

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*.

Thomas Forth

Submitted in accordance with the requirements for the degree of

Integrated PhD with Masters

The University of Leeds

The White Rose Doctoral Training Centre within the School of Physics and Astronomy

September, 2012

The candidate confirms that the work submitted is his own, except where work which has formed part of jointly-authored publications has been included. The contribution of the candidate and the other authors to this work has been explicitly indicated below. The candidate confirms that appropriate credit has been given within the thesis where reference has been made to the work of others.

Chapter 1 : Introduction. This chapter shares much of its content with the Wikipedia page written by the candidate on flux-balance analysis. The candidate remains the principal author of the page but recognises that the contributions of others — which are fully described in the "history" section of the Wikipedia page — are significant.

Chapter 2: MetNetMaker. The software described in this chapter was published as Forth T, McConkey GA, Westhead DR. MetNetMaker: a free and open-source tool for the creation of novel metabolic networks in SBML format. *Bioinformatics (Oxford, England)*. 2010;26(18):2352–3.

This paper was written by the candidate with guidance and correction provided principally by D. R Westhead.

This copy has been supplied on the understanding that it is copyright material and that no quotation from the thesis may be published without proper acknowledgement.

© 2012 The University of Leeds and Thomas Forth.

Acknowledgements

I am extremely grateful to my supervisors Prof. David Westhead and Dr. Glenn McConkey for suggesting this line of study and for giving me the freedom to explore so widely and so independently the rapidly changing field of systems biology. They have always been available to me and provided critical guidance, suggestions, and contacts at key moments.

Fewer and fewer doctoral students in the UK can expect to continue to academic careers and the balance of priorities in doctoral study has thus shifted away from purely academic achievement towards the broader acquisition skills, connections and experience for a rapidly changing and globally competitive future. I am well aware that this could present a conflict between the interests of a doctoral supervisor and the long-term interests of a doctoral student and I hugely appreciate that both of my supervisors and the White Rose doctoral training centre have always acted in my best interest.

Some of the most significant findings in this thesis come from a collaboration I developed with Prof. Julie Fisher's NMR metabolomics lab at the University of Leeds and I am greatly indebted to Prof. Fisher, Dr. Hayley Fenton-Saville and especially Dr. Cassey McRae for the time they spent sharing their experience and helping me develop the techniques described in this thesis.

Emese O'Donnell (Prandovszky) and Paul Bedingfield have been extremely generous with their time and expertise in the wet-lab and I could not have completed that section of this thesis without their continual support. I am similarly grateful for the hard-work of masters students I have worked with, particularly Jennifer Lake who worked exceptionally hard in the culture room and scoured early NMR spectra for the tiny peaks that led to a more serious metabolomics study of the malaria parasite.

Lastly I am grateful to the family, friends and colleagues who have provided inspiration, support and encouragement to continue with my work, especially at extremely difficult times in my first and in my most recent year at Leeds. That some of the greatest sources of my inspiration are no longer alive has not diminished the determination to succeed that they have passed on to me.

This research has been carried out by a team which has included Masters Students Cheng Ma, Jennifer Lake and Sara Zakutansky. My own contributions form the majority of the research and work performed in collaboration with others is explicitly and fully indicated in the thesis.

Abstract

Quantitative one-dimensional proton NMR metabolomics is performed on growth medium samples gathered at up to ten time-points during the *in vitro* culture of *P. falciparum* in human red blood cells. From this study, exchange fluxes between the parasite-host complex and the growth medium are calculated for glucose, lactate, glycerol, glutamine, hypoxanthine, valine, leucine, isoleucine, alanine, tyrosine and phenylalanine. Carbon-source exchange fluxes are added as constraints to a new model of malaria metabolism — built using my published MetNetMaker software — consisting of 249 reactions, 143 genes and a novel experimentally derived biomass function.

Analysis of this network including by flux-balance analysis and flux-variability analysis are projected onto a live map of the network providing the most accessible view of malaria metabolism to date. This model reproduces key phenotypes of the malaria parasite such as the unusual branched TCA cycle, and accurately predicts internal fluxes through the pentose-phosphate cycle and the low oxygen-dependence of the parasite's metabolism during its erythrocytic life stages. The model is carbon balanced and accurately predicts the parasite's growth-rate at measured glucose uptake rates. Furthermore, it accurately reproduces measured amino acid and purine-source exchange fluxes at the optimal solution and implies that the parasite digests 30% of its red blood cell host's haemoglobin but incorporates just 40% of the resulting freed amino acids into its proteome. Lethal single and double gene deletions are predicted and suggest potential drug and vaccine targets.

The metabolic model is available in MetNetMaker format for easy editing, SBML format including constraints for metabolic modelling and the independent reproduction of the reported results, and cytoscape format with metadata for visualisation of both the network and the results of simulations performed on it.

Contents

List of acronyms	xviii
Introduction	1
Metabolic networks	1
The malaria parasite.....	23
Network analysis	31
MetNetMaker	46
Building MetNetMaker	50
<i>Application structure</i>	51
<i>User interface</i>	61
Application outputs for further analysis.....	68
Visualisation and Network Reconstruction	71
KEGG projector and Cytoscape: one problem, two approaches	73
Four elements of my <i>P. falciparum</i> reconstruction	76
Example reconstruction: nucleotide metabolism.....	82
The size of the task.....	87
The final model	89
Encouraging model re-use.....	95
Experimental Methods	99
Malaria culture.....	99
<i>Observing cultures</i>	102
<i>Synchronisation of cultures</i>	106
<i>Saponin lysis and parasite biomass isolation</i>	111
<i>Problems working with large culture volumes</i>	112
Measurements techniques.....	113
<i>Used growth medium assays</i>	113
<i>Biomass analysis</i>	117
Accurate determination of <i>P. falciparum</i> growth-rate	120
Measuring <i>P. falciparum</i> growth-rate <i>in vitro</i> using blood smears	121
Development of flow cytometry	125
Single-channel flow cytometry.....	128

Dual-channel flow cytometry	132
Summary of flow cytometry	137
Experimental results.....	140
Towards a biomass function for malaria.....	140
<i>Protein content</i>	146
<i>DNA and RNA content</i>	149
<i>Best guesses and sanity checks</i>	150
<i>Details breakdown</i>	153
Measuring exchange fluxes.....	156
<i>Carbon balance</i>	181
<i>The absence of malate</i>	183
<i>Sanity checks and discussion of results</i>	186
<i>False starts and dead ends</i>	195
Metabolic modelling	199
Units of biomass and exchange fluxes	200
Defining the final biomass function.....	202
Applying constraints.....	203
ATP maintenance function	206
Results.....	209
Conclusions.....	221
<i>Final comparison of my final model to similar models</i>	225
Summary and future directions	227
References.....	234
Appendices.....	241
Appendix I : Dead-end compounds algorithm and implementation.....	242
Appendix II : Example reconstruction and evidence checker spreadsheet	246
Appendix III : RPMI components	248
Appendix IV : Free components of the biomass function	249
Appendix V : Measured exchange fluxes	250
Appendix VI: Key amino acid compositions.....	252
Appendix VII : Single gene deletions that affect growth-rate.....	254
Appendix VIII : Application for Synthetic Aesthetics funding	260
Appendix IX : Application for Gates Foundation funding.....	263

List of figures

Figure 1: Simplified depiction of the central dogma of molecular biochemistry (Daniel Horspool – Wikimedia Commons under license: CC BY-SA 3.0).....	4
Figure 2: KEGG reaction R04479, catalysed by enzyme 2.7.1.11, typically proceeds in the direction shown in part a . It can be considered a combination of two reactions, b and c with opposite energetically favoured directions.....	6
Figure 3: Visual representations of the Glycolysis I Pathway in <i>E. coli</i> (left) and Human (right) in Biocyc 15.1.....	8
Figure 4: The KEGG LIGAND compound entry for ATP, shortened from 3 pages. The simplicity of the ontology appeals to me.....	12
Figure 5: The MetaCyc compound entry for ATP, shortened from 17 pages. The extensive linking within the ontology has many advantages but is intimidating to work with.....	13
Figure 6: The tree structure of the EC number definition for 6-phosphofructokinase and the 10 KEGG reactions related to 2.7.1.11.....	15
Figure 7: The LIGAND reaction R01600 labelled with the key components required to define a reaction within a reconstruction framework. This is a still frame from the “Framework Video” on my project website ¹⁴	16
Figure 8: The key components required to define a reaction within a reconstruction framework. This is a still frame from the “Framework Video” on my project website ¹⁴	16
Figure 9: (top/ a) Reaction-centric graph. (left/ b) Compound-centric graph. (right/ c) Combined graph. These are all equivalent representations of the first six reactions in the glycolysis pathway. ATP and ADP nodes are aliased.	19
Figure 10: (left) A visual representation of a small reaction network. (right) The equivalent 8×6 stoichiometric matrix (S matrix) that completely defines the network. Reaction IDs and compound IDs and names are added as labels to the matrix.....	21
Figure 11 : The multiple life stages and host cells of Plasmodium. Parasite nuclei are shown in dark purple, cytosol as a lighter purple, as under Giemsa stain. Partially adapted from Dr. Wisner’s malaria course at Tulane University, FL, USA ³⁷	27
Figure 12 : The life-cycle red blood cell stage Plasmodium parasites with representative images of Giemsa stained infected red blood cells on thin blood smears. All images are my own.....	28
Figure 13: Cellular overview of <i>P. falciparum</i> network used for chokepoint analysis from Fig 2 in Yeh et al ⁷ . Reactions (edges) shown in grey are not present in the model.	34
Figure 14: a . A metabolic network on its own cannot simulate life. b . A metabolic network that is able to exchange metabolites with a growth medium and provide molecules for growth as part of a biomass function can.	35
Figure 15: The first six reactions in glycolysis prepared for FBA through the addition of an objective function (red) and the import and export of nutrients (ATP, ADP, BDG, ADG) across the system boundary (dashed green line). The full version shows the stoichiometric matrix both before and after the model is prepared for FBA.	36
Figure 16 : A metabolic network before (left) and after (right) flux-balance analysis. The ‘after’ picture conveys the optimal flux from the nutrient source to the biomass function as the weight of the lines representing each reaction. These weights correspond to the reaction flux vector, v , shown below in its transposed orientation v' . A constraint on the glucose uptake flux of 50 arbitrary units is shown as a pair of red bars constraining flux through these reactions.	37

Figure 17 : The seven elementary modes of the example system. Solutions 5,6 and 7 are equally valid in both directions.....	39
Figure 18 : Visual representation, with FBA and FVA solution vectors below, of flux-variability analysis (FVA) performed on the top six reactions of glycolysis. R03321 and R02740 were only allowed to run forwards to avoid an internal cycle. Constraint on R04779 of 50 units as shown places an upper limit on biomass production. Variability of fluxes (in light grey) shown at four times width with chevrons noting possible direction of flow. (left) Full network. (right) With non-lethal deletion of R02740.....	42
Figure 19: Representation of key parts of the application structure of MetNetMaker as published in Forth et al. ⁸³	51
Figure 20: The Entity Relation Diagram of the central database in MetNetMaker is central to implemented a rigorous reconstruction framework.....	53
Figure 21: The Entity Relation Diagram of the central database in MetNetMaker as represented within Microsoft Access.....	54
Figure 22: (top) Colour coded entry from the LIGAND database REACTION file for R00014. (bottom) The tables produced from parsing this entry and then used to populate tables in MetNetMaker's central database.....	58
Figure 23: (top) Colour-coded sample entry from the LIGAND database COMPOUND file for C00022. (bottom) The tables produced from parsing this entry and then used to populate tables in MetNetMaker's central database.....	59
Figure 24: Table produced during parsing to test whether reactions are mass balanced.....	60
Figure 25: Detachable window for the selected reactions list within MetNetMaker 1.5.2.....	63
Figure 26: Dead-end compounds report for an example set of reactions in the selected reactions table.....	63
Figure 27: Reaction Picker window within MetNetMaker 1.5.2. Reaction R00014 is currently highlighted. The ribbon UI at the top is always present.....	64
Figure 28: Reaction R00014 being edited within the Reaction Creator window of MetNetMaker 1.5.2. The ribbon UI at the top is always present.....	67
Figure 29: The cytoscape output form in MetNetMaker 1.5.2.....	69
Figure 30: Representation of key parts of the application structure of MetNetMaker 1.5.2 as currently available from www.metnetmaker.com and an overview of associated tools for further analysis of networks.....	70
Figure 31: <i>P. falciparum</i> metabolic model from Huthmacher et al. ⁵⁹ as visualised in SharkView ⁹⁰ with no compound aliasing. Squares represent reactions and circles represent compounds. The outline colour of each compound denotes the compartment it is found in. Isolated reactions are not connected to the network.....	72
Figure 32: <i>P. falciparum</i> metabolic model from Plata et al. ⁵⁸ as visualised in cytoscape ⁸⁶ with no compound aliasing. Red lines show compounds entering a reaction, green lines show compounds leaving a reaction. Compounds and reactions are arranged around the perimeter of the circle. Isolated reactions at the top-left are not connected to the network.....	72
Figure 33 : The simplified map of carbon metabolism from Olszewski et al. ⁸⁸ shows the unusual role of the mitochondrial TCA cycle in <i>P. falciparum</i> metabolism and the unusual role of the apicoplast in lipid synthesis and manipulation. Some of the visual simplifications are highly misleading, for example the reaction $\text{Glu} \Rightarrow \text{2OG}$ (glutamate \Rightarrow oxoglutarate) and the reaction $\text{OAA} \Leftrightarrow \text{Asp}$ (oxaloacetate \Leftrightarrow aspartate) are actually the same reaction and their linked behaviour radically changes the structure of the network. My final model on page 94 (figure 47) retains some of the design features of this model without making this important oversimplification.....	73
Figure 34 : The visualisation section of the ribbon in MetNetMaker v1.6+ includes new options for visualising reactions in the selected reactions table. These methods rely either on my KEGG projector tool (the two buttons on the left) or Cytoscape for visualisation.....	73

Figure 35 : (left) depiction of glycolysis from MPMP ⁸ (right) equivalent diagram within Cytoscape, created from a corrected version of the MPMP network using the ‘Visualise in Cytoscape’ button in MetNetMaker.....	76
Figure 36: R02086 as shown in MetNetMaker v1.0. This reaction is no longer present in current versions of LIGAND. No current LIGAND reaction involves C04256 but the compound remains in the ontology.	78
Figure 37: R08193 as shown in MetNetMaker v1.5+. This reaction replaces R02086 but specifies the product as C04501 in place of C04256. The two compounds are isomers.....	78
Figure 38: Entity relationship diagram for the ECinfo database as presented in Microsoft Access’s Relationships view.....	81
Figure 39: (left) The heavily shaded edges represent the 49 reactions of the nucleotide metabolism section of MPMP as projected onto the KEGG global map in KEGG projector (right) My reconstruction contains 17 more reactions but this is represented as only an additional 4 heavily shaded edges on the KEGG global map.....	83
Figure 40 : The report from MetNetMaker on the atom counts for R00575 showing that the reaction is unbalanced. R00575_balanced replaces it in my final model.....	85
Figure 41 : The chemical structure of KEGG compound C06198, also called P1,P4-Bis(5'-uridylyl) tetraphosphate or UppppU.	86
Figure 42: Projection onto the KEGGML pathway map ¹¹⁸ of all 1415 reactions with any evidence for existence in <i>P. falciparum</i> in my ECinfo database.....	88
Figure 43: Projection onto the KGML pathway map ¹¹⁸ of all 363 reactions related to an EC number with a consensus of evidence for existence in <i>P. falciparum</i> . Nucleotide metabolism (dark pink, top-right) appears more populated than in figure 39 because EC numbers associated with multiple reactions are projected onto multiple edges.	88
Figure 44 : Characteristics of the “full” metabolic network before reduction to form the final model.	91
Figure 45: Characteristics of the reduced “final” model used for flux-balance analysis.	92
Figure 46: Visual version of the final model ready for flux projection, as in figure 47.	94
Figure 47 : Final model with live flux projections of the optimal flux distribution given measured constraints. Annotations in red show key pathways and features of the network.	94
Figure 48 : The portion of the SBML export produced by MetNetMaker for the top reaction in table 33 (appendix II), R00156, showing the mechanism for encoding additional information in SBML level 2 format within the <notes> tag for COBRA compatibility.....	97
Figure 49: Thin blood smears are made from a single drop of blood pushed quickly across a glass slide. The red blood cells at the edge of the smear are one layer thick.	102
Figure 50: Categories used to define the life stages of <i>P. falciparum</i> in the microscopy time course. 1 : normal red cell. 2-10 : early ring trophozoites. 11-18 : late trophozoites. 19-26 : schizonts. 27-30 : gametocytes (not observed in our system of continuous culture). Image: Plate XLII, The primate malarias, 1971, G . R. Coatney ¹³⁵ . Similar illustrations by G. H. Nicholson for 26 other plasmodium species available from Images from the History of Medicine ¹³⁶	104
Figure 51: left: Rod-shaped bacteria are clearly visible in a Giemsa stained blood smear. right: A healthy ring-stage parasite in an infected culture shows that minor infections are not always fatal.	105
Figure 52: a) A second synchronisation performed 36 hours after the first leads to a tighter final synchronisation profile as shown in (b)	108
Figure 53: Progression of life stages for seven identical cultures in AlbuMAX I 100% growth medium synchronised at $t = 0$ hours and $t = -46$ hours. Lines modelling life stage progression are the convolution of figure 54a and figure 54b.....	110

Figure 54: a) Mathematical representation of a 14-hour synchronisation pulse killing 90% of the targeted life-stages. b) 48 hours of the idealised life-stage progression that best fits results in figure 53. This figure would continue infinitely in both directions.	110
Figure 55: An example of a glass stoppered NMR tube with no interface between the sample and the air as used in the FBS NMR service machines.	115
Figure 56 : Growth-rate of a population of <i>P. falciparum</i> grown in 100% AlbuMAX I media for two weeks as measured using Giemsa's stain on thin blood smears under an optical microscope. Here the natural logarithm of the parasitemia is plotted so that the exponential growth curve becomes linear. Shaded areas show the error in a least squares linear fit of the data after 4 (dark), 7 (lighter) and 10 (lightest) days.	123
Figure 57: Fluorescence microscope image of RBCs stained with Acridine Orange at 4.0 µg/ml . Green spots are areas of peak fluorescence under white light and are labelled as described. The fluorescent layer and the background layer were aligned in Photoshop due to a problem with the microscope optics.	127
Figure 58 : Comparison of my results and published results for single-channel flow cytometry. a. Sample result from Bhakdi et al. ¹⁵⁶ with M1 gate denoting uninfected cells and M2 denoting infected cells. Parasitemia ≈30-50%. b. My equivalent result for a sample with a known parasitemia of 4.6%. Gate R1 denotes infected cells.	128
Figure 59: FACS-determined parasitemia versus parasitemia determined by microscope inspection of Giemsa-stained blood smears. Removal of counts from un-dyed control samples means some FACS-determined parasitemias are below 0%.	129
Figure 60: Comparison of raw single-channel output from BD FACSCalibur acquisition of a control sample (0% infected RBCs in media dyed with 1.25 µg/ml of acridine orange) versus a sample with a known parasitemia of 0.18%. Overlapping section is darker. R1 shows gating optimal for control which is clearly not optimal for 0.18% case.	130
Figure 61 : Parasitemia measurements by single-channel flow cytometry using global and constant gating versus measurements by optical microscopy of Giemsa-stained blood smears. Samples were identical and prepared as described in the methods chapter and diluted with either 1.5ml or 2.0ml of PBS before acquisition. The point in red belongs to the 1.5ml dataset but is not included in the best fit line.	131
Figure 62: Parasitemia by single-channel flow cytometry using custom and variable gating versus optical microscopy of Giemsa-stained blood smears.	131
Figure 63 : Dual-channel flow cytometry (red and green) for the same sample as shown in figure 58b. The gated area shown is a naïve attempt to select infected cells. Samples were dyed with acridine orange to a final concentration of 1.25 µg/ml, left for at least 3 minutes, then diluted with 1.5ml of PBS before acquisition.	133
Figure 64 : Dual-channel output of a sample of known parasitemia of 3.5% showing the two distinct zones where it is suggested that fluorescence is the result of parasite infection or auto-fluorescence of an uninfected RBC. The zoning technique seems hard to justify as discussed in the thesis body.	134
Figure 65 : A. Stained uninfected cells from Bhakdi et al. (known parasitemia =0.0%, observed parasitemia = 0.09%) B. Stained mouse RBCs infected with <i>P. berghei</i> from Bhakdi et al. ¹⁵⁶ (known parasitemia ≈ observed parasitemia = 1.79%) C. Stained uninfected cells from my experiments. (known parasitemia =0.0%, observed parasitemia = 0.1%) D. Stained infected cells from my experiments. (known parasitemia = 3.5%, observed parasitemia = 2.0%)	135
Figure 66 : Parasitemia by dual-channel flow cytometry as demonstrated in figure 64 versus optical microscopy of Giemsa-stained blood smears. Samples prepared as in the methods chapter. Results are extremely variable across different runs.	136

Figure 67 : Dual-channel output of a sample of known parasitemia of 3.3% dyed with 0.25 ug/ml acridine orange. The infected fluorescence zone contains 0.78% of the cells counted.	137
Figure 68 : Serial dilutions of BSA standard and samples 3b (from 5mg/ml of pure biomass) and 4 (from 11mg/ml of pure biomass) in Bradford reagent. Note the bubbles in the second row for sample 3b; this data-point was excluded from analysis.	146
Figure 69 : Determination of protein concentration for sample 4 (see table 12) by the Bradford assay. There are two points at each concentration for each sample corresponding to acquisition at 15 minutes and 90 minutes.	147
Figure 70 : The components of the final biomass function as a percentage of dry weight.	155
Figure 71 : ¹ H-NMR spectra for the α -Glucose doublet at 5.25ppm from the Sara A set. (left) Positions as referenced to TMS at 0ppm. (right) Positions once manually aligned. Colours blue, green, grey, pink, red, correspond to t = 0, 9.2 18.5, 27.0, 36.5 hours and show the clear decrease in glucose concentration in the growth medium over time.	158
Figure 72 : The four ¹ H-NMR spectra in the Sara A set. Here the integrals for the selected spectrum (red, 30/06_00h20_A, t = 9.2 hours) are shown. The two symmetrical HEPES multiplets at 2.86 and 3.15 (integral referenced to 100) are visible. Colours green, grey, pink, red, correspond to t = 9.2 18.5, 27.0, 36.5 hours.	159
Figure 73 : Collection of ¹ H-NMR reference spectra used as a guide to peak identification displayed separately. From the top, spectra are of NADP, NAD, Leucine, Lactate, L-Valine, L-Glutamine, L-Glutamate, L-Alanine, Isoleucine, Hypoxanthine, Histidine, Glycerol, Glucose, Ehtanol, Isoleucine (included twice in error), Leucine (included twice in error), Malate.	161
Figure 74 : Collection of ¹ H-NMR reference spectra used as a guide to peak identification displayed together. Other reference spectra were downloaded and examined individually but determined not to be present in any spectra.	161
Figure 75 : HPLC measurements of dihydroorotate and orotate concentrations in the parasite biomass pellets extracted at the end of the Sara metbolomics set of experiments suggest that sub-lethal atovaquone inhibition (sample A) was successful in reducing the flux through R01867 leading to accumulation before, and reduced concentration after, the reaction.	168
Figure 76 : The concentration of lactate in growth medium over time. The starred red point was not used during numerical analysis and does not suggest reabsorption of lactate from the growth medium.	170
Figure 77 : The concentration of glucose ($\alpha + \beta$) in growth medium over time. The starred red point was not used during numerical analysis. Points at t =0 are only reported for blood nulls as discussed later.	171
Figure 78 : The concentration of glycerol in growth medium over time. The starred red point was not used during numerical analysis and does not suggest reabsorption of glycerol from the growth medium.	171
Figure 79 : The concentration of glutamine in growth medium over time. Points at t =0 are only reported for blood nulls as discussed later.	172
Figure 80 : Exchange fluxes of carbon compounds. Differences between the glycerol excretion and glutamine absorption rates in Sara A and Sara B sets are not significant. No glycerol was measured in the "Tom Repeat of Jenny" experiment.	174
Figure 81 : Amino acid sources and fates in a cultured infected RBC are more complex than with carbon sources due to the abundant proteins provided by haemoglobin digestion. It is unclear whether the free amino acid pool in the RBC is large enough to affect measured amino acid exchange fluxes between the parasite and growth medium significantly.	175
Figure 82 : The concentration of valine in the growth medium over time.	176
Figure 83 : The concentration of alanine in the growth medium over time.	177
Figure 84 : The concentration of leucine in the growth medium over time.	177

Figure 85 : The concentration of isoleucine in the growth medium over time.	178
Figure 86 : The concentration of tyrosine in the growth medium over time.	178
Figure 87 : The concentration of phenylalanine in the growth medium over time.	179
Figure 88 : The concentration of hypoxanthine in the growth medium over time.	179
Figure 89 : Amino acid and hypoxanthine exchange fluxes from infected RBCs. Positive values indicate the metabolite is released into the growth medium, negative values indicate that it is absorbed.	181
Figure 90 : The three peaks of the experimental ¹ H-NMR spectrum of malate from HMDB ¹⁷⁹ . There are no significant peaks outside of this range. The peak to the right is TSP at a chemical shift of zero.	184
Figure 91 : Exchange fluxes of the three metabolites potentially showing stage-specificity at the ring stage (early) and the trophozoite/schizont stages (late). The valine exchange flux and associated uncertainty is shown at 100 times real scale to be visible.	188
Figure 92 : Glycerol (red) and glucose (blue) peaks overlap slightly at 3.50ppm affecting the otherwise clear glycerol peak assigned to 3.56ppm. Spectra are from HMDB references and manually aligned with the TMS peak at 0ppm.	189
Figure 93 : Ethanol and isoleucine are both strong peaks that obscure the glycerol peak at 3.66ppm. Spectra are from HMDB references and manually aligned with the TMS peak at 0ppm.	189
Figure 94 : Spectra for the "Sara A" set annotated to show the set of peaks I assign to glycerol. Colours blue, green, pink, red, corresponds to t = 9.2 18.5, 27.0, 36.5 hours. Glucose concentration at the final time point is almost zero meaning the glycerol peak of the double doublet becomes visible. Due to the normalisation techniques used in ACD/labs the visual size of each peak is only a guide to its integral.	190
Figure 95 : Stacked bar charts showing the fluxes of amino acids to (negative) and from (positive) the pool of free amino acids (shown in figure 81) for the "Tom Repeat of Jenny" experiment. The blue bars represent the quantity of each amino acid freed from haemoglobin digestion (assuming 40% of the host cell's haemoglobin is digested). Red bars represent the quantity of each amino acid used for protein synthesis by the parasite. Green bars represent the measured fluxes of amino acids out of the parasite (into the parasite in the case of glutamine). The total mass of amino acids is conserved in this figure.	192
Figure 96 : Stacked bar charts showing the fluxes of amino acids to (negative) and from (positive) the pool of free amino acids (shown in figure 81) for the "Tom Repeat of Jenny" experiment. The blue bars represent the quantity of each amino acid freed from haemoglobin digestion (assuming 55% of the host cell's haemoglobin is digested). Red bars represent the quantity of each amino acid used for protein synthesis by the parasite. Green bars represent the measured fluxes of amino acids out of the parasite (into the parasite in the case of glutamine). The total mass of tyrosine and phenylalanine is conserved in this figure.	193
Figure 97: FBA-predicted fluxes through my early 2010 model of glycolysis, the pentose phosphate cycle and the TCA cycle using minimum and maximum constraints for glucose, lactate, glycerol and ethanol fluxes. Low energy small molecules are provided free. The objective function consumes high energy molecules or precursors (ATP, NADH, NADPH etc...) and is maximised up to a top limit. (left) with no flux allowed through additional ethanol generating reactions. (right) with flux allowed through additional ethanol generating reactions.	196
Figure 98 : Final annotated model with live flux projections of the optimal flux distribution given the applied experimentally derived constraints. Annotations in red show key pathways and features of the network. Flux weights are not linear and greatly exaggerate smaller fluxes. For example the three exchange fluxes at the top-left, α-D-glucose β-D-glucose and glutamine import have fluxes of 19.24, 19.24 and 0.98 respectively but appear much closer in width. Calculated using SurreyFBA ⁸⁷	199

Figure 99 : A visual representation of the constraints applied to the final model during flux-balance analysis. The biomass function is not shown as it is the optimisation goal and not constrained in the final model. Green = direction constrained, Red = experimentally measured or literature reported flux constraint applied, Blue = flux set to zero, Orange = flux limited to a reasonable rate that could not be measured (arginine, proline, glutamate exchange).....	204
Figure 100 : The effect of an increasing ATP maintenance flux on the rate of biomass formation. Calculated using SurreyFBA ⁸⁷	206
Figure 101 : The growth-rate predicted by FBA on the final model for a range of haemoglobin digestion fluxes. Calculated using SurreyFBA ⁸⁷ . (left) The formation of haemozoin from haem incurs no cost and the optimal solution is for the parasite to digest ever more haemoglobin. A reduction in the maximum allowed efflux of proline from the system places an upper limit on the haemoglobin digestion flux. (right) The addition of a 10 ATP cost for each molecule of haem incorporated into haemozoin moves the optimal haemoglobin digestion flux lower but still allows a large range of haemoglobin digestion fluxes that give a biomass production rate close to optimal.....	208
Figure 102 : Biomass production response to glucose uptake. The shaded region shows the measured glucose influx and its associated error (38.48 ± 14 mmole/gDW/hour). The lactate efflux of -60 mmole/gDW/hour is slightly higher than the best measured value (56) but well within the error on that measurement. Calculated using SurreyFBA ⁸⁷ ..	210
Figure 103 : Biomass production response to lactate efflux for three glucose uptake rates corresponding to the best-guess (38 ± 14 mmole/gDW/hour), minus and plus the associated error (24, 56). The shaded region shows the final lactate efflux and its associated error (60 ± 11 mmole/gDW/hour). Calculated using SurreyFBA ⁸⁷	210
Figure 104 : Measured flux of amino acids and hypoxanthine against FBA-predicted fluxes at the optimal FBA solution of my final model. Best fit line is forced through the origin.	213
Figure 105 : The relationship between the haemoglobin digestion flux and the percentage of the host's haemoglobin that is digested (blue line) and the percentage of the freed amino acids that are integrated into the parasite's proteins (red line).....	214
Figure 106 : The variability of fluxes that still give 100% of the optimal growth-rate. The very thick lines are three cycles that are futile and can have any value as they ultimately produce and consume nothing. The three exchange reactions involved in a futile cycle are H ₂ O, Orthophosphate and Diphosphate which combine with an internal reaction that converts Diphosphate to 2 Orthophosphate + H ₂ O to form a futile cycle. Aspartate, proline and arginine exchange fluxes are all variable within the optimal solution space and aspartate and proline exchange fluxes are constrained to a sensible maximum value in the final model. Calculated using the COBRA toolbox v1.3.3 ²⁰³	215
Figure 107 : The variability of fluxes that still give 99% of the optimal growth-rate. The flux through the mitochondrion via succinate becomes variable as does haemoglobin digestion and the pentose-phosphate cycle. Calculated using the COBRA toolbox v1.3.3 ²⁰³	216
Figure 108 : The variability of fluxes that still give 70% of the optimal growth-rate. Calculated using the COBRA toolbox v1.3.3 ²⁰³	217
Figure 109 : The location of essential reactions in the network. Calculated using SurreyFBA ⁸⁷	218
Figure 110 : Predicted growth-rate as a proportion of optimal for all pairwise deletions of fifteen genes that give interesting growth-rate predictions. The metabolic role of each gene within the network is shown at the right Growth-rate predictions made using COBRA toolbox 1.3.3 ²⁰³	219
Figure 111 : The succinate dehydrogenase (SDH) complex (left) and the mitochondrial glycerol-3-phosphate dehydrogenase (GPDH-M) complex (right) cannot share FAD/FADH ₂ since neither ever leaves the enzyme complex. Both enzyme complexes restore their FAD/FADH ₂ balance via external Coenzyme Q10 (Q/QH ₂) in the mitochondrial membrane space. Both images from Wikipedia (user: Johnhfst. user: Boghog2).	223

Figure 112 : (left) glycolysis, oxidative pentose-phosphate pathway, purine base synthesis and fatty acid elongation in the apicoplast in a simplified version of the complete final model. Haemoglobin digestion to scavenge amino acids is shown separately in the bottom-right. (right) the same model in SIMALARIA, glycolysis in orange, ribose sugar synthesis via the oxidative pentose phosphate cycle in green, nucleotide base synthesis via the purine synthesis pathway in light blue, fatty acid elongation in pink. Haemoglobin digestion requires no glucose flux but consumes energy in the game.	233
Figure 113 : The six queries composing the SQL portion of the dead-end compounds algorithm. All queries refer either to each other or to tables in MetNetMaker's central database shown in figure 20. Nested SQL queries break the algorithm down into distinct levels as shown in figure 116.	242
Figure 114 : Simplified version of the VBA code to run the SQL queries in figure 113 and output the results to a text file which is then opened with Notepad.	243
Figure 115 : Full version of the VBA code to run the SQL queries in figure 113 and output the results to a text file which is then opened with Notepad and displays the results as an on-screen notification.....	244
Figure 116 : Visual representation of the algorithm for returning dead-end compounds as represented as a set of SQL queries in figure 113.....	245

List of tables

Table 1: The purpose of text files produced by parsing the LIGAND REACTION file.....	56
Table 2: The purpose of text files produced by parsing the LIGAND COMPOUND and GLYCAN files.	56
Table 3 : Issues encountered during the digitisation of MPMP's depiction of pyrimidine metabolism.	77
Table 4 : The 29 consensus reactions associated with pyrimidine metabolism (KEGG pathway 00240) from the ECinfo database.	84
Table 5: Age of parasites since infection at which I have classified them at each distinct stage.	109
Table 6 : Haematocrit correction table for the "Sara metabolomics" set of experiments. 1ml of growth medium was removed every 3 hours during the experiment.	114
Table 7: The number of parasites counted in ten fields of seven slides each made from the same sample. The standard deviation rather than the standard error is calculated because during the experiment only a single slide ($n = 1$) is counted and $\sigma = \sigma/n$	122
Table 8 : Best-fit growth-rates for the three time periods depicted in figure 56.....	123
Table 9 : Counts of the different types of fluorescent spots in figure 57.	126
Table 10 : Summary of the gradient and intercepts of best-fit lines for points in figure 61 and figure 62.	132
Table 11 : Summary of major growth-rate and biomass experiments conducted in 2008 and 2009. The growth-rate experiments used to produce figure 56 on page 123 were performed up to 22/12/2008 and are not included in this summary chart but are present in my lab book.	142
Table 12 : Extraction methods and biomass weights of the four samples labelled "1/4" isolated at the end of the Tom Biomass 3 experiment.	145
Table 13 : Protein concentration of biomass samples extracted using the techniques summarised in table 12 Because of the bubbles visible in the second row of the 3b sample, see Figure 68, this point was omitted from analysis. Extractions by the methods used for samples 1 and 2 were clearly unsuccessful.	148
Table 14 : DNA concentration of biomass samples. Extractions by the methods used for samples 1 and 2 were clearly unsuccessful.....	149
Table 15 : RNA concentration of biomass samples. Extractions by the methods used for samples 1 and 2 were clearly unsuccessful.....	149
Table 16 : Amino acid frequency in the Plasmodium falciparum proteome, from Table 1 in Chanda et al. ¹⁷⁶ . Frequencies sum to 99% due to rounding in the original paper.	153
Table 17 : DNA nucleotide frequency in the P. falciparum genome.....	154
Table 18 : RNA nucleotide frequencies are calculated by taking into account both the ribosomal RNA nucleotide composition and the likely composition of mRNA from gene transcription. The rRNA:mRNA ratio is set at 1:1.	154
Table 19 : The components of the final biomass function as a percentage of dry weight.	155
Table 20 : Summary of metabolomics and biomass experiments performed in 2010 and 2011.	157
Table 21 : NMR peak assignment table ordered by observed peak position (ppm). The two unobserved malate peaks and the HEPES peak used as a concentration reference are included in addition to the 21 metabolite peaks representing 12 unique metabolites that were observed in at least some spectra. The hypoxanthine assignment is not very clear and should be treated with caution. In all cases the Madison Metabolomics Consortium Database ¹⁸⁰ was used as the definitive source of hydrogen counts for each peak.	162

Table 22 : Chemical shifts and integrals of the four HEPES peaks in my samples. Integrals are referenced to the clearest and most reliable peak at 3.15ppm.....	164
Table 23 : Mean and standard deviation for eight different metabolite concentrations for the five different $t = 0$ fresh medium samples across all experiments.....	169
Table 24 : Average standard deviation as a percentage of metabolite concentration for the six metabolites with multiple measurable peaks. *Lactate measurements at $t = 0$ were ignored as the standard deviation of the two measurements at 4.12(q) and 1.33(d) was extremely high as a percentage of the mean concentration at $t = 0$	169
Table 25 : Ratios of glucose uptake and lactate excretion ($10 - 10mg/ml$) from uninfected RBC and infected RBC populations in the three experiments with blood nulls. Results from the "Tom Repeat of Jenny" experiment should be treated with caution as the blood null line of best fit was calculated from only two points.	182
Table 26 : Concentration of six metabolites in RPMI 1640 (manufacturer's definitions) and the equivalent measurements for fresh growth medium ($t=0$) by 1H -NMR in Massimi et al. ¹⁸⁵ , in my NMR results and for glucose as reported in Jennifer Lake's Masters dissertation ¹⁴⁵	186
Table 27: Amino acid composition of 2 α -subunits + 2 β -subunits of the human haemoglobin molecule ordered by total weight per mole of haemoglobin for selected amino acids of interest, The full table is reproduced in Appendix VI as table 43.....	191
Table 28 : Amino acid use in the <i>P. falciparum</i> proteome, adapted from Chanda et al. ¹⁷⁶ for selected amino acids of interest, The full table is reproduced in Appendix VI as table 44.	191
Table 29 : A summary of the constraints applied to the reactions within the final model. The biomass is constrained in a forward direction until simulations are performed at which point it becomes the optimisation goal and is no longer constrained.....	204
Table 30 : Fluxes in mmole/gDW/hour of glucose towards the pentose phosphate cycle and nucleotide sugar formation. Figures from Atamna et al. ¹⁶⁷ are converted to equivalent units for comparison. Calculated using SurreyFBA ⁸⁷	211
Table 31 : Measured exchange fluxes of the six measured amino acids and hypoxanthine compared with the fluxes predicted by the optimal solution of my final model. Calculated using SurreyFBA ⁸⁷	213
Table 32 : Details on the 15 interesting double-deletion genes whose pair-wise deletions are lethal or growth-reducing (see figure 110) but whose deletions singly are not lethal.	220
Table 33 : Reactions for pyrimidine metabolism in <i>P. falciparum</i> in the evidence checker spreadsheet ready for import back into MetNetMaker for export and manipulation.	246
Table 34 : Evidence summary from the ECinfo database as presented in the evidence checker spreadsheet and used to help judge the existence of the reactions in table 33 and populate them with additional information for SBML export.....	247
Table 35 : Components of RPMI 1640 growth medium ²¹¹	248
Table 36 : Table created from calculations based on Table 1 in Teng. et al. Reported concentrations of metabolites have been converted from milliMolar concentration to $g/1015$ parasites. Note the very high standard deviations showing the inherent variability in the biology of the system.	249
Table 37 : Sara A carbon source flux calculations.	250
Table 38 : Sara B carbon source flux calculations.	250
Table 39 : Tom repeat of Jennifer carbon source flux calculations. No glycerol production was detected.	250
Table 40 : Sara A amino acid flux calculations. Hypoxanthine is listed last as its NMR assignment is unclear.	251
Table 41 : Sara B amino acid flux calculations. Hypoxanthine is listed last as its NMR assignment is unclear.	251

Table 42 : Tom repeat of Jennifer amino acid flux calculations. Hypoxanthine is listed last as its NMR assignment is unclear.....	251
Table 43 : Amino acid composition of 2 α -subunits + 2 β -subunits of the human haemoglobin molecule ordered by total weight per mol of haemoglobin. The molecular weights given are those of the amino acid after it has formed a protein.....	252
Table 44 : Amino acid use in the <i>P. falciparum</i> proteome, adapted from Chanda et al. ¹⁷⁶ . The molecular weights given are those of the amino acid after it has formed a protein.	253

List of acronyms

1H-NMR	Proton Nuclear Magnetic Resonance
AcCmpd	Generic Acetylated Compound
ADG	Alpha-D-Glucose
ADP	Adenosine Diphosphate
AMENDA	A text-mining method used to improve BRENDA results
AMP	Adenosine Monophosphate
AO	Acridine Orange
API	Application Programming Interface
A-T rich	Adenine-Thymine rich (referring to genomic nucleotide composition)
ATP	Adenosine Triphosphate
BDG	Beta-D-Glucose
BioPAX	Biological Pathways Exchange Language
BMRDB	Biological Magnetic Resonance Data Bank
BRENDA	Braunschweig Enzyme Database
BSA	Bovine Serum Albumen
cDNA	complementary Deoxyribonucleic acid
ChEBI	Chemical Entities of Biological Interest
COBRA	Constraints-Based Reconstruction and Analysis
CPMG	Carr-Purcell-Meiboom-Gill pulse sequence (see NMR)
CTP	Cytidine Triphosphate
dCTP	deoxy-Cytidine Triphosphate
DDT	Dichlorodiphenyltrichloroethane (an insecticide)
DEPC	Diethylpyrocarbonate
DHO	Dihydroorotate
DHODH	Dihydroorotate-dehydrogenase
DNA	Deoxyribonucleic Acid
dTTP	deoxy-Thymidine triphosphate
EC	Enzyme Commission (a number used to classify enzyme action)
EMBL	European Molecular Biology Laboratory
ER diagram	Entity Relationship diagram
ETC	Electron transport chain
FAC2H4unit	Fatty-Acid 2-carbon, 4-hydrogen subunit
FACS	Fluorescence-assisted Cell Sorting
FBA	Flux-balance Analysis
FBS	Faculty of Biological Sciences (University of Leeds)
FCS	Foetal Calf Serum
FRENDA	A text-mining method used to improve BRENDA results
FTP	File Transfer Protocol
FVA	Flux-variability Analysis
G-C content	Guanine-Cytosine content (referring to genomic nucleotide composition)
GDP	Guanosine Diphosphate
gDW	gramme of Dry Weight
GPR	Gene-Protein-Reaction
GSK	Glaxo-SmithKline
GTP	Guanosine Triphosphate
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HMDB	Human Metabolome Database
HMM	Hidden Markov Model
HMS	Hexose Monophosphate Shunt
HPLC	High-(performance/pressure) liquid chromatography

InChi	International Chemical Identifier
IQ	Intelligence Quotient
KEGG	Kyoto Encyclopedia of Genes and Genomes
KGML	KEGG Graphical Markup Language
LAMP	Liverpool Library of Apicomplexan Metabolic Pathways
LC-MS	Liquid chromatography–mass spectrometry
LINEST	Linear Estimation tool within Microsoft Excel and similar products
LISP	A programming language, historically List Processing
M	Molar, a concentration of one mole per litre
MIRIAM	Minimum Information Required In the Annotation of Models
MMCD	Madison Metabolomics Consortium Database
MPMP	Malaria Parasite Metabolic Pathways project
mRNA	messenger Ribonucleic acid
MySQL	an open-source database
NAD	Nicotinamide adenine dinucleotide (oxidised)
NADH	Nicotinamide adenine dinucleotide (reduced)
NADP	Nicotinamide adenine dinucleotide phosphate (oxidised)
NADPH	Nicotinamide adenine dinucleotide phosphate (reduced)
NEB	restriction endonuclease buffer
NeSC	National eScience Centre, Edinburgh
NMR	Nuclear Magnetic Resonance
NPO	Non-profit Organisation
OAA	Oxaloacetate
OG	Oxoglutarate
PARIS	Protein and RNA Isolation System
PBS	Phosphate-buffered Saline
PEP	Phosphoenolpyruvate
PERL	A programming language, Practical Extraction and Reporting Language
PPC	Pentose-phosphate Cycle
PROM	Probabilistic Regulation Of Metabolism
PRPP	5-Phosphoribosyl diphosphate
PSI-MI	Proteomics Standards Initiative - Molecular Interactions
RAST	Rapid Annotation using Subsystems Technology
RBC	Red Blood Cell
RID	Reaction Identifier
RNA	Ribonucleic Acid
RNA-seq	RNA-sequencing
RPM	Revolutions Per Minute
RPMI 1640	Roswell Park Memorial Institute 1640 growth medium
rRNA	ribosomal Ribonucleic Acid
SBGN	Systems Biology Graphical Notation
SBML	Systems Biology Markup Language
SBO	Systems Biology Ontology
SFAE	Single Fatty-Acid Elongation unit
Simpheny	Simulated Phenotype, software
SQL	Structured Query Language
SRI	Stanford Research International
SVG	Scalable Vector Graphics
TBS	Tris-Buffered Saline
TCA	Tricarboxylic Acid cycle. Also Krebs Cycle and Citric Acid Cycle
TF	Transcription Factor
TMS	Tetramethylsilane
tRNA	transfer Ribonucleic Acid
TTP	Thymidine triphosphate
UCSD	University of California, San Diego
UI	User Interface

UMP	Uridine monophosphate
UTP	Uridine triphosphate
VBA	Visual Basic for Applications
XML	eXtensible Markup Language

Introduction

Metabolic networks

Metabolism - the chemical processes that occur within a living organism in order to maintain life.

The Oxford English Dictionary

This thesis is an attempt to more fully understand the chemical processes that make up the metabolism of the parasite *Plasmodium falciparum*, the cause of malaria in humans.

Let us start at the beginning.

Life as a controlled network of chemical reactions

For the greater part of human history the realms of the living and the dead were considered to be largely separate. Greek philosophers, and many who followed, believed that the chemistry of animals and plants was inspired by a force greater than what drove reactions between minerals.

In 1828, Friedrich Wöhler began to dismantle the barrier between the living and the non-living by showing that urea could be made without the intervention of a living organism. The discovery that a compound found in the urine of animals could be made from two solutions not associated with life marked the birth of organic chemistry.

Meanwhile, in the fields of agriculture and biology, the increasing population of the world was quickly approaching the limits of the soil to produce food. In highest demand by farmers were the nitrites and similar nitrogenous compounds required for the production of proteins by plants. At this time, humanity had only one primary renewable source for these compounds; leguminous plants such as peas and beans which were known to accumulate nitrates in their roots. In 1888 Martinus Beijerinck had isolated and studied the bacteria that lived in the root nodules of these plants¹ and noted that they produced nitrites from the hydrogen and the nitrogen in the air but the mechanism of the reaction remained unknown. The limited amount of nitrogenous fertiliser, from manure, guano, or saltpetre mines in South America risked defining the limit of human population.

With such a constraint on production the price of nitrites grew ever higher and much attention was turned to performing for ourselves the same reaction that bacteria had mastered millennia before. The development of the Haber process in 1908 was the achievement of that goal. And yet it was striking that industrial production of nitrites required a temperature of 300°C and a pressure of 100 atmospheres whilst the same reaction in bacteria operates at ambient temperature and pressure. Although we had learned to mimic a reaction that bacteria had mastered it was clear that we still had a lot to learn from bacteria in the emerging field of biochemistry.

In this genomic era of biology we have moved beyond mimicking reactions that bacteria already perform and started to use bacteria for our own purposes. Insulin, a hormone produced by the pancreas, helps regulate blood sugar levels in humans. In patients with diabetes insulin is not produced properly by the patient but blood sugar levels can still be controlled by injecting insulin from an external source. For decades the only source of insulin was by isolating it from the pancreas of an animal an expensive and dangerously unsterile procedure. It was clear that it would be a great advantage to be able to synthesise pure insulin — a molecule over 300 times as large and enormously more

complex than ammonia — from basic materials but the existing chemistry was not up to the task.

The solution arrived at by Herbert Boyer in 1978 was to add a gene encoding insulin to a bacterium and let the intricately evolved and interoperable systems of life deal with the complexity of the synthesis. This application of biochemistry is now often called biotechnology and it is probably the fastest growing discipline in all of science. Today we are turning to biotechnology to solve problems in fields as diverse as medicine, environmental protection and energy supply.

In less than 200 years we have moved from single reactions between minerals to hijacking the thousands of controlled and overlapping reactions that let bacteria live to our own ends. Wöhler opened the world of organic chemistry to us with a single reaction. Haber developed a process that not only permitted a more difficult reaction but also linked together several reactions. Boyer harnessed the set of regulated reactions within bacteria to produce a chemical product of unprecedented industrial complexity, but he did so by using, not understanding, the complex systems of the *E. coli* bacterium. The future of biotechnology lies in extending our ability to harness the existing set of regulated reactions within an organism to shaping that metabolism to our own needs. Before we can do this we must first more fully understand the process of metabolism.

The chemical reactions we know to exist in the living world are no different to ones we could perform in isolation without the intervention of life. What makes life special is that it has evolved not to perform these reactions in isolation but rather to perform them together and most importantly to perform them as part of a controlled network with reactants and products separated and allowed to mix as and when required. The scale on which these networks operate and the variability and harshness of the conditions they are able to tolerate is astounding and it will take decades to understand them fully. However, even at this early stage it seems clear that much of the function and the resilience of the metabolic networks that maintain life can only be understood in terms of the relations between all the components of the network.

The gene-protein-reaction relationship and the importance of enzymes

Some of the reactions within a living organism occur spontaneously at a rate sufficient to permit life, and perhaps more frequently at a rate sufficient to cause death. Depending on how you look at it we should be either frustrated or grateful then, that the majority of reactions require the intervention of one enzyme or more to take part in the processes of life. It would be hard to make much progress in describing metabolism without first describing what has become known as the central dogma of molecular biochemistry.

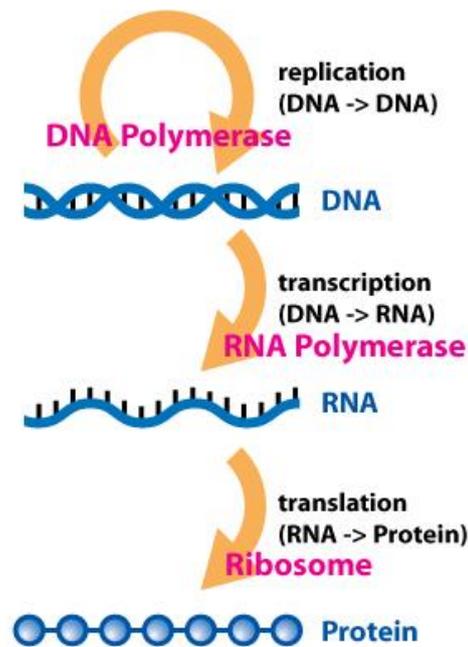


Figure 1: Simplified depiction of the central dogma of molecular biochemistry (Daniel Horspool – Wikimedia Commons under license: CC BY-SA 3.0).

The central dogma, as shown in figure 1, is that DNA is transcribed to RNA by RNA polymerase and then translated to a protein by a ribosome. In this simple model, the unit of transcription is the gene, with each gene directly related to one strand of RNA and each strand of RNA directly related to one protein. If the protein is an enzyme — and in study of metabolism almost all proteins of interest will be — then this enzyme catalyses a single reaction. This gene-protein-reaction (G-P-R) relationship, where the proteins act as enzymes, is at the core of studying metabolism.

In reality the links between gene and protein are complicated by alternative splicing, post-transcriptional modification and post-translational modification. Similarly the link between protein and reaction is often more complex than simply linking a single enzyme with a single chemical reaction. Nevertheless, the underlying concept of the central dogma and its application to the study of metabolism as a set of G-P-R relationships is valid and in the case of simple bacteria, unencumbered by many of the complexities of higher orders of life, can give a remarkably accurate representation of an organism.

Now that we know where they come from, let us turn to considering what enzymes do.

We can break enzyme function down into three parts. They speed up slow reactions, they allow reactions to occur that would otherwise take place so infrequently that they might as well be considered impossible and they let an organism control when and where reactions occur. Let's deal with these three features of enzymes in turn.

Enzymes speed up reactions

The rate at which a reaction proceeds is overwhelmingly governed by two factors, activation energy and temperature, as formalised in Arrhenius' law.

$$k = Ae^{-E_a/RT}$$

k = reactions/second, A = constant, E_a = activation energy, R = gas constant, T = temperature in Kelvin

The simplest way to speed up reactions is thus to increase the temperature and as a rule of thumb Arrhenius' law tells us that the rate of a reaction doubles for every 10°C increase in temperature. For industrial processes, increasing temperature is rarely a problem. Life on the other hand is restricted by both the temperature of its surroundings and the physical properties of water and even the hardiest bacteria cannot survive in boiling water for long.

Once the option of raising temperature to speed up reactions is exhausted*lowering a reaction's activation energy is the most effective alternative. Enzymes are the direct equivalents of the catalysts — typically made of heavy metals and often effective only at high temperatures — we use industrially for this purpose. In comparison, enzymes have evolved to operate in the conditions provided by the host organism and only occasionally incorporate rare metallic atoms to improve their effectiveness.

Enzymes and catalysts speed up reactions not by changing the equilibrium concentrations of the reactants and products or the free energy difference between them but rather by reducing the activation energy required for the reaction to proceed.

Enzymes permit extremely unlikely reactions

A naïve interpretation of the second law of thermodynamics can show life to be impossible in that it frequently appears to build order from disorder and structure from individual compounds. The flaw in this line of argument is that the Earth is not a closed system but rather driven by the heat and light delivered by the Sun. Still, it is essential to consider the mechanism by which the Sun is able to drive all the chemical processes of life on Earth given that only a tiny fraction of the reactions in a tiny fraction of the organisms on the planet are able to photosynthesise and thus directly harness the energy that the sun provides. These photosynthetic organisms are able to convert electromagnetic radiation into compounds with a large amount of easily accessible energy, usually adenosine triphosphate (ATP), frequently stored in a more stable form such as glucose for conversion back to ATP, or similar, as required. Compounds like ATP, with large amounts of available free energy, are what allow hundreds of other key enzyme-mediated metabolic reactions

* few warm-blooded animals exceed a body temperature of 40°C.

to occur. More specifically, ATP permits reaction to occur in a direction that would not otherwise satisfy the second law of thermodynamics.

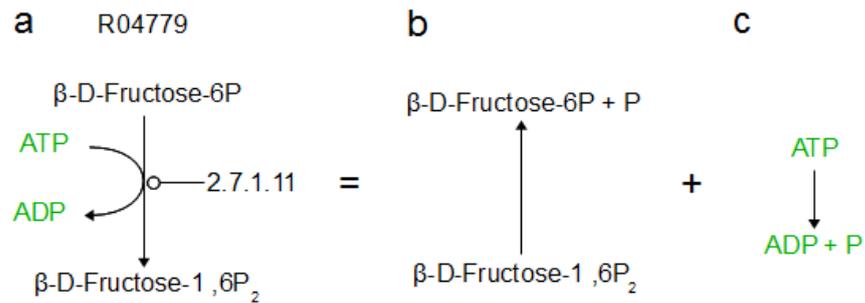


Figure 2: KEGG reaction R04479, catalysed by enzyme 2.7.1.11, typically proceeds in the direction shown in part **a**. It can be considered a combination of two reactions, **b** and **c** with opposite energetically favoured directions.

KEGG reaction R04479, shown in figure 2, is a good example of this property. The reaction shown in part **b**; $\beta\text{-D-Fructose-6P}$ to $\beta\text{-D-Fructose-1,6P}_2$ through the addition of a phosphate ion, is not energetically favourable and at equilibrium the population of $\beta\text{-D-Fructose-6P}$ will be much higher than the population of $\beta\text{-D-Fructose-1,6P}_2$. Similarly, but in the other direction, the equilibrium concentration of ATP is extremely low compared with the equilibrium concentration of ADP.

Because an enzyme ensures that both components of the desired reaction R04479 take place at the same time and in the same place it can also ensure that the free energy released by the dissociation of ATP and the high energy phosphate group can be used to phosphorylate the $\beta\text{-D-Fructose-6P}$ molecule. Furthermore, although the natural reaction rates of reactions **b** and **c** are much higher than the natural reaction rate of reaction **a**, the presence of the enzyme 2.7.1.11 specific to reaction **a** means that as long as ATP is supplied to the system, $\beta\text{-D-Fructose-1,6P}_2$ is produced more quickly than it is consumed by reaction **b**. For this reason we can, in the presence of the correct enzyme, draw reaction R04479 just as in figure 2a.

For almost all of this thesis this simplified picture will serve us adequately and all the analysis I have performed assumes that an enzyme-mediated reaction takes place only in the presence of the enzyme facilitating it. In all but a few cases I have ignored the kind of spontaneous decay shown in figure 2b and c. Nevertheless, it is important to appreciate the limitations of the simplification of all three sections of figure 2 to just the part shown in section **a** and be wary of situations where this simplification is no longer valid.

Enzymes let an organism control when and where a reaction takes place

The third key advantage to life of making use of enzyme-catalysed reactions is that the presence and location of these enzymes can be set to allow certain reactions to occur only at certain times within the life-cycle or within certain organelles and internal

compartments. Although this property is useful in bacteria and other prokaryotes it is particularly important to the functioning of mitochondria within eukaryotic cells. Mitochondria can only function because particular reactions occur in separate compartments with limited permeability between them. The inclusion of compartments within a metabolic model for *P. falciparum* is similarly essential as I will show later.

Metabolism is highly conserved

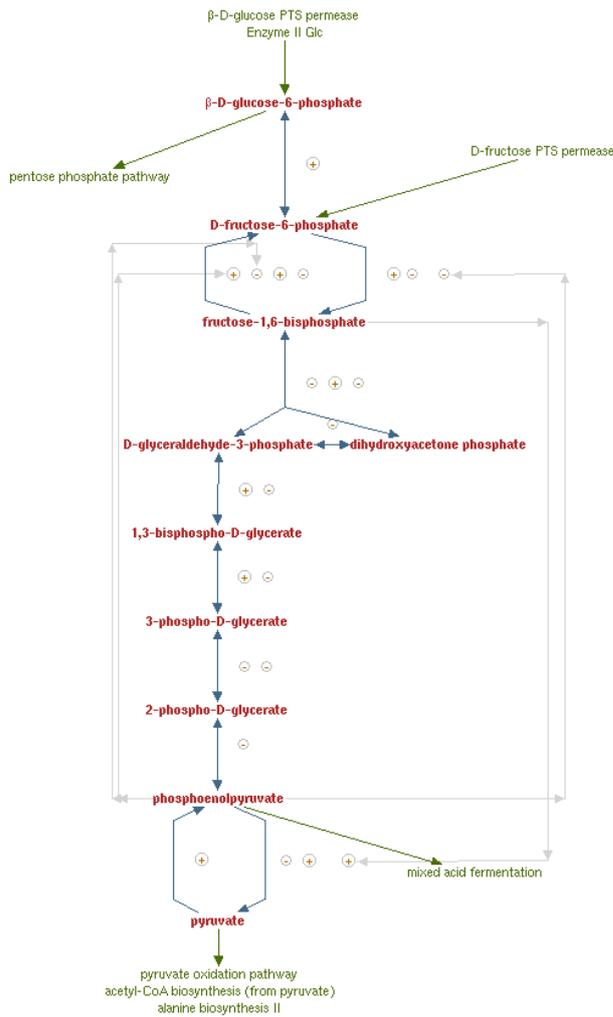
Now that we have seen how genes, enzymes and reactions are related and how enzymes thus govern the chemical reactions that make up metabolism, let us turn to what I consider one of the most beautiful results in all of science.

That the machinery of the central dogma of biochemistry is shared across all living organisms is magnificent evidence for the common ancestry of all life on Earth and it is this common language of DNA, RNA and proteins that lets us express human genes in bacteria. And yet I see absolutely no reason to extrapolate from this that the products of that machinery, the enzymes that govern the chemical reactions that make up our metabolism, should be as deeply shared. Figure 3 shows the collection of reactions that form the Glycolysis I metabolic pathway — the key link in the conversion of glucose to lactate that sports scientists call anaerobic respiration — in humans and in *E. coli*. It is striking in this image that the two pathways, despite around three billion years of separate evolution, are almost exactly the same. And not only are the reactions the same but the amino acid sequence of each pair of enzymes in this pathway is so similar that the DNA sequence of each enzyme-coding gene in *E. coli* can be found, albeit with a few modifications, in every human's genome.

In some ways it should not surprise us that some core pathways, and their component reactions (and thus enzymes and genes) are conserved. The process of evolution works in small steps, not in great leaps, and thus a successful system, even an imperfect one, is likely to be retained in future species. And yet in the 3-4 billion years since life began there has been plenty of time for changes to occur and for the different species, and certainly the different kingdoms, of life to diverge and find different enzymes, perhaps even different pathways, to perform identical roles. For this reason I think the finding of Peregrin-Alvarez *et al.*² that “half of the metabolic enzymes have homologs in all domains of life” is particularly beautiful. To someone like myself who entered the field of biochemistry after studying physics it is also surprising.

Glycolysis I Pathway in *E. coli*

K-12 substr. MG1655



Glycolysis I Pathway in *H. sapiens*

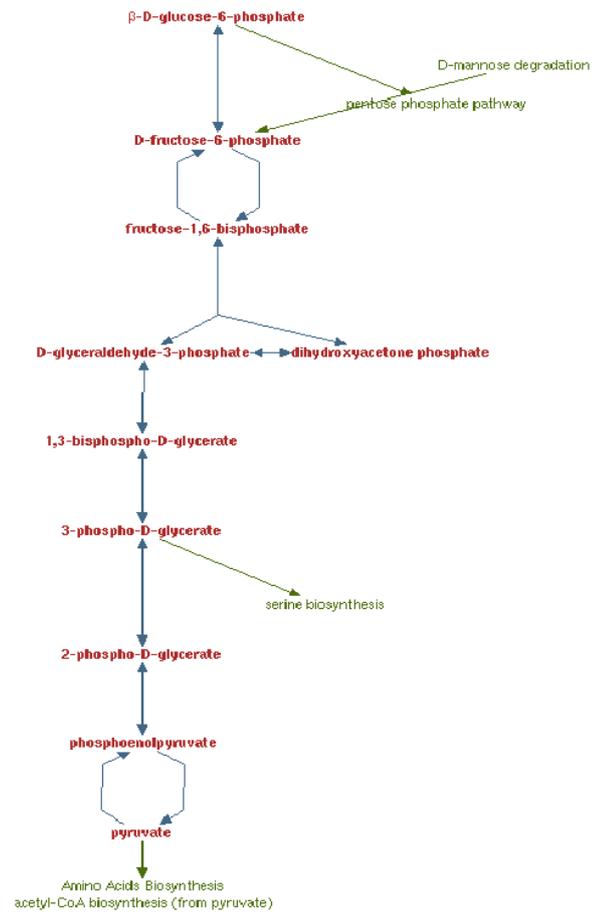


Figure 3: Visual representations of the Glycolysis I Pathway in *E. coli* (left) and Human (right) in Biocyc 15.1

The widely shared homology of metabolic enzymes has enormous practical benefits. It means that as long as we have the genome sequence of an organism — an increasingly affordable thing to discover — we can immediately find a significant proportion of its enzymes just by searching that genome for sequences homologous to known genes. In this way an enzyme whose associated gene and reaction has been discovered in one organism can be found in all other organisms where there is a homolog.

Existing ontologies in systems biology

KEGG LIGAND contains our knowledge on the universe of chemical substances and reactions that are relevant to life.

- **First sentence of the KEGG LIGAND site at www.genome.jp/ligand**

In order to define a set of chemical reactions we must first define a single reaction. This is not as easy as it sounds.

Identical reactions almost always have many different names and a single name is not guaranteed to refer to a single reaction. Furthermore, each compound making up a reaction will almost always have many different names and again a single name is not guaranteed to refer to a single compound. Clearly, in order to construct a metabolic network these naming problems have to be overcome and a single naming system must be defined. This defined naming system is called an ontology and within systems biology there are two leading ontologies that define both reactions and compounds, MetaCyc and KEGG.

MetaCyc

The MetaCyc ontology³ is maintained by Peter Karp's group at SRI International (previously Stanford Research Institute) and is available as a collection of flat-files for free to both academics and industry and as part of Pathway Tools for free to academics. It is tightly coupled to the Biocyc pathway/genome database collection and the pathway tools software package available from SRI International. As of 2011, MetaCyc contained a description of 9460 unique reactions and 9188 Compounds.

MetaCyc reactions and compounds are linked as part of a well-defined hierarchy stored by Pathway Tools in object-orientated structures written in the LISP programming language. This has the advantage of providing extensive links between entities within the ontology but the drawback that the data structure is extremely complex. Access to the MetaCyc ontology is available directly via Pathway Tools and via Java and Perl APIs for Pathway Tools all of which represent a steep learning curve for those new to the field. The ontology is also available as a collection of flat-files in a variety of formats although the structure of these files is necessarily complex to represent the extensive links between component entities.

KEGG LIGAND

The LIGAND ontology⁴ is part of the larger Kyoto encyclopedia⁵ of genes and genomes (KEGG) and was maintained until recently by its founder within the Kanehisa group.

Financial pressures mean that access is now provided via a non-profit organisation, NPO Bioinformatics Japan. Currently the ontology is mostly free to academics and industry* and looks likely to remain so but recent developments mean that access may become more difficult in the future. As of 2011, LIGAND describes 8146 unique reactions and 14762 Compounds (some deprecated, some similar).

The ontology is available directly on the internet and via the KEGG API with the complete ontology available in flat files. Each compound is given a unique ID starting with a C and followed by 5 digits (C00001 = water, C00002 = ATP, etc..) and each reaction is given a unique ID starting with an R and following by 5 digits (R00004 = diphosphate phosphohydrolase, R00028: maltose glucohydrolase, etc...). Although it is hard to find, KEGG provides a file named compound.inchi which provides International Chemical Identifier (InChI) codes — text strings that uniquely describe all chemical compounds — for all KEGG compounds.

LIGAND vs MetaCyc

With the exception of some notable efforts to link the two major ontologies of metabolism, some of which I mention later, they remain effectively incompatible. Therefore it is regrettably necessary to make a choice before embarking on a considerable effort in a given area. To this end I took the advice of many different people including my supervisor David Westhead, others in the lab including John Whitaker and Elizabeth Edwards and Markus Herrgard. Additionally, I attended a two-day workshop on Pathway Tools at King's College London in 2008⁶. The opinions I formed then, and since, on the differences between the two ontologies are summarised below in two parts.

Ontology take-up and completeness

Biocyc currently contains 1129 pathway/genome databases all based on the MetaCyc ontology of which 4 are intensively curated with almost all genes having been identified by an expert and the corresponding reaction assigned by that person. Thirty-two pathways are mildly curated with a computational model having been improved by an expert in the metabolism of the organism or in metabolism generally. The remaining metabolic networks have been created by computational annotation only.

KEGG currently contains 1536 pathway/genome databases all based on the LIGAND ontology the vast majority of which were assembled using computational annotation

* Direct internet and API access is unrestricted and can feasibly supply the whole ontology. Bulk downloads of the ontology as flat-files via FTP is now restricted to paid licensees as part of a larger package of services KEGG provides.

alone. The extent of manual curation on those pathways claiming to have been manually curated is extremely variable.

Both ontologies are widely used and capable of accurately representing a metabolic network. In fact both had been used for metabolic network reconstructions of *Plasmodium falciparum*, MetaCyc by Yeh *et al.*⁷ and LIGAND by Ginsberg⁸, before I started my PhD.

Ontology complexity

Although it is unwise to judge a scientific tool on its appearance, the best place to start comparing the complexity of the two ontologies is to look at the default representation of a single compound. In figure 4 and figure 5 on the following pages a condensed version of the compound descriptions for ATP in each of the ontologies is shown. Whilst both definitions uniquely define the compound there is a stark contrast between the LIGAND representation, three pages of simple but ordered information, and the MetaCyc representation, seventeen pages of extensive and well linked but cluttered information.

A closer inspection of these figures shows a key advantage of MetaCyc over LIGAND. In both ontologies compounds are well linked with reactions and pathways in addition to information on the physical properties of the compound. MetaCyc goes a step further by linking compounds with other compounds, a property of the ontology's structure that is extended to reactions. In the ATP example we see that ATP belongs to various nested classes of compounds, "a purine phosphate" which is a member of "a ribonucleoside triphosphate" which is a member of "a nucleoside triphosphate" and so on. This extra linking offers some attractive features. For example ambiguities in a metabolic network can be represented more simply and more flexibly within this system.

In my experience this contrast in styles extends across the full extent of both ontologies. LIGAND's use of unique IDs makes using the ontology simpler and easier to apply computationally and in my use there are only a few areas where the more complex hierarchical structure of MetaCyc makes sacrificing that simplicity worthwhile*. It is further worth noting that the ambiguities that the MetaCyc ontology permits in metabolic networks must be resolved; the subject of a fascinating talk⁹ by Jeremy Zucker of the Broad institute at the NeSC metabolic modelling workshop at the National eScience Centre, Edinburgh, 07/04/2010.

In summary then, my slight preference is for the LIGAND ontology. This is a particularly sensible choice because the most extensive manual curation of malaria metabolism to date — the malaria metabolic pathways database (MPMP)¹⁰ — uses the LIGAND ontology.

* Work by Mark Poolman and David Fell at Oxford Brookes University using the MetaCyc compound and reaction names within a simplified framework suggests that this trade-off is avoidable.

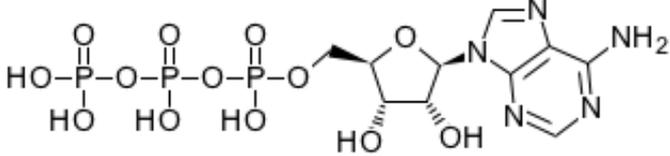
Entry	C00002 Compound
Name	ATP; Adenosine 5'-triphosphate
Formula	C10H16N5O13P3
Exact mass	506.9957
Mol weight	507.181
Structure	 <p>C00002</p> <p> Mol file KCF file DB search Jmol KegDraw </p>
Remark	Same as: D08646 BRITE hierarchy
Reaction	R00002 R00076 R00085 R00086 R00087 R00088 R00089 R00104 R00105 R00124 R00126 R00127 R00128 R00129 R00130 R00137 ... plus many more ...
Pathway	ko00190 Oxidative phosphorylation ko00195 Photosynthesis ... plus many more ...
Enzyme	1.2.1.30 1.2.1.31 1.3.7.8 1.13.12.7 1.14.99.19 (C) 1.17.4.1 (C) 1.17.4.2 1.17.4.2 (C) ... plus many more ...
Other DBs	CAS: 56-65-5 PubChem: 3304 ChEBI: 15422 KNAPSAcK: C00001491 3DMET: B01125 NIKKAJI: J10.680A
KCF data	Show

Figure 4: The KEGG LIGAND compound entry for ATP, shortened from 3 pages. The simplicity of the ontology appeals to me.

Compound: ATP

Synonyms: adenylypyrophosphate, adenosine-triphosphate, adenosine-5'-triphosphate

Superclasses: [a nucleic acid component](#) → [a nucleotide](#) → [a nucleoside triphosphate](#) → [a ribonucleoside triphosphate](#) → [a purine triphosphate](#)

[a nucleic acid component](#) → [a nucleotide](#) → [a ribonucleotide](#) → [a ribonucleoside triphosphate](#) → [a purine triphosphate](#)

[a nucleic acid component](#) → [a purine-related compound](#)

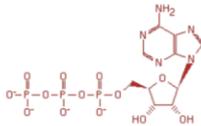
Component of: [DnaA-ATP transcriptional dual regulator](#) , [Malt-Maltotriose-ATP DNA-binding transcriptional activator](#)

In Mixture: [E.coli biomass \(example\)](#) , [E.coli biomass \(even smaller example\)](#) , [E.coli biomass \(smaller example\)](#)

Empirical Formula: C₁₀H₁₂N₅O₁₃P₃

Molecular Weight: 503.15 Daltons

Monoisotopic Molecular Weight: 506.99574515689994 Daltons



SMILES: C3(N(C1(C(C(O1)COP(OP(=O)([O-])OP(=O)([O-])OP(=O)([O-])O)O)C2(=C(C(=NC(=N2)N)N=3)))

Unification Links: CAS:56-65-5, [ChEBI:15422](#) , [KEGG:C00002](#) , [PubChem:5957](#)

Gibbs Energy of Formation (kcal/mol, estimated): -60.6

In Pathway Reactions as a Reactant:

[β-D-glucuronide and D-glucuronate degradation](#) :

[2-dehydro-3-deoxy-D-gluconate + ATP → 2-dehydro-3-deoxy-D-gluconate-6-phosphate + ADP + H⁺](#)

[1,4-dihydroxy-2-naphthoate biosynthesis I](#) :

[ATP + o-succinylbenzoate + coenzyme A → AMP + diphosphate + 4-\(2'-carboxyphenyl\)-4-oxobutyryl-CoA](#)

... plus many more ...

In Pathway Reactions as a Product:

[acetate formation from acetyl-CoA I](#) :

[acetylphosphate + ADP = acetate + ATP](#)

[allantoin degradation IV \(anaerobic\)](#) :

[carbamoyl-phosphate + ADP + H⁺ → ammonia + CO₂ + ATP](#)

... plus many more ...

In Transport Reactions:

[ATP + ferric enterobactin_{\[periplasmic space\]} + H₂O → ADP + phosphate + ferric enterobactin_{\[cytosol\]}](#) ,

[ATP + iron \(III\) hydroxamate complex_{\[periplasmic space\]} + H₂O → ADP + phosphate + iron \(III\) hydroxamate complex_{\[cytosol\]}](#) ,

... plus many more ...

In Reactions not Assigned to Pathways:

[a reduced flavodoxin + a ribonucleoside triphosphate = an oxidized flavodoxin + a deoxyribonucleoside triphosphate + H₂O](#) ,

[a nucleoside triphosphate + H₂O → a nucleoside monophosphate + diphosphate + H⁺](#) ,

... plus many more ...

This compound has been characterized as an enzyme activator. These enzymes are displayed based on the type of activation:

Activator (Allosteric) of: [ribonucleoside-triphosphate reductase](#) [[Eliasson94](#) , [Comment 1](#)] , [aspartate](#)

Activator (Mechanism unknown) of: [inosine kinase](#) [[HoveJensen89](#)] , [guanosine kinase](#) [[HoveJensen89](#)]

This compound has been characterized as an enzyme inhibitor. These enzymes are displayed based on the type of inhibition:

Inhibitor (Competitive) of: [D-mannonate oxidoreductase](#) [[MandrandBe77a](#)] , [malate oxidase](#) [[Comment 3](#)]

Inhibitor (Noncompetitive) of: [phosphate acetyltransferase](#) [[Suzuki69](#) , [CamposBerm10](#) , [Comment 7](#)]

Inhibitor (Allosteric) of: [6-phosphofructokinase](#) [[Guixe98](#) , [Kotlarz81](#)] , [adenylate cyclase](#) [[Yang83a](#)]

Inhibitor (Mechanism unknown) of: [GTPase](#) [[Jain09](#)] , [mannose isomerase](#) [[Itoh08](#)] , [thiamine phosphate](#)

Activates: [NtrC-P^{asp}](#) + H₂O → [phosphate](#) + [NtrC](#)

Transcription Units regulated by related protein DnaA-ATP transcriptional dual regulator (9 total):

Figure 5: The MetaCyc compound entry for ATP, shortened from 17 pages. The extensive linking within the ontology has many advantages but is intimidating to work with.

Other ontologies of note

Although KEGG and MetaCyc are the dominant ontologies for defining metabolism there are other ontologies that serve useful complementary functions which are worth describing briefly.

ChEBI

The chemical entities of biological interest database¹¹, ChEBI, is in its own words “a *freely available dictionary of molecular entities focused on 'small' chemical compounds*”. In aim this is no different to the compound databases found in MetaCyc, LIGAND or other competitors and its focus on compounds limits its usefulness for my purposes.

Despite this drawback, ChEBI has two distinct advantages. Firstly, its stated aim to stay completely free makes it a safer platform on which to build tools. Secondly, ChEBI is in my experience the best linked compound database with almost all compounds linking directly to international chemical identifiers (InCHI) codes and KEGG compound IDs amongst other identifiers.

ChEBI is of further note because the combined yeast model assembled at the 2007 Yeast Jamboree¹² — one the most comprehensive metabolic models assembled to date — used ChEBI identifiers as the preferred compound description format.

EC numbers

All the previous ontologies I have discussed deal with the reaction part of the gene-protein-reaction relationship. The enzyme commission’s nomenclature for the classifications of enzymes¹³, EC numbers*, are different in that they primarily describe the protein’s function.

An EC number consists of four numbers separated by a full-stop with the first number describing the action of the enzyme in very general terms and each number to the right describing a more detailed subset of that action. As an example, the enzyme involved in the reaction 6-phosphofructokinase, as shown in figure 2, is defined in the top part of figure 6.

* often referred to by the backronym enzyme classification numbers.

2 = Transferase

.7 = Transferase, **transferring Phosphorous-containing groups**

.1 = Transferase, transferring Phosphorous-containing groups, **with an alcohol group as acceptor**

.11 = Transferase, transferring Phosphorous-containing groups, with an alcohol group as acceptor, **6-phosphofructokinase**



2.7.1.11 =	R00756	ATP:D-fructose-6-phosphate 1-phosphotransferase
	R00767	CTP:D-fructose-6-phosphate 1-phosphotransferase
	R00769	UTP:D-fructose-6-phosphate 1-phosphotransferase
	R00770	ITP:D-fructose-6-phosphate 1-phosphotransferase
	R01843	ATP:Sedoheptulose 7-phosphate 1-phosphotransferase
	R03236	ATP:D-tagatose-6-phosphate 1-phosphotransferase
	R03237	CTP:D-Tagatose 6-phosphate 1-phosphotransferase
	R03238	UTP:D-Tagatose 6-phosphate 1-phosphotransferase
	R03239	UTP:D-Tagatose 6-phosphate 1-phosphotransferase
	R04779	ATP:D-fructose-6-phosphate 1-phosphotransferase*

**R00756 and R04779 are not identical, R04779 acts on the β form of the fructose*

Figure 6: The tree structure of the EC number definition for 6-phosphofructokinase and the 10 KEGG reactions related to 2.7.1.11.

An EC number could perfectly define of a reaction if it were unambiguous but in this, and many other cases, the single EC number 2.7.1.11 describes no fewer than ten separate KEGG reactions, each defined in terms of four of the twelve compounds present across all ten reactions. Whilst these reactions are very similar, when building a metabolic network where reactions must link exactly, very similar is not good enough.

Figure 6 shows us that the complexities of the real world and the requirement for an ontology to remove ambiguity means that the naïve model of metabolism, where EC numbers uniquely define both protein function and a single associated reaction, is insufficient. Still, EC numbers are probably the single most universally applicable descriptor of metabolism and an extremely valuable resource for linking genes and reactions.

Beyond the ontology: a rigorous reconstruction framework

The described ontologies are an important unambiguous vocabulary for describing the components of metabolic models. And yet in order to build functional models upon which we can perform analysis and simulations we need to assemble these components, usually from multiply ontologies, into models in a regular, unambiguous and structured way. In the absence of a generally accepted term I have called this system of assembly a reconstruction framework.

As part of my attempts to explore the requirements of such a framework I created the videos¹⁴ from which figure 7 and figure 8 are taken.

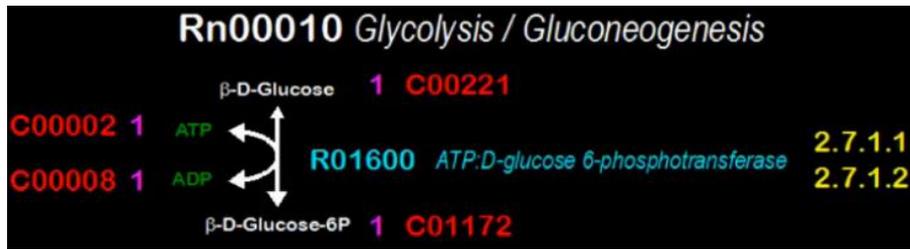


Figure 7: The LIGAND reaction R01600 labelled with the key components required to define a reaction within a reconstruction framework. This is a still frame from the “Framework Video” on my project website¹⁴.

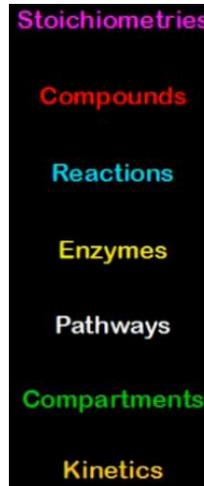


Figure 8: The key components required to define a reaction within a reconstruction framework. This is a still frame from the “Framework Video” on my project website¹⁴.

Figure 7 breaks down a single reaction within a reconstruction framework into components with each component colour co-ordinated as in figure 8. The stoichiometries, compounds (both IDs and names), reaction (both ID and name), and the associated pathway (both ID and name) are taken from the LIGAND ontology and fully define the reaction. Additionally, EC numbers associated with the reactions are included to provide a link to the enzyme. A gene could additionally be assigned to this reaction to complete the GPR definition. When further reactions are added to a model it is essential that they conform to this same structure and are defined according to the same ontologies in order to avoid confusion or ambiguity.

For the definition of metabolic models suitable for analysis two further properties are required for each reaction: compartment and kinetics. Since any given reaction can occur within different cellular compartments — the cytosol, mitochondria, nucleus, etc... — a compartment must be defined for each reaction to ensure it is unique and to ensure that its reactants and products are properly separated from those in other compartments. Kinetic properties are essential to creating a valid metabolic model upon which simulations can be run and in the scope of this thesis the only kinetic properties provided are limited to simple flux constraints.

The seven components of a reconstruction framework as shown in figure 8 are a guide rather than a prescription. As long as a rigorous framework is adhered to, each component of that framework could derive from a separate ontology. For the rest of this thesis the LIGAND ontology and the EC system, in addition to custom naming of compartments and arbitrary kinetic parameters, are used.

SBML: the standard for network exchange

A metabolic network can at its simplest be represented simply as a list of reactions on the condition that those reactions are properly defined somewhere as part of a rigorous framework. The first metabolic networks were thus lists, or, more usually, spreadsheets containing lists of reactions with additional information on each reaction contained in extra columns. Whilst this approach is perfectly adequate when a model is used within a single group using a single reconstruction framework it presents problems when models need to be shared. In these cases reactions are frequently not accompanied by a full definition of the stoichiometries and compounds from which they are formed leading to ambiguity in the metabolic model. Furthermore, without a standard for model interchange the order and meaning of entries in different columns of a spreadsheet are variable and unclear.

In April 2000, as this problem of model interchange within systems biology was becoming clear, “The 1st Workshop on Software Platforms for Systems Biology” was held at the California Institute of Technology¹⁵ and by 2003, the proposed solution, the systems biology markup language, SBML¹⁶, had been finalised. This was followed in 2004 by the biological pathways exchange language, BioPAX¹⁷, proposed by a group including Peter Karp at SRI and joined that year by PSI MI¹⁸, the Proteomics Standards Initiative’s molecular interaction XML format. Other interchange formats have been proposed but in the field of metabolic modelling SBML and BioPAX are the only two in wide enough use for me to have encountered them in the last four years.

Although SBML and BioPAX are far easier to convert between than LIGAND and MetaCyc it is still easiest to choose one when starting a large project and for this reason they must be compared. I can give no better a summary of the comparison between these two formats than that published in 2005,

“SBML is better on simulation-related properties [whilst] BioPAX provides the richest and most general representation. The richer hierarchy of BioPAX, which is a benefit with respect to representation of data, has a price with respect to computational complexity.”

From the conclusion of “Representations of molecular pathways: an evaluation of SBML, PSI MI and BioPAX”¹⁹ by Lena Strömbäck and Patrick Lambrich.

Since 2005, the scope of both SBML and BioPAX has increased as the standards have matured. Higher levels — currently at level 3 in both specifications²⁰⁻²² — have been added to allow a more complete description of the modelled system.

Choosing between these standards is difficult and many people in the systems biology community have very strong views on the subject. The hierarchy and flexibility of BioPAX means that almost every imaginable aspect of a model can be contained within the format but I have found the cost in terms of usability too high. SBML is much simpler to work with and in almost all cases it is good enough.

It seems that others agree with me and as of June 2011 the SBML software matrix lists 225 tools compatible with the format. To further prove the dominance of SBML, Pathway Tools — a project with close ties to the BioPAX working group — now supports SBML export of models in addition to the longer-standing option for BioPAX export. The interoperability of models afforded by SBML is an outstanding achievement and its value should not be underestimated. As long as the format is maintained and updated as rapidly as it is currently it will remain, as many argue it currently is²³, the *lingua franca* of biological modelling.

SBO (systems biology ontology)

The systems biology ontology²⁴, SBO, project is different to the other ontologies I have described in that it aims not to define the physical contents of a metabolic model but rather the mathematical properties of that model: what I called kinetics in figure 8. Currently models can be exchanged using SBML but there is no standard way to define things like enzyme properties, reaction kinetics or any of the more complex features of a mathematical model. This makes it difficult for one group to repeat simulations, and thus verify the results obtained by another group, performed on metabolic networks. This in turn makes it difficult for existing models to be expanded and adapted to future modelling techniques.

SBO is developed at the European Bioinformatics Institute and, although it can be used to mathematically define models in any format, it is tightly coupled with the maintenance and extension of the SBML standard. As the modelling techniques applied to metabolic, and other, models increases SBO is likely to become an increasingly important part of the SBML specification.

A less adventurous effort to standardise metabolic models with the goal of making their simulated behaviour reproducible is the MIRIAM²⁵ (minimum information required in the annotation of models) project which I discuss further at the end of this thesis.

Visual representations of metabolic networks

An intuitive feel for the process of metabolism can best be gained from a visual representation of it. For example, the visual representation in figure 3 is easily understandable to someone with some experience of chemistry and requires no special explanation. By comparison, the mathematical model underlying the visualisation would not be widely understood.

I strongly believe that visualisation techniques are essential to the wider understanding of metabolic models and the results of their analysis and I develop this idea throughout this thesis. To help with that it is important to define some basic principles of visualisation.

Networks of nodes (objects) connected by edges (lines) are called graphs and their study is called graph theory. As shown in figure 9, there are essentially three different types of graphs that can be used to display metabolic networks.

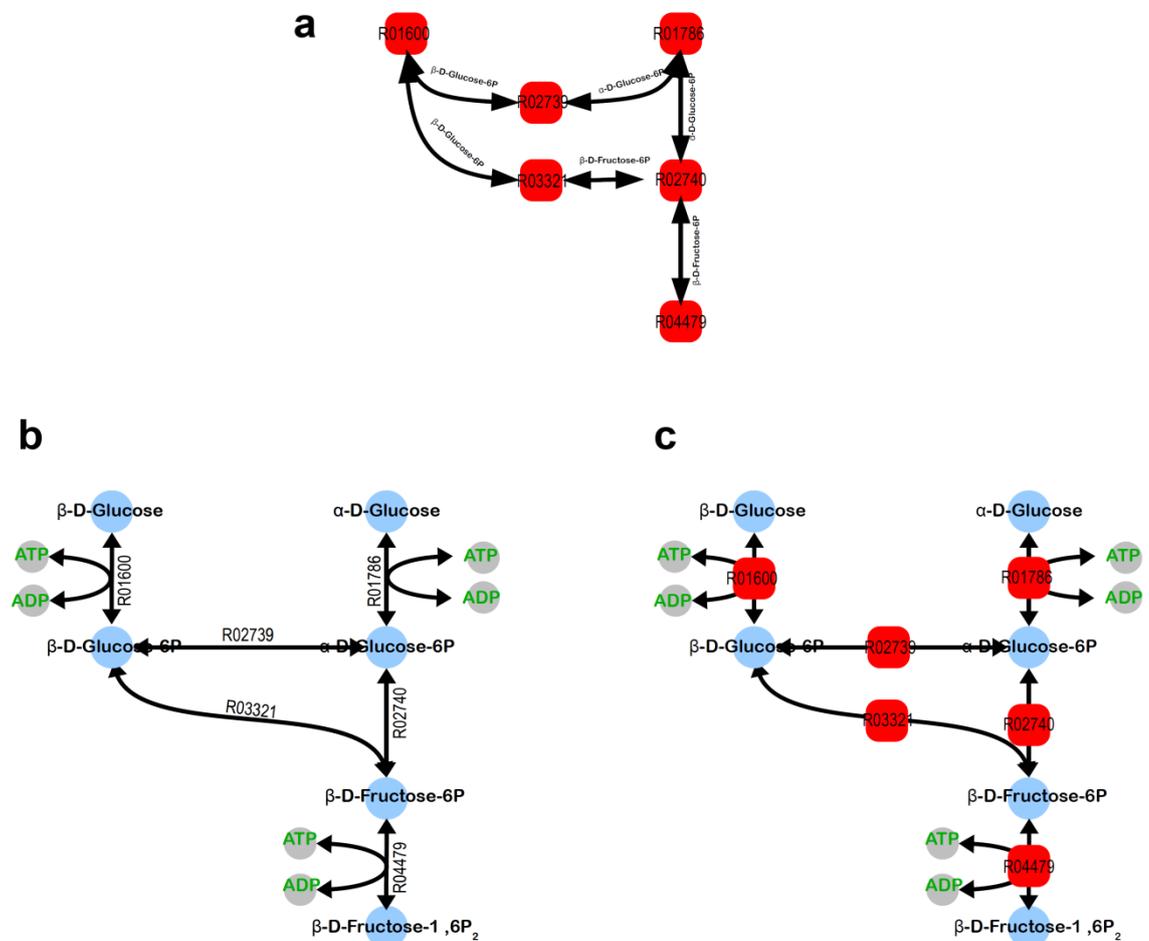


Figure 9: (top/a) Reaction-centric graph. (left/b) Compound-centric graph. (right/c) Combined graph. These are all equivalent representations of the first six reactions in the glycolysis pathway. ATP and ADP nodes are aliased.

Reaction-centric graph

In this type of graph, reactions are placed as nodes which are then connected by edges representing the compounds passed between reactions. In graph theory this is called a simple graph or a directed graph if a preferred direction for the compound flow is specified.

Compound-centric graph

Much more commonly the compounds in a metabolic network are chosen as the nodes connected by reactions shown as edges. Many visual representation of metabolism are compound-centric graphs such as those already shown in figure 3 and figure 7. Even though all reactions are reversible a preferred direction of each reaction is defined to ensure the correct coupling of reactants and products. At first it might be tempting to consider the compound-centric graph a directed graph within graph theory but in fact it is more like the combined graph below but with the reactions nodes omitted and instead used to label a group of edges representing the reaction.

Combined graph

A bipartite directed graph treats both compounds and reactions as nodes, with the edges showing only how compounds connect to each reaction. This is the graph type that most usefully represents a metabolic network although the increased complexity means that graphs like this can be confusing on a large scale.

Mathematical representations of metabolic networks

Visual representations of network are essential to our understanding of them but the true power of representing metabolic networks as graphs is the symmetry that these representations have with much more powerful mathematical representations.

Figure 10 shows us how we can represent a metabolic network mathematically by forming a stoichiometric matrix.

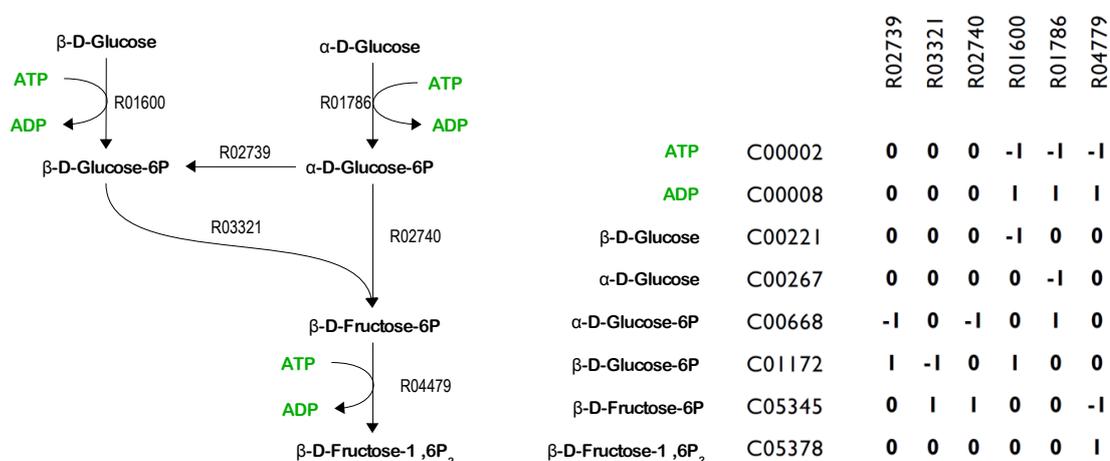


Figure 10: (left) A visual representation of a small reaction network. (right) The equivalent 8×6 stoichiometric matrix (S matrix) that completely defines the network. Reaction IDs and compound IDs and names are added as labels to the matrix.

