Reconstruction and systems analysis of metabolism in apicomplexan parasites *Toxoplasma gondii* and *Plasmodium falciparum*

THÈSE N^{O} 6692 (2015)

PRÉSENTÉE LE 11 AOÛT 2015

À LA FACULTÉ DES SCIENCES DE BASE
LABORATOIRE DE BIOTECHNOLOGIE COMPUTATIONNELLE DES SYSTÈMES
PROGRAMME DOCTORAL EN BIOTECHNOLOGIE ET GÉNIE BIOLOGIQUE

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

POUR L'OBTENTION DU GRADE DE DOCTEUR ÈS SCIENCES

PAR

Stepan TYMOSHENKO

acceptée sur proposition du jury:

Prof. M. Dal Peraro, président du jury
Prof. V. Hatzimanikatis, Prof. D. Soldati-Favre, directeurs de thèse
Prof. R. Mahadevan, rapporteur
Dr J. Macrae, rapporteur
Prof. J. McKinney, rapporteur

ÉCOLE POLYTECHNIQUE
FÉDÉRALE DE LAUSANNE

Suisse
2015

To all those who believe I can make it
You are those who make it possible

Acknowledgments

1600 days, is it much?

is it enough to make a difference?

to make something that really matters?

I believe, the answer is Yes. Although it is important to say that I did not make this journey alone. There are some really amazing people who one way or the other helped me to plot this journey and in many ways made it better than I could have ever imagined.

First of all, I would like to acknowledge those, who conceived this journey, my *academic parents* Vassily and Dominique. I know I was not an easy *kid* to deal with, so I am very grateful for your patience, guidance, believing in me as well as giving me so much freedom to explore, learn and network.

There were also several people who helped me to establish and expand my core competences, which DRAMAtically contributed to success of my work. Keng was the first one to introduce me to the world of systems biology, computational metabolic modeling and analysis. Rasmus created a wonderful instrument that greatly facilitated my struggles with reconstruction of the models and kindly agreed to give me a crash-course on using it. Rebecca is the one without whose extensive contribution and always-ready-to-help attitude the experimental part of my thesis would be hardly possible. Meriç contributed with his expertise in implementation of several algorithms I used in my studies. Thank you very much, guys, your help was of outmost importance to make this all happen. I also would like to thank Julien Limenitakis for the knowledge of physiology and metabolism of *T. gondii* he generously shared with me and for a valuable scientific advice. Thank you, Misko, for being always very supportive and a voice of reason in difficult situations.

My nearest and dearest office-mates Katerina and Noushin, I will always remember and miss those lovely days of *working* side-by-side with you. Anirikh and Anne-Claire, thank you for your friendship, hospitality and a little sunshine Aaheli-Caroline.

Specially, I would like to mention two young and courageous modelers of apicomplexan metabolism – Anush and Aarti, who took the challenge of refining and developing the outcomes of my rather explorative work. I am very proud of you.

I would also like to mention all the LCSB members with whom I did not interact much concerning work matters, however, whose company and/or friendship made all these years so great: Marianne, Stefano, Joana, Georgios F., Alex, Georgios S., and the younger generation – Tiziano, Yves, Tuure, Milenko, Daniel, Jasmin, Christian and Pierre. In particular, I would like to recognize a great job of Christine, who is always making sure things in the lab work like a charm. That was my big pleasure to meet you all, please, do maintain that nice and friendly ambience of the lab.

The time I spent at the MIMOL department in Geneva was as well a great experience both scientifically and socially, thank you, guys: Karine, Christina, Paco, Ruben, Kyoko, Pierre, Damien, Nicolo, Jean-Ba, Yonggen, George, Aarti, Anne-Marie, Sunil, Natasha, Arnault, Vanessa. I also thank the members of MIMOL for the kind and constructive feedback on my data session presentation. A special, giant thank you is for Hayley and Rebecca, who really helped me on several occasions with corrections of my writing, I very much appreciate your availability and willing to help.

I would also like to acknowledge (rather indirect but important) contribution of Tom Forth, who I have never met in person, but whose PhD thesis, computational and experimental results he unselfishly shared with the community deserve a special tribute.

Most importantly, this thesis exists thanks to my family whose unconditional love and care made me who I am and gave me strength and motivation to overcome the moments of self-doubts.

I would like to thank all of you, no matter whether you name is mentioned here explicitly or not, your inspiration and (maybe even unspoken) encouragement is what helped me to reach this far.

Finally, I am very grateful to my thesis jury members, who came back with a positive answer to my call and gave me most positive evaluation and a constructive feedback.

Abstract

Understanding of metabolism in disease-causing microorganisms promotes drug design through the identification of the enzymes whose activity is indispensable for important cellular functions of the pathogens. Nowadays such understanding arises from experimental as well as computational studies. These two approaches, long considered as rather orthogonal, in recent years began to converge and form a new field, where they are utilized as complementary. In this thesis I present my endeavors in bringing closer the fields of infection and systems biology with a particular focus on large-scale metabolic models and their analysis. Integrative, interdisciplinary nature of my project also included multiple experimental inputs as well as original experimental efforts on investigating model-derived hypotheses. In the scope of this thesis I explored metabolism of two of the most experimentally amenable apicomplexan species – human parasites *Plasmodium falciparum* and *Toxoplasma gondii*. As a foundation for the studies included in this thesis I used standard as well as recently developed computational algorithms, existing experimental datasets and innovative context-specific assumptions. I produced an extensive survey of the modeling efforts previously applied for studying metabolism of *P. falciparum* and available large-scale experimental datasets in comparison with the similar efforts made in other species. Further, I curated an existing model of metabolism in *P. falciparum* with respect to an up-to-date primary literature on metabolism of the parasite and addressed several important assumptions implicitly made in this model. Using a state-of-the-art approach, I reconstructed *de novo* a comprehensive metabolic model of *T. gondii*, and performed an extensive computational analysis to explore its metabolic needs and capabilities. I identified and classified the minimal set of substrates the parasite utilizes for growth, along with the genes and pairs of genes that are essential for cellular functions such as growth and energy metabolism. Subsequently, several of the model-driven hypotheses were confirmed experimentally, while for validation of the majority of the computational predictions forthcoming high-throughput approaches shall be used. Every confirmed hypothesis expands the scope of our knowledge on peculiarities of metabolism in apicomplexan parasites and hence can serve as an input for the pipeline of developing novel medicines.

Keywords: metabolism, constraint-based modeling, model reconstruction, genome-scale metabolic model, thermodynamics-based flux balance analysis, gene essentiality, *Toxoplasma gondii*, *Plasmodium falciparum, Apicomplexa*, acetyl-CoA, riboflavin, FMN, FAD.

Résumé

La compréhension du métabolisme des microorganismes pathogènes est une approche fonctionnelle pour identifier de nouvelles voies médicamenteuses. En effet l'identification d'enzymes dont l'activité est indispensable à la survie du pathogène contribue à la découverte de nouvelles cibles thérapeutiques. De nos jours, la compréhension du métabolisme d'un organisme provient aussi bien d'études expérimentales que computationnelles. Ces deux approches, bien que longtemps considérées comme concurrentes, se sont développées au cours des dernières années pour finalement converger et donner naissance à une nouvelle thématique de recherche, appelée «biologie des systèmes», où les deux approches sont exploitées de manière complémentaire voire synergétique. Au cours de ma thèse, je me suis efforcé de mettre en relation les domaines de la biologie de l'infection et de la biologie des systèmes, en portant l'accent sur les modèles métaboliques à grande échelle et leur analyse. La nature convergente et interdisciplinaire de mon projet m'a amené a travailler sur de nombreuses données expérimentales ainsi que le développement de méthodes innovantes pour répondre spécifiquement aux hypothèses soulevées par les modèles. Durant cette cette thèse, j'ai exploré le métabolisme de *Plasmodium falciparum* et de *Toxoplasma gondii,* deux des parasites les plus étudiés du phylum des Apicomplexes, et qui sont respectivement les agents pathogènes de la malaria et la toxoplasmose. Pour les travaux décrits dans cette thèse, j'ai utilisé de nombreux algorithmes computationnels considérés comme classiques, ainsi que d'autres développés plus récemment, en plus de données expérimentales pré-existantes. J'ai également été amené à faire un certain nombre d'hypothèses innovantes spécifiques au contexte de chaque modèle. J'ai effectué une analyse approfondie des modèles métaboliques précédemment établis et des données expérimentales à large échelle disponible pour *P. falciparum*. La mise rapport avec des méthodes et données existantes pour d'autres espèces a été un axe supplémentaire de ce travail. De plus, j'ai corrigé et amélioré un modèle métabolique de *P. falciparum* déjà publié en y intégrant les informations trouvées dans la littérature la plus récente, et ai également réévalué plusieurs hypothèses importantes faites implicitement et *a priori* dans ce modèle.

En utilisant une approche de pointe, j'ai construit *de novo* un modèle métabolique complet de *T. gondii* et réalisé une analyse computationnelle approfondie afin de déterminer les capacités et besoins métaboliques de ce parasite. J'ai pu identifier et classifier les ensembles de substrats minimaux nécessaires au développement du parasite. J'ai également déterminé les gènes et couples de gènes essentiels aux fonctions cellulaires, telles que la croissance ou le métabolisme énergétique.

Par la suite, plusieurs prédictions issues du modèle ont été confirmées expérimentalement. Toutefois la majorité des conclusions computationnelles devront être validées en utilisant des approches à haut-débit encore en cours de développement. Toutes ces prédictions confirmées contribuent à une meilleure compréhension des particularités du métabolisme des Apicomplexes, et potentiellement peuvent servir comme point de départ au développement de nouveaux médicaments anti-parasitiques. **Mots clés : métabolisme, modélisation par contrainte, reconstruction de modèles, modèles métaboliques à échelle du génome, analyse par conservation des flux et thermodynamique, essentialité des gènes, *Toxoplasma gondii*, *Plasmodium falciparum, Apicomplexes*, acetyl-CoA, riboflavine, FMN, FAD.**

Table of Contents

List of figures

List of tables

List of abbreviations

ACL − ATP-citrate lyase;

ACS − acetyl-CoA synthase;

ADP − adenosine 5'-diphosphate;

AMP − adenosine 5'-monophosphate;

ATP − adenosine 5'-triphosphate;

BCKDH − branched-chain keto-acid dehydrogenase;

CBM − constraint-based model;

cDNA − complementary DNA;

CoA − coenzyme A;

CRE − cyclization recombinase;

CTP − cytidine 5'-triphosphate;

dCTP − 2'-deoxycytidine 5'-triphosphate;

DHFR − dihydrofolate reductase;

DHODH − dihydroorotate dehydrogenase;

DMEM − Dulbecco's modified Eagle medium;

DNA − deoxyribonucleic acid;

dTTP − 2'-deoxythymidine 5'-triphosphate;

e-value − expect value;

E.C. − enzyme classification (nomenclature);

ER − endoplasmic reticulum;

ETC − electron transport chain;

FACS − fluorescence-activated cell sorting;

FAD − flavin-adenine dinucleotide;

FAS − fatty acid synthesis (pathway);

FBA − flux-balance analysis;

FCS − fetal calf serum;

FMN − flavin mononucleotide;

FMNAT − flavin mononucleotide adenylyltranserase;

FVA − flux variability analysis;

GABA − gamma-amino butyrate;

GBM − graph-based model;

GCM – group contribution method;

GDP − guanosine diphsphate;

GFP − green fluorescent protein;

GPR − gene-protein-reaction association;

GTP − guanosine 5'-triphsphate;

HFF − human foreskin fibroblasts;

HMM − hidden Markov model;

HXGPRT − hypoxanthine-guanosine phosphoribosyl transferase;

IFA − indirect immunofluorescence assay;

IMM − *in silico* minimal medium;

KEGG − Kyoto encyclopedia of genes and genomes;

LLAMP − Liverpool library of Apicomplexan metabolic pathways;

MECDPS – 2-C-methyl-D-erythritol 2,4-cyclodiphosphate synthase;

MEP/DOXP − 2-C-methyl-D-erythritol-4-phosphate/1-deoxy-D-xylulose-5-phosphate;

MPA – mycophenolic acid;

MPMP − malaria parasite metabolic pathways;

mRNA − messenger ribonucleic acid;

MS − mass-spectrometry;

NAD − nicotinamide-adenine dinucleotide (oxidized);

NADH − nicotinamide-adenine dinucleotide (reduced);

NADP − nicotinamide-adenine dinucleotide phosphate (oxidized);

NADPH − nicotinamide-adenine dinucleotide phosphate (reduced);

NMR − nuclear magnetic resonance;

ORF − open reading frame;

PBS − phosphate buffer saline;

PCR − polymerase chain reaction;

RK − riboflavin kinase;

RNA − ribonucleic acid;

RNAi − RNA interference;

TCA − tricarboxylic acid (Krebs) cycle;

TFBA/TMFA − thermodynamics-based flux balance analysis;

THF − tetrahydrofolate;

tRNA − transfer RNA;

TTP − thymidine 5'-triphosphate;

TVA − thermodynamic variability analysis;

WB − western blot (immunoblot).

Peer-reviewed publications included in the thesis

Functional genomics of *Plasmodium falciparum* using metabolic modelling and analysis by Tymoshenko S, Oppenheim RD, Soldati-Favre D, Hatzimanikatis V., Brief Funct Genomics. 2013 Jul;12(4):316-27, doi: 10.1093/bfgp/elt017

Metabolic needs and capabilities of *Toxoplasma gondii* through combined computational and experimental analysis by Tymoshenko S, Oppenheim RD, Agren R, Nielsen J, Soldati-Favre D, Hatzimanikatis V, PloS Computational biology, doi:10.1371/journal.pcbi.1004261

1. Introduction

The name Apicomplexa denotes the presence of apical organellar complex, a characteristic feature of apicomplexan parasites, which serves for invasion into a host cell. Apicomplexa is a phylum of eukaryotic pathogens, presumably all the members of the phylum are obligate intracellular parasites [1], some of which are of great medical and veterinary significance. *Plasmodium* and *Toxoplasma* species are causative agents of human malaria and toxoplasmosis respectively; *Theileria*, *Babesia*, *Neospora*, *Eimeria spp.* are veterinary pathogens infecting livestock, poultry, cattle, domesticated and wild animals. While the phylum encompasses several thousands of species only a few of them are actively studied and amenable for *in vitro* cultivation. *Plasmodium falciparum* and *Toxoplasma gondii*, the parasites my studies were focused at, are among the best studied apicomplexan pathogens. Both of the parasites have a complex life cycle, which involve intermediary and definitive host species and human is an intermediary host for both of the parasites (Figure 1, Figure 2).

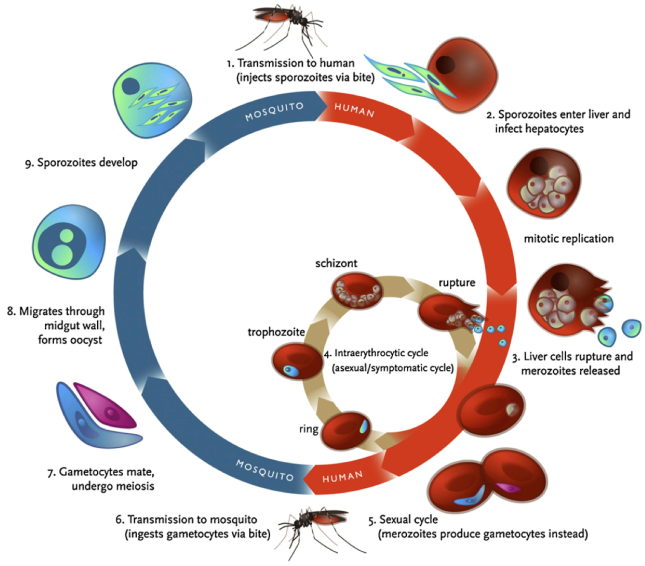


Figure 1. Schematic description of the lifecycle of malaria parasite *P. falciparum* (adapted from [2])

Definitive hosts, however, are different, namely, *Anopheles* mosquitos and feline species for *Plasmodium spp.* and *T. gondii* respectively. Sexual replication of the parasites occurs only in particular tissues of the definitive host species, while only asexual replication occurs in intermediary host organisms. Infection of intermediary hosts also occurs differently in case of malaria and toxoplasmosis. *T. gondii* and other coccidian pathogens are predominantly

ingested with infected tissues (e.g. meat contaminated with tissue cysts) or fecal contaminations of food by feline species (e.g. cat feces with infective oocysts). *Plasmodium* spp. as well as the other haemosporidian parasites are transmitted by arthropods (e.g. mosquitos or ticks).

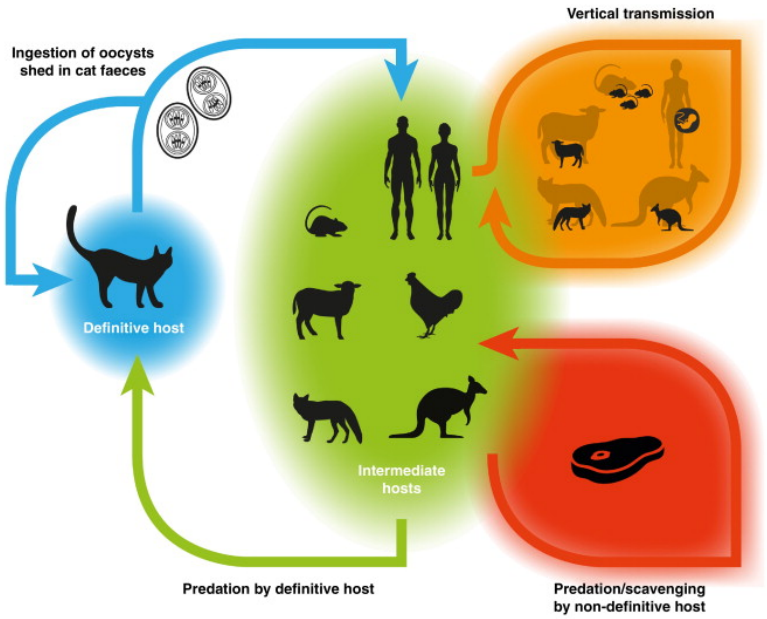


Figure 2. Schematic description of the lifecycle of *T. gondii* and different routes of transmission between the definitive and intermediate hosts (adapted from [3])

Toxoplasma is extremely versatile with respect to the intermediary hosts and can use almost any nucleated mammalian cell as a niche for an asexual replication. Notwithstanding, *Plasmodium* spp. are very specialized to the tissue type not only in the definitive but also in the intermediary host – five *Plasmodium* species can cause malaria in humans with the primary (asymptomatic) infection in liver and later release of merozoites to the bloodstream and infection of the red blood cells. It is possible that such specialization for a particular niche allowed further reduction of versatility in metabolic capabilities of *Plasmodium* spp. as compared to *T. gondii*. For example, *T. gondii* possesses fatty acid synthesis type I (eukaryotic), gluconeogenesis and methyl-citrate cycle enzymes that are apparently absent (potentially lost) in *Plasmodium* species. Despite these important differences there is also a large number of similarities in the metabolic capabilities of *P. falciparum* and *T. gondii*. One of the most prominent common features is the presence of the apicoplast – a chloroplast-like non-photosynthetic organelle, which harbors several metabolic pathways of bacterial origin. This organelle appears to be a source of isoprenoids and fatty acids in both of the parasites.

Importantly, experimental investigation of the role of the importance of apicoplast revealed that it is indispensable source of isoprenoids for both *P. falciparum* and *T. gondii*. Counter-intuitively, apicoplast-localized fatty acid synthesis pathway has been demonstrated as essential for asexual replication of *T. gondii* (despite the presence of the fatty acid synthesis type I pathway in the cytosol) but not for the blood stage *P. falciparum*. This rather unexpected difference is explained by a sufficient salvage of fatty acids by the latter parasite on the blood stage (but not on the preceding liver stage) and it highlights non-intuitive and context-dependent nature of gene essentiality. One of the possibilities to address this uncertainty in a rigorous manner is to reconstruct a computational metabolic model for a specie of interest and analyze redundancies in its metabolic capabilities through *in silico* simulations. Another option is to undertake a genome-wide experimental study on essentiality of the genes using available molecular biology techniques for gene knockout or knockdown. As such experimental techniques become available for *P. falciparum* and *T. gondii* only in the present days, this thesis is focused primarily on the computational approaches to predictably delineate essentiality of genes in these parasites.

1.1 What we do and do not know about metabolism in *Apicomplexa*?

The intracellular environment in which the parasites reside offers a relatively safe niche for evasion of the host immune response. However, this advantage comes with the challenge of acquiring sufficient metabolic resources for intracellular replication. Naturally, this requires an uptake of a significant amount of nutrients, as well as secretion of metabolic by-products within the environment of a host cell. In this competition for the limited resources of the host cell, the parasites appear to have evolved to keep the fine balance between meeting their metabolic needs and sustaining their host cells from a premature lysis. The obligate nature of parasitism is a hallmark of the physiology in *Apicomplexa*; to date virtually no attempts have succeeded in sustainable cultivation of these parasites axenically. This may be interpreted as evidence of some as yet undiscovered essential host-pathogen interactions beyond simple "supplier-consumer" relations.

While it is well known that *Apicomplexa* completely rely on the host cells and surrounding tissues for supply of the substrates that fuel their metabolism, the extent and flexibility of this dependence remain rather elusive. This is largely due to the complex nature of metabolism in general, as well as the technical challenges in distinguishing the metabolites of infected host

cells from those of the intracellular parasites. Studies included in this thesis represent a combination of using systems biology and experimental approaches for a better understanding of the intricate details about metabolism in the apicomplexan parasites *Toxoplasma gondii* and *Plasmodium falciparum*.

1.2 Do we need computational modeling of apicomplexan metabolism?

The ever-lowering cost of sequencing complete genomes creates an avalanche of genomic data. Extraction of knowledge from this data, however, proves to be more challenging than the data acquisition. This is, on the one hand, due to the challenge of functional annotation (i.e. assigning functions to the genes in newly sequenced genomes), which to date is solved only partially. On the other hand, an even bigger challenge is posed by the vast complexity and scale of the information encoded in genomes. Robust and exhaustive analyses of such an extent of information are often beyond the capabilities of human perception. Consequently, the "revolution of genomes" is followed by a necessary development of increasingly sophisticated computational methods for functional and comparative genomics. These methods facilitate analyses of large datasets and extraction of non-intuitive hypotheses for experimental validation. In particular, inhibition of metabolic enzymes is long known to be an efficient way to suppress growth of pathogenic microorganisms. In 2002, upon the very first sequencing of *P. falciparum* genome, it became evident, that this malaria parasite possesses a rather unusual metabolic network, consisting of pathways inherent to both prokaryotes (e.g. FAS type II, MEP/DOXP) and eukaryotes (Figure 1). The pathways of bacterial origin are likely inherited from the prokaryotic endosymbionts, which have evolved to become organelles of the eukaryotic pathogens (e.g. apicoplast, non-photosynthetic plastid-like organelle that is characteristic for Apicomplexa).

The enzymes and pathways that are not present in human metabolism may represent potential targets for novel antimalarials, but inhibition of the growth will only be observed if the targeted enzyme is *essential* for the parasite. The notion of *essentiality* in this context describes the absence of an alternative metabolic route that bypasses the targeted enzyme by using other enzymes and/or by salvage of additional substrates. Large-scale computational and experimental studies, however, indicate that metabolism is, in general, very redundant, and essential metabolic enzymes constitute only a modest fraction [4]. While metabolism of obligate pathogens is often significantly reduced when compared to free-living cells [5], this

reduction may be backed up by the capabilities of salvaging necessary metabolites from the host [6,7]. Due to the complexity and redundancy of metabolism the essentiality of an enzyme, as a rule, is neither apparent nor can be easily inferred from a genome sequence of an organism. For genetically tractable organisms, such as *Escherichia coli* and *Saccharomyces cerevisiae*, there already exist large genome-wide libraries of phenotypes observed upon deletion of every single gene, which indicate that only up to 18% of all the genes in these species are essential for their growth on a nutrient-rich medium [8-11]. Such genome-wide studies have not yet been done in any of the apicomplexan parasites. Nevertheless, genetic manipulations in several species are routinely possible in a low-throughput manner, which allows conditional and permanent disruption of individual genes [12]. This represents one of the important aspects where computational metabolic modeling can facilitate experimental efforts by predicting sets of genes and enzymes which are likely to be essential, as described below.

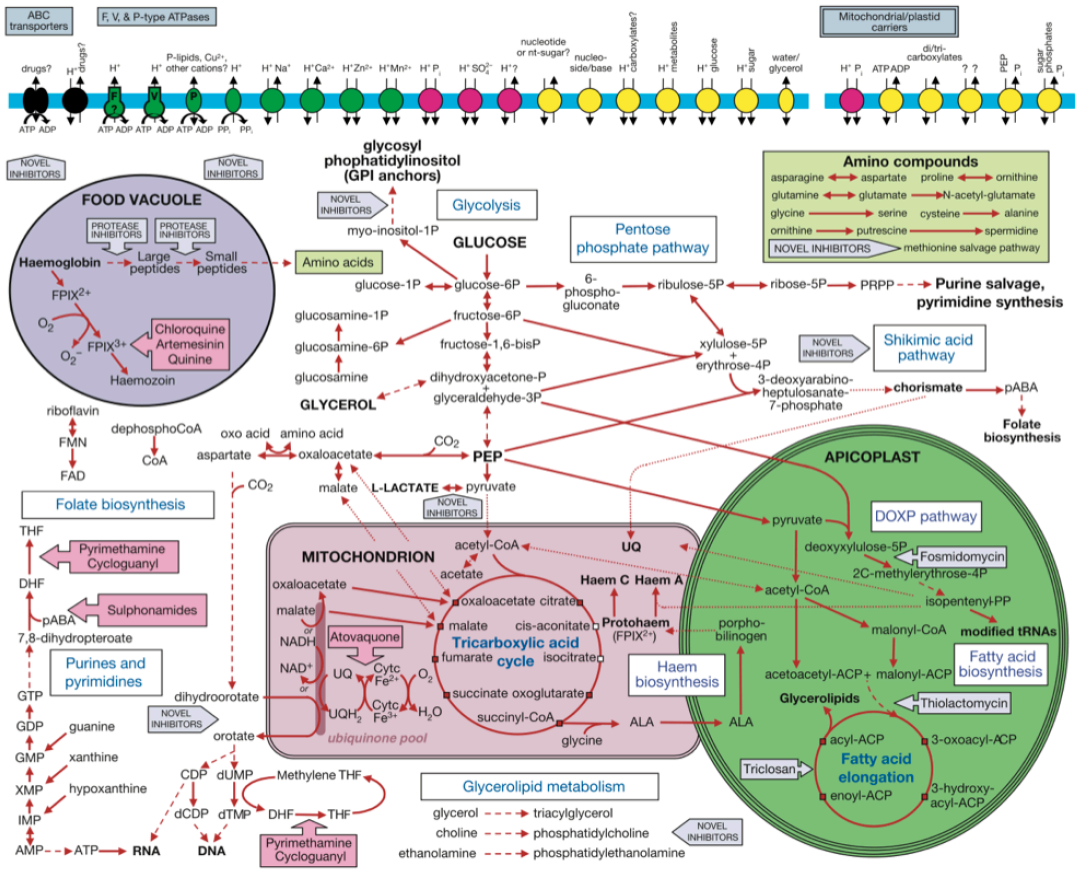


Figure 3. Graphical representation of key metabolic pathways in *P. falciparum* and antimalarial drugs targeting metabolic enzymes (adapted from [13]).

Apart from prediction of essential genes, the holistic approach of metabolic modeling also gives global insights into metabolism of a species of interest. These include but are not limited to: (1) identification of gaps or inconsistencies in the knowledge of particular aspects of metabolism; (2) improvement of functional annotations of metabolic enzymes; (3) elucidation of their putative subcellular localizations. Part of the input information for building large-scale models is produced in an automated and high-throughput manner (e.g. genome annotation), while the other part comes through results of more targeted research on some particular gene(s) or enzyme(s). Large-scale metabolic models are built to encompass and reconcile both types of the information in a concise and consistent manner to facilitate its further analyses.

1.3 What are genome-scale metabolic models?

Computational modeling of metabolic pathways is a rapidly developing field with numerous practical applications. One of the prominent directions in the field is reconstruction and analysis of large-scale constraint-based models (CBMs). This approach combines the advantages of modest prerequisites for reconstruction of a CBM and a relative simplicity of its computational analysis with a potential to integrate various available datasets as additional constraints. When a CBM is reconstructed with the aim to encompass a complete set of the metabolic capabilities encoded in a certain genome, such a model is termed a *genome-scale* CBM. In principle, genome-scale CBMs may contain only genes and biochemical reactions that correspond to the enzymes encoded by those genes. Additionally, there is also a plethora of methods for integration of various types of high-throughput data from the different levels [14], such as transcription for the enzyme-coding genes and abundances of enzymes and metabolites (Figure 2).

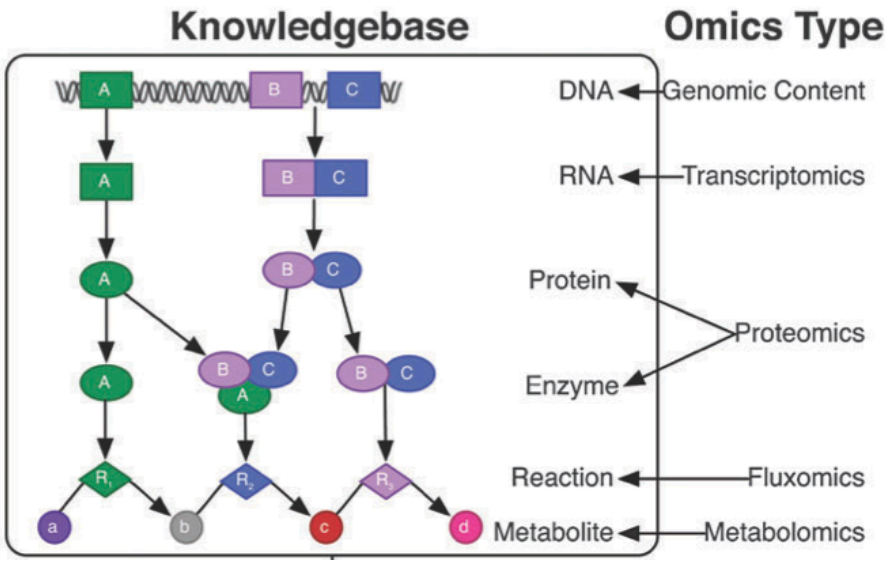


Figure 4. Genome-scale metabolic models include information from multiple levels (adapted from [15])

Genome-scale CBMs are, to some extent, a mathematical expression of the central dogma of molecular biology applied to metabolism of small molecules. These models encompass the knowledge about metabolism-related genetic content in a species of interest; the scope of this knowledge is based on the functional annotation of the genes. The genes are linked to biochemical knowledge represented by the reaction equations that describe the action of corresponding metabolic enzymes. The relations between genes and reactions in CBMs are denoted as gene-protein-reaction associations (GPRs), which capture the knowledge about the genes (and, implicitly, the proteins) necessary for each reaction to occur. GPRs can readily express simple one-to-one relations between gene-protein-reactions, as well as more complex cases, such as: (1) presence of several enzymes that can catalyze the same reaction (i.e. isoenzymes); (2) need for several different proteins to combine into a multi-enzyme complex for a certain reaction to occur. Built in such way, the models are repositories of the knowledge on the putative metabolic capabilities encoded in a genome of interest.

One of the most fundamental functions of metabolism is transformation of nutrients taken up by a cell into a specific set of molecules and energy, which are necessary for the replication of all the cellular components and, ultimately, cell division (Figure 3).

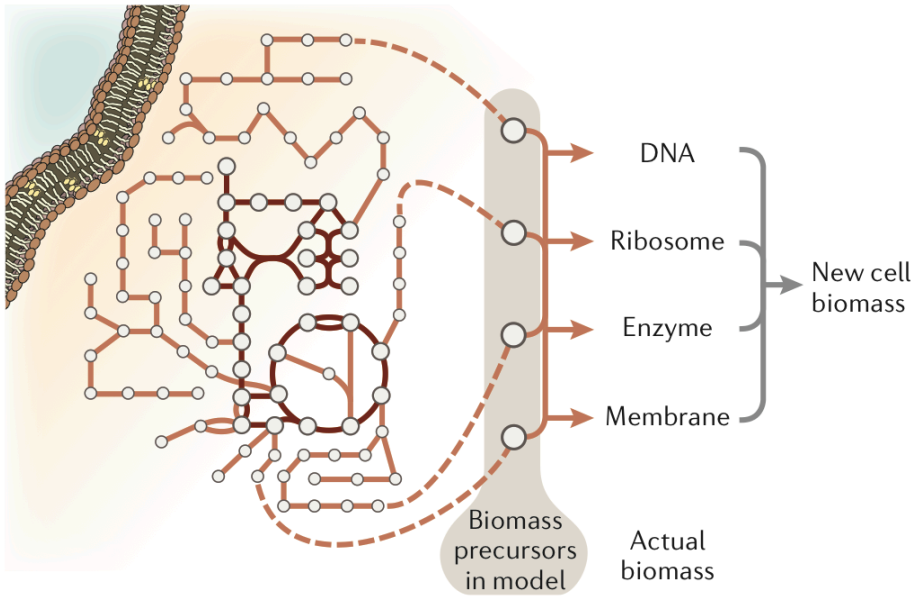


Figure 5. Schematic illustration of metabolism as machinery that transforms substrates into biomass building blocks, biological macromolecules and enables growth (adapted from [14]). The grey circles represent metabolites and the lines – corresponding metabolic reactions.

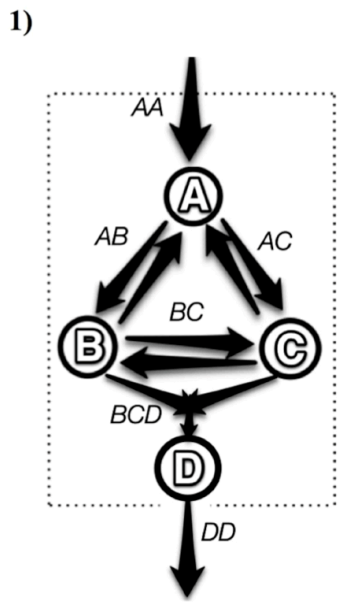
In the case of CBMs, this set of the molecules-precursors of major cellular components is (approximately) represented as a synthetic *biomass reaction*, which describes the metabolites necessary for the production of 1 gram of the dry-weight biomass. Such representation of growth by flux through *biomass reaction* allows simulation of gene deletions as explained below, and assessment of the dispensability of each enzyme and gene for growth in any defined conditions.

1.4 What is flux balance analysis?

In this chapter, I illustrate, using a simple example, the essence of the flux-balance analysis – a computational method used for studies on CBMs throughout this thesis. A more in-depth description of the method is published elsewhere [16] and provides further explanations and examples not covered herein.

For the sake of simplicity and clarity, let us consider a simplified metabolic model as shown in on Figure 4 with two reactions (AA and DD) that cross the boundaries of our fictional cell and four reactions (AB, BC, AC and BCD) that constitute an intracellular metabolic network. The reactions AA and DD are respectively an inlet and an outlet for the metabolites A and D

– such reactions are called *exchange reactions*. Herein, we consider that A is a unique single substrate that our cell takes up using reaction AA, and D is a product that the cell produces and excretes using reaction DD. Graphical representation of our simple metabolic model can be transformed into a two-dimensional matrix where every row corresponds to a metabolite and every column to a reaction (Figure 4, panel 2). As the reaction AB consumes one molecule of A it has a "-1" on the intersection of the row A and column AB, and positive "1" on the intersection of B and AB as one molecule of B is produced. Intersections on AB with C and D contain zeros because these metabolites do not participate in the reaction AB. In a similar way every reaction of the model is described and together they all form the so-called *stoichiometric matrix*. Assuming that all the intracellular reactions are reversible, there are at least three ways (*solutions*) that metabolite D can be produced from A (Figure 4, panel 3). These three solutions differ in utilization of the reaction BC; it is not used in the solution V_1 (zero at the vector denotes no flux through the reaction), while V_2 and V_3 use this reaction in its positive and negative direction respectively. Such multiplicity of routes that one product can be produced from a given substrate denotes flexibility of metabolic networks and creates the need for exhaustive accounting for all the possibilities in order to define whether a certain reaction (and corresponding gene/enzyme) is *essential*.



2)

	AA	AB	AC	BC	BCD	DD
A	1	-1	-1	0	0	0
B	0	1	0	-1	-1	0
C	0	0	1	1	-1	0
D	0	0	0	0	1	-1

3)

V_1	2	1	1	0	1	1
V_2	2	2	0	1	1	1
V_3	2	0	2	-1	1	1

Figure 6. The concept of flux balance analysis applied to a simple system.

Graphical description of a simplest metabolic model (panel 1), corresponding (stoichiometric) matrix representation for FBA (panel 2) and alternative solution vectors V_1, V_2, V_3 (panel 3).

Note that the product of the stoichiometric matrix S multiplied by each of the solution vectors (V_1, V_2 and V_3) gives a vector of zeros. This is because all the three solutions satisfy *quasi-steady-state assumption*, which implies no net change in the concentrations of intracellular metabolites. This assumption is necessary to simulate metabolism using large-scale CBMs in the absence of the information about kinetic properties for the majority of the enzymes. Biologically, *quasi-steady-state* denotes a homeostatic state of cellular metabolism at the timescale that exceed time at which cyclic fluctuations of metabolite concentrations occur.

Given the model described above, I now illustrate the concept of gene-essentiality simulations using flux-balance analysis. Suppose that production and secretion of metabolite D is a vital function of our imaginary cell. In the simplest case shown in panel 1 of Figure 5, only two reactions (AB and AC) require corresponding enzymes encoded by the genes *ab* and *ac*. As a result of the reversibility of the reaction BC, neither of these two genes is essential (see the alternative solutions V_2 and V_3 above). The same is not true in the case of reaction BC being irreversible, as shown in panel 2 of Figure 5. Now, when metabolite B can only be produced using reaction AB corresponding gene *ab* is essential, while metabolite C can still be produced by two reactions, thus, the gene *ac* is not essential. To illustrate gene essentiality in the case of more complex gene-protein-reaction associations we will make the assumptions schematically, as shown in panel 3 of Figure 5. There, the reaction AB requires a multi-enzyme complex that consists of two subunits encoded by two different genes, *ab1* AND *ab2*. The reaction AC is also associated to two different genes, but through the logical operator OR that denotes presence of isoenzymes encoded by two different genes *ac1* and *ac2*. Presence of either enzyme encoded by *ac1* or *ac2* is sufficient for the reaction AC to occur. In this case, both *ab1* and *ab2* are essential because deletion of either of them blocks reaction AB and makes production of B impossible. Conversely, neither *ac1* nor *ac2* is essential for the reason described above. Simulation of a pairwise withdrawal of the genes (i.e. *double gene deletion*) will identify the pair of *ab* and *ac* as double (or synthetically) essential in the case shown in panel 1 of Figure 5, while in the case shown in panel 2, this pair will be a *trivial* double essential, since *ab* is essential in its own right.

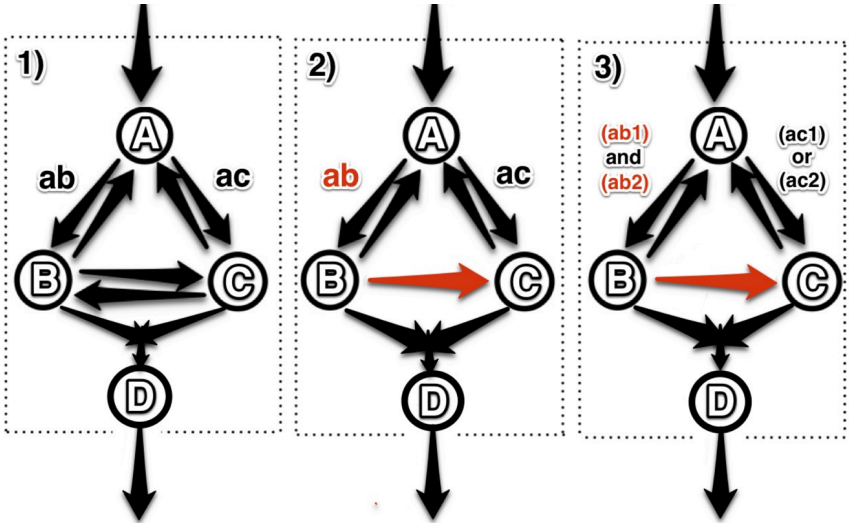


Figure 7. Illustration of the concepts of gene essentiality prediction using FBA

In a simple case, when gene-protein-reaction associations (GPR) are one-to-one, essentiality of reactions is equivalent to essentiality of the enzymes that uniquely catalyze these reactions as well as the genes that uniquely encode these enzyme. However, in the case of a GPR describing one-to-many (enzyme with broad substrate specificity) or many-to-one (heteropolymeric multienzyme complex) relations, essentiality of a reaction corresponding enzyme(s) and gene(s) may not coincide. Thus, GPRs, using Boolean operators "AND" and "OR", explicitly describe relations between genes and reactions in a model to enable systematic gene-knockout simulations. Technically, such simulation is performed by enforcing zero flux through reaction(s) associated with a knocked-out gene, provided that in the corresponding GPR there is no alternative gene assigned with the Boolean rule "OR". Double gene deletions are simulated in the same manner for all *non-trivial pairs* of genes: i.e. those that contain only the genes which are not essential on their own. Similarly to the single gene knockout simulations, each pair of genes' corresponding GPRs are evaluated and if there are reactions that exclusively depend on either of these genes (or on both of them) – fluxes through such reactions are blocked.

The simplified example we considered in this chapter only illustrates the concept of flux-balance analysis and gene deletion simulations. However, it does not demonstrate complexity of the analyses of a real genome-scale CBM where hundreds of genes, reactions and metabolites are involved as well as additional complexity arises from compartmentalization and other physiological constraints. In the following chapter, I review the metabolic

modeling efforts made in the malaria parasite *P. falciparum*, which are largely based on the concepts described above.

1.5 The landscape of functional genomics of studies in *P. falciparum* using metabolic modeling and analysis (based on the review manuscript by Tymoshenko et al., Brief Funct Genomics. 2013 Jul;12(4):316-27)

According to the recent "World Malaria Report" by the World Health Organization, malaria remains a major healthcare issue, being responsible for over 200 million infections and hundreds of thousands of deaths in 2010 alone [17]. An efficient and cost-effective artemisinin-based treatment is now available, but the emergence of resistance in malaria parasites, primarily as a result of drug treatments, urges the development of medicines with new targets and novel mechanisms of action. Despite tremendous research efforts and the passing of a decade since the publication of the *P. falciparum* genome sequence [13], about half of the genes still remain annotated as coding for 'hypothetical proteins' or 'conserved hypothetical proteins' [12]. These genes, and especially those restricted to *Plasmodium* or apicomplexan species, are of particular interest as their unique amino acid sequences may provide higher selectivity for new antiparasitic drugs. However, in the absence of tools for high-throughput gene knockdown, e.g. RNAi [18], the identification and validation of the essentiality to these genes for the parasite remains a major bottleneck [12]. Available experimental approaches for establishing gene essentiality in *P. falciparum* are cumbersome due to several unique properties of the pathogen. Firstly, it possesses an extremely AT-rich genome along with an unusually low frequency of homologous recombination, which makes it refractory to genetic manipulation such as gene replacement [12]. Secondly, *in vitro* cultivation is a delicate process [19] that is still mainly restricted to the intraerythrocytic stages. Thirdly, the limitation of using primates as an animal model makes the *in vivo* assessment of gene essentiality very limited and expensive. Nevertheless, emerging experimental breakthroughs show promise for cost-effective gene knockdown strategies of *P. falciparum* essential genes at a high-throughput scale [20,21]. In this context, there is a need to list genes of immediate interest, which should be validated as antimalarial targets once affordable experimental means are available.

Metabolic modeling is a modern approach of systems biology that, among several other applications, has been extensively exerted to predict gene essentiality in various bacteria, including a number of pathogenic species [22]. Computational (i.e. *in silico*) metabolic models offer a cost-effective pipeline to identify putatively indispensable metabolic functions that, in the case of pathogens, represent potential targets for medical intervention [23]. Eight years after the publication of the first computational model of a tentative metabolic network of *P. falciparum*, it has become evident that reconstruction and analysis of *in silico* models is a valuable tool for studying various aspects of the physiology of the pathogen [24-28]. In this study, we aim to review and provide an outlook on the current state and contribution of *in silico* metabolic modeling efforts to functional genomics studies on the malaria parasite.

1.5.1 The path from metabolic maps to "context for content" models

With the constantly decreasing cost of high-throughput measurements, the tendency to generate very large -omics datasets is emerging and this holds true for *P. falciparum* [29]. There are two distinct approaches to high-throughput measurements: *hypothesis-driven* studies generate large datasets in order to prove or falsify a hitherto existing hypothesis, while *hypothesis-free* studies primarily rely on a thorough analysis of datasets without presumptions aiming at the formulation of conclusions and testable hypotheses. Importantly, hypothesis-free approaches require an appropriate context, i.e. a framework of related prior knowledge, within which an obtained dataset can be interpreted. In the case of functional genomic studies, *in silico* metabolic models have been shown to provide such *context*, thus enabling researchers to use available datasets (i.e. *content*) to improve and challenge the models, as well as to derive additional, non-intuitive insights. In particular, this need for an explicit, up-to-date context is a driving force for the progress from a scope of generic biochemical knowledge to organism-specific (or life stage specific) metabolic networks, discussed further below.

In the last two decades, several databases of biochemical reactions and metabolic pathways have been developed in order to provide a systematic and comprehensive overview of metabolism. A pertinent example is the Kyoto Encyclopedia of Genes and Genomes (KEGG) Pathway Database [30], which has established itself as an encyclopedia of biochemical knowledge and is the first point of reference for numerous academic and industrial research. However, KEGG and similar large databases are more universal than organism-specific, thus they do not cover some crucial aspects of the *P. falciparum* metabolism, such as a cofactor

utilization, compartmentalization, and proteins involved in the transport of metabolites. The web-resource Malaria Parasite Metabolic Pathways (MPMP) is essentially the product of an extensive manual revision of KEGG maps according to known, as well as recently established, metabolic features of the *P. falciparum* intraerythrocytic life stage. Furthermore, integration with other web-resources facilitates quick access to additional information and related primary literature [31]. MPMP is probably the most comprehensive and up-to-date knowledge database on *P. falciparum's* metabolism, but unfortunately, the lack of any application-programming interface (API) limits the access of a broad research community to this high-quality data. Utility of MPMP for research purposes has been demonstrated via interpretation of the expression patterns of the genes involved in the pentose-phosphate pathway using the corresponding maps [32]. This led the authors to postulate that the oxidative part of the pathway is predominantly active during the early stage of the intraerythrocytic replication cycle, while the non-oxidative part is activated only during the later stages [32]. In contrast to this approach, further attempts to comprehend metabolic fluxes in the malaria parasite aimed at going beyond the study of a particular pathway and ultimately led to the reconstruction of large-scale models with hundreds of metabolites and interconnected reactions involved [24-28].

Graph-based models (GBMs) were built to comprehend and analyze the metabolic capabilities of *P. falciparum* in a systematic manner. GBMs represent metabolism as graphs with nodes (to denote metabolites) and links between them (respective metabolic reactions) [33]. The links only represent the possibility of interconversion for a particular metabolite into another one (justified by the correspondingly annotated genes), regardless of all the properties of the reaction, including its stoichiometry. A natural advantage of GBMs is that they require minimal input information and can assess well-annotated parts of a given metabolic network while skipping unclear ones. However, such representation of a metabolic network without capturing its mass-balance property is not a suitable framework for incorporation of experimental information, thus making the scope of GBM-based methods in metabolic modeling rather limited. Nowadays, graph-based metabolic reconstructions and topological analyses are largely obsolete and have been replaced by more comprehensive constraint-based approaches.

Over the past decade, **flux-balance analysis** (FBA) has been established as a leading approach for studying **constraint-based models** (CBMs) of cellular metabolism [34]. Constraint-based metabolic models for FBA are based on two-dimensional arrays, where

each row represents a metabolite and each column corresponds to a reaction which, as a rule, is linked to a certain enzyme and gene in the organism [35]. Each intersection of row and column contains a numerical value standing for the stoichiometric coefficient of a given metabolite in a given reaction. Such notation enables the explicit and quantitative description of a metabolic network and allows the imposing of the first basic constraint – *mass balance*. A model that only accounts for mass balance is largely undefined and further constraints are imposed to enrich the range of feasible solutions with biologically realistic ones. The scope of additional constraints is constantly growing and currently includes thermodynamic (TFBA/TMFA), regulatory (rFBA), and other constraints inherent in cellular metabolism (reviewed in [14]), as well as constraints inferred from experimental data.

CBMs, unlike GBMs, do require a pre-defined *objective* towards which utilization of available substrates should be optimized [36]. A common objective function for fast-growing cells is the *biomass reaction* (other plausible objectives are reviewed in [37]). It represents cellular replication as a reaction that consumes pre-defined amounts of metabolites referred to as *precursors of biomass*. Existence of a solution implies that the stoichiometric array describes at least one uninterrupted route that leads to the transformation of externally supplied substrates into the metabolites specified as biomass precursors. To satisfy this requirement, the model-building process often involves the inclusion of "orphan reactions", for which no enzyme-coding gene has been yet annotated [38].

A conventional reconstruction workflow shown in [38] clearly defines the list of enzymes that should be found in the genome to complete the pathways and meet experimentally observed metabolic behavior. With this approach, putative metabolic functions have been proposed for 17 genes of *Leishmania major*, which previously had no functional annotations [39]. Although these assignments, often based on moderate sequence identity, are not sufficient proofs of the suggested functions, they do represent a set of testable hypotheses for experimental validation. Such a list of orphan reactions provides invaluable guidance for functional genomics studies in *P. falciparum,* an organism that has over 2000 genes without even a putative functional annotation.

One of the most important applications for a CBM of pathogenic species is its capability to predict potential vulnerabilities in their metabolism [23]. Often metabolic networks are redundant and contain more than one chain of reactions to produce certain biomass precursors. To explore this redundancy, FBA enables an attempt at the simulation of growth

when each of the reactions in the model is removed in a one-at-a-time manner. Whenever production of biomass is blocked without a certain reaction, the latter is classified as *essential*. In a similar manner, FBA can simulate outcomes of withdrawal of enzymes or genes by attempting to simulate growth when utilization of all the reactions associated with an enzyme or gene are disabled. We refer the reader to a comprehensive publication on *in silico* essentiality studies in the CBM of *S. cerevisiae* [40] for further details and examples.

1.5.2 Outlook of *in silico* metabolic models for *P. falciparum*

Several studies to date have focused on the comprehensive reconstruction of the metabolic network of *P. falciparum* (Table 1). Early studies produced several GBMs [25,28,41] that considered the set of metabolic activities reported at the time without taking into account information about the compartments in which they occur or the differences in life stage specific metabolism of the pathogen. As mentioned before, without consistency in mass balance relationships, these models were not suitable for integration of experimental data and allowed only qualitative predictions of gene essentiality. Nevertheless, during the reconstruction process putative functions were suggested for hundreds of genes previously annotated as coding for a "hypothetical protein" [28]. These first modeling efforts gave a broad overview of the potential metabolic capabilities of *P. falciparum* annotated in its genome compared to the expected ones, and provided a solid foundation for building the modern, more comprehensive reconstructions discussed hereafter.

Using a constraint-based approach, Huthmacher and co-workers assembled the first compartmentalized, mass-balanced and life stage-specific model of *P. falciparum* metabolism [26]. Through *in silico* simulations, the authors identified enzymatic activities that are essential for proliferation of the parasite. Thirty of the *in silico* essential reactions were catalyzed by enzymes with no homologs in the human proteome (e-value > 0.075). These were ranked as targets of particular interest, based on evidence of activity during the multiple stages of the *P. falciparum* life cycle studied and/or the presence in the SuperTarget database [42] as candidates for treatment of other infections [26]. Furthermore, the natural environment of the parasite was simulated by embedding its CBM into the metabolic reconstruction of a human erythrocyte. This limited the substrate accessibility to only those available in the host cell milieu. Such constraints at the cellular interface and constraints on reaction fluxes, deduced from gene expression profiles (obtained in different life stages), not only allowed the model to retrieve known directions of metabolite exchanges between the

host cell and the parasite, but also identified several inconsistencies with experimental data, which need further investigation [26].

Authors (year of publication)	Information about the model*			
	Metabolites	Reactions	Genes	Compartments**
Yeh et al. (2004) [28]	525	696	–	–
Fatumo et al. (2009) [25]	554	575	–	–
Huthmacher et al. (2010) [26]	P: 1622 E: 566	P: 1375 E: 437	P: 579	P: c, m, a, n, r, v, g E: e, c
Plata et al. (2010) [27]	915	1001	366	P: e, c, m, a
Bazzani et al. (2012) [24]	P: 1622 H: 1149	P: 1395 H: 2539	P: 579 H: 704	P: c, m, a, n, r, v, g H: c, r, g, l, m, n, p, b, s

** - "P" denotes the model of the parasite, "E" - human erythrocyte, "H" - human hepatocyte*

*** - abbreviated names of compartments: e - extracellular space, c - cytosol, m - mitochondrion, a - apicoplast, n - nucleus, v - digestive vacuole, r - endoplasmic reticulum, g - Golgi complex, l - lysosome, p - peroxisome, b - bile canaliculus, s - sinusoidal space.*

Table 1. Comparison of the *in silico* metabolic reconstructions of *P. falciparum*.

The second CBM developed independently by Plata et al [27] also took account of compartmentalization of the intracellular space as well as mass-balance constraints. The results of gene deletions performed *in silico* were found to be in correspondence with reports in the primary literature: 100% agreement when compared to gene deletion studies and 70% in the case of enzymatic inhibition experiments. Comparison of the number of essential genes in the models of *P. falciparum* and *S. cerevisiae* confirmed the notion that the parasite likely possesses significantly lower metabolic flexibility to bypass single gene deletions as compared to free living organisms with similar genome sizes [27]. Forty genes were suggested as potential drug targets due to their *in silico* essentiality and extremely low or absent sequence identity to human proteins [27]. The essentiality of one of these genes encoding the nicotinate mononucleotide adenylyltransferase (NMNAT) was verified by using an experimental inhibitor, which caused an arrest of *P. falciparum* proliferation at an IC_{50} of 50 μM in *in vitro* culture [27]. In addition, Plata et al [27] were the first to report *in silico* double gene deletion simulations in *P. falciparum*, leading to the identification of 16 pairs of genes that were predicted as non-essential by single-gene knockout simulation but resulted in a dramatic impairment of the metabolism if targeted simultaneously.

Holzhutter and colleagues have developed the most recent CBM of the *P. falciparum* metabolic network [24] by updating their previous PlasmoNet1 model [26]. In PlasmoNet2 [24], they have added new transport reactions based on metabolomics data [43] and removed

one reaction according to an updated version of the KEGG database. Through integration of PlasmoNet2 with the CBM of the human hepatocyte [44], they evaluated *in silico* the essentiality of *P. falciparum* genes in the liver stage and assessed, using a "reduced fitness" approach, effects of targeting enzymes that are homologous and predicted to be essential both in the host and the pathogen [24].

Comparison of the essentiality predictions made by the afore-mentioned models is not a trivial task. Firstly, because the study by Plata et al aimed at predicting the essentiality of genes, while the other studies assessed the essentiality of enzymes, the predictions made do not overlap when gene-to-enzyme relations are not one-to-one. Secondly, a gene or enzyme may be absent from the list of "predicted as essential" not only because it has been predicted as non-essential but, possibly, also due to the fact that the corresponding metabolic process simply is not included in the model of interest. Thirdly, essentiality predictions in CBMs are directly dependent on the set of metabolites included into their biomass reaction, so that differences in assumed biomass composition directly affect the results of *in silico* simulations. For an overview of *P. falciparum* genes and enzymes predicted as essential in existing models see the table in Appendix 1. The table also provides the reader with literature references on the experimental assessment of genes/enzymes predicted to be essential by the different models.

An important simplification common to all of the aforementioned models is the *ad hoc* assignment of directionality to the reactions, which are pre-set either as all reversible [26] or assumed to have the same reversibility as in the metabolic models of non-related organisms [27]. In principle, the directionality of a reaction is subject to its thermodynamic properties and spurious *ad hoc* assignments might violate this fundamental constraint. This issue has been addressed rigorously in genome-scale metabolic networks of several organisms [45-47] by implementation of thermodynamic constraints on reaction directionality as an extension to conventional FBA methodology.

Overall, metabolic modeling of *P. falciparum* to date lags several years behind the similar efforts for the model eukaryote *S. cerevisiae*; there exist a few independently reconstructed models which often lack consistency with each other due to the differences in the reconstruction workflows, the sources of primary information, the level of complexity, and the varying degree of comprehensiveness. Similarly to the trends in modeling of the yeast metabolism, we expect the emergence of reconciliation efforts that will aim at obtaining a consensus, up-to-date CBM of *P. falciparum.* Recently developed workflows for manual [48]

and semi-automated [49] reconciliations of existing metabolic models may significantly facilitate these efforts. There are several other areas in which we foresee room for upcoming improvements: the first, as mentioned before, is a systematic implementation of thermodynamic constraints in the models of *P. falciparum;* the second is a deliberate revision of the objective function (i.e. biomass reaction) to make it consistent with the actual biomass composition of the parasite at the life stage of interest; the third is an experimental verification and quantification of the uptake fluxes present in the models. These improvements are likely to make *in silico* predictions of gene essentiality more reliable and also expand the number of metabolic functions currently known to be essential for *P. falciparum*.

Metabolic reconstruction efforts to date have summarized the results of decades of experimental research on the metabolism of *P. falciparum* in the form of *in silico* models, which not only reproduce the prior knowledge but also provide novel insights. Nevertheless, the fact that cultivation of the parasite in a fully-defined medium is still impossible clearly highlights that some important metabolic peculiarities remain to be discovered. The utilization of the CBMs as frameworks of current knowledge, which can then be challenged, refined and constrained by various high-throughput datasets is discussed in the following section.

1.5.3 High-throughput data and metabolic models: "content for context"

As mentioned above, CBMs hold a great potential to incorporate various types of experimental data as content for the context of computational metabolic networks. This section provides an outlook of the currently published studies that have integrated computational modeling and experimental research efforts for each type of high-throughput data introduced in Figure 6.

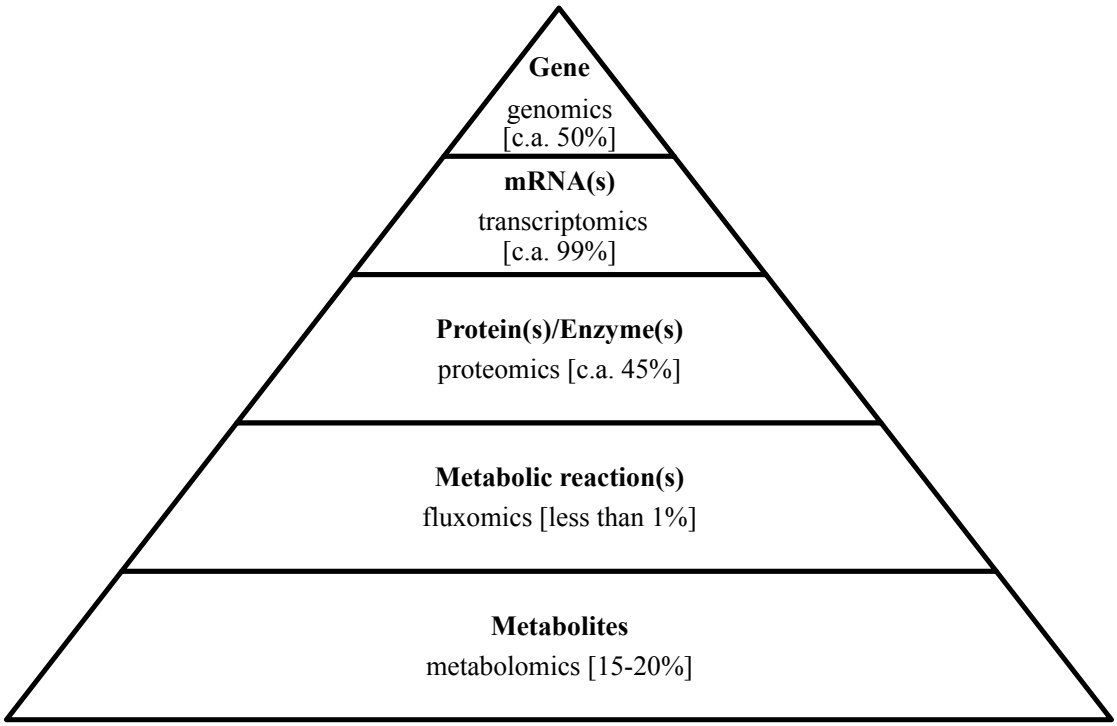


Figure 8. An overview of the high-throughput methods applied for functional genomics of *P. falciparum*

The maximal achieved to date coverage of the expected amount of information on each level is given in square brackets. The pyramid denotes the increasing likelihood of one-to-many relations on the route from a gene to its transcript(s), encoded enzyme(s), catalyzed reaction(s) and participating (transformed) metabolites.

Genomics

Genomics studies of *P. falciparum* have so far yielded several complete genome sequences (for 3D7 and IT strains [50]). However, their functional annotation remains far from complete [12]. Nonetheless, these partially annotated genomes gave rise to numerous genome-wide transcriptomics and large-scale proteomics studies, as well as making possible the reconstruction of the parasite's metabolic network *in silico*.

High-throughput functional genomics of *P. falciparum* is a nascent field in malaria research since only a limited number of studies have succeeded in generating single-gene mutant parasite clones at a large scale. The largest coverage reported to date was achieved with a forward genetics approach based on transposon mutagenesis using the transposable element *piggyBac* [51]. The collection regroups about 200 mutant parasite lines that cover only a modest part of the *P. falciparum* genome. Among those, 24 single-gene disruptions caused severe growth defects *in vitro*, which might lead to lethal phenotypes *in vivo* [51]. A major drawback of this method is its inability to reveal essential genes due to the haploid state of

the malaria parasite throughout most of its life stages. Although specific design of the transposon with an integrated inducible promoter has been proposed to assess essential genes [20], it still remains to be proved as being practicable. In contrast, *in vitro* essentiality data of the genetically tractable parasite *Trypanosoma brucei* has been obtained for nearly all the coding sequences of its genome using the RNA-interference target sequencing (RIT-seq) approach [52].

The value of the gene essentiality datasets is reinforced by the possibility of *in silico* simulation of similar genetic perturbations. This offers reciprocal benefits for experimental and computational research: CBMs can often suggest the underlying reason for viable/non-viable phenotypes of knockout strains or, otherwise, reveal the weak points where the model should be improved. Accordingly, the algorithm GrowMatch [53] has been applied for comparison of the high-throughput datasets of viable/non-viable single and double gene knockout mutants in *S. cerevisiae* with *in silico* gene essentiality predictions [40]. This approach has led to over a hundred corrections in the model (e.g. inclusion of additional reactions, compounds and genes as well as changes in biomass reaction, etc), each supported by literature evidence, and largely improved the consistency of the model with the existing experimental data [40]. Once a successful high-throughput forward or reverse genetic technique is established for *P. falciparum,* GrowMatch or a similar algorithm could validate the computational model, improving understanding of the obtained results and producing hypotheses for experimental investigation applicable to malaria drug research.

Transcriptomics

Transcriptomic profiling appears to be the one of most common high-throughput methods applied in malaria research; numerous gene expression datasets are available in PlasmoDB database (http://www.plasmodb.org) for different lineages of the parasite under various conditions. While the scope of gene expression data is growing, new increasingly advanced algorithms to integrate these datasets into CBMs are also being developed [54]. To date, two models of *P. falciparum* metabolism have incorporated available transcriptomics data as constraints [26,27] to represent life stage specific metabolic features of the parasite.

Huthmacher et al integrated their model with several life stage specific gene expression profiles [55-59] to avoid, whenever possible, utilization of reactions that are likely to be inactive based on the abundance of the corresponding mRNAs [26]. This approach has allowed the authors to infer plausible metabolite exchanges between the human erythrocyte

and the parasite, as well as to predict directionality of the pathways that cannot be easily inferred from gene-expression data alone. The predictions of the host-parasite metabolite exchanges were in significantly better agreement with the physiology reported in literature when fluxes were allowed through any reaction for which the mRNA was absent at the life stage specific transcriptome but had been present earlier (up to 12 preceding hours) [26]. This observation is consistent with an earlier study that found a significant time delay between maximal gene expression and peak of accumulation of the corresponding proteins [60].

Plata et al [27] also attempted to integrate gene expression data [43,61] into their model to predict the shifts in the extracellular (intraerythrocytic) abundance of some metabolites between ring to trophozoite and trophozoite to schizont stages. The authors assumed that a higher influx of a metabolite into the parasite would lower the abundance of the substrate molecule in the cytosol of the infected erythrocyte, whereas a higher efflux rate of a metabolite from the parasite will increase its intraerythrocytic abundance. The computational predictions were verified against the existing metabolomics data [43] and had an average accuracy of 70%.

With the major improvements in accuracy and sensitivity thresholds that can be achieved by modern RNA-seq approaches [62], utility of transcriptomics studies will increase significantly for experimental and *in silico* functional genomics.

Proteomics

The first two large-scale proteomics datasets published in 2002 [63,64] provided a relatively high coverage (c.a. 45% and 23% respectively) of the expected proteome of *P. falciparum* [65]. While delivering only semi-quantitative results, these studies provided the unique possibility to verify and correct the genome annotation in terms of assignments of open reading frames and splicing patterns, as well as confirming the presence of particular enzymes in different stages of infection.

