UNIVERSITY OF CALIFORNIA SAN DIEGO

Essential metabolic pathways in *Trypanosoma cruzi*

A Thesis submitted in partial satisfaction of the
requirements for the degree
Master of Science

in

Bioengineering

by

Isabel Souza Shiratsubaki

Committee in charge:

Professor Jair Lage De Siqueira Neto, Chair
Professor Bernhard O. Palsson, Co-Chair
Professor Shankar Subramaniam

2019

The Thesis of Isabel Souza Shiratsubaki as it is listed on UC San Diego Academic Records is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

_____

_____
Co-Chair

_____
Chair

University of California San Diego

2019

DEDICATION

To my family and husband Marty for all the support.

To all people suffering from Chagas disease.

EPIGRAPH

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

—Marie Curie

TABLE OF CONTENTS

LIST OF SUPPLEMENTAL FILES

shiratsubaki_supplementary_tables.xlsx
shiratsubaki_supplemental_figure1_metabolic_network_map.pdf
shiratsubaki_supplemental_figure2_FBA_epimastigote.pdf
shiratsubaki_supplemental_figure3_FBA_amastigote.pdf
shiratsubaki_supplemental_figure4_FBA_trypomastigote.pdf
shiratsubaki_supplemental_figure5_model_expansion.pdf

LIST OF FIGURES

ACKNOWLEDGEMENTS

I would first like to express my gratitude and appreciation for my advisor Dr. Jair Lage de Siqueira-Neto, who gave me the opportunity to join his lab and participate in this project. Despite persisting through medical issues that often pulled him away from his office, he was still able to offer guidance and help me throughout this process. Words cannot describe how thankful I am to him. He believed in the potential of this work, and made key contributions to improve and prepare it for publication. His positive attitude, insight, and guidance were invaluable.

I would also like to thank my co-advisor Dr. Bernhard Palsson for giving me the opportunity to join his lab and providing me the perfect work environment to run my simulations for my thesis. He provided me insightful and sincere feedback to improve my work for publication, and I am very grateful for that.

My thanks also go to Xin Fang, PhD student under Dr. Palsson. She was very helpful in guiding me during this entire process, in different phases of this work. She helped to develop this projects direction, get in touch with Dr. Palsson, and write the material to be published as first author as well. I have no words to describe how grateful I am for all her support.

I also have to thank Dr. Ariel Silber, from Universidade de Sao Paulo, and his Post-doctoral researcher Rodolpho Ornitz. Their expertise in *Trypanosoma cruzi* metabolism was very valuable for adding insightful discussion into this work. My thanks also go to Dr. Shankar Subramaniam, the third member of my committee, who provided me honest and fruitful feedback about this work. In addition, my MS journey could not have been happier without the amazing people I met in CDIPD and McKerrow Lab and Systems Biology Research Group. I am very grateful to have had the opportunity to work alongside them.

Furthermore, I have to thank all the great people I met in the Linguistics Department at UCSD. They gave me the opportunity to have a Teaching Assistantship that was able to help me financially during the two years of my MS program. A special thanks to Elke Riebeling and Stéphanie Gaillard, my 1st and 2nd-year French Coordinators, for all the guidance. I have to say

that I learned a lot from both of them.

Last but not least, I have to thank all people that emotionally supported me in the obstacles and challenges I faced during this journey. A very special thanks to my family: my parents Carmen and Marcos for being my life guiders and for providing me the best education they could afford, my sisters Taís and Luísa for all the emotional support and advice, my husband Marty for all the patience, support, advice, and companionship, and all my friends. A special thanks to my friends in the USA, Yen, Zack, Muyao, Bruno, and Luciana, my friends in Brazil, Aline, Karina, Lara, Laura, Paula, and Rodrigo, my friend in Belgium, Maria Olívia, and my friend in Portugal, Cibele. You all made this journey be funnier and lighter.

Chapter 2 (”Essential metabolic pathways in *Trypanosoma cruzi*”) and Chapter 3 (”Supplementary material”) are currently being prepared for publication by Shiratsubaki, Isabel S.; Fang, Xin; Siqueira-Neto, Jair L.; and Palsson, Bernhard . The thesis author was the primary investigator and one of the two primary authors of this paper.

VITA

2010-2011	Undergraduate Researcher Hydraulics Department, Universidade de Sao Paulo (USP), Brazil
2014	Bachelor of Science in Civil & Structural Engineering Universidade de Sao Paulo (USP), Brazil
2014-2015	Structural Engineer Formula Projetos Estruturais, Brazil
2016	Bachelor of Science & Master of Engineering in General Engineering Ecole Centrale Paris - Centrale Supelec (ECP), France
2017-2019	French Teaching Assistant & Instructor Linguistics Department, University of California San Diego (UCSD)
2017-2019	Research Assistant CDIPD and McKerrow Lab and Systems Biology Research Group, University of California San Diego (UCSD)
2019	Master of Science in Bioengineering University of California San Diego (UCSD)

ABSTRACT OF THE THESIS

Essential metabolic pathways in *Trypanosoma cruzi*

by

Isabel Souza Shiratsubaki

Master of Science in Bioengineering

University of California San Diego, 2019

Professor Jair Lage De Siqueira Neto, Chair
Professor Bernhard O. Palsson, Co-Chair

Chagas disease caused by a protozoan called *Trypanosoma cruzi* is a neglected tropical disease and a leading cause of heart failure in Latin America where it is endemic. Due to migration of asymptomatic infected population, it is now also present in North America, Europe, Japan and Australia. Drugs currently available to treat Chagas disease are limited to benznidazole and nifurtimox, both with severe side effects, and are more efficient when administered early in the course of the infection. Thus, there is a need for the discovery of alternative and more efficient drugs. In the identification process of potential drug targets, metabolic genes are good candidates, since they are critical for cellular growth and survival [1]. In addition, genome-scale

metabolic models (GEMs) [31] have been developed to accurately predict metabolic capabilities from genome sequences. This work developed an extended GEM, hereafter referred to as iIS312, of the published and validated *T. cruzi* CL Brener core metabolism model iSR215 [35]. The life cycle of the parasite is divided in three distinct stages: epimastigotes (replicative and insect-specific), trypomastigotes or metacyclic epimastigote (infective and non-replicative) and amastigotes (replicative and intracellular in humans). From iIS312, we built three stage-specific models using transcriptomics data [3] integration. Such models for other organisms (e.g. *Plasmodium falciparum*) have provided valuable insight of stage-specific changes in metabolism [1]. The stage-specific models were able to predict metabolic differences among the three *T. cruzi* stages, including essential genes and reactions, and flux distribution. The stage-specific models showed that epimastigotes present the most active metabolism among the stages, with least number of deactivated reactions. The trypomastigote model predicted the non-essentiality of pathways responsible for nucleic acid synthesis as this stage is non-replicative, and presented full metabolic activity in TCA for energy yield to move around. The amastigote model presented low activity in Glycolysis, as observed in previous studies, suggesting that this pathway is not essential in amastigotes. It also indicated that amastigotes take a shortcut in TCA, increasing the metabolic flux in the malic enzyme, as described in previous studies. Moreover, the gene essentiality predictions suggests potential drug targets, among which some have been proven lethal previously, including glutamate dehydrogenase [12], glucokinase and hexokinase [38]. Our results represents potentially a step forward towards the improvement of Chagas disease treatment. To our knowledge, these stage-specific models are the first GEM built for the stages amastigote and trypomastigote. This work is also the first to present an *in silico* GEM comparison among different stages in *T. cruzi* life-cycle.

Chapter 1

Introduction

Chagas disease is a neglected tropical disease affecting millions of people specially in Latin America where it is endemic [42]. The transmission can occur vertically from infected mother to baby during gestation, by blood transfusion or organ transplant, orally (by contaminated food), in laboratory accidents and most commonly by an insect vector from the family Reduviidae (genus Triatoma) popularly known as kissing bug that carries the parasite *Trypanosoma cruzi*, a protozoan of the family Trypanosomatidae. *T. cruzi* life cycle is composed by vertebrate hosts and invertebrate vectors, and three stage-specific forms: the epimastigote: axenic form that replicates in the midgut of the insect vector; trypomastigote: infectious non-replicative form present in the mammalian host blood circulation and in the proboscides of the insect vector (called metacyclic epimastigote in the latter case); and amastigote: intracellular replicative stage found in the cell of the mammalian host.

Although the stages were first defined by their morphological characteristics [8], they also present differences at the cellular level, including the composition of their surface and their energy metabolism [3]. Findings in the literature indicate that epimastigotes have an active metabolism (catabolism and anabolism), and use different nutrients as energy source (lipids, proteins and sugars) [5, 6]. In contrast, trypomastigotes present low levels of transcription and translation, and since their main function in *T. cruzi* cycle is infection, they are specialized in attachment and infection of the host cells [3]. Finally, very little information is available regarding amastigote metabolism. It is known that amastigotes are metabolically more active than trypomastigotes

but are not as responsive as epimastigotes to the change in the environment and use of different nutrients [16].

