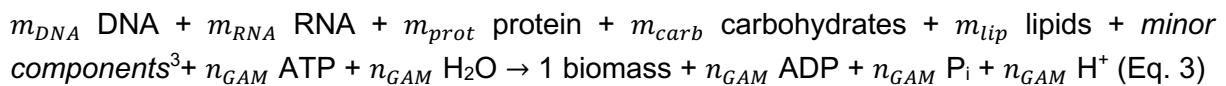
in lipid composition, protein and nucleic acid content, as well as cofactors and other essential metabolites, must be incorporated into the biomass equation.

In all reactions of the GEM, the stoichiometric coefficients are, in fact, *mol* (in the field, we typically operate in space, where expressing molar amounts in *mmol* is more convenient). The biomass equation is a special case, because for the dimensions to work out right (broader discussion on this available in [45]), the biomass constituents should be expressed as their abundance in the biomass, i.e. *mmol/gDW*. Multiple strategies exist to encoding the biomass formation in the GEM, and the critical sanity check in formulating the representation of the biomass is to make sure that the abundance of the biomass components is expressed in the correct dimensions.

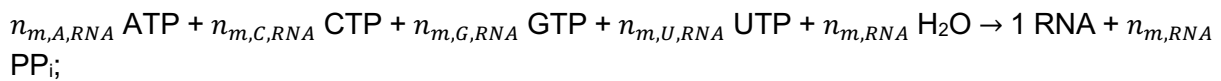
Here we will briefly describe our preferred approach, where we will create a simplified BOF which represents the bulk components in their mass fractions - this is typically what is experimentally measured. Then, for constituents formed of simpler components, we will create reactions representing production of these biomass precursors. This will end up in a set of reactions which are relatively easy for the modeler to interpret and are not cluttered as if all known biomass components were described in a single, very long reaction equation.

Designing a BOF

We will start with the simplified BOF. There, we describe the composition of the cells in terms of mass (*m*, in *g/gDW*) of high-level components (nucleic acids, protein, etc.). Moreover, we include the estimated GAM value in molar units (*n*, in *mmol/gDW*). The general representation of this equation will be as follows:



The next step is to add more detail, i.e. describe the molar composition of the bulk constituents, i.e. their molar contribution to the constituent mass n_m (in [*mmol/g constituent*]). This way, the “coupling”, or multiplication of, e.g., the moles of nucleoside triphosphates used to make 1 gram of RNA (*mmol/g RNA*) and the mass of the RNA in the BOF (*g RNA/gDW*), represents the demand of nucleoside triphosphates in *mmol/gDW*. We will take polymerizing bulk RNA from individual nucleoside triphosphates as an example:



$$\text{where } n_{m,RNA} = (n_{m,A,RNA} + n_{m,C,RNA} + n_{m,G,RNA} + n_{m,U,RNA}). \text{ (Eq. 4)}$$

This procedure is repeated for each biomass constituent. Below, we will describe how to obtain the stoichiometric coefficients to formulate the final BOF.

³ Referring to any known minor components of the biomass

Obtaining the separate biomass components for a lifecycle stage

Weight fractions for the biomass equation can be obtained from existing data in literature or from new measurements. It is important to note that each weight fraction needs to be normalised to the dry weight of the organism (i.e. g/gDW). The dry weight is not something that is often determined for parasites. Many studies will report the biomass used in assays as (i) the number of cells, (ii) optical density or (iii) protein amounts, and it will be important to know how much dry weight these measures of biomass represent. Therefore, we will here first provide a protocol for dry weight measurements, before we discuss the different weight fractions.

Dry weight measurements

For dry weight measurements intact cells are needed. As parasites do not have a cell wall they are more vulnerable to lysis than organisms like bacteria and yeast. To highlight the adjustments needed for parasites we here present an entire protocol that successfully been used for the sleeping-sickness parasite *Trypanosoma brucei* (data unpublished; personal communication, J.R. Haanstra).

Equipment and materials

- Filters: Nucleopore (TM) Track- Etch membrane; 0.4 μm ; diameter 25mm; Whatman(TM)
- Petri dish (\approx 92 mm)
- Small petri dishes (\approx 35 mm)
- Tweezers
- Vacuum manifold (e.g. VACUUBRAND diaphragm vacuum pump and Satorius manifold)
- 60°C incubator
- Analytical scale (e.g. Satorius Cubis® II Semi-Micro Configurable Lab Balances)
- Desiccator

Preparation of the filters

1. Dry filters by placing them in a petri dish in a 60 °C stove to remove water.
2. Take filters from the stove and keep them in a desiccator.
3. Weigh the filters (shiny side down) after at least 24 hours: measure weights on a milligram scale immediately after placement on the scale ($t=0$) and after 30 seconds of the filter placement ($t=30$). The relevant measurement for dry weight determination is after 30 seconds, since scales needs some time to adjust and to provide the most accurate measurement.
4. Place weighed filters in individual small petri dishes and write down their measurements.
5. Weighed filters can be stored at room temperature until use.

Preparation of the sample

1. Grow cells in the exponential phase *Note: The exponential growth phase is often the most relevant and theoretically stationary phase cells may alter their size (dry weight) and biomass composition*
2. Count the cells to know the number of cells that will be in the pellet.
3. Pellet the cells by centrifugation (*Note: centrifugation speed depends on the species - the cells should all be spun down but stay intact*)
4. Carefully resuspend the pellet in 1 mL of the growth medium, adjust volume if necessary.

Measurement

1. Place the dried filters (shiny side down) on the manifold holder and secure the filter with a clamp
2. Turn on the vacuum pump and wet the filters with growth 1 mL medium. *Note: In many protocols for other organisms cells are suspended in and washed with water or PBS to exclude proteins from the FCS-containing growth media. We have experienced that this is detrimental to fragile parasites and you lose a lot of biomass. Instead we do a control with medium alone to correct for carry-over of proteins from the medium*
3. Pipet the cell suspension slowly on a filter. *Note: avoid the liquid from touching the ring of the clamp, if it does touch, make sure to rinse it down in the next step.*
4. Flush the filter with 2 mL of growth medium.
5. Turn off the vacuum pump
6. Place each filter in their corresponding petri dish and place this in a 60°C incubator for at least 24 hours. [you need to identify the exact filter as they will have slightly different weights].
7. Weigh the filter on the same scale and in the same manner as in the preparation phase
8. As a control perform the same procedure for medium without cells [use the same volumes as for the cell suspension]

Calculation of the dry weight

1. Subtract the filter weight before the addition of the sample at $t=30s$ from the filter weight after the addition of the sample at $t=30s$.
2. If the medium control dry weight was not equal to 0, it had to be subtracted from each dry weight measurement
3. By dividing the dry weight by the number of cells in the sample (determined at Step 2 of 'preparation of the sample') you get to the dry weight/cell (or per number of cells - generally 10^6 or 10^8 cells)