Several approaches have been used to generate quantitative proteomics datasets. Nirmalan and coworkers have established a method for fully quantitative proteomics using [$^{13}C_6$ $^{15}N_1$]-labeled isoleucine to recognize *de novo* synthetized proteins [66]. Following this method, Prieto et al quantified 1253 proteins in *P. falciparum* trophozoites before and after exposure of the parasite to chloroquine and artemisinin, allowing identification of proteins involved into the parasite's response to treatment with these conventional antimalarial drugs [67]. A

recent alternative method for quantitative proteomics relies on externally supplied known amounts of proteins of interest, obtained using the QconCAT technique [68]. QconCAT-derived proteins are labeled with heavy isotopes and serve both as markers for identification of the similar unlabeled proteins as well as a scale for their quantification [68]. This approach offers a new level of sensitivity and holds promise for comprehensive high-resolution quantitative proteomics in *P. falciparum*.

Taking into account the complexity of the interplay between mRNA and protein abundances (thoroughly examined in [69]), the studies that combine both transcriptomic and proteomic measurements are of particular interest. As a result of such an integrated approach applied to *P. falciparum*, time-delayed correlation between peaks of transcription and maximal abundance of the proteins has been observed for all the glycolytic enzymes with the exception of enolase [60]. However, for several enzymes, accumulation of the mRNA did not correlate with the changes in the abundance of the corresponding enzymes over a complete intraerythrocytic replication cycle [60]. Mair et al demonstrated that in the *Plasmodium* species some genes can be transcribed but not expressed due to translational repressions [70].

Despite the availability of the data and necessary computational methods, proteomics has not been used systematically in the development and analysis of CBMs of *P. falciparum*. However, it has been extensively used for the intercellular pathogen *Trypanosoma cruzi*: Roberts et al have integrated in their CBM a proteomics dataset that constrains fluxes through reactions whenever the corresponding enzyme is not detected in life stage specific proteome [71]. Using this constraint the authors aimed at making the model as representative as possible of the metabolic state of the pathogen at a particular stage of its life cycle. Initially, this resulted into an over-constrained model that was unable to simulate growth, suggesting that, even though enzymes for some reactions were not detected in the proteome, they were likely to be present. On the other hand, constraints inferred from the proteomics data corrected those reaction essentiality predictions that were not in agreement with experimental data without these constraints [71].

While a complex interplay between concentrations of mRNAs and enzymes with the fluxes through the corresponding reactions remains to be elucidated, utilization of transcriptomics and proteomics information together can ensure higher confidence that an enzyme of interest is (or is not) present in the lifecycle stage of interest. For instance, it is reasonable to assume

that a reaction for which both enzyme and transcripts cannot be detected is most likely not to occur and should be tightly constrained in the stage-specific CBMs.

Metabolomics

Several recent reviews discuss in detail the methodologies and techniques that are currently employed for metabolomics of various organisms [72-74] and malaria parasites in particular [75]. The largest metabolomics profile to date was obtained by liquid chromatography coupled to a tandem mass-spectrometry (LC-MS/MS) analysis of both uninfected and *P. falciparum*-infected human erythrocytes [43]. Relative changes in the concentrations of about 90 metabolites were monitored in both the medium and cell lysates within a whole replication cycle, with measurements every 8 hours [43]. A further attempt to correlate these results with gene expression data obtained at the same time points of the infection revealed that, despite a common periodic pattern in gene expression, only less than a third of the measured metabolite abundances fluctuated periodically [43]. The coverage of the *P. falciparum* putative metabolome achieved in this study is between 15 and 20% (estimated similarly to [76]) and compares with the metabolomics of *Leishmania donovani* promastigotes [76]. Currently, no algorithm is available to directly incorporate the relative concentration values into CBMs. Even so, Plata et al used this data as a reference to verify their *in silico* predictions as discussed above. Furthermore, the dataset [43] gives a valuable insight into actively consumed and secreted metabolites, an aspect of the metabolic fate that can be utilized as a constraint by CBMs.

Quantification of absolute concentration values using MS is technically possible, although it is hampered by the need for a standard solution for each metabolite [77,78]. On the contrary, NMR techniques do not require external standards, either for identification or for quantification of metabolites, since the integral of the output signal is proportional to the concentration of the studied nuclei. Despite its lower sensitivity compared to the modern MS methods [74], ^{13}C-NMR can identify metabolites which are otherwise undetectable by MS due to their low ionization potential or lost during sample preparation (e.g. glycerol as reported in [79]). Using ^{1}H-NMR, the concentrations of over 50 metabolites were measured in cell extracts of *P. falciparum* trophozoites [80]. The unbiased nature of the method also enabled identification of some unexpected metabolites (e.g. amino butyric acid and the buffering agent HEPES) present in relatively high concentrations in lysates of the parasite cells [80].

An important drawback of the current metabolomics studies of *P. falciparum* is that, for metabolites present in more than one compartment (e.g. cytosolic and mitochondrial adenosine diphosphate), the measured concentration only reflects an average value, which may differ significantly from the actual concentration in each of the compartments. Metabolomics on individual organelles is an emerging field in functional genomics research, as exemplified in algea [81], that will enrich and verify current knowledge on the subcellular localization of various metabolic processes. The issue of compartmentalization is especially complex in the case of Apicomplexa. Indeed, these obligate intracellular parasites develop in either hepatocytes or erythrocytes in the intermediate host. Moreover, the parasite harbors two symbiotic organelles of prokaryotic origin: the mitochondrion and the apicoplast (relic of a plastid organelle acquired by engulfment of an alga), both of which host metabolic pathways that are crucial to the central carbon metabolism of *P. falciparum*.

Fluxomics

Fluxomics is a largely unexplored area in malaria research, while for experimentally amenable species it represents a relatively well established and rapidly developing field [82,83]. Incorporation of the measured values of flux through the reactions present in CBMs can significantly improve the accuracy of the models by reducing the uncertainty in the ranges and distributions of metabolic fluxes. To the best of our knowledge, the only fluxomics studies in *P. falciparum* to date are the assessments of influx rates for single substrates: glucose (in infected and non-infected human erythrocytes [84,85]), isoleucine [86], pantothenate [87] and inorganic phosphate [88]. Due to the indispensability of these substrates for the parasite, incorporation of these flux values may represent overriding constraints for the CBMs.

In the case of model organisms, e.g. *S. cerevisiae*, measured metabolic fluxes were included into the CBMs and allowed *in silico* resolution of experimentally observed metabolic features that could not be inferred otherwise - neither from transcriptomics nor proteomics data [83]. A relevant example is a large increase in glycolytic flux, which can be maintained by the yeast exposed at low levels of oxygen without changing the expression levels for involved genes [89]. This led to an important conclusion that fluxes in primary metabolism are perhaps more likely to be controlled via regulation of enzymatic activities and not by changes in gene expression as in secondary metabolism [83].

Overall, out of numerous methods developed for the integration of experimental data with CBMs, only a modest proportion have hitherto been applied to pathogenic organisms, and *P. falciparum* in particular. We argue that this is due not only to the complexity of the experimental study of pathogens, but also to the fact that a majority of the *in silico* methods discussed above was initially designed for free-living organisms. Although the methods are, in principle, applicable to intracellular pathogens, some aspects unique to the parasites may be crucial for obtaining relevant computational results. Examples of such aspects would be the common absence of a clearly defined set of substrates and by-products for the metabolism of intracellular parasites, the changes in composition of biomass across multiple stages of their life cycles or sub-optimal utilization of available substrates, etc. These issues, as well as the afore-mentioned experimental challenges, represent the area where we expect future improvements, leading to a better understanding of the metabolic peculiarities encoded in the genome of *P. falciparum,* and other parasitic pathogens.

2. Systematic curation of the genome-scale metabolic model of *P. falciparum*

Over the past decade, several computational approaches have been applied in the studies in *P. falciparum* to extract the most out of the available results by collecting, reconciling and analyzing them in the form of large-scale models [90]. Aiming at a whole-system view and a rigorous analysis, such models enabled generating of novel non-intuitive hypotheses as well as highlighted the points of controversy between the existing experimental results. Metabolic models also have been successfully employed for facilitation of the discovery of essential genes (i.e. potential drug targets) for a number of human pathogens, including several protozoan parasites [91,92].

Growing interest of the malaria research community in systems biology approaches led to reconstruction of at least five *in silico* metabolic models for *P. falciparum* seeking for better understanding of its metabolism and vulnerable points for novel drug intervention [25-28,93]. While being undoubtedly useful, all these models contain a number of important simplifications, which are, unfortunately, not always clearly stated and justified. Some of the simplifications are made to maintain computational feasibility of the simulations. However, there are also those, which reflect the lack of detailed knowledge about simplified aspects at the time a model was created. One of such simplifications, very common in genome-scale metabolic models, is a unidirectionality (irreversibility) pre-assigned to biochemical reactions and to the equations that describe transport of metabolites across cellular membranes. Notably, in a number of cases, such *ad hoc* assignment denotes an experimentally observed irreversibility; indeed, some enzymes facilitate transformation of metabolites exclusively in one direction (i.e. *catalytic irreversibility*). However, the majority of metabolic enzymes catalyze both forward as well as reverse reaction and, in principle, the direction, in which such a reaction occurs, is defined by its thermodynamic properties and conditions. As a simplification, rigorous assessment of the constraints imposed by thermodynamic laws are often omitted in metabolic models and replaced with *ad hoc* assignments of some reactions as unidirectional. Such assignments are important constraints, which may prevent fluxes through reactions in thermodynamically unfeasible directions or, if made spuriously, may lead to incorrect *in silico* predictions. One of the potential consequences of an incorrect assignment of reaction directionality is alteration of *in silico* gene essentiality predictions,

which may change if an irreversible reaction is assigned as reversible or *vise versa* [46]. In the case of well-studied model organisms, such as *E. coli* or *S. cerevisiae*, wealth of experimental information about reversibility of metabolic reactions *in vitro* can be found in primary literature as well as textbooks and databases. Notwithstanding, for an organism which is not readily amenable for protein isolation and orthologous gene expression (e.g. *P. falciparum*), the biochemical properties of a large amount of the enzymes remain unknown. As an implicit simplification, current metabolic models of *P. falciparum* contain unidirectionality constraints taken from the models of completely unrelated, but well-studied organisms [27] or assume all the reactions reversible [26]. In this study, we endeavor to address in a rigorous way the uncertainty about unidirectionality of the reactions in the models of *P. falciparum* using thermodynamics-based flux balance analysis (TFBA).

Further, we demonstrate how the model reconstruction capabilities of the RAVEN Toolbox [94] can aid in addressing another important simplification of the metabolic models – assignment of metabolic functions to genes without explicitly providing evidence or reliability measures. We apply the RAVEN Toolbox for an automated annotation of metabolic capabilities in four different malaria parasites and compare their metabolic capabilities.

2.1 Thermodynamics-based flux balance analysis of metabolism in *P. falciparum*

2.1.1 What is TFBA and why do we need it?

Unidirectionality or bidirectionality of metabolic reactions represent an important constraint in metabolic models. The importance of it has been recognized and first addressed in the context of the genome-scale metabolic model of *E. coli* iAF1260 by Feist et al [95]. In this study thermodynamic information was not part of the model *per se* but was treated as an additional point of reference for confirmation of unidirectionality assigned *ad hoc* to the reactions in the model. In the following years several algorithms for a systematic implementation of thermodynamics-based constraints in the genome-scale metabolic models were developed [96-100]. In this study, we used an automated algorithm developed by Soh & Hatzimanikatis [99] as an extension to the standard formulation of FBA in the COBRA Toolbox [101].

The model presented in this study was created based on iTH366, (http://www.ebi.ac.uk/biomodels-main/MODEL1007060000) by thorough manual curation with the aim to update it with respect to the current knowledge about metabolism of *P. falciparum*. A considerable amount of additional information (i.e. metabolite formula, charges, pKa values etc.) was included in the model to prepare it for incorporation of thermodynamic constraints. On the other hand, inactive ("out of the current scope") part of the model, blocked reactions detected by flux-variability analysis [102] with all the reactions set as reversible, had been removed, thus, leaving only the reactions that could carry flux. Some of the further changes correspond to the latest knowledge about metabolism of the parasite e.g., BCKDH-enzyme complex acts as a pyruvate dehydrogenase in mitochondrion thus providing a missing link between glycolytic pyruvate and tricarboxylic acid cycle [103] (for all the modifications see Table 4, Appendix 2). While all the reactions in iTH366 were pre-defined as uni- or bidirectional, only one out of the all sources of biochemical information mentioned by the authors, namely *iAF1260* model [95], contained thermodynamic information. This means that at best the unidirectionality constraints were copied from the model of unrelated organism with an implicit assumption that they are the same in *P. falciparum*. However, despite a certain degree of similarity in terms of the biochemical pathways present in *E. coli* and *P. falciparum,* directionality and reversibility of the reactions in these pathways may indeed differ between such non-related organisms. We foresee at least two reasons for such potential dissimilarities: firstly, non-orthologous enzymes that carry out the same enzymatic function may not have the same catalytic (ir)reversibility; secondly, even homologous enzymes may catalyze the same reaction in opposite directions in unrelated species. One such example is NADP-dependent isocitrate dehydrogenase (ICDH), which in many eukaryotes operates in a reductive direction (while NAD-dependent ICDH in the oxidative [104]); in *Apicomplexa* mitochondrial NAD-dependent ICDH is absent, thus, the NADP-dependent form catalyzes the reaction in an oxidative direction [105]. Such mode of action was observed for NADP-dependent ICDH under an oxidative stress conditions in *Mus musculus* embryo cells [106]. Thermodynamic constraints allow us to assess the *ad hoc* assignments that were brought to *iTH366* from the other models and use thermodynamic properties to define directionality and reversibility of the reactions.

According to the second principle of thermodynamics, any chemical reaction in order to occur must result in a negative change of Gibbs free energy. Standard Gibbs free energy

change of a reaction ($\mathbf{\Delta_r G^\circ}$) is a sum of the standard Gibbs free energy changes of formation ($\mathbf{\Delta_f G^\circ}$) for the participating molecules respectively multiplied by the stoichiometric coefficients with a positive sign for the reactants and a negative one for the products. In the case of non-standard conditions, $\mathbf{\Delta_r G}$ depends on the standard Gibbs free energy changes of formations for the reactants and the products ($\Delta_f \mathbf{G_i}^\circ$), their concentrations $\boldsymbol{c_i}$ as well as corresponding stoichiometric coefficients $\boldsymbol{\nu_i}$ and an absolute temperature $\boldsymbol{T}$ ($\boldsymbol{R}$ is a universal gas constant equal to 1.987 cal mol^{-1} K^{-1}):

$$\Delta_r G = \sum_i \nu_i \Delta_f G_i^0 + RT \ln(\prod_i c_i^{\nu_i})$$

Experimentally measured $\mathbf{\Delta_f G^\circ}$ values exist for a limited number of compounds. However, using a computational group-contribution method (GCM) [45,107,108], such values could be estimated for a very large number of metabolites, which encompasses a majority of the molecules present in genome-scale models. Notably, $\mathbf{\Delta_r G^\circ}$ imply 1 mol/l concentrations for the reactants and products, which is an unrealistically high concentration for the case of intracellular environments. Thus, in the previous studies, transformed values of $\mathbf{\Delta_r G'}$ were calculated for 1 mmol/l concentrations of the metabolites [95]. Ideally, $\mathbf{\Delta_r G'}$ should be estimated for experimentally measured concentrations of metabolites, however in the absence of such data, a generic broad range of experimentally observed concentrations can be assumed [45]. Due to this uncertainty in the concentrations, $\mathbf{\Delta_r G'}$ is estimated as a range and not a unique value. If the range of $\mathbf{\Delta_r G'}$ for a given reaction encompasses only negative values, this reaction is *thermodynamically irreversible* and occurs only in the forward direction under the given conditions. Conversely, if the range only includes positive values this reaction can exclusively occur in its reverse direction. In the case of the range spanning over positive and negative values the reaction is thermodynamically reversible and can occur in either direction depending on the concentrations of its reactants and products. In a similar way $\mathbf{\Delta_r G'}$ of transport reactions can be estimated subject to the concentration gradient of the transported metabolite(s) across a membrane, charge of the membrane and the gradient of pH [99,109].

In this study, we used a computational algorithm for automated implementation of thermodynamic information into genome-scale CBMs as well as TFBA [99,109]. The algorithm, based on data from GCM, estimated $\mathbf{\Delta_r G'}$ for the reactions included in our model

and formulated additional constraints that assured thermodynamic feasibility of the results obtained in this study.

2.1.2 TFBA of the *P. falciparum* metabolic model with unknown concentrations of metabolites

To date, only a part of the experimental information that is required for TFBA could be found in literature. For this study we used the conditions that are schematically described on the Figure 7. Acidity of the cytosol and the parasitophorous vacuole space in *P. falciparum* had been reported [110] as pH 7.3 and 6.8 respectively, while the values for the lumen of mitochondrion and a plastid-derived organelle (apicoplast) have not been assessed experimentally. Consequently, we assumed the pH in mitochondrion to be in the range 7 to 8, and in the apicoplast – 6 to 8 units. The mitochondrial membrane potential of *P. falciparum* also had not been measured experimentally, thus, we assumed the same potential of -180 mV as reported for mitochondria of *S. cerevisiae* [111]. More precise values can be readily incorporated in the model once they are available. The charge of the endoplasmic membrane of the parasite was set to 95 mV [112]. To reflect biologically realistic concentration of metabolites in the absence of experimental data, we assumed the same concentration ranges as previously used for TFBA by Henry et al [45], namely 10^{-3} to 20 mM for intracellular metabolites and 10^{-5} to 100 mM for extracellular ones.

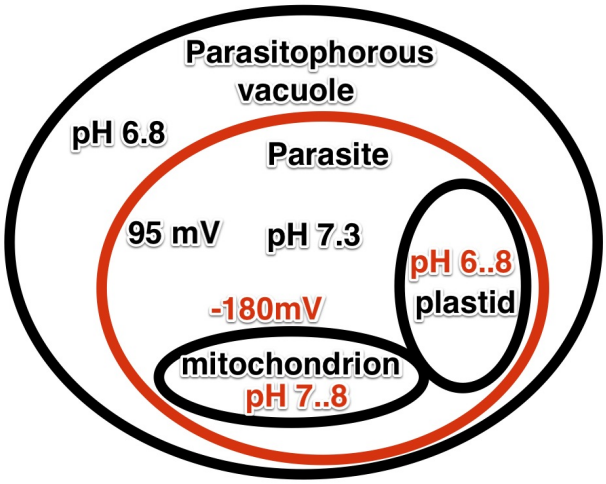


Figure 9. Physicochemical parameters used for different compartments of the model.

Plastid denotes an apicoplast compartment. Analysis of the influence of pH ranges (shown on the figure) on thermodynamic properties ($\Delta_r G'$) of the reactions in these compartments indicated that change of pH within the ranges do not affect $\Delta_r G'$ considerably.

Based on these concentration ranges and group-contribution method, the algorithm [99,109] estimated thermodynamic properties for 514 reactions present in the model (c.a. 80% of all the non-exchange reactions). Among these 514 reactions, 342 were metabolic reactions; 268 of them were *ad hoc* assigned as irreversible in the iTH366 model and only 74 as reversible (Figure 8). In order to compare these assignments with the thermodynamic reversibility we explored the ranges of $\Delta_r G'$ for these reactions using thermodynamics-based variability analysis (TVA [45]). According to the TVA results, only 27 out of 268 *ad hoc* irreversible reactions are thermodynamically irreversible given the assumed ranges of metabolite concentrations (exact values in the Table 5, Appendix 2). When considered each separately, over 90% of the 342 metabolic reactions can have both negative and positive free-energy change depending on the concentrations of their reactants and products within the ranges introduced above.

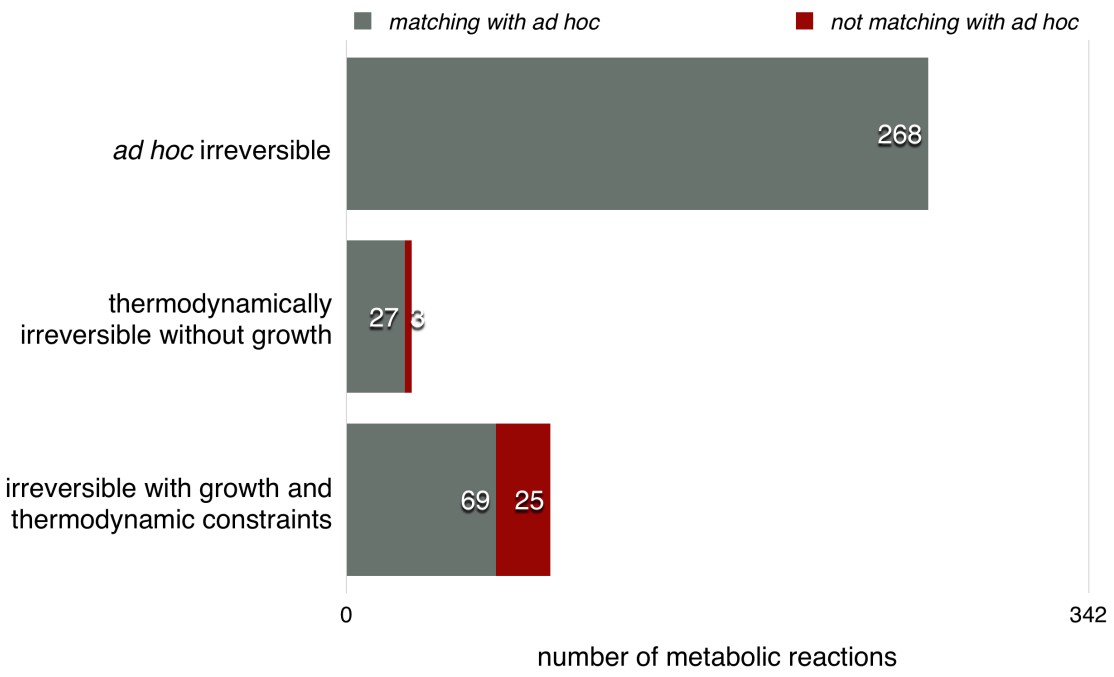


Figure 10. Comparison of the number of *ad hoc* reversible reactions, the irreversibility defined by ranges of free energy change ($\Delta_r G'$), and constraints imposed by simulation of growth

Importantly, 62 reactions, albeit thermodynamically reversible, appeared to be *essentially unidirectional* for production of biomass in the model and for 2 reactions $\Delta_r G'$ ranges shrunk to only negative values rendering them *thermodynamically irreversible* under this constraint

(Figures 8 and 9). Taken together, these results indicated that numerous *ad hoc* assignments of reactions as unidirectional in iTH366 were not supported by our thermodynamic estimates.

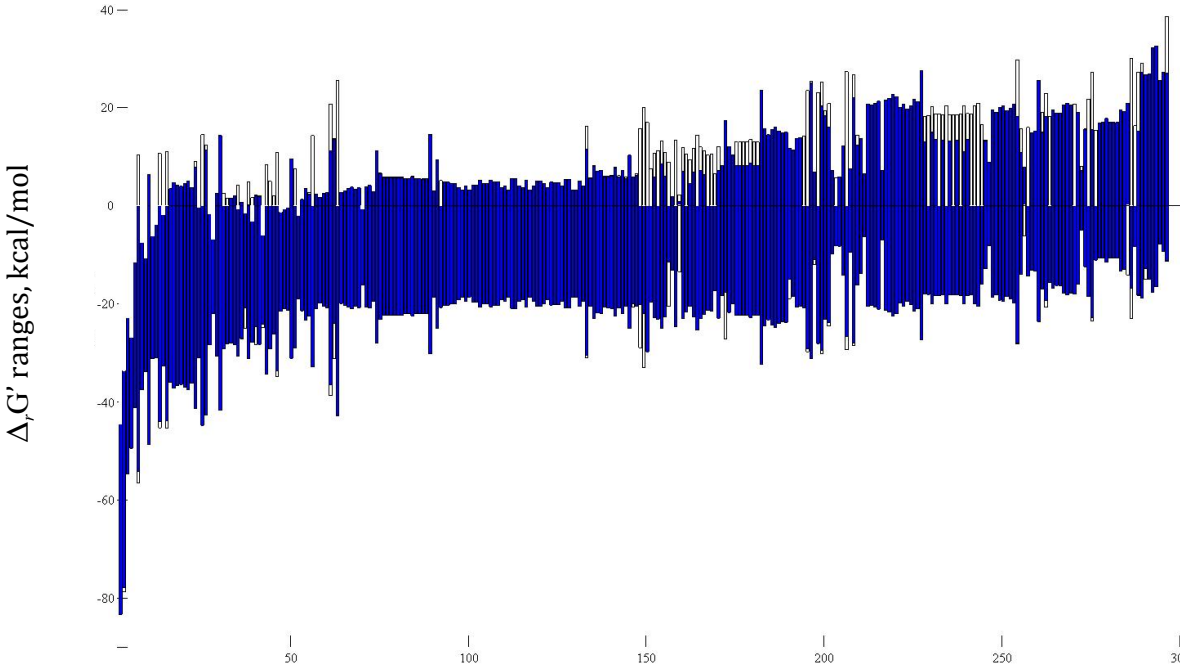


Figure 11. The ranges of free energy change ($\Delta_r G'$) for the cytosolic reactions with and without requirement of biomass production.

Each vertical bar corresponds to one of the reactions: white parts show the values feasible only when no growth is required, while the blue ones denote overlapping parts of the $\Delta_r G'$ ranges feasible with and without production of biomass. Requirement of growth forces a number of reactions to operate only in one direction (the ranges that are divided into blue and white part by x-axis).

One of the reactions in MEP/DOXP pathway (MECDPS_ap) localized in the apicoplast caused thermodynamic unfeasibility of isoprenoid production. This linear pathway produces key precursors of several biomass building blocks and thus is essential for growth of the parasite. Such inconsistency between the thermodynamics-based and growth-allowing directionality had been previously noticed in the model of *E. coli* [45]. The authors estimated standard $\Delta_r G^0$ of this reaction as a largely positive (25.09 ± 2.67 kcal/mol) that does not allow the reaction to proceed in its experimentally observed direction [113]. Herein, we report that this $\Delta_r G'$ remains positive within the whole span of the considered concentrations and ranges from 4.57 to 38.6 kcal/mol. Interestingly, the enzyme that catalyzes this reaction in *E. coli* (IspF) was found to be fused with another one, IspD, responsible for the one-before-preceding step of the pathway. Such a bifunctional enzyme might allow so-called

substrate channeling [114] and increase thermodynamic favorability of the reaction in the necessary direction. Nevertheless, further studies of similarly fused enzyme of *Agrobacterium tumifaciens* did not yield any evidence of substrate channeling between these two enzymatic steps [115]. Our analysis shows that negative value of Gibbs free energy change for this reaction could be achieved either when uncertainty bounds in the estimated standard Gibbs free energies of formation for the metabolites is increased by a factor of 1.6, or, alternatively, by allowing much higher concentrations for the reactants (0.05M instead of 0.02M) and much lower concentration of the product (2×10^{-8} M instead of 10^{-6} M). Such unprecedented assumptions are outside the scope of TFBA and, hence the question how MECDPS reaction occurs in a thermodynamically unfavorable direction remains open.

Using TVA, we estimated critical (the lowest and the highest) concentrations of all the metabolites that sustain thermodynamic feasibility of each separate reaction (in the directions necessary for growth), as well as a whole model. For several metabolites (18) these ranges were not equal to the assumed wide ranges of metabolite concentrations (Figure 10). These represent critical concentrations of the metabolites across the membranes; beyond these concentrations on each side of the respective membranes the transport process becomes thermodynamically unfeasible.

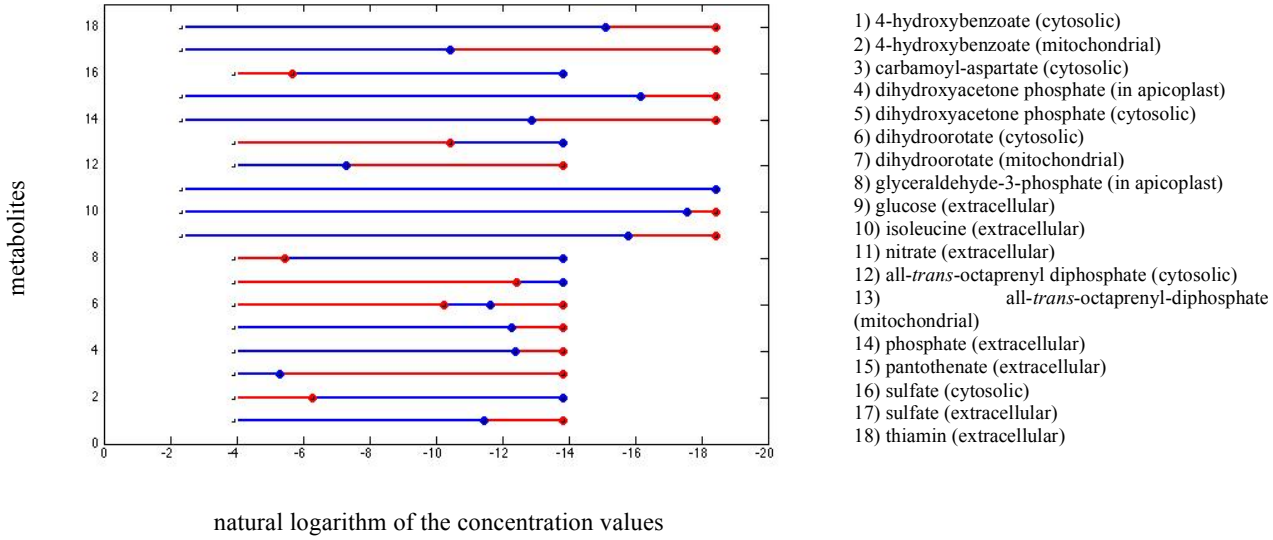


Figure 12. Logarithmic ranges of the metabolite concentrations, which do not reach lower or upper bound while maintaining thermodynamic feasibility of the model.

The blue bars highlight that part of the ranges that are compatible with production of biomass; conversely, the red parts are incompatible and together with the ranges in blue show the complete assumed concentration intervals.

Note that these results are subject to our assumptions on the charges of the membranes, the gradient of pH across them, as well as the transport mechanism specified for each of these molecules in the iTH366. These metabolites represent essential substrates that the model cannot obtain from any route that bypasses corresponding transport reactions (e.g. transport of 4-hydroxybenzoate between cytosol and mitochondrion).

In addition to the prediction of feasible metabolite concentration ranges, implementation of thermodynamic constraints also enables incorporation of the concentration values that were measured experimentally as described in the following section.

2.1.3 Integration of metabolomics data as constraints in the CBMs of *P. falciparum*

To explore the possibility to integrate metabolomics data as additional constraints, we used the results of H^1 NMR-based *in vitro* study on the blood stage infection by *P. falciparum*, where the authors quantified over 40 intracellular metabolites [80], the majority of which were present in our model. According to the authors, the metabolite extraction with perchloric acid was the most efficient, thus, we used the values acquired with this method allowing for the error ranges reported in the publication [80]. Based on these values, we estimated logarithmic ranges of the concentrations (Table 2) and used them in our model instead of the previously assumed wide ranges for these metabolites. An important assumption we made here was that the measured concentrations were representative of cytosolic as well as mitochondrial and apicoplast values.

Upon incorporation of the concentration ranges the production of biomass in the model remained to be thermodynamically feasible, which proves the model's consistency with this experimental data. Using TVA we again estimated concentration ranges of all metabolites which maintain thermodynamic feasibility of the biomass production in the model. Apart from the metabolomics-based experimental ranges, we also observed additional ranges that were no longer the same as generic ranges (Figure 11). This represents the phenomenon of propagation of the constraints in the network and a downstream effect of the incorporation of the metabolomics data.

Metabolite	Concentration value, C_i (mM)	Concentration error range, δ_i	Min. concentration, $(C_i - \delta_i)$	Max. concentration, $(C_i + \delta_i)$	$\log_e$ of min. concentration $\ln(C_i - \delta_i)$	$\log_e$ of max. concentration $\ln(C_i + \delta_i)$	Metabolite identifiers	
							SEED database	iTH366 model
cytosolic compartment								
Alanine	0.48	0.12	0.36	0.6	-7.9294	-7.4186	cpd00035	ala_L_c
Arginine	4.2	1.1	3.1	5.3	-5.7764	-5.2400	cpd00051	arg_L_c
Asparagine	0.7	0.4	0.3	1.1	-8.1117	-6.8124	cpd00132	asn_L_c
Aspartate	2	0.5	1.5	2.5	-6.5023	-5.9915	cpd00041	asp_L_c
4-Aminobutanoate	2.3	0.5	1.8	2.8	-6.3200	-5.8781	cpd00281	gaba_m
Glutamate	11	4	7	15	-4.9618	-4.1997	cpd00023	glu_L_c
Glutamine	1.34	0.3	1.04	1.64	-6.8685	-6.4131	cpd00053	gln_L_c
Glycine	0.43	0.08	0.35	0.51	-7.9576	-7.5811	cpd00033	gly_c
Histidine	0.16	0.04	0.12	0.2	-9.0280	-8.5172	cpd00119	his_L_c
Isoleucine	0.16	0.05	0.11	0.21	-9.1150	-8.4684	cpd00322	ile_L_c
Leucine	0.37	0.12	0.25	0.49	-8.2940	-7.6211	cpd00107	leu_L_c
Lysine	1.3	0.3	1	1.6	-6.9078	-6.4378	cpd00039	lys_L_c
Methionine							cpd00060	met_L_c
Phenylalanine	0.16	0.05	0.11	0.21	-9.1150	-8.4684	cpd00066	phe_L_c
Serine	–	–	–	–			cpd00054	ser_L_c
Threonine	0.43	0.1	0.33	0.53	-8.0164	-7.5426	cpd00161	thr_L_c
Tyrosine	0.19	0.04	0.15	0.23	-8.8049	-8.3774	cpd00069	tyr_L_c
Valine	0.29	0.08	0.21	0.37	-8.4684	-7.9020	cpd00156	val_L_c
Reduced glutathione	0.6	1.2	0.001	1.8	-13.8155	-6.3200	cpd00042	gthrd_c
Oxidised glutathione	1.7	0.7	1	2.4	-6.9078	-6.0323	cpd00111	gthox_m
myo-Inositol	0.24	0.08	0.16	0.32	-8.7403	-8.0472	cpd00121	inost_c
Phosphocholine	1.49	0.32	1.17	1.81	-6.7508	-6.3144	cpd00457	cholp_c
Phosphoethanolamine	4.5	1.5	3	6	-5.8091	-5.1160	cpd00285	ethamp_c
AMP	0.36	0.15	0.21	0.51	-8.4684	-7.5811	cpd00018	amp_c
ADP	1.1	0.4	0.7	1.5	-7.2644	-6.5023	cpd00008	adp_c
ATP	2	0.9	1.1	2.9	-6.8124	-5.8430	cpd00002	atp_c
NAD	2.2	0.5	1.7	2.7	-6.3771	-5.9145	cpd00003	nad_c
Acetate	0.36	0.12	0.24	0.48	-8.3349	-7.6417	cpd00029	ac_c
Formate	0.22	0.07	0.15	0.29	-8.8049	-8.1456	cpd00047	for_c
Fumarate	0.16	0.04	0.12	0.2	-9.0280	-8.5172	cpd00106	fum_c
α-Ketoglutarate	0.24	0.07	0.17	0.31	-8.6797	-8.0789	cpd00024	akg_c
Lactate	1.1	0.8	0.3	1.9	-8.1117	-6.2659	cpd00159	lac_L_c
Malate	1.8	0.5	1.3	2.3	-6.6454	-6.0748	cpd00130	mal_L_c
Succinate	0.41	0.07	0.34	0.48	-7.9866	-7.6417	cpd00036	succ_c
Putrescine	2.9	0.8	2.1	3.7	-6.1658	-5.5994	cpd00118	ptrc_c
Spermidine	5.7	1.6	4.1	7.3	-5.4968	-4.9199	cpd00264	spmd_c
mitochondrial compartment								
Alanine	0.48	0.12	0.36	0.6	-7.9294	-7.4186	cpd00035	ala_L_m
Aspartate	2	0.5	1.5	2.5	-6.5023	-5.9915	cpd00041	asp_L_m
Glutamate	11	4	7	15	-4.9618	-4.1997	cpd00023	glu_L_m
Glycine	0.43	0.08	0.35	0.51	-7.9576	-7.5811	cpd00033	gly_m
Tyrosine	0.19	0.04	0.15	0.23	-8.8049	-8.3774	cpd00069	tyr_L_m
Reduced	0.6	1.2	0.001	1.8	-13.8155	-6.3200	cpd00042	gthrd_m
ADP	1.1	0.4	0.7	1.5	-7.2644	-6.5023	cpd00008	adp_m
ATP	2	0.9	1.1	2.9	-6.8124	-5.8430	cpd00002	atp_m
NAD	2.2	0.5	1.7	2.7	-6.3771	-5.9145	cpd00003	nad_m
Fumarate	0.16	0.04	0.12	0.2	-9.0280	-8.5172	cpd00106	fum_m
α-Ketoglutarate	0.24	0.07	0.17	0.31	-8.6797	-8.0789	cpd00024	akg_m
Malate	1.8	0.5	1.3	2.3	-6.6454	-6.0748	cpd00130	mal_L_m
Succinate	0.41	0.07	0.34	0.48	-7.9866	-7.6417	cpd00036	succ_m
apicoplast compartment								
ADP	1.1	0.4	0.7	1.5	-7.2644	-6.5023	cpd00008	adp_ap
ATP	2	0.9	1.1	2.9	-6.8124	-5.8430	cpd00002	atp_ap
NAD	2.2	0.5	1.7	2.7	-6.3771	-5.9145	cpd00003	nad_ap

Table 2. Logarithmic concentration ranges estimated based on the metabolomics dataset [80]

The second and third columns contain the values of concentrations and error ranges respectively in mM, taken directly from the reported experimental results [80]. By addition and subtraction of the corresponding values of second and third column the values of minimal and maximal concentrations and corresponding logarithmic values were estimated.

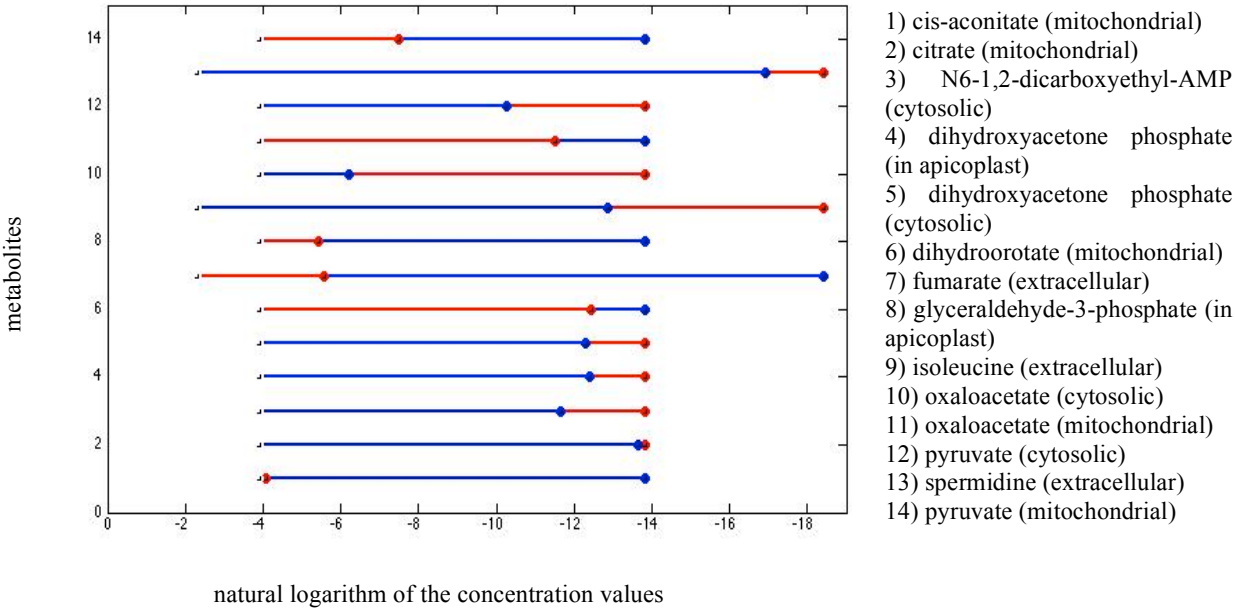


Figure 13. Logarithmic ranges of the metabolite concentrations that are constrained by incorporation of the metabolomics dataset.

The red and blue parts for each range add up to the complete allowed span (wider of extracellular metabolites than for intracellular ones). The blue parts denote thermodynamically feasible ranges and unfeasible parts are in red.

As a result of the tighter concentration ranges (both directly incorporated (Table 2) and indirectly affected ones, Figure 11), $\Delta_r G'$ spans shrank for the reactions that involved these metabolites: spans of $\Delta_r G'$ for 214 reactions changed (c.a. 41.6% of all the reactions with estimated thermodynamic properties in our model; exact values in the Table 5, Appendix 2). In several cases, the new ranges did not include both positive and negative values any more, implying that these reactions are thermodynamically irreversible in the given conditions (Figure 12). For instance, the reaction R00261 (glutamate carboxylase) could either produce glutamate from 4-aminobutanoate (GABA) and carbon dioxide or split a molecule of glutamate into these two compounds when we considered the generic concentration range for the reactants and products. This reaction becomes thermodynamically irreversible in the direction of glutamate decarboxylation upon incorporation of the measured concentrations of glutamate and GABA. However, the absence of known genes encoding 4-aminobutanoate aminotransferase and succinate-semialdehyde dehydrogenase (enzymes of the so-called GABA-shunt, both present and functional in *T. gondii*) in *P. falciparum* makes the downstream metabolic fate of this non-proteinogenic amino acid uncertain. In plants, glutamate decarboxylase reaction is used as a countermeasure in the case of acidification of

cytosol [116], but whether this might be a potential function of GABA in malaria parasites remains to be investigated.

Comparison of the gene essentiality predictions obtained using TFBA in the presence and absence of measured metabolite concentrations revealed several differences. For example, only upon integration of the metabolomics data did the model predict the following three enzymes of TCA cycle to be essential: citrate synthase (PF10_0218, PF3D7_1022500), aconitate hydratase (PF13_0229, PF3D7_1342100) and NADP-dependent isocitrate dehydrogenase (PF13_0242, PF3D7_1345700). Interestingly, the essentiality prediction for aconitate hydratase is consistent with the latest experimental results on assessment of essentiality of TCA cycle enzymes in *P. falciparum* [117]. This match represents an appealing instance for further investigation, as the reasons for an experimentally observed essentiality of this particular enzyme and not the preceding citrate synthase and the succeeding isocitrate dehydrogenase remain unclear.

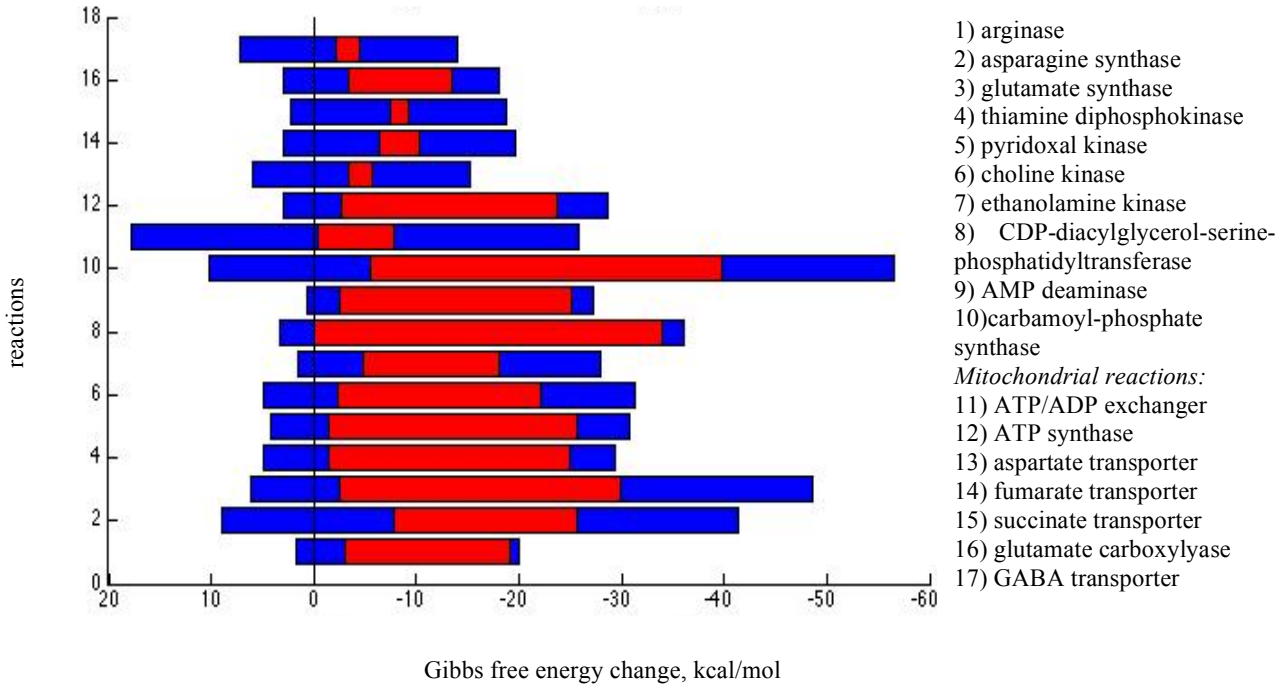


Figure 14. $\Delta_r G'$ spans of the reactions that were thermodynamically reversible with the assumed range of concentrations (blue and red parts of the bars together) yet became irreversible when experimentally measured concentrations were incorporated (red parts of the bars).

Notably, incorporation of the experimentally measured concentrations also restricted reversibility of several kinase reactions and mitochondrial transporters as well as prevented a bidirectionality in utilization of ATP/ADP exchanger and ATP-synthase.

As a summary of these TFBA studies on the iTH366-based, CBM we conclude that initially present *ad hoc* constraints were much more restrictive than thermodynamics-based ones given the wide concentration ranges assumed. This is similar to what had been previously observed in the comparison of *ad hoc* assigned unidirectionality with the thermodynamic estimates in the iAF1260 model, where the authors claimed that "*reversibility set forth for the reactions in the reconstruction is often based on the physiological behavior of the reactions in the cell, not using the relatively broad concentration range achievable for metabolites along with the uncertainty inherent in the utilized method* [for estimating thermodynamic properties]" [95]. Indeed, in the case when such experimental data is available for the organism of interest, this information may be used for adding *ad hoc* constraints. However, the same may not be valid when unidirectionality is pre-assigned based on experimental information for unrelated species. Thus, we emphasize that pre-assigned directionality of reactions in the models of *P. falciparum* needs careful consideration and TFBA analysis provides means for this. We demonstrated that incorporation of the metabolomics data as further constraints additionally shrinks the feasible concentration ranges for some metabolites and Gibbs free energy change spans for reactions, ultimately indicating that some of them are thermodynamically irreversible in these conditions. Recently published [75][114] as well as forthcoming larger metabolomics datasets will be instrumental for more precise assessment every *ad hoc* assignment in the model with respect to the thermodynamics-based reversibility.

2.2 Automated *de novo* reconstruction of genome-scale metabolic networks of apicomplexan species: comparison of malaria parasites

Using the reconstruction algorithm described below, I generated functional annotation that represents the metabolic capabilities of four malaria parasites: *P. falciparum*, *P. vivax*, *P. berghei*, *P. cynomolgi*. Among these *P. falciparum* and *P. vivax* are the most life-threatening human malaria parasites, while *P. berghei* and *P. cynomolgi* are of interest as the related (i.e. to some extent representative) animal-infecting species. The information about the metabolic capabilities of these parasites presented herein has at least two eventual applications: (1) it serves as an independent source of functional annotation of *P. falciparum* genes, which is invaluable input for reconciliation and building an up-to-date CBM of metabolism in this

parasite; (2) comparison of the enzymatic activities present in different malaria parasites gives knowledge about common, as well as unique, features of their metabolism, which highlights the aspects in which animal-borne parasites are not representative of their human counterparts.

Using the model reconstruction capabilities of the RAVEN Toolbox [94], I herein illustrate how we can address another concern inherent to all the models of *P. falciparum* published to date, namely, reliance on external sources of functional annotation of the genes. As a consequence, these models do not contain any information on reliability of the functional annotations they were built upon. Using only the amino acid sequences of the ORFs extracted from PlasmoDB for the aforementioned malaria species, I obtained an independent and metabolic-model oriented functional annotation. Note, that independent in this case means that the assignment of functions to the ORFs was completely independent from the functional annotation existing in PlasmoDB. These annotations are based on (and, thus, linked to) the KEGG database of biochemical pathways through KEGG orthology identifiers (K-ids).

For reconstruction of metabolic networks using the RAVEN Toolbox only a subset of all existing K-ids is selected, namely those which are associated with at least one fully-defined biochemical reaction (i.e. not marked as generic, incomplete or unclear). These K-ids are similar to E.C. numbers, yet more precise in the definition of the relation between a biochemical (catalytic) activity they denote with a type of enzyme that catalyzes this activity. Thus, as a rule, for one E.C. identifier there is one or more K-id in KEGG, each encompassing the genes that not only share the same catalytic activity but also have a common evolutionary origin (i.e. orthology). The information from the KEGG database is downloaded and treated such that for each K-id all the associated gene sequences from all the available species are assembled in individual files. The sequences in these files are further aligned between themselves, which is a computationally intensive process. However, the task of such alignment for each K-id is a completely independent process. Thus, for these tasks it is suitable to use parallel computing to perform alignments for each K-id independently and simultaneously. From the resulting multi-sequence alignments, hidden Markov models (HMM) are generated by HMMer software [118] for each K-id. These HMMs represented unique sequence "fingerprints" for each K-id (therefore an enzymatic activity) and are further used for functional annotation of a genome. The process of functional annotation relies on pairwise comparisons of every HMM with every ORF amino acid sequence present in a genome of interest. This step also requires significant computational resources and can be

expedited by means of a parallel computing. The result of the comparisons is an output file created for each K-id, which contains the identifiers of the ORFs matching to the corresponding HMM. These matches are described using two reliability measures – bit-score and expect value (e-value) for each ORF (Figure 13). A scope of these output files constitutes a new, independent, functional annotation of a genome, based on which a metabolic network can be reconstructed.

```
Query HMM:   K00953
Accession:   [none]
Description: [none]
  [HMM has been calibrated; E-values are empirical estimates]

Scores for complete sequences (score includes all domains):
Sequence        Description                              Score    E-value
--------        -----------                              -----    -------
PF3D7_1015000   | organism=Plasmodium_falciparum_3D7     293.1    3.2e-85
PF3D7_1359100   | organism=Plasmodium_falciparum_3D7      69.9    5.1e-18
PF3D7_0606800   | organism=Plasmodium_falciparum_3D7    -139.3        1.9
PF3D7_0505100   | organism=Plasmodium_falciparum_3D7    -140.8        2.5
PF3D7_0216400   | organism=Plasmodium_falciparum_3D7    -141.1        2.6


Query HMM:   K11753
Accession:   [none]
Description: [none]
  [HMM has been calibrated; E-values are empirical estimates]

Scores for complete sequences (score includes all domains):
Sequence      Description                                Score    E-value
--------      -----------                                -----    -------
PF3D7_1359100 | organism=Plasmodium_falciparum_3D7 |    -139.9     0.0024
PF3D7_1316600 | organism=Plasmodium_falciparum_3D7 |    -186.6        2.2
PF3D7_1009000 | organism=Plasmodium_falciparum_3D7 |    -187.6        2.6
PF3D7_0209000 | organism=Plasmodium_falciparum_3D7 |    -187.9        2.7
PF3D7_1103800 | organism=Plasmodium_falciparum_3D7 |    -194.2        6.8
```

Figure 15. Example of the output data of the annotation pipeline in the RAVEN Toolbox

Two extracts of the output files in the Figure 13 show the results of the matching process between the ORFs of *P. falciparum* (top 5 best hits shown) to the HMMs of two KEGG orthology groups. Both of the K-ids are associated to the E.C. number 2.7.7.2 (FMN adenylyltransferase activity), but K00953 groups monofunctional enzymes that catalyze only one reaction, while K11753 is a group of bifunctional enzymes that combine the activities E.C. 2.7.7.2 and E.C. 2.7.1.26 (riboflavin kinase) in one protein. As evident from the matching scores, the gene PF3D7_1015000 most likely encodes a monofunctional protein and not the bifunctional one. The second match (PF3D7_1359100) represents an interesting unclear case, where the gene shows significant similarity to two K-ids: K00953 (Figure 13) but also K00861 monofunctional riboflavin kinases (not shown, e-value of $4.1e^{-56}$). While this is implying a putative bifunctionality, there is no similarity detected to the orthology

group K11753 of bifunctional riboflavin kinases/FMN adenylyltransferases. Consistently with our findings, PF3D7_1359100 is annotated in PlasmoDB as "riboflavin kinase / FAD synthase family protein, putative". In this controversial situation, only isolation and biochemical characterization of the enzyme can definitively confirm presence or absence of the second catalytic function.

The network reconstruction process relies on a user-defined e-value threshold that differentiates between the reliable matches (i.e. resulting in annotation of an ORF with the function corresponding to the best matching K-id) and insignificant matches that result in no assignment of a functional annotation. There are also two other user-defined parameters that are necessary to process the cases of (i) multiple genes matching to the same K-id and (ii) one gene matching to multiple K-ids with different confidence scores. Based on these thresholds and the output files described above, a metabolic network is built by inclusion of the reactions associated in the KEGG database with the K-ids assigned to all the annotated ORFs.

All four malaria parasites considered in this study possess a remarkably similar set of metabolic enzymes, despite the difference in intermediate host organisms (Figure 14, Table 6). This observation is in accordance with the previous comparative genomics study [119]. Our results also corroborate an earlier report on the absence of several enzymes in *P. berghei*, which are present in primate malaria species [119]. Importantly, among these are phosphoethanolamine N-methyltransferase (phospholipid metabolism) and five enzymes necessary for the *de novo* biosynthesis of thiamine pyrophosphate. We also confirm that the enzyme aspyrase can be reliably annotated in the genome of *P. falciparum* (PF3D7_1431800 bit-score 436.7 e-value $1.9e^{-128}$), while in the three other species the best matching scores are over 10 folds lower and therefore do not allow trustworthy assignment of a function. All these cases of consistencies with the previously published findings represent the first validation of our independent annotation approach. Further, our comparative study is focused specifically on metabolism thus provides more exhaustive coverage of potential differences in metabolic capabilities. In particular, we report that *P. cynomolgi* apparently lacks cytochrome oxidase subunit I, which is encoded in the mitochondrial genomes of *P. falciparum*, *P. berghei* and *P. vivax*. We also could not detect genes encoding N-acetylglucosaminyl-phosphatidylinositol deacetylase, phosphatidylinositol bisphosphatase and 1-phosphatidylinositol-5-phosphate 4-kinase in *P. cynomolgi*, unlike in the three other species. Indeed, these examples are just a little snapshot of all the putative differences our comparison identified. Further reconstruction of the corresponding genome-scale metabolic

models is necessary to assess the consequences of these differences. Only a rigorous analysis with respect to an appropriate host cell environment can unravel the implications of these dissimilarities for the overall metabolic capabilities of the parasites.

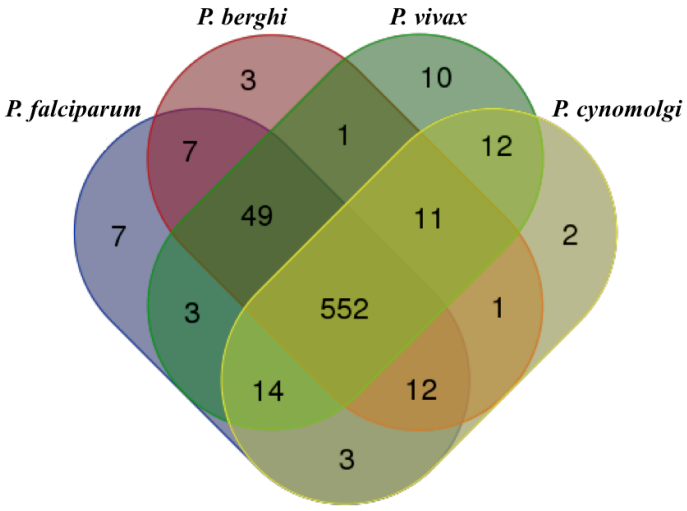


Figure 16. Comparison of the scopes of metabolic functions (expressed as KEGG orthology identifiers) annotated by the RAVEN Toolbox in the genomes of 4 *Plasmodium* spp.

In this comparison we used a permissive e-value cut-off equal to $1e^{-5}$; see Table 6 for a full list of the common and unique identifiers. This diagram was generated by the web-based software accessible at http://bioinformatics.psb.ugent.be/webtools/Venn/

Studies described in this chapter were aimed at curation and further improvement of the existing genome-scale metabolic models of *P. falciparum*. Using thermodynamics-based flux balance analysis, we demonstrated a power of the method to extend the scope of CBMs with incorporation of metabolomics datasets and to thoroughly re-evaluate the *ad hoc* reaction unidirectionality assignments. This, taken together with a careful revision of the metabolic capabilities of the parasite and understanding differences in these capabilities between malaria species, creates a foundation for next-generation CBMs to be reconstructed for *P. falciparum* and other *Plasmodium* spp.

3. Metabolic needs and capabilities of *T. gondii* through combined computational and experimental analysis

This publication contributes the first genome-scale computational model of metabolism in *T. gondii* coupled with rigorous computational analyses and an experimental validation of several model-based hypotheses.

The main outcomes of this work are summarized as follows:

1 ToxoNet1, a comprehensive metabolic model of *T. gondii* was built using a state-of-the-art semi-automated approach.
2 A number of aspects of the metabolism of the parasite that are not currently well understood were highlighted as a result of the reconstruction process.
3 Computational analyses of the model delivered several important insights into metabolic needs and capabilities of the parasite, as well as a list of genes that are putatively essential for its replication.
4 Experimental investigation of the routes leading to production of acetyl-CoA in the cytoplasm of the parasite clarified this important aspect of its metabolism and corroborated *in silico* predictions on individual dispensability yet synthetic essentiality of the enzymes ATP-citrate lyase and acetyl-CoA synthetase.

This study has broad implications for future studies as well, as a general understanding of the metabolism in the apicomplexan parasite *T. gondii,* which will ultimately lead to development of novel antiparasitic medicines.

Author contributions: VH and DSF conceived and designed the study; ST performed model reconstruction and computational studies and analyzed their results; ST, RDO, DSF, RA, JN, and VH helped curating the model; RDO generated and characterized the transgenic parasites and analyzed the phenotypes; ST, RDO, DSF and VH wrote the manuscript.

PLOS | COMPUTATIONAL BIOLOGY

RESEARCH ARTICLE

Metabolic Needs and Capabilities of *Toxoplasma gondii* Through Combined Computational and Experimental Analysis

Stepan Tymoshenko[1,2,3], Rebecca D. Oppenheim[1], Rasmus Agren[4¤], Jens Nielsen[4], Dominique Soldati-Favre[1]*, Vassily Hatzimanikatis[2,3]*

1 Department of Microbiology and Molecular Medicine, Faculty of Medicine, University of Geneva, CMU, Geneva, Switzerland, **2** Laboratory of Computational Systems Biotechnology, École Polytechnique Fédérale de Lausanne, EPFL, Lausanne, Switzerland, **3** Swiss Institute of Bioinformatics, Quartier Sorge, Batiment Genopode, Lausanne, Switzerland, **4** Department of Chemical and Biological Engineering, Chalmers University of Technology, Gothenburg, Sweden

¤ Current address: Novo Nordisk A/S, Bagsvaerd, Denmark
* Dominique.Soldati-Favre@unige.ch (DSF); Vassily.Hatzimanikatis@epfl.ch (VH)

Citation: Tymoshenko S, Oppenheim RD, Agren R, Nielsen J, Soldati-Favre D, Hatzimanikatis V (1969) Metabolic Needs and Capabilities of *Toxoplasma gondii* Through Combined Computational and Experimental Analysis. PLoS Comput Biol 0(0): e1004261. doi:10.1371/journal.pcbi.1004261

Editor: Costas D. Maranas, The Pennsylvania State University, UNITED STATES

Received: January 19, 2015

Accepted: March 31, 2015

Published: July 20, 1969

Data Availability Statement: All relevant data are within the paper and its Supporting Information files. The file containing the model data will be available at the Biomodels database: http://www.ebi.ac.uk/biomodels-main

Funding: ST is supported by an IPhD grant and the RTD grant MalarX, both within SystemsX.ch, the Swiss Initiative for Systems Biology evaluated by the Swiss National Science Foundation; RDO is supported by the iGE3 student salary award. DS is an international Scholar of the Howard Hughes Medical Institutes. VH is supported by the RTD grants

Abstract

Toxoplasma gondii is a human pathogen prevalent worldwide that poses a challenging and unmet need for novel treatment of toxoplasmosis. Using a semi-automated reconstruction algorithm, we reconstructed a genome-scale metabolic model, ToxoNet1. The reconstruction process and flux-balance analysis of the model offer a systematic overview of the metabolic capabilities of this parasite. Using ToxoNet1 we have identified significant gaps in the current knowledge of *Toxoplasma* metabolic pathways and have clarified its minimal nutritional requirements for replication. By probing the model via metabolic tasks, we have further defined sets of alternative precursors necessary for parasite growth. Within a human host cell environment, ToxoNet1 predicts a minimal set of 53 enzyme-coding genes and 76 reactions to be essential for parasite replication. Double-gene-essentiality analysis identified 20 pairs of genes for which simultaneous deletion is deleterious. To validate several predictions of ToxoNet1 we have performed experimental analyses of cytosolic acetyl-CoA biosynthesis. ATP-citrate lyase and acetyl-CoA synthase were localised and their corresponding genes disrupted, establishing that each of these enzymes is dispensable for the growth of *T. gondii*, however together they make a synthetic lethal pair.

Author Summary

Understanding the metabolism of disease-causing microorganisms can guide drug design through the identification of metabolic enzymes whose activity is indispensable for important cellular functions. Such understanding can come from the reconstruction and computational analysis of metabolic networks. In this study we have focused on the metabolism of an opportunistic human pathogen, *Toxoplasma gondii*, which chronically infects about 30% of humans worldwide. Through a semi-automatic reconstruction we have developed

MetaNetX and MalarX (both within SystemsX.ch), and the École polytechnique fédérale de Lausanne (EPFL). RA was supported by the Knut and Alice Wallenberg Foundation. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

ToxoNet1, a comprehensive metabolic model of *Toxoplasma gondii*, and we have performed an extensive computational analysis to explore the properties of *Toxoplasma* metabolism. In particular, we have identified and classified the minimal set of substrates the parasite utilizes for growth, along with the genes and pairs of genes that are essential for cellular functions such as growth and energy metabolism. We have validated our computational predictions that the genes encoding acetyl-CoA synthase and ATP-citrate lyase enzymes serve complementary function but their simultaneous disruption does not allow cell growth. This study presents a number of hypotheses generated using ToxoNet1, which can lead to the discovery of novel antiparasitic drug targets.

Introduction

The phylum of Apicomplexa comprises a large number of obligate intracellular parasites that can infect organisms across the whole animal kingdom. An important member of this phylum, *Toxoplasma gondii*, is a ubiquitous opportunistic pathogen responsible for one of the most common parasitic infections in humans and warm-blooded animals. It is estimated that up to 30% of the human population is chronically infected [1]. Toxoplasmosis is largely asymptomatic in healthy adults but can cause severe disease or even death in immunocompromised individuals and can lead to complications in development of the foetus, if primary infection occurs during pregnancy [2].

T. gondii possesses a complex life cycle, which is composed an asexual replicative stage in the intermediate host and a sexually replicative stage within the definitive feline host. During the asexual phase, *T. gondii* can switch from a fast-replicative tachyzoite form, which causes acute disease, to a slow-growing bradyzoite stage, which forms cysts that are characteristic of chronic infection. The encysted, slow growing form is resistant to commonly used drugs and immune system attack. Few efficient medicines are available to treat toxoplasmosis and they mainly treat the acute phase of the disease. Furthermore, poor tolerance of these drugs promotes the search for novel drug targets.

Unlike other notorious apicomplexan parasites that infect a narrow range of host cell types, such as the *Plasmodium* species, *T. gondii* is able to invade and asexually replicate within virtually any nucleated cell of warm-blooded animals. The broad range of cells amenable to infection by *T. gondii* reflects the plasticity of the parasite's metabolism and versatility in accessing and utilising nutrients to support its intracellular growth [3,4]. The complexity of decoupling the metabolic processes of this intracellular pathogen from those of the infected host limits the depth of our understanding about the metabolic capabilities of *T. gondii*.

Currently, no adequate experimental approaches exist to answer comprehensively the following important questions: 1) what substrates are available within the host cell that are necessary for *T. gondii* replication and which of these are dispensable; 2) in which intracellular compartments do the enzymatic activities annotated at the genome level occur; 3) which of these enzymatic activities are indispensable for replication or other vital processes of the parasite. While achievable, the application of high-throughput gene knockout or knockdown strategies to globally determine gene essentiality in *T. gondii* remains a major undertaking.

Computational (i.e. *in silico*) metabolic modelling coupled with systematic analyses facilitates the study of biological systems. This modern approach of systems biology has been extensively exerted to predict gene essentiality in various bacteria, including numerous pathogenic species [5]. *In silico* metabolic models offer a cost-effective pipeline to identify putatively indispensable metabolic processes that, in the case of pathogens, represent potential targets for

therapeutic intervention [6]. Recently, models have been constructed for eukaryotic pathogens including for members of the phylum Apicomplexa [7–10].

In this study we have reconstructed a genome-scale metabolic model of *T. gondii*, ToxoNet1, aiming to address the abovementioned questions in a systematic way and to the extent possible with the currently available knowledge. Using flux-balance analysis we have identified genes, reactions and pairs of genes for which deletion renders production of biomass components impossible. Furthermore, we have assessed which precursors are necessary for each of the biomass components and have defined minimal sets of such precursors that allow *in silico* simulation of growth. To illustrate the applicability of ToxoNet1 in filling knowledge gaps regarding parasite metabolism, we experimentally challenged the model prediction regarding the alternative routes for generation of cytosolic acetyl-CoA. We have confirmed *in vitro* the functional redundancy and synthetic lethality within the two biosynthetic routes that involve the ATP-citrate lyase and acetyl-CoA synthase enzymes.