In addition to the multi-stage aspect of *T. cruzi* cycle, the parasite can lead to two distinct clinical phases in the infected individual: acute, and chronic. The acute phase is characterized by the presence of parasites in the blood circulation. It is typically asymptomatic, but mild symptoms, including fever, fatigue, body aches, headaches and rash can occur and last for 8 to 12 weeks, when it is controlled by the hosts immune system. The chronic phase starts with an asymptomatic stage (called indeterminate chronic phase) that can last as long as 2 or 3 decades and is characterized by an undetectable parasitemia in the circulation. For reasons yet to be understood, about 30% of the infected individuals will manifest cardiac and about 5%, digestive complications. When the disease is clinically manifested in the chronic phase it is called determinate chronic phase. Current drug options to treat Chagas patients are limited to benznidazole and nifurtimox. Both drugs can cause severe gastrointestinal, dermatological, and neurological side effects, and are more efficient when administered in the acute phase than in the chronic phase. In the chronic phase, the efficacy of these drugs is questionable to prevent or treat the cardiac or digestive symptoms, regardless if they are used in indeterminate or determinate chronic phases [13, 17]. Thus, there is a clear unmet medical need and alternative and more efficient drugs must be discovered.

In the process of the identification of potential drug targets, metabolic genes are good candidates, since they are critical for cellular growth and survival [1]. In addition, genome-scale metabolic models (GEMs) [31] have been developed to accurately predict metabolic capabilities from genome sequences. Genome-scale metabolic network reconstructions are knowledge bases that map genotypes to phenotypes. Such reconstructions can then be converted to mathematical models - GEMs to enable computation of metabolic functions of the organism on a systems level. This study presents an extended genome-scale metabolic reconstruction model, hereafter referred to as iIS312, of the published and validated *T. cruzi* CL Brener core metabolism model iSR215 [35]. From the extended model iIS312, we built three stage-specific models using transcriptomics

data [3] integration (See Figure 2.1), as such models for other organisms (e.g. *Plasmodium falciparum*) have provided valuable insight of stage-specific changes in metabolism [1]. The resulting models of *T. cruzi* describe stage-specific differences in metabolism including essential genes, and consequently potential drug targets.

In Chapter 2, we present the results, discussion, and material and methods involved in building iIS312 (pan and the stage-specific models), and identifying essential genes in *T. cruzi* and main metabolic flux differences among the three stages of the parasite.

In Chapter 3, we present the supplementary material of the work "Essential metabolic pathways in *Trypanosoma cruzi*", presented in Chapter 2.

Lastly, Chapter 4 presents a conclusion of all the findings of this thesis. It also presents the ongoing work and future directions related to this work.

Chapter 2

Essential metabolic pathways in *Trypanosoma cruzi*

2.1 Results

2.1.1 Reconstructing and expanding the metabolic network of *Trypanosoma cruzi*, iIS312

The model reconstruction and curation of *T. cruzi* (Figure 2.1, Step 1) was developed using the genome annotation of the strain Dm28c [19]. The curation was based on the two pre-existing validated models: iSR215 (*T. cruzi* strain CL Brener) [35] and iAC560 (*Leishmania major* strain Friedlin) [9]. iAC560 model was the first metabolic network reconstruction built for a protozoan. In addition, *L. major* is a close-related organism to *T. cruzi*, as they both belong to the same family Trypanosomatidae. The cross-reference between genes in *T. cruzi* Dm28c and *T. cruzi* CL Brener and *L. major* Friedlin was done using the database TriTrypDB [2], which is a specific database for pathogens of the family Trypanosomatidae (*Leishmania* and *Trypanosoma* genera).

The expanded *T. cruzi* model iIS312 accounts for 519 reactions, 604 metabolites, 312 genes and 6 compartments (Figure 2.2), including the cytosol, mitochondria, glycosome, endoplasmic reticulum, nucleus, and the extracellular compartment. Among the 519 reactions, 328 (63%) are gene-association and 145 (27%) have no gene association (Figure 2.2a) due to lack of genome annotation or exchanges reactions that simulate nutrient uptake. In comparison with the old *T. cruzi* model iSR215, iIS312 significantly expanded the scope of the model (Figure 2.2c), including 83 new reactions in the lipid metabolism which is absent in iSR215. Figure 2.1c

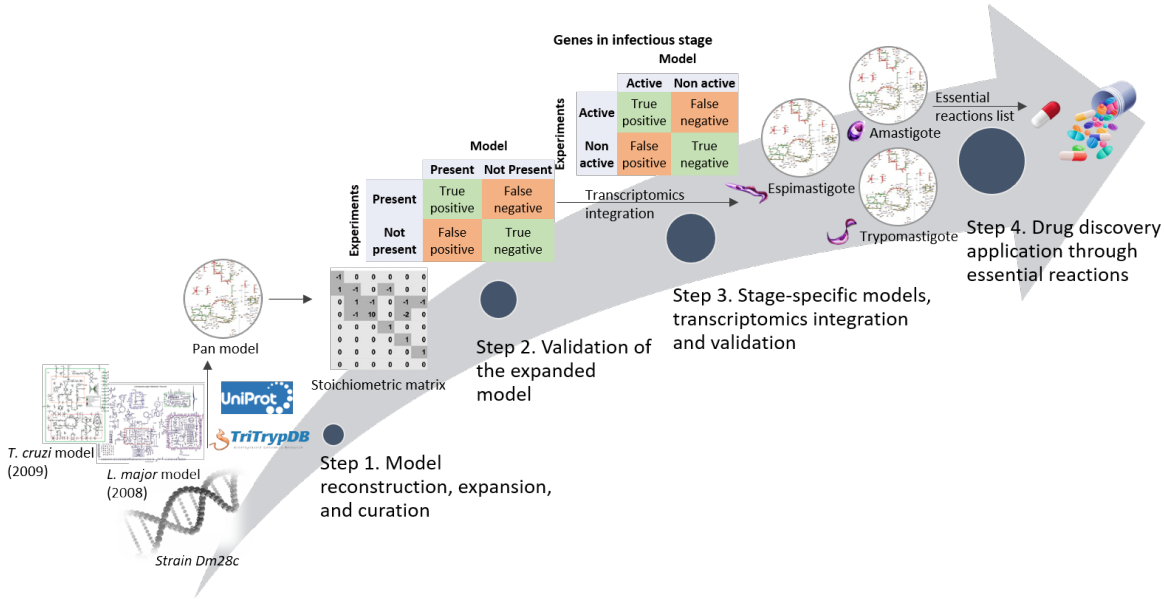

Figure 2.1: The workflow for the multi-stage reconstructions of *T. cruzi*. Step 1. The reconstruction of *T. cruzi* multi-stages used the genome annotation of the strain Dm28c [19], the models iSR215 (*T. cruzi* strain CL Brener) [35] and iAC560 (*L. major* strain Friedlin) [9], the literature, the genomic database for pathogens of the family Trypanosomatidae (TriTryDB) [2], and other genomic databases. Step 2. The expanded model iIS312 was validated using experimental data. Step 3. Stage-specific models (epimastigote, amastigote, and trypomastigote) were generated through transcriptomics integration and validated using experimental data from the infectious stage. Step 4. Analysis of metabolic functions and potential application in drug discovery.

indicated that the number of reactions increased for all subsystems in the updated model iIS312 (See Supplemental file shiratsubaki_supplemental_figure5_model_expansion.pdf for more details).

iS312 was validated by comparing its simulated excreted metabolites and experimental findings (See Figure 2.1 Step 2). Literature suggests that the byproducts of *T. cruzi* metabolism include [37, 40, 18, 7], alanine [37, 18], CO_2 [4], acetate [37, 40, 18], and glycine [11]. The byproducts predicted by iIS312 pan model fit the expected spectrum of metabolite secretion (Figure 2.2b), suggesting the validity of the model. We also compared the experimental observation with the predicted metabolite secretion by stage-specific models (details of stage-specific models discussed in the next section). Succinate, acetate, alanine, H_2O, and H^+ showed up as byproduct in all the iIS312 models. The trypomastigote model presented all the excreted metabolites expected. The excreted metabolites changed according to the constraints applied in

a

Properties	Count
Genes	312
Reactions	519
Gene associated (metabolic/transport)	328 (63%)
No gene association (metabolic)	13 (2.5%)
No gene association (transport)	132 (25%)
Exchange	45
Demand (biomass)	1
Metabolites	606
Compartments	6

b

Excreted metabolites	iIS312			
	Pan	Epis	Amas	Trypo
acetate	X	X	X	X
alanine	X	X	X	X
CO_2	X	X		X
glycine			X	X
H^+	X	X	X	X
H_2O	X	X	X	X
NH_4				X
succinate	X	X	X	X

c

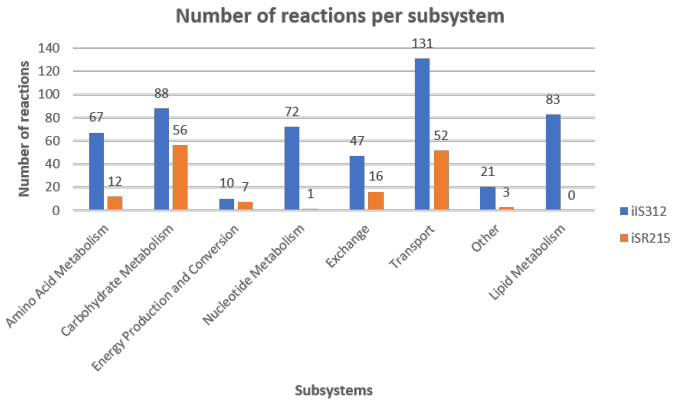

Figure 2.2: Content description, performance and validation of iIS312. (A) Properties of IS312, including number of genes, reactions, metabolites and compartments. (B) Validation of iIS312 predicted byproducts. (C) Distribution of reactions across subsystems in iIS312 and iSR215.

each model (e.g. deactivation of reactions through transcriptomics integration).

