DTU Library

A kinetic description of the Warburg effect in CHO cells

Cowie, Nicholas Luke

Publication date:
2023

Document Version
Publisher's PDF, also known as Version of record

[Link back to DTU Orbit](#)

A kinetic description of
the Warburg effect in CHO cells

Nicholas Luke Cowie
PhD Thesis

$V = E$

θ

Γ

Ψ

0
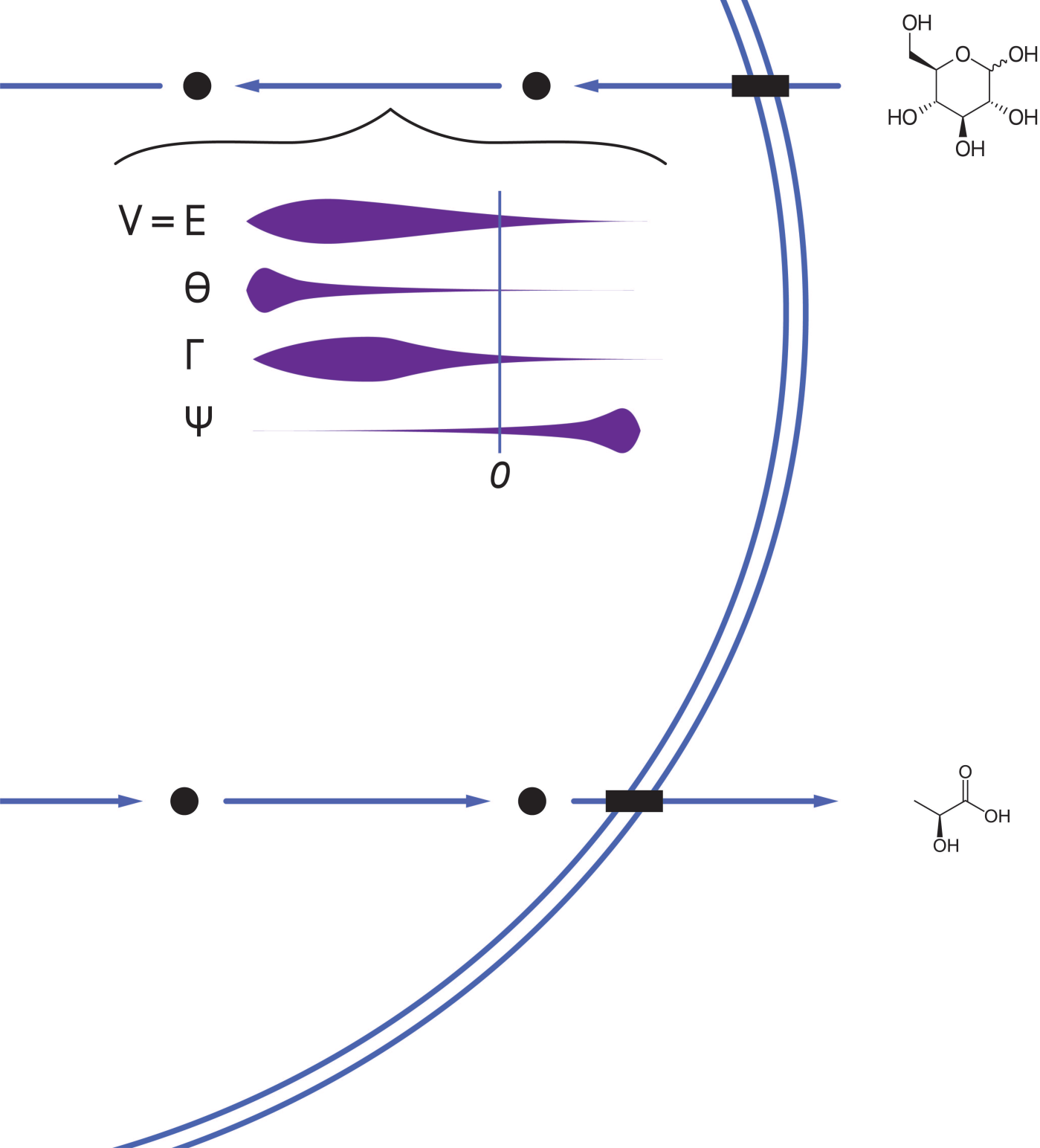
A kinetic investigation of the Warburg effect in CHO cells

PhD Thesis
June, 2023

By
Nicholas Luke Cowie

Acknowledgements

"It's simple, go in, do your PhD, and get out" are the words have thought ever when I arrived in Denmark. This was not the case, and you'd be hard pressed to find a PhD graduate came out with the same high intensity simplistic view of science as when they first arrived. My PhD has taught me two things: Science is agnostic of your efforts and wishes, and that the people around you are essential for offering support when science doesn't.

First of all I would like to thank my supervisor, Professor Lars Nielsen, for the opportunity to undertake this PhD. Lars, you're a great supervisor who's enthusiasm for scientific challanges has pushed my vision of science that made this a far more exciting PhD then it could have been. I'd also like to thank my three other supervisors and role models I've had throughout my time at CfB. Thank you Teddy for your support, knowledge, and positivity that you bring every day without fail. This thesis certainly would have not have been as enjoyable otherwise. Thank you Marta for taking me under your wing and showing me the deep dark and wonderful world of metabolic modelling. Your guidance and support kept me going at multiple points during my PhD journey. And thank you Pedro for our impromptu meetings about anything and everything kinetics.

Thank you to those who were my experimental half: David and Kristian. And also to those who turned those experiments Catarina, Tune, Kartsen and Stefano into the data which I wove my thesis around. My thesis literally would not have been possible without all of you. I'd also like to thank Rebeca, without which this program would have been an unnavigable jungle.

Looking back at my time at CfB there's been many people along the way I need to thank. So from the beginning, thank you Svetlana and Denis for being great colleageus, friends, and even better firewatch buddies. Thank you past and present memebers of the Nielsen group for making DTU such an enjoyable and welcoming space: Marinna, Areti, Pouyan, Hani, Igor, Ivan, Jorge, Shannara, Viktor, Hossein and Victor. Thank you Matthias, Philip, Svetlana, Emre, Pasquale, and Matthew for the countless chats and beverages we've had over the years.

COVID wasn't fun for anyone but to Ben, Anna, and Josh, thanks for making it far more enjoyable than it could have been. The amount of board games and late nights we had was certainly one of the highlights of my PhD.

I'd like to thank those at Amyris who made my stay in California such a fun experience: Alex, Niv, Oren, Ishita, Iris, Sammie, Lauren, Dmitry, and Mel.

None of this would have been possible without the support from both my Danish and Australian family, so thank you for being there for me throughout this entire journey. And finally, thank you Jacque. Thank you for making Denmark feel like home, you've supported me though countless late nights and are always in my corner. I would never have been able to do this without you.

Approval

This thesis is written as partial fulfilment of the requirements to obtain a PhD degree at the Technical University of Denmark, from September 2019 to January 2023. The work was carried out in the Quantative Modelling of Cell Metabolism group supervised by Professor Lars Nielsen, with co-supervisors Teddy Groves and Pedro Saa. The PhD study included research stay at the Amyris, Emeryville, CA, from February to May 2022 under the supervision of Dmitry Grapov, Alex Kilbo, and Niv Antonovsky.

Nicholas Luke Cowie

....
Signature

....
Date

Abstract

The Warburg effect has been known about for nearly 100 years. It is characterised by a high rate of glycolytic flux and production of lactate in highly proliferating cells, relative to their slower growing counterparts. This phenomenon is still observed in environments where there is sufficient oxygen to catabolise glucose into CO_2 via oxidative phosphorylation. High growth is a trait common among both cancer cells and cells engineered to produce biopharmaceuticals. The Warburg effect has often been described as an enigma due to it being less energy and carbon effecient than oxidative phosphorylation. This leads to the overall question of this thesis: *How do glycolytic enzymes regulate the Warburg effect?*.

The organism we are investigating the Warburg effect in are known as Chinese Hasmster Ovary (CHO) cells. CHO cells are a common cell line used in the production of biopharmaceuticals. The laboratory we are working in has developed a non-Warburg cell line, CHO-ZeLa, that has a similar growth rate to its parent cell line, CHO-S, which is Warburg positive. In this thesis we investigate this high growth, non-Warburg phenotype and compare it to its parent stain using a kinetic model.

We developed a Bayesian kinetic modelling framework called Maud that uses Hamiltonian Monte Carlo for posterior sampling. To evaluate our framework we carried out a case study of a representitive metabolic model, which compared full posterior sampling to an approximation of the posterior using point estimation. We showed that the posterior distribution is not well approximated by point estimation methods. Additionally, full Bayesian sampling translated to tighter confidence intervals and improved predictions. We made Maud available as a Python package that automates ODE construction, diagnostics, and posterior predictions.

We collected a dataset consisiting of 6 cell lines that included their proteome, metabolome and fluxome (measured using ^{13}C metabolic flux analysis). This dataset was crucial in developing a kinetic model of glycolysis in CHO cells that spaned the Embden-Meyerhof-Parnas (EMP) and pentose phosphate (PP) pathways.

Our comparative model of the CHO-S and CHO-ZeLa cell lines showed that hexokinase 2 is a strong regulator of glycolysis and may be mediated via mitochondrial binding, a mechanism not previously modelled. Furthermore, our results aligned with previous models of the Warburg effect in CHO cells that suggest that allosteric regulation of phosphofructokinase M is required for the Warburg effect. Kinetic modelling has the side effect of helping identify when mechanisms may be incorrectly understood. We were able to identify that the current understanding of triosphosphate isomerase and glyceraldehyde-3-phosphate dehydrogenase mechanisms are likely missing additional regulation, which was indicated by the surrounding metabolomics data not fitting the model.

The results also suggest that we require a method to account for phosphorylation in

metabolic models. We therefore developed a Monod-Wynman-Changeaux (MWC) analogous model of phosphorylation, and a combinatorial model of phosphorylation. These models will futher be used to model the hexokinase 2 and 6-phosphofructo-2-kinase/fructose 2,6-bisphosphatase mechansism, which were identified as essential to regulate the Warburg effect.

Sammenfatning

Warburg-effekten har været kendt i næsten 100 år. Warburg-effekten kan beskrives som en stigning i glycolyse med en høj produktion af laktat, på trods af at der er ilt til rådighed til at oxidere glukose fuldt ud til CO2. Warburg-effekten kan findes i hurtigt voksende celler, et træk der er almindeligt blandt cancerceller og celler designet til at producere biofarmaceutiske produkter. Denne fænotype er ofte blevet beskrevet som en gåde, fordi den er mindre energi – og kulstofseffektiv end oxidativ phosphorylering. I forsøget på at forbedre forståelse af hvorfor Warburg-effekten opstår, bliver vi nødt til at forstå mekanismen bag Warburg-effekten. Dette leder til spørgsmålet: Hvordan kan glycolyse regulere Warburg-effekten?

Organismen, vi undersøger er kendt som Chinese Hamster Ovary (CHO) celler og er en almindelig cellelinje, der bruges til produktion af biofarmaceutiske midler. Laboratoriet, vi arbejder i, har udviklet en non-Warburg-fænotype (CHO-ZeLa), der har en lignende vækst som sin forudgående cellelinje – en Warburg-positiv cellelinje (CHO-S). I denne ph.d. afhandling undersøger vi den høje vækst i non-Warburg-fænotype og sammenligner med den forudgående cellelinje (CHO-S) ved hjælp af kinetisk model.

Vi har udviklet en bayesiansk kinetisk modelleringsramme kaldet Maud, der bruger Hamiltonian Monte Carlo til at sample den posterior fordeling af de kinetiske parametre. Vores case-studie, der var en sammenligning mellem fuld posterior sampling og punktestimat-approksimationsmetoder, viste at den posterior fordeling ikke var godt approksimeret af punktestimater. Fuld bayesiansk sampling resulterer i strammere konfidensintervaller og forbedrede forudsigelser. For at forbedre tilgængeligheden, er Maud tilgængelig som en Python-pakke og automatiserer ODE-konstruktion, diagnostik og posterior forudsigelser.

Vi indsamlede et datasæt bestående af 6 cellelinjer, der inkluderede deres proteom, metabolom og fluxom (målt ved hjælp af 13C-metabolisk fluxanalyse). Dette datasæt var afgørende for udviklingen af en kinetisk model af glykolyse i CHO-celler, der omfattede Embden-Meyerhof-Parnas (EMP) og pentosefosfat (PP) veje.

Vores model af CHO-S og CHO-ZeLa cellelinjerne viste, at hexokinase 2 er en stærk regulator af glykolyse og kan formidles via mitokondriel binding; en mekanisme der ikke tidligere er blevet modelleret. Desuden stemte vores resultater overens med tidligere modeller af metabolisme, der antyder, at allosterisk regulering af phosphofructokinase M er nødvendig for Warburg-effekten. Kinetisk modellering har den fordel at identificere, hvornår mekanismer kan være fejlagtigt modelleret. Vi var i stand til at identificere, at den nuværende forståelse af triosefosfatisomarase- og glyceraldehyde-3-phosphat dehydrogenase-mekanismerne sandsynligvis mangler yderligere regulering, da de omgivende metabolomics data ikke passer til modellen.

Resultaterne antyder også, at vi har brug for en metode til at tage højde for phosphorylering i metaboliske modeller, hvilket fører til udviklingen af en Monod-Wynman-

Changeaux (MWC) analog model af phosphorylering og en kombinatorisk model af phosphorylering. Disse modeller vil yderligere blive brugt til at modellere hexokinase 2 og 6-Phosphofructo-2-Kinase/Fructose 2,6-Bisphosphatase-mekanismen, som blev identificeret som væsentlig for at regulere Warburg-effekten.

Forståelsen af metabolisme opnås bedst gennem kinetisk modellering, og det har vi vist ved at modellere Warburg-effekten i CHO-celler. Implikationerne af modelleringssoftwaren strækker sig ud over vores CHO-glycolysemodel og kan anvendes på hvilket som helst metabolisk netværk for at forstå hvordan reguleringen af disse ikke-lineære systemer forekommer. Det overordnede formål, som denne software stræber efter, er at skabe en informationsdatabase til fremtidig reference og værktøjsudvikling.

Publications

Included in thesis

- Chapters 2 and 4: **Nicholas Luke Cowie**, David Catalán Tatja, Kristian Lund Jensen and Lars Keld Nielsen. A kinetic investigation of the Warburg effect in CHO cells. (Manuscript in preparation)

- Chapter 3: **Nicholas Luke Cowie**[†], Teddy Groves[†] and Lars Keld Nielsen. Maud: a Bayesian kinetic modelling framework. (Manuscript in preparation)

- Chapter 5: **Nicholas Luke Cowie** and Lars Keld Nielsen. Developing a phosphorylation mechanism for kinetic models of metabolism. (Manuscript in preparation)

Not included in thesis

- Marta R A Matos, Pedro A Saa, **Nicholas Luke Cowie**, Svetlana Volkova, Marina de Leeuw, and Lars K Nielsen. GRASP: a computational platform for building kinetic models of cellular metabolism. Bioinform Adv. 2022 Sep 27;2(1):vbac066. doi: 10.1093/bioadv/vbac066. (Published)

[1][†]: equal contribution

---

[1]*

Contents

List of Figures

List of Tables

Abbreviations

	Kinetic Terms		Enzyme
k_cat	Turnover number	PGL	6-phosphogluconolactonase
K_m	Dissociation constant for substrates and products	GND	6-phosphogluconate dehydrogenase
K_i	Dissociation constant for competitive inhibitors	RPI	Ribose-5-phsphate isomerase
L_0	Transfer constant	RPE	Ribulose-5-phosphate epimerase
e_R	Dissociation constant for metabolites that bind to the active state	TKT1	Transketolase 1
e_T	Dissociation constant for metabolites that bind to the inactive state	TKT2	Transketolase 2
$\Delta_f G$	Gibbs energy of formation	TKTL1	Transketolase like-1
$\Delta_r G$	Gibbs energy of reaction	TALA	Transaldolase
MWC	Monod-Wynman-Changeux	AMPK	AMP-activated protein kinase
KNF	Koshland, Némethy and Filmer	Akt	Protein kinase B
ϵ	Regulatory enzyme ratio	PDH	Pyruvate dehydrogenase
Θ	Regulatory saturation ratio	LDH	Lactate dehydrogenase
Γ	Regulatory reversibility ratio	FCCP	carbonyl cyanide-p-trifluoromethoxyphenylhydrazone
Ψ	Regulatory allosteric ratio	GDH	glutamate decarboxylase 2
		PDK	Pyruvate dehydrogenase kinase
	Modelling Terms	PDP	Pyruvate dehydrogenase phosphatase
MFA	Metabolic flux analysis		
FBA	Flux balance analysis		metabolites
pFBA	parsimonious flux balance anlaysis	glc	glucose
S	Stoichiometric matrix	g6p	glucose-6-phosphate
v	reaction flux vector	atp	adenosine triphosphate
GEMMs	Genome scale metabolic models	adp	adenosine diphosphate
		amp	adenosine monophosphate
	Cell Culture	gdp	glucose-1,6-diphosphate
HEK	Human embryonic kidney	g1p	glucose-1-phosphate

CHO	Chinese hamster ovary	mit	proxy for mitochondrial binding
mAb	monoclonal antibodies	nad+	Nicotinamide adenine dinucleotide (oxidised)
wt	Wild type	nadh	Nicotinamide adenine dinucleotide (reduced)
AA	Anti	nadp+	Nicotinamide adenine dinucleotide phosphate (oxidised)
AC	Anti-clumping	nadph	Nicotinamide adenine dinucleotide phosphate (reduced)
HPLC	High-performance liquid-chromatography	f6p	fructose-6-phosphate
MS	Mass-spectrometry	fdp	fructose-1,6-bisphosphate
LC-MS/MS	liquid chromatography-tandem mass spectrometry	dhap	dihydroxyacetone phosphate
BCA	bicinchoninic acid	g3p	glyceraldehyde 3-phosphate
		pi	inorganic phosphate
	Pathways	13dpg	1,3-diphosphoglycerate
EMP	Embden-Meyerhof-Parnas	3pg	3-phosphoglycerate
PP	pentose phosphate	2pg	2-phosphoglycerate
oxPP	oxidative pentose phosphate	pep	phosphoenolpyruvate
noxPP	non-oxidative pentose phosphate	pyr	pyruvate
TCA	Tricarboxylic Acid	6pgl	6-phosphogluconolactonase
		6pgc	6-phospho-D-gluconate
	Enzymes	ru5p	ribulose-5-phosphate
HEX	Hexokinase	r5p	ribose-5-phosphate
HEX1	Hexokinase 1	xu5p	xylulose-5-phosphate
HEX2	Hexokinase 2	e4p	erythrose-4-phosphate
PFK	Phosphofructokinase	s7p	sedoheptulose-7-phosphate
PFKL	Phosphofructokinase type L	cit	citrate
PFKM	Phosphofructokinase type M	f26p	fructose-2,6-diphosphate
PFKFB	6-Phosphofructo-2-Kinase/Fructose 2,6-Bisphosphatase	DCA	Dichloroacetate
FBA	Fructose bisaldolase	lac	lactate
TPI	Triosephosphate isomerase	gln	glutamine

GAPD	Glyceraldehyde-3-phosphate dehydrogenase		
PGK	Phosphoglycerate kinase		Statistical Terms
PGM	Phosphoglycerate mutase	HMC	Hamiltonian Monte Carlo
ENO	Enolase	MCMC	Markov Chain Monte Carlo
PK	Pyruvate kinase	MAP	Maximum a posteriori
PKM1	Pyruvate kinase M1	MLE	Maximum likelihood estimate
PKM2	Pyruvate kinase M2	SSR	Sum of squares residual
G6PDH	Glucose-6-phosphate dehydrogenase		

A kinetic investigation of the Warburg effect in CHO cells

1   Introduction

1.1   The Warburg Effect

It has been nearly 100 years since Otto Warburg first reported the phenomenon of aerobic glycolysis in carcinoma cells (Warburg, 1924). Aerobic glycolysis, also known as the Warburg Effect, is a growth state where sufficient oxygen is available to metabolise glucose in the electron transport chain and tricarboxylic acid (TCA) cycle, a process known as oxidative phosphorylation, but instead glucose is converted into lactate (lac) (see figure 1.1). Aerobic glycolysis is significantly less energy efficient than oxidative phosphorylation, generating 4 adenosine triphosphate (atp) per glucose (glc), compared to the 36 available through oxidative phosphorylation (Young, 2013). Aerobic glycolysis also generates potentially harmful amounts of lactate, whereas oxidative phosphorylation does not accumulate toxic compounds. This leads to the question of why does a cell operate in a less efficient metabolic state that generates significant lactate waste when alternative mechanisms are available?

Warburg himself answered this question by postulating that aerobic glycolysis happens due to defective mitochondria that are unable to carry sufficient flux. A decrease in mitochondrial flux would prevent the conversion of nadh to nad+, and leave pyruvate (pyr) as a dead-end. Therefore, necessitating the production of lactate, as lactate production regenerates nicotinamide adenine dinucleotide oxidised (nad+) from its reduced form (nadh), and is a product of pyruvate, see Figure 1.1. This hypothesis has been disproved in numerous experiments that measured the flux through the TCA cycle in cells exhibiting aerobic glycolysis and found mitochondrial function to be operating normally. TCA cycle flux however, may not entirely be derived from glucose, but through glutaminolysis (Lee et al., 2019; Buchsteiner et al., 2018).

Recent experiments use dichloroacetate (DCA) as a method to divert flux from pyruvate to the TCA cycle by removing inhibition of pyruvate dehydrogenase (PDH) by pyruvate dehydrogenase kinase (PDK). This has been shown to increase the flux through the TCA cycle and decrease lactate secretion (Lee et al., 2019; Buchsteiner et al., 2018; Michelakis et al., 2008). PDH is an enzyme in the TCA cycle that is thought be rate limiting, as is shown in Buchsteiner et al., 2018. Experiments removing PDH inhibition suggest that it is possible for the TCA cycle to carry flux from pyruvate, and that another key mitochondrial function - the transport in of pyruvate - is fully functional in cells exhibiting aerobic glycolysis. Labelling experiments also show a clear incorporation of labelled TCA cycle intermediates from labelled glucose (Young, 2013).

According to a recent hypothesis, the limitation is located in the electron transport chain, which uses nadh to pump hydrogen ions that are then converted into atp by atp synthase. The hypothesis stipulates that the total cellular atp demand was superseded
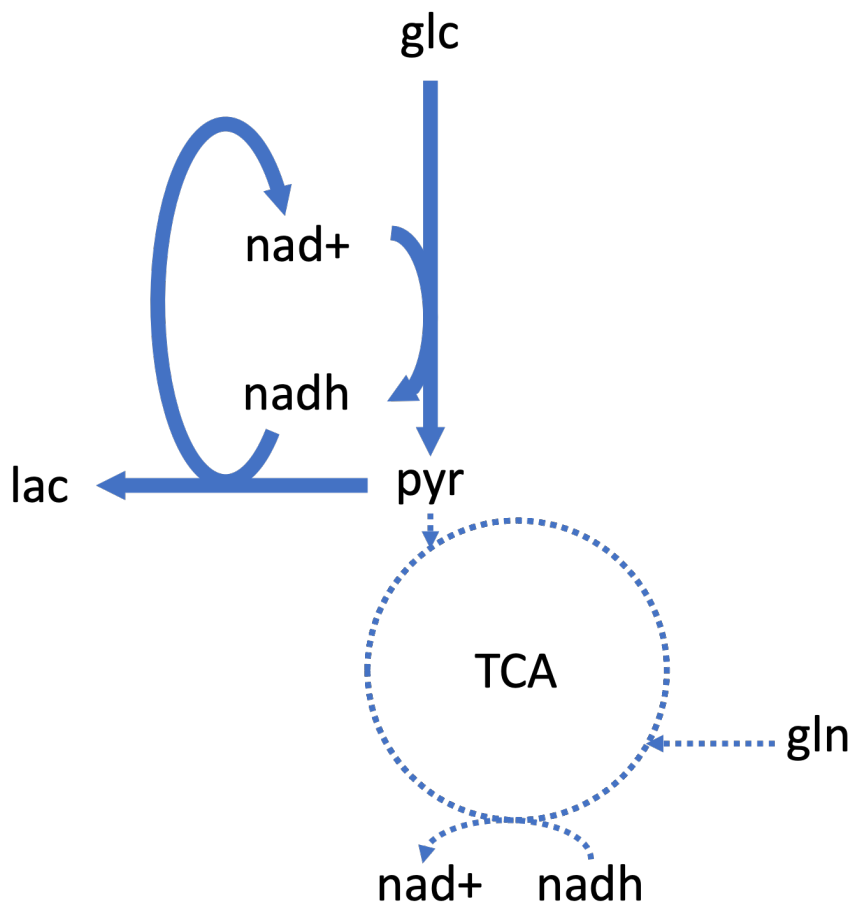
Figure 1.1: Schematic of the Warburg effect where the width of the arrow represents relative flux through the pathway. The Warburg effect is characterised by a large flux from glucose (glc) through to lactate (lac), with minimal conversion directly from pyruvate (pyr) through to the TCA cycle. The conversion of glucose to pyruvate is accomplished primarily through the Embden-Meyerhof-Parnas (EMP) pathway. The nadh generated via the EMP pathway is regenerated both through the tricarboxylic acid (TCA) cycle, and by the production of lac. In mammalian cells the TCA cycle is contained within the mitochondria with the EMP and glutaminolysis occurring in the cytosol. Anapleurotic reactions introduce glutamine (gln) into the TCA cycle, supplying the cell with an additional source of carbon.

by demand for nad+ (Luengo et al., 2021). This theory was supported by an experiment in which production of atp was uncoupled from nadh consumption through the introduction of the mobile ion pore carrier carbonyl cyanide-p-trifluoromethoxyphenylhydrazone (FCCP). FCCP transports hydrogen ions across the mitochondrial membrane and dissipates the driving force for atp production, this decreases atp production and increases the thermodynamic driving force regenerating nad+. Without FCCP, oxidative phosphorylation was promoted in a method similar to that of DCA, and prevented inhibition of PDH by inhibiting PDK which resulted in a decreased growth rate. This was partially alleviated through the inclusion of FCCP that increased the growth rate and restored

A kinetic investigation of the Warburg effect in CHO cells

the ratio of nad+ to nadh (Luengo et al., 2021). The hypothesis that cells exhibiting the Warburg effect is a result of insufficient nadh to nad+ flux is not unique to this article and is proposed by Fernandez-de-Cossio-Diaz and Vazquez, 2017, that oxidative phosphorylation linked glycolysis requires an increasing large proportion of mitochondria to achieve higher growth rates. Both of these articles however, do not attempt to answer how glycolysis is able to dramatically increase its glucose uptake rate, without a change in protein expression (Buchsteiner et al., 2018).

1.2   CHO Cells

Chinese hamster ovary (CHO) cells are the predominant producer of monoclonal antibodies (mAbs) in the biopharmacuetical industry. Their glycosylation patterns are favourable when compared with common producers of proteins such as yeast or *eschiera* coli. Compared to human cell lines, one of the most common being human embryonic kidney (HEK) cells, CHO cells are significantly higher producers. Unfortunately, the production of mAbs requires a large amount of cell culture due to their low production capabilities offered by CHO cells and also a large metabolic requirement. Therefore, mAbs are extremely expensive to produce with the average annual cost of antibody treatment being approximately $100000 (Hernandez et al., 2018). As a consequence of the effective biopharmaceuticals and difficulty of production the mAb industry is expected to grow to $300 billion by the year 2025 as mAbs provide a more targeted approach compared with their small molecule counterparts (Lu et al., 2020). Improving the productivity of CHO cells will decrease the cost for future mAb treatments and provide a benefit to patients who will have access to more options, and governments that may front a large part of the treatment cost.

The Warburg effect phenotype is shared by many fast growing mammalian cells, and is akin to other overflow metabolisms such as the Crabtree effect whereby yeast produces ethanol in high growth conditions. One well-known category of fast growing mammalian cells are cancer cells, but fast growth is also a preferred trait in the biopharmaceutical industry for mAb production. Consequently, the cells produce a large amount of lactate that decreases cell density, production titre and product quality (Dietmair et al., 2011).

CHO cells express the Warburg effect phenotype without significant changes in protein expression, and other than the increased nad+ to nadh ratio, metabolomics of cells exhibiting a decreased Warburg effect suggests that few metabolites in upper glycolysis have different intra-cellular concentrations (Buchsteiner et al., 2018). To explain how regulation of metabolism is achieved in CHO cells that exhibit the Warburg effect we require kinetic models.

1.3   Kinetic modelling of cellular metabolism

Cellular metabolism is the process by which a cell fuels cellular growth and any additional function through metabolic reactions. These reactions are carried out by en-

zymes, catalysts that dramatically increase the rate of reaction and have a high degree of specificity for their substrates. The rate of an enzyme-catalysed reaction can be described mathematically using a rate law; common rate laws include: Michaelis-Menten kinetics, the modular rate law, hill kinetics, and mass-action kinetics (Liebermeister et al., 2010; Michaelis and Menten, 1913; Jamshidi and Palsson, 2010; Hill, 1910). For example, Figure 1.2 presents four equations, each of which converts its substrate into its product according to the rate law f. The arguments to f include: substrate concentrations $[X_i]$, reaction flux v_i, total enzyme concentration E_i, and a maximal rate k_{cat}^i for each reaction i.

Rate laws can be any representation of an assumed mechanism and often include thermodynamic information, regulatory interactions, and pH-correction factors (P. Saa and Nielsen, 2015; Peskov et al., 2012). By combining together multiple rate laws it is possible to describe the behaviour of a network of interconnected metabolic reactions. In particular it is possible to describe how the concentrations of metabolites in the network change over time and in some instances systems will reach a so called steady state, where the metabolite concentrations are no longer changing and fluxes remain constant (see Figure 1.2). These systems of equations that describe cellular metabolism will be referred to as kinetic models.

$$v_{A_e \to A} = f(k_{cat}^1, [E_1], [A_e], K_m^A)$$

$$A_e \ \to A \ \to B \ \to C \to C_e$$

$$v_{A \to B} = f(k_{cat}^2, [E_2], [A], K_m^A)$$

$$v_{B \to C} = f(k_{cat}^3, [E_3], [B] \, K_m^B)$$

$$v_{C \to C_e} = f(k_{cat}^4, [E_4], [C], K_m^C)$$
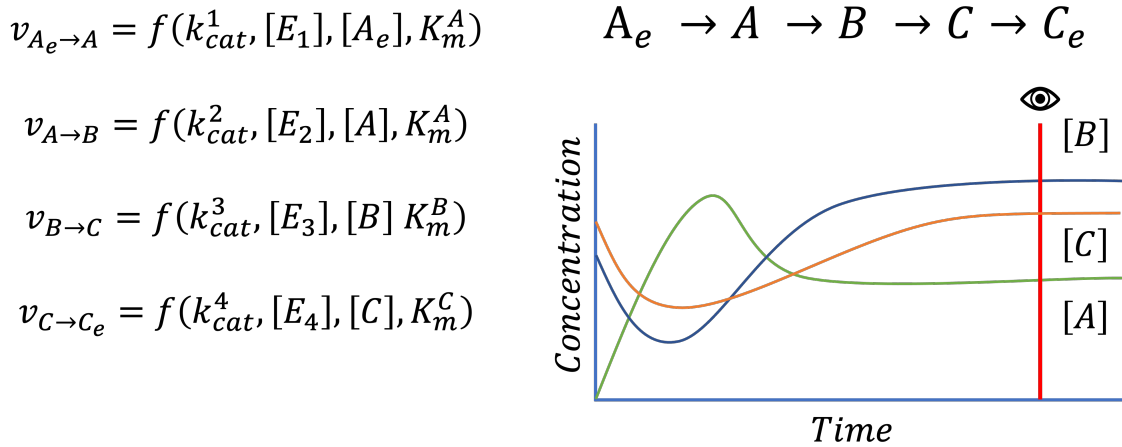
Figure 1.2: A description of a linear pathway that imports metabolite A, converts it to B, then C, before exporting metabolite C to the extracellular medium, as shown by the network description in the top right. The rate of each reaction is modelled as a function of the reaction's maximum catalytic rate k_{cat}^i, where i is the reaction being catalysed, the concentration E_i of the catalysing enzyme, the concentration of the reactant's substrate A, B, or C, and the dissociation constant K_m^X, where X is the given metabolite. The plot in the lower right depicts a hypothetical timecourse of the metabolite concentrations, A, B and C as they progresses to a steady state. The steady state often the metabolic state observed in the cell and is depicted by the red line.

The way in which reaction fluxes change metabolite concentrations in a metabolic network can be compactly represented as shown in equation 1.1 and is represented by $\frac{dC}{dt}$. In equation 1.1 S is a stoichiometric matrix that is an m by n matrix, where m is

the number of metabolites in the network, n is the number of reactions and S_{ij} is the stoichiometric coefficient of metabolite i in reaction j. The flux vector v is a vector of length n where v_j is the flux through reaction j; in the case of kinetic models, each entry to the flux vector is described by a rate law. The term $C \cdot \mu$ represents the dilution of metabolites due to growth, where C is a vector of concentrations of length m, and μ is the growth rate of an organism. In most metabolic networks where the metabolite concentrations are small relative to the flux of the surrounding network, the dilution term is negligible and can be ignored. It is typically safe to assume that dilution is negligible unless particularly concentrated metabolites are modelled, such as glycogen in the liver.

$$S \cdot v - C \cdot \mu = \frac{dC}{dt} \tag{1.1}$$

$$S \cdot v = \frac{dC}{dt} \tag{1.2}$$

Equation 1.1 defines the behaviour of metabolism at any point of a biological process. However, generating experimental data and measuring the rate of change of metabolites is particularly difficult experimentally. Instead it is common to investigate states that are at a so called pseudo-steady state, where the rate of change of metabolites is negligible and operate on timescales much faster than the change in enzyme concentrations. Making the psuedo-steady state assumption simplifies equation 1.1 to equation 1.3, where $\frac{dC}{dt} = 0$.

$$S \cdot v = 0 \tag{1.3}$$

The benefits that kinetic models provide are:

1. model prediction for experimentally untested states, effectively creating an *in silico* cell, and

2. If an approximately correct mechanistic representation is used, regulation of the system can be inferred.

However, constructing a kinetic model is a tedious process: literature reviews of all relevant enzymes and their regulation to construct rate laws, searches for parameter estimates, solver optimisation and multiomics data collection.

Realistic kinetic models often have no analytical solution to the steady-state, and consequently require numerical solvers. While there are many numerical solvers available, kinetic models typically present a large timescale separation between fast and slow reactions. Consequently, solvers that target stiff ODE systems tend to perform best for kinetic models (Garfinkel et al., 1977; Städter et al., 2021).

The most recent metabolic networks for mammalian cell metabolism as described in Recon3D, which considers 13,543 reactions Brunk et al., 2018. Numerically Solving an ODE system representing all of these reactions with realistic rate laws would be computationally infeasible and modelling endeavours must be targeted towards specific subsystems of interest.