Experimental methods to infer fractions of the different macromolecules (DNA, RNA, proteins, lipids)

Many of the bulk components of the cells can be measured using rather simple, typically spectrophotometric, assays (although ready-to-use kits are also available from many vendors). A successful example of covering most of the components is the biomass

quantification for *E. coli* [46]. Here we will highlight some straightforward measurement options for some of the macromolecules:

- DNA and RNA: can be isolated from cells separately and quantified by UV absorbance spectrophotometry according to standard biochemical protocols [e.g. RNA [47]]
- Protein: Several well-established protocols exist to measure proteins in lysates, such as e.g. BCA [48], Bradford [49]. In any case it will be extremely important to remove protein components from the culture medium in the assay.
- Lipids: Total amount of lipids in the biomass can be quantified colorimetrically using vanillin-phosphoric acid reagent [50]. However, determining the composition of the lipids (ratios of different phospholipid classes, and ratios of different fatty acids incorporated into lipids) requires technologically advanced methods (e.g. gas chromatography, mass spectrometry [51]).
- Carbohydrates: typical storage carbohydrates (like glycogen and trehalose), as well as total carbohydrate pool, can be measured using anthrone- [52] or phenol-sulfuric acid assays [53]. Alternatively, enzymatic assays can be used for, e.g. trehalose quantification [54].

Considerations for structural components: carbohydrates, lipids and special molecules produced by the organism of interest

Finally, specific structural metabolites specific or unique to the organism of interest must be incorporated into the biomass equation. Many organisms contain specific glycans and lipids, such as GPI anchors, N-glycans linked to proteins, and structures like the cell wall (in prokaryotes). These compounds are often involved in host-parasite interactions processes, making their biosynthesis essential for the parasite.

Minor constituents: vitamins and cofactors

Many organisms, including parasites, require vitamins and cofactors that are involved in enzymatic reactions and redox processes. Many of these compounds are critical to the parasite's survival but are not synthesized by the organism's metabolic network, and must be acquired from the host or extracellular environment. Identifying and quantifying these compounds for inclusion in the biomass equation is a challenging task. Experimentally, auxotrophies are identified using defined media [55]; essentiality studies can help guide the construction of the biomass equation. As an entry point, one might resort to reviewing the literature for information on characterized enzymes in the organism of interest that utilize specific cofactors. Typically, BOF only includes redox cofactors such as NADH, NADPH, glutathione, and trypanothione (for e.g. Kinetoplastea) due to their role in the bioenergetics and mitigating oxidative stress.

Connecting the macromolecular fractions to the network

Macromolecules are composed of simpler specific building blocks: the different nucleotides of the genomic DNA and in the RNA, the amino acids in the proteins and different carbon-chain fatty acids incorporated into the specific lipid classes. The final step in formulating a BOF is to determine the ratios of different building blocks coming together into the macromolecules. In brief, one has to count all the different building blocks in those datasets and determine their fractional contributions to the total.

For DNA, RNA and protein, the subfractions can be obtained from the genome composition and -omics datasets. Semi-quantitative lipidomics can be used to infer the composition of lipids. We refer the reader to the description of these computations [56] accompanying the semi-automated tool called BOFdat. Note that it is important to compute the macromolecule composition for different life-stages, as the fractional contributions will likely change during the lifecycle of your organism.

Quantifying growth-associated and non-growth associated maintenance parameters in trypanosomatids

The biomass equation implicitly includes the energy costs of producing individual cellular building blocks (DNA, RNA, proteins, set of membrane lipids, and other macromolecules), complementing the theoretical calculation of energy requirements (cf. the bloodstream form of a trypanosomatid *Trypanosoma brucei* [56]). It is important to realize that the ATP demand for biosynthesis arises from stoichiometric coupling in the reactions that make up the biomass precursors (e.g. each tRNA loading requires hydrolysis of 2 ATP units). Moreover, some energy might be required for functions that are not associated with making biomass. However, if we tracked the bioenergetics of trypanosomatids at different lifecycle stages, we would realize that biosynthesis and maintenance do not entirely cover the ATP production in these cells.

To address this, we typically decompose the ATP consumption (which is equal to production at steady-state) into

$$q_{ATP} = q_{ATP,GS} + q_{ATP,NGAM} \text{ (Eq. 5)}$$

where $q_{ATP,GS}$ (“GS” for “growth-scaling”) represents the ATP demand that scales with growth, i.e. the known ATP demand (demand for producing biomass precursors) and unspecified energy demands that are associated with growth (growth-associated ATP maintenance, GAM). The $q_{ATP,NGAM}$ (non-growth-associated ATP maintenance, NGAM) represents the energy demand for the processes in resting cells, e.g. maintaining membrane potential, DNA repair, proteome and transcriptome turnover, motility and other processes related to cellular organization and maintenance. This specific flux will set the lower bound through an ATP hydrolysis reaction (separate from the BOF!) in the model.

How to obtain these specific fluxes? Say, we grew cells in a defined medium and quantified the growth of the cells and consumption of major nutrients (e.g. glucose). Then, we can use the flux measurements to constrain the GEM we constructed, in this way inferring the specific production of ATP (q_{ATP}) in these cells. The inconsistency between what GEM perceives as “ATP demand” vs. actual ATP demand based on the experimental data (=fraction of unexplained q_{ATP} from the experiments) will allow us to estimate the ATP maintenance parameters to be put into the model: the n_{GAM} in the biomass equation and the value for $q_{ATP,NGAM}$. Below we describe the calculations⁴ needed to quantify growth associated maintenance and non-growth associated maintenance.

⁴ The BOFdat tool mentioned above can also automate the calculation of the growth associated maintenance and non-growth associated maintenance. However, we recommend performing calculations manually since the BOFdat tool was made with the intention to infer parameters for microbial GEMs from extensive (and precise) physiological data.

How to translate experimental measurements to model parameters?