2.1.2 Stage-specific models generated through integration of transcriptomics data

Stage-specific models were generated from the pan model iS312 and transcriptomics by deactivating genes with low expression levels and their corresponding reactions. By integrating gene expression data with the pan iIS312 model, we generated three stage-specific models: amastigote, epimastigote, and trypomastigote (See Figure 2.1, Step 3). During the integration of gene expression data, reactions were deactivated if all or some of their encoding genes (depending on the gene rule; See methods section) presented an expression level below a predefined threshold (See methods section). To validate the stage-specific models, we compared the deactivated reactions in the infectious stage (trypomastigotes) with a list of kinases and phosphatases that were

shown to be active during trypomastigote described by Mattos et al [27]. Among a list of 120 active genes obtained from the gene description in Mattos et al study [27], only 3 were not expressed in trypomastigote model after transcriptomics integration. Those genes are associated with reactions that are encoded by more than one gene. Those genes were TCDM_00385 (6-phosphofructo-2-kinase/fructose-2, also encoded by TCDM_00348 and TCDM_01740, both expressed in the Trypomastigote model), TCDM_02136 and TCDM_06298 (mitogen-activated protein kinase, also encoded by TCDM_02679, TCDM_07669, TCDM_09327, all of them expressed in the Trypomastigote model). See Supplemental file shiratsubaki_supplementary_tables.xlsx Table S2 and S10 for more details.

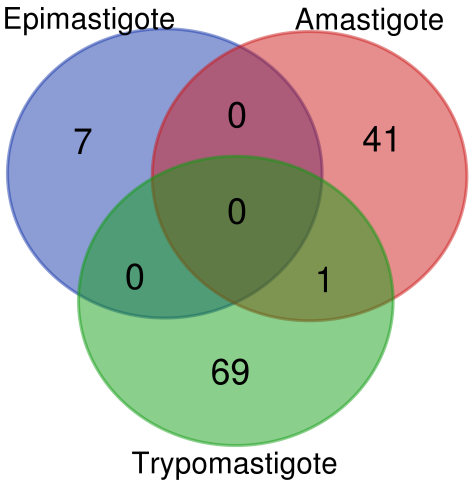

Figure 2.3: Venn diagram of deactivated reactions for each stage-specific iIS312 model (See Supplemental file shiratsubaki_supplementary_tables.xlsx Table S6 for more details).

The number of deactivated reactions differ for stage-specific models (Figure 2.3), showing consistency with literature finding regarding stage-specific metabolism activity. Out of all deactivated reactions, only 1 reaction overlaps between amastigote and trypomastigote models, suggesting substantial variation in metabolism between stages. The highest number of deactivated reactions was observed in the trypomastigote model (70 reactions), followed by amastigote (42 reactions) and epimastigote (7 reactions) models. The results are consistent with the findings in the literature regarding the metabolism of each stage: trypomastigote is the stage with the least

active metabolism, followed by amastigotes and epimastigotes.

2.1.3 Stage-specific models present differences among metabolic pathways

Variation in deactivated reactions across stages suggest stage-specific metabolic phenotypes. Most of the reactions deactivated in amastigotes (intracellular replicative stage) were involved in TCA cycle, Pentose Phosphate Pathway (PPP), Purine and Pyrimidine Metabolism. Literature [3] suggests that the carbohydrate catabolism is down-regulated in amastigotes, while the transmembrane transport, macromolecule metabolism and DNA replication are up-regulated. The weak simulated metabolic flux through some TCA cycle reactions in the amastigote model (See Figure 2.4) might suggest that amastigotes optimize their growth by using alternative shortcuts in TCA (e.g. through amino acid metabolic pathways as suggested by a previous studies) [39, 26]. In addition, the deactivations in Purine, and Pyrimidine Metabolism were mostly associated with lyases, which might be associated with the downregulation of nucleotide breakdown to favor DNA replication. In trypomastigotes (non-replicative and infective stage), the deactivations occurred in various subsystems, including Fatty Acid Biosynthesis, Glutamate metabolism, Glycerophospholipid metabolism, Glycolysis/Gluconeogenesis, Purine Metabolism, PPP, and Steroid Biosynthesis. Literature suggests that the nucleic acid metabolism is down-regulated in trypomastigotes, which is consistent with the deactivated reactions from Purine Metabolism and PPP in our model. In epimastigotes (replicative and insect-specific stage), the literature presented the downregulation of cell adhesion and none of any specific metabolic pathway. Our model does not show any prevalent subsystem deactivation, as it only includes reactions associated with metabolic functions but not cell adhesion.

In particular, we also observed that various essential genes (genes encoding enzymes that are required for growth) identified in the pan model were deactivated in stage-specific models, highlighting the significant changes in central metabolism across stages (See Figure 2.4). For amastigote, the 10 deactivated essential reactions are enriched in Glycolysis/Gluconeogenesis

(See Supplemental file shiratsubaki_supplementary_tables.xlsx Table S5), which is intriguing as the result suggest glycolysis is not essential for amastigote stage. This finding is consistent with a previous finding that glucose transporter activity was not detected in amastigote stage [39]. Therefore, it is likely that *T. cruzi* might have transporters for intermediate glycolytic metabolites, which explains the deactivation of reactions in Glycolysis in amastigotes. Confirmation of such transporters still needs further investigation in the future. Since our model simulates the parasite cell growth under steady-state, we kept the glycolytic genes active in the amastigote model. For trypomastigote, the deactivated reactions are mostly involved in Pentose Phosphate Pathway (PPP), which is responsible for nucleic acid synthesis and the production of NAHPH. This result is consistent with literature, since trypomastigotes are a non-replicative stage, which does not require nucleic acids for DNA synthesis. Therefore, we calibrate our biomass reaction for the trypomastigote model by removing metabolites involved in DNA synthesis (see Methods). For epimastigote, the only deactivated reaction is L-threonine dehydrogenase that is responsible for threonine metabolism, suggesting that threonine may not be essential for parasite growth, or alternative pathways that is missing in the genome annotation (See Supplemental file shiratsubaki_supplementary_tables.xlsx Table S5).

In addition to deactivated reaction analysis, the flux distribution map for each stage-specific model was generated (see methods section) to compare the metabolic differences among stages (See Supplemental files shiratsubaki_supplemental_figure2_FBA_epimastigote.pdf, shiratsubaki_supplemental_figure3_FBA_amastigote.pdf, and shiratsubaki_supplemental_figure4_FBA_trypomastigote.pdf for more details). Flux distribution map was generated using FBA with the objective set as optimal growth. As shown in Figure 2.4, the flux distribution in the central metabolism among the stages changed considerably. The reactions in gray are not necessarily inactive in the model, but they indicate weak simulated metabolic flux. While in the amastigote model the flux in TCA cycle takes a shortcut through proline and glutamate pathway (entering in TCA cycle through alpha-ketoglutarate), in the trypomastigote model we have opposite behavior,