There are frameworks that scale well for parameterising large networks, which is referred to as ensemble modelling (P. Saa and Nielsen, 2015; Gopalakrishnan et al., 2020; Tran et al., 2008; Chakrabarti et al., 2013). Ensemble modelling differs from traditionally kinetic model construction where multiomics data is sampled and is used to generate the parameter set that goes exactly through the sampled data. The consequence of this processes is that the solution is guaranteed only for the reference state. However, validation and parameter inference of these models still requires the solution to an ODE that becomes increasingly costly to evaluate as model size increases (Gopalakrishnan et al., 2020; P. A. Saa and Nielsen, 2016). A description of different modelling frameworks can be found in P. A. Saa and Nielsen, 2017 as well as in Chpater 3.

1.4   Thesis Motivation and Outline

This thesis is motivated by the need for a kinetic model that can explain glucose metabolism in CHO cells with sufficient multiomics data and answer the question: *How does glycolysis regulate metabolism at every single enzyme with an seemingly minimal change in its proteome and fluxome?*.

Firstly, we developed an experimental workflow allowing relatively high throughput collection of multiomics data for CHO cells (Chapter 2). Included in the cell lines was a strain of CHO that did not exhibit the Warburg effect and will be used as a contrasting cell line throughout this thesis. The 15 mL bioreactors in the Ambr15 (Sartorius) system allowed a single reactor to collect metabolomics, proteomics, and exometabolomics, and afterwards, possessed enough volume to allow us to conduct 13C metabolic flux analysis (13C-MFA). Comparative multiomics was performed and suggested that there was no correlation between glycolytic flux and enzyme concentration, and that there were minimal changes in metabolite concentration. Combined, these measurements indicate that that CHO cell glycolysis is a homeostatic system that relies on post-translational modifications.

Chapter 3 presents Maud, a new software application for analysing metabolic networks using kinetic models of metabolism that employs Bayesian statistics to infer parameters and for model prediction. Maud is a Python package that is integrated with Stan to perform Bayesian inference via Hamiltonian Monte Carlo (HMC) sampling. The kinetic model formalism is based on the modular rate law Liebermeister et al., 2010, complemented with allostery as described by Monod-Wynman-Changeaux (MWC) (P. Saa and Nielsen, 2015). In this work we show that point summary estimates are in-

sufficient to capture model behaviour and a full Bayesian evaluation is required. We also highlight the benefits of posterior sampling as generated properties require no additional assumptions.

Chapter 4 applies Maud to analyse the Warburg effect in CHO cells. We constructed a model of the Embden-Meyerhof-Parnas (EMP) and pentose phosphate (PP) pathways that captures model behaviour that and was validated using two conditions. Due to time constraints, we used the maximum *a posteriori* (MAP) estimate an an estimate of model behaviour rather than performing full HMC sampling. To explore regulation in the Warburg effect, the CHO-S and CHO-ZeLa strains were compared, as the CHO-ZeLa strain lacks the Warburg effect phenotype. We show that the Warburg effect requires allosteric activation of Hexokinase 2 (HEX2) and Phosphofructokinase M (PFKM) to accommodate the large increase in flux. Furthermore, we pinpoint regulatory interactions that are in need of further investigation such as glyceraldehyde-3-phosphate dehydrogenase (GAPD) and triosephosphate isomerase (TPI), which are poorly described by the multiomics data.

Regulation of the important enzymes PFKM and HEX2 is achieved through phosphorylation: whereby a set of enzymes known as kinases and phosphatases modulate the activity of metabolic enzymes through the addition of a phosphate group. The metabolic modelling field currently lacks an adequate toolset for mathematically describing this regulatory mechanism. The toolset to model phosphorylation is necessary to account for the regulation of PFKM and HEX2, and is currently lacking in the metabolic modelling field. Chapter 5 addresses this with proposed frameworks to model phosphorylation as a function of kinase and phosphatase activity. Of the three proposed mechanisms two cannot be discounted as they both describe cooperativity observed in literature. However, the most flexible approach was achieved by modelling phosphorylation similar to the allosteric events captured by Monod-Wynman-Changeaux (MWC) model.

Finally, Chapter 6 concludes this thesis with a discussion of the future prospects of kinetic modelling of the Warburg effect, cellular metabolism as a whole and the importance of multiomics datasets for metabolic modelling as a field.

References

Warburg, O. (1924). �ber den Stoffwechsel der Carcinomzelle. *Die Naturwissenschaften*, *12*(50), 1131–1137. https://doi.org/10.1007/BF01504608

Young, J. D. (2013). Metabolic flux rewiring in mammalian cell cultures. *Current Opinion in Biotechnology*, *24*(6), 1108–1115. https://doi.org/10.1016/j.copbio.2013.04.016

Lee, W. D., Mukha, D., Aizenshtein, E., & Shlomi, T. (2019). Spatial-fluxomics provides a subcellular-compartmentalized view of reductive glutamine metabolism in cancer cells. *Nature Communications*, *10*(1), 1351. https://doi.org/10.1038/s41467-019-09352-1

Buchsteiner, M., Quek, L.-E., Gray, P., & Nielsen, L. K. (2018). Improving culture performance and antibody production in CHO cell culture processes by reducing the Warburg effect. *Biotechnology and Bioengineering*, *115*(9), 2315–2327. https://doi.org/10.1002/bit.26724

Michelakis, E. D., Webster, L., & Mackey, J. R. (2008). Dichloroacetate (DCA) as a potential metabolic-targeting therapy for cancer. *British Journal of Cancer*, *99*(7), 989–994. https://doi.org/10.1038/sj.bjc.6604554

Luengo, A., Li, Z., Gui, D. Y., Sullivan, L. B., Zagorulya, M., Do, B. T., Ferreira, R., Naamati, A., Ali, A., Lewis, C. A., Thomas, C. J., Spranger, S., Matheson, N. J., & Vander Heiden, M. G. (2021). Increased demand for NAD+ relative to ATP drives aerobic glycolysis. *Molecular Cell*, *81*(4), 691–707.e6. https://doi.org/10.1016/j.molcel.2020.12.012

Fernandez-de-Cossio-Diaz, J., & Vazquez, A. (2017). Limits of aerobic metabolism in cancer cells. *Scientific Reports*, *7*(1), 13488. https://doi.org/10.1038/s41598-017-14071-y

Hernandez, I., Bott, S. W., Patel, A. S., Wolf, C. G., Hospodar, A. R., Sampathkumar, S., & Shrank, W. H. (2018). Pricing of monoclonal antibody therapies: Higher if used for cancer? *The American Journal of Managed Care*, *24*(2), 109–112.

Lu, R.-M., Hwang, Y.-C., Liu, I.-J., Lee, C.-C., Tsai, H.-Z., Li, H.-J., & Wu, H.-C. (2020). Development of therapeutic antibodies for the treatment of diseases. *Journal of Biomedical Science*, *27*(1), 1. https://doi.org/10.1186/s12929-019-0592-z

Dietmair, S., Nielsen, L. K., & Timmins, N. E. (2011). Engineering a mammalian super producer. *Journal of Chemical Technology & Biotechnology*, *86*(7), 905–914. https://doi.org/10.1002/jctb.2576

Liebermeister, W., Uhlendorf, J., & Klipp, E. (2010). Modular rate laws for enzymatic reactions: Thermodynamics, elasticities and implementation. *Bioinformatics*, *26*(12), 1528–1534. https://doi.org/10.1093/bioinformatics/btq141

Michaelis, L., & Menten, M. (1913). Die kinetik der invertinwirkung. *Biochem. z*, *49*, 333–369.

Jamshidi, N., & Palsson, B. Ø. (2010). Mass action stoichiometric simulation models: Incorporating kinetics and regulation into stoichiometric models. *Biophysical Journal*, *98*(2), 175–185. https://doi.org/10.1016/j.bpj.2009.09.064

Hill, A. (1910). PROCEEDINGS OF THE PHYSIOLOGICAL SOCIETY: January 22, 1910. *The Journal of Physiology*, *40*, iv–vii. https://doi.org/10.1113/jphysiol.1910.sp001386

Saa, P., & Nielsen, L. K. (2015). A general framework for thermodynamically consistent parameterization and efficient sampling of enzymatic reactions. *PLoS Computational Biology*, *11*(4), e1004195. https://doi.org/10.1371/journal.pcbi.1004195

Peskov, K., Mogilevskaya, E., & Demin, O. (2012). Kinetic modelling of central carbon metabolism in Escherichia coli. *The FEBS Journal*, *279*(18), 3374–3385. https://doi.org/10.1111/j.1742-4658.2012.08719.x

Garfinkel, D., Marbach, C. B., & Shapiro, N. Z. (1977). STIFF DIFFERENTIAL EQUATIONS. *Annual Review of Biophysics and Bioengineering*, *6*(1), 525–542. https://doi.org/10.1146/annurev.bb.06.060177.002521

Städter, P., Schälte, Y., Schmiester, L., Hasenauer, J., & Stapor, P. L. (2021). Benchmarking of numerical integration methods for ODE models of biological systems. *Scientific Reports*, *11*(1), 2696. https://doi.org/10.1038/s41598-021-82196-2

Brunk, E., Sahoo, S., Zielinski, D. C., Altunkaya, A., Dräger, A., Mih, N., Gatto, F., Nilsson, A., Preciat Gonzalez, G. A., Aurich, M. K., Prlić, A., Sastry, A., Danielsdottir, A. D., Heinken, A., Noronha, A., Rose, P. W., Burley, S. K., Fleming, R. M. T., Nielsen, J., … Palsson, B. O. (2018). Recon3D enables a three-dimensional view of gene variation in human metabolism. *Nature Biotechnology*, *36*(3), 272–281. https://doi.org/10.1038/nbt.4072

Gopalakrishnan, S., Dash, S., & Maranas, C. (2020). K-FIT: An accelerated kinetic parameterization algorithm using steady-state fluxomic data. *Metabolic Engineering*. https://doi.org/10.1016/j.ymben.2020.03.001

Tran, L. M., Rizk, M. L., & Liao, J. C. (2008). Ensemble modeling of metabolic networks. *Biophysical Journal*, *95*(12), 5606–5617. https://doi.org/10.1529/biophysj.108.135442

Chakrabarti, A., Miskovic, L., Soh, K. C., & Hatzimanikatis, V. (2013). Towards kinetic modeling of genome-scale metabolic networks without sacrificing stoichiometric, thermodynamic and physiological constraints. *Biotechnology Journal*, *8*(9), 1043–1057. https://doi.org/10.1002/biot.201300091

Saa, P. A., & Nielsen, L. K. (2016). Construction of feasible and accurate kinetic models of metabolism: A Bayesian approach. *Scientific Reports*, *6*, 29635. https://doi.org/10.1038/srep29635

Saa, P. A., & Nielsen, L. K. (2017). Formulation, construction and analysis of kinetic models of metabolism: A review of modelling frameworks. *Biotechnology advances*, *35*(8), 981–1003. https://doi.org/10.1016/j.biotechadv.2017.09.005

2 Multiomics analysis of CHO cells

2.1 Rationale

Eliminating the Warburg effect whilst maintaining productivity is one of the goals in CHO cell line engineering. As part of this effort, the CHO Bioprocess Laboratory at the Danish Technical University (DTU) succeeded in engineering a non-Warburg cell line called CHO-ZeLa, derived from the parent strain CHO-S. Characterisation of this cell lines metabolism is described in Monge et al., 2020.

CHO-ZeLa is a rationally designed cell line that alleviated the Warburg effect by knocking out the enzyme lactate dehydrogenase as well as the isoforms of pyruvate kinase dehydrogenase. This is the first successful knockout of lactate dehydrogenase in CHO cells and it exhibits several important differences in metabolism compared with the parent strain CHO-S. Specifically, CHO-ZeLa maintains comparable growth rate with CHO-S while completely avoiding lactate production and consuming 50% less glucose. CHO-ZeLa is particularly interesting in light of the hypothesis that lactate production is required in order to regenerate the essential cofactor nad+ at such high growth rates (Luengo et al., 2021; Fernandez-de-Cossio-Diaz and Vazquez, 2017).

The availability of both CHO-S and CHO-ZeLa provides a unique opportunity to study the metabolic processes that cause the Warburg effect. This is because the two cell lines are genetically identical except for relatively few known knockouts, making the comparisons of the two metabolisms easier. In addition, both cell lines were available for multiomics analysis at DTU Biosustain. We therefore conducted a detailed quantitative multimics analysis comparing CHO-ZeLa with several related strains and used the resulting datasets to fit a consolidated Bayesian kinetic model.

This chapter provides an account of the collected multiomics data and describes how it was acquired and analysed. The final result was a dataset that could be used to train a Bayesian kinetic model.

2.2 Designing a multiomics workflow for kinetic modelling applications

Designing a multiomics study for a kinetic modelling input has requirements that are specific with respect to data collection. To leverage the vast amounts of kinetic and thermodynamic information available in the literature, it is important to use units that correspond with the conventional units used in the literature. Therefore, all rates and concentrations were recorded per litre of cell volume.

Another consideration when designing a multiomics study for kinetic modelling is to collect the largest and most diverse dataset possible. A larger and more diverse train-

ing set will better constrain the model's posterior distribution and improve statistical inference. A larger and more diverse test set will allow for better validation of the model's predictive power, giving an idea about how well the model's conclusions will generalise.

For the purpose of a diverse dataset we selected six cell lines and performed a multiomics analysis on glycolysis with 4 replicates per strain. The additional cell lines were engineered to have either significant differences in metabolism, or are outliers from the CHO-S wild type behaviour. To our knowledge, this is the largest CHO glycolysis dataset that considers proteomics, endometabolomics, and fluxomics using 13C MFA.

I order to improve the quality of the data, the fermentations were conducted in the Ambr15 (Sartorius) platform which controlled pH, temperature, and dissolved oxygen. This control ensures constant environmental conditions throughout the fermentation, which reduces the impact of phenotype changes due to the environment.

We compared all cell lines during the early exponential growth phase for two reasons: 1) the Warburg effect is typically observed in the exponential phase, and 2) The cellular phenotype is assumed to be constant during the early exponential phase, facilitating modelling.

Cell lines

The cell lines characterised in this work are derivatives of CHO-S cells (Gibco) except for ExpiCHO (Gibco). The CHO-S strains used for this study were as follows: CHO-S wild type (CHO-S wt), CHO-ZeLa, CHO-Zen, CHO-ZenZeLa, and CHO-S 11 secretory knockout (CHO-S 11 KO).

A brief description of each cell line is as follows:

CHO-S wt: The wild type CHO-S cell line was used as a positive case of the Warburg effect. CHO-S wt is an adapted suspension cell line derived from the original CHO cell line (Puck et al., 1958). Compared to CHO-K1 and CHO-DG44, other cell lines derived from the original cell line, CHO-S wt has a higher growth rate and decreased specific productivity (Reinhart et al., 2019).

CHO-ZeLa: A clonal variant of the CHO-S wt that was rationally engineered to eliminate the Warburg effect phenotype through knockouts of lactate dehydrogenase A, and pyruvate dehydrogenase kinase isoenzymes. This cell line has abolished lactate production whilst maintaining a similar growth rate (Monge et al., 2020).

CHO-Zen: CHO-Zen was an effort to improve growth rates and production rates by decreasing ammonium production through disruption of glutamate decarboxylase 2 (GDH) (Ley et al., 2019). The altered amino-acid metabolism may impact the flux through glycolysis and hence should provide a contrasting cell line to train the model.

CHO-ZenZeLa: CHO-ZenZeLa is a cell line that combines the LDH and PDK knockouts of CHO-ZeLa with the GDH knockouts of CHO-Zen and is a recent project of the CHO Bioprocess Laboratory at DTU Center for Biosustainability. The objective of this cell line is to decrease the inhibitory effects of both lactate and ammonium production to achieve higher cell density and antibody production. This cell line is an intermediate cell line in the generation of a full CHO-ZenZeLa cell line as it has partially functional lactate production. Nevertheless, this cell line represents an interesting phenotype that should contrast with the CHO-S wt and CHO-ZeLa cell lines for model training.

ExpiCHO: The ExpiCHO expression system was developed by Gibco to dramatically increase protein expression in the CHO-S cell line. The cell line grows at a similar rate to CHO-S and exhibits protein expression up to 50-fold higher than Freestyle CHO-S expression systems (Jain et al., 2017). This cell line requires the ExpiCHO growth medium rather than the CD CHO medium used for the other cell lines. The ExpiCHO growth medium is propietary and owned by Gibco.

CHO-S 11 KO: CHO-S 11 KO is a variant of CHO-S wt with 11 host cell proteins knocked out. The cell line was characterised in Kol et al., 2020, 2020 and showed a higher growth rate than CHO-S wt with decreased host cell protein production. This cell line was included in order to provide a contrasting phenotype with CHO-S wt.

2.2.1 Fermentation

The cells were fermented in the ambr15 (Sartorius) system with a working volume of up to 15 mL and an rpm of 900. The base medium was CD-CHO (Gibco) except of the ExpiCHO cell line for which the ExpiCHO medium (Gibco) was used. All media were supplied with 8 mM glutamine (Sigma), antibiotic-antimycotic (AA) (Gibco), anti-clumping (AC) (Gibco), and anti-foam (AF) (Sigma). A schematic of the multiomics collected throughout the fermentation is shown in figure 2.1.
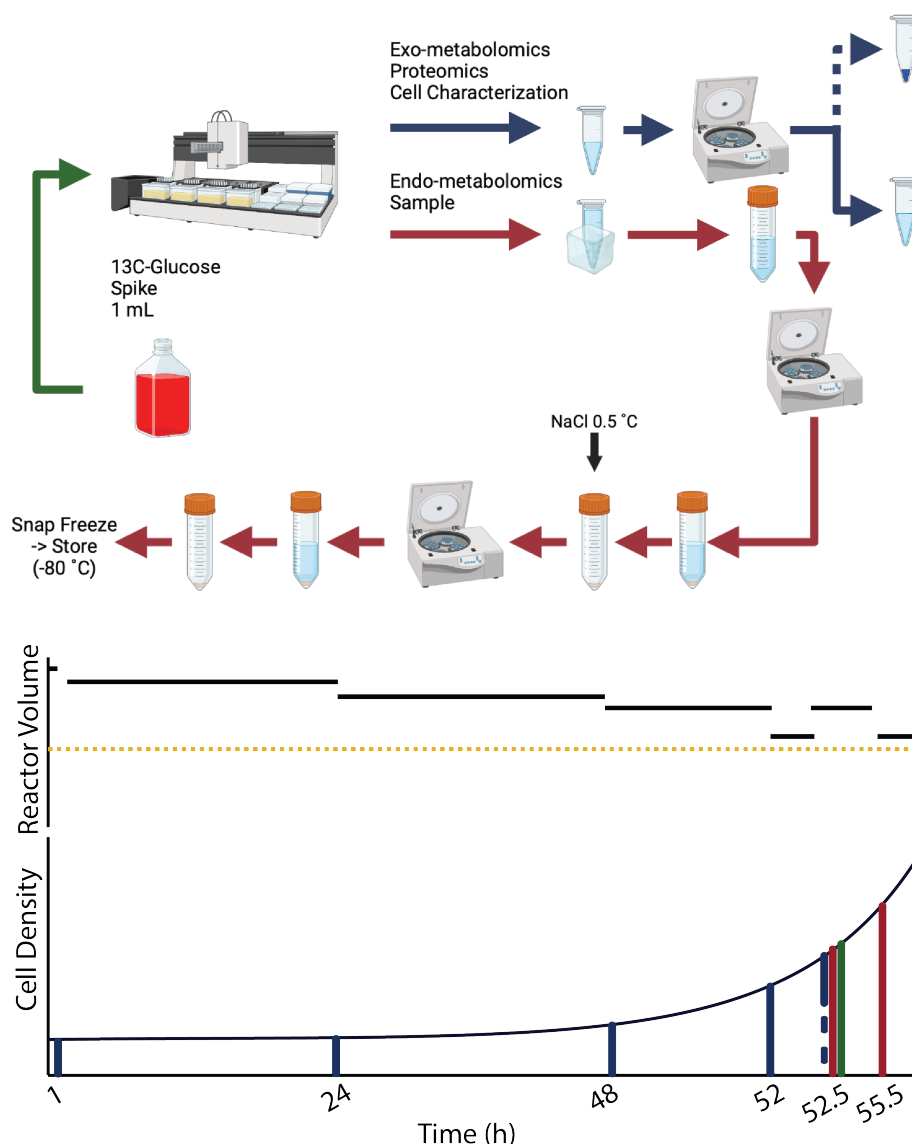
Figure 2.1: Schematic illustrating the multiomics workflow. The top picture represents the possible sample processing steps. The blue arrows are the steps used to collect samples for exometabolomics, proteomics, and for cell characterisation. A 500 μL sample of cell culture is taken from the bioreactor with 50 μL used to determine cell diameter, density, and viability. The remaining culture is centrifuged, with the supernatant used for exometabolomics and the pellet used for proteomics. The red arrows represent the steps used for labelled and unlabelled endometabolomics. Each requires a 1 mL sample that is directly quenched into ice cold saline solution, which is stored on a cold plate surrounded by an ice slurry. The sample then undergoes multiple centrifugation, decanting and washing steps before the pellet is stored at -80°C. Finally, the green arrow represents the labelled $[1, 2^{13}C - Glucose]$ addition that is conducted after the first 52 hours of fermentation. The bottom figure depicts the sampling schedule for each process shown in the top figure, with time on the x-axis of both plots. The lower plot shows a hypothetical cell density line corresponding to the early exponential growth phase. Coloured vertical lines matching the arrows in the top figure indicate the when the various sampling steps occurred. The upper plot has additional representative lines showing a typical reactor volume timecourse. The sampling method was optimised such that the minimal reactor volume (yellow dotted line) is never reached (8.5 mL)

This workflow was designed to collect 13C-endometabolomics, absolute endometabolomics, exometabolomics, and proteomics from a single bioreactor. The cell lines were fermented in the Ambr15 system in two distinct phases. The first phase occurs during the first 52 hours where the objective is to measure the endometabolomics, proteomics, and determine the exchange fluxes. The objective of the second phase was to collect labelled endometabolomics samples for 13C MFA. During the second phase the cell lines were spiked with $[1, 2^{13}C - Glucose]$; the labelled endometabolomics samples were taken 3 hours after spiking as suggested by Ahn and Antoniewicz, 2013, for CHO cells. A limitation of this workflow is that only one metabolic state can be examined at a time because introducing labelled glucose will shift the masses in mass spectrometry, preventing quantification of metabolite and protein abundances.

2.3 Results: a snapshot of CHO glycolysis multiomics

p After the fermentation was completed using the workflow defined in Figure 2.1, the samples were processed using the methods described in Section 2.4. With an overview of the mulitomics presented in the following section.

2.3.1 Identifying the flux map of CHO cells

The flux map of CHO cells is a result of the following components: the extracellular uptake and production rates (Figure 2.2), the growth rate (Figure 2.3), and the endo- and exo-metabolomics labelling states (Appendix D).

The intracellular fluxes are important for determining the metabolic regulation of CHO cells as glucose can either be processed through the EMP pathway or the PP pathway. We processed this data using the INCA software package that determines the best estimate with the 95% confidence intervals calculated using the Hessian (Young, 2014). The best estimate of the set of fluxes is determined by using the sum of squared residuals (SSR) as the objective function. Further analysis of the confidence interval is achieved through the profile likelihood function that is similar to the approach defined in Antoniewicz et al., 2006.

Significant differences based on the 95% confidence intervals were found between the EMP and PP pathways across cell lines. The statistical analysis suggested that four of the six strains were under-fit, having an SSR higher than that of the χ^2 distribution with the corresponding degrees of freedom, whilst the other two strains were within the expected bounds (Table 2.1). The associated mean fluxes and confidence intervals as calculated using the profile likelihood function in INCA are presented in Figure 2.5.

The following measurements were removed from the analysis: CHO-ZeLa (replicate B), CHO-S wt (replicate C) and CHO-Zen 6pgc (replicate A). These were removed from the study due to gross SSR measurements when fitting, and in the case of CHO-S wt replicate C, the tracer information was unavailable. Gross error was determined on a case-by-case basis.
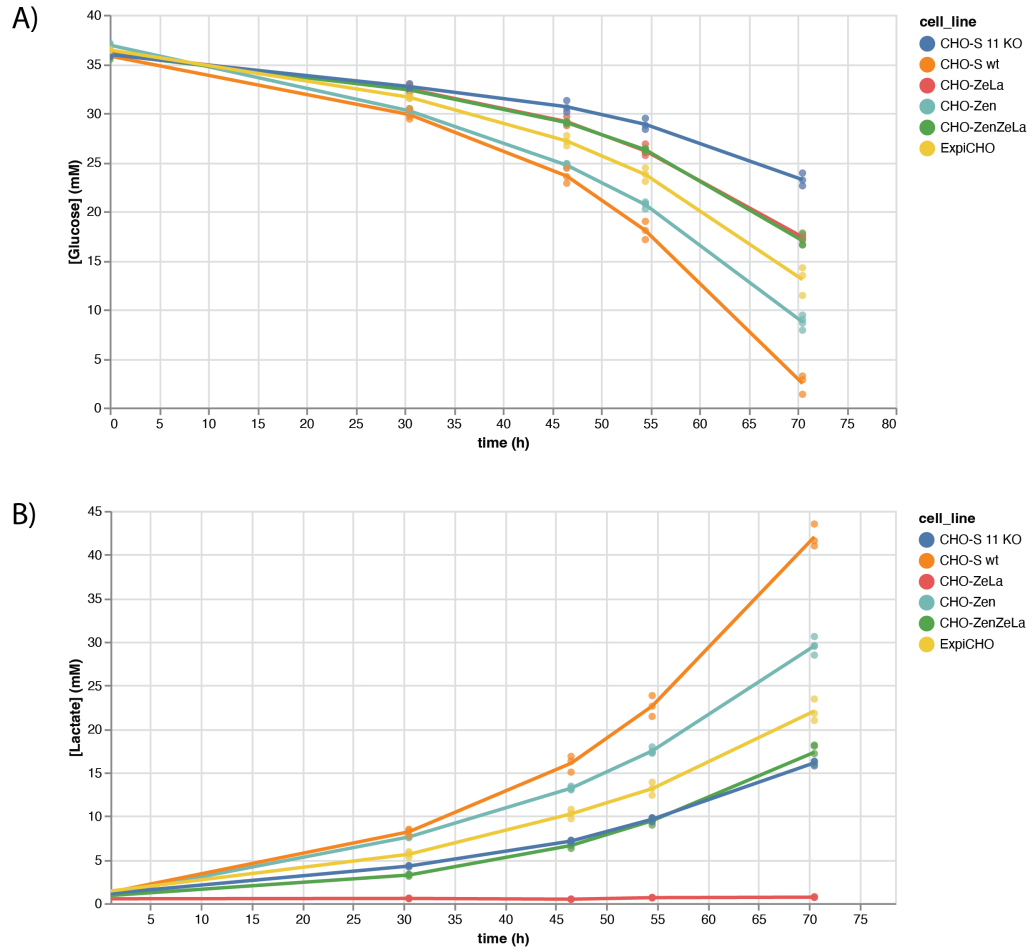
Figure 2.2: Timecourse exometabolomics measurements of glucose and lactate over a 72 hour period carried out in triplicate. A) The lines represent the mean of the exometabolomics measurements with the experimental dataset overlayed. B) Corresponding lactate productions for each cell line with the individual measurements and arithmetic means. The CHO-ZeLa cell line presented essentially zero lactate production over the entire timecourse.

Cell Line	SSR	SSR Range	DOF	Was SSR within Range
CHO-S wt	160.2	48.8 - 95.0	70	No
CHO-ZeLa	141.3	48.8 - 95	70	No
CHO-Zen	124	60.5 - 111.2	84	No
CHO-ZenZeLa	79.7	73.7 - 128.4	99	Yes
CHO-S 11 KO	69.1	39.7 - 82.1	75	Yes
ExpiCHO	150.9	61.4 - 112.4	85	No

Table 2.1: Fitting statistics of the fluxomics dataset *in silico* fitting of the fluxomics data. The Degree of Freedom (DOF) are determined by the number of measurements of each isotope for each cell line. Of the six fluxomics analysis conducted four of the datasets failed to coverage within the SSR range.

Growth rates are similar between all cell lines

Growth rates were estimated on a volumetric basis rather than cell number as CHO cells are known to change cell size throughout a fermentation (Pan et al., 2017). As

each reactor is a realisation of potential biological behaviour, we processed these observations using a combination of a linear regression to determine the reactor specific growth rate and a hierarchical model to calculate the 95 % interval of each cell line. All cell lines lines tested showed similar growth rates, with the exception of CHO-S 11 KO and ExpiCHO that that had a mean growth rate of 0.030 h^{-1} and 0.042 h^{-1} respectively (figure 2.3). The observed growth rates are similar to that of other CHO-S wt strains that had a measured growth rate of 0.037 h^{-1} (Széliová et al., 2020). Furthermore, our data also aligns with the observation that CHO-S and CHO-ZeLa have similar growth rates (Monge et al., 2020).
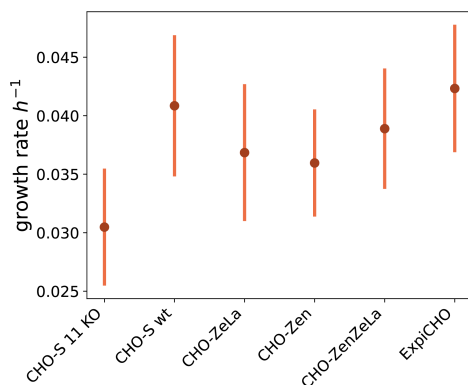
Figure 2.3: Growth rates of cell lines throughout the exponential phase (0-72 h). The points represent the mean of the sampling interval and the 2.5% to 97.5% confidence interval is denoted by the bars. Confidence intervals were calculated using Section 3.2.7

CHO-ZeLa varies in cell volume throughout the exponential phase

Throughout the exponential phase all cell lines apart from CHO-ZeLa maintained a constant cell volume as shown in figure 2.4. CHO cells are known to change volume throughout the growth phase, although this is observed primarily after 96 hours as the cells stop dividing and can increase the cellular volume three fold, yet this shift occurs during the initial phase of the fermentation (Pan et al., 2017). A possible explanation for this volume shift could be that changes in the reactor environment as the CHO-ZeLa cell line transitioned from the seedchain to reactor setting, e.g. pH is regulated in the bioreactor phase and not in the pH environment. These observations will not have an impact on the uptake rates as these parameters are standardised to per litre of cell volume.

EMP is significantly decreased in CHO-ZeLa compared to CHO-S wt

Comparing the EMP fluxes shown in figure 2.5 indicates that CHO-S wt has a significantly higher EMP flux than in CHO-ZeLa. This is consistent with previous comparisons of the two cell lines and the increased glucose uptake rate (Monge et al., 2020). Conversely, this increase in EMP flux does not translate to a higher anapleurotic rate from glucose, with the observed CHO-ZeLa flux being higher than the CHO-S cell line. This flux map however, does not consider the possible recycling of TCA cycle interme-
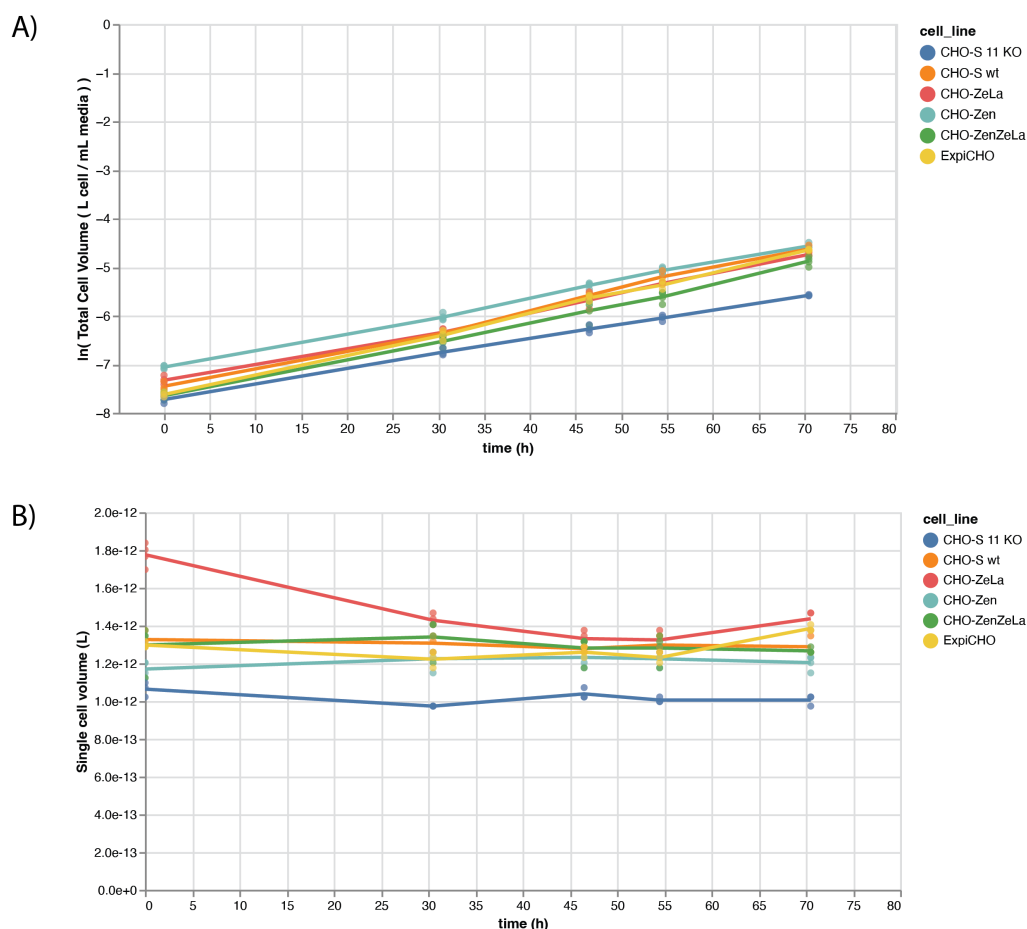
Figure 2.4: Cell growth characteristics throughout the fermentation. A) The volume of each reactor replicate represented on the natural logarithmic scale. B) The single cell volume for each reactor for each timepoint with the mean of each cell line plotted as a line.

diates back to pyruvate and require an extended metabolic map with additional tracer experiments to investigate further (Ahn and Antoniewicz, 2013; Quek et al., 2010).

CHO-ZeLa has a decreased oxidative PP (oxPPP) split ratio compared to CHO-S wt
The oxPPP split ratio is the amount of flux present in the oxPPP compared to the EMP pathway. Flux through this pathways is used to either generate PRPP for biosynthesis of nucleotides or to regenerate the reduced form of nicotinamide adenine dinucleotide phosphate (nadph), which regulates oxidative stress and is also used for biosynthesis of deoxyribonucleotides and lipids (Lunt and Vander Heiden, 2011). The oxPPP split ratio in CHO-ZeLa has an average of 13% and is lower than the 18% split ratio observed in CHO-S wt. Furthermore, the oxPPP flux carried in CHO-ZeLa is significantly lower than that of CHO-S wt suggesting differences in oxPPP regulation. The remaining cell lines encompass a range of split ratios and fluxes, with the highest average ratio of 27% observed in the CHO-Zen line and the lowest of 10% in the CHO-ZenZeLa cell line.