Results

Here we present the full genome-scale *in silico* reconstruction of metabolism in *T. gondii* with manually refined gene-reaction associations—ToxoNet1. The model reconstruction process required completion of the following major steps (schematically shown in Fig 1): (1) reconstruction of the draft metabolic network; (2) compartmentalization of the intracellular space; (3) verification of the metabolic capabilities and manual literature-based corrections; (4) representation of the plausible exchanges of metabolites between the infected host cell and the parasite. In general terms, the reconstruction process was consistent with the workflow that has been previously used for semi-automated reconstruction of a genome-scale metabolic model for *Penicillium chrysogenum* by means of the RAVEN Toolbox [11]. All the necessary manual corrections were made in accordance to the conventional model reconstruction protocol [12].

For the compartmentalization process we consulted the models of *Plasmodium falciparum* [9,13] and used sequence-based predictors of subcellular localization [14–16] as well as the ApiLoc (http://apiloc.biochem.unimelb.edu.au/apiloc/apiloc) database. The databases used for identification of "gap" reactions were KEGG [17] and LLAMP [18]. Recon 2 [19] was used as a model of host cell metabolism to define the list of putatively host-supplied substrates.

Reconstruction of the ToxoNet1

Reconstruction of the metabolic network was achieved by combination of the state-of-the-art algorithm for a semi-automated generation of metabolic networks [11] and a comprehensive manual curation based on the relevant primary literature. Details of the reconstruction process are provided in the materials and methods section and the most important steps are discussed below.

The reconstructed metabolic network accounted for 527 open-reading frames (ORFs) that were linked to 867 unique metabolic reactions present in the KEGG [17] database. Each functional annotation in the model was assigned with two estimates (namely a bit-score and e-value), which indicated confidence of association for a given ORF with a corresponding enzymatic function.

In ToxoNet1, the majority of reactions were associated with genes that encode corresponding metabolic enzymes (Fig 2A). Intracellular metabolic reactions not associated with any genes comprised only 9.6% of all the metabolic reactions present in the model (6.7% of all the reactions in the model); they include spontaneous reactions (not enzyme-catalysed) and so-called gap-filling reactions that were added for correct functioning of the model. Most of the

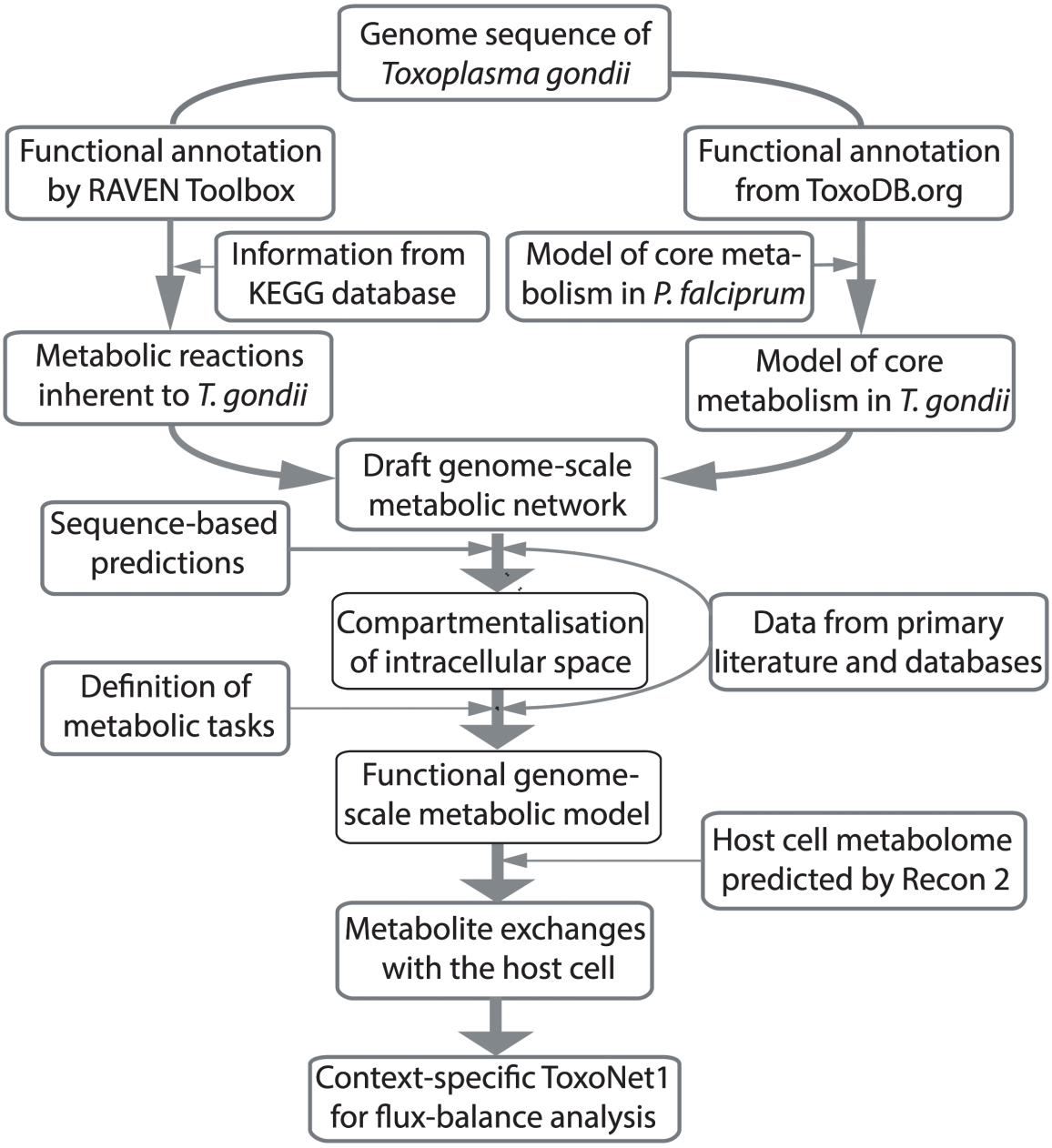


Fig 1. Reconstruction workflow for ToxoNet1 summarized as a flowchart illustration.

doi:10.1371/journal.pcbi.1004261.g001

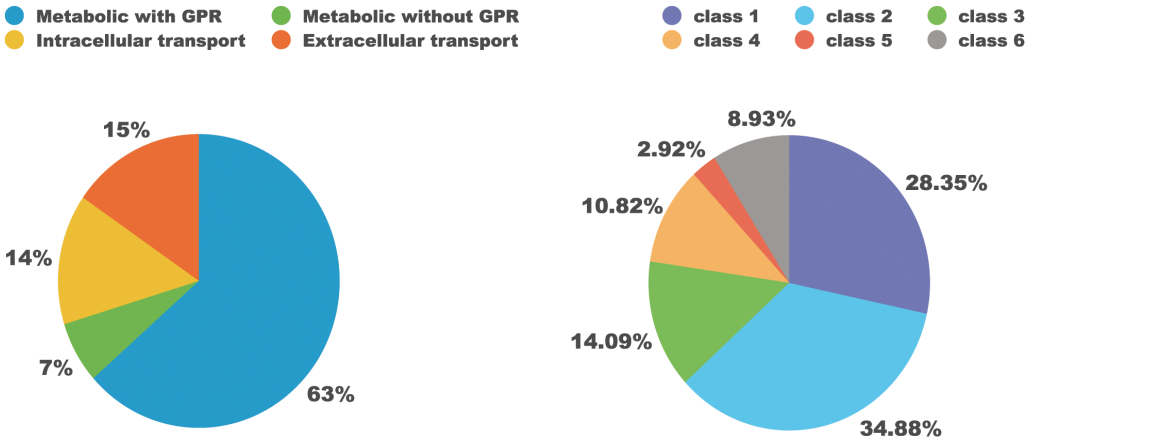


Fig 2. Breakdown of the metabolic network. A) By presence of gene-reactions association; B) By enzyme classes encoded in the genome of *T. gondii* (EC nomenclature, class 1: oxidoreductases, class 2: transferases, class 3: hydrolases, class 4: lyases, class 5: isomerases, class 6: ligases).

doi:10.1371/journal.pcbi.1004261.g002

metabolite transport reactions, which were added to connect pathways segregated between subcellular compartments, also lacked known gene associations.

Among the metabolic enzymes encoded in the genome of *T. gondii* and present within ToxoNet1, the transferases (class 2) and oxidoreductases (class 1) were the most numerous (over 60% of all the enzymes). Hydrolases (class 3), lyases (class 4) and ligases were significantly less represented (Fig 2B). The least frequent were isomerases (class 5), which is in accordance with the models of other pathogens, such as *Leishmania major* [20] and *Plasmodium falciparum* [8].

Compartmentalization of the ToxoNet1

Relatively straightforward experimental methods exist to determine the localization of proteins in *T. gondii* [21], however, it would be laborious and expensive to apply such methods for hundreds of enzymes. According to the ApiLoc database (http://apiloc.biochem.unimelb.edu.au/apiloc/apiloc), experimental data was available only for a limited number of proteins: 60 out of the 527 included in ToxoNet1. Thus the only reasonable option for building a compartmentalized metabolic model was to make use of sequence-based localization predictors. To define putative localizations of the enzymes with no experimental evidence we generated sequence-based predictions using three software algorithms: TargetP [16], MitoProt II [15], and ApicoAP [14]. We then manually reconciled output data of the three independent predicting algorithms and assigned the localization based on the manually determined consensus prediction, also considering recent primary literature whenever it was available (S1 Table).

Comparison of these predictions with 60 experimentally established subcellular localizations available from ApiLoc uncovered two issues: (i) not all the computational predictions matched their experimental data (highlighted in the S1 Table); (ii) some of the enzymes were reported to be present in two and, in one case (glutathione/thioredoxin peroxidase [22]), three different subcellular compartments defined in ToxoNet1. In these cases the sequence-based predictors were not efficient and suggested only one of the compartments. Therefore, we assigned compartments in a supervised manner considering all available evidence.

The information on compartmentalisation organised according to the global metabolic subsystems of ToxoNet1 is shown in the Table 1.

Table 1. Breakdown of ToxoNet1 by metabolic subsystems and subcellular compartments.

Metabolicsubsystems	Genes	Reactions	Reactions by compartments (non-unique)		
		(unique)	cytosol	mitochondrion	apicoplast
Carbohydrates	88	114	101	32	16
Amino acids	105	163	140	56	8
Nucleic acids	74	105	93	8	19
Fatty acids	50	99	59	41	45
Vitamins, Co-factors	78	114	111	29	22
Phospholipids	34	42	44	3	2
Miscellaneous	98	230	194	38	28
Total	527	867	742	207	140

doi:10.1371/journal.pcbi.1004261.t001

In ToxoNet1 a majority of the reactions (c.a. 69%) occur in the cytosolic compartment. We also considered this compartment as a default for enzymes with a subcellular localization remaining unclear from the *in silico* predictions. The mitochondrion accommodates 19% of all metabolic reactions of the model. The remaining 12% are localized to the apicoplast, a plastid-like non-photosynthetic organelle [23]. We connected these two compartments to the cytosol by 223 transport reactions. This enabled the corresponding number of metabolites to be transported across these boundaries, which delineate the organellar membranes (for further details see the materials and methods section). We allowed the metabolites present in the mitochondrion or apicoplast to be reversibly transported to and from the cytosol if they satisfied the following requirements: (i) metabolite has to be present (i.e. participate in at least one reaction) in both the cytosol and corresponding non-cytosolic compartment; (ii) it should be neither phosphorylated nor containing an acyl-carrier-protein or CoA moiety, unless supporting evidence of such transport is available. We also allowed: (i) import of phosphoenolpyruvate and dihydroxyacetone phosphate from the cytosol to the apicoplast [24]; (ii) export of ATP from the mitochondrion to the cytosol (TGME49_249900); (iii) export of isopentenyl pyrophosphate (IPP) and dimethylallyl pyrophosphate (DMAPP) from the apicoplast to the cytosol. The latter allowed us to observe an unusual feature of the isoprenoid biosynthesis pathway in *T. gondii* regarding the isopentenyl pyrophosphate isomerase (EC 5.3.3.2). This enzyme interconverts IPP into its more reactive isomer—DMAPP. In *T. gondii*, as well as in *Plasmodium spp*. this enzyme is absent [25]. Thus, both IPP and DMAPP have to be produced and transported from the apicoplast as separate entities for synthesis of long isoprenoid chains in the cytosol.

Verification of the metabolic capabilities of ToxoNet1

A major limitation of all algorithms for automatic reconstruction of metabolic models is the number of missing reactions (so-called gaps) that disrupt metabolic pathways. The RAVEN Toolbox offers a functionality called metabolic tasks (*checkTasks and fitTask* functions), which verifies that specific tasks are fulfilled and, if necessary, it fills the gaps in the metabolic pathways to meet the required functionality.

We created metabolic tasks for the synthesis of every metabolite included in the biomass reaction (complete formulation of all the tasks is provided in the S2 Table). The outcomes of the tasks were categorized in three groups shown in Table 2. The first group contains the metabolites that could be produced from glucose and inorganic compounds (i.e. *de novo* synthesis was possible in the model). The second group contains the metabolites whose synthesis required a precursor with a specific moiety (e.g. hypoxanthine or other source of purine moiety for purine nucleotides and their derivatives). The third group are the biomass constituents that could not

Table 2. Nutritional requirements of *T. gondii* predicted *in silico*.

	Produced *de novo* (from glucose and inorganic substrates)	Production requires specific precursors (secondary production)	No production from any of the available precursors
Amino acids	Alanine; Asparagine; Aspartate; Cysteine; Glutamate; Glutamine; Proline; Serine; Threonine	Tyrosine[1]; Glycine[2]; Lysine[3]; Valine[4]; Leucine[5]; Isoleucine[6]; Methionine[7,8]; Phenylalanine[8]	Arginine; Histidine; Tryptophan
Nucleotides	dCTP; dTTP; CTP; UTP	dATP[9]; dGTP[9]; ATP[9]; GTP[9]	
Fatty acids	Dodecanoate; Tetradecanoate; Octadecanoate; Hexadecanoate; Octadecenoate		Linoleate; Arachidonate
Cofactors	Pyridoxal phosphate; Protein-N[6]-(lipoyl)lysine[apicoplast][15]	Glutathione[2, 9]; CoA[9,10]; FMN[11]; NAD[9,12] NADP[9, 12]; Thiamin diphosphate[13]; S-Adenosyl-L-methionine[2, 7,9]; Protein N6-(lipoyl)lysine [mitochondrial][16]; Tetrahydrofolate[2]; Tetrahydrobiopterin[14]	
Other membrane precursors	UDP-N-acetyl D-glucosamine; Geranylgeranyl-PP; Undecaprenyl-PP; PhosphatidylethanolaminePhosphatidylserine	Phosphatidylcholine[17]; Sphingomyelin[17]; ADP-glucose[3] GDP-mannose[2]; 1-Phosphatidyl-D-myo-inositol[18]	Cholesterol

Required precursors:
[1]—phenylalanine or phenylpyruvate;
[2]—folic acid or its derivatives;
[3]—l-2-aminoadipate 6-semialdehyde;
[4]–3-methyl-2-oxobutanoic acid;
[5]–4-methyl-2-oxopentanoate;
[6] –(s)-3-methyl-2-oxopentanoic acid;
[7]—l-homocysteine;
[8]—phenylpyruvate;
[9]—a source of purine moiety: hypoxanthine, adenine, adenosine or similar;
[10]—pantothenate or valine+β-alanine;
[11]—riboflavin;
[12]—nicotinic acid or its derivatives;
[13]—thiamine;
[14]–4-hydroxybenzoate;
[15]—an undefined"sulfur donor" molecule;
[16]—lipoic acid;
[17]—choline;
[18]—myo-inositol.

doi:10.1371/journal.pcbi.1004261.t002

be produced even with the maximum set of host-supplied substrates, and therefore they were directly taken up from the host (e.g. choline, arachidonic acid, cholesterol).

According to ToxoNet1, *T. gondii* has the capability to produce *de novo* a number of biomass precursors such as pyrimidine nucleotides and their derivatives, fatty acids, isoprenoids and about half of the proteinogenic amino acids (Table 2). When tasks of *de novo* production for certain biomass constituents failed we attempted to produce them by supplying additional precursors. In some cases uptake of one or more molecule(s) from the list of host-supplied molecules, enabled production of the necessary biomass precursors in ToxoNet1. Nevertheless, several biomass building blocks could not be produced in the model even when all the 230 host-supplied substrates were provided simultaneously, indicating that the parasite is likely auxotrophic for these molecules.

In the cases when metabolic tasks could not be accomplished despite literature evidence (mainly acquired from the LLAMP database [18]), we performed gap-filling. This is a conventional part of building genome-scale metabolic models [12] and it implies inclusion of a

minimal number of reactions to complete pathways of interest. We used both the automated gap-filling function of the RAVEN Toolbox (*fillGaps*) and manual gap-filling based on the information from the LLAMP database [18]. For instance, the threonine biosynthesis pathway in ToxoNet1 was nearly complete but lacked a single enzyme, the homoserine dehydrogenase (E. C. 1.1.1.3). The corresponding reaction (R01773) was included in the model without a gene assigned (i.e. as a "gap-fill") justified by the presence of four other enzymes in the pathway and the fact that the immediate downstream enzyme was identified in a proteomics study [26]. ToxoNet1 also retrieved a known issue of the lysine biosynthesis pathway: the bacterial-type pathway consisting of 9 enzymatic steps lacks enzymes annotated for 4 sequential steps, making the presence of the functional pathway debatable. While this issue remains unresolved, we decided not to gap-fill this pathway in the model, thus, leaving lysine an essential amino acid. The full list of gap-fill reactions included in ToxoNet1 is provided in S3 Table.

Importantly, some of the information necessary for a metabolic model could not be easily deduced from the genome sequences alone. This is why we merged the draft output metabolic network from the RAVEN Toolbox with a small-scale model of central carbon metabolism in *T. gondii*. We manually built such a model prior to this study based on the genome annotation from ToxoDB and refined it according to the relevant primary literature. This model contained a number of recent experimental findings as well as manually assembled complex gene-reaction associations. Among these were the pyruvate dehydrogenase activity that can be carried out by the branched-chain keto-acid dehydrogenase complex (BCKDH) in the mitochondrion [27], mitochondrial pyruvate transporter [28–30] and the GABA-shunt connected to the tricarboxylic acid (TCA) cycle [3].

Substrate dispensability based on ToxoNet1

One of the major challenges in simulation of parasite metabolism *in silico* is the unknown range of metabolites available for the parasite within the host cell. As extracellular replication (axenic cultivation) of *T. gondii* is not possible, it remains undefined which of the substrates are dispensable for the parasite and which are not.

To address this issue *in silico*, we developed a method to enumerate all of the smallest/ minimal metabolite sets (further referred to as *in silico* minimal medium or IMM), which could enable simulation of growth in ToxoNet1. In brief, the algorithm applied iterative rounds of biomass production using the fewest number of substrates possible. After each of the iterations we added a constraint to ensure that the next IMM included at least one substrate uncommon from the preceding ones (further details on the formulation of the algorithm are described in the materials and methods section).

The results of these simulations showed that as few as 19 substrates were sufficient for simulation of *T. gondii* replication in ToxoNet1. We also observed a relatively large number (2592) of alternative sets of 19 substrates with 10 metabolites being constitutively present in all the IMMs. The other 9 substrates could be substituted by at least one other host-supplied metabolite (Table 3). Despite the very large theoretical number of combinations (in the case of 9 variable substrates picked from 220) we observed only 2592 alternative IMM of 19 substrates. This indicates the ability of ToxoNet1 to substitute one substrate with another from the set of 230 is rather limited. Indeed, the majority of the 2592 IMMs arise from flexibility in the carbon source (one out of 7 available) and a source of nicotinate moiety (one out of 3). The other metabolites are either non-substitutable (10) or substitutable with one single alternative. However, in this particular analysis, we excluded the possibility of substituting one metabolite with simultaneous uptakes of several others, as it would lead to more than the minimal number of substrates used.

Table 3. Composition of the minimal *in silico* media that allows ToxoNet1 to simulate replication of *T. gondii*.

Indispensable (biomass precursor for which it is necessary)	
5,10-methyltetrahydrofolate (THF); cholesterol (itself); choline (phosphatidylcholine); L-arginine (proteins); L-histidine (proteins); L-tryptophan (proteins); lipoate (lipoylation in mitochondrion); riboflavin (FAD); S-adenosyl-L-homocysteine (purine nucleotides, methionine, SAM); thiamine (TPP)	
Dispensable (multiple alternative precursors present in different minimal media sets)	
C6 or C5 carbon source	D-fructose or D-glucosamine or D-glucose or D-mannose or D-ribose or D-sorbitol or 2-Deoxy-D-ribose
NAD/NADP precursors	nicotinate or nicotinate-D-ribonucleoside or nicotinamide
Amino acids or precursors	L-lysine or L-2-aminoadipate 6-semialdehyde
	L-isoleucine or (S)-3-methyl-2-oxopentanoate
	L-valine or 3-methyl-2-oxobutanoate
	L-leucine or 4-methyl-2-oxo-pentanoate
	L-phenylalanine or phenylpyruvate
Source of inorganic iron	Fe^{2+} or heme
Source of inorganic phosphate	orthophosphate or diphosphate

Abbreviations: FAD—flavine-adenine dinucleotide, NAD—nicoticamide-adenine dinucleotide, NADP—NAD phosphate, TPP—thiamine pyrophosphate, SAM—S-adenosyl-methionine, THF—tetrahydrofolate.

doi:10.1371/journal.pcbi.1004261.t003

Gene essentiality predicted by ToxoNet1

We simulated *in silico* the outcomes of systematic removal of genes and reactions in ToxoNet1 to explore which of them represent indispensable metabolic functions. With the full set of 230 host-supplied substrates out of the 527 genes in ToxoNet1, 53 genes were predicted to be essential (Table 4 and literature evidence [31–38]). Considering that most of the transport reactions do not have known gene-reaction associations, we also simulated single reaction deletions. This allowed us to assess the dispensability of metabolite transport across the compartments, as well as metabolite exchanges between *T. gondii* and its host cell (Table 4 and S5 Table).

In addition to single gene and reaction essentiality, we also simulated double gene deletions to reveal the pairs in which genes are not essential for replication on their own, yet deleterious when disrupted together. A total of 20 pairs that caused such synergistic effect (synthetic lethality) are listed in S5 Table.

Gene essentiality predictions depend on the following important aspects of metabolism as represented in the model: (1) range of substrates that can enter the model (i.e. molecules that the parasite can take up from the infected host cell); (2) composition of the parasite cell represented as a set of biomass precursor molecules (the concept of biomass objective functions is explained in [39]); (3) the presence of alternative metabolic routes to produce biomass building blocks from the different substrates. Assuming a very permissive range of 230 substrates as potentially accessible for the parasite we predicted a minimal set of essential genes and reactions. In consequence, the number of genes predicted by ToxoNet1 as non-essential is likely to be overestimated. Yet with this assumption the probability of incorrect prediction of genes to be essential is lower.

ToxoNet1 as a framework for *in silico* assessment of experimental research questions

In order to evaluate whether ToxoNet1 predictions closely reflect the metabolic capabilities of *T. gondii* observed experimentally in tissue culture, we chose to assess the importance of the

Table 4. Gene essentiality predictions of ToxoNet 1 and available literature evidence.

Gene identifiers	Metabolic subsystem	E.C. identifier	Essentiality of ortholog(s) in *P. falciparum**	Essentiality evidence
TGME49_264780	Amino sugar and nucleotide sugar metabolism	2.7.7.23	no	NA
TGME49_264650		5.4.2.3		NA
TGME49_309730	Antioxidative metabolism	1.8.1.7	yes	NA
TGME49_298990	Biosynthesis of terpenoids	1.18.1.2	no	Inhibition of the pathway [31],[32]
TGME49_316770		2.5.1.-		NA
TGME49_269430		2.5.1.1	yes	Inhibition of the pathway [31],[32]
TGME49_227420		1.17.1.2		
TGME49_306260		2.7.7.60		
TGME49_255690		4.6.1.12		
TGME49_306550		2.7.1.148		
TGME49_262430		1.17.7.1		
TGME49_214850		1.1.1.267		
TGME49_208820		2.2.1.7		
TGME49_224490		2.5.1.1 2.5.1.10		
TGME49_290600	Citrate cycle(TCA cycle)	6.2.1.5	no	Growth reducing knockout [33]
TGME49_278910	Cysteine and methionine metabolism	2.5.1.47	no orthologs	NA
TGME49_225990	Fatty acid biosynthesis (apicoplast)	2.3.1.39	no (not essential on blood- and essential in liver stage)	Conditional knockout of apicoplast-targeted acyl-carrier protein [34]
TGME49_217740		1.1.1.100		
TGME49_206610		2.3.1.12		
TGME49_231890		2.3.1.180	yes	
TGME49_251930		1.3.1.9		
TGME49_321570		4.2.1.-		
TGME49_293590		2.3.1.179		
TGME49_221320		6.4.1.2 6.3.4.14		
TGME49_239710	Fructose and mannose metabolism	5.4.2.8	no	NA
TGME49_239250	Glycerolipid metabolism	2.7.1.107	no orthologs	NA
TGME49_310280	Glycerophospholipid	2.7.7.14	no	NA
TGME49_261480	metabolism	2.7.8.-		Growth arrest upon inhibition [35]
TGME49_212130		3.1.1.3 2.3.1.-	no orthologs	NA
TGME49_216930		2.7.7.15	yes	Growth reducing knockout res-cued by salvage from host [36]
TGME49_281980		2.7.7.41		NA
TGME49_233500	Glycolysis / Gluconeogenesis	5.3.1.1	yes	NA
TGME49_207710	Inositol phosphate metabolism	2.7.8.11	yes	NA
TGME49_315640	Lipoic acid biosynthesis	2.3.1.181	no	Conditional knockout of apicoplast-targeted acyl-carrier protein [34]
TGME49_226400		2.8.1.8		
TGME49_271820	Lipoic acid metabolism in mitochondrion	2.7.7.63	no	NA
TGME49_269800	Nicotinate and nicotinamide metabolism	6.3.5.1	yes	NA
TGME49_244700		2.7.1.23		
TGME49_224900	Purine metabolism	2.7.4.3	no	NA

(*Continued*)

Table 4. (*Continued*)

Gene identifiers	Metabolic subsystem	E.C. identifier	Essentiality of ortholog(s) in *P. falciparum**	Essentiality evidence
TGME49_233110		1.1.1.205	yes	NA
TGME49_242730		2.7.4.8		
TGME49_230450		6.3.5.2		
TGME49_257740	Pyrimidine metabolism	2.7.4.- 2.7.4.14	no	NA
TGME49_249180		1.5.1.3 2.1.1.45	yes	[37]
TGME49_306970		2.7.4.9		NA
TGME49_299210		6.3.4.2		NA
TGME49_305980	Pyruvate metabolism in apicoplast	1.8.1.4	no**	NA
TGME49_299070		2.7.1.40	yes	NA
TGME49_216740	Riboflavin metabolism	2.7.1.26	yes	NA
TGME49_214280		2.7.7.2		
TGME49_237200	Sphingolipid metabolism	1.14.-.-	no orthologs	NA
TGME49_316450		2.3.1.24	yes	
TGME49_215250	Thiamine metabolism	2.7.6.2	yes	NA

(NA denotes the cases when neither supporting nor contradicting literature reference could be found). Gene essentiality in *P. falciparum* is based on the Supplementary Table 1 of the review manuscript [38].

doi:10.1371/journal.pcbi.1004261.t004

two independent routes of cytosolic acetyl-CoA production in the rapidly dividing tachyzoite stage (Fig 3A). Acetyl-CoA is an important molecule in central carbon metabolism that is involved in many biochemical processes such as fatty acid synthesis (FAS type I pathway), fatty acid chain elongation and acetylation of proteins, in particular histones. ToxoNet1 identified two enzymes that can produce acetyl-CoA in the cytosol: (1) from acetate through the acetyl-CoA synthase (ACS) reaction (TGME49_266640), (2) from TCA-derived citrate by ATP-citrate lyase (ACL) (TGME49_223840), as shown in Fig 3A. Acetoacetate-CoA ligase (TGME49_219230) produces acetoacetyl-CoA that can be converted to acetyl-CoA by the acetyl-CoA acetyltransferase (TGME49_301120), which belongs to the fatty acid degradation pathway in mitochondrion. We thus concluded that this is unlikely to be producing cytosolic acetyl-CoA production due to the predicted mitochondrial localization of the two enzymes. In further support of the importance of cytosolic acetyl-CoA production is the presence of a putative ortholog of the human acetyl-CoA transporter (AT1) in the *T. gondii* genome [40]. This transporter was shown to localize at the endoplasmic reticulum (ER) membrane and to be essential for the survival of eukaryotic cells by allowing import of acetyl-CoA from the cytosol to the ER to acetylate proteins within this compartment. ToxoNet1 predicted ACS- and ACL-encoding genes to be fully dispensable when knocked out individually, however, their simultaneous knockout was predicted to be lethal (S5 Table).

Experimental investigation of acetyl-CoA biosynthesis in the cytosol and its essentiality

To determine the localization and level of expression of ACS and ACL in *T. gondii*, we modified the endogenous locus (knock-in) by introducing a 3xTy-epitope tag at the C-terminal end of both genes in the RH*ku80ko* (Ku80ko) background strain, which limits random integration in the genome hence facilitating recovery of homologous recombination events (Fig 3B). ACS

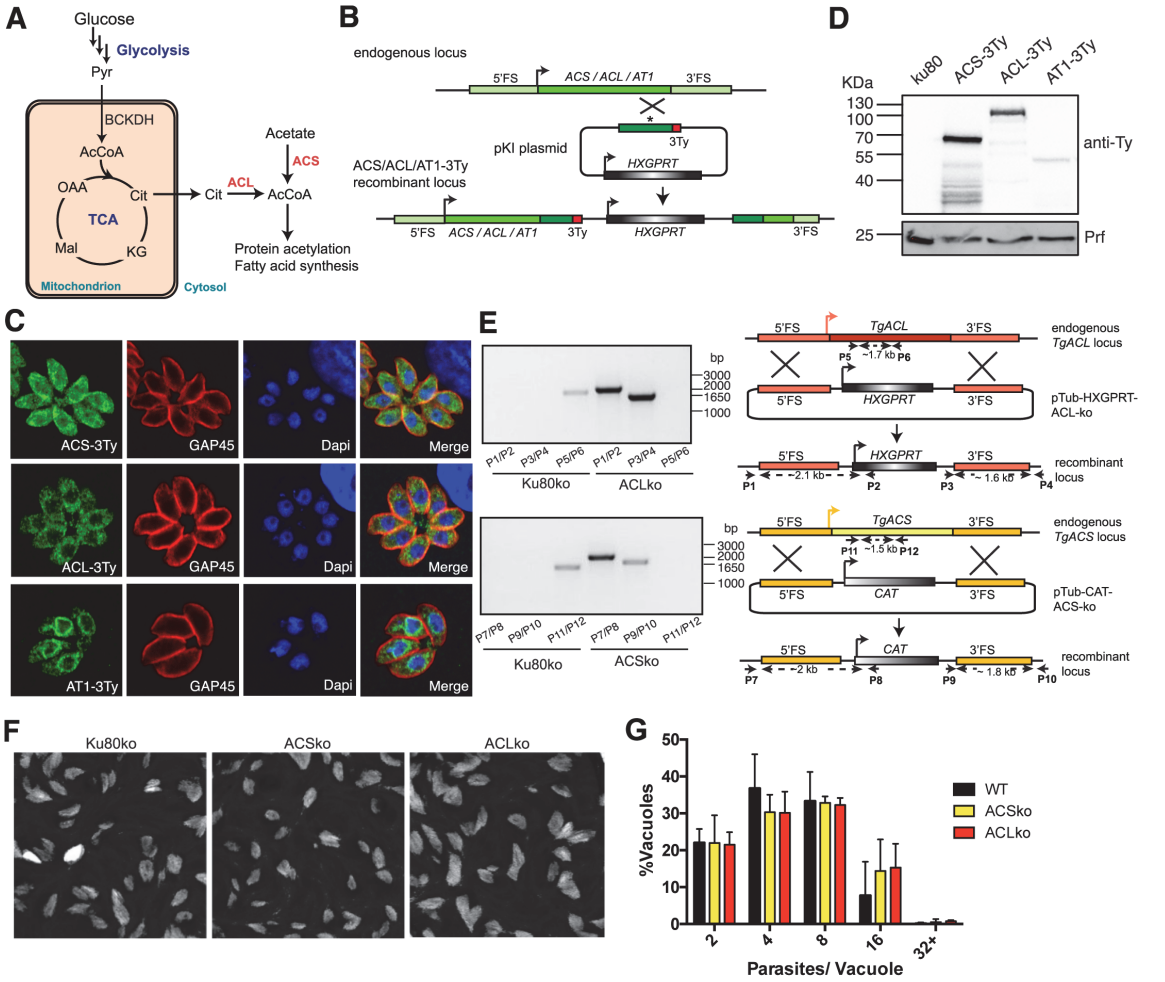


Fig 3. Both ACS and ACL are dispensable in the tachyzoite stage of *T. gondii*. (A) Schematic representation of the two pathways to produce acetyl-CoA in the cytosol of *T. gondii*. Abbreviations: AcCoA, acetyl-CoA; α-KG, α-ketoglutarate; Cit, citrate; Glc, glucose; Lac, lactate; Mal, malate; OAA, oxaloacetic acid; Pyr, pyruvate; Suc, succinate. Enzymes in red: ACL, ATP-citrate lyase; ACS, Acetyl-CoA synthetase. (B) Scheme of the knock-in strategy used to introduce a 3Ty-tag in the endogenous loci of ACS, ACL and AT1. (C) Localization of endogenous ACS, ACL and AT1 C-terminally Ty-tagged (ACS-3Ty, ACL-3Ty and AT1-3Ty) in the cytoplasm, cytosol and endoplasmic reticulum respectively of intracellular parasites using anti-Ty as well as anti-GAP45 that stains the periphery and DAPI which stains the nucleus of the parasite. (D) Immuno-blot of total lysates from Ku80ko parasites expressing the C-terminally Ty-tagged endogenous ACS, ACL and AT1 proteins by Western blot using anti-Ty antibodies. Anti-Profilin (Prf) represents a loading control. (E) Schematic representation of the direct knockout strategy by double homologous recombination where ACS was replaced by the chloramphenicol resistance cassette and *ACL* by the HXGPRT selection cassette. The position of the primers used to confirm the integration and the length of the PCR products are indicated. PCRs performed on genomic DNA extracted from Ku80ko, ACSko and ACLko strains to confirm the integration of the selection cassette and loss of the corresponding gene locus. The sequences of the primers can be found in S7 Table. (F) Plaque assays performed with Ku80ko, ACSko and ACLko parasite lines fixed after 7 days. No significant defect in the lytic cycle could be observed. (G) Intracellular growth assay performed on Ku80ko, ACSko and ACLko strains by determining the number of parasites per vacuole 24h post infection. Data are represented as mean ± SD from 3 biological replicates.

doi:10.1371/journal.pcbi.1004261.g003

is clearly cytosolic and nuclear whereas ACL appears to localize mostly to the cytosol, whilst a fainter nuclear staining can also be detected by indirect immunofluorescence assay (IFA) (Fig 3C). Localization of the acetyl-CoA transporter AT1 in the perinuclear region of the ER in *T. gondii* was validated using the same knock-in tagging strategy (Fig 3B and 3C). Expression of the epitope-tagged proteins was further validated by Western blot analyses, where ACS runs at the expected molecular weight of ~80 kDa, ACL at ~140 kDa and AT1 at ~65 KDa (Fig 3D). Comparative signal intensity observed on Western blots suggests that ACS is significantly more abundant than ACL and AT1.

To functionally assess the importance of both routes to produce cytosolic acetyl-CoA and challenge the predictions made by ToxoNet1, individual deletion of the genes encoding ACL and ACS were achieved using a double homologous recombination strategy (*ACLko* and *ACSko* respectively) in Ku80ko (Fig 3E). For both genes, transgenic parasites were readily obtained and cloned. Absence of the ACL and ACS ORFs and their replacement by a selection cassette in individual clones was validated by genomic PCR (Fig 3E), thus, supporting the prediction made by the model that these genes are both dispensable for the survival of *T. gondii* tachyzoites. No significant defect could be observed in the overall lytic cycle of these knockout mutants as represented by plaque assays performed in human foreskin fibroblast (HFF) monolayers. Indeed, both ACLko and ACSko parasites formed lysis plaques of similar sizes when compared to wild type (Ku80ko) parasites (Fig 3F). Moreover, after 24h of intracellular growth, most vacuoles of the ACLko strain contained 4 to 8 parasites, which is comparable to the number of parasites per vacuole for cells infected with Ku80ko or ACSko (Fig 3G). Finally, growth competition assays between mutants and wt parasites, which would detect mild loss of fitness, showed no significant defect either.

In order to assess whether a double knockout of ACS and ACL is lethal for *T. gondii* as suggested by ToxoNet1, we first attempted to disrupt ACL by single homologous recombination in the middle of the ORF in the ACSko parasite background (same strategy as presented in Fig 3B but leading to a truncation of the protein and removal of the catalytic site). While we were able to interrupt the ACL gene in Ku80ko parasites, we failed to generate such a mutant in the ACSko background. This result strongly suggested that ACS and ACL together are critical for the biosynthesis of cytosolic acetyl-CoA. To further confirm the synthetic lethality between ACS and ACL we generated a conditional knockdown of ACL in Ku80ko and in ACSko by U1 snRNP-mediated gene silencing with Cre-recombinase dependent positioning of U1, as has been recently developed in *T. gondii* [41] (generating ACL-lox and ACSko/ACL-lox respectively; Fig 4A). Following Cre-mediated recombination, the endogenous 3'untranslated region is excised and a U1 recognition site is placed adjacent to the termination codon. Consequently the ACL pre-mRNAs are cleaved at the 3'-end and degraded, leading to a highly efficient knockdown of the gene. Loss at the protein level can be assessed by immuno-detection of the C-terminal 3Ty-tag (Fig 4A). Correct integration of the construct was confirmed by genomic PCR (Fig 4B). Importantly, upon deletion of ACS the level of endogenous ACL protein was significantly increased compared to ACL-lox (Fig 4C). This change in level of ACL in the absence of ACS was reproducibly confirmed by generation of a second independent transgenic parasite line where the *ACS* gene was disrupted in the ACL-lox background strain (Fig 4D).

To conditionally disrupt ACL in the ACSko, the ACL-lox and ACSko/ACL-lox parasites were transfected with a plasmid transiently expressing the Cre recombinase. While ACL-lox excised parasites could be readily propagated in culture the ACSko/ACL-lox excised parasites were lost after the first passage as monitored by genomic PCR analyses of the two excised parasite populations. While genomic recombination of excised parasites could be readily detected in the original transfected parasites (P0), the signal was lost immediately after the first passage of these cultures (P1 and P2) (Fig 5A) indicating a rapid deleterious effect. Thirty hours post

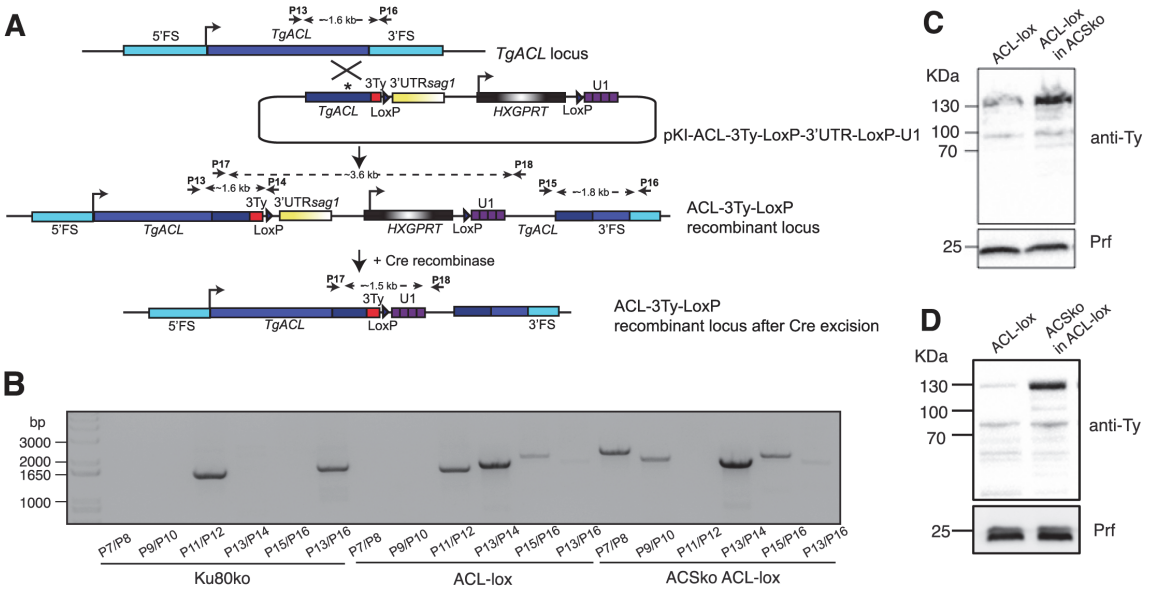


Fig 4. Generation of an inducible ACL knockdown in ACSko parasites. (A) Schematic representation of the U1 snRNP-mediated ACL gene silencing with Cre-recombinase dependent positioning of U1 in Ku80ko wildtype and ACSko parasites. (B) PCRs performed on genomic DNA extracted from Ku80ko, ACL-lox, ACSko/ACL-lox validating integration of the pKI-ACL-3TyLox3'UTRLoxU1 construct to knock down ACL in the different strains. The sequences of the primers can be found in (S7C and S7D) and S7 Table. Immuno-blot of total lysates from ACL-lox, ACSko/ACL-lox where ACL-lox was integrated in the ACSko strain or ACS was knocked out in ACL-lox. Both independent lines show increased levels of ACL when ACS is absent. Western blot was performed using anti-Ty antibodies. Anti-TgProfilin (Prf) represents a loading control.

doi:10.1371/journal.pcbi.1004261.g004

transfection of Cre recombinase, the loss of ACL-3Ty-tagged protein was evident by IFA in about 50% of the ACL-lox and ACSko/ACL-lox vacuoles. While ACL-lox parasites lacking ACL could be propagated in culture, parasites lacking ACL in the ACSko/ACL-lox strain were immediately lost in the first culture passage (Fig 5B). While ACL-lox excised parasites appeared normal, most vacuoles from excised ACSko/ACL-lox parasites exhibited a severe morphological defect and impairment in the parasite division process with a loss of pellicle integrity as seen by perturbation of GAP45 staining (Fig 5C). Furthermore these parasites appear to continue dividing their nuclei, mitochondrion and apicoplast, but fail to form daughter cells, resulting in organelle accumulation in the vacuolar space (Fig 5C).

To determine whether the deleterious phenotype and loss of pellicle integrity could be attributed to a block in type I fatty acid synthesis following depletion in cytosolic acetyl-CoA, the gene coding for *T. gondii* FASI was disrupted in wild type RH strain parasites by CRISPR/Cas9 mediated genome editing [42]. Trangenic parasites were obtained following double-stranded breaks generated by Cas9 at a position downstream of the FASI ATG. Two independent clones were sequenced to confirm the introduction of frame shift mutations (S1A Fig) No defect in the lytic cycle could be observed in these parasites (S1B Fig), which is in accordance with the ToxoNet1 prediction that FASI activity is not essential for parasite survival. Taken together, these data firmly support the predictions made by the model regarding the individual dispensability of ACS and ACL and the synthetic lethality of ACL in the absence of ACS.

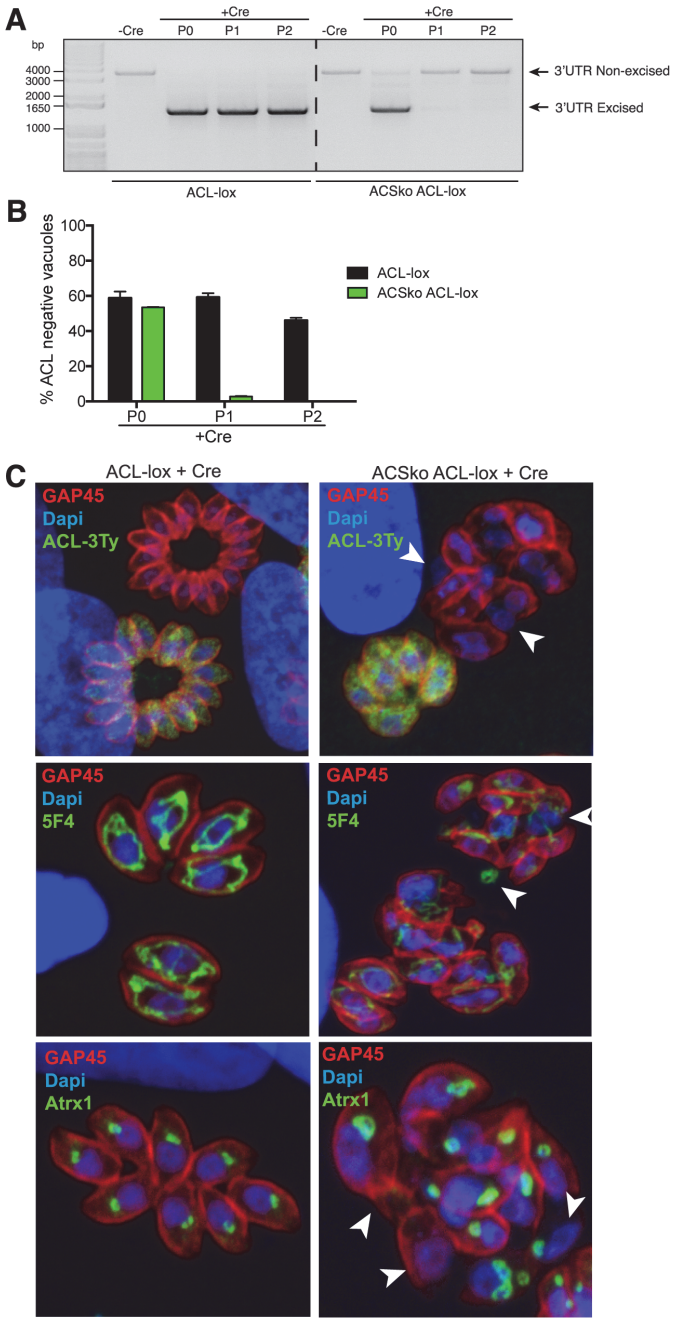


Fig 5. ACS and ACL are dually essential. (A) Following transfection of ACL-lox and ACSko/ACL-lox with Cre recombinase, excision and repositioning of U1 is followed by genomic PCRs over several culture

passages (every 48h) using primers P17/P18 as depicted in Fig 4A (P0, extracellular parasites 48h after Cre transfection; P1, extracellular parasites passaged once ~96h post transfection; P2, extracellular parasites passaged twice ~140h post transfection). (B) Histogram showing percentage of excised parasites over several passages (every 48h) in ACL-lox and ACSko/ACL-lox populations following transfection with Cre recombinase. Excised parasites were visualized by IFA looking for loss of ACL-3Ty signal. Due to fluctuation in transfection efficiency, data from one biological replicate is shown and represents mean ± SD from 3 technical replicates. 3 biological replicates were done and gave the same results. (P0, intracellular parasites 30h after Cre transfection; P1, intracellular parasites passaged once ~72h post transfection; P2, intracellular parasites passaged twice ~100h post transfection). (C) Immunofluorescence assay confirms the loss of ACL-3Ty and parasite pellicle integrity in a subset of vacuoles 30h following transfection of ACL-lox and ACSko/ACL-lox strains with a Cre recombinase expressing plasmid. IFAs were stained using anti-Ty, anti-GAP45 (Pellicle), anti-5F4 (mitochondrion) or anti-Atrx1 (apicoplast) antibodies and DAPI (nucleus). Arrowheads highlight nuclear and mitochondrial material lost in the vacuolar space and loss of the apicoplast due to loss of pellicle integrity.

doi:10.1371/journal.pcbi.1004261.g005

Discussion

ToxoNet1 constitutes a full genome-scale reconstruction of metabolism within *T. gondii*, which can simulate growth of the parasite *in silico* and infer essentiality of its genes. It considerably extends the scope of previous work [10] and contributes to better understanding of the limits in the metabolic capabilities of this opportunistic human and animal pathogen. While the modelling efforts of Song et al were aiming to reveal strain-type specific differences in metabolism of *T. gondii*, in the present study we reconstructed an independent model that represents the potential scope of the metabolic capabilities of the parasite independently of the life-stage and strain-type. Aiming at the most reliable gene essentiality predictions we used many alternative assumptions on the range of accessible substrates and transportability of metabolites. We have also achieved more comprehensive coverage of all the metabolic capabilities of the parasite and implemented functional gene-protein-reaction associations enabling rigorous gene deletions studies. Hereafter we provide a more in depth discussion on the key aspects of our approach.

There are two distinct approaches for model reconstruction process that are commonly referred to as top-down and bottom-up [43]. The top-down approach is usually applied when detailed experimental data about the majority of individual system components is scarce. Conversely, for the bottom-up approach, a significant body of relevant primary literature is a necessary prerequisite. These two approaches are very much complementary and this is why increasingly more studies combine them in order to achieve the best results. This was also the case in the ToxoNet1 reconstruction efforts: the first draft model was a hybrid of the output generated by RAVEN Toolbox and a significantly smaller manually reconstructed (and highly curated) model, which was built upon the similar model of *P. falciparum* [44] used as a template. This bottom-up small-scale model allowed capturing of the features that were not identified by functional annotations, such as formation of multi-enzyme complexes for certain enzymatic activities (e.g. the mitochondrial BCKDH complex that carries out the pyruvate dehydrogenase function [27]) as well as cofactor specificities for the enzymes.

Genes with metabolic functions in ToxoNet1 represent a modest fraction (c.a. 9%) of the *T. gondii* protein-coding genes, which is comparable but greater than in the genome-scale models of *P. falciparum* (7%) [8], *L. major* (6.7%) [20] and *Cryptosporidium hominis* (5.5%) [7]. ToxoNet1 provides a broader and more complete coverage of the metabolic capabilities of the parasite with 145 additional enzyme-coding genes (38% greater) compared to the earlier metabolic model of *T. gondii* [10]. Comparison of the number of metabolic reactions with the genome-scale models of the malaria parasite *P. falciparum* [8,9,13] confirms the common view that *T. gondii* possesses broader metabolic capabilities (Table 5). These differences in metabolic capabilities could be partially responsible for the observations that *T. gondii* can infect a very broad

Table 5. Comparison of the number of metabolites, reactions and genes between the models for *T. gondii* and *P. falciparum*.

Metabolic models of apicomplexan parasites	Information about the models			
	Metabolites	Reactions	Genes	Compartments
P. falciparum [9]	850	998	579	c, m, a, er, n, v, g
P. falciparum [8]	616	656	366	e, c, m, a
T. gondii [10]	384	400	382	c, m, a, er, im
T. gondii, this study	1019	1089	527	e, c, m, a

Abbreviations: e—extracellular space, c—cytosol, m—mitochondrion, a—apicoplast, n—nucleus, er—endoplasmic reticulum, g—Golgi complex, v—digestive vacuole, im—mitochondrial intermembrane space.

doi:10.1371/journal.pcbi.1004261.t005

range of cell types for asexual replication, while *Plasmodium* spp. can only replicate within specific tissues.

ToxoNet1 contains confidence estimates (e-values and bit-scores) for each gene annotated as encoding certain metabolic activities. This has allowed us to suggest a number of metabolic functions for genes that did not have annotations with E.C. numbers in the ToxoDB database (S4 Table). For example, TGME49_237140 (an "ethylene-inducible protein" in ToxoDB without an E.C. identifier) was annotated as pyridoxal 5'-phosphate synthase pdxS subunit (E.C. 4.3.3.6) with very high confidence estimates (e-value $5.30x10^{-141}$ and bit-score 479). Accordingly, an orthologous gene (PF3D7_0621200) is annotated as pyridoxine biosynthesis protein (PDX1) in *P. falciparum*, yet without an E.C. number assigned. There was also a set of genes with low confidence in their function, despite their annotation in ToxoDB (e.g. TGME49_305840 had a low sequence identity to known nicotinate-nucleotide adenylyltransferases with an e-value of $4.70x10^{-15}$ and a bit-score of 60.6). These two examples demonstrate the potential of the RAVEN Toolbox to improve genome annotation and produce models with confidence estimates that enable evaluation and, potentially, future corrections of the existing models.

ToxoNet1 contains 260 unique dead-end metabolites, which currently can be only produced or only consumed in the network. We kept them in the model in order to allow future developments and expansion of the scope of the model. Further definition of the metabolic routes can utilise these dead-end metabolites and contribute to a more complete understanding of the metabolic capabilities of the parasite. For instance, the tRNAs loaded with the respective amino acids were included to enable future extension of the model to the representation of protein synthesis.

We chose to impose relaxed constraints in terms of subcellular compartments due to rather high uncertainty in subcellular localisation of the enzymes as well as largely unknown capabilities of the parasite to transport metabolites across its organellar membranes. We assigned the enzymes with the corresponding reactions to their putative compartments and allowed a broad range of metabolites to be transported across the compartment boundaries. This approach to compartmentalisation ensures minimal bias in our results due to underestimated transport capabilities or incorrect assignment of enzymes to subcellular compartments. We expect subsequent refinements of ToxoNet1 with more stringent compartmentalisation as more reliable, high-throughput computational and experimental methods are becoming available for subcellular localisation of enzymes and functional annotation of transporter proteins in *T. gondii*. Our objective of this work is to provide ToxoNet1 as a resource for the community and therefore we avoided imposing constraints and hypotheses, which are not well-tested and confirmed that could contaminate the model and our results.

To understand the nutritional requirements of *T. gondii*, we implemented an algorithm that identifies and ranks the minimal number of substrates required for growth in ToxoNet1. The results indicate that uptake of as few as 19 of 230 substrates allows the model to produce all biomass building blocks. Moreover, we identified 2592 sets of 19 substrates that can allow growth, and 10 substrates common to all these sets (Table 3). These 10 substrates are compounds that are not synthesized *de novo* by *T. gondii* and are the precursors for biomass building blocks and essential cofactors. Interestingly, the uptake of S-adenosyl-homocysteine provided precursors for multiple biomass building blocks simultaneously (purine nucleotides, threonine, methionine and their derivatives). Among the alternative substrates we identified 7 carbon sources: five hexoses and two pentoses (ribose and deoxyribose), which can be incorporated into the metabolism by a ribokinase enzyme. Uptake of pentoses through a hexose transporter has been reported in the protozoan parasite *L. major* and may potentially represent an additional level of versatility in meeting the need of a carbon source, similarly to the previously observed in *T. gondii* utilization of glutamine [45,46]. We could envision such need as the parasite uses various host cells, where the set of available substrates may vary between cell types.

Validation of genome-scale models using metabolic tasks is an approach developed to evaluate metabolic capabilities of the models in a systematic manner. To date it has been applied to several models [11,47,48] for testing whether the major metabolic functions can be fulfilled and the biochemical pathways that support these functions were represented correctly. ToxoNet1 is the first metabolic model of an apicomplexan parasite where the reconstruction process involved the validation of metabolic tasks. Using this approach we explored the capability of this pathogen to produce *de novo* biomass building blocks from glucose and inorganic substances. Interestingly, almost all the cofactors of metabolic enzymes, with the exception of pyridoxal phosphate cannot be produced *de novo*. Their synthesis required precursors that contained certain chemical moieties (Table 2). This suggests that virtually all the metabolic functions of the parasite depend on an adequate supply of specific precursors by the infected host cell. Furthermore, we found that *T. gondii* requires uptake of almost all the amino acids, which are essential for the host, with the possible exception of threonine (discussed below). Notably, *T. gondii* lacks the pathway for arginine synthesis, which can be produced by human cells, however arginine is growth limiting for human cells because its *de novo* synthesis is insufficient and thus supplemental uptake is required [49]. This suggests that during infection and growth of the parasite, the biosynthetic capabilities of the host can be significantly compromised due to competition for essential and growth limiting amino acids and vitamins.

Threonine is the only amino acid essential for human cells and which *T. gondii* can potentially produce. However, this pathway includes the enzyme homoserine dehydrogenase (E.C. 1.1.1.3), the encoding gene for which is not identified to date in the genome of the parasite. Data from ToxoDB suggests that the genes encoding the enzymes of the threonine biosynthesis pathway, namely aspartokinase (TGME49_227090), threonine synthase (TGME49_220840), homoserine kinase (TGME49_216640) and aspartate-semialdehyde dehydrogenase (TGME49_205420), are expressed at low levels in the tachyzoite stage. Homoserine kinase has been detected in the proteome of oocysts [50], the infective forms of the parasite that can survive for extensive periods of time outside the host cell [51]. Taken together, these observations indicate that the whole pathway from aspartate to threonine might be not functional in certain life stages and therefore *T. gondii* could be a conditional threonine auxotroph. Should threonine biosynthesis also be active in the other life stages it could represent a selective drug target for inhibition of parasite replication without affecting the mammalian host cells. Maintenance of some metabolite pools in *T. gondii* may be a result of both uptake and *de novo* production. Hence, if the complete set of genes of a metabolic pathway is present in the genome of the parasite, this pathway is not necessarily active and will not meet the demands of the parasite

regarding the products of the pathway. Fox *et al.* showed that salvage of pyrimidine nucleotides from the host cell can support *in vitro* but not *in vivo* survival of *T. gondii* mutants with disrupted *de novo* pyrimidine synthesis pathways [52]. Therefore, accurate consideration of metabolism in the host and pathogen together is central for a better understanding of the metabolic needs and capabilities of *T. gondii*.

Our study represents the first model-based gene essentiality predictions for *T. gondii*. The earlier efforts to infer essentiality of genes in this parasite were reported by Gautam et al. [53]. Their approach for producing the list of putatively essential genes was largely dissimilar to ours and relied on conservation of the enzymes across parasitic and free-living species with further pruning based on the literature data. We believe that our modelling approach, which takes into account many more inputs and constraints, produces more reliable gene essentiality predictions, and it follows the standard modelling procedures established by the community [12,54]. Nevertheless, we consider that the part of the workflow proposed by Gautam et al. can be complementary to our study, which also includes larger number genes with the updated functional annotation we performed here. This future study will prioritise the putatively essential genes according to their apparent utility as drug targets with minimal risk of off-target effects on the host cell metabolism.

Using ToxoNet1 we predicted a minimal set of 53 enzyme-encoding genes to be indispensable for parasite replication within human cells. The majority of these genes (49 out of 53) have orthologs in the malaria parasite *P. falciparum* and 32 of them were predicted as essential in the existing metabolic models [38]. Evidence of essentiality for metabolic enzymes in *T. gondii* is currently scarce, thus, we could not test the majority of our essentiality predictions (Table 5). However, new technologies such as CRISPR/Cas9-mediated genome editing, hold promise of forthcoming high-throughput gene knockout studies in various organisms including *T. gondii* [42]. The dataset of experimentally established gene essentiality will provide an important validation of ToxoNet1 and a prerequisite for its further refinement and expansion.

We predicted synthetic essentiality of 20 pairs of genes, which represent two distinct cases. In the first case the pair of genes encode two isoenzymes that catalyse the same reaction in the same compartment. For example, the genes TGME49_318580 and TGME49_285980 encode isoenzymes, which catalyse the phosphoglucomutase reaction (E.C. 5.4.2.2) and both were experimentally localised to the cytosol [33]. This reaction produces glucose-1-phophate in the cytosol, which is an indispensable precursor for starch and nucleotide-sugar metabolism. Therefore, simultaneous deletion of these two genes is a synthetic lethal. In the second case of synthetic essentiality each gene encodes for an enzyme of a different reaction but with a common product, which is necessary for biomass synthesis. Thus, the two different enzymes can substitute for each other (sometimes indirectly) for production of an essential metabolite. For example, we established experimentally the synthetic lethality between the genes encoding acetyl-CoA synthase (ACS) and ATP-citrate lyase (ACL), both of which produce acetyl-CoA in the cytosol of the parasite. This central metabolite participates notably in fatty acid synthesis, fatty acid chain elongation and in acetylation of proteins. We were able to rule out experimentally that the severe phenotype observed upon depletion of cytosolic acetyl-CoA was due to an impact on FASI since parasites lacking FASI did not exhibit any significant impairment in the lytic cycle.

Moreover given that deletion of the genes coding for elongases previously reported [55] did not phenocopy the lethality observed when attempting to delete ACS and ACL simultaneously, we suspect that blockage of protein acylation might be responsible for the severe consequences of deletion of cytosolic acetyl-CoA. In this context, we experimentally confirmed that the *T. gondii* acetyl-CoA transporter (AT1) is localized in the membrane of the ER and is anticipated to deliver acetyl-CoA to the secretory pathway. It would be interesting to examine the

importance AT1 for parasite survival. In contrast, the mitochondrion and the apicoplast should not be affected since these compartments have their own routes for production of acetyl-CoA: (i) in the apicoplast it is produced by the pyruvate dehydrogenase (PDH) complex [56], and (ii) in the mitochondrion synthesis of acetyl-CoA is carried out by the BCKDH complex [27]. The predicted synthetic lethality between ACS and ACL suggests that these are the only two significant sources of acetyl-CoA in the cytosol, which was not assessed by previous experimental studies and is formally demonstrated here.

Furthermore this synthetic lethality suggests that there is no, or no significant, transport of acetyl-CoA to the cytosol from the mitochondrion or the apicoplast. This further confirms that, unlike acetate, acetoacetate could not be a source of cytosolic acetyl-CoA and therefore the metabolic role of acetoacetyl-CoA, considered to be a dead-end metabolite in ToxoNet1, requires further investigation. Unexpectedly, the deletion of the gene coding for ACS leads to a reproducible and immediate increase in abundance of ACL protein, a possible compensatory adaptation that has not been reported previously. It is unclear whether the capacity of the parasite to respond to the absence of ACS results from an increase at the level of transcription or protein stability but this intriguing mechanism of adaptation deserves further investigation. ToxoNet1 can assist in comprehensively embracing the various routes that *T. gondii* employs to produce acetyl-CoA in different subcellular compartments.

Beyond this study ToxoNet1 can be used as a global metabolic context for integration and interpretation of various high-throughput experimental data, similarly to the studies made in *P. falciparum* and other eukaryotic pathogens [38]. Of particular interest is the integration of experimental data collected on different life stages of *T. gondii* that will allow the model to yield context-specific predictions and, potentially, reveal valid drug targets as well as fundamental knowledge regarding the stages implicated in persistence and transmission of this important human and animal pathogen.

Materials and Methods

Draft reconstruction of ToxoNet1

The reconstruction process started with generation of a draft metabolic network of *T. gondii* based on the annotation of its protein sequences as extracted from the ToxoDB database [57]. Within the framework of the RAVEN Toolbox these protein sequences were compared to the hidden Markov models (HMM) generated for each KEGG [17] orthology group [11]. In the cases when e-values for the matches between a *T. gondii* protein and an HMM were smaller than the specified e-value cutoff (10^{-20}), the enzymatic reactions associated with the corresponding KEGG orthology ID were added to the draft metabolic network. As a result we have obtained a set of metabolic reactions linked to genetic loci of *T. gondii*. At this stage the model did not contain information about either subcellular compartments, or about transport of metabolites. We next merged this model with a manually curated, small-scale metabolic reconstruction that we previously built based on the *P. falciparum* model [44].

Similarly to Huthmatcher *et al* [9]., we removed reactions with generic metabolites (such as "protein", "dNTP") and replaced non-unique metabolite identifiers with unique ones when they corresponded to the biochemically equivalent entities (e.g. (S)-Lactate and L-lactate were replaced with L-lactate).

Compartmentalization

The parasitophorous vacuole (PV) space, which secludes *T. gondii* from the host cell cytosol, is the outermost compartment represented in ToxoNet1. Extracellular compartment of the model corresponds to the PV space. A eukaryotic organization of intracellular space of the

parasite also includes a number of compartments, among which we chose to represent in ToxoNet1 the most relevant to metabolism, namely: mitochondrion, apicoplast and cytosol. The following sequence-based localization predictors were used to suggest putative subcellular localization of the enzymes: TargetP [16] (version 1.1), MitoProt II [15] (version 1.101) and ApicoAP [14] (version 1)). As an input data we used sequences of ORFs extracted from ToxoDB (v.9, strain ME49) for the genes included in ToxoNet1. Computational localization predictions were reconciled with the literature evidence for *T. gondii* and *P. falciparum* in the cases when the latter were available (see S1 Table).

Transport reactions were included to link the majority of the metabolites in non-cytosolic intracellular compartments with their cytosolic counterparts; such transports were not created for phosphorylated metabolites or those that contained [acyl-carrier protein] ([ACP]) or Coenzyme A (CoA) moiety attached (with the exception of the apicoplast-to-cytosol transport of IPP, and DMAPP as well as the mitochondrion-to-cytosol transport of ATP).

Metabolic tasks

Testing metabolic tasks is a built-in functionality of the RAVEN Toolbox [11] meant for verification of correct metabolic capabilities in genome-scale models. Formulation of this function allows one to test production of a certain metabolite(s) (or flux through particular reactions) with strictly specified inlet and outlet of metabolites. For instance, a metabolic task "*de novo* synthesis of ATP" was formulated as following: given unlimited uptake of glucose, oxygen, inorganic phosphate, sulfate and ammonium, can the model produce ATP. The model failed to accomplish this task, which is consistent with the literature knowledge about auxotrophy of *T. gondii* for purines [58]. Thus, the task could be passed only when we added to the list of available compounds hypoxanthine, adenine or another molecule containing a purine moiety. In a similar manner we have tested over a 60 metabolic tasks (all in S2 Table) to ensure realistic metabolic capabilities of the model.