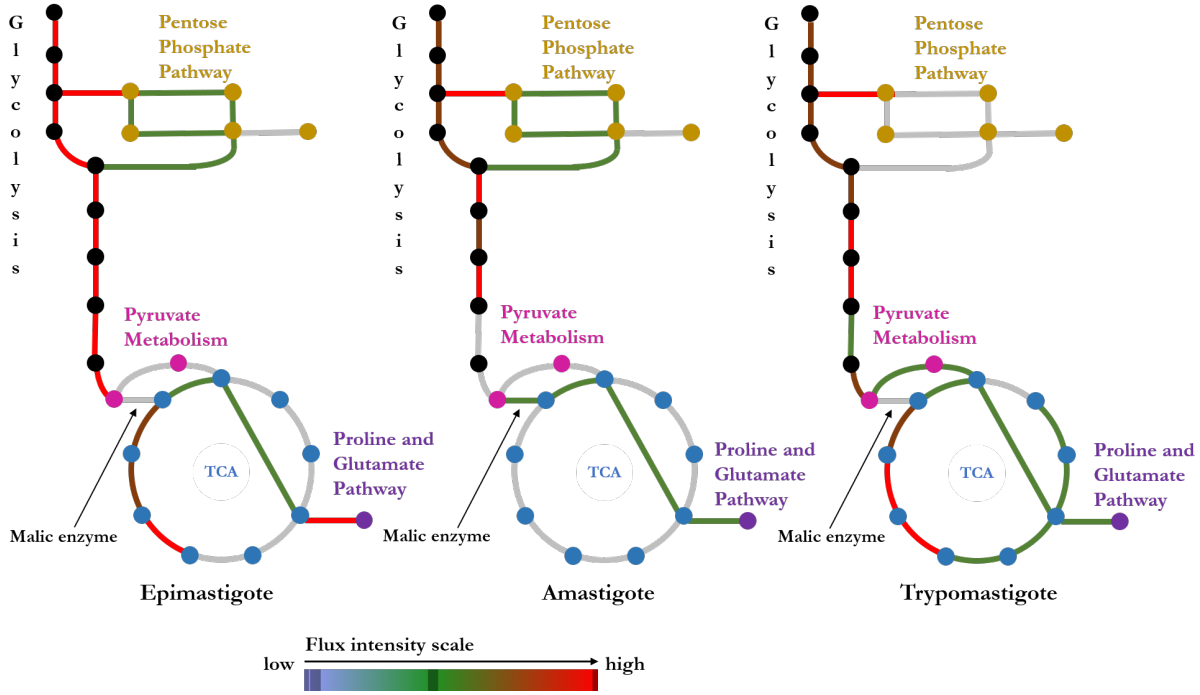

Figure 2.4: Flux distribution representation of each *T. cruzi* stage model in central metabolism (Glycolysis, Pentose Phosphate Pathway, TCA, and others). See Supplemental files shiratsubaki_supplemental_figure2_FBA_epimastigote.pdf, shiratsubaki_supplemental_figure3_FBA_amastigote.pdf, and shiratsubaki_supplemental_figure4_FBA_trypomastigote.pdf.

having almost all the TCA reactions being utilized. The metabolic flux through the malic enzyme is only present in the amastigote model to optimize the pyruvate flux outwards the TCA cycle and produce alanine. Trypomastigotes are the stage specialized in cell invasion and are not replicative. As a consequence, their PPP reactions are low-expressed and they require an active TCA for energy yield to move around. The epimastigote model presented an intermediate behavior for the flux distribution, showing a TCA less utilized than trypomastigote but more active than amastigote.

2.1.4 Stage-specific differences in essential genes and metabolic flux suggest potential drug targets

Through growth simulation of stage-specific models, we identified essential genes and reactions that could potentially be drug targets. Single reaction and gene deletion simulations were performed (See details in the methods section) for the pan and stage-specific models built from transcriptomics integration . We generated a list of potential essential genes and reactions for *T. cruzi* growth at different stages (See Supplemental file shiratsubaki_supplementary_tables.xlsx Tables S7, S8, and S9), which can be useful for drug discovery for Chagas disease. In this section, we show the stage-specific essential genes identified, and their corresponding essential reactions can be found in the Chapter 3 - Supplementary material (See Supplemental file shiratsubaki_supplementary_tables.xlsx Table S8 for more details).

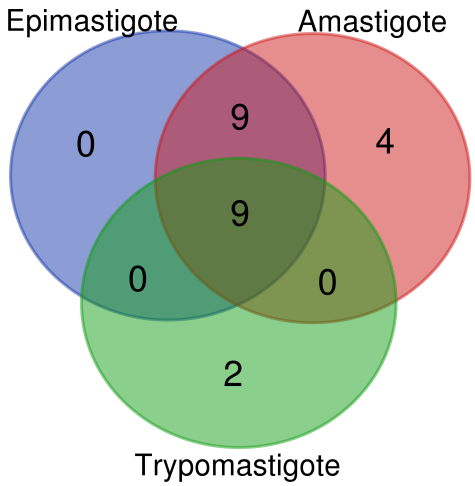

Figure 2.5: Venn diagram of essential single genes for each stage-specific iIS312 model (Supplemental file shiratsubaki_supplementary_tables.xlsx Table S8 for more details).

Genes that are essential for parasite survival differ for each stage, as the environment and objective change across stages. Stage-specific models predict that 10 genes were essential for all *T. cruzi* stages, among which 2 belong to the Glutamate Metabolism, 2 to Glycine, Serine, and Threonine Metabolism, and 6 to Glycolysis/Gluconeogenesis. We also found that 9 essential

genes were unique to epimastigote and amastigote (both replicative stages of *T. cruzi*). Mostly of these genes (7) belong to Pentose Phosphate Pathway, a pathway related to nucleic acid synthesis. Finally, 4 essential genes were unique to amastigote and only 2 for trypomastigote.

We also performed double gene deletions to get the list of lethal genes pairs that may potentially hinder cell survival if both deleted. This result provides additional drug targets that could be considered for Chagas disease treatment. (See Chapter 3 - Supplementary material and Supplemental file shiratsubaki_supplementary_tables.xlsx Table S9 for more details).

2.1.5 Potential drug targets and experiments

The single and double deletions of genes in the models provided us a list with potential essential genes and consequently potential drug targets. For single targets, all the stages showed more vulnerability in genes associated with Glutamate Metabolism, Glycine, Serine, and Threonine Metabolism, and Glycolysis/Gluconeogenesis. Some of the genes were also been found as lethal for other studies in *T. cruzi*, e.g. glutamate dehydrogenase (GLUDy) [12], glucokinase (GLUKg) and hexokinase (HEX) [38]. For hexokinase and other glycolytic enzymes, they can both be tricky targets for the amastigote stage. As it was discussed in previous section, findings in the literature showed that amastigotes can uptake glucose from host cytosol pools. However, it is still unknown if they present transporters for intermediate the glycolytic metabolites such as glucose-6-phosphate. Finally, the double deletion results can be also interesting for the simulation of double target experiments. We found 2603 cases of double gene deletions that might be lethal for all T. cruzi stages.

2.2 Discussion and conclusions

The development of drugs capable of targeting essential genes in multiple stages of *T. cruzi* might be the key for the improvement of Chagas disease treatment. In this work, we reconstructed the updated GEM of *T. cruzi*, iIS312, to incorporate the expansion of recent genome

annotation [19] of *T. cruzi*. Through integration of transcriptomics data, we generated three stage-specific models (epimastigote, amastigote, and trypomastigote) and validated them against experimental observations in previous studies. The reconstruction of the stage-specific GEM allowed us to investigate the variation in metabolic functions across stages, predict stage-specific essential genes and reactions, which can be experimentally validated through the disruption of gene or gene products (e.g. drugs for enzyme inhibition and gene knockouts).

Our model is significantly expanded and improved comparing to the existing *T. cruzi* model. The expanded iIS312 accounts for 516 reactions and 312 genes, as opposed to 162 and 215 in iSR215. Moreover, iIS312 includes more compartments and 83 new reactions in the lipid metabolism, which is absent in iSR215. Our GEM was validated by comparing its secretion predictions to findings in the literature, which fit the expected spectrum. In addition, we compared the essential reactions predicted by the pan model iIS312 with the experimental data used in iSR215 work [35] (See Supplementary material). Even though most of these experiments (more than 80%) were run in related organisms (*Trypanosoma brucei* and *Leishmania donovanni*, both pathogens belong to the family Trypanosomatidae), iIS312 predictions were able to reproduce 79% of the experimental data results (See Supplementary material). The use of experimental data in related organisms to validate *T. cruzi* GEMs is a common practice due to the lack of a specific CRISPR-Cas9 machinery for this pathogen.

Most importantly, the stage-specific models allowed us to predict and compare the differences in metabolic flux distribution and gene deactivation across stages. Interestingly, the deactivated reactions in the amastigote model suggested that glycolysis might not be essential for this stage. This result is consistent with a previous finding that glucose transporter activity was not detected in amastigotes [39]. A separate study [24] showed that the gene expression data in glycolytic genes drastically oscillated over time (0-72 hpi) - gene expression level of glycolysis reactions decreased when the infection of amastigotes in mammalian cells starts until reaching 6 hpi (0-6 hpi), and then increased between 6 to 72 hpi, possibly due to the inability of host to pro-

vide intermediate metabolites as it is being invaded by the parasite. These findings also reinforce the non-essentiality of glycolysis in amastigotes and might suggest the existence of transporters for intermediate glycolytic metabolites from the host cell, which explains the deactivation of some reactions in Glycolysis in our amastigote model. Confirmation of such transporters still needs further investigation in the future. In addition, the metabolic flux distribution predicted by our model revealed that amastigotes take a shortcut through proline and glutamate pathway, which enhances the flux metabolic activity in the malic enzyme. This behavior is also described in Marchese et al (2018) [26]. For trypomastigotes, our model predictions also resulted in expected behavior. The deactivated reactions were enriched in PPP pathway, which is consistent with findings in the literature as trypomastigotes are the stage specialized in cell invasion and are not replicative. Trypomastigotes also require an active TCA for energy yield to move around, and this behavior is observed in our prediction of metabolic flux distribution. For the epimastigote model, we found the least amount of gene deactivation, and the flux distribution revealed an intermediate TCA activity, as it utilizes Glycolysis and PPP, which align with findings in the literature [3].