Previous studies by Ahn and Antoniewicz, 2013 and Monge et al., 2020 have suggested that CHO cells have an oxPPP split ratio of approximately 0.5% and 8.7% respectively.
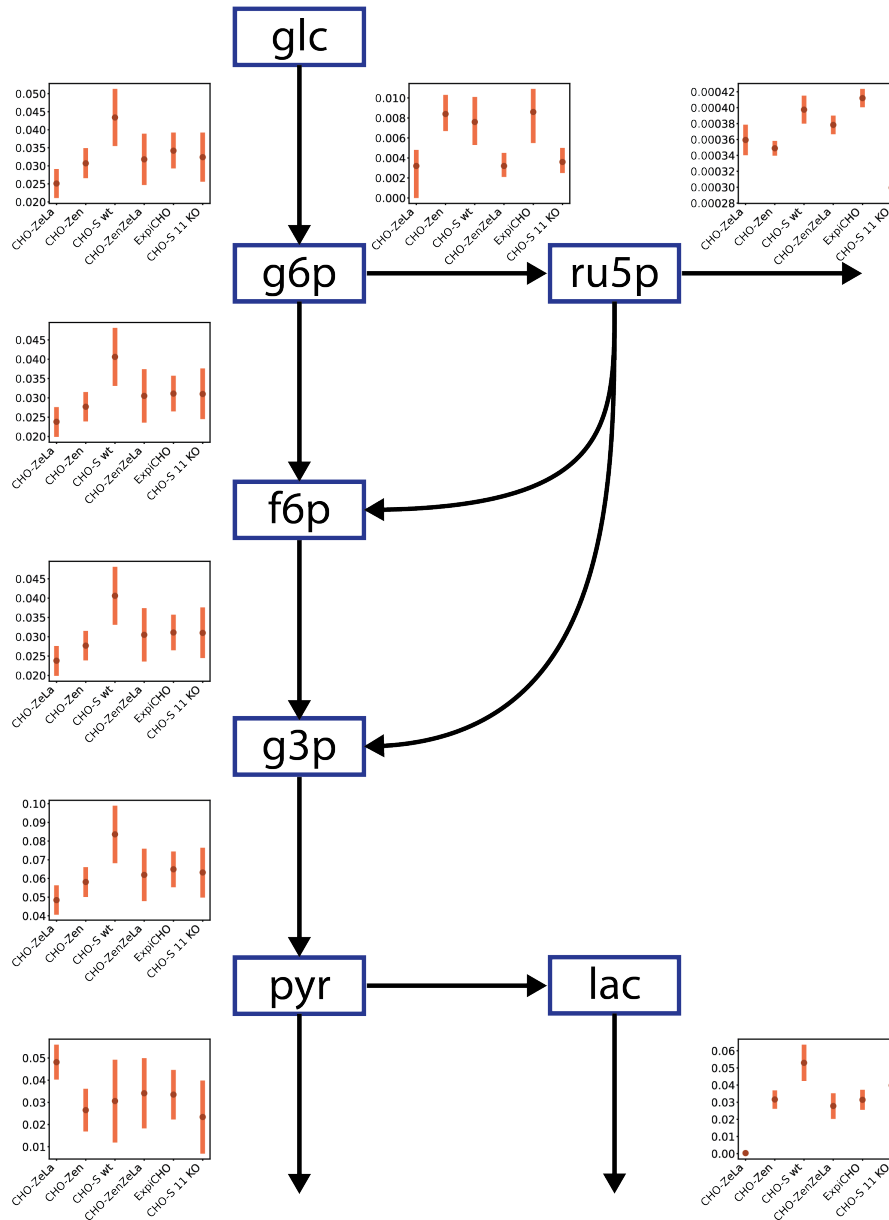
Figure 2.5: Comparison of estimated intracellular fluxes between cell lines using $^{13}C-MFA$. The means and 95% profile likelihoods for each cell line and each reaction are presented by point and bars respectively. The fluxes were estimated using in INCA (Young, 2014).

The oxPPP split ratio in CHO-S wt was found to be on average 18% in this study, which is significantly higher than previously measured for CHO cells. The difference from the study conducted by Ahn and Antoniewicz, 2013 could be explained by differences in cell lines, with Ahn and Antoniewicz, 2013 using CHO-K1 compared to CHO-S wt in this study. Additionally, we measured the isotope distribution on the PPP intermediates sedoheptulose-7-phosphate (s7p), 6-phospho-D-gluconate (6pgc) and ribulose-5-phosphate (ru5p) that may influence the split ratio in the fluxomics analysis.

2.3.2   Proteomics

The enzyme concentrations of the glycolytic intermediates were measured during the exponential phase

There is no correlation between protein expression and fluxes

Reactions closer to equilibrium are able to adapt to changes in flux by becoming more thermodynamically irreversible without changing protein concentration. Similarly, enzymes that are already far from equilibrium can modulate their flux by increasing substrate concentration, or through post-translational modifications. To investigate if enzyme expression is a mediator of flux in glycolysis we conducted a linear correlation for two groups, near equilibrium (reversible) and far from equilibrium (irreversible) reactions. After normalizing each protein expression and flux by their means and standard deviations we conducted a linear regression revealing that both groups displayed a slight negative but insignificant correlation as seen in figure 2.6.

This finding is consistent with previous studies which investigated overflow metabolism in yeast and concluded that glycolytic enzymes concentrations were not associated with an increase in fluxes (Xia et al., 2022). Another study was made in CHO cells where PDK was inhibited by dichloroacetate (DCA) decreasing flux through both the EMP and PPP pathways (Buchsteiner et al., 2018). Similar to the present study of CHO-ZeLa, no correlation in protein expression was observed. Thus our results support the conclusion of many previous studies which found that changes in glycolysis are not mediated by protein expression, but the post-translational modifications of its enzymes (Buchsteiner et al., 2018; Mulukutla et al., 2014).

2.3.3   Metabolomics

Compartmentalised metabolisms whereby reactions are localised to specific regions within a cell convolutes metabolomics where the measured whole cell concentration is not representative of the compartment specific concentration. A previous study by Lee et al., 2019 has shown that glycolytic metabolites are almost solely localized in the cytosol. This suggests that for the metabolites considered in glycolysis whole cell concentrations are consistent with the *in vivo* concentrations. To evaluate whether our method accurately captured the intracellular state, we used the energy charge ratio (ECR) as a metric as defined in equation 2.1. The ECR is the ratio of atp and adenosine diphosphate (adp) concentrations compared to the total adenine pool and decreases when the system is removed from its growth media (Talwalkar and Lester, 1973). The measured ecr varied between 0.79 and 0.87, which is in accordance with the ecr of 0.8 observed in directly extracted mammalian cells (Dietmair et al., 2010). Directly extracted cells stop metabolism faster than the quenching and then extraction process. This suggests that our method adequately quenched metabolism and that the measured metabolites should be representative of the intracellular state.

nadh concentration is significantly higher in CHO-ZeLa compared to CHO-S wt

Glycolysis is a major producer of nadh through GAPD generating 2 nadh for every glucose molecule. In an energy efficient system this generated nadh would be used to
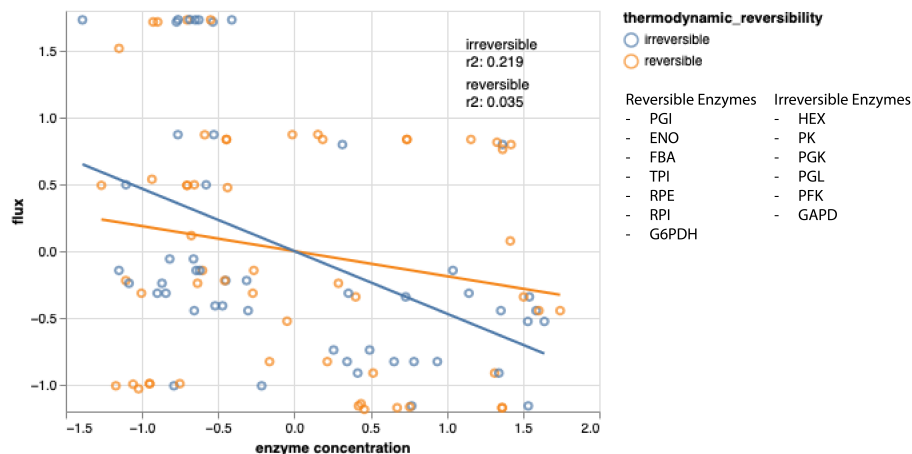
Figure 2.6: Normalised measured enzyme concentration compared to estimated fluxes. The fluxes and proteins were normalised using the following formula: $\hat{\mu}^{condition}_{normlised} = \frac{\hat{\mu}^{condition} - mean(\hat{\mu}^{condition})}{std(\hat{\mu}^{condition})}$. Where the $\hat{\mu}^{condition}$ is the mean of each condition, and the $mean(\hat{\mu}^{condition})$ is the mean of the means and the $std(\hat{\mu}^{condition})$ is the standard deviation of the means. Thermodynamic irreversibility was predetermined by the standard Gibbs free energy of reaction of $\Delta_r G = -10 \frac{kJ}{mol}$. The linear regression for each associated enzyme group as well as their r^2 values are shown. From the r^2 values we can see that there is a poor relation between a relative increase in enzyme concentration and an increase in flux of the corresponding reaction, irrespective of the reaction reversibility.

convert pyruvate to acetyl-coa through PDH to fuel the TCA cycle. However, cells that exhibit the Warburg effect instead perform overflow metabolism, consuming nadh by converting pyruvate to lactate. The inability to produce lactate in CHO-ZeLa removes a major sink for nadh and can be observed by the higher nadh concentrations that are only measurable in CHO-ZeLa (figure 2.7).

Substrates of irreversible reactions do not vary significantly

A comparison of the irreversible reaction substrates show that there is an insignificant change in metabolite concentrations relative to the change in flux (figure 2.7). These metabolties are fructose-6-phosphate (f6p) for the conversion of f6p into fructose-1,6-bisphosphate (fdp) catalysed by phosphofructokinase (PFK), and phosphoenolpyruvate (pep) for the conversion of pep into pyruvate (pyr) catalysed via pyruvate kinase (PK). Additionally, these metabolites are higher in concentration in the CHO-ZeLa phenotype than CHO-S wt. If the fluxes were mediated by enzyme saturation then we would expect an increase in concentrations to cause an increase in flux. This was not observed suggesting that these reactions are regulated by allostery.

2.3.4 Glycolysis must be an allosterically regulated process

The Warburg effect as observed when comparing CHO-S to CHO-ZeLa is a twofold increase in flux without any associated increase in protein expression and minimal changes in metabolite concentration. This comparison epitomises homeostasis in glycolysis, where flux is regulated via post-translational modifications. Previously, Mu-
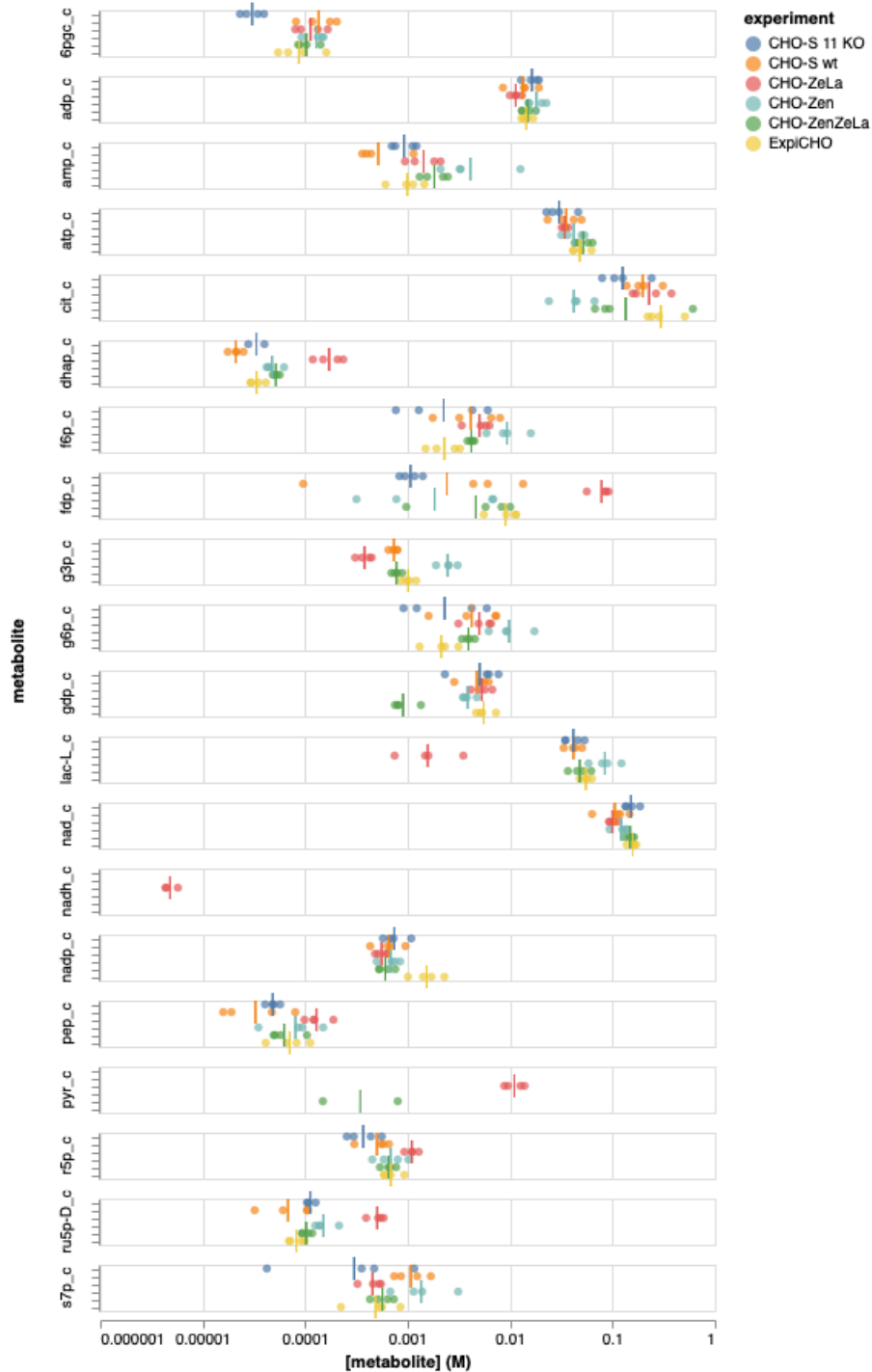
Figure 2.7: Absolute metabolomics of CHO cells during exponential phase measured at t=52.5 h. Each of the cell lines were operated in four separate bioreactors which are shown as seperate measurements (circles). Also shown is the mean estimate (bars) for each cell line and metabolite. Missing measurements such as the case with pyr and nadh suggest that the metabolite is likely below the lower limit of detection. the mean estimate presented is the logarithmic mean of the metabolite concentrations. the logarithmic mean is calculated by taking the mean estimate of the logarithmic transformed metabolite concentrations.

lukutla et al., 2014 have created a kinetic model of glycolysis in mammalian cells that captured high and low flux states without a change in protein expression. The authors incorporated allosteric regulation around two enzymes, PFK and PK, to generate bistability.

In this study we possess a contrasting CHO cell line that does not produce lactate, has similar enzyme and metabolite concentrations to CHO-S, and a significantly lower flux through glycolysis. Constructing a kinetic model is an obvious answer to the question of how does glycolysis regulate metabolism at every single enzyme with an seemingly minimal change in its proteome and fluxome. Kinetic models are mechanistic representations of metabolism which explicitly represent the thermodynamic, allosteric, saturation and expression contributions to flux (Liebermeister et al., 2010).

To leverage the multiomics dataset we generated for CHO cells and the prior knowledge of enzyme kinetics and regulation we developed a Bayesian modelling framework that infers kinetic parameters that are consistent with experimental measurements. We named this framework Maud, after Maud Menten, the co-contributer to the Michaelis-Menten equation (Michaelis and Menten, 1913). The development of Maud is described in Chapter 3.

2.4   Methods

The following section describes the methods used for the fermentation, sample analysis, and data processing. The following people were responsible for the sampling and analytical methods:

- **Fermentation and sampling**: David Catalán Tatjer and Kristian Lund Jensen

- **Proteomics sample preparation and analysis**: Tune Wulff, Shannara Taylor Parkins

- **Endometabolomics sample preparation and analysis**: Ana Catarina Sobral da Rocha

- **Exometabolomics sample preparation and analysis**: Karsten Steen Jensen

The remaining data analysis and integration were performed solely by Nicholas Luke Cowie.

Data availability The multiomics dataset can be found in the following repository: NicholasCowie/Thesis_Supplementaries

2.4.1   General sampling practices

The following procedures apply generally to all methods throughout the sampling process and will be described in detail here for clarity. This protocol required two people that could operate up to 24 reactors simultaneously.

Sampling

All sampling of bioreactors was performed using the built-in Ambr15 sampler (Sartorius). The samples are directly transferred to a cold plate that keeps samples at 4°C.

Centrifugation

To minimise changes in metabolism the centrifuge was pre-cooled for 20 minutes at 0.5°C. This reduces the amount of heat transferred to the samples during centrifugation.

2.4.2   Exometabolomics sampling

Data for the exometabolomics was sampled from the Ambr15 reactor at hours 0, 30.5, 46.5, 54.5, and 70.5. 0.5 mL was dispensed into a 1.5 mL Eppendorf tube where 50 μL was used for cell counts and characterisation. The remaining 450 μL was centrifuged at 1000 rcf for 1 minute with the cell pellet being stored for subsequent proteomics analysis, 300 μL was transferred into the Bioprofile Flex (Nova biomedical). To measure glucose, lactate, glutamine, glutamate, and ammonia, and 50 μL was transferred into a 500 μL Eppendorf tube For high-performance liquid chromatography (HPLC) analysis. The protein and HPLC samples were snap frozen by spraying ethanol over the samples and remained on dry ice until stored at -80°C awaiting analysis.

2.4.3   Endometabolomics sampling

When cells are removed from a bioreactor, their metabolism changes from the state present in the bioreactor; this shift can lead to biases in the quantitative metabolomics results. To minimise these biases we are required it is necessary to carry out quenching as fast as possible after sampling. Achieving this is difficult in the Ambr15 system as the liquid handler operates over a period of 1 minute allowing the cells to their metabolism on the designated cold plates that operate at 4°C. In order to minimise the impact of the liquid handler, 1 mL of cells at a cell density of approximately $3 \times 10^6 \frac{cells}{mL}$ was directly injected into a 5 mL Eppendorf tube containing 4 mL of 0.9% (w/v) NaCl as described in Dietmair et al., 2010. To ensure the cells remained cold the Eppendorf tubes were encased in an ice slurry of 15.6 (w/v) NaCl and placed in a freezer at -20°C an hour prior to sampling.

Immediately after the cells was sampled they were centrifuged at 1000 rcf for 1 minute at 0.5°C. The supernatant was discarded and 2.5 mL of 0.9% (w/v) 4°C NaCl was added to the Eppendorf tube and mixed via gentle inversion. The cells were then centrifuged at 1000 rcf for 1 minute at 0.5°C and the supernatant was discarded. The cell pellets were placed on dry ice and sprayed with ethanol to snap freeze before being stored at -80°C prior to extraction.

2.4.4   Metabolomics extraction

The cell pellets were extracted using the method described in Dietmair et al., 2010. 700 μL of an extraction buffer containing Acetonitrile:Methanol:Water in a ratio of 2:2:1 (v/v/v) as well as 0.1% Formic acid (v/v). Subsequently, 40 μL of internal standard (IS) (3:1 H2O:IS) was added to the frozen pellet and vortexed to mix. After resuspending

the thawed pellets, the samples were transferred to 2 mL Eppendorf tubes and kept on ice. Add 100 μL of extraction buffer to the Eppendorf tubes and vortexed to mix. The samples were then incubated at -80°C for 20 minutes. The centrifuge was pre-cooled for 20 minutes at -9°C before centrifuging the samples at 17000 rcf for 5 minutes. The supernatant was transferred to a new Eppendorf tube and stored at -80°C. Afterwards, the pellets were washed in 200 μL extraction buffer, vortexed and incubated at -80°C for 20 minutes before centrifuging at 17000 rcf for 5 more minutes with the supernatant being transferred to the same Eppendorf tube as before. The washing process was repeated one more time, resulting in a total sample volume of 1200 μL for each tube. The samples were dried at 30°C for approximately 8 hours before being stored at -80°C until analysis.

2.4.5 Absolute metabolomics quantification

In order to prepare them for analysis the samples were reconstituted in 120 μL of LC-MS H2O, vortexed and filtered through a 0.22 μm filter. The samples were then transferred to a 96-well Armadillo plate. 10, 100, and 1000 μL dilutions were prepared and analysed using the protocol detailed in McCloskey and Ubhi, 2014. The protocol utilises a reverse-phase ion-paring-liquid chromatography-tandem mass spectrometery (LC-MS/MS) method performed using a Sciex QTrap 5500 run in MRM mode as well as a Shimadzu LC-20/30 series LC system with a Waters XSelect HSS T3_2.5μm _2.1x150mm_XP column. Each injection was 10 μL and quantification was performed using SciexOS software with concentrations estimated based on peak area.

Calculation of energy charge ratio

The energy charge ratio was calculated using the following formula based on the mean concentrations of the measured concentrations of atp, adp and adenosine monophosphate (amp):

$$ecr = \frac{atp + 0.5adp}{atp + adp + amp} \tag{2.1}$$

2.4.6 Proteomics

Concurrently with the exometabolomics sampling defined in section 2.4.2 the cell pellets were stored at -80°C for proteomics analysis. After storage the cells were processed according to the protocol defined in Kozaeva et al., 2021. The cells were lysed in 6 M guanidinium•HCl, 5 mM tris(2-carboxethyl)phosphine, 10 mM chloroacetamide and 100 mM Tris•HCl and were agitated using a MM 400 Mixer Mill (Retsch) at 25 Hz for 5 min at room temperature and then heated in thermomixer at 95 $°C$ for another 10 min at 1800 rpm. The mixture was cetrifuged for 10 min at at 15000 rcf to produce a clarified supernatant. The total protein mass of the solution was quantified using the bicinchoninic acid (BCA) method and used to aliquot 20 μg of protein for tryptic digestion overnight. 1 μg of sample was injected into he Orbitrap Exploris 480 mass spectrometer (Thermo Scientific) operated in data dependent acquisition (DDA) mode.

The output files were analysed in Proteome Discoverer 2.4 in order to quantify the peptide counts. The false discovery rate was set at 0.1 %, the peptide length was set between 6 and 144, maximum missed cleavages sites was set to 2, the precursor mass tolerance was defined as 10ppm, and the fragment mass tolerance was set at 0.02 Da. The CHO reference proteome produce by Hilliard et al., 2020 and Rupp et al., 2018, was used for identification and annotation (RefSeq assembly accession: GCF_003668045.3).

Protein quantification was performed based on peptide measurements using the xTop algorithm (Mori et al., 2021) with data curation requiring that for any given peptide measurements are available from at least 3 samples. This generated a set of protein mass rations that were normalised using the total protein amount calculated from the BCA. The intra-cellular protein concentrations were calculated using equation 2.2 that converted the normalised per BCA mass to volume of cell.

$$[Protein_x] = \frac{A_x}{\sum_i MW_i \bullet A_i} \bullet TPA \bullet \frac{1}{V_c} \tag{2.2}$$

In equation 2.2 A_x is the abundance calculated from xTop, TPA is the total protein amount in g that was sampled from the bioreactor, MW_x is the molecular weight of x in $\frac{g}{mol}$, and V_c is the cellular volume in L. The molecular weights of proteins that contained unidentified proteins were imputed using the average molecular weight of an amino acid at 118.9 g/mol.

2.4.7 Labelled metabolomics

The quenching and extraction protocol for labelled metabolomics analysis was the same as described in Section 2.4.3. The samples were quenched and extracted in exactly the same way as for the absolute metabolomics samples with the exception that the samples were collected at a density of approximately $4-5 \times 10^6$ cells/mL with 3 mL sampled. Additionally, no internal standard was added during the extraction phase. Acquisition was performed using the protocol detailed in McCloskey et al., 2016 with a reduced isotopomer count, as the highest number of isotopomers measured was up to m+3 to increase the sensitivity of low abundance transitions by increasing the capture time.

2.4.8 Metabolic flux analysis

Metabolic flux analysis is the process of estimating the fluxes of a defined metabolic network. This analysis combines externally measured metabolites along with labelling experiments that provide information about the internal flux values. The kinetic model defined for this study directly describes only a fraction of the metabolism of a whole cell. Other parts of metabolism are represented in the model only by exogenous boundary conditions, including: biomass demand of ribose-5-phosphate (r5p) for nucleotide synthesis, serine demand for one-carbon metabolism, dihydroxyacetone phosphate

(dhap) flux for phospholipids, and pyruvate drains to a variety of cell processes. These metabolic pathways will be referred to as biomass drains and their behaviour has been characterised in genome scale metabolic models (GEMMs) (Széliová et al., 2020).

Combining GEMMs and exometabolomics we now have a representation of the external fluxes of the system, what remains is to estimate the internal flux split between the EMP and PP pathways. Inference of these fluxes is possible through the use of ^{13}C MFA, relying on the property that if labelling positions of a substrates carbons are known and the EMP and PP pathways produce different labelling configurations depending on their path, then the relative flux split can be estimated (Young, 2013; Antoniewicz, 2018).

For all rates, we will work in the units of $\frac{mol}{L_{cell} \bullet h}$, this ensures that the rates are comparable to the absolute metabolomics measurements, which provide thermodynamic information. A common standard among fluxomics is $\frac{mol}{g_{DCW} \bullet h}$, where g_{DCW} is the grams of dry cell weight, however, the standard for activity of thermodynamic measurements is volumetric rather than mass based, and would be unsuitable for a kinetic model.

To summarise, we use exometabolomics to determine uptake and production rates, a constraint based model to determine biomass drains, and ^{13}C MFA to determine internal fluxes.

2.4.9 Cell counting

Cell viability, number, and size were measured using the NucleoCounter NC-250 (ChemoMetec) using two fluorescent dyes, Acridine Orange and DAPI for total and dead cell populations.

2.4.10 Constructing strain specific uptake rates

Our objective was to compare different strains, and therefore, consolidate the replicate flux measurements into a defined distribution for each strain. We did this by constructing a hierarchical model where the observed rate μ of the true replicate rate $\hat{\mu}$ is a draw from the strain specific distribution μ_0 of rates which were observed with a known measurement error σ. The replicate specific uptake rates are assumed to vary from the per-strain average according to a normal distribution with standard deviation η_0, where η_0 is an unknown positive number that is assumed to be the same for a defined strain. This model specification is shown mathematically in equation 2.3 with the accompanying Stan code in A. The regression model was fit for each replicate using the scipy library and the hierarchical flux model was fit using cmdstan (version 1.0.8).

$$\hat{\mu} \sim Normal(\mu_0, \eta_0) \tag{2.3}$$

$$\mu \sim Normal(\hat{\mu}, \sigma) \tag{2.4}$$

2.4.11   Determining growth rate

The growth rate μ was estimated using measurements of volumetric growth rather than cell number. CHO cells are known to change cell size throughout a cultivation, so using cell number measurements would unnecessarily introduce a bias. Cellular diameter was measured using the NucleoCounter NC-250, and assuming a spherical cell shape, the cell volume was calculated using equation 2.5.

$$V_c = \frac{4}{3}\pi \left(\frac{d}{2}\right)^3 \tag{2.5}$$

Cellular growth was assumed to be exponential that follows equation 2.6:

$$n_t V_t = n_0 V_0 \bullet e^{\mu t} \tag{2.6}$$

Where t is the time, V_c is the cell volume, V_0 is the initial cell volume, n_0 is the initial cell count, and n_t is the cell count at time t.

$$log(n_t V_t) = log(n_0 V_0) + \mu t \tag{2.7}$$

by log transforming equation 2.6 the growth rate can be estimated using the gradient of the linear regression (equation 2.7). We performed the linear regression using scipy.stats.linregress (Virtanen et al., 2020) for each bioreactor. The resulting linear regression estimated the mean growth rate and standard deviation for each bioreactor. To estimate the strain specific growth rate a hierarchical model was used, see Equation 2.3. The hierarchical model was used to estimate the 95% interval determined by the 2.5 % and 97.5 % quantiles of the posterior distribution of the mean growth rate (μ_0 in the model).

2.4.12   Estimating extracellular fluxes

It is possible to measure glucose uptake rate and lactate production directly. Both of these quantities were estimated using the methods defined in Long and Antoniewicz, 2019 for a batch culture using the following formula:

$$\text{specific uptake/production rate} = \frac{\Delta[Metabolite]}{\Delta V_{cell}} \times \mu \tag{2.8}$$

2.4.13   Defining biomass constraints

Due to experimental constraints, nucleotide synthesis for biomass demand was not measured. Instead, the biomass demand was estimated based on previously measured DNA and RNA synthesis requirements by using the parsimonious flux balance analysis (pFBA) function in COBRApy and a CHO-S genome scale metabolic model (Széliová et al., 2020; Ebrahim et al., 2013; Lewis et al., 2010). To convert the biomass

Table 2.2: Cell density measurements from Széliová et al., 2020.

Volume	Mass
2168.4	286.7
2169.1	305.6
2169.0	287.1

specific nucleotide demand into $\frac{mol}{V_{cell} \bullet h}$ the conversion factor was also taken from Széliová et al., 2020 as shown in table 2.2.

2.4.14   13C Metabolic flux analysis

13C metabolic flux analysis (13C MFA) was used to estimate the intracellular fluxes of reactions in our model given boundary fluxes, endometabolomics, isotope abundances, and a defined glucose tracer. The assumptions used in constructing the 13C model are shown in table 2.3.

^{13}C-MFA model

The system was modelled using INCA version 2.0 (Young, 2014). The following setting were used for the entirely of the analysis: *relative convergence tolerance* = 1e-2, *Levenberg-Marquardt tau parameter* = 1e-6, *number of restarts* = 500. The model stoichiometry and atom map was defined as shown in table 2.4 and was based on the model in Ahn and Antoniewicz, 2013, with the following exceptions:

1. no TKTL1 reaction

2. biomass drain for r5p

3. lactate dehydrogenase (LDH) and pyr-drain split

Flux estimates were achieved by minimising the sum of squared residuals (SSR) between the measured and predicted isotope fractions and fluxes for n=4 isotope labelling experiments unless otherwise specified. The boundary fluxes — hexokinase (HEX), LDH, and r5p-drain — were estimated using external flux measurements and growth rates as explained in sections 2.4.12 and 2.4.13.

The measurements of mass isotopomer fractions are compositional; that is, each set of measurements $x \in \mathbb{R}^k$ lies on the k-simplex, where k is the number of carbon atoms in the metabolite were considered. A linear model assumes that all measurements are unconstrained, leading to biases when applied to compositional data. Ideally, a statistical method that accounts for compositional data would be used, as described in Aitchison, 1982. Unfortunately, compositional data analysis for 13C-MFA has not yet been implemented to the author's knowledge and developing such a method is beyond the scope of this project. However, it is possible to mitigate the effects of compositional data, which we did using the following approach. Isotopomer fractions were determined using equation 2.9. Only the isotopomers up to $m + (k - 1)$ were used as measurements. By fitting the linear model this reduced dataset we make it possible for the sum of the measurements not to equal 1, which brings the data closer to the

Table 2.3: Assumptions used when creating the ^{13}C-MFA model.

Assumption	Reasoning
Isotopic steady state achieved after 3 hours	Confirmed for upper glycolysis in Ahn and Antoniewicz, 2013
Serine enrichment into glycolysis is minimal	Non-essential amino acids may be consumed and enriched into glycolysis. However, of the papers that have considered additional amino acid streams none have shown significant in- or out-flows through glycolysis (Ahn and Antoniewicz, 2013; Quek et al., 2010)
Gluconeogenesis is supressed during the exponential growth phase	No labelling from $[U-^{13}C]$-glutamine was observed during the exponential phase in upper glycolytic intermediates in a study conducted by Ahn and Antoniewicz 2013, this suggests gluconeogenesis does not occur (Ahn and Antoniewicz, 2013).
Transketolase like-1 (TKTL1) activity is suppressed during exponential phase	TKTL1 that was identified in the stationary growth phase by Ahn et al., 2016 was not highly expressed during exponential growth across all cell lines.
Phosphofructokinase-2 (PFK2) carries a minimal flux during exponential phase	PFK2 produces f26bp, a sensory metabolite that is a regulator of glycolysis, flux diverted through PFK2 is not used for glycolysis and futile cycles to f6p expending atp. This reaction is unresolvable at steady state and would be extremely inefficient.

Table 2.4: ^{13}C-MFA atom transition map.

Reaction	Substrate	Directionality	Product
HEX	glc.e (abcdef)	→	g6p (abcdef)
GPI	g6p (abcdef)	↔	f6p (abcdef)
PFK	f6p (abcdef)	↔	fdp (abcdef)
FBA	fdp (abcdef)	↔	dhap (cba) + g3p (def)
TPI	dhap (abc)	↔	g3p (abc)
GAPD	GAP (abc)	↔	3PG (abc)
ENO	3PG (abc)	↔	PEP (abc)
PYK	PEP (abc)	→	Pyr (abc)
G6PDH_PGL	G6P (abcdef)	→	6PGC (abcdef)
GND	6PGC	→	Ru5P (bcdef)+CO2 (a)
RPE	Ru5P (abcde)	↔	X5P (abcde)
RPI	Ru5P (abcde)	↔	R5P (abcde)
X5P_EC2	X5P (abcde)	↔	EC2 (ab)+GAP (cde)
F6P_EC2	F6P (abcdef)	↔	EC2 (ab)+E4P (cdef)
S7P_EC2	S7P (abcdefg)	↔	EC2 (ab)+R5P (cdefg)
F6P_EC3	F6P (abcdef)	↔	EC3 (abc)+GAP (def)
S7P_EC3	S7P (abcdefg)	↔	EC3 (abc)+E4P (defg)
LDH	Pyr (abc)	→	Lac (abc)
pyr-drain	Pyr (abc)	→	Pyr.e (abc)
r5p-drain	R5P (abc)	→	R5P.e (abc)

assumptions of a linear model.

$$x_{i,j} = \frac{x_{i,j}}{\sum_i x_{i,j}} \qquad \text{where,} i = \text{isotopomer fraction} j \qquad = \text{metabolite} \qquad (2.9)$$

Isotope abundance error was calculated using the coefficient of variation (CV) of the measured isotope fraction in a quality control sample that was constructed by mixing all measurements into a single sample as shown in equation 2.10. The CV was measured using n=2 and the measurement error was calculated by multiplying the CV by the measured isotope fraction (equation 2.10). If the value was below 2 percentile points (0.02), a value of 0.02 was imputed. This is above the value suggested by Antoniewicz, 2018, however, for small isotope fractions that are common for the m+1 variants in this labelling experiment the suggested value of 0.01 resulted in extremely poor fits, as evaluated using the SSR (Antoniewicz, 2018).

$$CV_{i,j} = \frac{\sqrt{\bar{x}_{i,j} - x_{i,j}}}{\bar{x}_{i,j}} \qquad (2.10)$$

$$\sigma_{i,j,l} = CV_{i,j} \times x_{i,j,l} \qquad (2.11)$$

Table 2.5: Natural isotope abundance of carbon.