The range of available substrates

We simulated metabolism of the host cell using the most recent tissue-unspecific model of human metabolism [19]. The Recon2 model was modified in terms of available substrates to reflect growth on the defined minimal medium (Dulbecco's Modified Eagle's Medium with glutamine and glucose). The PV which secludes the parasite from the host cell cytosol had been reported as being permeable to small-molecule metabolites with molecular weights below 1500 Da [59] and thus imposes no relevant constraint to the metabolites we considered in ToxoNet1. Thus, all the molecules that could be produced from the medium components in the cytosol of the host cell were assumed to be potentially accessible for the parasite provided that they satisfy the following criteria: host-supplied substrates were only small molecules (below 1.5 kDa) that were not phosphorylated or bound to-CoA,-[ACP] or carnitine. We also assumed that the parasite could potentially dispose of a wide range of metabolic by-products from its cytosol into the host cell. Thus, we added sink reactions for the same 230 metabolites that were assumed to be host-supplied. As an exception we also allowed direct uptake into the cytosol of the following five generic metabolites: "reduced acceptor", "sulfur donor", "acyl-CoA", "carboxylate" and "1-acylglycerol"; two non-generic metabolites: selenite and myo-inositol, and the appearance of apoprotein in the mitochondrion and the apicoplast.

Biomass composition

In order to simulate replication of the parasite and assess the essentiality of genes, reactions and substrates, we assumed maximisation of flux through the biomass reaction to be the

objective function of ToxoNet1. It represents cellular replication as a reaction that consumes pre-defined amounts of metabolites defined as small-molecule biomass precursors as well as energy in the form of ATP. We used the biomass reaction from the previous study [10] as a template and introduced the following modifications: we extended this biomass reaction with the following cofactors—NAD, NADP, FAD and lipoylated protein necessary for pyruvate dehydrogenase (PDH) activity in the mitochondrion and the apicoplast; we changed the compartment for lipid precursors from endoplasmic reticulum (in Song et al. [10]) to the cytosol (no ER compartment in ToxoNet1), and for L-lysine from mitochondrial to cytosol. The presence of the mitochondrial pathway for L-lysine production remained obscure, thus in ToxoNet1 it is acquired from the host instead of gap-filled *de novo* production in the parasite and sequestered towards biomass from the cytosol.

Flux balance analysis (FBA) and substrate dispensability simulation

FBA is a standard computational approach for exploration of the metabolic capabilities represented in constraint-based models; principles, computational implementation of FBA as well as the key assumptions are extensively described elsewhere [60].

In the absence of clear knowledge on the scope of substrates that the parasite can take up from the host we made the following assumption: all the metabolites present in the host cell cytosol are potentially accessible for the parasite except those that are phosphorylated, bound to coenzyme A, acyl-carrier protein or carnitine. A list of molecules that satisfy these criteria was generated using the recent tissue-unspecific model of human metabolism Recon2 [19] (the list of the substrates is in S6 Table).

Minimal sets of metabolites necessary for the production of biomass were explored using an in-house developed mixed-integer linear programming algorithm. For each of the exchange reactions that allowed uptake of a substrate into the model (i.e. 230 host-supplied metabolites) we added one binary variable that denoted its utilization. Our algorithm solved the model subject to minimization of the sum of the binary variables thus yielding a minimal set of uptakes that allowed the model to simulate growth. After each iteration the algorithm added one new constraint to the model to assure that the following set would include at least one different uptake reaction compared to all the previously generated ones. The iterations were repeated until no more alternative sets of the same length could be found.

Essentiality studies

We performed simulation of gene deletion using a conventional approach [61] that implies evaluation of gene-reaction associations that include the gene of interest, preventing flux through corresponding metabolic reactions and an attempt to achieve a doubling time of 4.5 hours (specific growth rate of 0.15 h^{-1}). Similar approaches were used for double gene deletion studies—simultaneously blocking reactions associated with all pairs of genes that were not predicted as singularly essential. In reaction deletion simulations we blocked flux through every single reaction in ToxoNet1 one at a time. Subsequent attempts at achieving the wild-type doubling time in the absence of the reaction indicated whether the gene was dispensable or not.

T. gondii strains and culture

T. gondii tachyzoites (RH*ku80-ko* (Ku80ko), RH*ku80-ko/ACS-ko* (ACSko), RH*ku80-ko/ACL-ko* (ACLko), RHku80-ko/ACL3Ty-LoxP3'UTRLoxP-U1 (ACL-lox), RHku80-ko/ACS-ko/ACL3Ty-LoxP3'UTRLoxP-U1 (ACSko/ACL-lox), RH/FASIko (FASIko)) were grown in confluent Human Foreskin Fibroblasts (HFF) and maintained in Dulbecco's Modified Eagle's

Medium (DMEM, Life technology, Invitrogen) supplemented with 5% foetal calf serum, 2 mM glutamine and 25 μg/ml gentamicin in a humidified incubator at 37°C with 5% CO_2.

Cloning of DNA constructs

All amplifications were performed with LA Taq (TaKaRa) polymerase and primers used are listed in S7 Table.

Knock-in of ACS: *(pKI-ACS-3Ty)* The genomic fragment of ACS (TGME49_266640) was amplified using primers ACS-1 and ACS-2 prior to digestion with with *KpnI* and *SbfI* and subsequent cloning in the same sites of pTub8MIC13-3Ty-HXGPRT [62] to introduce 3 Ty-tags at the C-terminus of the endogenous locus. Before transfection pKI-ACS-3Ty was linearized with *SnaBI*.

Knock-in of AT1: *(pKI-AT1-3Ty)* The genomic fragment of AT-1 (TGME49_215940) was amplified using primers AT1-1 and AT1-2 prior to digestion with *KpnI* and *NsiI* and subsequent cloning in the same sites of pTub8MIC13-3Ty-HXGPRT [62] to introduce 3 Ty-tags at the C-terminus of the endogenous locus. Before transfection pKI-AT1-3Ty was linearized with *HindIII*.

Knock-in and knock-down of ACL: *(pKI-ACL-3Ty and pKI-ACL-3Ty-LoxP-3'UTR-LoxP-U1)* Genomic fragment of ACL (TGME49_223840) was amplified using primers ACL-1/ACL-2 and subsequently digested with *KpnI* and *NsiI* prior to cloning in the same sites of the pTub8-MIC13-3Ty-HXGPRT for 3Ty-tag knock-in [62], or the modified C-terminal destabilization vector pG152-3Ty-LoxP-3'UTRSag1-HXGPRT-LoxP-U1 [41]. Prior to transfection both plasmids were linearized with *XhoI*.

Knockout of ACS: *(pTub-CAT-ACS-ko)* around 1.5kb of the 5' and 3' flanking regions of ACS were amplified using primers ACS-3/ACS-4 and ACS-5/ACS-6 respectively. The 5' flanking region was then cloned between *KpnI* and *XhoI* restriction sites of the pTub5-CAT and the 3' flanking region between the *BamHI* and *NotI* sites. The plasmid was cut with *KpnI* and *NotI* prior to transfection.

Knockout of ACL: *(pTub-HXGPRT-ACL-ko)* around 1.5kb of the 5' and 3' flanking regions of *ACL* were amplified using primers ACL-3/ACL-4 and ACL-5/ACL-6 respectively. The 5' flanking region was then cloned between the *KpnI* and *XhoI* sites of the p2855-HXGPRT plasmid and the 3' flanking region between the *BamHI* and *NotI* sites. The plasmid was cut with *KpnI* and *NotI* restriction enzymes prior to transfection.

Frameshift knockout of FASI using CRISPR/CAS9 plasmid [42]: This vector has been generated using the Q5 site-directed mutagenesis kit (New England Biolabs) with the vector pSAG1::CAS9-U6::sgUPRT as template (a gift from Dr. L.D. Sibley). The UPRT-targeting gRNA was replaced by the FASI (TGME49_294820) specific gRNA using the primer pairs gRNA-FASI/gRNA-rev (gRNA highlighted in red in S7 Table).

Parasite transfection and selection of stable transformants

Parasite transfections were performed by electroporation as previously described [63].The *hxgprt* gene was used as a positive selectable marker in the presence of mycophenolic acid (25 μg/mL) and xanthine (50 μg/mL) for pKI-ACS-3Ty, pKI-AT1-3Ty, pKI-ACL-3Ty, pKI-ACL-3Ty-LoxP-3'UTR-LoxP-U1 and pTub-HXGPRT-ACL-ko vectors transfected in Ku80ko or ACSko, as previously described [64].

Ku80ko was transfected with pTub-CAT-ACS-ko and 20μM chloramphenicol was used to select resistant parasites.

Resistant parasites were cloned by limiting dilution in 96 well plates and clones were assessed by genomic PCR.

To efficiently disrupt the *FASI* locus, 30 ug of the *FASI* gRNA-specific CRISPR/CAS9 vector was transfected into wild type RH parasites. 24 hours after transfection, GFP positive parasites were sorted by flow cytometry and cloned into 96-well plates using a Moflo Astrios (Beckman Coulter). Individual clones were then analysed by sequencing.

Preparation of *T. gondii* genomic DNA

Genomic DNA was prepared from tachyzoites using the Wizard SV genomic DNA purification system (Promega). Correct integration of the different constructs into the genome of the various strains was determined by genomic PCR using GoTaq Green Master Mix (Promega).

Antibodies

The antibodies used in this study were described previously as follows: polyclonal rabbit anti-GAP45, rabbit anti-TgProfilin [65], monoclonal mouse anti-Ty (BB2), mouse monoclonal anti-F1-ATPase beta subunit (P. Bradley, unpublished) (5F4), mouse monoclonal anti-ATrx1 11G8 [66]. For Western blot analyses, secondary peroxidase conjugated goat anti-rabbit or mouse antibodies (Molecular Probes) were used. For immunofluorescence analyses, the secondary antibodies Alexa Fluor 488 and Alexa Fluor 594-conjugated goat anti-mouse or rabbit antibodies (Molecular Probes) were used.

Immunofluorescence assay (IFA) and confocal microscopy

Parasite-infected HFF cells seeded on cover slips were fixed with 4% paraformaldehyde/0.05% glutaraldehyde (PFA/Glu) in PBS. Fixed cells were then processed as previously described [67]. Confocal images were generated with a Zeiss (LSM700, objective apochromat 63x/1.4 oil) laser scanning confocal microscope at the Bioimaging core facility of the Faculty of Medicine, University of Geneva. Stacks of sections were processed with ImageJ and projected using the maximum projection tool.

Western blot analyses

Parasites were lysed in PBS-1% Triton X-100 and mixed with SDS–PAGE loading buffer under reducing conditions. The suspension was subjected to sonication on ice. SDS-PAGE was performed using standard methods. Separated proteins were transferred to nitrocellulose membranes and probed with appropriate antibodies in 5% non-fat milk in PBS-0.05% Tween20. Bound secondary peroxidase conjugated antibodies were visualized using the SuperSignal (Pierce).

Phenotypic analyses

Plaque assays: HFF monolayers were infected with parasites and let to develop for 7 days before fixation with PFA/GA and Crystal Violet staining to visualize plaques.

Intracellular growth assays: HFFs were inoculated with parasites, washed 4h post infection and coverslips were fixed at 24 h post-infection with 4% PFA/Glu and stained by IFA with rabbit anti-TgGAP45. Number of parasites per vacuole was counted in triplicate for each condition (n = 3). More than 200 vacuoles were counted per replicate.

Software and databases

Flux-balance analysis was performed using MATLAB (Version R2012b, The MathWorks) with CPLEX (ILOG IBM, version 12.51) and Mosek (version 7) solvers; RAVEN Toolbox was used within Matlab environment (version 1.07, downloaded from http://129.16.106.142/tools.php?c = raven).

Input data was taken from KEGG database (www.kegg.jp, up to date as of 18/11/2013), ToxoDB (www.toxodb.org, version 9), ApiLoc database (http://apiloc.biochem.unimelb.edu.au/apiloc/apiloc, version 3), and LLAMP portal (www.llamp.net, no evident version tracking).

Supporting Information

S1 Fig. (A) Sequencing from translational start (ATG) of 2 independent clones confirms disruption of the FASI ORF induced by CAS9. PCR for sequencing was done using primers FAS-1/FAS-2 (S7 Table). The gRNA used is written in red. (B) Plaque assays performed with RH, FASIko clone 8 and 10 parasite lines and fixed after 7 days. No significant defect in the lytic cycle could be observed.
(EPS)

S1 Table. Subcellular localization of enzymes in ToxoNet1.
(XLSX)

S2 Table. Metabolic tasks performed on the metabolic network of *T. gondii*.
(XLSX)

S3 Table. Gap-filling reactions introduced in the model and their description.
(XLSX)

S4 Table. New putative functional annotation for the genes not annotated with E.C. identifiers in ToxoDB.
(XLSX)

S5 Table. Essentiality predictions (genes, gene pairs, reactions).
(XLSX)

S6 Table. The list of 230 metabolites assumed as host-supplied substrates in ToxoNet1.
(XLSX)

S7 Table. Oligonucleotide primers used in this study for cloning and PCR analyses.
(PDF)

Acknowledgments

We thank Meriç Ataman for considerable assistance in implementation of the algorithm for simulation of minimal *in silico* media. We thank Dr. Peter Bradley for the 5F4 and Atrx1 antibodies and Dr. David Sibley (Washington University) for providing the CRISPR/CAS9 vector. We are grateful to Dr. Hayley Bullen for critical reading of the manuscript.

Author Contributions

Conceived and designed the experiments: VH DSF. Analyzed the data: ST RDO DSF VH. Contributed reagents/materials/analysis tools: DSF RA JN VH. Wrote the paper: ST RDO DSF RA JN VH. Curated the model: ST RDO DSF RA JN VH Performed model reconstruction: ST Generated and characterized the transgenic parasites and analyzed the phenotypes: RDO.

References

1. Tenter AM, Heckeroth AR, Weiss LM (2000) *Toxoplasma gondii*: from animals to humans. Int J Parasitol 30: 1217–1258. PMID: 11113252
2. Weiss LM, Dubey JP (2009) Toxoplasmosis: A history of clinical observations. Int J Parasitol 39: 895–901. doi: 10.1016/j.ijpara.2009.02.004 PMID: 19217908

3. Macrae JI, Sheiner L, Nahid A, Tonkin C, Striepen B, et al. (2012) Mitochondrial Metabolism of Glucose and Glutamine Is Required for Intracellular Growth of *Toxoplasma gondii*. Cell Host Microbe. pp. 682–692.
4. Polonais V, Soldati-Favre D (2010) Versatility in the acquisition of energy and carbon sources by the Apicomplexa. Biology of the Cell 102: 435–445. doi: 10.1042/BC20100005 PMID: 20586726
5. Kim HU, Sohn SB, Lee SY (2012) Metabolic network modeling and simulation for drug targeting and discovery. Biotechnol J 7: 330–342. doi: 10.1002/biot.201100159 PMID: 22125297
6. Pinney JW, Papp B, Hyland C, Wambua L, Westhead DR, et al. (2007) Metabolic reconstruction and analysis for parasite genomes. Trends in Parasitology 23: 548–554. PMID: 17950669
7. Vanee N, Roberts SB, Fong SS, Manque P, Buck GA (2010) A genome-scale metabolic model of *Cryptosporidium hominis*. Chem Biodivers 7: 1026–1039. doi: 10.1002/cbdv.200900323 PMID: 20491062
8. Plata G, Hsiao T-L, Olszewski KL, Llinás M, Vitkup D (2010) Reconstruction and flux-balance analysis of the *Plasmodium falciparum* metabolic network. Mol Syst Biol. pp. 408.
9. Huthmacher C, Hoppe A, Bulik S, Holzhütter H-G (2010) Antimalarial drug targets in *Plasmodium falciparum* predicted by stage-specific metabolic network analysis. BMC Syst Biol. pp. 120.
10. Song C, Chiasson MA, Nursimulu N, Hung SS, Wasmuth J, et al. (2013) Metabolic reconstruction identifies strain-specific regulation of virulence in *Toxoplasma gondii*. Mol Syst Biol. pp. 708.
11. Agren R, Liu L, Shoaie S, Vongsangnak W, Nookaew I, et al. (2013) The RAVEN Toolbox and Its Use for Generating a Genome-scale Metabolic Model for *Penicillium chrysogenum*. PLoS Comput Biol. pp. e1002980.
12. Thiele I, Palsson BO (2010) A protocol for generating a high-quality genome-scale metabolic reconstruction. Nature Protocols 5: 93–121. doi: 10.1038/nprot.2009.203 PMID: 20057383
13. Bazzani S, Hoppe A, Holzhütter H-G (2012) Network-based assessment of the selectivity of metabolic drug targets in *Plasmodium falciparum* with respect to human liver metabolism. BMC Syst Biol. pp. 118.
14. Cilingir G, Broschat SL, Lau AO (2012) ApicoAP: the first computational model for identifying apicoplast-targeted proteins in multiple species of Apicomplexa. PLoS One 7: e36598. doi: 10.1371/journal.pone.0036598 PMID: 22574192
15. Claros MG, Vincens P (1996) Computational method to predict mitochondrially imported proteins and their targeting sequences. European Journal of Biochemistry 241: 779–786. PMID: 8944766
16. Emanuelsson O, Brunak S, von Heijne G, Nielsen H (2007) Locating proteins in the cell using TargetP, SignalP and related tools. Nat Protoc 2: 953–971. PMID: 17446895
17. Ogata H, Goto S, Sato K, Fujibuchi W, Bono H, et al. (1999) KEGG: Kyoto Encyclopedia of Genes and Genomes. Nucleic Acids Res 27: 29–34. PMID: 9847135
18. Shanmugasundram A, Gonzalez-Galarza FF, Wastling JM, Vasieva O, Jones AR (2012) Library of Apicomplexan Metabolic Pathways: a manually curated database for metabolic pathways of apicomplexan parasites. Nucleic Acids Res. pp. D706-D713.
19. Thiele I, Swainston N, Fleming RMT, Hoppe A, Sahoo S, et al. (2013) A community-driven global reconstruction of human metabolism. Nat Biotechnol. pp. 419–425.
20. Chavali AK, Whittemore JD, Eddy JA, Williams KT, Papin JA (2008) Systems analysis of metabolism in the pathogenic trypanosomatid *Leishmania major*. Mol Syst Biol 4: 177. doi: 10.1038/msb.2008.15 PMID: 18364711
21. Woodcroft BJ, McMillan PJ, Dekiwadia C, Tilley L, Ralph SA (2012) Determination of protein subcellular localization in apicomplexan parasites. Trends Parasitol. pp. 546–554.
22. Pino P, Foth BJ, Kwok L-Y, Sheiner L, Schepers R, et al. (2007) Dual targeting of antioxidant and metabolic enzymes to the mitochondrion and the apicoplast of *Toxoplasma gondii*. PLoS Pathog. pp. e115.
23. Foth BJ, McFadden GI (2003) The apicoplast: a plastid in *Plasmodium falciparum* and other Apicomplexan parasites. Int Rev Cytol 224: 57–110. PMID: 12722949
24. Brooks CF, Johnsen H, van Dooren GG, Muthalagi M, Lin SS, et al. (2010) The toxoplasma apicoplast phosphate translocator links cytosolic and apicoplast metabolism and is essential for parasite survival. Cell Host Microbe 7: 62–73. doi: 10.1016/j.chom.2009.12.002 PMID: 20036630
25. Hunter WN (2011) Isoprenoid Precursor Biosynthesis Offers Potential Targets for Drug Discovery Against Diseases Caused by Apicomplexan Parasites. Current Topics in Medicinal Chemistry 11: 2048–2059. PMID: 21619509
26. Fritz HM, Bowyer PW, Bogyo M, Conrad PA, Boothroyd JC (2012) Proteomic analysis of fractionated *Toxoplasma* oocysts reveals clues to their environmental resistance. PLoS One 7: e29955. doi: 10.1371/journal.pone.0029955 PMID: 22279555

27. Oppenheim RD, Creek DJ, Macrae JI, Modrzynska KK, Pino P, et al. (2014) BCKDH: the missing link in apicomplexan mitochondrial metabolism is required for full virulence of *Toxoplasma gondii* and *Plasmodium berghei*. PLoS Pathog 10: e1004263. doi: 10.1371/journal.ppat.1004263 PMID: 25032958

28. Divakaruni AS, Murphy AN (2012) Cell biology. A mitochondrial mystery, solved. Science 337: 41–43. doi: 10.1126/science.1225601 PMID: 22767917

29. Bricker DK, Taylor EB, Schell JC, Orsak T, Boutron A, et al. (2012) A Mitochondrial Pyruvate Carrier Required for Pyruvate Uptake in Yeast, *Drosophila*, and Humans. Science. pp. 96–100.

30. Herzig S, Raemy E, Montessuit S, Veuthey J-L, Zamboni N, et al. (2012) Identification and Functional Expression of the Mitochondrial Pyruvate Carrier. Science. pp. 93–96.

31. Yardley V, Khan AA, Martin MB, Slifer TR, Araujo FG, et al. (2002) *In vivo* activities of farnesyl pyrophosphate synthase inhibitors against *Leishmania donovani* and *Toxoplasma gondii*. Antimicrobial Agents and Chemotherapy. pp. 929–931.

32. Ling Y, Sahota G, Odeh S, Chan JM, Araujo FG, et al. (2005) Bisphosphonate inhibitors of *Toxoplasma gondii* growth: *in vitro*, QSAR, and *in vivo* investigations. J Med Chem 48: 3130–3140. PMID: 15857119

33. Fleige T, Pfaff N, Gross U, Bohne W (2008) Localisation of gluconeogenesis and tricarboxylic acid (TCA)-cycle enzymes and first functional analysis of the TCA cycle in *Toxoplasma gondii*. International Journal for Parasitology. pp. 1121–1132.

34. Mazumdar J, H Wilson E, Masek K, A Hunter C, Striepen B (2006) Apicoplast fatty acid synthesis is essential for organelle biogenesis and parasite survival in *Toxoplasma gondii*. Proc Natl Acad Sci USA. pp. 13192–13197.

35. Gupta N, Zahn MM, Coppens I, Joiner KA, Voelker DR (2005) Selective disruption of phosphatidylcholine metabolism of the intracellular parasite *Toxoplasma gondii* arrests its growth. J Biol Chem 280: 16345–16353. PMID: 15708856

36. Sampels V (2012) Plasticity of the phosphatidylcholine biogenesis in the obligate intracellular Parasite *Toxoplasma gondii*: Humboldt-Universität zu Berlin, Mathematisch-Naturwissenschaftliche Fakultät I, PhD thesis, Available: http://edoc.hu-berlin.de/dissertationen/sampels-vera-2012-03-27/PDF/sampels.pdf

37. Anderson AC (2005) Targeting DHFR in parasitic protozoa. Drug Discov Today 10: 121–128. PMID: 15718161

38. Tymoshenko S, Oppenheim RD, Soldati-Favre D, Hatzimanikatis V (2013) Functional genomics of *Plasmodium falciparum* using metabolic modelling and analysis. Briefings in functional genomics. pp. 316–327.

39. Feist AM, Palsson BO (2010) The biomass objective function. Current Opinion in Microbiology 13: 344–349. doi: 10.1016/j.mib.2010.03.003 PMID: 20430689

40. Jonas MC, Pehar M, Puglielli L (2010) AT-1 is the ER membrane acetyl-CoA transporter and is essential for cell viability. Journal of Cell Science 123: 3378–3388. doi: 10.1242/jcs.068841 PMID: 20826464

41. Pieperhoff MS, Pall GS, Jimenez-Ruis E, Das S, Wong EH, et al. (2014) Conditional U1 Gene Silencing in *Toxoplasma gondii*, Preprint, Available: http://biorxiv.org/biorxiv/early/2014/09/01/008649.full.pdf. Accessed 01.11.2014

42. Shen B, Brown KM, Lee TD, Sibley LD (2014) Efficient Gene Disruption in Diverse Strains of *Toxoplasma gondii* Using CRISPR/CAS9. Mbio 5: e01114–01114. doi: 10.1128/mBio.01114-14 PMID: 24825012

43. Nielsen J, Jewett MC (2008) Impact of systems biology on metabolic engineering of *Saccharomyces cerevisiae*. FEMS Yeast Res 8: 122–131. PMID: 17727659

44. Forth T (2012) Metabolic systems biology of the malaria parasite: reconstruction, visualisation and analysis of an experimentally parameterised metabolic model of the human acute malaria parasite *Plasmodium falciparum*, PhD thesis, University of Leeds, Available: http://etheses.whiterose.ac.uk/3739/

45. Blume M, Rodriguez-Contreras D (2009) Host-derived glucose and its transporter in the obligate intracellular pathogen *Toxoplasma gondii* are dispensable by glutaminolysis. PNAS. pp. 12998–13003.

46. Lee IP, Evans AK, Yang C, Works MG, Kumar V, et al. (2014) *Toxoplasma gondii* Is Dependent on Glutamine and Alters Migratory Profile of Infected Host Bone Marrow Derived Immune Cells through SNAT2 and CXCR4 Pathways. Plos One 9: e109803. doi: 10.1371/journal.pone.0109803 PMID: 25299045

47. Gille C, Bölling C, Hoppe A, Bulik S, Hoffmann S, et al. (2010) HepatoNet1: a comprehensive metabolic reconstruction of the human hepatocyte for the analysis of liver physiology. Mol Syst Biol. pp. 411.

48. Mardinoglu A, Agren R, Kampf C, Asplund A, Uhlen M, et al. (2014) Genome-scale metabolic modelling of hepatocytes reveals serine deficiency in patients with non-alcoholic fatty liver disease. Nat Commun 5: 3083. doi: 10.1038/ncomms4083 PMID: 24419221

49. Reeds PJ (2000) Dispensable and indispensable amino acids for humans. J Nutr 130: 1835S–1840S. PMID: 10867060

50. Possenti A, Fratini F, Fantozzi L, Pozio E, Dubey JP, et al. (2013) Global proteomic analysis of the oocyst/sporozoite of *Toxoplasma gondii* reveals commitment to a host-independent lifestyle. BMC Genomics 14: 183. doi: 10.1186/1471-2164-14-183 PMID: 23496850

51. Torrey EF, Yolken RH (2013) *Toxoplasma* oocysts as a public health problem. Trends in Parasitology 29: 380–384. doi: 10.1016/j.pt.2013.06.001 PMID: 23849140

52. Fox BA, Bzik DJ (2010) Avirulent Uracil Auxotrophs Based on Disruption of Orotidine-5'-Monophosphate Decarboxylase Elicit Protective Immunity to *Toxoplasma gondii*. Infection and Immunity. pp. 3744–3752.

53. Gautam B, Singh G, Wadhwa G, Farmer R, Singh S, et al. (2012) Metabolic pathway analysis and molecular docking analysis for identification of putative drug targets in *Toxoplasma gondii*: novel approach. Bioinformation 8: 134–141. PMID: 22368385

54. Herrgard MJ, Swainston N, Dobson P, Dunn WB, Arga KY, et al. (2008) A consensus yeast metabolic network reconstruction obtained from a community approach to systems biology. Nature Biotechnology 26: 1155–1160. doi: 10.1038/nbt1492 PMID: 18846089

55. Ramakrishnan S, Docampo MD, MacRae JI, Pujol FM, Brooks CF, et al. (2012) Apicoplast and Endoplasmic Reticulum Cooperate in Fatty Acid Biosynthesis in Apicomplexan Parasite *Toxoplasma gondii*. Journal of Biological Chemistry. pp. 4957–4971.

56. Fleige T, Fischer K, Ferguson DJP, Gross U, Bohne W (2007) Carbohydrate Metabolism in the *Toxoplasma gondii* Apicoplast: Localization of Three Glycolytic Isoenzymes, the Single Pyruvate Dehydrogenase Complex, and a Plastid Phosphate Translocator. Eukaryotic Cell. pp. 984–996.

57. Gajria B, Bahl A, Brestelli J, Dommer J (2008) ToxoDB: an integrated *Toxoplasma gondii* database resource. Nucleic Acids Res. pp. D553-D556.

58. Perrotto J, Keister DB, Gelderma Ah (1971) Incorporation of Precursors into *Toxoplasma* DNA. Journal of Protozoology 18: 470–473. PMID: 5132320

59. Schwab JC, Beckers CJ, Joiner KA (1994) The parasitophorous vacuole membrane surrounding intracellular *Toxoplasma gondii* functions as a molecular sieve. Proc Natl Acad Sci USA. pp. 509–513.

60. Orth JD, Thiele I, Palsson BO (2010) What is flux balance analysis? Nat Biotechnol 28: 245–248. doi: 10.1038/nbt.1614 PMID: 20212490

61. Joyce AR, Palsson BØ (2008) Predicting gene essentiality using genome-scale *in silico* models. Microbial Gene Essentiality: Protocols and Bioinformatics: Springer. pp. 433–457.

62. Sheiner L, Santos JM, Klages N, Parussini F, Jemmely N, et al. (2010) *Toxoplasma gondii* transmembrane microneme proteins and their modular design. Molecular Microbiology 77: 912–929.

63. Soldati D, Boothroyd JC (1993) Transient transfection and expression in the obligate intracellular parasite *Toxoplasma gondii*. Science. pp. 349–352.

64. Donald RGK, Carter D, Ullman B, Roos DS (1996) Insertional tagging, cloning, and expression of the *Toxoplasma gondii* hypoxanthine-xanthine-guanine phosphoribosyltransferase gene—Use as a selectable marker for stable transformation. Journal of Biological Chemistry 271: 14010–14019. PMID: 8662859

65. Plattner F, Yarovinsky F, Romero S, Didry D, Carlier MF, et al. (2008) *Toxoplasma* profilin is essential for host cell invasion and TLR11-dependent induction of an interleukin-12 response. Cell Host & Microbe 3: 77–87.

66. DeRocher AE, Coppens I, Karnataki A, Gilbert LA, Rome ME, et al. (2008) A thioredoxin family protein of the apicoplast periphery identifies abundant candidate transport vesicles in *Toxoplasma gondii*. Eukaryotic Cell 7: 1518–1529. doi: 10.1128/EC.00081-08 PMID: 18586952

67. Hettmann C, Herm A, Geiter A, Frank B, Schwarz E, et al. (2000) A dibasic motif in the tail of a class XIV apicomplexan myosin is an essential determinant of plasma membrane localization. Molecular Biology of the Cell 11: 1385–1400. PMID: 10749937

4. Opening the black box of flavin-dependent metabolism in *Apicomplexa*

Cofactors are non-protein molecules that are required for biological activity of proteins, most frequently of enzymes. Most of the cofactors are derivatives of vitamins or pro-vitamins, which contain a partial or complete chemical moiety necessary for the biosynthesis of a cofactor. Enzymes in the inactive form, without necessary cofactor(s) attached, are called *apoenzymes*, as opposed to catalytically active *holoenzymes,* which contain the relevant cofactor(s). Certain enzymes and multi-enzyme complexes involve multiple cofactors. For instance, the pyruvate dehydrogenase complex requires thiamine pyrophosphate, lipoamide, flavin-adenine dinucleotide, coenzyme A and magnesium ions. Among these, there are cofactors that belong to two distinct groups – *coenzymes* and *prosthetic groups*. This distinction of cofactors denotes to the mode of their interaction with apoenzymes: when the holoenzyme exists only in the process of the reaction and easily dissociates liberating the cofactor, the latter is called a coenzyme. Coenzymes can be detected in a free form in the solution and participate in biochemical reactions in a manner similar to the reactants and products (for example, NAD, NADP, CoA, THF). Conversely, a cofactor is called a prosthetic group when it is permanently (sometimes even covalently) bound to its enzyme, forming a stable holoenzyme complex. Examples of such cofactors are biotin, lipoamide, heme, pyridoxal phosphate, flavin mononucleotide (FMN), flavin-adenine dinucleotide (FAD) etc. These molecules exist preferentially in the protein-bound form and generally are not part of the equations that describe biochemical reactions. Permanent binding of FMN and FAD to flavoenzymes is explained by their propensity to produce the reactive radicals necessary for the catalysis, yet highly deleterious for living cells if produced in an uncontrolled manner. Accordingly, recent experimental study on a human enzyme that produces FAD indicated that, upon its production, the coenzyme remains tightly bound to the enzyme and is likely passed directly to an acceptor apoenzyme rather than released in solution [120].

4.1 Metabolism of cofactors – potential source for drug targeting

Production of cofactors is a fundamental and an imperative function of metabolism in virtually any living organism. This is because activity of many metabolic enzymes strictly depends on availability of their active cofactors. Such availability is a result of either *de novo* production or salvage of necessary precursors (vitamins, pro-vitamins) with the subsequent transformation to an active form of the cofactor by phosphorylation, adenylation, oxidation, reduction, etc.

Organisms that evolved within niches which constitutively contained sufficient amounts of certain cofactor precursors (vitamins) acquired efficient uptake and storage mechanisms, while other species remain reliant on internal *de novo* production. However, it is also common that organisms possess metabolic capabilities of producing coenzymes *de novo* as well as from the precursors taken up from the surrounding milieu in a condition-specific manner. This is the case, for example, in *T. gondii* and some other *Apicomplexa,* which are capable of *de novo* synthesis and salvage of folates for the production of THF. Despite this evident redundancy, THF-dependent metabolism is a primary target of current anti-toxoplasmosis treatment. Detailed knowledge about the metabolism of folic acid and its derivatives in the parasite allowed design of a highly efficient combination therapy, where one component interferes with the *de novo* synthesis and the other with the reduction of folates, which are downstream of the salvaged precursors. In this manner, combined action of pyrimethamine and sulfonamides produces a synergistic detrimental effect on fast-replicating *T. gondii* tachyzoites. This case illustrates how detailed knowledge of a cofactor-dependent metabolism can be used for design of successful treatment strategies for apicomplexan infections. Yet many pathways and transport capabilities remain far less well studied and understood comparing to folate-dependent metabolism and potentially may hold promise for the discovery of further antiparasitic drug targets.

4.2 Role and peculiarities of flavin-dependent metabolism

Riboflavin (from the Latin *flatus* - yellow) is a polycyclic molecule, an essential part of two important cofactors – flavin mononucleotide (FMN) and flavin-adenine dinucleotide (FAD). Riboflavin and the flavin-cofactors contain an isoaloxasine moiety, a three-member, nitrogen-containing ring system, which is responsible for the redox activity (Figure 15).

Isoaloxasine can be a donor and an acceptor of one or two hydrogens and electrons, and thus, a vast majority (c.a. 90%) of flavoenzymes belongs to the oxidoreductases, with only a small part to the other classes of enzymes. Interestingly, presence of this moiety in riboflavin does not make it a redox-active cofactor. On the contrary, riboflavin is considered as an inert, and often a storage, form of flavins. FAD-dependent enzymes are generally more abundant than FMN-dependent ones, 75% and 25% respectively [121]. Nevertheless, it is thought that the AMP moiety, which differs between these two cofactors, does not play a direct role in catalysis, but only in stabilization of the protein-cofactor complex [122]. Despite the covalent binding to flavins being observed only in about 10% of the flavin-dependent enzymes, in all cases FMN and FAD are firmly attached to the proteins [121,123]. Consequently, they belong to prosthetic groups and not coenzymes as for example NAD and NADP.

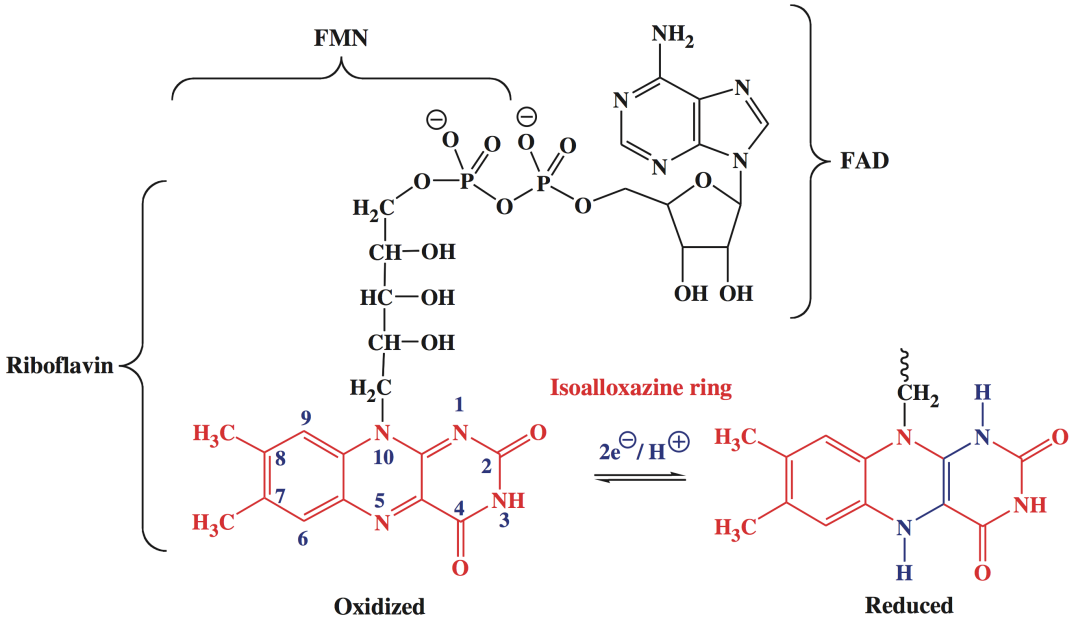


Figure 17. Structure of FAD, FMN and riboflavin with an oxidised and reduced form of the isoaloxasine moiety (adapted from [121])

The percentage of genes related to flavin-dependent metabolism varies between species and ranges from less than 0.1% to over 3% (Figure 16B) of all enzymes being dependent on either FMN or FAD [124]. *T. gondii* has a similar amount of flavoenzymes as *Homo sapiens,* however, in relative proportion to the overall number of genes, this constitutes nearly a 3-fold higher relative abundance of them in the parasite (Figure 16A). Both *T. gondii* and *P.*

falciparum seem to moderately rely on flavin-dependent enzymes, while some species reduced their repertoire of flavoenzymes to the most basic set, consisting of a dozen proteins (e.g. *Pyrococcus abyssi*, *Thermotoga maritima*).

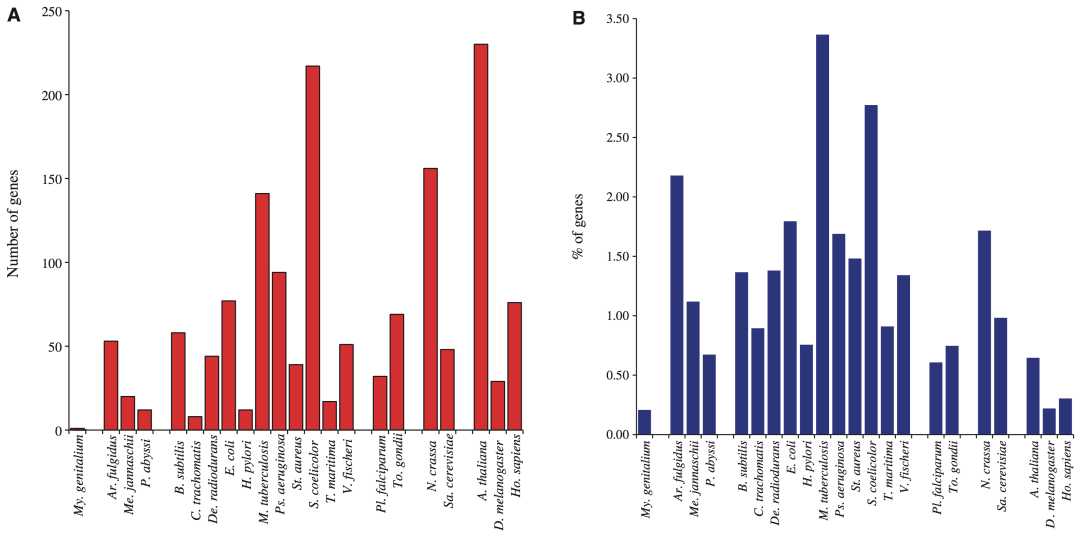


Figure 18. Distribution of the apparent number of flavoenzymes in different species (adapted from [121])

(A) number of all flavoenzymes annotated in genomes of different species;
(B) percent of flavoprotein-coding genes in the total number of genes in the organisms

Flavin-dependent metabolism is clearly an important part of the overall metabolic network in *Apicomplexa* and is present across at least three organellar compartments (Table 3). Therefore, interference with the metabolism of flavin cofactors should be expected to have a strong and multi-target effect on the parasites.

Among the flavoenzymes listed at the Table 3 of a special interest is ferredoxin reductase (FNR), an enzyme implicated in apicoplast-localized MEP/DOXP pathway of isoprenoid biosynthesis. The action of this enzyme is dependent two redox cofactors: FAD and NADP [125]. In photosynthetically active plastids, ferredoxin and FNR serve for electron transport from photosystem I to NADP, but reverse transfer of the reducing power is also possible and plays a key role in non-photosynthetic organelles such as apicoplast [126]. Unexpectedly, FNR and ferredoxin are also present in *Cryptosporidium* species, which lost their apicoplast, as well as the enzymes of the MEP/DOXP pathway. This may indicate that at least in these species, ferredoxin and FNR may have another, yet-to-discover role, resulting in preservation of them, even in the absence of an apicoplast.

Names of the enzymes	Metabolic subsystem	*Toxoplasma gondii*	*Plasmodium falciparum*	*Eimeria tenella*	*Babesia bovis*	*Theileria annulata*	*Cryptosporidium muris*
Cytosol							
glutathione reductase	antioxidative defense	TGME49_219130 TGME49_246920	PF3D7_1419800 (two isoforms)	ETH_00003395	absent?	absent?	absent?
thioredoxin reductase	synthesis of deoxyribo-nucleotides	TGME49_309730	PF3D7_0923800 (two isoforms)	ETH_00014915 ETH_00021540	BBOV_I002190	TA04645	CMU_002600
glycerol-3-phosphate dehydrogenase	shuttling of reducing power to ETC	TGME49_263730	PF3D7_0306400	ETH_00022360	BBOV_III000930	TA17925	absent
Mitochondrion							
dihydroorotate dehydrogenase (FMN-dependent)	synthesis of pyrimidine nucleobases	TGME49_210790	PF3D7_0603300	ETH_00004975	BBOV_II007190	TA11695	absent
E3 subunit of BCKDH and KDH	TCA cycle	TGME49_206470	PF3D7_1232200	ETH_00028950	BBOV_IV007190	TA03445	CMU_023430
malate-quinone oxidoreductase	TCA cycle	TGME49_288500	PF3D7_0616800	ETH_00035175	BBOV_III000580	TA18100	CMU_024300
succinate dehydrogenase	TCA cycle	TGME49_215590	PF3D7_1034400	ETH_00013335	BBOV_IV007210	TA03455	CMU_043190
protoporphyrinogen oxidase	heme biosynthesis	TGME49_272490	PF3D7_1028100	absent	absent	absent	absent
acyl-CoA dehydrogenase [127]	β-oxidation of fatty acids	TGME49_315480 TGME49_247500	absent	ETH_00032045 ETH_00015865	absent	absent	absent
glutathione reductase	antioxidative defense	TGME49_219130 TGME49_246920	PF3D7_1419800 (two isoforms)	ETH_00014915 ETH_00021540	absent?	absent?	absent?
Apicoplast							
E3 subunit of PDH	FAS II	TGME49_305980	PF3D7_0815900	ETH_00041205	absent	absent	absent
NADP-dependent ferredoxin reductase [125]	biosynthesis of isoprenoids (MEP/DOXP)	TGME49_298990	PF3D7_0623200	ETH_00024565	BBOV_IV011290/ BBOV_I000680	TA09580 TA15585	CMU_026730

Table 3. List of the flavoenzymes in *Apicomplexa* related to central metabolic pathways

This comparasitve table is based on the functional annotations of the respective apicomplexan genomes in EuPathDB database [50]

Plants, yeast and most prokaryotes can produce riboflavin *de novo,* while many animals – including humans – lack this ability and thus they rely on acquisition of riboflavin from nutrition. Similarly to animals, apicomplexan parasites are also riboflavin auxotrophs and are thought to salvage this vitamin from their host cells. Notwithstanding, *T. gondii* possesses the pathway for *de novo* synthesis of molybdopterin, a coenzyme structurally reminiscent of riboflavin and similarly implicated in oxidation/reduction processes (e.g. the enzyme sulfite reductase, TGME49_295720, is molybdopterin-dependent).

Production of the flavin cofactors from an exogenously supplied riboflavin requires several enzymatic steps. At the first step, riboflavin is phosphorylated by riboflavin kinase (RK) and becomes a redox-active cofactor – flavin mononucleotide (FMN). Subsequently, FMN adenylyltransferase (FMNAT) attaches an AMP moiety to FMN, forming flavin-adenine

dinucleotide (FAD), the second and the most common flavin-cofactor (Figure 17). Although both of the reactions use ATP, in the case of RK only phosphate is attached and ADP moiety is released as a product of the reaction, while FMNAT consumes AMP moiety to produce FAD, releasing pyrophosphate as a second product.

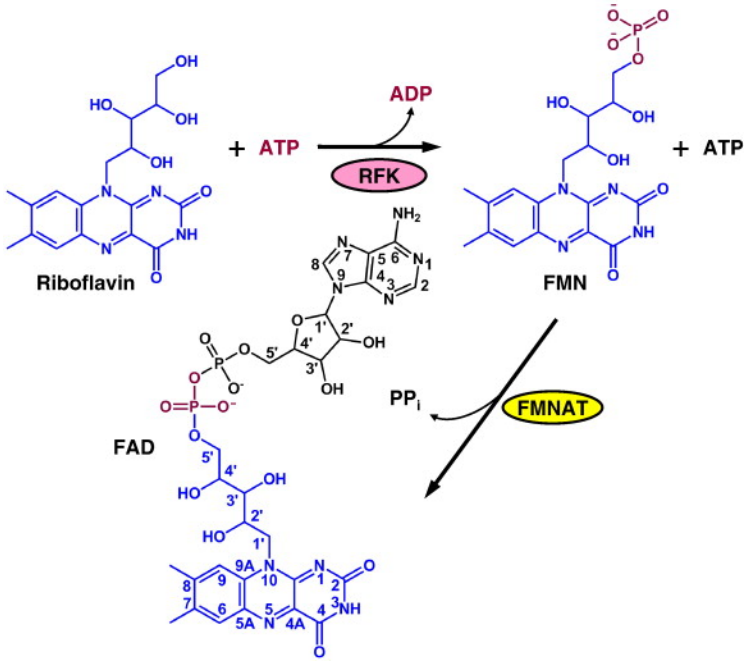


Figure 19. Transformation of riboflavin into active cofactors FMN and FAD through two enzymatic steps (adapted from [128]).

Unlike the reaction catalyzed by FMNAT, riboflavin kinase reaction is generally considered as irreversible [129]. It has been proposed that due to the concentration of PPi in intracellular conditions, FMNAT can also proceed only in its forward direction, as shown in the figure. Hence, for both reactions enzymes, that catalyze essentially the reverse reactions exist: namely FMN hydrolase (acid phosphatase) and FAD diphosphatase (FADase, hydrolyses FAD to FMN and AMP); shown on the Figure 18. Abbreviations: RFK – riboflavin kinase, FMNAT – FMN adenylyltransferase.

In eubacteria, RK and FMNAT enzymatic steps are commonly carried out by a single bifunctional enzyme, which has been proposed as a novel antimicrobial drug target [130]. This enzyme is often referred to as FAD synthase (or synthetase). However, due to a rather common use of this name as a synonym of FMNAT, I will not use it here to avoid any confusion. Similar bifunctional enzymes are also present in plants, but *Archeae*, animals and fungi possess separate enzymes for producing FMN and FAD. Presence of secluded organellar compartments in eukaryotes creates the need for either targeting of these nuclear-encoded enzymes to multiple compartments or inter-compartment transport of the flavin-cofactors across intracellular membranes (Figure 18). Although FMN and FAD are important cofactors of mitochondrial oxidation/reduction reactions, localization of the enzymes that

produce them remains not fully clarified, even in some experimentally amenable model species such as *S. cerevisiae* [123]. In human cells, the studies on the FMNAT enzyme revealed multiple isoforms of the enzyme transcribed from the same gene, which are responsible for this enzymatic activity in the cytosol and mitochondria [131,132]. Dual localization of this activity was also demonstrated in rat liver cells [133,134], *S. cerevisiae* [135,136] and *Nicotiana tabacum* [137].

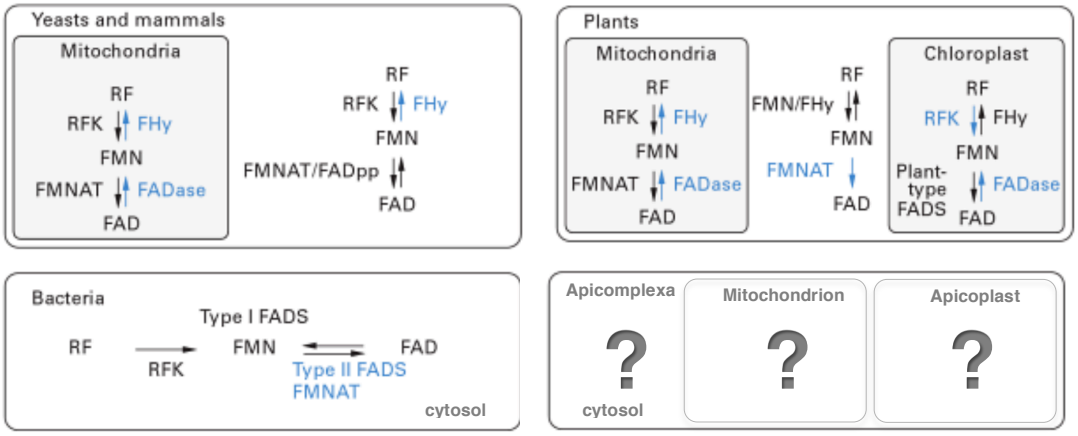


Figure 20. Intracellular localization of the flavin-cofactor producing enzymes in different species (adapted from [138]).

Abbreviations: RF – riboflavin, RFK – riboflavin kinase, FADS – FAD synthase (synthetase), FHy – FMN hydrolase, FADpp – FAD pyrophosphatase, RFK – riboflavin kinase, FMNAT – FMN adenylyltransferase.

In *Apicomplexa*, synthesis of FMN and FAD has not been studied in detail, despite their important role for a number of key metabolic pathways in the parasites. Furthermore, no transporter responsible for riboflavin uptake has been annotated in the genomes of *Apicomplexa*. Nevertheless, the apparent absence of the genes for its *de novo* synthesis suggests that a yet-to-discover uptake mechanism(s) for riboflavin is present. As for the intracellular transport of flavin-cofactors, it also remains elusive which protein(s) are responsible for this function. Our BLAST-search for the proteins similar to the yeast putative mitochondrial carrier of FAD (FLX1) revealed a significant match in *T. gondii* genome currently annotated as “mitochondrial carrier superfamily protein”. Apart from the matching region, the protein encoded by the gene TGME49_228680 has large extensions both on the N-terminal and C-terminal part as compared to the yeast counterpart (Figure 19). Indeed, as any sequence-based prediction, this match only represents a testable hypothesis that needs an experimental validation.

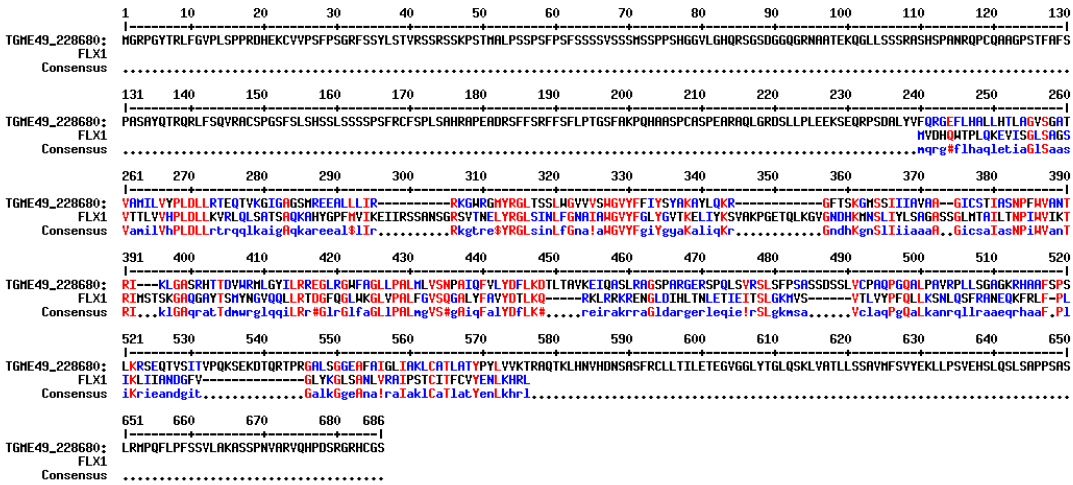


Figure 21. Alignment of the putative yeast mitochondrial carrier of FAD (FLX1) and *T. gondii* candidate TGME49_228680 annotated as "mitochondrial carrier superfamily protein"

In this study, I collected, produced and analyzed *in silico* and *in vitro* results to contribute to a clearer understanding of FMN and FAD-dependent metabolism in apicomplexan parasites using *T. gondii* as an experimentally amenable model species.

4.3 Subcellular localization and essentiality of FMN and FAD production in *T. gondii*

As highlighted by reconstruction and analysis of ToxoNet1, flavin-cofactors are necessary for metabolic processes in multiple sub-cellular compartments: at least in the cytosol, mitochondrion and apicoplast. Sequence-based predictions from Mitoprot [139] and TargetP [140] suggested presence of signal peptides in the genes encoding riboflavin kinase (TGME49_216740) and FMNAT (TGME49_214280), which was likely present there to target these enzymes to an organellar compartment. MitoProt [139] and ApicoAP [141] algorithms predicted that those bipartite signal peptides are leading the enzymes to the apicoplast, while TargetP predictions were in favor of the mitochondrial localization for both enzymes. Comparison of the RK and FMNAT sequences with the human orthologous enzymes revealed extra-large N-terminal extensions in both proteins, which were moderately conserved when comparing to a very closely related parasite *Neospora caninum* (Figures 20 and 21). Alignments of protein sequences here and thereafter were obtained using a web-interface of the Multalin algorithm [142].

Figure 22. Alignment of the sequences of TgRK with the orthologous genes of closely related pathogen *Neospora caninum* and human RK.

Sequences belonging to *T. gondii*, *N. caninum* and *H. sapiens* are labeled as T, N and H respectively. BLASTp-algorithm detected no putative conserved domains at the stretch of the amino acids preceding the sequence matching to the human enzyme; apparently, the sequence of this extra-large N-terminal extension is highly unique to *Toxoplasma*, *Neospora* and *Hammondia* spp.

In the ToxoNet1, we allocated both of the enzymes to the mitochondrion because of the apparent presence of N-terminal extensions, TargetP predictions, and the fact that mitochondria contain bulk of the flavin-dependent enzymes. Probing the model with metabolic tasks indicated its incapability to produce FAD in the mitochondrial compartment despite the presence of the whole set of reactions leading from uptake of riboflavin to FAD. In contrast, the task of production of mitochondrial FMN was passed. This was the first indication of an apparent gap in our current knowledge on how flavin-cofactors are actually produced in *Apicomplexa* and how the need for them in multiple compartments is satisfied. Detailed investigation of this issue indicated that net production of FAD in mitochondrion required a net transport of AMP, ADP or ATP to this compartment. Consistently with the fact

that such a transporter had not been reported in *Apicomplexa*, the corresponding transport reaction was absent in the ToxoNet1. A mitochondrial ATP/ADP exchanger present in the parasites (TGME49_300360) cannot serve this purpose, as it does not result in a net transport of adenylate moieties across the membrane. Therefore, it is more likely that FMNAT produces FAD in cytosol, where the need for AMP moieties can be met by the purine salvage pathway. FAD can then be transported to the mitochondrion using FLX1-like carrier of flavins or a similar transporter. This represents a model-derived hypothesis that could not be deduced from the structure of the metabolic network alone and only formulated upon reconstruction and analysis of the compartmentalized metabolic model.

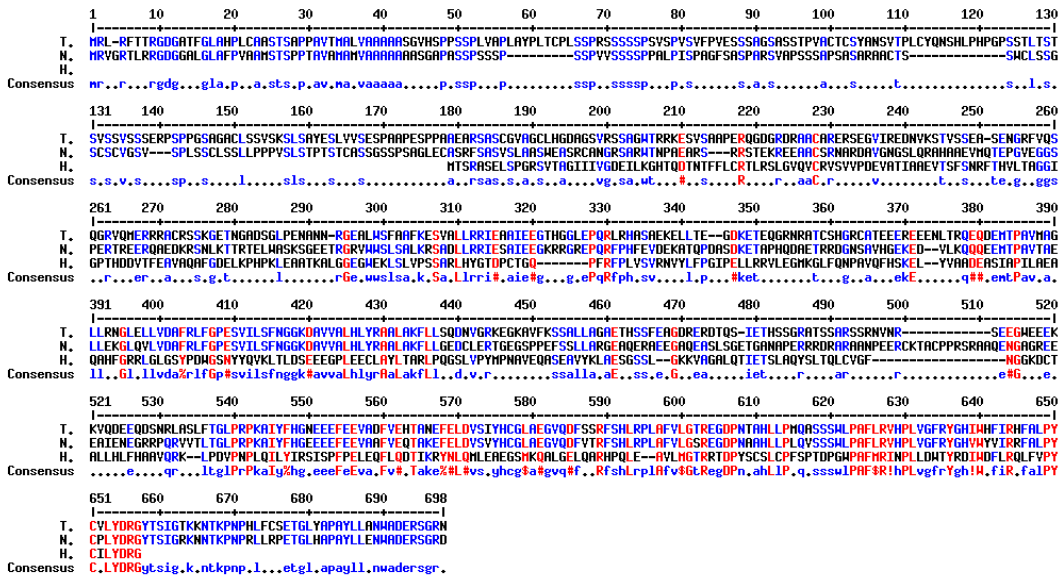


Figure 23. Alignment of the sequences of *TgFMNAT* with the orthologous genes of closely related pathogen *Neospora caninum* and human *FMNAT*.

Sequences belonging to *T. gondii*, *N. caninum* and *H. sapiens* are labeled as T, N and H respectively.

4.3.1 Localization and essentiality of TgRK

To assess subcellular localization of TgRK and TgFMNAT enzymes, we experimentally sought to modify corresponding genomic loci so that C-terminal parts of the proteins contained Ty-tag sequence. These tags, specifically recognized by anti-Ty antibodies, provide means to detect the proteins of interest by Western blotting (WB) and indirect immunofluorescence assay (IFA), as discussed in the following sections. Further, we attempted to knockdown and knockout FMN- and FAD-producing enzymes to validate predictions of ToxoNet1 on their essentiality.

We first used the two-step strategy for C-terminal tagging and subsequent knockdown of TgRK (essentially, as described in the previous chapter for ACS and ACL double knockout). This strategy appears to be especially suitable for studying the genes that might be essential. At the first step, the transfection plasmid is integrated in the gene of interest (GOI) such that the gene remains intact and can be expressed as in the wild-type strains. This step confirms that the GOI can be efficiently targeted and changed in the 3'-end; it also gives a possibility to attach tag-sequence at the extreme C-terminus of the protein for elucidating its subcellular localization (IFA) and expression level (WB). At the second step, part of the construct can be selectively excised leading to degradation of mRNA transcribed from the GOI and consequently knockdown of its expression.

We amplified a 3'-fragment of the gene encoding RK (c.a. 1100 nt) by PCR using primers 4869 and 4870 (see Appendix 4). This fragment subsequently was cut by endonucleases (SbfI and KpnI) and ligated with the pG152 plasmid containing 3 Ty-tag sequences, loxP-sequence, 3'-UTR of *sag1*, HXGPRT selection cassette, a second loxP-sequence, and U1-degradation sequence (Figure 22A). This region of homology to the 3'-end of the *TgRK* promoted integration of the plasmid in the targeted locus. The plasmid was linearized using NsiI enzyme, resulting in a single homology arm of c.a. 300 nt. Linearized plasmid was transfected into *Ku80ko HXGPRTko* tachyzoites of *T. gondii* by electroporation, essentially as described before [143]. After 24 hours of cultivation in a monolayer of HFF cells using DMEM medium with glucose and 5% FCS, the medium was replaced by fresh medium with mycophenolic acid (MPA) and xanthine added. Addition of the MPA results in an efficient inhibition of IMP-dehydrogenase activity in the parasites, which can be rescued by HXGPRT activity present as a selection marker on the plasmid. Consequently, only the parasites, which successfully integrated the plasmid *in cis* could survive prolonged (c.a. 10 days) cultivation on this selection medium. Following the selection period and full lysis of the host cells, we infected a 96-well plate with serial dilutions of the parasite suspension. 10 days later from this plate we selected wells with a single plaque formed by one parasite cell to obtain several pure clonal populations. Using PCR analysis (with the primer pairs 4950-p30a, 4851-M13F, 4950-4951, Figure 22B) with the genomic DNA extracted from these clonal populations we identified the clones which had successfully integrated the plasmid in the correct locus, as confirmed by PCR products of the expected sizes.

Further, we did an IFA to identify sub-cellular localization of the RK enzyme. For this we grew the parasites on a monolayer of HFF cells attached to thin glass slides. 24 hours post

infection, we fixed and permeabilized the cells with paraformaldehyde and glutaraldehyde treatment and applied corresponding primary antibodies (α-Ty for detection of RK, α-HSP70 as a mitochondrial marker and α-GAP45 to visualize parasite pellicle). Secondary antibodies against α-Ty (anti-mouse), α-GAP45 and α-HSP70 (both anti-rabbit) were bound to fluorescent dyes and allowed optical visualization of the corresponding proteins. IFA on the parasites with Ty-tagged TgRK repeatedly failed to detect any significant fluorescence above the background level. However, the hypothesis that the protein was in very low abundance was not concurred by immunoblot, which instead revealed at least two major bands (Figure 22C). One of the bands had an expected size of about 130 kDa, although another one had a molecular weight of ca. 55 kDa. This could be considered as an evidence of potentially two isoforms of TgRK one with and another without an extra-large N-terminal extension.

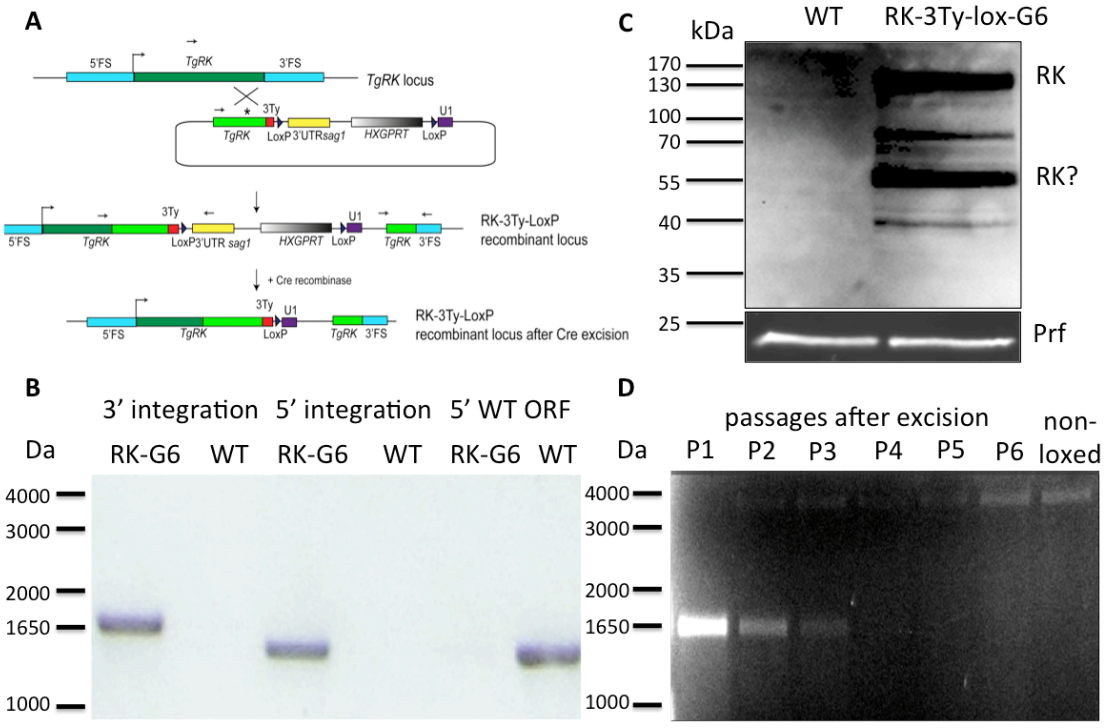


Figure 24. Two-step strategy used for C-terminal tagging-knockdown of the TgRK

(A) Wild-type and modified loci of TgRK, integrated (RK-G6) and excised construct; (B) PCR confirming correct integration of the construct in RK-G6 clone; (C) Immunoblot of total lysates from *Ku80ko* parasites (WT, control) and the clone with tagging/knockdown construct integrated; Western blot was performed using anti-Ty antibodies with anti-TgProfilin (Prf) as a loading control; (D) PCR on the gDNA extracted from the parasites prior to the transfection with Cre-recombinase encoding plasmid (RK-G6, control) and at each passage after the transfection; two independently transfected populations were controlled (with the primers 4950-M13F).

Despite the inability to detect TgRK by IFA, we could still perform the second step of the strategy – excise a part of the construct and study the effect of the gene knockdown. For this,

we transiently transfected the parasites with a plasmid expressing Cre-recombinase enzyme, which led to recombination and excision of the part of the construct that was between two loxP-sites. The transfected pool of the parasites was cultivated on the DMEM medium with glucose and 5% FCS until full lysis of the monolayer of host cells. Extracellular parasites were resuspended and a 96-well plate was infected with the serial dilutions of the suspension. We also extracted genomic DNA for PCR analysis and confirmed that the pool consisted of parasites with the excised loci and, undetectable levels of those with non-excised loci. Further, we performed the same extractions at each passage of the pool and observed slowly weakening signal of the excised locus (Figure 22D). The same PCR analysis performed on the individual clonal populations isolated from the pool did not detect any populations with the excised loci. This strongly suggests that the excision event and the knockdown of the *TgRK* expression was deleterious for the parasites and only the parasites with non-excised loci survived the prolonged cloning process. To further confirm that the knockdown of the *TgRK* is detrimental we used the Cre-recombinase expressing plasmid, which was also encoding green fluorescent protein (GFP). Presence of this fluorescent marker allowed us to avoid the time-consuming step of cloning by using cell-sorting (FACS, see Appendix 3) and select for analyses only the viable parasites that expressed the Cre-GFP plasmid. Consistently with our previous results, this approach also resulted in detection (by PCR with the 4869-M13F and 4950-M13F) of only clonal populations with non-excised *TgRK* locus. Although this strategy likely results in a knockdown rather than a knockout of the gene expression, our results are consistent with essentiality of TgRK in *T. gondii* tachyzoites. If the essentiality is confirmed by a direct or an inducible knockout this enzyme can be considered as a candidate target for development of new anti-toxoplasmosis treatment.