Our findings suggest glycolytic enzymes may not be the best drug targets for Chagas disease due to its non-essentially in amastigote stage. In addition, our predictions in gene and reaction essentiality, as well as metabolic flux distribution among the different stages might be a step forward towards the improvement of Chagas disease treatment. From what we know so far, these stage-specific models are the first GEM built for the stages amastigote and trypomastigote. This work is also the first to present an *in silico* GEM comparison among different stages in *T. cruzi* life-cycle.

Future work comprises the refinement of iIS312 through adjustments in the biomass reaction and better genome annotation. Our eukaryote cell GEM presents 519 reactions, and the published 2009 model iSR215, 162 reactions, while most recent prokaryotic cell GEMs present more than one thousand reactions [28]. The difference in the amount of reactions highlights the limitation in *T. cruzi* genome annotation and how little we still know about this pathogen.

Additional future work also comprises the construction of strain-specific *T. cruzi* models for a more accurate prediction of drug targets for different strains. It is important to note that the current GEM only delineates the metabolic capability, but can be expanded to a metabolic and gene expression model [25] and include structural information of metabolic enzymes [29] in the future.

2.3 Materials and methods

2.3.1 iIS312 metabolic network reconstruction

Multiple steps are involved in building expanding the metabolic network reconstruction of *T. cruzi*. The first step is curation of metabolic knowledge(Figure 2.1, Step 1), which consists of data collection from genome annotation and gene-protein-reaction databases, metabolic reaction list generation, and determination of gene-protein-reaction relations. To curate our model, we used the genome annotation of the strain Dm28c [19], TritrypDB [2], Uniprot [10], and the models iSR215 (core metabolism model reconstruction of *T. cruzi*) and iAC560 (genome-scale model reconstruction of *L. major*). Once the curation was accomplished (See Supplemental file shiratsubaki_supplementary_tables.xlsx Table S1 for more details), the list of metabolic reactions was translated into a stoichiometric matrix (S), which is the mathematical representation of the stoichiometric coefficients of all substrates and products of all the reactions. Palsson and collaborators [31] describe in their work how GEM can be useful to predict biological capabilities. Model reconstruction is performed using the toolbox COBRApy [14].

The model is named then following the rules introduced by Reed and colleagues [32]: model name starts with i to denote *in silico*, followed by the first authors first and last initials (IS), and by the number of genes in the model (312).

2.3.2 Biomass reaction

The biomass reactions for *T. cruzi* were built based on the cellular composition of *L. major*, a related protozoan parasite. The function of the biomass equation is to provide a drain of essential metabolites that are needed to support the growth of the metabolic system [9, 41, 15]. Since the cellular composition is likely to vary according to the physiological conditions, the biomass equation is an approximation [41]. On the other hand, findings in the literature have shown that perturbations in the coefficients of metabolites in the biomass should not affect the overall biomass yield significantly [41]. Biomass reaction is modified for trypomastigote model as it is the only non-replicative stage. Check Chapter 3 - Supplementary Material for a detailed explanation on the biomass equation derivation and stage-specific differences.

2.3.3 Growth simulation by flux balance analysis (FBA)

After the curation step and the construction of the stoichiometric matrix, the flux balance analysis (FBA) is applied in the model to simulate *T. cruzi* metabolism. FBA is a mathematical optimization problem whose goal is identifying a specific flux distribution that optimizes a given metabolic objective (in our case, the biomass reaction of the model). Palsson et al [30] discuss in their work what FBA is and its application in biochemical networks. Since our model is simulating cell growth, the quasi-steady state (S.v = 0) can be considered, as the time constants for this case are long (hours to days), which is different from the time constants for metabolic transients ($<$ tens of seconds) [35, 22]. The medium used for growth simulation for iIS312 was developed based on the ones used in *L. major* model iAC560 [9] and *T. cruzi* model iSR215 [35]. To define *in silico* growth conditions, we changed the lower bound of exchange reactions. For our FBA analysis, we allowed the intake and outtake of the following metabolites: glucose, H^+, phosphate, acetate, cysteine, alanine, arginine, asparagine, choline, citrate, aspartate, oxygen, succinate, glutamate, glycine, glycerol, threonine, carbon dioxide, proline, ammonium, deoxyribose, ergosterol, glutathione, guanine, hydrogen sulfide, histidine,

hypoxanthine, isoleucine, methionine, nicotinamide, nicotinic acid, ammonia, phenylalanine, serine, thymine, uracil, and valine.

2.3.4 Validation of the model expansion

The validation step (Figure 2.1, Step 2) consisted of comparing findings in the literature and our model predictions. The findings we used to validate the model were the byproducts secreted by *T. cruzi*. We also compared our pan model predictions to experiments involving targeted disruption of genes or gene products (gene knockouts, RNAi, and drugs for enzyme inhibition) in *T. cruzi* or related organisms (See supplementary material for details).

2.3.5 Identification of essential genes

The essential gene-reactions list was generated by using the single gene deletion function in COBRApy [14], a python-based tool to build metabolic network reconstruction. The function was applied in the pan iIS312 model to identify essential genes in the pan model. Essential genes were also identified in stage-specific models. Specifically, for the amastigote model, we included the genes TCDM_04067 (ENO enolase, subsystem Glycolysis/Gluconeogenesis) and TCDM_05454 (PPCKg glycosomal phosphoenolpyruvate carboxykinase, subsystem Glycolysis/Gluconeogenesis) into the list of essential genes to maintain a non-nul growth.

2.3.6 Transcriptomics integration and life cycle stage-specific models

We integrated transcriptomics data (Figure 2.1, Step 3) to build stage-specific models from the pan iIS312 model followed the method presented by Richelle et al [33]. The gene expression data used came from the work of Berná et al [3], where the authors generated transcriptomics for all the stages in *T. cruzi* life cycle (epimastigotes, amastigotes, and trypomastigotes).

We then used model extraction method (MEM) [34, 33] to generate stage-specific models through integration of transcriptomics data. For our models, we used the MBA-like MEM

[33, 34, 21], an algorithm that uses a set of core reactions that should not be removed, while removing other reactions with low gene expression. We defined the set of essential gene-reactions obtained from the pan model as the set of core reactions that should not be removed.

For the preprocessing of gene expression data, we attributed a gene activity score for each gene and defined a threshold to determine which genes are active in each *T. cruzi* stage. The gene threshold is usually defined by the mean of each gene expression level over all the sample stages coming from the same dataset. In addition, the threshold should to be higher or equal to the 25^{th} percentile of the overall gene expression value distribution and lower or equal to the 75^{th} percentile [33]. The gene score is given by:

$$GeneScore = 5.log\left(1 + \frac{ExpressionLevel}{Threshold}\right) \tag{2.1}$$

The gene scores are integrated into the model by parsing the Gene-protein-reaction association (GPR) rules (See Supplemental file shiratsubaki_supplementary_tables.xlsx Table S1 for more details) associated with each reaction. Since the GPR is a logical expression (i.e. AND and OR logical operators), the gene score for each reaction is defined by taking the minimum expression value among all the genes associated to an enzyme complex (AND rule) and the maximum expression value among all the genes associated to an isozyme (OR rule) [33, 20].

As stated previously, our MEM considers a set of essential gene-reactions that should not be deactivated. The gene scored of each gene was calculated using the given formula above. The choice of the threshold in the gene score formula was defined first as the 25^{th} percentile of the overall gene expression value distribution, given the variability in the transcriptomics data of *T. cruzi* stages. The gene scores were then compared to another threshold equals to 5log2, as recommended in the literature [33]. For each gene, we evaluated if the gene score was lower than 5log2 and if it belonged to the essential gene-reactions list. If these two conditions were satisfied, the gene was knockout in the model.

For refinements and validation of the MEM applied, we used experimental findings by

Mattos et al [27] regarding protein activation in *T. cruzi* cellular infection. Trypomastigote is the stage responsible for cellular infection. From the list of proteins that were activated in the infectious stage, we checked if the associated genes were also active in our trypomastigote model. This allowed us to refine the threshold of our MEM as the 22^{th} percentile of the overall gene expression value distribution (See Supplemental file shiratsubaki_supplementary_tables.xlsx Table S2 for more details) to minimize the inconsistency between our method and experimental findings.