We can best understand this stoichiometric matrix (S matrix) by inspecting its rows and columns separately.

Let us take as an example the fifth row of the S matrix. We see that $\alpha\text{-D-Glucose-6P}$ (C00668) is not involved in three of the six reactions, a single molecule is consumed by the forward direction of the reactions R02739 and R02740 and a single molecule is consumed by the forward direction of reaction R01786.

Inspecting a row of the S matrix is equivalent to inspecting the edges connected to that compound's node in the compound-centric visual representation shown in Figure 9b.

Likewise, inspecting a single column of the S matrix is equivalent to inspecting the edges connected to a reaction's node in the reaction-centric visual representation shown in Figure 9a.

Completing the symmetry between the rows/columns of the S matrix and the different possible visual representations we see that the combined graph in Figure 9c is equivalent to considering the nodes represented by the rows and columns of the S matrix at the same time.

SBGN: an emerging standard for network visualisation?

In March 2010 I attended the “Systems Biology of Microorganisms” conference at the Institut Pasteur in Paris. At the poster sessions of the meeting I took photos of 28 visual representations of biological networks. It was striking that the information in all 28 forms was conveyed in a completely different way. Whilst it was possible to understand each poster it took a long time to see what was being shown and it was almost impossible to make comparisons between similar systems shown in different ways. I thought that there had to be a better way to represent models and in fact the reason why the leading proposed solution was not yet being used was that it was less than a year old and had not yet been finalised.

The systems biology graphical notation²⁶ — SBGN — was proposed and promoted by much of the same group, led by Kitano and Hucka, that have worked on SBML for the last ten years. SBGN aims to standardise the major elements of the visual representation of metabolic models in the same way that SBML has standardised the major elements of their mathematical description. I think that the SBGN initiative is a worthwhile venture and many people, including myself, have already adopted elements of it.

A visualisation of a biological network is inherently more variable and more open to personal interpretation than the underlying structure of the network. It is also harder to build tools to create and interpret visual networks.

For these reasons I doubt SBGN will achieve the same universal acceptance as SBML. Nevertheless I have already noticed that some of its design elements and conventions are becoming more widely applied and I have adopted many for my own visualisation tool in this thesis. For example, my own visualisation efforts draw heavily on SBGN’s insistence on the use of the combined graph visual style, the use of rounded nodes for compounds and rectangular nodes for reactions, and the insistence that aliased compounds be represented differently to non-aliased compounds.

In time I’m sure that the efforts of the SBGN group will make visual representations of biological networks much clearer than they are now and I suspect that this — rather than the widespread adoption of their precise standard — was always their main aim.

The malaria parasite

Malaria is like no other disease, it isolates communities, people simply don't go there. It violates Adam Smith's precept of freedom of movement as being essential for economic growth. It sits on the people. It protects them from intruders but also denies them access to the world economy. I think that's probably the most important element of economic growth in Africa...

Interview with Prof Andrew Spielman, Professor of Tropical Public Health, Harvard School of Public Health ²⁷.

The background to some of the information in this section comes from the ninth, tenth and eleventh episodes of a podcast by Vincent Racaniello and Dickson Despommier called This Week in Parasitism, the tenth episode in particular is well worth listening to. Much of that material is in turn based upon Despommier's two books on the subject^{28,29}. An exceptional account of mosquito-borne diseases, including large sections on malaria, can be found in the excellent book *Mosquito* by the late Professor Andrew Spielman³⁰ a fascinating interview with whom can be found with the Vega Science Trust²⁷ from which the quote at the start of this section was taken.

Human impact

Malaria is a major cause of human suffering around the world with between one and three million directly attributable deaths each year. Direct deaths are principally the result of infection with *Plasmodium falciparum* and occur primarily in the young and in pregnant mothers. The secondary affects, caused primarily by the temporary inability of affected people to work, and the associated fear of travel and trade, but possibly extended into many other areas such as reduced IQ³¹, impose considerable economic costs and thus hamper projects to reduce malaria infection rates.

Perhaps the most significant long-term impacts of malaria are caused by its rapid evolution and the resulting specificity of acquired immunity. People growing up in a single affected village acquire immunity to the local variant of the malaria parasite with those who survive into their teens continually re-infected but with symptoms usually no worse than mild flu. Such is the variation of the parasite — and the specificity of their acquired immunity — that even travelling a distance on the order of ten miles can result in a debilitating or even deadly infection from a variant to which they have no immunity. In this way, sharing of goods and ideas, and even the establishment and maintenance of nations is cripplingly frustrated by malaria.

Malaria, alongside other mosquito-borne diseases, has critically shaped the course of human history. Malaria was a key part of the failure of the first Panama Canal project by the French and placed a huge burden on Egyptian, Greek and Roman empires. Quinine, the first drug to combat malaria, was first given by, and later stolen from, native South Americans when they came into contact with Europeans and remained an important treatment until over-use in the 20th century made it increasingly useless against resistant parasites. Derivatives of quinine³² that for a while proved more resilient to resistance are now unable to treat malaria in many parts of the world and there are worrying signs³³ that resistance to the last universally effective anti-malarial drug artemisinin may be emerging.

At one time in the 1970s it seemed as if malaria could be eradicated in the developing world as it had been in the developed world. This belief was so prevalent that much research into the parasite was stopped and many Universities greatly reduced the importance of the parasites in their teaching of both microbiologists and doctors. In addition to distributing bed nets, DDT was sprayed to target the mosquito vector whilst anti-malarial drugs were given as soon as symptoms emerged with the aim of both curing the illness and the parasite's spread. Despite early successes the discovery that DDT remained active in ecosystems far longer than predicted, notably in Rachel Carson's book

Silent Spring, and the corresponding reduction in its use marked the end of our last attempt, nearly 40 years ago, to rid the world of malaria.

The characteristics and life-cycle of the parasite

The primary host of parasites in the genus *Plasmodium* are mosquitoes, typically of the genera *Anopheles* or *Culex*. Over two hundred species of *Plasmodium* have been identified to date, with each usually specific to a given non-insect host and at least a specific genus of the insect host. Although further parasite species have not been classified, it is likely that many more mammal, reptile and bird species with territories overlapping a suitable mosquito vector play host to at least one corresponding *Plasmodium* parasite.

Humans are afflicted by five species of *Plasmodium*, most frequently by *falciparum* but also by *vivax*, *ovale*, *malariae*, and occasionally by the long-tailed macaque specific species *knowlesii*. Of these, *P. falciparum* reaches the highest population within the body and causes the most severe symptoms including severe anaemia and unconsciousness: frequently called cerebral malaria. By comparison *P. malariae* has far milder symptoms but can persist asymptotically for decades. *P. vivax* and *P. ovale* are distinguished by the persistence of hypnozoites in the liver stage which can later cause the recurrence of acute symptoms.

The complex life-cycle of *Plasmodium* is best understood in a diagram as shown in figure 11 on page 27 and described below.

Life stages within the human blood stream (erythrocytic stages)

1 – The **merozoite**, regardless of its source, attaches and induces the red blood cell to produce extra membrane and, enclosed in this membrane, enters the red blood cell.

2 – Early-stage trophozoites appear ring-shaped under Giemsa stain³⁴ and are easily recognised by clinicians for diagnosis. This is frequently called the **ring-stage**.

3 – Late-stage **trophozoites** grow to fill the host cell. In infections by *Plasmodium falciparum* these trophozoites induce the production of a histidine-rich protein (Pf HRP II) which causes red blood cells to clump around an infected cell and block capillaries³⁵. This coagulation ultimately causes unconsciousness called cerebral malaria*, making treatment and patient maintenance extremely difficult and thus greatly increasing the risk of death

* Many textbooks still explain that cerebral malaria is caused when clumped red blood cells induce the expression of tumor necrosis factor in the capillary walls. This in turn induces nitric acid production leading to unconsciousness. Newer evidence increasingly suggests that nitric acid frequently protects brain function²¹² in stressed conditions and that it can also protect against cerebral malaria²¹³. Current consensus seems to be shifting towards... (continued overleaf)

4 – Mitotic nuclear division takes place to form a **schizont** full of merozoites ready to enter the blood stream upon cell lysis. Toxic haemozoin — the crystallised accumulation of the haem groups left as a by-product of the parasite's haemoglobin digestion — is left in the lysed blood cells upon lysis.

5 and 6 – **Microgametocytes** (5, male) and **Macrogametocytes** (6, female) form from ring-stage parasites and remain inside un-lysed red blood cells in circulation. They are rarely seen in parasite cultures and the mechanism governing gametogenesis versus schizogony is not fully understood.

Life stages within the mosquito

7, 8 and 9 – Any parasites ingested by a female mosquito taking a blood meal are digested except for microgametocytes and macrogametocytes which are released from their surrounding red blood cell and activated by the mosquito's digestive tract. For the female gamete this activation is simple and the activated macrogametocyte becomes a **macrogamete** (9). At this point the activated microgameteocyte begins a process known as exflagellation, producing a free-living **microgamete** (8) capable of fertilising the **macrogamete** at which point the combined diploid organism is called the **zygote**.

10 – The zygote matures to become the mobile **ookinate** which embeds itself in the mosquito gut wall. Although exact details remain unclear it is likely that the zygote is the only diploid form of *Plasmodium*³⁶ with meiosis occurring during its maturation into the ookinate.

11 – The embedded ookinate matures into the **oocyst**.

12 and 13 – The oocysts divide asexually until individual **sporozoites** are mature.

14 – The **sporozoites** lyse the host cell and circulate within the mosquito until reaching the salivary gland where they settle, ready to be injected when the mosquito takes its next blood meal.

Life stage within the human liver (exoerythrocytic stages)

15 – Sporozoites injected into the human blood stream settle in the liver where they infect hepatocytes. At this stage the parasites are almost invisible to clinicians and are thus referred to as **cryptozoites**. In *P. vivax* and *P. malariae* these liver-stage parasites can become **hypnozoites**, capable of staying asymptotically in the liver for months or years before continuing their life-cycle and causing a recurrence of acute malaria.

cerebral malaria being caused by the breakdown of the blood-brain barrier and the ability of clumped/rosetted red blood cell groups to enter the brain^{214,215}.

16 – **Hepatic schizonts** form, the direct equivalent of the schizonts in red blood cells.

17 – The lysis of hepatic schizonts releases **merozoites** into the blood stream to commence the red-blood cycle.

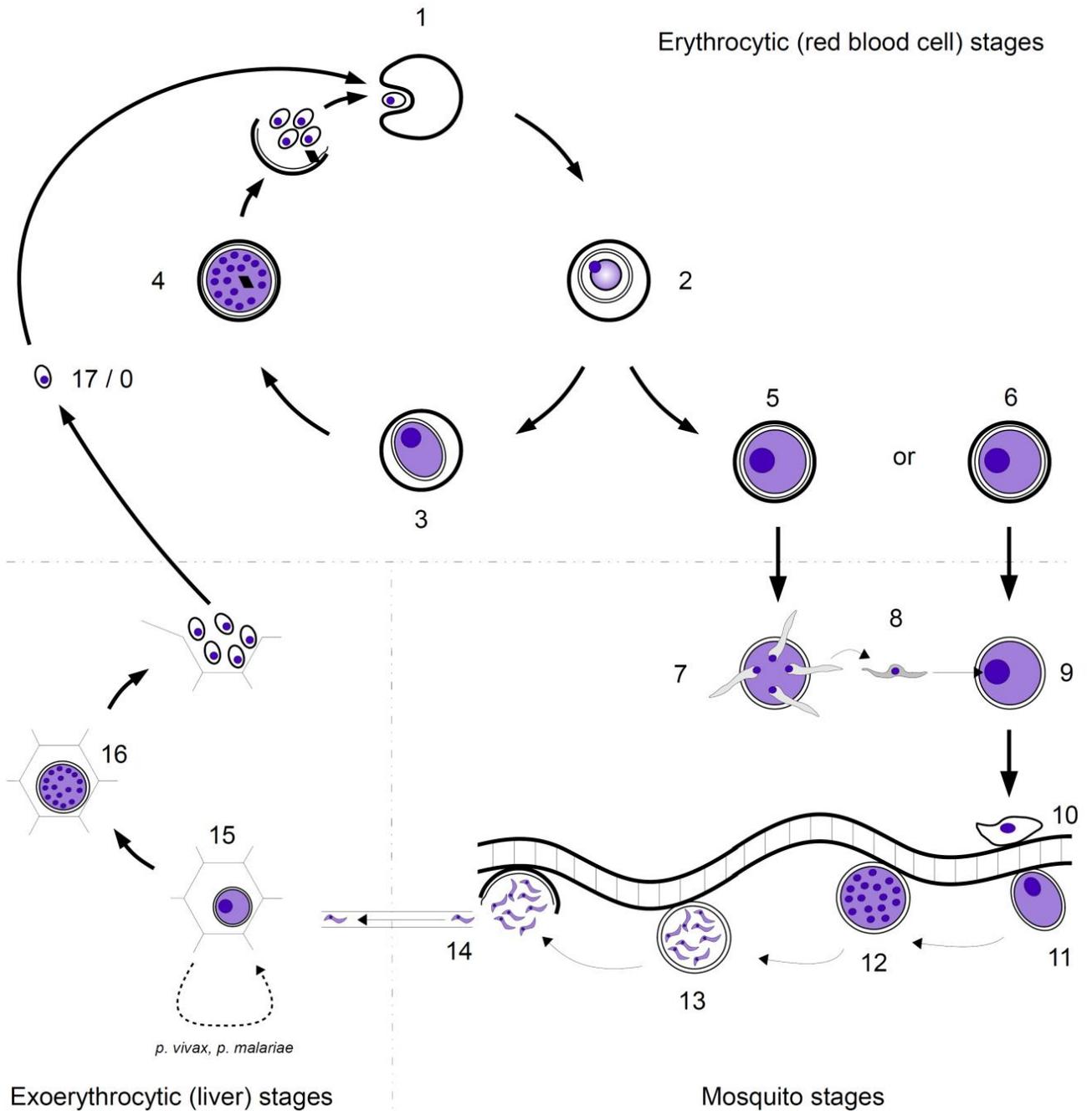


Figure 11 : The multiple life stages and host cells of Plasmodium. Parasite nuclei are shown in dark purple, cytosol as a lighter purple, as under Giemsa stain. Partially adapted from Dr. Wiser's malaria course at Tulane University, FL, USA³⁷.

Methods of study

Cell culture of erythrocytic stages

Human malaria parasites had proved impossible to culture until 1976 when Trager and Jensen realised that the low oxygen environment of coagulated red blood cells (RBCs) within the capillaries was essential to the progression of the *P. falciparum* life-cycle³⁸. Just a year later the technique had advanced to a system we would recognise today whereby frozen parasitised red blood cells could be thawed and grown in washed human RBCs in RPMI medium fortified with human serum all within culture flasks kept with a low oxygen internal environment³⁹. At the time the low oxygen environment was provided by burning a candle at the entrance of the culture flask to lower the oxygen concentration. Modern improvements to these methods have improved reproducibility; low-oxygen gas of known composition replaced candles and a standard culture medium of RPMI+AlbuMAX has largely replaced the variable composition of human serum⁴⁰.

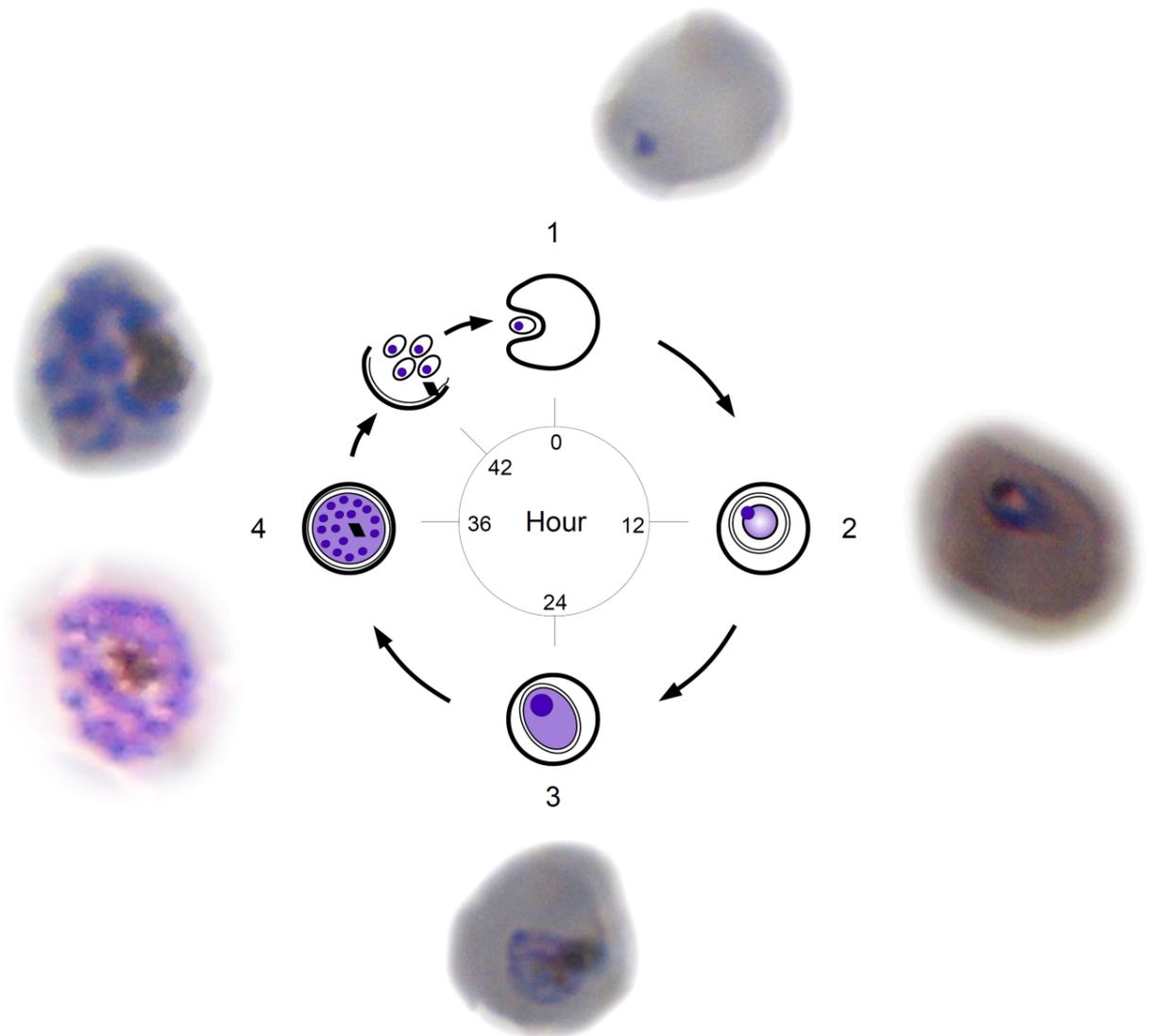


Figure 12 : The life-cycle red blood cell stage Plasmodium parasites with representative images of Giemsa stained infected red blood cells on thin blood smears. All images are my own.

The techniques I have used to manipulate *P. falciparum* cultures are detailed in the experimental methods chapter of this thesis. Additional methods of study that I have not conducted but which are worthy of a brief mention are,

Cell culture of liver stages

Limited success in cultivating liver stages of *Plasmodium vivax* and *malariae* to study their ability to persist as hyponozoites has been achieved^{41,42}. Furthermore, recent advances with *Plasmodium falciparum* are making it possible to culture the liver stages of the parasite in human liver cell lines⁴³. These techniques are currently very difficult but may offer a way to study the organism's metabolism more fully and may allow the development of new types of drugs, both prophylactic and as treatment, that target the liver-stage of the parasite's life cycle.

Other in vitro methods of study

Studies into *Plasmodium* continue in bird species, rats (*berghei*) and long-tailed macaques (*knowlesii*) as well as in the mosquito primary host and may provide valuable insights into human malaria. It is also possible to study the human species *in vivo* by infecting splenectomised, and thus immunodeficient, chimpanzees although these studies are both extremely expensive and ethically questionable.

Unique challenges to the study of *P. falciparum*

Although it is now relatively straightforward to culture erythrocytic life stages of *P. falciparum*, it remains more difficult than the culture of many other pathogens. Daily attention to cultures is usually required and unexplained culture death is not unusual. Furthermore, since only a third of the parasite's life cycle occurs in human RBCs the current life-cycle coverage of *in vitro* study remains low.

Although some initial efforts at gene silencing through RNA interference appeared successful in reducing growth-rate⁴⁴ further studies have shown that *P. falciparum* lacks the genetic machinery required for general application of RNAi⁴⁵.

The intracellular nature of the parasite means that direct manipulation typically involves lysing the host RBCs and conducting experiments on freed and thus dying parasites outside of their normal environment⁴⁶.

The genome of *P. falciparum* was published in 2002⁴⁷ and many hoped it would simplify the study of the parasite. By building a metabolic network of *P. falciparum*, based principally on its genome, and integrating it sufficiently with supporting experimental data we might reasonably expect to construct a model that would reproduce the inner

workings of the organism. Unfortunately the genome⁴⁷ presented many problems. Assembling the fragments from shotgun sequencing had been hard because the genome was known to be A-T rich but at 80.6% A-T, with around 90% A-T in introns and intergenic regions. The genome remains to my knowledge the most A-T rich genome ever sequenced. This unusual nucleotide composition presents further difficulties. Homology searches find fewer genes than expected making metabolic network reconstruction difficult and very few restriction enzymes are specific enough to perform genetic modifications on the genome.

A final impediment to creating an accurate metabolic reconstruction of the parasite is a result of it being an intracellular parasite and thus able to rely on its host, and even induce its host, to fill gaps in its own metabolism. Homology searches of the parasite genome will not identify these genes and the gaps must be filled with knowledge of red blood cell (RBC) metabolism.

Network analysis

We've got the human genome sequenced, we've got the parasite genome sequenced, we've got the mosquito genome sequenced. Somehow you should be able to work out some way of controlling everything...

Dickson Despommier, TWiP Episode 10

Flux-balance analysis is the primary metabolic analysis technique I will use in this thesis. It is well described in a book on the subject by Bernhard Palsson⁴⁸ — one of the early users of the technique — and in four papers charting the development of the technique over the last fifteen years⁴⁹⁻⁵² most recently in a summary published by J. Orth in 2010⁵³.

A note on Wikipedia

In 2010 the clearest academically published description of FBA was published in *Nature Biotechnology*⁵³ and, despite adding little to the field, was cited nearly fifty times in the following year. At this time, in frustration at the lack of an openly available, linkable and updatable resource describing FBA I decided to write what remains the current Wikipedia page on the subject.

The work on that page, including all figures, is almost completely my own and has been reviewed, graded and commented on by notable academics in the field, including Kieran Smallbone at the University of Manchester. Some sections are repeated in the following pages without citation because they were written by me.

In the eight years since I started as an undergraduate I have had numerous candid, and usually negative, discussions about Wikipedia. These have ranged from markers failing my lab reports for no other reason than they referenced Wikipedia, through lecturers instructing me to mark as errors in lab reports any reference to Wikipedia.. I have listened to, considered and engaged in the debates on whether knowledge should be open or correct and ultimately concluded that we can have both. I do not think that the current “solution” to the phenomenon of free and open information: rephrasing Wikipedia text and then — without reading the source — citing that text’s reference, is in interest of scientific advance or education.

For this reason I will in this section focus on the basics of FBA, in addition to the requirements, scope and limitations of the technique. The version online is both more complete and will already be more current than the limited version I reproduce here. It is reflective of the approach that academic institutions have taken to Wikipedia that in all the written guidance* given to PhD students and supervisors at the University of Leeds, the word does not occur once. I hope that my use of, and contribution to, this resource is in keeping with the requirements of the University of Leeds.

* Research Student Handbook 2010-2011, Ordinance and Regulations and Programmes of Study for Research Degrees 2010-2011, Guide for Research Degree Supervisors 2010-11, Guidance on Thesis Preparation, Copyright and Publication, General Academic Regulations.

Recognising the limitations of metabolic network reconstructions

Armed with the genome of an organism it is a relatively simple task to search for genes homologous to those in reference databases such as refseq⁵⁴. More advanced software such as SharkHunt⁵⁵ and RAST⁵⁶ can make this process even easier and more effective.

Yet no matter what technique is employed to find genes, and thus reactions, describing an organism's metabolism there are still significant limitations on these automated techniques. Just four of these limitations are,

- We haven't found all the enzymes required to build a complete set of reference sequences and we cannot search a genome for enzymes we don't yet know exist.
- We don't know the properties of most enzymes or their location or concentration within their host organism.
- We cannot predict the effect that small changes in mostly homologous proteins may have on their properties.
- We cannot predict how an enzyme's properties will change in a different organism.

Steps can be taken to improve the quality of a metabolic network by filling gaps in the automated annotation and I will discuss those in detail later. For now let's deal with another approach to dealing with incomplete metabolic networks; making simplifications when analysing them.

Techniques for the analysis of metabolic networks can be split as by Poolman *et al.*⁵⁷ into two groups: structural and kinetic. To this I add a further type, local analysis, which was considered too basic to be included in Poolman *et al.* and is arguably not true network analysis.

Local analysis (chokepoint analysis)

The first published metabolic network reconstruction and analysis for *P. falciparum* was conducted by Yeh *et al.*⁷ in 2004. Their core technique was chokepoint analysis where they define a chokepoint as "a reaction that either uniquely consumes a specific substrate or uniquely produces a specific product" and finding that 21 out of 24 known drug targets (enzymes coding for reactions disrupted by a drug) were chokepoints.

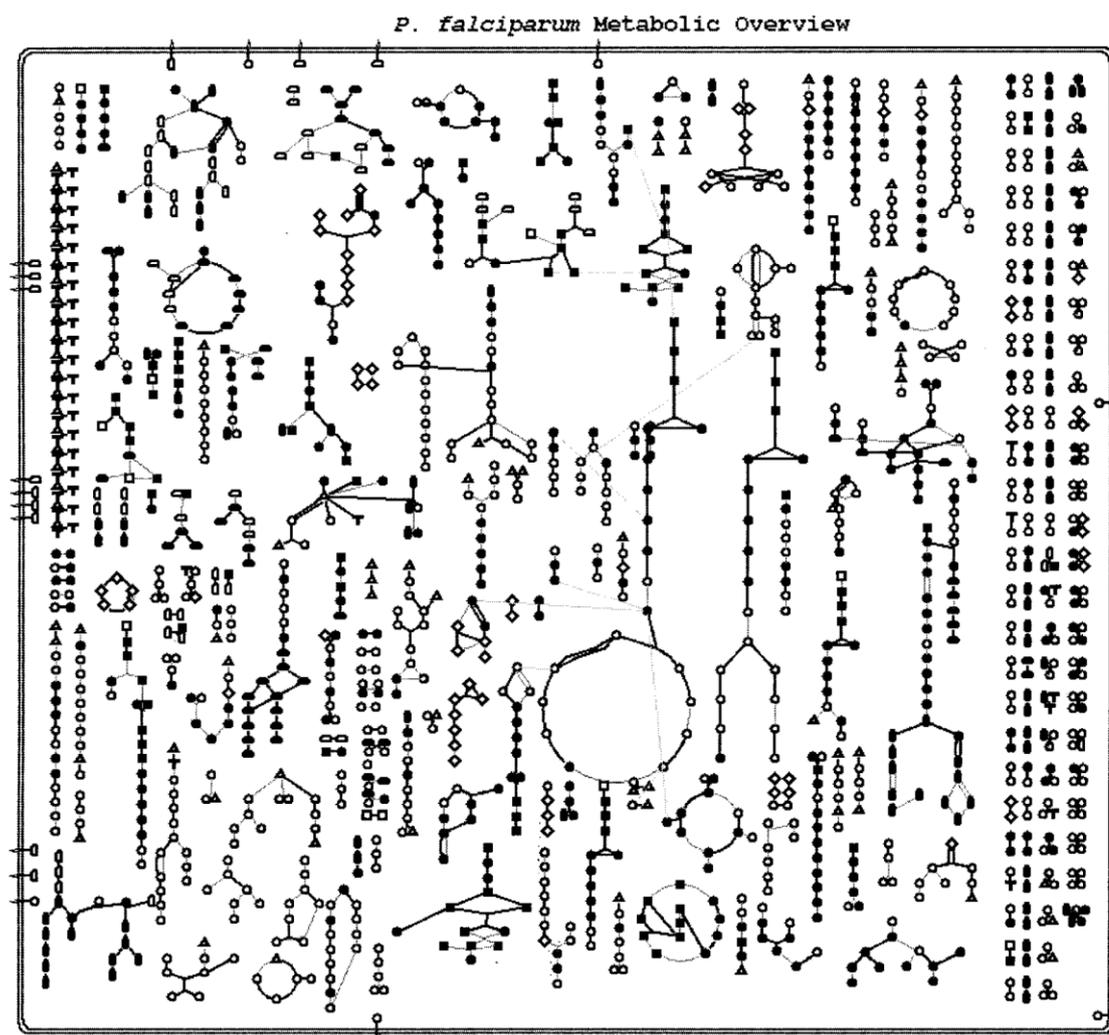


Figure 13: Cellular overview of *P. falciparum* network used for chokepoint analysis from Fig 2 in Yeh et al⁷. Reactions (edges) shown in grey are not present in the model.

Inspecting an overview of the reconstructed network, figure 13, to which chokepoint analysis was applied it becomes clear why the technique was adopted. Because of the large number of gaps in the reconstructed network an analysis technique that considered the flow of metabolites through the network could not be used. Since chokepoint analysis considers each reaction in turn it can deal with this problem with the associated downside that every reaction before and after a gap in the network is by definition a chokepoint.

In fact in the analysis of this basic network 216 out of 303 (71%) core metabolic enzymes were found to catalyse reactions that were chokepoints making it unsurprising that 21 out of 24 (87.5%) known drug targets were also chokepoints. The complexity of the network description — typical of a reconstruction in Pathway Tools — and the many reactions excluded from the chokepoint analysis ultimately make the results of this paper difficult to quantify and utility of chokepoint analysis unclear.

Structural analysis

Structural analysis differs from local analysis in two connected ways. Firstly, it examines reactions not in isolation but rather as a collection. Secondly — and partly as an extension of the first point — it operates not only on the metabolic network of a system but also on the interaction between that system and its surroundings; a key part of structural analysis is that there is some input and some output. In the examples of glycolysis shown back in figure 3 these are as simple as β -D-Glucose-6P as an input and pyruvate as an output. On the whole-organism scale, as represented in figure 14, the inputs are the metabolites absorbed from the environment and the outputs are both the metabolites excreted back to the environment and the metabolites used for growth and reproduction.

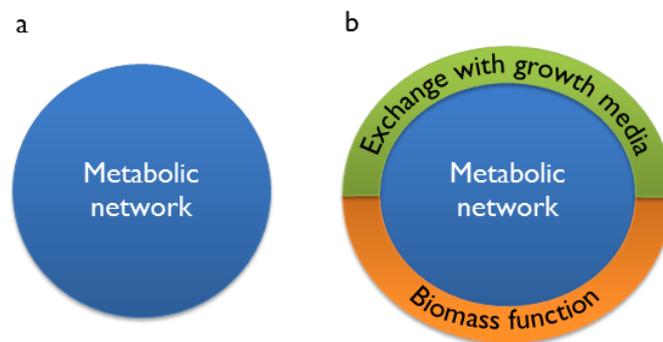


Figure 14: **a.** A metabolic network on its own cannot simulate life. **b.** A metabolic network that is able to exchange metabolites with a growth medium and provide molecules for growth as part of a biomass function can.

Because structural analysis requires inputs and outputs it invokes the concept of metabolites flowing through the network from input to output and for this reason it cannot operate on networks with key missing reactions. This requirement makes networks more difficult to prepare for structural analysis than for local analysis.

Flux-balance analysis (FBA)

A key form of structural analysis is FBA, which can be explained most concisely in mathematics as the solution to the flux-balance problem, where \mathbf{S} is the stoichiometric matrix and \vec{v} is the vector describing the collection of fluxes through the reactions in the network.

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

The result of multiplying the stoichiometric matrix with a vector of fluxes through each reaction in this way is a vector representing the rate of change in the concentration of all the compounds within the system. The condition that this is equal to zero is equivalent to the homeostatic condition, that the internal concentrations of compounds within a living organism stay constant over time.

Additional constraints on the minimum and maximum fluxes, either from thermodynamic considerations or experimental results, can be defined to ensure the model is realistic.

$$v_{\min_i} < v_i < v_{\max_i}$$

The constrained model, which will still usually have a large null space, is then solved to find the optimal solution according to some sensible optimisation goal. Typically this is biomass production, a single reaction, v_b , representing the consumption of metabolites in the correct proportion to represent growth of the organism.

$$\max_{\vec{v}} v_b \quad s. t. \quad \mathbf{S} \cdot \vec{v} = 0$$

Whilst the mathematical description of FBA is accurate it can be hard to comprehend and a purely mathematical understanding of the technique does not express the impressive ability of the technique to make insights into metabolism. Thankfully, as I showed earlier, mathematical and visual representations of metabolic networks can be completely equivalent and FBA can be best explained visually.

In figure 15 we see the process of preparing a metabolic network, in this case the first six reactions of the glycolysis pathway, for FBA. The process involves adding reactions to allow α -D-glucose, β -D-glucose, ATP and ADP to enter and leave the system from an infinite external pool. A biomass/objective function is added to remove β -D-Fructose-1-6P₂ from the system with this removal simulating growth in this highly simplified example.

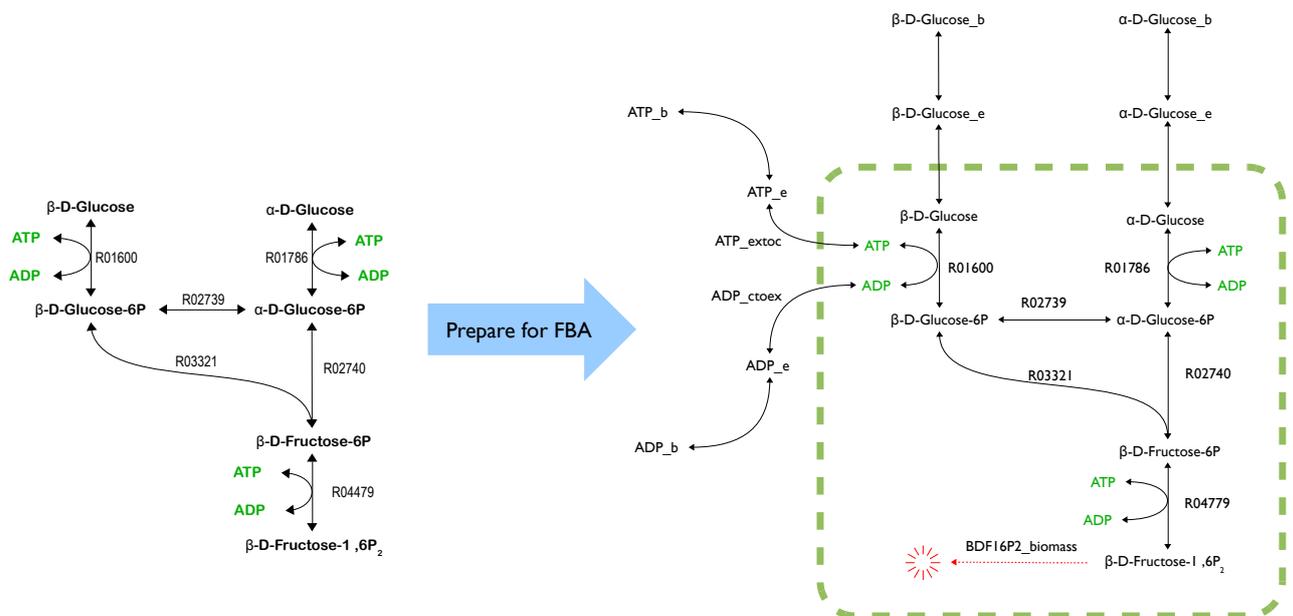


Figure 15: The first six reactions in glycolysis prepared for FBA through the addition of an objective function (red) and the import and export of nutrients (ATP, ADP, BDG, ADG) across the system boundary (dashed green line). The full version* shows the stoichiometric matrix both before and after the model is prepared for FBA.

* http://en.wikipedia.org/wiki/File:Prepare_for_FBA_maths.png

The result of performing FBA on the network shown in Figure 15 with constraints on the maximum uptake of glucose at 50 arbitrary units are shown both as a vector and represented pictorially in figure 16. In this representation the width of the edges representing each reaction is proportional to the calculated flux through that reaction. Applied constraints are represented as red bars limiting the flux on the entry of α -D-glucose and β -D-glucose to the system.

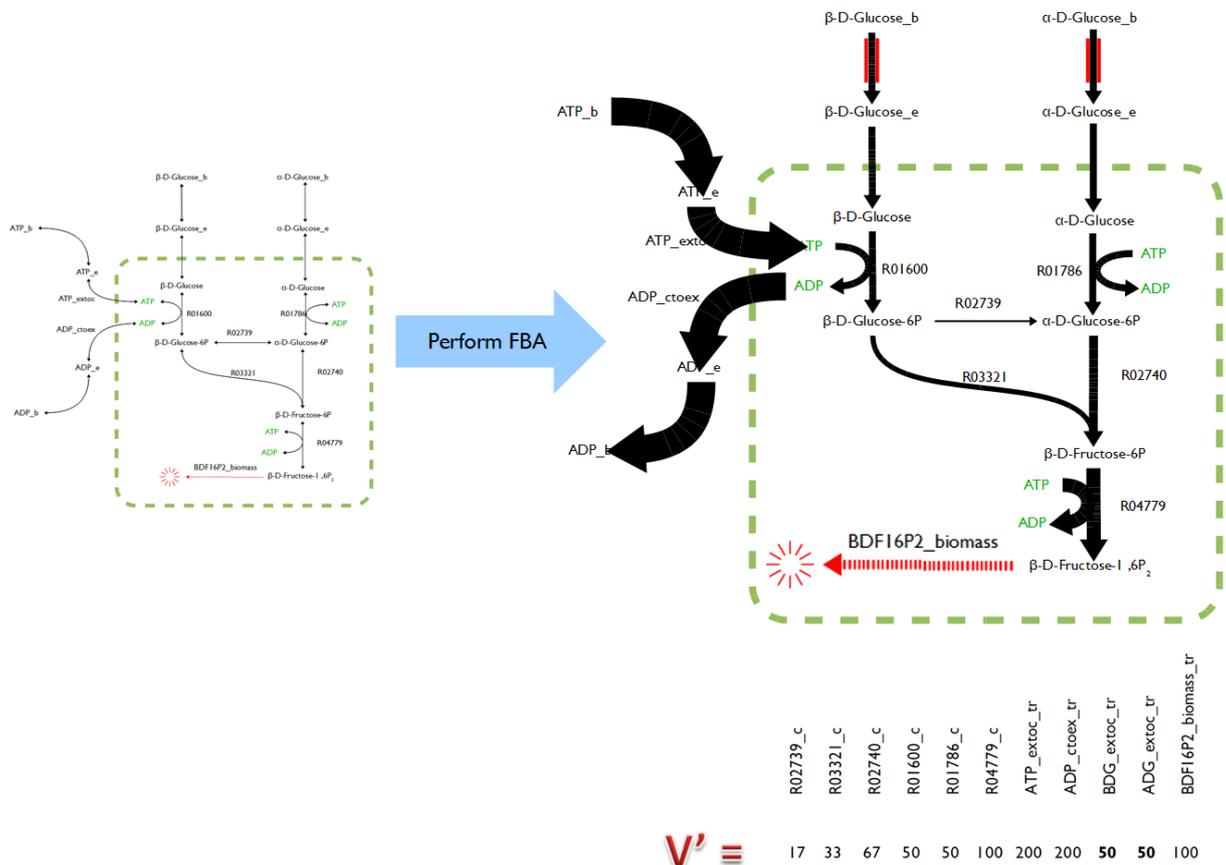


Figure 16 : A metabolic network before (left) and after (right) flux-balance analysis. The 'after' picture conveys the optimal flux from the nutrient source to the biomass function as the weight of the lines representing each reaction. These weights correspond to the reaction flux vector, v , shown below in its transposed orientation v' . A constraint on the glucose uptake flux of 50 arbitrary units is shown as a pair of red bars constraining flux through these reactions.

There is nothing surprising in this result and indeed it, or something similar, could be drawn without calculation by anyone familiar with the technique. FBA is important because it easily scales to large networks where human intuition on metabolite flows is useless. The efficiency of the technique is also a major advantage with analysis of a 1000 reaction network taking under a second on a desktop computer. This in turn allows further techniques such as studying the effect on network fluxes of simulated deletions of all single, and all pairs of, reactions to be carried out within minutes.

Flux-balance analysis of whole organism metabolic networks of *P. falciparum* have been used to predict essential reactions and thus drug targets^{58,59} and I will discuss these results in more detail later.

Elementary modes analysis

The solution to the flux-balance problem, the null-space of the stoichiometric matrix, is not singular. All solutions are made up of some linear combination of a set of basic solutions we call elementary modes which I have drawn for the example system used so far in figure 17.

Of these, solutions 1-4 that are valid solutions to the problem $\mathbf{S} \cdot \vec{v} = 0$ whilst also satisfying the objective function to produce biomass. Solutions 6 and 7 are valid solutions to $\mathbf{S} \cdot \vec{v} = 0$ but do not satisfy the objection function. Solution 5, is an example of an internal cycle which although a valid solution to the problem $\mathbf{S} \cdot \vec{v} = 0$ is of limited informative value and can be confusing when interpreting results.

Elementary modes are important because they help us understand how FBA works but also because they help us find and eliminate unwanted features of a network. The type of elementary mode that we call internal cycles are almost never biologically representative and can be eliminated by enforcing constraints on some of the components. For example, enforcing a forward direction on reactions R03321 and R02740 eliminates this cycle and is a sensible thermodynamic constraint to add to the model in this system. Beyond being of interest in their own right, elementary modes are key components of further analytic techniques such as extreme pathways⁶⁰ and minimal flux modes⁶¹.

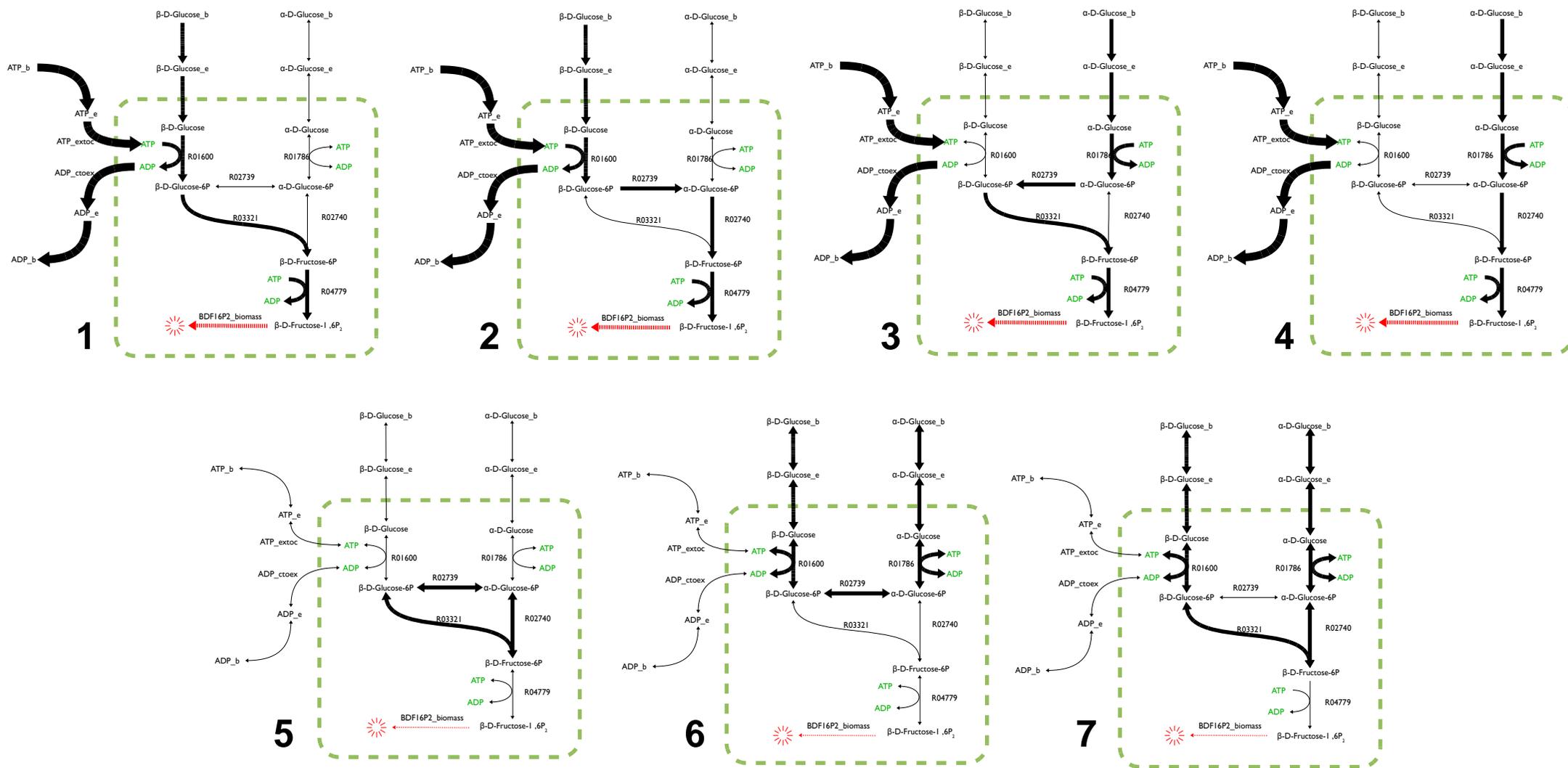


Figure 17 : The seven elementary modes of the example system. Solutions 5,6 and 7 are equally valid in both directions.

Kinetic modelling

Structural modelling techniques like FBA are an advantage over local modelling techniques like chokepoint analysis because they consider the whole network and thus interactions between all parts of the network. This advantage brings with it the associated disadvantage of requiring a network with no gaps in the chain of reactions between nutrient uptake and the production of compounds required for growth.

Kinetic modelling is an advantage over structural modelling techniques because it considers how the concentrations of compounds within a system evolve over time but requires a network even more complete than that required for structural analysis.

The key of FBA is that compounds are assumed to remain at a constant concentration within the organism. We wrote this earlier as $\mathbf{S} \cdot \vec{v} = 0$ with the right-hand side of the equation representing a zero vector of equal dimension to the number of unique compounds in the system. In kinetic modelling we no longer make the constant concentration assumption and this zero vector has in each position a differential equation describing the evolution of that compound's concentration with time.

$$\mathbf{S} \cdot \vec{v} = \begin{pmatrix} \frac{d[C]_1}{dt} \\ \frac{d[C]_2}{dt} \\ \vdots \\ \frac{d[C]_i}{dt} \end{pmatrix}$$

The exact form of these differential equations may be simplified but would most realistically consist of linked Michaelis-Menten kinetic equations with the metabolic products of one equation being the substrate of one of more other equations. These formulae require not only a good guess of each compound's initial concentration but also knowledge of the kinetic parameters for each enzyme. This is particularly difficult because the kinetic properties of enzymes are poorly predicted from homology and vary considerably in different conditions.

Once the metabolic model has been expanded to contain these required details solutions can be found by solving the fuller problem using software such as COPASI⁶².

To date kinetic modelling has been used mostly to study small subsystems within organisms. Attempts have been made on a larger scale and in 2010 an extremely impressive study of *E. coli* showed that kinetic modelling was able to reproduce many

complex behaviours of bacterial metabolism that had previously been thought to require sensing and regulation at a higher level⁶³.

Kinetic modelling can offer a more accurate description of an organism's metabolism than FBA but at the same time it is substantially more difficult. This difficulty means there is a larger scope for errors when constructing kinetic models and finding errors can take a long time. Furthermore, large problems can become computationally expensive and the resulting systems are frequently unstable and give results that are difficult to trust.

Kinetic modelling is a technique best applied either to small, simple or extremely well studied systems and I do not think it is currently suitable for malaria research.

Extensions of FBA

Many extensions have been built on top of FBA aimed at more deeply analysing the system being studied or attempting to mediate the limitations of the technique. Of these I will discuss the three of most interest to me and this thesis, flux-variability analysis (which I perform later), dynamic FBA (which I have not performed) and attempts to model the regulation of metabolism (which I discuss again in the final chapter as a possible extension of my work).

Flux variability analysis

The optimal solution to the flux-balance problem is rarely unique, with many possible and equally optimal, solutions existing. Flux variability analysis (FVA), as described by Mahadevan and Schilling⁶⁴, returns the boundaries for the fluxes through each reaction that can, when combined with the right combination of other fluxes, produce an optimal solution.

An example of the result of FBA performed on the top six reactions of the glycolysis pathway as shown in previous figures is shown in figure 18.

Reactions which can support a low variability of fluxes are likely to be of a higher importance to an organism and FVA is a promising technique for the identification of reactions that are highly important despite being non-essential. In this example we see that the essential reactions, R04779, ADP exchange and ATP exchange have no associated variability. Furthermore, once R01786 has been deleted we see that other reactions, notably R01600, that previously had an associated variability now have none, indicating the essentiality of these reactions within the new reduced network.

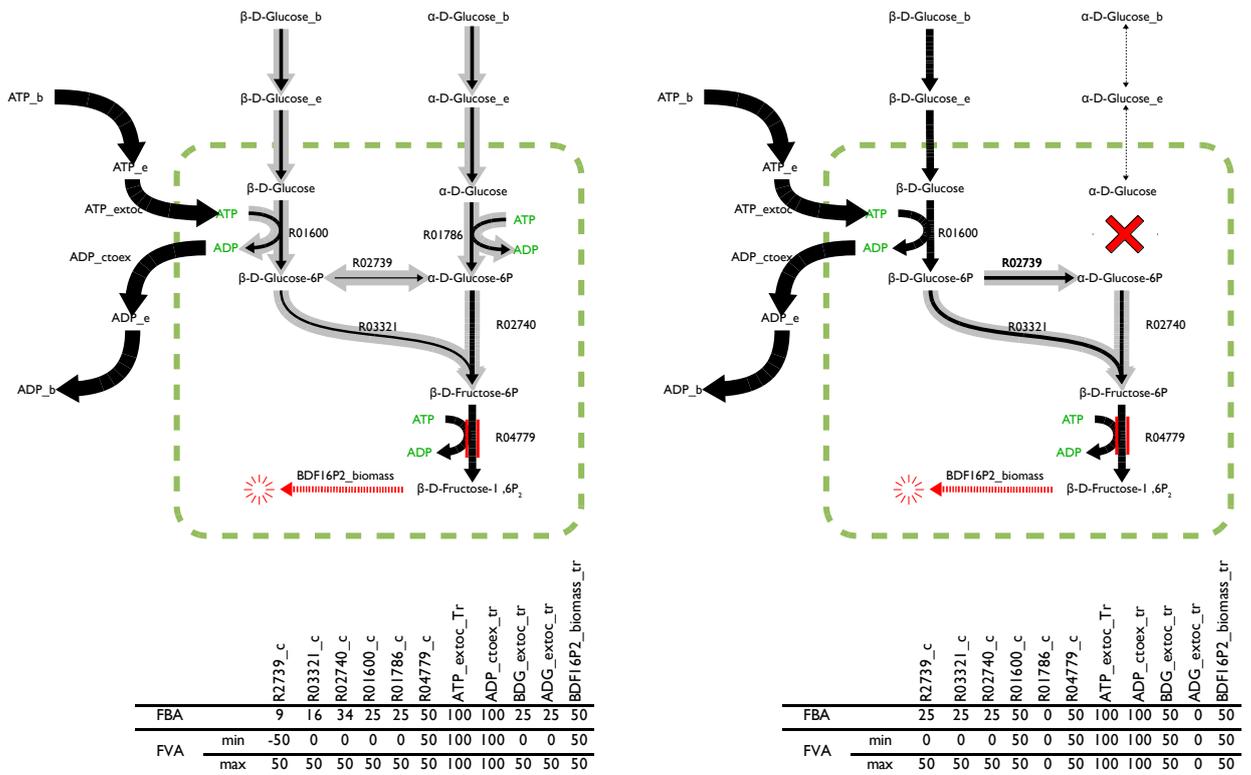


Figure 18 : Visual representation, with FBA and FVA solution vectors below, of flux-variability analysis (FVA) performed on the top six reactions of glycolysis. R03321 and R02740 were only allowed to run forwards to avoid an internal cycle. Constraint on R04779 of 50 units as shown places an upper limit on biomass production. Variability of fluxes (in light grey) shown at four times width with chevrons noting possible direction of flow. (left) Full network. (right) With non-lethal deletion of R02740.

Dynamic FBA

Dynamic FBA, as first described in Varma *et al.* 1994⁶⁵ and later expanded in Mahadevan *et al.* 2002⁶⁶, attempts to add the ability for models to change over time, thus avoiding the strict homeostatic condition of pure FBA but without requiring kinetic parameters for each reaction. Typically the technique involves running an FBA simulation, changing the model based on the outputs of that simulation, and rerunning the simulation. By repeating this process an element of feedback is achieved over time. I have not performed dynamic FBA in this thesis but it is a technique that I think could be applied to the model I have ultimately produced.

Regulation of metabolism

Finally, let me introduce one of the most interesting directions in which FBA is moving and a key area where improvements in its predictive power can be made. I mentioned at the beginning of this chapter that life is a controlled network of chemical reactions and yet I have barely discussed the element of control. Metabolism is not a fixed process and organisms must adapt to their changing surroundings or changing roles by changing how they live.

The ways that metabolism adapts to different conditions are varied. At the simplest conceptual level allosteric feedback loops within networks can increase or decrease flux through certain reactions. Referring back to Figure 18 we can imagine that if β -D-Fructose-6P bound to the enzyme catalysing reaction R01600 in such a way as to reduce the enzyme's effectiveness the network would limit its own use of β -D-Glucose and thus the organism's growth-rate. We can similarly imagine that β -D-Fructose-6P might bind to the enzyme catalysing reaction R03321 in such a way as to increase the enzyme's effectiveness and the network would increase the flux through that reaction at the expense of R02739 and R02740.

Complex networks of allosteric interactions can produce systems that adapt extremely well to changes in conditions, often frustrating attempts to disrupt metabolic networks with pharmaceuticals. Similarly — as we can see from the complexities of drug design — the effects of small molecules on enzymes are extremely difficult to predict and the huge diversity of chemistry combined with the variability of enzymes makes experimental screening of libraries of small molecules against target enzymes extremely expensive and often unsuccessful. This extremely difficult problem means that exhaustive lists of the small molecules and enzymes involved in allosteric regulation of metabolism are not available for anything larger than small systems within model organisms and even then are subject to substantial changes caused by small changes in conditions.

A more promising avenue for understanding the regulation of metabolism is via the more conceptually difficult but experimentally more accessible mechanism of gene regulation. Within the GPR relationship I have described, it follows logically that if the gene is not expressed, the protein is not synthesised and the reaction does not occur. Thus by turning genes on and off an organism can drastically change its metabolism. The mechanism of this gene regulation and the adaptability it allows is explained extremely well in Nessa Carey's book, "The Epigenetics Revolution"⁶⁷ but a brief explanation can give an idea of how we can improve FBA with this knowledge.

Gene expression data is primarily gathered in one of two ways. The more established technique is to use a micro-array with different DNA molecules (probes) attached at different grid positions. The contents of a growing cell are extracted and the mRNA isolated and reverse transcribed to cDNA, and then placed on the micro-array. cDNA sequences with sufficient complementarity to a probe bind to it, with the amount of cDNA bound to each probe signified by the intensity of the fluorescence at that position. The advent of much cheaper next-generation sequencing techniques now offers us a far more powerful way to measure the transcriptome (the set and quantity of transcribed genes), with the cDNA strands created by reverse transcription of the organism's mRNA and

sequenced directly. This technique, called RNA-seq, has the advantage of giving information for the whole transcriptome rather than just the parts for which probes were designed and placed on the micro-array. In addition to these techniques for measuring the transcriptome, techniques for measuring the collection of proteins, often called the proteome, exist. Methods based around mass spectrometry such as peptide mass fingerprinting are particularly useful when studying eukaryotes where transcribed genes and translated proteins are subject to a variety of post-transcriptional and post-translation modifications.

Whatever the method used to measure the transcriptome or the proteome the obvious applications to the study of metabolism are similar. By comparing the transcriptome or the proteome in different conditions we can see the set of reactions that are active within the organism in each condition. This approach extends to further studies of variation, for example seeing active metabolic reactions in different human cell types, seeing active metabolic reactions in different *P. falciparum* life stages or seeing how *P. falciparum* responds to known inhibitors⁶⁸.

Several transcriptome and proteome studies⁶⁹⁻⁷³ have measured which genes are transcribed or which proteins are present at different life stages in *P. falciparum* with the most notable probably being Bozech *et al.*⁷⁴. The work on *P. falciparum* by Huthmacher *et al.*⁵⁹ integrates this life-stage specific gene-expression/transcriptome data into their flux-balance analysis to simulate metabolism and predict metabolic fluxes at different life stages. This integration is achieved* by setting each gene to either “expressed” or “not expressed” depending on whether the gene-expression is above or below a defined threshold. The flux optimisation problem is then changed such that it seeks to maximise biomass production at each stage whilst minimising contradictions between whether flux flows through a reaction catalysed by a gene that is not expressed or does not flow through a reaction catalysed by a gene that is expressed. The results of this approach are interesting, with some evidence that the predicted metabolic variations at different life stages accurately represent changes in metabolism as measured by known exchanges of metabolites between the parasite and the RBC host. Despite this success, I think that this type of model is often more explanatory than it is predictive and the last extension of FBA I want to mention is an attempt to improve upon that.

The variable expression of the genes that largely determines the state of metabolism is itself governed by a network of molecules, usually proteins, called transcription factors which affect which genes are transcribed. Individual transcription factors increase or

* The last page of their paper explains this more precisely.

reduce the expression of a number of genes, often with related roles. For highly curated metabolic models such as *E. coli* and *S. cerevisiae* these gene-transcription factor networks have been painstakingly reconstructed and can be used to predict the effects of perturbations to the network, either in growth conditions of the organism or as a result of a knock-out of metabolic genes or transcription factors. The probabilistic regulation of metabolism (PROM) model proposed by Chandrasekaran and Price⁷⁵ is better able to predict the phenotype in response to varied perturbations of *E. coli* and *Mycobacterium tuberculosis* than other models by moving beyond the Boolean (express/not-expressed) gene regulation model in Huthmacher *et al.* towards a probabilistic model of gene regulation where the expression of a gene has a certain probability of being linked to a transcription factor. In this system the penalty paid for allowing flux to flow through a reaction linked to a non-expressed gene (and vice-versa) is proportional to the probability it was not expressed.

The collection of multi-layered and highly connected networks that explain the regulation of metabolic networks is hard to understand but it also gives good results as reported in Chandrasekaran and Price.

I'll finish this chapter by mentioning an example of where this multi-layered system of regulation explains a complex system very beautifully as reported in Kotte *et al.*⁷⁶ for *E. coli* metabolism. The paper shows that the ability of *E. coli* to sense and adapt to different carbon sources (acetate and glucose) can be explained, without specific sensing of the carbon sources, using a metabolic model of just "4 TFs, 17 transcriptional regulations and 28 enzymatic regulations". The achievements in that paper are beyond the scope of this thesis but I hope that it shows what will one day be possible in other organisms and eventually in *P. falciparum*.

MetNetMaker

MetNetMaker is a great program with a user-friendly interface that allows non-programming-literate people like me to enter the world of SBML.

Melinda Griffiths, MetNetMaker user creating SBML models of algae metabolism at the University of Cape Town

A paper on the subject of the software described in this section was published in Bioinformatics in 2010 as follows,

T. Forth, G. A. McConkey, D. R. Westhead. MetNetMaker: A free and open-source tool for the creation of novel metabolic networks in SBML format. *Bioinformatics (Oxford, England)*. 2010:2-3. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/20671147>.

The software can be downloaded at www.metnetmaker.com — both packaged as an installer with the Microsoft Access Runtime or as a standalone *.accdr runtime database (changeable to *.accdb to open with the full version of Access) — where tutorial videos are available and some frequently asked questions are answered. The tutorial videos and examples show all the key features described in this chapter and replace formal documentation.

Lessons learned from initial reconstruction attempts

My first attempt at building an SBML model of *P. falciparum* metabolism used results from an automated annotation of the raw genome by the SharkHunt⁷⁷ software previously developed in the group. This first effort selected those EC numbers returned by SharkHunt with a high confidence — which I defined as those with an E-value $< 10^{-30}$ — and looked up the KEGG reaction most closely linked to that EC number using the KEGG web API.

My software was written completely in Perl and output an SBML file containing the list of reactions as defined in the KEGG LIGAND ontology. Whilst this software produced a valid SBML file, the network contained a lot of holes (gaps in an otherwise complete series of reactions) and orphans (individual or small groups of reactions not linked to other reactions) and had a number of shortcomings that would not be present in a complete network. Some of these shortcomings were as follows.

- Reactions were not defined within pathways and so were difficult to link together.
- Pathways known from other sources to exist fully had missing reactions and orphaned reactions were common. Despite their frequency these problem were hard to spot as there was no way to see the network.
- The model was not compartmentalised and all reactions occurred in the cytosol of the parasite.
- The network could not be changed or added to except by direct manipulation of the SBML file.
- Information contained in the literature and in other manual annotations was not integrated into the network.

Overall then, my first attempt to create a metabolic network was a useful introduction to SBML, the KEGG database and the project in general but was far too simplistic a base on which to build a whole-network reconstruction. It was clear that I would need to use — or build — a more powerful system for tracking my metabolic network reconstruction and building it piece by piece.

Evaluation of existing solutions

The paucity of existing software for metabolic network reconstruction was summed up in “A protocol for generating a high-quality genome-scale metabolic reconstruction”⁷⁸ published in 2010 by Thiele & Palsson in which they list only Simpheny⁷⁹ as a suitable “reconstruction software package”. Whilst their assessment is both slightly biased and overly pessimistic, it is probably not far from the truth.

In the search for another option I spent a considerable amount of time researching existing methods for constructing whole-organism metabolic networks. This process included exchanging a number of emails with Markus Herrgard, discussing options with Elizabeth Edwards (Webb) and attending a training event with Peter Karp on Pathway Tools. I include a brief overview of my findings and conclusions below.

Simpheny

Elizabeth Edwards had spent several months prior to my arrival at Leeds working in Bernhard Palsson's group at UCSD and had constructed a small model of central carbon metabolism in *P. falciparum* using the Simpheny™ software package. Information on this software suite is hard to come by; the only published material is a single page abstract in a 2005 list of grantees produced by the genomics science program of the US department of energy ⁷⁹. Bernhard Palsson is currently listed as a "Board Observer" and a co-founder of Genomatica, the private company who develop and distribute the software. In conversations with former PhD students of his including Nathan Price, Ines Thiele and Markus Herrgard it has been suggested that Prof. Palsson is able to ensure free or cheap access to the Simpheny platform to researchers with personal affiliations to him as long as their use of the software does not compromise the interests of Genomatica.

My understanding is that Simpheny did not historically enforce a particular ontology — the many arbitrary reaction and compound names in Elizabeth Edwards's reconstruction reinforce this — but that the software has since evolved to use an ontology completely compatible with MetaCyc. Examining recent metabolic reconstructions made using the software it now seems that Simpheny has fully adopted the MetaCyc ontology. Simpheny has the advantage of being the only tool to allow all steps of the metabolic network reconstruction, analysis and visualisation process to occur within a single piece of software. Reactions can be added to a network either by choosing them from a list of pre-existing reactions or by independent definition. These reactions can then additionally be represented graphically in user-editable graphs. Flux-balance analysis can be performed within the programme and the results projected onto the edges of the network graph.

Given the work already done by Elizabeth Edwards in Simpheny and her opinion that it was a useful programme I contacted Genomatica in June 2008 with the hope of using their software for my project. I was quoted a fee of \$13,000 per year for use of the software, with additional fees for extra users and a worrying suggestion that Genomatica may retain certain rights to the metabolic network and simulation results produced using the software. Clearly we would have to find another solution.

Pathway Tools

The Pathway Tools software⁸⁰, at version 15.1 as of June 2011, is maintained by Peter Karp's group at Stanford Research International (SRI) and is available to academics free of charge. It continues to be well promoted and developed and is closely linked to the MetaCyc ontology maintained by the same group at SRI and described in the introduction.

The software is used for the curation of a group of networks all of which end with "cyc", including the *P. falciparum* network, plasmocyc, that we have already seen in relation to choke-point analysis. Pathway Tools includes tools for visualising networks and finding and filling gaps within networks, courtesy of the Pathologic module. When I originally investigated using the software it did not allow SBML export of models, nor did it allow flux-balance analysis to be performed on models. Both these features have now been added.

In the end my decision not to use Pathway Tools for my metabolic reconstruction was based on its extremely confusing user interface, its lack of features and flexibility, and its worrying tendency to revert to a LISP command-line interpreter at the first sign of trouble. That the software was only able to use the MetaCyc ontology was a further issue.

CellDesigner/Payao

CellDesigner⁸¹, and its collaborative online implementation Payao⁸², are interesting because they are primarily visual methods of creating valid SBML format metabolic models. As such, they can be a big help in creating joined up models without making errors and should be a good introduction for researchers looking to build their first models.

The promise of CellDesigner is rather let down by its implementation which is at the same time too complex — with an extremely cluttered user interface — and too simple; no ontology or reconstruction framework is offered or enforced. My experience with CellDesigner has been much like Markus Herrgard's who told me via email in July 2008 that, "*I've also used CellDesigner a bit, but it seems a bit too clumsy for my liking*".

Building MetNetMaker

Obviously, the more you can automate generating the input files the easier things will be.

Markus Herrgard's advice, July 2008 via email.

Having evaluated the available existing software, I decided that the best way to progress with my project was to build my own solution capable of six key tasks.

1. Store the LIGAND ontology* offline within a rigorous reconstruction framework.
2. Allow the user to generate new reactions not defined in the LIGAND ontology and insert them into this framework. Non-defined reactions are generally transfer reactions, spontaneous reactions, biomass reactions or simplifications of complex groups of reactions.
3. Help the user to re-use commonly custom-defined reactions, such as hydrogen ion transport across the cell membrane.
4. Output metabolic networks in SBML format complete with relevant constraints, optimisation objectives and kinetic properties for use with the COBRA toolbox, probably the most widely-used tool for flux-balance analysis.
5. Remain flexible and expandable enough to accommodate additional information. Amongst other things this additional information could consist of Gene-Protein-Reaction relations (GPRs), the chemical formulae of compounds, or references to evidence supporting the inclusion of a given reaction in a network.
6. Have an interface that makes creating new reactions, viewing existing reactions and adding reactions to a metabolic network extremely quick, easy, and reliable.

In the process of developing, publishing and updating MetNetMaker it has acquired a further key feature; the production of visual representations of the network being reconstructed.

* My reasons for choosing KEGG over MetaCyc are explained in the introduction.

Application structure

The structure of MetNetMaker version 1.0, as shown in figure 19, was my solution to these design challenges. MetNetMaker is built around a central reaction database which holds a rigorous reconstruction framework as described in the introduction. This database is initially populated by selected parts of the LIGAND ontology as parsed and imported by a comprehensive parsing script called the LIGAND loader which I describe later. Within MetNetMaker the reaction creator is used to create new reactions to add to the central database and the reaction picker is used to insert reactions from the central database into the selected reactions table in order to build up a metabolic network.

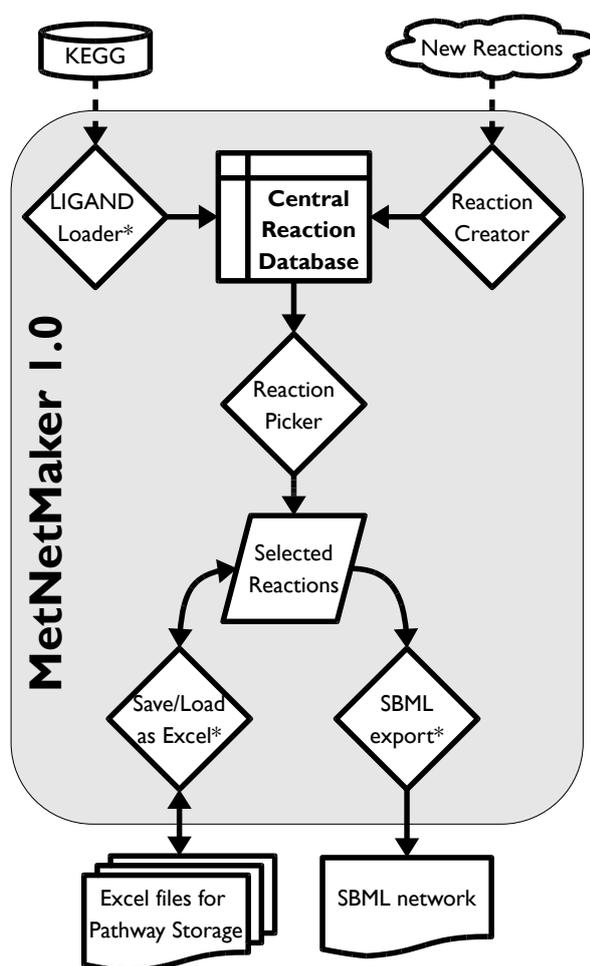


Figure 19: Representation of key parts of the application structure of MetNetMaker as published in Forth et al.⁸³

Excel spreadsheets are used to hold lists of selected reactions and can be considered MetNetMaker's internal file format. Multiple pathway files in Excel format can be joined within MetNetMaker and can additionally be exported as SBML files with extra markup for easy analysis in the COBRA toolbox.

The software can thus be considered in four main parts, the database structure, database population (ligand loader), database addition (reaction creator) and network assembly (reaction picker).

Software implementation

MetNetMaker is written using visual basic for applications (VBA) for the user interface elements and structured-query language (SQL) queries for database manipulation. These two elements of the design are brought together within the Microsoft Access Runtime which manages the central reaction database. The source code for the application (though not of the runtime) is available from within the full version of Microsoft Access.

Decision to use Microsoft Access

Since it has been the subject of much discussion let me briefly explain my decision to write MetNetMaker in Microsoft Access.

I made preliminary efforts to write my software as a webservice using html, css and javascript for the user interface and php scripts linked to a MySQL database to manage the data. This was unacceptable to me because of the complexity of installation if run locally and the poor responsiveness if run remotely.

I attempted to use the opensource database module of openoffice.org as the basis for MetNetMaker but quickly realised that the poor documentation and limited user interface tools available in openoffice.org would make this a very slow process.

Although at the time I was reluctant to write software that I knew could only be used on Windows I am extremely happy with the outcome. The software is extremely easy to install, completely portable and will be supported without any intervention on my part for at least ten years.

Database structure

A further advantage of using Microsoft Access was that the entity relation diagram I designed, figure 20, could be verified graphically as in figure 21. This made it much easier to add links between tables to ensure that additions and deletions of compounds and reactions to and from the reconstruction framework were handled correctly.

The entity relation structure ensures that each compound is referred to uniquely by a single ID and that each reaction — also referred to by a single ID — is built only of those compounds. Synonym tables mean that a reaction can be searched for using any combination of multiple names, metabolic pathways, associated EC numbers, or constituent compounds whilst retaining a single unique ID.

Final Entity Relation Diagram
Central Database

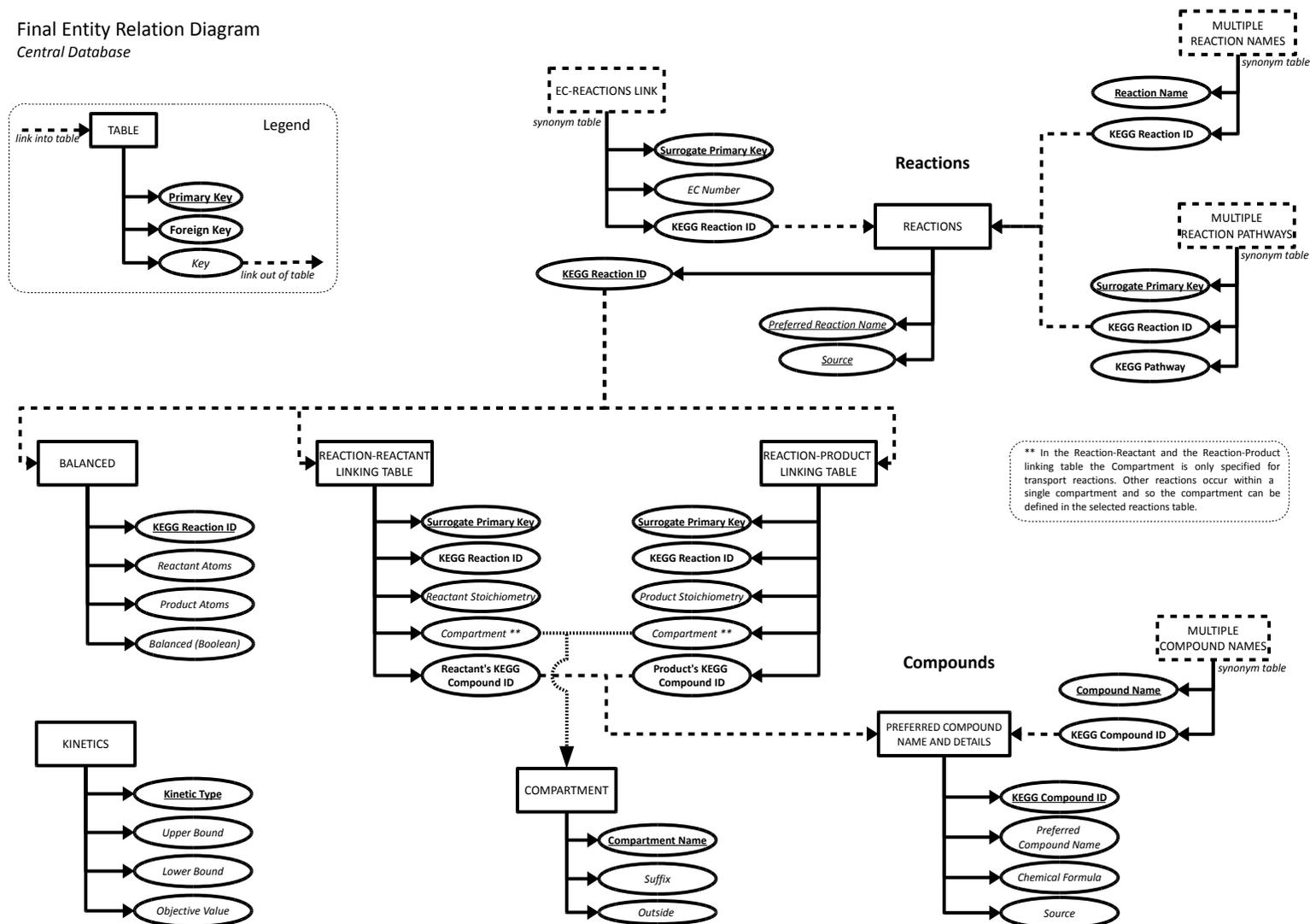


Figure 20: The Entity Relation Diagram of the central database in MetNetMaker is central to implemented a rigorous reconstruction framework.

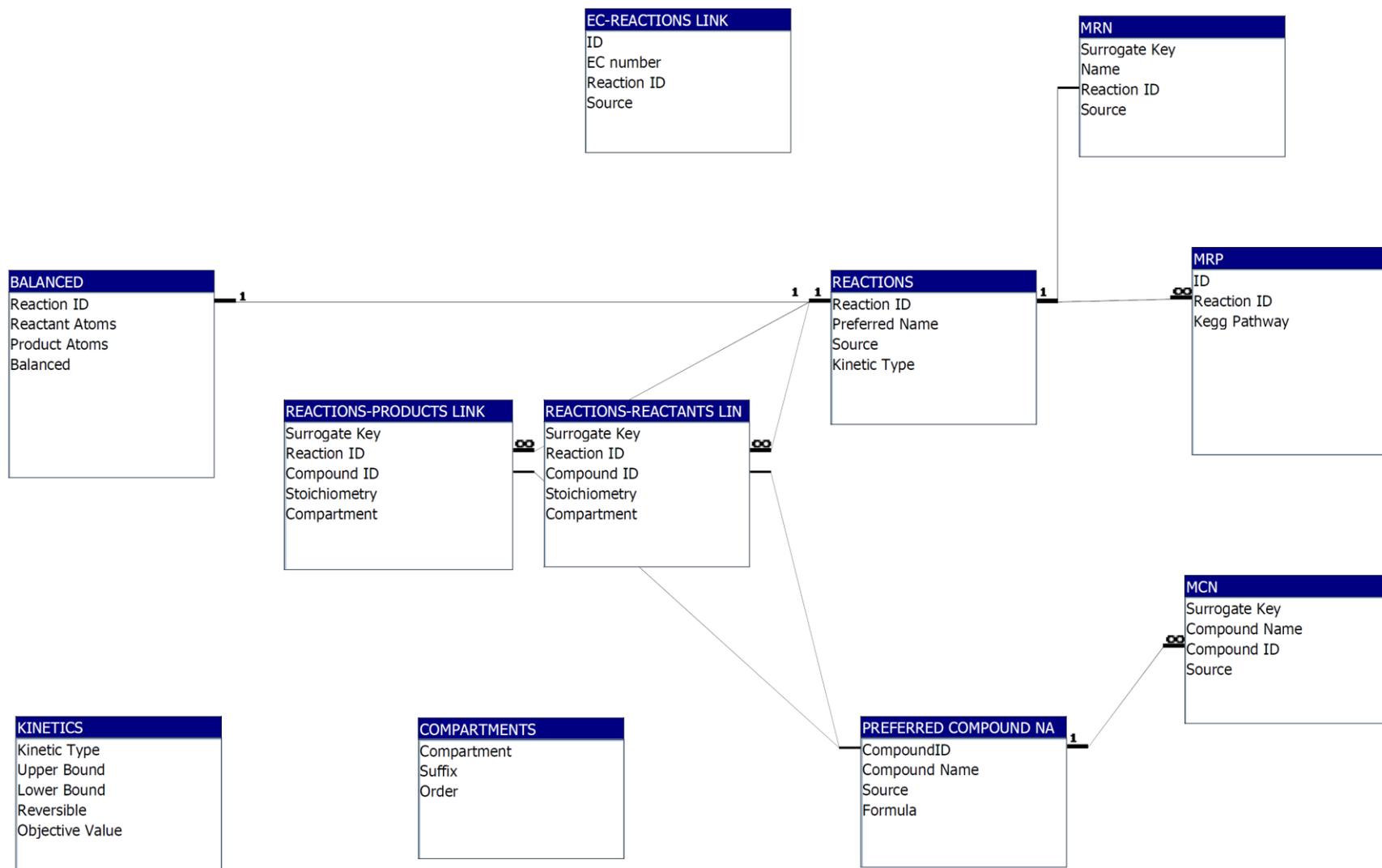


Figure 21: The Entity Relation Diagram of the central database in MetNetMaker as represented within Microsoft Access.

Database population

MetNetMaker's central database is populated with a parsed version of the LIGAND ontology, obtained either through direct download* of the following files,

- reaction from <ftp://ftp.genome.jp/pub/kegg/ligand/reaction/>
- compound from <ftp://ftp.genome.jp/pub/kegg/ligand/compound/>
- glycan from <ftp://ftp.genome.jp/pub/kegg/ligand/glycan/>

or via the KEGG API available at <http://www.genome.jp/kegg/soap/>.

Prior to version 1.5.2 these files were automatically downloaded and parsed on the first run of MetNetMaker but due to licensing changes by KEGG MetNetMaker now ships pre-populated with the last freely available (July 2011) version of the ftp-derived LIGAND ontology. This information — and subsequent ontology updates — could still be obtained legally using the KEGG API but would require some changes to be made to the scripts used to parse the ontology.

The parsing algorithm is written in Perl and notably makes fewer parsing errors than the kegg2sbml algorithm written for conversion of KEGG pathways to SBML⁸⁴. For convenience this perl script is compiled as a Windows executable file (`parsescript.exe`) for distribution with MetNetMaker so that the installation process has absolutely no dependencies but is also included uncompiled (`parsescript.pl`) with the standalone version of MetNetMaker (without installer) for inspection and re-use.

Since the structure and purpose of the glycan flat file is identical to the compound flat file, the approach to parsing the three input files from LIGAND (REACTION, COMPOUND and GLYCAN) can be explained in two parts. In figure 22 I show how a section of the REACTION flat file corresponding to reaction R00014 (pyruvate:thiamine diphosphate acetaldehydetransferase (decarboxylating)) is parsed and a summary of the total length and destinations of the files produced is shown below as table 1.

* Access to the FTP site is currently only available to paid subscribers.

Text file	Length	Purpose/Destination within MetNetMaker
parseKEGGreactionfileOUT.txt	8137 lines	Used to check for errors in the parsing algorithm.
multiplereactionpathways.txt	10781 lines	Imported to MRP table.
multipleECnumbers.txt	7626 lines	Imported to EC-REACTIONS LINK table.
preferredreactionpathway.txt	8137 lines	Created but not used.
preferredECnumber.txt	6836 lines	Created but not used.
multiplereactionnames.txt	8280 lines	Imported to MRN table.
preferredreactionname.txt	8137 lines	Imported to REACTION table.
reactants.txt	16821 lines	Imported to REACTIONS-REACTANTS LINK table.
products.txt	17358 lines	Imported to REACTIONS-PRODUCTS LINK table.

Table 1: The purpose of text files produced by parsing the LIGAND REACTION file.

The equivalent parsing approach for the COMPOUND and GLYCAN files is shown in figure 23 for C00022 (pyruvate) the first reactant of R00014 and a summary of the total length and destinations of the files produced is shown in table 2.

Text file	Length	Purpose/Destination within MetNetMaker
multiplecompoundnames.txt	17798 lines	Imported to MCN table.
preferredcompoundname.txt	10874 lines	Imported to PREFERRED COMPOUND NAME table.
reactionbalancedcheck.txt	8136 lines	Calculated for each reaction rather than parsed directly from LIGAND. Imported to BALANCED table.

Table 2: The purpose of text files produced by parsing the LIGAND COMPOUND and GLYCAN files.

A final text file (`reactionbalancedcheck.txt`) is created by the parsing script and is shown in figure 24. This file is slightly different from the other parsing outputs as it contains information that is calculated — rather than merely extracted — from the LIGAND flat files. The chemical formula information for compounds held in the `preferredcompoundname.txt` file is split into counts for each atom. These atom counts are then combined according to the compounds and stoichiometries for each reaction as given in the `reactants.txt` and `products.txt` file. By counting atoms on each side of the mass-balancing of each LIGAND reaction is checked and any violations are marked in the balanced column of the `reactionbalancedcheck.txt` file. Of the 8134 reactions imported from LIGAND, this methodology reports 6269 to be balanced.

Difficulties with mass and charge balance

It must be noted that although checking that reactions are balanced in this way is useful and avoids obvious mistakes, it is by no means perfect.

Mass balancing is made especially difficult where imprecise definitions of polymers are given such as the addition or subtraction of a single glucose molecule from starch in KEGG reaction R01790 (Starch + H₂O \rightleftharpoons alpha-D-Glucose + Starch). Another common area for problems in the LIGAND ontology is that most reactions involving RNA are also unbalanced, for example R00435 (ATP + RNA \rightleftharpoons Diphosphate + RNA).

Usually problems with mass balance can be resolved quite easily but resolving problems with charge-balance is typically much more difficult. Since metabolism mostly takes place in aqueous solutions, compounds can freely exchange protons and electrons with the solvent. Depending on the pH, many compounds will have different ionisation states from their defined state. LIGAND reactions are remarkably well balanced with respect to protons but electrons are not considered at all when counting atoms. For most pathways this is unimportant but in key areas, especially close to the electron transport chain, it is extremely important to check if reactions are properly balanced and avoid internal cycles capable of creating electrons from nowhere. It is perhaps fortunate that the electron transport chain plays such a minor role in *P. falciparum* metabolism that I have been able to omit this pathway from my reconstruction. The reasons for this simplification are given later.

```

///
ENTRY          R00014                      Reaction
NAME          pyruvate:thiamin diphosphate acetaldehydetransferase (decarboxylating)
DEFINITION    Pyruvate + Thiamin diphosphate <=> 2-(alpha-Hydroxyethyl) thiamine
              diphosphate + CO2
EQUATION      C00022 + C00068 <=> C05125 + C00011
COMMENT       2-oxoglutarate dehydrogenase complex: pyruvate dehydrogenase (see R01699),
              subsequently R03270
              TPP-dependent enzymatic reaction (see R00006), subsequently R03050
              TPP-dependent enzymatic reaction (see R00226), subsequently R04672
              TPP-dependent enzymatic reaction (see R00224), subsequently R00755
              TPP-dependent enzymatic reaction (see R08648), subsequently R04673
RPAIR         RP00074 C00068_C05125 main
              RP02918 C00022_C05125 trans
              RP05698 C00011_C00022 leave
ENZYME        1.2.4.1      2.2.1.6      4.1.1.1
PATHWAY       rn00010 Glycolysis / Gluconeogenesis
              rn00020 Citrate cycle (TCA cycle)
              rn00290 Valine, leucine and isoleucine biosynthesis
              rn00620 Pyruvate metabolism
              rn00650 Butanoate metabolism
              rn01100 Metabolic pathways
              rn01110 Biosynthesis of secondary metabolites
              rn01120 Microbial metabolism in diverse environments
ORTHOLOGY     K00161 pyruvate dehydrogenase E1 component subunit alpha [EC:1.2.4.1]
              K00162 pyruvate dehydrogenase E1 component subunit beta [EC:1.2.4.1]
              K00163 pyruvate dehydrogenase E1 component [EC:1.2.4.1]
              K01568 pyruvate decarboxylase [EC:4.1.1.1]
              K01652 acetolactate synthase I/II/III large subunit [EC:2.2.1.6]
              K01653 acetolactate synthase I/III small subunit [EC:2.2.1.6]
              K11258 acetolactate synthase II small subunit [EC:2.2.1.6]
///
parseKEGGreactionfileOUT.txt

```

Reaction ID	Name	Definition	Equation	Pathway
R00014	pyruvate:thiamin diphosphate acetaldehydetransferase (decarboxylating)	Pyruvate + Thiamin diphosphate <=> 2-(alpha-Hydroxyethyl)thiamine diphosphate + CO2	C00022 + C00068 <=> C05125 + C00011	rn00010 Glycolysis / Gluconeogenesis

multiplereactionpathways.txt

Reaction ID	Kegg Pathway
R00014	Glycolysis / Gluconeogenesis
R00014	Citrate cycle (TCA cycle)
R00014	Valine, leucine and isoleucine biosynthesis
R00014	Pyruvate metabolism
R00014	Butanoate metabolism
R00014	Metabolic pathways
R00014	Biosynthesis of secondary metabolites
R00014	Microbial metabolism in diverse environments

multipleECnumbers.txt

Reaction ID	EC number
R00014	1.2.4.1
R00014	2.2.1.6
R00014	4.1.1.1

preferredreactionpathway.txt
preferredECnumber.txt
Created but not used

multiplereactionnames.txt
preferredreactionname.txt

Reaction ID	Name	Source
R00014	pyruvate:thiamin diphosphate acetaldehydetransferase(decarboxylating)	KEGG

reactants.txt

Reaction ID	Compound ID	Stoichiometry
R00014	C00022	1
R00014	C00068	1

products.txt

Reaction ID	Compound ID	Stoichiometry
R00014	C05125	1
R00014	C00011	1

Figure 22: (top) Colour coded entry from the LIGAND database REACTION file for R00014. (bottom) The tables produced from parsing this entry and then used to populate tables in MetNetMaker's central database.

```

///
ENTRY          C00022                               Compound
NAME           Pyruvate;
               Pyruvic acid;
               2-Oxopropanoate;
               2-Oxopropanoic acid;
               Pyrroacemic acid

FORMULA        C3H4O3
MASS           88.016
REACTION        R00006 R00008 R00014 R00195 R00196 R00197 R00198 R00199
               R00200 R00203 R00205 R00206 R00207 R00208 R00209 R00210
               ...
               R08660 R08667 R08686 R08698 R08714 R09048 R09088 R09238
               R09254 R09366
PATHWAY         ko00010 Glycolysis / Gluconeogenesis
               ...
               ko00900 Terpenoid backbone biosynthesis
               map01060 Biosynthesis of plant secondary metabolites
               ...
               map01070 Biosynthesis of plant hormones
               ko01100 Metabolic pathways
               ...
               ko04930 Type II diabetes mellitus
ENZYME          1.1.1.27           1.1.1.28           1.1.1.38           1.1.1.39
               1.1.1.40           1.1.1.83           1.1.2.3            1.1.2.4
               ...
               4.4.1.13           4.4.1.15           4.4.1.16           4.4.1.24
               4.4.1.25           4.4.1.-            4.5.1.2            4.6.1.1 (C)
               6.4.1.1
DBLINKS         CAS: 127-17-3
               PubChem: 3324
               ChEBI: 32816
               LIPIDMAPS: LMFA01060077
               LipidBank: DFA0385
               KNApSAcK: C00001200
               PDB-CCD: PYR
               3DMET: B00006
               NIKKAJI: J2.015J
ATOM            6
               1   C1a C    12.6924  -17.4300
               2   C5a C    13.9049  -16.7300
               3   C6a C    15.1173  -17.4300
               4   O6a O    16.3318  -16.7288
               5   O5a O    13.9049  -15.3302
               6   O6a O    15.1173  -18.8300
BOND            5
               1     1     2  1
               2     2     3  1
               3     3     4  1
               4     2     5  2
               5     3     6  2
///
multiplecompoundnames.txt


| Compound ID | Compound Name       |
|-------------|---------------------|
| C00022      | Pyruvate            |
| C00022      | Pyruvic acid        |
| C00022      | 2-Oxopropanoate     |
| C00022      | 2-Oxopropanoic acid |
| C00022      | Pyrroacemic acid    |


preferredcompoundname.txt


| Compound ID | Compound Name | Formula |
|-------------|---------------|---------|
| C00022      | Pyruvate      | C3H4O3  |


```

Figure 23: (top) Colour-coded sample entry from the LIGAND database COMPOUND file for C00022. (bottom) The tables produced from parsing this entry and then used to populate tables in MetNetMaker's central database.

reactionbalancedcheck.txt			
Reaction ID	Reactant Atoms	Product Atoms	Balanced
R00014	C15 H23 N4 O10 P2 S1	C15 H23 N4 O10 P2 S1	True

Figure 24: Table produced during parsing to test whether reactions are mass balanced.

User interface

A key aim for me when writing MetNetMaker was that it should be usable by someone with a limited knowledge of metabolic network design. Ideally this would include people who spent a majority of their time in the wet lab performing experiments on the organism they were studying. A key part of achieving this goal was to have a friendly and familiar user interface capable of performing all tasks without ever presenting the user with a command line prompt. My decision to build features around the user interface, rather than the other way round, means that there are some useful features missing from MetNetMaker. This lack of features is a reasonable criticism of my software but I think my software fills a valuable niche below much more complex programmes like Metannogen⁸⁵ and I have been greatly encouraged by the feedback I have received from all around the world.

The user interface of MetNetMaker can be considered in two main parts, the reaction picker and the reaction creator, all tied together by the ribbon user-interface (UI). The ribbon — now common in modern Windows programmes — contains frequently used controls within the user interface.

Reaction Picker

The reaction picker, shown in figure 27 on page 64, is the immediately visible part of MetNetMaker and is itself divided into a number of subsections.

Reaction Filters

This section gives the user the chance to search for a reaction within the database by specifying reactants and/or products (by either full or partial name or ID), associated pathway, associated EC number and reaction name (including partial names). All of these parameters can be searched for in any combination.

Reaction Chooser

This section provides the user with a list of reactions that fit the search parameters chosen in the “Reaction Filters” section. In figure 27 the reaction chooser only shows reactions associated within the LIGAND ontology with the EC number 1.2.4.1. Reaction R00014 has been selected. Double-clicking on a selected reaction within the reaction chooser opens the relevant page of the LIGAND website with the most recent full description of the reaction.

Three additional buttons allow the user to delete existing reactions, open the reaction creator to create a new reaction from scratch, or edit an existing reaction by pressing the

“Edit Reaction” button. Selecting a reaction and pressing “Edit Reaction” opens the selected reaction within the reaction creator.

Reaction Viewer

Clicking on a reaction in the “Reaction Chooser” opens it for inspection in the “Reaction Viewer” section so that the user can be sure that they are ready to add the correct reaction to their metabolic network. Additional information is shown for each reaction, including whether it is balanced or not. Double-clicking on any compound within the reaction viewer opens the relevant page of the LIGAND website with the most recent full description of the compound.