There are two principal ways to rephosphorylate ADP into ATP in living cells: in substrate-level phosphorylation (SubPhos), an organic molecule is broken down and the energy released is used to phosphorylate ADP to ATP; or in oxidative phosphorylation (OxPhos), the energy from the transfer of electrons is used to create a H^+ gradient and the resulting proton-motive force is used to drive the ATP synthase. In the latter process, the electron needs a terminal acceptor, which is often molecular oxygen, but could also be organic molecules such as fumarate or nitrogen compounds in case of some bacteria. Thus, we can define the q_{ATP} in more metabolic terms as a sum of the ATP production fluxes through the two types of ATP production:

$$q_{ATP} = q_{ATP,OxPhos} + q_{ATP,SubPhos} \text{ (Eq. 6)}$$

Recap from the (Eq. 5) that the total ATP flux in growing cells is a combination of $q_{ATP,GS}$ and $q_{ATP,NGAM}$, so, putting the right-hand-sides of (Eq. 5) and (Eq. 6) together allow us to express the unknown specific fluxes $q_{ATP,GS}$ and $q_{ATP,NGAM}$ using terms $q_{ATP,OxPhos}$ and $q_{ATP,SubPhos}$, which we can infer by constraining the GEM with experimental data. However, the $q_{ATP,GS}$ has no meaning in the GEM by itself; instead it is a sum of two growth-related terms:

$$q_{ATP,GS} = (n_{Biomass} + n_{GAM})\mu, \text{ (Eq. 7)}$$

The $n_{Biomass} \times \mu$ is the fraction of the q_{ATP} that is accounted for by the GEM by stoichiometric description of biosynthesis of the biomass precursors at a given growth rate μ . The parameter n_{GAM} is the coefficient for ATP hydrolysis part of the BOF (Eq. 5), for which we want to get a value.

With this relation, (Eq. 5) can be rewritten as:

$$q_{ATP} = q_{ATP,NGAM} + (n_{Biomass} + n_{GAM})\mu \text{ (Eq. 8)}$$

Below we will discuss measurements and computational routines to infer $q_{ATP,NGAM}$ and n_{GAM} . At first, we perform measurements which would allow to estimate $q_{ATP,OxPhos}$ and $q_{ATP,SubPhos}$ (Eq. 6), and feed them into the GEM in order to turn the terms of (Eq. 6) into terms of (Eq. 5), out of which we can obtain the desired parameters.

Sometimes, precise measurements are not available for the organism of interest; in these cases, a good first approximation of the $q_{ATP,NGAM}$ and n_{GAM} values can be taken GEM from a closely related organism [57].

Measurement of the oxygen consumption rate

We will first address the $q_{ATP,OxPhos}$ term from the (Eq. 6). In principle, we can infer this value experimentally from measuring the specific oxygen consumption rate q_{O_2} . In turn, the total oxygen consumption is comprised of contributions by OxPhos and other contributions that do not contribute to ATP production (proton leak, alternative oxidase, or non-mitochondrial oxygen consumption):

$$q_{O_2} = q_{O_2,OxPhos} + q_{O_2,non-OxPhos} \text{ (Eq. 9)}$$

The $q_{O_2, OxPhos}$ flux will allow us to estimate the $q_{ATP, OxPhos}$; since the $q_{O_2, non-OxPhos}$ does not bear any mechanistic explanation in itself, the most straightforward way is to measure the total oxygen consumption flux q_{O_2} , and $q_{O_2, non-OxPhos}$ (i.e. residual oxygen consumption when OxPhos is inhibited).

Key considerations from mitochondrial physiology

In trypanosomatids with ‘canonical’ mitochondria (e.g., those retaining most respiratory complex activities), most q_{ATP} is generated through OxPhos. The electron transport chain (ETC) is a complex system with multiple entry-points, which leads to diverse functional capacity of this system. For instance, some *Leishmania* species possess a fully functional (proton-pumping) Complex I of the ETC, whereas *T. cruzi* and *T. brucei* lack critical subunits required for H^+ transfer across the membrane [58], and thus, reduced ATP yield on oxygen (P/O ratio). Electron entry via subsequent complexes (Complex II, glycerol-3-phosphate shuttle, electron transfer flavoprotein etc.) affects the P/O ratio in a similar way⁵.

However, it should be noted that in some organisms (and/or life-cycle stages), electrons may be diverted from entering the ETC, towards the water-forming alternative oxidase [AOX], as in the bloodstream form of *T. brucei* [59].

As explained above, we can obtain the $q_{O_2, OxPhos}$ by measuring the total oxygen consumption flux q_{O_2} , and the non-OxPhos part of the oxygen consumption $q_{O_2, non-OxPhos}$. The q_{O_2} can be robustly quantified through high-resolution respirometry assays conducted on intact cells in both growing (mid-log) and non-growing (early-stationary phase). If we inhibit the F_0F_1 -ATP synthase using, e.g., oligomycin [60], then the $q_{O_2, OxPhos}$ becomes 0 and the $q_{O_2} = q_{O_2, non-OxPhos}$. In the following, we describe a protocol to measure the oxygen consumption flux in both untreated (q_{O_2}) and treated ($q_{O_2} = q_{O_2, non-OxPhos}$) cells.

Materials

Liver infusion tryptose (LIT) medium supplemented with 10% fetal calf serum

Oligomycin A (5 mg mL⁻¹ and 0.5 mg mL⁻¹ stocks in ethanol)

High-resolution respirometer (e.g. Oroboros-O2K)

Hamilton microliter syringes

Cell Harvesting

1. Grow *T. cruzi* epimastigotes to the early-stationary phase in LIT medium (~10⁸ cells mL⁻¹).
2. Centrifuge the cells at 1000 × *g* for 10 min and save the supernatant⁶. Resuspend the cells in the saved supernatant (spent medium) to maintain physiological conditions, and adjust to the required density for each assay.

⁵ The P/O is ~2.5 ATP/O if electrons are passed onto proton-pumping Complex I; otherwise (Complex II/Complex I without a proton pump) the value is ~1.5.

⁶ The physiological concentrations at which the cells were collected will be maintained using this supernatant.

Measurement of respiration in intact cells

Cell Preparation and Instrument Setup

1. Adjust cell concentration to 1×10^9 cells mL^{-1} in the saved supernatant.
2. Set the equipment to the appropriate temperature for each cell type (28 °C for *T. cruzi* epimastigotes)
3. Transfer 2 mL of the retained supernatant medium into each respirometer chamber and wait 15 minutes for signal stabilization.

Cell Loading and Measurement

1. Add 50 μL of cell suspension to oxygraph chambers (final concentration: 2.5×10^7 cells mL^{-1}).
2. Wait 10 minutes for signal stabilization (routine respiration measurement).
3. Determine optimal oligomycin concentration for respiratory inhibition, performing stepwise titration by adding 1 μL oligomycin (0.5 mg mL^{-1} stock).
4. Allow 5 minutes between additions (leak respiration is established when additional oligomycin produces no further inhibitory effect). For subsequent replicates, use the minimum effective final concentration (use 5 mg mL^{-1} stock for this purpose).
5. Allow 5 minutes for signal stabilization.

Calculation of the oxygen flux to OxPhos

1. Subtract the leak from total oxygen flux:

$$q_{O_2, \text{OxPhos}} = q_{O_2} - q_{O_2, \text{non-OxPhos}} \text{ (Eq. 10)}$$

Oxygen flux rates derived from respirometric assays are expressed in $\text{pmol O}_2 \text{ s}^{-1}$ per cell number. To normalize via unit conversions for computational analyses, standardize cell counts to dry weight (see Section [Constraints for the stage-specific context](#)).