2.3.7 Flux distribution maps, metabolic network and gene-expression maps and venn diagrams

The flux distribution maps were generated by Escher-FBA [36]. This tool is a web application for interactive flux balance analysis (FBA) simulations within a pathway visualization. We uploaded our metabolic network maps and models to generate the output. In addition, all metabolic network and gene-expression maps were generated by Escher [23]. This tool is also a web application for building and generating data-rich visualizations for biological pathways. We also integrated transcriptomics into Escher map to facilitate the identification of trends in gene expression data. Finally, all the Venn diagrams were generated by a Venn diagram tool available at http://bioinformatics.psb.ugent.be/webtools/Venn/.

2.4 Acknowledgments

Chapter 2 (”Essential metabolic pathways in *Trypanosoma cruzi*”) and Chapter 3 (”Supplementary material”) are currently being prepared for publication by Shiratsubaki, Isabel S.; Fang, Xin; Siqueira-Neto, Jair L.; and Palsson, Bernhard . The thesis author was the primary investigator and one of the two primary authors of this paper.

Chapter 3

Supplementary material

3.1 Biomass calculations for *Trypanosoma cruzi*

The biomass reaction for *T. cruzi* was built using the similar cellular composition of *Leishmania major*, a close related organism belonging to the same family Trypanosomatidae [9]. The final biomass equation for *T. cruzi* iIS312 model is presented below:

(0.4281) ala-L + (0.2557) arg-L + (0.0932) asn-L + (0.1728) asp-L + (0.0672) cys-L + (0.1458) gln-L + (0.2134) glu-L + (0.2297) gly + (0.0960) his-L + (0.1060) ile-L + (0.3271) leu-L + (0.1184) lys-L + (0.0804) met-L + (0.1049) phe-L + (0.2059) pro-L + (0.3193) ser-L + (0.2134) thr-L + (0.2550) val-L + (0.1023) gmp + (0.0593) ump + (0.0565) ergst + (35.115) atp + (33.475) h2o + (14.405) nadph + (1.241) 3pg + (8.066) nh4 + (0.428) g3p + (0.397) e4p + (1.236) akg + (0.445) r5p + (1.785) oaa + (2.097) accoa + (0.642) pep + (0.712) g6p__B + (2.994) pyr + (3.015) nad $\rightarrow$ (35.115) adp + (33.475) h + (33.475) pi + (3.015) nadh + (2.852) co2 + (14.405) nadp + (2.097) coa

The changes made in the biomass reaction were supported by gap-filling for some metabolites. To check the gap-filling of the biomass metabolites, we used a package in COBRApy [14] called gapfilling. This tool compares two models with respect to a specific biomass metabolite and predicts which reactions are missing to enable growth. In our case, we compared our model iIS312 with iAC560, a larger reconstruction model. After getting the list of missing reactions for

each metabolite, we checked if the reactions are presented in *T. cruzi* by accessing TritrypDB [2] and Uniprot [10] databases. If we could not find strong evidences for the presence of those reactions, we set the metabolite coefficient in the biomass reaction as zero. We also added to this biomass formula the metabolites of the iSR215 biomass reaction. The overlapped metabolites that had their coefficient adjusted to the values in iSR215 biomass reaction are represented in blue. The metabolites in red represent the metabolites of iSR215 biomass reaction that were added into the iIS312 biomass formula.

3.2 Changes in the stage-specific trypomastigote model biomass reaction

As discussed in the main text, the results for deactivated genes and reactions and the fact that this stage is non-replicative suggest that nucleic acids are not important for trypomastigotes growth. For these reasons, we setted up the coefficients of Pentose Phosphate Pathway (PPP) intermediate metabolites and nucleotides as zero (damp, dcmp, dgmp, dtmp, amp, cmp, gmp, ump, g3p, e4p, and r5p). The metabolites in blue had their coefficients adjusted.

(0.4281) ala-L + (0.2557) arg-L + (0.0932) asn-L + (0.1728) asp-L + (0.0672) cys-L + (0.1458) gln-L + (0.2134) glu-L + (0.2297) gly + (0.0960) his-L + (0.1060) ile-L + (0.3271) leu-L + (0.1184) lys-L + (0.0804) met-L + (0.1049) phe-L + (0.2059) pro-L + (0.3193) ser-L + (0.2134) thr-L + (0.2550) val-L + (0.0565) ergst + (35.115) atp + (32.205) h2o + (13.135) nadph + (1.241) 3pg + (8.066) nh4 + (1.236) akg + (1.785) oaa + (2.097) accoa + (0.642) pep + (0.712) g6p__B + (2.994) pyr + (3.015) nad $\rightarrow$ (35.115) adp + (32.205) h + (33.475) pi + (3.015) nadh + (2.852) co2 + (13.135) nadp + (2.097) coa

3.3 Single reaction deletions

From the results (See Supplemental file shiratsubaki_supplementary_tables.xlsx Table S7), 42 reactions were essential for all the stage-specific models. Among them, 13 are exchange

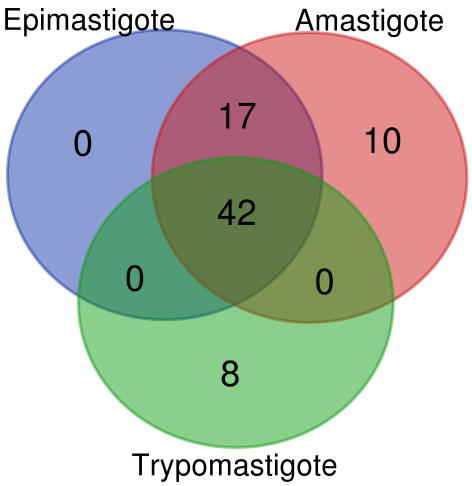

Figure 3.1: Venn diagram of essential reactions for each stage-specific iIS312 model (See Supplemental file shiratsubaki_supplementary_tables.xlsx Table S7 for more details).

reactions and are potentially related to essential nutrients for the parasite; 18 are transport reactions and no gene-associated; and 11 are metabolic gene-associated reactions and potential drug targets for Chagas disease. Among the 11 gene-associated essential reactions, they belong to Glutamate Metabolism (1 reaction), Glycine, Serine, and Threonine Metabolism (2 reactions) and Glycolysis/Gluconeogenesis (8 reactions). In addition, 17 reactions were unique for amastigote and epimastigote, both replicative stage of *T. cruzi*. Among these reactions, 12 are gene-associated and mostly belong to Pentose Phosphate Pathway (8 reactions). Two of them belong to Purine Metabolism. Both subsystems are associated with nucleic acid synthesis. Finally, 10 essential reactions were unique for amastigote and 8 for trypomastigote. Among the unique essential reactions for amastigote, 5 reactions are gene-associated metabolic reactions. For trypomastigote, only 2 reactions are gene-associated metabolic reactions.

3.4 Double gene deletions

We also did double gene deletions to get the list of double essential genes for each model. The results are described in Figure 3.2. From the results (See Supplemental file shiratsubaki_supplementary_tables.xlsx, Table S9 for more details). The results indicate that the

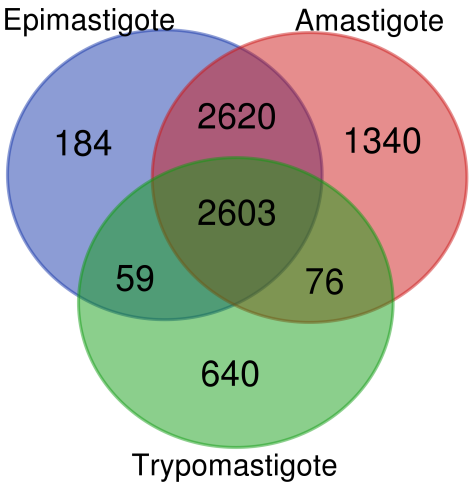

Figure 3.2: Venn diagram of essential reactions for each stage-specific iIS312 model (See Supplemental file shiratsubaki_supplementary_tables.xlsx, Table S7 for more details).

amastigotes are the most susceptible stage, presenting more lethal double deletions than the other stages.

3.5 Comparison with experiment findings

The comparison of iIS312 predictions with experimental data is shown in Supplemental file shiratsubaki_supplementary_tables.xlsx, Table S4. Our model simulations for the pan model were able to correctly reproduce respectively 79%, comparing to 79% for the full iSR215 model. Therefore, iIS312 kept prediction accuracy compared to iSR215.

3.6 Supplemental files

In the Supplemental Files for "Essential metabolic pathways in *Trypanosoma cruzi*", we present an excel file with all important tables generated from this work (Please see shiratsubaki_supplementary_tables.xlsx). The tabs in this file are: Table_S1: iIS312 content description, Table_S2: Transcriptomics data and integration, Table_S3: Excreted metabolites prediction, Table_S4: Comparison of model predictions and experimental data for lethal reactions in *T. cruzi* and related organisms, Table_S5: List of low-expressed essential genes among *T. cruzi* stages,

Tabl_S6: List of deactivated reactions after transcriptomics integration, Table_S7: List of essential reactions after single reaction deletion, Table_S8: List of essential genes after single gene deletion, Table_S9: List of essential pairs of genes after double gene deletion, and Table _S10: Comparison between transcriptomics results and list of kinases and phosphatases that were shown to be active during trypomastigote described by Mattos et al [27].