Isotope	Abundance relative to m+0
m+0	1
m+1	0.069
m+2	0.014
m+3	0.001

2.4.15   Determining tracer isotope ratio

The isotope label ratio of the tracer is a mixture of $[1, 2 -^{13} C]$ glucose and naturally abundant glucose. The ratio of labelled to unlabelled glucose was estimated by assuming the unlabelled (m+0) glucose to come entirely from the naturally labelled glucose fraction, which has an isotope distribution that is defined in Kim *et al.*, 2016, and is shown in table 2.5. The labelled glucose fraction and purity estimated was calculated according to equation 2.12. Where In which, x_i is the measured fraction of each isotope of glucose, and z_i is the measured abundance relative to m+0 of naturally labelled glucose.

$$frac_{[1,2-^{13}C]glc} = \sum_i x_i - x_0 \cdot z_i \tag{2.12}$$

$$purity = \frac{x_2}{frac_{[1,2-^{13}C]glc}} \tag{2.13}$$

References

Monge, I. M., Hefzi, H., Sanchez, P. C., Marín de Mas, I. B., Decker, M., Lecina, M., Cairó, J. J., Lewis, N. E., & Nielsen, L. K. (2020). *Lactate reduction in CHO cell cultures through metabolic analysis* (Poster). https://backend.orbit.dtu.dk/ws/portalfiles/portal/216677825/IMM_ESACT_2019_IG_LKN.pdf

Luengo, A., Li, Z., Gui, D. Y., Sullivan, L. B., Zagorulya, M., Do, B. T., Ferreira, R., Naamati, A., Ali, A., Lewis, C. A., Thomas, C. J., Spranger, S., Matheson, N. J., & Vander Heiden, M. G. (2021). Increased demand for NAD+ relative to ATP drives aerobic glycolysis. *Molecular Cell*, *81*(4), 691–707.e6. https://doi.org/10.1016/j.molcel.2020.12.012

Fernandez-de-Cossio-Diaz, J., & Vazquez, A. (2017). Limits of aerobic metabolism in cancer cells. *Scientific Reports*, *7*(1), 13488. https://doi.org/10.1038/s41598-017-14071-y

Puck, T. T., Cieciura, S. J., & Robinson, A. (1958). GENETICS OF SOMATIC MAMMALIAN CELLS. *Journal of Experimental Medicine*, *108*(6), 945–956. https://doi.org/10.1084/jem.108.6.945

Reinhart, D., Damjanovic, L., Kaisermayer, C., Sommeregger, W., Gili, A., Gasselhuber, B., Castan, A., Mayrhofer, P., Grünwald-Gruber, C., & Kunert, R. (2019). Bioprocessing of Recombinant CHO-K1, CHO-DG44, and CHO-S: CHO Expression Hosts Favor Either mAb Production or Biomass Synthesis. *Biotechnology Journal*, *14*(3), 1700686. https://doi.org/10.1002/biot.201700686

Ley, D., Pereira, S., Pedersen, L. E., Arnsdorf, J., Hefzi, H., Davy, A. M., Ha, T. K., Wulff, T., Kildegaard, H. F., & Andersen, M. R. (2019). Reprogramming AA catabolism in CHO cells with CRISPR/Cas9 genome editing improves cell growth and reduces byproduct secretion. *Metabolic Engineering*, *56*, 120–129. https://doi.org/10.1016/j.ymben.2019.09.005

Jain, N. K., Barkowski-Clark, S., Altman, R., Johnson, K., Sun, F., Zmuda, J., Liu, C. Y., Kita, A., Schulz, R., Neill, A., Ballinger, R., Patel, R., Liu, J., Mpanda, A., Huta, B., Chiou, H., Voegtli, W., & Panavas, T. (2017). A high density CHO-S transient transfection system: Comparison of ExpiCHO and Expi293. *Protein Expression and Purification*, *134*, 38–46. https://doi.org/10.1016/j.pep.2017.03.018

Kol, S., Ley, D., Wulff, T., Decker, M., Arnsdorf, J., Schoffelen, S., Hansen, A. H., Jensen, T. L., Gutierrez, J. M., Chiang, A. W. T., Masson, H. O., Palsson, B. O., Voldborg, B. G., Pedersen, L. E., Kildegaard, H. F., Lee, G. M., & Lewis, N. E. (2020). Multiplex secretome engineering enhances recombinant protein production and purity. *Nature Communications*, *11*(1), 1908. https://doi.org/10.1038/s41467-020-15866-w

Ahn, W. S., & Antoniewicz, M. R. (2013). Parallel labeling experiments with [1,2-(13)C]glucose and [U-(13)C]glutamine provide new insights into CHO cell metabolism. *Metabolic Engineering*, *15*, 34–47. https://doi.org/10.1016/j.ymben.2012.10.001

Young, J. D. (2014). INCA: A computational platform for isotopically non-stationary metabolic flux analysis. *Bioinformatics (Oxford, England)*, *30*(9), 1333–1335. https://doi.org/10.1093/bioinformatics/btu015

Antoniewicz, M. R., Kelleher, J. K., & Stephanopoulos, G. (2006). Determination of confidence intervals of metabolic fluxes estimated from stable isotope measurements. *Metabolic Engineering*, *8*(4), 324–337. https://doi.org/10.1016/j.ymben.2006.01.004

Pan, X., Dalm, C., Wijffels, R. H., & Martens, D. E. (2017). Metabolic characterization of a CHO cell size increase phase in fed-batch cultures. *Applied Microbiology and Biotechnology*, *101*(22), 8101–8113. https://doi.org/10.1007/s00253-017-8531-y

Széliová, D., Ruckerbauer, D. E., Galleguillos, S. N., Petersen, L. B., Natter, K., Hanscho, M., Troyer, C., Causon, T., Schoeny, H., Christensen, H. B., Lee, D.-Y., Lewis, N. E., Koellensperger, G., Hann, S., Nielsen, L. K., Borth, N., & Zanghellini, J. (2020). What CHO is made of: Variations in the biomass composition of Chinese hamster ovary cell lines. *Metabolic Engineering*, *61*, 288–300. https://doi.org/10.1016/j.ymben.2020.06.002

Quek, L.-E., Dietmair, S., Krömer, J. O., & Nielsen, L. K. (2010). Metabolic flux analysis in mammalian cell culture. *Metabolic Engineering*, *12*(2), 161–171. https://doi.org/10.1016/j.ymben.2009.09.002

Lunt, S. Y., & Vander Heiden, M. G. (2011). Aerobic Glycolysis: Meeting the Metabolic Requirements of Cell Proliferation. *Annual Review of Cell and Developmental Biology*, *27*(1), 441–464. https://doi.org/10.1146/annurev-cellbio-092910-154237

Xia, J., Sánchez, B. J., Chen, Y., Campbell, K., Kasvandik, S., & Nielsen, J. (2022). Proteome allocations change linearly with the specific growth rate of Saccharomyces cerevisiae under glucose limitation. *Nature Communications*, *13*(1), 2819. https://doi.org/10.1038/s41467-022-30513-2

Buchsteiner, M., Quek, L.-E., Gray, P., & Nielsen, L. K. (2018). Improving culture performance and antibody production in CHO cell culture processes by reducing the Warburg effect. *Biotechnology and Bioengineering*, *115*(9), 2315–2327. https://doi.org/10.1002/bit.26724

Mulukutla, B. C., Yongky, A., Daoutidis, P., & Hu, W.-S. (2014). Bistability in Glycolysis Pathway as a Physiological Switch in Energy Metabolism (P. Dzeja, Ed.). *PLoS ONE*, *9*(6), e98756. https://doi.org/10.1371/journal.pone.0098756

Lee, W. D., Mukha, D., Aizenshtein, E., & Shlomi, T. (2019). Spatial-fluxomics provides a subcellular-compartmentalized view of reductive glutamine metabolism in cancer cells. *Nature Communications*, *10*(1), 1351. https://doi.org/10.1038/s41467-019-09352-1

Talwalkar, R., & Lester, R. L. (1973). The response of diphosphoinositide and triphosphoinositide to perturbations of the adenylate energy charge in cells of Saccha-

romyces cerevisiae. *Biochimica et Biophysica Acta (BBA) - Lipids and Lipid Metabolism*, *306*(3), 412–421. https://doi.org/10.1016/0005-2760(73)90180-X

Dietmair, S., Timmins, N. E., Gray, P. P., Nielsen, L. K., & Krömer, J. O. (2010). Towards quantitative metabolomics of mammalian cells: Development of a metabolite extraction protocol. *Analytical Biochemistry*, *404*(2), 155–164. https://doi.org/10.1016/j.ab.2010.04.031

Liebermeister, W., Uhlendorf, J., & Klipp, E. (2010). Modular rate laws for enzymatic reactions: Thermodynamics, elasticities and implementation. *Bioinformatics*, *26*(12), 1528–1534. https://doi.org/10.1093/bioinformatics/btq141

Michaelis, L., & Menten, M. (1913). Die kinetik der invertinwirkung. *Biochem. z*, *49*, 333–369.

McCloskey, D., & Ubhi, B. K. (2014). Quantitative and qualitative metabolomics for the investigation of intracellular metabolism. *SCIEX Tech Note*, 1–11.

Kozaeva, E., Volkova, S., Matos, M. R., Mezzina, M. P., Wulff, T., Volke, D. C., Nielsen, L. K., & Nikel, P. I. (2021). Model-guided dynamic control of essential metabolic nodes boosts acetyl-coenzyme A–dependent bioproduction in rewired Pseudomonas putida. *Metabolic Engineering*, *67*, 373–386. https://doi.org/10.1016/j.ymben.2021.07.014

Hilliard, W., MacDonald, M. L., & Lee, K. H. (2020). Chromosome-scale scaffolds for the Chinese hamster reference genome assembly to facilitate the study of the CHO epigenome. *Biotechnology and Bioengineering*, *117*(8), 2331–2339. https://doi.org/10.1002/bit.27432

Rupp, O., MacDonald, M. L., Li, S., Dhiman, H., Polson, S., Griep, S., Heffner, K., Hernandez, I., Brinkrolf, K., Jadhav, V., Samoudi, M., Hao, H., Kingham, B., Goesmann, A., Betenbaugh, M. J., Lewis, N. E., Borth, N., & Lee, K. H. (2018). A reference genome of the Chinese hamster based on a hybrid assembly strategy. *Biotechnology and Bioengineering*, *115*(8), 2087–2100. https://doi.org/10.1002/bit.26722

Mori, M., Zhang, Z., Banaei-Esfahani, A., Lalanne, J.-B., Okano, H., Collins, B. C., Schmidt, A., Schubert, O. T., Lee, D.-S., Li, G.-W., Aebersold, R., Hwa, T., & Ludwig, C. (2021). From coarse to fine: The absolute *Escherichia coli* proteome under diverse growth conditions. *Molecular Systems Biology*, *17*(5). https://doi.org/10.15252/msb.20209536

McCloskey, D., Young, J. D., Xu, S., Palsson, B. O., & Feist, A. M. (2016). MID Max: LC–MS/MS Method for Measuring the Precursor and Product Mass Isotopomer Distributions of Metabolic Intermediates and Cofactors for Metabolic Flux Analysis Applications. *Analytical Chemistry*, *88*(2), 1362–1370. https://doi.org/10.1021/acs.analchem.5b03887

Young, J. D. (2013). Metabolic flux rewiring in mammalian cell cultures. *Current Opinion in Biotechnology*, *24*(6), 1108–1115. https://doi.org/10.1016/j.copbio.2013.04.016

Antoniewicz, M. R. (2018). A guide to 13C metabolic flux analysis for the cancer biologist. *Experimental & Molecular Medicine*, *50*(4), 19. https://doi.org/10.1038/s12276-018-0060-y

Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., Van Der Walt, S. J., Brett, M., Wilson, J., Millman, K. J., Mayorov, N., Nelson, A. R. J., Jones, E., Kern, R., Larson, E., … Vázquez-Baeza, Y. (2020). SciPy 1.0: Fundamental algorithms for scientific computing in Python. *Nature Methods*, *17*(3), 261–272. https://doi.org/10.1038/s41592-019-0686-2

Long, C. P., & Antoniewicz, M. R. (2019). High-resolution 13C metabolic flux analysis. *Nature Protocols*, *14*(10), 2856–2877. https://doi.org/10.1038/s41596-019-0204-0

Ebrahim, A., Lerman, J. A., Palsson, B. O., & Hyduke, D. R. (2013). COBRApy: COnstraints-Based Reconstruction and Analysis for Python. *BMC Systems Biology*, *7*(1), 74. https://doi.org/10.1186/1752-0509-7-74

Lewis, N. E., Hixson, K. K., Conrad, T. M., Lerman, J. A., Charusanti, P., Polpitiya, A. D., Adkins, J. N., Schramm, G., Purvine, S. O., Lopez-Ferrer, D., Weitz, K. K., Eils, R., König, R., Smith, R. D., & Palsson, B. Ø. (2010). Omic data from evolved E. coli are consistent with computed optimal growth from genome-scale models. *Molecular Systems Biology*, *6*, 390. https://doi.org/10.1038/msb.2010.47

Ahn, W. S., Crown, S. B., & Antoniewicz, M. R. (2016). Evidence for transketolase-like TKTL1 flux in CHO cells based on parallel labeling experiments and (13)C-metabolic flux analysis. *Metabolic Engineering*, *37*, 72–78. https://doi.org/10.1016/j.ymben.2016.05.005

Aitchison, J. (1982). *The Statistical Analysis of Compositional Data*. Wiley Online Library.

3 Maud

Abstract

This paper presents Maud, a command-line application that implements Bayesian statistical inference for kinetic models of biochemical metabolic reaction networks. Maud takes into account quantitative information from omics experiments and background knowledge, as well as structural information about kinetic mechanisms, regulatory interactions and enzyme knockouts. Below, we review the existing options in this area, explain how Maud improves on the state of the art, describe the intended modelling workflow and illustrate its use with an example application.

3.1 Introduction

A kinetic model of cellular metabolism aims to express what is known about a cellular process in the form of an *in silico* representation of the underlying network of chemical reactions. The resulting models have many important applications, including as improving production rates, elucidating regulation (Christodoulou et al., 2018) and identifying drug targets (DeBerardinis and Chandel, 2016; Liberti et al., 2017). However, kinetic models are only useful to the extent that they faithfully and completely represent the available information, including not only the qualitative structure and physical laws governing the network but also noisy experimental data and uncertain non-experimental background knowledge.

It is now recognised that Bayesian statistical inference makes it possible in principle to combine the structural information implicit in kinetic models of cell metabolism with background knowledge about latent metabolic parameters and information from multiomics measurements (P. A. Saa and Nielsen, 2016; Gopalakrishnan et al., 2020). However, kinetic models pose serious computational challenges for Bayesian inference (Gutenkunst et al., 2007; Raue et al., 2013), with the result that Bayesian kinetic models are not widely used in metabolic engineering.

Our software application Maud addresses this challenge by implementing Bayesian inference for a wide range of accurate kinetic models using a flexible and intuitive user interface, reliable and scalable computation and non-proprietary code. Maud is the first framework that verifiably performs full Bayesian inference for realistic kinetic models, taking into account quantitative multiomics data and producing results for non-trivial networks and datasets within a practically relevant time frame.

Maud uses realistic rate laws derived from Liebermeister et al., 2010 with augmentations based on Popova and Sel'kov, 1975 and our own research, guarantees thermodynamic feasibility and realistically represents information from multiomics measurements, kinetic parameter databases, summaries of thermodynamic information

like eQuilibator (Beber et al., 2022) and other sources. Maud provides an intuitive input format and simple command line interface, thereby avoiding the need for users to edit source code. Maud is cross-platform, depends only on freely available software, is available online under an open source licence and is actively used and maintained. Maud achieves excellent performance by using posterior sampling with adaptive Hamiltonian Monte Carlo (Hoffman and Gelman, 2011) as provided by Stan (Carpenter et al., 2017), incorporating fast stiff ordinary differential equation solving provided by the library CVODES (Serban and Hindmarsh, 2005).

Section 3.1 reviews kinetic modelling theory and assumptions, the state of the art, and the pressing needs that Maud addresses. Section 3.2 explains in detail Maud's intended workflow, its kinetic and statistical modelling assumptions and its software implementation. Section 3.3 applies Maud to a realistic artificial dataset, demonstrating that Maud can recover known parameters and produce actionable results in a plausible simulated research context. Additionally, we compare results obtained using Maud with an alternative method, highlighting the need for full Bayesian inference. Section 3.4 discusses how Maud can be used for strain engineering and therapeutic applications, how Maud improves on the state of the art, limitations and the prospects for future improvement.

3.1.1 Background

Cellular metabolism is the driving force that provides living organisms with energy and building blocks for growth and production. The mediators of metabolic networks are enzymes, which catalyse metabolic reactions in highly regulated networks.

Kinetic models

We previously have covered the fundamentals of kinetic modelling in Section 1.3 of this thesis. Here we will review the terminology and include additional information relevant to the Maud workflow.

Kinetic models describe cellular metabolism by specifying explicit functions that map enzyme and metabolite concentrations to fluxes through enzyme-catalysed reactions. The general form of this kind of model is shown below in equation 3.1.

$$v = f(C, E, K, \Delta_f G). \tag{3.1}$$

In equation 3.1 the term C represents a set of metabolite concentrations, E a set of enzyme concentrations, K a set of kinetic parameters and $\Delta_f G$ a set of metabolite formation energies.

A kinetic model is typically depicted as a matrix S representing the stoichiometric coefficient of each metabolite in each reaction, and the corresponding fluxes of each reaction v. A common assumption is that metabolism is at a steady-state as shown in equation 3.1.1 where the cellular concentrations and fluxes remain constant over time.

$$S \cdot v = \frac{dC}{dt}$$
$$= 0$$

Given a set of values for the K, $\Delta_f G$ and sufficient boundary conditions E and $C_{unbalanced}$, a set of steady state metabolite concentrations and fluxes can be found by solving equation 3.1.1 for C.

Kinetic models that are accurate in the sense that their flux functions correctly describe the underlying biochemical process make it possible to predict how cells will behave in conditions that have not been experimentally tested *in vivo*. This is useful because such *in silico* predictions can be generated much faster and at greater scale than *in vivo* experiments, and can produce insight into mechanisms that are often difficult to resolve experimentally.

Kinetic model development relies on access to diverse and accurate datasets that expose the model to regions exploring the non-linear nature of metabolism only reconciled by the true data generating process. The recent development of biofoundries promises access to high volumes of multiomics data such that it is realistic to construct accurate kinetic models (Casini et al., 2018). Many modelling frameworks are designed to automate the construction of accurate kinetic models and link them with multiomics data; these are discussed in Section 3.1.2.

3.1.2   The state of the art

Kinetic modelling is an inherently complex problem. Analytical methods for anything but the smallest of models are infeasible, the systems are notoriously non-linear as demonstrated by homeostasis, and measurements are plagued by noise or are often unavailable. This section reviews some modelling paradigms that have been developed in recent years in order to address the need to automate the construction of accurate omics-based kinetic models.

We consider the three following modelling paradigms with respect to kinetic models: prior only/ensemble models, Bayesian models without Markov Chain Monte Carlo (MCMC) sampling and Bayesian kinetic models using MCMC.

3.1.3   Prior Models

A broad category of prior only models exist, with no specific framework required. A typical example is a network constructed using kinetic parameters collected from *in vitro* analysis (Christodoulou et al., 2018). Whereby kinetic parameters are determined through *in vitro* assays or previous fitting experiments and sampled independently. These models are some of the most computationally efficient to generate as no parameter updating is required. Furthermore, they are the least constrained models and offer a scope of the potential behaviours encompassed by the prior information. Un-

fortunately, the computational efficiency of these models is at the cost of minimal statistical inference as parameters are not updated using the likelihood $p(y \mid \theta)$.

Prior models provide an insight into the potential behaviours of the system and the quality of the prior models predictions is as the name suggests, entirely dependent on the prior information of the parameter space.

Ensemble models

Ensemble models (Gopalakrishnan et al., 2020; Chakrabarti et al., 2013; P. Saa and Nielsen, 2015; Tran et al., 2008) assume a known steady state flux vector and given a defined kinetic mechanism are able to sample parameter configurations that exactly yield this flux. The resulting parameter sets then serve as the basis for further analysis.

Ensemble modelling is computationally efficient as it is possible to implement without the use of a numerical differential equation solver and can be useful for exploring the implications of prior models. Ensemble modelling has been used to restrict the prior space for further sampling methods (P. A. Saa and Nielsen, 2016) and machine learning approaches (Choudhury et al., 2022).

However, ensemble modelling does not address the need to quantify the information contained in multiple omics datasets. First, the true steady state flux is never known exactly, but rather is typically inferred from noisy measurements. Ignoring the uncertainty inherent in this measurement process can lead to inappropriate over-fitting. Second, ensemble modelling does not attempt to capture the information contained in metabolite and enzyme concentration measurements through an explicit measurement model. Since such measurements are highly informative, resorting to ensemble modelling entails a loss of predictive power.

Likelihood-free models

Likelihood-free kinetic modelling approaches aim to improve on ensemble modelling by incorporating information from measurements, but without representing this information explicitly in the form of a likelihood function $p(y \mid \theta)$ as would be required for full Bayesian inference (P. A. Saa and Nielsen, 2016; Gopalakrishnan et al., 2020; Hackett et al., 2016). This does not mean that the likelihood is not calculated as there is an implicit function in the form of rejection criteria or root mean square error. However, the lack of an explicit likelihood function assumes a measurement model behaviour that is not necessarily correct at the expense of convenience.

In addition to this general problem, some likelihood-free kinetic modelling approaches do not even implicitly specify a joint distribution over all possible data/parameter combinations. Ensembles of kinetic models generated using such approaches are difficult to interpret; in particular they are not germane to standard analyses of draws from the posterior distribution of a Bayesian statistical model.

Conversely, likelihood-free approaches that do implicitly specify joint data/parameter distributions are essentially a form of Bayesian kinetic model without MCMC, a mod-

elling approach that we discuss below in section 3.1.3.

Bayesian kinetic models without MCMC
As explained in P. A. Saa and Nielsen, 2016, the natural way to quantify the information contained in omics measurements is through Bayesian inference, i.e. by specifying a likelihood function $p(y \mid \theta)$ and posing questions about what is known given a dataset y in terms of the posterior distribution $p(\theta \mid y)$.

Kinetic models pose a particular computational challenge for Bayesian inference because of the need to solve the steady state problem in order to connect parameters with measurable quantities. For realistic kinetic models solving the steady state problem requires finding the roots of a large, complicated and non-linear equation system. Steady state finding approaches were made orders of magnitudes faster for ensemble modelling approaches through the work defined in Gopalakrishnan et al., 2020, however, this approach is only valid for elementary mass action kinetics. In addition, modern MCMC techniques such as Hamiltonian Monte Carlo (Neal, 2011) require the gradients of the steady state solution with respect to θ. See Timonen et al., 2022 and Betancourt, 2017 for discussion of this issue.

As a result of this computational difficulty, there have been efforts to implement Bayesian inference for kinetic models without the use of MCMC. Examples of alternative inference methods include variational inference as in St. John et al., 2019, rejection sampling and approximate Bayesian computation (P. A. Saa and Nielsen, 2016) and maximum *a posteriori* estimation with the use of the Fisher information matrix (Gopalakrishnan et al., 2020; Raue et al., 2013; Liebermeister and Noor, 2021; Stapor et al., 2018).

These approaches have limited utility because, in contrast with MCMC, they lack reliable diagnostic tools for verifying that their results genuinely approximate the target posterior distribution. This is a particularly serious problem for more realistic kinetic models, which tend to have highly correlated joint probability distributions (Gutenkunst et al., 2007; Stapor et al., 2018), making computation failures likely for any inference algorithm. In these circumstances it is essential to be able to detect failures reliably.

Bayesian kinetic models using MCMC
Markov Chain Monte Carlo, or MCMC, allows for posterior sampling with robust failure diagnostics (Vehtari et al., 2021). Modern MCMC algorithms such as Hamiltonian Monte Carlo allow for high-dimensional posterior distributions to be explored and can be combined with fast numerical solvers, making problems involving steady state problems potentially feasible. However, the kinetic modelling literature reports an aversion to MCMC, rooted mainly in concerns about sampling time and the presumed difficulty of implementing the required statistical model (P. A. Saa and Nielsen, 2016; Raue et al., 2013).

We are aware of two recent attempts to implement a Bayesian kinetic modelling ap-

proach using MCMC: St. John et al., 2019 and Stapor et al., 2018 feature detailed kinetic models but these are fit using inefficient MCMC algorithms that do not scale well to high dimensional parameter spaces. These approaches are therefore limited to fitting kinetic models with fewer parameters than could be modelled using, for example, adaptive Hamiltonian Monte Carlo.

St. John et al., 2019 Utilises an efficient sampling algorithm but uses a very approximate kinetic model, namely lin-log kinetics (Visser and Heijnen, 2003). As a result, this framework cannot reliably be used to infer cellular behaviour in experimental conditions that are not similar to a reference dataset.

We are not aware of any previous attempt to develop a general-purpose Bayesian kinetic modelling framework that features both detailed kinetic models and an efficient MCMC algorithm.

3.2 Methods

This section sets out the details of Maud, from its metabolic and statistical models to its code implementation and intended usage.

3.2.1 Workflow

Figure 3.1 visually summarises the intended workflow.

In order to use Maud, a user must first collate appropriate input information, represent it in files with Maud's required formats and put these files in a directory. These files and their formats are shown on the left hand side of figure 3.1 and discussed in more detail in section 3.2.2 below.

The user can now run Maud using a command line command that refers to the target directory. Maud then performs automated kinetic model construction and posterior sampling based on the provided input, as shown in the central column of figure 3.1. Maud produces as output a directory containing a copy of the input directory, debug information and files containing posterior draws and sampler diagnostic information, as shown on the right hand side of figure 3.1.

Maud provides functionality for analysing these outputs by converting draws to SBML files, generating out-of-sample predictions and analysing the contribution of different forms of regulation, as shown by the dotted lines in figure 3.1. Arbitrary further analysis can be performed by on Maud's output files as they use standard formats and include all generated samples - see section 3.3 below for an illustrative example.

3.2.2 File formats

The file formats for Maud's input information are custom. We chose to implement this custom input specification, rather than relying on existing standards, in order to achieve maximum flexibility in development and so as to limit the complexity and scope of Maud's core codebase. Re-purposing an existing format such as SBML would entail far more parsing and conversion and would still have required users to learn
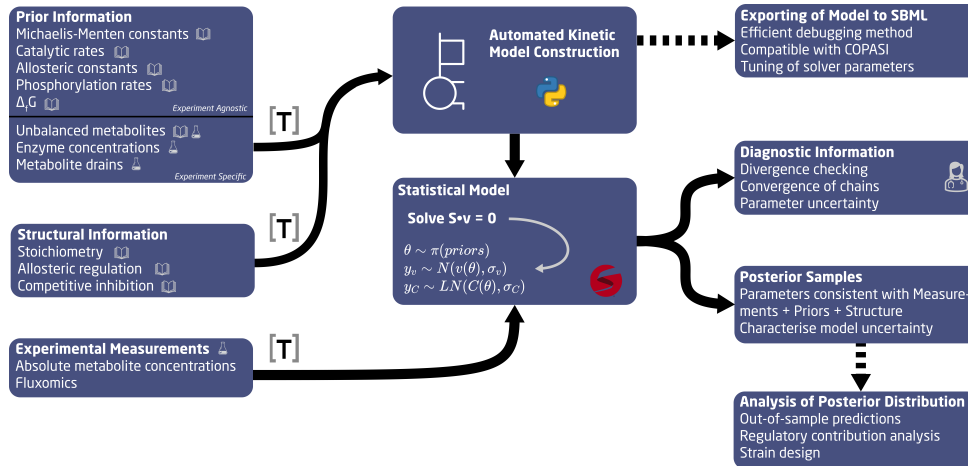
Figure 3.1: A workflow detailing the inputs, processes, outputs and downstream analysis using Maud. The inputs are defined in three separate files, the structural information and regulation is assumed to be constant for the defined model. Structural information is defined using a human readable `toml` input. The prior information for dissociation constants, catalytic rates, allosteric constants, phosphorylation rates, and Gibbs energies of formation is assumed to be experiment agnostic, whereas the unbalanced metabolite concentrations, enzyme concentrations and metabolite drains are defined per experiment. Priors are defined in a `toml` format and are defined either by their 1st and 99th percentile or the mean value and scale parameter. Finally, the experimental measurements of the balanced metabolite concentrations and fluxes are defined in a `toml` with their measured value and error. The model structure is processed in python and converted into a Stan input for sampling. The joint posterior $p(\theta, y)$ is sampled in Stan generating a diagnostics report checking divergences, convergence of chains, and MCMC error. Primarily, the output is a set of posterior samples including generated quantities such as regulatory contributions. The output can be further processed using the 'predict' command to do out-of-sample predictions. Furthermore, for compatability and visualisation purposes, Maud samples can be converted to SBML that is compatible with COPASI (Hoops et al., 2006).

many custom conventions because several of the phenomena that Maud models are not accommodated by existing formats.

The kinetic model file is specified in a `toml` file (Preston-Werner and Gedam, 2020, apr). This file must specify the following structural information:

- compartments and their volumes,

- metabolites,

- enzymes and their numbers of sub-units (if allosterically regulated),

- metabolite/compartment combinations and whether they are balanced,

- reactions and their mechanisms and stoichiometries,

- reaction/enzyme combinations,

- regulatory interactions including allosteric activation and inhibition, competitive

inhibition and phosphorylation.

Experiment-specific information other than measurement values—including temperature, knockouts and whether each experiment should be considered training or test data—are specified in a separate `toml` file.

Priors for parameters representing metabolites' standard-condition Gibbs free energy changes of formation can optionally be specified using a mean vector and covariance matrix, both supplied in `toml` files. Other priors must be specified in a single `toml` file. A separate `toml` file contains results of measurements.

Run-specific configuration is specified in a file which must be called `config.toml`. This file contains paths to the kinetic model file, priors files and experiment files, as well as ODE solver configuration, MCMC sampler configuration, a name and whether to run in prior-only mode.

Maud's output files are in standard formats, including the standard output of `cmdstanpy` (S. D. Team, 2023a), as well as an `arviz` (Kumar et al., 2019) `InferenceData` object saved in a `netcdf` file (Rew and Davis, 1990, July).

Full details about Maud's input and output specifications can be found at the documentation website https://maud-metabolic-models.readthedocs.io/.

3.2.3 maudtools

We have a separate project called maudtools which, among other goals, to convert popular systems biology file formats to skeleton Maud inputs and to generate such files from Maud outputs.

3.2.4 Metabolic model

Maud's metabolic model centres around the stoichiometric matrix S, defining the m compartment-specific metabolites contributing to the n edges of the metabolic pathway. 'Edge' is an umbrella term encompassing both typical biochemical reactions in the form of enzyme-reaction pairs, and also drain reactions. Each compartment-specific metabolite must be categorised by the user as balanced or unbalanced, with balanced metabolites being those whose concentrations are consumed and produced at the same rate when the system is in a steady state. Unbalanced metabolites' concentrations are unaffected by the system's fluxes, but can themselves influence the fluxes by changing its thermodynamics or through modification. The metabolic model is solved to a steady state solution defined by equation 3.1.1 with an n length flux vector v representing the rates of each edge.

To model the fluxes of enzyme-reaction pairs, Maud uses rate laws derived from Liebermeister et al., 2010 with augmentations for allosteric regulation based on Popova and Sel'kov, 1975, and equations describing regulation by phosphorylation based on our own work. To better illustrate each rate in meaningful components we decompose the reaction rate into the following structure 3.2, which we adapted from (Noor et al.,

2013).

$$v = [Enzyme] \cdot Reversibility \cdot Saturation \cdot Allostery \tag{3.2}$$

The term *Reversibility* in equation 3.2 represents the fraction of the forward flux that is catalysed to completion, a quantity that depends only on the system's thermodynamic properties (Noor et al., 2013). The term is expanded in equation 3.3. Where, $\Delta_r G$ is the Gibbs energy of reaction, RT is the universal gas constant multiplied by temperature, $\Delta_f G'$ is the Gibbs energy of formation, S is the stoichiometric matrix, S_{water}^T is the stoichiometry of water for each reaction, $\Delta_f G'_w ater$ is the Gibbs energy of water, *transported charge* is the number of electrons being transported across the membrane, ψ is the membrane potential, and F is the Faraday constant. The Gibbs energies of formation are transformed to the pH and temperature of the cell. Additionally, considering the solution occurs in an aqueous environment we have transformed the reactions that contain water assuming an activity of 1, using the Legendre transform as defined in Alberty, 2003, equation 4.1-1. The closer the reversibility term is to 0 the closer the reaction is to equilibrium.

$$Reversibility = 1 - \exp(\frac{\Delta_r G}{RT}) \tag{3.3}$$
$$\Delta_r G = S^T \Delta_f G' + S_{water}^T \cdot \Delta_f G'_{water} + \text{transported charge} \cdot \psi \cdot F$$

In equation 3.3 the term $\Delta_f G$ is an m-vector containing the standard-condition formation energy of the metabolite in each metabolite/compartment pair. The term R is the gas constant, the term T is the temperature and the term F is the Faraday constant. ψ is the membrane potential and $\Delta_f G'_{water}$ is the standard-condition formation energy of water. By calculating the Gibbs energies of reactions from compound formation energies, the possibility of free energy cycles is ruled out for any configuration of the model's unknown parameters.

This formulation complicates Maud's statistical model somewhat, as data from measurements of equilibrium constants are far more informative as to the values of non-trivial linear combinations of compound formation energies than their non-relative values. In order to capture the resulting correlations, information derived from such data must be represented in the form of a multivariate probability distribution. Luckily, this is possible using tools such as multiTFA (Mahamkali et al., 2021) or eQuilibrator (Beber et al., 2022).

$$Saturation = \frac{\prod_{i=1}^{M}(\frac{C_i}{K_i})^{S_i}}{\prod_{i=1}^{M}(\frac{C_i}{K_i} + 1)^{S_i} + (\prod_{j=1}^{P}(\frac{C_j}{K_j} + 1)^{abs(S_j)} - 1) + (\sum_{k=1}^{I} \frac{C_k}{K_k})} \tag{3.4}$$

The term *Saturation* in Equation 3.2 represents the fractional saturation of the enzyme by the substrates. We define this term in Equation 3.4 where i represents the substrate metabolite up to M, j is the product metabolite up to product P, C is the concentration of the metabolite, K is the dissociation constant of the metabolite, S is the reaction stoichiometry of the metabolite, and k is the competitive inhibitor metabolite up to inhibitor I. Additionally, the saturation term defined in Equation 3.4 captures competitive inhibition as described in Liebermeister et al., 2010. As competitive inhibitor and product concentrations increase, the saturation denominator increases, effectively decreasing the saturation of the substrate on the total enzyme pool. Conversely, as the substrate concentration increases this term approaches 1. Equation 3.4 changes whether or not the reaction is irreversible or not. The irreversible form entirely removes the terms within the parenthesis that contain the product terms, including the -1.

$$Allostery = \frac{1}{1 + L_0(e_R \cdot \frac{\prod_{i=1}^{t}\left(1+C_i/K_{T,i}\right)}{\prod_{j=1}^{r}\left(1+C_j/K_{R,j}\right)})^n} \tag{3.5}$$

The term *Allostery* as defined in Equation 3.5 represents the fraction of active enzyme in the MWC formulation (Popova and Sel'kov, 1975; Monod et al., 1965; Changeux, 2013; Popova and Sel'kov, 1979). Conformational changes are considered as in P. Saa and Nielsen, 2015 with the simplification of considering only a single binding domain for effector metabolites, this removes the sum term in P. Saa and Nielsen, 2015 leaving only a ratio of polynomial products between the activating and inhibiting effector metabolites. In Equation 3.5 L_0 represents the transfer constant, which determines the equilibrium between inactive and active states assuming no effectors are present, e_R is the free enzyme ratio and is the denominator in Equation 3.4, C is the effector metabolite concentration, K is the dissociation constant for the effector/enzyme complex, i is the inhibitor metabolites up to r, j is the effector metabolite up to r, and n is the number of subunits in the complex. This model considers the enzyme as either in an active (R) or inactive (T) state. The equilibrium between these states is determined by the binding of effector metabolites that stabilise either the R or T states. We consider a special instance of the MWC model where the T form carries no flux and hence cannot be bound to substrates. Additionally, we do not consider complexes which are combinations of different isoenzymes.