Reaction Picker

In this area the user chooses which compartment they want the reaction to take place in and the kinetic type of the reaction. Possible kinetic types are transfer, reversible (with varying maximum fluxes), forward, backward, special (if a user has defined the reaction themselves in the reaction creator they are prompted to choose this) and biomass. By clicking on the “Add Reaction ==>” button the reaction that is currently selected in the reaction chooser section is added to the “Selected Reactions” section with the relevant compartment and kinetic type attributes added to it. A selection of common compartments is pre-populated but non-standard custom compartments — such as the food vacuole in *P. falciparum* — can be added where necessary.

Selected Reactions

The selected reactions area shows the user a scrollable list of all the reactions that are currently in their metabolic network. Items can be deleted and edited from within this list to allow the user to build up a metabolic network. The selected reactions table can be detached, as shown in figure 25, and moved separately from the main window of MetNetMaker allowing changes to be made more easily. Additional columns of the selected reactions table leave space to tweak the composition of a model’s objective/biomass function, add information on the associated KEGG Pathway (Map Number) of each reaction, add information on a gene associated with each reaction and choose to reverse the preferred direction of the reaction during export.

Reaction ID	Compartment	EC Number	Max +ve Flux	Max -ve Flux	Objective	Map Number	Gene	Rev	Notes
OG_ctom	Special	spontaneous	500	-500	<input type="checkbox"/>			<input type="checkbox"/>	
P_btoc	Special	spontaneous	500	-500	<input type="checkbox"/>			<input type="checkbox"/>	
P_ctom	Special	spontaneous	500	-500	<input type="checkbox"/>			<input type="checkbox"/>	
PEP_ctoa	Special	unknown	25	0	<input type="checkbox"/>			<input type="checkbox"/>	
R00014	Apicoplast	1.2.4.1	500	-500	<input type="checkbox"/>	00620		<input type="checkbox"/>	
R00200	Cytosol	2.7.1.40	500	-500	<input type="checkbox"/>	00010		<input checked="" type="checkbox"/>	
R00200	Apicoplast	2.7.1.40	500	-500	<input type="checkbox"/>	00620		<input checked="" type="checkbox"/>	
R00267	Mitochondria	1.1.1.42	500	-500	<input type="checkbox"/>	00022		<input type="checkbox"/>	
R00341	Cytosol	4.1.1.49	500	-500	<input type="checkbox"/>	00620		<input type="checkbox"/>	
R00342	Mitochondria	1.1.1.37	500	-500	<input type="checkbox"/>	00028		<input checked="" type="checkbox"/>	
R00342	Cytosol	1.1.1.37	500	-500	<input type="checkbox"/>	00620		<input checked="" type="checkbox"/>	
R00345	Cytosol	4.1.1.31	500	-500	<input type="checkbox"/>	00620		<input type="checkbox"/>	
R00351	Mitochondria	2.3.3.1	500	-500	<input type="checkbox"/>	00029		<input type="checkbox"/>	
R00355	Cytosol	2.6.1.1	500	-500	<input type="checkbox"/>	00620		<input type="checkbox"/>	
R00405	Mitochondria	6.2.1.5	500	-500	<input type="checkbox"/>	00024		<input checked="" type="checkbox"/>	
R00408	Mitochondria	1.3.99.1	500	-500	<input type="checkbox"/>	00026		<input type="checkbox"/>	
R00432	Mitochondria	6.2.1.4	500	-500	<input type="checkbox"/>	00025		<input checked="" type="checkbox"/>	
R00621	Mitochondria	1.2.4.2	500	-500	<input type="checkbox"/>	00023		<input type="checkbox"/>	
R00658	Cytosol	4.2.1.11	500	-500	<input type="checkbox"/>	00010		<input type="checkbox"/>	
R00703	Cytosol	1.1.1.27	500	-500	<input type="checkbox"/>	00010		<input checked="" type="checkbox"/>	
R00842	Cytosol	1.1.1.8	500	-500	<input type="checkbox"/>	00010		<input checked="" type="checkbox"/>	
R00847	Cytosol	2.7.1.30	500	-500	<input type="checkbox"/>	00010		<input type="checkbox"/>	

Figure 25: Detachable window for the selected reactions list within MetNetMaker 1.5.2.

A further notable feature of the selected reactions table is the option to see dead-end compounds — the lists of compounds that only enter, or only leave the network — for the currently loaded set of reactions. Since this calculation requires knowledge of the preferred direction of each constituent reaction in the network any reactions known to be operating in the reverse of their default direction need to be reversed by selecting the “Rev” tickbox in the selected reactions table. The resulting dead-end compounds table, as shown in figure 26, can be extremely useful in solving small errors in a network such as missing inter-compartment transfer reactions.

A full description of the algorithm used to calculate the dead-end compounds within a list of selected reactions is included as appendix I.

This list may be truncated, a full copy has been opened in Notepad. Copy into a spreadsheet for correct formatting.

Compounds that only enter the network

Compound ID	Compartment	Compound Name
C00035	Mitochondria	GDP
C00049	Cytosol	L-Aspartate
C00049	Special	L-Aspartate
C00158	Mitochondria	Citrate
C00583	Cytosol	Propane-1,2-diol

Compounds that only exit the network

Compound ID	Compartment	Compound Name
C00044	Mitochondria	GTP
C00093	Cytosol	sn-Glycerol 3-phosphate
C00122	Mitochondria	Fumarate
C00577	Cytosol	D-Glyceraldehyde
C00794	Cytosol	D-Sorbitol

OK

Figure 26: Dead-end compounds report for an example set of reactions in the selected reactions table.

MetNetMaker

File MetNetMaker Ribbon

Import Excel Export Excel Export SBML (L2V1) Export SBML (L2V4) Visualise in Cytoscape Show Dead-End Compounds Reaction Picker Reaction Creator Selected Reactions (pop-up) Application Components Download and Parse KEGG LIGAND Parse KEGG LIGAND Populate Database Database Control Backup User Reactions Restore User Reactions Clear Database Clear Database Merge .xgml Visualisations Advanced Visualisation

Reaction Picker

Reaction Filters

Reaction Compound Pathway Clear Filters

Reaction ID:

Reaction Name:

EC number:

Filter !

Reaction Chooser

Reaction ID	Preferred Name
R00014	pyruvate:thiamin diphosphate acetaldehydetransferase(decaboxylating)
R00209	pyruvate dehydrogenase complex
R01699	pyruvate:[dihydrolypoyllysine-residueacetyltransferase]-lipoyllysine
R03270	NO NAME

Delete Reaction Edit Reaction Create New Reaction

Reaction Picker

EC Number:

1.2.4.1
2.2.1.6
4.1.1.1
spontaneous
unknown

Compartment: Edit

Cytosol

Initial Kinetic Type:

Reversible 500

Add Reaction ==>

Selected Reactions Open in a Seperate Window

Currently Loaded: Clear

Reaction ID	Compartment	EC Number	Max +ve Flux	Max -ve Flux
OG_btoc	Special	spontaneous	0	
OG_ctom	Special	spontaneous	500	
P_btoc	Special	spontaneous	500	
P_ctom	Special	spontaneous	500	
PEP_ctoa	Special	unknown	25	
R00014	Apicoplast	1.2.4.1	500	
R00200	Cytosol	2.7.1.40	500	
R00200	Apicoplast	2.7.1.40	500	
R00267	Mitochondria	1.1.1.42	500	
R00341	Cytosol	4.1.1.49	500	
R00342	Mitochondria	1.1.1.37	500	
R00342	Cytosol	1.1.1.37	500	
R00345	Cytosol	4.1.1.31	500	
R00351	Mitochondria	2.3.3.1	500	
R00355	Cytosol	2.6.1.1	500	
R00405	Mitochondria	6.2.1.5	500	
R00408	Mitochondria	1.3.99.1	500	
R00432	Mitochondria	6.2.1.4	500	
R00621	Mitochondria	1.2.4.2	500	
R00658	Cytosol	4.2.1.11	500	
R00703	Cytosol	1.1.1.27	500	

Record: 14 of 96 No Filter Search

Reaction Viewer

Reaction ID: Preferred Name: pyruvate:thiamin diphosphate acetaldehydetransferase(decaboxylating) Pathways: Glycolysis / Gluconeogenesis Citrate cycle (TCA cycle) Valine, leucine and isoleucine biosynthesis

Reactants: Balanced: Yes Products:

No.	ID	Compound Name	Compartment	No.	ID	Compound Name	Compartment
1	C00022	Pyruvate		1	C05125	2-Hydroxyethyl-ThPP	
1	C00068	TPP		1	C00011	CO2	

Record: 107 of 8241 Unfiltered Search

You can delete Reactions by clicking on the left side of the table to select them and pressing the Delete key.

Powered by Microsoft Access

Figure 27: Reaction Picker window within MetNetMaker 1.5.2. Reaction R00014 is currently highlighted. The ribbon UI at the top is always present.

Reaction Creator

The reaction creator, as shown in figure 28 on page 67, performs two functions. Firstly, it protects the reaction database and reconstruction framework from the user, and secondly it helps the user make additions to the reconstruction framework without creating inconsistencies or duplications. Like the reaction picker it can be considered in different subsections.

Compound Picker

Here the user can search — by either a partial or complete name or compound ID — for the compound they would like to add to the reaction. Because a single compound may have multiple names each compound may appear more than once in the possible compounds list, as in the example in figure 28.

Custom compounds, typically generic compounds such as “lipid” or “any tRNA”, can be added to the list of compounds here and are marked within the PREFERRED COMPOUND NAME table as being user generated so they can be exported by pressing the “Backup User Reactions” button in the ribbon UI. These exported custom compounds can be restored at a later date or transferred to other version of MetNetMaker as required.

Add Record

The user selects whether the compound selected in the compound picker is a reactant or a product and defines which, if any, compartment it is in. By clicking on the “Add New Record” button this compound is added to the selected compounds area. At this stage ambiguities from the compound picker area are removed and only the unique compound ID and its associated preferred name is added to the selected compound area.

Selected Compounds

This section holds lists and details of all the compounds in the reaction being created or edited. In figure 28 the reaction R00014 is being edited and the relevant reactants and products and their associated stoichiometries are present and editable in the selected compounds area.

Similar Reactions

The application checks that the reaction that is being defined has not already been created by warning of similar reactions (those containing the same reactants and products). In this example

it is warning that R00014 is already present within the reaction database so that it not defined twice with different names.

Reaction Naming

Despite efforts to automate naming of new reactions I have decided to offer suggestions on naming conventions but ultimately leave the user to name them. In this section the user defines a long name and a reaction ID before the reaction is added to the central reaction database.

Finalise and Create Reaction

Clicking this button finalises the reaction and adds it to the database. User-added reactions are marked “User” instead of “KEGG” in the source column of the REACTION table. These custom reactions are exportable in the same way as custom compounds by using the “Backup User Reactions” button in the ribbon UI.

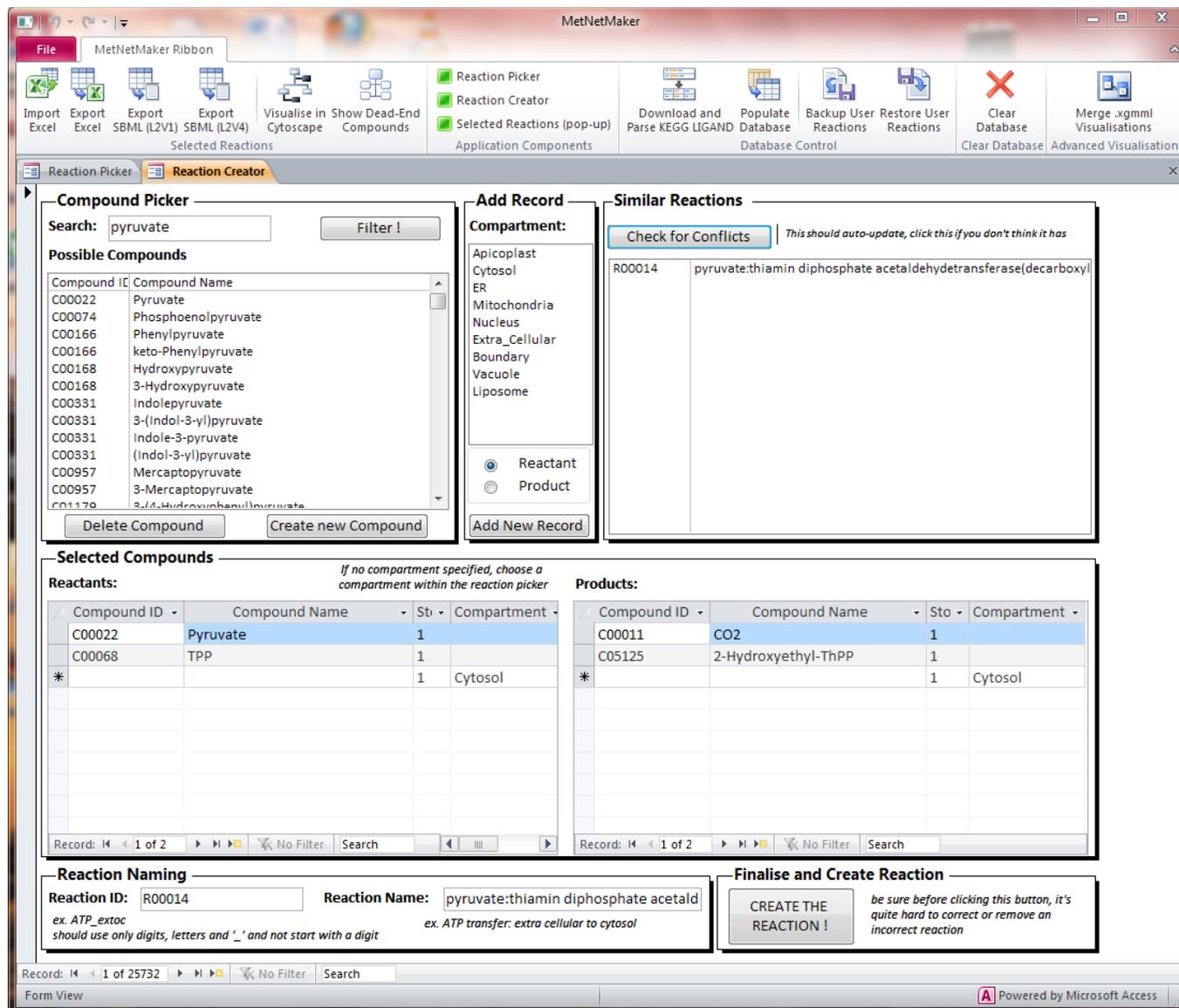


Figure 28: Reaction R00014 being edited within the Reaction Creator window of MetNetMaker 1.5.2. The ribbon UI at the top is always present.

Application outputs for further analysis

The most frequently used controls in the ribbon are those related to the import and export of models. Model export options are critical because further use of a model must take place in other programmes. MetNetMaker can export models to three destinations, Excel spreadsheets for internal saving, SBML files for sharing models, and Cytoscape⁸⁶ sessions for visualisation.

Excel spreadsheet import and export

Since MetNetMaker is written in Microsoft Access it is simple to export a list of reactions accumulated in the selected reactions table as an Excel spreadsheet using the “Export Excel” button. In this way, small sections of a complete network can be built over time and saved and shared easily.

These sets of reactions can later be loaded into MetNetMaker using the “Import Excel” button. Importing multiple sets of reactions into a single selected reactions table within MetNetMaker automatically joins them and in this way large networks can be quickly assembled from pre-prepared components.

SBML export

MetNetMaker offers two options for SBML export of the reactions currently defined in the Selected Reactions table. Files exported as level 2 version 1 have been fully tested to work with COBRA toolbox 1.3.3 and this option is retained for compatibility. Exported networks using the level 2 version 4 export button have been validated using the SBML validator at sbml.org/facilities/validator and are known to work with a wide variety of tools including CellDesigner⁸¹, SurreyFBA⁸⁷ and Cytoscape (SBML import).

Cytoscape export

Any valid SBML network can easily be imported into Cytoscape for visualisation. This would be a good method of network visualisation both for verification of the network during reconstruction and to view the results of analysis except for one problem. For useful visualisation of a network frequently occurring compounds need aliasing to allow the network to take a proper shape and neither SBML nor the cytoscape SBML import function provide a mechanism for defining and aliasing these currency metabolites. For this reason MetNetMaker includes a dedicated button to “Visualise in Cytoscape” which prompts the user to select which metabolites to alias before opening the network in cytoscape.

Since a metabolite may be a currency metabolite only in some parts of a complete network a related feature the “Merge .xgmmml visualisations” lets the user merge separate visual representations of networks in the same way as they merge separate excel files in the selected reactions table. This helps the user manually layout a large network in small pieces whilst specifying different currency metabolites for each section.

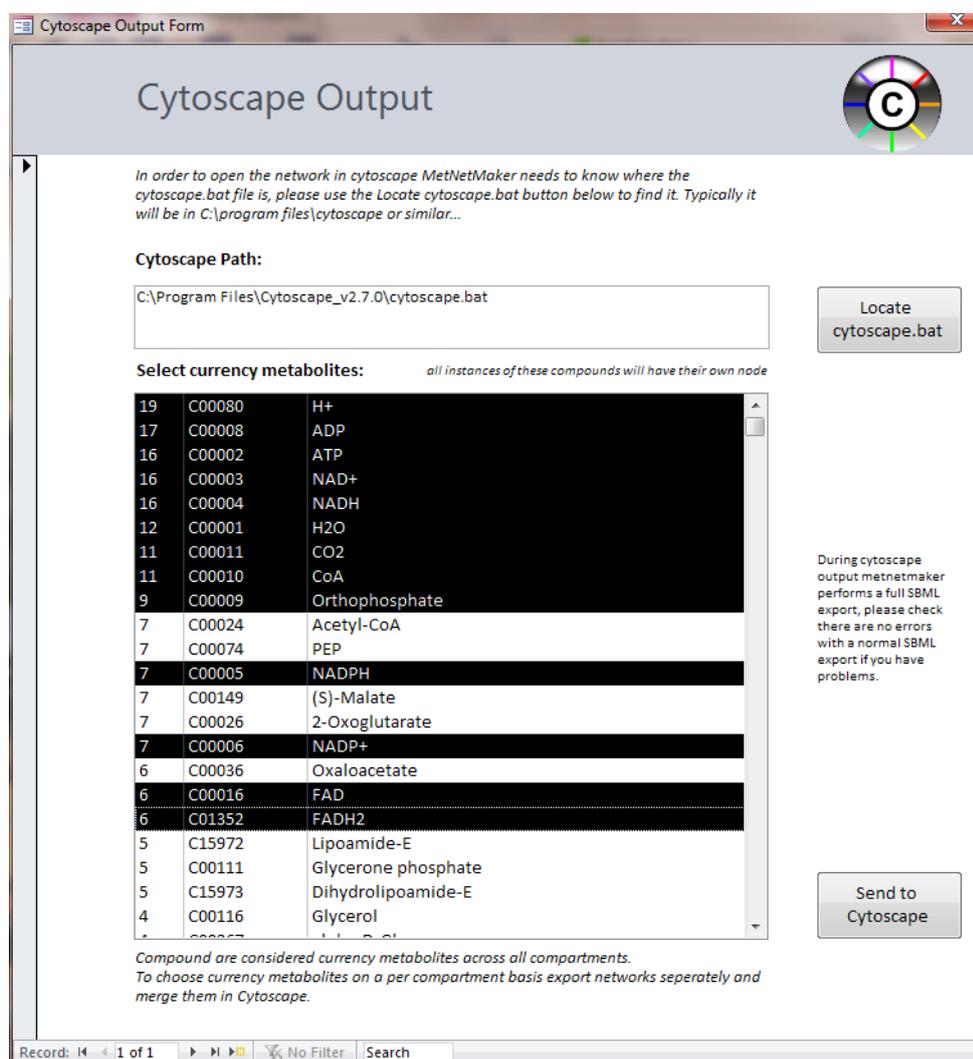


Figure 29: The cytoscape output form in MetNetMaker 1.5.2

Final structure and additional features

Because networks within MetNetMaker are stored as Excel spreadsheets they can easily be opened, edited and appended outside of MetNetMaker and then re-imported later. This flexibility is the key behind the evidence lookup process I have developed to help with my reconstruction of a metabolic network for *P. falciparum*. A key advantage of this approach is that organism-specific information is kept separate from the general reconstruction thus allowing a complete reconstruction without unnecessary complexity.

A thorough example of the reconstruction process, including all parts of the complete process diagram shown in figure 30, is contained in the next chapter of this thesis.

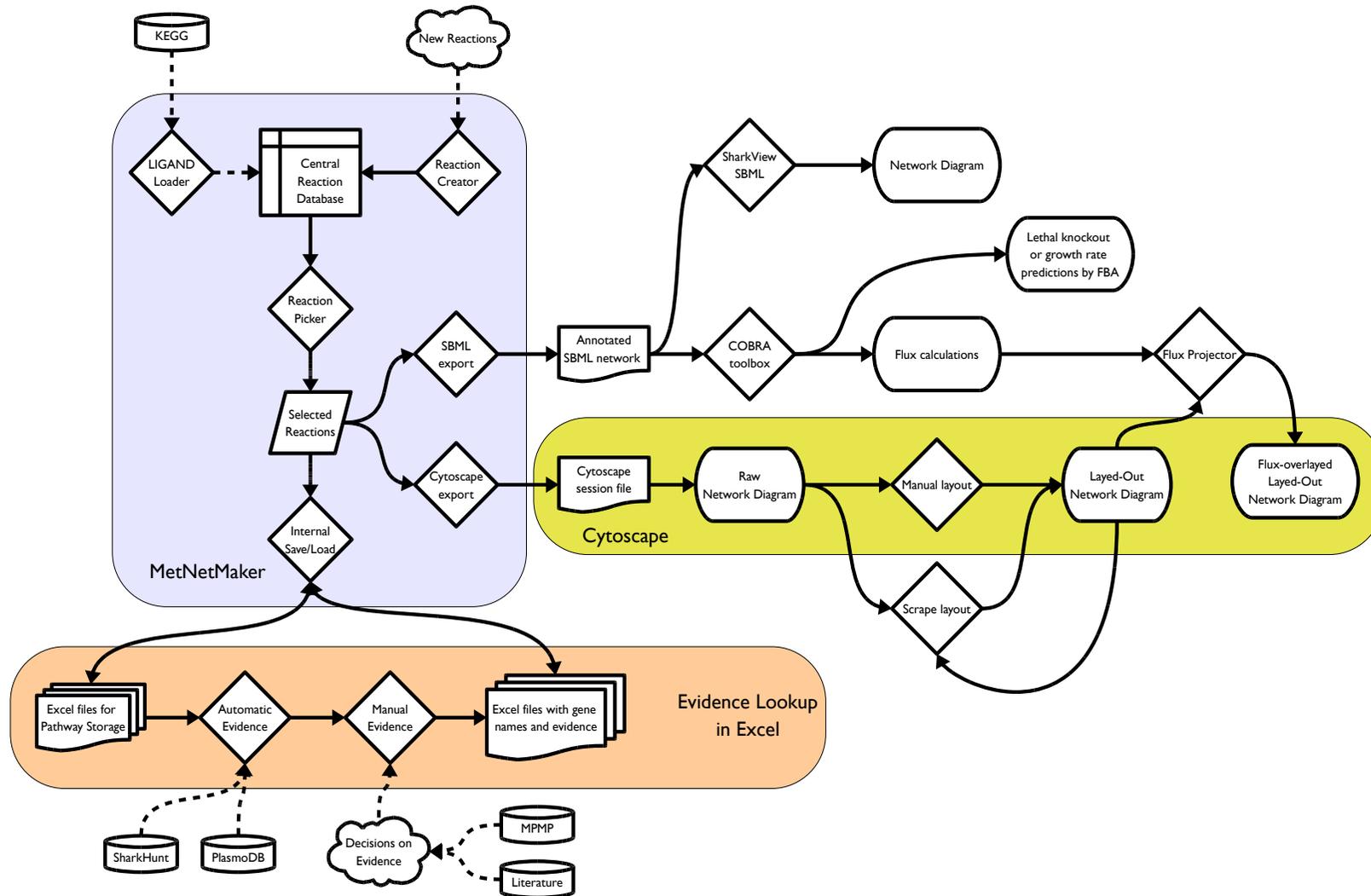


Figure 30: Representation of key parts of the application structure of MetNetMaker 1.5.2 as currently available from www.metnetmaker.com and an overview of associated tools for further analysis of networks.

Visualisation and Network Reconstruction

visit www.tomforth.co.uk/pfalnetwork for the network.

The case for manual curation: 25% of 4711 enzymatic activities are orphan reactions and 26% of Pfam families are of unknown function.*

Monica Munoz-Torres at ISB2012†

Why visualisation and network reconstruction in the same chapter?

I have shown in the introduction that a visual representation of a metabolic network is directly equivalent to the mathematical representation we can use for calculations and simulations. Previous reconstructions of the *P. falciparum* metabolic network have strongly focused on either the mathematical representation — Yeh *et al.*⁷, Hutchmacher *et al.*⁵⁹ and Plata *et al.*⁵⁸ — or the visual representation as in Ginsburg *et al.*¹⁰.

Another limited visual reconstruction is shown as figure 33 and is an extract from Olszewski *et al.*⁸⁸. No computational model of this network exists although my final model largely follows its template.

The focus on the mathematical approach to network reconstruction is likely to have contributed to some of the errors found in the Hutchmacher and Plata models, evidence of which is shown in figure 31 and figure 32. These errors are easily visible to humans when drawn but, as described in Gregorvyan *et al.*⁸⁹, are often difficult to uncover mathematically. Also described in Gregorvyan *et al.* are errors, such as unconserved metabolites, that are difficult to see but that can be discovered computationally and which I will return to later.

The visualisations of the networks in figure 31 and figure 32 use two different techniques to show a number of non-connected/orphan reactions. Since these network visualisations

* The Pfam database is a large collection of protein families, each represented by multiple sequence alignments and hidden Markov models (HMMs) – as defined at pfam.sanger.ac.uk.

† Biocuration 2012. The Conference of the International Society for Biocuration, April 2-4, 2012, Washington DC, USA.

have no aliased compounds these reactions are completely redundant and serve no possible purpose within the final networks. Visualisation lets us see these obvious errors but we must also consider that there are likely to be further orphan reactions, or non-orphan but incorrect reactions that are included in error within both models. The orphan reactions in both models provide good reasons why I think that network visualisation is a vitally important part of network reconstruction.

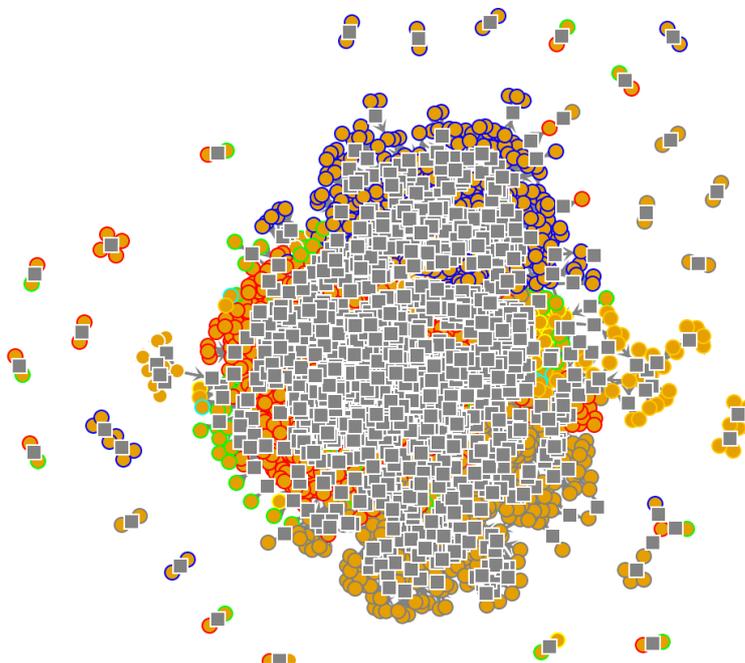


Figure 31: *P. falciparum* metabolic model from Huthmacher et al.⁵⁹ as visualised in SharkView⁹⁰ with no compound aliasing. Squares represent reactions and circles represent compounds. The outline colour of each compound denotes the compartment it is found in. Isolated reactions are not connected to the network.

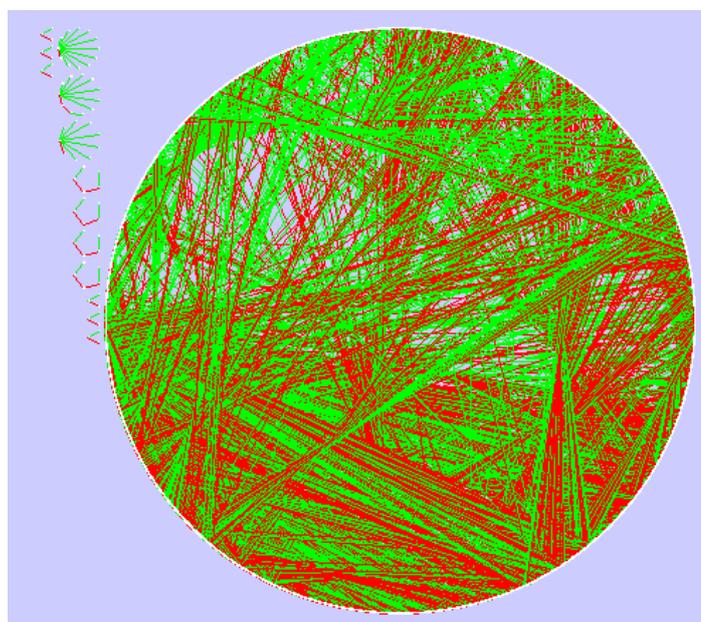


Figure 32: *P. falciparum* metabolic model from Plata et al.⁵⁸ as visualised in cytoscape⁸⁶ with no compound aliasing. Red lines show compounds entering a reaction, green lines show compounds leaving a reaction. Compounds and reactions are arranged around the perimeter of the circle. Isolated reactions at the top-left are not connected to the network.

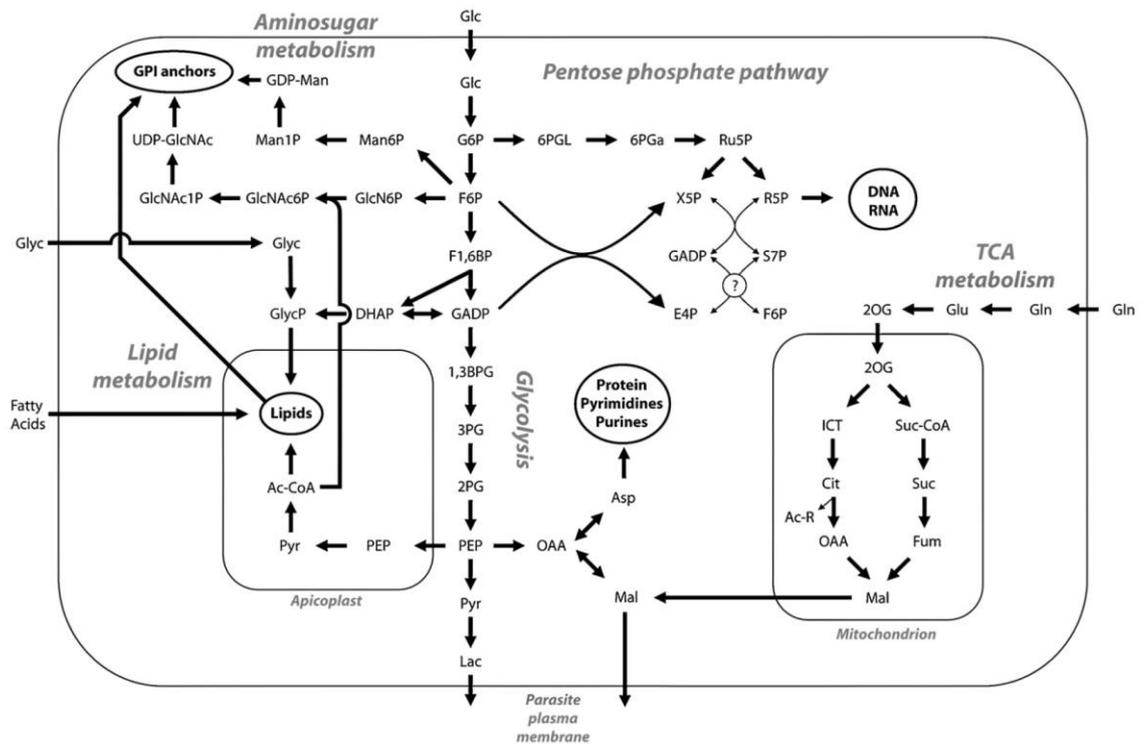


Figure 33 : The simplified map of carbon metabolism from Olszewski et al.⁸⁸ shows the unusual role of the mitochondrial TCA cycle in *P. falciparum* metabolism and the unusual role of the apicoplast in lipid synthesis and manipulation. Some of the visual simplifications are highly misleading, for example the reaction $\text{Glu} \Rightarrow \text{2OG}$ (glutamate \Rightarrow oxoglutarate) and the reaction $\text{OAA} \Leftrightarrow \text{Asp}$ (oxaloacetate \Leftrightarrow aspartate) are actually the same reaction and their linked behaviour radically changes the structure of the network. My final model on page 94 (figure 47) retains some of the design features of this model without making this important oversimplification.

KEGG projector and Cytoscape: one problem, two approaches

I have explained the general principles of metabolic network visualisation in the introduction and I have tried many techniques to implement those ideas. In the end I have largely settled for two network visualisation techniques which I have implemented in versions of MetNetMaker greater than v1.6. Access to both the KEGG projector and cytoscape visualisation techniques are shown in the user interface excerpt in figure 34. Both visualisation techniques are used widely in the rest of this thesis and it makes more sense to refer to later figures in this chapter as examples than to include them in this section out of context.



Figure 34 : The visualisation section of the ribbon in MetNetMaker v1.6+ includes new options for visualising reactions in the selected reactions table. These methods rely either on my KEGG projector tool (the two buttons on the left) or Cytoscape for visualisation.

KEGG projector: a visual tool for immediately gauging the extent of a network

KEGG projector relies on the KEGG markup language (KGML) global map available from the KEGG ftp site and described at www.kegg.jp/kegg/xml/ where an xml version of the global map can be downloaded. This xml file describes the KEGG global map⁹¹ in the scalable vector graphics (SVG) format and as such the visual map can be easily read by humans and manipulated by simple computer programmes called scripts.

The KEGG projector tool — available from within MetNetMaker and also at www.tomforth.co.uk/keggprojector — shows the reactions described by a list of KEGG reaction IDs or EC numbers projected onto the KEGG global map. An example of this for all the reactions for which I have found any evidence of existence in *P. falciparum* is shown on page 88 as figure 42. I have found the KEGG projector to be a very powerful tool because the KEGG global map is widely recognisable to people working in metabolism and because the maps are generated in seconds from basic input and without the need for any manual intervention.

There are some drawbacks to this very quick approach to visualisation. One is that the KGML global map only has 2048 edges, much fewer than the current 8135 active reactions and 6742 EC-linked reactions in the KEGG LIGAND ontology. This means that many different reactions correspond to the same edge within the visualisation and some reactions do not refer to any edge at all. A further problem arises as a side-effect of the simplifications made to create the KGML global map whereby only major connections between nodes are shown. Clearly currency metabolites need to be removed from this kind of visual overview, but at the level of simplification used in KGML important connections between major metabolites are also frequently omitted. An example of these simplifications is shown in context on page 83 as figure 39 where the addition of 18 reactions to a network is projected as just 7 extra paths, one of which appears — wrongly — to be orphaned from the other reactions in the network.

These and other limitations mean that KEGG projector is mostly useful to get a quick overview of a metabolic network and whilst I have written an extension of the tool that allows the thickness of each path to be weighted by calculated fluxes from FBA, the inherent disjointedness of the visualisation makes this a poor tool for visualising fluxes in a metabolic network.

Cytoscape: precise network display and flux projections

Cytoscape visualisation makes none of the simplifications that make the KEGG projector so easy to use and it therefore retains much more power and flexibility. Cytoscape visualisations have an extra advantage because they can contain a very large amount of metadata, allowing each node and edge to be precisely defined. The powerful model structure used by cytoscape allows this metadata to govern how data is displayed on the network using customisable and interchangeable Vizmaps. I have defined Vizmaps for showing fluxes (page 199, figure 98), flux constraints (page 204, figure 99), flux variabilities (page 216, figure 107) and essential reactions (page 218, figure 109).

Figure 35 shows how cytoscape visualisation can recreate — as a fully editable cytoscape session — the static representations in MPMP. These visual representations require currency metabolites to be defined during creation within MetNetMaker and then require manual layout making them manageable for no more than thirty reactions. Large visual models can be built up from these small networks by combining xgmml (extensible graph markup and modelling language) format files* representing each subnetwork. A script to merge *.xgmml metabolic networks is included in MetNetMaker v1.6+ and is available as the “merge .xgmml visualisations” button.

A key advantage of cytoscape visualisations is that they show all the reactions and metabolites within a model and the layout options make dead-ends and orphan reactions easy to see during model construction. The greatest strength of cytoscape visualisation is to provide a global map — like the one in figure 46 on page 94 — of the final model containing sufficient metadata to allow the projection of the results of simulations directly onto the model. This approach allows considerable insights into the system being studied.

* Cytoscape files (.cys) can be exported to xgmml files without loss of information from within cytoscape. Manipulated xgmml files can be reimported to cytoscape after merging. This round trip and merge retains all information within the network.

Four elements of my *P. falciparum* reconstruction

Whilst both the Huthmacher *et al.* model⁵⁹ and the Plata *et al.* model⁵⁸ have been published since I started my reconstruction, the majority of my reconstruction was complete and my reconstruction process developed before I was aware of these efforts. I have not used any significant elements from these reconstructions, largely because I did not need to but also because of difficulties re-using these models. At the end of this chapter I describe how I have learned from the problems I had trying to re-use these published models to make my model easier to re-use. In this section I will explain the four principal methods I used to complete my metabolic reconstruction, reusing and correcting MPMP, using existing sources of curated data, gap filling with literature and defining custom reactions.

Reusing and correcting MPMP

The malaria parasite metabolic pathways (MPMP) project¹⁰ remains one of the best sources of information on *P. falciparum* metabolism but the raw data is available only as linked pictures like the one shown in figure 35 alongside the accompanying digitised pathway. Digitising these drawings is cumbersome but could have been completed quickly using the advanced searching features in MetNetMaker were it not for a second problem.

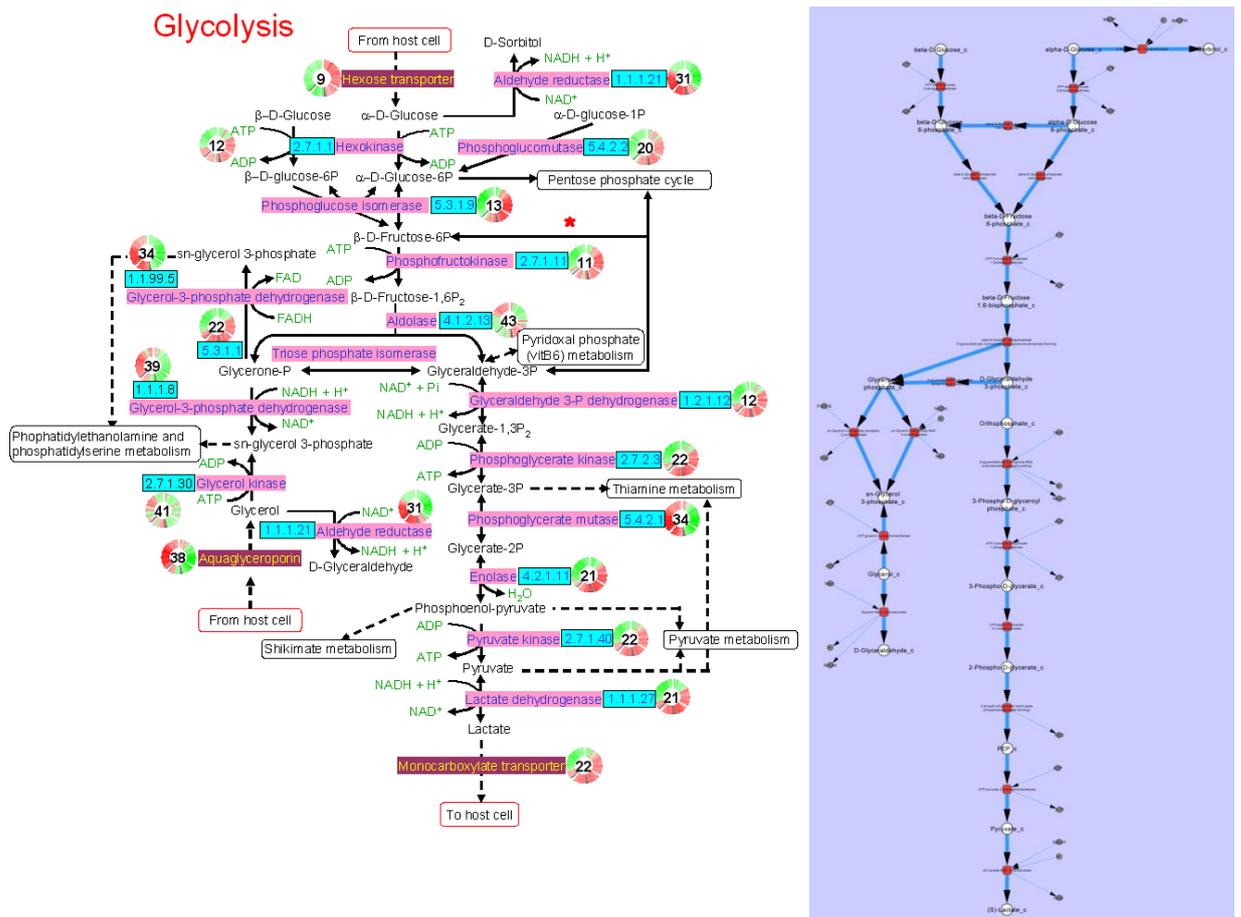


Figure 35 : (left) depiction of glycolysis from MPMP⁸ (right) equivalent diagram within Cytoscape, created from a corrected version of the MPMP network using the 'Visualise in Cytoscape' button in MetNetMaker.

In large part because it is assembled and defined graphically, MPMP is full of errors. As I digitised and corrected these errors I noted and categorised them according to severity. Counting these errors after model assembly showed there to be,

- 62 minor errors such as currency metabolites being wrongly defined or mass not being conserved for a single reaction. These problems could be fixed quickly.
- 57 serious errors such as mass and/or charge not being conserved across a larger set of reactions or required reactions within a pathway being omitted. A few of these problems could typically be fixed in a day, especially where information from other sources was available.
- 22 very serious problems that were difficult to resolve and were critical to the functioning of the network as a whole. These problems often required extensive simplifications or generalisations to be made to my network.

The full list of errors and corrections may no longer be accurate as corrections are made to MPMP frequently but to more fully explain the problem I include — as table 3 — an example of the kind of issues found when digitising an MPMP pathway.

Pyrimidine metabolism - MPMP digitisation issues		
Severity	EC Number	Note
serious	4.2.1.1 and 6.3.5.5	Carbonic Acid issue needs looking into seriously. Here 4.2.1.1 creates carbonic acid (H_2CO_3). This will disassociate by itself in solution to H^+ and HCO_3^- which is then used by 6.3.5.5. I should add this dissociation reaction implicitly.
minor	1.17.4.1	Extra H_2O in KEGG reaction
minor	2.7.4.14	KEGG reaction R02098 is a perfect match but has EC number 2.7.4.9
serious	2.7.4.9	This reaction is already in the pathway one step above. In FBA this is useless so I've omitted it.
minor	2.1.2.1	R00945 is a good match but produces an extra H_2O compared to MPMP.
minor	1.8.1.9 (occurs twice, second occurrence omitted)	R02016 is a good match but produces an extra H^+ compared to MPMP.
minor	1.17.4.1	R02023 is a good match but is much more complex than the MPMP reaction.
minor	6.3.4.2	R00573 is a good match but has an extra H_2O as a reactant.

Table 3 : Issues encountered during the digitisation of MPMP's depiction of pyrimidine metabolism.

Updating or redefining deprecated KEGG reactions

In the introduction I described how the simplicity of the KEGG LIGAND⁴ ontology played a large role in my decision to use it for this reconstruction but as I used it more I realised

that this simplicity comes at a price. As the ontology is updated some reactions are removed, often without a direct replacement. I first noticed this issue when manually entering reactions drawn in MPMP pathways which linked to numbered KEGG reactions that no longer existed. The issue came to light a second time when updating the first draft of my model two years after it was built. The first draft of my *P. falciparum* metabolic network was completed using MetNetMaker 1.0 in early 2010. Just over a year later seven reactions — R02086, R00415, R05919, R05921, R05922, R07461 and R07462 — that were valid at the time of reconstruction were no longer part of the updated KEGG LIGAND database.

Reaction Viewer

Reaction ID: Preferred Name: N-Acetyl-D-glucosamine 1-phosphate 1,6-phosphomutase Pathways:

Reactants:

No.	ID	Compound Name	Compartment	No.	ID	Compound Name	Compartment
1	C00357	N-Acetyl-D-glucosamine 6-phosphate		1	C04256	N-Acetyl-D-glucosamine 1-phosphate	

Figure 36: R02086 as shown in MetNetMaker v1.0. This reaction is no longer present in current versions of LIGAND. No current LIGAND reaction involves C04256 but the compound remains in the ontology.

Reaction Viewer

Reaction ID: Preferred Name: N-Acetyl-D-glucosamine 1-phosphate 1,6-phosphomutase Pathways:

Reactants: Products:

No.	ID	Compound Name	Compartment	No.	ID	Compound Name	Compartment
1	C00357	N-Acetyl-D-glucosamine 6-phosphate		1	C04501	N-Acetyl-alpha-D-glucosamine 1-phosphate	

Figure 37: R08193 as shown in MetNetMaker v1.5+. This reaction replaces R02086 but specifies the product as C04501 in place of C04256. The two compounds are isomers.

It is somewhat reassuring that where reactions are removed from LIGAND their reaction numbers seem to be permanently retired so at least a reaction does not change across different versions of the ontology. The extent of this problem is unclear but with the current LIGAND reaction numbering system going up to R09052 and with only 8135 distinct LIGAND reactions it seems likely that around 900 reactions have been removed from the LIGAND ontology since it was first published.

At first inspection it seems as if a similar purging of the ontology occurs with compounds; in the version of LIGAND used in MetNetMaker 1.5+ compounds are listed up to C13782 but there are only 10873 unique compounds. In considering this discrepancy we must remember that the compounds within the LIGAND database serve purposes greater than just defining reactions; they are also used to define drugs and inhibitors for example. In my experience no compounds that have ever been involved in a defined reaction are removed from the ontology, even if the reaction is removed.

A reaction that no longer exists within a model presents a problem. The ideal solution and the one I have used wherever possible is to update an outdated reaction to its modern equivalent. This often requires updating several reactions to ensure that the new reaction connects properly with its neighbours.

Where replacing reactions has not been possible I have taken advantage of the consistency of the defined compounds to redefine legacy reactions in the new version of the ontology using the original reaction ID. Because these redefined reactions are stored within MetNetMaker as user-defined rather than KEGG-defined reactions they are included in the user's custom reactions file if custom reactions are exported.

This problem is one of many unexpected and frustrating discoveries I have made over the last four years and worked around. Because of KEGG's decision to charge for access to the LIGAND ontology via its FTP service, all current and future version of MetNetMaker are pre-populated with the last free version of the ontology. This has the welcome side-effect that the problem I have just described is substantially reduced as compared to when each installation of MetNetMaker populated itself with the most up-to-date version of LIGAND when it was first run.

Using existing sources of curated data

SharkHunt⁷⁷ was run on version 2.1.4 of the *P. falciparum* 3D7 genome available from the European Molecular Biology Laboratory (EMBL) and returned 289 protein hits with an E-value under 1×10^{-20} (good) and 250 protein hits with an E-value under 1×10^{-30} (very good). These EC hits corresponded to 558 and 447 possible LIGAND reactions at each of the two confidence levels. SharkHunt returns more information than a list of EC hits including the sequence of each hit and the name of the reference sequence it was matched with but since this information is not required for network reconstruction I did not use it.

The list of genes and associated EC numbers was downloaded for the KEGG pfa model from the organism section of KEGG's FTP servers. This contained 616 genes corresponding to 646 unique EC numbers and 908 distinct possible LIGAND reactions. Initially I had used the output of KEGG2SBML⁸⁴ on the KEGG pfa model to convert EC numbers of reaction IDs but it was clear that a number of EC numbers were missed in that parsing script. Using the EC Number-Reaction ID linking table in my ECinfo database — described on page 81 — gave much better results.

The Braunschweig Enzyme Database (BRENDA) database⁹² contains 230 pieces of direct evidence, typically an experimentally isolated enzyme or experimentally confirmed reaction, covering 151 distinct EC numbers for enzymes that exist in *P. falciparum*. An

additional 451 distinct EC numbers are returned using the FRENDA text-mining method and 410 distinct EC numbers are returned using the AMENDA text-mining method. Hits by the three methods — direct evidence, FRENDA and AMENDA — overlap, meaning that in total BRENDA contains at least some evidence for 556 distinct EC numbers corresponding to 1053 distinct KEGG reactions.

In addition to reporting on the presence of enzymes in *P. falciparum* BRENDA contains information on the localisation of some of those enzymes. A BRENDA localisation search returned 107 unique EC numbers with an associated cellular localisation of which 47 were from direct evidence in *P. falciparum*, 18 from direct evidence in other species of *Plasmodium* and 42 from FRENDA text-mining. These 107 unique EC numbers correspond to 295 compartmentalised reactions; there may be evidence for a single reaction occurring in more than one compartment.

The version of PlasmoDB⁹³ — a centralised database of *Plasmodium* genes — in the ECinfo database contains 813 distinct genes associated with 452 distinct EC numbers in turn associated with 1190 distinct LIGAND reactions.

Extra reactions: gap filling with literature

Carbon metabolism^{94–96}, the role of the apicoplast^{97–99}, and amino acid metabolism^{100–103} are areas of metabolism where significant research has been conducted in *Plasmodium* and where ideas are still changing. Fatty acid metabolism is also an area where the behaviour of *P. falciparum* is complicated and I have incorporated results from Mi-ichi *et al.*¹⁰⁴ widely in this area. I have tried wherever possible to include information from these sources in my model.

Defining custom reactions

Sometimes reactions need to be added to metabolic models even when there is little or no evidence for their existence. There are seven reasons why these reactions may need defining.

1. Transport between compartments within the model. These make up the largest number of custom-defined reactions and are largely included without evidence.
2. Exchange reactions that move metabolites to and from the external pool of metabolites that simulates an inexhaustible growth medium.
3. The biomass function. Strictly the biomass function is a reaction in that it appears in the SBML model as such.

4. Reactions defined for simplification. An example is in the fatty acid biosynthesis and elongation pathways where long series of elongation reactions are simplified to the creation of a single fatty acid which is then added to the biomass function.
5. Balanced versions of reactions within LIGAND. In many cases unbalanced reactions in LIGAND can be fixed simply by defining a balanced equivalent.
6. Reactions which occur in the host red blood cell and whose mechanism is beyond the scope of my model.
7. Reactions added through necessity. Usually these reactions are strongly implied, for example where a single reaction is missing from a pathway that is known to be active.

Combining evidence in the ECinfo database

With evidence coming from many different sources it was a considerable challenge to join it together. The ECinfo database was my solution to this problem and it combines all the raw data used to create my final model. The structure of the database, as shown in figure 38, is much looser than that for MetNetMaker because there are none of the same requirements to preserve the integrity of a reconstruction framework. The database structure collects information from all the sources described by EC number and links this to the LIGAND ontology which — like the reconstruction framework within MetNetMaker — is based on Reaction IDs and pathways.

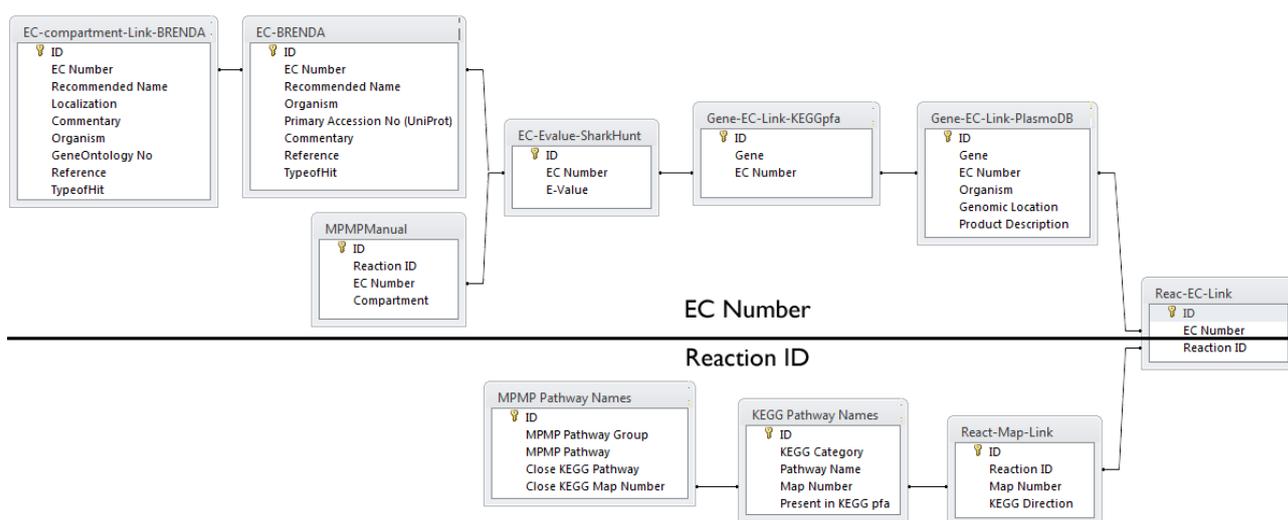


Figure 38: Entity relationship diagram for the ECinfo database as presented in Microsoft Access's Relationships view.

The Reaction ID – EC Number link

The key task of the ECinfo database is providing a reliable link between KEGG Reaction IDs and EC numbers. This is a many-many relationship with the conversion table being created when the KEGG LIGAND database is parsed and imported into MetNetMaker. In

total this table contains 7533 entries referring to 6742 distinct KEGG Reaction IDs. Incomplete EC numbers, eg (2.4.1.-) are linked only to those reactions specified in the KEGG LIGAND ontology and not to all the possible reactions that could be linked to an EC number if the dash, or dashes, in the EC number were treated as wildcards.

Additional information

Beyond the other described sources of information, the ECinfo database contains information from KEGG on the preferred direction, if specified, of each reaction. As discussed in the introduction, all chemical reactions are theoretically reversible and since metabolism in general often uses the same reaction both anabolically and catabolically the preferred direction of a reaction is sometimes related to the pathway it is found in. In order to allow this specification the KEGG pathways associated with each reaction are held in a table within ECinfo and I have additionally created my own table linking the non-standard MPMP pathways names to the KEGG pathway numbering/naming system.

The ECinfo database also contains the same information on whether a LIGAND reaction is balanced as contained in MetNetMaker so that problems with known unbalanced reactions within a network are highlighted twice.

Example reconstruction: nucleotide metabolism

A description of the process for choosing each reaction within my final model would serve little purpose but since the process is shared across the whole network a small example is useful in explaining how the four main reconstruction elements described so far in this chapter are combined. In *P. falciparum*, an excellent place to choose such an example is nucleotide metabolism which is an interesting group of pathways for two main reasons.

Firstly, because we know a great deal about both purine and pyrimidine metabolism, both in humans and in malaria parasites. Whilst humans can both salvage and synthesise *de novo* both purines and pyrimidines, it has been known for over three decades that *P. falciparum* parasites cannot perform *de novo* synthesis of purines¹⁰⁵ and cannot salvage pyrimidines¹⁰⁶. They must therefore rely on salvage of the host cell's purines and *de novo* synthesis of pyrimidines for nucleotide synthesis. These gaps in malaria metabolism are especially good drug targets since human metabolism can use the redundant pathway if one of the two mechanisms of nucleotide synthesis is perturbed. The widely used prophylactic drug atovaquone — used in combination with proguanil and marketed as Malarone® by GSK — perturbs pyrimidine metabolism and the drugs Coformycin, Bredinin, Hadacidin, Allopurinol and the iminoribitol group of compounds target purine metabolism¹⁰.

Secondly — and surprisingly given the importance of nucleotide metabolism — MPMP seems to omit parts of the pathway with considerable evidence for occurring. To highlight this, the left side of figure 39 shows — using the KEGG projector tool — the reactions I created from correcting and importing MPMP’s drawings for nucleotide metabolism. When compared to the equivalent section of the consensus reactions projection at the top right of figure 43 it is clear that there are considerable gaps that can be filled from our existing knowledge.

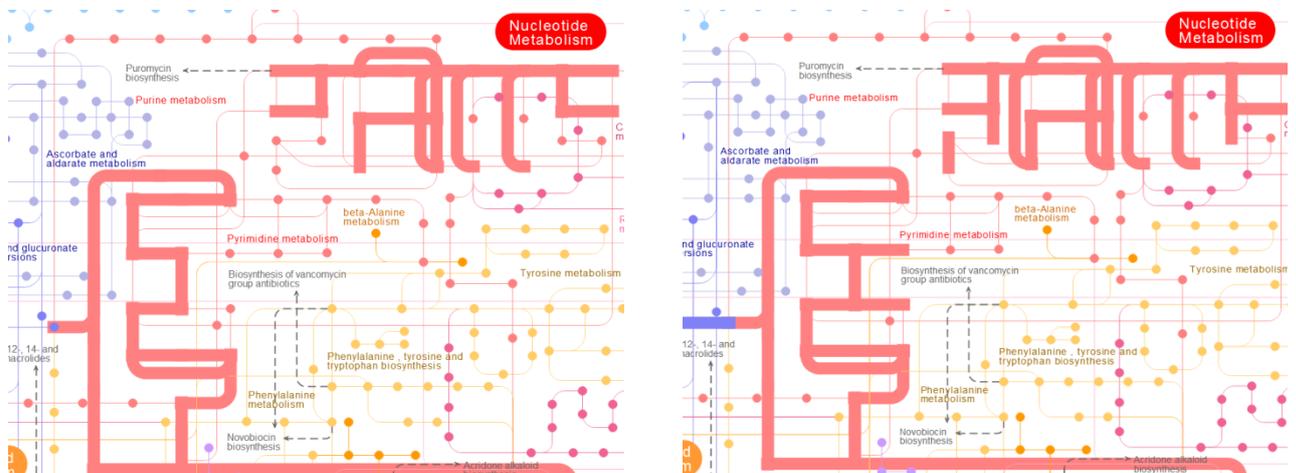


Figure 39: (left) The heavily shaded edges represent the 49 reactions of the nucleotide metabolism section of MPMP as projected onto the KEGG global map in KEGG projector (right) My reconstruction contains 17 more reactions but this is represented as only an additional 4 heavily shaded edges on the KEGG global map.

The following example of how I improved on MPMP’s model of pyrimidine is considerably more ordered than what I actually did but it is a good representation of what I would do if I started my reconstruction now with the knowledge I have gained during my PhD studies. The results of the improvements I have made, as shown on the right side of figure 39 seem small in the KEGG projector view of the network but are more considerable than they appear.

Pyrimidine metabolism as an example of my reconstruction technique

The starting point for my reconstruction is the list of reactions shown in table 4 and I assume that all reactions for which there is a consensus across all my data sources, as presented by the ECinfo database, are worthy of inclusion in my model.

Reaction ID	EC Number
R00139	2.7.4.6
R00156	2.7.4.6
R00158	2.7.4.14
R00377	2.7.7.7
R00378	2.7.7.7
R00442	2.7.7.6
R00443	2.7.7.6
R00512	2.7.4.14
R00570	2.7.4.6
R00571	6.3.4.2
R00573	6.3.4.2
R00575	6.3.5.5
R00965	4.1.1.23
R00969	3.6.1.17
R01665	2.7.4.14
R01867	1.3.3.1
R01870	2.4.2.10
R01876	2.4.2.3
R01993	3.5.2.3
R02016	1.8.1.9
R02018	1.17.4.1
R02024	1.17.4.1
R02093	2.7.4.6
R02094	2.7.4.9
R02098	2.7.4.9
R02100	3.6.1.23
R02101	2.1.1.45
R02326	2.7.4.6
R02331	2.7.4.6

Table 4 : The 29 consensus reactions associated with pyrimidine metabolism (KEGG pathway 00240) from the ECinfo database.

Using the ECinfo database's information on each reaction we immediately see that five of these reactions are unbalanced.

- R00377 and R00378 have both as reactants and products the generic compound DNA (C00039) and are the LIGAND ontology's representation of DNA elongation using dCTP (C00458) and dTTP (C00459) respectively. These reactions can be deleted so long as dCTP and dTTP are included in the biomass function and that the biomass function returns to the cytosol, for each nucleotide consumed, a molecule of diphosphate.
- R00442 and R00443 can be deleted for the similar reason that they refer to elongation of the generic compound RNA (C00046) from CTP and TTP bases respectively.
- R00575 (2 ATP + L-Glutamine + HCO₃⁻ + H₂O <=> 2 ADP + Orthophosphate + L-Glutamate + Carbamoyl phosphate) is an unbalanced reaction; the atom counts calculated by MetNetMaker are shown in figure 40. After checking these atom counts by hand I added a hydrogen ion as a reactant in R00575 to create R00575_balanced and include this in my model instead.

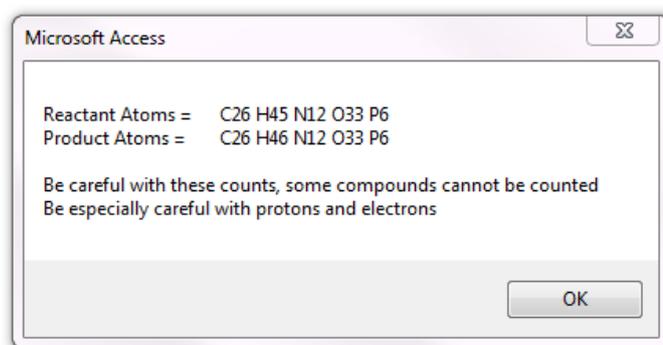


Figure 40 : The report from MetNetMaker on the atom counts for R00575 showing that the reaction is unbalanced. R00575_balanced replaces it in my final model.

The next step in network reconstruction is to visualise the network and see whether there are any orphaned reactions. Selecting Pi (Orthophosphate), P₂i (Diphosphate), H₂O, O₂, CO₂, PRPP (5-Phosphoribosyl diphosphate), ATP, ADP and H⁺ as currency metabolites in MetNetMaker's cytoscape export dialogue and applying cytoscape's circular layout to the resulting visualisation showed three reactions to be orphaned.

- R00139 shares an EC number with R00156 which is not orphaned and is exactly equivalent except for the use of thioredoxin as the charge carrier in place of R00139's ferredoxin. With thioredoxin used elsewhere in the network I chose to delete R00139 and have used thioredoxin as the preferred charge carrier in similar reactions throughout my final network.
- R01876 converts between Uracil (C00106) + Pi (C00009) and Uridine (C00299) + alpha-D-ribose 1-P (C00620). Since we know from the literature that pyrimidines cannot be salvaged this is entirely expected and we can assume that the gene for this reaction is vestigial or serves another purpose. R01876 is deleted.
- R02016 moves charge between thioredoxin and NADP and has no non-currency metabolites. This important reaction is kept since it only appears to be an orphaned reaction when NADP and thioredoxin are aliased and is actually well connected with the network.

During visualisation I observed a dead-end compound which I chose to remove from the network by removing the reaction that creates it. R00969 combines two phosphorylated uracil bases to create UppppU (C06198) in the reaction $UTP + UMP \rightleftharpoons C06198 + H_2O$. This compound, shown in figure 41, is a dead-end that I am not including in my biomass function and whose known biochemical properties¹⁰⁷ are beyond the scope of my model to simulate.

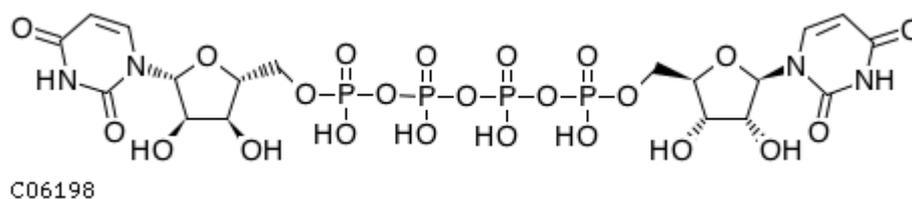


Figure 41 : The chemical structure of KEGG compound C06198, also called P1,P4-Bis(5'-uridylyl) tetraphosphate or UppppU.

At this point I turned to the digitisation of MPMP's depiction of pyrimidine metabolism which I had previously manually converted into 24 reactions. Table 3, earlier in this chapter, summarised the issues I encountered during digitisation of this pathway but for this example we are only interested in the five reactions from the MPMP digitisation that were to this stage not in my model.

- R00132 (4.2.1.1) has evidence for existence from both BRENDA text-mining and PlasmODB and is necessary to provide HCO_3^- from carbon dioxide dissolved in water. I added this reaction to my model.
- R01397 (2.1.3.2) provides N-Carbonyl-L-aspartate to R01993. The reaction did not appear in the consensus model because it is missing from BRENDA but it is present in all other sources and required for the pathway and so included in my model.
- R00945 (2.1.2.1) is omitted from the consensus model for this pathway because it is defined within other pathways (specifically 00670, 00460, 00260, 01100 and 00680) and in this case acts as a link to folate metabolism. It is added to my model.
- R02023 (1.17.4.1). There is only evidence for this reaction from BRENDA and these are text-mining hits. The reaction does not occupy an essential position in the pathway and although the reaction is included in MPMP I have not included it.
- R02325 (3.5.4.3). The only evidence for this reaction is from PlasmODB but given the reliability of this source it seems safe to include it. Without inclusion, conversion between dUTP and dCTP is only possible if the nucleotides are first oxidised which lends weight to its inclusion given the number of other reactions that would otherwise be redundant.

With these three reactions added, returning to the visual representation highlights a further problem. R00132 creates H_2CO_3 from carbon dioxide dissolved in water but the following reaction requires HCO_3^- which forms spontaneously in solution ($\text{H}_2\text{CO}_3 \rightleftharpoons \text{HCO}_3^- + \text{H}^+$). Defining this reaction as CarbonicAcid_disoc and adding it to the network solves this problem.

The next stage in the reconstruction is to flag the compounds in the pathway to include in my biomass function. In the case of pyrimidine metabolism these are UTP (C00075) and CTP(C00063) for RNA synthesis and dTTP(C00459) and dCTP(C00458) for DNA synthesis.

The last stage of reconstruction is to define exchange and transport reactions. For simplicity I have not modelled a separate nucleus or endoplasmic reticulum in the parasite which means that none of the compounds involved in pyrimidine metabolism need transport reactions between internal compartments defining. Exchange reactions simulating the entry and exit into the parasite of compounds are required.

A good guideline to the nutrients available to *P. falciparum* is provided by the exchange reactions in the *Leishmania major* metabolic network of Chavali *et al.*¹⁰⁸ which I examined closely. Further information on relevant exchange reactions is provided from a number of papers^{103,109–117}, from necessity and from my experimental results on nutrient exchange.

For this pathway I have defined exchange reactions for L-glutamate, L-glutamine, H₂O, CO₂, Orthophosphate, Diphosphate, L-aspartate, NH₃, and H⁺. Where exchange reactions are required for multiple pathways they only need defining once and it should be remembered that the fluxes through many of these exchange reactions — particularly those of the amino acids and carbon sources — are restricted by experimental results at the modelling stage.

The size of the task

The techniques I have documented for pyrimidine metabolism are similar to those for other pathways, each with their own difficulties. The scale of the challenge of creating a reconstruction can be shown as the difference between figure 42, showing the reactions for which there is some evidence of existence within *P. falciparum*, and figure 43, the EC numbers for which there is a consensus of evidence across all sources.

The challenge is thus to develop a model somewhere between these two extremes whilst adding the necessary transport and exchange reactions and any simplifications needed to keep the model functional when combined with experimentally measured fluxes.

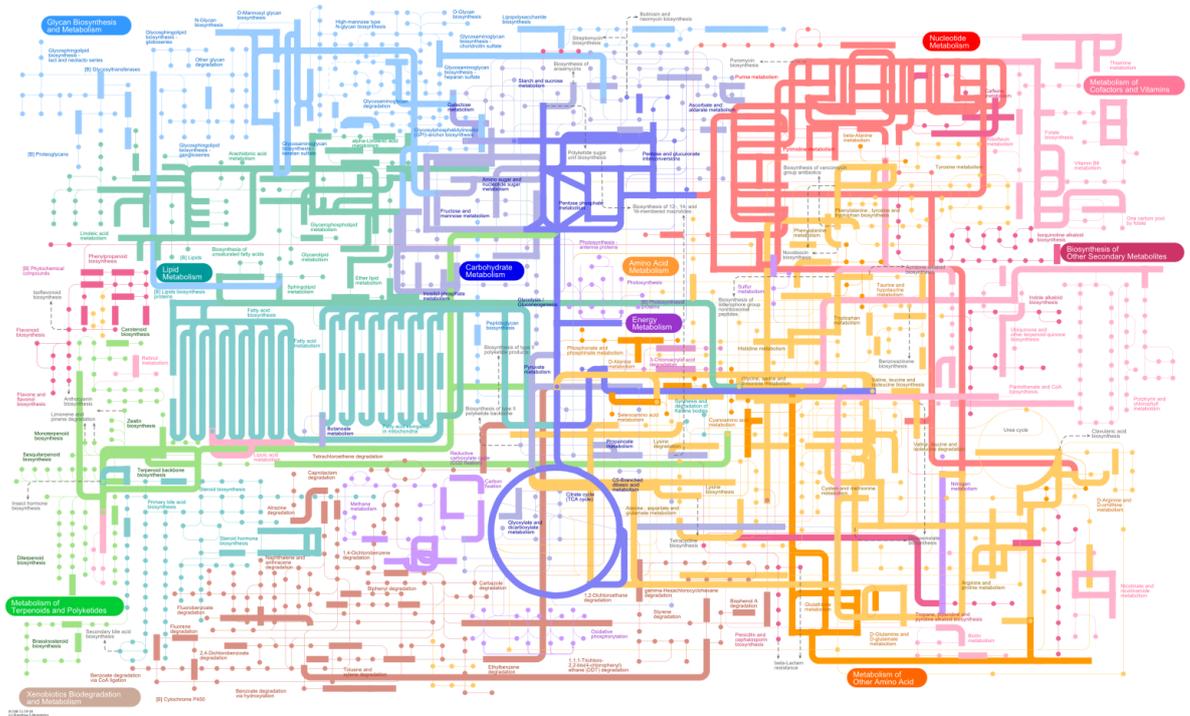


Figure 42: Projection onto the KEGGML pathway map¹¹⁸ of all 1415 reactions with any evidence for existence in *P. falciparum* in my ECinfo database.

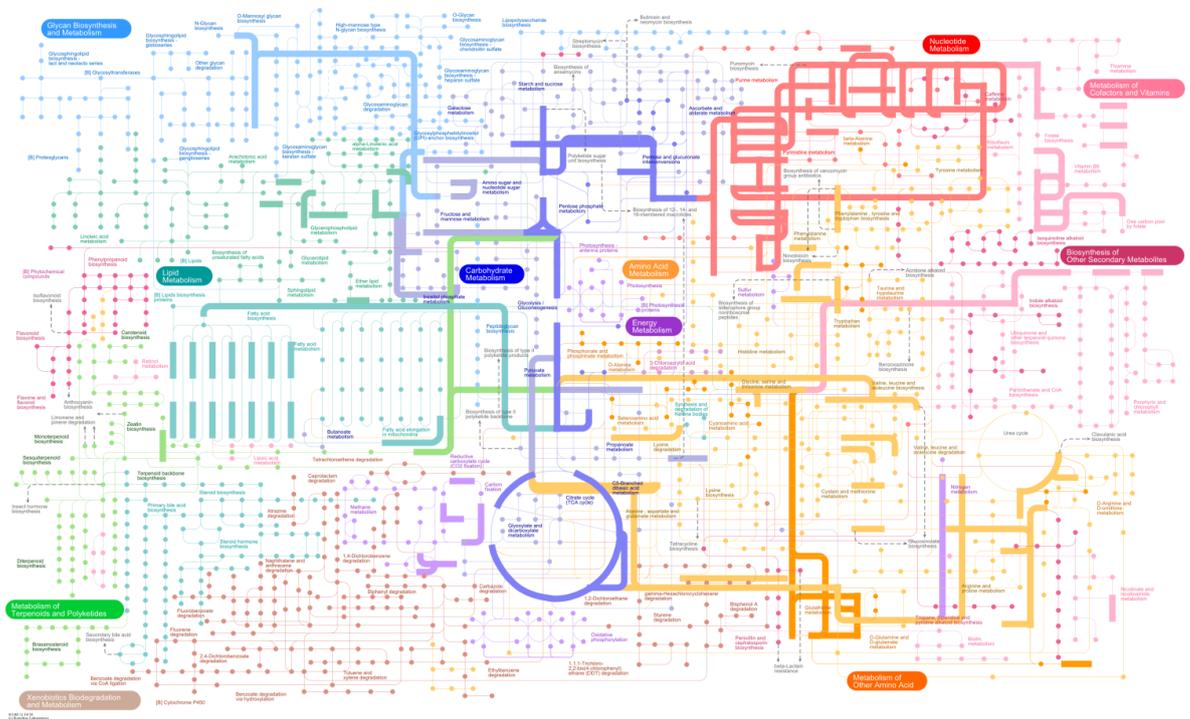


Figure 43: Projection onto the KEGML pathway map¹¹⁸ of all 363 reactions related to an EC number with a consensus of evidence for existence in *P. falciparum*. Nucleotide metabolism (dark pink, top-right) appears more populated than in figure 39 because EC numbers associated with multiple reactions are projected onto multiple edges.

The final model

Model creation ended up being a two-step process. First I built a large model with nearly 500 reactions which I have called the full model and whose properties are summarised in figure 44. I then began reducing it to a core network which I could usefully visualise and manually check to be free of errors.

Notable simplifications in this model are,

- All *de novo* fatty acid synthesis is reduced to a one-step creation of palmitic acid (16:0 fatty acid) in the apicoplast as a direct branch from glycolysis at phosphoenol pyruvate (PEP).
- The simplification of all fatty acid elongation to the addition of a single C₂H₄ unit to a generic fatty acid. This process occurs in the apicoplast as a branch from *de novo* fatty acid synthesis.
- The desaturation of all fatty acids is simplified to the desaturation of a generic fatty acid by a three step process in the cytosol. This step is thought to occur in the endoplasmic reticulum but from a modelling perspective moving it to the cytosol removes around a dozen reactions — without affecting the modelling results — and removes a compartment from the model.
- GDP-L-fucose is selected as the only compound to represent all glycosylation within the model.
- GDP-mannose is selected as the only compound to represent carbohydrate accumulation in the parasite.
- The majority of non-core metabolic reactions are removed from the network, including all reactions ultimately producing compounds not incorporated into my biomass function such as spermidine and putrescine.
- Thioredoxin replaces ferredoxin as the charge carrier for all reactions where either can be used.
- The electron transport chain is not included in the model even though it is known to be functional. This is justified because there is very strong evidence that it serves only to replenish ubiquinone levels to allow the synthesis of purines and not to provide energy via oxidative phosphorylation as per Painter *et al.* ¹¹⁹, “erythrocytic stages of the human malaria parasite *Plasmodium falciparum* seem to maintain an active mitochondrial electron transport chain to serve just one metabolic function: regeneration of ubiquinone required as the electron acceptor for dihydroorotate dehydrogenase, an essential enzyme for pyrimidine biosynthesis”
- Haemoglobin digestion in the food vacuole is substantially reduced in complexity to just five reactions; importing haemoglobin, releasing amino acids and forming

and exporting haemozoin. I took this decision because despite considerable study^{100,103,120-125} our understanding of the processes surrounding haemoglobin digestion remains unclear.

- Hydrogen peroxide, H_2O_2 is produced in the cytosol by pyrimidine synthesis and in the food vacuole by haemoglobin digestion. In many organisms, catalase (1.11.1.6) converts hydrogen peroxide to water and oxygen but there is no evidence that this protein class is expressed in *P. falciparum*. Catalase from the RBC is known to be ingested along with haemoglobin and is active within the parasite's food vacuole¹²⁶. The acidity of the food vacuole reduces the catalase's activity to about one third of its specific activity within the RBC¹²⁷ and the catalase appears to be broken down by proteases within the parasite. This means that some hydrogen peroxide is exported to the parasite's cytosol before being dealt with by catalase but I have not included this behaviour in my model. This behaviour could easily be simulated by forcing a flux of H_2O_2 from the vacuole to the cytosol and this reaction is included in the model for just this eventuality.

This simplified model, which I call the final model, is summarised in figure 45, and at 249 reactions and 143 genes is around half the size of the full model.

A

MODEL OVERVIEW	
Genes	305
Reactions	487
Gene-associated	306
Non-gene associated (intracellular)	89
Non-gene associated (transport)	50
Exchange	42
Input-output	41
Demand (biomass)	1
Metabolites	394
Subsystems (KEGG)	32
Compartments	4

B

METABOLITE CONNECTIVITY	
ATP (C00002)	110
H ₂ O (C00001)	97
ADP (C00018)	78
Diphosphate (C00013)	50
Orthophosphate (C00009)	43
L-Glutamate (C00025)	40
AMP (C00020)	38
H ⁺ (C00080)	35
NH ₃ (C00014)	22
L-Glutamine (C00064)	21
CO ₂ (C00011)	20

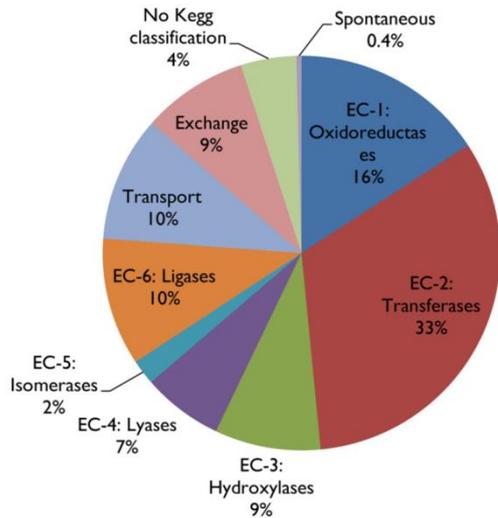
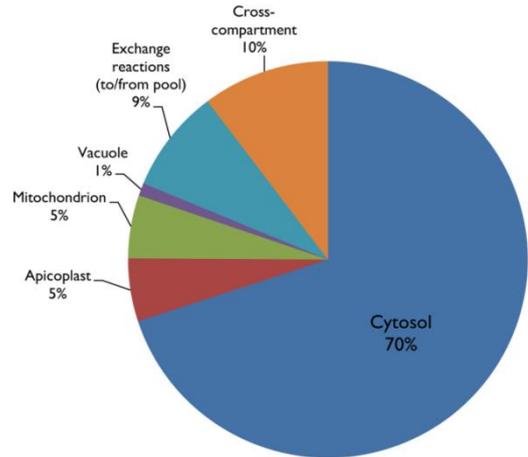
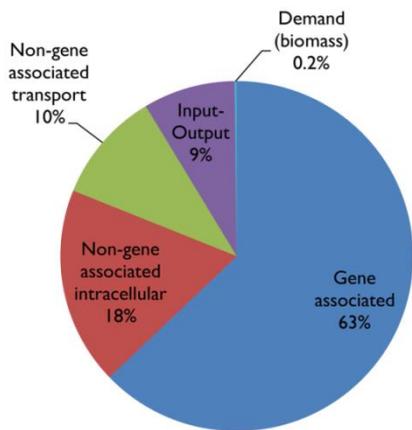
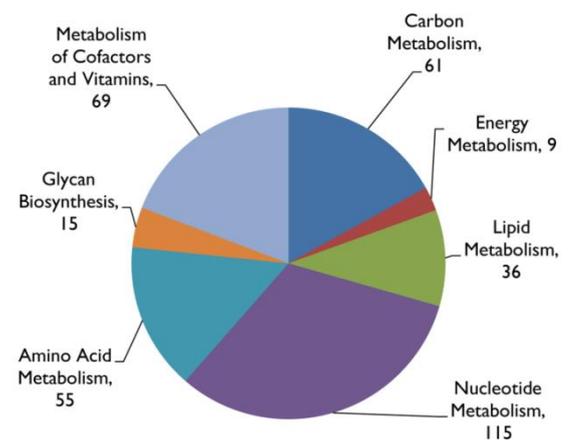
C Enzyme commission (EC) classifications**D Compartments****E Gene-Reaction associations****F Metabolic processes (KEGG subsystem)**

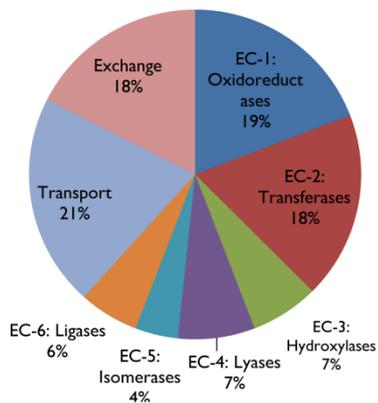
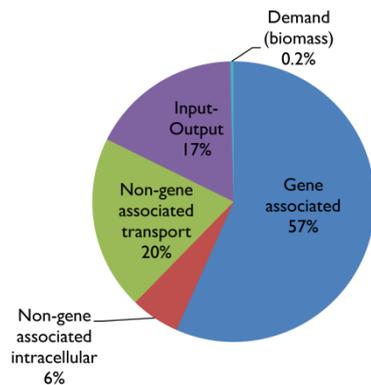
Figure 44 : Characteristics of the "full" metabolic network before reduction to form the final model.

A

MODEL OVERVIEW	
Genes	143
Reactions	249
Gene-associated	141
Non-gene associated (intracellular)	14
Non-gene associated (transport)	50
Exchange	44
Input-output	43
Demand (biomass) *	1
Metabolites (split by compartment)	265
Metabolites (not split by compartment)	163
Subsystems (KEGG)	32
Compartments**	4

* ATP maintenance could be considered an additional demand reaction

** Fatty-acid elongation occurs in the ER but is not placed there for simplicity

C Enzyme commission (EC) classifications**E Gene-Reaction associations****B**

METABOLITE CONNECTIVITY			
Metabolite	Compartment	Absolute	Relative (%)
H2O_c	Cytosol	42	4.7%
ATP_c	Cytosol	38	4.2%
ADP_c	Cytosol	34	3.8%
H+_c	Cytosol	30	3.3%
Orthophosphate_c	Cytosol	22	2.4%
NADH_c	Cytosol	17	1.9%
NAD+_c	Cytosol	17	1.9%
Diphosphate_c	Cytosol	14	1.6%
L-Glutamate_c	Cytosol	14	1.6%
NADPH_c	Cytosol	13	1.4%
NADP+_c	Cytosol	13	1.4%
CO2_c	Cytosol	11	1.2%
L-Glutamine_c	Cytosol	10	1.1%

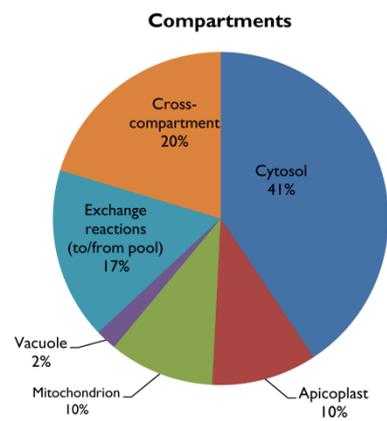
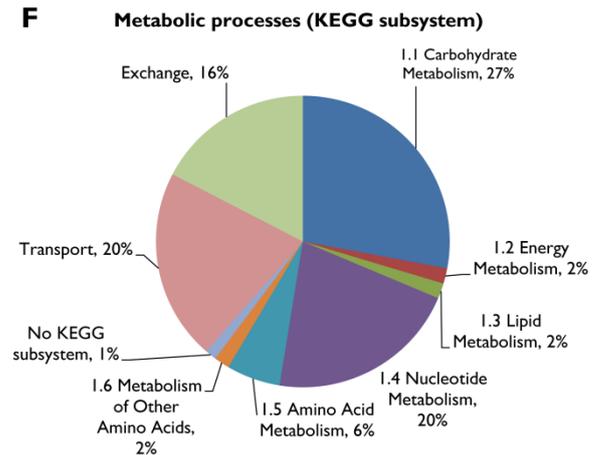
D**F**

Figure 45: Characteristics of the reduced "final" model used for flux-balance analysis.

Size vs accuracy

My final model is small but I think it has significant advantages over other models. That the Plata *et al.*⁵⁸ model contains 1001 reactions versus the 1812 for the combined parasite and red-blood cell model in Huthmacher *et al.*⁵⁹ is not a basis for comparing their quality or usefulness and in my opinion both are much larger than they should be considering our rapidly changing understanding of *P. falciparum* metabolism. Taking the size vs. accuracy argument to an extreme we could take the 1415 reactions shown in figure 42 and add the relevant transport and exchange reactions to create the largest model of *P. falciparum* metabolism to date but it seems unlikely that we would learn anything from doing that.

As discussed elsewhere in this thesis with reference to Olszewski *et al.* 2011⁸⁸ even the functioning of central carbon metabolism remains unclear in the most-studied life stage of the parasite. I think it is foolish to create an extremely large model and introduce further complexities at deeper levels when the basis of such a model is so unclear and those further levels would rest so heavily upon any errors at the core of the model. Likewise, it does not seem useful to define specific products of the metabolic network in the biomass function whilst it remains unclear what is produced by the parasite, not to mention where, when and how.

In comparing the number of exchange and transport reactions in my models with many others it is essential to realise that I have removed fifty reactions from my network. I have achieved this without side-effect by not separating exchange reactions (transfers of compounds between the extra cellular space and a pool external to the modelled system) and cell membrane transport reactions (import from the extra cellular space to the cytosol). Separating these two types of reaction is useful if there are reactions that occur in the extra cellular space but in my models no such reactions occur. Because of this a single reaction can be defined that moves a metabolite from an external pool directly into the cytosol, halving the number of exchange reactions.

Finally, when considering the size of my final model it is worth remembering the visibly redundant reactions shown in figure 31 and figure 32 for previously published *P. falciparum* metabolic models and considering that these and similar reactions increase the size of the network whilst adding nothing to — and sometimes diminishing — its quality. I am sure that my final network contains no completely disconnected reactions and I am almost certain it contains no similar internal errors. I think this level of confidence — combined with a powerful visualisation system — is essential to understanding the true functioning of a metabolic network.

Dynamic visualisation, compartments and fluxes

One of the reasons I can be so confident in the quality of the model is that I have assembled it both mathematically and visually. The full visual representation is shown in figure 46 and additionally with annotations of key features and an example flux projection in figure 47. These maps are fully editable within cytoscape and are coupled to SBML models representing the same system mathematically. Results of FBA can be shown on the map within seconds, highlighting problems that would otherwise escape attention.

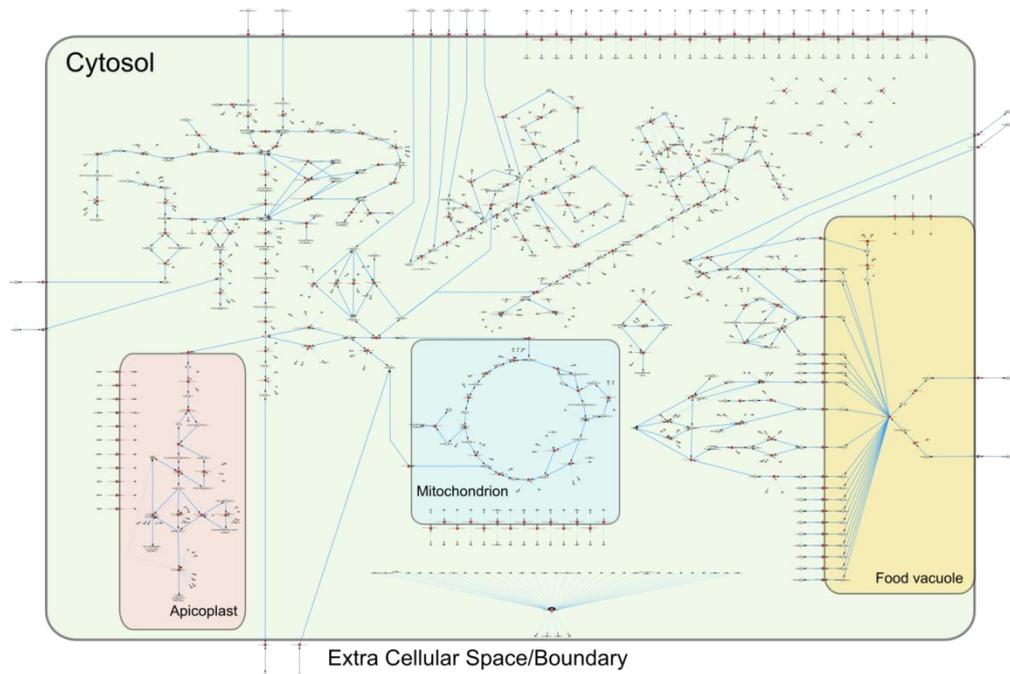


Figure 46: Visual version of the final model ready for flux projection, as in figure 47.

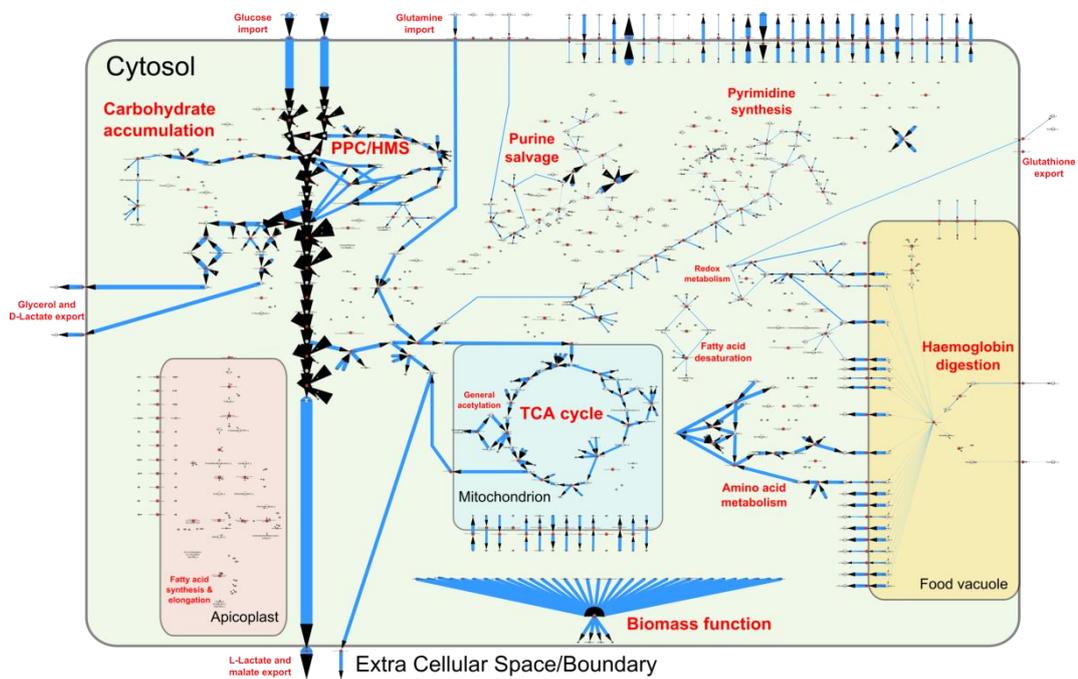


Figure 47 : Final model with live flux projections of the optimal flux distribution given measured constraints. Annotations in red show key pathways and features of the network.

Encouraging model re-use

This final model is arguably the fifth such model to be created for *P. falciparum* and I have spent some time examining how well existing models are being re-used to try and understand whether effort is being wasted by not building on previous achievements.

Plasmocyc⁷ was last updated in 2003 and is no longer relevant. The Huthmacher *et al.* model⁵⁹ is being re-used by the Hoppe group internally but, despite being well cited, a search on Google scholar suggests it has not yet been re-used for published work. Similarly, the Plata *et al.* model⁵⁸ has not, according to Google scholar, been re-used for published work outside of the group in which it was developed and the currently available model does not seem to incorporate some of the recent improvements published by the group. In fact, of all the existing models it is arguably the least easy to re-use, MPMP¹⁰, that has been most widely adopted.