Although the oxygen consumption flux is the key input to the model, optionally, you can compute the ATP produced by OxPhos ($q_{\text{ATP}, \text{OxPhos}}$) manually.

2. Calculate the number of protons pumped per O_2 consumed.

Example: If $q_{O_2, \text{OxPhos}} = 150 \text{ pmol O}_2 \text{ s}^{-1}$ per 2.5×10^7 cells and the electrons enter the ETC via the Q-junction, 6 H^+ are pumped per oxygen atom reduced (2e^-) (4 from CIII and 2 from CIV) so this results in $1,800 \text{ pmol H}^+ \text{ s}^{-1}$ per 2.5×10^7 cells. Since this is a measure of $q_{\text{ATP}, \text{OxPhos}}$, this value corresponds to the flux of protons re-entering the mitochondrial matrix through the $\text{F}_1\text{F}_\text{O}$ -ATP synthase.

3. Convert H^+ to ATP (P/O Ratio)

The P/O ratio (ATP produced per oxygen atom reduced)⁷ depends on the H⁺ required per ATP synthase: 4 H⁺/ATP (standard in mammalian mitochondria)⁸.

$$P/O = \frac{H \text{ pumped per } O}{H \text{ per ATP}} \text{ (Eq. 11)}$$

More on the P/O ratio and how to calculate them can be found in [61].

4. Calculate ATP Flux

$$q_{ATP, OxPhos} \text{ (pmol ATP s}^{-1} \text{ per } 2.5 \times 10^7 \text{ cells)} = q_{O_2, OxPhos} \times P/O \times 2 \text{ (Eq. 12)}^9$$

Example: If $q_{O_2, OxPhos} = 150 \text{ pmol O}_2 \text{ s}^{-1} \text{ per } 2.5 \times 10^7 \text{ cells}$ and the electrons entry point is the Q-junction, P/O is 1.5, so:

$$q_{OxPhos} = 150 \times 1.5 \times 2 = 450 \text{ pmol ATP s}^{-1} \text{ per } 2.5 \times 10^7 \text{ cells (Eq. 13)}$$

In general form, (Eq. 13) becomes:

$$q_{OxPhos} = (q_{O_2} - q_{O_2, non-OxPhos}) \times \left(\frac{H/O}{H/ATP} \right) \times 2 \text{ (Eq. 14)}$$

Measurement of the nutrients and byproducts of substrate-level phosphorylation

In addition to ATP production via OxPhos, an important fraction of q_{ATP} comes from glycolytic and mitochondrial substrate-level phosphorylation (collectively: SubPhos, $q_{SubPhos}$) (see Eq. 6). The contribution of $q_{SubPhos}$ to the total q_{ATP} can be inferred using the GEM, constrained with the fluxes of the excreted products. Glycolytic SubPhos typically yields 2 ATP molecules per hexose. Equivalently, the glycolytic substrate-level phosphorylation can be quantified as the excretion of pyruvate and alanine (product of glutamate:pyruvate transamination), with 1 ATP per Pyr/Ala. This might be a more accurate measurement if multiple hexoses are present in the media. The mitochondrial SubPhos refers to the production of ATP in the ASCT/SCS cycle¹⁰, with 1 ATP per acetate or propionate produced. Major sources of mitochondrial acetyl-CoA and propionyl-CoA are glycolysis (pyruvate), amino acid catabolism, and the beta-oxidation of fatty acids [62-65].

One of the most robust ways to quantify fluxes in complex systems is by tracking metabolite concentrations in time using ¹H-NMR. Here, we provide a protocol that allows preparing the cells for monitoring these fluxes by NMR using *T. cruzi* epimastigotes as an example.

⁷ For NADH: 10/4 = 2.5 ATP/O, for FADH₂ (Q-junction): 6/4=1.5 ATP/O

⁸ May vary (e.g., 3.3 H⁺/ATP in bacteria).

⁹ The factor 2 accounts for 2 oxygen atoms per O₂.

¹⁰ ASCT (acetyl-CoA:succinyl-CoA transferase) enzyme catalyzes the CoA transfer from acetyl-CoA or propionyl-CoA onto succinate. SCS (succinate-CoA ligase), a step of the canonical TCA cycle, catalyzes phosphorylation of ADP to ATP via succinate-CoA hydrolysis into succinate and free CoA.

Materials

Cell culture supernatant

3 kDa MWCO filter

Internal standard (e.g., TMSP-d₄)

NMR spectrometer (600 MHz)

Cell Harvesting

1. Grow epimastigote culture to early-stationary phase (10^8 cells mL⁻¹) in LIT medium.
2. Collect 1 mL sample from the culture on the first once the cells reach stationary density (10^8 cells mL⁻¹) and 24h after this point. Register the cell density of each sample for control.
3. Centrifuge the culture ($3000 \times g$, 10 min, 4 °C) to remove cells and cell debris.
4. Deproteinize: Filter supernatant (3 kDa MWCO filter)
5. Freeze the filtered supernatant at -80 °C until the day of data acquisition¹¹
6. Collect 300 μ L of sample, and mix it with 240 μ L deuterium oxide, 60 μ L of 5 mM the internal standard (TMSP-d₄), and transfer to a 5 mm glass tube for ¹H-NMR acquisition¹².

Data Preparation and Calculation

1. Input: ¹H-NMR peak integrals (from specialized software) for glucose, fructose, acetate, propionate, and the internal standard.
2. Normalize the concentration against the internal standard:

$$[\text{metabolite}] \text{ (mM)} = \frac{\text{analyte peak area} \times [\text{TMSP}]}{\text{TMSP peak area}} \text{ (Eq. 15)}$$

3. For flux calculations, see Subsection [“Constraints for the stage-specific context”](#).

Using the flux measurements to compute the ATP maintenance parameters

To get the values for $q_{ATP,NGAM}$ and n_{GAM} , we will need to use the flux measurements (and growth rate for growing life-cycle stages) to constrain the GEM. The basic principle is as follows: the constraints will constrain how much ATP can be produced in total (q_{ATP}), and with using multiple sets of experimental measurements, we will be able to decompose the q_{ATP} into its respective parts from (Eq. 7). Keep in mind that we need to perform the calculations sequentially in order to obtain the values.

¹¹ NMR signal acquisition is influenced by instrument-specific settings (e.g., receiver gain, scan number) and reference compound selection (e.g., TSP, formate); validate and adjust these parameters prior to data collection.

¹² After spectrum acquisition, the characteristic signals for each metabolite and the quantification standard must be integrated using specialized software. Metabolite concentrations shall then be calculated by comparison with the internal standard, applying appropriate normalization procedures.

Calculating the NGAM value

Using the consumption- and production fluxes computed from data for non-proliferative life-cycle stages, we can infer the value of $q_{ATP,NGAM}$: from (Eq. 8), in non-growing conditions $q_{ATP} = q_{ATP,NGAM}$, when $\mu = 0$.