We also present in the Supplemental Files some figures representing metabolic maps. Those files are: shiratsubaki_supplemental_figure1_metabolic_network_map.pdf (Figure S1: *Trypanosoma cruzi* Metabolic Network), shiratsubaki_supplemental_figure2_FBA_epimastigote.pdf (Figure S2: FBA in Epimastigote), shiratsubaki_supplemental_figure3_FBA_amastigote.pdf (Figure S3: FBA in Amastigote), shiratsubaki_supplemental_figure4_FBA_trypomastigote.pdf (Figure S4: FBA in Trypomastigote), and shiratsubaki_supplemental_figure5_model_expansion.pdf (Figure S5: *Trypanosoma cruzi* Metabolic Network Expansion).

3.7 Acknowledgments

Chapter 2 (”Essential metabolic pathways in *Trypanosoma cruzi*”) and Chapter 3 (”Supplementary material”) are currently being prepared for publication by Shiratsubaki, Isabel S.; Fang, Xin; Siqueira-Neto, Jair L.; and Palsson, Bernhard . The thesis author was the primary investigator and one of the two primary authors of this paper.

Chapter 4

Conclusions and Future Work

The development of drugs capable of targeting essential genes in multiple stages of *T. cruzi* might be the key for the improvement of Chagas disease treatment. In this work, we reconstructed the updated GEM of *T. cruzi*, iIS312, to incorporate the expansion of recent genome annotation of *T. cruzi*. Through integration of transcriptomics data, we generated three stage-specific models (epimastigote, amastigote, and trypomastigote) and validated them against experimental observations in previous studies. The reconstruction of the stage-specific GEM allowed us to investigate the variation in metabolic functions across stages, predict stage-specific essential genes and reactions, which can be experimentally validated through the disruption of gene or gene products (e.g. drugs for enzyme inhibition and gene knockouts).

Our model is significantly expanded and improved comparing to the existing *T. cruzi* model. The expanded iIS312 accounts for 516 reactions and 312 genes, as opposed to 162 and 215 in iSR215. Moreover, iIS312 includes more compartments and 83 new reactions in the lipid metabolism, which is absent in iSR215. Our GEM was validated by comparing its secretion predictions to findings in the literature, which fit the expected spectrum. In addition, we compared the essential reactions predicted by the pan model iIS312 with the experimental data used in iSR215 work [35] (See Supplementary material). Even though most of these experiments (more than 80%) were run in related organisms (*Trypanosoma brucei* and *Leishmania donovanni*, both pathogens belong to the family Trypanosomatidae), iIS312 model predictions were able to reproduce 79% of the experimental data results (See Supplementary material). The use of

experimental data in related organisms to validate *T. cruzi* GEMs is a common practice due to the lack of a specific CRISPR-Cas9 machinery for this pathogen.

Most importantly, the stage-specific models allowed us to predict and compare the differences in metabolic flux distribution and gene deactivation across stages. Interestingly, the deactivated reactions in the amastigote model suggested that glycolysis might not be essential for this stage. This result is consistent with a previous finding that glucose transporter activity was not detected in amastigotes [39]. A separate study [24] showed that the gene expression data in glycolytic genes drastically oscillated over time (0-72 hpi) - gene expression level of glycolysis reactions decreased when the infection of amastigotes in mammalian cells starts until reaching 6 hpi (0-6 hpi), and then increased between 6 to 72 hpi, possibly due to the inability of host to provide intermediate metabolites as it is being invaded by the parasite. These findings also reinforce the non-essentiality of glycolysis in amastigotes and might suggest the existence of transporters for intermediate glycolytic metabolites from the host cell, which explains the deactivation of some reactions in Glycolysis in our amastigote model. Confirmation of such transporters still needs further investigation in the future. In addition, the metabolic flux distribution predicted by our model revealed that amastigotes take a shortcut through proline and glutamate pathway, which enhances the flux metabolic activity in the malic enzyme. This behavior is also described in Marchese et al (2018) [26]. For trypomastigotes, our model predictions also resulted in expected behavior. The deactivated reactions were enriched in PPP pathway, which is consistent with findings in the literature as trypomastigotes are the stage specialized in cell invasion and are not replicative. Trypomastigotes also require an active TCA for energy yield to move around, and this behavior is observed in our prediction of metabolic flux distribution. For the epimastigote model, we found the least amount of gene deactivation, and the flux distribution revealed an intermediate TCA activity, as it utilizes Glycolysis and PPP, which align with findings in the literature [3].

Our findings suggest glycolytic enzymes may not be the best drug targets for Chagas disease due to its non-essentially in amastigote stage. In addition, our predictions in gene and

reaction essentiality, as well as metabolic flux distribution among the different stages might be a step forward towards the improvement of Chagas disease treatment. From what we know so far, these stage-specific models are the first GEM built for the stages amastigote and trypomastigote. This work is also the first to present an *in silico* GEM comparison among different stages in *T. cruzi* life-cycle.

Future work comprises the refinement of iIS312 through adjustments in the biomass reaction and better genome annotation. Our eukaryote cell GEM presents 519 reactions, and the published 2009 model iSR215, 162 reactions, while most recent prokaryotic cell GEMs present more than one thousand reactions [28]. The difference in the amount of reactions highlights the limitation in *T. cruzi* genome annotation and how little we still know about this pathogen. Additional future work also comprises the construction of strain-specific *T. cruzi* models for a more accurate prediction of drug targets for different strains. It is important to note that the current GEM only delineates the metabolic capability, but can be expanded to a metabolic and gene expression model [25] and include structural information of metabolic enzymes [29] in the future.

Bibliography

[1] A. M. Abdel-Haleem, H. Hefzi, K. Mineta, X. Gao, T. Gojobori, B. O. Palsson, N. E. Lewis, and N. Jamshidi. Functional interrogation of Plasmodium genus metabolism identifies species-and stage-specific differences in nutrient essentiality and drug targeting. *PLoS computational biology*, 14(1):e1005895, 2018.

[2] M. Aslett, C. Aurrecoechea, M. Berriman, J. Brestelli, B. P. Brunk, M. Carrington, D. P. Depledge, S. Fischer, B. Gajria, X. Gao, M. J. Gardner, A. Gingle, G. Grant, O. S. Harb, M. Heiges, C. Hertz-Fowler, R. Houston, F. Innamorato, J. Iodice, J. C. Kissinger, E. Kraemer, W. Li, F. J. Logan, J. A. Miller, S. Mitra, P. J. Myler, V. Nayak, C. Pennington, I. Phan, D. F. Pinney, G. Ramasamy, M. B. Rogers, D. S. Roos, C. Ross, D. Sivam, D. F. Smith, G. Srinivasamoorthy, C. J. Stoeckert, Jr, S. Subramanian, R. Thibodeau, A. Tivey, C. Treatman, G. Velarde, and H. Wang. Tritrypdb: a functional genomic resource for the trypanosomatidae. *Nucleic acids research*, 38(suppl_1):D457–D462, 2009.

[3] L. Berná, M. L. Chiribao, G. Greif, M. Rodriguez, F. Alvarez-Valin, and C. Robello. Transcriptomic analysis reveals metabolic switches and surface remodeling as key processes for stage transition in Trypanosoma cruzi. *PeerJ*, 5:e3017, 2017.

[4] F. Bringaud, L. Rivière, and V. Coustou. Energy metabolism of trypanosomatids: adaptation to available carbon sources. *Molecular and biochemical parasitology*, 149(1):1–9, 2006.

[5] J. J. Cazzulo. Protein and amino acid catabolism in Trypanosoma cruzi. *Comparative Biochemistry and Physiology Part B: Comparative Biochemistry*, 79(3):309–320, 1984.

[6] J.-J. Cazzulo. Aerobic fermentation of glucose by trypanosomatids. *The FASEB journal*, 6(13):3153–3161, 1992.

[7] J. J. Cazzulo, B. M. F. de Cazzulo, J. C. Engel, and J. J. Cannata. End products and enzyme levels of aerobic glucose fermentation in trypanosomatids. *Molecular and biochemical parasitology*, 16(3):329–343, 1985.

[8] C. Chagas. Nova tripanozomiaze humana: estudos sobre a morfolojia e o ciclo evolutivo do schizotrypanum cruzi n. gen., n. sp., ajente etiolojico de nova entidade morbida do homem. *Memórias do Instituto Oswaldo Cruz*, 1(2):159–218, 1909.