I have thought a lot about this and had useful discussions with Giancarlo Biaggini and Simon Wagstaff, who are currently building on my model at the Liverpool School of Tropical Medicine, to try and make my model easy to re-use in the following three ways.

Using the most standard LIGAND ontology possible

Because Plata *et al.* use the MetaCyc ontology, re-use is difficult in a field where much of metabolism has already been formalised using the LIGAND ontology. The Huthmacher *et al.* model is an improvement in this regard but the widespread simplifications they make to the ontology make the model more difficult to re-use. For this reason I have used the LIGAND ontology for my reconstruction and deviated from it as infrequently as possible, even where this might mean making the model slightly less accurate.

Providing the model in a variety of formats, both visual and mathematical and both as a complete network and in sections

The MPMP model is drawn pathway by pathway meaning it can be visually examined in manageable amounts. The drawback of this approach is that the details of connections between pathways are not well represented. In comparison, Plasmocyc is available only within Pathway tools and the Huthmacher and Plata models are easily available only as SBML files of the whole model without any corresponding visual representation of the network. I think that the lack of a visual representation for the two newest models is a huge barrier to their further use and maintenance.

For my model both individual pathways and pre-defined useful combinations are available in a variety of formats; as MetNetMaker-ready excel format spreadsheets, SBML files,

cytoscape visualisations and as KEGG projections. All of these files are accessible from a single website, currently www.tomforth.co.uk/pfalnetwork , and linked to from my published software's main website at www.metnetmaker.com .

Providing tools to easily edit and update the model

Since MPMP is provided as a collection of hyperlinked images it is not possible to re-use and update computationally. In discussions with Carola Huthmacher at ICSB 2011, Mannheim she mentioned that Hagai Ginsburg provided her with a list of the reactions in MPMP and I have had similar discussions with the LAMP (Liverpool Library of Apicomplexan Metabolic Pathways) consortium in Liverpool. It is unclear why MPMP is not made available publicly in a more computationally accessible format alongside the visual representations but I suspect that the errors that I encountered when manually digitising the pathways may provide some explanation.

The Huthmacher model was assembled from a spreadsheet using custom scripts written by Carola Huthmacher. There is no documented way to update the model except to work backwards from the SBML which itself is described in terms of a mostly undocumented ontology.

The Plasmocyc model can only be edited in Pathway Tools and I have already expressed my reservations with this software. As a result the Plasmocyc model has not been updated since 2003.

The Plata *et al.* model was built using Simpheny but the raw files are not provided and even if they were the software is not within the budget of most researchers.

By comparison my models are distributed ready to edit in MetNetMaker and because they are divided into pathways in the same way as MPMP, the choice of pathways is familiar to people working on malaria metabolism. My models are already being reused in Liverpool and I have tried my best to give them a good chance of being used even more widely in the future.

Links to MetNetMaker

A final goal of mine in constructing my model was that all forms of it should contain as much information as possible. MetNetMaker Excel files of the model contain the full set of information and much of this is exported to the Cytoscape visualisation file but when using the publicly available version of MetNetMaker most of the extra information for each reaction is omitted from the SBML export.

I designed the structure of MetNetMaker (figure 19 on page 51) so that it would be completely independent of the model being reconstructed and there are therefore no direct links between it and the ECinfo database. A link between the two databases is provided by the evidence checker spreadsheet which adds columns to the metabolic reconstruction from the ECinfo database. Examples of this process for the pyrimidine metabolism reconstruction used as an example in this chapter are found as table 33 and table 34 in appendix II.

A key feature of a MetNetMaker pathway file is that any number of extra columns can be added by the evidence checker spreadsheet and it remains readable. This means that adding evidence to pathways does not break the compatibility of the pathway files but it also means that the extra information cannot be exported from a normal version of MetNetMaker.

Since some of the evidence I have added to my model is specific to *P. falciparum* I wrote a special version of MetNetMaker to export the information in the extra columns of each pathway file as part of the SBML output. This took about half a day and I include an example of the extra information in the SBML description of R00156 below as figure 48.

```

<reaction id = "R00156_c" name = "ATP:UDP phosphotransferase" reversible = "true" >
  <notes>
    <html:p>GENE_ASSOCIATION: PF13_0349, PFF0275c</html:p>
    <html:p>PROTEIN_CLASS: 2.7.4.6</html:p>
    <html:p>SUBSYSTEM: 00240</html:p>
    <html:p>SHARKHUNT_EVALUATE: 4.2E-58</html:p>
    <html:p>PRESENT_IN_PLASMODB: True</html:p>
    <html:p>BRENDA_HIT: BRENDA</html:p>
  </notes>
  <listOfReactants>
    <speciesReference species="C00002_c" stoichiometry = "1"/>
    <speciesReference species="C00015_c" stoichiometry = "1"/>
  </listOfReactants>
  <listOfProducts>
    <speciesReference species="C00008_c" stoichiometry = "1"/>
    <speciesReference species="C00075_c" stoichiometry = "1"/>
  </listOfProducts>
  <listOfModifiers>
    <modifierSpeciesReference species="EC_2_7_4_6_7"/>
  </listOfModifiers>
  <kineticLaw>
    <listOfParameters>
      <parameter id="LOWER_BOUND" value="-500" units="mmol_per_gDW_per_hr"/>
      <parameter id="UPPER_BOUND" value="500" units="mmol_per_gDW_per_hr"/>
      <parameter id="OBJECTIVE_COEFFICIENT" value="0"/>
      <parameter id="FLUX_VALUE" value="0" units="mmol_per_gDW_per_hr"/>
    </listOfParameters>
  </kineticLaw>
</reaction>

```

Figure 48 : The portion of the SBML export produced by MetNetMaker for the top reaction in table 33 (appendix II), R00156, showing the mechanism for encoding additional information in SBML level 2 format within the <notes> tag for COBRA compatibility.

Keeping a model up to date

At the end of this chapter I would like to make it clear that I have made a compromise in keeping the model up-to-date. Where experimental results and publications have improved our understanding of *P. falciparum* I have made sure to update my network accordingly and in these contentious areas my network includes data published as recently as the end of 2011. SharkHunt results and the KEGG pfa reconstruction are largely static and do not require updating but PlasmoDB and BRENDA are being constantly updated. I initially used data from these two sources from mid-2009 and updated this information most recently towards the end of 2010.

Experimental Methods

In addition to precise details on the experimental techniques used in this thesis this chapter includes some discussion of the development and assessment of those techniques. Details of actual measurements taken and their analysis is contained within the chapter on experimental results.

Malaria culture

Preparing human red blood cells for culture

Erythrocytes for parasite culture are isolated from rejected donations from the national blood service at St. James' Hospital, Leeds. Blood is still thoroughly screened and donations have typically been rejected for being underweight or having too much or too little of a measured component such as fat, alcohol or salt. Before use, the RBCs must be separated from other blood components which would complicate culture or lead to variability in the nutrient content of the growth medium. White blood cells and platelets are removed using a filter provided with the blood as it is transferred from a donation bag into sterile glassware. In this state the blood partially separates as the serum rises and the RBCs sink. Washing these RBCs aims to isolate them whilst minimising the content of other components.

1. 10ml of blood placed in a 50ml plastic conical tube (falcon tube).
2. Falcon tube filled to 50ml with RPMI 1640 growth medium and shaken.
3. Falcon tube centrifuged at 30-45k RPM in bench-top centrifuge for 5 minutes.
4. A clear divide between the supernatant and compacted RBCs should be clear, the supernatant is removed and discarded.
5. Repeat steps 2, 3 and 4: typically three times.
6. Dilute the compacted RBCs (defined as being close to 100% haematocrit at this point) to double their volume with RPMI 1640 and store at 4°C labelled as "washed RBCs, 50% haematocrit" with the date of the initial blood stock.

Culture maintenance

Plasmodium parasites in culture typically require daily attention but this can be reduced to every other day if the percentage of red blood cells infected with parasites is kept low

(below 1%). A culture with a parasitemia below 1% left in a good condition on Friday night can be expected to be in a recoverable condition on Monday morning but cultures are unlikely to handle such neglect frequently and the following protocol for media change should ideally be completed daily, especially during experiments.

1. Red blood cells are separated from the growth media by leaving the culture flask tilted within its tray and waiting for separation or, especially for larger cultures, by transferring to falcon tubes and centrifuging at 30-40k RPM for 5 minutes.
2. The supernatant is removed. For analysis of used growth medium this is immediately frozen and labelled appropriately.
3. Fresh growth medium, made as described below, is added back to the culture flask.
4. Red blood cells are re-suspended in growth medium.

Complete growth medium

Complete growth medium is prepared by combining the following ingredients in sterile conditions and then filter-sterilising by passing through a filter with a pore size of 0.22µm.

- 45ml RPMI 1640 (Gibco) (optionally supplemented with 0.1g/L hypoxanthine).
- 5ml human serum.
- 0.1g sodium bicarbonate powder.
- 50µL gentamicin solution (10mg/ml stock).

Fresh complete growth medium is a colour best described as salmon and is slightly cloudy. It keeps for one or two days in at 4°C before components of it oxidise, lowering the pH and changing it to a brighter pink colour. By storing prepared medium in full containers with less available oxygen it can keep for twice as long.

AlbuMAX I growth medium

AlbuMAX I growth medium is prepared by combining the following ingredients in sterile conditions and then filter-sterilising by passing through a filter with a pore size of 0.22µm.

- 50ml RPMI 1640 (Gibco) (supplemented with 0.1g/L hypoxanthine).
- 0.25g AlbuMAX I powder.
- 0.1g sodium bicarbonate powder.
- 50µL gentamicin solution (10mg/ml stock).

AlbuMAX I growth medium looks and behaves almost identically to complete growth medium and has the advantage of being identical across batches and thus more likely to give reproducible results across different experiments and laboratories. There are also some problems with using AlbuMAX I growth medium, notably that it moves the *in vitro*

system further away from the *in vivo* system we are aiming to reproduce. Evidence of this has been both in previous studies¹²⁸ although no reduction in growth-rate was reported in the paper which first reported the use of AlbuMAX II growth medium⁴⁰. The composition of AlbuMAX I is unknown and unspecified beyond being a lipid-rich fraction of bovine serum albumen (BSA).

Note on maximum parasitemia and culture “crash”

Cultures with a parasitemia higher than 8% tend to die. In some cases this occurs where glucose levels become critically low and in discussions with Giancarlo Biaggini at the Liverpool School of Tropical Medicine he has mentioned that parasites die in the absence of glucose within 20 minutes. Whilst low glucose levels cause culture death this can be avoided by supplementing the growth medium with glucose or changing the growth medium more frequently. Even where this is done cultures still crash at parasitemia greater than about 8% for unknown reasons; proposed mechanisms include toxin accumulation and even inter-parasite chemical communication/quorum sensing^{129,130,131}.

Note on the requirement for supplementary hypoxanthine

Work on the requirements of a minimal growth medium for cultured *Plasmodium falciparum* by Geary *et al.*^{101,102,132} published in 1985 had already found that,

“hypoxanthine was the preferred purine source for the parasite over adenine, guanine, inosine, adenosine and guanosine although all supported growth equally.”

Geary *et al.*^{101,102,132}

This finding was also reported in 1995⁴⁰ in the first paper on substitution of human serum for AlbuMAX where hypoxanthine was required for parasite growth in AlbuMAX growth medium and the addition of a small amount of hypoxanthine to some examples of human serum that would otherwise not support parasite growth allowed parasites to grow.

“One important finding was the absolute requirement for hypoxanthine in RPMI-A [AlbuMAX II growth medium] to enable good parasite growth.”

Cranmer *et al.* 1995⁴⁰

There is a further discussion of the parasite’s uptake of hypoxanthine from the growth medium in the next chapter.

Transition of cultures from complete medium to AlbuMAX I

Mixtures of complete medium and AlbuMAX I medium are defined according to the percentage of the AlbuMAX I medium. For example, growth medium referred to as 50% AlbuMAX is half complete medium and half AlbuMAX I medium. For simplicity of

preparation, 90% AlbuMAX can be prepared by adding 0.5ml of human serum to prepared AlbuMAX growth medium.

Immediately transitioning cultures from complete medium to 100% AlbuMAX I medium has been observed to slow parasite growth-rate and even cause the death of a culture. The reasons for this are unknown but can be avoided if the amount of human serum within the growth medium is reduced to nothing over the course of a week. The observations of *Plasmodium falciparum* cultures adapting to changing conditions over a period greater than a full life cycle seems bizarre. Beyond the set of transcribed genes and expressed proteins inherited by merozoites from their mother cell I can think of no good mechanism for what appears to be some signalling between generations in culture and yet this property of slowly adapting to new conditions is well known to occur. A good published example is of parasites recovering from being deprived of hypoxanthine in their growth medium.

"It required weeks to months of meticulous culture for the parasites to adapt successfully and to restore good growth kinetics in the absence of hypoxanthine."

Cranmer et al. 1995⁴⁰

Observing cultures

Blood smears

Cultures of *P. falciparum* are still observed in much the same way as Alphonse Laveran first observed the parasite in Algeria around twenty years before his discovery and subsequent work won him the 1907 Nobel Prize¹³³.

"Films of blood are stained by Leishman or Giemsa stains and the red cells are examined..."

Medical Microbiology, G. Thomas, 1963¹³⁴

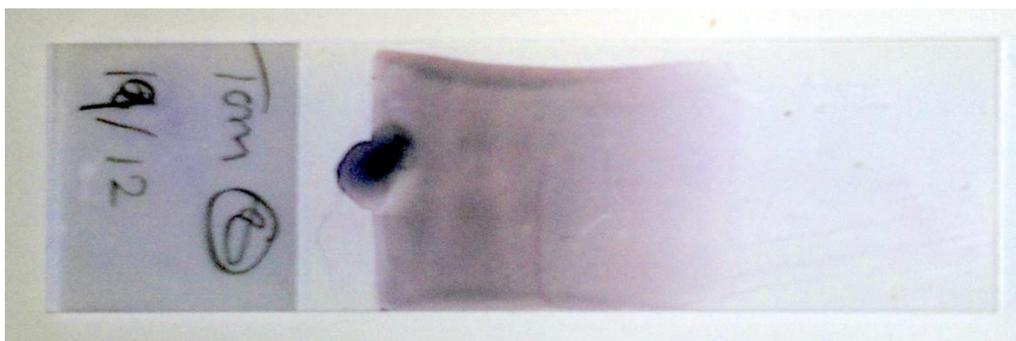


Figure 49: Thin blood smears are made from a single drop of blood pushed quickly across a glass slide. The red blood cells at the edge of the smear are one layer thick.

In our lab we take thin blood smears and use Giemsa stain, which binds specifically to the phosphate groups in the parasite's DNA — red blood cells (RBCs) have no DNA — whilst being able to easily diffuse across multiple membranes to reach the parasite nucleus. The protocol for making blood smears is the same as for changing the growth medium and the two are usually performed at the same time.

1. The culture flask is tilted and left until the RBCs have mostly separated from the growth medium.
2. If the growth medium is being changed it is removed.
3. The smallest possible drop of packed RBCs is removed from the flask and placed on a labelled glass slide.
4. A second glass slide is used to quickly push the drop of blood out into a thin smear with only a single layer of RBCs towards the edges of the smear.
5. Once dry, the smear is fixed in 100% methanol and can be stained immediately or left for staining later.
6. The fixed slide is stained for 10 minutes in Giemsa's stain, diluted 1 part in 5 parts Sorenson's buffer. Slides are held upside-down in a staining tray so that any Giemsa crystals falling out of solution do not stick to the slide.
7. Once staining is complete, slides are immediately de-stained under running water for as long as necessary, usually a few minutes.
8. Dry slides are ready to be examined under an optical microscope.

The parasitemia of a culture is defined as the number of parasites divided by the number of red blood cells in that smear. Under the oil-immersion microscope used in our lab (x100 object lens magnification in series with an x6 magnification eyepiece lens = x600 total magnification) a tightly-packed monolayer of red blood cells contains approximately 250 red blood cells. Usually ten fields are counted though more may be necessary to get an accurate reading at low parasitemias. For all reported parasitemias all slides were read blind, that is to say without knowledge of the expected parasitemia or life stages present. I found this invaluable in preventing a positive bias creeping into my results and it is a technique we have all adopted in the lab when taking measurements for analysis.

Where it is necessary to count not only the number of parasites but also the number at each approximate life-stage the generalisations set out by Coatney in "The primate malarial" ¹³⁵ are generally adopted, These classifications are shown in figure 50 where drawings 1 through 26 might be interpreted as the progression of the parasite through equal time steps over the full 48 hour life cycle. Since each person counts differently these life-stage drawings can only be considered a guide and I report an accurate measurement

of my counting later in this chapter in the section on “objectively re-assessing sorbitol synchronisation”.

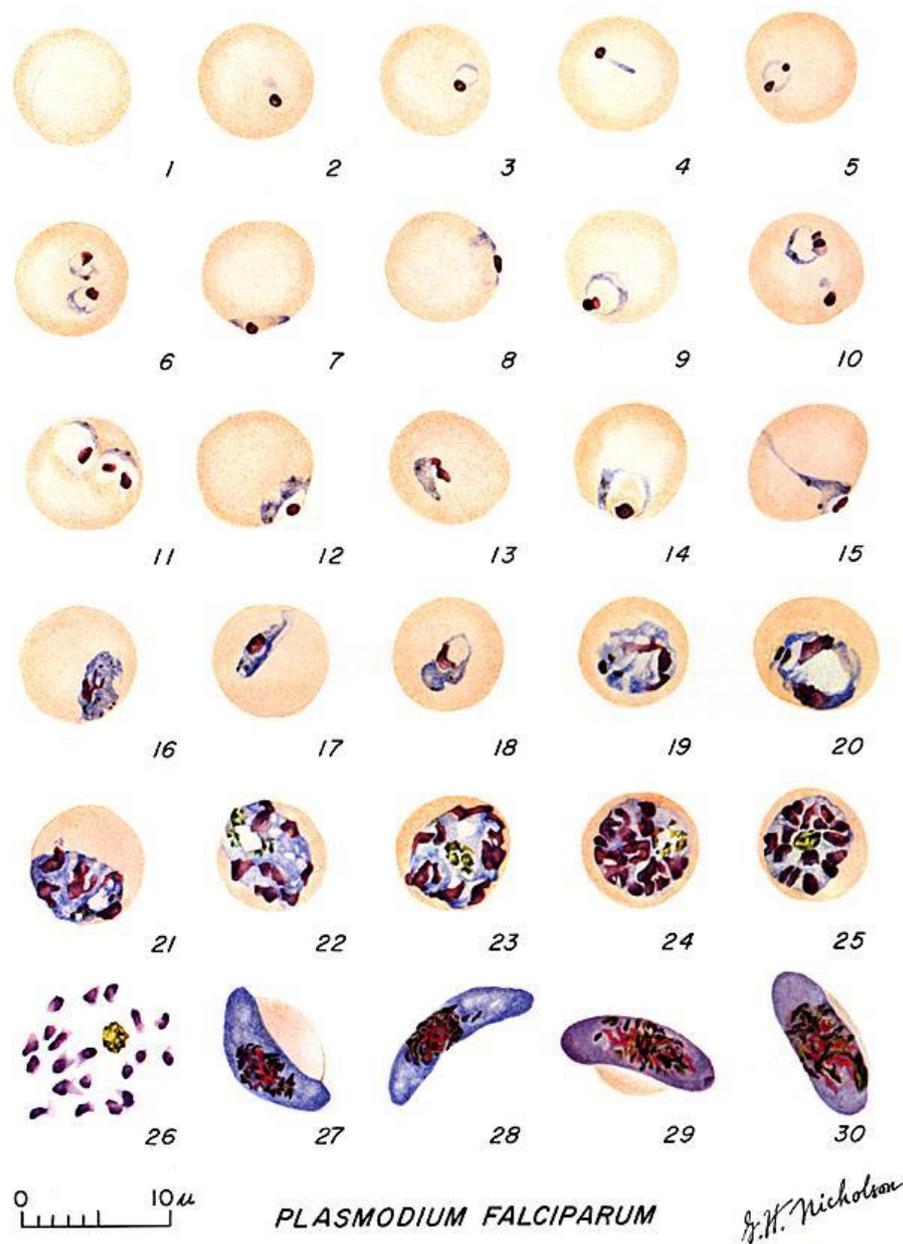


Figure 50: Categories used to define the life stages of *P. falciparum* in the microscopy time course. **1:** normal red cell. **2-10:** early ring trophozoites. **11-18:** late trophozoites. **19-26:** schizonts. **27-30:** gametocytes (not observed in our system of continuous culture). Image: Plate XLII, *The primate malarias*, 1971, G. R. Coatney¹³⁵. Similar illustrations by G. H. Nicholson for 26 other plasmodium species available from *Images from the History of Medicine*¹³⁶.

Potential problems visible in blood smears

Previously in our lab we have had problems with airborne yeast infections which were not visible in stained blood smears. I was fortunate to avoid these problems and since thoroughly disinfecting the sterile culture extraction hood these problems have not returned.

Many problems in culturing can be diagnosed from examining blood smears, in particular bacterial infection as shown in figure 51. Bacteria quickly kill the parasites within a culture but can be tolerated at very low levels. Infections can often be removed if they are caught early by adding gentomycin and are less likely to occur if gentomycin is used continually. Although gentomycin has no known effect on *P. falciparum* metabolism it is often preferable to take great care to keep sterile when handling cultures rather than add a further variable to an experiment. The cultures for all the metabolomics experiments reported in this thesis were cultured without gentamycin in strictly aseptic conditions.

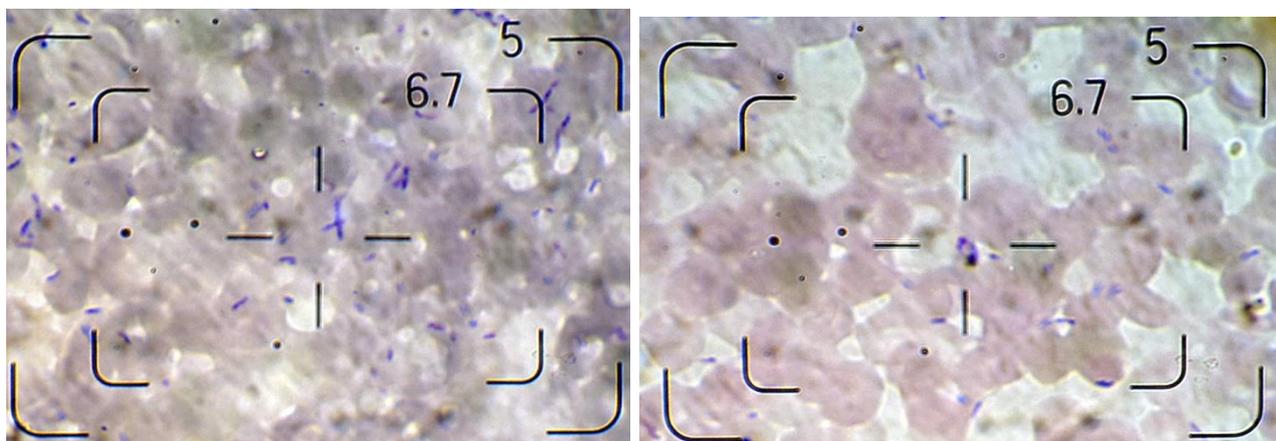


Figure 51: left: Rod-shaped bacteria are clearly visible in a Giemsa stained blood smear. right: A healthy ring-stage parasite in an infected culture shows that minor infections are not always fatal.

Optical fluorescence microscopy

Optical fluorescence microscopy was used only to confirm that acridine orange dyed infected RBCs selectively. Slides were prepared according to the following protocols.

Unfixed fluorescent slides

1. Start with 100 μ l of mixed culture (5% haematocrit, 95% growth medium).
2. Add 1 μ l of acridine orange dye at 50 μ g/ml (diluted x200 from stock at 10mg/ml) to a final dilution of 0.5 μ g/ml.
3. Place on a glass microscope slide and cover with a glass cover slip.

Hybrid fixed fluorescent slides (limited success)

1. Start with 100 μ l of mixed culture (5% haematocrit, 95% growth medium).
2. Add 1 μ l of acridine orange dye at 50 μ g/ml (diluted x200 from stock at 10mg/ml).
3. Incubate for 3 minutes in the dark.
4. Make long smears (slowly push the dyed culture over a slide), dry in the dark and fix with methanol.
5. Stain with Giemsa.

Flow cytometry

Preliminary experiments using flow cytometry were performed on a PRS III flow cytometer which proved effective but extremely unreliable. All reported experiments were performed on a BD FACSCalibur flow cytometer from Beckton Dickinson Biosciences. Acquired data was analysed at the time using the software supplied with the relevant flow cytometer as set up by Gareth Howell, director of the FBS flow cytometry facility. Post-acquisition analysis, and the preparation of figures for this thesis, was performed using the Cytomation Summit 4.2 software from Dako.

The development of a technique to measure parasitemia by flow cytometry is discussed in detail in the next chapter. Although we had limited success it seems likely that with the right experimental setup the technique would be more successful. Our experiments suggest the following protocol is most likely to yield good results.

1. Start with 40µl of mixed culture (5% haematocrit, 95% growth medium).
2. Add 1µl of acridine orange dye at 50µg/ml (diluted x200 from stock at 10mg/ml).
3. Incubate in the dark for 3 minutes.
4. Add 1.5ml PBS (phosphate-buffered saline) and acquire.

If an accurate and reliable system for counting parasites using flow cytometry could be developed it should be possible to use fluorescence-assisted cell sorting (FACS) to create extremely high parasitemia cultures for — amongst other potential uses — high-precision metabolomics. The drawback of this technique is that the effects of the intercalating acridine orange dye on metabolism are unknown and the technique seems considerably more complex than the magnetic isolation technique of parasite concentration which I discovered late in my project and describe later in this chapter.

Synchronisation of cultures

Female *Anopheles* mosquitoes typically bite in the evening and for this reason it would seem advantageous for *P. falciparum* to time the production of gametocytes so they were at their highest population within the bloodstream at that time. Some support for this hypothesis is provided by recent studies¹³⁷ showing that *Plasmodium* parasites in mice (*P. chabaudi*) grow best when their 48-hour circadian rhythm is synchronised with the 24-hour circadian rhythm of their host.

No mechanism for this possible host-parasite circadian rhythm synchronisation has been proved but is clear that whether or not such a system exists *in vivo* it does not exist *in vitro*. Cultured parasites show no natural synchronicity of life stages and life-stage-

synchronised cultures return to this state in a matter of weeks. Most studies are performed on synchronised cultures where the majority of parasites are at the same stage in their 48-hour life cycle together.

Synchronisation presents advantages in two major areas for study. Firstly, when trialling new drugs or growth conditions, it allows the stage-specific impacts of any perturbations to be observed. Secondly, if metabolite exchange between the growth medium and the parasites is being measured it is possible to minimise contamination of the growth medium with the contents of lysing RBCs by ensuring measurements are not taken during parasite release.

There are many ways to synchronise parasite cultures and excellent reviews on the subject exist^{138,139}. Here I include details on the sorbitol incubation method we use in the lab and on a magnetic isolation method which we have not used but which holds considerable promise for further research.

Sorbitol incubation

The sorbitol incubation method as described by Lambros and Vandenburg in 1979¹⁴⁰ works by selectively killing schizonts and late-stage trophozoites by submitting them to the high osmotic pressure of otherwise non-toxic high concentrations of sorbitol. For short periods, early-stage trophozoites (rings) can withstand this osmotic pressure. Sorbitol synchronisation is performed according to the following protocol.

1. Transfer the whole culture to be synchronised to a 50ml falcon tube.
2. Centrifuge at 30-45k RPM in a benchtop centrifuge for 5 minutes.
3. Remove and discard the supernatant.
4. Add half the volume of the removed medium of 5% w/w sorbitol solution in water to the compacted RBCs.
5. Incubate for exactly 5 minutes at room temperature.
6. Add excess growth medium to restore osmotic pressure.
7. Centrifuge the falcon tube at 30-45k RPM in a benchtop centrifuge for 5 minutes.
8. Remove supernatant and replace with growth medium for continued culture.
9. Transfer the mixture back to a culture flask and treat as usual (gas and incubate).

Synchronisation typically kills slightly over half the parasites in a mixed culture and parasitemia will not increase until around 40 hours later, at which point it will increase by the usual daily growth-rate squared. Further synchronisations, as shown in figure 52 can produce a tighter synchronisation window at the cost of the extra time needed to achieve a high parasitemia.

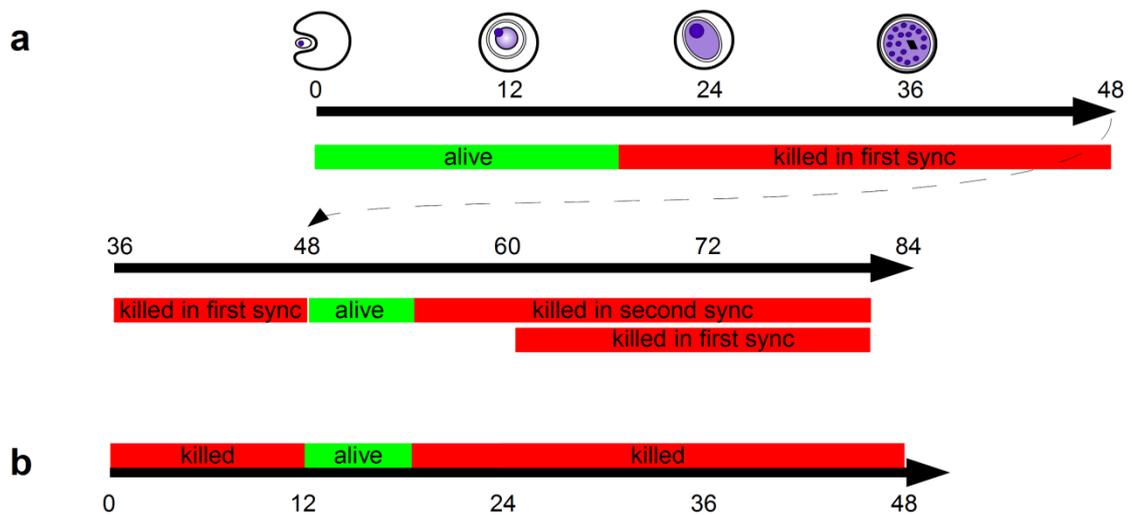


Figure 52: **a)** A second synchronisation performed 36 hours after the first leads to a tighter final synchronisation profile as shown in **(b)**.

Objectively re-assessing sorbitol synchronisation

Initial observations by Lambros and Vandenburg on what proportion of — and at precisely what times in the life-cycle — parasites are killed by sorbitol incubation are vague,

Cultures examined immediately after sorbitol treatment consisted almost entirely of single and multiple ring-form infections, and uninfected RBCs.

Lambros and Vandenburg¹⁴⁰

That paper later suggests a second synchronisation 34 hours after the first as the best time to fine tune and achieve a narrower synchronisation window, suggesting that they experienced survival of parasites up to 16-20 hours post invasion. Since the existing literature is unclear on this I decided to make a more accurate estimation based on a large experiment I conducted in December 2010. Seven identical cultures were grown in parallel and synchronised at $t = -46$ hours, and $t = 0$ hours. Blood smears of all cultures were taken approximately 24, 46 and 86 hours later with a selection of cultures being examined at times in between. The parasitemias of the cultures was measured and the proportion of each life-stage — rings, late-stage trophozoites or schizonts — was determined.

The results of these tests are shown in figure 53 and the results are consistent — as shown by the dashed-line fits — with the combination of the two synchronisations killing the majority of parasites outside of a 14-hour window as shown in figure 54a. Since the second synchronisation took place 46 hours after the first synchronisation the 14-hour

period is consistent with a single synchronisation window being approximately 16 hours, at the low end of the range suggested in Lambros and Vandenburg¹⁴⁰.

The dashed-line fits in figure 53 are particularly interesting to me as an example of a biological system that seem complicated but whose behaviour can be explained with extremely simple rules. These dashed lines are just the convolution of the two functions shown in figure 54a and figure 54b performed using a simple script and the `conv()` function in Matlab. The convolution function exactly represents the progress through time of an applied synchronisation window. The following parameters in the source functions started with a sensible guess and were adjusted by hand to provide a good fit of the data,

- Duration of the synchronisation window,
- Smoothness of the edges of the synchronisation window representing the imperfection of the synchronisation technique,
- Time at which the transition between each of the four distinguished life stages occurs.

It is likely that a slightly better fit could be made by further refining the parameters but within the uncertainties of the measurements this would provide us with no useful information. In addition to letting us elegantly quantify the progression of a population of *P. falciparum* through its life stages and giving us a more accurate measurement of the effectiveness of sorbitol synchronisation these results provide an objective measure of what times in a parasite's life I have classified them to each named stage.

Life-stage	Visible period
Merozoites*	0-1 hours
Rings	2-16 hours
Trophozoites	17-37 hours
Schizonts	38-47 hours

Table 5: Age of parasites since infection at which I have classified them at each distinct stage.

*We know that merozoites re-invade extremely quickly after they are released so this 2 hour period is actually best interpreted as the time after successful invasion but at which the ring form is not yet sufficiently mature to be visible in a blood smear.

After exiting the host, merozoites recognize, attach, and enter RBCs, and this process occurs rapidly, probably on the order of 60s.

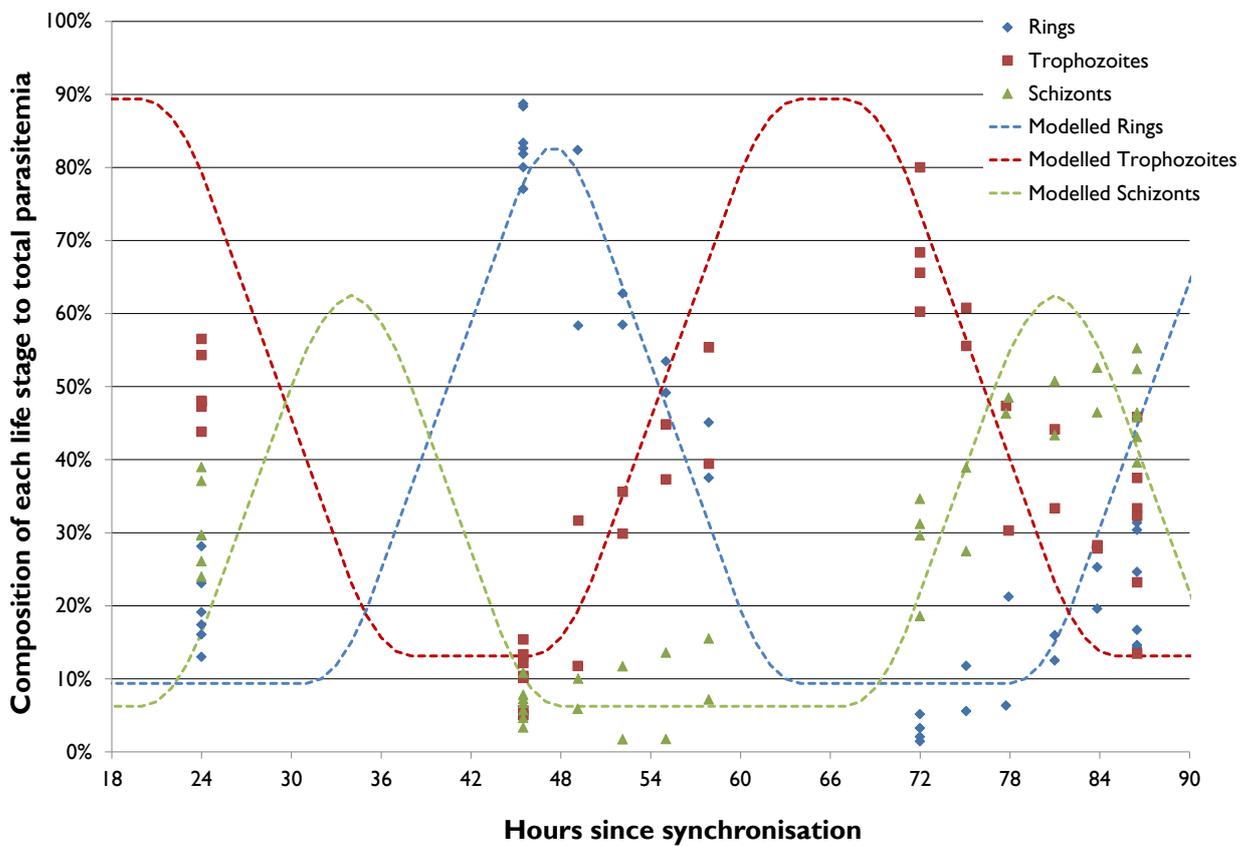


Figure 53: Progression of life stages for seven identical cultures in AlbuMAX I 100% growth medium synchronised at $t = 0$ hours and $t = -46$ hours. Lines modelling life stage progression are the convolution of figure 54a and figure 54b.

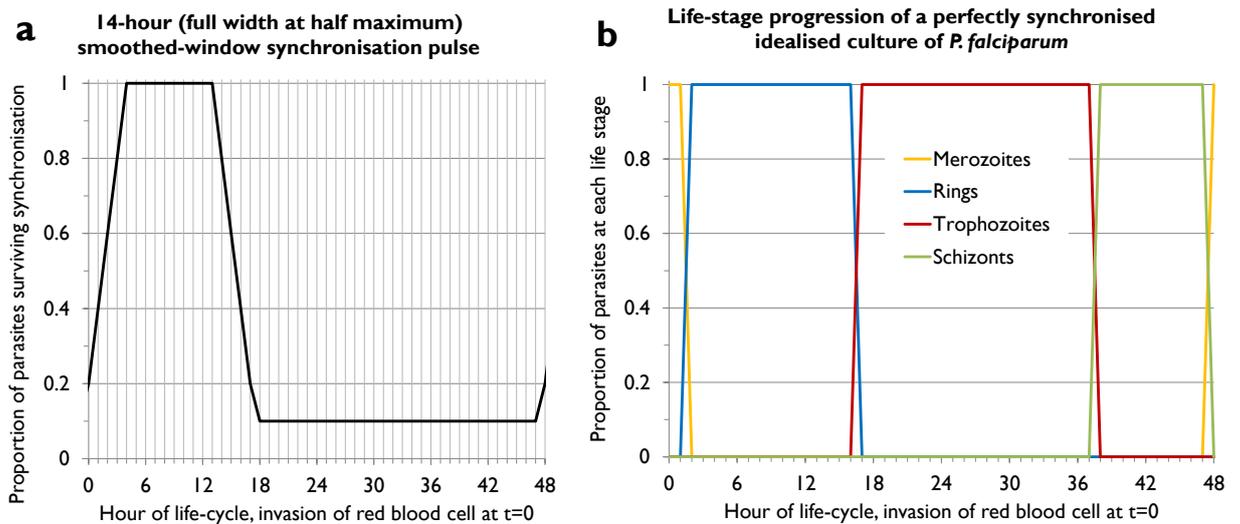


Figure 54: **a)** Mathematical representation of a 14-hour synchronisation pulse killing 90% of the targeted life-stages. **b)** 48 hours of the idealised life-stage progression that best fits results in figure 53. This figure would continue infinitely in both directions.

Synchronisation and concentration of parasitised red blood cells using magnetic isolation

I was introduced to the magnetic isolation method as described in Ahn *et al.* 2008¹⁴² by PhD student Murad Mubarak and his supervisor Giancarlo Biaggini at the Liverpool School of Tropical Medicine. The technique works because the crystallised haem groups (haemozoin) formed as a by-product of haemoglobin digestion by the parasites at their later life stages is paramagnetic¹²⁰ and therefore late-stage trophozoites and schizont infected RBCs respond differently to an applied magnetic field than uninfected RBCs and early-stage trophozoite infected RBCs.

Magnetic isolation is particularly interesting because in addition to creating a highly synchronous culture it can also concentrate infected RBCs up to a parasitemia of around 90%. At this parasitemia cultures are liable to crash quickly but the increased concentration of parasites allows for more sensitive measurements of parasite metabolism. I do not include a protocol for this procedure beyond the reference to Ahn *et al.* as I have not used it and mention it only for completeness and because it may be useful for future experiments.

Saponin lysis and parasite biomass isolation

Parasite biomass must be separated from red blood cell biomass before any analysis of it can be performed. This process — as described in Ashong *et al.*¹⁴³ — consists broadly of three steps, lysing the red blood cells using saponin, centrifuging the resultant mix at very high speed to separate parasite biomass and then washing away the red blood cell membrane components. Tubes were made RNA-safe by washing with phosphate-buffered saline (PBS) with 0.1% diethylpyrocarbonate (DEPC) added, the tubes were then autoclaved to ensure sterility and deactivate the DEPC before use. The following protocol was used.

1. Place culture to lyse in a 50ml RNA-safe falcon tube and spin at 4500RPM on the benchtop centrifuge for 5 minutes.
2. Remove the growth medium and re-suspend the pellet in 5 times the pellet's volume of Tris-buffered Saline (TBS).
3. Add stock saponin (15% w/w in water) to a final dilution of 1 in 100 and agitate for \approx 30s until the culture goes from a red colour to a deep purple, signifying lysis has occurred.
4. Centrifuge for 10 minutes at 10k RPM in the large Beckmann centrifuge cooled to 4°C using RNA-safe tubes.

5. Remove supernatant and discard.
6. Add a small amount of TBS and gently wash away the white-coloured red blood cell ghosts, remove supernatant and discard.
7. Resuspend all remaining parasite components in a small amount of TBS and transfer to RNA-safe 2ml micro-centrifuge (eppendorf) tubes.
8. Centrifuge eppendorf tubes containing parasite biomass at 10k RPM for 10 minutes in the benchtop centrifuge.
9. Remove supernatant and any remaining red blood cell ghosts.
10. Add ten times the pellet volume of RNAlater®(Invitrogen) and freeze at -80°C immediately.

Problems working with large culture volumes

Initially I planned to measure biomass components using the adapted TRI REAGENT (Sigma) protocol described in Daniel Opi's masters thesis¹⁴⁴. It seems likely that a large reason why he was unable to make accurate measurements of the RNA, DNA and protein content of the total parasite biomass was because the quantities of biomass he was measuring were too small. To overcome this problem I started examining options for culturing with up to 1 litre total culture volume and around 180ml in each culture flask.

The principal problems I encountered with culturing in very large flasks were,

- Tilting flasks to separate haematocrit and growth medium took a long time and I often removed cultures from their flasks and centrifuged them to ensure that growth medium was properly changed without accidentally removing haematocrit.
- The gassing procedure — where air with only 3% oxygen is injected into flasks to replace atmospheric air — took much longer and seemed less successful.
- Infections seemed more likely, probably due to the greater difficulty in dealing with the cultures and the longer times they were manipulated for.
- Growth seemed to slow even further, from a 1.6 daily growth-rate to closer to 1.4.
- Preparing and filtering large quantities of growth medium required special equipment including a pump-attached filtration bottle.

Measurements techniques

Used growth medium assays

Collection of samples

My usual technique for taking used growth medium samples was as follows,

1. Shake flask to homogenise the culture.
2. Remove 0.5-1.0ml of culture. The precise amount depends on the number of time points and the volume of the culture. As little as 0.1ml of culture could easily be used now that the techniques for sample analysis are established. Gas the culture and place back in the incubator.
3. Centrifuge to separate pellet from supernatant whilst avoiding precipitating any supernatant contents. Usually 5k RPM for 3 minutes on a small benchtop centrifuge will do this but higher speeds for higher durations are unlikely to cause any dissolved contents to fall out of the supernatant.
4. Remove the supernatant and freeze at -80°C ready for later analysis.
5. Re-suspend the pellet in the remaining supernatant and make a blood smear.
6. If required, saponin lyse the pellet to isolate parasite biomass and freeze at -80°C for HPLC analysis of biomass components.

Special case for Sara metabolomics set of experiments

The above technique has the advantage of maintaining the same haematocrit concentration throughout but involves disturbing the culture every time a sample is collected. In the Sara metabolomics set of experiments — see table 20 on page 157 for a full summary of metabolomics experiments — I modified the technique to increase sensitivity and avoid disturbing the culture. In this modified technique, at each time point the culture flask was tilted for 15 minutes to allow the haematocrit and supernatant to fully separate before 1ml of growth medium was carefully removed without disturbing the haematocrit. From that point the technique continued as previously described.

There is no evidence that leaving the culture separated and thus the haematocrit undisturbed made any difference but this technique did increase sensitivity as the percentage haematocrit increased over the course of the experiment. The difficulty came when adjusting the results to take into account of this. The most elegant solution to this problem was to assume a constant exchange rate of metabolites and substitute the haematocrit at each time point where a sample was taken with the mean haematocrit of all previous time points as shown in table 6.

Hour of experiment	Medium (ml)	RBCs (ml)	Haematocrit	Mean haematocrit up to now
0	25	1.25	5.0%	5.0%
3	24	1.25	5.2%	5.0%
6	23	1.25	5.4%	5.1%
9	22	1.25	5.7%	5.2%
12	21	1.25	6.0%	5.3%
15	20	1.25	6.3%	5.5%
18	19	1.25	6.6%	5.6%
21	18	1.25	6.9%	5.7%
24	17	1.25	7.4%	5.9%
27	16	1.25	7.8%	6.0%
30	15	1.25	8.3%	6.2%
33	14	1.25	8.9%	6.4%
36	13	1.25	9.6%	6.6%

Table 6 : Haematocrit correction table for the "Sara metabolomics" set of experiments. 1ml of growth medium was removed every 3 hours during the experiment.

This method has the drawback that where the change in a metabolite's concentration stops being constant it can produce strange results. There are examples of this which I deal with in the experimental results chapter, notably figure 76 on page 170 and figure 78 on page 171.

Chemical assays to measure lactate and glucose levels

We hoped to measure glucose and lactate levels in used growth medium immediately after acquiring it and as part of Jennifer Lake's Masters thesis¹⁴⁵ we developed protocols to do so. The results of our preliminary tests and our reasons for choosing not to pursue the techniques are summarised here and described more fully in that Masters thesis. In all cases measurement were made according to the suggested protocols and calibration routines included with the kits. The glucose and lactate levels of samples of growth medium were either acquired immediately or thawed from -80°C immediately prior to testing.

The (SIGMA GAGO-20) glucose oxidase kit was used for most glucose level measurement, some tests were run using the glucose hexokinase assay (SIGMA GAHK-20). The absorptions, at 540nm and 340nm respectively, of the samples from both kits were acquired using an Ultrospec 2100 pro spectrophotometer.

The chosen lactate assay kit (Source BioScience AutoGen ABE2467) measured L -Lactate as this has been reported as accounting for 93-94% of the total lactate produced by *P. falciparum*⁹⁴.

NMR Analysis of used growth medium

The results from NMR were much better than from the chemical assay kits we used and those results are described in the chapter dedicated to experimental results.

Sample preparation

Samples of used growth medium for NMR analysis were thawed from -80°C storage and prepared as follows,

- 270 µl sample
- 27 µl D₂O
- 3 µl TSP

These amounts were placed in NMR tubes, either open topped (Julie Fisher lab) or sealed with a glass stopper meaning there was no interface between the air in the tube and the sample (FBS NMR service) as shown in figure 55. By reducing edge effects, the glass stopper increased the sensitivity of the measurements with the drawback that sample preparation was more difficult.

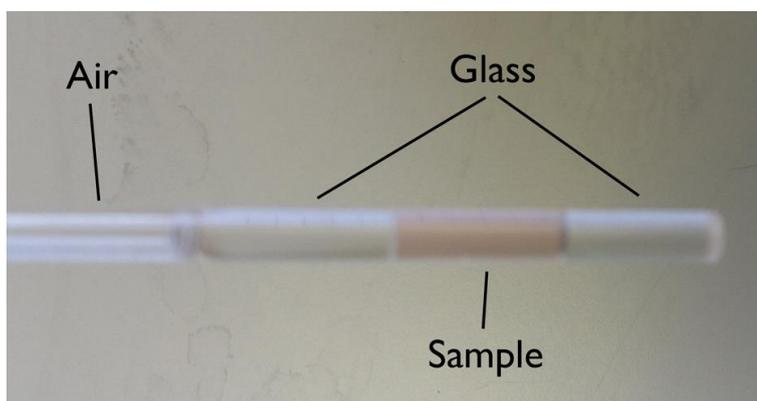


Figure 55: An example of a glass stoppered NMR tube with no interface between the sample and the air as used in the FBS NMR service machines.

Analysis in the Julie Fisher lab, Chemistry, The University of Leeds

The machine used was a Varian Unity Inova 500 spectrometer operating at 499.97 MHz proton frequency. All NMR reported in this thesis was proton (¹H) NMR.

The Carr-Purcell-Meiboom-Gill (CPMG) pulse sequence was used: [RD – 90°x - (t – 180°y - t)]n – acq] with the following parameters;

relaxation delay (RD) 3 seconds

t = 1.5 ms

n = 150

During the relaxation delay the decoupler was applied to selectively irradiate the water resonance

Decoupler power = 4

Spectral width = 6499.84

Acq time = 1.26034s

Acquired complex points = 8192

Transients = 256

Power = 57

Samples were run at 20°C.

Analysis by the NMR service, Faculty of Biological Sciences, The University of Leeds

Minor adjustments to the protocol developed in chemistry were made by Arnout Kalverda at the NMR service in FBS. These attempted to account for the slightly different machines used for acquisitions but the general protocol remained the same.

Data extraction from *.fid files

Analysis of *.fid files produced by the spectrometers was performed using the freeware ACD/NMR Processor Academic Edition available at acdlabs.com/resources/freeware/.

The protocol for analysis was broadly similar for the *.fid files produced by each machine,

1. Load *.fid file.
2. Apply a Lorentzian-Gaussian window function with the following parameters
Julie Fisher lab machine = (LB = -0.3, GF = 0.25),
FBS NMR service machine = (LB = -0.3, GF = 0.1).
3. Zero-fill the spectrum to increase the number of acquired points.
Julie Fisher lab machine = 8192 → 32768,
FBS NMR service machine = 8192 → 65536.
4. Perform a Fourier Transform of the data.
5. Adjust phase, first using the auto function and then manually using the central water peak at around 4.80ppm for reference.
6. Apply a baseline correction.
Julie Fisher lab machine = Auto simple,
FBS NMR service machine = Auto simple, followed by manual adjustments and then a further automatic baselining is usually required.

7. Reference to TMS at 0ppm chemical shift.
8. Manually integrate the HEPES (4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid) signal (5 overlapping peaks) at 3.15ppm and reference to 100.

These prepared spectra were saved as ACD/labs *.esp format files. Their further analysis is described in the experimental results chapter.

Biomass analysis

Biomass component extraction

Extraction of biomass components from the biomass pellet was performed by manually disrupting/grinding the pellet within the eppendorf tube and vortexing the mixtures in the selected solvent for as long as necessary (often 5-10 minutes) to achieve as homogeneous a solution as possible. This solution was always a dark grey colour with fine — but clearly visible — suspended black particles.

Samples were clarified by either centrifugation or spin filtration using the 40 micron spin filters supplied with the PARIS™ kit (AMBION).

Determining protein concentration

Protein content of biomass extractions was determined using the Bradford assay with a known concentration of bovine serum albumen (BSA) at seven x2 serial dilutions in the extraction solvent starting at 4mg/ml for calibration. 190µl of Bradford reagent was added to 10µl of sample in a well of a 96-well plate and incubated for at least 10 minutes before the absorption at 595nm was measured in a Spectramax 340 pc microplate reader (*Molecular Devices*).

Determining DNA and RNA concentration

DNA and RNA concentrations were measured using a nanodrop photometer blanked with Tris-Buffered Saline following the same extraction and purification steps described for protein content determination.

NMR analysis of biomass components

¹H-NMR (Proton-NMR) analysis of parasite biomass components extracted in Tris-buffered saline was attempted but produced no useful results. We lacked the NMR time to perform a thorough investigation of different extraction techniques and peak identification methods and once Teng *et al.*¹⁴⁶ published a thorough ¹H-NMR analysis in 2009 no further efforts were made in this area. A full analysis of the findings of Teng *et al.* (appendix IV) suggests that repeating their techniques on a solubilised biomass pellet may yield good results in the future.

Analysis of lactate and glucose in biomass using chemical assays

Analysis of parasite biomass components extracted in Tris-buffered saline using the glucose and lactate kits described for used growth medium analysis was attempted but produced no useful results.

HPLC analysis of biomass components

A fuller description of the development of the HPLC techniques used for biomass analysis — as reported briefly in figure 75 on page 168 — are contained within Sara Zakutansky's Masters thesis¹⁴⁷. This section contains only those details required to repeat the experiments.

Sample preparation

All HPLC extractions were performed on parasite biomass pellets that had been frozen at -80°C in either RNAlater™ (AMBION) or PBS.

Thaw sample, remove supernatant and remove any remaining red-blood cell ghosts.

Transfer the parasite pellet to a pre-weighed eppendorf tube, reweigh the tube and calculate the pellet mass.

Add 500µl of 60% methanol, disrupt the pellet manually with a pipette tip and vortex until the pellet is as dissolved as possible.

Speed vac the pellet until dry.

-----Steps 5 and 6 are required only if alkaline phosphatase digestion is required.

Re-suspend the pellet in 200µl restriction endonuclease buffer (NEB) and pass through a 40micron spin filter to clarify (PARIS kit spin filters used).*

Add 50U alkaline phosphatase and incubate at room temperature for 2 hours.

Dilute in Ammonium Acetate HPLC running buffer to volume required (typically 500µl final volume).

Samples were frozen at -20°C if they could not be acquired immediately.

HPLC calibration

Tests were performed on a Dionex 3000 series liquid chromatograph with a 200 mm × 4.6 mm Thermo Scientific Hypersil GOLD column with 5 µm particle size and 175 Å pore size.

The column eluate was measured at 254 nm and the resulting chromatographs were analysed in the Chromeleon software package.

To calibrate the column, Sara Zakutansky measured the retention times and peak sizes of dihydroorotate, orotate, guanine, guanidine, uracil, uridine, deoxyuridine, hypoxanthine, xanthine, thymine, thymidine, inosine, and adenosine run both separately and as a mixture at concentrations ranging from 0.5 to 50 μM .

Accurate determination of *P. falciparum* growth-rate

The parasite culturing for the fluorescence microscopy and flow cytometry in this chapter was performed by a Masters student, Cheng Ma. All reported experiments were designed and performed together. All the analysis reported in this section is my own.

Cheng Ma's Master's thesis¹⁴⁸ is available from the University of Leeds and offers a broadly similar, though perhaps more optimistic, analysis of our results. I am indebted to him for his efforts.

Do we need a better way of measuring growth-rate?

Well-parameterised genome-scale metabolic models of model organisms like *E. coli* and *S. cerevisiae* are able to accurately predict changes in a population's growth-rate in response to perturbations^{149,150} including changes in available nutrients. These model organisms are typically grown in chemostats where conditions are precisely controlled, with population, and thus growth-rate, measured via the dilution rate required to keep the growth medium's optical density constant. This approach also works for more challenging organisms such as *M. tuberculosis* but there is currently no widely-used equivalent test for any species of *Plasmodium*. The closest equivalent, the substantially more difficult technique of measuring the uptake of radioactive hypoxanthine¹⁵¹, is described in the final section of this chapter where I argue that it is not suitable for the experiments in this thesis.

Measuring *P. falciparum* growth-rate *in vitro* using blood smears

The parasitemia of cultures is determined using optical microscopy of stained blood smears and the growth-rate of a culture can be calculated from the rate at which parasitemia increases. When measuring changes in the growth-rate of cultures an additional complexity must be considered.

P. falciparum is notoriously difficult to culture outside of live animals and so it is difficult to make changes to the growth medium without killing the parasites. When measuring a change in growth-rate due to a change in growth conditions — such as the addition of a known inhibitor or the restriction of a required nutrient — we have to balance the need for the change in growth-rate to be big enough to be measurable and small enough not to kill the cultures. Work on this compromise is not well reported in the literature and the basic background measurements to make that assessment are required.

Uncertainty in the parasitemia of a single blood smear

The uncertainty in a measurement of the parasitemia of a culture from a single blood smear can be estimated by taking a number of blood smears from the same sample and comparing the parasite counts.

Parasites counted in ten fields of seven slides	Mean	Standard Deviation	Implied percentage error in a single slide
---	------	--------------------	--

Table 7: The number of parasites counted in ten fields of seven slides each made from the same sample. The standard deviation rather than the standard error is calculated because during the experiment only a single slide ($n = 1$) is counted and $\sigma = \sigma/\sqrt{n}$.

In the example shown in table 7 the parasitemia is $3.6 \pm 0.2\%$ * taking into account all 7 slides with the uncertainty increasing to ± 0.5 if just a single blood smear were read. As a percentage, uncertainty in the measurement of a single slide is at least 10%. To improve the readability of the graph in figure 56, and other graphs in this thesis, I have not included these large errors on individual measurements. This is because once a sufficient number of measurements has been taken over time the uncertainties on the measurements at each time point become unimportant.

Uncertainty in the growth-rate over multiple readings

I grew a standard culture of *P. falciparum* for two weeks, splitting the parallel cultures as necessary so that the parasitemia stayed in the safe range between 0.5% and 6%. I measured the parasitemia of the cultures each day using thin blood smears and scaled up the recorded parasitemias according to how diluted with respect to the culture at the start of the trial the cultures had become. This meant that by the end of day 14 the average parasitemia of the 5 final parallel cultures was 4.6% but reported as 1251.1% after applying the dilution factor on day 5 with respect to day 0 of 303.5.

* the error here is an absolute value and not a percentage. The % refers to parasitemia which is always reported as the percentage of red-blood cells that are infected.

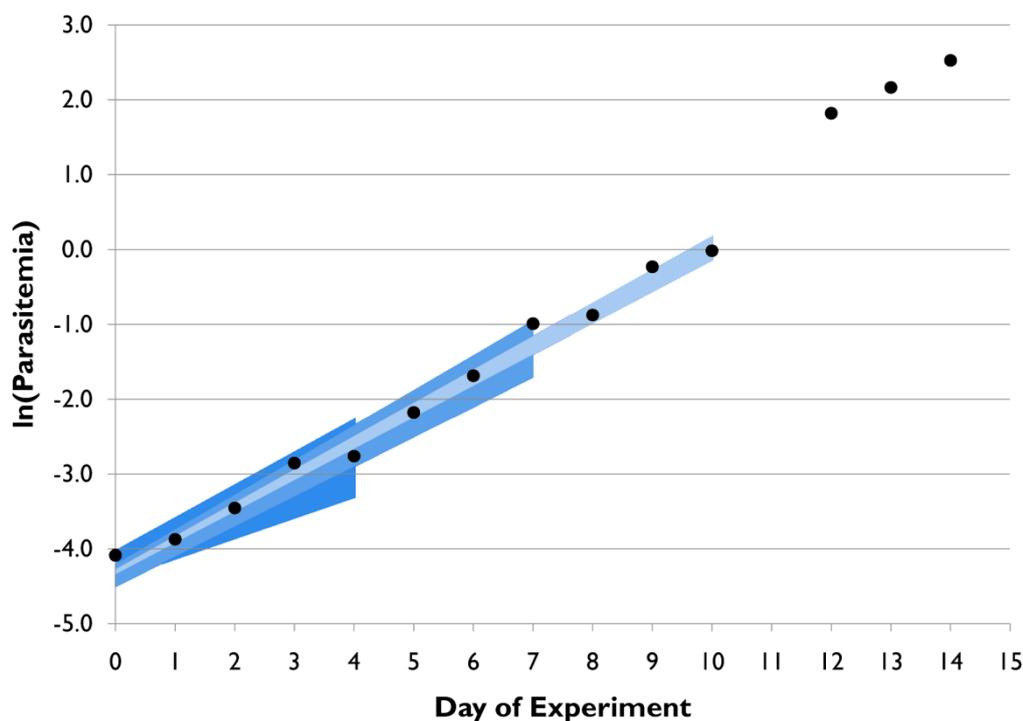


Figure 56 : Growth-rate of a population of *P. falciparum* grown in 100% AlbuMAX I media for two weeks as measured using Giemsa's stain on thin blood smears under an optical microscope. Here the natural logarithm of the parasitemia is plotted so that the exponential growth curve becomes linear. Shaded areas show the error in a least squares linear fit of the data after 4 (dark), 7 (lighter) and 10 (lightest) days.

Growth, as shown in figure 56, is exponential, with the population increasing by a factor of about 1.55 per day (d^{-1}). At day 14 all cultures were harvested and the parasite biomass extracted and separated from red blood cells (RBCs) and frozen for further analysis.

The blue shaded areas in figure 56 show the maximum and minimum possible growth-rate calculated by a least-squares fit and associated error calculations as provided by the linest function in Microsoft Excel after four days, seven days and ten days respectively. The results are as we would expect with the error on the best fit line decreasing as the experiment progresses. As the error decreases we can see that we would be able to distinguish between two samples with ever closer growth-rates but we need to look at the lines in a numerical form, as in table 8, to see this most clearly.

Experiment duration	Number of measurements	Calculated growth-rate (per day/ d^{-1})	Error in growth-rate
4 days	5	1.44	± 0.12
7 days	8	1.55	± 0.05
10 days	11	1.55	± 0.025

Table 8 : Best-fit growth-rates for the three time periods depicted in figure 56.

Looking at the numbers in table 2 we see that if we aim to measure a change in growth from 1.55 to 1.45 d⁻¹ then we would need to maintain the two cultures for 10 days to be sure that they are different (after 7 days 1.45+0.05 d⁻¹ is not distinguishable from 1.55-0.05 d⁻¹). This is certainly possible but every extra day of culturing introduces a possibility that a culture could be affected by an unknown factor such as infection by bacteria or yeast or an unexpected crash.

The advantages of making more frequent measurements

These results give us a rough idea of the duration of an experiment needed to measure a change in growth-rate if measurements of parasitemia are taken every day during normal culture maintenance. If we assume that growth-rates remain exponential whatever their magnitude — sensible for an organism with no resource limits and that reproduces through asexual fission — we can use these results as a guide to the number of measurements required to accurately determine growth-rate over a shorter period.

Taking more frequent measurements of a culture has a direct advantage, the growth-rate can be determined in less time, but also a secondary advantage, samples can be taken to observe properties of the culture on a shorter timescale. My work on the metabolism of *P. falciparum* over a single life-cycle relied on taking a minimum of five measurements over a period of around 40 hours.

Note on the reproducibility of growth-rates across experiments and laboratories

The reproducibility of growth-rates has been a problem throughout this project. Asynchronous cultures grown in growth medium containing over 10% human serum have been measured to have a typical growth-rate of around 2 d⁻¹ (parasitemia doubling each day), in keeping with previous experience and the literature.

In contrast, almost all my experiments with parasites grown in 100% AlbuMAX medium show a growth-rate of nearer 1.55 d⁻¹. The work of a 2008 Master's degree student, Cheng Ma¹⁴⁸, reproduced a growth-rate of around 1.6 in 90% AlbuMAX I although previous experiments by both myself and Daniel Opi¹⁴⁴ have achieved growth-rates of 2 in 90% AlbuMAX I. Considerable efforts to ensure that all components of the growth medium were prepared properly have been unable to solve this problem and in more recent experiments I too have seen a growth-rate of around 1.6 d⁻¹ in 90% AlbuMAX I growth medium.

These observations are worrying but not completely unexpected as slower growth-rates in AlbuMAX growth medium have been reported previously¹²⁸. Since we are typically interested in measuring changes — rather than absolute values — of the metabolic

activity and the growth-rate of cultures submitted to perturbations, we can work around this issue.

It is also of note that variability in growth-rate is widely reported, with early attempts at optimising growth conditions¹⁵² achieving a near doubling of parasite growth-rate by making small changes to the growth medium.

Development of flow cytometry

Thin blood smears are often time consuming to stain and clumsy to store and counting parasites by hand introduces large errors on individual readings. My finding that — with daily culture maintenance and observation — it would take two weeks of culturing to measure a change in growth-rate from 1.55 to 1.45 d⁻¹ led me to think about better ways of measuring parasite growth-rate. After reading the review “*Cytometry in malaria: moving beyond Giemsa*” by Shapiro *et al.*¹⁵³, I decided to investigate flow cytometry as a way to improve the measurement of growth-rate.

Selection of a suitable fluorescent dye

Flow cytometry works by staining cells with a dye that selectively binds to only those cells we want to count. In this case any DNA-binding dye will bind only to parasites as human RBCs contain no DNA. A quick review of the literature provided us with three dyes to consider, Hoescht^{153,154}, Acridine Orange^{155,156} and syto-16¹⁵⁷. Our discussions with Dr. Gareth Howell in the Astbury centre at The University of Leeds led us not to consider using the technique described in the paper using syto-16 and preliminary tests with Hoescht dye led to an unexplained sedimentation in our samples. Acridine orange became the best option for continued study and we set ourselves the goal of repeating the results found in “*Re-Evaluating acridine orange for rapid flow cytometric enumeration of parasitemia in malaria-Infected rodents*” by Bhakdi *et al.*¹⁵⁶.

Fluorescence microscopy to confirm selective staining of infected RBCs

Our first task was to test that acridine orange selectively stained infected RBCs by examining a stained sample of live cells under a fluorescence microscope. Unfortunately the only fluorescence microscope available to us has a broken lens at maximum magnification which caused us some of the following problems.

- The illumination of the sample area was uneven, meaning that some areas within each field of view were harder to see than others.
- Each field contained around 1500 RBCs making it hard to count the cells.
- The cells were not fixed and thus tended to move slightly.

- Cells tended to stick to either the glass slide or the coverslip but the microscope could only focus on one plane at a time.

Despite these difficulties we were able to confirm that acridine orange was selectively staining parasite-infected RBCs from our cultures. Figure 57 is a typical capture of the centre of a field of infected RBCs with a known parasitemia of 3.8%, as measured by Giemsa staining a thin blood smear under an optical microscope. Within this picture we found it useful to define two different types of fluorescent spots. Type 1 spots are clearly inside RBCs and of a similar size to a parasite whereas the type 2 spots are outside RBCs and often appear smaller. I have also highlighted a single stain with 2* which looks like it may be either part of the nearby cell or a parasite in a cell in another focal plane.

Table 9 shows the results of counting the cells in figure 57 and we see that type 1 stains imply a parasitemia of around 4%, the same as determined by reading the blood smear, suggesting that type 1 stains are infected parasites.

	Number	Inferred parasitemia
Red blood cells	187*	-
Type 1 stains	8	4.3%
Type 2 stains	15	8.0%
Combined	23	12.3%

**estimate based on number of cells crossing the field edges (17x11)*

Table 9 : Counts of the different types of fluorescent spots in figure 57.

The presence of the type 2 stains is unfortunate since it seems possible that they could be falsely counted as positive in a flow cytometer if the threshold of fluorescence were not adequately set. These spots could be aggregates of waste products and dead parasites in solution that have been stained by acridine orange or aggregates of dye particles. The tempting explanation that they are due to auto-fluorescence of unstained RBCs as described later in this chapter seems unlikely given their high frequency and the fact that they seem to occur outside of the RBCs.

My original tests were carried out with a dye concentration of 4.0 µg/ml whereas future test were carried out at around 0.5-1.0 µg/ml. At these lower dye concentrations, fewer type 2 stains are visible under the fluorescence microscope suggesting that the problem of type 2 staining may be reduced at the concentrations we use in flow cytometry.

Nevertheless, we still need to consider these staining artefacts when interpreting results from flow cytometry.

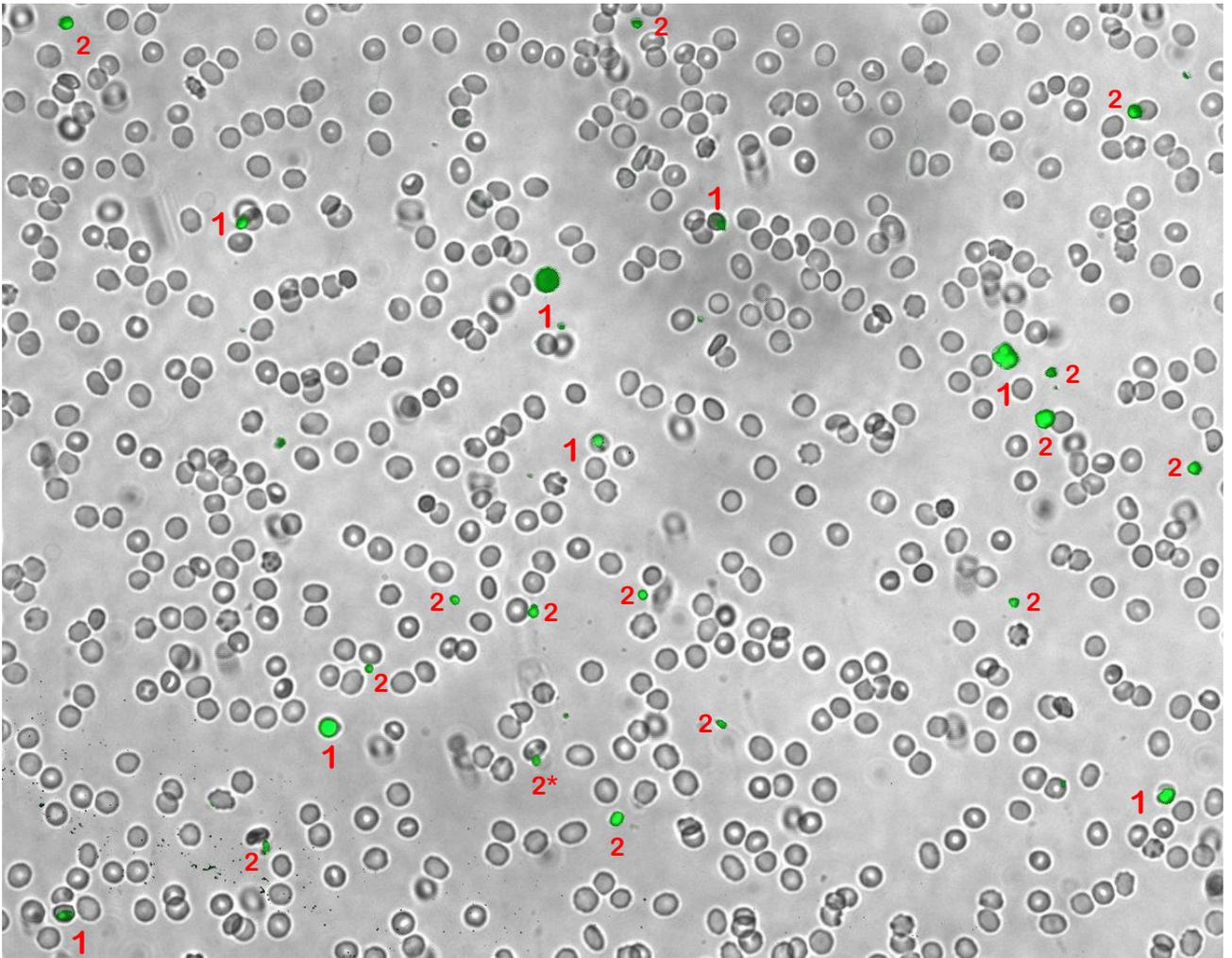


Figure 57: Fluorescence microscope image of RBCs stained with Acridine Orange at 4.0 $\mu\text{g/ml}$ *. Green spots are areas of peak fluorescence under white light and are labelled as described. The fluorescent layer and the background layer were aligned in Photoshop due to a problem with the microscope optics.

Having confirmed that acridine orange was able to stain infected RBCs whilst leaving uninfected cells unstained we moved on to seeing whether stained and unstained cells could be distinguished within a flow cytometer.

* Bhakdi et al.²⁰ quote their dye concentrations after samples have been diluted by half in PBS. As a result a dye concentration of 4.0 $\mu\text{g/ml}$ is defined as being 2.0 $\mu\text{g/ml}$ in their paper. My results are completely equivalent as long as this factor of 2 is taken into account.

Single-channel flow cytometry

The simplest type of flow cytometry is single-channel analysis, in our case using the green channel. This uses the confirmed property that acridine orange fluoresces green under illumination once bound to DNA and sets a cut-off intensity of fluorescence above which a cell is considered infected and below which a cell is considered uninfected. In order to test this method we prepared serial dilutions of a culture with an optically measured parasitemia of 4.6%. These assays were prepared by adding acridine orange to a final concentration of 1.25 µg/ml and then diluting with 1.5ml of phosphate-buffered saline solution (PBS) before acquisition. The threshold above which cells would be considered infected was set by running a control of stained uninfected cells through the flow cytometer and setting the gate to count a minimum of these cells.

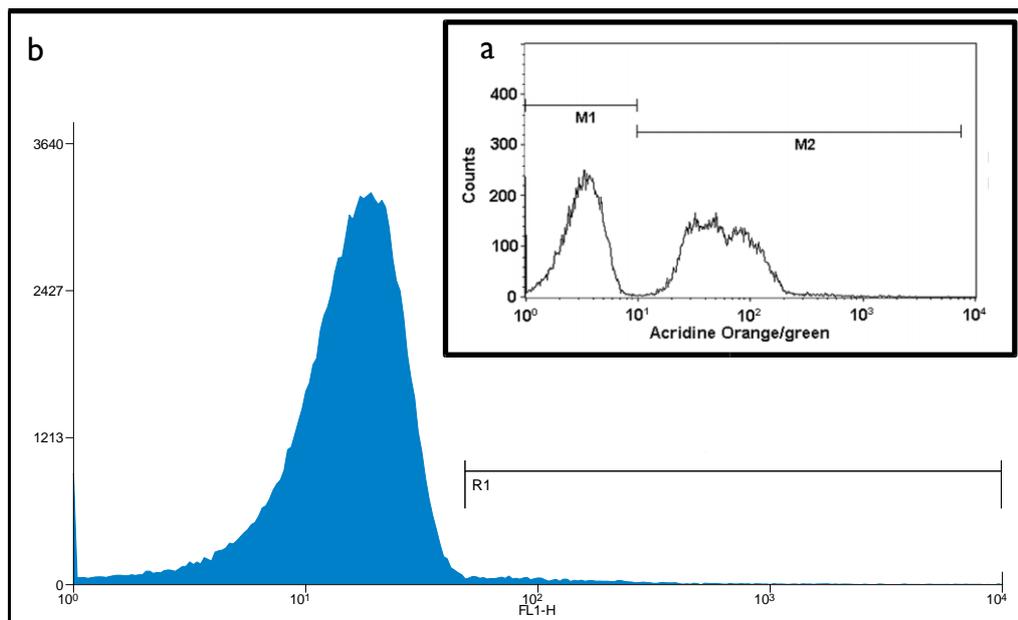


Figure 58 : Comparison of my results and published results for single-channel flow cytometry. **a.** Sample result from Bhakdi *et al.*¹⁵⁶ with M1 gate denoting uninfected cells and M2 denoting infected cells. Parasitemia ≈30-50%. **b.** My equivalent result for a sample with a known parasitemia of 4.6%. Gate R1 denotes infected cells.

Figure 58 shows an example of gate-setting to distinguish infected and uninfected populations. The excerpt in section **a** shows an idealised case where the distinction is easy to make whereas section **b** shows my real results with the gated area defining infected cells. In the ideal case the histogram for a control population (known parasitemia of 0.0%) would be identical to the histogram seen in **b** but with no counts in the section gated as R1. Any counts in the R1 gated section of this histogram for this control population should be discounted from all other samples and discarded as background noise.

The very different parasitemias of the samples in figure 58 are a result of the limitation caused by culture “crashes” if culturing above 8% parasitemia. As Bhakdi *et al.*'s samples

came from living hamsters infected with *Plasmodium berghei* they were not constrained by this limitation.

Determining an ideal dye concentration

With a gate threshold set we moved on to trying to determine ideal parameters to use before sample acquisition. To this end we ran extensive tests to examine the effects on the measured parasitemia of,

- the time the sample was left to incubate with the acridine orange before dilution in PBS,
- keeping the samples in the dark during incubation,
- the amount of PBS the sample was diluted with before acquisition,
- the time the sample was left to incubate with the acridine orange after dilution
- the concentration of acridine orange used.

An example of the results from the last of these tests is shown below in figure 4 and shows a large variation between the measured parasitemia of samples stained with slightly different concentrations of acridine orange.

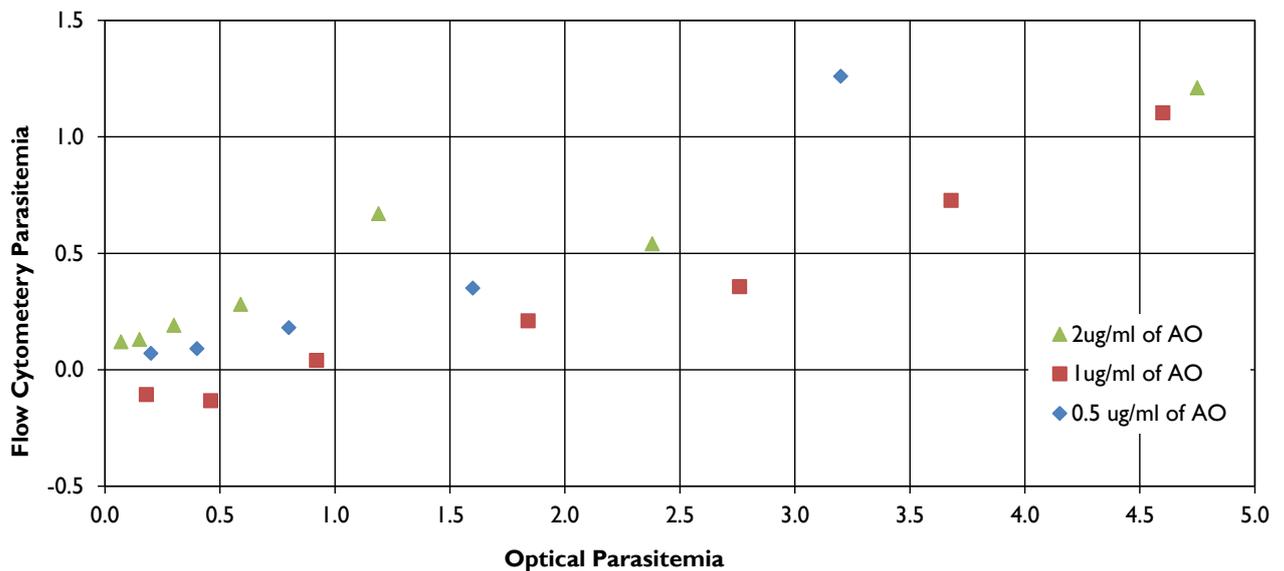


Figure 59: FACS-determined parasitemia versus parasitemia determined by microscope inspection of Giemsa-stained blood smears. Removal of counts from un-dyed control samples means some FACS-determined parasitemias are below 0%.

Calibration issues

The effects of varying the acridine orange concentration on the results of flow cytometry were not encouraging and the results for the other variables were no better. Further tests showed that part of our problems came from the raw data coming from the flow

cytometer. A clear example of one of these problems, setting a threshold for distinguishing infected from uninfected cells, is shown in figure 60 .

These are histograms for two sequential runs using identical parameters on the same flow cytometer. The control sample consists of uninfected RBCs in media with 1.25 $\mu\text{g}/\text{ml}$ (before dilution in PBS) of acridine orange added. The 0.18% parasitemia sample is identical but 2% of the total volume was from a culture at 4.6% parasitemia. Clearly this addition of a tiny amount of culture to an identical sample should not move the whole fluorescence peak since the peak in both cases corresponds to uninfected RBCs from an identical source.

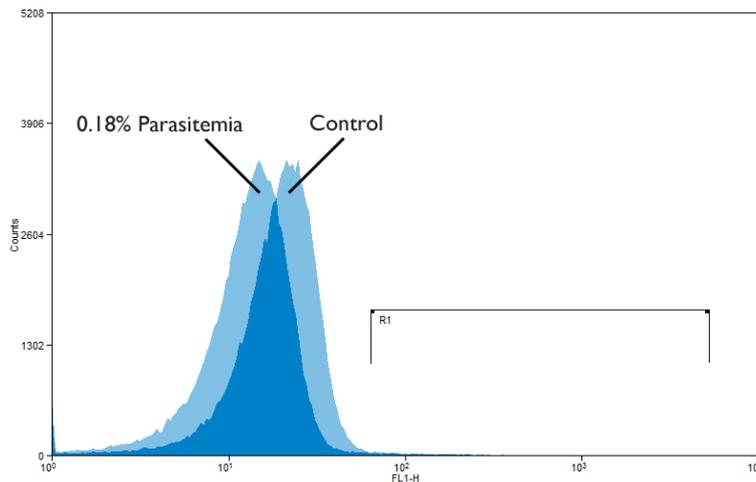


Figure 60: Comparison of raw single-channel output from BD FACSCalibur acquisition of a control sample (0% infected RBCs in media dyed with 1.25 $\mu\text{g}/\text{ml}$ of acridine orange) versus a sample with a known parasitemia of 0.18%. Overlapping section is darker. R1 shows gating optimal for control which is clearly not optimal for 0.18% case.

This peak-shifting made it extremely difficult to set a gate that could consistently distinguish infected and uninfected cells. In our preliminary tests we never saw — no matter what the dye concentration used — as distinct a separation between the fluorescence of infected and uninfected RBCs as shown in figure 58a. This closeness in the magnitude of the fluorescence of the two populations meant that the placement of the gate had a big influence on the results.

I had some success in overcoming the problems with the raw output of the flow cytometer by setting the gate position in different ways for different experiments as I will now explain. Where I have set a single gate and applied it to all samples I have called this gating technique “global and constant”.

Where peak-shifting meant that a threshold could not be set this way I was forced to set a gate for each sample by eye. This meant estimating the position of the edge of the peak for each result and manually drawing a gate. I have called this gating technique “custom and variable” since the gate is different for each sample.

Figure 61 shows the results of a test to measure the parasitemia of samples by flow cytometry using global and constant gating where the parasitemia had already been determined by optical microscopy. The results appear impressive, with a good linear response and the flow cytometer picking up over a third of the parasites. Unfortunately the results are not nearly as positive when we look at them in detail. The control sample — uninfected RBCs stained with acridine orange — has a parasitemia of 2.10%, far higher than expected. Furthermore, even a tiny and insignificant change in the preparation of the samples, adding 2.0ml of PBS instead of 1.5ml of PBS before running the samples, causes a large and unpredictable change in the results.

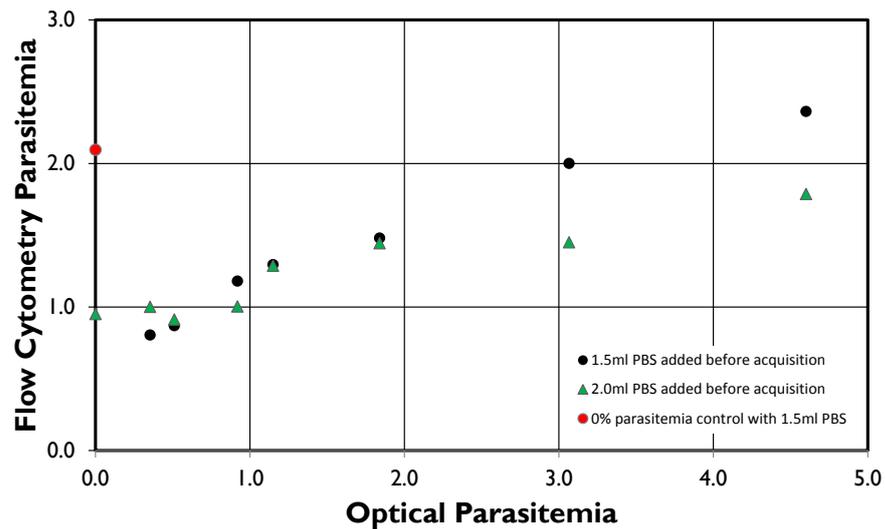


Figure 61 : Parasitemia measurements by single-channel flow cytometry using global and constant gating versus measurements by optical microscopy of Giemsa-stained blood smears. Samples were identical and prepared as described in the methods chapter and diluted with either 1.5ml or 2.0ml of PBS before acquisition. The point in red belongs to the 1.5ml dataset but is not included in the best fit line.

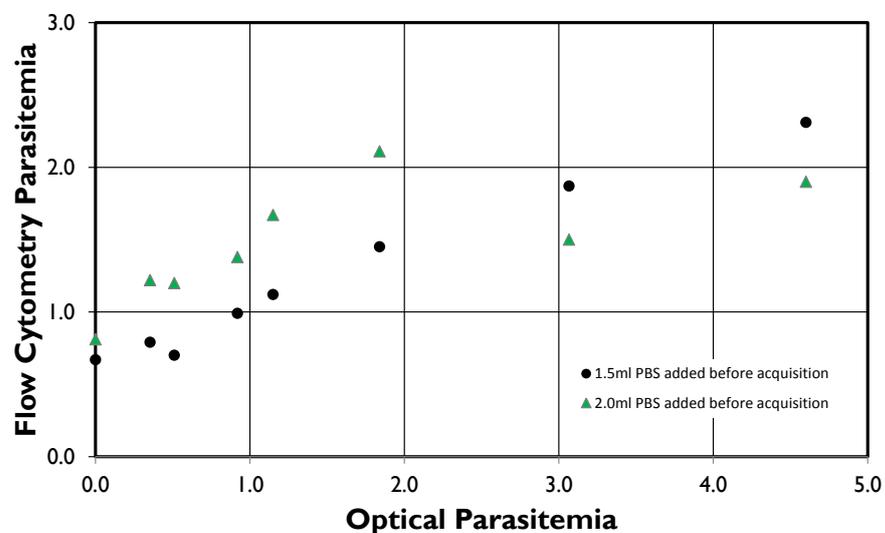


Figure 62: Parasitemia by single-channel flow cytometry using custom and variable gating versus optical microscopy of Giemsa-stained blood smears.

Figure 62 shows the results from analysing exactly the same raw data from the flow cytometer as used to create figure 61 but instead determining by eye a custom gate for each sample. The combined numerical results from figure 61 and figure 62 are summarised below in table 10.

Gating	PBS added before acquisition (ml)	Gradient	Intercept	Parasitemia of control
Global and constant	1.5	0.37	0.78	2.10%
Custom and variable	1.5	0.38	0.65	0.67%
Global and constant	2	0.19	0.93	0.95%
Custom and variable	2	0.13	1.33	0.81%

Table 10 : Summary of the gradient and intercepts of best-fit lines for points in figure 61 and figure 62.

Summary of single-channel flow cytometry

Table 4 raises some issues that are very difficult to resolve. The measured relation between optical parasitemia and parasitemia as measured by flow cytometry seems erratic, possibly owing to tiny changes in sample preparation but more probably owing to instabilities within the flow cytometer. The peak-shifting issue highlighted in figure 60 means we are forced to analyse samples by hand but even this does not give us a reliable technique for analysis. Custom and variable gating has a further drawback in that it removes the speed and objectivity of the process and thus both main advantages of flow cytometry over optical microscopy. Our failure to develop a better method for counting parasites using single-channel flow cytometry is perhaps not surprising given that examples in the literature have only proved the technique for the very high parasitemias found in *in vivo* infections.

Dual-channel flow cytometry

Acridine orange has a very useful property, it not only fluoresces green when bound to DNA and but it also fluoresces red when bound to RNA. Instead of a histogram of counts against intensity of fluorescence in a single colour axis, dual-channel flow cytometry produces a three dimensional plot with each measured cell placed according to the intensity of its fluorescence in the two different colours. By extracting this information from our samples we hoped to overcome the problems we experienced using single-channel analysis.

Figure 63 is the output of dual-channel flow cytometry from exactly the same raw data as shown in figure 58. In other words, if we projected all the points in figure 63 onto the x-axis and counted them to produce a histogram we would exactly reproduce figure 58. The peak shifting that meant we had to use a different gate for the control than for the rest of the samples in the single-channel case is also the case in dual-channel analysis and so this technique cannot offer any improvement in this regard and custom and variable is still likely to be required.

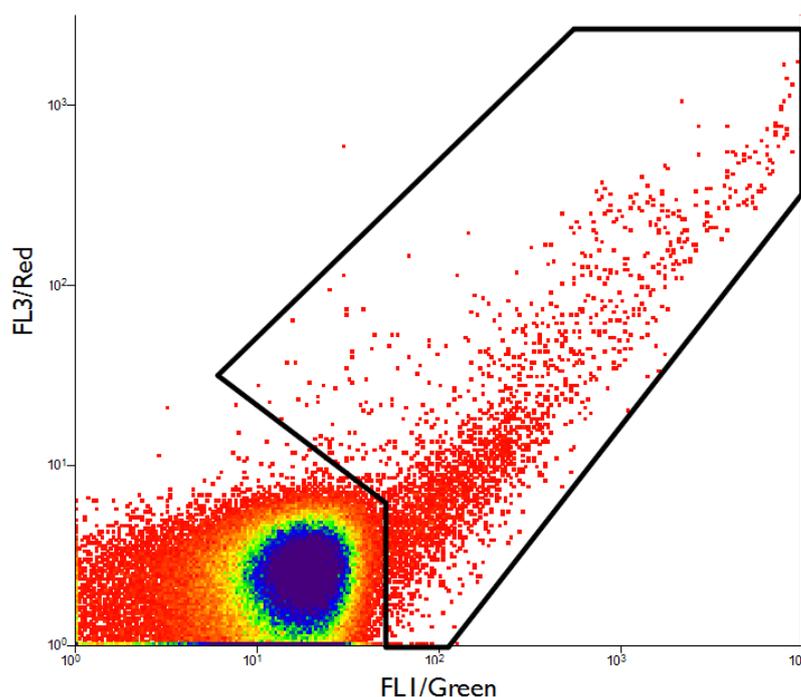


Figure 63 : Dual-channel flow cytometry (red and green) for the same sample as shown in figure 58b. The gated area shown is a naïve attempt to select infected cells. Samples were dyed with acridine orange to a final concentration of 1.25 $\mu\text{g/ml}$, left for at least 3 minutes, then diluted with 1.5ml of PBS before acquisition.

A more advanced application of dual-channel flow cytometry is to specify zones of the plot which may correspond to different features of the cells measured. An excellent example of this can be found in Grimberg *et al.*¹⁵⁸ where the authors were able to distinguish between all life stages of *Plasmodium falciparum* with a two-dye technique using Hoescht and thiazole orange and dual-channel flow cytometry. Whilst this technique certainly warrants further investigation, neither of the flow cytometers I have been able to use at Leeds have working filters to measure Hoescht fluorescence. We would also have to overcome a problem we had during our preliminary experiments with precipitates forming in our samples after adding Hoescht to them. We are still unsure what caused this precipitation but it seems likely that it was due to some reaction between the solvent used in our Hoescht stock and components within the growth medium.

In Bhakdi *et al.*¹⁵⁶ the zoning technique described in Grimberg *et al.*¹⁵⁸ for distinguishing life stages was used only to distinguish between cells that were fluorescing due to parasite

infection and those exhibiting simple auto-fluorescence. With acridine orange this seems likely to be our only option. Figure 64 is an illustration of the zones in a dual-channel plot that we might suggest are due to these two different types of fluorescence.

The significantly increased number of points in the infected fluorescence zone where parasitemia is high (3.5%) versus low (0.18%), as in figure 8, support this distinction. That the number of points in the auto fluorescence section of the plot goes down is in keeping with the observations of Bhakdi *et al.*¹⁵⁶ but is hard to explain.

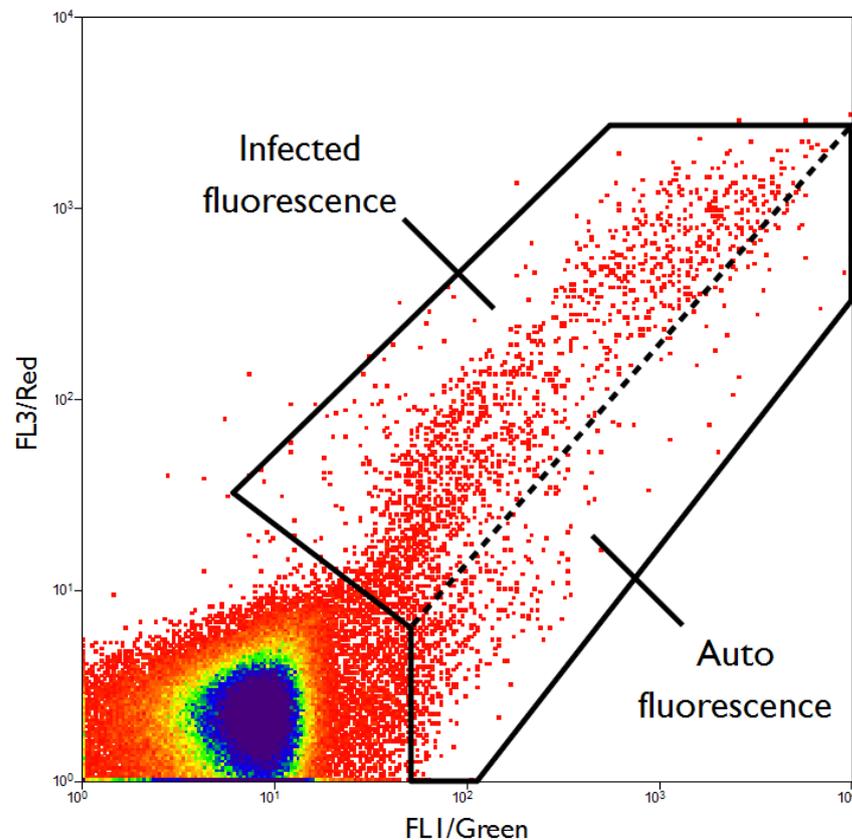


Figure 64 : Dual-channel output of a sample of known parasitemia of 3.5% showing the two distinct zones where it is suggested that fluorescence is the result of parasite infection or auto-fluorescence of an uninfected RBC. The zoning technique seems hard to justify as discussed in the thesis body.