1. Create the reaction of ATP hydrolysis that we will call ATPM¹³, where ATP is hydrolyzed into ADP and inorganic phosphate ($ATP + H_2O \rightarrow ADP + P_i + H^+$). Set this reaction as an objective flux (maximize).
2. Set the lower- and upper flux bounds of the biomass equation to [0, 0].
3. Set the exchange fluxes bounds for known consumption- and excretion fluxes (oxygen, glucose consumption; pyruvate, acetate excretion etc.).
4. Optimize the model: the flux through the ATPM reaction will be $q_{ATP,NGAM}$.
5. Set the $q_{ATP,NGAM}$ as the lower flux bound for the ATPM reaction. Proceed to estimating the n_{GAM} .

Calculating the GAM coefficient in the BOF

In the previous section we have determined the $q_{ATP,NGAM}$ term of the (Eq. 8). To obtain the n_{GAM} to be input into the BOF, we will need to apply a similar routine as described above, but this time with the data coming from growing life-cycle stage. Note: the lower flux bound set in the Step 5 of the previous section should remain set!

1. Set the growth rate of the cells: constrain the BOF to the experimentally observed growth rate (i.e. $[\mu, \mu]$). Do **not** set the BOF as the objective.
2. Set the exchange fluxes bounds for known consumption- and excretion fluxes.
3. Optimize the model: the resulting flux through the NGAM reaction will be $q_{ATP} - n_{Biomass} \times \mu$, because the $n_{Biomass} \times \mu$ is already set by the flux bounds of the BOF.
4. To extract the $n_{GAM} \times \mu$ (Eq. 8), subtract the $q_{ATP,NGAM}$ from the obtained flux value. Divide the value by μ to obtain the n_{GAM} value.
5. Add the ATP hydrolysis part to the BOF to bring it to the shape referred to in (Eq. 3).

Alternative objective function(s) for non-growing stages

Non-replicative stages require sufficient energy and nutrients to maintain cell viability, enabling the parasite to evade immune responses, infect new hosts and continue its life cycle. The objective function for these non-replicative stages can be rendered as the set of maintenance processes required to maintain cellular integrity and viability. Typically, ATP hydrolysis (v_{ATPM}) is rendered as a sufficient description of non-replicative states. Identifying cellular and metabolic processes that are more active during the non-replicative phases of these organisms can help in the formulating such an objective function.

When using ATP maintenance as the objective, a practical approach is to set the lower bound of the ATP hydrolysis reaction to the measured NGAM value, and perform minimization of the flux through the reaction (essentially, to obtain a flux distribution, in which ATP is hydrolyzed into ADP and inorganic phosphate).

¹³ ATPM is for “ATP maintenance”. Some GEMs also call this reaction “NGAM” to explicitly specify which kind of ATP maintenance this reaction represents.

Algorithms and software to obtain lifestage-specific models

Data pre-processing usually is the most labor-intensive task to obtain the lifestage-specific models. The actual process is conceptually rather straightforward: in the end, we would like to have a lifestage-specific GEM (more broadly, a context-specific GEM) that contains only reactions for which an [active] enzyme is present at the said lifestage. In other words, metabolic functionalities without the support from the data should be removed from the model.

You might consider coding your own algorithm, or leveraging a software package that is publicly available. It is best to make a (semi-)automated pipeline for the stage specific models that can be redone easily: in case you modify the pan-GEM, the stage-specific models have to be rebuilt as well. Depending on the exact formulation of your algorithm (or the choice of a publicly available one), the process might become computationally heavy because of memory- and computing power-demanding tasks (e.g. solving mixed-integer linear programs [MILPs]). In either case, the process should translate the expression (or enzyme activity) data into an “action plan” to remove reactions from the GEM to obtain a data-informed, reduced model.

Due to multiple confounded factors, like detection limits in -omics data, incomplete annotations of metabolic enzymes, and moonlighting functions, the initial context-specific GEM will probably lack essential reactions to perform most of the metabolic tasks/objectives. This is the usual outcome irregardless of the method (custom- or publicly available) used, and threshold applied to discriminate between “present” and “absent” enzymes. Thus, algorithms usually also include a “mitigation” step, where reactions, critical to performing metabolic tasks/objectives, are added back to the GEM.

To summarize the algorithm of turning a GEM into a lifestage-specific model:

1. Match the GPRs with the entries in the expression dataset (of the lifestage of interest) to score the GPRs (levels of enzymes) in the GEM.
2. Remove reactions for which GPRs are scored below the predefined threshold value.
3. Check if the model still can perform the metabolic tasks/objectives (e.g. produce biomass, consume and/or produce specific metabolites).
If yes: **go to 5**. Else: **go to 4**.
4. Determine a reaction (or a reaction set) to add back to the model, and **return to 3**.
Note that different algorithms take different (computational) strategies to identify the reaction(s) to add back.
5. Obtain the final lifestage-specific model.

In the last 15 years, multiple software packages have been developed to automate this task and get to models that can perform desired metabolic functions. Two popular algorithms for generating context-specific models are GIMME [66] and tINIT [67]; ftINIT [68], and we refer the reader to a recent review/benchmark of some of these methods [69] for a more in-depth discussion.

Naming the lifestage-specific models

The naming of the lifestage-specific models should follow the naming of the PAN-model (see Section 2). We propose to append that name with an underscore and a three-letter

abbreviation that reflects the lifestage (e.g. **_try** for trypomastigote or **_bsf** for bloodstream form).

4. Validation of the lifecycle-stage models using experimental data

Through the workflow presented above, the lifecycle-stage model will have been made with the prerequisite that it can function, *i.e.* that it can satisfy the constraints and objective(s) that was included in the workflow. So, if biomass formation was set there to the experimentally determined growth rate, the resulting model should be able to sustain that growth rate. But it may do this through any flux distribution through the network, which does not necessarily correspond to experimentally measured metabolic fluxes.

To determine whether the predicted flux distributions make sense, the stage-specific models can be further validated with experimental data from that life-stage. The experimental validation of GEMs must be designed according to the specific objective function that was defined in its construction. In most cases, the objective function is a simulation of proliferation, or biomass production as the growing life-cycle stages have been more extensively studied. Below, we will describe some common experimental data(sets) that can be used to validate the model.

Known auxotrophies

Many protistan parasites are auxotrophic for certain essential metabolites, such as amino acids and nitrogenous bases, and are thus unable to synthesize them (instead, relying on uptake from the host or extracellular environment). Refined metabolic models must accurately reproduce these auxotrophies *in silico* as part of the validation routine. For this, one should restrict the lower exchange bound for the essential metabolite to zero (*i.e.* do not allow uptake); the model should report no growth [70].