Our formulation, with each term isolated from the others, provides no computational benefit to the model. However, the separation of terms is scientifically insightful, with the mechanisms providing understanding and guiding engineering techniques (Hackett et al., 2016; Noor et al., 2013). Furthermore, tailored rate laws can always be introduced into Maud, with these providing a solid foundation for most modelling needs. The decision to focus on rate laws that have macroscopic parameters rather than mi-

croscopic parameters (Matos, 2018) was influenced by the amount of available information to use as priors (Chang et al., 2021).

3.2.5 Regulatory comparison

In Section 3.2.4 we presented the flux for each enzyme as a product of separable terms 3.2. As these terms are separable and mechanistically informative we can extend the work of Noor et al., 2013 to include the allosteric term defined in Equation 3.5. We can then compare these terms between conditions to determine the exact mechanism by which the enzyme was able to change flux.

$$\frac{v_2}{v_1} = \epsilon_{\frac{2}{1}} \cdot \Theta_{\frac{2}{1}} \cdot \gamma_{\frac{2}{1}} \cdot \Psi_{\frac{2}{1}} \tag{3.6}$$

Where, $\tag{3.7}$

$$\epsilon = \frac{[Enzyme_2]}{[Enzyme_1]} \tag{3.8}$$

$$\Theta = \frac{Saturation_2}{Saturation_1} \tag{3.9}$$

$$\Gamma = \frac{Reversibility_2}{Reversibility_1} \tag{3.10}$$

$$\Psi = \frac{Regulatory_2}{Regulatory_1} \tag{3.11}$$

We define the flux ratio $\frac{v_2}{v_1}$ between condition 2 and 1 as the product of regulatory ratios in Equation 3.6. These value of each term shows how each regulatory component changed between conditions and can be used to make statements surrounding any component of the mechanism, e.g., allosteric control between two flux states. This method is most informative when the flux between two conditions is in the same direction as the saturation and reversibility terms lose interpretability as what is considered the substrate in one direction is the product in the other. However, this is a powerful tool for understanding regulation of complex networks.

3.2.6 Solving the steady state problem

In order to connect a parameterised kinetic model with measurements, it is necessary to find a configuration of balanced compartment-specific metabolite concentrations that leave the system in steady state, or in other words to solve the equation $S_{balanced} \cdot v = 0$ for balanced compartment-specific metabolite concentrations. Due to the complex and non-linear rate equations in Maud's metabolic model this problem is difficult to solve for realistic networks.

Maud addresses this problem by numerically simulating the system from a fixed starting point for a fixed period of time, and then verifying whether a steady state has been reached by checking whether $S_{balanced} \bullet v = 0$ is approximately true. For numerical ODE solving Maud uses the backward differentiation formula solver provided by the library CVODES (Serban and Hindmarsh, 2005) through Stan. To confirm that the system

achieves steady state Maud performs both an absolute and relative tolerance check defined in equation 3.12.

$$
\frac{\frac{dC}{dt}}{C} < relative\ tolerance
$$
$$
\frac{dC}{dt} < absolute\ tolerance
$$

$$(3.12)$$

This way of solving the steady state problem is somewhat inelegant and could perhaps be improved by formulating the steady state problem as an algebraic equation system. However it is reliable in practice and has historically been popular in this field

3.2.7   Statistical model

Maud follows the standard statistical modelling practice of representing information from measurements using a measurement model in the form of a probability function $l : Y \times \hat{Y} \to [0, 1]$ that describes the conditional probability density of any set of measurements $y \in Y$ given true values $\hat{y} \in \hat{Y}$ of measured quantities such as fluxes, metabolite concentrations and protein concentrations.

Maud's metabolic model, together with the steady state equation 3.1.1, provides a rule determining the values of measurable quantities based on a set of unknown parameters θ, which lies in the space $\mathcal{T}$. Non-experimental information is represented by a prior model $p : \mathcal{T} \to [0, 1]$ that describes the pre-experimental probability density of any possible parameter configuration $\theta \in \mathcal{T}$. Together the measurement and prior models determine a joint probability function $\pi : Y \times \mathcal{T} \to [0, 1]$ that assigns a probability density to any combination of measurements and parameters.

Questions about the kinetic model can be expressed as integrals of the joint probability function. For example, suppose we have a dataset y and are interested in whether a kinetic parameter θ_j is greater than a certain value v. Given a Bayesian statistical model encapsulated by the joint probability function $\pi : Y \times \mathcal{T} \to [0, 1]$, the answer to this question corresponds to the quantity $\pi(\theta_j > v | y = y) = \int_{i=v}^{\infty} \pi(\theta_j = i | y = y) d\theta_j$.

Parameters

Maud's statistical model has unknown parameters corresponding to quantities in the metabolic model that are assumed to be unknown, other than steady state balanced metabolite/compartment concentrations and fluxes, which are derived from the values of other parameters by solving the steady state problem as set out in Section 3.2.6. See Table 3.1 for a description of all the parameters in Maud's statistical model and their dimensions.

Note that Maud's metabolic model includes some quantities that are not treated as parameters in its statistical model, including temperatures, compartment volumes and the formation energy of water. Maud treats these quantities as if they were known precisely: they can be configured by the user or default values can be used. Although in

Table 3.1: Parameters of Maud's statistical model

Parameter	Modelled quantity	Dimensions
ΔG_f	Formation energies	Metabolites
K_m	Dissociation constants	Substrates of all enzyme/reactions and products of reversible enzyme/reactions
K_i	Inhibition constants	Inhibiting metabolite/compartments of enzyme/reactions exhibiting competitive inhibition
K_{cat}	Rate constants	Enzyme/reactions
e_j^T	T dissociation constant	Modifying metabolite of allosteric interactions for metabolite j
e_j^R	R dissociation constant	Modifying metabolite of allosteric interactions for metabolite j
L_0	Transfer constant	Allosteric interactions
v_{drain}	Drain fluxes	Drains, experiments
C_{enzyme}	Enzyme concentrations	Enzymes, experiments
$C_{unbalanced}$	Unbalanced metabolite/compartment concentrations	Unbalanced metabolite/compartments, experiments
ψ	Membrane potentials	Experiments

practice there can be considerable uncertainty regarding these quantities, we chose to disregard this uncertainty in the interest of simplicity.

Measurement model

Maud's statistical model takes into account three types of measurement: metabolite concentration measurements, enzyme concentration measurements and flux measurements, represented by vectors y^{conc}, y^{enz} and y^{flux} respectively.

All measurements are specific to an experimental condition - that is, a case where the true state of the network, including knockouts, boundary conditions and state variables as well as kinetic and thermodynamic parameters, can safely be assumed to be the same. Maud's statistical model allows for arbitrarily many experimental conditions, and for any measurable quantity to be measured any number of times in any condition, including zero.

Metabolite and enzyme measurements are intended to represent the results of quantitative metabolomics and proteomics experiments. The likelihood functions for such measurements are shown in equations 3.13 and 3.14.

$$y_i^{conc} \sim LN(\ln \hat{y}_i^{conc}, \sigma_i^{conc}) \tag{3.13}$$

$$y_i^{enz} \sim LN(\ln \hat{y}_i^{enz}, \sigma_i^{enz}) \tag{3.14}$$

Both equations are log-normal generalised linear models with a standard link function (the natural logarithm ln) and known standard deviation σ^{conc}. The use of this measurement model is motivated by the consideration that concentrations are by definition non-negative, so the measurement model should avoid assigning positive probability mass to negative metabolite concentration values. In addition, we expect the precision of most metabolomics and proteomics experiments to be roughly proportional to the value of the true measured quantity, which supports a measurement model with constant coefficient of variation. The measurement standard deviations σ^{conc} and σ^{enz} are assumed to be known exactly for simplicity, and can be elicited by considering the likely coefficient of variation of the measuring apparatus.

Analysis of metabolomics and proteomics data often assumes a regression model with normally-distributed errors, whether explicitly through the use of a standard linear model or implicitly through the use of ordinary least squares fitting. This is undesirable because the assumption of normal errors implies that the measured quantity could in principle be negative, which is false by definition, and because it implies an additive underlying random process, while a multiplicative process tends to better describe real data.

The use of independent measurement models carries an implicit assumption that there are no systematic correlations in the measurement errors. This choice was motivated by simplicity - it would be better to use a model with potentially correlated measurements, but such a model would be more complicated and harder to fit.

Flux measurements representing the results of quantitative fluxomics analyses are taken into account using a likelihood function from a standard linear model, as shown in equation 3.15.

$$y_i^{flux} \sim N(\hat{y}_i^{flux}, \sigma_i^{flux}) \tag{3.15}$$

There are two important considerations that arise in connection with Maud's measurement model for fluxes. First, it is important only to specify measurements for a network's free fluxes, as the values of some steady state fluxes in a metabolic network are constrained by others, with the result that dependent fluxes cannot typically be measured separately. If measurements of multiple dependent fluxes are entered, information will inappropriately be double-counted.

Second, whereas quantitative metabolomics and fluxomics measurements, though complicated, produce results that approximately correspond to the given likelihoods, this is not the case for fluxomics measurements, which typically derive from labelling experiments. In order to derive quantitative flux values from such results, an upstream modelling step must be performed, for example as described in Young, 2014.

Ideally Maud's measurement model for fluxes would extend from fluxes to the results

of potential labelling experiments, thereby removing the need for upstream analysis and avoiding any double-counting. This option has not yet been pursued due to its complication.

Prior model

With the exception of metabolites' standard condition Gibbs energy changes of formation, Maud uses independent normal prior distributions for parameters that can in principle be both negative and positive. For parameters that are by definition positive, Maud uses independent log-normal distributions. Formation energy parameters have a multivariate normal prior distribution.

Posterior sampling - Hamiltonian Monte Carlo

Although integrals of the joint probability model for kinetic models are typically analytically intractable, they can be approximated numerically using Markov Chain Monte Carlo (MCMC) and other methods. Maud uses MCMC primarily because there exist many methods for verifying that MCMC samples really do approximate the target probability distribution: see Vehtari et al., 2021 and Talts et al., 2018 for discussion of this point. In addition, there are several examples of successful Bayesian kinetic modelling projects using MCMC including St. John et al., 2019 and Xing et al., 2010. The following section will explain the theory surrounding Hamiltonian Monte Carlo (HMC) and its necessity in high dimensional problems such as those experienced in Maud: this section is based on the overview of HMC by Betancourt, 2017 and Gelman et al., 2013.

Bayes' theorem solves the problem of "What is the probability of my parameter set given the observed data" and has the following mathematical form: $P(\theta|y)$, and will be referred to as the posterior distribution. To calculate the posterior distribution we can use Equation 3.16.

$$P(\theta|y) = \frac{P(y|\theta)P(\theta)}{P(y)} \tag{3.16}$$

Where, $P(y|\theta)$ is the likelihood of observing the data given the parameter set θ, and $P(\theta)$ is the probability of observing θ. The denominator of Equation 3.16 refers to the probability of the data $P(y)$ and is constant over the parameter space, hence we can rewrite Equation 3.16 to be proportional to the numerator terms, this is shown in Equation 3.17.

$$P(\theta|y) \propto P(y|\theta)P(\theta) \tag{3.17}$$

Evaluating the right hand side of Equation 3.17 requires evaluation of the integral defined in Equation 3.18

$$P(y|\theta)P(\theta) = \int_\theta p(y|\theta)p(\theta)d\theta \tag{3.18}$$

Equation 3.18 is often not analytically tractable and can be approximated using MCMC. The underlying process of MCMC is to generate a series of parameter sequences $\theta^1, \theta^2, \theta^3, \ldots, \theta^n$, each based on the previous point where the next point is drawn around the previous point using a transition distribution $T_i(\theta^i|\theta^{i-1})$. However, these sequences, also known as Markov chains, are not just a series of consecutive jumps that diffuse freely throughout the parameter space. The sequence only accepts the next based based on the posterior probability ratio, as shown in Equation 3.19.

$$ r = \frac{P(\theta^{proposed})P(y|\theta^{proposed})}{P(\theta^{i-1})P(y|\theta^{i-1})} \tag{3.19} $$

In Equation 3.19, the proposed parameters $\theta^{proposed}$ accepted with the probability of r and run for a sufficiently long period of time the Markov chain will approximate the posterior distribution. And the result of a successful MCMC simulation is a set of draws from the posterior distribution that can be used for parameter inference.

However, approaching the target distribution in N dimensional space becomes more difficult as N increases, a problem known as the *curse of dimensionaility*. This emerges because as you increase the number of dimensions, the probability of choosing a poor guess, $\theta^{propsed}$, increases, and the efficacy of the MCMC sampler deteriorates. Simply put, the higher the dimensionality, the more difficult it is to sample the posterior distribution, and the longer the Markov chain needs to be before you converge to the posterior distribution.

Hamiltonian Monte Carlo (HMC), solves the issue of taking poor guesses of the posterior distribution by using the information provided at the the current state θ^i. If the gradient of the log posterior density $log\, p(\theta|y)$ is taken with respect to the current parameter set θ^i, then the next guess can be limited to areas that are more likely to be in areas of high probability density. This algorithm is defined and explored in further detail in (Betancourt, 2017; S. D. Team, 2023a; Gelman et al., 2013). However, the central principle is that HMC increases the number of effective samples per step and makes high N problems computationally feasible. The implementation of HMC used throughout this work was made possible by the Stan developers (S. D. Team, 2023b).

3.2.8 Implementation

Maud is written in Python (Rossum and Drake, 2010), with versions greater than 3.8 fully supported, and depends on popular Python libraries and the probabilistic programming framework Stan. It supports the operating systems Windows, Linux and macOS. It is kept under active development at https://github.com/biosustain/Maud.

This section explains how Maud works from a software point of view.

Installation

Maud is registered on the Python package index and can be installed from the command line with the command `pip install maud-metabolic-models`.

Due to the dependency on Cmdstan, an additional post-installation step is required in order to use Maud. This is straightforward because after installing Maud a convenience command `install_cmdstan` is accessible from the command line thanks to the already-installed Python dependency `cmdstanpy` (S. D. Team, 2023a).

Command line interface

Maud uses the Python 3 library `click` (Developers, 2022) to implement a command line interface with the general format `maud [function] [path]`, where `[path]` is a path from the present working directory to an input or output directory.

The available commands are as follows:

- `maud --help`

- `maud sample [PATH TO INPUT DIRECTORY]`

- `maud simulate [PATH TO INPUT DIRECTORY]`

- `maud predict [PATH TO OUTPUT DIRECTORY]`

- `maud variational [PATH TO INPUT DIRECTORY]`

Parsing, data model and validation

Maud uses Python 3 and standard scientific libraries to parse input files, define a data model representing the input abstractly and perform validation. Specifically, the following libraries are used: `numpy` (Harris et al., 2020), `scipy` (Virtanen et al., 2020), `sympy` (Meurer et al., 2017), `pandas` (T. P. D. Team, 2023), `toml` (Preston-Werner and Gedam, 2020, apr) and `pydantic` (P. developers, 2022).

Statistical model definition

Maud's statistical model is defined by files with the extension `.stan`, written in the probabilistic programming language Stan (Carpenter et al., 2017). Specifically, Maud has three main Stan files:

- `functions.stan`, containing functions representing rate equations.

- `model.stan`, a file specifying Maud's main statistical model.

- `out_of_sample_model.stan` , a Stan program for generating out-of-sample predictions.

Creating outputs

To produce an output from a valid input, Maud uses the program `CmdStan` (S. D. Team, 2023b) via the Python interface `cmdstanpy` (S. D. Team, 2023c). The results are then converted to `InferenceData` and saved as a `netcdf` file using the Python library `arviz` (Kumar et al., 2019).

Tests

Maud uses the Python libraries `isort` (I. developers, 2022), `black` (B. developers, 2022) `flake8` (flake8 developers, 2022), `flake8-docstrings` (flake8-docstrings developers, 2022) and `flake8-bugbear` (flake8-bugbear developers, 2022) for style checking, the

library `safety` (safety developers, 2022) to check for known security vulnerabilities and the library `pytest` (pytest developers, 2022) to run custom unit tests. Maud uses the library `tox` (tox developers, 2022) to create test environments in all supported Python versions. All tests are run automatically on pull requests using Github actions.

Documentation

Maud's documentation is written in RestructuredText format (reStructuredText developers, 2001) built using the Python library Sphinx (Georg and team, 2022) and hosted at https://maud-metabolic-models.readthedocs.io.

3.2.9 Calculating the Maximum A Posteriori (MAP) estimate and covariance

As a comparison for HMC, we will use the Laplace approximation to estimate the posterior distribution. The Laplace approximation is a method for approximating the parameter variance of a posterior distribution by fitting a multivariate Gaussian distribution centred at the MAP estimate. Calculating the covariance of this Gaussian distribution requires the Hessian of the log posterior distribution (Equation 3.21) evaluated at the MAP estimate (Equation 3.20).

$$\hat{theta} = argmax_\theta p(\theta|y) \tag{3.20}$$

$$H(\theta) = \frac{\delta^2}{\delta\theta_i\delta\theta_j} \log p(\theta|y) \tag{3.21}$$

In Equation 3.21 $H(\theta)$ is the hessian at parameter set θ, i and j are indices of the parameter set θ, and $p(\theta|y)$ is the posterior probability density. We can use the Hessian as an estimate for the covariance of the posterior probability density.

$$Var(\hat{\theta}_{MATP}) = -H(\hat{\theta}_{MAP})^{-1} \tag{3.22}$$

In Equation 3.22 $Var(\hat{\theta}_{MAP})$ is the covariance matrix of the parameters θ at the MAP estimate $\hat{\theta}_{MAP}$. The resulting uncertainty of the parameters using the Laplace approximation is given by Equation 3.23.

$$\theta \propto Normal(\hat{\theta}_{MAP}, -H(\hat{\theta}_{MAP})^{-1}) \tag{3.23}$$

Simultaneous estimation of the MAP

3.3   Case study

To illustrate how Maud is used in practice we performed inference on a kinetic model of the methionine cycle. We first used Maud to generate fake measurements from known

parameter values in a range of training and testing conditions, then ran Maud in sampling mode to extract information from the training conditions. To showcase Maud's predictive power we used it to place probability distributions over boundary conditions and compared these with the testing dataset. To illustrate the improvement from using a fully Bayesian approach as opposed to augmented maximum a posteriori estimation, we also fit our methionine cycle model using the latter method and compared the results to those obtained using Maud.

Finally, to show how Maud can be used in scientific problems to guide experimental design, we defined a simple methodology for choosing the best experiment to perform next given some data and applied it to our case study.

3.3.1 The Methionine cycle

The methionine cycle is a fundamental pathway in human metabolism, whose intermediate metabolites participate in a variety of mechanisms which must compete for the same resources. Due to this competition, as well as the fact that all the functions occur simultaneously, the methionine cycle is highly regulated, with 6 known allosteric effectors. This complex regulation makes the methionine cycle a good illustration of Maud's capabilities, as inferences based on less detailed metabolic models are likely to be unreliable in this case.

We constructed a kinetic model of the methionine cycle in Maud's format using the description in Korendyaseva et al., 2008. Figure 3.2 shows the reactions and allosteric interactions that we modelled.

The ordinary differential equation system describing this model is shown in equation 3.24.

$$
\begin{aligned}
\frac{d[met]}{dt} &= v_{Influx} - v_{PROT} - v_{MAT} + v_{MS} + v_{BHMT} \\
\frac{d[amet]}{dt} &= v_{MAT} - v_{GNMT} - v_{METH} \\
\frac{d[ahyc]}{dt} &= v_{GNMT} + v_{METH} - v_{AHC} \\
\frac{d[hyc-L]}{dt} &= v_{AHC} - v_{CBS} - v_{MS} - v_{BHMT} \\
\frac{d[5mthf]}{dt} &= v_{MTHFR} - v_{MS}
\end{aligned}
\tag{3.24}
$$

3.3.2 Measurement Errors

For simplicity, we specified a measurement error of 10% for all proteomics and metabolomics measurements. The accuracy of real quantitative metabolomics and proteomics experiments can be highly variable due to matrix effects, equipment noise and. Our choice of 10% is somewhat optimistic, but is a reasonable first approximation to a realistic set of measurements.
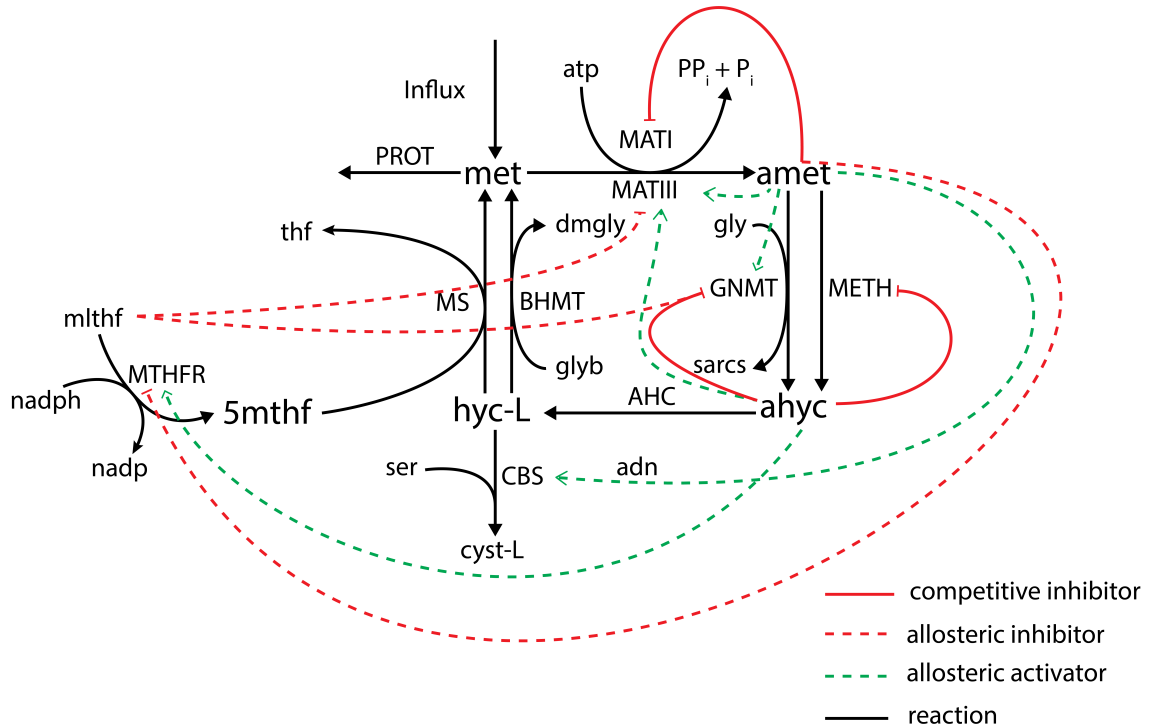
Figure 3.2: The methionine cycle as modelled, with the solid black lines representing the reactions, the green lines representing allosteric interaction, and the red lines representing allosteric inhibition. The capitalised fonts are the reaction names and the lower-case font represents the metabolites.

3.3.3 Priors

The prior distributions and corresponding true parameter values used in our case study are shown in table B.1. The first two columns show the 1% and 99% quantiles of each marginal prior distribution. True parameter value are shown in column three, and the last column shows the z-score on log scale of the true parameter value according the marginal prior distribution. As shown in Appendix B there are 7 parameters for which the true value is outside of 1%-99% range. This is desirable, making the case study more realistic, because extreme deviance from the prior distribution is likely to occur in practice due to *in vivo* to *in vitro* measurement differences.

3.3.4 Training testing split

The training testing split was selected to achieve a large difference between the fluxes of the training and testing dataset. The split was determined as we are interested in showing how our model can fit to varied conditions, and conditions closer to the training set are likely to be predicted well without necessarily learning the system.

3.3.5 Diagnostic checks

We first confirmed that the sampling process was valid using diagnostic tools provided by Maud via `cmdstanpy` and `arviz`. All sampled variables had improved $\hat{R}$ statistics less than 0.01 from 1, indicating appropriate mixing within and between Markov chains. See Vehtari et al., 2021 for more about this diagnostic statistic. Additionally, the num-

ber of effective samples was high ($>$1000), indicating that we generated enough posterior samples to support inferences about the bulks of the distributions of the sampled parameters. Furthermore, we observed no post-warmup divergent transitions that were not a result of ODE failure in any of the chains, indicating that the sampler was able to transform the log-posterior distribution so as to avoid any regions with excessive curvature that might frustrate exploration via HMC. Overall, we concluded from investigation of the diagnostics that our samples were are reasonable approximation to the target distribution.

3.3.6 Evaluation of model fit

Once we were satisfied with the diagnostic checks, we performed posterior predictive checks to see if our model's predictions about measurable quantities matched the observed data, and looked at the marginal and joint distributions of parameters of interest to get an idea of what the model learned from the measurements. Figure 3.3 shows a characteristic example of these investigations.

The model achieved a good fit to the provided synthetic fluxomics and metabolomics measurements, with the observed value generally falling within the 5%-95% posterior prediction interval, both for the training and test experiments, as can be seen the top row of figure 3.3. This suggests that given our model definition we are able to fit the data and also conduct inference on conditions outside of our training space.

We observed correlations in the pairwise marginal distributions of some parameters, as shown by the pair plot in the bottom left of figure 3.3, according to which there is a roughly banana-shaped relationship between the marginal posterior distributions of two k_m parameters. More common, however, were uncorrelated parameter pairs as seen in the bottom-right plot in figure 3.3.

These findings align with those of Gutenkunst et al., 2007, according to which parameterised non-linear biological systems are typically "sloppy" in the sense that many parameter configurations result in the same steady-state. Without carefully targeted measurements it is rare that relationships between pairs of parameters, as opposed to, say, three-or-more-way relationships will tightly constrain the system's behaviour, as would be required in order to decorrelate their pair plots. Similarly, the 'one-way' marginal posterior distributions for parameters in sloppy systems tend to be approximately the same as the corresponding marginal priors, despite there being large differences between the joint prior and posterior distributions overall, and between the marginal prior and posterior distributions of state variables (Poirier, 1998).

Our example model does indeed exhibit sloppiness, as very few parameters have substantially different marginal prior and posterior distributions. However, the overall posterior is very different from the prior, as can be seen in the different marginal distributions for measurable quantities like concentrations and fluxes. Further, the model makes accurate predictions about out-of-sample measurements, which are far better
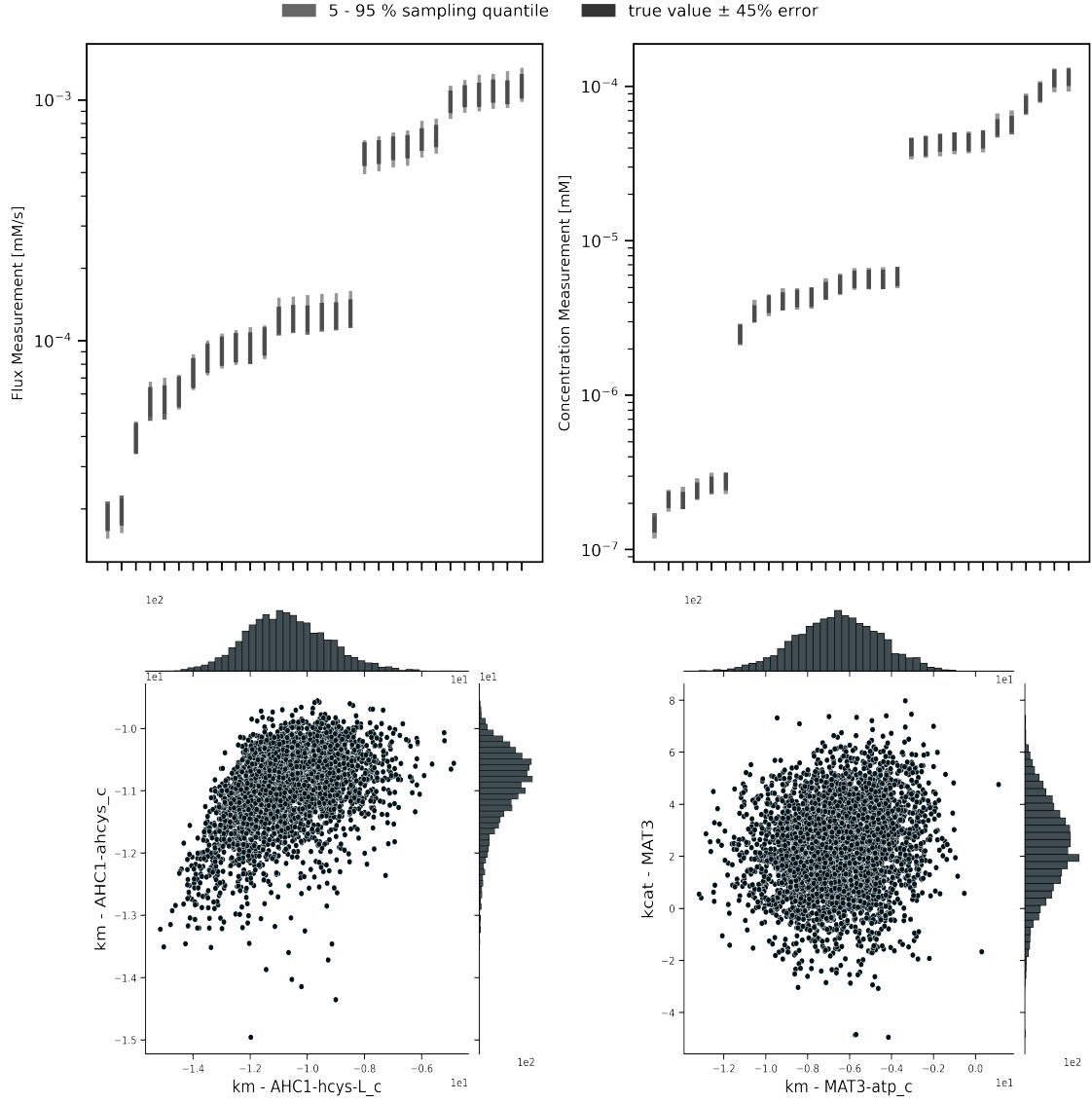
Figure 3.3: **(Top Row)** Comparison of posterior predictive intervals with true values of measured quantities indicate that the training set was able to capture the true model behaviour given their measurement model. **(Bottom Left)** An example of a correlated pair of marginal posterior distribution observed between $K_m^{AHC1,ahcys_c}$ and $K_m^{AHC1,hcys_c}$. **(Bottom Right)** An uncorrelated parameter set within MAT3 between the k_{cat}^{MAT3} and K_m^{MAT3,atp_c}

than predictions generated using only the prior model.

3.3.7 Missing experimental data

A common occurrence in multiomics studies is not having access to all measurements. For instance, measuring all metabolites in a pathway can be infeasible because of limitations of mass spectrometers, availability of standards, column effects, and compartmentalisation. However, provided that sufficient information is available from other sources, our approach can produce sensible results from incomplete metabolomics data.
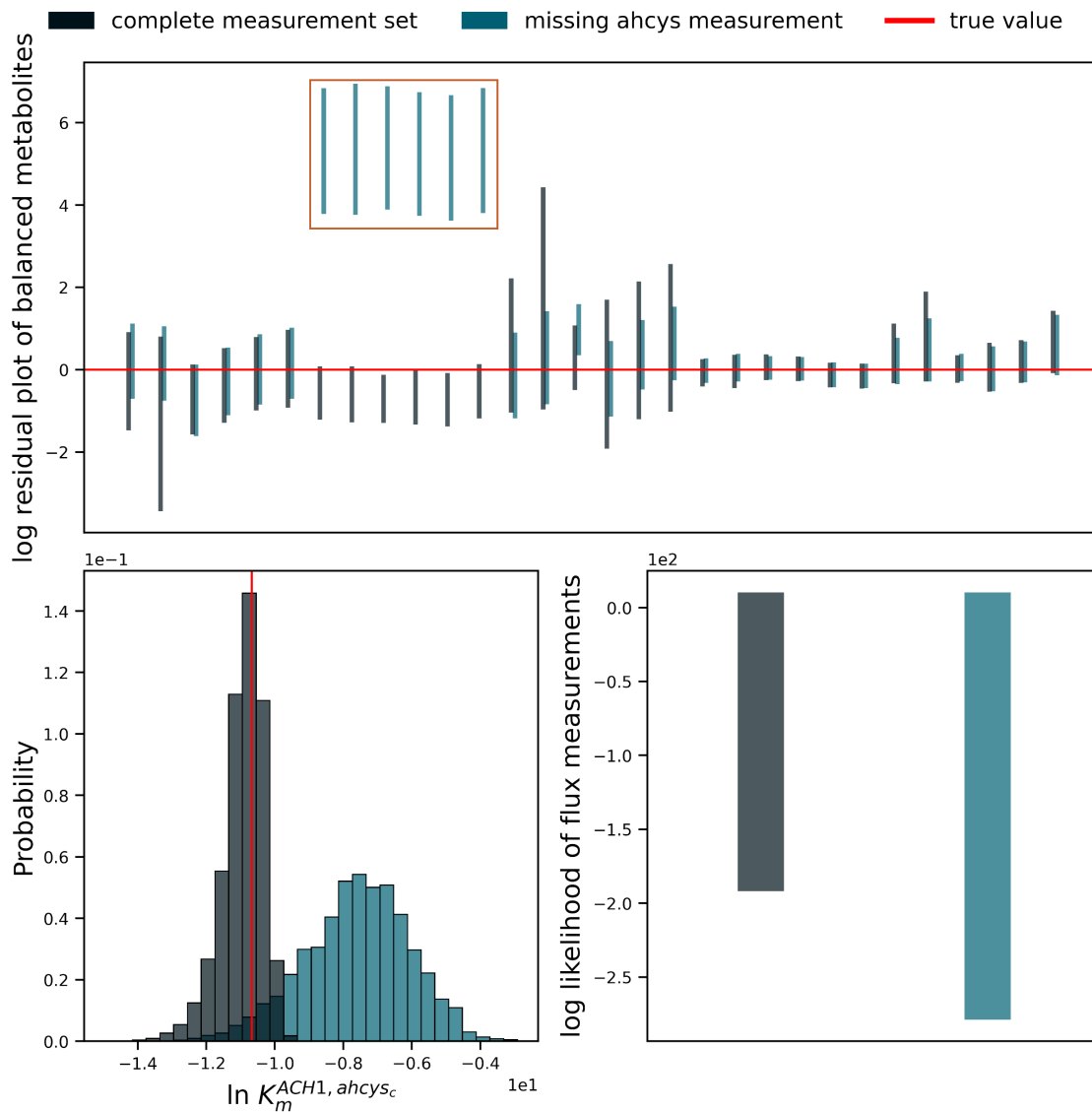
Figure 3.4: **(Top)** A comparison between the full measurement dataset and the dataset missing measurements for $ahcys_c$. This is a plot of the 5% to 95% sampling interval log residuals for all balanced metabolite concentrations, where a value of 0 represents a prediction of the true value. These results show significant deviation in the missing ahcys measurement set when predicting the concentrations of ahcys. The ahcys estimates of the model lacking the ahcys measurement are highlighted by the orange box. **(Bottom Left)** The marginal posterior distribution of $K_m^{AHC1,ahcys_c}$ comparing the convergence to the true parameter value (red line) when the measurement for ahcys is available and not. As seen in this figure, the true value is only recovered when the metabolite is available. **(Bottom Right)** The 5% to 95% interval of the sum of the log-likelihoods for the flux measurements within the testing set comparing the complete measurement set to the dataset missing ahcys measurements. As observed there is a significant overlap between the two flux predictions that suggests missing the ahcys measurement did not cause catastrophic failure for the model. However, the prediction interval is overall better when the full measurement set is available.