In order in investigate further the possibility of multiple TgRK isoforms being targeted to several sub-cellular compartments, we used a fractionation assay (Appendix 3). We prepared and analyzed by immunoblot total cell lysates of the parasites with C-terminally-tagged RK, as well as their cytosolic and organellar fractions separately (Figure 23). No signal could be detected by anti-Ty antibodies in the organellar fraction in two independent trials performed with different concentrations of the detergent. However, a strong indication of cytosolic localization is apparent. Both the upper band of the size corresponding to TgRK and previously observed (Figure 22C) lower band were detected only in the total lysates and cytosolic fractions.

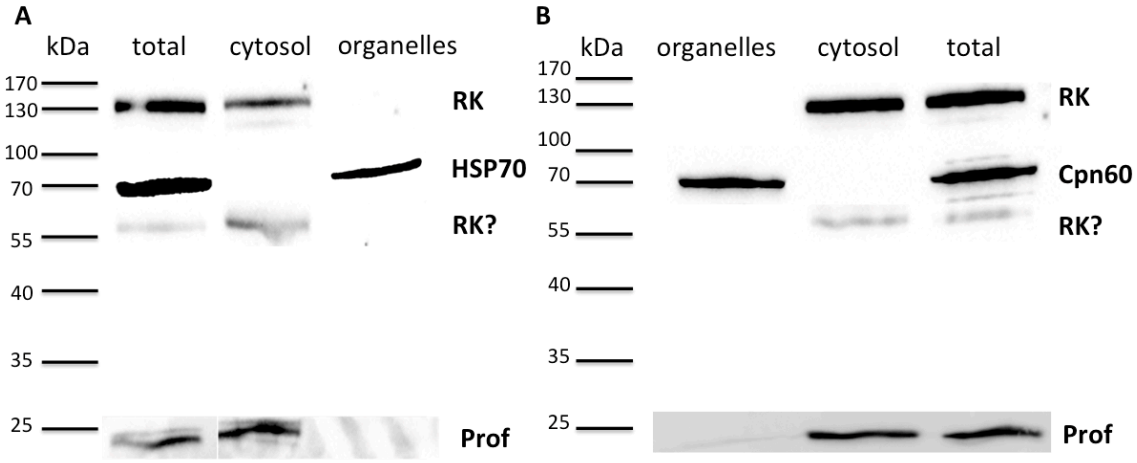


Figure 25. Fractionation assay detects no indications of organellar localization of TgRK. Extracellular tachyzoites were treated with digitonin to permeabilize their membranes and obtain soluble cytosolic fraction of proteins (by a short-term exposure to the detergent) separately from the organellar fraction (protected by additional membranes) isolated by centrifugation. Prolonged treatment with the detergent resulted in permeabilization of all the membranes and yielded a total extract. Anti-Ty antibodies were used to detect tagged RK enzyme; anti-HSP70 and anti-Cpn60 – as markers of successful separation of the organellar fraction; anti-profilin (Prof) – as loading control for the total and cytosolic fractions (see Appendix 3).

These results all but confirm cytosolic localization of *TgRK*, however also raise further questions: (1) why there is an extra-large N-terminal extension in a cytosolic protein? (2) why there is more than one band reproducibly detected by anti-Ty antibodies? Additional studies are necessary to address these questions and ultimately solve the puzzle of N-terminally extended cytosolic enzymes in *Apicomplexa*.

4.3.2 Localization and essentiality of TgFMNAT

The two-step knockdown strategy described above for *TgRK* repeatedly failed when applied to *TgFMNAT*. Due to an unknown reason we could achieve integration the pG152-based construct at the 3'-end of the ORF (3'-fragment of TgFMNAT was obtained by PCR in the same way as for TgRK, but using the primers 4871-4872). Repeatedly, after the selection period, all the isolated clones had PCR products (obtained using primers 4952-p30a, 4953-M13F, 4952-4953) inconsistent with a correct integration of the plasmid. To overcome this issue, and in view of the failure to detect TgRK by IFA, we employed two alternative and separate strategies for knockout and localization of the enzyme encoded by *TgFMNAT*.

To elucidate sub-cellular localization of the enzyme, we produced a mutant strain with an additional copy of the gene encoding a C-terminally tagged version of the protein. We first used RT-PCR to amplify cDNA encoding TgFMNAT (with the primers 4993-4994). We obtained a single PCR product of an expected molecular weight, and digested it with EcoRI and NsiI enzymes. We then ligated this cDNA with the knockin plasmid containing the

pTub8-promotor immediately upstream of the *TgFMNAT* cDNA sequence and 3 Ty-tags, as well as a downstream HXGPRT selection cassette. This plasmid was amplified in competent bacteria, isolated, and checked by sequencing for absence of mistakes in the cDNA region. Subsequently, we transfected *T. gondii* tachyzoites with this plasmid by electroporation (Appendix 3).

IFA of the transiently transfected pool of the parasites showed significant level of cytosolic second-copy tagged FMNAT (Figure 24), which did not co-localize with the fluorescence from the mitochondrial marker. These results indicate cytosolic presence of the FMNAT, yet does not allow us to draw conclusion about the organellar localization of the enzyme.

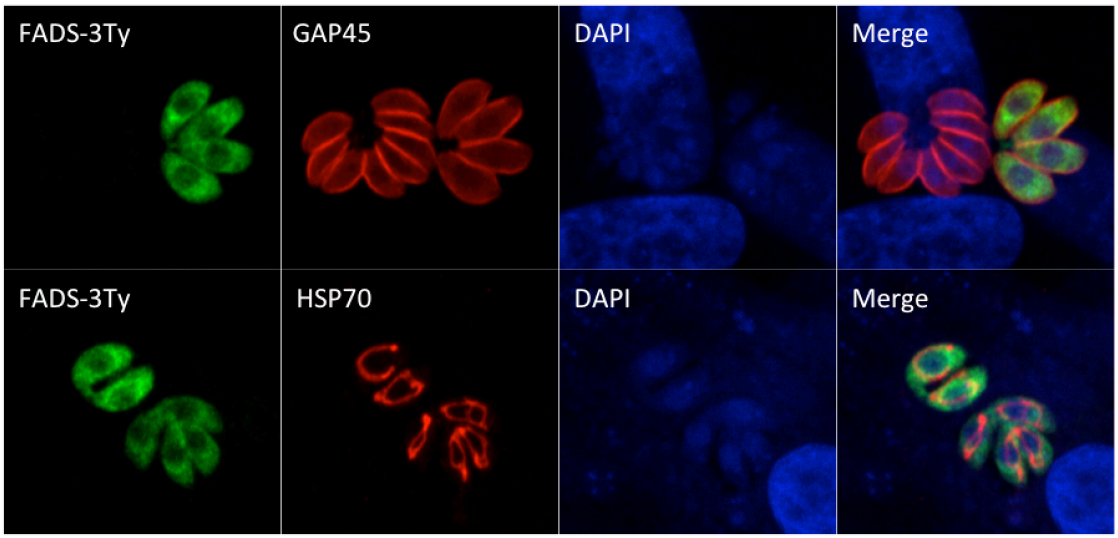


Figure 26. Cytosolic localization of the transiently expressed second copy of TgFMNAT

IFAs were made using anti-Ty, anti-GAP45 (pellicle), anti-HSP70 (mitochondrion) antibodies and DAPI (nucleus).

To rule out possible artefacts in the localization of the enzyme second copy due to high expression levels, we isolated from the pool 12 clonal populations of the parasites using the HXGPRT selection process. For this, we cultivated transfected pool of the parasites on HXGPRT selection medium starting from 24 hours after transfection. A selection period lasting c.a. 10 days ensured that the parasites resistant to MPA was not due to HXGPRT expressed *in trans* but *in cis* (integrated in the nuclear genome). We then isolated clonal populations of the parasites on a 96 well plate by cultivation of the limiting dilutions of the parasite suspension obtained after the selection process.

Unexpectedly, IFA of the clonal populations showed very low (and occasionally moderate) abundance of the tagged protein in most of the parasites. Expression levels were also variable

between the parasites within the clonal populations. This may indicate that prolonged overexpression of *TgFMNAT* causes a deleterious effect for the parasites and hence we can only observe viable cells, which inhibited the over-expression of the enzyme. Notably, overexpression of the gene encoding RK, the preceding enzyme, was observed to suppress the growth of *S. cerevisiae* [144]. However, we cannot rule out the possibility that this phenomenon could be related to the one we observed for the overexpression of *TgFMNAT*. Based on the results of the IFA, we cannot draw conclusions about the absence of the TgFMNAT in the non-cytosolic compartments. Western blot results confirmed very low abundance of the enzyme, even in the case of *in trans* expression driven by pTub8 promotor. Otherwise, a fractionation assay could be instrumental in resolving the possibility of mitochondrial and apicoplast localization of the enzyme in addition to the cytosol.

We also assessed the possibility of *TgFMNAT* translation starting at a site different to the one annotated in ToxoDB [144]. An extremely large N-terminal extension of *TgFMNAT* could be a result of mis-annotation of the starting cite in the ORF. Importantly, correct localization of the second copy of *TgFMNAT* to the relevant subcellular compartment(s) was heavily relying on the fact that the transcription start site had been assigned correctly. We performed PCR analysis of the N-terminal part of the *TgFMNAT* using the forward primer upstream of all the stop codons. As reverse primers, we used one in the middle of the ORF and at the end of it (primers 5104-4994 and 5104-4995). We sequenced obtained PCR products with the forward primer 5104 used for the PCR. The analysis of sequencing results suggested that the start codon annotated as the start of translation in ToxoDB was likely correct. However, there was also a possibility that the subsequent start codon could be an alternative transcription start. Interestingly, sequence-based localization predictions by TargetP are very different for the two proteins produced from these two transcription start sites. The longer polypeptide (680 amino acids, as it is annotated in ToxoDB) is predicted to have mitochondrial-targeting peptide (score 0.559 using plant networks and 0.843 with the non-plant ones), while the putative shorter one (648 amino acids) is predicted to be targeted to the plastid (score 0.831, plant networks). Presence of the N-terminal extensions in mRNA transcribed from *TgRK* and *TgFMNAT* was independently confirmed by RNA-sequencing results obtained by Sunil Kumar Dogga (personal communication, University of Geneva). Thus, this unusual extra-large extension is unlikely a rudiment of incorrect annotation of the ORF and possibly plays a specific role, which remains to be discovered. A similar kind of an extension had been previously reported in thiamine pyrophosphokinase of *P. falciparum* and, as in our case, this

enzyme appeared to be cytosolic [145].

To assess essentiality of *TgFMNAT* we used CRISPR/Cas9-mediated genome editing technique [146] (conceptually shown on the Figure 25). The strategy used in this study, albeit similar to the one we used for introducing a frame-shift mutation in a TgFASI-coding gene (see the previous chapter), has an important distinction. In this case, we exploited homologous recombination to introduce DHFR-selection cassette in the place where a double-strand DNA break was made by Cas9. In order to promote this site-specific insertion, we introduced 20 base-long homology regions on each side of the insert (Appendix 4). The DHFR-insert had a promotor in the opposite direction to the *TgFMNAT* ORF, in order to disrupt transcription of the gene once this selection cassette was integrated.

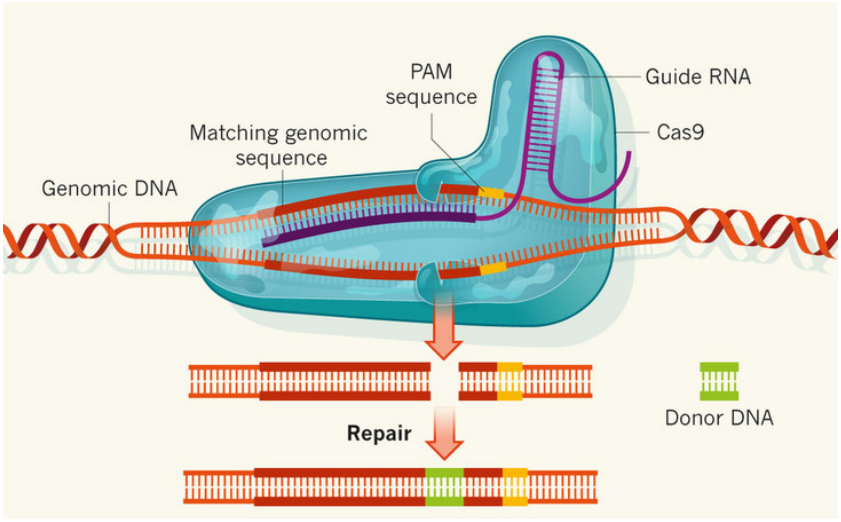


Figure 27. Illustration of the concept of CRISPR/Cas9 methodology (adapted from [147])

Wild type RH parasites were transfected with 8 µg of the CRISPR/Cas9 vector containing gRNA sequence specific to *TgFMNAT* (4997, Table 7, Appendix 4) together with 30 µg of the DHFR selection cassette with two 28 nt homology arms. We used GFP-fluorescence of the CRISPR/Cas9 plasmid to apply cell sorting after the transfection and so select only those in which the plasmid was expressed. 48 hours after transfection, GFP positive parasites were sorted by flow cytometry and cloned into 96-well plates using a Moflo Astrios (Beckman Coulter) FACS-machine. After cultivation of such parasites on the DMEM medium with pyrimethamine (0.25 uM), we isolated 6 clonal populations of the parasites which survived the selection process. To confirm that integration of the cassette occurred in the *TgFMNAT* coding sequence, we performed PCR analyses with the primers specific to this region (2087-

4996 and 2018-4872). In all the analyzed clones, PCR repeatedly failed to amplify a product that would confirm integration of the selection cassette in the *TgFMNAT* ORF. This suggests that the disruption of the *TgFMNAT* sequence was highly deleterious or lethal for the parasites and only those that repaired double strand break without integration of the selection cassette remained viable.

Based on the presented evidence, we speculate that *T. gondii* is sensitive to the expression level of the *TgFMNAT* and knockout of the gene encoding this enzyme is apparently lethal. *T. gondii* is one of the most experimentally amenable organisms among the *Apicomplexa*, yet even in this parasite studying the genes and enzymes responsible for the production of FMN and FAD proved to be very challenging. In the presented studies we admittedly did not exhaust all the available approaches and further efforts are necessary to substantiate and expand the body of evidence we obtained.

4.4 RK and FMNAT – putative targets against apicomplexan infections

Many animal species, such a mammals, solely rely on uptake of riboflavin from the diet, while other organisms (e.g. plants, yeast) are capable of *de novo* production from GTP and ribulose 5-phosphate. However, in the both cases enzymatic action of RK and FMNAT remains indispensable for production on the flavin-cofactors FMN and FAD. Accordingly, *S. cerevisiae* (riboflavin prototroph) becomes inviable upon knockout of either of these enzymes [148]. RNAi-mediated knockdown of FMNAT in *C. elegans* (riboflavin auxotroph) causes severe phenotypic abnormalities instigated by reduced levels of FMN, FAD, ATP, drastically perturbed proteome, and increased oxidative stress [149].

Apicomplexa, similar to their vertebrate hosts, are riboflavin auxotrophs, and as such rely on the salvage of this vitamin form the host cells. Thus, selective targeting of the flavin-dependent metabolism in the parasites requires precise knowledge about the dissimilarities in this aspect with the host cells. So far, there are two mechanisms known for antimicrobial action of riboflavin analogs. Firstly, a direct interference with flavoenzymes when they bind a riboflavin analog that is not catalytically active and thus lose enzymatic activity. The permanent nature of binding flavin-cofactors favors of such a mode of inhibition. However, the inhibitor should be designed to selectively bind the flavoenzymes of the parasites and not the human enzymes. Another mechanism of action is interference with FMN-dependent

riboswitches, which regulate metabolism of flavin nucleotides in some bacteria and fungi [150]. These riboswitches are sensitive to the concentration of FMN in cells and suppress *de novo* production of the riboflavin when FMN concentration is high. Riboflavin analogs, such as the naturally occurring antibiotic roseoflavin, can be recognized by riboswitches and lead to suppressed riboflavin production in addition to the first mode of action, described above. While riboswitches are not known to be present in *Apicomplexa*, successful growth suppression through interference with the flavin-dependent metabolism had been reported in several parasites of the phylum as discussed below.

Riboflavin deficiency is long known to have a protective effect against malaria in humans [151]. It has also been shown that the red blood cells infected with *P. falciparum* have a significant increase in the uptake of riboflavin from the medium, the magnitude of which was proportional to parasitemia [152]. This is consistent with the hypothesis of riboflavin auxotrophy of malaria parasites and *Apicomplexa* in general.

The antimalarial action of riboflavin analogs was first proposed in 1946 [153]. Several decades later, suppression of the malaria caused by *Plasmodium vinkey* was reported *in vivo* upon administration of synthetic 10-(4'-chlorophenyl)-3-substituted flavins without an apparent toxicity for the animals [154,155]. As for the other *Apicomplexa*, Graham et al. reported an anticoccidial effect of 5-deazariboflavin and several other riboflavin analogues [156], some of which were subsequently patented as effective inhibitors of *Eimeria* infection for poultry [157]. A profound inhibitory effect of 5-deazariboflavin and other flavin-derivatives was also observed *in vitro* in blood-stage cultures of *P. falciparum* [158,159].

While these scarce studies provide clues on potential suitability of flavoenzymes as drug targets in *Apicomplexa*, our study presents the preliminary endeavor for a more global understanding of the extent and sub-cellular localization of flavin-dependent metabolism in the parasites. Furthermore, herein we report for the very first time genetic evidence of essentiality of FMN and FAD-producing enzymes in *Apicomplexa*. These results imply that it is unlikely that the salvage of FMN or FAD from host cell is possible and thus RK and FMNAT may represent promising antiparasitic drug targets.

5. Outlook and future prospective

This thesis describes the studies on metabolism of the apicomplexan parasites *P. falciparum* and *T. gondii* using a combination of systems biology and molecular biology approaches. At the beginning of these studies genomes of multiple strains of those parasites were sequenced, annotated and publically available through EuPathDB online repository [160]. Sequencing and annotation of the genomes, among other insightful applications, scratched the surface of the metabolic capabilities of the parasites. Further, based on the genomics data and a body of primary literature, several genome-scale metabolic models of *P. falciparum* had been reconstructed before the beginning of this project [26,27]. On the contrary, no such model existed at the time for metabolism of *T. gondii*. Thus, the aims of this project were to utilize the state-of-the-art approaches to improve and extend the scope of the existing metabolic models of *P. falciparum* and to reconstruct *de novo* a metabolic model of *T. gondii.*

Genome-scale metabolic models have various applications focused on thorough analysis of possible distributions of metabolic fluxes in species of interest. As metabolism maintains a number of vital functions in living cells, understanding of metabolism of pathogenic species can serve as a foundation for development of new anti-infective medicines. Various types of metabolic modeling and analyses exist to represent and explore metabolic needs and capabilities of living cells. In this thesis I focused on a genome-scale approach as particularly GSMs were shown to be suitable for predicting essentiality of genes and enzymes. A major advantage of a genome-scale approach is a holistic manner of the analysis, encompassing a maximum of the metabolic capabilities experimentally observed or inferred to be present in the specie. As a downside, genome-scale approach requires a number of simplifications, scope of which slowly shrinks with the development of innovative computational and experimental methods. In this thesis I applied two of state-of-the-art computational methods to address simplifications made in *P. falciparum* models that concern reaction uni-/bidirectionality and functional annotations of the genome. I also used one of these methods to facilitate reconstruction of ToxoNet1, the first genome-scale metabolic model of *T. gondii*.

In view of the rather significant challenges *P. falciparum* poses for the global health, whole arsenal of experimental and computational methods must be engaged in the major endeavor

of malaria eradication. In the first chapter of this thesis, I review evolution of the metabolic models of this parasite, which is by far the best-studied malaria species. Over the decade since the first reconstruction and analysis of a graph-based metabolic network of *P. falciparum* in 2004, we witnessed increasingly more elaborate and sophisticated systems biology approaches applied to study metabolic peculiarities of this parasite. These approaches are aimed at bridging the area of computational metabolic modeling and the area of experimental research on metabolism. Such efforts have a synergistic effect and enable deduction of additional insights, which could not revealed by computational or experimental studies alone. Although, metabolic modeling efforts in *Plasmodium spp.* are not limited to large-scale models and also include small-scale mechanistic (i.e. kinetic) models of particular pathways [161,162], I chose to focus my attention on the large-scale approaches. From the side of experimental studies, my primary focus was also on those, which produced large-scale, high-throughput datasets that are the most complimentary to the large-scale modeling.

One important learning point is that the coverage of estimated proteome and metabolome of *P. falciparum* by the existing high-throughput studies is far from being complete and exhaustive as, for example, in the case of transcriptomics studies. While the analytical challenges remain to be resolved, more and more methods are being developed for utilization of transcriptomics data as a source of additional constraints for metabolic models. This relies on an implicit assumption that there is a correlation between expression levels of the genes encoding metabolic enzymes and the fluxes through corresponding metabolic reactions. Considering the outmost importance of this assumption for development of future metabolic models I am discussing it in further details below.

Indeed, if expression of a particular enzyme-coding gene is reproducibly undetectable under certain conditions one can assume absence of a flux though corresponding metabolic reaction(s). Ideally, the fact of absence of that enzyme under given conditions should be independently confirmed by proteomics analysis. However, for the majority of metabolic enzymes non-zero level of expression of their corresponding genes can be detected across multiple conditions. In this case reaction fluxes cannot be simply blocked as in the case of no expression. This constitutes the major challenge of translating gene transcription levels (e.g. RPKM values from RNA-seq analysis) into relevant constraints for their corresponding reaction fluxes. To date, there is no universal description of this correlation and all the existing methods use different (discrete or continuous) approaches to describe it. Many of the methods rely on setting *ad hoc* thresholds, which define “highly expressed” and “low-

expression" genes. An alternative approach could be a comparison of the expression levels of the same gene across as many different conditions as possible to deduce whether the expression of this gene is high or low individually, without applying generic thresholds to all genes. Even partial proteomics and metabolomics data can further improve the accuracy of the inferred constraints by indicating whether the higher/lower expression of a gene indeed translates into higher/lower abundance of the protein and if that causes a perturbation in the stead-state concentration of the molecules this enzyme acts upon. To this end, metabolic models can serve as a scaffold for consolidation of all the available pieces of data and further analysis within the appropriate context. Such integrative studies, with multiple complementary –omics datasets used, still remain to be undertaken for *P. falciparum* and other apicomplexan species. Necessary prerequisites for such studies, however, exist, and thus, one should expect such efforts as a next logical step in the area of metabolic modeling of *Apicomplexa*.

From another perspective, incorporation of constraints based on high-throughput experimental datasets can be also viewed as an opportunity for reconstruction of context-specific rather than standard genome-scale models. Such models would simulate a particular state of metabolism under the conditions of interest, and, consequently, can be used for a rigorous analysis of particular metabolic state(s) rather then general metabolic capabilities.

As I pointed out above, a genome-scale approach is used with a number of simplifications some of which were necessary at the time the models were reconstructed and analyzed. Within the studies described at the second chapter I used two recently established computational methods to address simplifications made in the most recent at the time model of *P. falciparum*.

Several prominent examples of the simplifications currently used for the analysis of genome-scale metabolic models are: (1) quasi-steady state approximation that allows us to analyze distributions of metabolic fluxes in the absence of the detailed knowledge about the kinetic properties of metabolic enzymes and concentrations of metabolites; (2) subcellular localization of metabolic enzymes and transportability of intracellular metabolites across subcellular compartments in the absence of reliable high-throughput methods to define these properties; (3) assumptions on uni-/bidirectionality of metabolic and transport reactions pre-assigned *ad hoc*.

The latter two may not appear to be simplifications *per se* as enzyme localizations, metabolite transportability and uni-/bidirectionality of reactions are unambiguously defined in the models. However, these definitions indeed constitute implicit simplifications as in majority of the cases they are not backed up by any conclusive evidence.

In the presented studies I attempted to address the simplification concerning uni-/bidirectionality of the reactions through estimation of their thermodynamic properties to confirm or reject possibility of bidirectionality based on the natural driving force of (bio)chemical reactions – change of the Gibbs free energy. Constraints based on such thermodynamic estimates appeared to be significantly less restrictive then the *ad hoc* assigned constraints. This largely reflects our assumption about the very wide range of metabolite concentrations used for the estimates. Accordingly, once I replaced these ranges with the experimentally measured concentrations from metabolomics dataset, additional reactions appeared to be unidirectional. Indeed, the number of metabolites, concentrations on which were measured is small when compared to the number of metabolites in the genome-scale models of *P. falciparum*. Thus, while the method is present for addressing in a rigorous way the uncertainty in uni-/bidirectionality of the reactions, more information about the actual concentrations of metabolites is necessary to eliminate the need for the *ad hoc* assignments.

Acquisition of such data requires resolution of at least three major challenges: (1) reliable technique for separation of the parasites from their host cells that excludes mixing of the parasite and host cell metabolomes; (2) metabolomics technique that enables detection and quantification of several hundreds of intracellular metabolites; (3) advanced extraction technique for metabolites, which are present in more than one subcellular compartment, to estimate their concentration in each of the compartments.

To date only *P. falciparum* was in the main focus of metabolic modeling of malaria parasites, while this disease in humans also can be caused by four other *Plasmodium* spp. (*P. vivax*, *P. ovale*, *P. malariae* and monkey malaria parasite *P. knowlesi*). No metabolic models have been developed to date for these species, which highlights an implicit assumption that the differences in their metabolic capabilities compared to *P. falciparum* are minor. *P. falciparum* is also responsible for the majority of malaria-associated mortality, thus, particularly this species is in the main focus of experimental research efforts. This, together

with the possibility of *in vitro* cultivation of the blood stage of *P. falciparum*, resulted in a significant body of primary literature being available particularly for this parasite, leaving the other species significantly less studied. Nonetheless, outside sub-Saharan Africa region malaria is predominantly caused by species other than *P. falciparum* and infections with *P. vivax* have been also reported to cause acute malaria albeit less frequently [163]. The other significant reason why research on *P. vivax* should gain more attention of the malaria research community is the ability of this species to establish dormant forms (hypnozoites) in liver of the infected humans. These dormant forms cause a long-lasting chronic infection with recurring relapses of the disease. Currently there is only one drug is known to act on the parasites at hypnozoite stage (primaquine and its derivatives) and the mechanism of its action is not well understood. Experimental research in *P. vivax* is hampered by inability to culture the parasites *in vitro* in the way similar to *P. falciparum*. Consequently, limited information we can currently access comes from the analysis of the samples of humans, who contracted infection with this parasite. Taken all together, this creates a clear need for using systems biology approaches to consolidate scarce experimental information in the form of a model, which can be further analyzed in an independent as well as comparative (with *P. falciparum*) manner to gain new insights about the metabolism in *P. vivax*. Importantly, representation of the environment in which the pathogen replicates is crucial for modeling growth of the blood stage *P. vivax* parasites. This is due to the fact that this malaria parasite can only use reticulocytes (immature red blood cells) and not erythrocytes (mature ones) as a niche for replication [164]. One of the fundamental differences between these two cell types is significantly more active metabolism in reticulocytes and presence of metabolic enzymes, which are degraded upon their maturation into erythrocytes. Ideally, to study blood stage of *P. vivax* a combined host-pathogen metabolic model should be developed to account for metabolic capabilities of not only the parasite but also of its host cell. This, however, will require reconstruction of the first metabolic model of human reticulocytes.

P. berghei is a rodent malaria parasite, which is commonly used as a model specie for *in vivo* studies. *P. berghei* infects rats and mice, and, thus significant body of experimental data is available for liver and mosquito stages, which are rather inaccessible in the case of human parasites. No metabolic model exists to date for *P. berghei*, while comparative genomics studies revealed several differences in the metabolic capabilities of rodent and human malaria species [119]. Consequently, there is a need for development of such model, which can be

either based on existing models of *P. falciparum* with necessary modifications or reconstructed *de novo* (e.g. using metabolic capabilities of the RAVEN Toolbox).

Conversely to the development of new models, there is also a clear need for reconciliation of the two existing lineages of the models that have been independently reconstructed to study metabolism of *P. falciparum*. The RAVEN Toolbox can facilitate such reconciliation efforts by producing a template metabolic network with confidence scores for every functional annotation. As an independent dataset, it would facilitate the decisions whether each of the genes and enzymatic activities present in the existing models should be retained in the consensus model or rejected as such that have no sufficient evidence to support their presence. Consensus model should ideally contain confidence scores and literature evidence, based on which each of the reactions was included in the model, set as unidirectional and localized in a certain subcellular compartment. Presence of a high-quality consensus model for *P. falciparum* will expedite development of the models for other *Plasmodium* spp. as well as other closely related haemosporidia.

It is also likely that the first libraries of single gene-knockout phenotypes will be soon created for *T. gondii* and *P. berghei*. Technologies that enable such studies are already available [146,165-167] yet remain rather tedious and expensive to be used in a high-throughput manner. Importantly, both of these technologies abolish (knockout) expression of the corresponding genes completely, unlike RNAi-based methods that only suppress (knockdown) expression of the genes to a certain extent. Only complete knockout of a gene encoding a metabolic enzyme can be directly translated to exclusion of the corresponding metabolic activity from the CBM. Reconciliation of CBMs with such upcoming genome-wide gene essentiality studies is possible using the algorithms such as GrowMatch [53]. Based on the similar studies done in *S. cerevisiae* [40], it is clear that such reconciliation will considerably improve accuracy of the predictions made by the model. On the other hand, it will also provide a set of hypotheses on potential reasons for essentiality of the genes, knockout of which was observed to be lethal or detrimental for the parasites. Moreover, the models reconciled with the single gene essentiality dataset are likely to predict more accurately the sets of synthetically essential pairs of genes, which will facilitate further exploration of the limits in the metabolic capabilities of the parasites.

In the third chapter of this thesis, I describe development of ToxoNet1 model, which illustrates how reconstruction capabilities of the RAVEN Toolbox facilitate model-building efforts. The completeness I aimed for when reconstructing the metabolic network of *T. gondii* led to identification of a number of "gray zones" in our current knowledge about the metabolic capabilities of this parasite. In all these cases, I made assumptions based on currently available information but each of these assumptions needs further targeted research for its validation. Further, under the assumptions made for the human host cell environment, I analyzed ToxoNet1 to predict minimal nutritional requirements for replication of the parasite as well as the genes which are essential and the synthetically lethal. I used two complementary approaches to study flexibility in nutritional requirements of the parasite by enumeration of all: (1) alternative precursor molecules necessary for production of each biomass building block separately; (2) alternative compositions of "*in silico* minimal media" that allow production of all the biomass building blocks simultaneously. The results of these studies contribute to better understanding of the limits in metabolic capabilities of the parasite and to obtaining a shortlist of indispensable metabolic functions to be considered as potential drug targets.

ToxoNet1 model is largely similar to the model by Song et al. [168] but our studies were focused on scrutinizing different properties of the same system using rather distinct approaches. In particular, our primary aim was to achieve reliable gene essentiality predictions. Thus, we used different assumptions on the range of accessible substrates, transportability of metabolites, and aimed for as complete as possible coverage of the metabolic capabilities. For the very same purpose we also made efforts to implement gene-protein-reactions associations in ToxoNet1 that were not present in the pervious study. Our gene essentiality predictions represent a conservative minimal set of essential genes, reactions and substrates that should be common for all the strain types. Additional, strain- or stage- specific gene essentiality predictions can be made using ToxoNet1 in future upon integration of the corresponding experimental data as further constraints (e.g. transcriptomics, proteomics, metabolomics etc.). To this end, an immediate extension of this work is the development and analysis of context-specific metabolic models for the apicomplexan parasite. ToxoNet1 can serve as a scaffold for integration of additional constraints, for

instance, the ones inferred from stage-specific transcriptomics and metabolomics data. Algorithms for such integration and analyses of the context-specific models are readily available (e.g. EXAMO [169] uses transcriptomics data and performs condition-specific assessment of gene essentiality).

Of a particular interest is reconstruction of the context-specific models that reflect metabolism of *T. gondii* on the life-stages other than tachyzoites, and, consequently, are less experimentally accessible. Transcriptomics and proteomics datasets are available for multiple life-stages of the parasite and freely accessible online in the ToxoDB database. With the reservations discussed above, this data can be used to infer context-specific constraints for ToxoNet1, which will make it representing metabolism on merozoite or bradyzoite stage. The latter one, however, will also require some careful re-evaluation of the objective function of the model, as no active proliferation is observed on bradyzoite stage. Alternative, and, potentially, multiple, objective(s) should be defined to describe the "maintenance of viability" phenotype of bradyzoites. The constraints inferred from the stage-specific transcriptomics data may also shed light on the identity of the carbon sources that *T. gondii* uses at this life-stage. Such knowledge about the pathways, which are active and essential for survival of bradyzoites, can be used for design of novel antiparasitic treatments effective against the dormant apicomplexan infection.

The workflow used for the reconstruction of the ToxoNet1 model using the RAVEN Toolbox, is in principle, applicable to any apicomplexan parasite whose genome is fully sequenced. Even in the absence of a functional annotation, a draft metabolic network can be reconstructed in a matter of weeks. Indeed, further steps, such as compartmentalization, inclusion of transporters and model refinement still remain time-consuming and laborious. The time and quality of reconstruction of the models can improve considerably when an existing model from very closely related species can be used as a template [94]. For example, ToxoNet1 can be readily used as such template for expedited reconstruction of a genome-scale CBM of *Neospora caninum* or other closely related coccidian species.

Experimental validation of hypotheses obtained through computational modeling is a necessary part of gaining confirmed knowledge. Accordingly, in the third and fourth chapters

of this thesis, several prominent model-derived hypotheses were assessed experimentally. Namely, localization and synthetic essentiality of ACS and ACL enzymes together with the localization of apparently the only acetyl-CoA transporter fill a gap of knowledge about acetyl-CoA biosynthesis in *T. gondii*. While this study provides a missing piece of knowledge anout the sources of acetyl-CoA in cytosol and ER, it nevertheless leaves several important questions open. In particular, additional studies are necessary to rule out the primary cause of the lethal outcome upon depletion of the cytoplasmic pool of acetyl-CoA. Such potential reasons can be: (1) impossibility of fatty acid synthesis through type I pathway in cytosol; (2) impossibility of the elongation of fatty acids; (3) interference with the acetylation of proteins and, particularly, histones. It also remains to be elucidated what are the sources of the substrates for ACS and ACL enzymes in cytoplasm. Could acetate and citrate be taken up from the host cell resources and if so how much such salvage contributes to the cytosolic pools of these molecules in the parasite. Finally, the reasons for accumulation of the ACL upon knockout of ACS are also not intuitive and need further targeted research.

Experimental assessment of RK and FMNAT and their corresponding genes described in the fourth chapter of this thesis contribute to better understanding of flavin-dependent metabolism in *T. gondii*. These enzymes were long assumed as potential drug targets in the different species, but until present studies had not been validated as essential in *Apicomplexa*. I demonstrated that the knockdown of RK and the knockout of FMNAT were deleterious and potentially lethal for *T. gondii* tachyzoites. This constitutes the very first genetic evidence for the essentiality of FMN- and FAD-producing enzymes in an apicomplexan parasite. Further, antiparasitic activity of natural and synthetic riboflavin analogs (e.g. roseoflavin and deazariboflavin) should be evaluated on a range of *Apicomplexa* to assess whether these enzymes can be selectively inhibited in the parasites without significant side effects for the uninfected tissues.

Taken together, these efforts illustrate the complete cycle from a model reconstruction, curation and analysis to generation of a model-based hypothesis, its experimental assessment and acquisition of knowledge, which can be incorporated by the model as well as used for practical applications.

Studies included in this thesis summarize and extend the scope of our current knowledge of metabolism in the apicomplexan parasites *T. gondii* and *P. falciparum*. Admittedly, many of the presented hypotheses and predictions remain to be assessed experimentally, due to the absence of suitable experimental resources. Validation of these model-based hypotheses, particularly on gene essentiality, will foster the efforts for the identification of novel and potent targets for the treatment of *Apicomplexa*-caused infections.

Bibliography

1. Adl SM, Leander BS, Simpson AG, Archibald JM, Anderson OR, et al. (2007) Diversity, nomenclature, and taxonomy of protists. Syst Biol 56: 684-689.
2. Klein EY (2013) Antimalarial drug resistance: a review of the biology and strategies to delay emergence and spread. Int J Antimicrob Agents 41: 311-317.
3. Worth AR, Lymbery AJ, Thompson RC (2013) Adaptive host manipulation by *Toxoplasma gondii*: fact or fiction? Trends Parasitol 29: 150-155.
4. Papp B, Pal C, Hurst LD (2004) Metabolic network analysis of the causes and evolution of enzyme dispensability in yeast. Nature 429: 661-664.
5. Nerima B, Nilsson D, Maser P (2010) Comparative genomics of metabolic networks of free-living and parasitic eukaryotes. BMC Genomics 11: 217.
6. Martin RE, Henry RI, Abbey JL, Clements JD, Kirk K (2005) The 'permeome' of the malaria parasite: an overview of the membrane transport proteins of *Plasmodium falciparum*. Genome Biol 6: R26.
7. Lau AO (2009) An overview of the *Babesia*, *Plasmodium* and *Theileria* genomes: a comparative perspective. Mol Biochem Parasitol 164: 1-8.
8. Giaever G, Chu AM, Ni L, Connelly C, Riles L, et al. (2002) Functional profiling of the *Saccharomyces cerevisiae* genome. Nature 418: 387-391.
9. Tong X, Campbell JW, Balazsi G, Kay KA, Wanner BL, et al. (2004) Genome-scale identification of conditionally essential genes in *E. coli* by DNA microarrays. Biochem Biophys Res Commun 322: 347-354.
10. Jacobs MA, Alwood A, Thaipisuttikul I, Spencer D, Haugen E, et al. (2003) Comprehensive transposon mutant library of *Pseudomonas aeruginosa*. Proc Natl Acad Sci U S A 100: 14339-14344.
11. Winzeler EA, Shoemaker DD, Astromoff A, Liang H, Anderson K, et al. (1999) Functional characterization of the *S. cerevisiae* genome by gene deletion and parallel analysis. Science 285: 901-906.
12. Limenitakis J, Soldati-Favre D (2011) Functional genetics in *Apicomplexa*: potentials and limits. FEBS Lett 585: 1579-1588.
13. Gardner MJ, Hall N, Fung E, White O, Berriman M, et al. (2002) Genome sequence of the human malaria parasite *Plasmodium falciparum*. Nature 419: 498-511.
14. Lewis NE, Nagarajan H, Palsson BO (2012) Constraining the metabolic genotype-phenotype relationship using a phylogeny of *in silico* methods. Nature Reviews Microbiology 10: 291-305.
15. Hyduke DR, Lewis NE, Palsson BO (2013) Analysis of omics data with genome-scale models of metabolism. Mol Biosyst 9: 167-174.
16. Orth JD, Thiele I, Palsson BØ (2010) What is flux balance analysis? Nat Biotechnol: Nature Publishing Group. pp. 245-248.
17. Aregawi M, World Health Organization. Global Malaria Programme. (2010) World malaria report 2010. Geneva: World Health Organization. xxix, 204 p. p.
18. Baum J, Papenfuss AT, Mair GR, Janse CJ, Vlachou D, et al. (2009) Molecular genetics and comparative genomics reveal RNAi is not functional in malaria parasites. Nucleic Acids Res 37: 3788-3798.
19. Visvesvara GS, Garcia LS (2002) Culture of protozoan parasites. Clin Microbiol Rev 15: 327-328.
20. Balu B (2012) Moving "Forward" in *Plasmodium* Genetics through a Transposon-Based Approach. J Trop Med 2012: 829210.

21. Augagneur Y, Wesolowski D, Tae HS, Altman S, Ben Mamoun C (2012) Gene selective mRNA cleavage inhibits the development of *Plasmodium falciparum*. Proceedings of the National Academy of Sciences of the United States of America 109: 6235-6240.
22. Kim HU, Sohn SB, Lee SY (2012) Metabolic network modeling and simulation for drug targeting and discovery. Biotechnol J 7: 330-342.
23. Pinney JW, Papp B, Hyland C, Wambua L, Westhead DR, et al. (2007) Metabolic reconstruction and analysis for parasite genomes. Trends in Parasitology 23: 548-554.
24. Bazzani S, Hoppe A, Holzhutter HG (2012) Network-based assessment of the selectivity of metabolic drug targets in *Plasmodium falciparum* with respect to human liver metabolism. BMC Syst Biol 6: 118.
25. Fatumo S, Plaimas K, Mallm JP, Schramm G, Adebiyi E, et al. (2009) Estimating novel potential drug targets of *Plasmodium falciparum* by analysing the metabolic network of knock-out strains *in silico*. Infect Genet Evol 9: 351-358.
26. Huthmacher C, Hoppe A, Bulik S, Holzhutter HG (2010) Antimalarial drug targets in *Plasmodium falciparum* predicted by stage-specific metabolic network analysis. BMC Syst Biol 4: 120.
27. Plata G, Hsiao TL, Olszewski KL, Llinas M, Vitkup D (2010) Reconstruction and flux-balance analysis of the *Plasmodium falciparum* metabolic network. Mol Syst Biol 6: 408.
28. Yeh I, Hanekamp T, Tsoka S, Karp PD, Altman RB (2004) Computational analysis of *Plasmodium falciparum* metabolism: organizing genomic information to facilitate drug discovery. Genome Res 14: 917-924.
29. Hayes CN, Wheelock AM, Normark J, Wahlgren M, Goto S, et al. (2006) Enlistment of omics technologies in the fight against malaria: Panacea or Pandora's Box? Journal of Pesticide Science 31: 263-272.
30. Ogata H, Goto S, Sato K, Fujibuchi W, Bono H, et al. (1999) KEGG: Kyoto Encyclopedia of Genes and Genomes. Nucleic Acids Res 27: 29-34.
31. Ginsburg H (2006) Progress in *in silico* functional genomics: the malaria Metabolic Pathways database. Trends in Parasitology 22: 238-240.
32. Bozdech Z, Ginsburg H (2005) Data mining of the transcriptome of *Plasmodium falciparum*: the pentose phosphate pathway and ancillary processes. Malaria Journal 4.
33. Stelling J (2004) Mathematical models in microbial systems biology. Current Opinion in Microbiology 7: 513-518.
34. Orth JD, Thiele I, Palsson BO (2010) What is flux balance analysis? Nat Biotechnol 28: 245-248.
35. Palsson B (2006) Systems biology : properties of reconstructed networks. New York: Cambridge University Press. xii, 322 p. p.
36. Feist AM, Palsson BO (2010) The biomass objective function. Current Opinion in Microbiology 13: 344-349.
37. Schuetz R, Kuepfer L, Sauer U (2007) Systematic evaluation of objective functions for predicting intracellular fluxes in *Escherichia coli*. Mol Syst Biol 3: 119.
38. Thiele I, Palsson BO (2010) A protocol for generating a high-quality genome-scale metabolic reconstruction. Nature Protocols 5: 93-121.
39. Chavali AK, Whittemore JD, Eddy JA, Williams KT, Papin JA (2008) Systems analysis of metabolism in the pathogenic trypanosomatid *Leishmania major*. Mol Syst Biol 4.
40. Zomorrodi AR, Maranas CD (2010) Improving the iMM904 *S. cerevisiae* metabolic model using essentiality and synthetic lethality data. BMC Syst Biol 4.

41. Fatumo S, Plaimas K, Adebiyi E, Konig R (2011) Comparing metabolic network models based on genomic and automatically inferred enzyme information from *Plasmodium* and its human host to define drug targets *in silico*. Infect Genet Evol 11: 201-208.
42. Hecker N, Ahmed J, Eichborn J, Dunkel M, Macha K, et al. (2012) SuperTarget goes quantitative: update on drug-target interactions. Nucleic Acids Res 40: D1113-D1117.
43. Olszewski KL, Morrisey JM, Wilinski D, Burns JM, Vaidya AB, et al. (2009) Host-parasite interactions revealed by *Plasmodium falciparum* metabolomics. Cell Host Microbe 5: 191-199.
44. Gille C, Bolling C, Hoppe A, Bulik S, Hoffmann S, et al. (2010) HepatoNet1: a comprehensive metabolic reconstruction of the human hepatocyte for the analysis of liver physiology. Mol Syst Biol 6: 411.
45. Henry CS, Broadbelt LJ, Hatzimanikatis V (2007) Thermodynamics-based metabolic flux analysis. Biophysical Journal 92: 1792-1805.
46. Soh KC, Hatzimanikatis V (2010) Network thermodynamics in the post-genomic era. Current Opinion in Microbiology 13: 350-357.
47. Soh KC, Miskovic L, Hatzimanikatis V (2012) From network models to network responses: integration of thermodynamic and kinetic properties of yeast genome-scale metabolic networks. Fems Yeast Research 12: 129-143.
48. Oberhardt MA, Puchalka J, dos Santos VAPM, Papin JA (2011) Reconciliation of Genome-Scale Metabolic Reconstructions for Comparative Systems Analysis. Plos Computational Biology 7.
49. Chindelevitch L, Stanley S, Hung D, Regev A, Berger B (2012) MetaMerge: scaling up genome-scale metabolic reconstructions with application to *Mycobacterium tuberculosis*. Genome Biology 13.
50. Aurrecoechea C, Heiges M, Wang H, Wang Z, Fischer S, et al. (2007) ApiDB: integrated resources for the apicomplexan bioinformatics resource center. Nucleic Acids Res 35: D427-D430.
51. Balu B, Singh N, Maher SP, Adams JH (2010) A Genetic Screen for Attenuated Growth Identifies Genes Crucial for Intraerythrocytic Development of *Plasmodium falciparum*. Plos One 5.
52. Alsford S, Turner DJ, Obado SO, Sanchez-Flores A, Glover L, et al. (2011) High-throughput phenotyping using parallel sequencing of RNA interference targets in the African trypanosome. Genome Res 21: 915-924.
53. Kumar VS, Maranas CD (2009) GrowMatch: An Automated Method for Reconciling *In Silico*/*In Vivo* Growth Predictions. Plos Computational Biology 5.
54. Blazier AS, Papin JA (2012) Integration of expression data in genome-scale metabolic network reconstructions. Front Physiol 3: 299.
55. Bozdech Z, Llinas M, Pulliam BL, Wong ED, Zhu JC, et al. (2003) The transcriptome of the intraerythrocytic developmental cycle of *Plasmodium falciparum*. Plos Biology 1: 85-100.
56. Le Roch KG, Zhou YY, Blair PL, Grainger M, Moch JK, et al. (2003) Discovery of gene function by expression profiling of the malaria parasite life cycle. Science 301: 1503-1508.
57. Sacci JB, Ribeiro JMC, Huang FY, Alam U, Russell JA, et al. (2005) Transcriptional analysis of *in vivo Plasmodium yoelii* liver stage gene expression. Mol Biochem Parasitol 142: 177-183.
58. Daily JP, Scanfeld D, Pochet N, Le Roch K, Plouffe D, et al. (2007) Distinct physiological states of *Plasmodium falciparum* in malaria-infected patients. Nature 450: 1091-U1015.

59. Tarun AS, Peng X, Dumpit RF, Ogata Y, Silva-Rivera H, et al. (2008) A combined transcriptome and proteome survey of malaria parasite liver stages. Proceedings of the National Academy of Sciences of the United States of America 105: 305-310.
60. Le Roch KG, Johnson JR, Florens L, Zhou Y, Santrosyan A, et al. (2004) Global analysis of transcript and protein levels across the *Plasmodium falciparum* life cycle. Genome Res 14: 2308-2318.
61. Llinas M, Bozdech Z, Wong ED, Adai AT, DeRisi JL (2006) Comparative whole genome transcriptome analysis of three *Plasmodium falciparum* strains. Nucleic Acids Res 34: 1166-1173.
62. Otto TD, Wilinski D, Assefa S, Keane TM, Sarry LR, et al. (2010) New insights into the blood-stage transcriptome of *Plasmodium falciparum* using RNA-Seq. Molecular Microbiology 76: 12-24.
63. Florens L, Washburn MP, Raine JD, Anthony RM, Grainger M, et al. (2002) A proteomic view of the *Plasmodium falciparum* life cycle. Nature 419: 520-526.
64. Lasonder E, Ishihama Y, Andersen JS, Vermunt AMW, Pain A, et al. (2002) Analysis of the *Plasmodium falciparum* proteome by high-accuracy mass spectrometry. Nature 419: 537-542.
65. Wastling JM, Xia D, Sohal A, Chaussepied M, Pain A, et al. (2009) Proteomes and transcriptomes of the *Apicomplexa* - Where's the message? International Journal for Parasitology 39: 135-143.
66. Nirmalan N, Sims PFG, Hyde JE (2004) Quantitative proteomics of the human malaria parasite *Plasmodium falciparum* and its application to studies of development and inhibition. Molecular Microbiology 52: 1187-1199.
67. Prieto JH, Koncarevic S, Park SK, Yates J, Becker K (2008) Large-Scale Differential Proteome Analysis in *Plasmodium falciparum* under Drug Treatment. PLoS ONE 3(12): e4098
68. Southworth PM, Hyde JE, Sims PFG (2011) A mass spectrometric strategy for absolute quantification of *Plasmodium falciparum* proteins of low abundance. Malaria Journal 10.
69. Foth BJ, Zhang N, Chaal BK, Sze SK, Preiser PR, et al. (2011) Quantitative Time-course Profiling of Parasite and Host Cell Proteins in the Human Malaria Parasite *Plasmodium falciparum*. Molecular & Cellular Proteomics 10.
70. Mair GR, Braks JAM, Garver LS, Wiegant LCAG, Hall N, et al. (2006) Regulation of sexual development of *Plasmodium* by translational repression. Science 313: 667-669.
71. Roberts SB, Robichaux JL, Chavali AK, Manque PA, Lee V, et al. (2009) Proteomic and network analysis characterize stage-specific metabolism in *Trypanosoma cruzi*. BMC Syst Biol 3: 52.
72. Veenstra TD (2012) Metabolomics: the final frontier? Genome Med 4: 40.
73. Nicholls AW (2012) Realising the potential of metabolomics. Bioanalysis 4: 2195-2197.
74. Wishart DS (2010) Computational approaches to metabolomics. Methods Mol Biol 593: 283-313.
75. Lakshmanan V, Rhee KY, Daily JP (2011) Metabolomics and malaria biology. Mol Biochem Parasitol 175: 104-111.
76. t'Kindt R, Jankevics A, Scheltema RA, Zheng L, Watson DG, et al. (2010) Towards an unbiased metabolic profiling of protozoan parasites: optimisation of a *Leishmania* sampling protocol for HILIC-orbitrap analysis. Analytical and Bioanalytical Chemistry 398: 2059-2069.
77. Paglia G, Hrafnsdottir S, Magnusdottir M, Fleming RMT, Thorlacius S, et al. (2012) Monitoring metabolites consumption and secretion in cultured cells using ultra-

performance liquid chromatography quadrupole-time of flight mass spectrometry (UPLC-Q-ToF-MS). Analytical and Bioanalytical Chemistry 402: 1183-1198.
78. Duy SV, Besteiro S, Berry L, Perigaud C, Bressolle F, et al. (2012) A quantitative liquid chromatography tandem mass spectrometry method for metabolomic analysis of *Plasmodium falciparum* lipid related metabolites. Analytica Chimica Acta 739: 47-55.
79. Lian LY, Al-Helal M, Roslaini AM, Fisher N, Bray PG, et al. (2009) Glycerol: An unexpected major metabolite of energy metabolism by the human malaria parasite. Malaria Journal 8.
80. Teng RW, Junankar PR, Bubb WA, Rae C, Mercier P, et al. (2009) Metabolite profiling of the intraerythrocytic malaria parasite *Plasmodium falciparum* by H-1 NMR spectroscopy. Nmr in Biomedicine 22: 292-302.
81. Oikawa A, Matsuda F, Kikuyama M, Mimura T, Saito K (2011) Metabolomics of a Single Vacuole Reveals Metabolic Dynamism in an *Alga Chara australis*. Plant Physiology 157: 544-551.
82. Tang YJ, Martin HG, Myers S, Rodriguez S, Baidoo EEK, et al. (2009) Advances in Analysis of Microbial Metabolic Fluxes Via C-13 Isotopic Labeling. Mass Spectrometry Reviews 28: 362-375.
83. Winter G, Kromer JO (2012) Fluxomics - connecting 'omics analysis and phenotypes. Environ Microbiol. Jul;15(7):1901-16
84. Mehta M, Sonawat HM, Sharma S (2005) Malaria parasite-infected erythrocytes inhibit glucose utilization in uninfected red cells. FEBS Lett 579: 6151-6158.
85. Mehta M, Sonawat HM, Sharma S (2006) Glycolysis in *Plasmodium falciparum* results in modulation of host enzyme activities. J Vector Borne Dis 43: 95-103.
86. Martin RE, Kirk K (2007) Transport of the essential nutrient isoleucine in human erythrocytes infected with the malaria parasite *Plasmodium falciparum*. Blood 109: 2217-2224.
87. Saliba KJ, Kirk K (2001) H+-coupled pantothenate transport in the intracellular malaria parasite. J Biol Chem 276: 18115-18121.
88. Saliba KJ, Martin RE, Broer A, Henry RI, McCarthy CS, et al. (2006) Sodium-dependent uptake of inorganic phosphate by the intracellular malaria parasite. Nature 443: 582-585.
89. Wiebe MG, Rintala E, Tamminen A, Simolin H, Salusjarvi L, et al. (2008) Central carbon metabolism of *Saccharomyces cerevisiae* in anaerobic, oxygen-limited and fully aerobic steady-state conditions and following a shift to anaerobic conditions. FEMS Yeast Res 8: 140-154.
90. Tymoshenko S, Oppenheim RD, Soldati-Favre D, Hatzimanikatis V (2013) Functional genomics of *Plasmodium falciparum* using metabolic modelling and analysis. Briefings in functional genomics. pp. 316-327.
91. Chavali AK, Whittemore JD, Eddy JA, Williams KT, Papin JA (2008) Systems analysis of metabolism in the pathogenic trypanosomatid *Leishmania major*. Mol Syst Biol 4: 177.
92. Vanee N, Roberts SB, Fong SS, Manque P, Buck GA (2010) A genome-scale metabolic model of *Cryptosporidium hominis*. Chem Biodivers 7: 1026-1039.
93. Forth T (2012) Metabolic systems biology of the malaria parasite : reconstruction, visualisation and analysis of an experimentally parameterised metabolic model of the human acute malaria parasite *Plasmodium falciparum* (PhD Thesis) University of Leeds
94. Agren R, Liu L, Shoaie S, Vongsangnak W, Nookaew I, et al. (2013) The RAVEN Toolbox and Its Use for Generating a Genome-scale Metabolic Model for *Penicillium chrysogenum*. PLoS Comput Biol. pp. e1002980.

95. Feist AM, Henry CS, Reed JL, Krummenacker M, Joyce AR, et al. (2007) A genome-scale metabolic reconstruction for *Escherichia coli* K-12 MG1655 that accounts for 1260 ORFs and thermodynamic information. Mol Syst Biol 3: 121.
96. Fleming RM, Thiele I (2011) von Bertalanffy 1.0: a COBRA toolbox extension to thermodynamically constrain metabolic models. Bioinformatics 27: 142-143.
97. Noor E, Bar-Even A, Flamholz A, Lubling Y, Davidi D, et al. (2012) An integrated open framework for thermodynamics of reactions that combines accuracy and coverage. Bioinformatics 28: 2037-2044.
98. Kummel A, Panke S, Heinemann M (2006) Systematic assignment of thermodynamic constraints in metabolic network models. BMC Bioinformatics 7: 512.
99. Soh KC, Hatzimanikatis V (2014) Constraining the flux space using thermodynamics and integration of metabolomics data. Methods Mol Biol 1191: 49-63.
100. Martinez VS, Quek LE, Nielsen LK (2014) Network thermodynamic curation of human and yeast genome-scale metabolic models. Biophys J 107: 493-503.
101. Schellenberger J, Que R, Fleming RMT, Thiele I, Orth JD, et al. (2011) Quantitative prediction of cellular metabolism with constraint-based models: the COBRA Toolbox v2.0. Nature Protocols 6: 1290-1307.
102. Mahadevan R, Schilling CH (2003) The effects of alternate optimal solutions in constraint-based genome-scale metabolic models. Metab Eng 5: 264-276.
103. Oppenheim RD, Creek DJ, Macrae JI, Modrzynska KK, Pino P, et al. (2014) BCKDH: the missing link in apicomplexan mitochondrial metabolism is required for full virulence of *Toxoplasma gondii* and *Plasmodium berghei*. PLoS Pathog 10: e1004263.
104. Sazanov LA, Jackson JB (1994) Proton-translocating transhydrogenase and NAD- and NADP-linked isocitrate dehydrogenases operate in a substrate cycle which contributes to fine regulation of the tricarboxylic acid cycle activity in mitochondria. FEBS Lett 344: 109-116.
105. Danne JC, Gornik SG, Macrae JI, McConville MJ, Waller RF (2013) Alveolate mitochondrial metabolic evolution: dinoflagellates force reassessment of the role of parasitism as a driver of change in apicomplexans. Mol Biol Evol 30: 123-139.
106. Jo SH, Son MK, Koh HJ, Lee SM, Song IH, et al. (2001) Control of mitochondrial redox balance and cellular defense against oxidative damage by mitochondrial NADP+-dependent isocitrate dehydrogenase. J Biol Chem 276: 16168-16176.
107. Mavrovouniotis ML (1990) Group contributions for estimating standard gibbs energies of formation of biochemical compounds in aqueous solution. Biotechnol Bioeng 36: 1070-1082.
108. Jankowski MD, Henry CS, Broadbelt LJ, Hatzimanikatis V (2008) Group contribution method for thermodynamic analysis of complex metabolic networks. Biophys J 95: 1487-1499.
109. Soh KC (2013) Computational Studies on Cellular Bioenergetics. PhD thesis EPFL.
110. Beitz E (2007) Jammed traffic impedes parasite growth. Proc Natl Acad Sci U S A 104: 13855-13856.
111. Ludovico P, Sansonetty F, Corte-Real M (2001) Assessment of mitochondrial membrane potential in yeast cell populations by flow cytometry. Microbiology 147: 3335-3343.
112. van Schalkwyk DA, Chan XW, Misiano P, Gagliardi S, Farina C, et al. (2010) Inhibition of *Plasmodium falciparum* pH regulation by small molecule indole derivatives results in rapid parasite death. Biochem Pharmacol 79: 1291-1299.

113. Rohmer M (1999) The discovery of a mevalonate-independent pathway for isoprenoid biosynthesis in bacteria, algae and higher plants. Natural Product Reports 16: 565-574.
114. Gabrielsen M, Bond CS, Hallyburton I, Hecht S, Bacher A, et al. (2004) Hexameric assembly of the bifunctional methylerythritol 2,4-cyclodiphosphate synthase and protein-protein associations in the deoxy-xylulose-dependent pathway of isoprenoid precursor biosynthesis. J Biol Chem 279: 52753-52761.
115. Lherbet C, Pojer F, Richard SB, Noel JP, Poulter CD (2006) Absence of substrate channeling between active sites in the *Agrobacterium tumefaciens* IspDF and IspE enzymes of the methyl erythritol phosphate pathway. Biochemistry 45: 3548-3553.
116. Bouche N, Fromm H (2004) GABA in plants: just a metabolite? Trends Plant Sci 9: 110-115.
117. Ke H, Lewis IA, Morrisey JM, McLean KJ, Ganesan SM, et al. (2015) Genetic Investigation of Tricarboxylic Acid Metabolism during the *Plasmodium falciparum* Life Cycle. Cell Rep 11: 164-174.
118. Eddy SR (1998) Profile hidden Markov models. Bioinformatics 14: 755-763.
119. Frech C, Chen N (2011) Genome comparison of human and non-human malaria parasites reveals species subset-specific genes potentially linked to human disease. PLoS Comput Biol 7: e1002320.
120. Torchetti EM, Bonomi F, Galluccio M, Gianazza E, Giancaspero TA, et al. (2011) Human FAD synthase (isoform 2): a component of the machinery that delivers FAD to apo-flavoproteins. FEBS J 278: 4434-4449.
121. Pedrolli DB, Jankowitsch F, Schwarz J, Langer S, Nakanishi S, et al. (2013) Riboflavin Analogs as Antiinfectives: Occurrence, Mode of Action, Metabolism and Resistance. Current Pharmaceutical Design 19: 2552-2560.
122. Dym O, Eisenberg D (2001) Sequence-structure analysis of FAD-containing proteins. Protein Sci 10: 1712-1728.
123. Gudipati V, Koch K, Lienhart W-D, Macheroux P (2013) The flavoproteome of the yeast *Saccharomyces cerevisiae*. Biochim Biophys Acta. pp. 535-544.
124. Macheroux P, Kappes B, Ealick SE (2011) Flavogenomics--a genomic and structural view of flavin-dependent proteins. FEBS J 278: 2625-2634.
125. Balconi E, Pennati A, Crobu D, Pandini V, Cerutti R, et al. (2009) The ferredoxin-NADP+ reductase/ferredoxin electron transfer system of *Plasmodium falciparum*. FEBS J 276: 3825-3836.
126. Vollmer M, Thomsen N, Wiek S, Seeber F (2001) Apicomplexan parasites possess distinct nuclear-encoded, but apicoplast-localized, plant-type ferredoxin-NADP+ reductase and ferredoxin. J Biol Chem 276: 5483-5490.
127. Possenti A, Fratini F, Fantozzi L, Pozio E, Dubey JP, et al. (2013) Global proteomic analysis of the oocyst/sporozoite of *Toxoplasma gondii* reveals commitment to a host-independent lifestyle. BMC Genomics 14: 183.
128. Huerta C, Borek D, Machius M, Grishin NV, Zhang H (2009) Structure and mechanism of a eukaryotic FMN adenylyltransferase. J Mol Biol 389: 388-400.
129. Kearney EB, Englard S (1951) The enzymatic phosphorylation of riboflavin. J Biol Chem 193: 821-834.
130. Serrano A, Ferreira P, Martinez-Julvez M, Medina M (2013) The prokaryotic FAD synthetase family: a potential drug target. Curr Pharm Des 19: 2637-2648.
131. Torchetti EM, Brizio C, Colella M, Galluccio M, Giancaspero TA, et al. (2010) Mitochondrial localization of human FAD synthetase isoform 1. Mitochondrion 10: 263-273.

132. Brizio C, Galluccio M, Wait R, Torchetti EM, Bafunno V, et al. (2006) Over-expression in *Escherichia coli* and characterization of two recombinant isoforms of human FAD synthetase. Biochem Biophys Res Commun 344: 1008-1016.
133. Barile M, Brizio C, Valenti D, De Virgilio C, Passarella S (2000) The riboflavin/FAD cycle in rat liver mitochondria. Eur J Biochem 267: 4888-4900.
134. Barile M, Passarella S, Bertoldi A, Quagliariello E (1993) Flavin adenine dinucleotide synthesis in isolated rat liver mitochondria caused by imported flavin mononucleotide. Arch Biochem Biophys 305: 442-447.
135. Pallotta ML, Brizio C, Fratianni A, De Virgilio C, Barile M, et al. (1998) *Saccharomyces cerevisiae* mitochondria can synthesise FMN and FAD from externally added riboflavin and export them to the extramitochondrial phase. FEBS Lett 428: 245-249.
136. Bafunno V, Giancaspero TA, Brizio C, Bufano D, Passarella S, et al. (2004) Riboflavin uptake and FAD synthesis in *Saccharomyces cerevisiae* mitochondria: involvement of the Flx1p carrier in FAD export. J Biol Chem 279: 95-102.
137. Salcedo-Sora JE, Ward SA (2013) The folate metabolic network of *Falciparum* malaria. Mol Biochem Parasitol 188: 51-62.
138. Hille R, Miller S, Palfey B (2013) Handbook of flavoproteins. Berlin ; Boston: De Gruyter.
139. Claros MG (1995) MitoProt, a Macintosh application for studying mitochondrial proteins. Comput Appl Biosci 11: 441-447.
140. Emanuelsson O, Brunak S, von Heijne G, Nielsen H (2007) Locating proteins in the cell using TargetP, SignalP and related tools. Nat Protoc 2: 953-971.
141. Cilingir G, Broschat SL, Lau AO (2012) ApicoAP: the first computational model for identifying apicoplast-targeted proteins in multiple species of *Apicomplexa*. PLoS One 7: e36598.
142. Corpet F (1988) Multiple sequence alignment with hierarchical clustering. Nucleic Acids Res 16: 10881-10890.
143. Soldati D, Boothroyd JC (1993) Transient transfection and expression in the obligate intracellular parasite *Toxoplasma gondii*. Science. pp. 349-352.
144. Gajria B, Bahl A, Brestelli J, Dommer J (2008) ToxoDB: an integrated *Toxoplasma gondii* database resource. Nucleic Acids Res. pp. D553-D556.
145. Eschbach M-L, Müller IB, Gilberger T-W, Walter RD, Wrenger C (2006) The human malaria parasite *Plasmodium falciparum* expresses an atypical N-terminally extended pyrophosphokinase with specificity for thiamine. Biological chemistry 387: 1583-1591.
146. Shen B, Brown KM, Lee TD, Sibley LD (2014) Efficient Gene Disruption in Diverse Strains of *Toxoplasma gondii* Using CRISPR/CAS9. Mbio 5: e01114-01114.
147. Charpentier E, Doudna JA (2013) Biotechnology: Rewriting a genome. Nature 495: 50-51.
148. Cherry JM, Hong EL, Amundsen C, Balakrishnan R, Binkley G, et al. (2012) *Saccharomyces* Genome Database: the genomics resource of budding yeast. Nucleic Acids Res 40: D700-705.
149. Liuzzi VC, Giancaspero TA, Gianazza E, Banfi C, Barile M, et al. (2012) Silencing of FAD synthase gene in *Caenorhabditis elegans* upsets protein homeostasis and impacts on complex behavioral patterns. Biochim Biophys Acta 1820: 521-531.
150. Nudler E, Mironov AS (2004) The riboswitch control of bacterial metabolism. Trends Biochem Sci 29: 11-17.
151. Das BS, Das DB, Satpathy RN, Patnaik JK, Bose TK (1988) Riboflavin deficiency and severity of malaria. Eur J Clin Nutr 42: 277-283.