Known metabolic end products

For many parasites there will be known end-products that they produce and/or metabolic maps for the specific lifecycle stage (e.g. [71]). In addition, quantitative exo-metabolomic data are a source of information about the organism's metabolism. The model should be able to reproduce these findings. To evaluate this, flux distributions for the model can be calculated, with the medium (and potentially measured uptake fluxes) as constraints. With FBA, a flux distribution can be found at the minimal or maximal value of the objective function, e.g. at the maximal growth rate. In addition, FVA can predict the flux ranges that are possible for all reactions at a certain value of the objective function, e.g. for the growth rate that is experimentally observed. These simulated metabolite profiles can then be compared to the experimental data.

Gene essentiality screens

The information from gene essentiality screens can be used to measure, in a certain way, the consistency of the reconstruction. One can simulate gene knockdowns by lowering the upper flux bound of a reaction; in a similar vein, setting the flux bounds of a reaction to zero will mimic a knock-out. The effect on the objective function value can be then compared to datasets from genetically modified cells (e.g. by using methods such as CRISPR/Cas9 (see e.g. [72] for *T. brucei* and [73] for *T. cruzi* or RNAi (for *T. brucei*: [74])).

In a gene-essentiality analysis, the genes fall into one of four groups: True positive (TP), True negative (TN), False positive (FP) and False negative (FN). A “true” (TP and TN) means that computational prediction matches the experimental data: the gene is essential for growth (TP) or not (TN). In the FP situation, when non-essential genes are predicted to be essential, it is important to check the model for missing reactions to identify potential metabolic routes that bypass the reaction using the enzyme encoded by the FP gene. Lastly, the FNs - the genes that are known to be essential, but are predicted to be non-essential - must be evaluated with caution. This suggests that a bypass might exist in the model unless the model is sufficiently constrained with flux data.

There are several points to consider when comparing the model predictions to the experimental dataset. First, it is important how essentiality is defined. Is a gene essential for growth when upon depletion the cells die or when growth is reduced below a threshold (say, below 5% of the original growth rate)? Model predictions should be evaluated with the same threshold in mind. Furthermore, be aware of the size of the experimental dataset when reporting the fraction of correct predictions (and comparing this to other models). If the experimental dataset is small, the percentage of correct predictions may be artificially high.

5. Scientific questions for the lifecycle-stage specific models

After the construction and validation of the lifecycle stage specific model, the model can be used for *in silico* studies. Here we name a few options - often inspired from outside the parasite field.

Comparing network structures between lifecycle-stages

The different lifecycle stage models for a parasite (and between parasites) can be compared. This can both be done for the entire network and for subsystems (pathways). The *T. cruzi* lifecycle-stage models showed significant differences among metabolic pathways [11]. Enzymes or even whole metabolic pathways that were not present in some lifecycle stage models usually resulted in alternative flux distributions. It is possible to use the models to systematically inspect these flux changes and to investigate *in silico* to which differences these changes can be attributed. For this, the absence of an enzyme or a pathway can be simulated in a stage in which it is present, or the opposite, enzymes or pathways that are absent in a stage, can be included *in silico* to inspect their impact in the flux distribution.

Prediction of intracellular fluxes

The model can predict which (parts of) pathways are operational inside the cells. The more constraints on the extracellular fluxes, the more reliable the predicted intracellular fluxes will be. In liver cancer, the predicted intracellular fluxes revealed why glutamate was excreted as a metabolic end-product [75].

Predicting essential genes beyond the validation dataset

The gene essentiality analysis focused on the genes that were tested in experimental datasets, so a model-wide gene essentiality test may reveal novel drug targets (e.g. [11]). The reactions that turn out to be non-essential for growth may be further evaluated in a double-knockout set-up, where all possible reaction pairs are selected and both reactions of the pair are turned off at the same time [76].

Design of optimized culture media

Some cells may be limited in their growth because the medium does not contain all components they need in their optimal concentrations. Through a reduced cost analysis, it can be identified which metabolite(s) currently limit(s) growth. This can steer medium design. Interestingly, such an analysis may also show which chemical elements are scarce in the medium and can identify which metabolites could potentially provide this element given the organisms' metabolic capabilities. In case of several alternative metabolites, availability and costs may be taken into account for the final choice. Such an analysis has been previously done to allow bulk production of the whooping cough pathogen *Bordetella pertussis* [31].

The above are just a few examples of what is possible. The power of GEMs is that they give access to the genome-encoded metabolic potential of the organism. The power of lifestage-specific GEMs is that they go beyond the genome sequence to explore the relevant submaps that are expressed at a particular stage in the life cycle.

Acknowledgements

This work was supported by Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) grant 2021/12938-0 (awarded to A.M.S.) and Conselho Nacional de Pesquisas Científicas e Tecnológicas (CNPq) grant 307487/2021-0 (awarded to A.M.S.). P.G. also thanks SURF (<https://www.surf.nl>) for computational support through the access to the Dutch National Supercomputer Snellius.