The idea that measuring auto-fluorescence can be eliminated using this method is very interesting but we have to remember that our results from single-channel flow cytometry already measured too few parasites rather than too many. Another worry for me is that Bhakdi *et al.*¹⁵⁶ does not fully justify the applied zoning technique or adequately explain why different types of fluorescence should occur in the different zones. The position of the infected fluorescence zone would suggest that more RNA is present in infected RBCs, which would make sense, but Bhakdi *et al.*¹⁵⁶ makes no mention this. These issues are particularly troubling given that Grimberg *et al.*¹⁵⁸ makes no mention of auto-fluorescence in unfixed cells and makes no effort to filter out these cells in their paper.

A further reason that splitting the zone of higher fluorescence into two distinct sections may not be useful comes from inspecting the dual-channel plot in figure 63. Here we see that almost all the fluorescence in this plot would be counted as auto-fluorescence even though the known parasitemia is 4.6%. Experience with single-channel flow cytometry makes it clear that any technique that attempts to remove auto fluorescence from results must be calibrated using a negative control or dyed uninfected RBCs. When we consider the peak-shifting effect that we noticed in the green channel of the available equipment, and is likely to also exist in the red channel, we see that this is far from easy.

Despite these problems, when dual-channel flow cytometry works it can give us good results. Figure 65 shows that at least some of our results do seem to agree with the results in Bhakdi *et al.*¹⁵⁶ and the results shown in figure 66 show that flow cytometry using this method can measure over half of the true number of parasites.

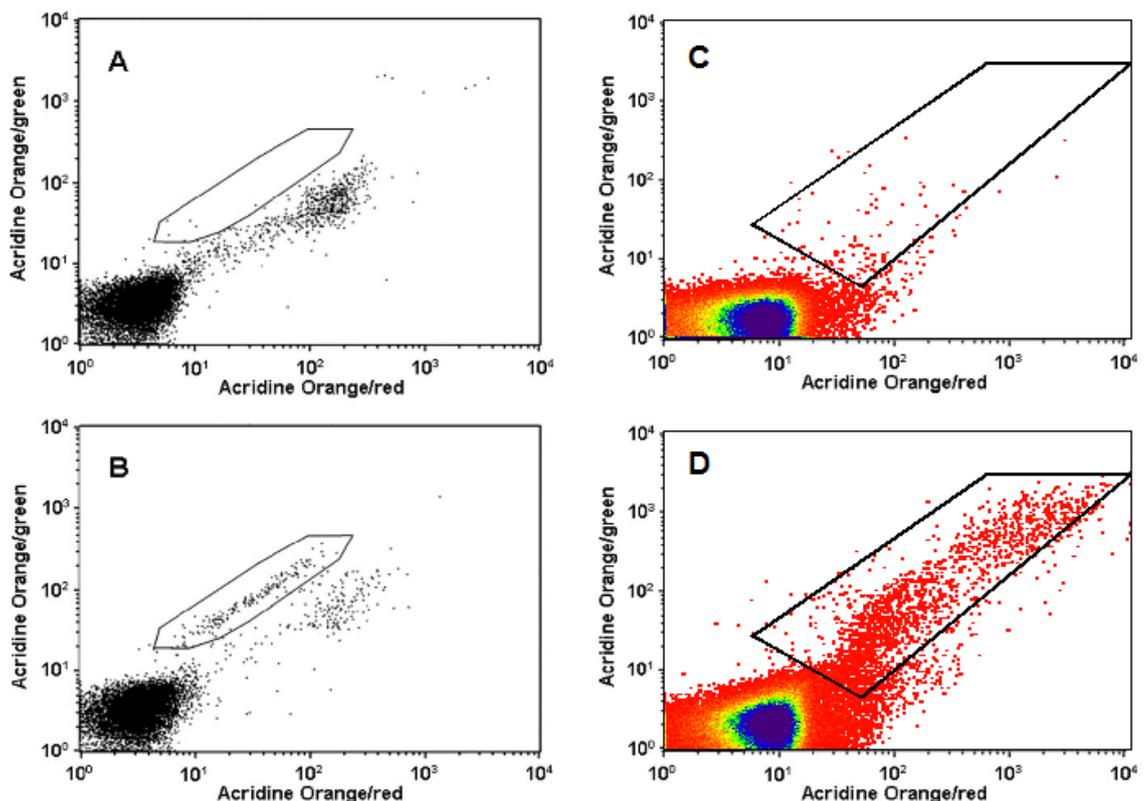


Figure 65 : A. Stained uninfected cells from Bhakdi *et al.* (known parasitemia =0.0%, observed parasitemia = 0.09%)
 B. Stained mouse RBCs infected with *P. berghei* from Bhakdi *et al.*¹⁵⁶ (known parasitemia \approx * observed parasitemia = 1.79%)
 C. Stained uninfected cells from my experiments. (known parasitemia =0.0%, observed parasitemia = 0.1%)
 D. Stained infected cells from my experiments. (known parasitemia = 3.5%, observed parasitemia = 2.0%)

* This measurement by Giemsa stain/optical microscopy is not given in the paper but a later figure shows a near perfect 1:1 correspondence between optical measurements and measurements using flow cytometry.

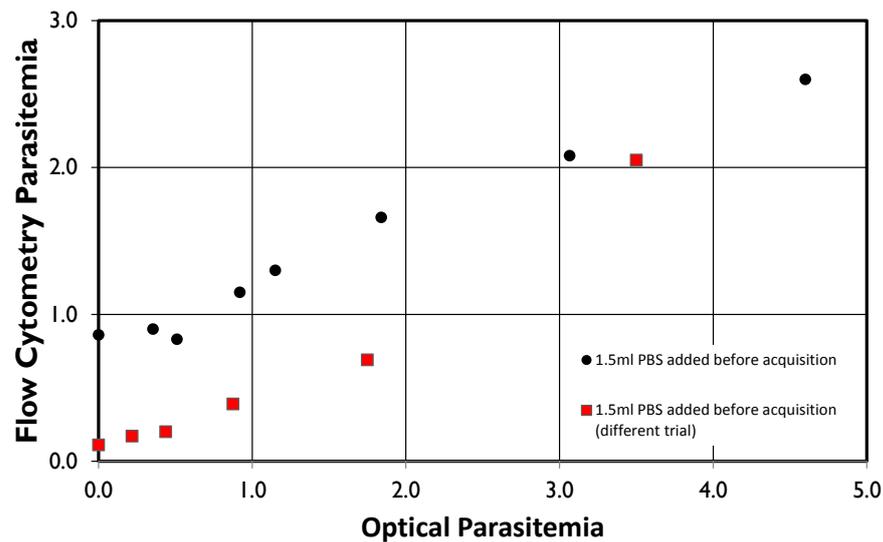


Figure 66 : Parasitemia by dual-channel flow cytometry as demonstrated in figure 64 versus optical microscopy of Giemsa-stained blood smears. Samples prepared as in the methods chapter. Results are extremely variable across different runs.

Disruptive variables

We carried out a huge number of tests whilst trying to develop a robust protocol for measuring parasitemias of samples using flow cytometry. Many of our problems in the beginning were caused by machine issues but we also noticed that the results of flow cytometry were extremely difficult to predict and tiny changes in protocol — or just different runs of the same machine with the exact same settings — seemed to lead to massively different outcomes. We can see this in figure 66 where samples run in exactly the same conditions by the same machine give very different results. The peak shifting effect I have described for single-channel flow cytometry is also visible in dual-channel flow cytometry with the peak in the green channel in very different positions in figure 63 and figure 64.

These problems meant that calibration of the output from the machines we used was almost impossible and required manual adjustments which removed many of the advantages of the technique. Even after rigorous calibration results were unpredictable.

An interesting example of the variability of the experiments is shown below in figure 67. We noticed that at lower concentrations of acridine orange the two zones of fluorescence described previously became more and more distinct. At first we were very excited by this until we counted the number of cells in the infected fluorescence zone and saw that it accounted for less than a quarter of the known parasitemia, even fewer than measured at normal dye concentrations.

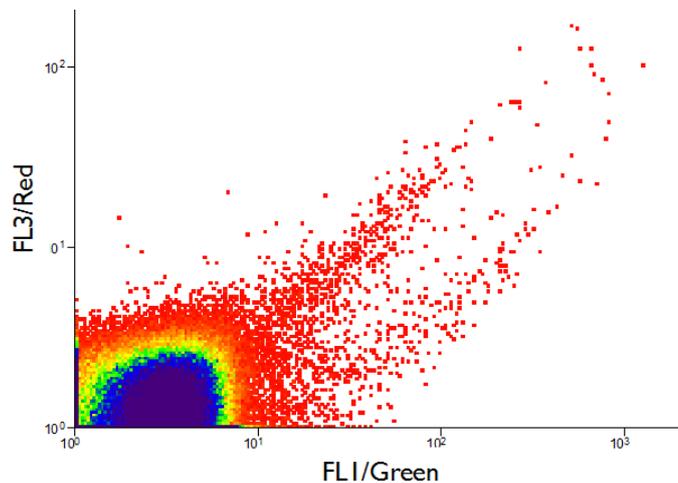


Figure 67 : Dual-channel output of a sample of known parasitemia of 3.3% dyed with 0.25 ug/ml acridine orange. The infected fluorescence zone contains 0.78% of the cells counted.

Summary of flow cytometry

Flow cytometry could have given us a quicker and more accurate way to measure the parasitemia of a culture and helped to more easily and objectively measure growth-rates of large cultures of *P. falciparum* grown in parallel under different conditions. In the end we have struggled to reproduce the published results as we would have liked to, only picking up between a quarter and a half of the infected cells as detected by Giemsa staining and optical microscopy. The unreliability of the equipment casts further doubt on the technique as despite a huge number of tests we were never able to isolate or eliminate sources of variability. It is an interesting hypothesis that we measure only a proportion of the true parasitemia because we are picking up only infections where the parasite is at its largest stage (schizont) but our experiments with a synchronised culture failed to show this*. My feeling is that the results of Bhakdi *et al.*¹⁵⁶ were idealised and not ready for integration into a lab setting. That the paper was published in a flow cytometry journal rather than a parasitology journal, and the knowledge that Barbara Kappes and her group in Heidelberg have also tried and failed to use flow cytometry for parasitemia measurements[†], supports this view. It seems possible that with the freedom to more accurately control and calibrate the flow cytometer many of the problems we had could be overcome but this was not possible within the restrictions of this project.

Useful applications of flow cytometry to studies of *Plasmodium falciparum* using both dual-channel and dual stain techniques have recently been published by Jiménez-Díaz *et al.* in 2009¹⁵⁷ and, most applicably to my project and already mentioned, by Grimberg *et al.* in 2008¹⁵⁸.

* These tests were carried out on the PRS III machine and should not be considered reliable. They are not reported.

† Discussed in private communications with Glenn McConkey.

The inherent limits of “growth-rate” measurements in *P. falciparum* cultures.

The final section of this chapter is a common point of discussion at all conferences on metabolism; what does growth-rate really mean?

The growth-rate of an organism can mean either the rate at which biomass is produced, or the rate at which the population of the organism increases. In batch cultures of ideal organisms like *E. coli* and *S. cerevisiae* these two definitions are interchangeable since the population times the average mass of a single organism — which is constant — is equal to the biomass. This simplification no longer holds when the rate at which biomass is accumulated is no longer related to the rate at which the population of the organism increases and this is exactly the case with *Plasmodium* grown in culture. This discrepancy arises because, before bursting out of their containing RBCs, mature schizonts appear to have between 8 and 16 (16 - 32 according to Keeley *et al.*¹⁵⁹) separate nuclei that will become merozoites upon lysis. Since only between 2 and 4 (equivalent to a daily growth-rate of between 1.4 and 2.0) of these merozoites will successfully invade a new RBC as much as 7/8 of the accumulated biomass is lost with every life cycle thus breaking the link between biomass created and current population.

In this example, all the processes of metabolism were still needed to create the biomass that was lost. However, since determining the exact proportion lost is extremely difficult, the simplification between the organism’s population, which we can easily measure, and the rate at which biomass is produced, which we cannot easily measure but which is most important for flux-balance analysis, is lost. The simplest solution to this problem is to only study biomass accumulation within a single life cycle and most my work has thus been limited to the 40-hour period in highly synchronised cultures where no RBC lysis occurs. This also requires the assumption that the biomass of the invading merozoites was negligible compared to the biomass of the final mature schizonts which seems sensible in comparison to other larger sources of uncertainty.

A further discussion of this complexity and how it relates to the units of flux in my final model forms the first section of the chapter on metabolic modelling.

Limits of the radioactive hypoxanthine assay

As briefly mentioned at the start of this chapter, a standard objective method for measuring *Plasmodium* growth-rate is the radioactive hypoxanthine assay. In this assay parasites are cultured with radioactive hypoxanthine in the growth-medium and the resulting radioactivity of haematocrit is used to calculate the amount of hypoxanthine incorporated by the parasites and thus their growth-rate.

The assay works because the parasite is unable to perform *de novo* synthesis of purines which must be imported from the growth medium and/or the host cell. Given a selection of purines it has been shown that hypoxanthine is taken up preferentially¹⁶⁰ and we know — both from personal experience and the literature — that it can be extremely difficult to maintain cultures in the absence of hypoxanthine⁴⁰. Furthermore, in the presence of known inhibitors hypoxanthine uptake is reduced the least of the alternative purine sources: adenine, guanine, inosine, adenosine and guanosine¹⁶⁰. More recent work agrees that hypoxanthine is taken up in preference to other purine sources¹⁶¹ but confusingly notes that,

“the rate of adenosine uptake was considerably higher than the rate of hypoxanthine uptake in infected human RBCs”

Quashie et al. 2010¹⁶¹

More worryingly for the assumption that hypoxanthine uptake is a proxy for growth-rate,

*“In both uninfected and *P. falciparum*-infected erythrocytes, hypoxanthine uptake was completely blocked in the presence of 1 mM adenine.”*

Quashie et al. 2010¹⁶¹

These findings must be treated with caution and considerable doubt is cast on the conclusions of Quashie *et al.*¹⁶¹ relating to purine uptake by Kirk *et al.*¹⁶². This disagreement leaves me with the impression that we do not yet fully understand the mechanism of hypoxanthine uptake or the effect on it of changing growth conditions.

Despite these complexities I think that it is probably reasonable to presume that as long as the growth medium contains an excess of hypoxanthine and there are no major changes to metabolism or growth conditions, then the amount of hypoxanthine incorporated by the parasites is likely to be more or less proportional to biomass formation. This linear relationship has indeed been shown over extremely short (≈ 10 s) time periods¹⁶¹ and periods up to five days¹⁵¹.

Still, we need to think carefully about what hypoxanthine uptake assays measure and how that might be affected by changes to metabolism not linked to growth-rate. If we want to perform experiments on the metabolism of the malaria parasite in response to perturbation that may change the behaviour of purine salvage — and they are areas of interest in this thesis and its possible expansions — I would suggest that the hypoxanthine uptake assay is not an acceptable measurement of growth-rate.

Experimental results

No account of the biochemistry of Plasmodium would be complete without some mention of the inherent pitfalls and problems encountered.

Biochemistry of Plasmodium (Malarial Parasites), I. W. Sherman¹⁰⁶, 1979

Towards a biomass function for malaria

Beyond a gap-free metabolic network, FBA requires both a biomass function and some constraints on the rate of exchange of compounds between the organism and its environment to accurately simulate metabolism. *P. falciparum* has been shown to be adaptable to different nutrient sources, and a minimal controllable growth medium¹⁶⁰ is used to reduce the variability between experimental repeats. Nevertheless, the requirement to use human RBCs — necessarily of different ages or from different donors across experiments — introduces an uncontrollable variability into all experiments. This variability is visible later in my measured exchange fluxes and is also reported in a large body of knowledge on human RBCs.

There is no published experimentally measured biomass function for *P. falciparum*. In explaining this, and perhaps light-heartedly, Giancarlo Biagini once suggested to me that a lot of the basic biochemistry was never done on malaria parasites because it couldn't be cultured when such experiments were in fashion, and by the time *Plasmodium* could be cultured people were no longer interested in publishing basic biochemistry. Irwin Sherman's¹⁰⁶ 1979 publication in *Microbiological Reviews* remains the most comprehensive general work but is not comprehensive enough to build a biomass function from. More recently a number of more specialised papers have been published which fill in some gaps in our knowledge about free nucleotide and amino acid composition¹⁴⁶ and fatty acid composition⁹⁹ but these papers still do not reliably report the whole-organism concentrations for the key biomass components: protein, DNA, RNA, lipids and carbohydrates.

At a push we could use existing data to build a biomass function for our own FBA analysis and in fact Chavali *et al.*'s 2008 analysis of *Leishmania Major*¹⁰⁸ includes a biomass function adapted from *E. coli* using published information on essential components of growth media for other species of *Leishmania*. This is not a reason to ignore the

predictions of that model and even for model organisms such as *E. coli*, debates continue as to the importance and relevance of an accurate biomass function¹⁶³.

Still, for *P. falciparum* none of the existing solutions to the biomass problem are attractive. There are good reasons to think that the biomass function of a human parasite would be very different to that of a bacterium and in any case the biomass functions of different species of bacteria vary enormously^{164,165}. Furthermore, even though the biomass function is probably not the most representative objective function of real metabolism⁷⁶ I think it remains — within an experimentally constrained model — a reasonable approximation of “life” and in any case the best available tool.

A first effort to measure the biomass of parasites grown in our own minimal media was a Masters project carried out in the year I started my PhD by Daniel Opi¹⁴⁴. We worked together to convert his results into a preliminary biomass function but in addition to solving some early problems this served as a warning that biomass measurement would be much more difficult than expected. At the time, the lab’s technique for measuring parasite biomass components was to use Sigma’s Tri-Reagent system to separate the components of a large amount of biomass in different solvents and then dry and weigh each solvent separately. This method returned negligible amounts of DNA and RNA and a protein concentration of 89% of total biomass in the best case and over 100% in the worst case.

By repeating the difficult large-batch cultures that Daniel had used for his experiments I was able to gather the large amounts of biomass needed to develop the techniques in this thesis. Ultimately, the new extraction technique I describe in this chapter can accurately determine the composition of the biomass extracted from small cultures and large-batch culturing is not required.

The best place to start in this discussion of my experimental results is with table 11 which summarises the major wet-lab experiments that led to the final biomass function calculations I report in this thesis. The majority of the culturing in these experiments was performed by the person whose name is written first in the experiment name and I am extremely grateful to Cheng Ma, Sara Zakutansky and Jennifer Lake for their work and achievements in their Masters degree projects.

A summary of experiments performed

Growth-rate and Biomass Experiments (2008 - 2009)

Name	Tom Biomass 1	Cheng FACS	Tom Biomass 2	Tom Biomass 3
Culture End Date	22nd December 2008	July 2009	18th November 2009	2nd December 2009
Extracted Culture Volume (ml)		-	540	1080
Extracted Parasitemia		-	3.2%	1.7%
Extracted Pellet Volume (ml)		-	18	30
Biomass Extracted	Yes	-	Yes	Yes
Description	Biomass samples used for preliminary experiments on biomass measurement (NMR metabolomics, glucose measurements by biochemical kit, Bradford Assay, RNA/DNA measurement).	21/03/09 fluorescence microscopy. I did not do this culturing.	Extraction 18/11/2009. Culturing continued to December 2009, Biomass samples used for Sara Masters Project technique development and calibration of HPLC.	Extraction 02/12/2009 used for final biomass measurements. Cultures were saponin-lysed together then the lysate split by four and centrifuged and isolated to 4 eppendorf tubes labelled "1/4"

Table 11 : Summary of major growth-rate and biomass experiments conducted in 2008 and 2009. The growth-rate experiments used to produce figure 56 on page 123 were performed up to 22/12/2008 and are not included in this summary chart but are present in my lab book.

Sources of uncertainty in biomass recovery

As summarised in table 11 — and table 20 in the metabolomics section later — biomass was recovered from a number of experiments. The development of the techniques for measuring biomass components, also part of Sara Zakutansky's and Jennifer Lake's Masters dissertations, used the biomass samples from the experiments named Tom Biomass 1 and Tom Biomass 2. The extracted samples from Tom Biomass 3 — all of which had been RNA-safe acquired and stored at -80°C with *RNAlater*[®] (*Ambion*) — were acquired at the late schizont stage before lysis and were used to gain the final results reported in this section.

Preliminary experiments allowed methods for determining biomass content to be improved with particular attention paid to four key areas,

Separating parasite biomass from RBC biomass

A key step in the isolation of parasite biomass is the separation of parasite and RBC components after the saponin lysis described in the methods chapter. The pellet visible after high-speed centrifugation of a lysed culture consists of a black core with a white covering. The black core is the parasite biomass, the white covering is formed of the remnants of the lysed red-blood cells and is called the RBC ghost. That the protein, DNA and RNA concentrations measured in table 13, table 14 and table 15 are within the expected range suggests that the separation of largely intact parasites and lysed red-blood cells is effective but it still seems likely that the two components of the pellet are not completely distinct. A more significant source of uncertainty in the isolated biomass weight is the process of separating the black and the white pellet section. RBC ghosts are “washed away” gently in TBS but this process is an inherently variable one with too much washing discarding some parasite biomass and too little leaving RBC ghosts in the parasite pellet.

There is an elegant method for quantifying the scale of this uncertainty. Following saponin lysis of the infected RBC pellet from the Tom Biomass 3 experiment the homogenous lysate was split equally into four centrifuge tubes for parasite isolation. The measured weights of the dried parasite biomass from each of these tubes (1,2,3,4) should be identical but we see in table 12 (page 145) that they are not. This variance lets us calculate a mean dry biomass of 13.4mg with a standard error (σ/\sqrt{n}) of 1.7mg: most easily expressed as a percentage error of 13%.

One suggestion that is often made for simplifying the process of getting the *P. falciparum* biomass function is that the metabolic model for *P. falciparum* is inserted into the existing

FBA-ready metabolic model for a human RBC¹⁶⁶. In this case, the biomass function for the RBC on its own would be substituted for the biomass composition of the RBC and the parasite together. Whilst this avoids the difficult step during biomass measurement of lysing RBCs and purifying the released parasites I suspect that this idea would not work. It is well documented^{94,113,127,167-170} that infected RBCs have massively altered metabolism and it has not yet been shown that any existing model for RBC metabolism is capable of reproducing these changes.

Weighing pellets

Two problems frustrated my first attempts at the simple task of weighing parasite biomass pellets. Firstly, I did not at first consider that eppendorf tubes vary considerably in mass; using one empty tube to tare a set of scales used to weigh another tube containing a biomass pellet produced highly misleading and useless results. As an example, some of the empty tubes used in the extractions summarised in table 12 weighed 0.8671g, 0.9005g, 0.8937g and 0.8996g, a considerable variation when compared to the biomass pellet weight. Secondly, I did not at first consider that the vigorous vortexing needed to re-suspend the isolated biomass pellet would grind away an appreciable amount of plastic from each eppendorf tube.

Careful weighing at every step of biomass handling solved these problems and any remaining uncertainty is included in the 13% error described in the previous consideration of parasite and RBC biomass separation.

Clarifying solutions before photometric assays

All the successful techniques for measuring the composition of parasite biomass have been photometric but these techniques are made impossible by the dark grey colour of the re-suspended parasite biomass. For each of the four samples and corresponding extraction methods I attempted, I clarified the samples before acquisition by two methods, **a** and **b**. Samples labelled **a** — and sample 4 — were clarified by centrifugation in a 2ml eppendorf tube at 5000RPM for 5 minutes except for 3a which was clarified by centrifugation at 13000RPM for 5 minutes. No noticeable change in protein or nucleotide concentration arose from this change. Samples labelled **b** were clarified by passing through the 40 micron filter cartridges provided with the PARIS(*Ambion*) kit in a benchtop centrifuge at 13000RPM for 5 minutes.

Solutions were visibly slightly clearer after filtration but the statistically indistinguishable results for samples 3a and 3b suggests that both techniques of clarification are equivalent.

Extracting biomass components from parasite pellets

In part because of the overestimation of protein content in Daniel Opi's thesis I wanted to avoid extraction of parasite biomass components using sodium dodecyl sulfate (SDS) as a detergent. The results reported in Teng *et al.*¹⁴⁶ made me hopeful that simple extraction techniques such as vigorous agitation in methanol followed by drying and re-suspension in TBS (methanol/TBS, sample 1) would be sufficient. I also wanted to try extracting biomass components through vigorous agitation in TBS alone (sample 2). Extraction using phenol/chloroform and perchloric acid were not performed partly because they did not perform well in Teng *et al.*¹⁴⁶ and partly because I felt uncomfortable performing them. A further extraction technique I tried was the cell lysis buffer component of the PARIS kit (sample 3 and sample 4).

A summary of the samples and their various treatments is shown in table 12 and the following results show that the PARIS kit's cell lysis buffer proved a simple and effective method for isolating biomass components.

Extraction Method	Methanol/TBS	TBS	PARIS cell lysis buffer	PARIS cell lysis buffer
Sample names	1a and 1b	2a and 2b	3a and 3b	4 (no filtration)
wet biomass weight (mg)	47.4	54.5	35.1	52.6
dry biomass weight (mg)	17.1	15.5	10.0	11.0
dry weight/wet weight	36%	28%	28%	21%
volume of extraction solvent (ml)	2	2	2	1
dry biomass concentration (mg/ml)	8.6	7.8	5.0	11.0

Table 12 : Extraction methods and biomass weights of the four samples labelled "1/4" isolated at the end of the Tom Biomass 3 experiment.

How much biomass is recovered

In total, 30ml of haematocrit at a parasitemia of 1.7% was extracted from the Tom Biomass 3 experiment. The dry weight of the total parasite biomass was $53.6 \pm 6.8\text{mg}$ with this mass coming from an RBC pellet of 30ml at 1.7% parasitemia.

From this — and the known concentration of RBCs in haematocrit (10^{10}RBCs/ml) — we can calculate N_p , the total number of parasites, to be,

$$N_p = 30 \times 0.017 \times 10^{10} = 5.1 \times 10^9 = 5.1 \text{ billion parasites}$$

The dry weight per parasite DW_{parasite} is thus,

$$DW_{\text{parasite}} = \frac{53.6 \times 10^{-3}}{5.1 \times 10^9} = 10.5 \pm 1.3 \times 10^{-12} \text{ g/parasite}$$

Protein content

The typical method of determining protein content with the Bradford assay is to create a calibration curve using a protein standard, BSA in my case, and then determine the protein content of a sample by comparing it to this curve. Preliminary tests with this technique were less accurate than I hoped and I settled upon another method — most clearly shown in figure 68 — where serial dilutions of both the BSA standard and the samples prepared by each method were measured for absorbance at 595nm.

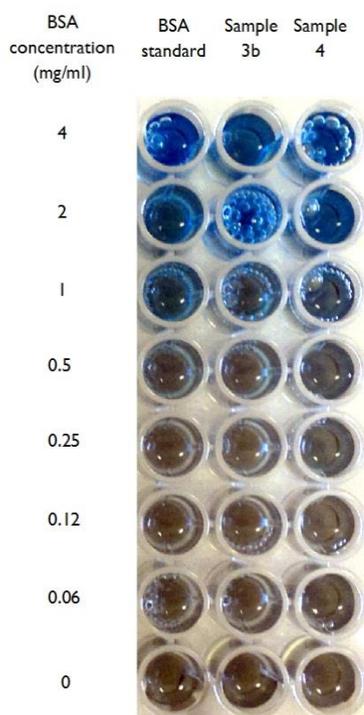


Figure 68 : Serial dilutions of BSA standard and samples 3b (from 5mg/ml of pure biomass) and 4 (from 11mg/ml of pure biomass) in Bradford reagent. Note the bubbles in the second row for sample 3b; this data-point was excluded from analysis.

Figure 69 shows the results of these measurements for sample 4 and the BSA standard with the gradient of lines of best fit shown. The graph is illustrative only and the points at 4mg/ml were discarded during numerical analysis because they were outside of the linear range of the assay.

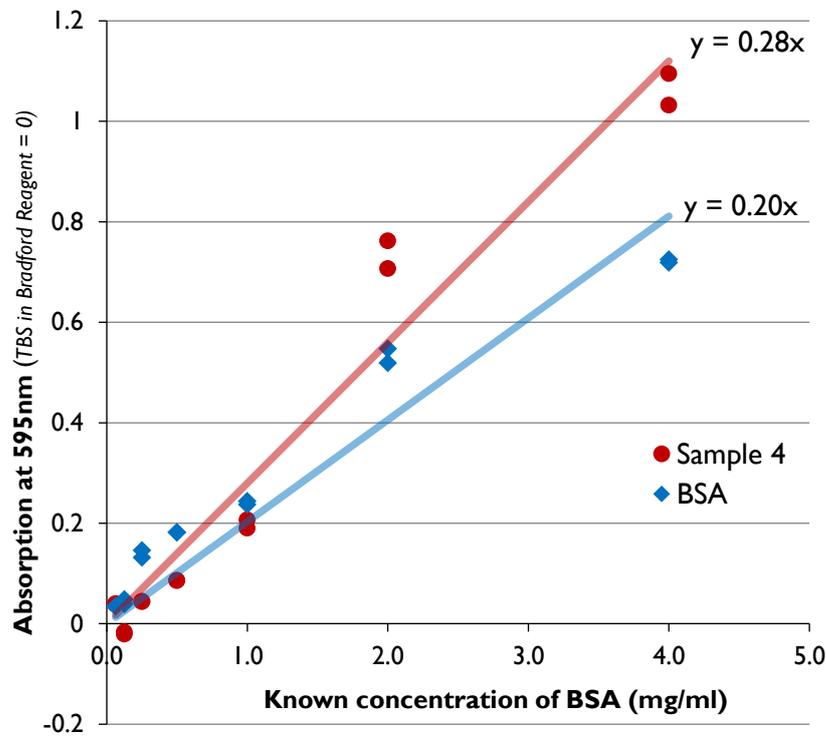


Figure 69 : Determination of protein concentration for sample 4 (see table 12) by the Bradford assay. There are two points at each concentration for each sample corresponding to acquisition at 15 minutes and 90 minutes.

Given the following algebraic identities,

P = protein as a percentage by weight of total biomass

C_S = concentration of biomass in pure sample (*mg of biomass/ml*)

C_B = concentration of pure BSA (*4mg/ml*)

m_S = Gradient of sample on BSA scale $\left(\frac{dA_{\text{Sample at 595nm}}}{dC_B} \right)$

m_B = Gradient of BSA $\left(\frac{dA_{\text{BSA at 595nm}}}{dC_B} \right)$

The percentage protein can be calculated because the ratio of the lines of best fit is equal to the inverse of the ratios of the protein concentrations in the two samples, thus,

$$P = \frac{C_B m_S}{C_S m_B}$$

Error in protein content determination

Errors for m_S and m_B (δm_S and δm_B) are calculated by the LINEST least-squares regression function in Microsoft Excel; also available in most other spreadsheets (eg. Libreoffice). The error in C_S is 13% as previously shown ($\delta C_S = C_S \times 0.13$) and the error in C_B is negligible and therefore ignored. These component errors are combined according

to the standard formula for the propagation of uncertainty from multiple sources¹⁷¹ to give an uncertainty in the final protein content. In this case this combination is,

$$\delta P = \sqrt{\left(\frac{\partial P}{\partial m_B} \cdot \delta m_B\right)^2 + \left(\frac{\partial P}{\partial m_S} \cdot \delta m_S\right)^2 + \left(\frac{\partial P}{\partial C_B} \cdot \delta C_B\right)^2}$$

Or explicitly,

$$\delta P = \sqrt{\left(\frac{C_B m_S}{C_S m_B^2} \cdot \delta m_B\right)^2 + \left(\frac{C_B}{C_S m_B} \cdot \delta m_S\right)^2 + \left(\frac{C_B m_S}{C_S^2 m_B} \cdot \delta C_B\right)^2}$$

The resulting protein content determinations and associated percentage-point errors for the seven samples are shown in Table 13.

	1a	1b	2a	2b	3a	3b*	4
% Protein	4%	4%	0%	-1%	54%	46%	43%
percentage-point error	1%	1%	1%	1%	8%	7%	7%

Table 13 : Protein concentration of biomass samples extracted using the techniques summarised in table 12. Because of the bubbles visible in the second row of the 3b sample, see Figure 68, this point was omitted from analysis. Extractions by the methods used for samples 1 and 2 were clearly unsuccessful.

The three separate samples, **3a**, **3b**, and **4** have the same protein content within the uncertainties of the experiment and can be combined to create a single best estimate of the protein content of *P. falciparum*'s late-stage schizont biomass. The uncertainties in this measurement are the combination — according to the previously described general formula for combining uncertainties — of the component percentage-point errors and the standard error (σ/\sqrt{n}) arising from the variance of the three constituents. This gives a final best-estimate of the protein proportion of late-stage schizont biomass of *P. falciparum* to be,

$$P = 48 \pm 9\%$$

It is important to remember that the quoted uncertainty is a percentage-point error and not a percentage error, in this case meaning that the protein content has been measured to lie somewhere between 39% and 57%.

* If z is a function of x and y (with uncertainties δx and δy), the uncertainty in z (δz) is given by,

$$\delta z = \sqrt{\left(\frac{\partial z}{\partial y} \cdot \delta y\right)^2 + \left(\frac{\partial z}{\partial x} \cdot \delta x\right)^2 + \dots}$$

with the ... representing the continuation of the pattern for all further variables.

DNA and RNA content

The DNA and RNA contents of the biomass extractions were measured using a nanodrop spectrophotometer and calculated as follows.

Given the algebraic identities,

D = DNA as a percentage by weight of total biomass

C_S = concentration of biomass in pure sample (*mg of biomass/ml*)

C_{DNA} = measured concentration of DNA in extraction

The DNA as a percentage of total biomass by weight is given by,

$$D = \frac{C_{DNA}}{C_S}$$

And — following the same formula for combining errors as for proteins — the error in D is given by,

$$\delta D = \sqrt{\left(\frac{1}{C_S} \cdot \delta C_{DNA}\right)^2 + \left(\frac{C_{DNA}}{C_S^2} \cdot \delta C_S\right)^2}$$

The resulting DNA content determinations and associated percentage-point errors for the seven samples are shown in table 14.

	1a	1b	2a	2b	3a	3b	4
% DNA	0.4%	0.4%	0.2%	0.1%	7.6%	8.3%	4.1%
percentage-point error	0.2%	0.1%	0.1%	0.0%	1.3%	1.1%	0.7%

Table 14 : DNA concentration of biomass samples. Extractions by the methods used for samples 1 and 2 were clearly unsuccessful.

Applying exactly the same treatment for the RNA measurements gives us table 15.

	1a	1b	2a	2b	3a	3b	4
% RNA	0.5%	0.7%	0.1%	0.1%	7.4%	6.6%	3.6%
percentage-point error	0.5%	0.6%	0.1%	0.1%	1.1%	0.9%	0.5%

Table 15 : RNA concentration of biomass samples. Extractions by the methods used for samples 1 and 2 were clearly unsuccessful.

Like with the protein level determination it is clear that the DNA and RNA content of samples **3a** and **3b** are the same within the uncertainties of the experiment but in this case it is unclear that sample **4** is part of the same set. To check that this was not because

sample 4 had a nucleotide concentration above the range of the nanodrop spectrophotometer I tested dilutions of the raw sample. These scaled correctly, showing that this was not the cause of the issue. Despite this I can see no reason not to combine the nucleotide determinations from samples 3a, 3b and 4 into a single reading. In calculating the error in this estimate the percentage-point errors reading was combined with the standard error (σ/\sqrt{n}) arising from the variance of the three constituents. This gives a final best-estimate of the DNA (D) and RNA (R) proportion of late-stage schizont biomass of *P. falciparum* to be,

$$D = 6.7 \pm 2.5\%$$

$$R = 5.9 \pm 2.1\%$$

As with the protein figure, the quoted uncertainty is a percentage-point error and not a percentage error.

Best guesses and sanity checks

One of the difficult but interesting parts of moving to a new scientific field is that you lack some very basic knowledge of techniques and expected outcomes. It is frequently disappointing, sometimes exciting and always very useful to check that results seem sensible. I made biologically sensible comparisons in the following four areas.

Wet weight/Dry weight ratio

Haemoglobin makes up around 92% of the dry weight¹⁷² of an RBC and around 35% of the wet weight of an RBC¹⁷³ thus inferring a dry weight/wet weight ratio of 38%. The equivalent figure for *E. coli* is around 30%¹⁶⁴. By comparison, the average dry weight/wet weight ratio for my parasite biomass extractions — see table 12 — is 28%: which seems sensible.

Nucleotides as a percentage of dry weight

Martin *et al.*¹¹⁶ report a final RNA mass of $\approx 40 \mu\text{g}/10^8$ parasites but do not report the mean dry weight of parasites in their study. If we use the dry weight I have measured, this figure corresponds to an RNA percentage of around 4%. The biomass function of Arvind Chavali's *L. major* model (sup. 1)¹⁰⁸ contains 1.6% DNA and 11% RNA as a percentage of dry weight. The biomass function of the *S. cerevisiae* in Förster *et al.* (sup. 3)¹⁷⁴ contains 0.4% DNA and 6.3% RNA as a percentage of dry weight.

My measured RNA figure of $5.9 \pm 2.1\%$ seems within the expected range but my DNA measurement of $6.7 \pm 2.5\%$ initially seems very high. Once we remember that the biomass was harvested from synchronised cultures at the most mature stage of schizogony my

results are more sensible. Since merozoites have been formed, the total amount of DNA will have increased by a factor of 8-32³⁶ from the level before schizogony began. The lower DNA content as a percentage of dry weight of the *L. major* biomass function is despite a larger genome of 32.8 million base pairs. This is to be expected in part because individual amastigotes of that species reproduce by simple division to produce two daughter cells and in part because the host macrophage is substantially larger meaning the minimum required amount of DNA represents a smaller percentage of total dry weight.

To try and see if this difference in number of daughter cells explains the amount of DNA I have measured I have performed a basic calculation to calculate the number of complete genomes the measured DNA mass represents.

The total DNA (DNA_{total}) in the parasite's biomass is formed of two components, free DNA nucleotides (DNA_{free}) and chromosomal DNA ($DNA_{chrom.}$) which add up to give the total DNA measured by my assays.

$$DNA_{total} = DNA_{free} + DNA_{chrom.} = DW_{parasite} \cdot D$$

An approximation of the amount of free DNA can be calculated from Teng *et al.*¹⁴⁶ (table 36 in appendix IV) giving a total of 0.9×10^{-10} mg/parasite of free nucleotides of which it seems reasonable to assume that around a third* are/were intended for DNA synthesis.

$$DNA_{free} = 30 \times 10^{-15} g/parasite$$

We know that the *P. falciparum* nuclear genome (G) is 22,853,764 base pairs. Taking the mass of an average base pair to be ($M_{bp} = 618 g/mol$) we can calculate the total mass of DNA in a single parasite genome (M_{DNA}) to be,

$$M_{DNA} = \frac{G \cdot M_{bp}}{N_A}$$

$$M_{DNA} = \frac{22853764 \times 618}{6.02 \times 10^{23}} = 23.5 \times 10^{-15} g/genome$$

The genome is haploid in all of the erythrocytic life-stages so the total mass of chromosomal DNA is given by multiplying the mass of a single genome by the number of genomes ($N_{genomes}$) to give,

$$DNA_{chrom.} = N_{genomes} \times M_{DNA}$$

* Here I split DNA, RNA and energy (ATP, ADP etc...) populations equally for a very rough calculation. Free DNA is a small enough proportion of total DNA that any error in this figure should have only a small effect on these calculations.

Combining all these formulas and rearranging for the number of genomes gives us,

$$N_{genomes} = \frac{(DW_{parasite} \cdot D) - DNA_{free}}{M_{DNA}} = \frac{(10.5 \times 10^{-12} \cdot 0.067) - 30 \times 10^{-15}}{23.5 \times 10^{-15}} = 28.6$$

This means that the best-estimate measurement of DNA content is within the maximum sensible figure of 32 genome copies per late-stage schizont. My feeling is that the most accurate figure is towards the lower end of the estimate for D with the equivalent calculation giving an average of 17.5 genome copies per late-stage schizont but the optical microscopy and staining techniques described in this thesis cannot reliably distinguish between 8, 16 and 32 nuclei in schizonts so I have little evidence either way.

Protein as a percentage of dry weight

The biomass function of Arvind Chavali's *L. major* model (sup. 1)¹⁰⁸ contains 45% protein as a percentage of dry weight, exactly the same as in Förster *et al.*'s *S. cerevisiae* model. This is in good agreement with my measurement of a 48% protein content.

Dry weight per parasite

The dry weight per parasite $DW_{parasite}$ is,

$$DW_{parasite} = \frac{53.6 \times 10^{-3}}{5.1 \times 10^9} = 10.5 \times 10^{-12} \text{ g/parasite}$$

By comparison the dry mass of an RBC is given as 45×10^{-12} g/RBC in Mysliwski and Korczak¹⁷⁵ and 32×10^{-12} g/RBC in Williams Hematology¹⁷³. My measurements and these figures suggest that the dry weight of a mature *P. falciparum* schizont is around 20-30% of the dry weight of a human RBC.

As previously noted, 92% of the dry weight of a human RBC is composed of haemoglobin. The principal source of amino acids in mature *P. falciparum* schizonts is from the digestion of the host erythrocyte's haemoglobin¹⁰⁰ and the best independent estimate from Krugliak *et al.* 2002¹⁰³ is that 65% of the infected RBC's haemoglobin is digested by the parasite of which up to 16% is retained. Multiplying these three percentages suggests that the protein mass of a mature *P. falciparum* schizont is around 10% of the RBC mass. My measurements show that protein makes up around half of the parasite's biomass and so the expected total biomass is just over 20% of the RBC mass; within the measured range of 20-30% of RBC mass.

Combining information from so many sources and confirming that my measurements are sensible is extremely pleasing but it raises a more serious point. I intended to use the biomass components from Teng *et al.*¹⁴⁶ to provide information on relative proportion of amino acids and nucleotides in my biomass function but it is clear from the total mass

returned in their paper that their techniques measure only free metabolites, not those incorporated in polymers. A better solution is required.

Details breakdown

In order to create a detailed biomass function the measured amounts of DNA, protein and RNA need breaking down into their constituent nucleotides and amino acids and the lipids and carbohydrates need to be added to the biomass composition to make it up to 100%.

Amino acid frequencies could naïvely be measured from the composition of translated proteins but a more complex analysis where the relative abundance of each protein is considered is already provided by Chanda *et al.*¹⁷⁶ and adapted as table 16.

Amino Acid	Frequency (%)
Aspartate	12.2%
Lysine	11.6%
Isoleucine	9.3%
Leucine	8.1%
Glutamate	7.0%
Serine	6.3%
Aspartate	6.0%
Tyrosine	5.5%
Phenylalanine	4.6%
Threonine	4.2%
Valine	4.1%
Glycine	3.1%
Arginine	2.9%
Glutamine	2.8%
Alanine	2.4%
Proline	2.2%
Histidine	2.2%
Methionine	2.2%
Cysteine	1.8%
Tryptophan	0.5%

Table 16 : Amino acid frequency in the *Plasmodium falciparum* proteome, from Table 1 in Chanda *et al.*¹⁷⁶. Frequencies sum to 99% due to rounding in the original paper.

DNA nucleotide composition is easy to add since we know that the majority of DNA nucleotides are incorporated in the chromosomes and we know the exact proportion of DNA nucleotides from the genome sequence⁴⁷.

DNA nucleotide Frequency (%)

dAMP	40.3%
TMP	40.3%
dGMP	9.7%
dCMP	9.7%

Table 17 : DNA nucleotide frequency in the *P. falciparum* genome.

Another contribution to total RNA comes from gene transcripts in the form of messenger RNA (mRNA) and from Gardener *et al.*⁴⁷ we know that around 76.3% of the nucleotides in introns are A or T meaning the mRNA composition should be about 76.3% T and U, giving the final RNA nucleotide composition shown in table 18 .

RNA nucleotide Frequency (%)

AMP	38.7%
UMP	38.7%
GMP	11.3%
CMP	11.3%

Table 18 : RNA nucleotide frequencies are calculated by taking into account both the ribosomal RNA nucleotide composition and the likely composition of mRNA from gene transcription. The rRNA:mRNA ratio is set at 1:1.

A complexity that I have omitted in my final model with regard to the RNA content of the biomass function is that, in quickly growing cells, ribosomal RNA makes up the major part of the total RNA; typically 85% in *E. coli* and 81% in *S. cerevisiae*¹⁶⁵. From Vezza *et al.*¹⁷⁷ we know that the ribosomal RNA (rRNA) of *P. falciparum* has around a 35-37% G-C content and the AMP and UMP composition of the biomass may well be correspondingly lower. I have not made this correction since the proportion of RNA which is ribosomal is unknown. In any case, making such a change results in negligible changes to the model's behaviour and predictions.

No serious attempts were made to quantify the carbohydrate and lipid composition of the biomass and so these remaining components were added in the same ratio (carbohydrate 27:15 lipids) as they are including the *L. major* biomass function of Chavali *et al.*¹⁰⁸. The breakdown of lipid and carbohydrate components of the biomass function are discussed in the next chapter of this thesis as the pathways that generate are considerably simplified in the final model.

All of this gives the final biomass composition shown in table 19 and figure 70.

Component	% dry weight
DNA	6.7%
RNA	5.9%
Protein	48.0%
Lipids	14.1%
Carbohydrate	25.3%
Total	100.0%

Table 19 : The components of the final biomass function as a percentage of dry weight.

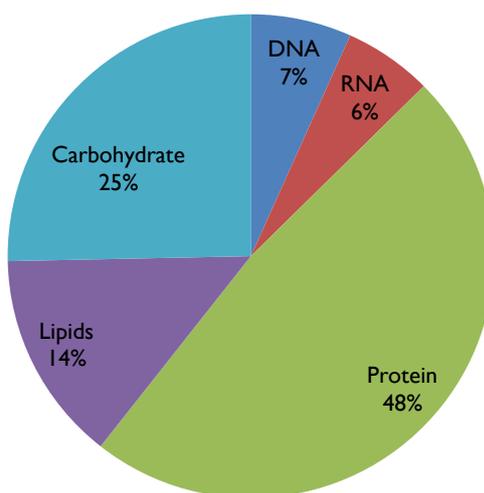


Figure 70 : The components of the final biomass function as a percentage of dry weight.

Measuring exchange fluxes

What started as an idea to extend Jennifer Lake's Masters project evolved into a key part of my thesis when I spoke at the White Rose Doctoral Training Centre internal conference in Sheffield in late 2009. I was pitching my ideas to perform metabolomics on the malaria parasite using liquid chromatography linked to mass spectrometry (LC-MS) techniques as described widely for many other organisms. The specialisms and experience at Leeds were not ideal for trying these types of experiments but Cassey McRae and Hayley Fenton in Julie Fisher's group in Chemistry at Leeds said that what I was looking to do was very similar to the work they were doing on the metabolomics of cancer patients, looking at the composition of body fluids.

Jennifer had been measuring the rate at which glucose was taken up by the parasite and the rate at which lactate was produced using biochemical kits which were inaccurate and unreliable*. Preliminary tests we performed using ¹H-NMR (Proton NMR) metabolomics proved to be more reliable and were additionally able to distinguish more metabolites. A summary of these preliminary trials and the three major metabolomics experiments that followed are contained in table 20. In this section I will explain the results and the techniques I have developed.

The methods and data analysis are identical for all compounds and so I start with a detailed description of the techniques. The results are then split into carbon compounds and amino acids. Glutamine is considered as a carbon compound rather than an amino acid due to its recently clarified role as the principal carbon source — via conversion to oxoglutarate — driving the mitochondrial TCA cycle⁹⁵ as shown earlier as figure 33 on page 73. The magnitude and direction of the exchange fluxes measured also supports glutamine being primarily considered as a carbon compound.

Splitting the metabolites this way makes sense because the exchange fluxes measure two different things. The exchange fluxes for carbon compounds primarily tell us how the parasite produces the energy it needs to survive, with the magnitude and choice of pathway providing further information in areas such as redox metabolism. The amino acid exchanges tell us how the parasite acquires the amino acids it needs to synthesise proteins and the way it deals with the haemoglobin it digests.

Hypoxanthine exchange fluxes are included with the amino acid analysis simply because the magnitudes of the fluxes are similar. The NMR assignment of hypoxanthine is unclear and the results should be treated with even greater caution than their high uncertainties and wider variability suggest.

* Protocols are in the methods section, a full evaluation is contained in Jennifer's thesis¹⁴⁵.

Metabolomics and Biomass Experiments (2010-2011)

Name	Jenny Metabolomics	Tom repeat of Jenny Metabolomics	Sara Metabolomics of Inhibition	Tom repeat of Sara Metabolomics of Inhibition
Culture End Date	July 2010	17th November 2010	June 2011	22nd December 2011
Extracted Culture Volume (ml)		120	2 x 13	8
Extracted Parasitemia		4.1%	3.5% (uninhibited) 4.5% (inhibited)	2.6%
Haematocrit volume at extraction (ml)		3.75	2 x 1.25	0.32
Biomass Extracted Successfully	No	Partially	Yes	No
NMR Acquisition Month	29/07/2010	21/02/2011	July and August 2011	January 2012
Time Points Harvested (including fresh medium)	2	11	13	5
Time Points Measured	2	11	5	5
End Time	24	41		
Blood Nulls	No	Yes, insufficient	Yes, sufficient	No
Notes	Ethanol production detected.	Methanol sterilisation. No ethanol production detected.	Full set of blood nulls both with and without atovaquone inhibition, Normal growth and sublethal atovaquone inhibition time course of parasitised red blood cells	No blood nulls. Significantly lower base metabolism.
Description	Fresh Albumex growth medium was measured for the first time. Three biological replicates of used growth medium harvested after 24 hours were measured.	Single blood null at 14.50 hours not accurate enough. Ethanol production result from Jenny's results not reproduced. 11 time-points taken for one sample. 1 time-point taken at 31.58 hours for a biological replicate largely agrees with main sample	Strange experimental technique means changing haematocrit percentage. Abnormally high haematocrit at end of experiment may affect results. Strange technique has advantage of increasing sensitivity. Sample B is uninhibited. Sample A is inhibited with atovaquone at 3.2nM.	Sample 1 is uninhibited. Samples 2 is inhibited with atovaquone at 3.2nM

Table 20 : Summary of metabolomics and biomass experiments performed in 2010 and 2011.

Collecting spectra

I create spectra for each growth medium sample using the $^1\text{H-NMR}$ techniques described in the methods section. These spectra are then grouped by experiment like in the example shown in figure 72 for the Sara A set from the “Sara metabolomics of inhibition, June 2011” set of experiments summarised in table 20.

The exact position of peaks is not constant across different experiments even if the data processing is identical. It is therefore extremely important to align each peak manually in the ACD/labs software to ensure that the range of integration is consistent across spectra and for this reason it is best to integrate spectra that will later be compared in sets.

An example of the visual alignment required — in this case the prominent α -Glucose doublet at 5.25ppm — is shown in figure 71 and each peak requires its own alignment. This figure also gives an excellent visual example of how glucose is consumed by the parasite over time, leading to a large decrease in its concentration in the growth medium. A final note on this image is that whilst the sizes of the peaks in this case largely correspond to the concentration of glucose in the medium this is often misleading due to the visual normalisation techniques in the ACD/labs software. Although this kind of visual inspection can be useful for peak identification, the reported area under each curve — and not the visual representation of the area — is the only reliable way to assess the concentration of each metabolite.

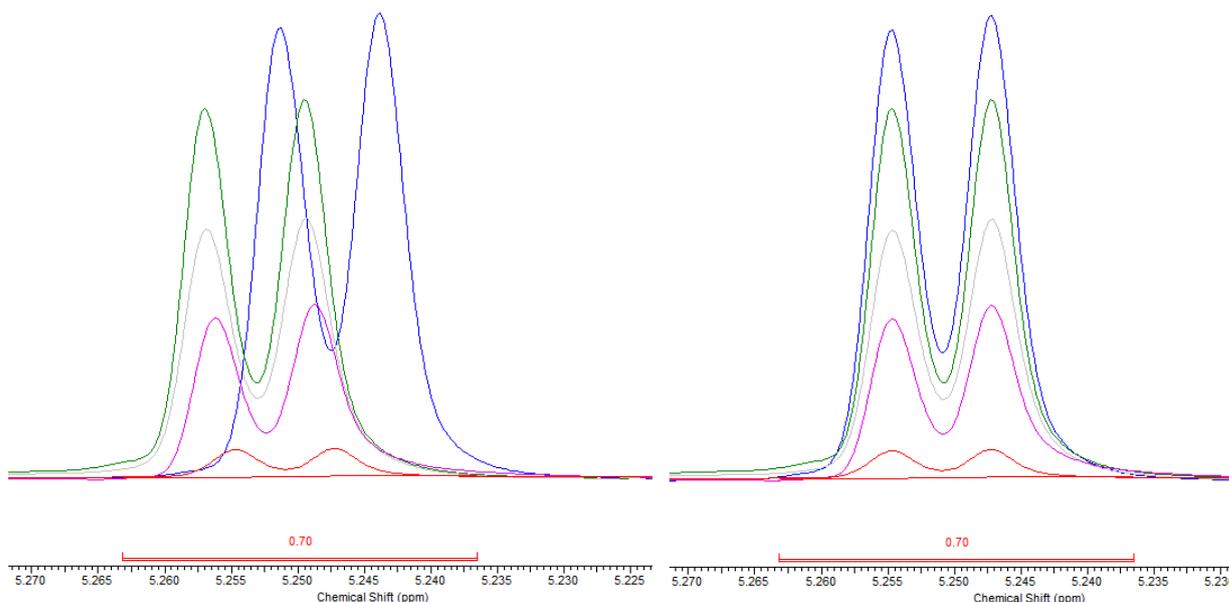


Figure 71 : $^1\text{H-NMR}$ spectra for the α -Glucose doublet at 5.25ppm from the Sara A set. (left) Positions as referenced to TMS at 0ppm. (right) Positions once manually aligned. Colours blue, green, grey, pink, red, correspond to $t = 0, 9.2, 18.5, 27.0, 36.5$ hours and show the clear decrease in glucose concentration in the growth medium over time.

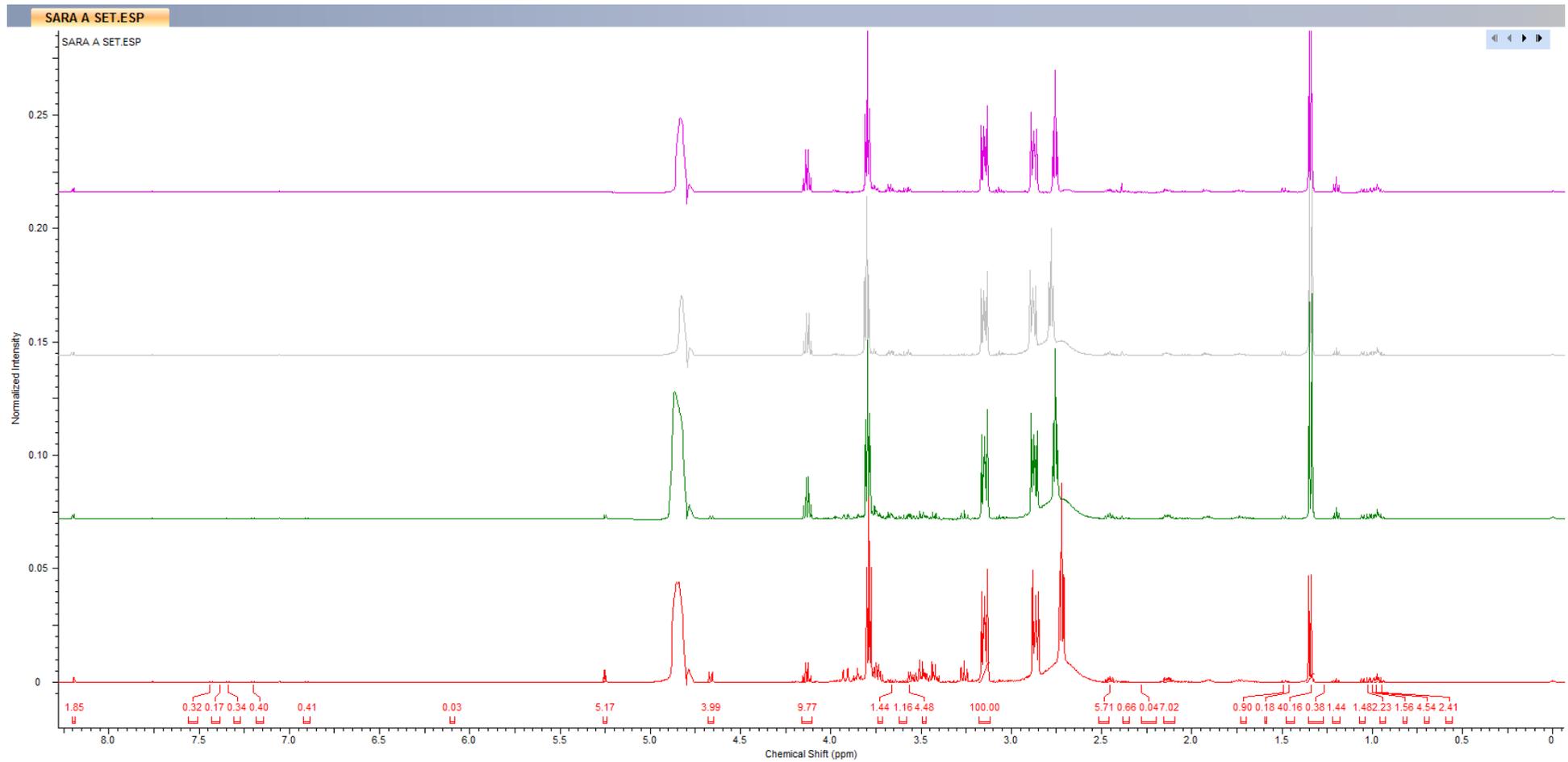


Figure 72 : The four ¹H-NMR spectra in the Sara A set. Here the integrals for the selected spectrum (red, 30/06_00h20_A, t = 9.2 hours) are shown. The two symmetrical HEPES multiplets at 2.86 and 3.15 (integral referenced to 100) are visible. Colours green, grey, pink, red, correspond to t = 9.2, 18.5, 27.0, 36.5 hours.

Picking peaks

With the spectra of all the acquired samples collected into sets for each experiment we need to identify the peaks within the spectra that correspond to metabolites of interest. Initially we drew heavily on the time and experience of Cassey McRae, a PhD student in Julie Fisher's lab who showed us how to identify and interpret spectra. My method for identifying peaks has evolved but in my final full review of my results was as follows.

1. Visually move through the set of spectra and note the positions and forms (singlet, doublet, multiplet, etc...) of visible peaks. In total 34 peaks or sets of peaks were found.
2. Look for peaks at the same position in Cassey's suggested reference paper, "750 MHz ¹H and ¹H-¹³C NMR spectroscopy of human blood plasma" by Nicholson et al.¹⁷⁸.
3. Search for other peaks using the human metabolome database¹⁷⁹ (HMDB) available at www.hmdb.ca and the Biological Magnetic Resonance Data Bank (BMRDB) at <http://www.bmrb.wisc.edu/>.
4. Verify proton assignments for each peak using the interactive ¹H-NMR viewer at <http://mmcd.nmrfam.wisc.edu/>, part of the Madison Metabolomics Consortium Database¹⁸⁰ (MMCD).

With a large number of the peaks identified I downloaded the reference spectra from HMDB for the following metabolites,

Malate, leucine, isoleucine, ethanol, glucose, glycerol, histidine, hypoxanthine, alanine, glutamate, glutamine, valine, lactate, NAD, NADP, phenylalanine, tyrosine.

These reference spectra, as shown separately in figure 73 and combined in figure 74, were used to check that the peaks identified in my samples were at the correct position and of the same shape as expected. Metabolites that were known to be present in the growth medium were not included in the reference set of spectra where the only measurable peaks overlapped with other strong signals. For example asparagine is present in RPMI medium but its principal peaks at 2.8–3.0 ppm are obscured by a HEPES peak at 2.84–2.90 ppm. Arginine's triplets at 3.23ppm and 3.76ppm are obscured by glucose and glucose/glutamate/glutamine respectively. There are techniques to separate overlapping signals in NMR spectra but they introduce considerable error into the results¹³⁰, especially when used by someone with no experience of them and when one signal is considerably stronger than the other.

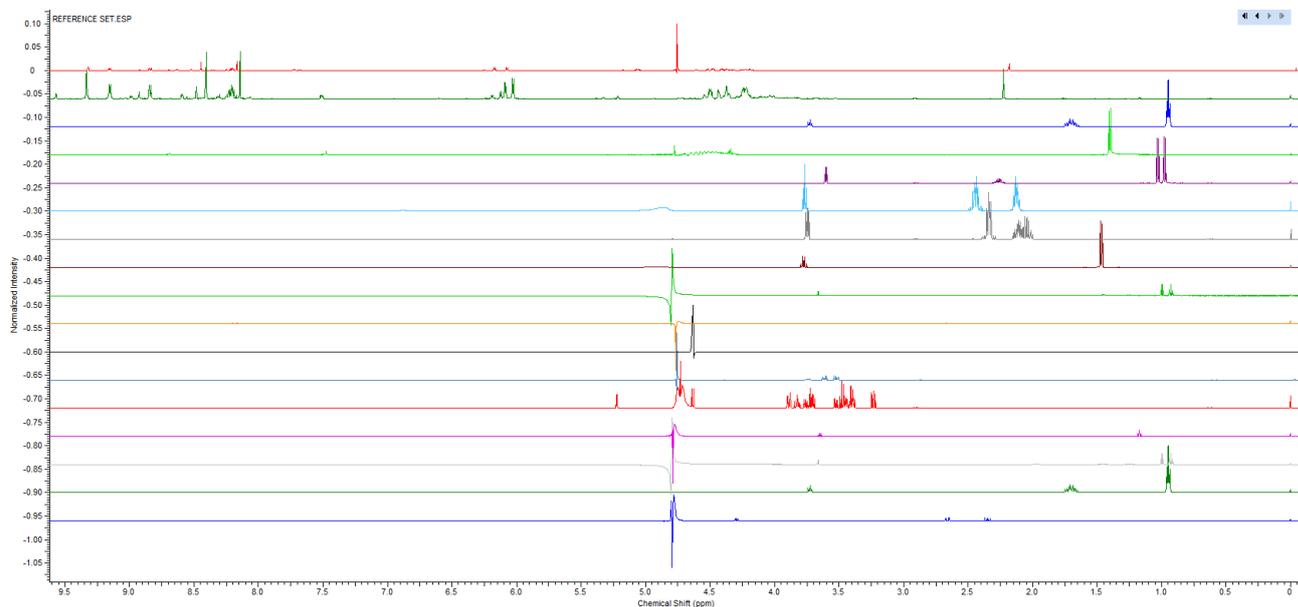


Figure 73 : Collection of ^1H -NMR reference spectra used as a guide to peak identification displayed separately. From the top, spectra are of NADP, NAD, Leucine, Lactate, L-Valine, L-Glutamine, L-Glutamate, L-Alanine, Isoleucine, Hypoxanthine, Histidine, Glycerol, Glucose, Ehtanol, Isoleucine (included twice in error), Leucine (included twice in error), Malate.

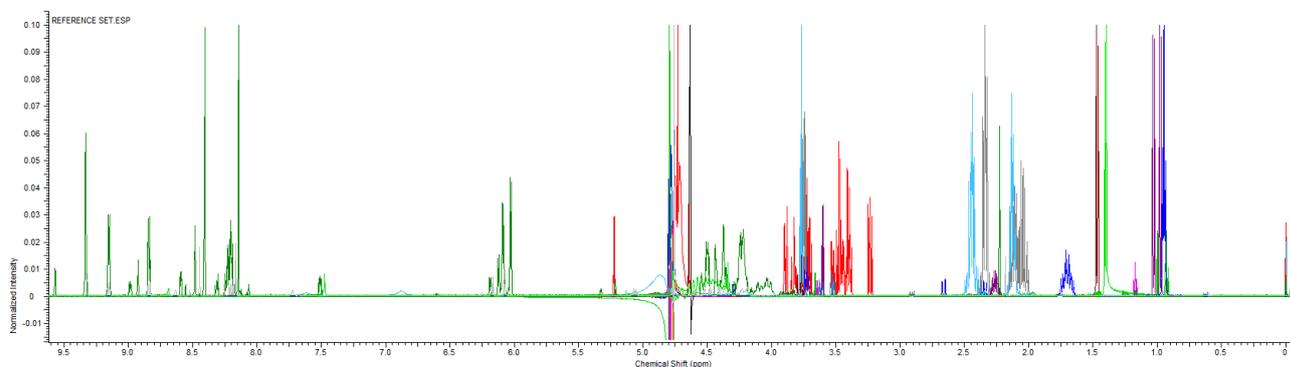


Figure 74 : Collection of ^1H -NMR reference spectra used as a guide to peak identification displayed together. Other reference spectra were downloaded and examined individually but determined not to be present in any spectra.

A total of 21 peaks represent 12 unique metabolites (counting α and β -glucose together^{*}) were identified and are summarised in table 21 with the majority of these metabolites visible in all 42 spectra acquired. The glycerol assignments and associated results should be treated with caution and I discuss this further at the end of this chapter.

* The α and β isomers of glucose interconvert spontaneously and the distinction between the two forms within my model could be removed without affecting my results. Restrictions on glucose flux in the modelling section of this thesis are always on the combined flux of both forms.

	Observed Position (ppm)	Peak Type	Nicholson Position (ppm)	Hydrogens	Molecular Weight (g/mol)
Isoleucine	0.95	t	0.93	3	131.17
Leucine	0.97	t	0.97	6	131.17
Valine	0.99	d	0.97	3	117.15
Isoleucine	1.02	d	1.00	3	131.17
Valine	1.05	d	1.02	3	117.15
Ethanol	1.16	t	-	3	46.07
Lactate	1.33	d	1.34	3	90.08
Alanine	1.46	d	1.46	3	89.09
Glutamine	2.13	m	-	2	146.14
Glutamine	2.45	m	-	2	146.14
HEPES	3.15	m	-	2	238.30
Glycerol	3.56	dd	3.56	2	92.09
Glycerol	3.66	dd	3.64	2	92.09
Lactate	4.12	q	4.12	1	90.08
β -Glucose	4.65	d	4.66	1	180.16
α -Glucose	5.24	d	5.24	1	180.16
Tyrosine	6.90	m	6.87	2	181.19
Tyrosine	7.20	m	7.17	2	181.19
Phenylalanine	7.33	m	7.33	2	165.19
Phenylalanine	7.38	m	7.38	1	165.19
Phenylalanine	7.43	m	7.43	2	165.19
Hypoxanthine	8.20	d	-	2	136.11

Table 21 : NMR peak assignment table ordered by observed peak position (ppm). The two unobserved malate peaks and the HEPES peak used as a concentration reference are included in addition to the 21 metabolite peaks representing 12 unique metabolites that were observed in at least some spectra. The hypoxanthine assignment is not very clear and should be treated with caution. In all cases the Madison Metabolomics Consortium Database¹⁸⁰ was used as the definitive source of hydrogen counts for each peak.

With the peaks identified, the area under each of them was integrated within the ACD/labs software with the integral expressed as a percentage of the HEPES peak at 3.15ppm.

Converting metabolite integrals to metabolite concentrations

At the level of NMR analysis used in this thesis, the integral of a peak (I_n) is proportional to the number of molecules in solution (proportional to the milliMolar concentration ($[n]_{mM}$)) times the number of hydrogen atoms (H_n) in each molecule that produce the integrated peak. For all metabolites in solution,

$$I_n \propto H_n \times [n]_{mM}$$

And thus

$$\frac{I_n}{H_n [n]_{mM}} = \text{constant}$$

From this relationship we can calculate the milliMolar concentration of a metabolite from the known concentration of any other metabolite. Often an external concentration standard is added to samples before acquisition but this requires extremely accurate pipetting and sample preparation at the small acquisition volumes ($\approx 300\mu\text{l}$) of our NMR system. Instead we make use of the extremely clear peaks of the HEPES* buffer used at very high concentration (see table 35 in appendix III) in RPMI 1640 medium. In particular I have chosen to use the multiplet at 3.15ppm produced by two hydrogens.

$$\frac{I_n}{H_n[n]_{mM}} = \frac{I_{HEPES}}{H_{HEPES}[HEPES]_{mM}}$$

Isolating $[n]_{mM}$ gives us,

$$[n]_{mM} = \frac{I_n \cdot H_{HEPES} \cdot [HEPES]_{mM}}{I_{HEPES} \cdot H_n}$$

This milliMolar concentration can be converted to the more useful mg/ml using the molar mass of the metabolite (M_n),

$$[n]_{\text{mg/ml}} = [n]_{mM} \times \frac{M_n}{1000}$$

The integrals of different peaks within the reference spectra for each metabolite available from the human metabolome database¹⁷⁹ show that small deviations from this basic theory occur but applying these corrections is difficult and unnecessary since the uncertainties are small compared to biological variance.

HEPES as a concentration reference

Using HEPES as an internal concentration reference is not completely new¹³⁰ but it is rare enough that documentation is not widely available. Indeed the technique has its drawbacks, namely that,

... the use of protonated, organic buffers (e.g., Tris, MOPS or HEPES) should be avoided. Signals from these additives can mask important metabolite resonances or can be mistaken as "unknown" endogenous metabolites.

Quantitative metabolomics using NMR, David. S. Wishart, 2008¹⁸¹

Since HEPES exists in the growth medium already there is no easy way to avoid this problem and it makes sense to use the HEPES signals if possible.

* 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid.

The NMR spectrum of HEPES is strange in that it changes considerably with concentration* and is extremely poorly predicted by the best commercial NMR spectrum prediction software (MestReNova®, analysis performed in collaboration with a fellow doctoral student, Simon England at the University of Sheffield). The BMRDB assignment table for HEPES points to there being five distinct peaks, of various multiplicity, at 2.851, 2.936, 3.056, 3.151 and 3.851ppm with hydrogen assignments of 4, 2, 2, 6 and 2 respectively and in light of this Cassey McRae assigned the peak at 3.15ppm to 6 hydrogens.

The problem with these assignments is that the peaks I observe in my samples, table 22, do not agree with them.

Chemical shift (ppm)	Peak type	Integral
2.72	t	88.7
2.86	m	92.2
3.15	m	100.0
3.78	t	105.2
	Mean	96.5
	St. dev.	6.4

Table 22 : Chemical shifts and integrals of the four HEPES peaks in my samples. Integrals are referenced to the clearest and most reliable peak at 3.15ppm.

My observed four peaks with equal integrals are far more consistent with the HEPES spectrum in MMCD (Madison Metabolomics Consortium Database¹⁸⁰). In that spectrum the four peaks are the same shape as I observe but only the 3.15ppm peak is in the same position. That the peaks are in different positions is not ideal but it seems reasonable given the way the reference spectra vary with HEPES concentration.

There is a second concern with using the 5958mg/L (25.03mM) of HEPES in RPMI 1640 growth medium as a concentration reference. In Teng *et al.*¹⁴⁶ it is reported that HEPES accumulates in the parasite biomass to a final concentration of at least $12 \pm 2 \text{ mM} = 80 \pm 13 \text{ g}/10^{15}$ parasites . This is the highest concentration of any measured metabolite accumulated in the parasite biomass and significantly higher than the three next most significantly accumulated metabolites: glutamate at around $30 \text{ g}/10^{15}$ parasites and glutathione and ATP at around $20 \text{ g}/10^{15}$ parasites. This high absorption of HEPES by the parasite will lower its concentration in the used growth medium and could thus falsely suggest an increase in the concentrations of the other metabolites. An estimate of the percentage error that this problem could introduce can be calculated as follows.

* The spectra on the Biological Magnetic Resonance Data Bank (BMRDB) at <http://www.bmrwisc.edu/> show this very clearly.

The initial concentration of HEPES in the growth medium and the amount absorbed by the parasites is roughly,

$$[HEPES]_{\text{initial}} = 6.0\text{g/L}$$

$$[HEPES]_{\text{parasite_final}} = 80\text{g}/10^{15}\text{parasites} = 80 \times 10^{-15}\text{g/parasite}$$

Assuming a typical haematocrit of 5%, parasitemia of 5% and packed RBC density of 10^{10} RBCs/ml of haematocrit¹⁸² we can calculate the number of parasites per litre of culture.

$$[Parasites] = 1000 \times 10^{10} \times 0.05 \times 0.05 = 2.5 \times 10^{10}\text{parasites/L}$$

The total amount of HEPES absorbed by these parasites is thus,

$$M_{\text{absorbed HEPES}} = [HEPES]_{\text{parasite_final}} \times [Parasites]$$

$$M_{\text{absorbed HEPES}} = 2.5 \times 10^{10} \times 80 \times 10^{-15} = 2 \times 10^{-3}\text{g/L} = 2\text{mg/L}$$

The percentage change in HEPES concentration in the growth medium due to absorption by parasites is thus,

$$\text{Percentage of HEPES absorbed} = \frac{M_{\text{absorbed HEPES}}}{[HEPES]_{\text{initial}}} = 0.3\%$$

A 0.3% underestimation of metabolite concentrations is tiny compared to other uncertainties and can be ignored. Accumulation of HEPES by uninfected RBCs has not been reported as a major concern. The effect of HEPES absorption may become an issue in experiments conducted at higher haematocrit concentrations and higher parasitemias.

A further concern using HEPES as the stated reference concentration of 5958mg/L is that it is diluted by 5-10% when the RPMI medium is combined with the haematocrit for culturing. It might be sensible to make a correction to account for this but for the unrelated finding in Lewis *et al.*¹³⁰ that metabolite concentrations calculated from ¹H-NMR spectra and referenced to HEPES were typically underestimated by a similar amount.

A note on the graphs and axes

I am convinced that the graphs on the next few pages are the best way to show my results but I am also aware that they are heavy on information and require detailed explanation. The four sets of experiments summarised in table 20 are all displayed on the same graphs (figure 76 — figure 79 and figure 82 — figure 88) with the experiments grouped according to shape of marker;

- Diamonds — Jenny preliminaries.
- Dots — Tom repeat of Jenny.
- Squares — Sara metabolomics.
(Set A is inhibited with sub-lethal dose of atovaquone, Set B is grown in normal growth medium, Sets C and D are uninfected blood nulls for set A and B respectively.)
- Triangles — Tom repeat of Sara.

The colour of each series is different except for blood nulls — parallel cultures of identical but uninfected RBCs — which are always shown in blue.

The y-axes of the graphs have units of mg per 10 billion RBCs which is a product of the measured concentration of the metabolites in the growth medium in mg/ml and the volume of homogenous culture containing an average of 10 billion RBCs (ml/10 billion RBCs). For metabolites such as lactate and glycerol that are not present at detectable levels in the growth medium at $t = 0$ this is a perfect unit for the y-axis. For other metabolites with a significant concentration at $t = 0$ the choice of y-axis can be confusing. This is because an equal concentration will have a different value depending on the haematocrit concentration in the culture. The major upside of the choice of y-axis unit is that gradients are directly comparable across all experiments and correspond to the exchange rates of the compound between the growth medium and the population of infected and uninfected RBCs.

A final explanation needs to be made of the starred points present in figure 76, figure 77 and figure 78. As described on page 114 of the methods chapter the Jenny metabolomics set of experiments used a complex technique that required an equivalent haematocrit to be calculated for each measured point. These calculations assume a constant exchange of metabolites but in the case of the Sara A set the glucose in the growth medium was exhausted and could no longer be absorbed by the parasite. This had a knock-on effect on lactate and glycerol entering the growth medium and the assumption of constant exchange rate in the haematocrit correction no longer held.

Although figure 76 appears to show lactate being reabsorbed by the RBC population in the final time step the actual levels in the growth medium increased over this period, from 4.65mg/ml to 4.67mg/ml. Likewise with glycerol, figure 78 greatly exaggerates the reduction in glycerol concentration. In the last three time points the concentrations of glycerol in the growth medium were 2.25mg/ml, 3.14mg/ml and 2.82mg/ml. Whilst the idea that glycerol produced from glucose may be reabsorbed by the parasite once the glucose is exhausted is fascinating — and in keeping with models of *E. coli* metabolism⁷⁶ — it is impossible to suggest this based on a single point in a single experiment with such large underlying uncertainties.

The experimental setup used to get the results in the Sara metabolomics set of results is undoubtedly complex and in appendix IX to this thesis I discuss much simpler ways to achieve even better results. Despite its many drawbacks this overly complex system allowed high sensitivity assays to be conducted and has produced some fascinating results.

Atovaquone inhibition

Two of the sets of experiments, Sara metabolomics and Tom repeat of Sara metabolomics, had one time series where the growth medium contained 3.2nM of the known inhibitor atovaquone in addition to an experiment in normal conditions. This was the concentration of the compound measured by Sara Zakutansky in her Masters thesis to reduce the parasite population's growth-rate to 1; no increase or decrease in population over time. We choose a sub-lethal concentration of the inhibitor in the hope of observing not the death of a culture — which could lead to the complete breakdown of metabolism — but rather a highly stressed but still functional version of metabolism.

The mechanism of action of atovaquone is well known, with precise details in Olliaro¹²⁶ amongst others. At its most simple it is an inhibition of the enzyme dihydroorotate dehydrogenase which catalyses reaction R01867 as shown in figure 75. Dihydroorotate (DHO) is primarily produced from aspartate and the conversion of glutamine to glutamate. Orotate goes on to form pyrimidine nucleotides. The results shown in figure 75 show that with a 3.2nM concentration of atovaquone in the growth medium the flux through R01876 is slowed, leading to an accumulation of dihydroorotate before the reaction and a reduction of the amount of orotate after the reaction.

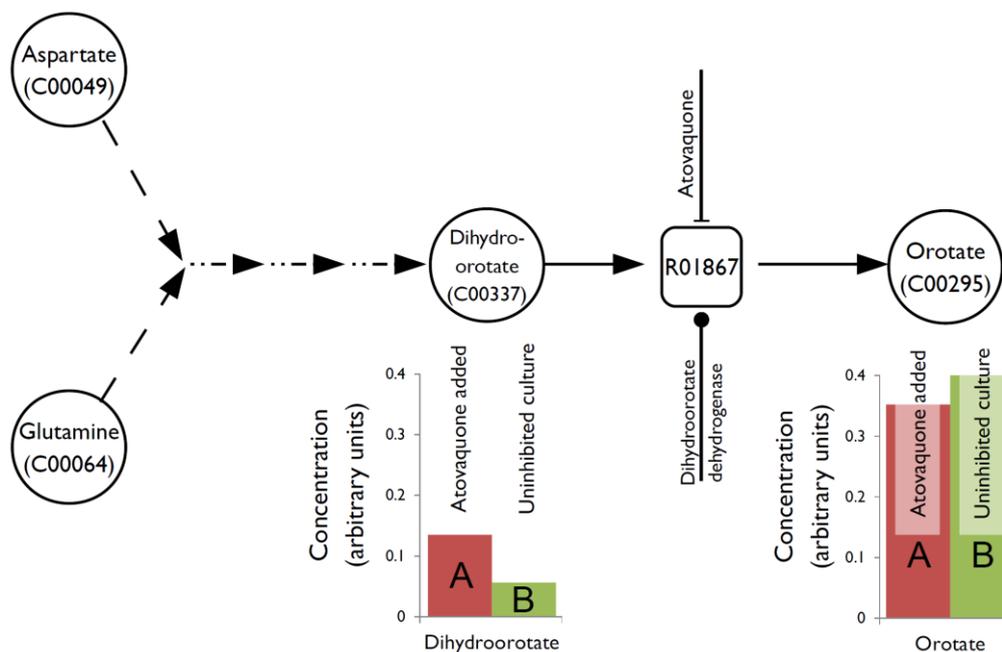


Figure 75 : HPLC measurements of dihydroorotate and orotate concentrations in the parasite biomass pellets extracted at the end of the Sara metabolomics set of experiments suggest that sub-lethal atovaquone inhibition (sample A) was successful in reducing the flux through R01867 leading to accumulation before, and reduced concentration after, the reaction.

The results in figure 75 are presented briefly because the parasite biomass extractions in these experiments were performed using the methanol/water technique I later found to be ineffective in freeing the contents of isolated biomass pellets*.

Uncertainties on individual points

The graphs in the remainder of this chapter are already full of information so I have not included estimates of the errors in individual readings even though I have made efforts estimate these. As I discuss later, no two batches of RBCs have the same metabolic properties so I cannot calculate errors by comparing results from different experiments; the variability of RBC metabolism across different experiments is far greater than the technical error in each of my measurements.

One way of estimating the technical uncertainties in individual points is by making use of the multiple measurements of fresh growth medium since these samples have never been in contact with RBCs or parasites.

* A more complete set of results and detailed methods are contained in Sara Zakutansky's Masters thesis¹⁴⁷ but are the subject to the same caveat.

	Valine	Leucine	Isoleucine	Tyrosine	Phenylalanine	Glutamine	Glucose	Hypoxanthine
Mean Concentration (mg/ml)	0.02	0.03	0.04	0.03	0.02	0.25	1.85	0.24
Standard Deviation	0.002	0.009	0.004	0.006	0.006	0.05	0.48	0.05
Standard Deviation as percentage of mean concentration	10%	30%	9%	22%	28%	20%	26%	21%

Table 23 : Mean and standard deviation for eight different metabolite concentrations for the five different $t = 0$ fresh medium samples across all experiments.

The result of this error analysis is in table 23 but we need to be wary of the results. My feeling is that a large part of the uncertainty calculated in this measurement is actually variability in the growth medium across experiments from differing amounts of AlbuMAX I powder used in preparation, different composition of any remaining serum not washed from the RBCs, etc...

A better way to estimate the size of the uncertainty in single readings is suggested when examining table 21. Many metabolites have multiple peaks identified and since these are merely different reports of the same metabolite's concentration any variance in the results gives an idea of the uncertainty in calculations of metabolite concentration from the NMR results. By calculating the standard deviations from the two — or three for phenylalanine — integrals for each of these metabolites and expressing this as a percentage of the average concentration for the metabolite in that spectrum we produce the uncertainty estimates in table 24.

	Valine	Isoleucine	Tyrosine	Phenylalanine	Glutamine	Lactate*
Average standard deviation as a percentage of concentration	7%	10%	5%	17%	10%	24%

Table 24 : Average standard deviation as a percentage of metabolite concentration for the six metabolites with multiple measurable peaks. *Lactate measurements at $t = 0$ were ignored as the standard deviation of the two measurements at 4.12(q) and 1.33(d) was extremely high as a percentage of the mean concentration at $t = 0$.

Whilst this method is only possible for the few metabolites where multiple peaks are available for measurement it makes more sense to me as a way to quantify the technical error in the NMR method of measuring metabolite concentrations.

The errors derived from this technique are used later in the numerical analysis of the "Tom Repeat of Jenny Blood Null" series where an error in the straight line fit cannot be returned from the linear regression as there are only two points. The error is instead calculated by assuming an error of 10% in both directions for both points and propagating this error to the calculated gradient. This assumed error is the average of table 24 if lactate is ignored and is in keeping with other studies of HEPES-normalised quantitative NMR¹³⁰ for isolated peaks.

We are now ready to look at the next four graphs (figure 76 — figure 79) which show how the concentrations of carbon compounds change within the growth medium over the course of an approximately 40 hour experiment. In all cases the cultures are synchronised and the $t = 0$ point of the experiment is just after the merozoites have infected their host cells. The end of the experiments is just before the first late-stage schizonts start to lyse and exit their host cells.

Carbon compound results

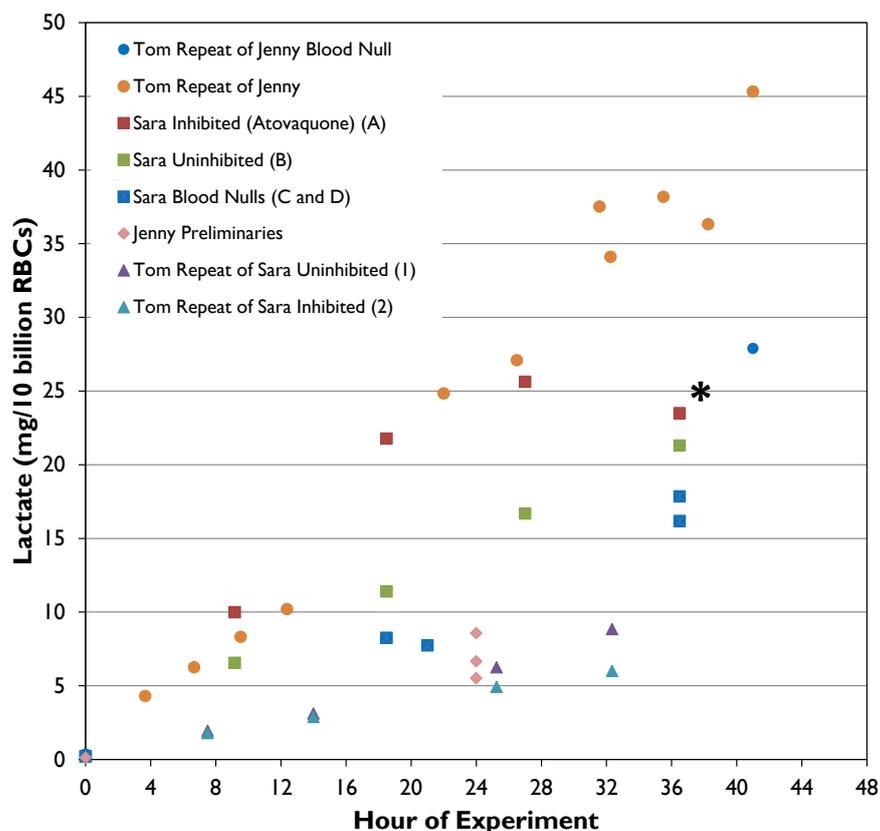


Figure 76 : The concentration of lactate in growth medium over time. The starred red point was not used during numerical analysis and does not suggest reabsorption of lactate from the growth medium*.

* see page 112 for a complete explanation of the origin of this misleading point and why it is omitted from further analysis.

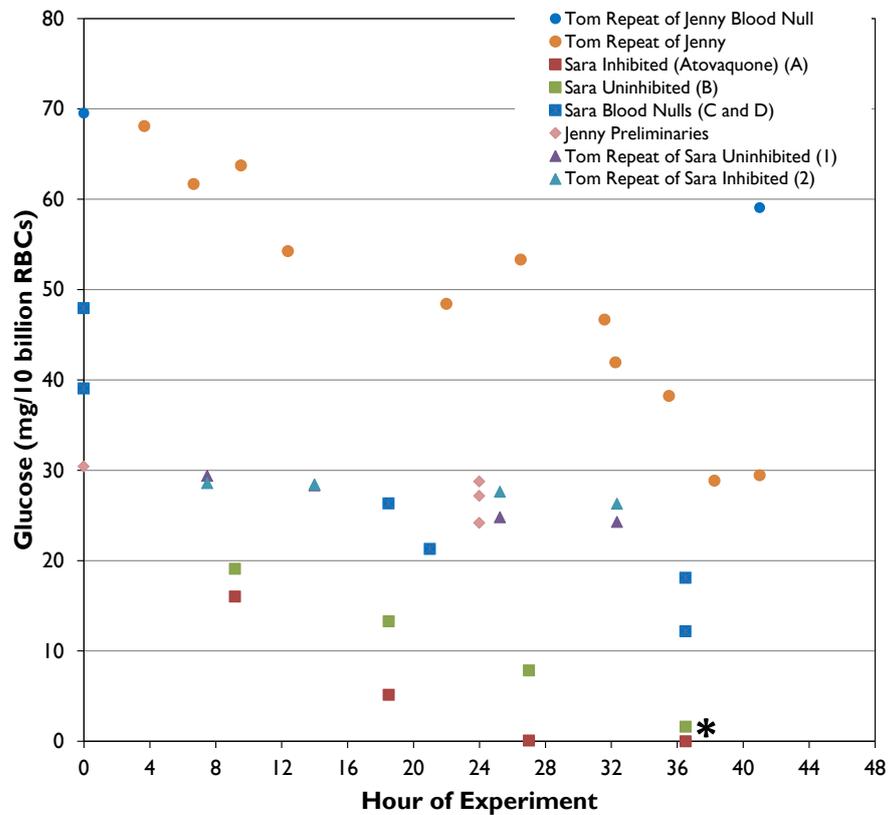


Figure 77 : The concentration of glucose ($\alpha + \beta$) in growth medium over time. The starred red point was not used during numerical analysis. Points at $t=0$ are only reported for blood nulls as discussed later.

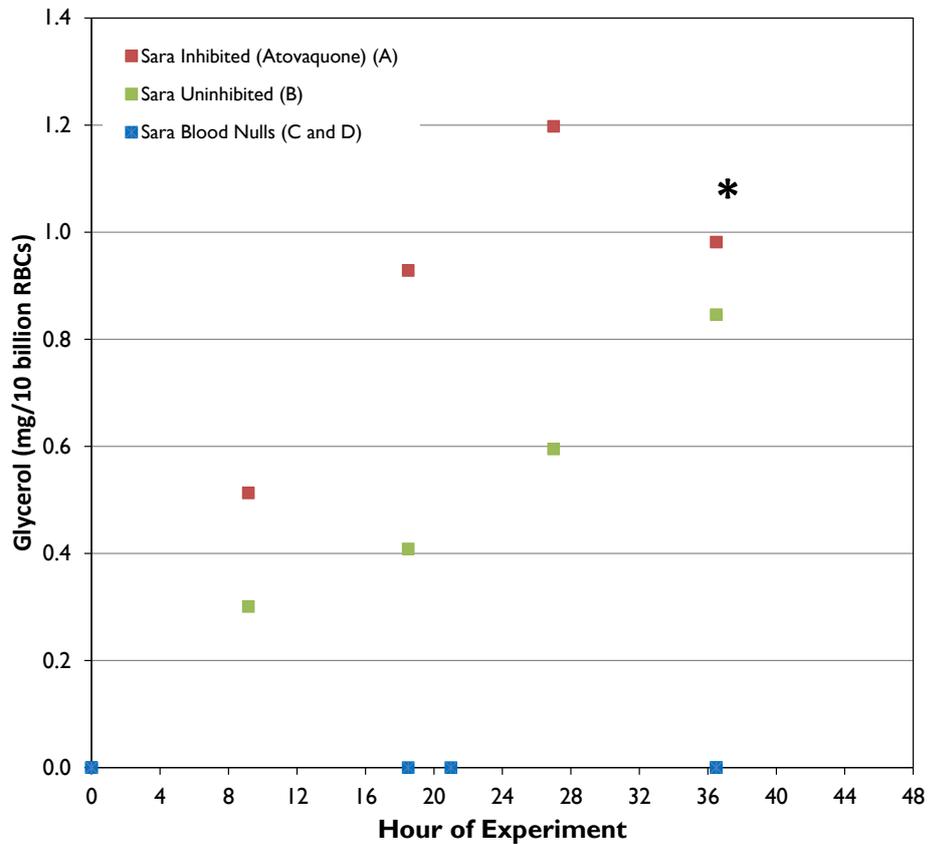


Figure 78 : The concentration of glycerol in growth medium over time. The starred red point was not used during numerical analysis and does not suggest reabsorption of glycerol from the growth medium.