[9] A. K. Chavali, J. D. Whittemore, J. A. Eddy, K. T. Williams, and J. A. Papin. Systems analysis of metabolism in the pathogenic trypanosomatid Leishmania major. *Molecular systems biology*, 4(1):177, 2008.

[10] U. Consortium. The universal protein resource (uniprot). *Nucleic acids research*, 35(suppl_1):D193–D197, 2006.

[11] G. Cross, R. A. Klein, and D. J. Linstead. Utilization of amino acids by trypanosoma brucei in culture: L-threonine as a precursor for acetate. *Parasitology*, 71(2):311–326, 1975.

[12] G. J. Crowther, D. Shanmugam, S. J. Carmona, M. A. Doyle, C. Hertz-Fowler, M. Berriman, S. Nwaka, S. A. Ralph, D. S. Roos, W. C. Van Voorhis, and F. Agüero. Identification of attractive drug targets in neglected-disease pathogens using an in silico approach. *PLoS neglected tropical diseases*, 4(8):e804, 2010.

[13] A. L. S. S. de Andrade, F. Zicker, R. M. de Oliveira, S. A. e Silva, A. Luquetti, L. R. Travassos, I. C. Almeida, S. S. de Andrade, J. G. de Andrade, and C. M. Martelli. Randomised trial of efficacy of benznidazole in treatment of early Trypanosoma cruzi infection. *The Lancet*, 348(9039):1407–1413, 1996.

[14] A. Ebrahim, J. A. Lerman, B. O. Palsson, and D. R. Hyduke. Cobrapy: constraints-based reconstruction and analysis for python. *BMC systems biology*, 7(1):74, 2013.

[15] J. Edwards and B. Palsson. The escherichia coli mg1655 in silico metabolic genotype: its definition, characteristics, and capabilities. *Proceedings of the National Academy of Sciences*, 97(10):5528–5533, 2000.

[16] J. C. Engel, B. M. F. de Cazzulo, A. O. Stoppani, J. J. Cannata, and J. J. Cazzulo. Aerobic glucose fermentation by Trypanosoma cruzi axenic culture amastigote-like forms during growth and differentiation to epimastigotes. *Molecular and biochemical parasitology*, 26(1-2):1–10, 1987.

[17] S. S. Estani, E. L. Segura, A. M. Ruiz, E. Velazquez, B. M. Porcel, and C. Yampotis. Efficacy of chemotherapy with benznidazole in children in the indeterminate phase of chagas' disease. *The American journal of tropical medicine and hygiene*, 59(4):526–529, 1998.

[18] B. FRYDMAN, C. de los SANTOS, J. J. CANNATA, and J. J. CAZZULO. Carbon-13 nuclear magnetic resonance analysis of [1-13c] glucose metabolism in trypanosoma cruzi: Evidence of the presence of two alanine pools and of two co2 fixation reactions. *European journal of biochemistry*, 192(2):363–368, 1990.

[19] E. C. Grisard, S. M. R. Teixeira, L. G. P. de Almeida, P. H. Stoco, A. L. Gerber, C. Talavera-López, O. C. Lima, B. Andersson, and A. T. R. de Vasconcelos. Trypanosoma cruzi clone dm28c draft genome sequence. *Genome Announc.*, 2(1):e01114–13, 2014.

[20] P. A. Jensen, K. A. Lutz, and J. A. Papin. Tiger: Toolbox for integrating genome-scale metabolic models, expression data, and transcriptional regulatory networks. *BMC systems biology*, 5(1):147, 2011.

[21] L. Jerby, T. Shlomi, and E. Ruppin. Computational reconstruction of tissue-specific metabolic models: application to human liver metabolism. *Molecular systems biology*, 6(1):401, 2010.

[22] A. R. Joyce and B. Ø. Palsson. Toward whole cell modeling and simulation: comprehensive functional genomics through the constraint-based approach. In *Systems Biological Approaches in Infectious Diseases*, pages 265–309. Springer, 2007.

[23] Z. A. King, A. Dräger, A. Ebrahim, N. Sonnenschein, N. E. Lewis, and B. O. Palsson. Escher: a web application for building, sharing, and embedding data-rich visualizations of biological pathways. *PLoS computational biology*, 11(8):e1004321, 2015.

[24] Y. Li, S. Shah-Simpson, K. Okrah, A. T. Belew, J. Choi, K. L. Caradonna, P. Padmanabhan, D. M. Ndegwa, M. R. Temanni, H. C. Bravo, N. M. El-Sayed, and B. A. Burleigh. Transcriptome remodeling in trypanosoma cruzi and human cells during intracellular infection. *PLoS pathogens*, 12(4):e1005511, 2016.

[25] C. J. Lloyd, A. Ebrahim, L. Yang, Z. A. King, E. Catoiu, E. J. OBrien, J. K. Liu, and B. O. Palsson. Cobrame: A computational framework for genome-scale models of metabolism and gene expression. *PLoS computational biology*, 14(7):e1006302, 2018.

[26] L. Marchese, J. Nascimento, F. Damasceno, F. Bringaud, P. Michels, and A. Silber. The uptake and metabolism of amino acids, and their unique role in the biology of pathogenic trypanosomatids. *Pathogens*, 7(2):36, 2018.

[27] E. C. Mattos, G. Canuto, N. C. Varón, R. D. Magalhães, T. W. Crozier, D. J. Lamont, M. F. Tavares, W. Colli, M. A. Ferguson, and M. J. M. Alves. Reprogramming of trypanosoma cruzi metabolism triggered by parasite interaction with the host cell extracellular matrix. *PLoS neglected tropical diseases*, 13(2):e0007103, 2019.

[28] D. McCloskey, B. Ø. Palsson, and A. M. Feist. Basic and applied uses of genome-scale metabolic network reconstructions of escherichia coli. *Molecular systems biology*, 9(1):661, 2013.

[29] J. M. Monk, C. J. Lloyd, E. Brunk, N. Mih, A. Sastry, Z. King, R. Takeuchi, W. Nomura, Z. Zhang, H. Mori, A. M. Feist, and B. O. Palsson. iML1515, a knowledgebase that computes escherichia coli traits. *Nature biotechnology*, 35(10):904, 2017.

[30] J. D. Orth, I. Thiele, and B. Ø. Palsson. What is flux balance analysis? *Nature biotechnology*, 28(3):245, 2010.

[31] E. J. OBrien, J. M. Monk, and B. O. Palsson. Using genome-scale models to predict biological capabilities. *Cell*, 161(5):971–987, 2015.

[32] J. L. Reed, T. D. Vo, C. H. Schilling, and B. O. Palsson. An expanded genome-scale model of escherichia coli k-12 (i jr904 gsm/gpr). *Genome biology*, 4(9):R54, 2003.

[33] A. Richelle, A. W. Chiang, C.-C. Kuo, and N. E. Lewis. Increasing consensus of context-specific metabolic models by integrating data-inferred cell functions. *bioRxiv*, 2018.

[34] S. Robaina Estvez and Z. Nikoloski. Generalized framework for context-specific metabolic model extraction methods. *Frontiers in Plant Science*, 5:491, 2014.

[35] S. B. Roberts, J. L. Robichaux, A. K. Chavali, P. A. Manque, V. Lee, A. M. Lara, J. A. Papin, and G. A. Buck. Proteomic and network analysis characterize stage-specific metabolism in Trypanosoma cruzi. *BMC systems biology*, 3(1):52, 2009.

[36] E. Rowe, B. O. Palsson, and Z. A. King. Escher-fba: a web application for interactive flux balance analysis. *BMC systems biology*, 12(1):84, 2018.

[37] M. Sanchez-Moreno, M. Fernandez-Becerra, J. Castilla-Calvente, and A. Osuna. Metabolic studies by 1h nmr of different forms of trypanosoma cruzi as obtained by in vitroculture. *FEMS microbiology letters*, 133(1-2):119–125, 1995.

[38] C. E. Sanz-Rodríguez, J. L. Concepción, S. Pekerar, E. Oldfield, and J. A. Urbina. Bisphosphonates as inhibitors of trypanosoma cruzi hexokinase kinetic and metabolic studies. *Journal of Biological Chemistry*, 282(17):12377–12387, 2007.

[39] A. M. Silber, R. R. Tonelli, C. G. Lopes, N. Cunha-e Silva, A. C. T. Torrecilhas, R. I. Schumacher, W. Colli, and M. J. M. Alves. Glucose uptake in the mammalian stages of trypanosoma cruzi. *Molecular and biochemical parasitology*, 168(1):102–108, 2009.

[40] A. G. Tielens and J. J. Van Hellemond. Differences in energy metabolism between trypanosomatidae. *Parasitology Today*, 14(7):265–272, 1998.

[41] A. Varma and B. O. Palsson. Metabolic capabilities of escherichia coli ii. optimal growth patterns. *Journal of Theoretical Biology*, 165(4):503–522, 1993.

[42] World Health Organization. Chagas disease (American trypanosomiasis), 2019.