References

1. Judge, A. and M.S. Dodd, *Metabolism*. Essays Biochem, 2020. **64**(4): p. 607-647.
2. David D. van Niekerk, G.P., François du Toit, Jacky L. Snoep, Barbara M. Bakker, Jurgen R. Haanstra, *Attacking Blood-Borne Parasites with Mathematics*, in *Comprehensive Analysis of Parasite Biology: From Metabolism to Drug Discovery: From Metabolism to Drug Discovery*, R.C. Sylke Müller, Ovidiu Radulescu, Editor. 2016, Wiley-VCH Verlag GmbH & Co. KGaA. p. 513-541.
3. Stelling, J., *Mathematical models in microbial systems biology*. Curr Opin Microbiol, 2004. **7**(5): p. 513-8.
4. Bakker, B.M., P.A. Michels, F.R. Opperdoes, and H.V. Westerhoff, *Glycolysis in bloodstream form Trypanosoma brucei can be understood in terms of the kinetics of the glycolytic enzymes*. J Biol Chem, 1997. **272**(6): p. 3207-15.
5. Kerkhoven, E.J., et al., *Handling uncertainty in dynamic models: the pentose phosphate pathway in Trypanosoma brucei*. PLoS Comput Biol, 2013. **9**(12): p. e1003371.
6. Saavedra, E., et al., *Kinetic modeling can describe in vivo glycolysis in Entamoeba histolytica*. FEBS J, 2007. **274**(18): p. 4922-40.
7. Bexkens, M.L., et al., *The unusual kinetics of lactate dehydrogenase of Schistosoma mansoni and their role in the rapid metabolic switch after penetration of the mammalian host*. Int J Parasitol, 2024. **54**(7): p. 367-378.
8. van Niekerk, D.D., et al., *A detailed kinetic model of glycolysis in Plasmodium falciparum-infected red blood cells for antimalarial drug target identification*. J Biol Chem, 2023. **299**(9): p. 105111.
9. Montagud, A., et al., *Reconstruction and analysis of genome-scale metabolic model of a photosynthetic bacterium*. BMC Syst Biol, 2010. **4**: p. 156.
10. Chiappino-Pepe, A., V. Pandey, and O. Billker, *Genome reconstructions of metabolism of Plasmodium RBC and liver stages*. Curr Opin Microbiol, 2021. **63**: p. 259-266.
11. Shiratsubaki, I.S., et al., *Genome-scale metabolic models highlight stage-specific differences in essential metabolic pathways in Trypanosoma cruzi*. PLoS Negl Trop Dis, 2020. **14**(10): p. e0008728.
12. Terzer, M., N.D. Maynard, M.W. Covert, and J. Stelling, *Genome-scale metabolic networks*. Wiley Interdiscip Rev Syst Biol Med, 2009. **1**(3): p. 285-297.
13. Bruggeman, F.J. and H.V. Westerhoff, *The nature of systems biology*. Trends Microbiol, 2007. **15**(1): p. 45-50.
14. Bordbar, A., J.M. Monk, Z.A. King, and B.O. Palsson, *Constraint-based models predict metabolic and associated cellular functions*. Nat Rev Genet, 2014. **15**(2): p. 107-20.
15. Cuevas, D.A., et al., *From DNA to FBA: How to Build Your Own Genome-Scale Metabolic Model*. Front Microbiol, 2016. **7**: p. 907.
16. Thiele, I. and B.O. Palsson, *A protocol for generating a high-quality genome-scale metabolic reconstruction*. Nat Protoc, 2010. **5**(1): p. 93-121.
17. Machado, D., S. Andrejev, M. Tramontano, and K.R. Patil, *Fast automated reconstruction of genome-scale metabolic models for microbial species and communities*. Nucleic Acids Res, 2018. **46**(15): p. 7542-7553.
18. Alstott, J., et al., *Modeling the impact of lesions in the human brain*. PLoS Comput Biol, 2009. **5**(6): p. e1000408.
19. Mendoza, S.N., B.G. Olivier, D. Molenaar, and B. Teusink, *A systematic assessment of current genome-scale metabolic reconstruction tools*. Genome Biol, 2019. **20**(1): p. 158.
20. Agren, R., et al., *The RAVEN toolbox and its use for generating a genome-scale metabolic model for Penicillium chrysogenum*. PLoS Comput Biol, 2013. **9**(3): p. e1002980.
21. Gu, C., et al., *Current status and applications of genome-scale metabolic models*. Genome Biol, 2019. **20**(1): p. 121.

22. Monk, J.M., et al., *iML1515, a knowledgebase that computes Escherichia coli traits*. Nat Biotechnol, 2017. **35**(10): p. 904-908.
23. Marinos, G., C. Kaleta, and S. Waschina, *Defining the nutritional input for genome-scale metabolic models: A roadmap*. PLoS One, 2020. **15**(8): p. e0236890.
24. Orth, J.D., I. Thiele, and B.O. Palsson, *What is flux balance analysis?* Nat Biotechnol, 2010. **28**(3): p. 245-8.
25. Herrmann, H.A., et al., *Flux sampling is a powerful tool to study metabolism under changing environmental conditions*. NPJ Syst Biol Appl, 2019. **5**: p. 32.
26. Prieto-de Lima, T.S., K. Rojas-Jimenez, and C. Vaglio, *Strategy for Optimizing Vitamin B12 Production in Pseudomonas putida KT2440 Using Metabolic Modeling*. Metabolites, 2024. **14**(11).
27. Pelicaen, R., D. Gonze, L. De Vuyst, and S. Weckx, *Genome-scale metabolic modeling of Acetobacter pasteurianus 386B reveals its metabolic adaptation to cocoa fermentation conditions*. Food Microbiol, 2020. **92**: p. 103597.
28. Lieven, C., et al., *MEMOTE for standardized genome-scale metabolic model testing*. Nat Biotechnol, 2020. **38**(3): p. 272-276.
29. Raman, K., et al., *FROG Analysis Ensures the Reproducibility of Genome Scale Metabolic Models*. bioRxiv, 2024: p. 2024.09.24.614797.
30. Orth, J.D., et al., *A comprehensive genome-scale reconstruction of Escherichia coli metabolism--2011*. Mol Syst Biol, 2011. **7**: p. 535.
31. Branco Dos Santos, F., et al., *Probing the Genome-Scale Metabolic Landscape of Bordetella pertussis, the Causative Agent of Whooping Cough*. Appl Environ Microbiol, 2017. **83**(21).
32. Zhang, C., et al., *Yeast9: a consensus genome-scale metabolic model for S. cerevisiae curated by the community*. Mol Syst Biol, 2024. **20**(10): p. 1134-1150.
33. Wang, H., et al., *Genome-scale metabolic network reconstruction of model animals as a platform for translational research*. Proc Natl Acad Sci U S A, 2021. **118**(30).
34. Anton, M., et al., *standard-GEM: standardization of open-source genome-scale metabolic models*. bioRxiv, 2023: p. 2023.03.21.512712.
35. Lahtee, P.J., et al., *Protein turnover forms one of the highest maintenance costs in Lactococcus lactis*. Microbiology (Reading), 2014. **160**(Pt 7): p. 1501-1512.
36. Elsemman, I.E., et al., *Whole-cell modeling in yeast predicts compartment-specific proteome constraints that drive metabolic strategies*. Nat Commun, 2022. **13**(1): p. 801.
37. Michels, P.A.M., et al., *Carbohydrate metabolism in trypanosomatids: New insights revealing novel complexity, diversity and species-unique features*. Exp Parasitol, 2021. **224**: p. 108102.
38. Liu, Y., A. Beyer, and R. Aebersold, *On the Dependency of Cellular Protein Levels on mRNA Abundance*. Cell, 2016. **165**(3): p. 535-50.
39. Zikova, A., et al., *A paradigm shift: The mitoproteomes of procyclic and bloodstream Trypanosoma brucei are comparably complex*. PLoS Pathog, 2017. **13**(12): p. e1006679.
40. Shameer, S., et al., *TrypanoCyc: a community-led biochemical pathways database for Trypanosoma brucei*. Nucleic Acids Res, 2015. **43**(Database issue): p. D637-44.
41. Doyle, M.A., et al., *LeishCyc: a biochemical pathways database for Leishmania major*. BMC Syst Biol, 2009. **3**: p. 57.
42. Stouthamer, A.H. and C. Bettenhausen, *Utilization of energy for growth and maintenance in continuous and batch cultures of microorganisms. A reevaluation of the method for the determination of ATP production by measuring molar growth yields*. Biochim Biophys Acta, 1973. **301**(1): p. 53-70.
43. Haanstra, J.R., et al., *Proliferating bloodstream-form Trypanosoma brucei use a negligible part of consumed glucose for anabolic processes*. Int J Parasitol, 2012. **42**(7): p. 667-73.