152. Dutta P (1991) Enhanced uptake and metabolism of riboflavin in erythrocytes infected with *Plasmodium falciparum*. J Protozool 38: 479-483.
153. Neeman MM (1946) Syntheses of 9-substituted flavins as antimalarials. J Chem Soc: 811-813.
154. Halladay PK, Hunt NH, Butcher GA, Cowden WB (1990) Antimalarial action of flavin analogues seems not be due to inhibition of glutathione reductase of host erythrocytes. Biochem Pharmacol 39: 1063-1065.
155. Cowden WB, Halladay PK, Cunningham RB, Hunt NH, Clark IA (1991) Flavins as potential antimalarials. 2. 3-Methyl-10-(substituted-phenyl) flavins. J Med Chem 34: 1818-1822.
156. Graham DW, Brown JE, Ashton WT, Brown RD, Rogers EF (1977) Anticoccidial riboflavine antagonists. Experientia 33: 1274-1276.
157. Graham DW, Rogers EF, Ashton WT (1978) 5-Deazariboflavin and its derivatives. Accessed at http://www.google.com/patents/US4079069
158. Geary TG, Divo AA, Jensen JB (1985) Nutritional requirements of *Plasmodium falciparum* in culture. II. Effects of antimetabolites in a semi-defined medium. J Protozool 32: 65-69.
159. Cowden WB, Butcher GA, Hunt NH, Clark IA, Yoneda F (1987) Antimalarial activity of a riboflavin analog against *Plasmodium vinckei in vivo* and *Plasmodium falciparum in vitro*. Am J Trop Med Hyg 37: 495-500.
160. Harb OS, Roos DS (2015) The Eukaryotic Pathogen Databases: a functional genomic resource integrating data from human and veterinary parasites. Methods Mol Biol 1201: 1-18.
161. Sen P, Vial HJ, Radulescu O (2013) Kinetic modelling of phospholipid synthesis in *Plasmodium knowlesi* unravels crucial steps and relative importance of multiple pathways. BMC Syst Biol 7: 123.
162. Penkler G, du Toit F, Adams W, Rautenbach M, Palm DC, et al. (2015) Construction and validation of a detailed kinetic model of glycolysis in *Plasmodium falciparum*. FEBS J 282: 1481-1511.
163. Abdallah TM, Abdeen MT, Ahmed IS, Hamdan HZ, Magzoub M, et al. (2013) Severe *Plasmodium falciparum* and *Plasmodium vivax* malaria among adults at Kassala Hospital, eastern Sudan. Malar J 12: 148.
164. Thomson-Luque R, Scopel KK (2015) Immature reticulocytes as preferential host cells and the challenges for *in vitro* culture of *Plasmodium vivax*. Pathog Glob Health 109: 91-92.
165. Schwach F, Bushell E, Gomes AR, Anar B, Girling G, et al. (2015) PlasmoGEM, a database supporting a community resource for large-scale experimental genetics in malaria parasites. Nucleic Acids Res 43: D1176-1182.
166. Ghorbal M, Gorman M, Macpherson CR, Martins RM, Scherf A, et al. (2014) Genome editing in the human malaria parasite *Plasmodium falciparum* using the CRISPR-Cas9 system. Nat Biotechnol 32: 819-821.
167. Zhang C, Xiao B, Jiang Y, Zhao Y, Li Z, et al. (2014) Efficient editing of malaria parasite genome using the CRISPR/Cas9 system. MBio 5: e01414-01414.
168. Song C, Chiasson MA, Nursimulu N, Hung SS, Wasmuth J, et al. (2013) Metabolic reconstruction identifies strain-specific regulation of virulence in *Toxoplasma gondii*. Mol Syst Biol 9: 708.
169. Rossell S, Huynen MA, Notebaart RA (2013) Inferring metabolic states in uncharacterized environments using gene-expression measurements. PLoS Comput Biol 9: e1002988.

170. Saliba KJ, Horner HA, Kirk K (1998) Transport and metabolism of the essential vitamin pantothenic acid in human erythrocytes infected with the malaria parasite *Plasmodium falciparum*. J Biol Chem 273: 10190-10195.
171. Banerjee T, Jaijyan DK, Surolia N, Singh AP, Surolia A (2012) Apicoplast triose phosphate transporter (TPT) gene knockout is lethal for *Plasmodium*. Mol Biochem Parasitol 186: 44-50.
172. Webster HK, Whaun JM (1982) Antimalarial properties of bredinin. Prediction based on identification of differences in human host-parasite purine metabolism. J Clin Invest 70: 461-469.
173. Jomaa H, Wiesner J, Sanderbrand S, Altincicek B, Weidemeyer C, et al. (1999) Inhibitors of the nonmevalonate pathway of isoprenoid biosynthesis as antimalarial drugs. Science 285: 1573-1576.
174. Cassera MB, Merino EF, Peres VJ, Kimura EA, Wunderlich G, et al. (2007) Effect of fosmidomycin on metabolic and transcript profiles of the methylerythritol phosphate pathway in *Plasmodium falciparum*. Mem Inst Oswaldo Cruz 102: 377-383.
175. Odom AR, Van Voorhis WC (2010) Functional genetic analysis of the *Plasmodium falciparum* deoxyxylulose 5-phosphate reductoisomerase gene. Mol Biochem Parasitol 170: 108-111.
176. Razakantoanina V, Nguyen Kim PP, Jaureguiberry G (2000) Antimalarial activity of new gossypol derivatives. Parasitol Res 86: 665-668.
177. Vivas L, Easton A, Kendrick H, Cameron A, Lavandera JL, et al. (2005) *Plasmodium falciparum*: stage specific effects of a selective inhibitor of lactate dehydrogenase. Exp Parasitol 111: 105-114.
178. Biagini GA, Fisher N, Shone AE, Mubaraki MA, Srivastava A, et al. (2012) Generation of quinolone antimalarials targeting the *Plasmodium falciparum* mitochondrial respiratory chain for the treatment and prophylaxis of malaria. Proc Natl Acad Sci U S A 109: 8298-8303.
179. Soulere L, Delplace P, Davioud-Charvet E, Py S, Sergheraert C, et al. (2003) Screening of *Plasmodium falciparum* iron superoxide dismutase inhibitors and accuracy of the SOD-assays. Bioorg Med Chem 11: 4941-4944.
180. Vinayak S, Sharma YD (2007) Inhibition of *Plasmodium falciparum* ispH (lytB) gene expression by hammerhead ribozyme. Oligonucleotides 17: 189-200.
181. Chakrabarti D, Schuster SM, Chakrabarti R (1993) Cloning and characterization of subunit genes of ribonucleotide reductase, a cell-cycle-regulated enzyme, from *Plasmodium falciparum*. Proc Natl Acad Sci U S A 90: 12020-12024.
182. Munro JB, Silva JC (2012) Ribonucleotide reductase as a target to control apicomplexan diseases. Curr Issues Mol Biol 14: 9-26.
183. Vaughan AM, O'Neill MT, Tarun AS, Camargo N, Phuong TM, et al. (2009) Type II fatty acid synthesis is essential only for malaria parasite late liver stage development. Cell Microbiol 11: 506-520.
184. Yu M, Kumar TR, Nkrumah LJ, Coppi A, Retzlaff S, et al. (2008) The fatty acid biosynthesis enzyme FabI plays a key role in the development of liver-stage malarial parasites. Cell Host Microbe 4: 567-578.
185. Ramya TN, Mishra S, Karmodiya K, Surolia N, Surolia A (2007) Inhibitors of nonhousekeeping functions of the apicoplast defy delayed death in *Plasmodium falciparum*. Antimicrob Agents Chemother 51: 307-316.
186. Nagaraj VA, Arumugam R, Prasad D, Rangarajan PN, Padmanaban G (2010) Protoporphyrinogen IX oxidase from *Plasmodium falciparum* is anaerobic and is localized to the mitochondrion. Mol Biochem Parasitol 174: 44-52.

187. Deng X, Gujjar R, El Mazouni F, Kaminsky W, Malmquist NA, et al. (2009) Structural plasticity of malaria dihydroorotate dehydrogenase allows selective binding of diverse chemical scaffolds. J Biol Chem 284: 26999-27009.
188. Gujjar R, El Mazouni F, White KL, White J, Creason S, et al. (2011) Lead optimization of aryl and aralkyl amine-based triazolopyrimidine inhibitors of *Plasmodium falciparum* dihydroorotate dehydrogenase with antimalarial activity in mice. J Med Chem 54: 3935-3949.
189. Painter HJ, Morrisey JM, Mather MW, Vaidya AB (2007) Specific role of mitochondrial electron transport in blood-stage *Plasmodium falciparum*. Nature 446: 88-91.
190. Baldwin J, Michnoff CH, Malmquist NA, White J, Roth MG, et al. (2005) High-throughput screening for potent and selective inhibitors of *Plasmodium falciparum* dihydroorotate dehydrogenase. J Biol Chem 280: 21847-21853.
191. Jiang L, Lee PC, White J, Rathod PK (2000) Potent and selective activity of a combination of thymidine and 1843U89, a folate-based thymidylate synthase inhibitor, against *Plasmodium falciparum*. Antimicrob Agents Chemother 44: 1047-1050.
192. Massimine KM, McIntosh MT, Doan LT, Atreya CE, Gromer S, et al. (2006) Eosin B as a novel antimalarial agent for drug-resistant *Plasmodium falciparum*. Antimicrob Agents Chemother 50: 3132-3141.
193. Biagini GA, Viriyavejakul P, O'Neill P M, Bray PG, Ward SA (2006) Functional characterization and target validation of alternative complex I of *Plasmodium falciparum* mitochondria. Antimicrob Agents Chemother 50: 1841-1851.
194. Uyemura SA, Luo S, Vieira M, Moreno SN, Docampo R (2004) Oxidative phosphorylation and rotenone-insensitive malate- and NADH-quinone oxidoreductases in *Plasmodium yoelii yoelii* mitochondria *in situ*. J Biol Chem 279: 385-393.
195. Biot C, Bauer H, Schirmer RH, Davioud-Charvet E (2004) 5-substituted tetrazoles as bioisosteres of carboxylic acids. Bioisosterism and mechanistic studies on glutathione reductase inhibitors as antimalarials. J Med Chem 47: 5972-5983.
196. Zhang YA, Hempelmann E, Schirmer RH (1988) Glutathione reductase inhibitors as potential antimalarial drugs. Effects of nitrosoureas on *Plasmodium falciparum in vitro*. Biochem Pharmacol 37: 855-860.
197. Luersen K, Walter RD, Muller S (2000) *Plasmodium falciparum*-infected red blood cells depend on a functional glutathione *de novo* synthesis attributable to an enhanced loss of glutathione. Biochem J 346 Pt 2: 545-552.
198. Grellier P, Maroziene A, Nivinskas H, Sarlauskas J, Aliverti A, et al. (2010) Antiplasmodial activity of quinones: roles of aziridinyl substituents and the inhibition of *Plasmodium falciparum* glutathione reductase. Arch Biochem Biophys 494: 32-39.
199. Krnajski Z, Gilberger TW, Walter RD, Cowman AF, Muller S (2002) Thioredoxin reductase is essential for the survival of *Plasmodium falciparum* erythrocytic stages. J Biol Chem 277: 25970-25975.
200. Witola WH, El Bissati K, Pessi G, Xie C, Roepe PD, et al. (2008) Disruption of the *Plasmodium falciparum* PfPMT gene results in a complete loss of phosphatidylcholine biosynthesis via the serine-decarboxylase-phosphoethanolamine-methyltransferase pathway and severe growth and survival defects. J Biol Chem 283: 27636-27643.
201. Nduati E, Hunt S, Kamau EM, Nzila A (2005) 2,4-diaminopteridine-based compounds as precursors for *de novo* synthesis of antifolates: a novel class of antimalarials. Antimicrob Agents Chemother 49: 3652-3657.

202. Banerjee AK, Arora N, Murty USN (2012) Aspartate carbamoyltransferase of *Plasmodium falciparum* as a potential drug target for designing anti-malarial chemotherapeutic agents. Medicinal Chemistry Research 21: 2480-2493.
203. Cassera MB, Gozzo FC, D'Alexandri FL, Merino EF, del Portillo HA, et al. (2004) The methylerythritol phosphate pathway is functionally active in all intraerythrocytic stages of *Plasmodium falciparum*. J Biol Chem 279: 51749-51759.
204. Santiago TC, Zufferey R, Mehra RS, Coleman RA, Mamoun CB (2004) The *Plasmodium falciparum* PfGatp is an endoplasmic reticulum membrane protein important for the initial step of malarial glycerolipid synthesis. J Biol Chem 279: 9222-9232.
205. Nicolas O, Margout D, Taudon N, Wein S, Calas M, et al. (2005) Pharmacological properties of a new antimalarial bisthiazolium salt, T3, and a corresponding prodrug, TE3. Antimicrob Agents Chemother 49: 3631-3639.
206. Gerold P, Schwarz RT (2001) Biosynthesis of glycosphingolipids *de-novo* by the human malaria parasite *Plasmodium falciparum*. Mol Biochem Parasitol 112: 29-37.
207. Surolia N, Padmanaban G (1992) *de novo* biosynthesis of heme offers a new chemotherapeutic target in the human malarial parasite. Biochem Biophys Res Commun 187: 744-750.
208. Varadharajan S, Dhanasekaran S, Bonday ZQ, Rangarajan PN, Padmanaban G (2002) Involvement of delta-aminolaevulinate synthase encoded by the parasite gene in *de novo* haem synthesis by *Plasmodium falciparum*. Biochem J 367: 321-327.
209. Waller RF, Ralph SA, Reed MB, Su V, Douglas JD, et al. (2003) A type II pathway for fatty acid biosynthesis presents drug targets in *Plasmodium falciparum*. Antimicrob Agents Chemother 47: 297-301.
210. Waller RF, Keeling PJ, Donald RG, Striepen B, Handman E, et al. (1998) Nuclear-encoded proteins target to the plastid in *Toxoplasma gondii* and *Plasmodium falciparum*. Proc Natl Acad Sci U S A 95: 12352-12357.
211. Lee PJ, Bhonsle JB, Gaona HW, Huddler DP, Heady TN, et al. (2009) Targeting the fatty acid biosynthesis enzyme, beta-ketoacyl-acyl carrier protein synthase III (PfKASIII), in the identification of novel antimalarial agents. J Med Chem 52: 952-963.
212. Kicska GA, Tyler PC, Evans GB, Furneaux RH, Schramm VL, et al. (2002) Purine-less death in *Plasmodium falciparum* induced by immucillin-H, a transition state analogue of purine nucleoside phosphorylase. J Biol Chem 277: 3226-3231.
213. Cassera MB, Hazleton KZ, Merino EF, Obaldia N, 3rd, Ho MC, et al. (2011) *Plasmodium falciparum* parasites are killed by a transition state analogue of purine nucleoside phosphorylase in a primate animal model. PLoS One 6: e26916.
214. Seymour KK, Lyons SD, Phillips L, Rieckmann KH, Christopherson RI (1994) Cytotoxic effects of inhibitors of *de novo* pyrimidine biosynthesis upon *Plasmodium falciparum*. Biochemistry 33: 5268-5274.
215. Krungkrai SR, Aoki S, Palacpac NM, Sato D, Mitamura T, et al. (2004) Human malaria parasite orotate phosphoribosyltransferase: functional expression, characterization of kinetic reaction mechanism and inhibition profile. Mol Biochem Parasitol 134: 245-255.
216. Dawson PA, Cochran DA, Emmerson BT, Gordon RB (1993) Inhibition of *Plasmodium falciparum* hypoxanthine-guanine phosphoribosyltransferase mRNA by antisense oligodeoxynucleotide sequence. Mol Biochem Parasitol 60: 153-156.
217. Li CM, Tyler PC, Furneaux RH, Kicska G, Xu Y, et al. (1999) Transition-state analogs as inhibitors of human and malarial hypoxanthine-guanine phosphoribosyltransferases. Nat Struct Biol 6: 582-587.

218. Artz JD, Wernimont AK, Dunford JE, Schapira M, Dong A, et al. (2011) Molecular characterization of a novel geranylgeranyl pyrophosphate synthase from *Plasmodium* parasites. J Biol Chem 286: 3315-3322.
219. Triglia T, Menting JG, Wilson C, Cowman AF (1997) Mutations in dihydropteroate synthase are responsible for sulfone and sulfonamide resistance in *Plasmodium falciparum*. Proc Natl Acad Sci U S A 94: 13944-13949.
220. Zhang Y, Meshnick SR (1991) Inhibition of *Plasmodium falciparum* dihydropteroate synthetase and growth *in vitro* by sulfa drugs. Antimicrob Agents Chemother 35: 267-271.
221. Haider N, Eschbach ML, Dias Sde S, Gilberger TW, Walter RD, et al. (2005) The spermidine synthase of the malaria parasite *Plasmodium falciparum*: molecular and biochemical characterisation of the polyamine synthesis enzyme. Mol Biochem Parasitol 142: 224-236.
222. Becker JV, Mtwisha L, Crampton BG, Stoychev S, van Brummelen AC, et al. (2010) *Plasmodium falciparum* spermidine synthase inhibition results in unique perturbation-specific effects observed on transcript, protein and metabolite levels. BMC Genomics 11: 235.
223. Fritz-Wolf K, Becker A, Rahlfs S, Harwaldt P, Schirmer RH, et al. (2003) X-ray structure of glutathione S-transferase from the malarial parasite *Plasmodium falciparum*. Proc Natl Acad Sci U S A 100: 13821-13826.
224. Harwaldt P, Rahlfs S, Becker K (2002) Glutathione S-transferase of the malarial parasite *Plasmodium falciparum*: characterization of a potential drug target. Biol Chem 383: 821-830.
225. Liebau E, Bergmann B, Campbell AM, Teesdale-Spittle P, Brophy PM, et al. (2002) The glutathione S-transferase from *Plasmodium falciparum*. Mol Biochem Parasitol 124: 85-90.
226. Perbandt M, Burmeister C, Walter RD, Betzel C, Liebau E (2004) Native and inhibited structure of a Mu class-related glutathione S-transferase from *Plasmodium falciparum*. J Biol Chem 279: 1336-1342.
227. Roberts F, Roberts CW, Johnson JJ, Kyle DE, Krell T, et al. (1998) Evidence for the shikimate pathway in apicomplexan parasites. Nature 393: 801-805.
228. Specht S, Sarite SR, Hauber I, Hauber J, Gorbig UF, et al. (2008) The guanylhydrazone CNI-1493: an inhibitor with dual activity against malaria-inhibition of host cell pro-inflammatory cytokine release and parasitic deoxyhypusine synthase. Parasitol Res 102: 1177-1184.
229. Fletcher S, Cummings CG, Rivas K, Katt WP, Horney C, et al. (2008) Potent, *Plasmodium*-selective farnesyltransferase inhibitors that arrest the growth of malaria parasites: structure-activity relationships of ethylenediamine-analogue scaffolds and homology model validation. J Med Chem 51: 5176-5197.
230. Choubey V, Maity P, Guha M, Kumar S, Srivastava K, et al. (2007) Inhibition of *Plasmodium falciparum* choline kinase by hexadecyltrimethylammonium bromide: a possible antimalarial mechanism. Antimicrob Agents Chemother 51: 696-706.
231. Ke H, Morrisey JM, Ganesan SM, Mather MW, Vaidya AB (2012) Mitochondrial RNA polymerase is an essential enzyme in erythrocytic stages of *Plasmodium falciparum*. Mol Biochem Parasitol 185: 48-51.
232. Lauer SA, Ghori N, Haldar K (1995) Sphingolipid synthesis as a target for chemotherapy against malaria parasites. Proc Natl Acad Sci U S A 92: 9181-9185.
233. Crooke A, Diez A, Mason PJ, Bautista JM (2006) Transient silencing of *Plasmodium falciparum* bifunctional glucose-6-phosphate dehydrogenase- 6-phosphogluconolactonase. FEBS J 273: 1537-1546.

234. Preuss J, Maloney P, Peddibhotla S, Hedrick MP, Hershberger P, et al. (2012) Discovery of a *Plasmodium falciparum* glucose-6-phosphate dehydrogenase 6-phosphogluconolactonase inhibitor (R,Z)-N-((1-ethylpyrrolidin-2-yl)methyl)-2-(2-fluorobenzylidene)-3-oxo-3,4-dihydr o-2H-benzo[b][1,4]thiazine-6-carboxamide (ML276) that reduces parasite growth *in vitro*. J Med Chem 55: 7262-7272.
235. Ogwan'g R, Mwangi J, Gachihi G, Nwachukwu A, Roberts CR, et al. (1993) Use of pharmacological agents to implicate a role for phosphoinositide hydrolysis products in malaria gamete formation. Biochem Pharmacol 46: 1601-1606.
236. Hanada K, Palacpac NM, Magistrado PA, Kurokawa K, Rai G, et al. (2002) *Plasmodium falciparum* phospholipase C hydrolyzing sphingomyelin and lysocholinephospholipids is a possible target for malaria chemotherapy. J Exp Med 195: 23-34.
237. Brand V, Koka S, Lang C, Jendrossek V, Huber SM, et al. (2008) Influence of amitriptyline on eryptosis, parasitemia and survival of *Plasmodium berghei*-infected mice. Cell Physiol Biochem 22: 405-412.
238. Taylor CJ, McRobert L, Baker DA (2008) Disruption of a *Plasmodium falciparum* cyclic nucleotide phosphodiesterase gene causes aberrant gametogenesis. Mol Microbiol 69: 110-118.
239. Kitade Y, Kozaki A, Gotoh T, Miwa T, Nakanishi M, et al. (1999) Synthesis of S-adenosyl-L-homocysteine hydrolase inhibitors and their biological activities. Nucleic Acids Symp Ser: 25-26.
240. Shuto S, Minakawa N, Niizuma S, Kim HS, Wataya Y, et al. (2002) New neplanocin analogues. 12. Alternative synthesis and antimalarial effect of (6'R)-6'-C-methylneplanocin A, a potent AdoHcy hydrolase inhibitor. J Med Chem 45: 748-751.
241. Bujnicki JM, Prigge ST, Caridha D, Chiang PK (2003) Structure, evolution, and inhibitor interaction of S-adenosyl-L-homocysteine hydrolase from *Plasmodium falciparum*. Proteins 52: 624-632.
242. Seymour KK, Yeo AE, Rieckmann KH, Christopherson RI (1997) dCTP levels are maintained in *Plasmodium falciparum* subjected to pyrimidine deficiency or excess. Ann Trop Med Parasitol 91: 603-609.
243. Krungkrai J, Krungkrai SR, Phakanont K (1992) Antimalarial activity of orotate analogs that inhibit dihydroorotase and dihydroorotate dehydrogenase. Biochem Pharmacol 43: 1295-1301.
244. Tyler PC, Taylor EA, Frohlich RF, Schramm VL (2007) Synthesis of 5'-methylthio coformycins: specific inhibitors for malarial adenosine deaminase. J Am Chem Soc 129: 6872-6879.
245. Ho MC, Cassera MB, Madrid DC, Ting LM, Tyler PC, et al. (2009) Structural and metabolic specificity of methylthiocoformycin for malarial adenosine deaminases. Biochemistry 48: 9618-9626.
246. Assaraf YG, Golenser J, Spira DT, Bachrach U (1984) Polyamine levels and the activity of their biosynthetic enzymes in human erythrocytes infected with the malarial parasite, *Plasmodium falciparum*. Biochem J 222: 815-819.
247. Das Gupta R, Krause-Ihle T, Bergmann B, Muller IB, Khomutov AR, et al. (2005) 3-Aminooxy-1-aminopropane and derivatives have an antiproliferative effect on cultured *Plasmodium falciparum* by decreasing intracellular polyamine concentrations. Antimicrob Agents Chemother 49: 2857-2864.
248. Muller IB, Das Gupta R, Luersen K, Wrenger C, Walter RD (2008) Assessing the polyamine metabolism of *Plasmodium falciparum* as chemotherapeutic target. Mol Biochem Parasitol 160: 1-7.

249. Scott HV, Gero AM, O'Sullivan WJ (1986) *In vitro* inhibition of *Plasmodium falciparum* by pyrazofurin, an inhibitor of pyrimidine biosynthesis *de novo*. Mol Biochem Parasitol 18: 3-15.
250. Krungkrai SR, DelFraino BJ, Smiley JA, Prapunwattana P, Mitamura T, et al. (2005) A novel enzyme complex of orotate phosphoribosyltransferase and orotidine 5'-monophosphate decarboxylase in human malaria parasite *Plasmodium falciparum*: physical association, kinetics, and inhibition characterization. Biochemistry 44: 1643-1652.
251. Meza-Avina ME, Wei L, Buhendwa MG, Poduch E, Bello AM, et al. (2008) Inhibition of orotidine 5'-monophosphate decarboxylase and its therapeutic potential. Mini Rev Med Chem 8: 239-247.
252. Wright PS, Byers TL, Cross-Doersen DE, McCann PP, Bitonti AJ (1991) Irreversible inhibition of S-adenosylmethionine decarboxylase in *Plasmodium falciparum*-infected erythrocytes: growth inhibition *in vitro*. Biochem Pharmacol 41: 1713-1718.
253. Ramya TN, Surolia N, Surolia A (2006) Polyamine synthesis and salvage pathways in the malaria parasite *Plasmodium falciparum*. Biochem Biophys Res Commun 348: 579-584.
254. Wanidworanun C, Nagel RL, Shear HL (1999) Antisense oligonucleotides targeting malarial aldolase inhibit the asexual erythrocytic stages of *Plasmodium falciparum*. Mol Biochem Parasitol 102: 91-101.
255. Reungprapavut S, Krungkrai SR, Krungkrai J (2004) *Plasmodium falciparum* carbonic anhydrase is a possible target for malaria chemotherapy. J Enzyme Inhib Med Chem 19: 249-256.
256. Ghosh AK, Coppens I, Gardsvoll H, Ploug M, Jacobs-Lorena M (2011) *Plasmodium* ookinetes coopt mammalian plasminogen to invade the mosquito midgut. Proc Natl Acad Sci U S A 108: 17153-17158.
257. Bonday ZQ, Dhanasekaran S, Rangarajan PN, Padmanaban G (2000) Import of host delta-aminolevulinate dehydratase into the malarial parasite: identification of a new drug target. Nat Med 6: 898-903.
258. Sharma SK, Kapoor M, Ramya TN, Kumar S, Kumar G, et al. (2003) Identification, characterization, and inhibition of *Plasmodium falciparum* beta-hydroxyacyl-acyl carrier protein dehydratase (FabZ). J Biol Chem 278: 45661-45671.
259. McRobert L, McConkey GA (2002) RNA interference (RNAi) inhibits growth of *Plasmodium falciparum*. Mol Biochem Parasitol 119: 273-278.
260. Bulusu V, Srinivasan B, Bopanna MP, Balaram H (2009) Elucidation of the substrate specificity, kinetic and catalytic mechanism of adenylosuccinate lyase from *Plasmodium falciparum*. Biochim Biophys Acta 1794: 642-654.
261. Thornalley PJ, Strath M, Wilson RJ (1994) Antimalarial activity *in vitro* of the glyoxalase I inhibitor diester, S-p-bromobenzylglutathione diethyl ester. Biochem Pharmacol 47: 418-420.
262. Urscher M, More SS, Alisch R, Vince R, Deponte M (2012) Tight-binding inhibitors efficiently inactivate both reaction centers of monomeric *Plasmodium falciparum* glyoxalase 1. FEBS J 279: 2568-2578.
263. Crane CM, Kaiser J, Ramsden NL, Lauw S, Rohdich F, et al. (2006) Fluorescent inhibitors for IspF, an enzyme in the non-mevalonate pathway for isoprenoid biosynthesis and a potential target for antimalarial therapy. Angew Chem Int Ed Engl 45: 1069-1074.
264. Barker RH, Jr., Metelev V, Rapaport E, Zamecnik P (1996) Inhibition of *Plasmodium falciparum* malaria using antisense oligodeoxynucleotides. Proc Natl Acad Sci U S A 93: 514-518.

265. Bodley AL, Cumming JN, Shapiro TA (1998) Effects of camptothecin, a topoisomerase I inhibitor, on *Plasmodium falciparum*. Biochem Pharmacol 55: 709-711.
266. Gamage SA, Tepsiri N, Wilairat P, Wojcik SJ, Figgitt DP, et al. (1994) Synthesis and *in vitro* evaluation of 9-anilino-3,6-diaminoacridines active against a multidrug-resistant strain of the malaria parasite *Plasmodium falciparum*. J Med Chem 37: 1486-1494.
267. Chavalitshewinkoon-Petmitr P, Pongvilairat G, Auparakkitanon S, Wilairat P (2000) Gametocytocidal activity of pyronaridine and DNA topoisomerase II inhibitors against multidrug-resistant *Plasmodium falciparum in vitro*. Parasitol Int 48: 275-280.
268. Noonpakdee W, Pothikasikorn J, Nimitsantiwong W, Wilairat P (2003) Inhibition of *Plasmodium falciparum* proliferation *in vitro* by antisense oligodeoxynucleotides against malarial topoisomerase II. Biochem Biophys Res Commun 302: 659-664.
269. Salcedo E, Cortese JF, Plowe CV, Sims PF, Hyde JE (2001) A bifunctional dihydrofolate synthetase--folylpolyglutamate synthetase in *Plasmodium falciparum* identified by functional complementation in yeast and bacteria. Mol Biochem Parasitol 112: 239-252.
270. Wang P, Wang Q, Yang Y, Coward JK, Nzila A, et al. (2010) Characterisation of the bifunctional dihydrofolate synthase-folylpolyglutamate synthase from *Plasmodium falciparum*; a potential novel target for antimalarial antifolate inhibition. Mol Biochem Parasitol 172: 41-51.
271. Platel DF, Mangou F, Tribouley-Duret J (1999) Role of glutathione in the detoxification of ferriprotoporphyrin IX in chloroquine resistant *Plasmodium berghei*. Mol Biochem Parasitol 98: 215-223.
272. Meierjohann S, Walter RD, Muller S (2002) Regulation of intracellular glutathione levels in erythrocytes infected with chloroquine-sensitive and chloroquine-resistant *Plasmodium falciparum*. Biochem J 368: 761-768.
273. Raman J, Mehrotra S, Anand RP, Balaram H (2004) Unique kinetic mechanism of *Plasmodium falciparum* adenylosuccinate synthetase. Mol Biochem Parasitol 138: 1-8.
274. McConkey GA (2000) *Plasmodium falciparum*: isolation and characterisation of a gene encoding protozoan GMP synthase. Exp Parasitol 94: 23-32.
275. Flores MV, Atkins D, Wade D, O'Sullivan WJ, Stewart TS (1997) Inhibition of *Plasmodium falciparum* proliferation *in vitro* by ribozymes. J Biol Chem 272: 16940-16945.
276. Sumanadasa SD, Goodman CD, Lucke AJ, Skinner-Adams T, Sahama I, et al. (2012) Antimalarial activity of the anticancer histone deacetylase inhibitor SB939. Antimicrob Agents Chemother 56: 3849-3856.
277. Andrews KT, Tran TN, Fairlie DP (2012) Towards histone deacetylase inhibitors as new antimalarial drugs. Curr Pharm Des 18: 3467-3479.
278. Pieperhoff MS, Pall GS, Jimenez-Ruis E, Das S, Wong EH, et al. (2014) Conditional U1 Gene Silencing in *Toxoplasma gondii*.
279. Donald RGK, Carter D, Ullman B, Roos DS (1996) Insertional tagging, cloning, and expression of the *Toxoplasma gondii* hypoxanthine-xanthine-guanine phosphoribosyltransferase gene - Use as a selectable marker for stable transformation. Journal of Biological Chemistry 271: 14010-14019.
280. Plattner F, Yarovinsky F, Romero S, Didry D, Carlier MF, et al. (2008) *Toxoplasma* profilin is essential for host cell invasion and TLR11-dependent induction of an interleukin-12 response. Cell Host & Microbe 3: 77-87.
281. DeRocher AE, Coppens I, Karnataki A, Gilbert LA, Rome ME, et al. (2008) A thioredoxin family protein of the apicoplast periphery identifies abundant candidate transport vesicles in *Toxoplasma gondii*. Eukaryotic Cell 7: 1518-1529.

282. Hettmann C, Herm A, Geiter A, Frank B, Schwarz E, et al. (2000) A dibasic motif in the tail of a class XIV apicomplexan myosin is an essential determinant of plasma membrane localization. Molecular Biology of the Cell 11: 1385-1400.

Curriculum Vitae

+41 78 635 25 35
stepan.tymoshenko@epfl.ch
http://people.epfl.ch/stepan.tymoshenko


Photo by David Schweizer

Stepan Tymoshenko

Avenue de Rosemont 3,
Lausanne, CH-1006
Vaud, Switzerland

Profile

- Solid competence in biochemistry, metabolism and computational metabolic modelling
- Interdisciplinary research expertise, PhD in systems biology and infection biology
- Intercultural experience: work on the interface of two multinational research teams

Looking for a challenging project in the areas of medical biotechnology or metabolism

Skills

Computational:
programming in Matlab environment
optimisation (LP, MILP using IBM CPLEX)
high-performance computing (SLURM)
processing of large datasets (Matlab, Excel)
Basic familiarity: PERL, MS VBA, R

Metabolic modelling:
reconstruction of metabolic networks
flux-balance analysis (FBA)
thermodynamics-based FBA
integration of metabolomics data
gene/enzyme essentiality prediction

Wet-lab skills:
cell culturing (biosafety level P2 lab)
gene knock-in/out (via homologous recombination or CrisprCas9-mediated)
PCR, gel-electrophoresis, gel-doc
DNA, protein extraction, Western blot
light and fluorescence microscopy
keeping lab records

Soft skills:
creativity techniques (brainstorm, flashlight)
scientific communication (oral presentations, presentation of science for general audiences)
preparation of scientific grant applications
scientific writing (reports, publications)
establishing a productive teamwork
maintenance of community relations

Experience

2011-Present **Research assistant at EPFL, Laboratory of Computational Systems Biotechnology and University of Geneva, Department of Microbiology and Molecular Medicine**

My interdisciplinary PhD project aimed at facilitation of developing novel antimalarial and anti-toxoplasmosis medicines. I reconstructed the first genome-scale metabolic model of *Toxoplasma gondii*. Among the other aspects of the pathogen's physiology this model predicts which metabolic enzymes are indispensable for replication of the pathogen. I validated several of my *in silico* predictions using genetic approaches to knock-out corresponding genes. This experimentally confirmed essentiality represents a valuable input for design of new anti-toxoplasmosis treatment. I have also reconciled existing metabolic models of the malaria parasite *Plasmodium falciparum* and extended their scope with integration of thermodynamics-based constraints and quantitative metabolomics data. This model as well as other results produced by me within this PhD project laid the groundwork for a large interdisciplinary project (***MalarX.ch***) that aims at addressing the biggest challenge in eradication of malaria - dormancy.

"Functional genomics of Plasmodium falciparum using metabolic modelling and analysis" ***Tymoshenko S****, Oppenheim RD, Soldati-Favre D, Hatzimanikatis V., Brief Funct Genomics. 2013 Jul;12(4):316-27. doi: 10.1093/bfgp/elt017.*

"Metabolic needs and capabilities of Toxoplasma gondii through combined computational and experimental analysis" Tymoshenko et al., manuscript under peer-review

"Thermodynamics-based flux balance analysis of metabolism in Plasmodium falciparum" Tymoshenko et al., manuscript in preparation

2009-2010 **Junior research assistant** (full-time) at the Department of Genetics of Microorganisms, Institute of Microbiology and Virology, Kyiv, Ukraine.

My aim on this position was to investigate a putative regulatory mechanism of landomycin E production in *Streptomyces globisporus*. I performed screening of over 50 uncharacterised species of streptomycetes for ability to trigger production of landomycin E in the strain with a defect in regulation. Among those I identified a number of effective strains and chromatographically isolated from them the substances, which enabled biosynthesis of the antibiotic in the otherwise non-producing strain *S. globisporus* B2. My contribution concurred the proposed mechanism and further allowed elucidation of the chemical structure of this novel master-regulator molecule in streptomycetes.

*"Screening and characteristics of regulators of landomycin E biosynthesis in Streptomyces globisporus" Matseliukh BP, **Tymoshenko SH**, Bambura OI, Kopeiko OP. Mikrobiol Z. 2011 Sep-Oct;73(5):16-20. Article in Ukrainian*

2007-2009 **Junior research assistant** (part-time) at the Department of Genetics of Microorganisms, Institute of Microbiology and Virology, Kyiv, Ukraine.

Within my master project I studied biosynthesis of carotenoid lycopene in a mutant strain of a soil bacterium *Streptomyces globisporus.* As a result of systematic trials of various compositions of the growth medium I increased the yield of the carotenoid by 2.8-fold. I also investigated the oxidative stress resistance in this bacterium and the correlation of an oxidative stress with production of lycopene.

*"Influence of carbon and nitrogen sources on biosynthesis of lycopene by Streptomyces globisporus 4Lcp" Holembiovs'ka SL, **Tymoshenko SH**, Matseliukh BP. Mikrobiol Z. 2010 Nov-Dec;72(6):46-51. Article in Ukrainian.*

Education

2011-2015 École Polytechnique Fédérale de Lausanne (Swiss Federal Institute of Technology in Lausanne), Doctoral school of Biotechnology and Bioengineering, PhD degree expected in June 2015

2009-2010 Zabolotny Institute of Microbiology and Virology, National Academy of Sciences of Ukraine Doctoral student ("aspirant") at the Department of Genetics of Microorganisms.

2004-2009 Bachelor and Master program at the National University of Food Technologies, Kyiv, Ukraine; Master Degree in Biotechnology, honours diploma

Languages

English (Professional working proficiency)

French (Limited proficiency, B1 level),

Ukrainian (Native or bilingual proficiency)

Russian (Native or bilingual proficiency)

Extracurricular activities

Natural curiosity and willing to go beyond limits make me a passionate traveller. Fitness is a mean of challenging myself and training the habit of making regular incremental efforts towards long-term aims.

My top 5 strengths (StrengthsFinder 2.0 by Tom Rath):

Ideation, Focus, Futuristic, Command, Learner (short description here: goo.gl/U1vN8R)

Appendix 1. Single gene/enzyme deletions predicted to cause impairment of metabolic network of *P. falciparum*

Gene ID (old/new)	Enzyme	E.C. number	Yeh et al.	Fatumo et al.	Huthmacher et al.	Plata et al.	Bazzani et al.	Experimental anlaysis based on drugs or genetic approaches	Influence
PF11_0059/PF3D7_1104800	Pantothenate transporter	–	–	–	–	+	–	[170]	Drug
PF11_0169/PF3D7_1116200.1	Pyridoxine/pyridoxal 5-phosphate biosynthesis enzyme	–	–	–	–	+	–	No evidence	
PFE0410w/PF3D7_0508300	Dihydroxyacetone phosphate transporter (apicoplast)	–	–	–	–	+	–	[171]	Genetic (Essential)
PFE1510c/PF3D7_0530200	Phosphoenolpyruvate transporter (apicoplast)	–	–	–	–	+	–	[171] non essential	Genetic (non- essential)
PFF0450c/PF3D7_0609100	Iron transporter	–	–	–	–	+	–	No evidence	
MAL13P1.206/PF3D7_1340900	Phosphate transporter (cytoplasm)	–	–	–	–	**SGR**	–	[88]	
PFI1020c/PF3D7_0920800	IMP dehydrogenase	1.1.1.205	–	–	–	–	+	[172]	Drug
ORPHAN	Shikimate dehydrogenase	1.1.1.25	–	–	–	+	–	No evidence	
PF14_0641/PF3D7_1467300	1-deoxy-D-xylulose 5-phosphate reductoisomerase	1.1.1.267	–	–	+	+	–	[173-175]	Drug - Genetic
PF13_0141/PF3D7_1324900 | PF13_0144 (new annotated PF3D7_1325200 oxidoreductase)	lactate dehydrogenase	1.1.1.27	–	–	–	+	+	[176] [177]	Drug (Highly toxic drug)
PF14_0520/PF3D7_1454700	6-phosphogluconate dehydrogenase (decarboxylating)	1.1.1.44	–	–	–	**SGR**	–	No evidence	
PF14_0373/PF3D7_1439400 & PF10_0120/PF3D7_1012300 & PF14_0248/PF3D7_1426900	mitochondrial ubiquinol-cytochrome C reductase	1.10.2.2	–	–	+	–	+	[178]	Drug
ORPHAN	2-octaprenylphenol hydroxylase	1.14.13.-	–	–	+	+	–	No evidence	
PFF1130c/PF3D7_0623500 | PF08_0071/PF3D7_0814900	superoxide dismutase	1.15.1.1	–	–	+	–	+	[179]	Drug

PFA0225w/PF3D7_0104400	4-hydroxy-3-methylbut-2-enyl diphosphate reductase	1.17.1.2	−	−	−	+	−	[180]	
PF14_0352/PF3D7_1437200 & PF10_0154/PF3D7_1015800 | PF14_0053/PF3D7_1405600	ribonucleotide reductase	1.17.4.1	+	+	+	+	+	[181, 182]	Drug
PF10_0221/PF3D7_1022800	(E)-4-hydroxy-3-methylbut-2-enyl-diphosphate synthase	1.17.7.1	−	−	−	+	−	No evidence	
PF14_0598/PF3D7_1462800	Glyceraldehyde-3-phosphate dehydrogenase (phosphorylating)	1.2.1.12	−	−	−	+	−	No evidence	
PF13_0070/PF3D7_1312600 | PFE0225w/PF3D7_0504600	3-Methyl-2-oxobutanoate dehydrogenase (BCKDH)	1.2.4.4	−	+	−	−	−	No evidence	
PFF0730c/PF3D7_0615100	enoyl-ACP-reductase	1.3.1.9	+	−	+	−	−	[183, 184]	Genetic
PF11_0436/PF3D7_1142400	Coproporphyrinogen oxidase	1.3.3.3	−	−	+	+	−	No evidence	
PF10_0275/PF3D7_1028100	Protoporphyrinogen oxidase	1.3.3.4	−	−	−	+	−	[185, 186]	Drug
PFF0160c/PF3D7_0603300	Dihydroorotate oxidase (fumarate)/dihydroorotate dehydrogenase	1.3.3.1/1.3.98.1	−	−	+	+	−	[187-190]	Drug
PFD0830w/PF3D7_0417200	dihydrofolate reductase	1.5.1.3	−	+	−	+	+	[191, 192]	Drug
PFF1490w/PF3D7_0630700	Methylenetetrahydrofolate dehydrogenase (NADP+), methenyltetrahydrofolate cyclohydrolase	1.5.1.5, 3.5.4.9	−	−	−	+	−	[192].	Drug
PFI0735c/PF3D7_0915000	NADH dehydrogenase	1.6.99.3	+	+	−	−	+	[193, 194]	Drug
PF14_0192/PF3D7_1419800.1	glutathione reductase	1.8.1.7	−	−	+	−	+	[195-198]	Drug
PFI1170c/PF3D7_0923800.1	thioredoxin reductase	1.8.1.9	−	−	+	+	+	[199]	Genetic
multiple genes	cytochrome C oxidase	1.9.3.1	−	−	−	+	+	No evidence	

PFL1780w/PF3D7_1236800	Protein S-isoprenylcysteine-O-methyltransferase	2.1.1.100	–	–	+	–	–	No evidence	
MAL13P1.214/PF3D7_1343000	phosphoethanolamine N-methyltransferase	2.1.1.103	–	–	–	–	+	[200]	Genetic
PFD0830w/PF3D7_0417200	thymidylate synthase	2.1.1.45	+	+	+	+	+	[191, 201]	Drug
MAL7P1.130/PF3D7_0724300 | PFI0815c/PF3D7_0916600	3-Demethylubiquinone-9,3-O-methyltransferase	2.1.1.64	–	+	+	–	–	No evidence	
PF14_0534/PF3D7_1456100 | PFL1720w/PF3D7_1235600	serine hydroxymethyltransferase	2.1.2.1	–	–	–	–	+	No evidence	
MAL13P1.221/PF3D7_1344800	aspartate carbamoyltransferase	2.1.3.2	–	–	+	+	+	[202]	Drug
PFF0530w/PF3D7_0610800	Transketolase	2.2.1.1	–	–	–	+	–	No evidence	
MAL13P1.186/PF3D7_1337200	1-deoxy-D-xylulose-5-phosphate synthase	2.2.1.7	–	–	–	+	–	[174, 203]	Drug
PFL0620c/PF3D7_1212500 | PF13_0100/PF3D7_1318200	Glycerol 3-phosphate O-Acyltransferase	2.3.1.15	–	+	–	–	+	[204, 205]	Drug
Putatively: PFE0405c/PF3D7_0508200 or PF14_0034/PF3D7_1403700	sphingosine N-acyltransferase	2.3.1.24	–	–	+	–	+	[206]	Drug
PFL2210w/PF3D7_1246100	5-aminolevulinate synthase	2.3.1.37	–	–	+	+	+	[207, 208]	Drug
PFB0505c/PF3D7_0211400 | PFF1275c/PF3D7_0626300	beta-ketoacyl-ACP synthase	2.3.1.41	+	+	+	–	+	[183, 209-211]	Drug
PF14_0155/PF3D7_1415700	serine C-palmitoyltransferase	2.3.1.50	–	–	+	–	+	[206]	Drug
PFE0660c/PF3D7_0513300	purine nucleoside phosphorylase	2.4.2.1	+	+	–	**SGR**	+	[212, 213]	Drug
PFE0630c/PF3D7_0512700	orotate phosphoribosyltransferase	2.4.2.10	–	–	+	+	+	[214, 215]	Drug
PFF1410c/PF3D7_0629100	Nicotinate phosphoribosyltransferase	2.4.2.11	–	+	–	–	–	No evidence	
PF10_0121/PF3D7_1012400	hypoxanthine guanine phosphoribosyltransferase	2.4.2.8	+	+	–	+	+	[216, 217]	Drug
PF11_0295/PF3D7_1128400	Geranyltranstransferase, dimethylallyltranstransferase	2.5.1.10, 2.5.1.1	–	–	–	+	–	[218]	Drug

PF08_0095/PF3D7_0810800	dihydropteroate synthase	2.5.1.15	−	+	+	+	−	[219, 220]	Drug
PF11_0301/PF3D7_1129000	Spermidine synthase	2.5.1.16	−	+	−	+	+	[221, 222]	Drug
PF14_0187/PF3D7_1419300	glutathione-S transferase	2.5.1.18	−	+	+	−	−	[223-226]	Drug
PFB0280w/PF3D7_0206400	3-Phosphoshikimate 1-carboxyvinyltransferase	2.5.1.19	−	+	+	+	−	[227]	Drug
ORPHAN	Farnesyl-diphosphate famesyltransferase	2.5.1.21	+	+	−	−	−	No evidence	
PF14_0125/PF3D7_1412600	deoxyhypusine synthase	2.5.1.46	−	+	+	−	−	[228]	Drug
ORPHAN	3-deoxy-7-phosphoheptulonate synthase	2.5.1.54	−	−	−	+	−	No evidence	
PFF0120w/PF3D7_0602500& PF11_0483/PF3D7_1147500	protein farnesyltransferase	2.5.1.58 (2.5.1.59)	−	−	+	−	−	[229]	Drug
PFI1090w/PF3D7_0922200	S-adenosylmethionine synthetase	2.5.1.6	−	−	−	+	+	No evidence	
PFB0200c/PF3D7_0204500	Aspartate transaminase, tyrosine transaminase, phenylalanine(histidine) transaminase	2.6.1.1, 2.6.1.5, 2.6.1.58	−	−	−	+	−	No evidence	
PFI1100w/PF3D7_0922400	aminodeoxychorismate synthase	2.6.1.85	−	−	−	+	−	No evidence	
PFF1155w/PF3D7_0624000	hexokinase	2.7.1.1	−	+	−	−	+	No evidence	
PFE0150c/PF3D7_0503100	4-diphosphocytidyl-2C-methyl-D-erythritol kinase	2.7.1.148	−	−	+	+	−	No evidence	
Putatively: PFI0650c/PF3D7_0913300	NAD+ kinase	2.7.1.23	−	−	−	+	−	No evidence	
PF14_0415/PF3D7_1443700	Dephospho-CoA kinase	2.7.1.24	−	−	−	+	−	No evidence	
PF14_0020/PF3D7_1401800	choline kinase	2.7.1.32	−	+	+	−	+	[230]	Drug
PF10_0363/PF3D7_1037100	Pyruvate kinase	2.7.1.40	−	−	−	+	−	No evidence	
PFI1105w/PF3D7_0922500	Phosphoglycerate kinase	2.7.2.3	−	−	−	+	−	No evidence	
PFI1420w/PF3D7_0928900	Guanylate kinase	2.7.4.8	−	−	−	+	−	No evidence	
PFL2465c/PF3D7_1251300	dTMP kinase	2.7.4.9	−	+	−	+	−	No evidence	
PF13_0157/PF3D7_1327800 | PF13_0143/PF3D7_1325100	phosphoribosyl pyrophosphate synthase	2.7.6.1	−	−	−	+	+	No evidence	
PFI1195c/PF3D7_0924300	Thiamine diphosphokinase	2.7.6.2	−	−	−	+	−	No evidence	

MAL13P1.86/PF3D7_1316600	choline-phosphate cytidylyltransferase	2.7.7.15	−	−	+	**SGR**	+	No evidence	
PF13_0159/PF3D7_1327600	Nicotinate-nucleotide adenylyltransferase	2.7.7.18	−	−	−	+	−	[27]	
MAL13P1.292/PF3D7_1359100	FAD synthetase, riboflavin kinase	2.7.7.2, 2.7.1.26	−	+	+	+	−	No evidence	
PF07_0018/PF3D7_0704700	Pantetheine-phosphate adenylyltransferase	2.7.7.3	−	−	+	+	−	No evidence	
PF14_0097/PF3D7_1409900	phosphatidate cytidylyltransferase	2.7.7.41	−	−	−	**SGR**	−	No evidence	
multiple genes	RNA polymerase	2.7.7.6	+	−	−	−	−	[231]	
PFA0340w/PF3D7_0106900	2-C-methyl-D-erythritol 4-phosphate cytidylyltransferase	2.7.7.60	−	−	+	+	−	No evidence	
multiple genes	DNA polymerase	2.7.7.7	+	−	−	−	−	No evidence	
PFF1375c/PF3D7_0628300	Ethanolaminephosphotransferase, diacylglycerol cholinephosphotransferase	2.7.8.1, 2.7.8.2	−	−	−	**SGR**	−	No evidence	
MAL13P1.82/PF3D7_1315600	CDP-diacylglycerol-inositol 3-phosphatidyltransferase	2.7.8.11	−	+	−	**SGR**	−	No evidence	
PFF1215w/PF3D7_062510	Sphingomyelin synthase	2.7.8.27	−	−	−	**SGR**	−	[232]	
PFF1210w/PF3D7_0625000.1	sphingomyelin synthase/phosphatidic acid phosphatase check the new annotation	2.7.8.3	−	−	+	−	+	[206]	Drug
PF14_0511/PF3D7_1453800	6-phosphogluconolactonase, glucose-6-phosphate dehydrogenase	3.1.1.31, 1.1.1.49	−	−	−	**SGR**	−	[233, 234]	RNAi - Drug
multiple genes	lysophospholipase	3.1.1.5	+	−	−	−	−	Zidovetzki R, Sherman IW, Prudhomme J, Parasitology. 1994 Apr;108 (Pt 3):249-55	Drug
MAL8P1.151/PF3D7_0802500 | PF11_0122/PF3D7_1111600(endonuclease/exonuclease/phosphatase family protein) | PF13_0285/PF3D7_1354200	Inositol-1,4,5-trisphosphate 5-phosphase	3.1.3.56	−	+	−	−	−	[235]	Drug
PFL1870c/PF3D7_1238600	sphingomyelinase	3.1.4.12	+	+	−	**SGR**	−	[236, 237]	Drug
multiple genes	3',5'-cyclic-nucleotide phosphodiesterase	3.1.4.17	−	+	−	−	−	[238]	Genetic
PFE1050w/PF3D7_0520900	S-adenosyl-L-homocysteine hydrolase	3.3.1.1	+	+	+	−	+	[239-241]	Drug
PFF1190c/PF3D7_0624700 | PFI0535w/PF3D7_0911000	N-acetylglucosaminylphosphotidylinositol deacetylase	3.5.1.89	−	−	+	−	−	No evidence	

PF14_0697/PF3D7_1472900	dihydroorotase	3.5.2.3	–	–	+	+	+	[242, 243]	Drug
PF13_0259/PF3D7_1349400	dCTP deaminase	3.5.4.13	–	–	–	**SGR**	–	No evidence	
PF10_0289/PF3D7_1029600	adenosine deaminase	3.5.4.4	–	+	–	+	+	[244, 245]	Drug
PFE1035c/PF3D7_0520600	bis(5'-nucleosyl)-tetraphosphatase	3.6.1.17 (3.6.1.41)	–	+	–	–	–	No evidence	
PFF0370w/PF3D7_0607500	3-octaprenyl-4-hydroxybenzoate carboxy-lyase	4.1.1.-	–	–	–	+	–	No evidence	
ORPHAN	3-octaprenyl-4-hydroxybenzoate carboxy-lyase	4.1.1.-	–	–	–	+	–	No evidence	
PF10_0322/PF3D7_1033100	ornithine decarboxylase	4.1.1.17	–	+	–	+	+	[246-248]	Drug
PF10_0225/PF3D7_1023200	orotidine-5'-phosphate decarboxylase	4.1.1.23	–	+	+	+	+	[249-251]	Drug
MAL8P1.81/PF3D7_0816100	Phosphopantothenoylcysteine decarboxylase	4.1.1.36	–	–	–	+	–	No evidence	
PFF0360w/PF3D7_0607300	Uroporphyrinogen decarboxylase	4.1.1.37	–	–	+	+	–	No evidence	
PF10_0322/PF3D7_1033100	S-adenosylmethionine decarboxylase	4.1.1.50	+	+	–	+	+	[247, 248, 252, 253]	Drug
PF14_0425/PF3D7_1444800	aldolase	4.1.2.13	+	+	–	**GR**	+	[254]	antisense ODN
ORPHAN	Chorismate lyase	4.1.3.40	–	–	–	+	–	No evidence	
PF11_0411/PF3D7_1140000 (PF11_0410/PF3D7_1140000)	carbonate dehydratase	4.2.1.1	–	–	+	+	+	[255]	Drug
ORPHAN	3-dehydroquinate dehydratase	4.2.1.10	–	–	–	+	–	No evidence	
PF10_0155/PF3D7_1015900	Phosphopyruvate hydratase / enolase	4.2.1.11	–	–	–	+	–	[256]8.	Antibody
PF14_0381/PF3D7_1440300	delta-aminolevulinate dehydratase	4.2.1.24	–	+	+	+	+	[185, 257]	Drug

PF13_0128/PF3D7_1323000	3-hydroxyacyl-ACP dehydratase	4.2.1.58 – 61	–	–	+	+	–	[183, 258]	Drug - Genetic
PFL0480w/PF3D7_1209600	Uroporphyrinogen-III synthase	4.2.1.75	–	–	–	+	–	No evidence (reported by Plata et al as an orpahn we assign gene based on PMID: 17962188)	
ORPHAN	3-dehydroquinate synthase	4.2.3.4	–	–	–	+	–	No evidence	
PFF1105c/PF3D7_0623000	chorismate synthase	4.2.3.5	+	+	–	+	–	[259]	RNAi
PFB0295w/PF3D7_0206700	adenylosuccinate lyase	4.3.2.2	–	–	–	+	+	[260]	Drug
PFF0230c/PF3D7_0604700 | PF11_0145/PF3D7_1113700	glyoxylase	4.4.1.5	+	+	+	–	–	[261, 262]	Drug
PFB0420w/PF3D7_0209300	2C-methyl-D-erythritol 2,4-cyclodiphosphate synthase	4.6.1.12	–	–	+	+	–	[263]	Drug
MAL13P1.326/PF3D7_1364900	Ferrochelatase	4.99.1.1	–	–	–	+	–	No evidence	
PFL0960w/PF3D7_1219900	Ribulose-phosphate 3-epimerase	5.1.3.1	–	–	–	**GR**	–	No evidence	
PF14_0378 / PFC0831w	triosephosphate isomerase	5.3.1.1	–	–	–	**GR**	+	[264]	antisense oligodeoxynucleotides
PF14_0341/PF3D7_1436000	Glucose-6-phosphate isomerase	5.3.1.9	–	–	–	**SGR**	–	No evidence	
PF10_0122/PF3D7_1012500	Phosphoglucomutase	5.4.2.2	–	–	–	**SGR**	–	No evidence	
PFE0520c/PF3D7_0510500	topoisomerase I	5.99.1.2	–	–	+	–	–	[265]	Drug
PFE0520c/PF3D7_0510500 | PF13_0251/PF3D7_1347100	topoisomerase II	5.99.1.3	–	–	+	–	–	[266-268]	Drug
PF11_0270/PF3D7_1126000	threonine-tRNA ligase	6.1.1.3	–	+	+	–	–	No evidence	
PF13_0354/PF3D7_1367700	alanine-tRNA ligase	6.1.1.7	–	+	+	–	–	No evidence	
PFF1350c/PF3D7_0627800	acetyl-CoA synthetase	6.2.1.1	–	–	–	–	+	No evidence	
multiple genes	acyl-CoA synthetase	6.2.1.3	–	+	–	–	+	No evidence	
PF13_0140/PF3D7_1324800	dihydrofolate synthase	6.3.2.12	–	–	–	+	–	[269]	Drug
PF13_0140/PF3D7_1324800	folylpolyglutamate synthase	6.3.2.17	–	–	–	+	+	[269, 270]	Drug

PFI0925w/PF3D7_0918900	gamma-glutamylcysteine synthetase	6.3.2.2	+	+	+	−	+	[197, 271, 272]	Drug
PF14_0100/PF3D7_1410200	CTP synthase	6.3.4.2	−	−	−	−	+	No evidence	
PF13_0287/PF3D7_1354500	adenylosuccinate synthase	6.3.4.4	−	−	+	+	+	[273]	
PFI1310w/PF3D7_0926700	NAD+ synthase (glutamine-hydrolysing)	6.3.5.1	−	−	−	+	−	No evidence	
PF10_0123/PF3D7_1012600	GMP synthase	6.3.5.2	−	−	−	−	+	[274]	
PF13_0044/PF3D7_1308200	carbamoyl phosphate synthetase	6.3.5.5	+	+	+	+	+	[275]	Ribo -zyme
PF10_0409/PF3D7_1026900 | PF14_0664/PF3D7_1469600	acetyl-CoA carboxylase	6.4.1.2	−	−	+	−	+	[209]	Drug
multiple genes	histone deacetylase	none	+	−	−	−	−	[276, 277]	Drug

Abbreviations: GR: Growth Reducing, SGR: Slight Growth Reducing, "+": predicted as essential, "−": predicted as non-essential or not present in the model; " | ": delimiter for genes that code isozymes, "&": delimiter for genes that code subunits of multienzyme complex, "ORPHAN": no gene encoding this enzyme has been identified yet.