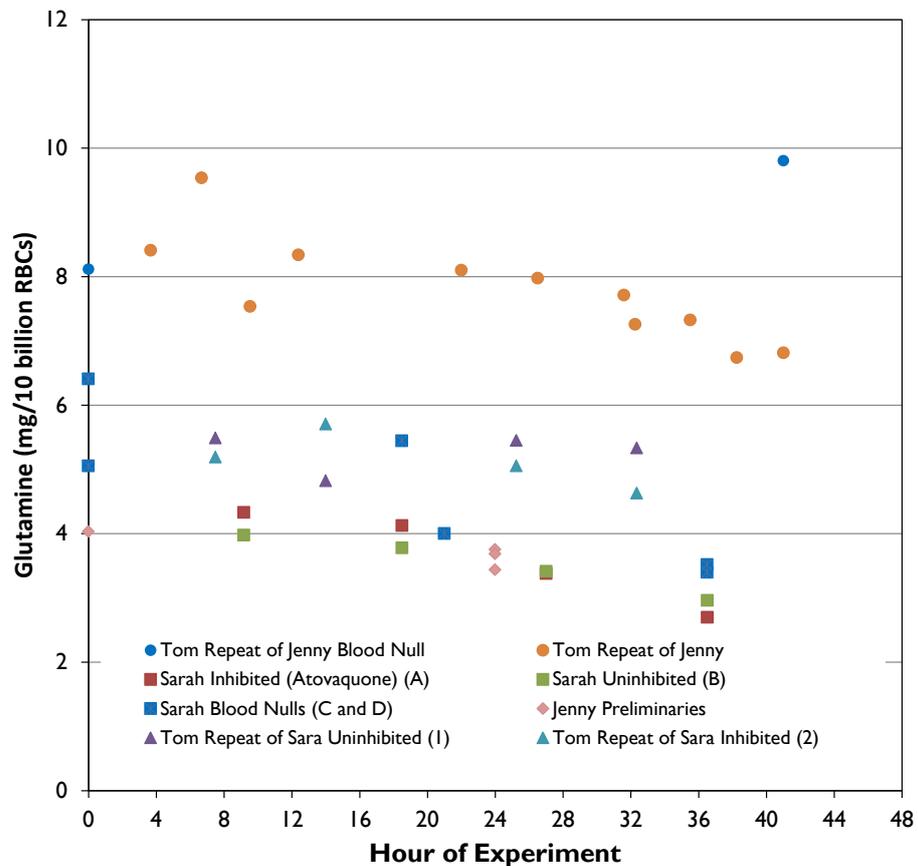


Figure 79 : The concentration of glutamine in growth medium over time. Points at $t=0$ are only reported for blood nulls as discussed later.

We can make some important observations from these graphs which I list in the same order as the graphs.

- Lactate is not detectable in fresh growth medium (at $t = 0$) and is produced at a constant rate in all experiments, including by blood nulls.
- The rate of lactate production varies considerably across different experiments with blood nulls (uninfected RBCs) in one experiment producing more lactate than infected RBCs in other experiments.
- In experiments with paired blood null series we see that infected RBCs produce more lactate than blood nulls.
- Glucose is consumed at a constant rate which is similar to the rate at which lactate is produced.
- The rate of glucose consumption varies considerably across different experiments.
- Glucose is consumed more quickly by infected RBCs than uninfected RBCs.
- Glycerol is not detectable in fresh growth medium (at $t = 0$) and is produced at a constant rate by infected RBCs. It is not produced by uninfected RBCs.
- Glycerol production was only detected in the highest sensitivity experiment and even this assignment is not without controversy.

- Glycerol production is at a rate about a tenth of the rate of lactate production.
- Glutamine appears to be consumed at a constant rate although this rate is considerably slower ($\approx 20 - 40$ times) than glucose.
- The observed production of glutamine by the blood null series in the “Tom repeat of Jenny Blood Null” experiment seems unlikely and is an example of why multiple measurements of blood null series are required.

Numerical analysis of carbon compound exchange

Whilst looking at the graphs of how carbon source concentrations in the growth medium change over time gives us a good feeling for what is happening, a more rigorous numerical analysis is required to extract numbers that we can use with our metabolic model as constraints for flux-balance analysis. Since the blood nulls show significant and variable levels of metabolism we can only make calculations of parasite metabolism for those experiments where we have a parallel blood null. In these case we calculate for each metabolite a least-squares linear regression using the LINEST function which returns a gradient (m_m) and an associated error (e_m) in that gradient. A gradient (m_{RBC}) and an associated error (e_{RBC}) for the blood null was similarly calculated.

The blood null gradient is subtracted from metabolite gradient to give the gradient caused by infected RBCs (m_p) whilst being careful to remember that parasites replace, in the same proportion as the culture’s parasitemia (p), rather than add to the number of uninfected RBCs.

$$m_p = m_m - ((1 - p) \cdot m_{RBC})$$

The uncertainties combine to give e_p ,

$$e_p = \sqrt{e_m^2 + e_{RBC}^2}$$

The gradient m_p is the result of the small percentage (p) of the RBCs that are infected and so to calculate the exchange flux in units of mg/10 billion infected RBC/hour or more sensibly 10^{-10} mg/h per parasite (F_p)

$$F_p = \frac{m_p}{p} = \frac{m_m - ((1 - p) \cdot m_{RBC})}{p}$$

The results of this analysis for the three series with associated blood nulls is summarised table 37, table 38 and table 39 in appendix V but shown most easily in figure 80 below.

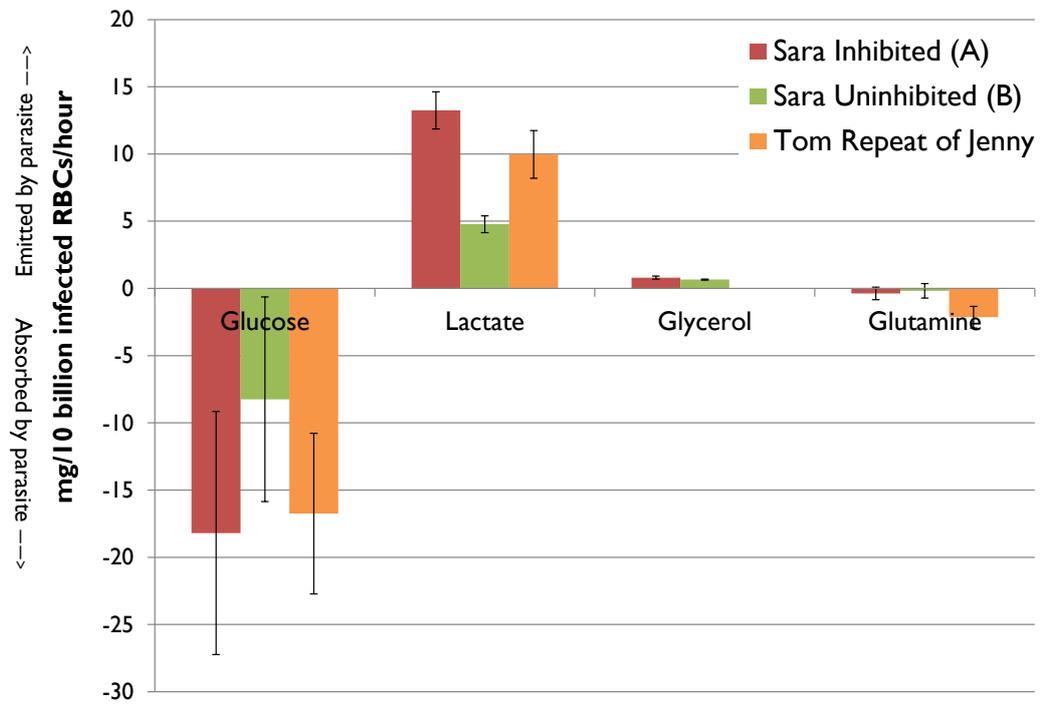


Figure 80 : Exchange fluxes of carbon compounds. Differences between the glycerol excretion and glutamine absorption rates in Sara A and Sara B sets are not significant. No glycerol was measured in the "Tom Repeat of Jenny" experiment.

Amino acids and hypoxanthine results

A key complexity of working with malaria parasites is the interplay between RBC host and the parasite itself. This is important in many areas but nowhere as strongly as with regard to amino acid metabolism since the parasite does not synthesise amino acids itself⁴⁷ but rather obtains them from haemoglobin digestion or from the host cell. Measuring the content of the growth medium tells us the exchange of metabolites between the infected RBC (including the internal parasite) and the extracellular space. Putting aside for a moment the significant changes to RBC metabolism provoked by parasite invasion we can safely subtract the background effect of the RBC's metabolism for carbon compounds. This is because the quantities of glucose, lactate and glycerol contained within the RBC at the moment of parasite infection are insignificant compared with the amounts exchanged with the growth medium over the subsequent 40 hours of parasite growth. The same cannot be said for amino acids because — as mentioned when considering the biomass function measurements earlier — around 92% of an RBC's dry weight¹⁷² at the moment of parasite infection is protein: almost exclusively haemoglobin.

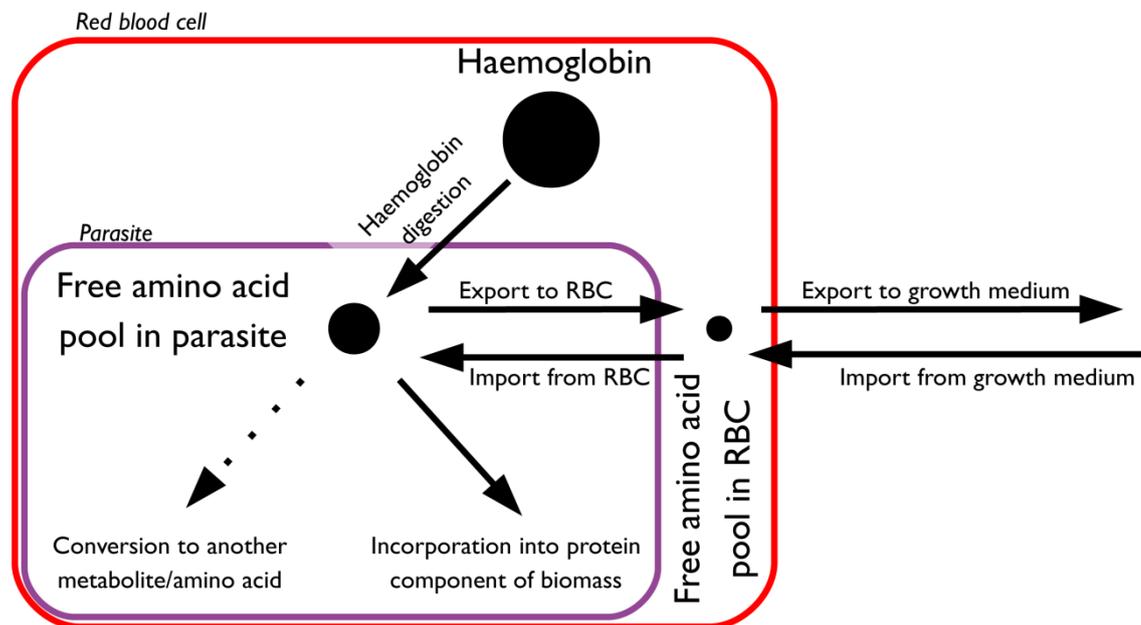


Figure 81 : Amino acid sources and fates in a cultured infected RBC are more complex than with carbon sources due to the abundant proteins provided by haemoglobin digestion. It is unclear whether the free amino acid pool in the RBC is large enough to affect measured amino acid exchange fluxes between the parasite and growth medium significantly.

In the case of amino acids we only measure the exchanges between the infected RBCs and the extra-cellular space even though only the exchange of isoleucine — an essential amino acid not present in haemoglobin — is required for growth*. Furthermore, the dynamics of amino acid exchange are already known to be extremely complex with the parasite known

* Parasites cannot grow without isoleucine and enter a state of hibernation until isoleucine is adequately provided¹²⁴.

to digest more haemoglobin than it needs for growth, export excess amino acids¹⁰³ and, bizarrely, to grow more slowly if certain amino acids that should be adequately supplied by the digestion of haemoglobin are not present in the growth medium¹⁰⁰. Finally, because the fluxes of amino acids are smaller than for the carbon compounds the relative size of the internal pool within the RBC but not within the parasite may no longer be insignificant as a buffer between a compound entering/leaving the parasite and entering/leaving the growth-medium.

These issues, displayed in figure 81, make the meaning of measured amino acid exchange fluxes between the infected RBC and the growth medium more challenging to interpret than with carbon compounds. The data from which these exchange fluxes are calculated is shown on this and the next three pages as figure 82 — figure 88.

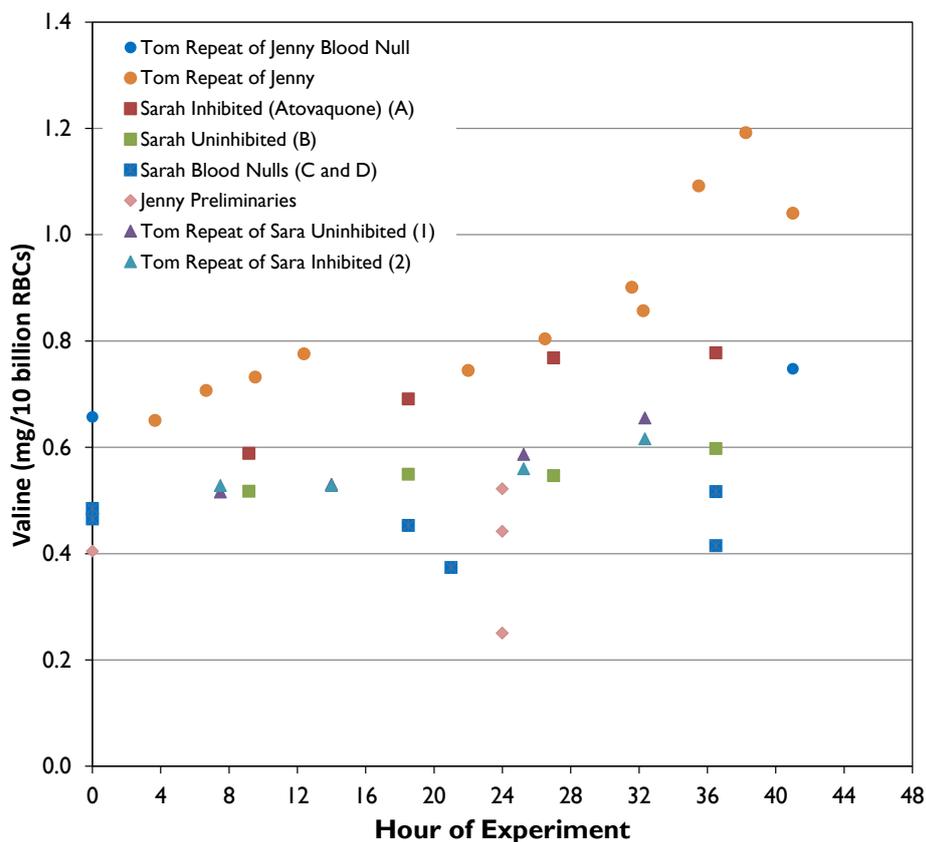


Figure 82 : The concentration of valine in the growth medium over time.

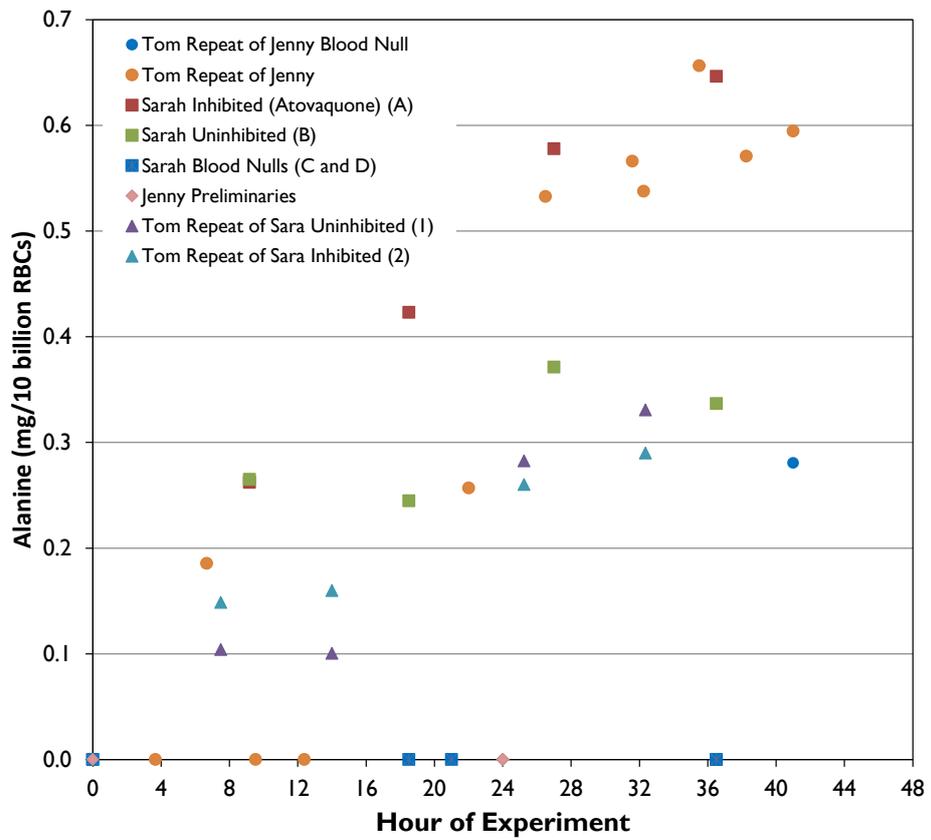


Figure 83 : The concentration of alanine in the growth medium over time.

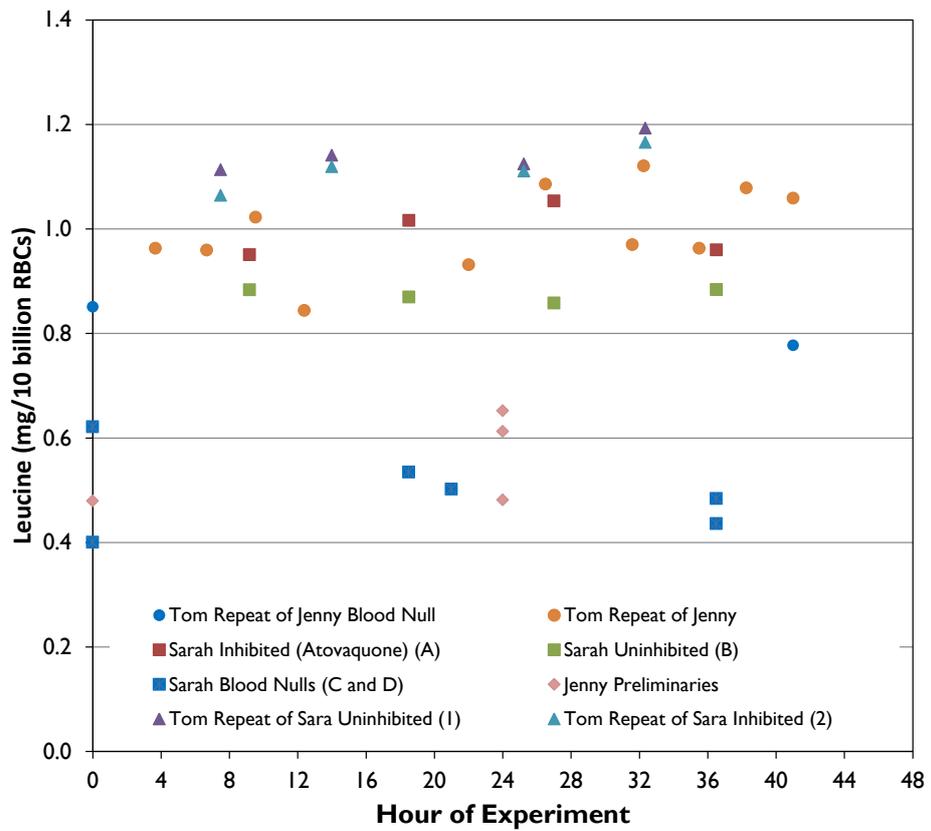


Figure 84 : The concentration of leucine in the growth medium over time.

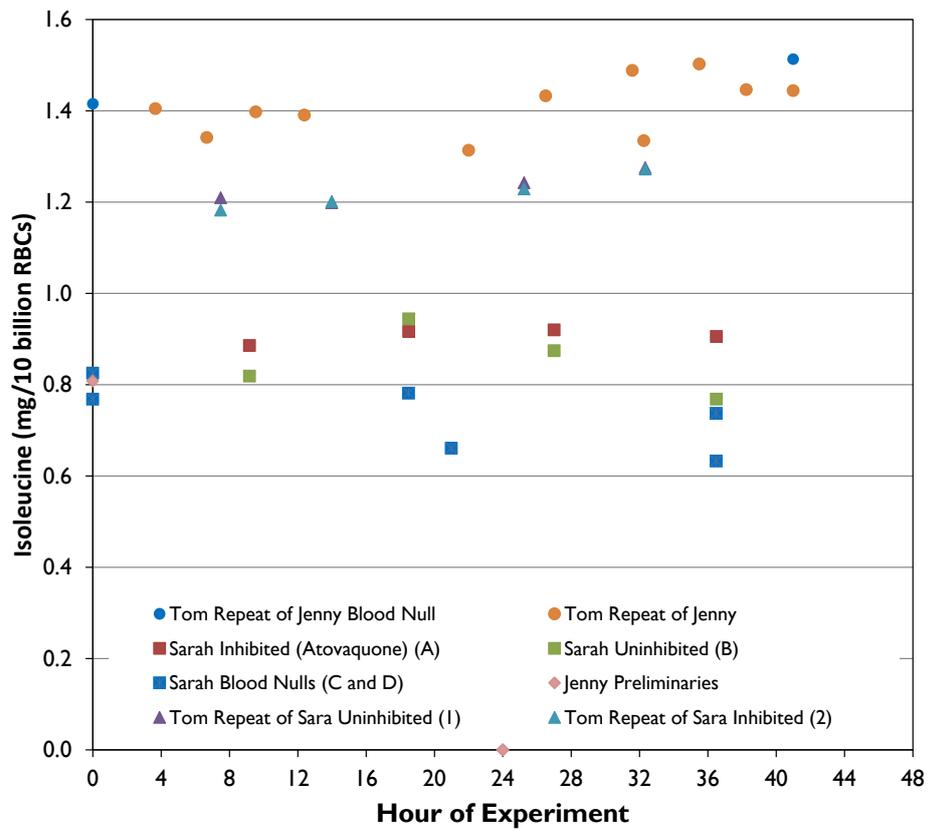


Figure 85 : The concentration of isoleucine in the growth medium over time.

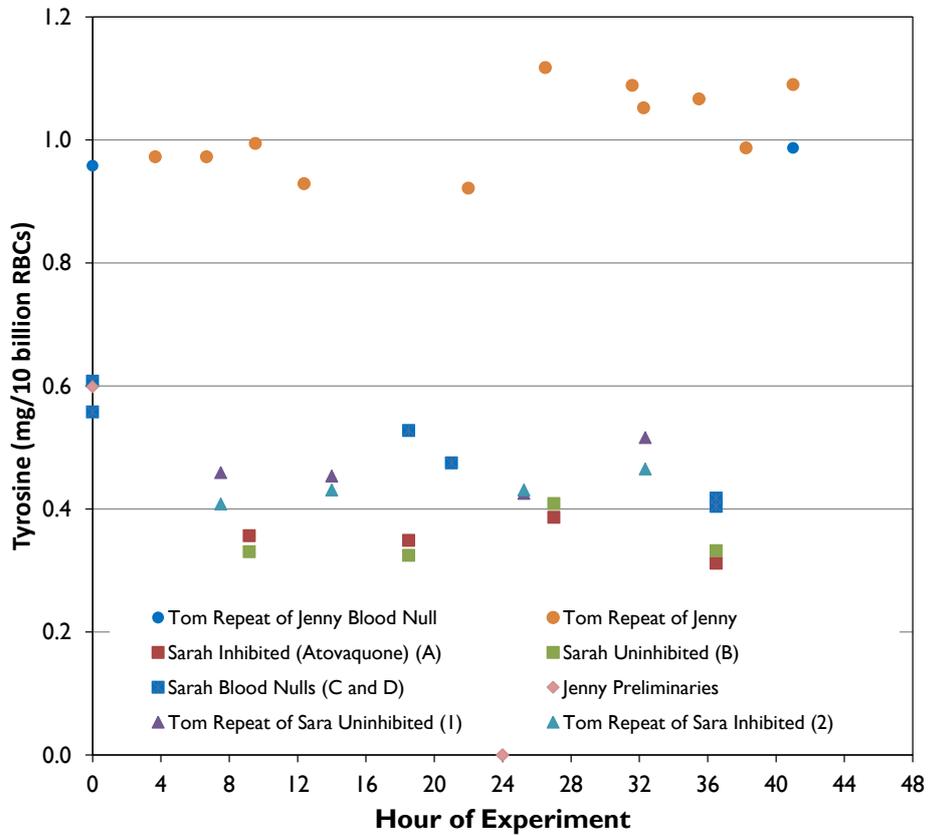


Figure 86 : The concentration of tyrosine in the growth medium over time.

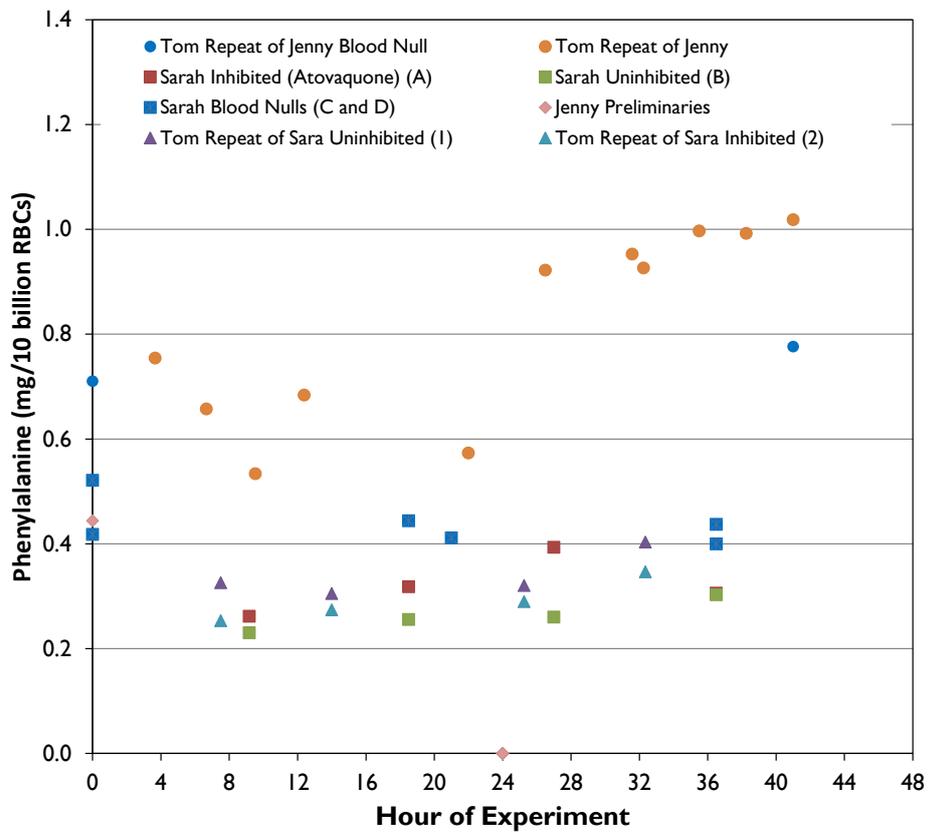


Figure 87 : The concentration of phenylalanine in the growth medium over time.

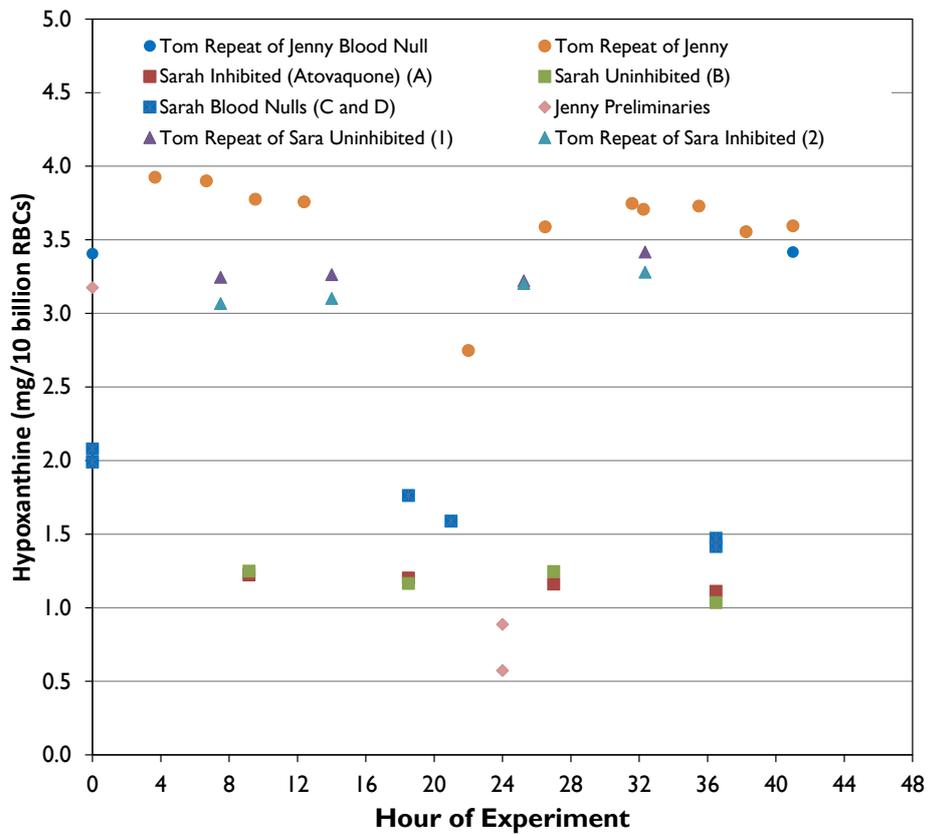


Figure 88 : The concentration of hypoxanthine in the growth medium over time.

As with the carbon compounds it is useful to make some general observations from the graphs.

- Valine, alanine and phenylalanine are clearly exported from the infected RBC into the growth medium.
- Valine seems to be exported more quickly in the second half of the life cycle.
- Alanine may be exported more quickly in the second half of the life cycle but the low concentration of alanine at early time points means it cannot be reliably detected by NMR metabolomics.
- Leucine seems to be exported into the growth medium but it is not clear.
- Isoleucine and tyrosine seem to be exchanged at the same rate between RBCs and the growth medium in both parasite-infected cultures and the uninfected blood null series.
- The considerably lower exchange fluxes measured for the amino acids and hypoxanthine make the accurate quantification of the fluxes more difficult.
- The considerable variability of the two blood null sets makes the interpretation of results difficult and increases the uncertainty in my findings.

A more complex but extremely important point to make is that the series of infected cultures when traced back to $t = 0$ frequently do not seem to start at the same concentration as the blood nulls. This is not something we see in the carbon source graphs and is a particular problem for the Sara metabolomics experiments because measurements were only taken at five time points.

I haven't been able to uncover the origins of this discrepancy but the possible interpretation that in the first time point a large exchange of amino acids occurs between infected RBCs and the growth medium that does not occur between uninfected RBCs and the growth medium is not plausible. Without a good suggestion for why this incongruity might be occurring I am forced to deal with it as well as possible. In this case I have done so by ignoring the $t = 0$ points when calculating the gradients of the infected RBC lines of best fit. In the Sara metabolomics set of experiments this makes a considerable difference. In the case of the "Tom Repeat of Jenny" metabolomics set of experiments this makes almost no difference as a sufficient number of time points are sampled to make a single point relatively unimportant.

Numerical analysis of amino acid exchange

With the exception of this change the numerical analysis of the amino acid exchange fluxes is identical and produces table 40, table 41 and table 42 in the appendix V and as figure 89 below.

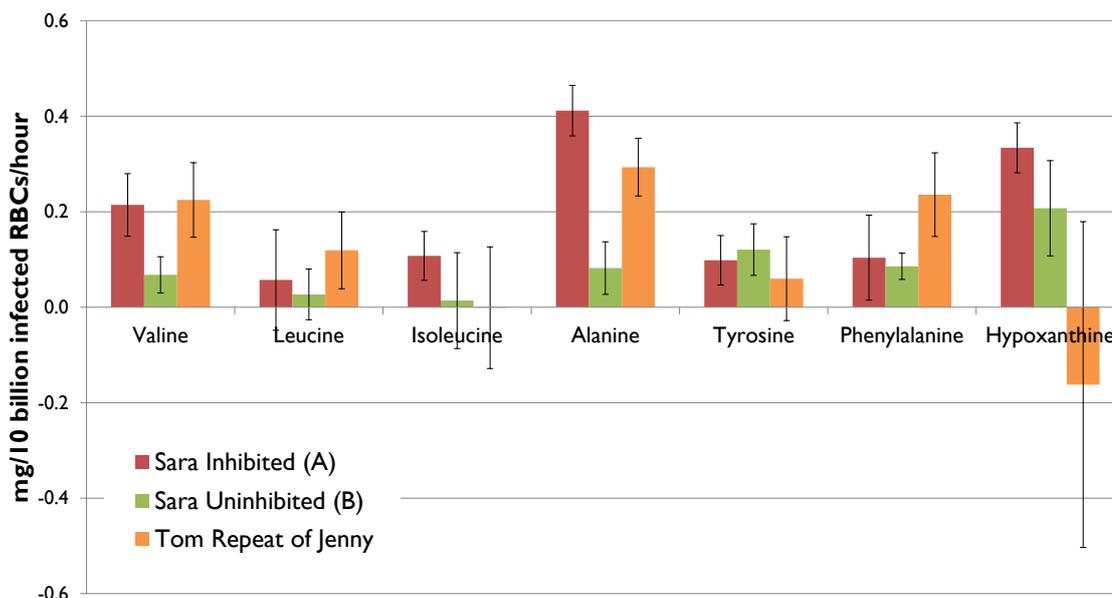


Figure 89 : Amino acid and hypoxanthine exchange fluxes from infected RBCs. Positive values indicate the metabolite is released into the growth medium, negative values indicate that it is absorbed.

Carbon balance

Examining figure 80 showing carbon compound exchange fluxes we see that, within error, carbon compounds are balanced within the system. The phrase “within error” is very important because in each case the glucose consumption is higher than the lactate production and when averages of the three experiments are taken the uncertainties reduce such that the difference becomes significant. This incomplete conversion of glucose to lactate has been observed before with Krugliak *et al.*¹⁰³ stating that,

Most of the glucose consumed (60–70%) by Plasmodium falciparum is incompletely oxidized to lactic acid and excreted although the exact percentage varies between different Plasmodium species and the atmospheric culture conditions used. This glucose consumption contrasts with the >90% glucose-to-lactate conversion observed in uninfected RBCs and reflects the increased flux of glucose carbon into biomass (nucleic acids, lipids, glycosylated proteins) required for proliferating parasites.

The first part of this agrees with my results as summarised in table 25 with a measured glucose flux to lactate flux mass ratio* of between 1.38:1.00 (72% conversion) and 1.71:1.00 (58% conversion). My results agree far less well with Krugliak *et al.*'s¹⁰³ second finding, with the only reliable blood null experiment I conducted showing a glucose flux to lactate flux mass ratio for uninfected RBCs of 1.76:1.00 (56% conversion).

Sara A	Glucose	Lactate	Ratio
uninfected RBCs	-0.79	0.45	1.76
infected RBCs	-18.2	13.2	1.38
Ratio	23.0	29.3	

Sara B	Glucose	Lactate	Ratio
uninfected RBCs	-0.79	0.45	1.76
infected RBCs	-8.2	4.8	1.71
Ratio	10.4	10.7	

Tom repeat of Jenny	Glucose	Lactate	Ratio
uninfected RBCs*	-0.26	0.68	0.38
infected RBCs	-16.7	10.0	1.68
Ratio	64.4	14.7	

Table 25 : Ratios of glucose uptake and lactate excretion (10^{-10} mg/ml) from uninfected RBC and infected RBC populations in the three experiments with blood nulls. Results from the "Tom Repeat of Jenny" experiment should be treated with caution as the blood null line of best fit was calculated from only two points.

My results also do not completely agree with the suggestion that the origin of the incomplete measured conversion of glucose to lactate is extra biomass formation from carbon compounds. The mean carbon accumulation I have measured in parasites across all three experiments is 5.48×10^{-10} mg/parasite/hour so over 40 hours the unaccounted mass of carbon compounds is 220×10^{-10} mg/parasite. This is twice the average parasite dry weight of $105 \pm 13 \times 10^{-10}$ mg/parasite showing that biomass formation cannot fully explain the imbalance.

It is known that *P. falciparum* can produce acetate and experiments with other species of *Plasmodium* have measured the production of formate and other volatile carbon compounds^{183, 184} though there are reasons discussed in Krugliak *et al.* to think that this may not occur in *P. falciparum* cultured in more modern conditions. It is known that D-lactate is additionally produced at a rate of around 5% that of L-lactate and this is not measured. It is important to note that precise carbon balance is not achieved at the most best estimate for the carbon compound fluxes but that carbon balance is achieved well within the uncertainties in those measurements.

* Glucose to lactate mass ratio of 2:1 (180.16:90.08) is exactly equivalent to the carbon mass ratio of 2:1 (6:3).

In summary, my measured carbon compound fluxes achieve carbon balance but not at their most likely rates. It is likely that a combination of biomass formation, retention of lactate and glycerol within the parasite and the RBC, the production of carbon compounds that I have not measured, and most importantly uncertainties in my measured fluxes can explain what might seem to be a carbon imbalance. My metabolic model confirms this,

The absence of malate

I cannot measure some metabolites which I know to present in the medium at visible concentrations because they are obstructed by other peaks. A good example of this is L-arginine whose principal peaks at 3.23(t) and 3.76(t) are both overwhelmed by glucose peaks at the same position.

This does not seem to be the case for malate, which is important because one of the most interesting results in malaria metabolism since I started this project was published in Nature in August 2010 by Olsewski *et al.*⁹⁵. One of the many findings — figure 3 in their manuscript — shows that *P. falciparum* parasites produce and export malate to the growth medium using a highly adapted version of the TCA cycle fed with glutamine. In the paper, extracellular concentrations of malate reach a concentration of about 60 μ M 40 hours after synchronisation. This compares with the 90 μ M concentration of L-phenylalanine and the 110 μ M concentration of L-tyrosine that NMR analysis of my growth medium can detect, giving me hope that we too would measure malate production. The possibility that we might be able to detect malate and confirm this result seems even more likely when we examine the experimental technique in the paper in more detail.

Olszewski *et al.*'s⁹⁵ experiments continued to hour 40 (t_{ols}) and were conducted at an unusually low haematocrit (h_{ols}) of 0.4% and at 6% parasitemia (p_{ols}). My equivalent experiments named “Tom repeat of Jenny Metabolomics” were conducted at a final measured haematocrit concentration of 3.1% (h_{TF}) and a parasitemia of 4.1% (p_{TF}). From this we can calculate the expected final concentration of malate in my used growth medium, $[M]_{TF}$ from the concentration in Olszewski *et al.* $[M]_{ols}$.

$$[M]_{TF} = [M]_{ols} \times \frac{t_{TF}}{t_{ols}} \times \frac{h_{TF}}{h_{ols}} \times \frac{p_{TF}}{p_{ols}}$$

$$[M]_{TF} = 60 \mu\text{M} \times \frac{41}{40} \times \frac{3.1}{0.4} \times \frac{4.1}{6} = 217 \mu\text{M} \approx 330 \mu\text{M}$$

From this predicted concentration we can calculate the expected integral value for each of malate's three single proton (nH = 1) peaks with reference to a single (nH = 2) HEPES peak defined as having an integral of 100.

$$I_{malate} = I_{HEPES} \times \frac{nH_{malate}}{nH_{HEPES}} \times \frac{[M]}{[HEPES]}$$

$$I_{malate} = 100 \times \frac{1}{2} \times \frac{0.330}{25.03} = 0.66$$

The set of experiments named “Sara Metabolomics” which were conducted to a final time of 36.5 hours at an average haematocrit of 7.5% and a parasitemia of 4.5% giving an equivalent predicted concentration of 770 μ M and a predicted integral of 1.53. For comparison, integrals for phenylalanine and tyrosine are registered in all samples with magnitudes typically between 0.20 and 0.40 so we would expect to see malate signals at this predicted intensity.

Before we can try to find malate we need to know what we’re looking for. The principal multiplet peaks of the malate NMR spectrum, as shown in figure 90, are summarised in the human metabolome database (HMDB)¹⁷⁹ in the following order of maximum relative intensity,

- A double-doublet produced by a single proton with a maximum relative intensity of 1.00 at 2.66ppm referenced to TSP at 0.00ppm.
- A double-doublet produced by a single proton with a maximum relative intensity of 0.91 at 2.36ppm referenced to TSP at 0.00ppm.
- A double-doublet produced by a single proton with a maximum relative intensity of 0.25 at 4.29ppm referenced to TSP at 0.00ppm.

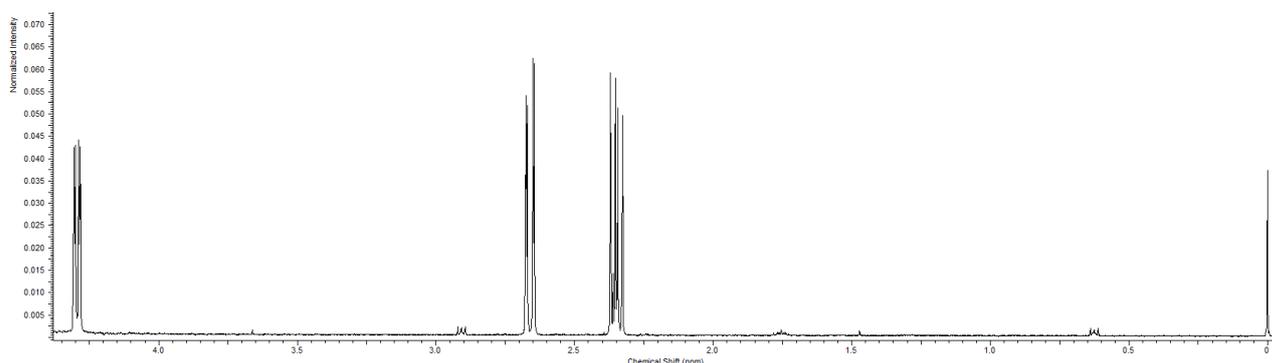


Figure 90 : The three peaks of the experimental ¹H-NMR spectrum of malate from HMDB¹⁷⁹. There are no significant peaks outside of this range. The peak to the right is TSP at a chemical shift of zero.

Of the three malate peaks the one at 2.66ppm is close to the HEPES signal at 2.72ppm making it difficult to identify and the one at 2.36ppm is obscured by an unreliable, but present, glutamate multiplet at the same location. The peak at 4.29ppm is flanked by two larger unidentified signals making the assignment unclear. There is a small signal at 4.29ppm in the Sara metabolomics set of experiments but if this is malate then it is not visible in the “Tom Repeat of Jenny” set of experiment and its integral of 0.07 (35 μ M,

0.005mg/ml) in the final time point is around one twentieth of the production measured in Olszewski *et al.*⁹⁵. Furthermore, the shape of the peak at 4.29ppm is not convincingly malate.

A clue to explaining this discrepancy between our results is given in another paper by the same group, in which they suggest why malate may be exported,

“in cell culture, aspartate levels are relatively high and this flux [from oxaloacetate to aspartate] may be redirected to malate, which is then excreted due to overflow metabolism”

Central carbon metabolism of Plasmodium parasites, Olszewski *et al.*⁸⁸

To consider this idea we need to look at three things we know about aspartate,

1. It is present at 20mg/L (0.15mM) in RPMI medium (see appendix III). I cannot measure changes in its concentration with my NMR technique since the three double doublets at 2.66, 2.80 and 3.89ppm are obscured by HEPES, HEPES and glucose respectively.
2. It is present in haemoglobin at about the same level as phenylalanine and makes up a similar proportion of the proteome composition. We know that phenylalanine is exported from the parasite due to excess haemoglobin digestion but we need to remember that phenylalanine does not have the additional roles of aspartate in metabolism so aspartate is unlikely to be in excess simply due to haemoglobin digestion.
3. It is produced instead of lactate by a branch of glycolysis at phosphoenolpyruvate and additionally, according to Olszewski *et al.*⁹⁵, from glutamine in the mitochondrial TCA cycle*. Furthermore, it is consumed by both nucleotide and amino acid synthesis. We do not know what percentage of the aspartate consumed comes from each possible source but it seems unlikely that the requirement is exceeded by haemoglobin digestion and the branched TCA cycle. If the glycolytic flux were channelled to aspartate production this would easily produce a large excess of the metabolite.

The system that Olszewski proposes as to why malate may be produced and exported is plausible but to me it seems odd. Malate is most likely produced in part because of an overflow from the main branch of glycolysis and yet lactate is clearly the main overflow metabolite. Channelling glycolysis to malate seems to offer no advantage over the

* The figure showing this in Olszewski *et al.* is reproduced earlier in this thesis as Figure 33.

production of lactate and costs a molecule of ATP in comparison to lactate production and export.

I revisit the malate issue in the final chapter with respect to a fuller metabolic model which provides some answer, and yet more questions. Certainly we can say that we have not measured the same malate production and efflux as previously reported.

Sanity checks and discussion of results

Absolute metabolite concentration

The composition of RPMI 1640 is very accurately known — see table 35 in appendix III —, and our only additions are AlbuMAX I, sodium bicarbonate and hypoxanthine as described in the methods chapter. This provides an excellent benchmark with which to test the metabolite concentrations measured by our unusual HEPES-referenced NMR technique. We can also compare this technique with the results of Massimi *et al.* (sup. 1)¹⁸⁵ who used a similar quantitative NMR approach to measure the components of a growth medium consisting of RPMI 1640 medium (*Sigma*) supplemented with 10% foetal calf serum (FCS) and 2 mM (0.29mg/ml) L-glutamine. A final check of the measured concentrations is provided by the biochemical assays of fresh growth medium we performed and which are reported fully in Jennifer Lake's Masters dissertation¹⁴⁵. All these sources of information and the equivalent averages from all my NMR measurements of fresh growth medium are shown for the overlapping metabolites in table 26.

	RPMI 1640	Massimi <i>et al.</i>	My measured mean concentrations, fresh growth medium		Biochemical assays
	Concentration (mg/ml)	Concentration (mg/ml)	Concentration (mg/ml)	Standard Deviation	Concentration (mg/ml)
Isoleucine	0.050	0.10	0.042	0.004	-
Valine	0.020	0.04	0.022	0.006	-
Alanine	0.000	0.02	0.002	0.010	-
Glutamine	0.300	0.38	0.245	0.15	-
Glucose	2.000	1.03	1.853	1.44	2.33
Phenylalanine	0.015	0.05	0.020	0.017	-

Table 26 : Concentration of six metabolites in RPMI 1640 (manufacturer's definitions) and the equivalent measurements for fresh growth medium ($t=0$) by $^1\text{H-NMR}$ in Massimi *et al.*¹⁸⁵, in my NMR results and for glucose as reported in Jennifer Lake's Masters dissertation¹⁴⁵.

The higher amino acid concentrations in Massimi *et al.*¹⁸⁵ could be explained by their use of unprocessed FCS unlike my use of AlbuMAX which has the majority of free amino acids removed. but their observation of a far lower glucose and glutamine level than would be expected in RPMI supplemented with 0.29mg/ml L-glutamine is more puzzling and

suggests that my HEPES-referenced method is in many ways superior to their more complex techniques. My results vindicate the decision to assign the reference peak of HEPES to 2 hydrogens as the concentrations of the comparable metabolites are in good agreement with the composition of RPMI 1640 and the independent measurement of glucose levels by a biochemical assay. The underestimation of metabolite concentration at higher concentrations is in keeping with the results in Wishart¹⁸¹.

The composition of AlbuMAX I is unknown but from these results we can broadly suggest that it is principally albumen with a considerable amount of fatty acids and/or lipids and negligible free amino acids and carbohydrates.

Known fluxes and stage specificity

Lactate production by infected RBCs has been reported previously, with Pfaller *et al.*¹⁸⁶ measuring a mean lactate production rate of $800-1100 \times 10^{-8}$ nmol/RBC/hour* with a higher rate of $1100-2100 \times 10^{-8}$ nmol/RBC/hour at the schizont stage. Vander Jagt *et al.*⁹⁴ report higher but comparable L-lactate production rates† for ring stages of 1340×10^{-8} nmol/RBC/hour and 2050×10^{-8} nmol/RBC/hour for trophozoite/schizont stages.

My results seem reasonable in comparison to these previously reported results with the best estimate of lactate production in the range $530 - 1470 \times 10^{-8}$ nmol/RBC/hour. The slightly lower figure that I measure is likely due to the RBC population variability I discuss later and my observed slower growth-rate in AlbuMAX-based medium as opposed to the 10% serum medium used in Pfaller *et al.*¹⁸⁶ and presumably used in Vander Jagt *et al.*‡.

The stage specificity of lactate production in these two papers is presented — as far as I can see — without a statistical test and in the case of Vander Jagt with an experimental technique my results show to be unreliable; two different experiments were used to calculate early and mature stage lactate production rates meaning background RBC metabolism was different in each experiment. Of my experiments the “Tom Repeat of Jenny” set is the one where enough time points were sampled to make observing stage-specificity most likely. Taking time points before 17h to be early stage (ring) and those after 17h to be late stage (trophozoite/schizont) — these times are accurately derived on page 110 of this thesis — I calculated the stage-specific exchange fluxes for glucose, lactate

* The normal equivalent unit (Eq) used in the paper is rare and for lactate $1\text{nEq} = 1\text{nmol}$.

† D-lactate is produced at a rate about 5% of L-lactate. The D-lactate NMR peak at 1.40ppm (d) is not visible or obscured by the 1.33(d) L-lactate peak so I have not measured it.

‡ Culture details are in referenced paper to which I do not have access.

and valine as shown in figure 91. For all three metabolites the exchange flux is higher in the latter life stage but only in the case of valine is this statistically significant.

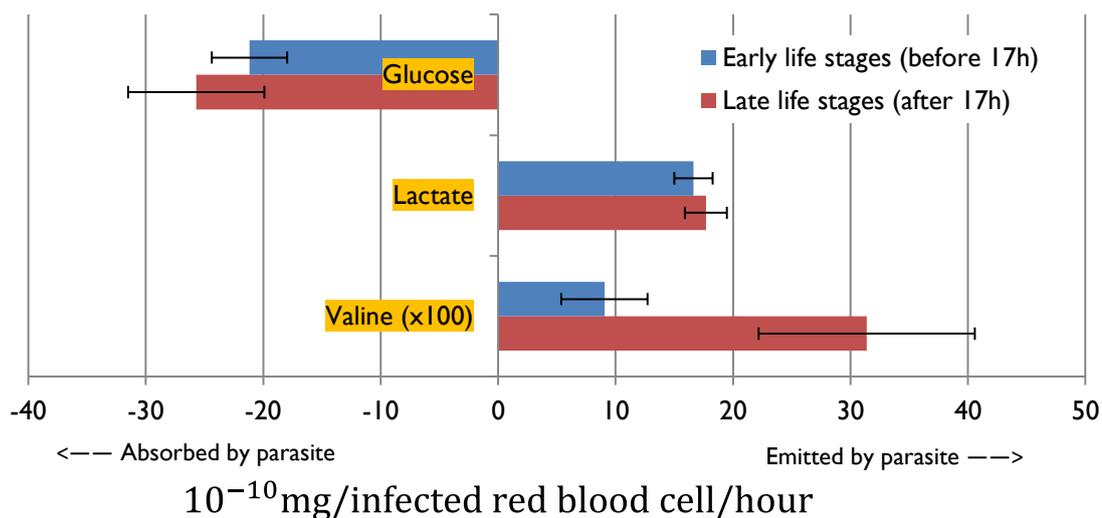


Figure 91 : Exchange fluxes of the three metabolites potentially showing stage-specificity at the ring stage (early) and the trophozoite/schizont stages (late). The valine exchange flux and associated uncertainty is shown at 100 times real scale to be visible.

It is worth noting that in figure 91 the exchange fluxes of glucose and lactate are higher when life stages are split than they are for the whole series. This is because the two lines of best fit are not forced to meet at $t = 17h$; the origins of this strange result can be seen by examining figure 76 and figure 77. Forcing the two lines of best fit to meet near $t = 17h$ does not change the results; the gap between early and late life stage exchange fluxes reducing slightly but the gap in valine emission remains statistically significant.

Glycerol production

The metabolism and exchange of glycerol by *P. falciparum* is an area of considerable disagreement. Lian *et al.*⁴⁶ measured production and export of glycerol by the infected RBC and Olszewski *et al.*¹⁰⁹ measured absorption and use of glycerol by the parasite. Absorption of glycerol would be impossible to measure with my NMR technique since the growth medium contains no NMR-measurable free glycerol but this does not make the findings of Olszewski *et al.* implausible. Firstly, their measurements are more sensitive — though less accurate — and secondly it is possible that some glycerol becomes available to the parasite due to some degradation of the glycerophospholipids likely to be present in the growth medium and the host RBC.

Nevertheless, if my NMR assignments to glycerol are correct then production and export of glycerol was seen clearly in the “Sara A” and “Sara B” metabolomics experiments with no measurable background production of glycerol by uninfected RBCs. The only doubt I have with these results is that the glycerol peaks are close to glucose peaks and I saw no

measurable production of glycerol by infected RBCs in the other less sensitive experiments*. Since this result is of such importance I have spent a considerable amount of time examining the NMR glycerol assignment which I justify here.

Figure 92 shows glycerol's two principal peaks and how the double-doublet at 3.56ppm overlaps slightly with a distinctive glucose signal. Apart from this interference, this glycerol peak is not affected by any other identified metabolites.

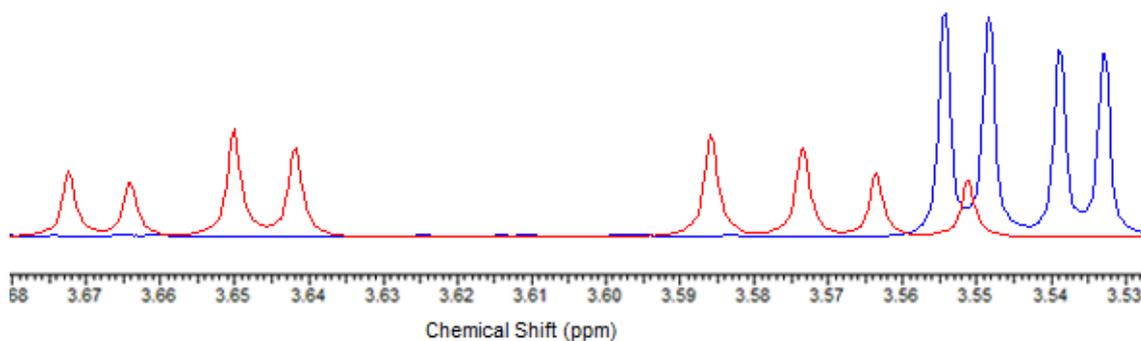


Figure 92 : Glycerol (red) and glucose (blue) peaks overlap slightly at 3.50ppm affecting the otherwise clear glycerol peak assigned to 3.56ppm. Spectra are from HMDB references and manually aligned with the TMS peak at 0ppm.

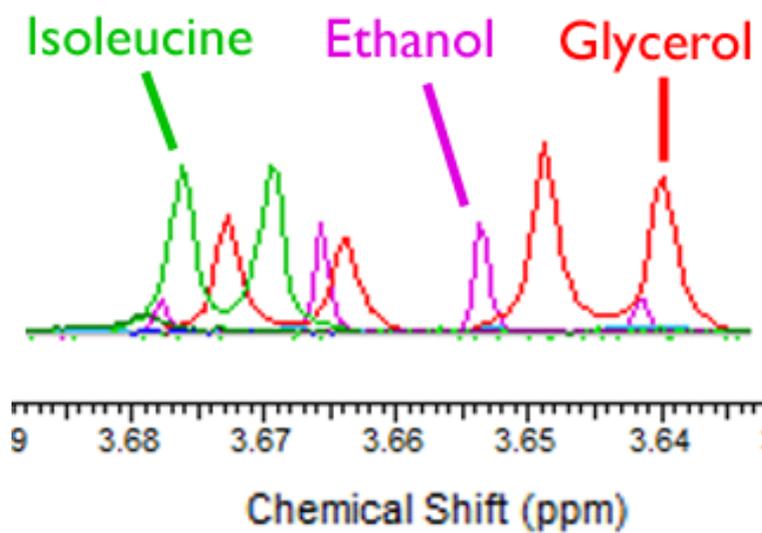


Figure 93 : Ethanol and isoleucine are both strong peaks that obscure the glycerol peak at 3.66ppm. Spectra are from HMDB references and manually aligned with the TMS peak at 0ppm.

As we can see in figure 93, the same cannot be said for the glycerol peaks (dd) at 3.66ppm. These peaks are obscured by isoleucine and ethanol, both known to exist in the growth medium at high concentrations and clearly seen in the experimental results shown in figure 94.

I am confident enough to make the glycerol assignment at 3.56ppm but it is clear that only the left two peaks of the double-doublet are isolated enough to be reliably measured. In

* The Sara metabolomics set of experiments were more sensitive because of the higher parasitemia, higher average haematocrit and because the spectra were acquired by the FBS NMR service with the improvement in sensitivity mentioned in the methods chapter on page 115.

the reference spectra for glycerol the integrals of the left pair of peaks and the right pair of peaks have integrals in the ratio 3:2 and so in all cases I have measured the left pair of peaks and multiplied by $\frac{5}{3}$ to estimate the true integral.

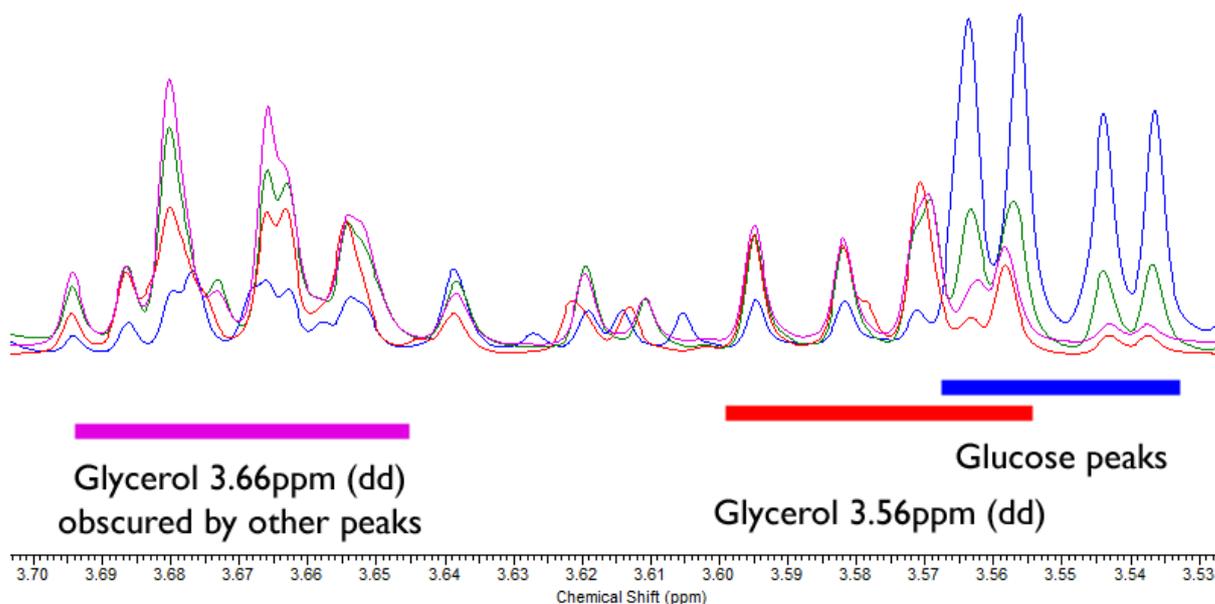


Figure 94 : Spectra for the "Sara A" set annotated to show the set of peaks I assign to glycerol. Colours blue, green, pink, red, corresponds to $t = 9.2, 18.5, 27.0, 36.5$ hours. Glucose concentration at the final time point is almost zero meaning the glycerol peak of the double doublet becomes visible. Due to the normalisation techniques used in ACD/labs the visual size of each peak is only a guide to its integral.

Amino acid fluxes

As mentioned previously, the amino acid fluxes cannot be understood without a substantial amount of background information. An essential tool for helping with the interpretation of these results is the amino acid composition of the haemoglobin protein as calculated from the amino acid sequences of haemoglobin subunit α (P69905) and haemoglobin subunit β (P68871) in UniProt¹⁸⁷ and shown for selected amino acids in table 27.

Name	Molecular Weight (g/mol)	Count	Weight per mole of haemoglobin (g)	Percent by Weight	Percent by amino acid count
Leucine	113	72	8142	13.0%	12.5%
Valine	99	62	6142	9.8%	10.7%
Alanine	71	72	5115	8.2%	12.5%
Phenylalanine	147	30	4412	7.1%	5.2%
Aspartate	115	30	3451	5.5%	5.2%
Tyrosine	163	12	1957	3.1%	2.1%
Glutamine	128	8	1024	1.6%	1.4%
Isoleucine	113	0	0	0.0%	0.0%

Table 27: Amino acid composition of 2 α -subunits + 2 β -subunits of the human haemoglobin molecule ordered by total weight per mole of haemoglobin for selected amino acids of interest, The full table is reproduced in Appendix VI as table 43.

The composition of haemoglobin gives us the ratios of amino acids that the parasite acquires through haemoglobin digestion. The next important quantity is the ratio of amino acids that the parasite uses to create its biomass; already extracted from Chanda *et al.*¹⁷⁶ as table 16 on page 153. Here, the same table for the selected amino acids of interest is reproduced as table 28.

Amino Acid	Percent by weight	Percent by amino acid count
Isoleucine	8.5%	9.3%
Glutamate	8.3%	7.0%
Leucine	7.4%	8.1%
Tyrosine	7.3%	5.5%
Aspartate	6.5%	6.0%
Phenylalanine	5.5%	4.6%
Valine	3.3%	4.1%
Glutamine	2.9%	2.8%
Alanine	1.4%	2.4%

Table 28 : Amino acid use in the *P. falciparum* proteome, adapted from Chanda *et al.*¹⁷⁶ for selected amino acids of interest, The full table is reproduced in Appendix VI as table 44.

We've measured the exchange fluxes between the infected RBC and the growth medium and so referring back to figure 81 outlining the sources and fates of amino acids we see

that we now have the three key fluxes in the system. If we temporarily ignore the pool of free amino acids in the RBC*, and the possibility of conversion between amino acids, we can calculate some very interesting conclusions from our results.

Figure 95 shows how these three factors balance when it is assumed that 40% — the amount required to achieve an average net balance of the six amino acids to the left of the dotted line — of the haemoglobin in the host RBC is digested by the parasite.

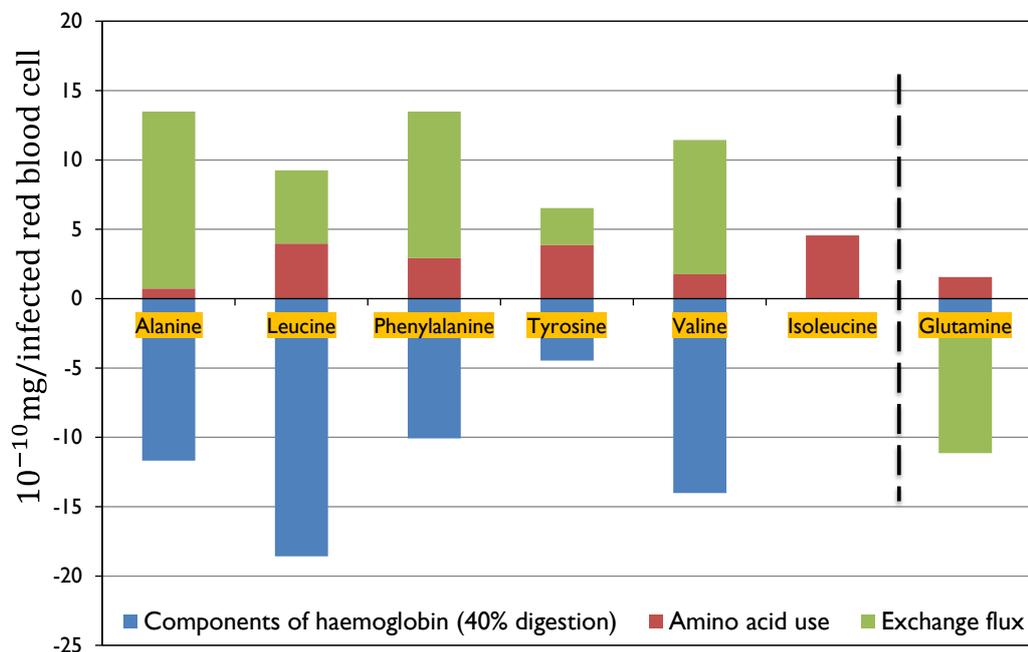


Figure 95 : Stacked bar charts showing the fluxes of amino acids to (negative) and from (positive) the pool of free amino acids (shown in figure 81) for the "Tom Repeat of Jenny" experiment. The blue bars represent the quantity of each amino acid freed from haemoglobin digestion (assuming 40% of the host cell's haemoglobin is digested). Red bars represent the quantity of each amino acid used for protein synthesis by the parasite. Green bars represent the measured fluxes of amino acids out of the parasite (into the parasite in the case of glutamine). The total mass of amino acids is conserved in this figure.

We can see clearly in this figure that glutamine is used in a way quite unlike the other amino acids, confirming that it is better considered as a carbon source. We can also clearly see the problem we face in explaining why no isoleucine uptake flux from the growth medium was measured. A possible explanation for the problem with the isoleucine result is suggested by Elford *et al.*¹⁸⁸ who report that isoleucine is taken up equally by uninfected and infected red blood cells. More recent work at higher precision than my method and that employed by Elford *et al.*¹⁸⁸ has measured an increase in isoleucine uptake by infected red blood cells¹⁸⁹.

A drawback of this amino acid balancing approach is that the solution does not provide enough tyrosine, alanine or phenylalanine to explain the amounts incorporated into proteins within the organism and exported by the infected RBC. Figure 96 shows the result

* There is some evidence¹⁰³ that infected RBCs are made extremely permeable by the parasite and thus unable to retain a significant free amino acid pool.

of a different calculation, ensuring that the amino acids tyrosine and phenylalanine are supplied from haemoglobin digestion in adequate amounts. This calculation suggests that at least 55% of the host RBC's haemoglobin is digested by the parasite.

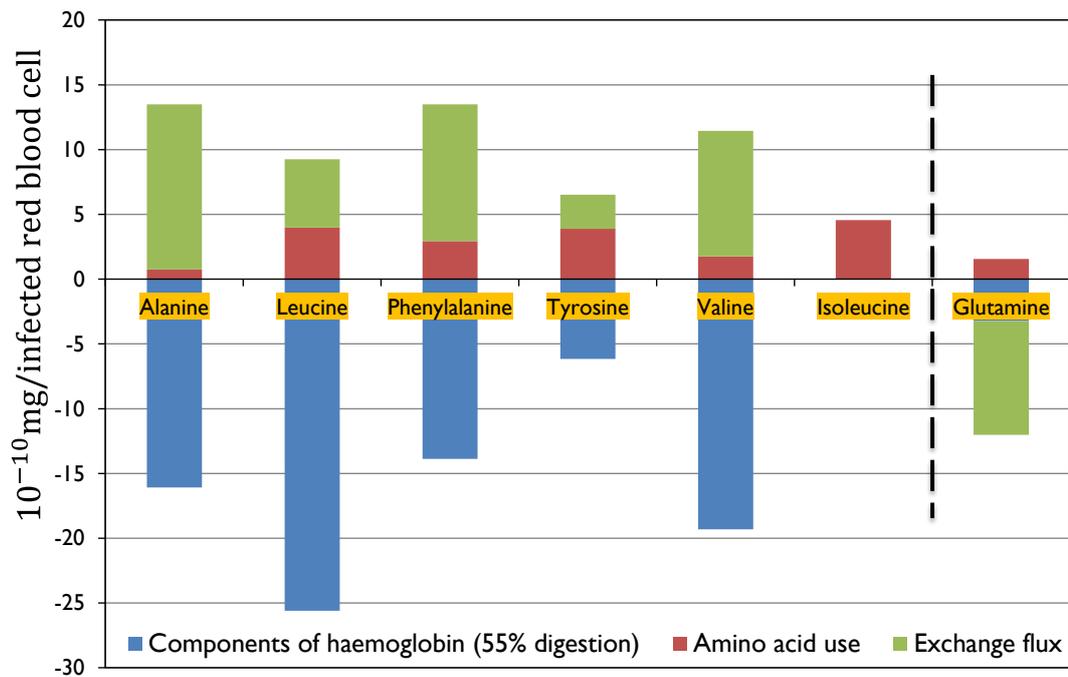


Figure 96 : Stacked bar charts showing the fluxes of amino acids to (negative) and from (positive) the pool of free amino acids (shown in figure 81) for the "Tom Repeat of Jenny" experiment. The blue bars represent the quantity of each amino acid freed from haemoglobin digestion (assuming 55% of the host cell's haemoglobin is digested). Red bars represent the quantity of each amino acid used for protein synthesis by the parasite. Green bars represent the measured fluxes of amino acids out of the parasite (into the parasite in the case of glutamine). The total mass of tyrosine and phenylalanine is conserved in this figure.

As always I am interested in checking that my measurements and calculations are reasonable and these results let us calculate a figure to compare with the unexpected haemoglobin digestion strategy employed by *P. falciparum*, specifically that,

"The parasite digested up to 65% of the host cell's haemoglobin but utilized only up to about 16% of the amino acids derived from haemoglobin digestion."

Krugliak et al.¹⁰³

Our minimum figure of 55% of haemoglobin being digested to supply adequate amounts of phenylalanine and tyrosine to explain protein formation and the measured excretion fluxes agrees well with this figure, deduced through a very different method. For the amino acids for which I have flux information — alanine, leucine, phenylalanine, tyrosine and valine — I calculate that 16% of the amino acids obtained through haemoglobin digestion are incorporated into the parasite's proteins. If isoleucine and glutamine are included then

this number increases to 23%: still in good agreement with the previous measurements by Krugliak *et al.*¹⁰³.

Lastly I should note that figure 95 and figure 96 are presented for the "Tom Repeat of Jenny" set of experiments because the larger number of acquired time points makes the amino acid exchange flux figures from this experiment more reliable. Similar results are observed for the Sara A and Sara B set of experiments but are not shown.

Possibility of stress response

In treating the Sara A set of experiments with a sub-lethal but active dose of atovaquone we might have expected to observe a slow-down in carbon use and haemoglobin digestion/amino acid excretion in keeping with the predicted slowing of the culture's growth-rate to zero. Instead we observe the opposite. Similar stress responses have been observed in other organisms and recently in *P. falciparum* grown in large-scale cultures where a similar stress response was observed,

"Once a continual decline in parasitemia was observed, however, glucose consumption and lactate production per parasite increased"

Preechapornkul *et al.*¹²⁹

It is tempting to imagine that a sensing mechanism within the parasite — perhaps similar to the distributed metabolic sensing observed in *E. coli*⁷⁶ — might allow the parasite to respond to the inhibited production of pyrimidines caused by atovaquone treatment by increasing other fluxes to try and force more flux through the affected reaction. Disappointingly the repeat of this experiment, "Tom repeat of Sara", shows the opposite with less growth and less metabolism in response to the same atovaquone inhibition.

Hypoxanthine

There is considerable discussion on the role and importance of hypoxanthine to parasites at the very of my chapter on measuring *P. falciparum* growth-rates and we expect to observe hypoxanthine being taken up by infected RBCs from the growth medium.

The NMR metabolomics techniques that I have used are good at detecting large changes in metabolite concentration caused by large carbon source fluxes due to their use as energy sources and large amino acid fluxes in considerable part due to excess haemoglobin digestion and the disposal of surplus amino acids. In comparison the calculations in this chapter on HEPES absorption show us that the NMR techniques I have used are not particularly well suited to measuring the smaller amounts of metabolite uptake caused by

the simple accumulation of a metabolite as is the case for hypoxanthine. A further complexity in interpreting the hypoxanthine results is that the assignment is unclear. In table 21 I assign a doublet at 8.20ppm to hypoxanthine but the shape of this doublet is variable within my spectra and often not convincingly in keeping with the reference spectra's shape.

The uncertainty around the hypoxanthine fluxes is reflected in the wide variability and extremely large errors for hypoxanthine exchange shown in figure 89. Clearly the results are not exactly what we would have expected but the issues I have just mentioned and a closer inspection of the time-series data in figure 88 suggests that measurements of hypoxanthine being produced and excreted from infected RBCs in the Sara metabolism set of experiments are incorrect. In the "Tom Repeat of Jenny" set of experiments where a larger number of time-points were acquired a considerable hypoxanthine uptake is measured albeit with a very large error in the best fit line.

False starts and dead ends

Measuring metabolism using kits

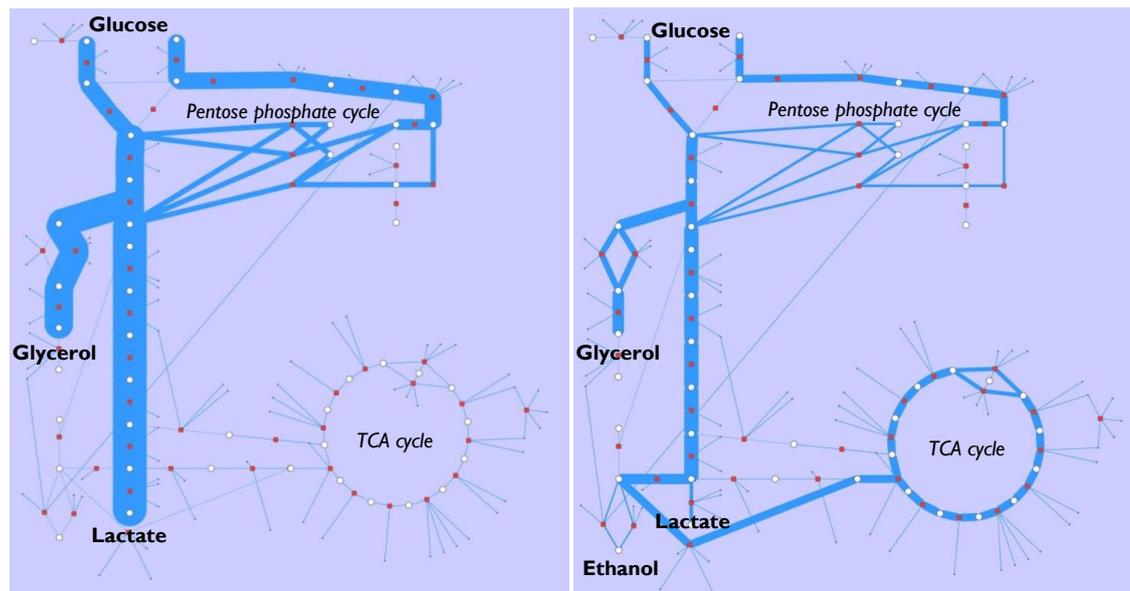
The use of biochemical assays to measure glucose consumption and lactate production by infected RBCs was largely a failure and is fully reported in Jennifer Lake's Masters dissertation¹⁴⁵. It was however a failure in the right direction and the glucose measurements were useful in confirming the validity of the NMR technique we moved on to and which has been very successful.

Fermentation of glucose to ethanol in *P. falciparum*

All metabolomics experiments — except "Tom repeat of Jenny metabolomics" where methanol was used instead of ethanol in the spray used to maintain sterility in the culture hood — show a steady accumulation of ethanol in the growth medium. When we first observed this it was very interesting because ethanol production could play a significant role in achieving a more satisfactory carbon balance. It would also be a completely new observation in *Plasmodium* although not completely unbelievable since the metabolism of glucose to ethanol has been observed in other apicomplexan parasites such as *Cryptosporidium parvus*¹⁹⁰.

Section 4.4 of Jennifer Lake's Masters dissertation¹⁴⁵ details the possible genes required to allow this and how we found some evidence for most of these within the *P. falciparum* genome. Figure 97 below is included as an example of the role of FBA in assessing the

metabolic changes that can result from small changes to a metabolic network and a first introduction to the tools and visual representation in the final chapter.



Lactate

Figure 97: FBA-predicted fluxes through my early 2010 model of glycolysis, the pentose phosphate cycle and the TCA cycle using minimum and maximum constraints for glucose, lactate, glycerol and ethanol fluxes. Low energy small molecules are provided free. The objective function consumes high energy molecules or precursors (ATP, NADH, NADPH etc...) and is maximised up to a top limit.

(left) with no flux allowed through additional ethanol generating reactions.
 (right) with flux allowed through additional ethanol generating reactions.

Ultimately, the much reduced and variable ethanol content of the growth medium when methanol was used as the sterilising agent strongly argued against fermentation occurring in *P. falciparum* and the experience serves as a warning that when you want to find genes coding for reactions in an organism they are worryingly easy to find and justify.

Variability of RBC metabolism

If you had asked me at the beginning of this project what metabolic processes were performed by an RBC I would have guessed that — apart from exchange of CO₂ and O₂ — they did nothing. This is very clearly not the case and I have measured considerable exchanges of many different metabolites between uninfected RBCs and their environment. Perhaps more forgivably I felt that the background metabolism of isolated and washed human RBCs would be largely constant across different experiments as long as the conditions were kept the same. For this reason — and largely because of the extreme difficulty in getting the time to acquire samples using the NMR machines — I have made poor use of much of my experimental time by not running sufficient blood nulls in parallel with my experiments with infected RBCs.

The variability of RBC metabolism, as shown most clearly in the wide variation of gradients in figure 76, still surprises me. Looking into RBC variability further it quickly

becomes clear that tiny changes in temperature¹⁷², age of donor¹⁷⁵, diet of donor¹⁹¹ and other unidentified or random sources of variability¹⁸² all lead to RBCs with very different metabolic profiles. A further complication is that the malaria parasite does not just feed on the RBC it infects, it changes its metabolism to its own requirements, digests and disposes of haemoglobin and the toxic haem group, and provides nutrients to sustain its host. Not only does the background metabolism of uninfected RBCs change considerably from experiment to experiment but so too does the parasite's ability and approach to adapting these variable hosts to its own metabolic requirements.

It is always dangerous to assume that anything can be left unmeasured and uncontrolled for between experiments with malaria.

Finally, as if the variability of RBCs was not enough of a challenge, there is even some good evidence in Mehta *et al.*¹³¹ that "*malaria parasite-infected erythrocytes inhibit glucose utilization in uninfected red cells*".

Solving the RBC variability problem?

Discussions with Prof. D. Fell have suggested some possibilities for avoiding the worst of the RBC variability problems that are worth considering. It is known that storage of RBCs leads to the depletion of internal metabolites — notably ATP and 2,3-Bisphosphoglycerate^{192,193} — and it seems likely that the variable storage times of RBCs before use makes this a significant source of variability. Additionally, degradation of RBC morphology is known to occur during storage¹⁹⁴.

As reported in Watanabe *et al.*¹⁹⁴, incubation in inosine restores ATP to near pre-storage concentrations and incubation with glucose and citrate restores RBC morphology to close to normal. Incubation in all three compounds restores both morphology and ATP levels, though less effectively than when performed separately.

No previous studies of *P. falciparum* metabolism report using any of these techniques to standardise the host RBCs' metabolic properties before experiments began and discussions with Dr. Glenn McConkey suggest this is not widely, if ever, used. It would be worth conducting further experiments using metabolically restored RBCs to see whether some of the variability observed in my work can be reduced. If this technique could allow experiments conducted at different times to be compared it would greatly increase the scope of comparative metabolomics of the malaria parasite.

Two precautions would need to be taken if following these RBC preparation techniques. Firstly, the impact of adding additional glucose to the system would need closely monitoring but should cause no problems as long as the glucose level was constant, or low

enough to be insignificant, at the time the RBCs were used. Secondly, the addition of inosine to the system — and to some extent the restoration of erythrocytic ATP — provides a supplementary purine source to hypoxanthine and further casts doubt on the applicability of radioactive hypoxanthine methods of growth-rate measurement. Assuming erythrocytic ATP was restored to the same level across all experiments this should cause no problems.

Metabolic modelling

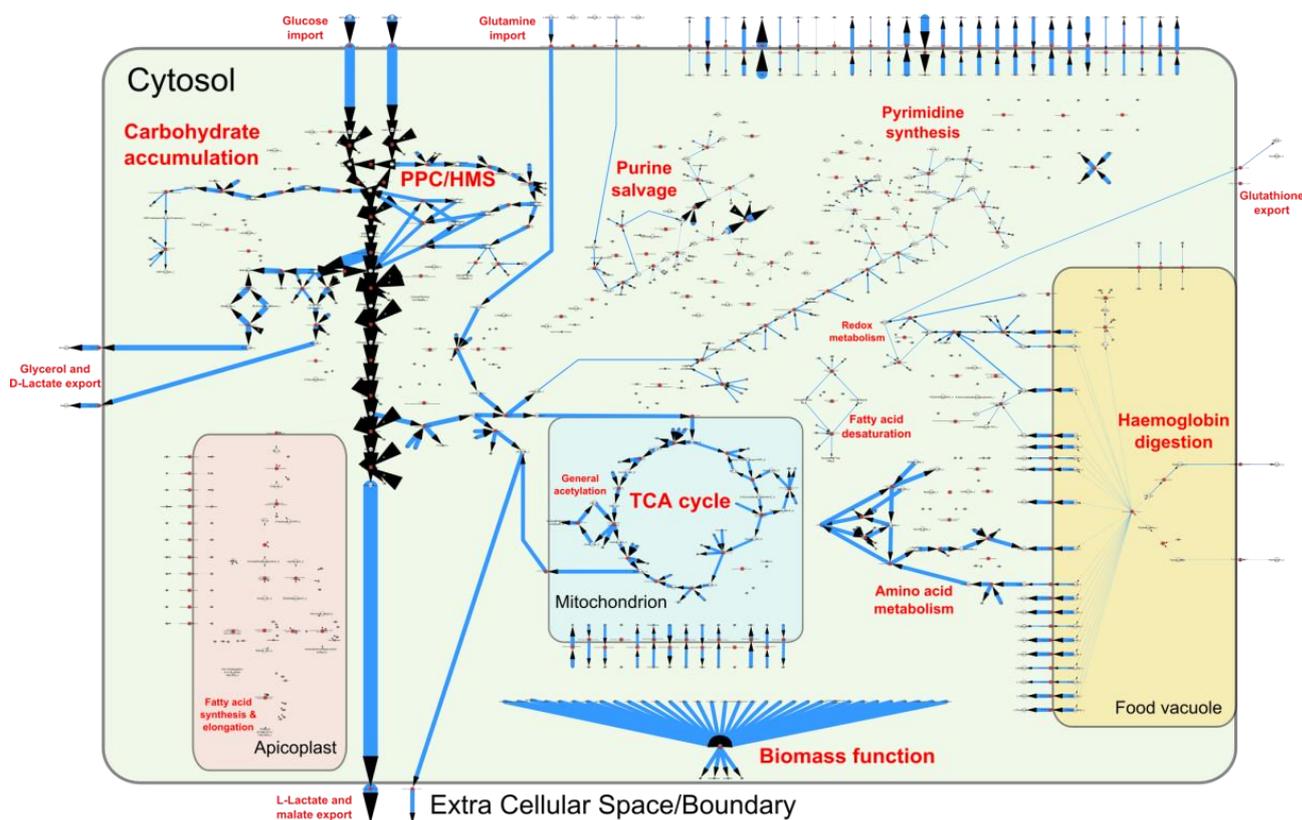


Figure 98 : Final annotated model with live flux projections of the optimal flux distribution given the applied experimentally derived constraints. Annotations in red show key pathways and features of the network. Flux weights are not linear and greatly exaggerate smaller fluxes. For example the three exchange fluxes at the top-left, α -D-glucose β -D-glucose and glutamine import have fluxes of 19.24, 19.24 and 0.98 respectively but appear much closer in width. Calculated using SurreyFBA⁸⁷.

The above figure shows the final result of all my modelling; it is my best representation of the metabolic fluxes in *P. falciparum*. This chapter describes how I combined everything I've described in the rest of this thesis to arrive at this answer and then explores what the model tells us about the parasite.

Creating the above model has taken four years and has been a process of trying, failing and improving as often as possible. The constraints, objectives and simplifications described in this chapter were developed alongside the development of the model rather in stark contrast to the way they are presented alone at the end of this thesis. The final results are heavily influenced by many layers of earlier results. This iterative process was made possible because of the technique of visual representation I have developed and use heavily in this chapter; I believe it is an important technique in metabolic modelling.

I will begin with a description of some of the key concepts, constraints and objectives that define the model.

Units of biomass and exchange fluxes

The problem of units in my metabolic model can be best outlined in the following two statements.

- In idealised batch-cultures of *E. coli* the rate of biomass production, the amount of biomass, and the fluxes in the system are all proportional.
- In synchronised cultures of *P. falciparum* the rate of biomass production is proportional to the fluxes in the system but neither of these is proportional to the amount of biomass. We know this because, as shown in figure 91 on page 188, no metabolite except valine shows a strong stage-specificity/time dependency of exchange flux even though biomass increases by a factor of 8-32 during the life cycle.

This key difference makes comparing the flux-balance analysis solutions of my model with other organisms very difficult and requires further explanation so that possible solutions to the problem can be fully understood. The results in this thesis are presented in units that are understood throughout the FBA community and a full understanding of these additional complexities is not necessary to appreciate my results.

The goal of my metabolic model is to simulate the creation of biomass and thus the growth of the parasite. Any units for flux and biomass formation can be defined within the SBML model but the convention is to express biomass creation in grams of dry weight per hour ($\Delta\text{gDW}/\text{hour}$) and fluxes in units of millimoles per hour per gram of dry weight of biomass ($\text{mmole}/\text{hour}/\text{gDW}$) usually rearranged as $\text{mmole}/\text{gDW}/\text{hour}$.

These units work well for unsynchronised cultures of bacteria or yeast in chemostats where the rate of consumption of nutrients and thus the rate of biomass formation is proportional to the biomass of the subject organism in the system and where no mass is lost from the system at the moment of replication. This condition is met for *E. coli*, *S. cerevisiae* and even *Leishmania major* where reproduction produces daughter cells which all survive. In these systems the rate of biomass production by the subject organism is equal to the rate at which the biomass of the system increases.

The cultures of *P. falciparum* that I have worked with are more complex than this. Since the majority of merozoites released when the schizont lyses its host do not successfully infect a new RBC the majority of biomass accumulated by the parasite over the 48 hour life cycle is lost to the system. We can estimate this since we know that the doubling time of a culture is around 24 hours and thus after 48 hours the biomass of an unsynchronised culture will have increased by a factor of 4. In this same time a single parasite starting a

merozoite and bursting out from a red blood cell at the 48th hour will have increased its biomass by a factor of between 8 and 32 depending on the number of daughter cells it produced. We can see that in an unsynchronised culture this means that the rate of biomass creation is far greater than the rate at which the biomass of the system increases and the condition described in the previous paragraph is no longer met.

A second complexity with *P. falciparum* is that my cultures are synchronised and remain synchronised far more strongly than batch cultures of bacteria or yeast. Although some variability of fluxes in line with life-stage has been observed elsewhere I measure little or none for the majority of the exchange fluxes I have been able to measure.

Since the biomass of the system is expanding rapidly over the life cycle — a schizont presumably weighs 8-32 times as much as a merozoite — the fluxes, if measured in units of mmole/gDW/hour with the gDW representing the dry weight of biomass at that moment, decrease by a factor of 8-32 over the course of the life cycle. Because of this it makes much more sense to measure fluxes with respect to the rate at which biomass is formed rather than the overall biomass of the system. This means that the unit of flux within the network is $(\text{mmole/hour})/(\Delta\text{gDW/hour}) = \text{mmole}/\Delta\text{gDW}$ with the fluxes defined in proportion to the rate of biomass accumulation.

Using these units would be fine except for the fact that all the tools for model analysis and the assumptions of all researchers in the field are set up differently. Thankfully there is an elegant hack that lets us pretend we're using the usual unit of flux.

We just have to imagine that we are examining a system with a mass equal to the amount of biomass produced in one hour. At this point the biomass of the system (gDW) is equal to rate of change of the biomass ($\Delta\text{gDW/hour}$) and the flux values are identical whether in units of mmole/gDW/hour or mmole/ ΔgDW . This simplification is the one that I have made throughout this chapter and that I use to report all of my results.

Where comparisons need to be made with other organisms we ignore the problem of lost biomass at reproduction and consider an unsynchronised culture with an average parasite age of 24 hours. Assuming this and dividing all the fluxes in my model — except the biomass flux — by 24 gives us absolute fluxes in mmole/gDW/hour for comparison with other flux-balance models. Following discussions with Prof. David Fell this is the system of units I will use for publishing and sharing my final model.

Within the final model fluxes are scaled and the biomass function calculated such that a flux through the biomass function of 1 gDW/hour is equal to the experimentally measured rate of biomass production.

Defining the final biomass function

The calculations in the experimental results chapter provide the percentages of dry weight that are used to calculate the biomass function. In these calculations the proportion of each component by dry weight (amino acids, DNA nucleotides, RNA nucleotides, fat, carbohydrate) is broken down into a proportion by stoichiometry and then further to stoichiometries of each of the subcomponents.

Amino acid stoichiometries are calculated as per Chanda *et al.*¹⁷⁶ (table 44 in appendix VI) and for each amino acid in the biomass function 2 molecules of GTP and one molecule of ATP are converted to 2 molecules of GDP and 1 molecule of AMP plus 2 molecules of orthophosphate and one molecule of diphosphate. This energy cost of protein formation makes up for the simplifications in my model whereby amino acids are not attached to tRNA molecules nor is the process of polymerisation modelled.

Nucleotides, both DNA and RNA, are included in the biomass function in the cytosol as the nucleus is not modelled. For each nucleotide consumed by the biomass function a molecule of diphosphate is released back into the system.

The carbohydrate composition of *P. falciparum* is unclear. In the distantly related parasite *L. major*, 90% of the carbohydrate content is known to be mannan¹⁹⁵, a polymer of mannose. In my model GDP-mannose (C00096) forms 90% of the carbohydrate component of the biomass function, with GDP-L-fucose (C00325) making up the remainder and acting as a surrogate for all glycosylation within the cell. For each molecule of GDP-mannose or GDP-L-fucose incorporated into the biomass function a molecule of GDP is returned to the system.

Fatty acids are incorporated into the biomass function according to the observations in Tarun *et al.*¹⁹⁶ and Mi-ichi *et al.*¹⁰⁴.

... Plasmodium parasites depend on de novo fatty acid synthesis only for liver-stage development.

A. Tarun *et al.*¹⁹⁶

... the parasite's overall fatty acid composition reflects that of the medium, although the parasite has a limited capacity to desaturate and elongate serum-derived fatty acids.

Mi-ichi *et al.*¹⁰⁴

Table 2 in Mi-ichi *et al.*¹⁰⁴ lets us calculate that the vast majority of fatty acids in the final parasite are unchanged from those absorbed from the growth medium and only 0.64% are elongated by a single C₂H₄ unit and 10.8% are desaturated. Elongation is modelled by the reaction 'single fatty-acid elongation' (SFAE) which is the simplification and generalisation of KEGG reactions R01626, R04952, R04953, R04954 and R04955 to create the custom compound 'FAC2H4unit' which is incorporated into the biomass function. Desaturation is modelled by 'R02222_generalised' and included in the biomass function as a custom compound 'DesatFA'. Whilst *de novo* synthesis of fatty acids is left in the model it is not included in the biomass as per Tarun *et al.*¹⁹⁶.

Figures on the process of fatty acid absorption and incorporation in *P. falciparum* are not available and the energy costs of this process are considered only as part of the ATP maintenance function. The general acetylation purpose of the branched-TCA cycle as described in Olszewski *et al.*⁹⁵ is modelled by the inclusion of 'Acetylated generic compound' (AcCmpd) in the biomass function at a stoichiometry equal to 10% of the fatty-acids portion of the biomass function.

Applying constraints

The largest number of constraints applied to the network are thermodynamic in origin and applied to the reversibility and direction of reactions. These restrictions were determined iteratively: by performing FBA, projecting the results onto the network map and then examining reactions where fluxes were far higher than reasonable. Where loops occurred that produced free ATP, NADPH or NADH, they were removed by adding constraints on the direction of reactions or limited to sensible fluxes in the case of arginine and proline conversion. In cases where reactions in pairs formed loops but the net flux direction for the pair of reactions was unclear one of the reactions was restricted to zero flux. Reaction R04125 — part of glycine and serine interconversion — was restricted to zero flux because it increased amino acid influx and efflux without improving biomass production. This was only a problem when using the SurreyFBA software for network analysis and could be removed if another solver were used.

Exchange flux constraints for measured carbon sources were applied to limit the model to the experimentally determined solution space. During model development constraints were applied to amino acid fluxes to try and match them to experimental parameters but this was not necessary in the final model.

A breakdown of the number of reactions to which constraints are applied and the reasons for those constraints is provided in table 29 with their locations represented on the visual map in figure 99.