44. Pineda, E., et al., *Glycerol supports growth of the Trypanosoma brucei bloodstream forms in the absence of glucose: Analysis of metabolic adaptations on glycerol-rich conditions*. PLoS Pathog, 2018. **14**(11): p. e1007412.
45. Teusink, B., et al., *Analysis of growth of Lactobacillus plantarum WCFS1 on a complex medium using a genome-scale metabolic model*. J Biol Chem, 2006. **281**(52): p. 40041-8.
46. Simensen, V., et al., *Experimental determination of Escherichia coli biomass composition for constraint-based metabolic modeling*. PLoS One, 2022. **17**(1): p. e0262450.
47. Benthin, S., J. Nielsen, and J. Villadsen, *A simple and reliable method for the determination of cellular RNA content*. Biotechnology Techniques, 1991. **5**(1): p. 39-42.
48. Smith, P.K., et al., *Measurement of protein using bicinchoninic acid*. Anal Biochem, 1985. **150**(1): p. 76-85.
49. Bradford, M.M., *A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding*. Anal Biochem, 1976. **72**: p. 248-54.
50. Van Handel, E., *Rapid determination of total lipids in mosquitoes*. J Am Mosq Control Assoc, 1985. **1**(3): p. 302-4.
51. Kawai, Y. and A. Moribayashi, *Characteristic lipids of Bordetella pertussis: simple fatty acid composition, hydroxy fatty acids, and an ornithine-containing lipid*. J Bacteriol, 1982. **151**(2): p. 996-1005.
52. Van Handel, E., *Rapid determination of glycogen and sugars in mosquitoes*. J Am Mosq Control Assoc, 1985. **1**(3): p. 299-301.
53. Masuko, T., et al., *Carbohydrate analysis by a phenol-sulfuric acid method in microplate format*. Anal Biochem, 2005. **339**(1): p. 69-72.
54. Parrou, J.L. and J. Francois, *A simplified procedure for a rapid and reliable assay of both glycogen and trehalose in whole yeast cells*. Anal Biochem, 1997. **248**(1): p. 186-8.
55. Saldanha, A.J., M.J. Brauer, and D. Botstein, *Nutritional homeostasis in batch and steady-state culture of yeast*. Mol Biol Cell, 2004. **15**(9): p. 4089-104.
56. Lachance, J.C., et al., *BOFdat: Generating biomass objective functions for genome-scale metabolic models from experimental data*. PLoS Comput Biol, 2019. **15**(4): p. e1006971.
57. Grigaitis, P., et al., *A Computational Toolbox to Investigate the Metabolic Potential and Resource Allocation in Fission Yeast*. mSystems, 2022. **7**(4): p. e0042322.
58. Oppendoes, F.R. and P.A. Michels, *Complex I of Trypanosomatidae: does it exist?* Trends Parasitol, 2008. **24**(7): p. 310-7.
59. Clarkson, A.B., Jr., E.J. Bienen, G. Pollakis, and R.W. Grady, *Respiration of bloodstream forms of the parasite Trypanosoma brucei brucei is dependent on a plant-like alternative oxidase*. J Biol Chem, 1989. **264**(30): p. 17770-6.
60. Gnaiger Erich, M.T.G. *Mitochondrial physiology*. Bioenerg Commun 2020 [cited 2020; 1-44].
61. Hinkle, P.C., *P/O ratios of mitochondrial oxidative phosphorylation*. Biochim Biophys Acta, 2005. **1706**(1-2): p. 1-11.
62. Alencar, M.B., et al., *The role of l-serine and l-threonine in the energy metabolism and nutritional stress response of Trypanosoma cruzi*. mSphere, 2025. **10**(3): p. e0098324.
63. Alencar, M.B., et al., *Energy metabolism in Trypanosoma cruzi: the validated and putative bioenergetic and carbon sources*. mBio, 2025: p. e0221524.
64. Faria, J.D.N., et al., *Inhibition of L-threonine dehydrogenase from Trypanosoma cruzi reduces glycine and acetate production and interferes with parasite growth and viability*. J Biol Chem, 2025. **301**(2): p. 108080.
65. Souza, R.O.O., et al., *Fatty acid oxidation participates in resistance to nutrient-depleted environments in the insect stages of Trypanosoma cruzi*. PLoS Pathog, 2021. **17**(4): p. e1009495.

66. Becker, S.A. and B.O. Palsson, *Context-specific metabolic networks are consistent with experiments*. PLoS Comput Biol, 2008. **4**(5): p. e1000082.
67. Agren, R., et al., *Identification of anticancer drugs for hepatocellular carcinoma through personalized genome-scale metabolic modeling*. Mol Syst Biol, 2014. **10**(3): p. 721.
68. Gustafsson, J., et al., *Generation and analysis of context-specific genome-scale metabolic models derived from single-cell RNA-Seq data*. Proc Natl Acad Sci U S A, 2023. **120**(6): p. e2217868120.
69. Molversmyr, H., O. Oyas, F. Rotnes, and J.O. Vik, *Extracting functionally accurate context-specific models of Atlantic salmon metabolism*. NPJ Syst Biol Appl, 2023. **9**(1): p. 19.
70. Renz, A., L. Widerspick, and A. Drager, *First Genome-Scale Metabolic Model of Dolosigranulum pigrum Confirms Multiple Auxotrophies*. Metabolites, 2021. **11**(4).
71. Muller, M., et al., *Biochemistry and evolution of anaerobic energy metabolism in eukaryotes*. Microbiol Mol Biol Rev, 2012. **76**(2): p. 444-95.
72. Kovarova, J., et al., *CRISPR/Cas9-based precision tagging of essential genes in bloodstream form African trypanosomes*. Mol Biochem Parasitol, 2022. **249**: p. 111476.
73. Picchi-Constante, G.F.A., et al., *Efficient CRISPR-Cas9-mediated genome editing for characterization of essential genes in Trypanosoma cruzi*. STAR Protoc, 2022. **3**(2): p. 101324.
74. Alsford, S., et al., *High-throughput phenotyping using parallel sequencing of RNA interference targets in the African trypanosome*. Genome Res, 2011. **21**(6): p. 915-24.
75. Nilsson, A., et al., *Quantitative analysis of amino acid metabolism in liver cancer links glutamate excretion to nucleotide synthesis*. Proc Natl Acad Sci U S A, 2020. **117**(19): p. 10294-10304.
76. Goldstein, Y.A. and A. Bockmayr, *Double and multiple knockout simulations for genome-scale metabolic network reconstructions*. Algorithms Mol Biol, 2015. **10**(1): p. 1.