Appendix 2. Supplementary data for curation of *Plasmodium* models

modification name	change as a Cobra Toolbox command	reversibility assignments	results in reference to the previous change		
			biomass yield	gene essentiality	reaction essentiality
PCt	model=addReaction(model, 'PCt', 'atp[c] + h2o[c] + pc[e] <=> adp[c] + h[c] + pc[c] + pi[c]');	rev	0.5% decrease	No effect	No effect
PIOHex_mt	model=addReaction(model,'PIOHex_mt', 'h[c] + pi[c] <=> h[m] + pi[m]');	irrev	52.2% increase	+3 essential PF10_0155, PF14_0598	5 less essential Rxns: 'ENO', 'GAPD', 'PGK', 'PGM'
PIOHex_mt	model=changeRxnBounds(model, 'PIOHex_mt', -1000, 'l');	rev	0.01% increase	+1 essential PF13_0121	1 less essential Rxn: 'AKGDH_mt'
H2Ot_mt	model=addReaction(model, 'H2Ot_mt', 'h2o[c] <=> h2o[m]');	rev	No effect	No effect	No effect
Pyr_mt	model=addReaction(model,'Pyr_mt', 'pyr[c] <=> pyr[m]');	rev	No effect	No effect	No effect
PDH_mt	model=addReaction(model,'PDH_mt', 'nad[m] + coa[m] + pyr[m] <=> accoa[m] + co2[m] + nadh[m]');	irrev	No effect	No effect	4 less essential Rxn: 'LDH_L', 'EX_lac_L(e)', 'L_LACt2r', 'Htmt'
PDH_mt	model=changeRxnBounds(model, 'PDH_mt', 0, 'l');	not applicable	No effect	No effect	No effect
ACCOAtm	model=changeRxnBounds(model,'ACCOAtm',0,'b');	not applicable	No effect	No effect	2 additional essential Rxns: 'COAtap', 'Htmt'
ACCOAtap	model=changeRxnBounds(model,'ACCOAtap', 0, 'b');	not applicable	No effect	No effect	1 additional essential Rxn: 'PItap'
R00261	model=addReaction(model, 'R00261', 'glu_L[c] <=> 4abut[c] + co2[c]');	rev	3.27% increase	No effect	No effect
R00713	model=addReaction(model, 'R00713', 'h2o[m] + nad[m] + sucsal[m] <=> 2 h[m] + nadh[m] + succ[m]');	rev	No effect	No effect	No effect
R01648	model=addReaction(model, 'R01648', '4abut[m] + akg[m] <=> sucsal[m] + glu_L[m]');	rev	No effect	No effect	No effect
GABAt_mt	model=addReaction(model, 'GABAt_mt', '4abut[c] <=> 4abut[m]');	rev	No effect	No effect	1 less essential Rxn: 'Htmt'
CYOOm_mt	model=addReaction(model, 'CYOOm_mt', '4 focytc[m] + 6 h[m] + o2[m] -> 4 ficytc[m] + 2 h2o[m] + 2 h[c]');	irrev	No effect	No effect	No effect
CYOR_u6m_mt	model=addReaction(model, 'CYOR_u6m_mt', '1.5 h[m] + 2 ficytc[m] + q8h2[m] -> 3.5 h[c] + 2 focytc[m] + q8[m]');	irrev	No effect	No effect	No effect
ReactionsDeleted (look)	[Dups, nDups, model] = checkDuplicateRxn(model, 'true');	not applicable	No effect	No effect	No effect
DAGCPT	model=addReaction(model,'DAGCPT', 'cdpchol[c] + dag[c] => cmp[c] + pc[c]');	irrev	0.12% decrease	No effect	No effect
DAGCPT	model=changeRxnBounds(model, 'DAGCPT', -1000, 'l');	rev	No effect	No effect	No effect
PINOS	model=addReaction(model,'PINOS', 'cdpdag[c] + inost[c] <=> cmp[c] + ptd1ino[c]');	irrev	0.42% decrease	+2 essential MAL13P1_206	2 additional essential Rxns: 'CTPS2', 'PIt2r'
PINOS	model=changeRxnBounds(model, 'PINOS', -1000, 'l');	rev	0.77% increase	+2 essential MAL13P1_206	2 less essential Rxns: 'CTPS2', 'PIt2r'
R07456	model=addReaction(model,'R07456', 'g3p[c] + gln_L[c] + ru5p_D[c] <=> glu_L[c] + pydx5p[c] + 3 h2o[c] + pi[c]');	rev	No effect	No effect	No effect
ETHAPT	model=addReaction(model, 'ETHAPT', 'cdpea[c] +	rev	No effect	No effect	No effect

	dag[c] <=> cmp[c] + pe[c]');				
2_7_8_3	model=addReaction(model,'2_7_8_3', 'cdpchol[c] + crm[c] <=> cmp[c] + sphmyln[c]');	irrev	No effect	No effect	No effect
2_7_8_3	model=changeRxnBounds(model, '2_7_8_3', -1000, 'l');	rev	No effect	No effect	No effect
SMPD3l	model=addReaction(model, 'SMPD3l' , 'h2o[c] + sphmyln[c] <=> cholp[c] + crm[c]');	irrev	No effect	No effect	No effect
SMPD3l	model=changeRxnBounds(model, 'SMPD3l', -1000, 'l');	rev	No effect	No effect	No effect
PTHPS2	model=addReaction(model,'PTHPS2','ahdt[c] + h2o[c] -> 6hmhpt[c] + pppi[c] + ac[c]');	irrev	No effect	No effect	No effect
PTHPS2	model=changeRxnBounds(model, 'PTHPS2', -1000, 'l');	rev	No effect	No effect	No effect
255 Dead-end reactions	[blocked]=findBlockedReactions(model*); newmodel = removeRxns(model,blocked);	all reactions set reversible	No effect	No effect	No effect
change bounds to -100/100	[A,B]=ismember(newmodel.lb,-1000); newmodel.lb(find(A))=-100;	not applicable	90% decrease	No effect	No effect
71 blocked reactions	[blocked]=findBlockedReactions(newmodel); model=removeRxns(newmodel,blocked);	not applicable	No effect	No effect	No effect
MET_SLV	model=removeRxns(model,'MET_SLV');	not applicable	No effect	No effect	+2 essential rx: 'EX_spmd(e)', 'SPMDt2'
4 blocked reactions	[blocked]=findBlockedReactions(model); model=removeRxns(model,blocked);	not applicable	No effect	No effect	No effect
THZPSN_mt	model=addReaction(model,'THZPSN_mt', 'cys_L[m] + dxyl5p[m] + tyr_L[m] -> 4hba[m] + 4mpetz[m] + ala_L[m] + co2[m] + h2o[m]');	irrev	No effect	-1 essential: PFI1195c	2 additional essential Rxns: 'TMDPK', 'EX_thm(e)', 'THMt3'
THZPSN_mt	[~,b]=ismember('THZPSN_mt', model.rxns); model.grRules(b)={'((MAL7P1_150) or (PF07_0068)) and ((PFF1335c) or (PFL1920c))'};	not applicable	No effect	No effect	No effect
4MPETZ_mt	model=addReaction(model,'4MPETZ_mt', '4mpetz[m] <=> 4mpetz[c]'); [~,b]=ismember('4MPETZ_mt', model.rxns); model.lb(b)=-100;	rev	No effect	+1 essential: PFI1195c	2 less essential Rxns: 'TMDPK', 'EX_thm(e)', 'THMt3'
2 blocked reactions	[blocked]=findBlockedReactions(model); model=removeRxns(model,blocked);	not applicable	No effect	No effect	No effect
metSEEDID vector	[metSEEDIDs]=mapMetIDs(model,DB_AlbertyUpdate); model.metSEEDID=metSEEDIDs;	not applicable	No effect	No effect	No effect
metCompSymbol vector	model.metCompSymbol=metCompSymbol;	not applicable	No effect	No effect	No effect
CompartmentData	created	not applicable	No effect	No effect	No effect
Preparation to tFBA	tmodel = prepModelforTFBA(newmodel,DB_AlbertyUpdate, newmodel.CompartmentData);	not applicable	No effect	No effect	No effect
Conversion to tFBA	tmodel = convToTFBA(tmodel,DB_AlbertyUpdate,1,true);	not applicable	No effect	No effect	No effect
Removing one thermoconstrain	[~,b]=ismember(={'MECDPS_ap';'Htap'}π,tmodel.rxns); tmodel.rxnThermo(b)=0;	not applicable	No effect	No effect	No effect
Conversion to tFBA	tmodel = convToTFBA(tmodel,DB_AlbertyUpdate,1);	not applicable	No effect	No effect	1 less essential Rxn: 'Htap'
FBA final versus the initial model			84.3% increase	-4 essential PF10_0155, PF13_0121,	15 less essential Rxn: 'ENO', 'GAPD', 'LDH_L', 'PGK',
TFBA solution (versus FBA final)			No effect	+1 essential: PF13_0121	

Table 4. Table of the modifications made in iTH366 model prior to the TFBA studies

The table and contained information was produced by Anush Chiappino Pepe based on the initial studies done by Stepan Tymoshenko

Reaction ID	*ad hoc*	without biomass production			with biomass production			with metabolomics data		
		$\Delta_rG'_{min}$	$\Delta_rG'_{max}$		$\Delta_rG'_{min}$	$\Delta_rG'_{max}$		$\Delta_rG'_{min}$	$\Delta_rG'_{max}$	
ARGN	=>	-19.96	1.66	<=>	-19.96	1.66	<=>	-19.18	-3.10	=>
G5SADr	<=>	-22.04	-6.92	=>	-22.04	-6.92	=>	-22.04	-6.92	=>
ORNTA	=>	-18.34	8.08	<=>	-18.34	8.08	<=>	-10.63	4.86	<=>
P5CR	=>	-27.29	17.33	<=>	-17.82	17.25	<=>	-17.82	17.25	<=>
ASNS1	=>	-41.36	8.88	<=>	-41.36	7.76	<=>	-25.72	-7.78	=>
GLUSx	=>	-48.69	6.18	<=>	-48.69	6.18	<=>	-29.85	-2.50	=>
CBPSam	=>	-38.72	20.62	<=>	-36.51	11.17	<=>	-28.67	4.16	<=>
GLNS	=>	-20.79	22.79	<=>	-19.36	18.00	<=>	-11.49	7.70	<=>
GLUDxi	=>	-17.81	32.11	<=>	-17.81	32.10	<=>	-13.41	19.99	<=>
GLUDy	<=>	-11.40	38.48	<=>	-11.21	26.88	<=>	-8.00	19.17	<=>
GF6PTA	=>	-32.82	-1.93	=>	-32.82	-1.93	=>	-26.10	-6.22	=>
GLUCYS	=>	-24.58	20.72	<=>	-24.00	15.93	<=>	-19.39	7.11	<=>
GLYCL_mt	=>	-40.39	34.08	<=>	-40.39	33.96	<=>	-37.04	26.20	<=>
ADMDC	=>	-25.07	9.24	<=>	-25.07	9.23	<=>	-25.07	9.23	<=>
METAT	=>	-32.94	14.18	<=>	-32.94	-0.01	=>	-31.80	-0.01	=>
ORNDC	=>	-18.76	2.93	<=>	-18.76	2.93	<=>	-14.23	1.93	<=>
SPMS	=>	-31.17	9.40	<=>	-31.17	9.40	<=>	-25.24	4.28	<=>
1_7_1_1_mt	=>	-56.15	-13.26	=>	-56.15	-13.26	=>	-51.74	-14.45	=>
1_7_1_3_mt	=>	-62.11	-19.27	=>	-62.11	-19.27	=>	-51.84	-19.27	=>
NTRIR2x_mt	=>	-141.96	-41.12	=>	-141.96	-41.12	=>	-128.74	-44.68	=>
ENO	<=>	-8.34	5.75	<=>	-8.34	5.75	<=>	-8.34	5.75	<=>
FBA	<=>	-12.01	6.78	<=>	-11.11	6.78	<=>	-11.11	6.78	<=>
G3PD_mt	=>	-23.15	22.02	<=>	-23.15	22.02	<=>	-23.15	22.02	<=>
GAPD	<=>	-18.85	28.98	<=>	-18.85	27.00	<=>	-17.66	22.60	<=>
GLYK	=>	-20.57	9.31	<=>	-20.57	4.45	<=>	-15.55	0.66	<=>
HEX1	=>	-23.20	11.14	<=>	-23.20	-0.01	=>	-18.17	2.52	<=>
LDH_L	<=>	-15.07	26.66	<=>	-12.96	26.66	<=>	-10.38	18.88	<=>
PFK	=>	-22.74	10.71	<=>	-22.74	5.89	<=>	-17.72	2.11	<=>
PGI	<=>	-9.56	7.41	<=>	-9.56	7.41	<=>	-9.56	7.41	<=>
PGK	<=>	-14.13	15.51	<=>	-14.13	10.71	<=>	-9.11	6.95	<=>
PGM	<=>	-6.50	6.39	<=>	-6.50	6.39	<=>	-6.50	6.39	<=>
SBTR	=>	-31.27	13.53	<=>	-24.12	13.45	<=>	-22.66	13.45	<=>
TPI	<=>	-5.03	7.95	<=>	-5.03	7.05	<=>	-5.03	7.05	<=>
TPI_ap	<=>	-5.03	7.95	<=>	-4.97	-0.01	=>	-4.97	-0.01	=>
G3PD1ir	=>	-23.22	19.18	<=>	-23.22	18.28	<=>	-18.81	17.09	<=>
G3PD1ir_ap	=>	-23.73	18.66	<=>	-23.73	17.82	<=>	-19.33	16.64	<=>
HEX7	=>	-22.94	10.51	<=>	-22.94	5.71	<=>	-17.92	1.94	<=>
PYK	=>	-19.02	11.61	<=>	-15.01	11.61	<=>	-9.40	6.59	<=>
PYK_ap	=>	-22.45	14.38	<=>	-22.45	-0.01	=>	-16.77	-0.01	=>
HEX8	=>	-22.80	11.53	<=>	-22.80	6.69	<=>	-17.78	2.91	<=>
MAN6PI	<=>	-9.96	7.02	<=>	-9.96	7.02	<=>	-9.96	7.02	<=>
G6PDH2r	<=>	-18.36	27.10	<=>	-18.17	15.00	<=>	-18.17	15.50	<=>
GND	=>	-23.59	27.04	<=>	-22.96	15.44	<=>	-22.96	15.44	<=>
PGL	=>	-13.32	1.69	<=>	-13.32	1.69	<=>	-13.32	1.69	<=>
PGMT_2	<=>	-8.31	8.69	<=>	-8.31	8.69	<=>	-8.31	8.69	<=>

PRPPS	<=>	-21.68	10.40	<=>	-21.68	-0.01	=>	-17.37	-0.01	=>
RPE	<=>	-7.00	7.00	<=>	-7.00	7.00	<=>	-7.00	7.00	<=>
RPI	<=>	-6.30	7.71	<=>	0.01	7.71	<=	0.01	7.71	<=
TALA	<=>	-16.29	14.27	<=>	-16.29	12.24	<=>	-16.29	12.24	<=>
TKT1	<=>	-13.37	15.06	<=>	-13.37	15.06	<=>	-13.37	15.06	<=>
TKT2	<=>	-18.63	11.33	<=>	-18.63	11.33	<=>	-18.63	11.33	<=>
ACACT1r	<=>	-16.57	32.44	<=>	-16.57	32.44	<=>	-16.57	32.44	<=>
ACCOAL	=>	-28.66	26.63	<=>	-28.08	21.84	<=>	-21.43	15.02	<=>
ACS	=>	-31.33	25.31	<=>	-31.33	24.97	<=>	-24.56	14.74	<=>
MDH2_mt	=>	-32.89	10.46	<=>	-32.89	10.46	<=>	-31.61	1.71	<=>
PPC	=>	-20.51	6.35	<=>	-20.51	6.16	<=>	-15.59	6.16	<=>
PPCK	=>	-25.10	13.06	<=>	-25.10	8.35	<=>	-20.08	0.04	<=>
ASPTA	<=>	-14.33	12.04	<=>	-14.33	12.04	<=>	-0.47	4.50	<=>
ASPTA_mt	<=>	-14.32	12.04	<=>	-14.32	12.04	<=>	-5.38	-0.01	=>
FUM_mt	<=>	-7.11	5.86	<=>	-7.11	5.86	<=>	-0.13	1.75	<=>
MDHc	<=>	-15.16	26.77	<=>	-15.16	26.77	<=>	-7.78	16.23	<=>
PDH_ap	=>	-42.94	25.45	<=>	-42.94	25.45	<=>	-41.75	21.05	<=>
SUCOAS_mt	<=>	-34.17	29.39	<=>	-34.17	29.39	<=>	-26.94	20.26	<=>
FRD_mt	=>	-34.97	11.15	<=>	-34.97	11.15	<=>	-28.79	6.11	<=>
DPCOAK_ap	=>	-29.39	20.01	<=>	-28.43	-0.01	=>	-22.62	-0.01	=>
PNTK	=>	-21.42	10.19	<=>	-21.42	-0.01	=>	-16.39	-0.01	=>
PPCDC	=>	-20.86	5.03	<=>	-20.86	-0.01	=>	-20.86	-0.01	=>
PPNCL2	=>	-28.17	22.98	<=>	-28.17	-0.01	=>	-28.17	-0.01	=>
PTPATi	=>	-18.12	20.59	<=>	-18.12	-0.01	=>	-16.97	-0.01	=>
DHFR	<=>	-33.03	19.91	<=>	-22.04	-0.01	=>	-19.62	-0.01	=>
DHFS	=>	-29.86	23.34	<=>	-29.27	-0.01	=>	-24.67	-0.01	=>
DHPS2	=>	-29.15	7.44	<=>	-29.15	-0.01	=>	-29.15	-0.01	=>
GTPCI	=>	-37.58	-7.54	=>	-37.58	-7.54	=>	-34.62	-10.05	=>
HPPK2	=>	-24.76	13.31	<=>	-24.76	-0.01	=>	-20.45	-0.01	=>
MTHFC	<=>	-8.38	16.18	<=>	-8.38	-0.01	=>	-8.38	-0.01	=>
MTHFD	<=>	-23.10	29.86	<=>	-16.99	-0.01	=>	-16.99	-0.01	=>
ADCL	=>	-49.51	-26.94	=>	-49.51	-26.94	=>	-47.40	-26.94	=>
ADCS	=>	-31.05	-0.43	=>	-31.05	-0.43	=>	-24.33	-4.71	=>
NADK	=>	-34.91	10.73	<=>	-33.75	-0.01	=>	-27.54	-2.27	=>
NADS2	=>	-44.86	14.39	<=>	-44.86	-0.01	=>	-29.42	-0.01	=>
NAMNPP	=>	-45.34	10.53	<=>	-44.01	-0.01	=>	-40.31	-0.01	=>
NNAM	=>	-22.61	2.66	<=>	-22.61	2.31	<=>	-22.61	2.31	<=>
NNATr	<=>	-18.58	21.64	<=>	-18.58	-0.01	=>	-17.43	-0.01	=>
THD2pp	=>	-28.31	29.61	<=>	-28.12	18.01	<=>	-23.72	16.83	<=>
NADTRHD	=>	-35.35	22.56	<=>	-27.76	22.49	<=>	-26.57	18.09	<=>
ALAS_mt	=>	-29.84	16.85	<=>	-29.84	-0.01	=>	-27.67	-0.01	=>
HMBS_ap	=>	-37.91	42.01	<=>	-37.91	-0.01	=>	-37.91	-0.01	=>
PPBNGS_ap	=>	-17.49	6.35	<=>	-17.49	-0.01	=>	-17.49	-0.01	=>
ACP1	=>	-23.05	8.08	<=>	-23.05	7.90	<=>	-20.11	7.90	<=>
FMNAT	=>	-34.42	8.31	<=>	-34.42	-0.01	=>	-33.27	-0.01	=>
RBFK	=>	-25.49	14.25	<=>	-25.49	-0.01	=>	-20.47	-0.01	=>
CHORS	=>	-31.28	-6.06	=>	-31.28	-6.25	=>	-31.28	-6.25	=>

DDPA	=>	-26.28	1.97	<=>	-26.28	-0.01	=>	-26.28	-0.01	=>
DHQS	=>	-33.88	-10.60	=>	-33.88	-10.79	=>	-33.88	-10.79	=>
DHQTi	=>	-16.19	-0.76	=>	-16.19	-0.76	=>	-16.19	-0.76	=>
PSCVT	<=>	-14.41	15.84	<=>	-14.41	-0.01	=>	-14.41	-0.01	=>
SHK3Dr	<=>	-29.05	15.61	<=>	-20.19	-0.01	=>	-20.19	-0.01	=>
SHKK	=>	-21.59	10.41	<=>	-21.59	-0.01	=>	-16.56	-0.01	=>
HETZK	=>	-22.27	11.05	<=>	-22.27	6.24	<=>	-17.25	2.47	<=>
HMPK1	=>	-23.03	11.95	<=>	-23.03	7.14	<=>	-18.00	3.37	<=>
PMPK	=>	-18.08	18.12	<=>	-18.08	12.86	<=>	-13.05	8.94	<=>
THMDP	=>	-30.77	2.46	<=>	-30.77	2.46	<=>	-30.77	2.34	<=>
THMP	=>	-13.87	13.65	<=>	-13.87	13.65	<=>	-13.87	13.22	<=>
TMDPK	=>	-29.31	4.87	<=>	-29.31	-0.01	=>	-25.00	-1.45	=>
TMPK	=>	-17.11	18.99	<=>	-17.11	14.98	<=>	-12.09	11.48	<=>
TMPPP	=>	-25.10	10.19	<=>	-25.10	10.19	<=>	-24.78	10.19	<=>
CHRPL	=>	-52.16	-29.79	=>	-49.70	-29.79	=>	-47.59	-29.79	=>
HBZOPT_mt	=>	-38.31	3.89	<=>	-33.04	-0.01	=>	-33.04	-0.01	=>
OMMBLHX_mt	=>	-55.40	-22.23	=>	-55.40	-22.23	=>	-55.40	-22.23	=>
OMMBLHX3_mt	=>	-73.32	47.89	<=>	-73.32	47.89	<=>	-62.09	32.12	<=>
OPHBDC_mt	=>	-23.40	12.74	<=>	-23.40	-0.01	=>	-23.40	-0.01	=>
OPHHX_mt	=>	-54.73	-22.93	=>	-54.73	-22.93	=>	-54.73	-22.93	=>
HYPOE	=>	-12.94	13.31	<=>	-12.94	13.13	<=>	-12.94	13.13	<=>
PDXPP	=>	-20.64	5.66	<=>	-20.64	5.47	<=>	-20.64	5.47	<=>
PYDAMK	=>	-30.73	4.15	<=>	-30.73	-0.67	=>	-25.71	-4.44	=>
PYDXK	=>	-23.09	11.81	<=>	-23.09	6.99	<=>	-18.07	3.22	<=>
PYDXNK	=>	-23.07	11.84	<=>	-23.07	7.02	<=>	-18.05	3.24	<=>
PYDXPP	=>	-20.60	5.68	<=>	-20.60	5.49	<=>	-20.60	5.49	<=>
R07456	<=>	-78.83	-33.48	=>	-77.99	-33.67	=>	-72.11	-37.95	=>
PTD1INOt	<=>	-18.70	11.32	<=>	-18.70	11.32	<=>	-18.70	11.32	<=>
FEROpp	=>	-31.62	29.15	<=>	-31.62	-0.01	=>	-31.62	-0.01	=>
ACCOAC_ap	=>	-33.45	30.04	<=>	-33.44	30.04	<=>	-26.68	22.97	<=>
KAT1	=>	-30.27	14.39	<=>	-30.27	14.39	<=>	-30.27	14.39	<=>
MI1PP	=>	-20.21	5.38	<=>	-20.21	5.19	<=>	-17.21	2.74	<=>
MI1PS	=>	-10.10	7.43	<=>	-10.10	7.43	<=>	-10.10	7.43	<=>
CDPMEK_ap	=>	-27.24	16.61	<=>	-26.32	-0.01	=>	-20.47	-0.01	=>
DMPPS_ap	=>	-43.70	-0.63	=>	-43.70	-0.63	=>	-43.70	-0.63	=>
DXPRIi_ap	=>	-32.71	10.41	<=>	-32.71	-0.01	=>	-32.71	-0.01	=>
DXPS_ap	=>	-20.58	6.81	<=>	-18.94	-0.01	=>	-18.94	-0.01	=>
IPDPS_ap	=>	-44.41	-1.25	=>	-44.41	-1.25	=>	-44.41	-1.25	=>
MECDPDH2_ap	=>	-69.24	-26.16	=>	-69.24	-26.16	=>	-69.24	-26.16	=>
MEPCT_ap	=>	-15.78	18.04	<=>	-15.78	-0.01	=>	-15.78	-0.01	=>
CHLPCTD	=>	-9.49	27.10	<=>	-9.49	27.10	<=>	-8.06	22.92	<=>
CHOLK	=>	-31.29	4.83	<=>	-31.29	0.02	<=>	-22.08	-5.18	=>
DAGK120	=>	-22.40	12.88	<=>	-22.40	8.02	<=>	-17.38	4.24	<=>
DAGK140	=>	-22.40	12.88	<=>	-22.40	8.02	<=>	-17.38	4.24	<=>
DAGK141	=>	-22.88	13.36	<=>	-22.88	8.50	<=>	-17.85	4.71	<=>
DAGK160	=>	-22.40	12.88	<=>	-22.40	8.02	<=>	-17.38	4.24	<=>
DAGK161	=>	-22.40	12.88	<=>	-22.40	8.02	<=>	-17.38	4.24	<=>

DAGK180	=>	-22.40	12.88	<=>	-22.40	8.02	<=>	-17.38	4.24	<=>
DAGK181	=>	-22.40	12.88	<=>	-22.40	8.02	<=>	-17.38	4.24	<=>
DASYN120	=>	-18.30	18.96	<=>	-18.30	18.96	<=>	-18.30	18.96	<=>
DASYN140	=>	-18.80	19.46	<=>	-18.80	19.46	<=>	-18.80	19.46	<=>
DASYN141	=>	-18.61	19.27	<=>	-18.61	19.27	<=>	-18.61	19.27	<=>
DASYN160	=>	-19.33	20.00	<=>	-19.33	20.00	<=>	-19.33	20.00	<=>
DASYN161	=>	-19.08	19.75	<=>	-19.08	19.75	<=>	-19.08	19.75	<=>
DASYN180	=>	-19.90	20.56	<=>	-19.90	20.56	<=>	-19.90	20.11	<=>
DASYN181	=>	-19.60	20.26	<=>	-19.60	20.26	<=>	-19.60	20.11	<=>
ETHAK	=>	-27.90	1.53	<=>	-27.90	-3.26	=>	-18.14	-7.75	=>
PETHCT	=>	-7.95	25.46	<=>	-7.95	25.46	<=>	-7.23	20.72	<=>
PSD120	=>	-10.78	16.87	<=>	-10.78	16.87	<=>	-10.78	16.87	<=>
PSD140	=>	-10.78	16.87	<=>	-10.78	16.87	<=>	-10.78	16.87	<=>
PSD141	=>	-11.58	17.67	<=>	-11.58	17.67	<=>	-11.58	17.67	<=>
PSD160	=>	-10.78	16.87	<=>	-10.78	16.87	<=>	-10.78	16.87	<=>
PSD161	=>	-10.78	16.87	<=>	-10.78	16.87	<=>	-10.78	16.87	<=>
PSD180	=>	-13.38	19.48	<=>	-13.38	19.47	<=>	-13.38	19.47	<=>
PSD181	=>	-13.02	19.12	<=>	-13.02	19.11	<=>	-13.02	19.11	<=>
PSSA120	=>	-36.02	3.35	<=>	-36.02	3.35	<=>	-33.87	-0.09	=>
PSSA140	=>	-36.52	3.85	<=>	-36.52	3.85	<=>	-34.37	0.41	<=>
PSSA141	=>	-36.33	3.66	<=>	-36.33	3.66	<=>	-34.18	0.22	<=>
PSSA160	=>	-37.06	4.38	<=>	-37.06	4.38	<=>	-34.91	0.95	<=>
PSSA161	=>	-36.81	4.13	<=>	-36.81	4.13	<=>	-34.66	0.70	<=>
PSSA180	=>	-37.62	4.95	<=>	-37.62	4.95	<=>	-35.47	1.51	<=>
PSSA181	=>	-37.32	4.65	<=>	-37.32	4.65	<=>	-35.18	1.22	<=>
GPDDA2pp	=>	-21.07	-0.80	=>	-21.07	-0.80	=>	-21.07	-0.80	=>
GPDDA4pp	=>	-21.38	-0.50	=>	-21.38	-0.50	=>	-21.38	-0.50	=>
LPLIPAL1A120pp	=>	-20.03	2.64	<=>	-20.03	2.64	<=>	-20.03	2.64	<=>
LPLIPAL1A140pp	=>	-20.37	2.99	<=>	-20.37	2.99	<=>	-20.37	2.99	<=>
LPLIPAL1A141pp	=>	-21.00	2.30	<=>	-21.00	2.30	<=>	-21.00	2.30	<=>
LPLIPAL1A160pp	=>	-20.72	3.34	<=>	-20.72	3.34	<=>	-20.72	3.34	<=>
LPLIPAL1A161pp	=>	-20.58	3.20	<=>	-20.58	3.20	<=>	-20.58	3.20	<=>
LPLIPAL1A180pp	=>	-21.07	3.69	<=>	-21.07	3.69	<=>	-21.07	3.69	<=>
LPLIPAL1A181pp	=>	-20.90	3.52	<=>	-20.90	3.52	<=>	-20.90	3.52	<=>
LPLIPAL1E120pp	=>	-19.72	4.09	<=>	-19.72	4.09	<=>	-19.72	4.09	<=>
LPLIPAL1E140pp	=>	-20.07	4.44	<=>	-20.07	4.44	<=>	-20.07	4.44	<=>
LPLIPAL1E141pp	=>	-20.04	4.41	<=>	-20.04	4.41	<=>	-20.04	4.41	<=>
LPLIPAL1E160pp	=>	-20.42	4.78	<=>	-20.42	4.78	<=>	-20.42	4.78	<=>
LPLIPAL1E161pp	=>	-20.28	4.65	<=>	-20.28	4.65	<=>	-20.28	4.65	<=>
LPLIPAL1G120pp	=>	-20.02	4.39	<=>	-20.02	4.39	<=>	-20.02	4.39	<=>
LPLIPAL1G140pp	=>	-20.37	4.74	<=>	-20.37	4.74	<=>	-20.37	4.74	<=>
LPLIPAL1G141pp	=>	-20.34	4.71	<=>	-20.34	4.71	<=>	-20.34	4.71	<=>
LPLIPAL1G160pp	=>	-20.72	5.09	<=>	-20.72	5.09	<=>	-20.72	5.09	<=>
LPLIPAL1G161pp	=>	-20.58	4.95	<=>	-20.58	4.95	<=>	-20.58	4.95	<=>
LPLIPAL1G180pp	=>	-21.07	5.43	<=>	-21.07	5.43	<=>	-21.07	5.43	<=>
LPLIPAL1G181pp	=>	-20.90	5.27	<=>	-20.90	5.27	<=>	-20.90	5.27	<=>
LPLIPAL2E120	=>	-19.72	4.09	<=>	-19.72	4.09	<=>	-19.72	4.09	<=>

LPLIPAL2E140	=>	-20.07	4.44	<=>	-20.07	4.44	<=>	-20.07	4.44	<=>
LPLIPAL2E141	=>	-20.70	3.75	<=>	-20.70	3.75	<=>	-20.70	3.75	<=>
LPLIPAL2E160	=>	-20.42	4.78	<=>	-20.42	4.78	<=>	-20.42	4.78	<=>
LPLIPAL2E161	=>	-20.28	4.65	<=>	-20.28	4.65	<=>	-20.28	4.65	<=>
LPLIPAL2E180	=>	-20.77	5.13	<=>	-20.77	5.13	<=>	-20.77	5.13	<=>
LPLIPAL2E181	=>	-20.60	4.97	<=>	-20.60	4.97	<=>	-20.60	4.97	<=>
LPLIPAL2G120	=>	-20.02	4.39	<=>	-20.02	4.39	<=>	-20.02	4.39	<=>
LPLIPAL2G140	=>	-20.37	4.74	<=>	-20.37	4.74	<=>	-20.37	4.74	<=>
LPLIPAL2G141	=>	-21.00	4.05	<=>	-21.00	4.05	<=>	-21.00	4.05	<=>
LPLIPAL2G160	=>	-20.72	5.09	<=>	-20.72	5.09	<=>	-20.72	5.09	<=>
LPLIPAL2G161	=>	-20.58	4.95	<=>	-20.58	4.95	<=>	-20.58	4.95	<=>
LPLIPAL2G180	=>	-21.07	5.43	<=>	-21.07	5.43	<=>	-21.07	5.43	<=>
LPLIPAL2G181	=>	-20.90	5.27	<=>	-20.90	5.27	<=>	-20.90	5.27	<=>
PLIPA2A120pp	=>	-18.72	3.09	<=>	-18.72	3.09	<=>	-18.72	3.09	<=>
PLIPA2A140pp	=>	-18.72	3.09	<=>	-18.72	3.09	<=>	-18.72	3.09	<=>
PLIPA2A141pp	=>	-19.62	2.67	<=>	-19.62	2.67	<=>	-19.62	2.67	<=>
PLIPA2A160pp	=>	-18.72	3.09	<=>	-18.72	3.09	<=>	-18.72	3.09	<=>
PLIPA2A161pp	=>	-18.72	3.09	<=>	-18.72	3.09	<=>	-18.72	3.09	<=>
PLIPA2A180pp	=>	-18.72	3.09	<=>	-18.72	3.09	<=>	-18.72	3.09	<=>
PLIPA2A181pp	=>	-18.72	3.09	<=>	-18.72	3.09	<=>	-18.72	3.09	<=>
PLIPA2E120pp	=>	-19.29	3.66	<=>	-19.29	3.66	<=>	-19.29	3.66	<=>
PLIPA2E140pp	=>	-19.29	3.66	<=>	-19.29	3.66	<=>	-19.29	3.66	<=>
PLIPA2E141pp	=>	-20.61	2.34	<=>	-20.61	2.34	<=>	-20.61	2.34	<=>
PLIPA2E160pp	=>	-19.29	3.66	<=>	-19.29	3.66	<=>	-19.29	3.66	<=>
PLIPA2E161pp	=>	-19.29	3.66	<=>	-19.29	3.66	<=>	-19.29	3.66	<=>
PLIPA2G120pp	=>	-19.59	3.96	<=>	-19.59	3.96	<=>	-19.59	3.96	<=>
PLIPA2G140pp	=>	-19.59	3.96	<=>	-19.59	3.96	<=>	-19.59	3.96	<=>
PLIPA2G141pp	=>	-20.91	2.64	<=>	-20.91	2.64	<=>	-20.91	2.64	<=>
PLIPA2G160pp	=>	-19.59	3.96	<=>	-19.59	3.96	<=>	-19.59	3.96	<=>
PLIPA2G161pp	=>	-19.59	3.96	<=>	-19.59	3.96	<=>	-19.59	3.96	<=>
PLIPA2G180pp	=>	-19.59	3.96	<=>	-19.59	3.96	<=>	-19.59	3.96	<=>
PLIPA2G181pp	=>	-19.59	3.96	<=>	-19.59	3.96	<=>	-19.59	3.96	<=>
PGSA120	=>	-23.36	14.36	<=>	-23.36	14.36	<=>	-23.36	14.36	<=>
PGSA140	=>	-23.86	14.86	<=>	-23.86	14.86	<=>	-23.86	14.86	<=>
PGSA141	=>	-23.67	14.67	<=>	-23.67	14.67	<=>	-23.67	14.67	<=>
PGSA160	=>	-24.40	15.40	<=>	-24.40	15.40	<=>	-24.40	15.40	<=>
PGSA161	=>	-24.15	15.15	<=>	-24.15	15.15	<=>	-24.15	15.15	<=>
PGSA180	=>	-24.96	15.96	<=>	-24.96	15.96	<=>	-24.96	15.96	<=>
PGSA181	=>	-24.66	15.66	<=>	-24.66	15.66	<=>	-24.66	15.66	<=>
PGPP120pp	=>	-22.47	5.78	<=>	-22.47	5.59	<=>	-22.47	5.59	<=>
PGPP140pp	=>	-22.47	5.78	<=>	-22.47	5.59	<=>	-22.47	5.59	<=>
PGPP141pp	=>	-23.24	6.54	<=>	-23.24	6.36	<=>	-23.24	6.36	<=>
PGPP160pp	=>	-22.47	5.78	<=>	-22.47	5.59	<=>	-22.47	5.59	<=>
PGPP161pp	=>	-22.47	5.78	<=>	-22.47	5.59	<=>	-22.47	5.59	<=>
PGPP180pp	=>	-22.47	5.78	<=>	-22.47	5.59	<=>	-22.47	5.59	<=>
PGPP181pp	=>	-22.47	5.78	<=>	-22.47	5.59	<=>	-22.47	5.59	<=>
DMATT	=>	-28.18	1.41	<=>	-28.18	-0.01	=>	-28.18	-0.01	=>

GRTT	=>	-29.17	2.41	<=>	-29.17	-0.01	=>	-29.17	-0.01	=>
OCTDPS	=>	-115.52	-18.31	=>	-110.47	-18.31	=>	-110.47	-18.31	=>
LPLIPAL2ATE120	=>	-19.66	19.66	<=>	-19.66	19.66	<=>	-19.66	19.66	<=>
LPLIPAL2ATE140	=>	-20.63	20.63	<=>	-20.63	20.63	<=>	-20.63	20.63	<=>
LPLIPAL2ATE141	=>	-20.27	20.27	<=>	-20.27	20.27	<=>	-20.27	20.27	<=>
LPLIPAL2ATE160	=>	-21.62	21.62	<=>	-21.62	21.62	<=>	-21.62	21.62	<=>
LPLIPAL2ATE161	=>	-21.17	21.17	<=>	-21.17	21.17	<=>	-21.17	21.17	<=>
LPLIPAL2ATE180	=>	-22.62	22.62	<=>	-22.62	22.62	<=>	-22.62	22.62	<=>
LPLIPAL2ATE181	=>	-22.10	22.10	<=>	-22.10	22.10	<=>	-22.10	22.10	<=>
LPLIPAL2ATG120	=>	-19.84	19.84	<=>	-19.84	19.84	<=>	-19.84	19.84	<=>
LPLIPAL2ATG140	=>	-20.48	20.48	<=>	-20.48	20.48	<=>	-20.48	20.48	<=>
LPLIPAL2ATG141	=>	-20.58	20.58	<=>	-20.58	20.58	<=>	-20.58	20.58	<=>
LPLIPAL2ATG160	=>	-21.12	21.12	<=>	-21.12	21.12	<=>	-21.12	21.12	<=>
LPLIPAL2ATG161	=>	-20.82	20.82	<=>	-20.82	20.82	<=>	-20.82	20.82	<=>
LPLIPAL2ATG180	=>	-21.78	21.78	<=>	-21.78	21.78	<=>	-21.78	21.78	<=>
LPLIPAL2ATG181	=>	-21.43	21.43	<=>	-21.43	21.43	<=>	-21.43	21.43	<=>
PAPA120pp	=>	-22.07	5.38	<=>	-22.07	5.29	<=>	-22.07	5.29	<=>
PAPA140pp	=>	-22.07	5.38	<=>	-22.07	5.37	<=>	-22.07	5.37	<=>
PAPA141pp	=>	-22.55	5.86	<=>	-22.55	5.75	<=>	-22.55	5.75	<=>
PAPA160pp	=>	-22.07	5.38	<=>	-22.07	5.37	<=>	-22.07	5.37	<=>
PAPA161pp	=>	-22.07	5.38	<=>	-22.07	5.37	<=>	-22.07	5.37	<=>
PAPA180pp	=>	-22.07	5.38	<=>	-22.07	5.37	<=>	-22.07	5.37	<=>
PAPA181pp	=>	-22.07	5.38	<=>	-22.07	5.37	<=>	-22.07	5.37	<=>
ADA	=>	-28.10	1.45	<=>	-28.10	1.44	<=>	-27.98	1.44	<=>
ADK1	<=>	-16.05	16.37	<=>	-16.05	-0.01	=>	-4.97	-0.01	=>
ADNCYC	=>	-20.09	10.31	<=>	-20.09	10.31	<=>	-18.94	5.74	<=>
ADSL1r	<=>	-11.18	15.26	<=>	-11.18	-0.01	=>	-4.60	-0.01	=>
AMPDA	=>	-27.21	0.65	<=>	-27.02	0.65	<=>	-24.19	-2.61	=>
DADA	=>	-28.50	1.85	<=>	-28.50	1.85	<=>	-28.50	1.85	<=>
DGK1	<=>	-20.01	20.16	<=>	-20.01	14.90	<=>	-14.99	10.98	<=>
GK1	<=>	-20.13	20.28	<=>	-20.13	-0.01	=>	-15.11	-0.01	=>
GUACYC	=>	-21.86	11.99	<=>	-21.86	11.99	<=>	-21.86	11.99	<=>
GUAPRT	=>	-20.60	13.82	<=>	-20.60	13.82	<=>	-20.60	13.82	<=>
HXPRT	=>	-20.86	14.08	<=>	-20.86	-0.01	=>	-20.86	-0.01	=>
IMPD	=>	-23.71	25.47	<=>	-23.71	25.47	<=>	-22.52	20.97	<=>
NDP3	=>	-28.45	2.10	<=>	-24.73	1.91	<=>	-21.29	1.91	<=>
NDP4	=>	-28.34	1.99	<=>	-28.34	1.80	<=>	-28.34	1.80	<=>
NTD10pp	=>	-22.04	7.10	<=>	-22.04	7.10	<=>	-22.04	7.10	<=>
NTD11pp	=>	-21.87	6.93	<=>	-21.87	6.93	<=>	-21.87	6.77	<=>
NTD1pp	=>	-20.87	5.93	<=>	-20.87	5.75	<=>	-20.87	5.75	<=>
NTD2pp	=>	-21.06	6.12	<=>	-21.06	5.93	<=>	-21.06	5.93	<=>
NTD3pp	=>	-20.94	6.00	<=>	-20.94	5.82	<=>	-20.94	5.82	<=>
NTD4pp	=>	-31.08	-3.77	=>	-31.08	-3.95	=>	-31.08	-3.95	=>
NTD5pp	=>	-20.81	5.87	<=>	-20.81	5.68	<=>	-20.81	5.68	<=>
NTD7pp	=>	-20.68	5.83	<=>	-20.01	5.83	<=>	-17.18	1.89	<=>
NTD8pp	=>	-22.64	7.70	<=>	-22.64	7.51	<=>	-22.64	7.51	<=>
NTD9pp	=>	-22.09	7.15	<=>	-22.09	7.15	<=>	-22.09	7.00	<=>

PDE1	=>	-19.06	-1.36	=>	-19.06	-2.03	=>	-15.31	-4.86	=>
PDE4	=>	-21.51	1.18	<=>	-21.51	1.18	<=>	-21.18	1.18	<=>
PPA	=>	-21.57	-0.98	=>	-21.57	-1.35	=>	-21.57	-1.35	=>
PUNP1	<=>	-16.91	19.34	<=>	-16.91	19.34	<=>	-16.43	18.53	<=>
PUNP2	<=>	-18.31	20.78	<=>	-18.31	20.78	<=>	-18.31	20.78	<=>
PUNP3	<=>	-16.34	18.76	<=>	-16.34	18.76	<=>	-16.34	18.76	<=>
PUNP4	<=>	-17.88	20.35	<=>	-17.88	20.35	<=>	-17.88	20.35	<=>
PUNP5	<=>	-16.34	18.76	<=>	-16.34	18.76	<=>	-16.34	18.76	<=>
PUNP6	<=>	-17.89	20.36	<=>	-17.89	20.36	<=>	-17.89	20.36	<=>
PUNP7	<=>	-16.16	18.94	<=>	-16.16	18.94	<=>	-16.16	18.94	<=>
XPRT	=>	-20.78	13.64	<=>	-20.78	13.64	<=>	-20.78	13.64	<=>
ADSS	=>	-29.49	27.20	<=>	-26.78	-0.01	=>	-16.35	-0.01	=>
GMPS2	=>	-42.72	12.30	<=>	-42.72	11.18	<=>	-31.68	0.50	<=>
NDPK2	<=>	-18.42	18.61	<=>	-18.42	-0.01	=>	-13.40	-0.01	=>
NDPK3	<=>	-18.48	18.68	<=>	-18.42	13.42	<=>	-13.40	9.49	<=>
NDPK4	<=>	-18.19	18.38	<=>	-18.19	-0.01	=>	-13.17	-0.01	=>
NDPK5	<=>	-20.01	20.20	<=>	-20.01	-0.01	=>	-14.98	-0.01	=>
NDPK6	<=>	-18.25	18.44	<=>	-18.25	13.18	<=>	-13.23	9.26	<=>
NDPK7	<=>	-18.31	18.51	<=>	-18.31	-0.01	=>	-13.29	-0.01	=>
NDPK8	<=>	-20.60	20.80	<=>	-20.60	-0.01	=>	-15.58	-0.01	=>
NDPK1	<=>	-20.12	20.32	<=>	-20.12	10.98	<=>	-14.85	7.27	<=>
ATPH1e	=>	-41.22	-11.52	=>	-41.22	-11.62	=>	-35.69	-18.20	=>
ATPH2e	=>	-25.05	-1.48	=>	-20.87	-1.66	=>	-18.32	-8.07	=>
ATPM	=>	-24.89	-1.32	=>	-24.30	-6.11	=>	-19.86	-9.69	=>
CTPS2	=>	-45.36	10.89	<=>	-43.87	-0.01	=>	-33.56	-1.76	=>
CYTK1	<=>	-18.48	18.63	<=>	-18.48	13.37	<=>	-13.46	9.45	<=>
CYTK2	<=>	-18.31	18.46	<=>	-18.31	13.20	<=>	-13.29	9.28	<=>
DCTPD	=>	-23.40	3.45	<=>	-23.40	3.45	<=>	-23.40	3.45	<=>
DHORTS	<=>	-13.55	2.12	<=>	0.01	0.83	<=	0.01	0.83	<=
DTMPK	<=>	-18.18	18.33	<=>	-18.18	-0.01	=>	-13.16	-0.01	=>
DUMPK	<=>	-18.23	18.38	<=>	-18.23	13.13	<=>	-13.21	9.21	<=>
DUTPDP	=>	-28.49	-1.71	=>	-28.49	-1.71	=>	-28.49	-1.71	=>
OMPDC	=>	-19.78	7.35	<=>	-19.78	-0.01	=>	-19.78	-0.01	=>
ORPT	<=>	-14.20	20.76	<=>	0.01	20.76	<=	0.01	20.76	<=
UMPK	<=>	-18.42	18.56	<=>	-18.42	-0.01	=>	-13.40	-0.01	=>
ASPCT	=>	-20.62	8.78	<=>	-11.65	-0.01	=>	-10.41	-0.01	=>
CBPS	=>	-56.50	10.20	<=>	-54.29	-0.01	=>	-39.74	-10.54	=>
GHMT2r	<=>	-18.33	18.44	<=>	-15.97	14.60	<=>	-10.35	9.00	<=>
HCO3E	<=>	-8.19	5.53	<=>	-8.19	-0.01	=>	-8.19	-0.01	=>
TMDS	=>	-37.16	5.40	<=>	-37.16	-0.01	=>	-37.16	-0.01	=>
DHORD2_mt	=>	-36.77	11.41	<=>	-31.72	-0.01	=>	-31.72	-0.01	=>
DURIPP	<=>	-15.92	18.39	<=>	-15.92	18.39	<=>	-15.92	18.39	<=>
SUCD2_u6m_mt	<=>	-24.83	19.99	<=>	-24.83	19.99	<=>	-19.78	13.82	<=>
SUCD3_u6m_mt	<=>	-46.07	17.41	<=>	-46.07	17.41	<=>	-46.07	17.41	<=>
NADH2_u6cm_mt	=>	-47.12	12.26	<=>	-47.12	12.26	<=>	-42.71	11.08	<=>
ACONTam_mt	<=>	-5.78	8.95	<=>	-5.78	8.95	<=>	-5.78	-0.01	=>
ACONTbm_mt	<=>	-6.99	6.65	<=>	-6.99	6.65	<=>	-6.23	-0.01	=>

AKGDH_mt	=>	-42.99	25.50	<=>	-42.99	-0.01	=>	-39.33	-0.01	=>
CS_mt	=>	-28.08	11.16	<=>	-28.08	11.16	<=>	-22.79	-0.01	=>
SUCOAS1m_mt	<=>	-33.65	28.74	<=>	-33.65	28.74	<=>	-31.44	25.29	<=>
ICDHyr_mt	<=>	-20.94	28.94	<=>	-20.94	28.94	<=>	-17.90	-0.01	=>
ACCOAL2	=>	-30.23	25.12	<=>	-29.65	20.33	<=>	-25.21	16.75	<=>
ACCOALm	=>	-32.49	23.39	<=>	-32.49	23.39	<=>	-28.18	16.48	<=>
GTHOr_mt	<=>	-41.78	14.19	<=>	-41.78	14.19	<=>	-30.26	7.25	<=>
GTHP_mt	<=>	-83.38	-44.60	=>	-83.38	-44.60	=>	-76.44	-45.86	=>
SPODMc	=>	-30.29	-1.84	=>	-30.29	-1.84	=>	-30.29	-1.84	=>
GTHS	=>	-31.04	16.04	<=>	-30.56	11.34	<=>	-23.84	2.91	<=>
ABUTt2r	<=>	-15.76	11.65	<=>	-15.76	11.65	<=>	-8.26	9.68	<=>
ACt2r	<=>	-13.19	14.83	<=>	-13.19	14.83	<=>	-7.71	11.10	<=>
AKGMAL	<=>	-27.39	27.42	<=>	-27.39	27.42	<=>	-16.06	20.12	<=>
AKGt2r	<=>	-10.77	16.78	<=>	-10.77	16.78	<=>	-5.63	12.62	<=>
CHOLtu	<=>	-15.70	14.32	<=>	-15.70	14.32	<=>	-15.70	14.32	<=>
CO2t	<=>	-12.51	15.51	<=>	-12.51	15.50	<=>	-12.51	15.50	<=>
ETHAt2pp	=>	-17.16	10.03	<=>	-17.16	10.03	<=>	-17.16	10.03	<=>
F6Pt6_2pp	=>	-50.49	31.57	<=>	-19.63	31.57	<=>	-19.63	31.57	<=>
FE2t	=>	-16.89	11.13	<=>	-16.89	-0.01	=>	-16.89	-0.01	=>
FORt	<=>	-10.32	17.70	<=>	-10.32	17.70	<=>	-5.31	13.47	<=>
FRUt1r	<=>	-13.51	16.51	<=>	-13.51	16.51	<=>	-13.51	16.51	<=>
FUMt2r	<=>	-11.44	17.45	<=>	-3.12	17.45	<=>	0.01	12.85	<=
G6Pt6_2pp	=>	-50.46	31.60	<=>	-19.60	31.60	<=>	-19.60	31.60	<=>
GAM6Pt	<=>	-14.64	15.38	<=>	-14.64	15.38	<=>	-14.64	15.38	<=>
GLCt1r	<=>	-13.51	16.51	<=>	-13.51	-0.01	=>	-13.51	16.51	<=>
GLYCt	<=>	-15.40	12.40	<=>	-15.40	12.40	<=>	-15.40	12.40	<=>
GTHRDti	<=>	-12.04	17.98	<=>	-12.04	17.98	<=>	-12.04	15.57	<=>
H2O2t	<=>	-12.51	15.51	<=>	-12.51	15.51	<=>	-12.51	15.51	<=>
Ht	<=>	2.87	2.87	<=	2.87	2.87	<=	2.87	2.87	<=
L_LACt2r	<=>	-12.70	14.33	<=>	-12.70	14.33	<=>	-6.99	11.98	<=>
MALt2r	<=>	-10.61	16.65	<=>	-10.61	16.65	<=>	-3.44	14.49	<=>
MANt1r	<=>	-13.51	16.51	<=>	-13.51	16.51	<=>	-13.51	16.51	<=>
NH4t	<=>	-14.70	13.32	<=>	-14.70	13.31	<=>	-14.70	13.31	<=>
NO2t2rpp	<=>	-13.19	14.83	<=>	-13.19	14.83	<=>	-13.19	14.83	<=>
NO3t2rpp	<=>	-13.19	14.83	<=>	-13.19	14.83	<=>	-13.19	14.83	<=>
O2St	<=>	-15.51	12.51	<=>	-15.51	12.51	<=>	-15.51	12.51	<=>
O2t	<=>	-12.51	15.51	<=>	-12.50	15.51	<=>	-12.50	15.51	<=>
ORNtiDF	=>	-14.72	13.34	<=>	-14.72	13.34	<=>	-14.72	13.34	<=>
OXAHCOtex	=>	-42.77	39.29	<=>	-42.77	39.29	<=>	-42.77	39.29	<=>
PIt2r	<=>	-10.60	17.42	<=>	-10.60	-0.01	=>	-10.60	-0.01	=>
PTRCt2	=>	-19.82	8.31	<=>	-19.82	8.31	<=>	-12.17	6.62	<=>
PYRt2r	<=>	-12.79	14.43	<=>	-12.79	14.43	<=>	-9.23	14.43	<=>
SBT_Dt	<=>	-13.49	16.49	<=>	-13.49	16.49	<=>	-13.49	16.49	<=>
SO4HCOtex	=>	-42.77	39.29	<=>	-42.77	19.39	<=>	-42.77	19.39	<=>
SO4OXAtex2	=>	-35.15	46.91	<=>	-35.15	9.11	<=>	-35.15	9.11	<=>
SO4ti	=>	-8.13	19.89	<=>	-8.13	-0.01	=>	-8.13	-0.01	=>
SPMDt2	=>	-22.96	7.06	<=>	-22.96	-0.01	=>	-14.64	-0.01	=>

SUCCt2r	<=>	-10.54	16.61	<=>	-10.54	16.61	<=>	-4.71	12.88	<=>
TTDCAt	=>	-11.32	18.71	<=>	-11.32	18.71	<=>	-11.32	18.71	<=>
UREAt	<=>	-12.51	15.51	<=>	-12.51	15.51	<=>	-12.51	15.51	<=>
ALAGLNexR	<=>	-27.82	27.82	<=>	-27.82	27.82	<=>	-19.43	17.37	<=>
ALAGLYexR	<=>	-27.09	27.09	<=>	-27.09	27.09	<=>	-17.54	17.73	<=>
ALAt2r	<=>	-14.97	12.22	<=>	-14.97	12.22	<=>	-9.09	8.72	<=>
ARGLYSex	=>	-28.02	28.02	<=>	-28.02	28.02	<=>	-17.46	19.79	<=>
ARGt2r	<=>	-18.58	11.45	<=>	-18.58	11.45	<=>	-10.54	10.12	<=>
ASNt2r	<=>	-15.56	12.81	<=>	-15.56	12.81	<=>	-9.86	9.91	<=>
ASPt2r	<=>	-13.56	13.84	<=>	-13.56	13.84	<=>	-6.24	11.76	<=>
CYSGLUexR	<=>	-28.74	27.31	<=>	-28.74	27.31	<=>	-26.23	20.36	<=>
CYSGLYexR	<=>	-27.19	28.61	<=>	-27.18	28.61	<=>	-21.32	24.94	<=>
CYSt2r	<=>	-16.49	12.32	<=>	-16.49	12.32	<=>	-16.49	12.32	<=>
CYSTGLUex	=>	-29.53	26.51	<=>	-29.53	26.51	<=>	-29.24	17.66	<=>
CYSTSERex	=>	-28.02	28.02	<=>	-28.02	28.02	<=>	-24.39	22.22	<=>
GLNt2r	<=>	-15.60	12.85	<=>	-15.60	12.85	<=>	-8.65	10.35	<=>
GLUt2r	<=>	-13.63	13.90	<=>	-13.63	13.90	<=>	-4.77	13.61	<=>
GLYt2r	<=>	-14.87	12.12	<=>	-14.86	12.12	<=>	-9.00	8.45	<=>
HISt2r	<=>	-16.39	13.64	<=>	-16.39	13.64	<=>	-11.60	9.03	<=>
ILEt2r	<=>	-15.28	12.53	<=>	-15.28	-0.01	=>	-10.58	-0.01	=>
LEUt2r	<=>	-15.28	12.53	<=>	-15.28	12.53	<=>	-9.76	8.82	<=>
LYSt2r	<=>	-17.69	10.56	<=>	-17.69	10.56	<=>	-10.79	8.04	<=>
METLEUex	=>	-28.02	28.02	<=>	-28.02	15.27	<=>	-24.31	14.52	<=>
METt2r	<=>	-16.02	13.27	<=>	-16.02	13.27	<=>	-16.02	13.27	<=>
PHEt2r	<=>	-16.04	13.29	<=>	-16.04	13.29	<=>	-11.34	8.73	<=>
PROt2r	<=>	-14.08	15.71	<=>	-14.08	15.71	<=>	-14.08	15.71	<=>
SERGLNexR	<=>	-27.86	27.86	<=>	-27.86	27.86	<=>	-19.56	17.28	<=>
SERGLYexR	<=>	-27.13	27.13	<=>	-27.13	27.13	<=>	-17.66	17.64	<=>
SERt2r	<=>	-15.01	12.26	<=>	-15.01	12.26	<=>	-9.21	8.63	<=>
THRGLNexR	<=>	-28.02	28.02	<=>	-28.02	28.02	<=>	-25.52	21.07	<=>
THRGLYexR	<=>	-27.33	27.33	<=>	-27.33	27.33	<=>	-23.67	21.48	<=>
THRt2r	<=>	-15.22	12.47	<=>	-15.22	12.47	<=>	-15.22	12.47	<=>
TRPt2r	<=>	-16.39	13.64	<=>	-16.39	13.64	<=>	-16.39	13.64	<=>
TYRt2r	<=>	-16.19	13.44	<=>	-16.19	13.44	<=>	-11.18	8.97	<=>
VALt2r	<=>	-15.26	12.51	<=>	-15.26	12.51	<=>	-9.91	8.52	<=>
DPCOAtap	<=>	-11.90	11.90	<=>	0.01	11.90	<=	0.01	11.90	<=
NADPtap	<=>	-12.61	11.20	<=>	-12.61	-0.01	=>	-12.61	-0.01	=>
NADtap	<=>	-11.90	11.90	<=>	-11.90	11.90	<=>	-2.46	2.46	<=>
NH4tap	<=>	-10.90	10.90	<=>	-10.90	-0.01	=>	-10.90	-0.01	=>
O2tap	<=>	-10.90	10.90	<=>	-10.89	10.90	<=>	-10.89	10.90	<=>
PA120tap	<=>	-11.90	11.90	<=>	-11.90	11.90	<=>	-11.90	11.90	<=>
PA140tap	<=>	-11.90	11.90	<=>	-11.90	11.90	<=>	-11.90	11.90	<=>
PA141tap	<=>	-11.90	11.90	<=>	-11.90	11.90	<=>	-11.90	11.90	<=>
PA160tap	<=>	-11.90	11.90	<=>	-11.90	11.90	<=>	-11.90	11.90	<=>
PA161tap	<=>	-11.90	11.90	<=>	-11.90	11.90	<=>	-11.90	11.90	<=>
PA180tap	<=>	-11.90	11.90	<=>	-11.90	11.90	<=>	-11.90	11.90	<=>
PA181tap	<=>	-11.90	11.90	<=>	-11.90	11.90	<=>	-11.90	11.90	<=>

PItap	<=>	-10.68	11.12	<=>	0.01	8.39	<=	0.01	8.39	<=
PPItap	<=>	-11.20	10.61	<=>	-11.20	-0.01	=>	-11.20	-0.01	=>
ACCOAtap	<=>	-12.20	11.61	<=>	-12.20	11.61	<=>	-12.20	11.61	<=>
ADPtap	<=>	-12.10	11.71	<=>	-12.10	11.71	<=>	-2.96	2.57	<=>
ATPtap	<=>	-12.01	11.80	<=>	-12.01	11.80	<=>	-3.08	2.86	<=>
CMPtap	<=>	-12.19	11.62	<=>	-12.19	-0.01	=>	-12.19	-0.01	=>
CO2tap	<=>	-10.90	10.90	<=>	-10.90	-0.01	=>	-10.90	-0.01	=>
COA3tap	<=>	-11.09	10.28	<=>	0.01	10.28	<=	0.01	10.28	<=
COAtap	<=>	-11.93	11.88	<=>	-11.93	-0.01	=>	-11.93	-0.01	=>
CTPtap	<=>	-11.99	11.82	<=>	0.01	11.82	<=	0.01	11.82	<=
DHAPtap	=>	-20.84	22.07	<=>	-8.39	-0.01	=>	-8.39	-0.01	=>
DMPPtap	<=>	-11.23	10.80	<=>	-11.23	-0.01	=>	-11.23	-0.01	=>
FE2tap	<=>	-10.90	10.90	<=>	-10.90	10.90	<=>	-10.90	10.90	<=>
HCO3tap	<=>	-10.84	10.96	<=>	-10.84	10.96	<=>	-10.84	10.96	<=>
IPDPtap	<=>	-11.27	10.85	<=>	-11.27	-0.01	=>	-11.27	-0.01	=>
NADHtap	<=>	-11.90	11.90	<=>	-11.90	11.90	<=>	-11.90	11.90	<=>
NADPHtap	<=>	-12.20	11.60	<=>	0.01	11.60	<=	0.01	11.60	<=
PEPPItap	=>	-21.35	22.27	<=>	-10.83	-0.01	=>	-10.83	-0.01	=>
ADEt	<=>	-13.51	16.51	<=>	-13.51	16.51	<=>	-13.51	16.51	<=>
ADNt	<=>	-13.51	16.51	<=>	-13.51	16.51	<=>	-13.51	16.51	<=>
CYTDt	<=>	-13.51	16.51	<=>	-13.51	16.51	<=>	-13.51	16.51	<=>
DADNt4	<=>	-13.51	16.51	<=>	-13.51	16.51	<=>	-13.51	16.51	<=>
DCYTt	<=>	-13.51	16.51	<=>	-13.51	16.51	<=>	-13.51	16.51	<=>
DGSNt	<=>	-13.51	16.51	<=>	-13.51	16.51	<=>	-13.51	16.51	<=>
DINt	<=>	-13.52	16.50	<=>	-13.52	16.50	<=>	-13.52	16.50	<=>
DURIt	<=>	-13.51	16.51	<=>	-13.51	16.51	<=>	-13.51	16.51	<=>
GSNt	<=>	-13.51	16.51	<=>	-13.51	16.51	<=>	-13.51	16.51	<=>
GUAt	<=>	-13.51	16.51	<=>	-13.51	16.51	<=>	-13.51	16.51	<=>
HYXNt	<=>	-13.52	16.50	<=>	-13.52	16.50	<=>	-13.52	16.50	<=>
INSt	<=>	-13.52	16.50	<=>	-13.52	16.50	<=>	-13.52	16.50	<=>
THYMDt1	=>	-13.51	16.51	<=>	-13.51	16.51	<=>	-13.51	16.51	<=>
URAt	<=>	-13.51	16.51	<=>	-13.51	16.51	<=>	-13.51	16.51	<=>
URIt	<=>	-13.51	16.51	<=>	-13.51	16.51	<=>	-13.51	16.51	<=>
XANt	<=>	-13.58	16.44	<=>	-13.58	16.44	<=>	-13.58	16.44	<=>
NACUP	=>	-11.32	18.70	<=>	-11.32	18.70	<=>	-11.32	18.70	<=>
PNTOt2	<=>	-13.89	15.53	<=>	-13.89	-0.01	=>	-13.89	-0.01	=>
RIBFLVt2	=>	-17.01	13.01	<=>	-17.01	-0.01	=>	-17.01	-0.01	=>
THMt3	<=>	-12.83	17.19	<=>	-12.83	-0.01	=>	-12.83	-0.01	=>
EX_nicotinamide2	<=>	-13.51	16.51	<=>	-13.51	16.51	<=>	-13.51	16.51	<=>
EX_folate2	<=>	-13.51	16.51	<=>	-13.51	16.51	<=>	-13.51	16.51	<=>
HDCAtr	<=>	-11.32	18.71	<=>	-11.32	18.71	<=>	-11.32	18.71	<=>
HDCEAtr	<=>	-11.32	18.70	<=>	-11.32	18.70	<=>	-11.32	18.70	<=>
INSTt2r	<=>	-16.39	13.64	<=>	-16.39	13.64	<=>	-11.31	9.50	<=>
OCDCAtr	<=>	-11.32	18.71	<=>	-11.32	18.71	<=>	-11.32	18.71	<=>
OCDCEAtr	<=>	-11.32	18.70	<=>	-11.32	18.70	<=>	-11.32	18.70	<=>
PEFLIP	=>	-34.21	15.38	<=>	-33.74	-0.01	=>	-29.19	-0.01	=>
DHAPtmt	<=>	-15.10	6.00	<=>	-13.57	6.00	<=>	-13.57	6.00	<=>

ATPADPex_mt	=>	-25.87	17.75	<=>	-24.55	17.75	<=>	-7.79	-0.33	=>
ATPtmt	=>	-28.69	3.04	<=>	-28.69	3.04	<=>	-23.67	-2.64	=>
PIOHex_mt	=>	-6.56	15.24	<=>	-6.56	15.24	<=>	-6.56	15.24	<=>
4HBAtmt	<=>	-11.68	11.68	<=>	-11.68	11.68	<=>	-11.68	11.68	<=>
4HBZtmt	<=>	-15.84	7.54	<=>	0.01	7.54	<=	0.01	7.54	<=
5AOPtmt	<=>	-11.09	10.28	<=>	-10.67	-0.01	=>	-10.67	-0.01	=>
AKGtmt	=>	-21.20	21.19	<=>	-21.20	21.19	<=>	-2.56	2.55	<=>
ALALtmt	<=>	-10.49	10.49	<=>	-10.49	10.49	<=>	-1.10	1.10	<=>
ASPtmt	<=>	-15.15	6.04	<=>	-15.15	6.04	<=>	-5.76	-3.36	=>
CO2tmt	<=>	-10.90	10.90	<=>	-10.90	-0.01	=>	-10.90	-0.01	=>
COAtmt	<=>	-24.38	-0.57	=>	-24.38	-0.57	=>	-24.38	-0.57	=>
CYSLtmt	<=>	-11.73	10.87	<=>	-11.73	10.87	<=>	-11.73	10.87	<=>
DHORtmt	<=>	-16.05	7.75	<=>	0.01	1.39	<=	0.01	1.39	<=
DXYL5Ptmt	<=>	-15.31	6.43	<=>	-15.31	6.43	<=>	-15.31	6.43	<=>
FUMtmt	<=>	-19.64	3.04	<=>	-19.64	3.04	<=>	-10.25	-6.36	=>
GDPtmt	<=>	-20.42	3.39	<=>	-20.42	3.39	<=>	-20.42	3.39	<=>
GLUASPtmt	=>	-21.25	21.25	<=>	-21.25	21.25	<=>	-2.71	2.72	<=>
GLYC3Ptmt	<=>	-15.10	6.22	<=>	-15.10	6.22	<=>	-15.10	6.22	<=>
GLYtmt	<=>	-10.39	10.39	<=>	0.01	10.39	<=	0.01	0.86	<=
GTPtmt	<=>	-24.44	-0.64	=>	-24.44	-0.64	=>	-24.44	-0.64	=>
H2O2tmt	<=>	-10.90	10.90	<=>	-10.90	10.90	<=>	-10.90	10.90	<=>
Htmt	<=>	3.74	3.74	<=	3.74	3.74	<=	3.74	3.74	<=
MLTHFtmt	<=>	-20.21	3.60	<=>	-20.21	3.60	<=>	-20.21	3.60	<=>
NH4tmt	<=>	-6.75	15.05	<=>	-6.75	15.04	<=>	-6.75	15.04	<=>
NO2tmt	<=>	-15.05	6.75	<=>	-15.05	6.75	<=>	-15.05	6.75	<=>
NO3tmt	<=>	-15.05	6.75	<=>	-15.05	6.75	<=>	-15.05	6.75	<=>
O2tmt	<=>	-10.90	10.90	<=>	0.01	10.90	<=	0.01	10.90	<=
OAAtmt	<=>	-18.90	2.30	<=>	-18.90	2.30	<=>	0.01	2.30	<=
OCTDPtmt	<=>	-20.42	3.39	<=>	0.01	3.39	<=	0.01	3.39	<=
OROTtmt	<=>	-16.07	7.73	<=>	-16.07	-0.01	=>	-16.07	-0.01	=>
PPItmt	<=>	-19.50	2.30	<=>	-19.50	-0.01	=>	-19.50	-0.01	=>
SUCCtmt	<=>	-18.76	2.18	<=>	-18.76	2.18	<=>	-9.20	-7.38	=>
THFtmt	<=>	-20.21	3.60	<=>	-20.21	3.60	<=>	-20.21	3.60	<=>
TYRLtmt	<=>	-11.71	11.71	<=>	-11.71	11.71	<=>	-2.23	2.23	<=>
4AHMMPtr	<=>	-13.51	16.51	<=>	-13.51	16.51	<=>	-13.51	16.51	<=>
DXYL5Ptap	<=>	-11.16	10.58	<=>	-11.16	10.58	<=>	-11.16	10.58	<=>
4MHETZtmt	<=>	-11.90	11.90	<=>	-11.90	11.90	<=>	-11.90	11.90	<=>
MTAADA	=>	-28.80	2.15	<=>	-28.80	2.15	<=>	-28.80	2.15	<=>
PUNP8	=>	-18.53	20.95	<=>	-18.52	20.95	<=>	-18.52	20.95	<=>
ACCOAtm	<=>	0.85	24.65	<=	0.85	24.65	<=	0.85	24.65	<=
Pyr_mt	<=>	-6.35	14.65	<=>	-6.35	14.65	<=>	-6.35	-0.01	=>
PDH_mt	<=>	-42.94	25.45	<=>	-42.94	19.47	<=>	-39.31	-0.01	=>
R00261	<=>	-18.05	2.93	<=>	-18.05	2.92	<=>	-13.44	-3.49	=>
R00713	<=>	-30.68	10.98	<=>	-30.68	10.98	<=>	-26.05	4.36	<=>
R01648	<=>	-14.36	11.66	<=>	-14.36	11.66	<=>	-5.49	4.01	<=>
GABAt_mt	<=>	-13.93	7.26	<=>	-13.93	7.26	<=>	-4.47	-2.20	=>

Table 5. Table of **Δ,G'** ranges obtained as a result of TVA studies

** − minimal growth yield was fixed to 0.042 that was 1% of the maximal yield model predicted with unconstrained uptake rates*