In order to demonstrate this capability, we amended our case study to reflect a representative case of missing metabolite measurements, fitting Maud to the same 6 experimental datasets, but with measurements of the metabolite *S-Adenosyl-L-homocysteine*, or "ahcys" removed. Since ahcys regulates three enzymes in the methionine cycle, including one enzyme which is also thermodynamically regulated, we expected the removal of these measurements to yield interesting results. We then compared the results to those previously generated based on a full metabolomics dataset; the results of the comparison are shown in figure 3.4.

As we can see in figure 3.4, both models predicted the out of sample concentrations adequately, except that the missing metabolite model was unable to infer any ahcys concentrations. However, despite the inferred ahcys concentrations being far from the measured value, the remaining measured metabolites were still well predicted, suggesting that information about the measured network is still preserved with missing measurements. To extend this observation to measurements of fluxes, we calculated the 5% - 95% interval for these, which indicated that the missing measurement did not result in catastrophic model failure.

The missing measurements did affect Maud's ability to infer parameter values correctly. As we can see in the lower left plot of figure 3.4, the model with full metabolomics accurately learned the region where the true value was drawn from, despite this value being far from the mean of the corresponding marginal prior distribution, whereas the model with missing ahcys measurements stayed in the neighbourhood of the prior.

A kinetic investigation of the Warburg effect in CHO cells

3.3.8 Point estimates compared with MCMC sampling
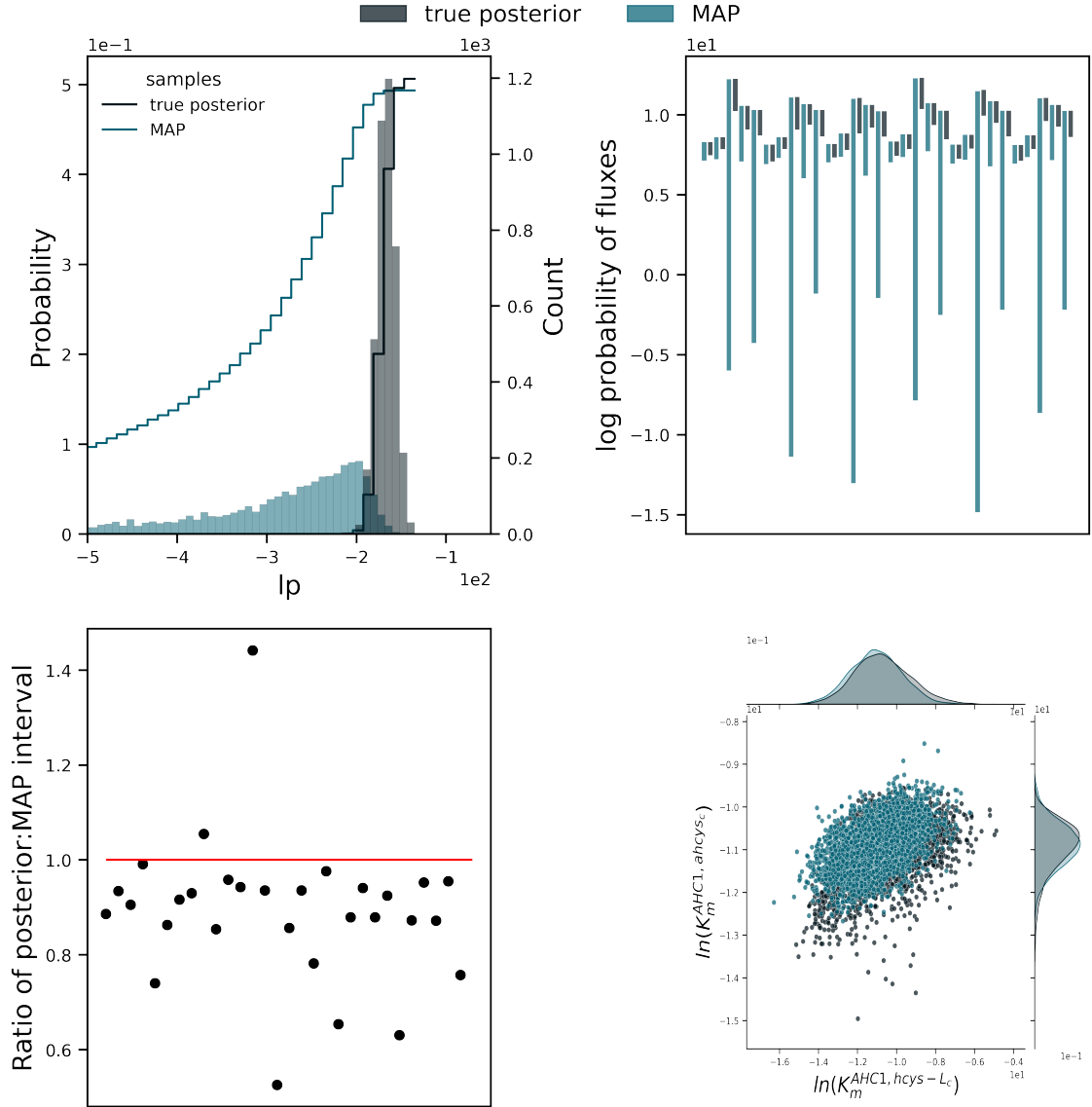
Figure 3.5: **(Top Left)** probability density function (PDF) and cumulative distribution function (CDF) of the total log probabilities comparing the true posterior and the maximum a posterior approximation. This comparison shows that the MAP does not approximate the true posterior. **(Top Right)** Shows the 5% to 95% interval of the log-likelihoods of the sampled fluxes for the training dataset, comparing the true posterior sampled in Maud to that of the MAP approximation. As shown, the sampled fluxes are significantly worse in the MAP approximation compared to the true posterior. **(Bottom Left)** Comparing the prediction intervals of the posterior and the MAP approximation. We plotted the ratio of the 5% to 95% quantile width given by the true posterior divided by the equivalent quantile width of the MAP approximation. All but two of these ratios are lower than 1 indicating that the MAP was interval too wide. The approximation using MAP leads toa bias where posterior intervals are estimated to be inappropriately wide compared with the true posterior. **(Bottom Right)** Comparing the log joint marginal distribution of $K_m^{AHC1,hcys-L_c}$ and $K_m^{AHC1,ahcys_c}$ for the true posterior and the MAP approximation. The MAP distribution does not encompass the full sampling space obtained in the true posterior distribution and hence is a poor method for parameter inference.

A common approach to fitting parameters is to use a point estimate such as the maximum likelihood estimate (MLE) or maximum *a posteriori* estimate (MAP). These approaches aim to find a single optimal parameter configuration: either a configuration that makes the observed data maximally likely given the measurement model or a configuration that maximises the joint posterior density given a Bayesian statistical model. In order to achieve some degree of uncertainty quantification, the optimal parameter configuration is often supplemented with a multivariate normal distribution over the unconstrained parameter space, derived from the gradients of the objective function at the optimal parameter configuration. This distribution can be used as an approximation to the true posterior distribution, potentially avoiding the need for costly alternative posterior approximation methods like MCMC.

In order to evaluate whether this alternative method provides a viable alternative to MCMC for Bayesian kinetic models, we compared the results of our case study with results derived from the approach based on MAP plus a gradient-based distribution. Figure 3.5 summarises the results. As can be seen from the top left plot, the MAP method does not provide a good approximation to the true posterior distribution, as the marginal distribution of the total log probability density is clearly different. This was confirmed using the Kolmogorov-Smirnov test, which is a test to differentiate two empirical uni-variate distributions. We found that the distributions are significantly different with a p-value indistinguishable from 0, and hence the MAP-plus-gradient-based-distribution method is not an appropriate approximation of the posterior distribution in this case.

The difference between the MAP-plus output and the true posterior distribution manifests itself not only in the parameter space, but also in the measurement space. Figure 3.5 frame B shows the 5% to 95% interval for log probability densities for flux measurements in the HMC and MAP case. Lower log probabilities indicate that the sample and the measured values are further away. The sampled values for all fluxes are significantly worse for MAP when compared to HMC. This highlights that model intervals presented by MAP methods do not estimate model behaviour even within the training set. Most of the discrepancy is located within

To evaluate the practical consequences of this finding, we observed that the 5%-95% posterior predictive intervals for fluxes tended to be substantially narrower in the true posterior distribution compared to the MAP approximation, as seen in the top right cell of figure 3.5. This suggests that, perhaps due to the phenomenon of sloppiness as mentioned above, systems like the one we investigated cannot be well understood based on a single parameter configuration and its gradients. Unfortunately, the whole joint posterior distribution must be investigated.

To further explore why this is the case we compared the pairwise marginal distributions of two dissociation constants - $K_m^{AHC1,hcys-L_c}$ and $K_m^{AHC1,ahcys_c}$ - with samples generated by both HMC via Maud and the MAP-plus approach. This comparison is

shown in the bottom right cell of figure 3.5. Overlaying the samples in this way demonstrates that the MAP method is not able to capture the correct relationships between parameters' distributions, even for merely two-way parameter relationships.

3.3.9 Application to regulatory understanding
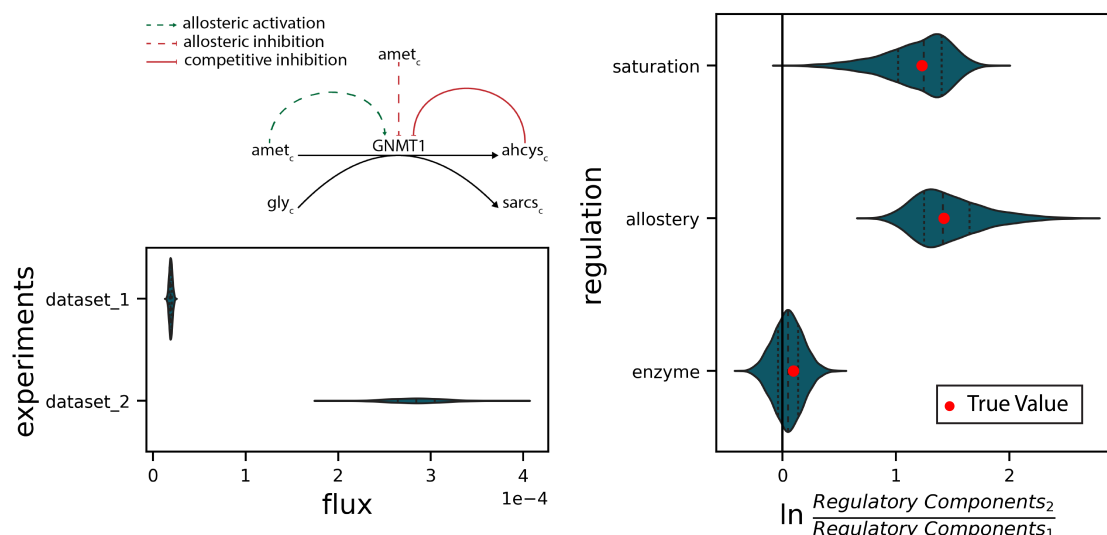
Figure 3.6: **(Top Left)** A schematic of the regulatory interactions associated with GNMT1. Where, dashed green represents allosteric activation, dashed red represents allosteric inhibition, and solid red represents competitive inhibition. **(Bottom Left)** Comparing the marginal posteriors of the sampled fluxes for datasets 1 and 2. **(Right)** Log ratios of the regulatory elements defined in section 3.2.4. These plots identify the reason why flux increases in dataset 2 compared to 1: the flux increase is due to allostery and saturation with no control from enzyme concentration changes.

Constructing a kinetic model, augmenting it with a statistical model, and then producing and investigating the results is not often a straightforward task, even using Maud. It is therefore important to motivate the use of such models by showing how they can be used to achieve specific scientific or engineering goals. In this section we show, in the context of our case study, how Maud can be used interrogate the regulation of the enzyme Glycine N-methyltransferase, or GMNT1 in a way that is relevant for strain engineering and drug targeting purposes. We do this by evaluating the regulatory ratio of GMNT1 between dataset 2 and 1, the regulatory interactions will inform how the flux dramatically increases in the second dataset.

Glycine N-methyltransferase is an irreversible enzyme that is regulated by two inhibitors and an activator. Between condition 1 and 2 there is a significant change in flux through the enzyme without a change in enzyme concentration. Figure 3.6 plots ratios between the two conditions of some of the flux components discussed in section 3.2.5, elucidating how the increase in flux is achieved. A value of zero suggests that there is no change in regulation between conditions, and the probability of regulation in a particular direction is given by the total area above or below zero. The primary

regulators appear to be both the allosteric and saturation coefficients, which the model thinks both up-regulate the flux. Furthermore, comparing the true simulated values to that of the inferred values we were able to encompass the true value well within the high-density region of the posterior. This shows that Maud not only predicts fluxes and metabolite concentrations, but can also resolve the regulatory behaviour of networks.

3.3.10 Conclusion

This study demonstrated, in a realistic kinetic modelling scenario, that Maud can fit a plausible dataset, successfully recover true parameter values and make accurate out of sample predictions, as shown in figure 3.3. Furthermore, Maud continues to make reasonable predictions when challenged with missing experimental data, as shown in figure 3.4. This is important due to difficulties in obtaining metabolomics for all metabolites for a given pathway.

As shown in figure 3.5, it is not possible to achieve comparable results using augmented maximum *a posteriori* estimation. Finally, we showed that Maud not only be used to generate accurate predictions, but can also yield causal insight that would not be available without the use of a mechanistic model, and as we show in section 3.4.1, could potentially be used to guide strain engineering, cellular understanding and even drug targeting approaches.

3.4 Discussion

3.4.1 Applications of kinetic models

Here we discuss how Maud can be used in three important biochemical engineering contexts, namely strain design, drug target selection and model selection.

Strain engineering   As discussed in section 3.2.4, Maud decomposes the flux through each enzyme-catalysed reaction into the product of regulatory components and the enzyme's k_{cat} parameter. Strain engineers can use the values of these regulatory coefficients in Maud's samples to determine how much capacity is remaining for each mode of regulation (Liebermeister et al., 2010; Hackett et al., 2016; Noor et al., 2013) and to suggest modes of action for further increases or decreases in flux. As an example, if the reversibility of a flux is close to zero, this suggests that the reaction is in equilibrium, in which case the enzyme is carrying a large amount of forward and reverse flux and little net flux. Therefore, an intervention that only increases the enzyme concentration will not dramatically increase flux and other engineering targets that improve the thermodynamic driving force should be considered first.

Drug target selection   Flux decompositions are also useful when targeting cells that drastically change metabolism such as cancer cells that exhibit the Warburg effect (Warburg, 1956; Warburg, 1924). Current research suggests that targeting metabolism (Wang et al., 2010; Kuo et al., 2000; Kroemer et al., 2018) may be a treatment option for cancer cells (Vander Heiden, 2011). Kinetic models that are *in silico* representations of

cancer metabolism can be used to screen metabolic targets for cancer treatment. *In silico* screening offers a tractable method for hypothesis testing and can be conducted orders of magnitude faster than *in vivo* testing.

Model selection   It is often important to take into account not only quantitative uncertainty as to the values of model parameters and state variables, but also qualitative uncertainty as to the nature of the underlying reaction network and interactions. These cases can be understood in a Bayesian statistical modelling context as model selection problems, with the models under consideration differing only in their metabolic models.

The canonical way to resolve a model selection problem, as discussed in Vehtari et al., 2017, is to score every model under consideration using an appropriate validation method. This approach is theoretically correct and can be achieved using Maud. However, when the number of models under consideration is high, or the validation method requires a large number of model fits (as, for example with exact leave-one-out cross-validation), the computational cost makes this solution infeasible. Methods that reduce the model space may make this methodology more tractable for future model selection work (Foster et al., 2022; Diether et al., 2019).

3.4.2   Limitations

Modelling approximations

Like all models, Maud makes simplifying assumptions that can lead to inaccurate inferences in certain circumstances. This section reviews these approximations and discusses when and how they might cause problems.

Steady state   Maud assumes that the metabolite concentrations and fluxes are measured at a steady state. This assumption limits the scope of potential data and problems to those that are at a steady-state, and prevents modelling oscillatory systems or dynamic experiments

There are no model assumptions or Bayesian principles preventing the inclusions of dynamic model fitting, apart from the practical implementation into Maud.

Channelling   Maud's metabolic model cannot accommodate the phenomenon of channelling, whereby an enzyme complex catalysing two separate reactions is formed and the intermediate metabolite does not dissociate into the bulk media. When channelling occurs, the concentration of the intermediate metabolite will be different in the enzyme—intermediate—enzyme complex compared with the intermediate—bulk media solution. Maud will therefore tend to assume the wrong metabolite concentration and will therefore model the thermodynamics of the channelled reaction incorrectly.

A workaround for using Maud to model networks involving enzyme channelling while avoiding incorrect thermodynamics is to represent only one reaction per complex, ig-

noring all intermediate metabolites.

Maud's prior model   Maud's prior model uses mostly independent normal and log-normal distributions to describe the non-omics information. This is inappropriate in circumstances where these distributions do not adequately summarise the available non-experimental information. Situations that may make this aspect of Maud's prior model inappropriate include:

- **Non-independence** Sometimes non-experimental information pertains to the sum, difference or product of two parameters. In order to represent this information a prior model is needed where the two parameters are not independent, which can be achieved either through re-parameterisation or through the use of a non-independent probability distributions such as the multivariate normal distribution. Maud allows multivariate normal priors for formation energy parameters but not for other non-independent parameters.

- **Multimodality** It can be known that a parameter's value probably lies outside a certain continuous range, but not on which side of the range. An appropriate marginal prior distribution for this kind of parameter will have modes on both sides of the range. For example, some instruments are only capable of measuring values inside a certain range; if no measurement occurs it can be reasonable to assume that the true value lies outside of that range. Maud allows only unimodal marginal prior distributions, so this kind of information cannot be represented correctly.

- **Likely outliers** Maud's prior model also cannot represent non-experimental information according to which a parameter's value is very likely to lie far away from its most likely value. In this case a heavy-tailed marginal prior distribution like the student-T distribution would be more appropriate than Maud's normal distributions.

A further problem with Maud's prior model is that it does not provide a way to represent prior information about state variables like fluxes and concentrations of balanced metabolite-compartment pairs. For example, it might be known in advance that the steady-state concentration of a certain metabolite is almost certainly below a particular value. While in principle it is possible to specify a prior model that assigns non-negligible probability mass only to parameter configurations that lead to the correct values, this is difficult to achieve in practice and Maud does not provide a way to easily choose such a prior specification. We recommend the use of prior push-forward checking in order to verify that the prior model's implications for state variables are broadly plausible, but precise calibration with prior information remains a challenge. See Betancourt, 2021 for discussion of prior push-forward checking.

The problem of prior information about likely outliers will be addressed in a future version of Maud by allowing users to choose heavy-tailed distributions instead of log-

normal distributions where appropriate. However we do not plan to allow further non-independent or multi-modal prior distributions, nor to implement automatic model calibration in order to allow explicit input of information about state variables. This is because we judge that the benefit of adding these features would not be worth the additional complexity in Maud's user interface and code. However, since Maud is fully open source and well documented and tested it would be possible for a capable user to add these features to their own fork if they proved indispensable for a particular application.

Maud's measurement model   Maud's measurement model uses a log-normal generalised linear model with known measurement error for measurements of quantities that are known to be non-negative, and a linear model with known measurement error for other measurements. This is inappropriate in cases where the measurement error is unknown, where it is known that outliers are likely and where the values of certain measurements are known to be correlated. The first problem could be addressed by changing measurement errors from user-provided data to model parameters. However this would entail extra complexity and impact model performance, and therefore is not currently planned. The problem of a known high probability of outliers could be addressed by allowing the user to specify alternative measurement error distributions. This is a relatively straightforward change and is planned for a future version of Maud.

The final problem—known relationships between measurements—is the most interesting. This can happen because a measurement pertains to a combination of latent quantities: for example a metabolomics measurement might not be able to distinguish between compartments, and will therefore correspond to the sum of the volume-weighted concentration of the metabolite in all compartments where it is present. This important special case is planned to be addressed in a future version of Maud. However there are further reasons why omics measurements can be related. One is that metabolomics and proteomics measurements typically involve separate processes to calculate the absolute and relative values. In some cases no attempt at absolute quantification is made, and only relative values are reported. This issue can potentially be addressed by extending Maud's measurement model so that the separate relative and absolute parts of the measurement process are represented, allowing a general bias to be taken into account as well as independent measurement-specific errors. However extending the model in this way requires careful consideration of the details of the measurement process to make sure that the new, more detailed model is accurate; as a result there are no current plans to modify Maud in this way. Finally, fluxomics measurements are typically inferred from metabolomics data, together with an atom model, stoichiometric model and split-ratio data from isotope labelling experiments, as exemplified by Young, 2014. Rather than treating flux measurements as independent as in the current version of Maud, it would clearly be preferable to expand Maud's measurement model so that split-ratio data could be modelled directly. This would

require users to input an atom model and would significantly increase Maud's complexity, but this may well be worthwhile in a future version.

Performance

Maud's performance is mainly limited by the need to solve the augmented steady state problem at every leapfrog step of Stan's sampler. Solving this problem entails finding the steady state balanced metabolite-compartment concentration vector corresponding to the current parameter configuration, as well as the gradients of this vector with respect to the parameters. Since many leapfrog steps are required in order to explore the complex posterior distributions typical in practice, a large number of augmented steady state problem solutions must typically be solved in order to obtain a single MCMC draw. The difficulty of approximately solving this problem increases with a greater number of state variables and parameters (Städter et al., 2021). Despite the fact that Maud uses state of the art sampling and equation solving algorithms, as well as a relatively efficient parameterisation, current computational resources practically limit inference to kinetic models with at most around 30 state variables. For context real metabolisms involve thousands of balanced metabolites and reconstructions usually include several hundred.

A number of strategies for improving Maud's performance, thereby allowing inference for larger models, are under active consideration, including better numerical methods, parallelisation, GPU use and re-parameterisation.

3.5   Conclusion

This paper presented Maud, a software package implementing Bayesian inference for kinetic models of cell metabolism, explaining the relevant context, demonstrating its use with a case study and comparing Maud with the state of the art. Maud can now be applied to the multiomics dataset collected in Chapter 2 and modelled using the kinetic model presented in Chapter 4. We have developed Maud as a tool for interpretable kinetic models and with the flux comparison presented in Section 3.2.5 we have the capability to answer how each enzyme is regulated.

References

Christodoulou, D., Link, H., Fuhrer, T., Kochanowski, K., Gerosa, L., & Sauer, U. (2018). Reserve Flux Capacity in the Pentose Phosphate Pathway Enables Escherichia coli's Rapid Response to Oxidative Stress. *Cell Systems*, *6*(5), 569–578.e7. https://doi.org/10.1016/j.cels.2018.04.009

DeBerardinis, R. J., & Chandel, N. S. (2016). Fundamentals of cancer metabolism. *Science Advances*, *2*(5), e1600200. https://doi.org/10.1126/sciadv.1600200

Liberti, M. V., Dai, Z., Wardell, S. E., Baccile, J. A., Liu, X., Gao, X., Baldi, R., Mehrmohamadi, M., Johnson, M. O., Madhukar, N. S., Shestov, A. A., Chio, I. I., Elemento, O., Rathmell, J. C., Schroeder, F. C., McDonnell, D. P., & Locasale, J. W. (2017). A Predictive Model for Selective Targeting of the Warburg Effect through GAPDH Inhibition with a Natural Product. *Cell Metabolism*, *26*(4), 648–659.e8. https://doi.org/10.1016/j.cmet.2017.08.017

Saa, P. A., & Nielsen, L. K. (2016). Construction of feasible and accurate kinetic models of metabolism: A Bayesian approach. *Scientific Reports*, *6*, 29635. https://doi.org/10.1038/srep29635

Gopalakrishnan, S., Dash, S., & Maranas, C. (2020). K-FIT: An accelerated kinetic parameterization algorithm using steady-state fluxomic data. *Metabolic Engineering*. https://doi.org/10.1016/j.ymben.2020.03.001

Gutenkunst, R. N., Waterfall, J. J., Casey, F. P., Brown, K. S., Myers, C. R., & Sethna, J. P. (2007). Universally sloppy parameter sensitivities in systems biology models. *PLoS Computational Biology*, *3*(10), 1871–1878. https://doi.org/10.1371/journal.pcbi.0030189

Raue, A., Kreutz, C., Theis, F. J., & Timmer, J. (2013). Joining forces of Bayesian and frequentist methodology: A study for inference in the presence of non-identifiability. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *371*(1984), 20110544. https://doi.org/10.1098/rsta.2011.0544

Liebermeister, W., Uhlendorf, J., & Klipp, E. (2010). Modular rate laws for enzymatic reactions: Thermodynamics, elasticities and implementation. *Bioinformatics*, *26*(12), 1528–1534. https://doi.org/10.1093/bioinformatics/btq141

Popova, S. V., & Sel'kov, E. E. (1975). Generalization of the model by Monod, Wyman and Changeux for the case of a reversible monosubtrate reaction SR,TP. *FEBS Letters*, *53*(3), 269–273. https://doi.org/10.1016/0014-5793(75)80034-2

Beber, M. E., Gollub, M. G., Mozaffari, D., Shebek, K. M., Flamholz, A. I., Milo, R., & Noor, E. (2022). eQuilibrator 3.0: A database solution for thermodynamic constant estimation. *Nucleic Acids Research*, *50*(D1), D603–D609. https://doi.org/10.1093/nar/gkab1106

Hoffman, M. D., & Gelman, A. (2011). The No-U-Turn Sampler: Adaptively Setting Path Lengths in Hamiltonian Monte Carlo. https://doi.org/10.48550/ARXIV.1111.4246

Carpenter, B., Gelman, A., Hoffman, M. D., Lee, D., Goodrich, B., Betancourt, M., Brubaker, M., Guo, J., Li, P., & Riddell, A. (2017). *Stan* : A Probabilistic Programming Language. *Journal of Statistical Software*, *76*(1). https://doi.org/10.18637/jss.v076.i01

Serban, R., & Hindmarsh, A. C. (2005). CVODES: The Sensitivity-Enabled ODE Solver in SUNDIALS. *Volume 6: 5th International Conference on Multibody Systems, Nonlinear Dynamics, and Control, Parts A, B, and C*, 257–269. https://doi.org/10.1115/DETC2005-85597

Casini, A., Chang, F.-Y., Eluere, R., King, A. M., Young, E. M., Dudley, Q. M., Karim, A., Pratt, K., Bristol, C., Forget, A., Ghodasara, A., Warden-Rothman, R., Gan, R., Cristofaro, A., Borujeni, A. E., Ryu, M.-H., Li, J., Kwon, Y.-C., Wang, H., … Gordon, D. B. (2018). A pressure test to make 10 molecules in 90 days: External evaluation of methods to engineer biology. *Journal of the American Chemical Society*, *140*(12), 4302–4316. https://doi.org/10.1021/jacs.7b13292

Chakrabarti, A., Miskovic, L., Soh, K. C., & Hatzimanikatis, V. (2013). Towards kinetic modeling of genome-scale metabolic networks without sacrificing stoichiometric, thermodynamic and physiological constraints. *Biotechnology Journal*, *8*(9), 1043–1057. https://doi.org/10.1002/biot.201300091

Saa, P., & Nielsen, L. K. (2015). A general framework for thermodynamically consistent parameterization and efficient sampling of enzymatic reactions. *PLoS Computational Biology*, *11*(4), e1004195. https://doi.org/10.1371/journal.pcbi.1004195

Tran, L. M., Rizk, M. L., & Liao, J. C. (2008). Ensemble modeling of metabolic networks. *Biophysical Journal*, *95*(12), 5606–5617. https://doi.org/10.1529/biophysj.108.135442

Choudhury, S., Moret, M., Salvy, P., Weilandt, D., Hatzimanikatis, V., & Miskovic, L. (2022). Reconstructing Kinetic Models for Dynamical Studies of Metabolism using Generative Adversarial Networks. *Nature Machine Intelligence*, *4*(8), 710–719. https://doi.org/10.1038/s42256-022-00519-y

Hackett, S. R., Zanotelli, V. R. T., Xu, W., Goya, J., Park, J. O., Perlman, D. H., Gibney, P. A., Botstein, D., Storey, J. D., & Rabinowitz, J. D. (2016). Systems-level analysis of mechanisms regulating yeast metabolic flux. *Science*, *354*(6311). https://doi.org/10.1126/science.aaf2786

Neal, R. M. (2011). *Handbook of Markov chain Monte Carlo* (S. Brooks, Ed.). CRC Press OCLC: 740896891.

Timonen, J., Siccha, N., Bales, B., Lähdesmäki, H., & Vehtari, A. (2022). An importance sampling approach for reliable and efficient inference in Bayesian ordinary differential equation models. https://doi.org/10.48550/ARXIV.2205.09059

Betancourt, M. (2017). A Conceptual Introduction to Hamiltonian Monte Carlo. https://doi.org/10.48550/ARXIV.1701.02434

St. John, P. C., Strutz, J., Broadbelt, L. J., Tyo, K. E. J., & Bomble, Y. J. (2019). Bayesian inference of metabolic kinetics from genome-scale multiomics data (C. D. Maranas,

Ed.). *PLOS Computational Biology*, *15*(11), e1007424. https://doi.org/10.1371/journal.pcbi.1007424

Liebermeister, W., & Noor, E. (2021). Model Balancing: A Search for In-Vivo Kinetic Constants and Consistent Metabolic States. *Metabolites*, *11*(11), 749. https://doi.org/10.3390/metabo11110749

Stapor, P., Weindl, D., Ballnus, B., Hug, S., Loos, C., Fiedler, A., Krause, S., Hroß, S., Fröhlich, F., Hasenauer, J., & Wren, J. (2018). PESTO: Parameter estimation toolbox. *Bioinformatics*, *34*(4), 705–707. https://doi.org/10.1093/bioinformatics/btx676

Vehtari, A., Gelman, A., Simpson, D., Carpenter, B., & Bürkner, P.-C. (2021). Rank-Normalization, Folding, and Localization: An Improved R^ for Assessing Convergence of MCMC (with Discussion). *Bayesian Analysis*, *16*(2). https://doi.org/10.1214/20-BA1221

Visser, D., & Heijnen, J. J. (2003). Dynamic simulation and metabolic re-design of a branched pathway using linlog kinetics. *Metabolic Engineering*, *5*(3), 164–176. https://doi.org/10.1016/s1096-7176(03)00025-9

Hoops, S., Sahle, S., Gauges, R., Lee, C., Pahle, J., Simus, N., Singhal, M., Xu, L., Mendes, P., & Kummer, U. (2006). COPASI—a COmplex PAthway SImulator. *Bioinformatics*, *22*(24), 3067–3074. https://doi.org/10.1093/bioinformatics/btl485

Preston-Werner, T., & Gedam, P. (2020, apr). *TOML Specification* (Version 1.0.0-Rc.1). https://toml.io/en/v1.0.0-rc.1/

Team, S. D. (2023a). *Stan Modeling Language Users Guide and Reference Manual* (Version 2.28). https://mc-stan.org

Kumar, R., Carroll, C., Hartikainen, A., & Martin, O. (2019). ArviZ a unified library for exploratory analysis of Bayesian models in Python. *Journal of Open Source Software*, *4*(33), 1143. https://doi.org/10.21105/joss.01143

Rew, R. K., & Davis, G. P. (1990, July). NetCDF: An Interface for Scientific Data Access. *IEEE Computer Graphics and Applications*, *10*(4), 76–82.

Noor, E., Flamholz, A., Liebermeister, W., Bar-Even, A., & Milo, R. (2013). A note on the kinetics of enzyme action: A decomposition that highlights thermodynamic effects. *FEBS Letters*, *587*(17), 2772–2777. https://doi.org/10.1016/j.febslet.2013.07.028

Alberty, R. A. (2003). *Thermodynamics of biochemical reactions*. Wiley-Interscience.