249	total reactions
194	unconstrained and reversible reactions
35	non-reversible reactions (zero in one direction, unconstrained in the other)
9	exchange fluxes limited to known values (glucose uptake, glutamine uptake, etc...)
4	internal fluxes set to zero to stop thermodynamically unlikely cycles
3	limited to sensible fluxes where fluxes were not measured (arginine, proline, aspartate)
2	Adenine and adenosine fluxes set to zero to force hypoxanthine uptake
1	ATP maintenance
1	biomass function

Table 29 : A summary of the constraints applied to the reactions within the final model. The biomass is constrained in a forward direction until simulations are performed at which point it becomes the optimisation goal and is no longer constrained.

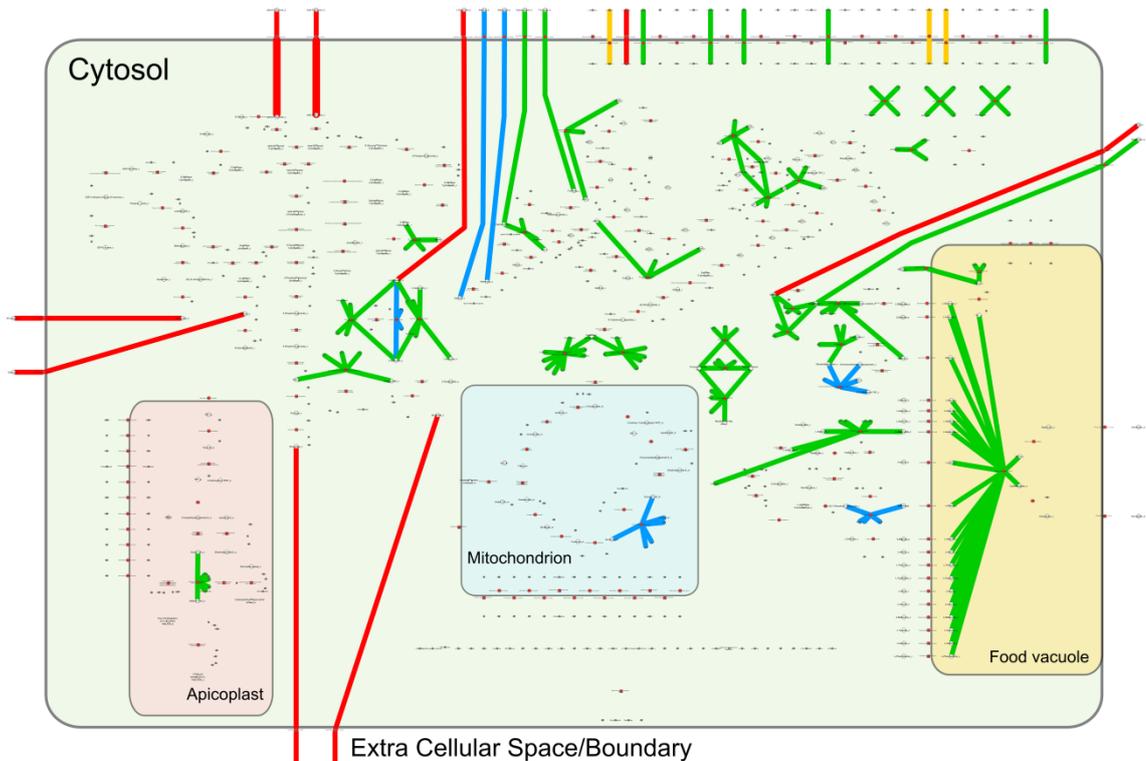


Figure 99 : A visual representation of the constraints applied to the final model during flux-balance analysis. The biomass function is not shown as it is the optimisation goal and not constrained in the final model. Green = direction constrained, Red = experimentally measured or literature reported flux constraint applied, Blue = flux set to zero, Orange = flux limited to a reasonable rate that could not be measured (arginine, proline, glutamate exchange).

Restriction on purine sources

The decision to restrict adenine and adenosine uptake as purine sources to zero is justifiable given the findings of Quashie *et al.* that,

“Hypoxanthine was taken up with 12-fold higher efficiency than adenosine.”

Quashie *et al.*¹⁹⁷

However, there are a number of papers including the appropriately named “*Erythrocytic Adenosine Monophosphate as an Alternative Purine Source in Plasmodium falciparum*” by Cassera

*et al.*¹⁹⁸ that suggest purine uptake may be more complex than previously discussed. The reactions allowing the uptake of the alternative purine sources are left in the model and the restrictions on the fluxes through them can be loosened if desired.

NADH/NADPH interconversion

The importance of NADH fates within the parasite⁴⁶ means that NADH/NADPH interconversion could play an important role in the parasite's metabolism. Direct interconversion via R00112 (NADPH:NAD⁺ oxidoreductase, EC number 1.6.1.1 and/or 1.6.1.2) is part of the network but a limited additional capacity may also be provided by two loops between Glutathione and GSSG (reduced glutathione) and between Ferricytochrome B and Ferrocycytochrome B. So that NADH/NADPH interconversion can be monitored by observing a single flux these loops are disallowed in the network and then provided for by R00112, with a precautionary note placed in the SBML file.

Alternatives to internal constraints

Applying constraints to stop loops and to limit thermodynamically unfeasible reactions works but requires a judgement to be made on what the behaviour of the system would be. Since the model is designed to show us the behaviour of the system this clearly reduces the objectivity of the model's predictions.

Two ways that similar restrictions can be placed on the model without a subjective decision being made are to consider the thermodynamic cost and to couple reactions.

The thermodynamic consideration involves calculating the free energy change of each reaction within the model and adding an optimisation condition to the analysis such that the system as a whole is thermodynamically balanced. This condition would limit fluxes through costly reactions — such as loops that create free ATP — without absolutely stopping them and a discussion of how to implement a system like this is found in Henry *et al.*¹⁹⁹.

A simpler solution to the problem is to link the direction of pairs or groups of reactions that may cause an unnatural loop such that net flow in either direction is allowable but the thermodynamically unacceptable creation of free ATP is disallowed. This is an approach adopted in Feist *et al.*²⁰⁰ and elsewhere.

Adopting either of these — or other similar — approaches can improve the predictive power and objectivity of the metabolic model but has the side effect of requiring more complex analysis, often using custom software which makes the resulting model hard to re-use and its predictions hard to reproduce.

ATP maintenance function

It is an inconvenient fact that FBA of even the best current metabolic models for model organisms like *E. coli* predicts far higher growth-rates than are experimentally measured. The solution to this problem is to add an ATP maintenance function to the final model that reduces the growth-rate to an experimentally measured or expected rate in ideal conditions. The ATP maintenance is typically split into two parts with the majority of the ATP flux integrated with the biomass function and called the growth-associated maintenance function and a lesser part added to the network as a standalone reaction converting ATP and H₂O to ADP and orthophosphate and called the non-growth-associated maintenance function.

Since all the fluxes in my network are proportional to growth the distinction between non-growth associated and growth-associated maintenance is unnecessary and I have implemented it as a separate reaction (ATPmaint_c) within my final network. Figure 100 shows the effect of varying the flux forced through the ATP maintenance function on biomass production and justifies my choice of 30 mmole/gDW/hour to limit the biomass production in optimal conditions to close to the measured rate of 1 gDW/hour.

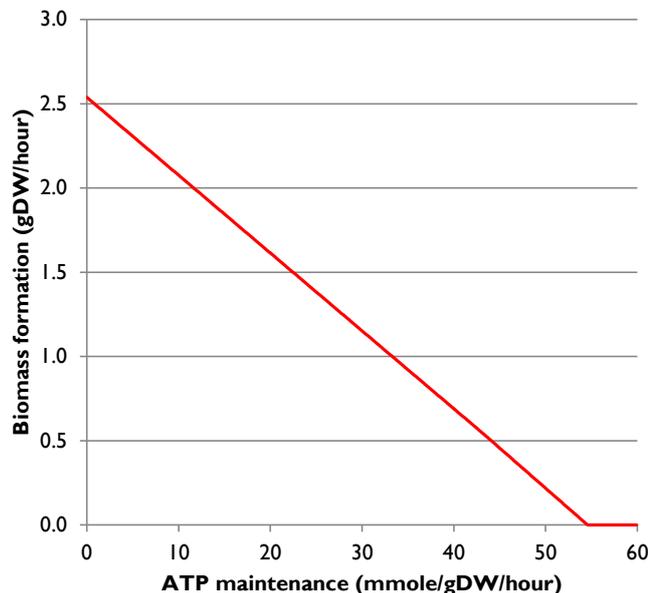


Figure 100 : The effect of an increasing ATP maintenance flux on the rate of biomass formation. Calculated using SurreyFBA⁸⁷.

If the ATP maintenance function is integrated into the biomass function it changes the growth-rate predicted by my model by less than 5% and makes no difference to other fluxes. Given the cost in terms of lost flexibility of incorporating the ATP maintenance function into biomass function and the limited physiological ranges of *P. falciparum* growth-rates I have chosen to keep the ATP maintenance function separate from the biomass function for all the results discussed in this thesis.

The ATP maintenance function is the “least bad” solution to the problem that growth-rates are otherwise over estimated and in conversations with Pedro Mendes at the University of Manchester and Fiona Achcar at the University of Glasgow it is clear that in the majority of metabolic models there is very limited experimental justification for the value eventually chosen. My chosen ATP maintenance flux of 30 mmole/gDW compares to 59.81mmol /gDW in Adam Feist’s 2007 *E. coli* reconstruction²⁰⁰, 32.26mmol/gDW in Chavali’s *L. major* reconstruction¹⁰⁸ and 60.01mmol/gDW in Plata’s *P. falciparum* reconstruction⁵⁸. The Huthmacher *P. falciparum* model implements the ATP maintenance reaction separately from the biomass function in the same way as I do but the flux forced through it is not included in the SBML model as a non-standard analysis technique is used.

I prefer to keep the ATP maintenance function completely separate from any identifiable use of ATP but this is not the case in many other models. If the ATP maintenance flux is integrated with the biomass function and added to the identified ATP and GTP costs of amino acid polymerisation the final ATP/GTP growth-associated maintenance flux is 42.31mmole/gDW; even closer to that of the similar models I have just mentioned.

Malaria-specific ATP exchange complexities

There are additional difficulties and distractions in the literature when considering the ATP drains in the malaria parasite. In Choi *et al.*²⁰¹, isolated parasites — freed from their hosts RBC by saponin lysis —import ATP from their surroundings. This result should be considered with caution since freed parasites may well have extremely perturbed metabolism. An intriguing result in Kanaani *et al.* 1989¹¹⁷ suggests that during the RBC stage of its life cycle the parasite exports ATP to its host, presumably to stop the host from dying. I have considerable doubts about this finding given that the paper relies on the assumption that mitochondrial inhibitors reduce the amount of ATP production in the parasite. Significant research, most notably Fry *et al.*²⁰² and Painter *et al.*¹¹⁹, strongly suggest that the mitochondrion does not play a significant role in ATP production in *P. falciparum*.

The uncertainty and complexity of the ATP exchange between parasite and host is such that I have not included it in my final model and the ATP maintenance reaction is sufficient to contain any intriguing ATP transfer behaviour performed by the parasite that I have not added to my model.

Adding an additional cost to haemoglobin digestion

My final model’s considerably simplified haemoglobin digestion pathway within the food vacuole serves its primary purposes of supplying amino acids to the system but presents a

problem. This problem is shown most clearly in the graph on the left of figure 101 where we see that a higher rate of biomass formation occurs where more haemoglobin is digested. An upper bound on this effect is only imposed by limiting proline efflux to an acceptable rate*.

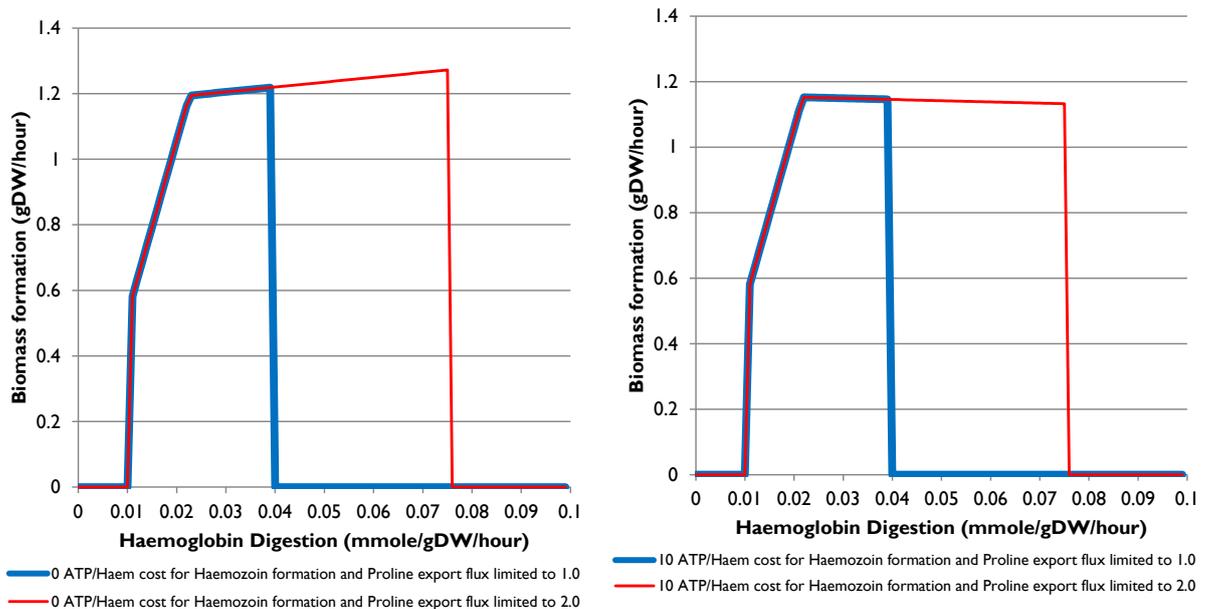


Figure 101 : The growth-rate predicted by FBA on the final model for a range of haemoglobin digestion fluxes. Calculated using SurreyFBA⁸⁷.

(left) The formation of haemozoin from haem incurs no cost and the optimal solution is for the parasite to digest ever more haemoglobin. A reduction in the maximum allowed efflux of proline from the system places an upper limit on the haemoglobin digestion flux.

(right) The addition of a 10 ATP cost for each molecule of haem incorporated into haemozoin moves the optimal haemoglobin digestion flux lower but still allows a large range of haemoglobin digestion fluxes that give a biomass production rate close to optimal.

The solution to the problem of non-convergent haemoglobin digestion is to add an ATP cost to the process of crystallising the haem molecules released by haemoglobin digestion to haemozoin. We can justify this cost as representing what we know is a metabolically expensive process but its magnitude is chosen for no other reason than that it moves the solution space of the model to give the graph on the right of figure 101. Here the system's optimum rate of haemoglobin digestion does not require limiting by an arbitrary constraint on proline efflux.

At the maximum haemoglobin digestion flux of 0.075 mmole/gDW/hour shown in figure 101 the 10 ATP/mole of haem crystallised as haemozoin (40 ATP/mole of haemoglobin digested) represents an addition of 3 mmole/gDW/hour to the total ATP maintenance.

* see Figure 105 for a justification of this rate.

Results

The creation, parameterisation and constraint of my metabolic model has taken up the majority of this thesis and the majority of my time over the past four years. Getting the results from the final model was a considerably shorter and more instantly rewarding experience. I report the main findings here.

The model accurately predicts the growth-rate of *P. falciparum*

The carbon-source flux constraints in my model are set from my experimental values such that the experimentally measured biomass production rate is 1 gDW/hour. Before the addition of the ATP maintenance function my model produces a biomass production rate of around 2.5 gDW/hour which — whilst higher than what is experimentally measured — is of the same magnitude. If the predicted growth-rate was an order of magnitude higher than what we had measured we would be able to infer that the parasite had an energy source other than that included in the model, presumably either the pathways required for aerobic respiration or a direct uptake of ATP from the host RBC.

That the predicted growth-rate is reduced to the experimental growth-rate when the ATP maintenance function is set to a value close to that of other similar organisms is further evidence that the model accurately predicts growth-rate.

Measured glucose influx and lactate efflux rates achieve carbon balance and optimal growth

We know that the primary energy source for *P. falciparum* is the creation of lactate from glucose via the glycolysis pathway and therefore the influx of glucose is closely linked to the efflux of lactate. Figure 102 shows how biomass production varies for the model's final allowed efflux rate of lactate of -60 mmole/gDW/hour and has some very interesting features. At low levels of glucose uptake the model produces no biomass at all as the organism's ATP maintenance requirements are not met. This is entirely in keeping with widespread knowledge that *P. falciparum* parasites die extremely rapidly in low glucose environments. Once the maintenance requirements of the parasite are met growth is highly dependent on glucose influx until a peak is reached where no further lactate efflux is allowed. The flexibility of the model means that carbon balance can be achieved at no cost to growth for a brief plateau of increasing glucose influx before the glucose is used inefficiently by the parasite and leads to reduced growth.

The shaded region in the graph is centred on the measured glucose influx rate for the Tom Metabolomics set of experiments and extends as far as the uncertainty on that measurement in either direction.

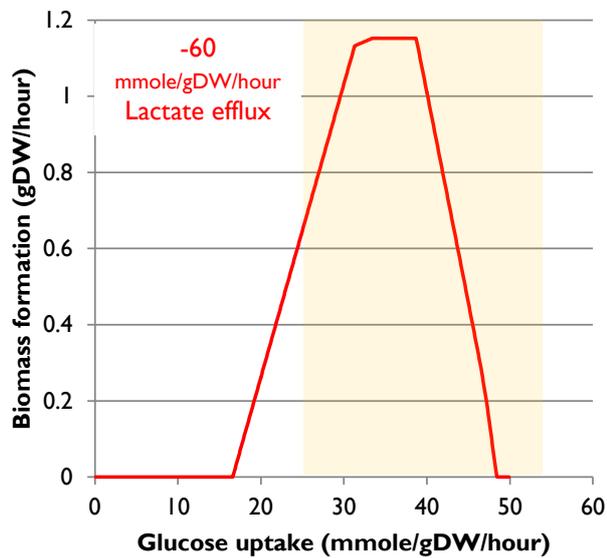


Figure 102 : Biomass production response to glucose uptake. The shaded region shows the measured glucose influx and its associated error (38.48 ± 14 mmole/gDW/hour). The lactate efflux of -60 mmole/gDW/hour is slightly higher than the best measured value (56) but well within the error on that measurement. Calculated using SurreyFBA⁸⁷.

Figure 103 is the equivalent for lactate of figure 102 with the shaded region this time centred on the final lactate efflux rate of -60 mmole/gDW/hour and extending as far as the uncertainty on that measurement in either direction. The dependency of biomass formation rate on the increasing lactate efflux shows that energy produced by glycolysis is a primary restriction on the growth-rate of the parasite.

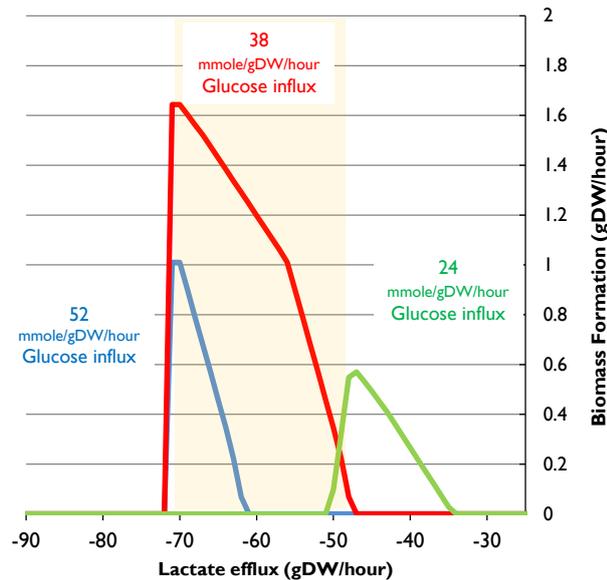


Figure 103 : Biomass production response to lactate efflux for three glucose uptake rates corresponding to the best-guess (38 ± 14 mmole/gDW/hour), minus and plus the associated error (24, 56). The shaded region shows the final lactate efflux and its associated error (60 ± 11 mmole/gDW/hour). Calculated using SurreyFBA⁸⁷.

What we see from these two graphs is that my measurements of glucose and lactate fluxes accurately predict growth-rate — the experimentally measured growth-rate given the glucose and lactate flux constraints is 1gDW/hour — and achieve carbon balance well

within the uncertainties in their measurements. Optimal growth-rates of the model are achieved within the measured ranges of glucose influx and lactate efflux.

I experimented with displaying the dependency of the rate of biomass formation as a function of allowed lactate efflux and glucose influx on a three-dimensional plot but found these hard to interpret once printed.

The model accurately predicts internal fluxes

Large-scale quantitative measurements of internal metabolic fluxes in *P. falciparum* has to my knowledge been limited to the work of Olszewski *et al.*¹⁰⁹ focusing on the TCA cycle. A far more limited resource is provided by Atamna *et al.*¹⁶⁷ which measures the diversion of the glycolytic flux towards the oxidative pentose-phosphate cycle* (PPC) and towards nucleotide metabolism through the formation of ribose sugars.

This provides an excellent opportunity to compare the predictions of my metabolic model with experimentally measured fluxes and the agreement shown in table 30 is encouraging.

Path Taken	Flux from Atamna 1994 (mmole/gDW/hour)	FBA predicted Flux (mmole/gDW/hour)
Glycolysis	41.90	37.54
Oxidative PPC	1.50	1.42
Nucleotide	0.06	0.24

Table 30 : Fluxes in mmole/gDW/hour of glucose towards the pentose phosphate cycle and nucleotide sugar formation. Figures from Atamna *et al.*¹⁶⁷ are converted to equivalent units for comparison. Calculated using SurreyFBA⁸⁷.

Whilst glucose and oxidative PPC fluxes are in excellent agreement, my model predicts a higher flux to nucleotides than reported in Atamna *et al.*. This is consistent with my suspicion that the true DNA content of my biomass function should be towards the lower end of the uncertainties in the values I've measured.

Another explanation for some of the discrepancy in the glucose to nucleotide flux is that that my model forces the exclusive use of hypoxanthine as a purine source. This requires a glucose flux to nucleotides in the form of PRPP (C00119) or Ribose 1-phosphate (C00620) whereas an uptake of adenosine — as known to be possible from Cassera *et al.*¹⁹⁸ — would not require this diversion of flux as PPC-synthesised ribose sugars would only be required for pyrimidine synthesis.

* The hexose-monophosphate shunt (HMS) is referred to in the paper. These are the same thing.

The model accurately predicts the low-oxygen metabolism of *P. falciparum*.

Another area in which the model accurately reproduces the known metabolism of *P. falciparum* is in relation to oxygen consumption. Oxygen and carbon dioxide are both taken up from the environment but neither O₂ nor CO₂ exchange is essential for parasite growth. If exchange of oxygen is limited to zero, growth continues but is slowed to 67% of the optimal growth-rate, consistent with the knowledge that *P. falciparum*'s optimum culturing conditions are a low — but not zero — oxygen environment.

The optimal growth-rate is achieved when the parasite takes in CO₂ at 0.009 mmole/gDW/hour and O₂ at 0.178 mmole/gDW/hour. This compared with the Adam Feist *et al.*²⁰⁰ *E. coli* model which takes up 18.2 mmole/gDW/hour of oxygen. If the factor of 24 correction discussed at the start of this chapter is applied then *P. falciparum* has an oxygen requirement for optimal growth of just 0.074 mmole/gDW/hour of oxygen: around 250 times less than *E. coli*.

By comparison, Adam Feist's *E. coli* model has a glucose import flux of 8-11 mmole/gDW/hour compared to the glucose import flux of my *P. falciparum* model of 38.4 mmole/gDW/hour, or at the 24th hour of the life stage, 1.6 mmole/gDW/hour. That the comparable glucose uptake rate of *E. coli* is six times higher (8-11 compared to 1.6) than *P. falciparum* but its oxygen use is 250 times higher (0.178 compared to 0.074) shows how inefficiently *P. falciparum* uses glucose by relying on anaerobic respiration whilst *E. coli* can perform aerobic respiration.

The model accurately predicts measured amino acid exchange rates

Although loose constraints are placed on proline and arginine exchange fluxes these are not reached in the model unless the rate of haemoglobin digestion would lead to more haemoglobin being digested than is available*. None of the other amino acid efflux rates are constrained in the final model and so the predicted amino acid exchange fluxes are the result of the model's haemoglobin digestion rate, amino acid use for the protein synthesis part of the biomass and any interconversion of amino acids within the model.

The FBA-predicted exchange fluxes and my NMR-measured fluxes for the five amino acids plus hypoxanthine I was able to measure are shown in table 31.

* see Figure 105 for a justification of this rate.

Compound	Measured flux (mmole/gDW/hour)	FBA predicted Flux (mmole/gDW/hour)
Valine	0.63	1.16
Leucine	0.35	1.18
Isoleucine	0.00	-0.44
Alanine	1.25	1.46
Tyrosine	0.13	0.26
Phenylalanine	0.54	0.44
Hypoxanthine	-0.45	-0.13

Table 31 : Measured exchange fluxes of the six measured amino acids and hypoxanthine compared with the fluxes predicted by the optimal solution of my final model. Calculated using SurreyFBA⁸⁷.

The agreement of these two sets of results can be best seen as a graph as in figure 104 where a linear fit shows predicted fluxes to be on average 1.26 times larger than measured fluxes. If we scale this figure to take into account the discrepancy between the measured growth-rate of 1 gDW/hour and the predicted growth-rate of 1.15 gDW/hour the gradient of the line of best fit becomes 1.10: an excellent agreement between the model and experimental results.

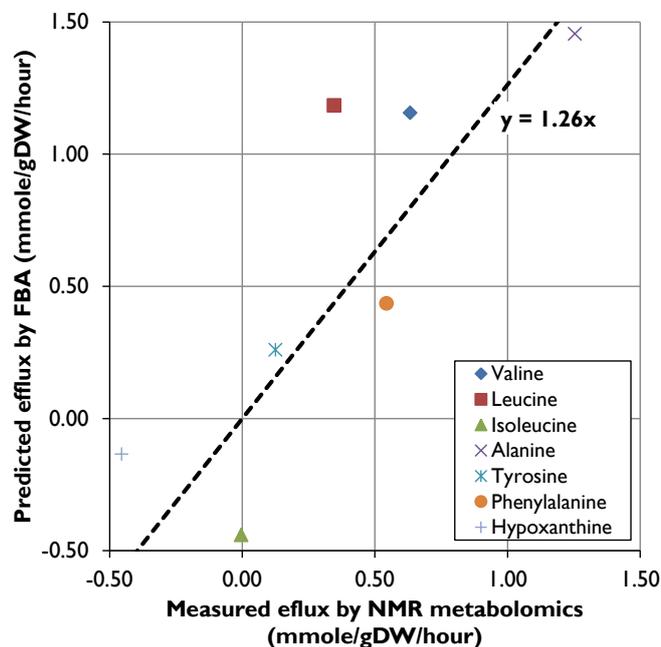


Figure 104 : Measured flux of amino acids and hypoxanthine against FBA-predicted fluxes at the optimal FBA solution of my final model. Best fit line is forced through the origin.

As a result of the complex nature of my final model the relationship between the rate of haemoglobin digestion and the rate of biomass formation is complicated, but flux-balance analysis easily calculates it. From these two pieces of information and knowledge of the haemoglobin content of an uninfected red blood cell and the mass and protein content of a mature parasite we can calculate — for a single parasite — the percentage of the host

RBC's haemoglobin that is digested and the percentage of the freed amino acids that are integrated into the parasite's proteins as in figure 105.

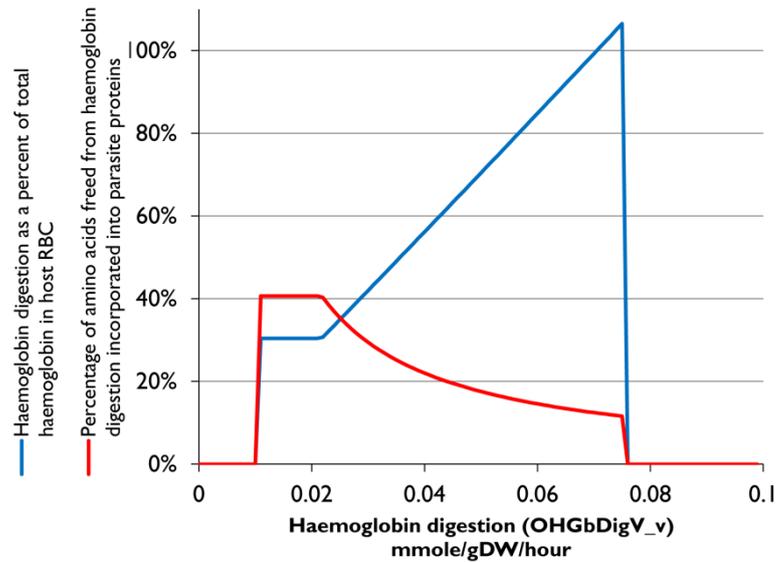


Figure 105 : The relationship between the haemoglobin digestion flux and the percentage of the host's haemoglobin that is digested (blue line) and the percentage of the freed amino acids that are integrated into the parasite's proteins (red line).

At the optimal solution — giving the numbers in table 31 — the haemoglobin digestion flux of 0.022 mmole/gDW/hour is equivalent to the digestion of 30% of the host's haemoglobin and incorporation of 40% of the freed amino acids into the parasite's protein. The upper limit on haemoglobin digestion set by a limit on proline efflux approximately limits haemoglobin digestion to the amount present in the host red blood cell.

A haemoglobin digestion flux of 0.046 mmole/gDW/hour is equivalent to digestion of 65% of the host's haemoglobin and incorporation of 19% of the freed amino acids into the parasite's protein and is closer to the maximum equivalent values measured by Krugliak *et al.*¹⁰³ of 65% digestion and 21% incorporation respectively. As previously shown in figure 101 the range of haemoglobin digestion fluxes possible whilst only slightly reducing the predicted growth-rate is considerable and so Krugliak *et al.*'s results are consistent with my model. However, it is important to note that this level of haemoglobin digestion, whilst energetically possible, does not agree with my measured NMR exchange fluxes as shown in table 31.

Flux variability analysis (FVA) at the optimal growth-rate

I have shown how the level of haemoglobin can vary considerably whilst still producing a near-optimal growth-rate but it is informative to perform flux-variability analysis on all the reactions in the network.

Figure 106 shows the reactions that can vary whilst still producing the optimal growth-rate where the thickness of each edge corresponds to the range of fluxes allowed through the corresponding reaction. The four futile loops that produce and consume nothing in a complete cycle are noted in the figure's caption and are not interesting except to note that the fluxes through these reactions are not informative in the optimal solution.

Of more interest in figure 106 is to notice that the variances of the amino acid interconversions at the bottom right of the figure and the variances of the reactions in glycolysis are the same, suggesting that the two are linked. L-malate export is variable between zero and its maximum allowed value for the optimal solution meaning that both the measured malate efflux in Olszewski *et al.*⁹⁵ and my measured lack of a malate efflux are consistent with the predictions of my model.

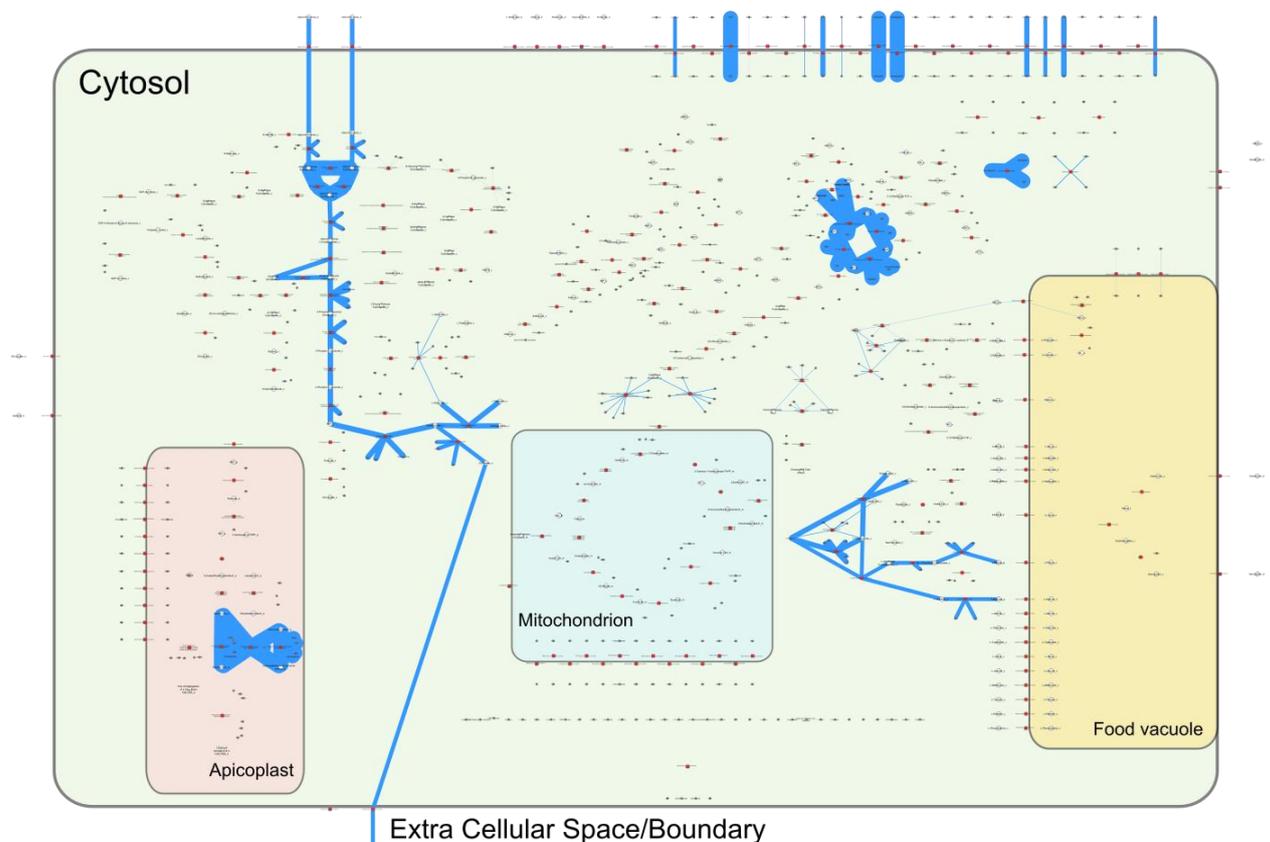


Figure 106 : The variability of fluxes that still give 100% of the optimal growth-rate. The very thick lines are three cycles that are futile and can have any value as they ultimately produce and consume nothing. The three exchange reactions involved in a futile cycle are H_2O , Orthophosphate and Diphosphate which combine with an internal reaction that converts Diphosphate to 2 Orthophosphate + H_2O to form a futile cycle. Aspartate, proline and arginine exchange fluxes are all variable within the optimal solution space and aspartate and proline exchange fluxes are constrained to a sensible maximum value in the final model. Calculated using the COBRA toolbox v1.3.3²⁰³.

Flux variability analysis (FVA) for a solution producing at least 99% of the optimal growth-rate

FVA at the 99% level, shown in figure 107, shows the range of values that each flux can take within a solution that can still produce biomass at 99% of the optimal rate. The interesting progressions from the 100% FVA solution are that

- The flux through the branched TCA cycle via succinate becomes highly variable, suggesting that this branch plays a minimal role in any modelled aspect of growth. The proposed role of this branch of the branched TCA cycle in Olszewski *et al.*⁹⁵ is not completely clear and needs further investigation.
- The pentose-phosphate cycle (PPC) and NADPH:NADH interconversion (R00112. X-shape at the top-right of figure 107) become variable, suggesting that changes in NADPH production by the oxidative PPC can be compensated for by R00112 without significantly affecting growth.
- Haemoglobin digestion rate becomes variable, as expected from the results in figure 101.

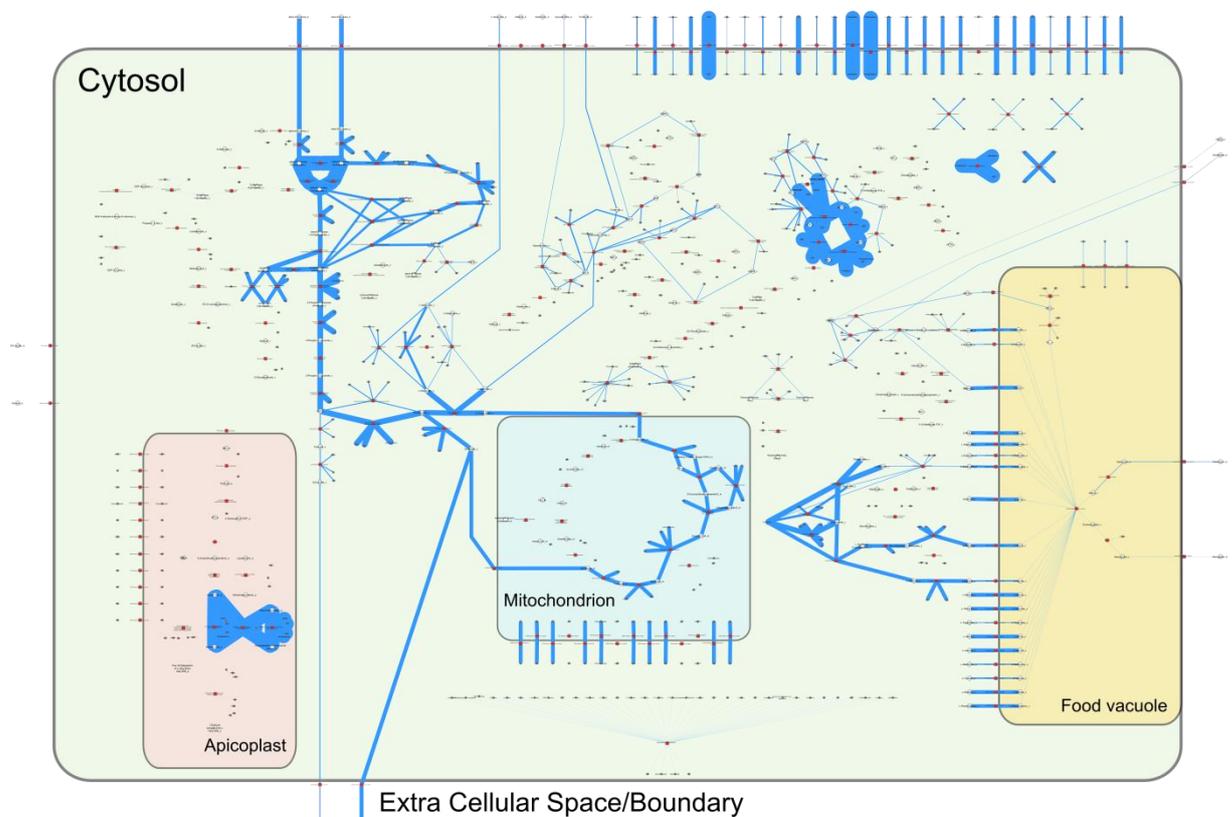


Figure 107 : The variability of fluxes that still give 99% of the optimal growth-rate. The flux through the mitochondrion via succinate becomes variable as does haemoglobin digestion and the pentose-phosphate cycle. Calculated using the COBRA toolbox v1.3.²⁰³

Flux variability analysis (FVA) for a solution producing at least 70% of the optimal growth-rate

The range that each flux can take and still form a solution that produces biomass at 70% of the optimal rate — shown in figure 108 — is extremely large and covers almost all the network. The reactions that are least variable in this solution are those that are not capable of carrying any flux and those that produce a component of the biomass along a metabolic pathway for which there are no alternative routes. For example, adenine and adenosine uptake and the subsequent reactions required to use these purine sources for nucleotide synthesis have no flux variability because they are disallowed in the model and thus always have a flux of zero. The uptake of isoleucine and the branch of the TCA cycle via citrate has a low variability since isoleucine and the general acetylation product produced in the mitochondrion are made available only through a set of non-redundant reactions.

With relation to earlier observations of the low oxygen and carbon dioxide uptake rates of the model it is interesting to note that the parasite is much more able to deal with changes to oxygen uptake than with changes to carbon dioxide uptake. At the FVA 70% level, oxygen absorption ranges between 0.009 — 0.202 mmole/gDW/hour whilst carbon dioxide absorption ranges between 0.006 — 0.009 mmole/gDW/hour.

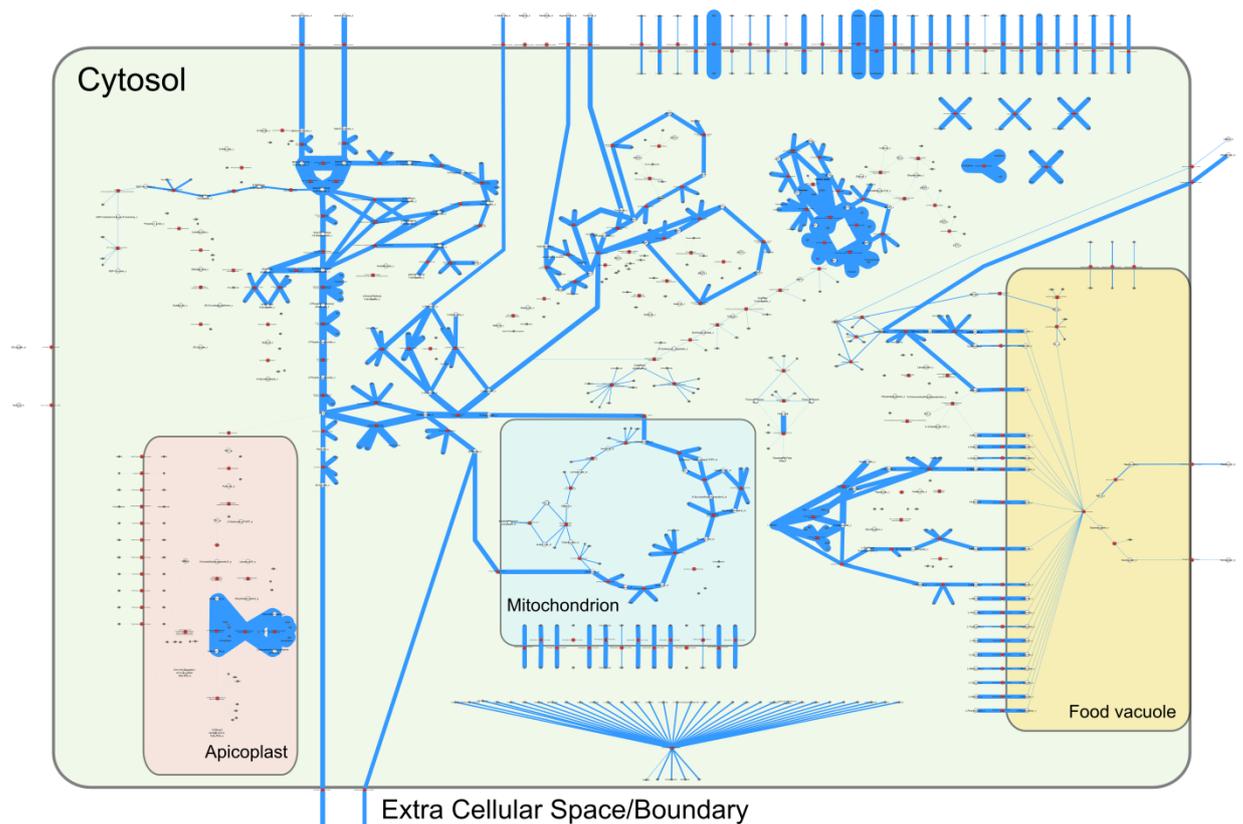


Figure 108 : The variability of fluxes that still give 70% of the optimal growth-rate. Calculated using the COBRA toolbox v1.3.3²⁰³.

Essential reactions and predicted lethal and sub-lethal single-gene deletions

Out of 249 reactions in the network, 220 are capable of carrying some flux. Of these “live reactions” 117 are essential and reduce growth-rate to zero if removed. The location of these reactions within the network is shown in figure 109.

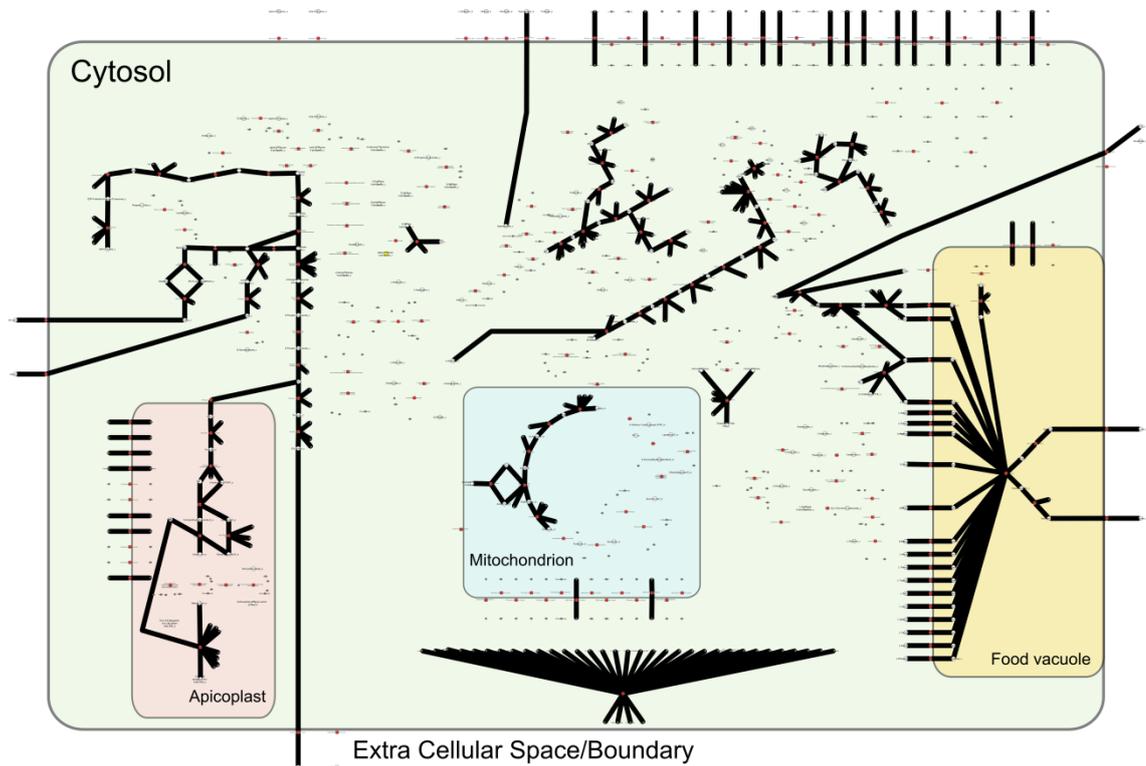


Figure 109 : The location of essential reactions in the network. Calculated using SurreyFBA⁸⁷.

Since more than one gene, or indeed no gene, can be linked to a reaction and more than one reaction can be linked to a gene the 117 essential reactions correspond to 79 lethal gene deletions. A further 19 gene deletions cause a reduced growth-rate but are not lethal with these genes largely associated with the oxidative PPC cycle and the succinate branch of the TCA cycle. These are, as expected, the parts of metabolism that show considerable flux-variability within the FVA 99% result shown in figure 109.

Another interesting group of genes that cause a reduced growth-rate but that are not lethal are those associated with the oxidative stress response. This set of genes includes those coding for reactions that oxidise hydrogen peroxide via the reduction of glutathione and genes associated with the maintenance of NADPH/NADH levels needed to restore glutathione to its oxidised state.

A summary of the genes whose deletions affect growth-rate is provided in the table making up appendix VII. Of these genes it is interesting to note that of the 98 predicted lethal or growth-reducing single-gene deletions only 22 are predicted as lethal in Plata *et al.*⁵⁸. More worryingly, of the 45 genes whose deletion has no effect on growth-rate, 8

are predicted as lethal in Plata *et al.*⁵⁸. All single gene-deletion simulations were carried out in the COBRA toolbox v1.3.3²⁰³.

Lethal and sub-lethal double-gene deletions

Any double gene deletion involving a gene that is lethal when deleted on its own is also lethal and the vast majority of double-gene deletions are trivially lethal in this way and therefore uninteresting. The second largest group of double-gene deletions are those where both genes are non-lethal when deleted singly and non-lethal when deleted in pairs. The most interesting set of double-gene deletion behaviours are those where both genes when deleted singly are not-lethal but that are lethal when deleted together. The predicted growth-rate — as a proportion of the optimal growth-rate — for fifteen genes and thus 105 gene pairs is shown in figure 110. Full details on these genes is presented in table 32.

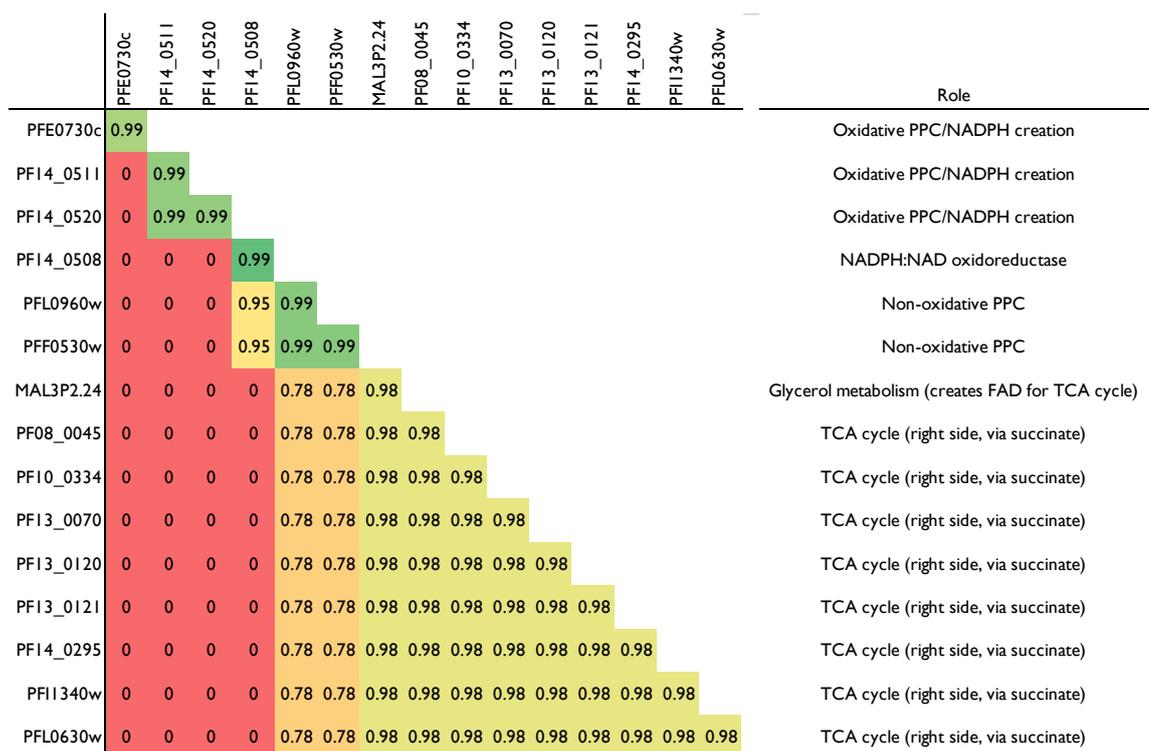


Figure 110 : Predicted growth-rate as a proportion of optimal for all pairwise deletions of fifteen genes that give interesting growth-rate predictions. The metabolic role of each gene within the network is shown at the right. Growth-rate predictions made using COBRA toolbox 1.3.3²⁰³.

Gene	Associated Reaction(s)	Reaction Name(s)	Associated EC number	Role
PFE0730c	R01056	D-ribose-5-phosphate aldose-ketose-isomerase	5.3.1.6	Oxidative PPC/NADPH creation
PF14_0511	R02035	6-Phospho-D-glucono-1,5-lactone lactonohydrolase	3.1.1.31	Oxidative PPC/NADPH creation
	R02736	beta-D-Glucose-6-phosphate:NADP+ 1-oxoreductase	1.1.1.49	
PF14_0520	R01528	6-phospho-D-gluconate:NADP+ 2-oxidoreductase (decarboxylating)	1.1.1.44	Oxidative PPC/NADPH creation
PF14_0508	R00112	NADPH:NAD+ oxidoreductase	1.6.1.1	NADPH:NAD oxidoreductase
PFL0960w	R01529	D-Ribulose-5-phosphate 3-epimerase	5.1.3.1	Non-oxidative PPC
PFF0530w	R01641	Sedoheptulose-7-phosphate:D-glyceraldehyde-3-phosphateglycolaldehyde transferase	2.2.1.1	Non-oxidative PPC
	R01830	beta-D-Fructose 6-phosphate:D-glyceraldehyde-3-phosphateglycolaldehyde transferase	2.2.1.1	
MAL3P2.24	R00848	sn-Glycerol-3-phosphate:(acceptor) 2-oxidoreductase	1.1.5.3	Glycerol metabolism (creates FAD for TCA cycle)
PF08_0045	R00621	NO NAME	1.2.4.2	TCA cycle (right side, via succinate)
	R03316	NO NAME	1.2.4.2	
PF10_0334	R00408	Succinate:(acceptor) oxidoreductase	1.3.99.1	TCA cycle (right side, via succinate)
	R00432	Succinate:CoA ligase (GDP-forming)	6.2.1.4	
PF13_0070	R00621	NO NAME	1.2.4.2	TCA cycle (right side, via succinate)
	R03316	NO NAME	1.2.4.2	
PF13_0120	R02570	succinyl-CoA:enzyme N6-(dihydrolipoyl)lysine S-succinyltransferase	2.3.1.61	TCA cycle (right side, via succinate)
PF13_0121	R02570	succinyl-CoA:enzyme N6-(dihydrolipoyl)lysine S-succinyltransferase	2.3.1.61	TCA cycle (right side, via succinate)
PF14_0295	R00405	Succinate:CoA ligase (ADP-forming)	6.2.1.5	TCA cycle (right side, via succinate)
PF11340w	R01082	(S)-malate hydro-lyase (fumarate-forming)	4.2.1.2	TCA cycle (right side, via succinate)
PFL0630w	R00408	Succinate:(acceptor) oxidoreductase	1.3.99.1	TCA cycle (right side, via succinate)

Table 32 : Details on the 15 interesting double-deletion genes whose pair-wise deletions are lethal or growth-reducing (see figure 110) but whose deletions singly are not lethal.

Conclusions

The final model contains more information than would be useful to present in this thesis. I have taken great care to ensure that questions that future researchers may wish to pose of the model can be asked easily and the answers understood. A number of further findings are,

- D-lactate is not produced by the model unless forced to but the measured level as presented in Vander Jagt *et al.*⁹⁴ contributes to the carbon balance of the model.
- Glycerol is produced by the model but not at the rate I have measured unless forced. The consumption of FAD in the succinate half of the branched TCA cycle is met by the production of glycerol and its efflux from the parasite. See additional note on the next page.
- Forcing glycerol and D-lactate efflux, combined with the level of carbohydrate accumulation from Chavali's¹⁰⁸ *L. major* metabolic model and the level of fatty acid modification in Mi-ichi *et al.*¹⁰⁴ creates a carbon-balanced model if the L-lactate efflux is raised slightly from the best estimate of 56 mmole/gDW/hour to 60 mmole/gDW/hour.
- L-malate is produced only if the allowed lactate efflux is not sufficient to achieve carbon balance. Flux-variability analysis (FVA) shows that any value of L-malate efflux between 0 and the maximum allowed can form part of a solution that give the optimal growth-rate so both my experimental results and those in Olszewski *et al.*⁹⁵ are in agreement with my model.
- Aspartate is similarly variable within the optimal solution, strongly supporting the view presented in Olszewski *et al.*⁸⁸ that L-malate and aspartate exchanges are balanced by the parasite to achieve carbon balance. It is likely that the different culture conditions used by me and Olszewski *et al.*⁹⁵ led me not to measure L-malate production where they did.
- The predicted flow of glucose around the pentose-phosphate cycle matches measurements in the literature and the predicted flow to nucleotides is close to measurements in the literature.
- The ATP maintenance cost that gives the measured growth-rate is in keeping with both Chavali *et al.*¹⁰⁸'s *L. major* and Plata *et al.*⁵⁸'s *P. falciparum* model.
- My model predicts that — over the complete life cycle — between 31% and 100% of the host RBC's haemoglobin can be digested whilst remaining within 95% of the optimal growth-rate. The predicted percentage of the digested amino acids used by the parasite at these two extremes is 40% and 13% respectively. At the maximum 65% digestion suggested in Krugliak *et al.*¹⁰³ my FBA model predicts that the

parasite uses 21% of the available amino acids compared to the 16% measured in that paper: an excellent agreement within the uncertainties of the model and the measurements.

- For the optimal solution of my model, corresponding to digestion of 31% of the RBC's haemoglobin, the amino acid exchanges predicted by my model are consistent are within 10% of those measured by NMR metabolomics: an excellent agreement.
- The branched TCA-cycle reported in Olszewski *et al.*⁹⁵ is the optimal flux solution although FVA shows that the behaviour of the cycle can vary considerably and still produce a growth-rate of 99% of optimal.

The complexities of glycerol metabolism

With respect to the second of my conclusions on the previous page a clarification is required. The role of protein complexes is not always elegantly contained within the KEGG LIGAND ontology and my network only contains the first of the reactions performed by the succinate dehydrogenase (SDH) enzyme complex (Complex II) shown in figure 111. The FADH_2 produced by the conversion of succinate to fumarate is retained within the enzyme complex and is oxidised back to FAD by the reduction of Coenzyme Q_{10} (Q) (ubiquinone) to ubiquinol (QH_2), both of which are free within the inner mitochondrial space — a compartment which, for simplicity and due to the limited function of the mitochondria in *P. falciparum*, I omit from my model.

The mitochondrial glycerol-3-phosphate dehydrogenase (GPDH-M) complex works similarly via Coenzyme Q_{10} but in the opposite direction and whilst the two enzyme complexes cannot in reality exchange FAD or FADH_2 directly as they do my model, they can do so in effect via the intermediate Coenzyme Q_{10} . This means that the simplification in my model of allowing FAD and FADH_2 to be exchanged between the SDH complex and the GPDH-M complex is unrealistic but also unlikely to have any effect on the predictions of my model.

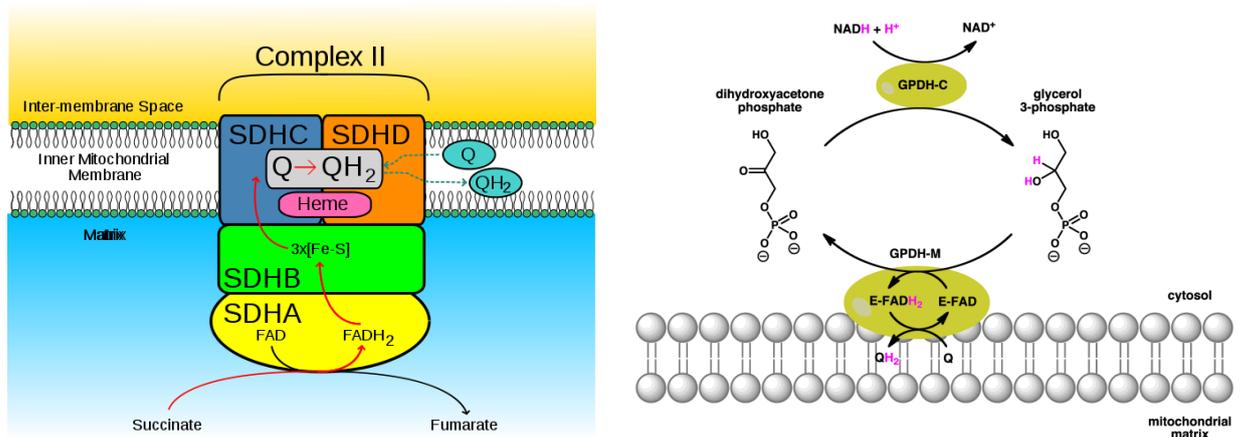


Figure 111 : The succinate dehydrogenase (SDH) complex (left) and the mitochondrial glycerol-3-phosphate dehydrogenase (GPDH-M) complex (right) cannot share FAD/FADH₂ since neither ever leaves the enzyme complex. Both enzyme complexes restore their FAD/FADH₂ balance via external Coenzyme Q₁₀ (Q/QH₂) in the mitochondrial membrane space. Both images from Wikipedia (user: Johnhfst. user: Boghog2).

The functioning of these complexes is additionally interesting because of the unusual role of Coenzyme Q₁₀ in *P. falciparum* — as reported on page 89 of Painter *et al*¹¹⁹.

“erythrocytic stages of the human malaria parasite Plasmodium falciparum seem to maintain an active mitochondrial electron transport chain to serve just one [essential] metabolic function: regeneration of ubiquinone [Coenzyme Q₁₀] required as the electron acceptor for dihydroorotate dehydrogenase, an essential enzyme for pyrimidine biosynthesis”.

Within my model, the action of dihydroorotate dehydrogenase (DHODH) is simplified to completely remove the complexity of including the electron transport chain (ETC) in my model. I have justified this because the flux through the ETC is tiny compared to that in an aerobically respiring organism. The DHODH reaction in my model instead represents the combined role of the ETC/ubiquinone interaction with a conversion of water to hydrogen peroxide simulating the ETC’s creation of superoxide ions leading to oxidative stress.

The Coenzyme Q₁₀ that is regenerated by *P. falciparum*’s ETC could be used to drive the succinate dehydrogenase enzyme complex in the citrate cycle but there are good reasons why this might not be the case. Painter *et al*¹¹⁹. state that,

“maintenance of mitochondrial membrane potential, ... was essential in these parasites”

and that,

“maintenance of electropotential across the mitochondrial inner membrane is a critical function of the electron transport chain”.

Typically electropotential is maintained largely by ATP synthase and yet despite the recent discovery of a *P. falciparum* ATP synthase²⁰⁴ it remains unclear what function it serves given that any ATP derived in this way has not been observed to play any role in the energy metabolism of the parasite in the RBC life stages.

These results further convince me that the unusual — and disputed — metabolic fate of glycerol in *P. falciparum* plays an important role in explaining the role of the mitochondria in RBC metabolism. Returning to Lu-Yun Lian *et al.*⁴⁶'s findings about glycerol.

“it is hypothesized ... that glycerol production by the malaria parasite is the result of a metabolic adaptation to growth in O₂-limited (and CO₂ elevated) conditions by the operation of a glycerol-3-phosphate shuttle for the re-oxidation of assimilatory NADH.”

The results of my model and experimental results agree with this hypothesis but make a key addition. Since the glycerol-3-phosphate shuttle in this instance is unbalanced — with excess glycerol exported rather than used to re-oxidise NADH — thus maintaining mitochondrial membrane potential, it might be not be wholly useful to refer to it as a shuttle, but rather a partially branched shuttle.

Final comparison of my final model to similar models

At the end of this chapter I want to make one final comparison between by model, the two other leading malaria metabolic models and the model and work which has served as my example for much of this thesis: Arvind Chavali's *Leishmania major* metabolic model¹⁰⁸. I already referred to the problem with orphaned reactions in the Huthmacher *et al.*⁵⁹ and Plata *et al.*⁵⁸ models on page 72 but I want now to focus on another two problems: unconserved metabolites and reproducibility of modelling results.

Unconserved metabolites

Consider two reactions, $A \rightleftharpoons B + C$ and $A \rightleftharpoons B$. Individually these reactions may seem reasonable but together they allow the infinite creation, or destruction, of the metabolite C. In this example we refer to C as an unconserved metabolite. The mathematical solution to this problem is to assign compound C a mass of zero but in more complex loops a solution to similar problems can sometimes be found by assigning to the same compound a negative mass in some reaction and a positive mass in other reactions. Indeed, the documentation of SurreyFBA⁸⁷ describes unconserved metabolites as, “those whose molecular masses cannot be simultaneously assigned positive values”. A more formal definition of an unconserved metabolite, including this basic example and algorithms for detecting them within networks, is found in Nikolaev²⁰⁵ with significant development in Gevorgyan⁸⁹.

The importance of unconserved metabolites is that whilst mathematical solutions to these problems exist it is never biochemically acceptable to create or destroy mass.

Unconserved metabolites are almost always a symptom of stoichiometric inconsistencies within a metabolic network.

In light of the impact unconserved metabolites can have on the accuracy of a model I used SurreyFBA⁸⁷ to search for them in the three models of interest.

As a result of the balanced reactions checker in MetNetMaker and my visual model assembly process my final model contains no unconserved metabolites. In comparison the Huthmacher *et al.*⁵⁹ *P. falciparum* model has 7 unconserved metabolites and the Chavali *et al.*¹⁰⁸ *L. major* model has 319 unconserved metabolites. The Plata *et al.*⁵⁸ *P. falciparum* model could not even be loaded into SurreyFBA and checked for unconserved metabolites*.

* I tried searching for unconserved metabolites by first manually parsing the model and SurreyFBA reported none but it is unclear whether the manually parsed model was analysed correctly.

Reproducibility of results

A second area for comparison between models is on the reproducibility of results. Since all models are constraints-based reconstructions designed for metabolic analysis using at least FBA it is reasonable to expect that opening the models in commonly used FBA software and running a simple simulation to maximise biomass production would yield a result. I tried this with all three models.

The Huthmacher *et al.*⁵⁹ model cannot produce biomass in SurreyFBA⁸⁷ or COBRA v1.3.3²⁰³. Constraints are not included in the model and since the analysis techniques used to create the published results are implemented in custom software that is not distributed they are almost impossible to reproduce. There are a huge number of reactions and no visual model, which — combined with the almost unreproducible computational results — makes the model less useful than it could be.

The Plata *et al.*⁵⁸ model produces biomass in COBRA v1.3.3²⁰³ but cannot load into SurreyFBA⁸⁷. The model seems to contain no constraints which may limit the biological accuracy of its predictions. The very large number of reactions and lack of a visual model makes it hard to examine this network.

The Chavali *et al.*¹⁰⁸ model cannot produce biomass in SurreyFBA or COBRA 1.3.3²⁰³. The visual model is excellent and is clearly produced in Simpheny which may provide excellent analysis and visualisation tools*. In speaking with Arvind Chavali at ICSB 2010 in Edinburgh he mentioned that he was continuing to use his model but that a considerable amount of information had not made it across from Simpheny to the SBML format model he published.

* I am told that the software has matured since then but seems effectively restricted to former PhD students of Bernhard Palsson.

Summary and future directions

... in 10 human metabolic pathway databases, none of the descriptions (of TCA) is entirely correct and consensus exists on only 3 reactions.

Stobbe et al.²⁰⁶ via Ben Heaver*

I have drawn conclusions on my results in each chapter and this final section does not repeat them. Instead I look at some of the larger issues and recent advances in systems biology and how these have affected and could improve my work. It is likely that many of the opinions in these final few pages will turn out to be wrong.

Whilst all the ideas are my own they have often come about from discussions, largely on twitter, with people including, but not limited to, Fiona Achcar at the University of Glasgow, Natalie Stanford at the Humboldt University of Berlin, Christian Priesnitz at the University of Saarbrücken, Nicolas le Novère at the European Bioinformatics Institute, Cambridge, Mark Hucka at the California Institute of Technology, Ben Heaver at the Institute for Systems Biology, Seattle and Pedro Mendes, Neil Swainston and Kieran Smallbone all at the University of Manchester. I find twitter increasingly valuable for both engaging in, and observing at a distance, discussions between systems biologists all around the world.

* Postdoctoral fellow at the Institute for Systems Biology (Price group), Seattle

The utility of my model

I have been stunned by the agreement between the predictions and my model and experimentally measured knowledge of *P. falciparum* metabolism. That internal fluxes, amino acid exchange fluxes and growth-rate are all accurately reproduced by a relatively simple model analysed by a simple technique still amazes me. This is true even though the technique has been shown to work in many other organisms.

In both cases where my experimental measurements disagree with literature — on malate efflux and haemoglobin digestion rate — the model shows that differences that seem considerable are found in areas of high variability within the modelled metabolism of *P. falciparum*. It is very possible that even where my measured results and previously published results do not agree, they are both correct; they certainly both agree with my model of malaria metabolism.

There's a lot more that can be learned from my model as long as the limit of its utility is appreciated. I am hopeful that the work of the ParaMet consortium* will build on my model by using Carbon-13 fluxomics — directly measuring more internal fluxes to refine and confirm the predictions of the model — and by adding a model of gene activity and regulation similar to that reported in Chandrasekaran and Price⁷⁵. A promising avenue for further research is the possibility of building upon my preliminary studies of the metabolic response to known inhibitors of *P. falciparum* that I reported on page 168 with respect to atovaquone inhibition. An application for funding which I wrote to continue these metabolomics studies and improve their integration with an ever improving metabolic model is included as appendix IX and gives some further details on this.

The limits of existing models

The observation from Stobbe *et al.*²⁰⁶ on the previous page is an excellent condensation in 136 characters of the greatest weakness and the greatest strength of current tools in metabolic systems biology. Existing metabolic models are often excellent for their designed purpose but poor in other situations and frequently in complete disagreement with other models of the same system built for a slightly different purpose.

I am not convinced that enough care is being taken with existing models to explore the limits of their usefulness and I hope that by ensuring that a visualisation of each reaction, connection and connection within my network is provided, an idea of my network's limits can also be appreciated. Much more rigorous work on the limitations of models and the

* project descriptions at <http://bit.ly/S0v7HE>

impact of uncertainty on the conclusions that can be drawn from their analysis is being done by Fiona Achcar²⁰⁷ at the University of Glasgow.

Standards beyond SBML, ensuring models can be reused

SBML has been a great achievement in systems biology, allowing the wide interchange and re-use of metabolic models. But as I have shown in this thesis it is a step forward rather than a solution to the problems of model re-use. The goal of the MIRIAM²⁵ (Minimum Information Required In The Annotation of Models) project and the associated identifiers.org²⁰⁸ project is to allow ontologies for reactions, compounds and simulation parameters of biological models to be defined such that the models can be re-used and their results reproduced. In the words of the Wikipedia page, *“MIRIAM is a community-level effort to standardize the annotation and curation processes of quantitative models of biological systems.”*

My model is not MIRIAM compliant and as shown in the following tweets the worthy goal of MIRIAM is, as yet, unrealised.

“MIRIAM support is really hard. I just annotated a large model and what was hard was to find the sources to annotate.”

“Good #sbml annotation has to be done carefully, and the ontologies were far from optimal: lots of elements missing.”

“Software tools (mine included) have a long way to go to really help with MIRIAM annotation... many cannot even do it.”

Tweets from Pedro Mendes to Mike Hucka and Neil Swainston in 2012 on the subject of model re-use.

Tools for FBA and FVA are unreliable; higher level analysis is hard to trust

One of the most worrying findings in this PhD has been that basic tools that large amounts of published work are based upon are not reliable. The basic techniques of FBA and FVA are well-established and a simulation run on one piece of software should give the same result on another piece of software. This is not the case!

The development of my final model took place largely in combination with analyses conducted in SurreyFBA⁸⁷, with constraints added to the model to improve the results of FBA in that software. SurreyFBA is an excellent piece of software in that it can be run either from the command line or from an excellent small user interface called JyMet. It has a further considerable advantage of being free so I could install and run it easily on any

computer. The final flux-variability analysis and gene deletion predictions were performed in the COBRA toolbox v1.3.3 as these types of analysis were difficult to perform in SurreyFBA.

It was by using two pieces of software on the same model that I noticed some problems. As shown in figure 106 a large number of fluxes can vary and still give the optimal solution and there is no reason that two different programs should give the same solution to a problem as long as all fluxes lie within the allowed variabilities. The problem is that the COBRA toolbox v1.3.3 actually predicts a slightly (<2%) higher growth-rate than SurreyFBA when analysing the same model. Upon closer inspection the model was able to grow without any flux through the purine salvage pathways, a prediction that was clearly incorrect. The complexity of installing the COBRA toolbox v1.3.3 and its dependencies was such* that I have not tried using a newer version of the COBRA toolbox to see if this problem still exists. In discussions with members of John Pinney's group at Imperial College London I have heard that they too have got strange results from the COBRA toolbox, this time with COBRApy v2.0.5. In this case the exact same problem would give a different optimal growth-rate each time it was solved.

My point in raising these concerns is not to criticise existing software but to re-iterate a feeling I have had throughout my PhD. Systems biology is already very complex and advances like the *Mycoplasma genitalium* whole-cell model²⁰⁹ signal a future where currently separate models of different processes in organisms are joined together. Since errors in biological simulations have a tendency to multiply rather than fade away when scaled up and joined together the success of these systems may hinge on these small details.

For this reason I have made considerable efforts — as discussed frequently in this thesis — to ensure that my model is re-usable and that the predictions it makes are both reproducible and, through visualisation, easy to understand and play with. I have also tried hard to get some idea of the uncertainties in many of the measurements I have made to ensure that the conclusions I have drawn are reasonable.

Sharing ideas as well as models

When I started my PhD the Wikipedia pages on computational biology didn't exist and those on malaria were largely limited to the human impact rather than the science of the parasite. Since then these great introductory resources have improved markedly, in part thanks to the efforts of people like Alex Bateman at the Sanger Institute who have outlined

* I spent nearly 4 months trying to get Matlab installed, compiling and installing dependencies and testing the final installation before I could perform any simulations.

areas for improvement and encouraged researchers to engage with the platform. Pages on metabolic engineering, isotopic labelling and synthetic biology have all been created and improved in the last 18 months and are now valuable starting points for students and cross-disciplinary scientist like me to start exploring new fields. On a personal level I am extremely happy that the page I wrote on flux-balance analysis is viewed at least 120 times a month compared to 30 times per month before I rewrote it and that is now available in three languages. I am very glad that the complete dismissal of tools for online learning and sharing of knowledge that I experienced six years ago as an undergraduate has given way to an appreciation of their advantages and limitations.

Better metabolic visualisation

I've shown how combining mathematical and visual approaches to network reconstruction can help avoid errors that are hard to see when a reconstruction is performed predominantly in one of those two ways. I hope I have also caused the reader to pose the same questions that I have in trying to understand the behaviour of these networks. I am convinced that the simplification of metabolism to the pathway metaphor in biochemistry risks obscuring the highly connected nature of metabolism but I am also convinced that software tools could solve that problem and let scientists see and play with the complex interactions within metabolic networks. I have made efforts to solve this problem with my MetNetMaker software and my visual metabolic reconstruction but I think that outside help from the world of design will be needed to really solve the problem. I have tried to get this assistance via synthetic aesthetics, without success, and via the Wellcome Trust, with more success.

Synthetic aesthetics

In 2010 I submitted a letter of interest for funding and support from the Synthetic Aesthetics project led by the University of Edinburgh and Stanford University whereby scientists would be paired with designers to try and solve similar problems. I include the full letter of interest, which includes my ideas on how we might fix these problems, as appendix IX to this thesis but three sentences of the letter provide a good summary,

In the exciting and rapidly progressing new fields of synthetic biology and systems biology people are saving time by using old network design principles, often taken from electronic engineering, rather than thinking about the real requirements of the problems they face. My fear is that by reusing these existing design tools in the world of biology we risk obscuring a truer understanding of biological systems. To put it simply, I think that we are drawing biological networks just like we draw electronic networks not because the two are similar but because we haven't taken the time to design a better way of doing things.

Extract from my Letter of Interest for Synthetic Aesthetics²¹⁰ funding

Simalaria

Although my application for synthetic aesthetics funding was unsuccessful many of the concepts were re-used in an application for funding from the Wellcome Trust which was successful. This short project put me in touch with a Bristol based games development company called MobilePie to develop a game based on my PhD research. The game is a scientifically accurate representation of flux-balance analysis where the user must balance energy creation via glycolysis and glucose consumption — and the associated energy cost — of DNA, Fat and Protein synthesis by the control of flux gates shown as multi-coloured nodes in figure 112. In two days we were unable to create more than a basic representation of the malaria parasite but I'm very pleased with the game's role of proving that many of my ideas for better metabolic network visualisation are achievable.

The best way to understand SIMALARIA is to play it, currently at <http://media.mobilepie.com/simalaria/simalaria.html> (currently the second hit on google for 'simalaria') and playable through a web browser on Mac and Windows.

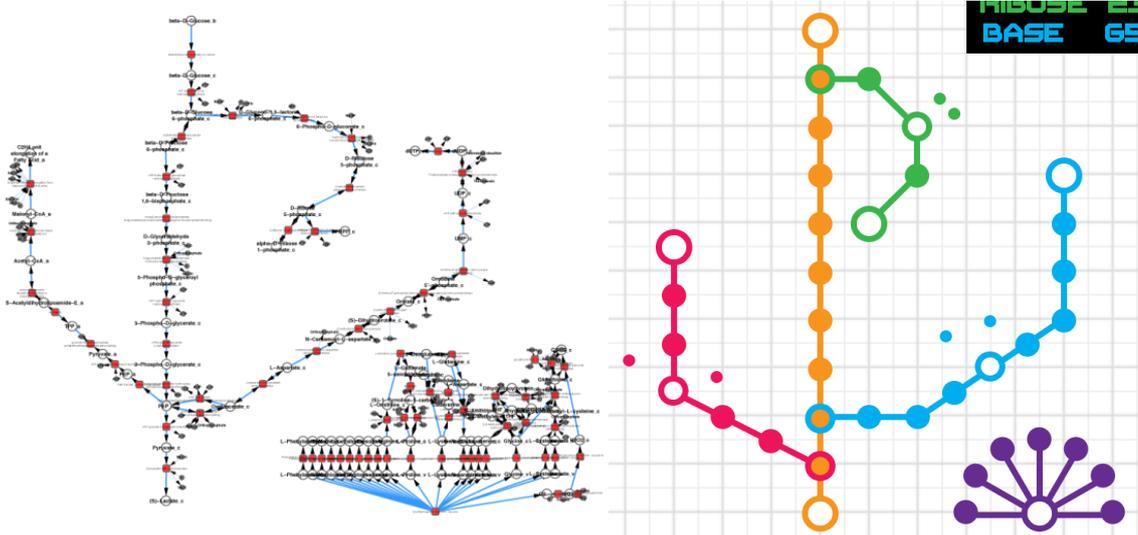


Figure 112 : (left) glycolysis, oxidative pentose-phosphate pathway, purine base synthesis and fatty acid elongation in the apicoplast in a simplified version of the complete final model. Haemoglobin digestion to scavenge amino acids is shown separately in the bottom-right. (right) the same model in SIMALARIA, glycolysis in orange, ribose sugar synthesis via the oxidative pentose phosphate cycle in green, nucleotide base synthesis via the purine synthesis pathway in light blue, fatty acid elongation in pink. Haemoglobin digestion requires no glucose flux but consumes energy in the game.

References

1. Hirsch, A. M. *Brief History of the Discovery of Nitrogen-fixing Organisms Rhizobia Actinomycetes*. In *Situ* 1–5 (2009).at <<http://www.mcdb.ucla.edu/Research/Hirsch/imagesb/HistoryDiscoveryN2fixingOrganisms.pdf>>
2. Peregrin-Alvarez, J. M., Tsoka, S. & Ouzounis, C. A. The phylogenetic extent of metabolic enzymes and pathways. *Genome research* **13**, 422–7 (2003).
3. Caspi, R. MetaCyc: a multiorganism database of metabolic pathways and enzymes. *Nucleic Acids Research* **34**, D511–D516 (2006).
4. Goto, S., Okuno, Y., Hattori, M., Nishioka, T. & Kanehisa, M. LIGAND: database of chemical compounds and reactions in biological pathways. *Nucleic acids research* **30**, 402–4 (2002).
5. Ogata, H. *et al.* KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic acids research* **27**, 29–34 (1999).
6. Karp, P. D. BioCyc / Pathway Tools Tutorial: September 22-23, 2008 London, England. (2008).at <<http://brg.ai.sri.com/ptools/tutorial/2008-09-London.html>>
7. Yeh, I., Hanekamp, T., Tsoka, S., Karp, P. D. & Altman, R. B. Computational analysis of Plasmodium falciparum metabolism: organizing genomic information to facilitate drug discovery. *Genome research* **14**, 917–24 (2004).
8. Ginsberg, H. Malaria Parasite Metabolic Pathways (<http://sites.huji.ac.il/malaria/>). (2007).at <<http://sites.huji.ac.il/malaria/>>
9. Zucker, J. Lessons learned from the Genome-scale metabolic reconstruction and curation of Neurospora crassa. *National e-Science Centre meeting Edinburgh, Theme 14 Workshop 3: Metabolic modelling and networks* at <<http://www.nesc.ac.uk/action/esi/download.cfm?index=4826>>
10. Ginsburg, H. Progress in in silico functional genomics: the malaria Metabolic Pathways database. *Trends in parasitology* **22**, 238–40 (2006).
11. De Matos, P. *et al.* Chemical Entities of Biological Interest: an update. *Nucleic acids research* **38**, D249–54 (2010).
12. Herrgard, M. J. *et al.* A consensus yeast metabolic network reconstruction obtained from a community approach to systems biology. *Nature Biotechnology* **26**, 1155–1160 (2008).
13. NC-IUBMB & Webb, E. C. *Enzyme Nomenclature 1992: Recommendations of the Nomenclature Committee of the International Union of Biochemistry and Molecular Biology on the Nomenclature and Classification of Enzymes*. 862 (Academic Press: 1992).
14. Forth, T. Framework Video on Tom Forth's Ph.D project page. at <<http://www.tomforth.co.uk/phd/project>>
15. Events/Workshops/The 1st Workshop on Software Platforms for Systems Biology - SBML.org. at <http://sbml.org/Events/Workshops/The_1st_Workshop_on_Software_Platforms_for_Systems_Biology>
16. Hucka, M. The systems biology markup language (SBML): a medium for representation and exchange of biochemical network models. *Bioinformatics* **19**, 524–531 (2003).
17. The BioPAX workgroup BioPAX – Biological Pathways Exchange Language Level 1, Version 1.0 Documentation. *Structure* **39** (2004).
18. Hermjakob, H. *et al.* The HUPO PSI's molecular interaction format—a community standard for the representation of protein interaction data. *Nature biotechnology* **22**, 177–83 (2004).
19. Strömbäck, L. & Lambrix, P. Representations of molecular pathways: an evaluation of SBML, PSI MI and BioPAX. *Bioinformatics (Oxford, England)* **21**, 4401–7 (2005).
20. Demir, E. *et al.* The BioPAX community standard for pathway data sharing. *Nature Biotechnology* **28**, 935–942 (2010).
21. Hucka, M., Keating, S. M. & Wilkinson, D. J. The Systems Biology Markup Language (SBML): Language Specification for Level 3 Version 1 Core. (2009).
22. Hucka, M. *et al.* Systems Biology Markup Language (SBML) Level 2: Structures and Facilities for Model Definitions. *Nature Precedings* (2008).doi:10.1038/npre.2008.2715.1
23. Hucka, M. *et al.* Evolving a lingua franca and associated software infrastructure for computational systems biology: the Systems Biology Markup Language (SBML) project. *Systems Biology* **1**, 41 (2004).
24. Novere, N. Le, Courtot, M. & Laibe, C. Adding Semantics in Kinetics Models of Biochemical Pathways. *Proceedings of the 2nd International ESCEC Symposium on Experimental Standard Conditions on Enzyme Characterizations*. 137–154 (2007).
25. Le Novère, N. *et al.* Minimum information requested in the annotation of biochemical models (MIRIAM). *Nature biotechnology* **23**, 1509–15 (2005).
26. Novère, N. Le *et al.* The Systems Biology Graphical Notation. *Nature Biotechnology* **27**, 735–742 (2009).
27. Andrew Spielman Malaria. *Vega Science Trust* (2004).at <<http://vega.org.uk/video/programme/87>>
28. Knirsch, D. G. H. *Parasitic Diseases, Fifth Edition*. 375 (Apple Trees Productions, LLC: 2005).
29. Despommier, D. D. & Karapelou, J. W. *Parasite Life Cycles*. 127 (Springer: 1987).
30. D'Antonio, M. & Spielman, D. A. *Mosquito: The Story of Man's Deadliest Foe*. 256 (Faber and Faber: 2002).
31. Fernando, S. D., Rodrigo, C. & Rajapakse, S. The "hidden" burden of malaria: cognitive impairment following infection. *Malaria journal* **9**, 366 (2010).

32. E.Wellems, T. Plasmodium Chloroquine Resistance and the Search for a Replacement Antimalarial Drug. *Science* **298**, 124–126 (2002).
33. Towie, N. Malaria breakthrough raises spectre of drug resistance. *Nature* **440**, 852–853 (2006).
34. Fleischer, B. Editorial: 100 years ago: Giemsa's solution for staining of plasmodia. *Tropical medicine & international health : TM & IH* **9**, 755–6 (2004).
35. Howard, R. J. Secretion of a malarial histidine-rich protein (Pf HRP II) from Plasmodium falciparum-infected erythrocytes. *The Journal of Cell Biology* **103**, 1269–1277 (1986).
36. McKenzie, F. E., Ferreira, M. U., Baird, J. K., Snounou, G. & Bossert, W. H. Meiotic recombination, cross-reactivity, and persistence in Plasmodium falciparum. *Evolution; international journal of organic evolution* **55**, 1299–307 (2001).
37. Wiser, M. Supplementary material to Malaria (TRMD 782). Mark F. Wiser at Tulane University (1999).at <http://www.tulane.edu/~wiser/malaria/>
38. Trager, W. & Jensen, J. Human malaria parasites in continuous culture. *Science* **193**, 673–675 (1976).
39. Jensen, J. B. & Trager, W. Plasmodium falciparum in culture: use of outdated erythrocytes and description of the candle jar method. *The Journal of Parasitology* **63**, 883–886 (1977).
40. Cranmer, S. L., Magowan, C., Liang, J., Coppel, R. L. & Cooke, B. M. An alternative to serum for cultivation of Plasmodium falciparum in vitro. *Transactions of the Royal Society of Tropical Medicine and Hygiene* **91**, 363–5 (1995).
41. Millet, P., Collins, W. E., Fisk, T. L. & Nguyen-Dinh, P. In vitro cultivation of exoerythrocytic stages of the human malaria parasite Plasmodium malariae. *The American journal of tropical medicine and hygiene* **38**, 470–3 (1988).
42. Hollingdale, M. R., Collins, W. E., Campbell, C. C. & Schwartz, A. L. In vitro culture of two populations (dividing and nondividing) of exoerythrocytic parasites of Plasmodium vivax. *The American journal of tropical medicine and hygiene* **34**, 216–22 (1985).
43. Sattabongkot, J. *et al.* Establishment of a human hepatocyte line that supports in vitro development of the exo-erythrocytic stages of the malaria parasites Plasmodium falciparum and P. vivax. *The American journal of tropical medicine and hygiene* **74**, 708–15 (2006).
44. McRobert, L. & Mcconkey, G. A. RNA interference (RNAi) inhibits growth of Plasmodium falciparum. *Mol Biochem Parasitol* **119**, 273–278 (2002).
45. Baum, J. *et al.* Molecular genetics and comparative genomics reveal RNAi is not functional in malaria parasites. *Nucleic acids research* **37**, 3788–98 (2009).
46. Lian, L.-Y. *et al.* Glycerol: An unexpected major metabolite of energy metabolism by the human malaria parasite. *Malaria Journal* **8**, 38 (2009).
47. Gardner, M. J. *et al.* Genome sequence of the human malaria parasite Plasmodium falciparum. *Nature* **419**, 498–511 (2002).
48. Palsson, B. O. *Systems Biology: Properties of Reconstructed Networks*. 334 (Cambridge University Press: Cambridge, 2006).
49. Varma, A. & Palsson, B. Metabolic Flux Balancing: Basic Concepts, Scientific and Practical Use. *Nature Biotechnology* **12**, 994–998 (1994).
50. Edwards, J. S., Covert, M. & Palsson, B. Metabolic modelling of microbes: the flux-balance approach. *Environmental Microbiology* **4**, 133–140 (2002).
51. Kauffman, K. J., Prakash, P. & Edwards, J. S. Advances in flux balance analysis. *Current Opinion in Biotechnology* **14**, 491–496 (2003).
52. Lee, J. M., Gianchandani, E. P. & Papin, J. A. Flux balance analysis in the era of metabolomics. *Briefings in bioinformatics* **7**, 140–50 (2006).
53. Orth, J. D., Thiele, I. & Palsson, B. Ø. What is flux balance analysis? *Nature Biotechnology* **28**, 245–248 (2010).
54. Pruitt, K. D., Tatusova, T., Klimke, W. & Maglott, D. R. NCBI Reference Sequences: current status, policy and new initiatives. *Nucleic acids research* **37**, D32–6 (2009).
55. Hyland, C., Pinney, J. W., Mcconkey, G. A. & Westhead, D. R. metaSHARK: a WWW platform for interactive exploration of metabolic networks. *Nucleic Acids Res* **34**, 0–8 (2006).
56. Aziz, R. K. *et al.* The RAST Server: rapid annotations using subsystems technology. *BMC genomics* **9**, 75 (2008).
57. Poolman, M. G., Assmus, H. E. & Fell, D. a Applications of metabolic modelling to plant metabolism. *Journal of experimental botany* **55**, 1177–86 (2004).
58. Plata, G., Hsiao, T.-L., Olszewski, K. L., Llinás, M. & Vitkup, D. Reconstruction and flux-balance analysis of the Plasmodium falciparum metabolic network. *Molecular Systems Biology* **6**, (2010).
59. Huthmacher, C., Hoppe, A., Bulik, S. & Holzhütter, H.-G. Antimalarial drug targets in Plasmodium falciparum predicted by stage-specific metabolic network analysis. *BMC Systems Biology* **4**, 120 (2010).
60. Wiback, S. J. & Palsson, B. O. Extreme Pathway Analysis of Human Red Blood Cell Metabolism. *Biophysical Journal* **83**, 808–818 (2002).
61. Hoffmann, S., Hoppe, A. & Holzhütter, H.-G. Composition of metabolic flux distributions by functionally interpretable minimal flux modes (MinModes). *Genome informatics. International Conference on Genome Informatics* **17**, 195–207 (2006).
62. Hoops, S. *et al.* COPASI—a Complex PATHway Simulator. *Bioinformatics (Oxford, England)* **22**, 3067–74 (2006).
63. Kotte, O., Zaugg, J. B. & Heinemann, M. Bacterial adaptation through distributed sensing of metabolic fluxes. *Molecular systems biology* **6**, 355 (2010).
64. Mahadevan, R. & Schilling, C. H. The effects of alternate optimal solutions in constraint-based genome-scale metabolic models. *Metabolic Engineering* **5**, 264–276 (2003).