Species	Number	KEGG orthology identifiers
P. berghi P. cynomolgi P. falciparum P. vivax	552	K01698 K01711 K00764 K00991 K00939 K01718 K00997 K01681 K10250 K10780 K11204 K03861 K01736 K10105 K00507 K00813 K04567 K00025 K00822 K13644 K00079 K01613 K00643 K00956 K01756 K02372 K00259 K00031 K00789 K01658 K00068 K01583 K01677 K04487 K00006 K10248 K01908 K00889 K00978 K05288 K01113 K05551 K05283 K01803 K11717 K00645 K00150 K11262 K02301 K00823 K16794 K01657 K01662 K09699 K03857 K01676 K01783 K13024 K00116 K08232 K01519 K16876 K15019 K01480 K00763 K01623 K11186 K12256 K12447 K12407 K01961 K07757 K05605 K00529 K00560 K15778 K00034 K00850 K00324 K01763 K00216 K00239 K10534 K00038 K01440 K00245 K01703 K11142 K00600 K08681 K11185 K07432 K08692 K07751 K16871 K13950 K01584 K05710 K12420 K00208 K00796 K05780 K01520 K00827 K01764 K00868 K00919 K03858 K14455 K01830 K00973 K01011 K00033 K06127 K00325 K05296 K00806 K01582 K01868 K01902 K15018 K11207 K11541 K00528 K13507 K00630 K00981 K13497 K00955 K00363 K00757 K01251 K03800 K01255 K01710 K01874 K01899 K03179 K01120 K11140 K01101 K05954 K02257 K12503 K12319 K01892 K00394 K14634 K13370 K01930 K00854 K06132 K03472 K01477 K00323 K00975 K02434 K12406 K00895 K00167 K00968 K14379 K00568 K15728 K10246 K08683 K01784 K00943 K11263 K00164 K01869 K00949 K00111 K00244 K08316 K01595 K01664 K17363 K00972 K00036 K00844 K04710 K00888 K00994 K00626 K00261 K13774 K00873 K13524 K00099 K15232 K00586 K00284 K00866 K00254 K04116 K02552 K00591 K00940 K13755 K14655 K00240 K00627 K10807 K10703 K13509 K09759 K14163 K15407 K00902 K01870 K01939 K02437 K06131 K00648 K00980 K01904 K03779 K09698 K00016 K00802 K07151 K15888 K00380 K00271 K12466 K00650 K06129 K01769 K01689 K07548 K00162 K01903 K12505 K16363 K05361 K00922 K01867 K00969 K12318 K00901 K04283 K00635 K00995 K00915 K00605 K02275 K01956 K03691 K00942 K11778 K00914 K13038 K00266 K13998 K11147 K03526 K01678 K01893 K14656 K01880 K11754 K00658 K07823 K00166 K00832 K01117 K04042 K03860 K01426 K01858 K00065 K00057 K01665 K13777 K00226 K14066 K00953 K01875 K01770 K00721 K13935 K01553 K15303 K01920 K01749 K04110 K02259 K01905 K03784 K17450 K15233 K07513 K14156 K00797 K01948 K05783 K00950 K09458 K11419 K01965 K07547 K01895 K01953 K10258 K01959 K00364 K11433 K00218 K04709 K00039 K11422 K00105 K05861 K00286 K10012 K02377 K10251 K01896 K11188 K11121 K13704 K00609 K12410 K10207 K10705 K11428 K01490 K00880 K07542 K07508 K01659 K09680 K05284 K05552 K01599 K01886 K15734 K01768 K14760 K00383 K00639 K03366 K01772 K13800 K02615 K00235 K02523 K04105 K00604 K00944 K01913 K11153 K01619 K13939 K14454 K10249 K07509 K16792 K03918 K03780 K03183 K00655 K00864 K15779 K15866 K11381 K06123 K11809 K05715 K04708 K13789 K01054 K00948 K05355 K00894 K00024 K15372 K01958 K00632 K01964 K07250 K00322 K00652 K15785 K00927 K05857 K09065 K01514 K12504 K00382 K00260 K16066 K00999 K12345 K00326 K02433 K00064 K00820 K08729 K05356 K01881 K01509 K00278 K15762 K05286 K00859 K07535 K00760 K00791 K00993 K01807 K00974 K01834 K01885 K07753 K01069 K00030 K13938 K06167 K15669 K08680 K00762 K01091 K14467 K00787 K11528 K00019 K01057 K00654 K00274 K15371 K01907 K01889 K00432 K05275 K10708 K17100 K02428 K07253 K01001 K00819 K13937 K05830 K09251 K01512 K00840 K00647 K14675 K00161 K01581 K00879 K13761 K01915 K00637 K01809 K00821 K01835 K00088 K01890 K14080 K01465 K00134 K03342 K00805 K14067 K03801 K00811 K03148 K08081 K03859 K01737 K00966 K02226 K05310 K01950 K00818 K00328 K06125 K16881 K06215 K17195 K01845 K01610 K00967 K00228 K01705 K01591 K13787 K08730 K03644 K01911 K00262 K00611 K13373 K00532 K01450 K00861 K15013 K01008 K11155 K01598 K11540 K14268 K04568 K01476 K00836 K01840 K01837 K01937 K03331 K00026 K01810 K06970 K01897 K01883 K00384 K01955 K01922 K00833 K00287 K00234 K01647 K14674 K01491 K00023 K03341 K01522 K00526 K05955 K12506 K01518 K01951 K03851 K08295 K01488 K11152 K01900 K05298 K01887 K13941 K15022 K01876 K00615 K01836 K10760 K01866 K03527 K01873 K01586 K00773 K15231
P. berghi P. falciparum P. vivax	49	K04748 K00472 K14582 K15238 K14373 K10247 K02827 K14746 K03434 K15237 K15373 K00081 K10617 K07441 K09882 K13070 K13657 K11991 K11146 K02276 K10620 K08281 K08723 K13775 K01511 K11151 K00920 K11150 K13606 K03524 K01510 K01609 K13298 K00231 K12343 K05711 K01106 K01500 K02274 K08261 K02502 K13367 K15730 K16269 K12344 K09186 K01495 K13712 K09841
P. berghi P. cynomolgi P. falciparum	12	K01954 K01256 K00264 K11420 K03688 K00525 K07758 K00265 K10808 K01872 K01968 K10526
P. cynomolgi P. falciparum P. vivax	14	K00941 K00788 K11157 K15095 K05929 K10245 K11423 K14154 K05928 K00878 K00911 K02613 K00877 K07537
P. berghi P. cynomolgi P. vivax	11	K14338 K10203 K12454 K05297 K15634 K16343 K01577 K17218 K02361 K05607 K02636
P. berghi P. falciparum	7	K01097 K08690 K01674 K13043 K01916 K01672 K01099
P. falciparum P. vivax	3	K00263 K00072 K10205
P. cynomolgi P. falciparum	3	K00292 K10244 K07252
P. berghi P. vivax	1	K05858
P. berghi P. cynomolgi	1	K00679
P. cynomolgi P. vivax	12	K04075 K07546 K13767 K00547 K00826 K01457 K01715 K01906 K07511 K00795 K01692 K03367
P. falciparum	7	K16951 K00608 K01759 K13675 K13427 K12305 K10797
P. berghi	3	K15016 K00011 K13510
P. vivax	10	K00963 K10837 K13239 K01934 K08098 K01052 K00042 K13699 K09591 K13766
P. cynomolgi	2	K03272 K16880

Table 6. Comparative table of the metabolic capabilities in four malaria species based on the functional annotation by the RAVEN Toolbox

List of the KEGG orthology identifiers that were annotated not in all four malaria species

K04748 nitric-oxide reductase NorQ protein [EC:1.7.99.7]
K00472 prolyl 4-hydroxylase [EC:1.14.11.2]
K14582 cis-1,2-dihydro-1,2-dihydroxynaphthalene/dibenzothiophene dihydrodiol dehydrogenase [EC:1.3.1.29 1.3.1.60]
K15238 2,5-dichloro-2,5-cyclohexadiene-1,4-diol dehydrogenase 2 [EC:1.3.1.-]
K14373 C-5 ketoreductase
K10247 elongation of very long chain fatty acids protein 1
K02827 quinol oxidase polypeptide I [EC:1.9.3.-]
K14746 (S)-1-phenylethanol dehydrogenase [EC:1.1.-.-]
K03434 N-acetylglucosaminylphosphatidylinositol deacetylase [EC:3.5.1.89]
K15237 2,5-dichloro-2,5-cyclohexadiene-1,4-diol dehydrogenase 1 [EC:1.3.1.-]
K15373 sulfoacetaldehyde reductase [EC:1.1.1.313]
K00081 prostaglandin-E2 9-reductase [EC:1.1.1.189]
K10617 p-cumic alcohol dehydrogenase
K07441 beta-1,4-N-acetylglucosaminyltransferase [EC:2.4.1.141]
K09882 cobaltochelatase CobS [EC:6.6.1.2]
K13070 momilactone-A synthase [EC:1.1.1.295]
K13657 alpha-1,3-mannosyltransferase [EC:2.4.1.-]
K11991 tRNA-specific adenosine deaminase [EC:3.5.4.-]
K11146 dehydrogenase/reductase SDR family member 3 [EC:1.1.-.-]
K02276 cytochrome c oxidase subunit III [EC:1.9.3.1]
K10620 2,3-dihydroxy-2,3-dihydro-p-cumate dehydrogenase [EC:1.3.1.58]
K08281 nicotinamidase/pyrazinamidase [EC:3.5.1.19 3.5.1.-]
K08723 5'-nucleotidase [EC:3.1.3.5]
K13775 citronellol/citronellal dehydrogenase
K01511 ectonucleoside triphosphate diphosphohydrolase 5/6 [EC:3.6.1.6]
K11151 retinol dehydrogenase 10 [EC:1.1.1.-]
K00920 1-phosphatidylinositol-5-phosphate 4-kinase [EC:2.7.1.149]
K11150 retinol dehydrogenase 8 [EC:1.1.1.-]
K13606 chlorophyll(ide) b reductase [EC:1.1.1.294]
K03524 BirA family transcriptional regulator, biotin operon repressor / biotin-[acetyl-CoA-carboxylase] ligase [EC:6.3.4.15]
K01510 apyrase [EC:3.6.1.5]
K01609 indole-3-glycerol phosphate synthase [EC:4.1.1.48]
K13298 dual 3',5'-cyclic-AMP and -GMP phosphodiesterase 11 [EC:3.1.4.17 3.1.4.35]
K00231 protoporphyrinogen oxidase [EC:1.3.3.4]
K12343 3-oxo-5-alpha-steroid 4-dehydrogenase 1 [EC:1.3.99.5]
K05711 2,3-dihydroxy-2,3-dihydrophenylpropionate dehydrogenase [EC:1.3.1.-]
K01106 inositol-1,4,5-trisphosphate 5-phosphatase [EC:3.1.3.56]
K01500 N/A
K02274 cytochrome c oxidase subunit I [EC:1.9.3.1]
K08261 D-sorbitol dehydrogenase (acceptor) [EC:1.1.99.21]
K02502 ATP phosphoribosyltransferase regulatory subunit
K13367 non-specific polyamine oxidase [EC:1.5.3.17]
K15730 cytosolic prostaglandin-E synthase [EC:5.3.99.3]
K16269 cis-1,2-dihydrobenzene-1,2-diol dehydrogenase [EC:1.3.1.19]
K12344 3-oxo-5-alpha-steroid 4-dehydrogenase 2 [EC:1.3.99.5]
K09186 histone-lysine N-methyltransferase MLL1 [EC:2.1.1.43]
K01495 GTP cyclohydrolase I [EC:3.5.4.16]
K13712 phosphatidylinositol-4-phosphate 5-kinase-like protein 1 [EC:2.7.1.68]
K09841 xanthoxin dehydrogenase [EC:1.1.1.288]
K01954 carbamoyl-phosphate synthase [EC:6.3.5.5]
K01256 aminopeptidase N [EC:3.4.11.2]
K00264 glutamate synthase (NADPH/NADH) [EC:1.4.1.13 1.4.1.14]
K11420 euchromatic histone-lysine N-methyltransferase [EC:2.1.1.43]
K03688 ubiquinone biosynthesis protein
K00525 ribonucleoside-diphosphate reductase alpha chain [EC:1.17.4.1]
K07758 pyridoxal phosphatase [EC:3.1.3.74]
K00265 glutamate synthase (NADPH/NADH) large chain [EC:1.4.1.13 1.4.1.14]
K10808 ribonucleoside-diphosphate reductase subunit M2 [EC:1.17.4.1]
K01872 alanyl-tRNA synthetase [EC:6.1.1.7]
K01968 3-methylcrotonyl-CoA carboxylase alpha subunit [EC:6.4.1.4]
K10526 OPC-8:0 CoA ligase 1 [EC:6.2.1.-]
K00941 hydroxymethylpyrimidine/phosphomethylpyrimidine kinase [EC:2.7.1.49 2.7.4.7]
K00788 thiamine-phosphate pyrophosphorylase [EC:2.5.1.3]
K11157 patatin-like phospholipase domain-containing protein 4 [EC:3.1.1.-]
K15095 (+)-neomenthol dehydrogenase [EC:1.1.1.208]
K05929 phosphoethanolamine N-methyltransferase [EC:2.1.1.103]
K10245 fatty acid elongase 2 [EC:2.3.1.-]
K11423 histone-lysine N-methyltransferase SETD2 [EC:2.1.1.43]
K14154 thiamine-phosphate diphosphorylase / hydroxyethylthiazole kinase [EC:2.5.1.3 2.7.1.50]
K05928 tocopherol O-methyltransferase [EC:2.1.1.95]
K00878 hydroxyethylthiazole kinase [EC:2.7.1.50]
K00911 1D-myo-inositol-triphosphate 3-kinase [EC:2.7.1.127]
K02613 phenylacetic acid degradation NADH oxidoreductase
K00877 hydroxymethylpyrimidine/phosphomethylpyrimidine kinase [EC:2.7.1.49 2.7.4.7]
K07537 cyclohexa-1,5-dienecarbonyl-CoA hydratase [EC:4.2.1.100]
K14338 cytochrome P450 / NADPH-cytochrome P450 reductase [EC:1.14.14.1 1.6.2.4]
K10203 elongation of very long chain fatty acids protein 6 [EC:2.3.1.-]
K12454 CDP-paratose 2-epimerase [EC:5.1.3.10]
K05297 rubredoxin-NAD+ reductase [EC:1.18.1.1]
K15634 probable phosphoglycerate mutase [EC:5.4.2.12]
K16343 calcium-independent phospholipase A2 [EC:3.1.1.4]
K01577 oxalyl-CoA decarboxylase [EC:4.1.1.8]
K17218 sulfide:quinone oxidoreductase [EC:1.8.5.-]
K02361 isochorismate synthase [EC:5.4.4.2]
K05607 methylglutaconyl-CoA hydratase [EC:4.2.1.18]
K02636 cytochrome b6-f complex iron-sulfur subunit [EC:1.10.99.1]
K01097 N-acylneuraminate-9-phosphatase [EC:3.1.3.29]
K08690 cis-2,3-dihydrobiphenyl-2,3-diol dehydrogenase [EC:1.3.1.56]
K01674 carbonic anhydrase [EC:4.2.1.1]
K13043 N-succinyl-L-ornithine transcarbamylase [EC:2.1.3.11]
K01916 NAD+ synthase [EC:6.3.1.5]
K01672 carbonic anhydrase [EC:4.2.1.1]
K01099 phosphatidylinositol-bisphosphatase [EC:3.1.3.36]
K00263 leucine dehydrogenase [EC:1.4.1.9]
K00072 sepiapterin reductase [EC:1.1.1.153]
K10205 elongation of very long chain fatty acids protein 2 [EC:2.3.1.-]
K00292 saccharopine dehydrogenase (NAD+, L-glutamate forming) [EC:1.5.1.9]
K10244 elongation of very long chain fatty acids protein 5 [EC:2.3.1.-]
K07252 dolichyldiphosphatase [EC:3.6.1.43]
K05858 phospholipase C, beta [EC:3.1.4.11]
K00679 phospholipid:diacylglycerol acyltransferase [EC:2.3.1.158]
K04075 tRNA(Ile)-lysidine synthase [EC:6.3.4.-]
K07546 E-phenylitaconyl-CoA hydratase [EC:4.2.1.-]
K13767 enoyl-CoA hydratase [EC:4.2.1.17]
K00547 homocysteine S-methyltransferase [EC:2.1.1.10]
K00826 branched-chain amino acid aminotransferase [EC:2.6.1.42]
K01457 allophanate hydrolase [EC:3.5.1.54]
K01715 3-hydroxybutyryl-CoA dehydratase [EC:4.2.1.55]
K01906 6-carboxyhexanoate--CoA ligase [EC:6.2.1.14]
K07511 enoyl-CoA hydratase [EC:4.2.1.17]
K00795 farnesyl diphosphate synthase [EC:2.5.1.1 2.5.1.10]
K01692 enoyl-CoA hydratase [EC:4.2.1.17]
K03367 D-alanine--poly(phosphoribitol) ligase subunit 1 [EC:6.1.1.13]
K16951 anaerobic sulfite reductase subunit B
K00608 aspartate carbamoyltransferase [EC:2.1.3.2]
K01759 lactoylglutathione lyase [EC:4.4.1.5]
K13675 UDP-glucose:O-linked fucose beta-1,3-glucosyltransferase [EC:2.4.1.-]
K13427 nitric-oxide synthase, plant [EC:1.14.13.39]
K12305 ectonucleoside triphosphate diphosphohydrolase 4 [EC:3.6.1.6]
K10797 2-enoate reductase [EC:1.3.1.31]
K15016 enoyl-CoA hydratase / 3-hydroxyacyl-CoA dehydrogenase [EC:4.2.1.17 1.1.1.35]
K00011 aldehyde reductase [EC:1.1.1.21]
K13510 lysophosphatidylcholine acyltransferase / lyso-PAF acetyltransferase [EC:2.3.1.23 2.3.1.67]
K00963 UTP--glucose-1-phosphate uridylyltransferase [EC:2.7.7.9]
K10837 O-phosphoseryl-tRNA(Sec) kinase [EC:2.7.1.164]
K13239 peroxisomal 3,2-trans-enoyl-CoA isomerase [EC:5.3.3.8]
K01934 5-formyltetrahydrofolate cyclo-ligase [EC:6.3.3.2]
K08098 phosphatidylinositol glycan, class Z [EC:2.4.1.-]
K01052 lysosomal acid lipase/cholesteryl ester hydrolase [EC:3.1.1.13]
K00042 2-hydroxy-3-oxopropionate reductase [EC:1.1.1.60]
K13699 abhydrolase domain-containing protein 5 [EC:2.3.1.51]
K09591 probable steroid reductase DET2 [EC:1.3.99.-]
K13766 methylglutaconyl-CoA hydratase [EC:4.2.1.18]
K03272 D-beta-D-heptose 7-phosphate kinase / D-beta-D-heptose 1-phosphate adenosyltransferase [EC:2.7.1.- 2.7.7.-]
K16880 2-oxoglutaroyl-CoA hydrolase

Appendix 3. Methods, materials and protocols

Cloning of DNA constructs

All amplifications were performed with LA Taq (TaKaRa) polymerase and primers used are listed in Table 7, Appendix 4.

Knockin and knockdown of TgRK (pKI-TgRK-3Ty-LoxP-3'UTR-LoxP-U1): Genomic fragment of *TgRK* (TGME49_216740) was amplified using primers 4869/4770 and subsequently digested with *KpnI* and *SbfI* prior to cloning in the same sites of the modified C-terminal destabilization vector pG152-3Ty-LoxP-3'UTRSag1-HXGPRT-LoxP-U1 [278]. Prior to transfection the plasmid was linearized with *NsiI.*

Knockin and knockdown of TgFMNAT (pKI-TgFMNAT-3Ty-LoxP-3'UTR-LoxP-U1): Genomic fragment of *TgFMNAT* (TGME49_214280) was amplified using primers 4871/4772 and subsequently digested with *KpnI* and *SbfI* prior to cloning in the same sites of the modified C-terminal destabilization vector pG152-3Ty-LoxP-3'UTRSag1-HXGPRT-LoxP-U1 [278]. Prior to transfection the plasmid was linearized with *BssHII.*

Knockout of TgFMNAT using CRISPR/CAS9 plasmid [146] : This vector has been generated using the Q5 site-directed mutagenesis kit (New England Biolabs) with the vector pSAG1::CAS9-U6::sgUPRT as template (a gift from Dr. L. D. Sibley). The UPRT-targeting gRNA was replaced by the *TgFMNAT* (TGME49_214280) specific gRNA using the primer pair 4997/4883 (gRNA sequence is underlined in Table 7).

Parasite transfection and selection of stable transformants

Parasite transfections were performed by electroporation as previously described [143].The *hxgprt* gene was used as a positive selectable marker in the presence of mycophenolic acid (25 µg/mL) and xanthine (50 µg/mL) for *pKI-TgRK-3Ty-LoxP-3'UTR-LoxP-U1* and *pKI-TgFMNAT-3Ty-LoxP-3'UTR-LoxP-U1* vectors transfected in Ku80ko tachyzoites as previously described [279].

Resistant parasites were cloned by limiting dilution in 96 well plates and clones were assessed by PCR using their genomic DNA.

To efficiently disrupt the *TgFMNAT* locus, 8 µg of the *TgFMNAT* gRNA-specific Crispr/Cas9 vector and with 30 µg of DHFR selection cassette were transfected into wild

type RH parasites. This DHFR selection cassette obtained by PCR (using p2851 plasmid as a template, primers 4998-4999) and contained point mutations, which confer resistance to pyrimethamine; it was also flanked by 28 nt long homology arms to *TgFMNAT* on each side. 48 hours after transfection, GFP positive parasites were sorted by flow cytometry and cloned into 96 well plates using a Moflo Astrios (Beckman Coulter). Medium for the cloning contained 0.25 μM of pyrimethamine for selection of the parasites that repaired the double strand break made by Cas9 using homologous template with mutated DHFR.

Preparation of *T. gondii* genomic DNA

Genomic DNA was prepared from tachyzoites using the Wizard SV genomic DNA purification system (Promega). Correct integration of the different constructs into the genome of the various strains was determined by genomic PCR using GoTaq Green Master Mix (Promega).

Antibodies

The antibodies used in this study were described previously as follows: polyclonal rabbit anti-GAP45, rabbit anti-TgProfilin [280], monoclonal mouse anti-Ty (BB2), mouse monoclonal anti-F1-ATPase beta subunit (P. Bradley, unpublished) (5F4), mouse monoclonal anti-ATrx1 11G8 [281]. For Western blot analyses, secondary peroxidase conjugated goat anti-rabbit or mouse antibodies (Molecular Probes) were used. For immunofluorescence analyses, the secondary antibodies Alexa Fluor 488 and Alexa Fluor 594-conjugated goat anti-mouse or rabbit antibodies (Molecular Probes) were used.

Immunofluorescence assay (IFA) and confocal microscopy

Parasite-infected HFF cells seeded on cover slips were fixed with 4% paraformaldehyde/0.05% glutaraldehyde (PFA/Glu) in PBS. Fixed cells were then processed as previously described [282]. Confocal images were generated with a Zeiss (LSM700, objective apochromat 63x/1.4 oil) laser scanning confocal microscope at the Bioimaging core facility of the Faculty of Medicine, University of Geneva. Stacks of sections were processed with ImageJ and projected using the maximum projection tool.

Western blot analyses

Parasites were lysed in PBS-1% Triton X-100 and mixed with SDS–PAGE loading buffer under reducing conditions. The suspension was subjected to sonication on ice. SDS-PAGE

was performed using standard methods. Separated proteins were transferred to nitrocellulose membranes and probed with appropriate antibodies in 5% non-fat milk in PBS-0.05% Tween20. Bound secondary peroxidase conjugated antibodies were visualized using the SuperSignal (Pierce).

Fractionation analyses

Parasites were harvested upon lysis of confluent monolayer of HFF cells from one 6 cm plate, precipitated by centrifugation at 3000 rpm for 5 min, and washed once with PBS. The pellets were resuspended in 300 µl of SoTE (20 mM Tris-HCl, pH 7.5, 0.6 M Sorbitol, 2mM EDTA) and split in 3 aliquots of 100 µl. 100 ul of 0.01% digitonin was added to the first aliquot, the same volume of the 0.05% digitonin to the second and 100 µl of SoTE to the third. The content of the tubes with digitonin was mixed by inverting once and precipitated by centrifugation at 8000 rpm for 10 min. Collected supernatant contained the proteins of the cytosolic fraction. The pellets containing organellar fraction of the proteins were resuspended in 200 µl of SoTE. 40 µl of the sample buffer and 10 µl of DTT were added to the suspensions of cytosolic, organellar fractions and the one not treated with digitonin. Volume of the untreated sample was adjusted by adding SoTE to be equal to the treated samples. For separation of the proteins electrophoresis in 10% acrylamide gel was used followed by wet transfer of the proteins from the gel to a membrane. The first immunoblot was performed essentialy as described above with the anti-Ty and anti-TgProfilin primary antibodies, and anti-mouse HRP and anti-rabbit YFP secondary antibodies for detection. Subsequently, the antibodies were stripped from the membrane by treatment 30 min treatment with 10 ml of 2% SDS-PBS and 60 µl of beta-mercaptoethanol at room temperature. After washing the membrane with PBS-Tween, anti-Cpn60 and anti-HSP70 primary antibodies were applied. For visualization secondary anti-Rabbit-YFP antibody was used.

Appendix 4. Primers and sequencing results

Cloning	
Identifier	Sequence 5'→3'
4869	GCCGGTACCCTTACACACCTTGCCGAACG
4870	GCCCCTGCAGGGAGACTCGAAGTGTGTGGCG
4871	GCCGGTACCCACTCTCGATCTTCGCTTCTG
4872	GCCCCTGCAGGGTTTCTTCCAGATCTTTCATCTGCCC
5104	CTTCTCGTCTCCCTGTATCC
4994	GGCATGCATTTCTTCCAGATCTTTCATCTGCCC
4995	GCCATGCATATGGCGCCGCGCTAACTGATTC
4997	GGAACAAGACTCTAACCGCTGTTTTAGAGCTAGAAATAGC
4883	AACTTGACATCCCCATTTAC
4993	GGCGAATTCCGACTTCAATGCGATTGCGTTTCACG
4998	GAAGGGCGACATCGTCGGCGAGGAGTTCGCGGCCGCTCTA
4999	CGAGTTCGAACTCGTTTGCAGTGTGCTCGCGGAAGATCCG
PCR analyses	
	Sequence 5'→3'
4950	GAACACTTCGGTGACCATTG
4951	CTCTCTTGTGTCTCGCTTCC
p30a	CAGTTTCTTTATAATGGGGC
M13F	GTAAAACGACGGCCAGT
4952	TGAGGCCCGCTGGTATTTC
4953	AGAGTGATAACGGTCGAGC
p30a	CAGTTTCTTTATAATGGGGC
M13F	GTAAAACGACGGCCAGT
4872	GCCCCTGCAGGGTTTCTTCCAGATCTTTCATCTGCCC
4996	GGCGAATTCATGGCGCCAGAGAGACAGGGCGATG
2087	ACTGCCTGGAATCCTGCAGCGC
2018	CTTGGGGGTCATCGCGACGACCAGAC

Table 7. Primers used in the experimental study on TgRK and TgFMNAT

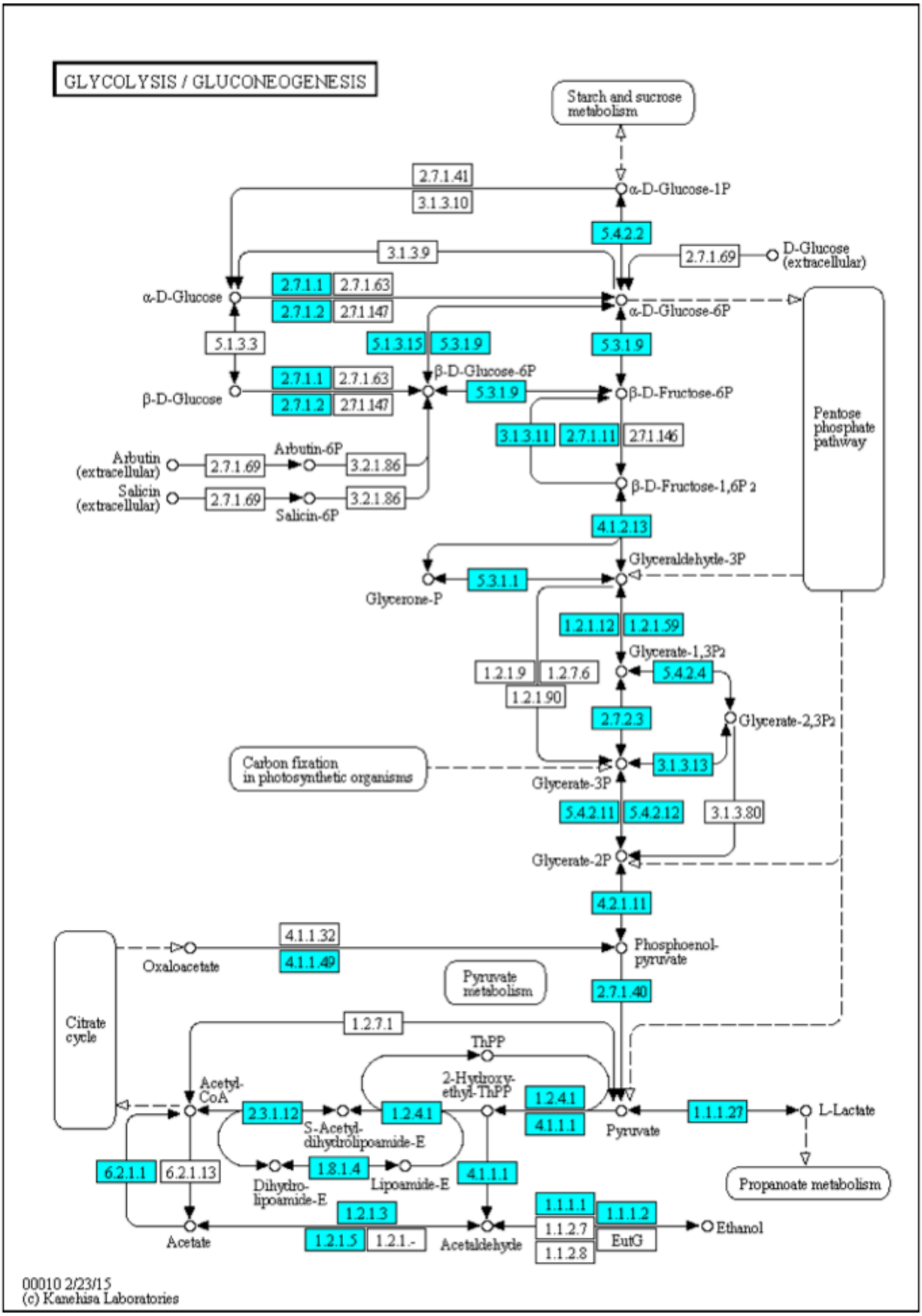
Appendix 5. Visualization of ToxoNet1 using KEGG mapper

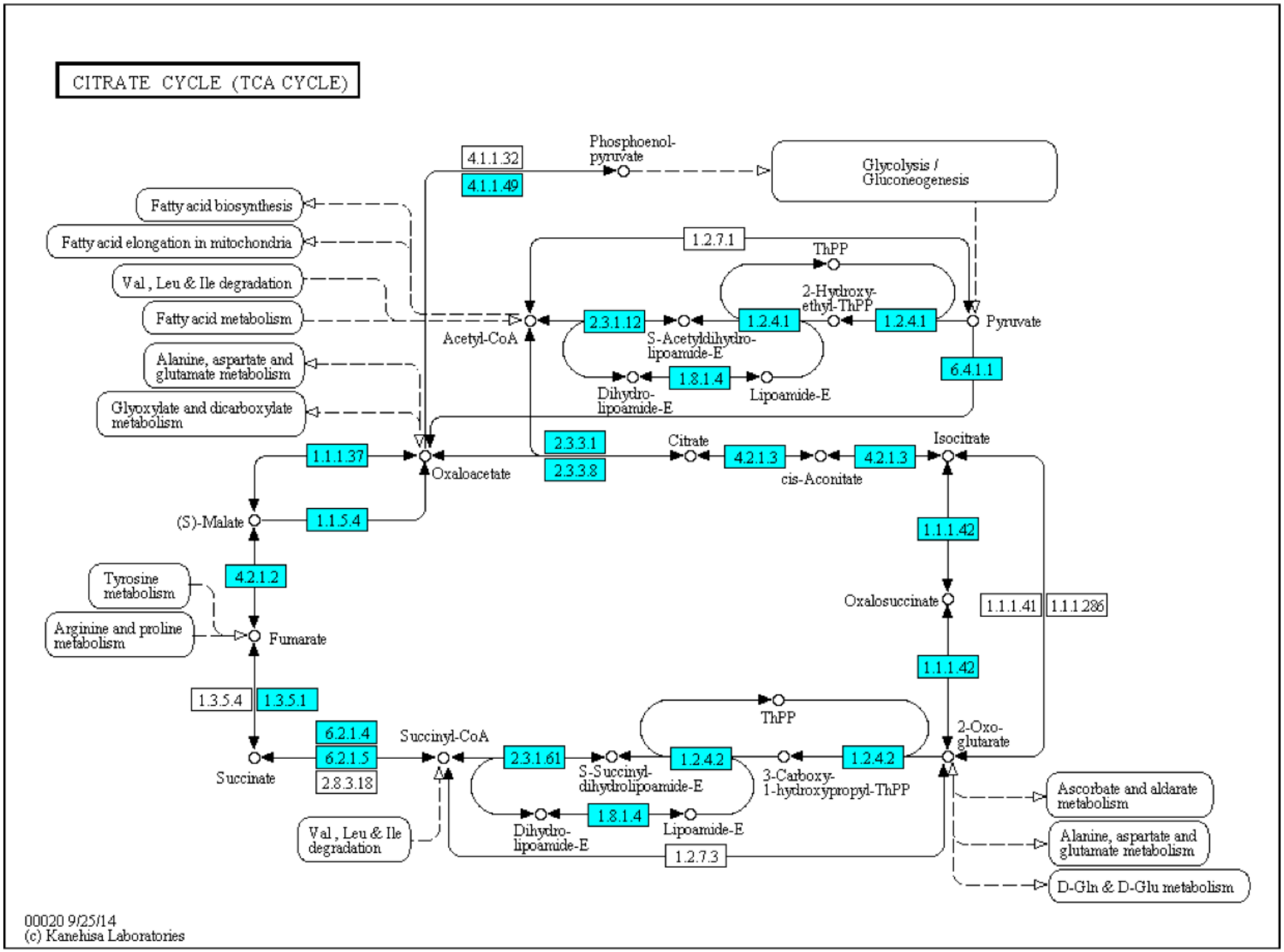
One of the current limitations of large-scale metabolic models is the lack of a visual and intuitive way to illustrate model's content and specific flux distributions. Herein I provide a snapshot of a partial solution to this issue. KEGG mapper allows visualization of the list of E.C. entries, K-identifiers or KEGG reaction identifiers by highlighting them on KEGG metabolic maps. Such lists of the entries are present for example in ToxoNet1 can and be easily extracted, treated such that no duplicated entries are present and visualized using KEGG mapper.

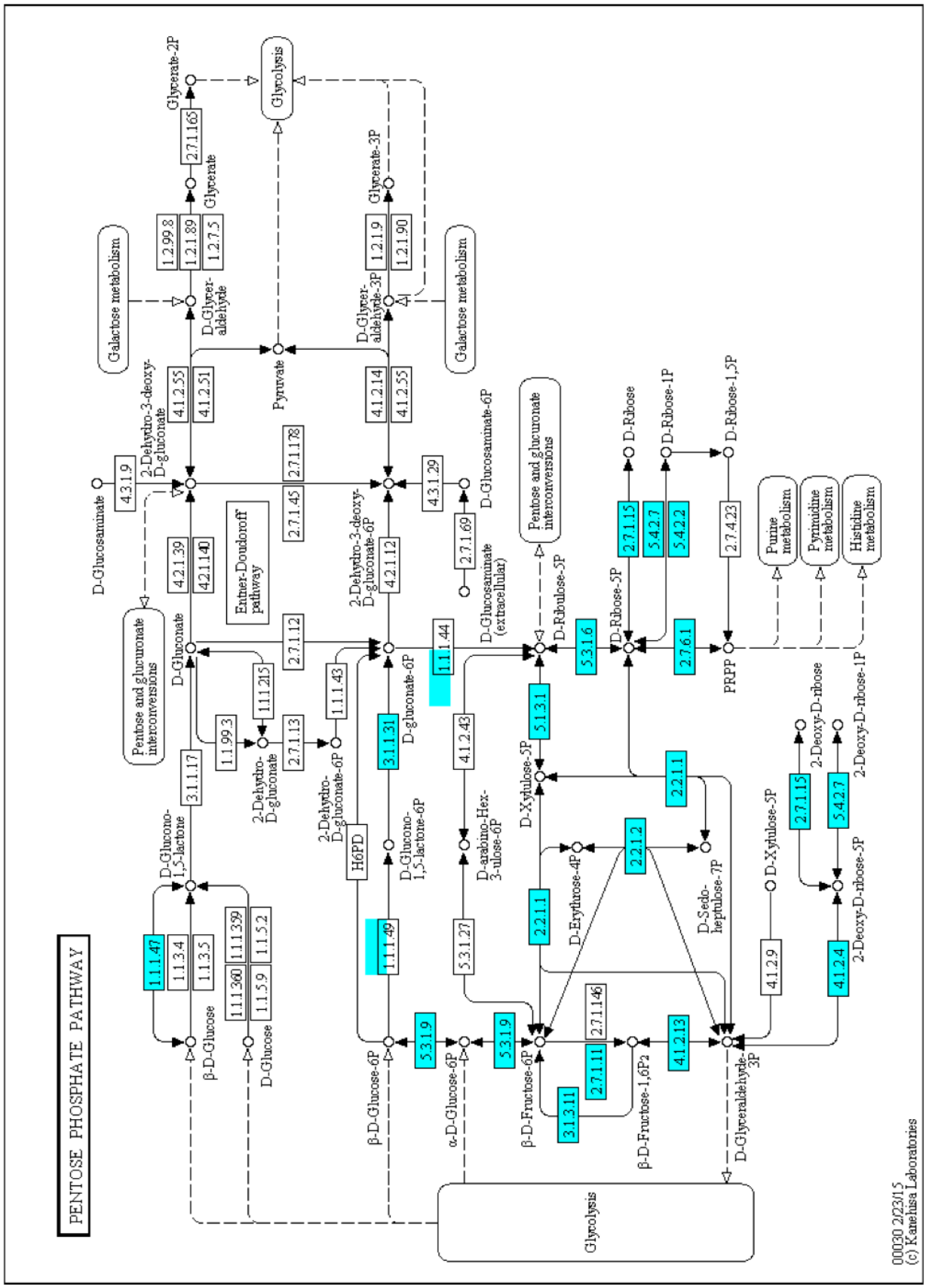
This web-based software is available at http://www.genome.jp/kegg/tool/map_pathway2.html

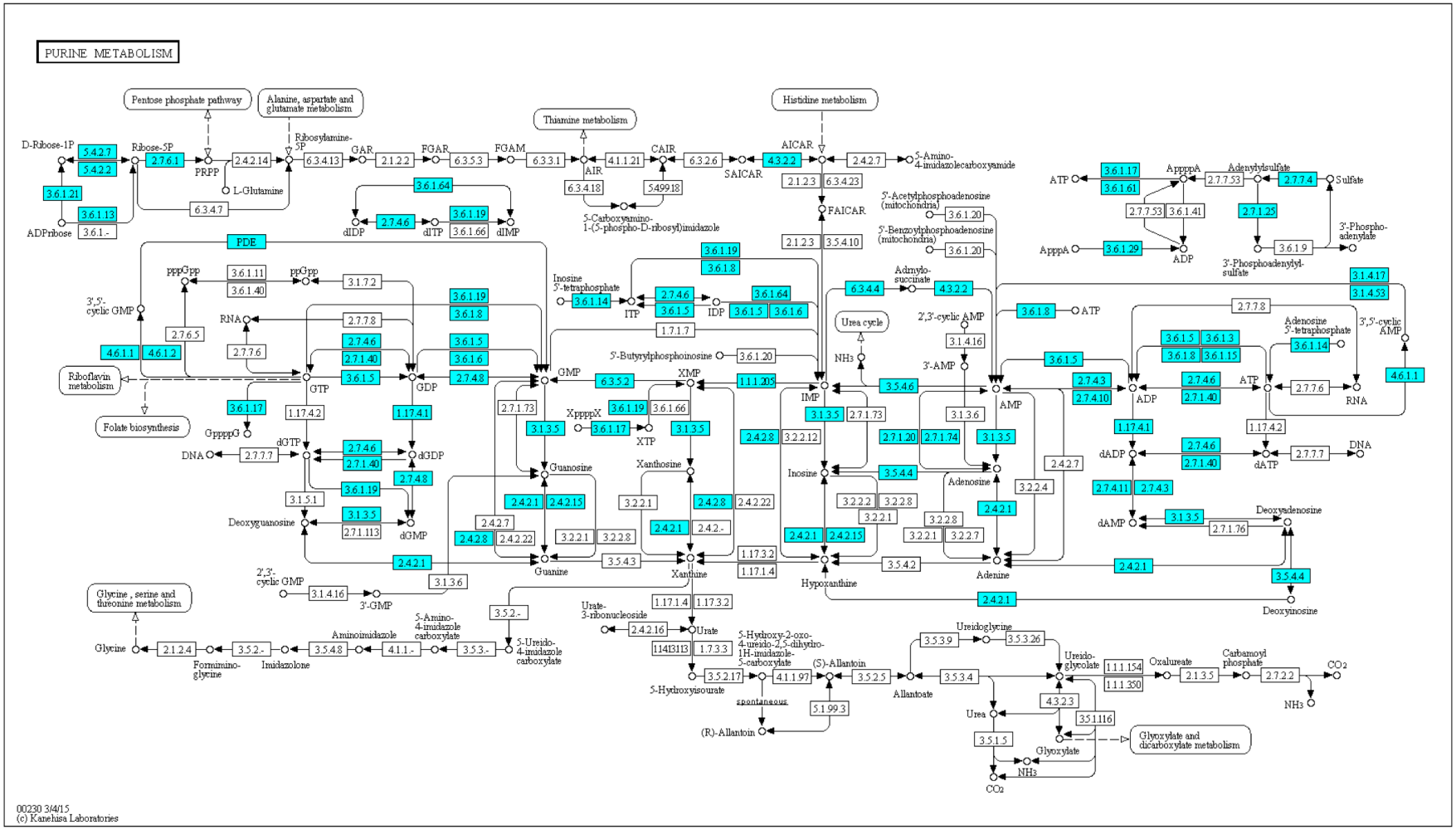
Herein I provide the list of unique E.C. numbers, which can be used as an input for the KEGG Mapper to obtain a visual representation of the metabolic capabilities of *T. gondii* present in ToxoNet1 (see 5 selected figures below).

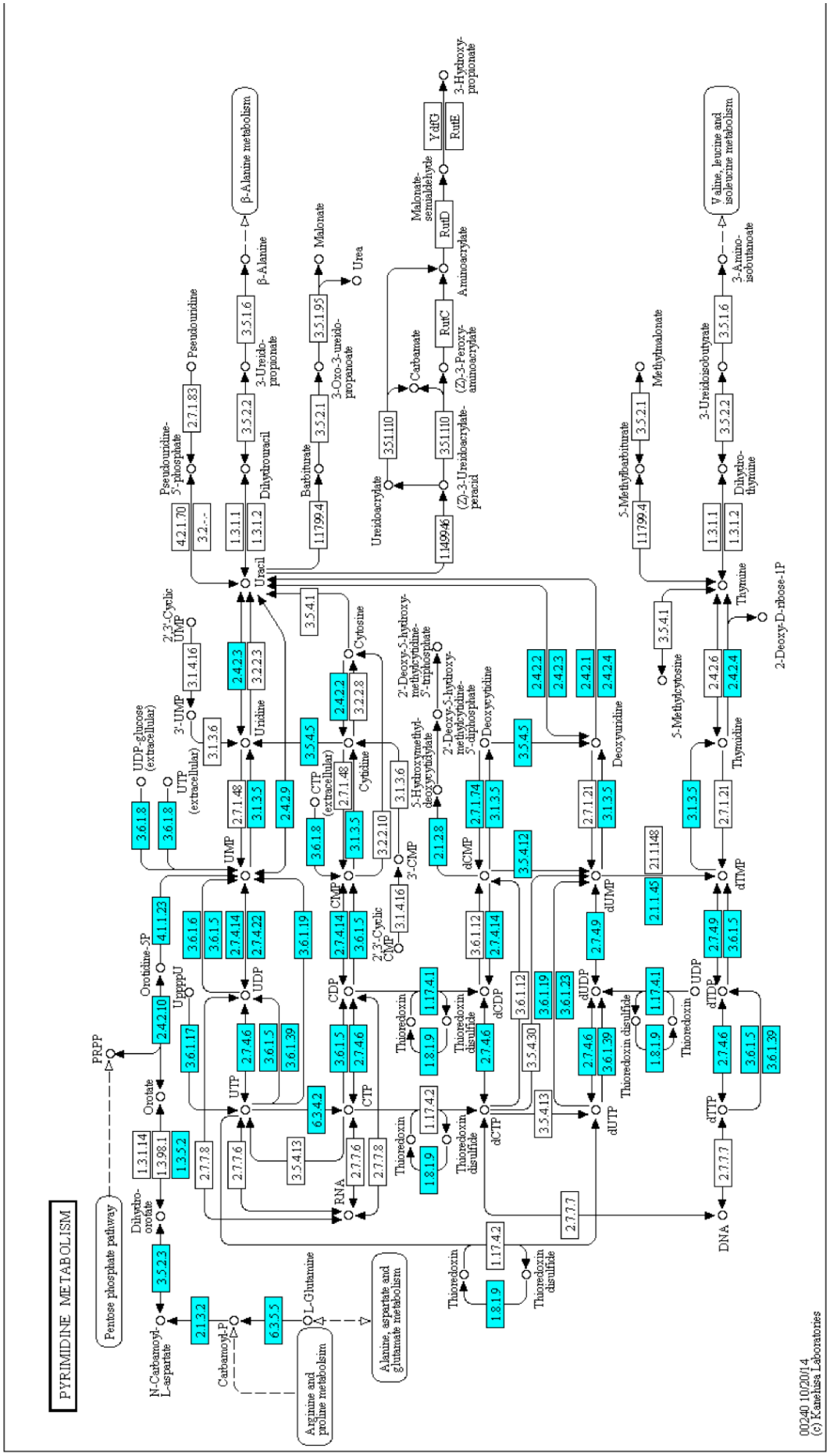
1.1.1.- 1.1.1.1 1.1.1.100 1.1.1.102 1.1.1.105 1.1.1.125 1.1.1.127 1.1.1.146 1.1.1.153 1.1.1.157 1.1.1.169 1.1.1.178 1.1.1.184 1.1.1.189 1.1.1.2 1.1.1.205 1.1.1.21 1.1.1.211 1.1.1.220 1.1.1.25 1.1.1.26 1.1.1.267 1.1.1.27 1.1.1.271 1.1.1.272 1.1.1.282 1.1.1.283 1.1.1.284 1.1.1.29 1.1.1.299 1.1.1.3 1.1.1.300 1.1.1.307 1.1.1.31 1.1.1.310 1.1.1.330 1.1.1.35 1.1.1.351 1.1.1.36 1.1.1.37 1.1.1.42 1.1.1.44 1.1.1.47 1.1.1.49 1.1.1.51 1.1.1.60 1.1.1.62 1.1.1.71 1.1.1.72 1.1.1.78 1.1.1.79 1.1.1.8 1.1.1.81 1.1.1.94 1.1.1.95 1.1.4.1 1.1.5.3 1.1.5.4 1.1.99.1 1.10.2.2 1.11.1.- 1.11.1.15 1.11.1.21 1.11.1.6 1.11.1.7 1.11.1.9 1.14.-.- 1.14.11.2 1.14.12.- 1.14.12.1 1.14.12.11 1.14.12.18 1.14.12.19 1.14.12.3 1.14.13.- 1.14.16.1 1.14.16.2 1.14.16.4 1.14.19.1 1.14.19.2 1.15.1.1 1.16.1.8 1.17.1.2 1.17.1.8 1.17.4.1 1.17.7.1 1.18.1.1 1.18.1.2 1.18.1.3 1.18.1.4 1.18.1.6 1.2.1.- 1.2.1.11 1.2.1.12 1.2.1.16 1.2.1.18 1.2.1.19 1.2.1.24 1.2.1.27 1.2.1.29 1.2.1.3 1.2.1.36 1.2.1.39 1.2.1.4 1.2.1.41 1.2.1.5 1.2.1.53 1.2.1.59 1.2.1.79 1.2.1.8 1.2.1.88 1.2.4.1 1.2.4.2 1.2.4.4 1.3.1.10 1.3.1.20 1.3.1.22 1.3.1.33 1.3.1.38 1.3.1.9 1.3.1.93 1.3.3.1 1.3.3.3 1.3.3.4 1.3.3.6 1.3.5.1 1.3.5.2 1.3.8.1 1.3.8.4 1.3.8.6 1.3.8.7 1.3.8.8 1.3.8.9 1.3.99.- 1.3.99.1 1.3.99.12 1.3.99.22 1.4.1.1 1.4.1.13 1.4.1.14 1.4.1.2 1.4.1.3 1.4.1.4 1.4.3.5 1.4.4.2 1.5.1.2 1.5.1.3 1.5.1.34 1.5.1.5 1.5.1.7 1.5.1.9 1.5.99.8 1.6.1.1 1.6.1.2 1.6.2.2 1.6.2.4 1.6.5.3 1.6.5.4 1.7.1.1 1.7.1.2 1.7.1.3 1.7.1.4 1.8.1.2 1.8.1.4 1.8.1.7 1.8.1.9 1.8.3.1 1.8.3.5 1.8.3.6 1.9.3.1 2.1.1.10 2.1.1.114 2.1.1.13 2.1.1.201 2.1.1.222 2.1.1.43 2.1.1.45 2.1.1.59 2.1.1.60 2.1.1.64 2.1.1.98 2.1.2.1 2.1.2.10 2.1.2.11 2.1.2.8 2.1.2.9 2.1.3.2 2.10.1.1 2.2.1.1 2.2.1.2 2.2.1.3 2.2.1.6 2.2.1.7 2.3.-.- 2.3.1.- 2.3.1.12 2.3.1.15 2.3.1.155 2.3.1.16 2.3.1.168 2.3.1.174 2.3.1.179 2.3.1.180 2.3.1.181 2.3.1.199 2.3.1.20 2.3.1.24 2.3.1.26 2.3.1.31 2.3.1.37 2.3.1.38 2.3.1.39 2.3.1.4 2.3.1.41 2.3.1.47 2.3.1.50 2.3.1.51 2.3.1.61 2.3.1.85 2.3.1.86 2.3.1.9 2.3.3.1 2.3.3.3 2.3.3.5 2.3.3.8 2.4.1.- 2.4.1.1 2.4.1.117 2.4.1.132 2.4.1.141 2.4.1.142 2.4.1.15 2.4.1.18 2.4.1.198 2.4.1.256 2.4.1.257 2.4.1.265 2.4.1.267 2.4.1.41 2.4.1.83 2.4.2.- 2.4.2.1 2.4.2.10 2.4.2.15 2.4.2.2 2.4.2.23 2.4.2.29 2.4.2.3 2.4.2.4 2.4.2.8 2.4.2.9 2.4.99.18 2.5.1.- 2.5.1.1 2.5.1.10 2.5.1.15 2.5.1.19 2.5.1.29 2.5.1.30 2.5.1.31 2.5.1.39 2.5.1.47 2.5.1.48 2.5.1.49 2.5.1.54 2.5.1.58 2.5.1.6 2.5.1.61 2.5.1.65 2.5.1.75 2.5.1.81 2.5.1.82 2.5.1.83 2.5.1.84 2.5.1.85 2.5.1.91 2.6.1.1 2.6.1.13 2.6.1.16 2.6.1.19 2.6.1.2 2.6.1.23 2.6.1.39 2.6.1.4 2.6.1.42 2.6.1.44 2.6.1.45 2.6.1.5 2.6.1.51 2.6.1.52 2.6.1.57 2.6.1.6 2.6.1.67 2.6.1.83 2.6.1.85 2.6.1.9 2.7.1.1 2.7.1.107 2.7.1.11 2.7.1.137 2.7.1.144 2.7.1.148 2.7.1.15 2.7.1.151 2.7.1.2 2.7.1.20 2.7.1.23 2.7.1.24 2.7.1.25 2.7.1.26 2.7.1.30 2.7.1.31 2.7.1.32 2.7.1.33 2.7.1.34 2.7.1.35 2.7.1.39 2.7.1.4 2.7.1.40 2.7.1.56 2.7.1.6 2.7.1.67 2.7.1.68 2.7.1.7 2.7.1.71 2.7.1.74 2.7.1.8 2.7.1.82 2.7.1.90 2.7.2.11 2.7.2.3 2.7.2.4 2.7.4.10 2.7.4.11 2.7.4.12 2.7.4.14 2.7.4.15 2.7.4.22 2.7.4.24 2.7.4.25 2.7.4.3 2.7.4.4 2.7.4.6 2.7.4.8 2.7.4.9 2.7.6.1 2.7.6.2 2.7.6.3 2.7.7.1 2.7.7.13 2.7.7.14 2.7.7.15 2.7.7.18 2.7.7.2 2.7.7.23 2.7.7.27 2.7.7.3 2.7.7.4 2.7.7.41 2.7.7.44 2.7.7.60 2.7.7.63 2.7.7.64 2.7.7.73 2.7.7.9 2.7.8.- 2.7.8.1 2.7.8.11 2.7.8.15 2.7.8.2 2.7.8.27 2.7.8.29 2.7.8.3 2.7.8.5 2.7.8.7 2.7.9.3 2.8.1.1 2.8.1.2 2.8.1.7 2.8.1.8 2.9.1.2 3.1.1.- 3.1.1.13 3.1.1.3 3.1.1.31 3.1.1.4 3.1.1.47 3.1.1.5 3.1.1.92 3.1.2.14 3.1.2.2 3.1.2.21 3.1.2.22 3.1.2.4 3.1.2.6 3.1.3.- 3.1.3.1 3.1.3.11 3.1.3.12 3.1.3.13 3.1.3.2 3.1.3.25 3.1.3.3 3.1.3.35 3.1.3.37 3.1.3.4 3.1.3.41 3.1.3.5 3.1.3.7 3.1.3.73 3.1.3.74 3.1.3.81 3.1.4.11 3.1.4.17 3.1.4.2 3.1.4.35 3.1.4.46 3.1.4.53 3.1.4.55 3.2.1.106 3.2.1.33 3.2.1.49 3.2.1.84 3.3.1.1 3.4.11.1 3.4.11.2 3.4.11.23 3.4.13.- 3.4.22.- 3.4.24.84 3.5.1.1 3.5.1.111 3.5.1.19 3.5.1.2 3.5.1.3 3.5.1.31 3.5.1.38 3.5.1.47 3.5.1.63 3.5.1.89 3.5.2.3 3.5.3.1 3.5.4.- 3.5.4.12 3.5.4.16 3.5.4.17 3.5.4.4 3.5.4.5 3.5.4.6 3.5.4.9 3.5.5.4 3.5.99.6 3.6.1.- 3.6.1.1 3.6.1.13 3.6.1.14 3.6.1.15 3.6.1.17 3.6.1.19 3.6.1.21 3.6.1.23 3.6.1.29 3.6.1.3 3.6.1.39 3.6.1.42 3.6.1.5 3.6.1.53 3.6.1.6 3.6.1.61 3.6.1.64 3.6.1.7 3.6.1.8 3.6.3.1 3.6.5.1 3.6.5.2 3.6.5.3 4.1.1.- 4.1.1.1 4.1.1.15 4.1.1.18 4.1.1.19 4.1.1.20 4.1.1.23 4.1.1.31 4.1.1.36 4.1.1.37 4.1.1.49 4.1.1.65 4.1.1.81 4.1.2.13 4.1.2.25 4.1.2.4 4.1.3.26 4.1.3.30 4.1.3.38 4.1.3.4 4.1.3.40 4.1.99.18 4.2.1.1 4.2.1.10 4.2.1.107 4.2.1.11 4.2.1.116 4.2.1.134 4.2.1.17 4.2.1.18 4.2.1.2 4.2.1.22 4.2.1.24 4.2.1.3 4.2.1.4 4.2.1.47 4.2.1.55 4.2.1.59 4.2.1.74 4.2.1.75 4.2.1.79 4.2.1.96 4.2.1.99 4.2.3.1 4.2.3.12 4.2.3.3 4.2.3.4 4.2.3.5 4.3.2.2 4.3.3.6 4.3.3.7 4.4.1.1 4.4.1.11 4.4.1.16 4.4.1.17 4.4.1.2 4.4.1.5 4.4.1.8 4.6.1.1 4.6.1.12 4.6.1.2 4.99.1.1 5.1.1.7 5.1.3.1 5.1.3.15 5.1.3.2 5.3.1.1 5.3.1.6 5.3.1.8 5.3.1.9 5.3.2.1 5.3.99.3 5.4.2.1 5.4.2.11 5.4.2.12 5.4.2.2 5.4.2.3 5.4.2.4 5.4.2.5 5.4.2.7 5.4.2.8 6.1.1.1 6.1.1.10 6.1.1.11 6.1.1.12 6.1.1.14 6.1.1.15 6.1.1.16 6.1.1.17 6.1.1.18 6.1.1.19 6.1.1.2 6.1.1.20 6.1.1.21 6.1.1.22 6.1.1.23 6.1.1.24 6.1.1.3 6.1.1.4 6.1.1.5 6.1.1.6 6.1.1.7 6.1.1.9 6.2.1.- 6.2.1.1 6.2.1.16 6.2.1.17 6.2.1.3 6.2.1.36 6.2.1.4 6.2.1.5 6.3.1.2 6.3.2.1 6.3.2.12 6.3.2.17 6.3.2.2 6.3.2.3 6.3.2.5 6.3.3.2 6.3.4.14 6.3.4.16 6.3.4.2 6.3.4.21 6.3.4.4 6.3.5.1 6.3.5.2 6.3.5.4 6.3.5.5 6.3.5.6 6.3.5.7 6.4.1.1 6.4.1.2 6.4.1.3 6.4.1.4


GLYCOLYSIS / GLUCONEOGENESIS
Starch and sucrose metabolism
2.7.1.41
3.1.3.10
α-D-Glucose-1P
5.4.2.2
3.1.3.9
2.7.1.69
D-Glucose (extracellular)
α-D-Glucose
2.7.1.1
2.7.1.63
2.7.1.2
2.7.1.147
α-D-Glucose-6P
5.1.3.3
5.1.3.15
5.3.1.9
β-D-Glucose-6P
β-D-Glucose
β-D-Fructose-6P
Pentose phosphate pathway
3.1.3.11
2.7.1.11
2.7.1.146
Arbutin (extracellular)
Arbutin-6P
3.2.1.86
Salicin (extracellular)
Salicin-6P
β-D-Fructose-1,6P 2
4.1.2.13
Glyceraldehyde-3P
Glycerone-P
1.2.1.12
1.2.1.59
Glycerate-1,3P2
1.2.1.9
1.2.7.6
1.2.1.90
5.4.2.4
2.7.2.3
Glycerate-2,3P2
Carbon fixation in photosynthetic organisms
3.1.3.13
Glycerate-3P
5.4.2.11
5.4.2.12
3.1.3.80
Glycerate-2P
4.2.1.11
4.1.1.32
4.1.1.49
Oxaloacetate
Phosphoenol-pyruvate
Pyruvate metabolism
2.7.1.40
Citrate cycle
1.2.7.1
ThPP
Acetyl-CoA
2-Hydroxy-ethyl-ThPP
1.2.4.1
2.3.1.12
S-Acetyl-dihydrolipoamide-E
4.1.1.1
Pyruvate
1.1.1.27
L-Lactate
6.2.1.1
6.2.1.13
1.8.1.4
Dihydro-lipoamide-E
Lipoamide-E
4.1.1.1
Propanoate metabolism
1.2.1.3
1.1.1.1
1.1.1.2
1.1.1.7
EutG
1.1.1.8
Ethanol
Acetate
1.2.1.5
1.2.1.-
Acetaldehyde
00010 2/23/15
(c) Kanehisa Laboratories


CITRATE CYCLE (TCA CYCLE)
Phosphoenol-pyruvate
4.1.1.32
4.1.1.49
Glycolysis / Gluconeogenesis
Fatty acid biosynthesis
Fatty acid elongation in mitochondria
Val, Leu & Ile degradation
Fatty acid metabolism
Alanine, aspartate and glutamate metabolism
Glyoxylate and dicarboxylate metabolism
1.2.7.1
ThPP
2-Hydroxy-ethyl-ThPP
2.3.1.12
1.2.4.1
1.2.4.1
Pyruvate
Acetyl-CoA
S-Acetyldihydro-lipoamide-E
6.4.1.1
1.8.1.4
Dihydro-lipoamide-E
Lipoamide-E
1.1.1.37
2.3.3.1
2.3.3.8
Citrate
4.2.1.3
4.2.1.3
Isocitrate
Oxaloacetate
cis-Aconitate
(S)-Malate
1.1.5.4
1.1.1.42
Tyrosine metabolism
4.2.1.2
Oxalosuccinate
1.1.1.41
1.1.1.286
Arginine and proline metabolism
Fumarate
1.1.1.42
1.3.5.4
1.3.5.1
ThPP
6.2.1.4
Succinyl-CoA
2-Oxo-glutarate
6.2.1.5
2.3.1.61
1.2.4.2
1.2.4.2
Succinate
2.8.3.18
S-Succinyl-dihydrolipoamide-E
3-Carboxy-1-hydroxypropyl-ThPP
Ascorbate and aldarate metabolism
1.8.1.4
Dihydro-lipoamide-E
Lipoamide-E
Val, Leu & Ile degradation
Alanine, aspartate and glutamate metabolism
1.2.7.3
D-Gln & D-Glu metabolism
00020 9/25/14
(c) Kanehisa Laboratories


PENTOSE PHOSPHATE PATHWAY
Glycolysis
Galactose metabolism
Pentose and glucuronate interconversions
 Entner-Doudoroff pathway
Pyruvate
Glycerate
Glycerate-2P
Glycerate-3P
D-Glycer-aldehyde
D-Glycer-aldehyde-3P
2-Dehydro-3-deoxy-D-gluconate
2-Dehydro-3-deoxy-D-gluconate-6P
D-Glucosaminate
D-Gluconate
D-Glucosaminate-6P
D-Glucosaminate (extracellular)
2-Dehydro-D-gluconate
2-Dehydro-D-gluconate-6P
D-Gluconate-6P
D-Glucono-1,5-lactone
D-Glucono-1,5-lactone-6P
H6PD
D-arabino-Hex-3-ulose-6P
D-Xylulose-5P
D-Ribulose-5P
D-Ribose-5P
D-Ribose
D-Ribose-1P
D-Ribose-1,5P
PRPP
D-Xylulose
2-Deoxy-D-ribose
2-Deoxy-D-ribose-1P
2-Deoxy-D-ribose-5P
D-Erythrose-4P
D-Sedo-heptulose-7P
β-D-Glucose
D-Glucose
β-D-Glucose-6P
α-D-Glucose-6P
β-D-Fructose-6P
β-D-Fructose-1,6P2
D-Glyceraldehyde-3P
Purine metabolism
Pyrimidine metabolism
Histidine metabolism
00030 2/23/15
(c) Kanehisa Laboratories


PURINE METABOLISM
Pentose phosphate pathway
Alanine, aspartate and glutamate metabolism
Thiamine metabolism
Histidine metabolism
D-Ribose-1P
Ribose-5P
PRPP
L-Glutamine
Ribosylamine-5P
GAR
FGAR
FGAM
AIR
CAIR
AICAR
SAICAR
5-Amino-4-imidazolecarboxyamide
5-Carboxyamino-1-(5-phospho-D-ribosyl)imidazole
FAICAR
ADPribose
dIDP
dITP
dIMP
PDE
pppGpp
ppGpp
3',5'-cyclic GMP
RNA
GTP
GDP
GMP
Inosine 5'-tetraphosphate
ITP
IDP
5'-Butyrylphosphoinosine
Riboflavin metabolism
Folate biosynthesis
GppppG
dGTP
dGDP
DNA
Deoxyguanosine
dGMP
XMP
XpppX
XTP
Guanosine
Xanthosine
IMP
Inosine
Guanine
Xanthine
Hypoxanthine
2',3'-cyclic GMP
3'-GMP
Urea cycle
NH3
2',3'-cyclic AMP
3'-AMP
5'-Benzoylphosphoadenosine (mitochondria)
5'-Acetylphosphoadenosine (mitochondria)
Adenylo-succinate
AMP
Adenosine
Adenine
ATP
ADP
dADP
dATP
dAMP
Deoxyadenosine
Deoxyinosine
AppppA
ApppA
Adenylylsulfate
Sulfate
3'-Phospho-adenylate
3'-Phosphoadenylyl-sulfate
Adenosine 5'-tetraphosphate
3',5'-cyclic AMP
Glycine , serine and threonine metabolism
Glycine
Formimino-glycine
Imidazolone
Aminoimidazole carboxylate
5-Amino-4-imidazole carboxylate
5-Ureido-4-imidazole carboxylate
Urate-3-ribonucleoside
Urate
5-Hydroxy-2-oxo-4-ureido-2,5-dihydro-1H-imidazole-5-carboxylate
5-Hydroxyisourate
spontaneous
(S)-Allantoin
(R)-Allantoin
Allantoate
Ureidoglycine
Ureido-glycolate
Oxalureate
Carbamoyl phosphate
CO2
NH3
Urea
Glyoxylate
Glyoxylate and dicarboxylate metabolism
00230 3/4/15
(c) Kanehisa Laboratories


PYRIMIDINE METABOLISM
Pentose phosphate pathway
PRPP
Orotate
Orotidine-5P
UPpppU
UDP-glucose (extracellular)
UTP (extracellular)
UTP
UDP
UMP
3'-UMP
2',3'-Cyclic UMP
Uridine
Uracil
Pseudouridine-5'-phosphate
Pseudouridine
Dihydrouracil
3-Ureido-propionate
β-Alanine
β-Alanine metabolism
Barbiturate
3-Oxo-3-ureido-propanoate
Malonate
Urea
Ureidoacrylate
(Z)-3-Ureidoacrylate-peracid
(Z)-3-Peroxy-aminoacrylate
Carbamate
Aminoacrylate
Malonate-semialdehyde
3-Hydroxy-propionate
UTP
CTP (extracellular)
CTP
CMP
CDP
dCTP
dCDP
dCMP
2'-Deoxy-5-hydroxy-methylcytidine-5'-triphosphate
2'-Deoxy-5-hydroxy-methylcytidine-5'-diphosphate
5-Hydroxymethyl-deoxycytidylate
2'-CMP
2',3'-Cyclic CMP
Cytidine
Cytosine
Deoxycytidine
Deoxyuridine
dUTP
dUDP
dUMP
dTTP
dTDP
dTMP
Thymidine
Thymine
5-Methylcytosine
5-Methylbarbiturate
Methylmalonate
3-Ureidoisobutyrate
3-Amino-isobutanoate
Dihydro-thymine
Valine, leucine and isoleucine metabolism
2-Deoxy-D-ribose-1P
RNA
DNA
Thioredoxin
Thioredoxin disulfide
Thioredoxin reductase
L-Glutamine
Alanine, aspartate and glutamate metabolism
Arginine and proline metabolism
N-Carbamoyl-L-aspartate
Carbamoyl-P
Dihydro-orotate
RutC
RutD
RutE
RutF
YdfG
00240 10/20/14
(c) Kanehisa Laboratories