Mahamkali, V., McCubbin, T., Beber, M. E., Noor, E., Marcellin, E., & Nielsen, L. K. (2021). multiTFA: A Python package for multi-variate thermodynamics-based flux analysis (A. Valencia, Ed.). *Bioinformatics*, *37*(18), 3064–3066. https://doi.org/10.1093/bioinformatics/btab151

Monod, J., Wyman, J., & Changeux, J. P. (1965). On the nature of allosteric transitions: A plausible model. *Journal of Molecular Biology*, *12*, 88–118. https://doi.org/10.1016/S0022-2836(65)80285-6

Changeux, J.-P. (2013). 50 years of allosteric interactions: The twists and turns of the models. *Nature Reviews Molecular Cell Biology*, *14*(12), 819–829. https://doi.org/10.1038/nrm3695

Popova, S. V., & Sel'kov, E. E. (1979). [Description of the kinetics of the two substrate re-actions S1+S2 goes to and comes from S3+S4 by a generalized Monod, Wyman, Changeux model]. *Molekuliarnaia Biologiia, 13*(1), 129–139. Retrieved May 27, 2020, from https://www.ncbi.nlm.nih.gov/pubmed/156878

Matos, M. R. (2018). *Mass Action Stoichiometric. Simulation for Cell Factory Design* (THE-SIS.DOCTORAL). Novo Nordisk Foundation Center for BiosustainabilityResearch GroupsQuantitative Modeling of Cell MetabolismiLoop. Retrieved July 1, 2020, from https://orbit.dtu.dk/en/publications/mass-action-stoichiometric-simulation-for-cell-factory-design

Chang, A., Jeske, L., Ulbrich, S., Hofmann, J., Koblitz, J., Schomburg, I., Neumann-Schaal, M., Jahn, D., & Schomburg, D. (2021). BRENDA, the ELIXIR core data resource in 2021: New developments and updates. *Nucleic Acids Research, 49*(D1), D498–D508. https://doi.org/10.1093/nar/gkaa1025

Young, J. D. (2014). INCA: A computational platform for isotopically non-stationary metabolic flux analysis. *Bioinformatics (Oxford, England), 30*(9), 1333–1335. https://doi.org/10.1093/bioinformatics/btu015

Talts, S., Betancourt, M., Simpson, D., Vehtari, A., & Gelman, A. (2018). Validating Bayesian Inference Algorithms with Simulation-Based Calibration. https://doi.org/10.48550/ARXIV.1804.06788

Xing, Z., Bishop, N., Leister, K., & Li, Z. J. (2010). Modeling kinetics of a large-scale fed-batch CHO cell culture by Markov chain Monte Carlo method. *Biotechnology Progress*, NA–NA. https://doi.org/10.1002/btpr.284

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., & Rubin, D. B. (2013, November 27). *Bayesian Data Analysis* (0th ed.). Chapman and Hall/CRC. https://doi.org/10.1201/b16018

Team, S. D. (2023b). *CmdStanPy*. https://github.com/stan-dev/cmdstanpy

Rossum, G. van, & Drake, F. L. (2010). *The Python language reference* (Release 3.0.1 [Repr.]). Python Software Foundation.

Developers, C. (2022). *Click: {{Python}} Composable Command Line Interface Toolkit*. https://pypi.org/project/click/

Harris, C. R., Millman, K. J., Van Der Walt, S. J., Gommers, R., Virtanen, P., Cournapeau, D., Wieser, E., Taylor, J., Berg, S., Smith, N. J., Kern, R., Picus, M., Hoyer, S., Van Kerkwijk, M. H., Brett, M., Haldane, A., Del Río, J. F., Wiebe, M., Peterson, P., … Oliphant, T. E. (2020). Array programming with NumPy. *Nature, 585*(7825), 357–362. https://doi.org/10.1038/s41586-020-2649-2

Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., Van Der Walt, S. J., Brett, M., Wilson, J., Millman, K. J., Mayorov, N., Nelson, A. R. J., Jones, E., Kern, R., Larson, E., … Vázquez-Baeza, Y. (2020). SciPy 1.0: Fundamental algorithms for scientific computing in Python. *Nature Methods, 17*(3), 261–272. https://doi.org/10.1038/s41592-019-0686-2

Meurer, A., Smith, C. P., Paprocki, M., Čertík, O., Kirpichev, S. B., Rocklin, M., Kumar, Am., Ivanov, S., Moore, J. K., Singh, S., Rathnayake, T., Vig, S., Granger, B. E., Muller, R. P., Bonazzi, F., Gupta, H., Vats, S., Johansson, F., Pedregosa, F., … Scopatz, A. (2017). SymPy: Symbolic computing in Python. *PeerJ Computer Science*, *3*, e103. https://doi.org/10.7717/peerj-cs.103

Team, T. P. D. (2023, April 24). *Pandas-dev/pandas: Pandas* (Version v2.0.1). https://doi.org/10.5281/ZENODO.3509134

developers, P. (2022). *Pydantic*. https://pypi.org/project/pydantic/

Team, S. D. (2023c). *CmdStan*. https://github.com/stan-dev/cmdstanpy

developers, I. (2022). *Isort*. https://pypi.org/project/isort/

developers, B. (2022). *Black*. https://pypi.org/project/black/

flake8 developers. (2022). *Flake8*. https://pypi.org/project/flake8/

flake8-docstrings developers. (2022). *Flake8-Docstrings*. https://pypi.org/project/flake8-docstrings/

flake8-bugbear developers. (2022). *Flake8-Bugbear*. https://pypi.org/project/flake8-bugbear/

safety developers. (2022). *Safety*. https://pypi.org/project/safety/

pytest developers. (2022). *Pytest*. https://pypi.org/project/pytest/

tox developers. (2022). *Tox*. https://pypi.org/project/tox/

reStructuredText developers. (2001). *{{reStructuredText}}, the Easy to Read, Easy to Use, What-You-See-Is-What-You-Get Plaintext Markup Language.* https://docutils.sourceforge.io/rst.html#reference-documentation

Georg, B., & team, S. (2022). *Sphinx*. https://www.sphinx-doc.org/

Korendyaseva, T. K., Kuvatov, D. N., Volkov, V. A., Martinov, M. V., Vitvitsky, V. M., Banerjee, R., & Ataullakhanov, F. I. (2008). An Allosteric Mechanism for Switching between Parallel Tracks in Mammalian Sulfur Metabolism. *PLOS Computational Biology*, *4*(5), e1000076. https://doi.org/10.1371/journal.pcbi.1000076

Poirier, D. J. (1998). REVISING BELIEFS IN NONIDENTIFIED MODELS. *Econometric Theory*, *14*(4), 483–509. https://doi.org/10.1017/S0266466698144043

Warburg, O. (1956). On the Origin of Cancer Cells. *Science*, *123*(3191), 309–314. https://doi.org/10.1126/science.123.3191.309

Warburg, O. (1924). �ber den Stoffwechsel der Carcinomzelle. *Die Naturwissenschaften*, *12*(50), 1131–1137. https://doi.org/10.1007/BF01504608

Wang, J.-B., Erickson, J. W., Fuji, R., Ramachandran, S., Gao, P., Dinavahi, R., Wilson, K. F., Ambrosio, A. L., Dias, S. M., Dang, C. V., & Cerione, R. A. (2010). Targeting Mitochondrial Glutaminase Activity Inhibits Oncogenic Transformation. *Cancer Cell*, *18*(3), 207–219. https://doi.org/10.1016/j.ccr.2010.08.009

Kuo, W.-y., Lin, J.-y., & Tang, T. K. (2000). Human glucose-6-phosphate dehydrogenase (G6PD) gene transforms NIH 3T3 cells and induces tumors in nude mice. *International Journal of Cancer*, *85*(6), 857–864. https://doi.org/10.1002/(SICI)1097-0215(20000315)85:6<857::AID-IJC20>3.0.CO;2-U

Kroemer, G., López-Otín, C., Madeo, F., & de Cabo, R. (2018). Carbotoxicity—Noxious Effects of Carbohydrates. *Cell*, *175*(3), 605–614. https://doi.org/10.1016/j.cell. 2018.07.044

Vander Heiden, M. G. (2011). Targeting cancer metabolism: A therapeutic window opens. *Nature Reviews Drug Discovery*, *10*(9), 671–684. https://doi.org/10.1038/nrd3504

Vehtari, A., Gelman, A., & Gabry, J. (2017). Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. *Statistics and Computing*, *27*(5), 1413–1432. https://doi.org/10.1007/s11222-016-9696-4

Foster, C., Boorla, V. S., Dash, S., Gopalakrishnan, S., Jacobson, T. B., Olson, D. G., Amador-Noguez, D., Lynd, L. R., & Maranas, C. D. (2022). Assessing the impact of substrate-level enzyme regulations limiting ethanol titer in Clostridium thermocellum using a core kinetic model. *Metabolic Engineering, 69*, 286–301. https://doi.org/ 10.1016/j.ymben.2021.12.012

Diether, M., Nikolaev, Y., Allain, F. H., & Sauer, U. (2019). Systematic mapping of protein-metabolite interactions in central metabolism of Escherichia coli. *Molecular Systems Biology*, *15*(8), e9008. https://doi.org/10.15252/msb.20199008

Betancourt, M. (2021). *Prior {{Modeling}}*. https://github.com/betanalpha/knitr_case_ studies/tree/master/prior_modeling

Städter, P., Schälte, Y., Schmiester, L., Hasenauer, J., & Stapor, P. L. (2021). Benchmarking of numerical integration methods for ODE models of biological systems. *Scientific Reports*, *11*(1), 2696. https://doi.org/10.1038/s41598-021-82196-2

4   Exploring the Warburg Effect

4.1   Introduction

The question we posed at the beginning of this thesis is the following: *How can glycolysis regulate the Warburg effect?* To develop a foundation to answer this question; we have 1) collected and presented data of a contrasting phenotype between CHO-S epitomising the high growth and high lactate secretion of the Warburg effect, and CHO-ZeLa a cell line engineered to not produce lactate yet maintain the high growth rate of CHO-S, and 2) developed a modelling toolbox that fits kinetic parameters using information contained both in multiomics experiments and in the vast wealth of kinetic data. This chapter summarises our efforts at understanding the regulatory mechanisms of the Warburg effect using a kinetic model.

We will detail the construction of our kinetic model, evaluate the model fit, and perform a comparison between the regulation of glycolysis in the CHO-S cell line and the CHO-ZeLa cell line. Performing this comparison on a per enzyme basis we can make statements about the main mechanisms by which the high flux state in CHO-S is achieved. These results will answer our overarching question of how glycolysis is regulated in the Warburg effect.

There are models of the Warburg effect in mammalian cells that already exist, and even a model applicable to CHO cells. We first explored what is already presented in the literature, what have been identified as key regulatory components, and what we considered differently in our models.

4.1.1   Previous models of CHO metabolism

Compared with other biological phenomena, there are relatively few published kinetic models of CHO glycolysis or more generally the Warburg effect (Shestov et al., 2014; Mulukutla et al., 2014; Mulukutla et al., 2015; O'Brien et al., 2019; Mulquiney and Kuchel, 1999; Hagrot et al., 2019). The kinetics surrounding the Warburg effect are intrinsically complicated because of the number of isoenzymes involved in this pathway, the high degree of interconnectivity, and the number of regulatory interactions at the phosphofructokinase (PFK) node. This makes a general model of glycolysis among all mammalian cells infeasible. And furthermore, a definitive explanation of how the Warburg effect regulates flux difficult to answer.

The most prominent kinetic models of the Warburg effect in CHO cells were published in 2014, prominent models being the set of models developed by Mulukutla et al., 2014 (Mulukutla et al., 2014; Mulukutla et al., 2015; O'Brien et al., 2019). Originally developed for HeLa cells, this was the first kinetic model to demonstrate the high and low flux modes through glycolysis that characterise the Warburg effect. This model

achieved this via the two control loops surrounding the phosphofructokinase (PFK) and Pyruvate Kinase (PKM) reactions.

A limitation of the model by Mulukutla et al., 2014 when applied to the current CHO cell lines is that it does not account for different isoenzyme expression. As an example, their model in predominantly considers the pyruvate kinase M2 (PKM2) isoenzyme which it supposes is allosterically activated by the metabolites f16bp and pep. However, expression data from CHO cells suggests that the pyruvate kinase M1 (PKM1) isoform is an order of magnitude higher in concentration than pyruvate kinase M2 (PKM2). PKM1 and PKM2 are both tetramers, however, PKM2 is an allosterically mediated enzyme that exhibits low activity without the allosteric activators; PKM1 however, is constitutively active and does not undergo allosteric changes. Therefore, if we are to construct a model that describes glycolysis of CHO cells, we must consider both the expression of PKM1 and PKM2.

Mulukutla et al., 2014 showed that if there was no feedforward activation of PKM2, which is analogous to having only PKM1, the EMP pathway would achieve a high flux state. In their model, low flux states could still be achieved, however, this was only possible in the low glucose regimes. This is in contrast with the experimental data where we observed both low and high flux states between CHO-ZeLa and CHO-S wt and it can be assumed that both strains would have glycolysis modelled according to the Mulukutla et al., 2014 as a high flux regime, considering the similar growth rates and glucose concentrations in the media. The model described in Mulukutla et al., 2014 was created to model the bistability of HeLa glucose uptake rates and succeeded in reproducing this phenomenon. However, given the new information provided by the multiomics of apparent low and high flux modes unrelated to glucose concentration or expression, it is clear that a revised kinetic model of the Warburg effect in mammalian cells is needed.

4.1.2   Defining the scope of our model

In our effort to understand how CHO cells attain high and low flux states our model must encompass the both the EMP and PP pathways. It is not necessary to include the TCA cycle to decompose the reactions into their regulatory components. The model developed in this work has the objective of determining the regulation mediating each enzyme in two specific strains. If the objective was to construct a predictive model for any enzyme concentration then we would be required to include the TCA cycle in our model. This is because the flux through the EMP and PP pathways influences the flux in the TCA cycle that contains metabolites that have feedback regulation of the EMP and PP pathways, such as citrate (cit). For our purposes it is sufficient to model these feedback metabolites as external boundary conditions that constant throughout the simulation. This limits the predictive capabilities of our model, and to a degree is less constrained than if we were to include the TCA cycle, however, the large increase in simulation time makes including the TCA cycle impractical with little detriment to the

fitting result.

A detailed description of each regulator of glycolytic flux is provided in Appendix C with associated prior information. However, a summary is provided in Figure 4.1. Here we present the considered flux map as well as the allosteric and competitive regulation considered. A brief description of the hexokinase 2 regulatory node and the fructose-2,6-bisphosphate (f26bp) node are provided below, these are highlighted due to their perceived importance and reliance on auxiliary process that are not considered in this model.
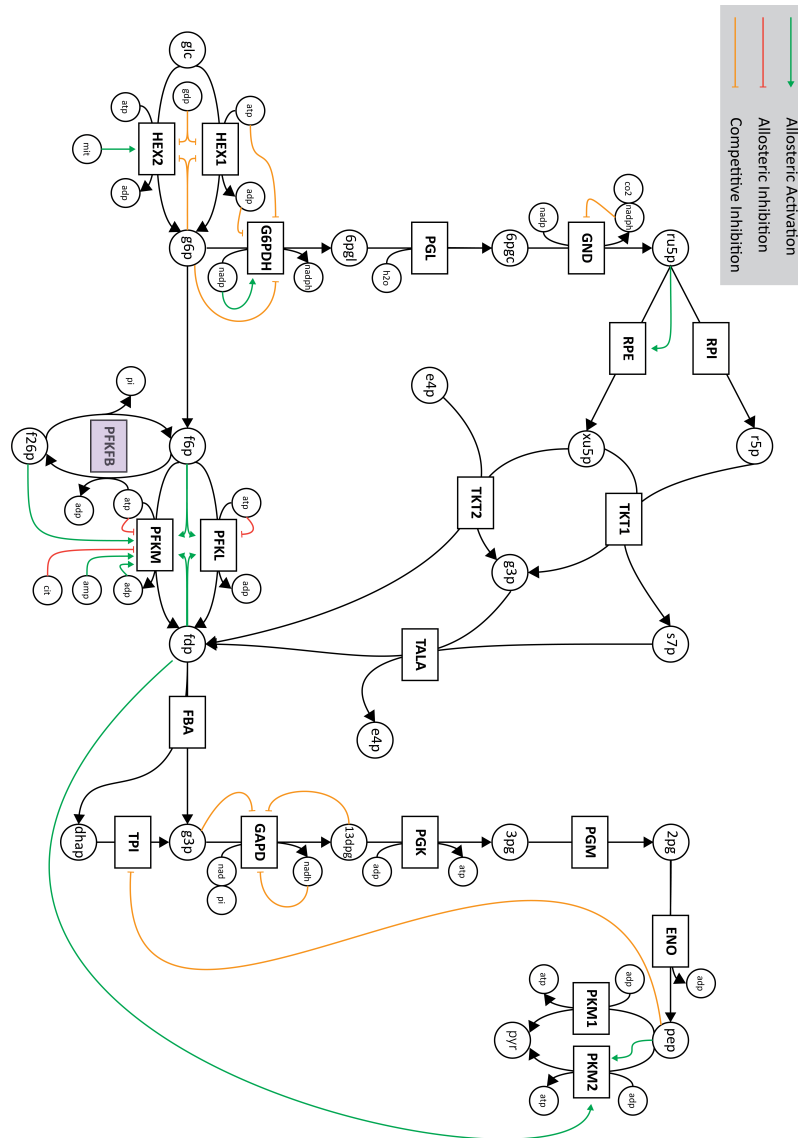
Figure 4.1: Reaction and regulatory network modelled. The green arrows indicate allosteric activation of the source metabolite to the target reaction. The red arrows indicate allosteric inhibition of the source metabolite of the target reaction. Finally, the orange arrows denote competitive inhibition by the source metabolites on the target reaction. This map encompasses the EMP and PP pathways and is entirely located in the cytosol. PFKFB is shown in purple, and it not included in the kinetic model, but was included in this graphic to illustrate the high degree of regulation within glycolysis

Regulation of HEX2

The first committal step in glycolysis is the essentially irreversible phosphorylation of glucose into glucose-6-phosphate by the enzyme hexokinase. Since hexokinase 2 (HEX2) operates far from equilibrium, meaning that most of the substrate converted to product remains as production, control cannot be mediated by the thermodynamic driving force and only through the following mechanisms: expression, saturation, or allostery. The expression of HEX2 is unlikely to be a major regulator of glycolysis as the enzyme expression is not observed to be correlated with flux as seen in Chapter 2. Regulation via saturation can be achieved by increasing glucose concentration or by increasing the concentration of the competitive inhibitors glucose-6-phosphate (g6p) and glucose-1,6-diphosphate (gdp) (Ahn et al., 2009; Beitner, 1979). Regulation by varied glucose concentration is unlikely due to the high concentration of glucose in the media and high growth rate. A study of the permeability of glucose in red blood cells by Britton, 1964 found that the external to internal glucose transport was nearly equal and would suggest being close to equilibrium. Finally, Bustamante and Pedersen, 1977 found that hexokinase activity was distributed around rat hepatoma cells with the majority being associated with the mitochondria Bustamante and Pedersen, 1977 with a 3-fold higher flux compared to non-mitochondrial bound regions. It has been shown that HEX2 couples to the production of atp by ATPase through the association with voltage dependent anion channels (VDACs) following phosphorylation from serine/thrionine kinase (Akt) (Gottlob et al., 2001; Heiden et al., 1999).

To model the regulation of HEX2, we will consider the standard competitive inhibition by g6p and gdp. In addition to competitive inhibtion we will model the regulation by VDACs by considering the mitochondria as an allosteric activator of HEX2. The quantitiy of interest is the allosteric component in equation 3.2 that can vary between 0 and 1. Whilst not mechanistic in its implementation, allosteric activation will be achieved via the inclusion of a metabolite *mit*, referring to mitochondrial binding.

Regulation by fructose-2,6-bisphosphate

Fructose-2,6-bisphosphate (f26bp) is produced when the enzyme phosphofructokinase 2 (PFK2) phosphorylates fructose-6-phosphate. f26bp is a potent allosteric activator of the enzyme phosphofructokinase M (PFKM) and is hypothesised to act as a sensory metabolite as is suggested to be a key regulator of glycolysis (Telang et al., 2006; Zancan et al., 2007). Regulation of f26bp steady state concentration is achieved through a balance of PFK2 and fructose-2,6-bisphosphatase (FBAse) activtity that simultaneously phosophrylate f6p and dephosphorylate f26bp. This process is achieved via the bifunctional enzyme phosphofructokinase 2/fructose-2,6-bisphosphatase (PFKFB). There are four different isoenzymes, being PFKFB 1-4, with the most dominant expression in CHO-S identified by rna expression being PFKFB3 (Singh et al., 2018). PFKFB3 is inhibited by pep and activated by AMP-activate protein kinase (AMPK) (Manes and El-Maghrabi, 2005). More specifically, AMPK phosphorylation was shown to prevent allosteric inhibition by pep on the kinase domain of PFKFB3 (Manes and El-Maghrabi,

2005).

To address regulation of PFKM, one of the key enzymes in glycolysis, we must include f26bp as a parameter in the model. This is more difficult than posed however, as neither f26bp or any PFKFB isoenzyme was able to be measured. Rather than including the PFKFB3 reaction cycle, we will consider f26bp as an unbalanced boundary metabolite. Not including the PFKFB3 reactions in the model will have no impact on the mass balance as the metabolism is assumed to be at steady state and PFKFB is a futile cycle (see Figure 4.1). Assuming f26bp is an unbalanced metabolite again limits the predictive power of this model, and will negatively impact the validation results. However, the lack of measurements surrounding this node obscure any current attempt at fitting an accurate model of f26bp metabolism.

4.1.3 Overview of paper

These mechanisms along with the other regulation described in Figure 4.1 were modelled using Maud, the framework developed in Chapter 3. We then compared the CHO-S and CHO-ZeLa cell lines using the fitted model to determine how regulation can be achieved through glycolysis. This study is unique as it is the first study to compare a Warburg expressing cell line with one that is undergoing high growth but no lactate production and also because it was the first study to emphasise the importance of HEX2 control over glycolysis. This was made possible due to the multiomics data we collected in Chapter 2.

We fitted the model on the CHO-S, CHO-ZeLa, CHO-Zen and CHO-ZenZeLa data and performed validation on the CHO-S 11 KO and ExpiCHO strains. This inspected the fitting capabilities and if our model was able to generalise to conditions outside of the defined training set. Afterwards we then used the model to determine regulation along glycolysis.

We used CHO-S as the cell line that epitomises the Warburg effect and compared it to the CHO-ZeLa cell line that does exhibit the Warburg effect. These states were compared using the kinetic model with the regulatory comparison as shown in Section 3.2.5. The result of this comparison stated the mechanism of how flux was increased for every enzyme between the CHO-S and CHO-ZeLa cell lines. with more positive values indicating an increase of that mechanism, whilst negative values indicating a decrease. This overview provided a detailed understanding of the allosteric, thermodynamic, enzymatic, and saturation changes observed and will move us a step close to understanding the root regulatory mechanisms that effect glycolysis.

4.2 Results

The results Section is broken into two parts. The first is an evaluation of the model fit, determining how well we were able to capture the underlying process. Secondly, we use the trained kinetic model to investigate the regulation of the Warburg effect by comparing the CHO-S to the CHO-ZeLa cells.

To demonstrate that our model was able to capture the underlying process, we first conducted a prior check using the fitted parameters from the trained model and left out validation datasets. We showed that the model was able to capture the fluxes and metabolite concentration trends of the left out datasets.

We then approached understanding how CHO-S increases the flux through the EMP pathway by applying the method used in 3.2.5 in which the two reference states are comparing the flux increase of CHO-S to that of CHO-ZeLa. The results elucidate the most important regulatory components of how the Warburg effect by using the regulatory coefficients of each enzyme.

4.2.1   Model fit

Figure 4.2 indicates that the trained datasets showed a reasonable agreement with the measured values with 75% of the fluxes and 81% of metabolite concentrations falling within $\pm$ 20% of the measured values (Figure 4.2). The largest differences from the measured values were the CHO-S and CHO-ZeLa oxPPP fluxes, which were under- and over-estimated respectively. Despite this, the flux was still significantly higher in the CHO-S cell lines and within the 95% confidence interval determined by 13C-MFA (Chapter 2). This suggests that the optimal parameter set was able to sufficiently capture the fluxes observed in the training dataset.

The largest difference in metabolite concentrations was observed primarily around the glyceraldehyde 3-phosphate (g3p) and dhap metabolites (Figure 4.2), suggesting that the model was unable to capture the behaviour around TPI. Although these metabolites were measured, the measurements were discarded as the model could not accommodate them, leading to computation errors. These difficulties indicate either that the metabolite measurements were biased or that the model's description of TPI was incorrect. Future analyses should therefore prioritise correctly describing TPI and ensuring that g3p and dhap are measured accurately. Difficulties in model fit are indicative of an incorrect mechanism and suggests a reconsideration of the TPI mechanism.

Similar to the training dataset shown in Figure 4.2, the oxidative PP pathway was poorly predicted in the validation dataset with estimates showing a lower flux compared to the measurement. Consequently, all of the metabolites part of the non-oxidative PP (noxPP) pathway were under-predicted due to the reduced thermodynamic driving force as a result of the lower flux. However, the validation dataset suggested that the mean parameter estimate was adequate at fitting the EMP pathway. HEX fluxes in both the ExpiCHO and CHO-S 11 KO datasets were contained within a weighted error of ± 60 % of the measured value.

Overall, the predictions were fairly accurate, and we concluded that the model results would likely be instructive as to the real mechanisms. The PPP predictions and therefore inference are likely biased. This is indicated by the consistent under-prediction in flux in the ExpiCHO cell line, and the metabolite concentrations not being predicted
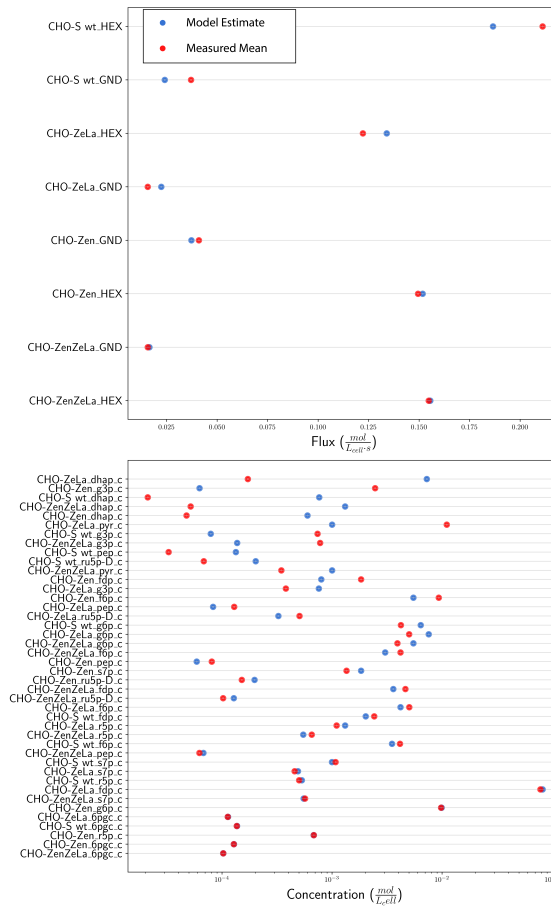
Figure 4.2: Comparison of the modelled fluxes and metabolite concentrations with the mean of the measured values for the training dataset. The modelled values represent the mode of the posterior density. The fluxes are on the linear scale and the metabolites are on the natural logarithmic scale. The fluxes and metabolite concentrations are ordered top to bottom by largest difference between the measured and modelled values.

around their measured values. This is in contrast to the EMP pathway, whereby the fluxes and metabolites were contained within the prior estimates.

4.2.2   Interpreting the regulation within glycolysis

In this section we are going to be presenting the results comparing CHO glycolysis in CHO-ZeLa and CHO-S cells using the regulatory decomposition between the two cell lines, with CHO-ZeLa being the reference state, as described using Equation 3.6. Figure 4.4 shows each isoenzyme with the corresponding terms for the regulatory decomposition of enzymatic, thermodynamic, allosteric, and saturation control. Each of these terms indicate how glycolysis is regulated in the Warburg effect, and the how flux is able to achieve the lower flux state in the CHO-ZeLa cells.

The noxPP pathway is a near equilibrium pathway
Unlike the majority of glycolysis, the noxPP pathway is driven by an increase in the thermodynamic driving force (Figure 4.4). The fitted model suggests that the non-oxidative PPP in CHO-S cells has a flux that is 7.4% higher than in CHO-ZeLa cells (Fig-
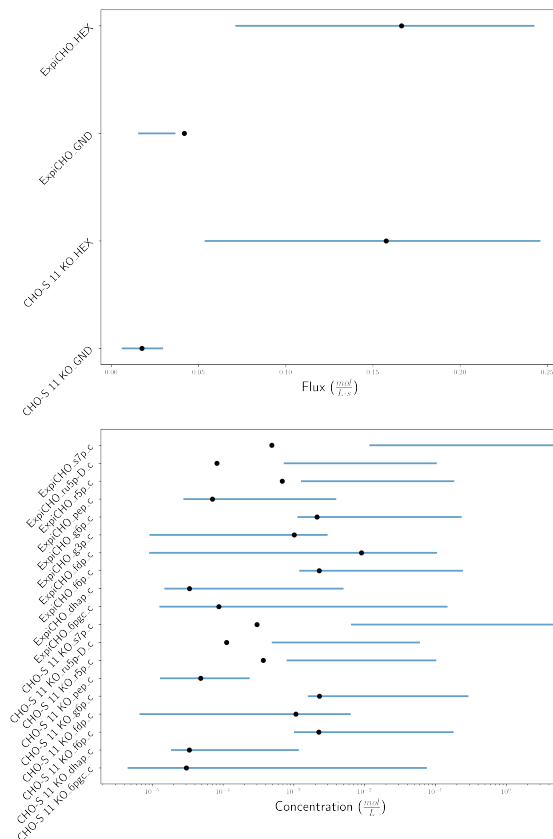
Figure 4.3: The blue forest plot represents the 95% prior predictive interval of the fluxes and metabolite concentrations using the method as defined in Section 4.3.3. The means of the measured values are represented by the black points, and the 95% sampling intervals are the blue bars. Flux plots are on the normal scale and the metabolite concentrations are on the base-10 logarithmic scale.

ure 4.2). Despite the increase in flux there was a decrease in enzyme expression across the pathway and general decrease in saturation (Figure 4.4). However, the noxPPP is able to achieve the modelled flux increase through changing the thermodynamic driving force of the pathway.

The thermodynamic driving force is a direct measure of enzyme efficiency. At equilibrium, more correctly referred to as dynamic equilibrium in this instance. An enzyme would operate both in the forward and reverse directions equally as fast. Therefore, no net flux would be observed despite the enzyme operating in both the forward and reverse directions. As the thermodynamic force is increased in the forward direction, less of the reverse reaction occurs relative to the forward, and the net flux increases. Figure 4.4 shows that the saturation of the enzyme can decrease, meaning there is less substrate available, however, the increase in thermodynamics allows the enzyme to operate more efficiently and at a higher flux.

There are two benefits of operating close to equilibrium. Firstly, the pathway does not require fine tuned regulation of enzyme expression, as each enzyme in the pathway
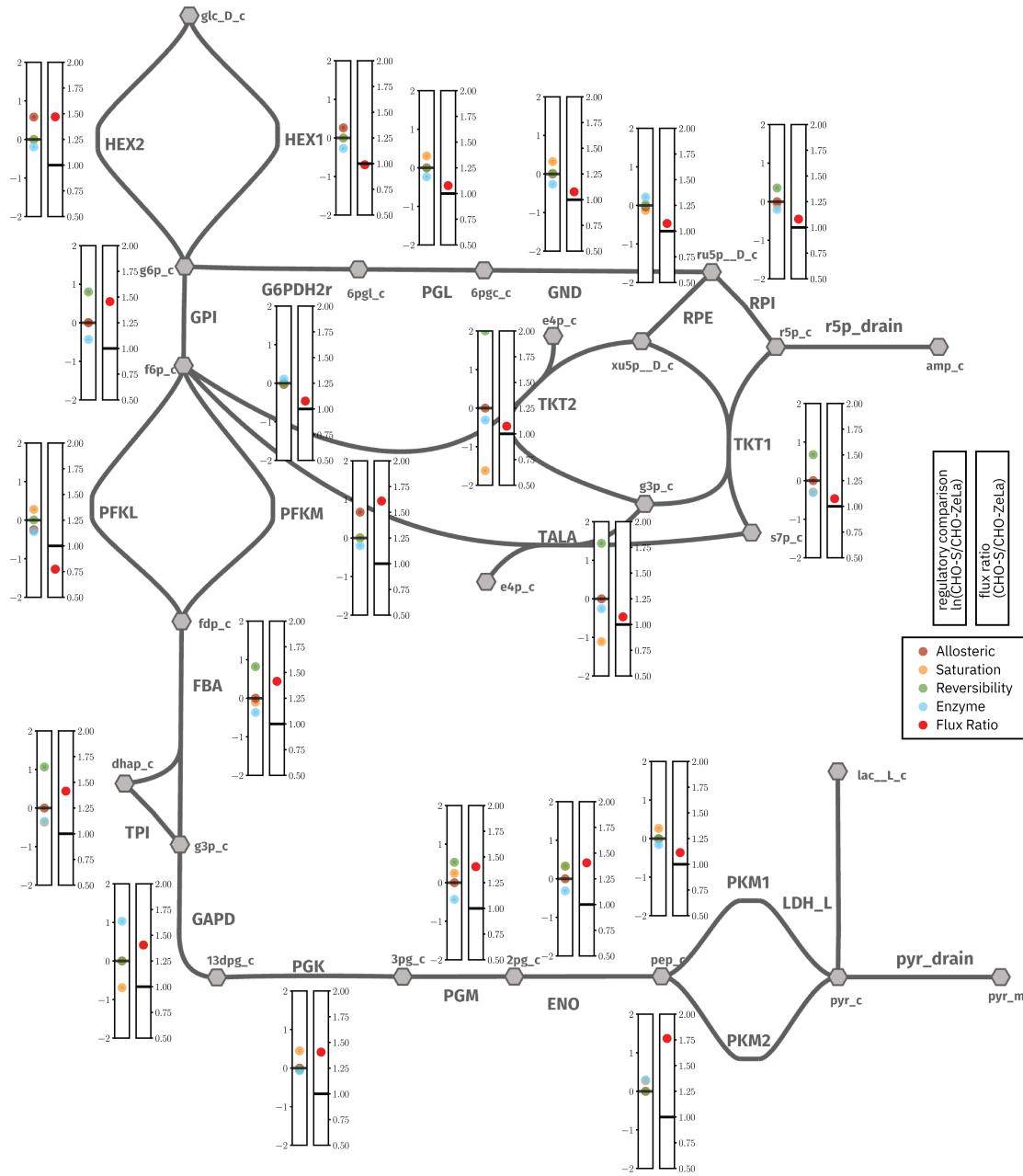
Figure 4.4: Regulatory comparison of glycolysis between the CHO-S and CHO-ZeLa cell lines that were estimated by the mode of the posterior. The metabolic map is generated using the Shu software package Muriel et al., 2023. For the nearest enzyme there is presented the regulatory decomposition of using the formula defined in Section 3.2.5, that is, the log ratios of the allosteric (maroon), saturation (yellow), reversibilty (green), and enzyme (green) decomposition the CHO-S being the numerator and CHO-ZeLa the denominator. To the right is the flux ratio ($\frac{v_{CHO-S}}{v_{CHO-ZeLa}}$ as suggested by the optimised parameters of the kinetic model.

can increase its flux by operating more efficiently. And secondly, as the transketolases and transaldolase operate close to equilibrium (Figure 4.4, the cell is more robust to changes in biomass demand. This makes it possible to switch directions from sourcing

carbon for r5p from the oxPPP, to utilising the EMP intermediates g3p and f6p. This would not be possible if the pathway was operating far from equilibrium as the changes in metabolite concentration are expected to be too large to operate within a typically homeostatic cell.

The Warburg effect requires allosteric activation of PFK and HEX

Across the model suggest that within the EMP pathway there is a decrease in expression from the CHO-ZeLa to the CHO-S strains in 11 of the 13 enzymes expressed (Figure 4.4). Nonetheless, the CHO-S strain was able to drive a higher flux through the EMP pathway: according to the kinetic model, this was possible as a result of the increased allosteric activation of PFKM and HEX2 (Figure 4.4). Additionally, the isoenzymes phosphofructokinase type L (PFKL) and hexokinase 1 (HEX1) carry a lower flux in the CHO-ZeLa phenotype due to the decreased enzyme expression.