65. Varma, a & Palsson, B. O. Stoichiometric flux balance models quantitatively predict growth and metabolic by-product secretion in wild-type *Escherichia coli* W3110. *Applied and environmental microbiology* **60**, 3724–31 (1994).
66. Mahadevan, R., Edwards, J. S. & Doyle, F. J. Dynamic flux balance analysis of diauxic growth in *Escherichia coli*. *Biophysical Journal* **83**, 1331–40 (2002).
67. Carey, N. *The Epigenetics Revolution: How Modern Biology is Rewriting Our Understanding of Genetics, Disease and Inheritance*. 320 (Icon Books Ltd: 2012).
68. Hu, G. *et al.* Transcriptional profiling of growth perturbations of the human malaria parasite *Plasmodium falciparum*. *Nature biotechnology* **28**, (2009).
69. Patankar, S., Munasinghe, A., Shoaibi, A., Cummings, L. M. & Wirth, D. F. Serial analysis of gene expression in *Plasmodium falciparum* reveals the global expression profile of erythrocytic stages and the presence of anti-sense transcripts in the malarial parasite. *Molecular biology of the cell* **12**, 3114–25 (2001).
70. Le Roch, K. G. Global analysis of transcript and protein levels across the *Plasmodium falciparum* life cycle. *Genome Research* **14**, 2308–2318 (2004).
71. Florens, L. *et al.* A proteomic view of the *Plasmodium falciparum* life cycle. *Nature* **419**, 520–526 (2002).
72. Hayward, R. E. *et al.* Shotgun DNA microarrays and stage-specific gene expression in *Plasmodium falciparum* malaria. *Molecular microbiology* **35**, 6–14 (2000).
73. Nirmalan, N., Sims, P. F. G. & Hyde, J. E. Quantitative proteomics of the human malaria parasite *Plasmodium falciparum* and its application to studies of development and inhibition. *Molecular Microbiology* **52**, 1187–1199 (2004).
74. Bozdech, Z. *et al.* The transcriptome of the intraerythrocytic developmental cycle of *Plasmodium falciparum*. *PLoS biology* **1**, E5 (2003).
75. Chandrasekaran, S. & Price, N. D. Probabilistic integrative modeling of genome-scale metabolic and regulatory networks in *Escherichia coli* and *Mycobacterium tuberculosis*. *Proceedings of the National Academy of Sciences of the United States of America* **107**, 17845–50 (2010).
76. Kotte, O., Zaugg, J. B. & Heinemann, M. Bacterial adaptation through distributed sensing of metabolic fluxes. *Molecular systems biology* **6**, 355 (2010).
77. Pinney, J. W., Shirley, M. W., Mcconkey, G. A. & Westhead, D. R. metaSHARK: software for automated metabolic network prediction from DNA sequence and its application to the genomes of *Plasmodium falciparum* and *Eimeria tenella*. *Nucleic Acids Res* **33**, 1399–1409 (2005).
78. Thiele, I. & Palsson, B. Ø. A protocol for generating a high-quality genome-scale metabolic reconstruction. *Nature protocols* **5**, 93–121 (2010).
79. Schilling, C. H., Thakar, R., Travnik, E., Dien, S. Van & Wiback, S. SimPheny™: A Computational Infrastructure for Systems Biology. *US Department of Energy, Genomic Science Program publications* (2005).at <<http://citesex.ist.psu.edu/viewdoc/summary?doi=10.1.1.107.8797>>
80. Karp, P. D., Paley, S. & Romero, P. The Pathway Tools software. *Bioinformatics* **18**, 225–232 (2002).
81. Funahashi, A. *et al.* CellDesigner 3.5: A Versatile Modeling Tool for Biochemical Networks. *Proceedings of the IEEE* **96**, 1254–1265 (2008).
82. Matsuoka, Y., Ghosh, S., Kikuchi, N. & Kitano, H. Payao: a community platform for SBML pathway model curation. *Bioinformatics (Oxford, England)* **26**, 1381–3 (2010).
83. Forth, T., McConkey, G. A. & Westhead, D. R. MetNetMaker: a free and open-source tool for the creation of novel metabolic networks in SBML format. *Bioinformatics (Oxford, England)* **26**, 2352–3 (2010).
84. Funahashi, A., Jouraku, A. & Kitano, H. Converting the KEGG Pathway Database to SBML. *8th Annual International Conference on Research in Computational Molecular Biology (RECOMB 2004)* at <<http://sbml.org/images/d/df/Kegg2sbmlRECOMB2004.pdf>>
85. Gille, C., Hübner, K., Hoppe, A. & Holzhütter, H.-G. METANNOGEN: Annotation of biological reaction networks. *Bioinformatics (Oxford, England)* **1**, 2007–2008 (2011).
86. Smoot, M. E., Ono, K., Ruscheinski, J., Wang, P.-L. & Ideker, T. Cytoscape 2.8: new features for data integration and network visualization. *Bioinformatics (Oxford, England)* **27**, 431–2 (2011).
87. Gevorgyan, A., Bushell, M. E., Avignone-Rossa, C. & Kierzek, A. M. SurreyFBA: a command line tool and graphics user interface for constraint-based modeling of genome-scale metabolic reaction networks. *Bioinformatics (Oxford, England)* **27**, 433–4 (2011).
88. Olszewski, K. L. & Llinás, M. Central carbon metabolism of *Plasmodium* parasites. *Molecular and biochemical parasitology* **175**, 95–103 (2011).
89. Gevorgyan, A., Poolman, M. G. & Fell, D. A. . Detection of stoichiometric inconsistencies in biomolecular models. *Bioinformatics* **24**, 2245–2251 (2008).
90. Pinney, J. W., Starkovich, J. M., Li, P. & Robertson, D. L. SHARKview: a tool for the visualization of systems biology data. *BMC Systems Biology* **1**, P33 (2007).
91. KEGG KEGG global map. at <<http://www.kegg.jp/kegg/atlas/?01100>>
92. Schomburg, I. *et al.* BRENDA, the enzyme database: updates and major new developments. *Nucleic acids research* **32**, D431–3 (2004).
93. Aurrecoechea, C. *et al.* PlasmoDB: a functional genomic database for malaria parasites. *Database* **37**, 539–543 (2009).
94. Jagt, D. L. Vander, Hunsaker, L. A., Campos, N. M. & Baack, B. R. D-Lactate production in erythrocytes infected with *Plasmodium falciparum*. *Molecular and Biochemical Parasitology* **42**, 277–284 (1990).
95. Olszewski, K. L. *et al.* Branched tricarboxylic acid metabolism in *Plasmodium falciparum*. *Nature* **466**, 774–778 (2010).
96. Vander Jagt, D. L., Hunsaker, L. a, Campos, N. M. & Baack, B. R. D-lactate production in erythrocytes infected with *Plasmodium falciparum*. *Molecular and biochemical parasitology* **42**, 277–84 (1990).

97. Van Dooren, G. G. *et al.* Development of the endoplasmic reticulum, mitochondrion and apicoplast during the asexual life cycle of *Plasmodium falciparum*. *Molecular microbiology* **57**, 405–19 (2005).
98. Ralph, S. A. *et al.* Metabolic maps and functions of the *Plasmodium falciparum* apicoplast. *Nature Reviews Microbiology* **2**, 203–216 (2004).
99. Jayabalasingham, B., Menard, R. & Fidock, D. a Recent insights into fatty acid acquisition and metabolism in malarial parasites. *F1000 biology reports* **2**, 2–5 (2010).
100. Liu, J., Istvan, E. S., Gluzman, I. Y., Gross, J. & Goldberg, D. E. *Plasmodium falciparum* ensures its amino acid supply with multiple acquisition pathways and redundant proteolytic enzyme systems. *Proceedings of the National Academy of Sciences of the United States of America* **103**, 8840–5 (2006).
101. Geary, T. G., Divo, A. A. & Jensen, J. B. Nutritional Requirements of *Plasmodium falciparum* in Culture. II. Effects of Antimetabolites in a Semi Defined Medium. *The Journal of Eukaryotic Microbiology* **32**, 65–69 (1985).
102. Divo, A. A., Geary, T. G., Davis, N. L. & Jensen, J. B. Nutritional Requirements of *Plasmodium falciparum* in Culture. I. Exogenously Supplied Dialyzable Components Necessary for Continuous Growth. *The Journal of Eukaryotic Microbiology* **32**, 59–64 (1985).
103. Krugliak, M., Zhang, J. & Ginsburg, H. Intraerythrocytic *Plasmodium falciparum* utilizes only a fraction of the amino acids derived from the digestion of host cell cytosol for the biosynthesis of its proteins. *Molecular and Biochemical Parasitology* **119**, 249–256 (2002).
104. Mi-Ichi, F., Kita, K. & Mitamura, T. Intraerythrocytic *Plasmodium falciparum* utilize a broad range of serum-derived fatty acids with limited modification for their growth. *Parasitology* **133**, 399–410 (2006).
105. König, E. Salvage syntheses and their relationship to nucleic acid metabolism. *Bulletin of the World Health Organization* **55**, 249–52 (1977).
106. Sherman, I. W. Biochemistry of *Plasmodium* (Malarial Parasites). *Microbiological Reviews* **43**, 453–495 (1979).
107. Maria-Pia Abbraccio, Jean-Marie Boeynaems, José L. Boyer, Geoffrey Burnstock, Stefania Ceruti, Marta Fumagalli, Christian Gachet, Rebecca Hills, Robert G. Humphries, Kenneth A. Jacobson, Charles Kennedy, Brian F. King, Davide Lecca, Maria Teresa Miras-Po, G. A. W. P2Y receptors: P2Y2. *IUPHAR database (IUPHAR-DB)* (2012).at <<http://www.iuphar-db.org/DATABASE/ObjectDisplayForward?objectId=324>>
108. Chavali, A. K., Whittemore, J. D., Eddy, J. A., Williams, K. T. & Papin, J. A. Systems analysis of metabolism in the pathogenic trypanosomatid *Leishmania major*. *Molecular Systems Biology* **4**, 1–19 (2008).
109. Olszewski, K. L. *et al.* Host-parasite interactions revealed by *Plasmodium falciparum* metabolomics. *Cell host & microbe* **5**, 191–9 (2009).
110. Liu, J., Istvan, E. S., Gluzman, I. Y., Gross, J. & Goldberg, D. E. *Plasmodium falciparum* ensures its amino acid supply with multiple acquisition pathways and redundant proteolytic enzyme systems. *Proceedings of the National Academy of Sciences* **103**, 8840–8845 (2006).
111. Baldwin, S. A., Mcconkey, G. A., Cass, C. E. & Young, J. D. Nucleoside Transport as a Potential Target for Chemotherapy in Malaria. *Current Pharmaceutical Design* **13**, 569–580 (2007).
112. Kirk, K. & Saliba, K. J. Targeting nutrient uptake mechanisms in *Plasmodium*. *Curr Drug Targets* **8**, 75–88 (2007).
113. Kanaani, J. & Ginsburg, H. Metabolic interconnection between the human malarial parasite *Plasmodium falciparum* and its host erythrocyte. Regulation of ATP levels by means of an adenylate translocator and adenylate kinase. *J Biol Chem* **264**, 3194–3199 (1989).
114. Joet, T. & Krishna, S. The hexose transporter of *Plasmodium falciparum* is a worthy drug target. *Acta Trop* **89**, 371–374 (2004).
115. Pouvelle, B. *et al.* Direct access to serum macromolecules by intraerythrocytic malaria parasites. *Nature* **353**, 73–75 (1991).
116. Martin, R. E., Henry, R. I., Abbey, J. L., Clements, J. D. & Kirk, K. The “permeome” of the malaria parasite: an overview of the membrane transport proteins of *Plasmodium falciparum*. *Genome biology* **6**, R26 (2005).
117. Kanaani, J. & Ginsburg, H. Metabolic Interconnection between the Human Malarial Parasite *Plasmodium falciparum* and Its Host Erythrocyte. *The Journal of Biological Chemistry* **264**, 3194–3199 (1989).
118. Kanehisa, M., Goto, S., Sato, Y., Furumichi, M. & Tanabe, M. KEGG for integration and interpretation of large-scale molecular data sets. *Nucleic acids research* **40**, D109–14 (2012).
119. Painter, H. J., Morrissey, J. M., Mather, M. W. & Vaidya, A. B. Specific role of mitochondrial electron transport in blood-stage *Plasmodium falciparum*. *Nature* **446**, 88–91 (2007).
120. Francis, S. E., Sullivan, D. J. & Goldberg, D. E. Hemoglobin metabolism in the malaria parasite *Plasmodium falciparum*. *Annual review of microbiology* **51**, 97–123 (1997).
121. Pishchany, G. & Skaar, E. P. Taste for blood: hemoglobin as a nutrient source for pathogens. *PLoS pathogens* **8**, e1002535 (2012).
122. Lew, V. L., Tiffert, T. & Ginsburg, H. Excess hemoglobin digestion and the osmotic stability of *Plasmodium falciparum*-infected red blood cells. *Analysis* **101**, 4189–4194 (2003).
123. Elliott, D. A. *et al.* Four distinct pathways of hemoglobin uptake in the malaria parasite *Plasmodium falciparum*. *Proceedings of the National Academy of Sciences* **105**, 2463–2468 (2008).
124. Ledbetter, S. The Unconventional Amino Acid Starvation Response of the Malaria Parasite, *Plasmodium falciparum*. (2012).at <<http://openscholarship.wustl.edu/etd/708/>>
125. Coppens, I., Sullivan, D. J. & Prigge, S. T. An update on the rapid advances in malaria parasite cell biology. *Trends in parasitology* **26**, 305–10 (2010).
126. Olliaro, P. Mode of action and mechanisms of resistance for antimalarial drugs. *Pharmacology & therapeutics* **89**, 207–19 (2001).

127. Atamna, H. & Ginsburg, H. Origin of reactive oxygen species in erythrocytes infected with *Plasmodium falciparum*. *Molecular and Biochemical Parasitology* **61**, 231–241 (1993).
128. Ringwald, P., Meche, F. S., Bickii, J. & Basco, L. K. In vitro culture and drug sensitivity assay of *Plasmodium falciparum* with nonserum substitute and acute-phase sera. *Journal of clinical microbiology* **37**, 700–5 (1999).
129. Preechapornkul, P. *et al.* Optimizing the culture of *Plasmodium falciparum* in hollow fiber bioreactors. *The Southeast Asian journal of tropical medicine and public health* **41**, 761–9 (2010).
130. Lewis, I. A. *et al.* Method for determining molar concentrations of metabolites in complex solutions from two-dimensional ¹H-¹³C NMR spectra. *Analytical chemistry* **79**, 9385–90 (2007).
131. Mehta, M., Sonawat, H. M. & Sharma, S. Malaria parasite-infected erythrocytes inhibit glucose utilization in uninfected red cells. *FEBS letters* **579**, 6151–8 (2005).
132. Geary, T. G., Divo, a a, Bonanni, L. C. & Jensen, J. B. Nutritional requirements of *Plasmodium falciparum* in culture. III. Further observations on essential nutrients and antimetabolites. *The Journal of protozoology* **32**, 608–13 (1985).
133. Alphonse Laveran - Biography. *nobelprize.org* at <http://www.nobelprize.org/nobel_prizes/medicine/laureates/1907/laveran-bio.html>
134. Thomas, C. G. . *Medical Microbiology*. 435 (Baillière Tindall: London, 1964).
135. Coatney, G. R. *The primate malarias*. (Division of Parasitic Diseases, Centers for Disease Control and Prevention.: Atlanta, GA, USA, 1971).
136. Images from the History of Medicine (NLM) - What: *Plasmodium* Who: Nicholson, Gertrude H., illustrator. at <<http://ihm.nlm.nih.gov/luna/servlet/view/all/who/Nicholson,+Gertrude+H.,+illustrator./what/Plasmodium/>>
137. O'Donnell, A. J., Schneider, P., McWatters, H. G. & Reece, S. E. Fitness costs of disrupting circadian rhythms in malaria parasites. *Proceedings. Biological sciences / The Royal Society* **278**, 2429–36 (2011).
138. Hoppe, H. C., Verschoor, J. A. & Louw, A. I. *Plasmodium falciparum*: A comparison of synchronisation methods for in vitro cultures. *Experimental Parasitology* **72**, 464–467 (1991).
139. Naughton, J. A. & Bell, A. Studies on cell-cycle synchronization in the asexual erythrocytic stages of *Plasmodium falciparum*. *Parasitology* **134**, 331–7 (2007).
140. Lambros, C. & Vanderberg, J. P. Synchronization of *Plasmodium falciparum* erythrocytic stages in culture. *The Journal of parasitology* **65**, 418–20 (1979).
141. Cowman, A. F. & Crabb, B. S. Invasion of red blood cells by malaria parasites. *Cell* **124**, 755–66 (2006).
142. Ahn, S.-Y. *et al.* Magnetic separation: a highly effective method for synchronization of cultured erythrocytic *Plasmodium falciparum*. *Parasitology research* **102**, 1195–200 (2008).
143. Ashong, J. O. O., Blench, P., Warhurst, D. C. & Blench, I. P. The composition of haemozoin from *Plasmodium falciparum*. *Transactions of the Royal Society of Tropical Medicine and Hygiene* **83**, 167–172 (1989).
144. Opi, D. H. Determination of the biomass composition of *plasmodium falciparum* to aid in the construction of a genome-scale metabolic model. 45 (2008).
145. Lake, J. Metabolic flux during growth of malaria parasite: towards a model of parasite metabolism. (2010).
146. Teng, R. *et al.* Metabolite profiling of the intraerythrocytic malaria parasite *Plasmodium falciparum* by (¹H) NMR spectroscopy. *NMR in biomedicine* **22**, 292–302 (2009).
147. Zakutansky, S. Metabolic implications of atovaquone inhibition in *Plasmodium falciparum* as determined by HPLC and NMR analysis. (2011).
148. Ma, C. Fluorescent-based diagnostic techniques and inhibitor testing for anti-malaria development. (2009).
149. Edwards, J. S., Ibarra, R. U. & Palsson, B. O. In silico predictions of *Escherichia coli* metabolic capabilities are consistent with experimental data. *Nature Biotechnology* **19**, 125–130 (2001).
150. Edwards, J. S. & Palsson, B. Metabolic flux balance analysis and the in silico analysis of *Escherichia coli* K-12 gene deletions. *BMC Bioinformatics* (2000).doi:10.1186/1471-2105/1/1
151. McConkey, G. A., Ittarat, I., Meshnick, S. R. & McCutchan, T. F. Auxotrophs of *Plasmodium falciparum* Dependent on p-Aminobenzoic Acid for Growth. *Proceedings of the National Academy of Sciences* **91**, 4244–4248 (1994).
152. Zolg, J. W., Macleod, A. J., Dickson, I. H. & Scaife, J. G. Modifications of the In vitro Culture Conditions Improving Parasitic Yields. *The Journal of Parasitology* **68**, 1072–1080 (1982).
153. Shapiro, H. M. & Mandy, F. Cytometry in malaria: moving beyond Giemsa. *Cytometry. Part A : the journal of the International Society for Analytical Cytology* **71**, 643–5 (2007).
154. Vianen, P. H. Van, Ponnudurai, T. & Overdulve, J. P. *Plasmodium* Species: Flow Cytometry and Microfluorometry Assessments of DNA Content and Synthesis. *Experimental Parasitology* **64**, 88–94 (1987).
155. McCarthy, L. R. & Senne, J. E. Evaluation of acridine orange stain for detection of microorganisms in blood cultures. *Journal of clinical microbiology* **11**, 281–5 (1980).
156. Bhakdi, S. C. *et al.* Re-Evaluating acridine orange for rapid flow cytometric enumeration of parasitemia in malaria-infected rodents. *Cytometry Part A* **71A**, 662–667 (2007).
157. Jiménez-Díaz, M. B. *et al.* Quantitative measurement of *plasmodium* -infected erythrocytes in murine models of malaria by flow cytometry using bidimensional assessment of syto-16 fluorescence. *Cytometry Part A* **75A**, 225–235 (2009).

158. Grimberg, B. T., Erickson, J. J., Sramkoski, R. M., Jacobberger, J. W. & Zimmerman, P. a Monitoring Plasmodium falciparum growth and development by UV flow cytometry using an optimized Hoechst-thiazole orange staining strategy. *Cytometry. Part A : the journal of the International Society for Analytical Cytology* **73**, 546–54 (2008).
159. Keeley, A. & Soldati, D. The glideosome: a molecular machine powering motility and host-cell invasion by Apicomplexa. *Trends in cell biology* **14**, 528–32 (2004).
160. Geary, T. G., Divo, A. A., Bonanni, L. C. & Jensen, J. B. Nutritional requirements of Plasmodium falciparum in culture. III. Further observations on essential nutrients and antimetabolites. *The Journal of protozoology* **32**, 608–13 (1985).
161. Quashie, N. B., Ranford-Cartwright, L. C. & De Koning, H. P. Uptake of purines in Plasmodium falciparum-infected human erythrocytes is mostly mediated by the human equilibrative nucleoside transporter and the human facilitative nucleobase transporter. *Malaria journal* **9**, 36 (2010).
162. Kirk, K., Howitt, S. M., Bröer, S., Saliba, K. J. & Downie, M. J. Purine uptake in Plasmodium: transport versus metabolism. *Trends in parasitology* **25**, 246–9 (2009).
163. Schuetz, R., Kuepfer, L. & Sauer, U. Systematic evaluation of objective functions for predicting intracellular fluxes in Escherichia coli. *Molecular systems biology* **3**, 119 (2007).
164. Robertson, B., Button, D. & Koch, A. Determination of the biomasses of small bacteria at low concentrations in a mixture of species with forward light scatter measurements by flow cytometry. *Applied and environmental microbiology* **64**, 3900–9 (1998).
165. Karpinets, T. V., Greenwood, D. J., Sams, C. E. & Ammons, J. T. RNA:protein ratio of the unicellular organism as a characteristic of phosphorous and nitrogen stoichiometry and of the cellular requirement of ribosomes for protein synthesis. *BMC biology* **4**, 30 (2006).
166. Kinoshita, A., Nakayama, Y. & Tomita, M. Towards Simulation of Whole Metabolic Pathways in Human Erythrocyte. *Science And Technology* **313**, 312–313 (2001).
167. Atamna, H., Pascarmona, G. & Ginsburg, H. Hexose-monophosphate shunt activity in intact Plasmodium falciparum-infected erythrocytes and in free parasites. *Molecular and Biochemical Parasitology* **67**, 79–89 (1994).
168. Atamna, H. & Ginsburg, H. The malaria parasite supplies glutathione to its host cell--investigation of glutathione transport and metabolism in human erythrocytes infected with Plasmodium falciparum. *European journal of biochemistry / FEBS* **250**, 670–9 (1997).
169. Palacpac, N. M. Q. Developmental-stage-specific triacylglycerol biosynthesis, degradation and trafficking as lipid bodies in Plasmodium falciparum-infected erythrocytes. *Journal of Cell Science* **117**, 1469–1480 (2004).
170. Elliott, J. L., Saliba, K. J. & Kirk, K. Transport of lactate and pyruvate in the intraerythrocytic malaria parasite, Plasmodium falciparum. *Biochem J* **355**, 733–739 (2001).
171. Taylor, J. R. *Introduction To Error Analysis: The Study of Uncertainties in Physical Measurements*. 448 (University Science Books: 1997).
172. Stadler, A. M. et al. Hemoglobin dynamics in red blood cells: correlation to body temperature. *Biophysical journal* **95**, 5449–61 (2008).
173. Kaushansky, K. et al. *Williams Hematology, Eighth Edition*. 2460 (McGraw-Hill Professional: 2010).
174. Förster, J., Famili, I., Fu, P., Palsson, B. Ø. & Nielsen, J. Genome-scale reconstruction of the Saccharomyces cerevisiae metabolic network. *Genome research* **13**, 244–53 (2003).
175. Mysliwski, A. & Korczak, A. Increase of size and dry mass of human erythrocytes depending on age of donors. *Mechanisms of Ageing and Development* **34**, 111–115 (1986).
176. Chanda, I., Pan, A. & Dutta, C. Proteome Composition in Plasmodium falciparum: Higher Usage of GC-Rich Nonsynonymous Codons in Highly Expressed Genes. *Journal of Molecular Evolution* **61**, 513–523 (2005).
177. Vezza, A. C. & Trager, W. Preliminary characterization of the major RNA species from Plasmodium falciparum. *Molecular and biochemical parasitology* **4**, 149–62 (1981).
178. Nicholson, J. K., Foxall, P. J., Spraul, M., Farrant, R. D. & Lindon, J. C. 750 MHz 1H and 1H-13C NMR spectroscopy of human blood plasma. *Analytical chemistry* **67**, 793–811 (1995).
179. Wishart, D. S. et al. HMDB: a knowledgebase for the human metabolome. *Nucleic acids research* **37**, D603–10 (2009).
180. Cui, Q. et al. Metabolite identification via the Madison Metabolomics Consortium Database. *Nature biotechnology* **26**, 162–4 (2008).
181. Wishart, D. S. Quantitative metabolomics using NMR. *TrAC Trends in Analytical Chemistry* **27**, 228–237 (2008).
182. Fairbanks, V. F. et al. Measurement of blood volume and red cell mass: re-examination of 51Cr and 125I methods. *Blood cells, molecules & diseases* **22**, 169–86; discussion 186a–186g (1996).
183. Scheibel, L. & Pflaum, W. Carbohydrate metabolism in Plasmodium knowlesi. *Comparative Biochemistry and Physiology* **37**, 543–553 (1970).
184. Ali, S. N. & Fletcher, K. A. Carbohydrate metabolism of malarial parasites--I. Metabolism of lactate in Plasmodium knowlesi infected monkey erythrocytes. *Comparative biochemistry and physiology. B, Comparative biochemistry* **80**, 725–9 (1985).
185. Massimi, M. et al. Effects of resveratrol on HepG2 cells as revealed by (1)H-NMR based metabolic profiling. *Biochimica et biophysica acta* **1820**, 1–8 (2012).
186. Pfaller, M., Krogstad, D., Parquette, A. & Nguyen-Dinh, P. Plasmodium falciparum: Stage-specific lactate production in synchronized cultures. *Experimental Parasitology* **54**, 391–396 (1982).
187. Consortium, T. U. Reorganizing the protein space at the Universal Protein Resource (UniProt). *Nucleic acids research* **40**, D71–5 (2012).

188. Elford, B. C., Haynes, J. D., Chulay, J. D. & Wilson, R. J. M. Selective stage-specific changes in the permeability to small hydrophilic solutes of human erythrocytes infected with *Plasmodium falciparum*. *Molecular and Biochemical Parasitology* **16**, 43–60 (1985).
189. Martin, R. E. & Kirk, K. Transport of the essential nutrient isoleucine in human erythrocytes infected with the malaria parasite *Plasmodium falciparum*. *Blood* **109**, 2217–24 (2007).
190. Abrahamsen, M. S. *et al.* Complete genome sequence of the apicomplexan, *Cryptosporidium parvum*. *Science (New York, N.Y.)* **304**, 441–5 (2004).
191. Jackson, A. a, Gibson, N. R., Lu, Y. & Jahoor, F. Synthesis of erythrocyte glutathione in healthy adults consuming the safe amount of dietary protein. *The American journal of clinical nutrition* **80**, 101–7 (2004).
192. Dawson, R. B. Blood storage XXII. Improvement in red blood cell 2,3-DPG levels at six weeks by 20 mM PO₄ in CPD-adenine-inosine. *Transfusion* **16**, 450–4 (1977).
193. Dawson, R. Blood storage XXIV: red blood cell 2,3-DPG and ATP maintenance for six weeks in CPD-adenine with higher phosphate, pyruvate, and dihydroxyacetone. *Transfusion* **17**, 242–247 (1977).
194. Watanabe, S., Yamamoto, R., Ogata, M. & Murakami, T. Adenosine triphosphate restoration and discocytic transformation of stored human erythrocytes. *Acta medica Okayama* **39**, 239–46 (1985).
195. Ralton, J. E. *et al.* Evidence that intracellular beta1-2 mannan is a virulence factor in *Leishmania* parasites. *The Journal of biological chemistry* **278**, 40757–63 (2003).
196. Tarun, A. S., Vaughan, A. M. & Kappe, S. H. I. Redefining the role of de novo fatty acid synthesis in *Plasmodium* parasites. *Trends in parasitology* **25**, 545–50 (2009).
197. Quashie, N. B. *et al.* A comprehensive model of purine uptake by the malaria parasite *Plasmodium falciparum*: identification of four purine transport activities in intraerythrocytic parasites. *The Biochemical journal* **411**, 287–95 (2008).
198. Cassera, M. B. *et al.* Erythrocytic adenosine monophosphate as an alternative purine source in *Plasmodium falciparum*. *The Journal of biological chemistry* **283**, 32889–99 (2008).
199. Henry, C. S., Broadbelt, L. J. & Hatzimanikatis, V. Thermodynamics-based metabolic flux analysis. *Biophysical journal* **92**, 1792–805 (2007).
200. Feist, A. M. *et al.* A genome-scale metabolic reconstruction for *Escherichia coli* K-12 MG1655 that accounts for 1260 ORFs and thermodynamic information. *Molecular systems biology* **3**, 121 (2007).
201. Choi, I. & Mikkelsen, R. B. *Plasmodium falciparum*: ATP/ADP transport across the parasitophorous vacuolar and plasma membranes. *Experimental parasitology* **71**, 452–62 (1990).
202. Fry, M., Webb, E. & Pudney, M. Effect of mitochondrial inhibitors on adenosinetriphosphate levels in *plasmodium falciparum*. *Comparative Biochemistry & Physiology* **96**, 775–782 (1990).
203. Becker, S. A. *et al.* Quantitative prediction of cellular metabolism with constraint-based models: the COBRA Toolbox. *Nature protocols* **2**, 727–38 (2007).
204. Balabaskaran Nina, P. *et al.* ATP synthase complex of *Plasmodium falciparum*: dimeric assembly in mitochondrial membranes and resistance to genetic disruption. *The Journal of biological chemistry* **286**, 41312–22 (2011).
205. Nikolaev, E. V., Burgard, A. P. & Maranas, C. D. Elucidation and structural analysis of conserved pools for genome-scale metabolic reconstructions. *Biophysical journal* **88**, 37–49 (2005).
206. Stobbe, M. D., Houten, S. M., Van Kampen, A. H. C., Wanders, R. J. A. & Moerland, P. D. Improving the description of metabolic networks: the TCA cycle as example. *FASEB journal : official publication of the Federation of American Societies for Experimental Biology* **26**, 3625–36 (2012).
207. Achcar, F., Kerkhoven, E. J., Bakker, B. M., Barrett, M. P. & Breitling, R. Dynamic modelling under uncertainty: the case of *Trypanosoma brucei* energy metabolism. *PLoS computational biology* **8**, e1002352 (2012).
208. Juty, N., Le Novère, N. & Laibe, C. Identifiers.org and MIRIAM Registry: community resources to provide persistent identification. *Nucleic acids research* **40**, D580–6 (2012).
209. Karr, J. R. *et al.* A Whole-Cell Computational Model Predicts Phenotype from Genotype. *Cell* **150**, 389–401 (2012).
210. Synthetic Aesthetics. at <<http://www.syntheticaesthetics.org/>>
211. Invitrogen RPMI 1640 Formulation. *Product Technical Resources* at <http://www.invitrogen.com/site/us/en/home/support/Product-Technical-Resources/media_formulation.188.html>
212. Levett, D. Z. *et al.* The role of nitrogen oxides in human adaptation to hypoxia. *Scientific Reports* **1**, 1–8 (2011).
213. Chiwakata, C. B., Hemmer, C. J. & Dietrich, M. High levels of inducible nitric oxide synthase mRNA are associated with increased monocyte counts in blood and have a beneficial role in *Plasmodium falciparum* malaria. *Infection and immunity* **68**, 394–9 (2000).
214. Carlson, J., Helmby, H., Wahlgren, M. & Hill, A. Human cerebral malaria: association with erythrocyte rosetting and lack of anti-rosetting antibodies. *The lancet* **336**, 1457–1460 (1990).
215. Adams, S., Brown, H. & Turner, G. Breaking down the blood-brain barrier: signaling a path to cerebral malaria? *Trends in parasitology* **18**, 360–6 (2002).

Appendices

Appendix I : Dead-end compounds algorithm and implementation

The algorithm behind the functioning of the dead-end compounds button can be expressed either as a diagram as in figure 116 or as code as in figure 113. As mentioned in chapter 2 : MetNetMaker the SQL queries are run and the query results output to files and displayed on screen using VBA code. The full code of the ExComps() subroutine that does this is included as figure 115 with a simplified version in figure 114 more suitable for understanding how the code works.

ReactantsQuery

```
SELECT DISTINCT [REACTIONS-PRODUCTS LINK].[Compound ID], [Selected Reactions].Compartment
FROM [Selected Reactions] , [REACTIONS-PRODUCTS LINK]
WHERE [Selected Reactions].[Reaction ID] = [REACTIONS-PRODUCTS LINK].[Reaction ID]
AND [Selected Reactions].Rev=False
UNION
SELECT DISTINCT [REACTIONS-REACTANTS LINK].[Compound ID], [Selected Reactions].Compartment
FROM [Selected Reactions] , [REACTIONS-REACTANTS LINK]
WHERE [Selected Reactions].[Reaction ID] = [REACTIONS-REACTANTS LINK].[Reaction ID]
AND [Selected Reactions].Rev=True
```

ProductsQuery

```
SELECT DISTINCT [REACTIONS-PRODUCTS LINK].[Compound ID], [Selected Reactions].Compartment
FROM [Selected Reactions] , [REACTIONS-PRODUCTS LINK]
WHERE [Selected Reactions].[Reaction ID] = [REACTIONS-PRODUCTS LINK].[Reaction ID]
AND [Selected Reactions].Rev=True
UNION
SELECT DISTINCT [REACTIONS-REACTANTS LINK].[Compound ID], [Selected Reactions].Compartment
FROM [Selected Reactions] , [REACTIONS-REACTANTS LINK]
WHERE [Selected Reactions].[Reaction ID] = [REACTIONS-REACTANTS LINK].[Reaction ID]
AND [Selected Reactions].Rev=False
```

CompoundsInOnly

```
SELECT ReactantsQuery.[Compound ID], ReactantsQuery.Compartment
FROM ReactantsQuery
LEFT JOIN ProductsQuery ON ReactantsQuery.[Compound ID] = ProductsQuery.[Compound ID]
WHERE ProductsQuery.[Compound ID] Is Null
```

CompoundsOutOnly

```
SELECT ProductsQuery.[Compound ID], ProductsQuery.Compartment
FROM ProductsQuery
LEFT JOIN ReactantsQuery ON ProductsQuery.[Compound ID] = ReactantsQuery.[Compound ID]
WHERE ReactantsQuery.[Compound ID] Is Null
```

CompoundsInOutputQuery

```
SELECT CompoundsInOnly.[Compound ID], CompoundsInOnly.[Compartment],
[PREFERRED COMPOUND NAME].[Compound Name]
FROM CompoundsInOnly , [PREFERRED COMPOUND NAME]
WHERE CompoundsInOnly.[Compound ID] = [PREFERRED COMPOUND NAME].CompoundID
```

CompoundsOutOutputQuery

```
SELECT CompoundsOutOnly.[Compound ID], CompoundsOutOnly.[Compartment],
[PREFERRED COMPOUND NAME].[Compound Name]
FROM CompoundsOutOnly , [PREFERRED COMPOUND NAME]
WHERE CompoundsOutOnly.[Compound ID] = [PREFERRED COMPOUND NAME].CompoundID
```

Figure 113 : The six queries composing the SQL portion of the dead-end compounds algorithm. All queries refer either to each other or to tables in MetNetMaker's central database shown in figure 20. Nested SQL queries break the algorithm down into distinct levels as shown in figure 116.

Simplified Export Subroutine (no screen display, no formatting, no error handling)

```
Sub ExComps()  
    OutputPath = Application.CurrentProject.Path & "\excompstemp.txt"  
    Open OutputPath For Output As #5  
  
    Set ta20 = CurrentDb.OpenRecordset(CompoundsInOutputQuery, dbOpenSnapshot)  
    While Not ta20.EOF 'Compounds that only enter the network  
        Print #5, ta20.[Compound ID] & Chr(9) & ta20.[Compartment] & Chr(9) & ta20.[Compound Name]  
        ta20.MoveNext  
    Wend  
  
    Set ta21 = CurrentDb.OpenRecordset(CompoundsOutOutputQuery, dbOpenSnapshot)  
    While Not ta21.EOF 'Compounds that only exit the network  
        Print #5, ta21.[Compound ID] & Chr(9) & ta21.[Compartment] & Chr(9) & ta21.[Compound Name]  
        ta21.MoveNext  
    Wend  
  
    Close #5  
    VBA.Shell ("Notepad.exe " & Application.CurrentProject.Path & "\excompstemp.txt") 'Opens the file in notepad  
End Sub
```

Figure 114 : Simplified version of the VBA code to run the SQL queries in figure 113 and output the results to a text file which is then opened with Notepad.

Full Export Subroutine

```

Sub ExComps()
    Dim SQLstring As String
    Dim ta20 As Object
    Dim OutputPath As String
    Dim MessageText As String
    Dim CompartmentHolder As String
    Dim IDHolder As String

    Forms("Reaction Picker").Requery 'Refresh before

    MessageText = "This list may be truncated, a full copy has been opened in Notepad. Copy into a spreadsheet for correct formatting." + vbNewLine + vbNewLine

    OutputPath = Application.CurrentProject.Path & "\excompstemp.txt"
    Open OutputPath For Output As #5 'The numbers here shouldn't matter, chosen to avoid any chance of conflict
    SQLstring = "SELECT CompoundsInOnly.[Compound ID], CompoundsInOnly.[Compartment], [PREFERRED COMPOUND NAME].[Compound Name] FROM
    CompoundsInOnly INNER JOIN [PREFERRED COMPOUND NAME] ON CompoundsInOnly.[Compound ID] = [PREFERRED COMPOUND NAME].CompoundID;"
    Set ta20 = CurrentDb.OpenRecordset(SQLstring, dbOpenSnapshot)
    Print #5, "Compounds that only enter the network"
    MessageText = MessageText + "Compounds that only enter the network" + vbNewLine
    Print #5, "-----"
    MessageText = MessageText + "-----" + vbNewLine
    Print #5, ""
    MessageText = MessageText + "" + vbNewLine
    Print #5, "ID" & Chr(9) & "Compartment" & Chr(9) & "Name"
    MessageText = MessageText + "Compound ID" & Chr(9) & "Compartment" & Chr(9) & "Compound Name" + vbNewLine

    While Not ta20.EOF
        CompartmentHolder = ta20.[Compartment]
        IDHolder = ta20.[Compound ID]
        If Len(CompartmentHolder) < 10 Then 'Hack to double the number of tabs if compartment is short to make the output align
            CompartmentHolder = CompartmentHolder + Chr(9)
        End If
        If Len(IDHolder) < 10 Then
            IDHolder = IDHolder + Chr(9)
        End If
        Print #5, ta20.[Compound ID] & Chr(9) & ta20.[Compartment] & Chr(9) & ta20.[Compound Name]
        MessageText = MessageText + IDHolder & Chr(9) & CompartmentHolder & Chr(9) & ta20.[Compound Name] + vbNewLine
        ta20.MoveNext
    Wend

    Print #5, ""
    MessageText = MessageText + "" + vbNewLine
    SQLstring = "SELECT CompoundsOutOnly.[Compound ID], CompoundsOutOnly.[Compartment], [PREFERRED COMPOUND NAME].[Compound Name] FROM
    CompoundsOutOnly INNER JOIN [PREFERRED COMPOUND NAME] ON CompoundsOutOnly.[Compound ID] = [PREFERRED COMPOUND NAME].CompoundID;"
    Set ta20 = CurrentDb.OpenRecordset(SQLstring, dbOpenSnapshot)
    Print #5, "Compounds that only exit the network"
    MessageText = MessageText + "Compounds that only exit the network" + vbNewLine
    Print #5, "-----"
    MessageText = MessageText + "-----" + vbNewLine
    Print #5, ""
    MessageText = MessageText + "" + vbNewLine
    Print #5, "ID" & Chr(9) & "Compartment" & Chr(9) & "Name"
    MessageText = MessageText + "Compound ID" & Chr(9) & "Compartment" & Chr(9) & "Compound Name" + vbNewLine

    While Not ta20.EOF
        CompartmentHolder = ta20.[Compartment]
        IDHolder = ta20.[Compound ID]
        If Len(CompartmentHolder) < 10 Then 'Hack to double the number of tabs if compartment is short to make the output align
            CompartmentHolder = CompartmentHolder + Chr(9)
        End If
        If Len(IDHolder) < 10 Then
            IDHolder = IDHolder + Chr(9)
        End If
        Print #5, ta20.[Compound ID] & Chr(9) & ta20.[Compartment] & Chr(9) & ta20.[Compound Name]
        MessageText = MessageText + IDHolder & Chr(9) & CompartmentHolder & Chr(9) & ta20.[Compound Name] + vbNewLine
        ta20.MoveNext
    Wend

    MsgBox MessageText, vbDefaultButton1, "Dead-End Compounds"
    Close #5
    VBA.Shell ("Notepad.exe " & Application.CurrentProject.Path & "\excompstemp.txt") 'Opens the file in notepad
End Sub

```

Figure 115 : Full version of the VBA code to run the SQL queries in figure 113 and output the results to a text file which is then opened with Notepad and displays the results as an on-screen notification.

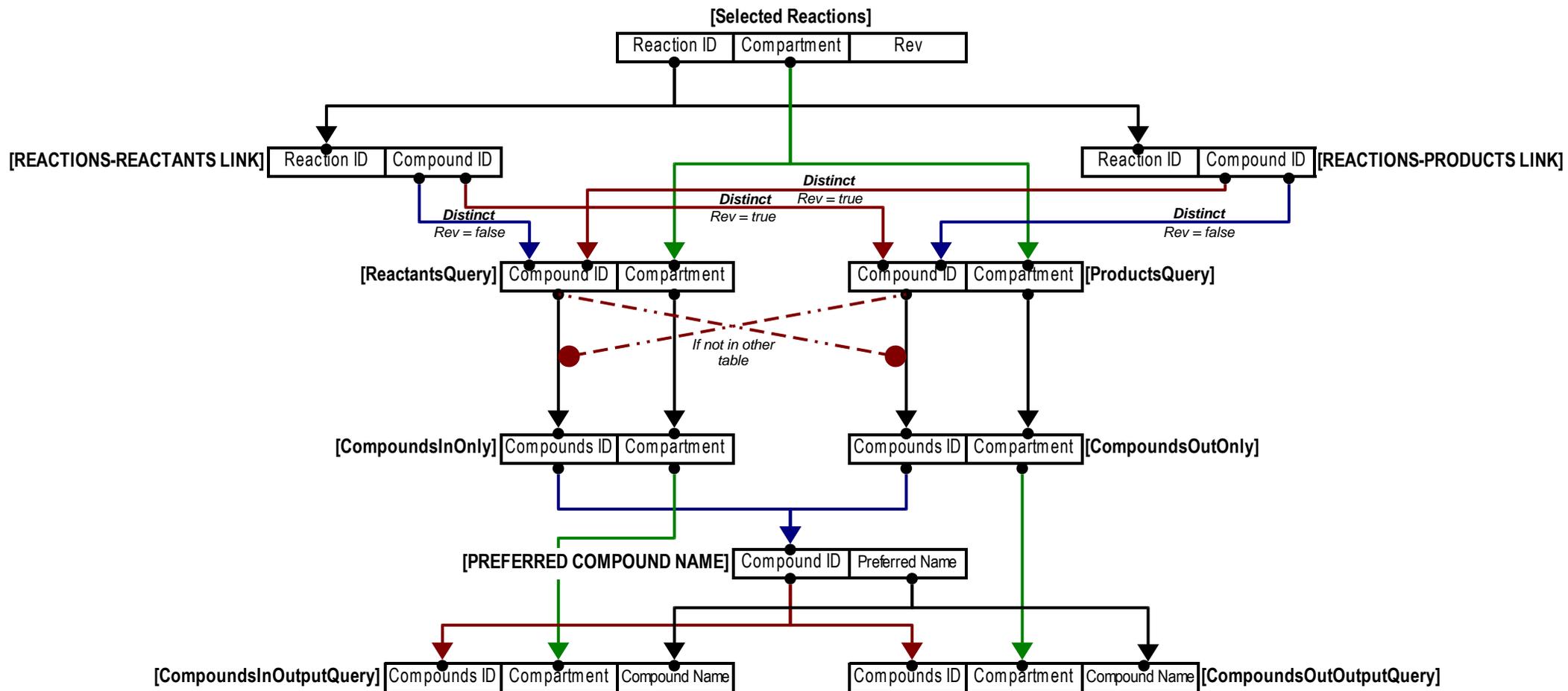


Figure 116 : Visual representation of the algorithm for returning dead-end compounds as represented as a set of SQL queries in figure 113.

Appendix II : Example reconstruction and evidence checker spreadsheet

Reaction Id	Compartment	EC Number	Max +ve Flux	Max -ve Flux	Objective	Map Number	Gene	Rev	SharkHunt	PlasmoDB	BRENDA	KEGG Direction	Notes
R00156	Cytosol	2.7.4.6	500	-500	FALSE	00240	PF13_0349, PFF0275c	FALSE	4.2E-58	yes	BRENDA	reversible	
R00158	Cytosol	2.7.4.14	500	-500	FALSE	00240	PFA0555c	FALSE	0.001	yes	BRENDA	reversible	
R00512	Cytosol	2.7.4.14	500	-500	FALSE	00240	PFA0555c	FALSE	0.001	yes	BRENDA	reversible	not in MPMP
R00570	Cytosol	2.7.4.6	500	-500	FALSE	00240	PF13_0349, PFF0275c	FALSE	4.2E-58	yes	BRENDA	reversible	
R00571	Cytosol	6.3.4.2	500	-500	FALSE	00240	PF14_0100	FALSE	0	yes	DIRECT EVIDENCE	forward	not in MPMP
R00573	Cytosol	6.3.4.2	500	-500	FALSE	00240	PF14_0100	FALSE	0	yes	DIRECT EVIDENCE	forward	
R00575_balanced	Cytosol	6.3.5.5	500	-500	FALSE	00240	PF13_0044	FALSE	0.0000019	yes	BRENDA		not in MPMP
R00965	Cytosol	4.1.1.23	500	-500	FALSE	00240	PF10_0225	FALSE	2.8E-14	yes	DIRECT EVIDENCE	forward	
R01665	Cytosol	2.7.4.14	500	-500	FALSE	00240	PFA0555c	FALSE	0.001	yes	BRENDA	forward	not in MPMP
R01867	Cytosol	1.3.3.1	500	-500	FALSE	00240	PFF0160c	FALSE	2E-88	yes	DIRECT EVIDENCE	reversible	
R01870	Cytosol	2.4.2.10	500	-500	FALSE	00240	PFE0630c	FALSE	5.5E-32	yes	DIRECT EVIDENCE	reversible	
R01993	Cytosol	3.5.2.3	500	-500	FALSE	00240	PF14_0697	FALSE	1.9E-45	yes	DIRECT EVIDENCE	reversible	
R02016	Cytosol	1.8.1.9	500	-500	FALSE	00240	PF11170c	FALSE	0.036	yes	DIRECT EVIDENCE	forward	
R02018	Cytosol	1.17.4.1	500	-500	FALSE	00240	PF10_0154, PF14_0053, PF14_0352	FALSE	0	yes	DIRECT EVIDENCE	forward	
R02024	Cytosol	1.17.4.1	500	-500	FALSE	00240	PF10_0154, PF14_0053, PF14_0352	FALSE	0	yes	DIRECT EVIDENCE	forward	
R02093	Cytosol	2.7.4.6	500	-500	FALSE	00240	PF13_0349, PFF0275c	FALSE	4.2E-58	yes	BRENDA	reversible	
R02094	Cytosol	2.7.4.9	500	-500	FALSE	00240	PFL2465c	FALSE	1.4E-48	yes	DIRECT EVIDENCE	reversible	
R02098	Cytosol	2.7.4.9	500	-500	FALSE	00240	PFL2465c	FALSE	1.4E-48	yes	DIRECT EVIDENCE	reversible	
R02100	Cytosol	3.6.1.23	500	-500	FALSE	00240	PF11_0282	FALSE	3.2E-19	yes	DIRECT EVIDENCE	forward	
R02101	Cytosol	2.1.1.45	500	-500	FALSE	00240	PFD0830w	FALSE	6E-61	yes	DIRECT EVIDENCE	forward	
R02326	Cytosol	2.7.4.6	500	-500	FALSE	00240	PF13_0349, PFF0275c	FALSE	4.2E-58	yes	BRENDA	reversible	
R02331	Cytosol	2.7.4.6	500	-500	FALSE	00240	PF13_0349, PFF0275c	FALSE	4.2E-58	yes	BRENDA	reversible	
R00132	Cytosol	4.2.1.1	500	-500	FALSE	00240	PF11_0410	FALSE	3.8E-10	yes	BRENDA		
R01397	Cytosol	2.1.3.2	500	-500	FALSE	00240	MAL13P1.221	FALSE	3.9E-89	yes	no	forward	
R02325	Cytosol	3.5.4.13	500	-500	FALSE	00240	PF13_0259	FALSE	no	yes	no	forward	
R00945	Cytosol	2.1.2.1	500	-500	FALSE	00240	PF14_0534, PFL1720w	FALSE	9E-146	yes	DIRECT EVIDENCE		
CarbonicAcid_disoc	Cytosol	spontaneous	500	-500	FALSE	00240		FALSE					

Table 33 : Reactions for pyrimidine metabolism in *P. falciparum* in the evidence checker spreadsheet ready for import back into MetNetMaker for export and manipulation.

Reaction Id	SharkHunt	PlasmoDB		MPMP	BRENDA HIT		BRENDA Localisation		KEGG pfa		KEGG direction	Balanced
		By EC number -> Gene	By Reaction ID -> EC		By EC number (best)	By Reaction ID (best)	By EC number	By Reaction ID (best)	By EC number	By Reaction ID		
R00156	4.2E-58	PF13_0349, PFF0275c	2.7.4.6	R00156	BRENDA	BRENDA			PF13_0349, PFF0275c	PF13_0349	reversible	True
R00158	0.001	PFA0555c	2.7.4.14	R00158	BRENDA	BRENDA			MAL1P2.40	MAL1P2.40	reversible	True
R00512	0.001	PFA0555c	2.7.4.14	no	BRENDA	BRENDA			MAL1P2.40	MAL1P2.40	reversible	True
R00570	4.2E-58	PF13_0349, PFF0275c	2.7.4.6	R00570	BRENDA	BRENDA			PF13_0349, PFF0275c	PF13_0349	reversible	True
R00571	0	PF14_0100	6.3.4.2	no	DIRECT EVIDENCE	DIRECT EVIDENCE			PF14_0100	PF14_0100	forward	True
R00573	0	PF14_0100	6.3.4.2	R00573	DIRECT EVIDENCE	DIRECT EVIDENCE			PF14_0100	PF14_0100	forward	True
R00575_balanced	0.0000019	PF13_0044	no	no	BRENDA	no			PF13_0044	#N/A		#N/A
R00965	2.8E-14	PF10_0225	4.1.1.23	R00965	DIRECT EVIDENCE	DIRECT EVIDENCE			PF10_0225	PF10_0225	forward	True
R01665	0.001	PFA0555c	2.7.4.14	no	BRENDA	BRENDA			MAL1P2.40	MAL1P2.40	forward	True
R01867	2E-88	PFF0160c	1.3.3.1	R01867	DIRECT EVIDENCE	DIRECT EVIDENCE			PFF0160c	PFF0160c	reversible	True
R01870	5.5E-32	PFE0630c	2.4.2.10	R01870	DIRECT EVIDENCE	DIRECT EVIDENCE	soluble	soluble	PFE0630c	PFE0630c	reversible	True
R01993	1.9E-45	PF14_0697	3.5.2.3	R01993	DIRECT EVIDENCE	DIRECT EVIDENCE			PF14_0697	PF14_0697	reversible	True
R02016	0.036	PF11170c	1.8.1.9	R02016	DIRECT EVIDENCE	DIRECT EVIDENCE			PF11170c	PF11170c	forward	True
R02018	0	PF10_0154, PF14_0053, PF14_0352	1.17.4.1	R02018	DIRECT EVIDENCE	DIRECT EVIDENCE			PF10_0154, PF14_0053, PF14_0352	PF10_0154	forward	True
R02024	0	PF10_0154, PF14_0053, PF14_0352	1.17.4.1	R02024	DIRECT EVIDENCE	DIRECT EVIDENCE			PF10_0154, PF14_0053, PF14_0352	PF10_0154	forward	True
R02093	4.2E-58	PF13_0349, PFF0275c	2.7.4.6	R02093	BRENDA	BRENDA			PF13_0349, PFF0275c	PF13_0349	reversible	True
R02094	1.4E-48	PFL2465c	2.7.4.9	R02094	DIRECT EVIDENCE	DIRECT EVIDENCE			PFL2465c	PFL2465c	reversible	True
R02098	1.4E-48	PFL2465c	2.7.4.9	R02098	DIRECT EVIDENCE	DIRECT EVIDENCE			PFL2465c	PFL2465c	reversible	True
R02100	3.2E-19	PF11_0282	3.6.1.23	R02100	DIRECT EVIDENCE	DIRECT EVIDENCE	intracellular	intracellular	PF11_0282	PF11_0282	forward	True
R02101	6E-61	PFD0830w	2.1.1.45	R02101	DIRECT EVIDENCE	DIRECT EVIDENCE			PFD0805w	PFD0805w	forward	True
R02326	4.2E-58	PF13_0349, PFF0275c	2.7.4.6	R02326	BRENDA	BRENDA			PF13_0349, PFF0275c	PF13_0349	reversible	True
R02331	4.2E-58	PF13_0349, PFF0275c	2.7.4.6	R02331	BRENDA	BRENDA			PF13_0349, PFF0275c	PF13_0349	reversible	True
R00132	3.8E-10	PF11_0410	4.2.1.1	R00132	BRENDA	BRENDA			no	#N/A		True
R01397	3.9E-89	MAL13P1.221	2.1.3.2	R01397	no	no			MAL13P1.221	MAL13P1.221	forward	True
R02325	no	PF13_0259	3.5.4.13	R02325	no	no			no	#N/A	forward	True
R00945	9E-146	PF14_0534, PFL1720w	2.1.2.1	R00945	DIRECT EVIDENCE	DIRECT EVIDENCE	cytosol	cytosol	MAL13P1.67, PFL1720w	MAL13P1.67		True
CarbonicAcid_disoc												

Table 34 : Evidence summary from the ECinfo database as presented in the evidence checker spreadsheet and used to help judge the existence of the reactions in table 33 and populate them with additional information for SBML export.

Appendix III : RPMI components

COMPONENTS	Molecular Weight	Concentration (mg/L)	mM
Amino Acids			
Glycine	75	10	0.133
L-Arginine	174	200	1.15
L-Asparagine	132	50	0.379
L-Aspartic acid	133	20	0.15
L-Cystine	240	20	0.0833
L-Glutamic Acid	147	20	0.136
L-Glutamine	146	300	2.05
L-Histidine	155	15	0.0968
L-Hydroxyproline	131	20	0.153
L-Isoleucine	131	50	0.382
L-Leucine	131	50	0.382
L-Lysine hydrochloride	146	40	0.274
L-Methionine	149	15	0.101
L-Phenylalanine	165	15	0.0909
L-Proline	115	20	0.174
L-Serine	105	30	0.286
L-Threonine	119	20	0.168
L-Tryptophan	204	5	0.0245
L-Tyrosine	181	20	0.11
L-Valine	117	20	0.171
Vitamins			
Biotin	244	0.2	0.00082
Choline chloride	140	3	0.0214
D-Calcium pantothenate	477	0.25	0.000524
Folic Acid	441	1	0.00227
Niacinamide	122	1	0.0082
Para-Aminobenzoic Acid	137	1	0.0073
Pyridoxine hydrochloride	206	1	0.00485
Riboflavin	376	0.2	0.000532
Thiamine hydrochloride	337	1	0.00297
Vitamin B12	1355	0.005	3.7E-06
i-Inositol	180	35	0.194
Inorganic Salts			
Calcium nitrate (Ca(NO ₃) ₂ 4H ₂ O)	236	100	0.424
Magnesium Sulfate (MgSO ₄ ·7H ₂ O)	246	100	0.407
Potassium Chloride (KCl)	75	400	5.33
Sodium Bicarbonate (NaHCO ₃)	84	2000	23.81
Sodium Chloride (NaCl)	58	5500	94.83
Sodium Phosphate dibasic (Na ₂ HPO ₄) anhydrous	142	800	5.63
Other Components			
D-Glucose (Dextrose)	180	2000	11.11
Glutathione (reduced)	307	1	0.00326
HEPES	238	5958	25.03
Phenol Red	376.4	5	0.0133

Table 35 : Components of RPMI 1640 growth medium²¹¹.

Appendix IV : Free components of the biomass function

In my analysis of Table 1 in Teng *et al.*¹⁴⁶ I took the maximum measured concentration of each metabolite from each of the extraction methods and converted from a milliMolar concentration to $g/10^{15}$ parasites. The total mass measured in their experiment is $536.8 g/10^{15}$ parasites, around a twentieth of the total mass that I have measured of total biomass per parasite. It is clear that the paper only reports free metabolites, not the composition of the biomass. It is possible that digesting the proteins and DNA/RNA polymers to monomers and repeating their analysis techniques could accurately determine the biomass composition of *P. falciparum*.

Metabolite	Maximums from all methods ($g/10^{15}$ parasites)	
	Best estimate	Standard Deviation
Alanine (Ala)	5.2	2.0
GABA	6.6	1.4
Arginine (Arg)	20.5	5.4
Asparagine (Asn)	9.6	4.8
Aspartate (Asp)	7.5	1.9
Glutamate (Glu)	45.3	16.5
Glutamine (Gln)	7.0	1.6
Glycine (Gly)	3.4	1.3
Histidine (His)	2.6	1.0
Isoleucine (Ile)	4.0	2.2
Leucine (Leu)	10.3	4.0
Lysine (Lys)	13.9	4.9
Methionine (Met)	1.8	1.2
Phenylalanine (Phe)	4.2	1.9
Serine (Ser)	4.4	2.4
Threonine (Thr)	6.0	3.0
Tyrosine (Tyr)	4.1	1.5
Valine (Val)	4.6	1.6
Total amino acids	160.9	58.6
Reduced (GSH)	46.5	10.3
Oxidised (GSSG)	29.2	12.0
myo-Inositol (Ins)	1.2	0.4
Phosphocholine (PC)	7.7	1.6
Phosphoethanolamine (PE)	17.8	5.9
AMP	5.3	3.2
ADP	13.2	4.8
ATP	28.4	12.8
UMP, CMP, (U/C MP)	11.8	7.3
UDP, CDP, UTP, CTP, (U/C D/T P)	31.7	10.2
NAD+	53.9	9.3
Acetate	1.3	0.8
Formate	0.3	0.1
Fumarate	0.7	0.2
alpha-Ketoglutarate (alpha-KG)	2.0	1.6
Lactate (Lac)	2.8	2.0
Malate (Mal)	6.8	1.9
Succinate (Succ)	1.4	0.2
Ethanol	0.6	0.4
HEPES	80.1	13.3
Putrescine (Put)	7.2	2.0
Spermidine (Spd)	23.2	6.5
Spermine (Spm)	2.9	0.6
Total	536.8	224.8

Colours refer to extraction methods used,

perchloric acid (n=13)
methanol/water (n=12)
methanol/chloroform/water (n=10)

Table 36 : Table created from calculations based on Table 1 in Teng. *et al.* Reported concentrations of metabolites have been converted from milliMolar concentration to $g/10^{15}$ parasites. Note the very high standard deviations showing the inherent variability in the biology of the system.

Appendix V : Measured exchange fluxes

Carbon source exchange flux calculations

Units	Value	Glucose	Lactate	Glycerol	Glutamine
10^{-10} mg/RBC/hour	m_{RBC}	-0.79	0.45	0.00	-0.06
	m_m	-1.57	1.03	0.04	-0.08
	e_{RBC}	0.13	0.02	0.00	0.02
	e_m	0.39	0.06	0.01	0.01
	e_p	0.41	0.06	0.01	0.02
	m_p	-0.82	0.60	0.04	-0.02
10^{-10} mg/infected RBC/hour	F_p	-18.2	13.2	0.8	-0.4
	e_F	50%	10%	14%	125%
10^{-8} nmol/infected RBC/hour	F_p	-1010	1470		

Table 37 : Sara A carbon source flux calculations.

Units	Value	Glucose	Lactate	Glycerol	Glutamine
10^{-10} mg/RBC/hour	m_{RBC}	-0.79	0.45	0.00	-0.06
	m_m	-1.05	0.60	0.02	-0.07
	e_{RBC}	0.13	0.02	0.00	0.02
	e_m	0.23	0.01	0.00	0.00
	e_p	0.27	0.02	0.00	0.02
	m_p	-0.29	0.17	0.02	-0.01
10^{-10} mg/infected RBC/hour	F_p	-8.2	4.8	0.7	-0.2
	e_F	92%	13%	5%	302%
10^{-8} nmol/infected RBC/hour	F_p	-458	530		

Table 38 : Sara B carbon source flux calculations.

Units	Value	Glucose	Lactate	Glutamine
10^{-10} mg/RBC/hour	m_{RBC}	-0.26	0.68	0.04
	m_m	-0.93	1.06	-0.05
	e_{RBC}	0.22	0.07	0.03
	e_m	0.10	0.03	0.01
	e_p	0.25	0.07	0.03
	m_p	-0.69	0.41	-0.09
10^{-10} mg/infected RBC/hour	F_p	-16.74	9.97	-2.14
	e_F	36%	18%	38%
10^{-8} nmol/infected RBC/hour	F_p	-929	1107	

*Table 39 : Tom repeat of Jennifer carbon source flux calculations.
No glycerol production was detected.*

Amino acid exchange flux calculations

Units	Value	Valine	Leucine	Isoleucine	Alanine	Tyrosine	Phenylalanine	Hypoxanthine
10^{-10} mg/RBC/hour	m_{RBC}	0.000	-0.001	-0.003	0.000	-0.005	-0.001	-0.016
	m_m	0.007	0.001	0.001	0.014	-0.001	0.002	-0.004
	e_{RBC}	0.002	0.002	0.002	0.000	0.001	0.001	0.002
	e_m	0.002	0.003	0.001	0.002	0.002	0.003	0.001
	e_p	0.002	0.004	0.002	0.002	0.002	0.003	0.002
	m_p	0.008	0.002	0.004	0.014	0.003	0.004	0.012
10^{-10} mg/infected RBC/hour	F_p	0.21	0.06	0.11	0.41	0.10	0.10	0.33
	e_F	31%	184%	47%	13%	53%	86%	16%

Table 40 : Sara A amino acid flux calculations.
Hypoxanthine is listed last as its NMR assignment is unclear.

Units	Value	Valine	Leucine	Isoleucine	Alanine	Tyrosine	Phenylalanine	Hypoxanthine
10^{-10} mg/RBC/hour	m_{RBC}	0.000	-0.001	-0.003	0.000	-0.005	-0.001	-0.016
	m_m	0.003	0.000	-0.002	0.004	0.001	0.002	-0.006
	e_{RBC}	0.002	0.002	0.002	0.000	0.001	0.001	0.002
	e_m	0.001	0.001	0.004	0.002	0.002	0.001	0.004
	e_p	0.002	0.002	0.005	0.002	0.002	0.001	0.004
	m_p	0.003	0.001	0.001	0.004	0.005	0.004	0.009
10^{-10} mg/infected RBC/hour	F_p	0.07	0.03	0.01	0.08	0.12	0.09	0.21
	e_F	56%	198%	721%	67%	45%	32%	48%

Table 41 : Sara B amino acid flux calculations.
Hypoxanthine is listed last as its NMR assignment is unclear.

Units	Value	Valine	Leucine	Isoleucine	Alanine	Tyrosine	Phenylalanine	Hypoxanthine
10^{-10} mg/RBC/hour	m_{RBC}	0.002	-0.002	0.002	0.007	0.001	0.002	0.000
	m_m	0.011	0.003	0.002	0.019	0.003	0.011	-0.006
	e_{RBC}	0.002	0.003	0.005	0.001	0.003	0.003	0.012
	e_m	0.002	0.002	0.001	0.002	0.001	0.003	0.008
	e_p	0.003	0.003	0.005	0.002	0.004	0.004	0.014
	m_p	0.009	0.005	0.000	0.012	0.002	0.010	-0.007
10^{-10} mg/infected RBC/hour	F_p	0.22	0.12	0.00	0.29	0.06	0.24	-0.16
	e_F	35%	67%	12088%	21%	147%	37%	211%

Table 42 : Tom repeat of Jennifer amino acid flux calculations.
Hypoxanthine is listed last as its NMR assignment is unclear.

Appendix VI: Key amino acid compositions

Key amino acid compositions, full tables

Name	Molecular Weight (g/mol)	Count	Weight per mole of haemoglobin (g)	Percent by Weight	Percent by amino acid count
Leucine	113	72	8142	13.0%	12.5%
Valine	99	62	6142	9.8%	10.7%
Lysine	128	44	5636	9.0%	7.6%
Histidine	137	38	5208	8.3%	6.6%
Alanine	71	72	5115	8.2%	12.5%
Phenylalanine	147	30	4412	7.1%	5.2%
Aspartate	115	30	3451	5.5%	5.2%
Threonine	101	32	3234	5.2%	5.5%
Glutamate	129	24	3097	5.0%	4.2%
Serine	87	32	2785	4.5%	5.5%
Proline	97	28	2717	4.4%	4.8%
Asparagine	114	20	2281	3.7%	3.5%
Glycine	57	40	2281	3.7%	6.9%
Tyrosine	163	12	1957	3.1%	2.1%
Arginine	156	12	1873	3.0%	2.1%
Methionine	131	10	1310	2.1%	1.7%
Tryptophan	186	6	1116	1.8%	1.0%
Glutamine	128	8	1024	1.6%	1.4%
Cysteine	103	6	618	1.0%	1.0%
Isoleucine	113	0	0	0.0%	0.0%

Table 43 : Amino acid composition of 2 α -subunits + 2 β -subunits of the human haemoglobin molecule ordered by total weight per mol of haemoglobin. The molecular weights given are those of the amino acid after it has formed a protein.

Amino Acid	Percent by weight	Percent by amino acid count
Asparagine	13.1%	12.2%
Lysine	12.0%	11.6%
Isoleucine	8.5%	9.3%
Glutamate	8.3%	7.0%
Leucine	7.4%	8.1%
Tyrosine	7.3%	5.5%
Aspartate	6.5%	6.0%
Phenylalanine	5.5%	4.6%
Serine	4.4%	6.3%
Arginine	4.1%	2.9%
Threonine	3.4%	4.2%
Valine	3.3%	4.1%
Glutamine	2.9%	2.8%
Histidine	2.8%	2.2%
Methionine	2.7%	2.2%
Proline	2.0%	2.2%
Glycine	1.9%	3.1%
Cysteine	1.8%	1.8%
Alanine	1.4%	2.4%
Tryptophan	0.8%	0.5%

Table 44 : Amino acid use in the *P. falciparum* proteome, adapted from Chanda et al.¹⁷⁶. The molecular weights given are those of the amino acid after it has formed a protein.

Appendix VII : Single gene deletions that affect growth-rate

Gene	Relative Growth	Associated Reaction(s)	Reaction Name(s)	Associated EC number
MALI3PI.221	0	R01397	carbamoyl-phosphate:L-aspartate carbamoyltransferase	2.1.3.2
MALI3PI.40	0	R01512	ATP:3-phospho-D-glycerate 1-phosphotransferase	2.7.2.3
MAL8PI.156	0	R01819	D-mannose-6-phosphate aldose-ketose-isomerase	5.3.1.8
PF07_0072	0	R04779	ATP:D-fructose-6-phosphate 1-phosphotransferase	2.7.1.11
PF08_0062	0	R00127	ATP:AMP phosphotransferase	2.7.4.3
PF08_0066	0	R03815	dihydrolipoylprotein:NAD ⁺ oxidoreductase	1.8.1.4
		R07618	enzyme N6-(dihydrolipoyl)lysine:NAD ⁺ oxidoreductase	1.8.1.4
		R07618	enzyme N6-(dihydrolipoyl)lysine:NAD ⁺ oxidoreductase	1.8.1.4
PF08_0071	0	R00275	superoxide:superoxide oxidoreductase	1.15.1.1
PF08_0077	0	R00888	GDP-mannose 4,6-hydro-lyase(GDP-4-dehydro-6-deoxy-D-mannose-forming)	4.2.1.47
PF10_0086	0	R00127	ATP:AMP phosphotransferase	2.7.4.3
PF10_0123	0	R01231	Xanthosine-5'-phosphate:L-glutamine amido-ligase (AMP-forming)	6.3.5.2
PF10_0137	0	R05692	GDP-L-fucose:NADP ⁺ 4-oxidoreductase (3,5-epimerizing)	1.1.1.271
PF10_0154	0	R02017	2'-Deoxyadenosine 5'-diphosphate:oxidized-thioredoxin2'-oxidoreductase	1.17.4.1
		R02018	2'-Deoxyuridine 5'-diphosphate:oxidized-thioredoxin2'-oxidoreductase	1.17.4.1
		R02019	2'-Deoxyguanosine 5'-diphosphate:oxidized-thioredoxin2'-oxidoreductase	1.17.4.1
		R02023	2'-Deoxyuridine 5'-triphosphate:oxydized-thioredoxin2'-oxidoreductase	1.17.4.1
		R02024	2'-Deoxycytidine diphosphate:oxidized-thioredoxin 2'-oxidoreductase	1.17.4.1
PF10_0155	0	R00658	2-phospho-D-glycerate hydro-lyase (phosphoenolpyruvate-forming)	4.2.1.11
PF10_0169	0	R01818	D-Mannose 6-phosphate 1,6-phosphomutase	5.4.2.8
PF10_0218	0	R00351	acetyl-CoA:oxaloacetate C-acetyltransferase (thioester-hydrolysing)	2.3.3.1
PF10_0225	0	R00965	orotidine-5'-phosphate carboxy-lyase (UMP-forming)	4.1.1.23
PF10_0363	0	R00200	ATP:pyruvate 2-O-phosphotransferase	2.7.1.40
		R00200	ATP:pyruvate 2-O-phosphotransferase	2.7.1.40
PF10_0407	0	R02569	acetyl-CoA:enzyme N6-(dihydrolipoyl)lysine S-acetyltransferase	2.3.1.12
PF10_0409	0	R00742_balanced	Acetyl-CoA:carbon-dioxide ligase (ADP-forming)_balanced	6.4.1.2
		R04386_balanced	Acetyl-CoA:carbon-dioxide ligase (ADP-forming)_balanced	6.4.1.2

Gene	Relative Growth	Associated Reaction(s)	Reaction Name(s)	Associated EC number
PF11_0145	0	R02530	(R)-S-Lactoylglutathione methylglyoxal-lyase (isomerizing)	4.4.1.5
PF11_0157	0	R00842	sn-Glycerol-3-phosphate:NAD+ 2-oxidoreductase	1.1.1.8
PF11_0208	0	R01518	2-Phospho-D-glycerate 2,3-phosphomutase	5.4.2.1
PF11_0256	0	R00014	pyruvate:thiamin diphosphate acetaldehydetransferase(decarboxylating)	1.2.4.1
		R03270	NO NAME	1.2.4.1
PF11_0294	0	R04779	ATP:D-fructose-6-phosphate 1-phosphotransferase	2.7.1.11
PF13_0044	0	R00149	Carbon-dioxide:ammonia ligase(ADP-forming,carbamate-phosphorylating)	6.3.4.16
		R00575_balanced	hydrogen-carbonate:L-glutamine amido-ligase (ADP-forming,carbamate-phosphorylating)_balanced	6.3.5.5
PF13_0141	0	R00703	(S)-Lactate:NAD+ oxidoreductase	1.1.1.27
PF13_0143	0	R01049	ATP:D-ribose-5-phosphate diphosphotransferase	2.7.6.1
PF13_0144	0	R00703	(S)-Lactate:NAD+ oxidoreductase	1.1.1.27
PF13_0157	0	R01049	ATP:D-ribose-5-phosphate diphosphotransferase	2.7.6.1
PF13_0229	0	R01325	citrate hydro-lyase (cis-aconitate-forming)	4.2.1.3
		R01900	isocitrate hydro-lyase (cis-aconitate-forming)	4.2.1.3
PF13_0242	0	R00267	Isocitrate:NADP+ oxidoreductase (decaboxylating)	1.1.1.42
PF13_0269	0	R00847	ATP:glycerol 3-phosphotransferase	2.7.1.30
PF13_0349	0	R00124	ATP:ADP phosphatransferase	2.7.4.6
		R00156	ATP:UDP phosphotransferase	2.7.4.6
		R00330	ATP:GDP phosphotransferase	2.7.4.6
		R00570	ATP:CDP phosphotransferase	2.7.4.6
		R01137	ATP:dADP phosphotransferase	2.7.4.6
		R01857	ATP:dGDP phosphotransferase	2.7.4.6
		R02093	ATP:dTDP phosphotransferase	2.7.4.6
		R02326	ATP:dCDP phosphotransferase	2.7.4.6
		R02331	ATP:dUDP phosphotransferase	2.7.4.6

Gene	Relative Growth	Associated Reaction(s)	Reaction Name(s)	Associated EC number
PF14_0053	0	R02017	2'-Deoxyadenosine 5'-diphosphate:oxidized-thioredoxin2'-oxidoreductase	1.17.4.1
		R02018	2'-Deoxyuridine 5'-diphosphate:oxidized-thioredoxin2'-oxidoreductase	1.17.4.1
		R02019	2'-Deoxyguanosine 5'-diphosphate:oxidized-thioredoxin2'-oxidoreductase	1.17.4.1
		R02023	2'-Deoxyuridine 5'-triphosphate:oxydized-thioredoxin2'-oxidoreductase	1.17.4.1
		R02024	2'-Deoxycytidine diphosphate:oxidized-thioredoxin 2'-oxidoreductase	1.17.4.1
PF14_0100	0	R00573	UTP:L-glutamine amido-ligase (ADP-forming)	6.3.4.2
PF14_0341	0	R02739	alpha-D-Glucose 6-phosphate ketol-isomerase	5.3.1.9
		R02740	alpha-D-Glucose 6-phosphate ketol-isomerase	5.3.1.9
		R03321	beta-D-Glucose 6-phosphate ketol-isomerase	5.3.1.9
PF14_0352	0	R02017	2'-Deoxyadenosine 5'-diphosphate:oxidized-thioredoxin2'-oxidoreductase	1.17.4.1
		R02018	2'-Deoxyuridine 5'-diphosphate:oxidized-thioredoxin2'-oxidoreductase	1.17.4.1
		R02019	2'-Deoxyguanosine 5'-diphosphate:oxidized-thioredoxin2'-oxidoreductase	1.17.4.1
		R02023	2'-Deoxyuridine 5'-triphosphate:oxydized-thioredoxin2'-oxidoreductase	1.17.4.1
		R02024	2'-Deoxycytidine diphosphate:oxidized-thioredoxin 2'-oxidoreductase	1.17.4.1
PF14_0378	0	R01015	D-glyceraldehyde-3-phosphate aldose-ketose-isomerase	5.3.1.1
PF14_0425	0	R01070	beta-D-fructose-1,6-bisphosphate D-glyceraldehyde-3-phosphate-lyase(glycerone-phosphate-forming)	4.1.2.13
PF14_0441	0	R00014	pyruvate:thiamin diphosphate acetaldehydetransferase(decarboxylating)	1.2.4.1
		R03270	NO NAME	1.2.4.1
PF14_0534	0	R00945	5,10-Methylenetetrahydrofolate:glycine hydroxymethyltransferase	2.1.2.1
PF14_0598	0	R01061	D-glyceraldehyde-3-phosphate:NAD+ oxidoreductase (phosphorylating)	1.2.1.12
PF14_0664	0	R00742_balanced	Acetyl-CoA:carbon-dioxide ligase (ADP-forming)_balanced	6.4.1.2
		R04385	biotin-carboxyl-carrier-protein:carbon-dioxide ligase (ADP-forming)	6.3.4.14
		R04386_balanced	Acetyl-CoA:carbon-dioxide ligase (ADP-forming)_balanced	6.4.1.2
PF14_0697	0	R01993	(S)-dihydroorotate amidohydrolase	3.5.2.3
PF14_0774	0	R00885	GTP:alpha-D-mannose-1-phosphate guanylyltransferase	2.7.7.13
PFA0530c	0	R00127	ATP:AMP phosphotransferase	2.7.4.3

Gene	Relative Growth	Associated Reaction(s)	Reaction Name(s)	Associated EC number
PFA0555c	0	R00158	ATP:UMP phosphotransferase	2.7.4.14
		R02569	acetyl-CoA:enzyme N6-(dihydrolipoyl)lysine S-acetyltransferase	2.3.1.12
PFC0170c	0	R00621	NO NAME	1.2.4.2
		R03316	NO NAME	1.2.4.2
PFC0275w	0	R00842	sn-Glycerol-3-phosphate:NAD+ 2-oxidoreductase	1.1.1.8
PFC0831w	0	R01015	D-glyceraldehyde-3-phosphate aldose-ketose-isomerase	5.3.1.1
PFD0311w	0	R01736	(R)-S-Lactoylglutathione hydrolase	3.1.2.6
PFD0660w	0	R01518	2-Phospho-D-glycerate 2,3-phosphomutase	5.4.2.1
PFD0755c	0	R00127	ATP:AMP phosphotransferase	2.7.4.3
		R00939	5,6,7,8-Tetrahydrofolate:NADP+ oxidoreductase	1.5.1.3
PFD0830w	0	R02101	5,10-Methylenetetrahydrofolate:dUMP C-methyltransferase	2.1.1.45
PFE0555w	0	R02222_generalised	Generalised Single Unit Fatty Acid Desaturation	1.1.4.19.1
PFE0605c	0	R00497	gamma-L-glutamyl-L-cysteine:glycine ligase (ADP-forming)	6.3.2.3
PFE0630c	0	R01870	Orotidine-5'-phosphate:diphosphatephospho-alpha-D-ribosyl-transferase	2.4.2.10
PFF0160c	0	R01867	(S)-Dihydroorotate:oxygen oxidoreductase	1.3.3.1
PFF0230c	0	R02530	(R)-S-Lactoylglutathione methylglyoxal-lyase (isomerizing)	4.4.1.5
		R00124	ATP:ADP phosphatransferase	2.7.4.6
		R00156	ATP:UDP phosphotransferase	2.7.4.6
		R00330	ATP:GDP phosphotransferase	2.7.4.6
		R00570	ATP:CDP phosphotransferase	2.7.4.6
PFF0275c	0	R01137	ATP:dADP phosphotransferase	2.7.4.6
		R01857	ATP:dGDP phosphotransferase	2.7.4.6
		R02093	ATP:dTDP phosphotransferase	2.7.4.6
		R02326	ATP:dCDP phosphotransferase	2.7.4.6
		R02331	ATP:dUDP phosphotransferase	2.7.4.6
PFF0455w	0	R00351	acetyl-CoA:oxaloacetate C-acetyltransferase (thioester-hydrolysing)	2.3.3.1

Gene	Relative Growth	Associated Reaction(s)	Reaction Name(s)	Associated EC number
PFF0895w	0	R00342	(S)-malate:NAD ⁺ oxidoreductase	1.1.1.37
		R00342	(S)-malate:NAD ⁺ oxidoreductase	1.1.1.37
PFF1130c	0	R00275	superoxide:superoxide oxidoreductase	1.15.1.1
		R01326	ATP:D-mannose 6-phosphotransferase	2.7.1.1
PFF1155w	0	R01600	ATP:beta-D-glucose 6-phosphotransferase	2.7.1.1
		R01786	ATP:alpha-D-glucose 6-phosphotransferase	2.7.1.1
		R00200	ATP:pyruvate 2-O-phosphotransferase	2.7.1.40
PFF1300w	0	R00200	ATP:pyruvate 2-O-phosphotransferase	2.7.1.40
PFI0755c	0	R04779	ATP:D-fructose-6-phosphate 1-phosphotransferase	2.7.1.11
PFI0925w	0	R00894	L-glutamate:L-cysteine gamma-ligase (ADP-forming)	6.3.2.2
PFI1020c	0	R01130	IMP:NAD ⁺ oxidoreductase	1.1.1.205
PFI1105w	0	R01512	ATP:3-phospho-D-glycerate 1-phosphotransferase	2.7.2.3
		R08539_balanced	NADPH:ferricytochrome-b5 oxidoreductase	1.6.2.4
PFI1140w	0	R00100	NADH:ferricytochrome-b5 oxidoreductase	1.6.2.2
PFI1170c	0	R02016	NADPH:oxidized-thioredoxin oxidoreductase	1.8.1.9
PFI1420w	0	R00332	ATP:GMP phosphotransferase	2.7.4.8
PFL0285w	0	R01736	(R)-S-Lactoylglutathione hydrolase	3.1.2.6
PFL0595c	0	R00274	glutathione:hydrogen-peroxide oxidoreductase	1.1.1.9
PFL0675c	0	R00885	GTP:alpha-D-mannose-1-phosphate guanylyltransferase	2.7.7.13
PFL0780w	0	R00842	sn-Glycerol-3-phosphate:NAD ⁺ 2-oxidoreductase	1.1.1.8
		R03815	dihydrolipoylprotein:NAD ⁺ oxidoreductase	1.8.1.4
PFL1550w	0	R07618	enzyme N6-(dihydrolipoyl)lysine:NAD ⁺ oxidoreductase	1.8.1.4
		R07618	enzyme N6-(dihydrolipoyl)lysine:NAD ⁺ oxidoreductase	1.8.1.4
PFL1720w	0	R00945	5,10-Methylenetetrahydrofolate:glycine hydroxymethyltransferase	2.1.2.1
		R02094	ATP:dTMP phosphotransferase	2.7.4.9
PFL2465c	0	R02098	ATP:dUMP phosphotransferase	2.7.4.9
PFI4_0334	0.9665	R00093	L-glutamate:NAD ⁺ oxidoreductase (transaminating)	1.4.1.14

Gene	Relative Growth	Associated Reaction(s)	Reaction Name(s)	Associated EC number
MAL3P2.24	0.982	R00848	sn-Glycerol-3-phosphate:(acceptor) 2-oxidoreductase	1.1.5.3
PF08_0045	0.982	R00621	NO NAME	1.2.4.2
		R03316	NO NAME	1.2.4.2
PF10_0334	0.982	R00408	Succinate:(acceptor) oxidoreductase	1.3.99.1
		R00432	Succinate:CoA ligase (GDP-forming)	6.2.1.4
PF13_0070	0.982	R00621	NO NAME	1.2.4.2
		R03316	NO NAME	1.2.4.2
PF13_0120	0.982	R02570	succinyl-CoA:enzyme N6-(dihydrolipoyl)lysine S-succinyltransferase	2.3.1.61
PF13_0121	0.982	R02570	succinyl-CoA:enzyme N6-(dihydrolipoyl)lysine S-succinyltransferase	2.3.1.61
PF14_0295	0.982	R00405	Succinate:CoA ligase (ADP-forming)	6.2.1.5
PFI1340w	0.982	R01082	(S)-malate hydro-lyase (fumarate-forming)	4.2.1.2
PFL0630w	0.982	R00408	Succinate:(acceptor) oxidoreductase	1.3.99.1
PF14_0192	0.9861	R00094	glutathione:NAD ⁺ oxidoreductase	1.8.1.7
		R00115	glutathione:NADP ⁺ oxidoreductase	1.8.1.7
PFE0730c	0.9874	R01056	D-ribose-5-phosphate aldose-ketose-isomerase	5.3.1.6
PF14_0511	0.9893	R02035	6-Phospho-D-glucono-1,5-lactone lactonohydrolase	3.1.1.31
		R02736	beta-D-Glucose-6-phosphate:NADP ⁺ 1-oxoreductase	1.1.1.49
PF14_0520	0.9893	R01528	6-phospho-D-gluconate:NADP ⁺ 2-oxidoreductase (decarboxylating)	1.1.1.44
PFF0530w	0.9903	R01641	Sedoheptulose-7-phosphate:D-glyceraldehyde-3-phosphateglycolaldehyde transferase	2.2.1.1
		R01830	beta-D-Fructose 6-phosphate:D-glyceraldehyde-3-phosphateglycolaldehyde transferase	2.2.1.1
PFL0960w	0.9903	R01529	D-Ribulose-5-phosphate 3-epimerase	5.1.3.1
PF14_0508	0.9934	R00112	NADPH:NAD ⁺ oxidoreductase	1.6.1.1
MAL13PI.146	0.9975	R00181	AMP aminohydrolase	3.5.4.6
PF13_0259	0.9994	R02325	dCTP aminohydrolase	3.5.4.13

Appendix VIII : Application for Synthetic Aesthetics funding

The synthetic aesthetics project was run by the University of Edinburgh and the University of Stanford and aimed to bring together designers and scientists to solve problems with design in synthetic biology. More details on the project and some of the results of the funding are online at www.syntheticaesthetics.org. My proposal, as outlined in the letter of interest below, was not funded but remains an interest of mine.

Letter of interest,

For me, synthetic biology is all about networks. From standardising biological building blocks to creating new organisms, it is the interactions rather than the components that are most fascinating. In the exciting and rapidly progressing new fields of synthetic biology and systems biology people are saving time by using old network design principles, often taken from electronic engineering, rather than thinking about the real requirements of the problems they face. My fear is that by reusing these existing design tools in the world of biology we risk obscuring a truer understanding of biological systems. To put it simply, I think that we are drawing biological networks just like we draw electronic networks not because the two are similar but because we haven't taken the time to design a better way of doing things.

Recently, I took pictures of all the networks that were printed on posters at the Systems Biology of Microorganisms conference in Paris. There were reaction networks, regulatory networks, gene expression networks and protein interaction networks galore. Almost without exception the aesthetics of the network drawings were poor but the real problem in the drawings was their underlying design. I saw that all the networks were drawn with large components connected by identical thin lines even though the components were largely unimportant and the connections were both critical and varied.

My specialism is metabolism but all biological networks are similar in that they are rarely static collections of connected components. The flows within them are changeable and the networks themselves reorganise far more dynamically than their electrical counterparts. Biological networks pulse with life, they adapt to changes and most importantly they grow and shrink with time. All of these properties can conceivably be described in terms of statically connected components within highly connected networks but they could be

much more meaningfully described by concentrating more on the connections and less on the components.

I am confident that my vision as outlined on the next page can be realised, but I know from experience that it will not be easy. Recently I completed a full reconstruction of the metabolic network of human malaria but found the task frustrating and unnecessarily difficult. In talking with other people I found that the problems I had were common and as a newcomer to the field of systems biology I was surprised that many of the problems had not been resolved. I decided to try and help by designing MetNetMaker, probably the first metabolic network model creator with an easy to use interface and no enormous nested menus or command prompts. In doing this I met with a surprising amount of resistance. I was frequently told, "it would be quicker to just design your network" or asked "why do you want the tool to look nice? You're the only one who'll use it" and occasionally mocked for "wasting my PhD" spending more of my time designing the tool than doing what systems biologists consider to be real work.

The Synthetic Aesthetics project offers me a chance to show that good design in science is worth spending time on. I firmly believe that time spent on good design will repay itself many times over but it is rare to get the opportunity to try and prove that conviction. I am well aware that my vision has large gaps as well as unseen problems; it needs a fresh input of ideas and someone to look at it who can ask the questions I haven't thought of yet. I'd like help to find someone with new ideas and a different perspective on design and science and I'd really like to grasp this rare opportunity to justify design within science as something more than just an eccentric passion. I have friends who are designers who might be interested to work with me on this project but I think it might be more creative and original to take advantage of the skills and interests of the designers who are applying for this project so that together we can look at these problems and change how we think about, and design, metabolic networks. I'm not proposing that we create the software to do this, but I would certainly like to create a compelling visual template for building such a system that we could use to tempt students, amateur coders and funding bodies into helping make our improved vision a reality.

Vision

I imagine metabolic networks where the consumption of a metabolite is shown as a visible whirlpool draining from the system and the production of a metabolite is shown as a pulsing fountain. Currency metabolites flow in different coloured streams between their multiple occurrences and feedback, both negative and positive, visually impedes or magnifies flow within the network. In this vision, dead-end metabolites are immediately visible as obstructions to flow, redundant reactions stick out like rocks in a river, the role of cellular compartments as barriers to flow is clear and the effects of changes in the structure of the network are immediately reflected in the flows within it. The network itself would be largely static, in keeping with existing design principles in SBML (sbml.org) and SBGN (sbgn.org), and efforts would be made to preserve recognisable design elements such as the layout of the KEGG metabolism map and the shapes of the TCA cycle, glycolysis and fatty-acid metabolism. The goal must be to move networks away from mathematical abstraction and towards living entities that designers of biological systems can play with and explore.

Appendix IX : Application for Gates Foundation funding

Section I

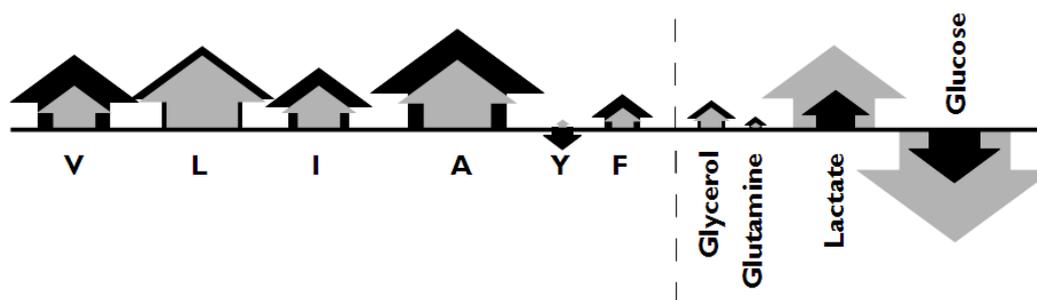
The genetic origins of emerging drug resistance in malaria are being uncovered more quickly but our understanding of how existing drugs and drug-candidates work has not kept pace. We propose a mix of computation and experiment to improve existing models of *P. falciparum* metabolism with the aim of discovering how anti-malarial compounds kill parasites.

There is no published process for finding out how existing anti-malarial drugs affect parasite metabolism. Even for those drugs with a known mechanism of action, the wider metabolic affects — and thus areas of metabolism where synergistic drug combinations could be targeted — have not been systematically explored. We think a process we have developed and have good plans to improve can address this problem.

Existing models of malaria metabolism (Plata *et al.* 2010; Huthmacher *et al.* 2010) lack key links with experimentally measured properties of malaria metabolism and lack the associated tools and simplicity needed to keep them flexible, up-to-date, and accessible to wet-lab scientists. Significant recent advances that move us closer to an accurate biomass composition for *P. falciparum* are not integrated with these models nor have measurements of metabolite exchange between parasite, host and growth medium.

Using our MetNetMaker software (Forth *et al.*, *Bioinformatics*, 2010) we have created a simple and adaptable metabolic model of *P. falciparum* that allows flux-balance analysis (FBA) and similar analysis to be performed on an improved version of the most widely used, most frequently updated and most discussed metabolic network for malaria (Ginsburg 2006). This manually curated model has not previously been available for computational analysis. As a proof of concept we have conducted an initial trial comparing the metabolism of parasites grown in normal culture conditions and those grown in growth medium containing a sub-lethal dose of atovaquone, a widely used prophylactic anti-malarial. NMR analysis of the used growth medium in this trial was able to quantify the exchange flux of metabolites between the growth medium and parasitised red blood cells for the four principle carbon sources (Glucose, Lactate, Glycerol, Glutamine) and a further six amino acids (Valine, Isoleucine, Leucine, Alanine, Tyrosine, Phenylalanine) over forty hours. HPLC analysis of parasites isolated at the end of this period confirmed that the critical step in pyrimidine synthesis was partially inhibited and provided information to

update the nucleotide portion of the biomass function for the parasite under these conditions.



Production (up) and consumption (down) of metabolites by malaria parasites in normal conditions (black) and when inhibited by atovaquone (gray). Amino acids (left) are on a different scale to carbon sources (right).

Exchange fluxes between the parasites and their growth medium largely agree with existing knowledge and, with the exception of lipids, cover all major metabolite destinations branching from the most accurate current model of *P. falciparum* central carbon metabolism. Applying these measured fluxes as constraints on the metabolic model and performing flux-balance analysis shows that inhibition of a single reaction involved in pyrimidine synthesis significantly affects pathways as distant as glycerol metabolism. Currently atovaquone is used in combination with pro-guanyl and marketed by GSK as Malarone®, but the clear effect it has on glycerol metabolism suggests a new possibility for synergy that could overcome developing resistance.

Significant issues have been resolved and we feel ready to apply our process to a larger number of known anti-malarial compounds and candidate compounds with the aim of proving that the effects on core metabolism of many of these drugs can be systematically deduced. The limited success that we have had to date suggests that this is possible and we are keen to see if this technique can be useful with no prior knowledge of a drug's mechanism of action.

Section II

In phase 1 we will study 20% of the drug-like compounds in the open-access malaria box, choosing the forty with the lowest EC50 concentration, plus the antimalarials atovaquone, artemesinin and doxycycline. Each round of experiments will analyse the metabolism of parasitised red blood cells in the presence of the EC50 concentration of six compounds. An

identically treated culture of red blood cells and another of parasitised red blood cells in growth medium without an anti-malarial compound added act as the required controls.

1. 0 hours— synchronise 200ml of high parasitemia ($\approx 5\%$) culture by soribitol incubation.
2. 24 hours — concentrate schizonts to the $\approx 90\%$ parasitemia by magnetic isolation.
3. 24 hours — create 6 parallel 6ml cultures at 10% parasitemia and 10% haematocrit by diluting with red blood cells and standard growth medium supplemented with extra glucose and the relevant test compound. Also prepare the two controls.
4. 34, 44, 54 hours — remove 0.25ml of culture and centrifuge to isolate the growth medium.
5. 54 hours — isolate the resultant parasite biomass for HPLC analysis.

Extracted growth medium samples and biomass samples will be stored at -80°C and retained for future analysis by LC-MS or other techniques at a later date or by another research group. We aim to acquire NMR spectra, calculate exchange fluxes and measure biomass components by HPLC for the forty-three compounds in six batches over the course of the year. Integration with our metabolic model will be ongoing and may extend slightly after this.

	British Pounds	US Dollars
<i>Grade 7 postdoctoral researcher salary + pension + NI + University costs (10%)</i>	$31\,591 + 5\,055 + 2\,441 + 3\,908$ <i>= £42 995</i>	<i>\$69 214</i>
<i>NMR acquisition costs</i>	<i>£12 000</i>	<i>\$19 317</i>
<i>Malaria culturing costs</i>	<i>£5 000</i>	<i>\$8 050</i>
<i>HPLC running costs</i>	<i>£1 000</i>	<i>\$1 609</i>
Total		<i>\$98 190</i>