The increase in flux of PFK was achieved by decreasing the citrate inhibition of PFKM rather than activation by f26bp. This was identified as the concentration of the allosteric activator f26bp was similar between CHO-S and CHO-ZeLa, but significantly decreased in the CHO-S state (Table E.5). However, regulating PFKM via citrate is not the only mechanism, as the model suggest that 24% of PFKM is in the active state in the CHO-ZeLa cell line E.2 there is still significant potential for allosteric regulation via activation from f26bp. The current state is a result of the model being unable to distinguish between two potential regulatory mechanisms, and further experiments are required to resolve this ambiguity.

The regulation network surrounding HEX2 is less ambiguous than PFKM, as HEX2 has the sole allosteric regulator of mitochondrial binding, see Figure 4.4. The model therefore suggests that allosteric activation of HEX2 is a requirement for the increase in flux through glycolysis. This is supported by experiments conducted by Bustamante and Pedersen, 1977, who's experiments showed that hexokinase associated with the mitochondria had a higher activity than cytosolic hexokinase. It is noted that the psudeometabolite mit, and allosteric binding are not mechanistic, but solely a proxy to model mitochondrial binding.

An unexpected result was the dramatic increase in flux through the PYK2 isoenzyme according to the model in the high flux state (Figure 4.4). According to the simulation this was not achieved through allosteric activation but by an increase in enzyme expression and an increase in saturation. Allosteric activation between the states did not change, this is because the dissociation constant of the activator fdp is significantly low relative to the concentration of the fitted fdp concentration and the enzyme was always active at physiological concentrations. The measurement of pep and fdp in the CHO-S condition was measured inaccurately, see Figure 4.2, and because of this the model has an increased freedom to fit the data.

The EMP pathway is dependent on GAPD concentration

To achieve the flux increase through the EMP pathway in the CHO-S cell line, the model had to assume that the concentration of the enzyme GAPD was much higher than its measured value. As a consequence, a wider prior had to be used; otherwise, the model would not be able to reproduce the high flux state characteristic of the observed strain. A similar finding was discovered in Shestov et al., 2014, where the authors identified GAPD as the limiting enzyme using metabolic control analysis (MCA).

4.3   Discussion

This is the first study that attempts to elucidate the regulatory mechanisms of the Warburg effect in CHO cells with a multiomics dataset including integrated fluxomics, metabolomics and proteomics data. We have shown that allosteric regulation on phosphofructokinase and hexokinase is required in order to attain the high and low flux states of the EMP pathway. Furthermore, 11 of the 13 enzymes in the EMP pathway were predicted by the model to not increase in concentration to achieve the high flux state observed in CHO-S cell. This aligns with previous experiments that were unable to identify differentially expressed proteins between high and low flux states (Buchsteiner et al., 2018). Our model validation of strains not contained within our training set highlights the model's ability to estimate EMP behaviour and indicates that further work is required to model the PPP.

For example, the hypothesis that the measurements are incorrect because the assumed mechanism of TPI is wrong, rather than GAPDH and fructose bisaldolase (FBA), is supported by the following reasoning. Firstly, there is an increase in the measurement substrate of TPI, dhap, in the CHO-ZeLa cell line and a decrease in the product concentration, g3p (Chapter 2). This increases both the saturation and thermodynamic driving forces of the TPI reaction. Secondly, the enzyme concentrations are similar based on measurements, eliminating enzyme expression regulation as a method of control. Considering that enzymatic, thermodynamic, and saturation effects are higher in the CHO-ZeLa cell line whilst simultaneously having a lower flux, the current mechanism will be unable to capture this change in flux. Therefore, the mechanism of TPI is not fully understood and requires further investigation.

In this work we identified both regulatory requirements and areas where fundamental understanding of reaction mechanisms are lacking. For example, in both conditions the model suggests that PYK2 was fully activated, and much like GAPD, the model compensated by increasing the enzyme concentration. The increase in enzyme concentration was in contrast to the multiomics measurements in Chapter 2 suggesting that allosteric regulation of GAPD and PYK2 is a requirement for the Warburg effect. This insight was assisted by the the decomposition of rate laws adapted from Liebermeister et al., 2010 and Noor et al., 2013, as it would not be possible to identify the lack of allosteric control otherwise.

The model was designed to explain how each enzyme regulates flux in CHO cells. With few exceptions we were able to model these changes consistently with experimental data and with the assumed model. The following points limit our inference and require either further modelling or experiments: boundary metabolite concentrations surrounding the PFK and hexokinase reactions are unmeasured and unbalanced; ordered mechanisms are unaccounted for; and, additional situations for phosphorylation of HEX2 are required.

4.3.1 Improvements required

In order to determine the relative contributions of citrate and f26bp on the allosteric control of PFKM we require accurate measurements of the metabolites themselves and more detailed model that considers the metabolism surrounding these metabolites. As mentioned in Section 4.1.2, PFKFB3 was not included as there were no differentiating measurements of the metabolite f26bp or the enzyme PFKFB3. Secondly, no viable measurements of citrate can be made without a fractionation experiment such as that conducted by Lee et al., 2019 in HeLa cells. Furthermore, measurements of PFKFB3 phosphorylation would improve model inference as the impact of AMPK phosphorylation could be modelled, as suggested in (Bando et al., 2005).

We have approached modelling the increased activity of mitochondria bound hexokinase as an allosteric process that is activated by the pseudo-metabolite mit, see Section 4.1.2. Even though the regulatory mode is a proxy, the allosteric parameter is not confounded by any other regulatory parameter and can be considered directly. Improvements to this model could be made by considering the phosphorylation of HEX2 by protein kinase B (Akt), as suggested by (Gottlob et al., 2001).

Modelling phosphorylation requires a model of phosphorylation, which has not been implemented into the current Maud version, proposed models are discussed in Chapter 5. Supposing a model of phosphorylation was implemented there are two measurements that would improve the model. First, conducting the same fractionation experiment as in Bustamante and Pedersen, 1977 for the different conditions would give the distribution of nexokinase that is activated. Secondly, performing phosphoproteomics to identify the phoosphorylated fraction of hexokinase would improve inference of the Akt interaction.

Measurements of PFKFB3 and Akt would improve model inference. The current proteomics workflow described in Chapter 2 was unable to measure either of the proteins. Investigating the raw peptide measurements provided no evidence of any identified peptide being measured. This was not unique to this analysis and was also observed in a colleague's analysis (data not shown) using DIA that is a more sensitive technique for identifying peptides. The existence of these proteins was confirmed by transcriptomics conducted by Singh et al., 2018 suggesting that targeted proteomics technique is required to identify these kinases.

When comparing the model estimated to the measured concentration it would appear that a further investigation is required into the regulation of GAPD. A study by Blatnik et al., 2008, found that fumerate was a significant allosteric inhibitor of GAPD. The hypothesis that regulation of GAPD is achieved by fumerate inhibition rather than by changes in enzyme concentration is supported a previous study conducted by (Monge et al., 2020). In their study, intracellular fumerate concentration was found to be almost an order of magnitude higher in CHO-ZeLa than CHO-S during the exponential phase. The differential concentrations together with the hypothesised allosteric mechanism, could explain the increased flux without conflicting with the measured data. In light of our model's inability to fit the measurements without altering the enzyme concentrations, together with the results from the literature, future models of CHO central carbon metabolism should consider allosteric regulation of GAPD by fumerate.

This study relied on generic rate laws that do not fully capture the mechanisms of all the modelled enzymes. The most relevant cases are the enzymes GAPD and PYK, which are known to have ordered mechanisms, whereas the modular rate law most accurately describes enzymes with random mechanisms. This problem was mitigated by the addition of substrate inhibition, but this is not sufficient to fully describe the true mechanism that lack higher order substrate inhibition interactions. Secondly, models for enzyme phosphorylation by other enzymes is not yet implemented, but required to account for PFKFB3 and HEX2 activities. Future development of Maud will amend these missing regulatory components to improve the accuracy of these models.

We have shown that information about metabolic networks is best stored as a kinetic model. They represent a concise summary of all applicable multiomics data and are capable of testing supposed hypothesis, and indicate when current assumptions are incorrect. Although full Bayesian inference was not possible due to time constraints, maximum a posteriori estimation produced useful insights into the target system. Sampling of the model will result in parameter intervals that can provide certainty and explore model degeneracies such as the discussed competing allosteric regulation of PFKM.

In the following chapter we will introduce a general modelling framework to simulate phosphorylation of enzymes applicable to metabolic networks. This will improve our ability to model metabolism and make predictions further from the training datasets.

Methods

4.3.2   Maximum a posteriori point estimation

To calculate the maximum a posteriori parameter configuration, optimisation was performed on the log posterior distribution. This was implemented in Maud (Chapter 3), and was performed using the stan.optimize() function available in cmdstan (Team, 2023). The parameters used for optimising the log-posterior are defined in Table 4.1.

Table 4.1: Optimisation parameters used for maximum a posteriori point estimation

parameter	value	definition
algorithm	LBFGS	Optimization algorithm used for MAP estimation.
iter	2000	Number of iterations before optimization terminates if convergence conditions are not satisfied.
init_alpha	1e-4	Step size of first iteration.
tol_param	1e-12	Convergence tolerances for changes in parameter values.
history_size	10	Number of update vectors used for Hessian approximation.

Kinetic model assumptions

The following kinetic model is based on previous literature exploring mechanisms and regulators. Simplifications were made to these mechanisms to fit the selected kinetic modelling framework defined in Chapter 3. The assumptions are defined below in table 4.2.

Table 4.2: Assumptions used when constructing CHO metabolic model

Assumption	Reasoning
reactions are in rapid equilibrium	It is assumed that the binding and dissociation rates of reactions are significantly higher than that of the slow catalytic step.
Intracellular glucose concentration is at 20 mM.	Studies in human erythrocytes identified that the influx and efflux of glucose were approximately equal with a timescale of 4 seconds (Britton, 1964).
Metabolism is at steady state	Metabolomics, proteomics and fluxomics are measured well before the transition to cell size increase (Pan et al., 2017).

Allosteric regulation

Rather than considering allostery as defined in Chapter 3, we adapted equation 3.5 to not include the free enzyme concentration e_R. This assumes that the allosteric modifiers are able to bind to the enzyme regardless of its binding state, e.g., bound to substrates or products. This assumption improved model fit, and is more consistent with the *prior art* of modelling glycolysis in CHO cells (Mulukutla et al., 2014).

$$Allostery = \frac{1}{1 + L_0\left(\frac{\prod_{i=1}^{t}\left(1 + C_i/e_{T,i}\right)}{\prod_{j=1}^{r}\left(1 + C_j/e_{R,j}\right)}\right)} \tag{4.1}$$

In Equation 4.1 L_0 is the allosteric constant that determines the equilibrium between

the active and inactive enzyme without any allosteric modifiers, C_i and C_j is the concentration of the allosteric inhibitor and activator respectively, and $e_{T,i}$ and $e_{R,j}$ are the dissociation constants for the allosteric inhibitor and activator respectively.

Priors

The prior distributions for the kinetic model are shown in Table E.1 and were determined based on a literature search summarised in Appendix C. For parameters with multiple esimtates, a range was taken that encompassed their measured values with 98% of the probability density. For singularly estimated kinetic parameters we used a defined uncertainty 1 unit on the log-normal scale for k_{cat}, 2 for enzyme dissociation (L_0) parameters, and 0.5 units for binding constants (K_m, K_i, e_R, e_T), all centered around the measured estimate. For the transfer constants which are generally unknown, we applied a prior with a value of 1 to give the possibility of shifting the equilibrium between favouring the active or inactive enzyme.

4.3.3 Evaluating model

In order to determine how well our model captured the underlying process we separated the data into a training set and a validation set, then predicted the fluxes and metabolite concentrations of the validation set based on the training set. Since measurements were available for the validation set it was theoretically possible to evaluate the model using standard Bayesian model evaluation techniques (See Chapters 6 and 7 of Gelman et al., 2013).

The training set included the following cell lines: CHO-S, CHO-ZeLa, CHO-ZenZeLa and CHO-Zen. The validation set included the following cell lines: CHO-S 11 KO and ExpiCHO. These cell lines were chosen as, based on the multiomics dataset in Chapter 2, their central carbon fluxes were different than that of CHO-S and CHO-ZeLa (Figure 2.5. Predicting the validation set measurements despite the metabolic difference would suggest that the model is a good approximation of the underlying process.

In this work we were unable to conduct full Bayesian sampling, and instead performed MAP estimation as described in Section 4.3.2. MAP estimation generates a point estimate rather than an ensemble of parameter configurations; as a result we cannot generate an ensemble of simulated measurements as required for standard Bayesian posterior predictive checking. We therefore conducted an alternative model evaluation method that was feasible using the available information and is targeted at the scientific question at hand.

The MAP fit generates a set parameters that we can differentiate into two separate groups: those that are experiment specific, and those that are the same for all experiments. The experiment specific parameters are the boundary conditions: that is, enzyme concentrations and unbalanced metabolite concentrations. The non-experiment specific parameters are the kinetic and thermodynamic parameters that describe the underlying properties of the modelled enzymes.

To perform model validation we generated an ensemble of boundary conditions parameters for the validation set by simulating from these parameters' prior distribution [1]. The result was a set of 1000 parameter configurations reflecting the measured values of the boundary conditions of the validation data. We then solved the steady state problem for each configuration in this ensemble using the MAP estimate for non-experiment-specific parameters resulting in a set of fluxes and balanced metabolite concentrations for each validation condition. These form the basis of the evaluation.

The reason for using this model evaluation method is to check that the MAP estimate for the non-experiment-specific parameters produced reasonable metabolic states. While it would be preferable to conduct full posterior sampling for the out of sample data, this method is an indicator that the fitted parameters are not just artefacts of the training data and can produce reasonable estimates over a range of plausible boundary conditions.

The workflow is as follows:

1. Find the optimal parameter set $\hat{\theta}$ using MAP on the training data.

2. Take a subset of $\hat{\theta}$ that are the same for all experiments, leaving only:

   - k_{cat}

   - K_m

   - K_i

   - L_0

   - e_R

   - e_T

   - $\Delta_f G$.

   This is the optimal non-experiment-specific parameter set and will be used for the remaining simulations

3. Randomly generate a value from the prior distribution for experiment specific parameters in the validation dataset, referred to as $\theta_{validation}$. These include:

   - $[Enzyme]$

   - $[Unbalanced\ Metabolites]$

   - $Flux\ of\ measured\ biomass\ drains$

4. Generate steady state fluxes and balanced metabolite concentrations using $\hat{\theta}$ and $\theta_{validation}$.

5. Repeat steps 3 and 4 1000 times.

---

[1] Note that the prior distributions we chosen to reflect the measurements of these parameters

Relative error

The relative error (RE) is a metric used to evaluate the distance to the measured value. The relative error was used as a metric to account for the varying orders of magnitude. The RE is calculated as defined in equation 4.2

$$\text{WE} = \frac{y_i - \hat{y}_i}{\hat{y}_i} \tag{4.2}$$

$$\text{where} \quad y_i = \text{measured value} \tag{4.3}$$

$$\hat{y}_i = \text{predicted value} \tag{4.4}$$

References

Shestov, A. A., Liu, X., Ser, Z., Cluntun, A. A., Hung, Y. P., Huang, L., Kim, D., Le, A., Yellen, G., Albeck, J. G., & Locasale, J. W. (2014). Quantitative determinants of aerobic glycolysis identify flux through the enzyme GAPDH as a limiting step. *eLife*, *3*. https://doi.org/10.7554/eLife.03342

Mulukutla, B. C., Yongky, A., Daoutidis, P., & Hu, W.-S. (2014). Bistability in Glycolysis Pathway as a Physiological Switch in Energy Metabolism (P. Dzeja, Ed.). *PLoS ONE*, *9*(6), e98756. https://doi.org/10.1371/journal.pone.0098756

Mulukutla, B. C., Yongky, A., Grimm, S., Daoutidis, P., & Hu, W.-S. (2015). Multiplicity of steady states in glycolysis and shift of metabolic state in cultured mammalian cells. *Plos One*, *10*(3), e0121561. https://doi.org/10.1371/journal.pone.0121561

O'Brien, C., Allman, A., Daoutidis, P., & Hu, W.-S. (2019). Kinetic model optimization and its application to mitigating the Warburg effect through multiple enzyme alterations. *Metabolic Engineering*. https://doi.org/10.1016/j.ymben.2019.08.005

Mulquiney, P. J., & Kuchel, P. W. (1999). Model of 2,3-bisphosphoglycerate metabolism in the human erythrocyte based on detailed enzyme kinetic equations: Equations and parameter refinement. *The Biochemical Journal, 342 Pt 3*, 581–596.

Hagrot, E., Oddsdóttir, H. Æ., Mäkinen, M., Forsgren, A., & Chotteau, V. (2019). Novel column generation-based optimization approach for poly-pathway kinetic model applied to CHO cell culture. *Metabolic Engineering Communications*, *8*, e00083. https://doi.org/10.1016/j.mec.2018.e00083

Ahn, K. J., Kim, J., Yun, M., Park, J. H., & Lee, J. D. (2009). Enzymatic properties of the N- and C-terminal halves of human hexokinase II. *BMB Reports*, *42*(6), 350–355. Retrieved February 15, 2019, from https://www.ncbi.nlm.nih.gov/pubmed/19558793

Beitner, R. (1979). The role of glucose 1,6-bisphosphate in the regulation of carbohydrate metabolism in muscle muscle. *Trends in Biochemical Sciences*, *4*(10), 228–230. https://doi.org/10.1016/0968-0004(79)90124-5

Britton, H. G. (1964). Permeability of the human red cell to labelled glucose. *The Journal of Physiology*, *170*(1), 1–20. https://doi.org/10.1113/jphysiol.1964.sp007310

Bustamante, E., & Pedersen, P. L. (1977). High aerobic glycolysis of rat hepatoma cells in culture: Role of mitochondrial hexokinase. *Proceedings of the National Academy of Sciences*, *74*(9), 3735–3739. https://doi.org/10.1073/pnas.74.9.3735

Gottlob, K., Majewski, N., Kennedy, S., Kandel, E., Robey, R. B., & Hay, N. (2001). Inhibition of early apoptotic events by Akt/PKB is dependent on the first committed step of glycolysis and mitochondrial hexokinase. *Genes & Development*, *15*(11), 1406–1418. https://doi.org/10.1101/gad.889901

Heiden, M. G., Chandel, N. S., Schumacker, P. T., & Thompson, C. B. (1999). Bcl-xL Prevents Cell Death following Growth Factor Withdrawal by Facilitating Mito-

chondrial ATP/ADP Exchange. *Molecular Cell*, *3*(2), 159–167. https://doi.org/10. 1016/S1097-2765(00)80307-X

Telang, S., Yalcin, A., Clem, A. L., Bucala, R., Lane, A. N., Eaton, J. W., & Chesney, J. (2006). Ras transformation requires metabolic control by 6-phosphofructo-2- kinase. *Oncogene*, *25*(55), 7225–7234. https://doi.org/10.1038/sj.onc.1209709

Zancan, P., Almeida, F. V. R., Faber-Barata, J., Dellias, J. M., & Sola-Penna, M. (2007). Fructose-2,6-bisphosphate counteracts guanidinium chloride-, thermal-, and ATP-induced dissociation of skeletal muscle key glycolytic enzyme 6-phosphofructo- 1-kinase: A structural mechanism for PFK allosteric regulation. *Archives of Bio- chemistry and Biophysics*, *467*(2), 275–282. https://doi.org/10.1016/j.abb.2007. 08.032

Singh, A., Kildegaard, H. F., & Andersen, M. R. (2018). An Online Compendium of CHO RNA-Seq Data Allows Identification of CHO Cell Line-Specific Transcriptomic Signatures. *Biotechnology Journal*, *13*(10), 1800070. https://doi.org/10.1002/ biot.201800070

Manes, N. P., & El-Maghrabi, M. R. (2005). The kinase activity of human brain 6-phosphofructo- 2-kinase/fructose-2,6-bisphosphatase is regulated via inhibition by phospho- enolpyruvate. *Archives of Biochemistry and Biophysics*, *438*(2), 125–136. https:// doi.org/10.1016/j.abb.2005.04.011

Muriel, J. C., Cowie, N., Mansouvar, M., Groves, T., & Nielsen, L. K. (2023). Shu: Visu- alization of high dimensional biological pathways. https://doi.org/10.48550/ ARXIV.2304.07178

Buchsteiner, M., Quek, L.-E., Gray, P., & Nielsen, L. K. (2018). Improving culture per- formance and antibody production in CHO cell culture processes by reducing the Warburg effect. *Biotechnology and Bioengineering*, *115*(9), 2315–2327. https: //doi.org/10.1002/bit.26724

Liebermeister, W., Uhlendorf, J., & Klipp, E. (2010). Modular rate laws for enzymatic re- actions: Thermodynamics, elasticities and implementation. *Bioinformatics*, *26*(12), 1528–1534. https://doi.org/10.1093/bioinformatics/btq141

Noor, E., Flamholz, A., Liebermeister, W., Bar-Even, A., & Milo, R. (2013). A note on the kinetics of enzyme action: A decomposition that highlights thermodynamic effects. *FEBS Letters*, *587*(17), 2772–2777. https://doi.org/10.1016/j.febslet. 2013.07.028

Lee, W. D., Mukha, D., Aizenshtein, E., & Shlomi, T. (2019). Spatial-fluxomics provides a subcellular-compartmentalized view of reductive glutamine metabolism in cancer cells. *Nature Communications*, *10*(1), 1351. https://doi.org/10.1038/ s41467-019-09352-1

Bando, H., Atsumi, T., Nishio, T., Niwa, H., Mishima, S., Shimizu, C., Yoshioka, N., Bu- cala, R., & Koike, T. (2005). Phosphorylation of the 6-Phosphofructo-2-Kinase/Fructose 2,6-Bisphosphatase/PFKFB3 Family of Glycolytic Regulators in Human Cancer.

Clinical Cancer Research, *11*(16), 5784–5792. https://doi.org/10.1158/1078-0432.CCR-05-0149

Blatnik, M., Frizzell, N., Thorpe, S. R., & Baynes, J. W. (2008). Inactivation of glyceraldehyde-3-phosphate dehydrogenase by fumarate in diabetes: Formation of S-(2-succinyl)cysteine, a novel chemical modification of protein and possible biomarker of mitochondrial stress. *Diabetes*, *57*(1), 41–49. https://doi.org/10.2337/db07-0838

Monge, I. M., Hefzi, H., Sanchez, P. C., Marín de Mas, I. B., Decker, M., Lecina, M., Cairó, J. J., Lewis, N. E., & Nielsen, L. K. (2020). *Lactate reduction in CHO cell cultures through metabolic analysis* (Poster). https://backend.orbit.dtu.dk/ws/portalfiles/portal/216677825/IMM_ESACT_2019_IG_LKN.pdf

Team, S. D. (2023). *CmdStanPy*. https://github.com/stan-dev/cmdstanpy

Pan, X., Dalm, C., Wijffels, R. H., & Martens, D. E. (2017). Metabolic characterization of a CHO cell size increase phase in fed-batch cultures. *Applied Microbiology and Biotechnology*, *101*(22), 8101–8113. https://doi.org/10.1007/s00253-017-8531-y

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., & Rubin, D. B. (2013, November 27). *Bayesian Data Analysis* (0th ed.). Chapman and Hall/CRC. https://doi.org/10.1201/b16018

5 Phosphorylation of metabolic enzymes

5.1 Introduction

Our previous analysis of the Warburg effect indicated that phosphorylation of Hexokinase 2, and PFKFB were essential for regulation of glycolysis. Phosphorylation is mediated by kinases that typically attach a phosphate group to serine/thrionine residues on a protein and typically act to deactivate the protein. Inactivation is reversed by phophatases that remove the phosphate group (see Figure 5.1). The combined action of the kinases and phosphatases regulates the amount of enzyme through a concerted change in the enzymes structure.

Mechanisms for phosphorylation have been considered heavily in signal transduction pathways Salazar and Höfer, 2006, 2009. However, research into phosphorylation applied to metabolic models remains a largely unexplored field. There are several studies outside of this work the emphasise the importance of phosphorylation in metabolic modelling. For example, Xia et al., 2022, observed a decrease in enzyme fraction of glycolytic enzymes and in increase in in glycolytic flux that was hypothesised to be a result of phosphorylation. Despite the fractional decrease in enzyme concentrations that they observed, phosphoproteomics would reveal a general decrease in phosphorylation occupation, in other words, an increase in enzyme activity. Another study by Oliveira et al., 2012, observed a linear correlation between the unphosphorylated enzyme fraction and flux in yeast central carbon metabolism. It is clear that phosphorylation is an important regulator of metabolism and in order to use kinetic models we require a concise and consistent model of phosphorylation.

This chapter presents a novel approach to modelling phosphorylation in the context of kinetic modelling. where we consider three possible models of phosphorylation with assumptions that make them unique to metabolic networks. The objective of this chapter is to identify a model that is capable of explaining the observed behaviour of phosphorylation that is mechanistically sound without requiring a large number of kinetic parameters.
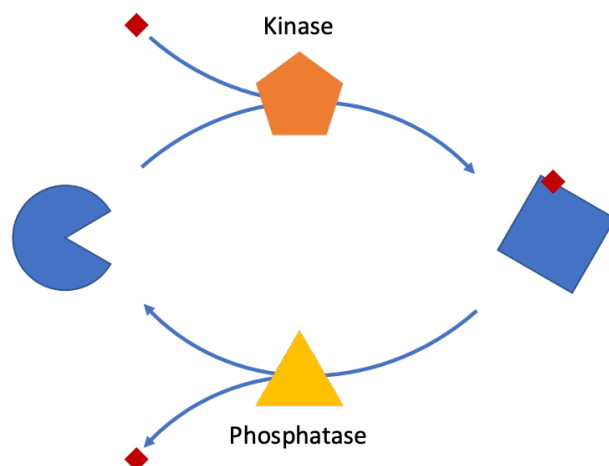
Figure 5.1: Schematic of phosphorylation, where the phosphate group (red diamond), is added to the active enzyme (blue circle) by the kinase (orange pentagon) that converts the enzyme into its inactive (blue square) form. This process is reversed by the pohsphatase (yellow triangle) that removes the phosphate group.

5.1.1   The three proposed phosphorylation models

The ideal parameterisation minimises the number of parameters fit, can display co-operativity, and does not violate mass balance relationships. Fewer parameters are desirable as they reduce the number of degrees of freedom, making parameter inference easier. Cooperativity is a characteristic feature of phosphorylation whereby the activity of the phosphorylated enzyme changes disproportionately in response to a comparatively small change in phosphorylation fraction. And finally, mass balance relationships must always be conserved in order to ensure an accurate model.

Considering the defined criteria we propose three models: 1) a linear model whereby the rate is proportional to the unphosphorylated enzyme fraction 2) a model analogous to the MWC model described in Section 3.2.4 where the phosphorylated fraction acts as the bound inhibitor, and 3) a combinatorial model, which represents a special instance of allostery where any phosphorylated subunit causes a concerted change to the inactive state in the multimeric enzyme. psi Before further elaboration of these model we will present fundamental assumptions consistent for all three models as presented in Table 5.1.

Given these assumptions the phosphorylation fraction can be described as a function of the phosphatase and kinase concentrations, together with the corresponding k_{cat} parameters as shown in Equation 5.1. This functional relationship represents the steady state phosphorylation fraction that is determined by the proportion of kinase flux to the total flux.

Table 5.1: Fundamental assumptions for all three proposed phosphorylation models

Assumption	Justification
Phosphorylating enzymes are in small concentrations relative to the target enzyme	When considering metabolic enzymes it can be hypothesised that regulation of the enzymes shouldn't be hindered by the binding of the regulators themselves. This hypothesis is supported by Madhukar et al., 2015, where metabolic enzymes in glycolysis were significantly higher in concentration that most other proteins. Additionally, the proteins we identified as kinases (PFKFB3, AMPK, Akt) were difficult to measure compared to the metabolic enzymes (Chapter 4).
Rate is independent of atp, P_i and adp concentrations	The concentrations of these metabolites are typically constant across conditions (Chapter 2)
Binding constants are significantly smaller than the concentrations of the kinase and phosphatases. And therefore, rates are proportional to kinase and phosphatase concentration	This was assumed under the reasoning that regulation of metabolic enzymes will be as energetically efficient as possible. If the binding constants are on the same order of magnitude as the kinases and phosphatases then an increase in concentration will result in a less than linear change in flux because not all of the modulating enzymes will be bound.

$$\psi = \frac{v_k}{v_k + v_p} \tag{5.1}$$

where,

$$v_k = k_{cat,k} \cdot [K] \tag{5.2}$$

$$v_p = k_{cat,p} \cdot [P] \tag{5.3}$$

Linear

In the paper by Oliveira et al., 2012, the authors note a linear relationship between the unphosphorylated enzyme fraction and the flux. This relationship can be described using Equation 5.4.

$$v \propto (1 - \psi) \tag{5.4}$$

MWC-analogous model

By considering each phosphorylation event as an allosteric inhibitor and each unphosphorylated site as an allosteric activator we can incorporate phosphorylation fraction into the generalised MWC equation (see Equation 5.5).

The MWC formalism specifies that there is an equilibrium between the active and inactive state of an enzyme, see Figure 5.2. The binding event of a negative effector will then stabilise the inactive state shifting the equilibrium. The opposite effect oc-

curs when an activating effector binds, shifting the equilibrium towards the active state Popova and Sel'kov, 1975.

Rather than considering the binding of independent active and inactive effectors, we considered the phosphorylated site as an allosteric inhibitor and the inactive site as the allosteric activator. This is analogous to the MWC model defined in Saa and Nielsen, 2015 and in Chapter 3, with the difference being that the activators and inhibitors are the coupled ratio of unphosphorylated and phosphorylated sites respectively. The coupling ratio $\frac{\psi}{1-\psi}$ permits a greater range of allostery than if either were considered independently.

$$v \propto \frac{1}{1 + L_0(1 + \frac{\psi}{1-\psi})^N} \tag{5.5}$$

Where, L_0 represents the equilibrium between the T and R states without effector present and N is the number of subunits that form the complex.
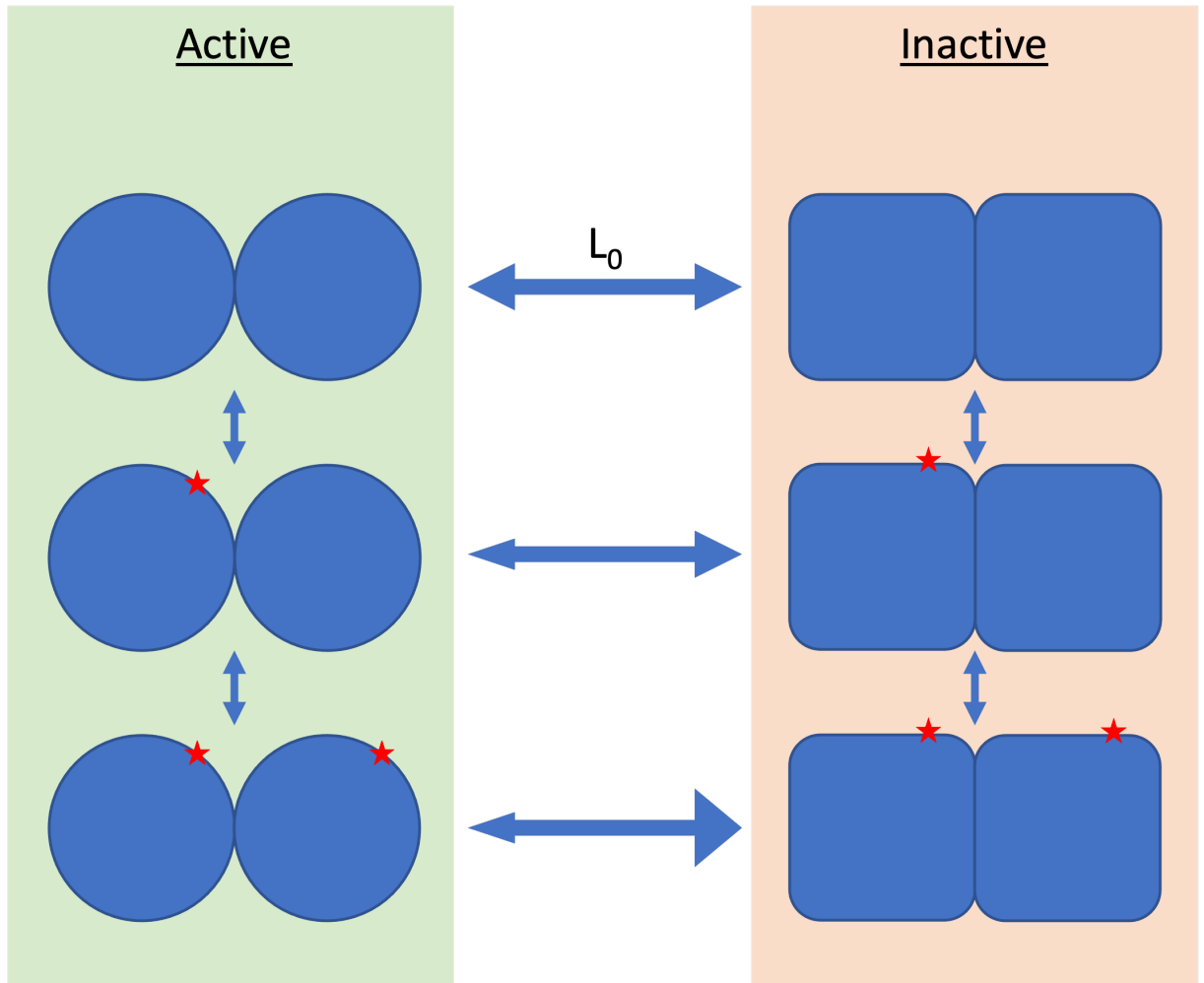
Figure 5.2: Phosphorylation modelled as an MWC model. Inactive enzymes are shown as squares, and active enzymes are circles. Phosphorylation of single subunits are indicated by a red star. The equilibrium between states is shown the by arrows, with a larger arrowhead representing a shift in equilibrium towards that state.

Combinatorial model

To develop an equation for the combinatorial model we assume that each phosphorylation site can be either in a phosphorylated or unphosphorylated state. Further, we assume that the probability of each site being phosphorylated is the same and probabilistically independent of the states of the other sites. Based on these assumptions we can calculate the probability of any combination of phosphorylation states for sites on the same enzyme. This is represented by the binomial expression in Equation 5.6. Where n is the number of phosphorylated sites, N is the total number of sites, and ψ is the probability of a single site being phosphorylated.

$$Pr(n) \propto \binom{N}{n} (\psi)^n (1 - \psi)^{(N-n)} \tag{5.6}$$

Finally, we assume that any single binding will lead to a concerted change in structure that completely inactivates the enzyme. Therefore, the overall rate is proportional to the fraction of enzymes with no phosphorylated sites. We can represent this mathematically as shown in Equation 5.8.

$$v \propto \binom{N}{0} (\psi)^0 (1 - \psi)^N \tag{5.7}$$

$$\propto (1 - \psi)^N \tag{5.8}$$
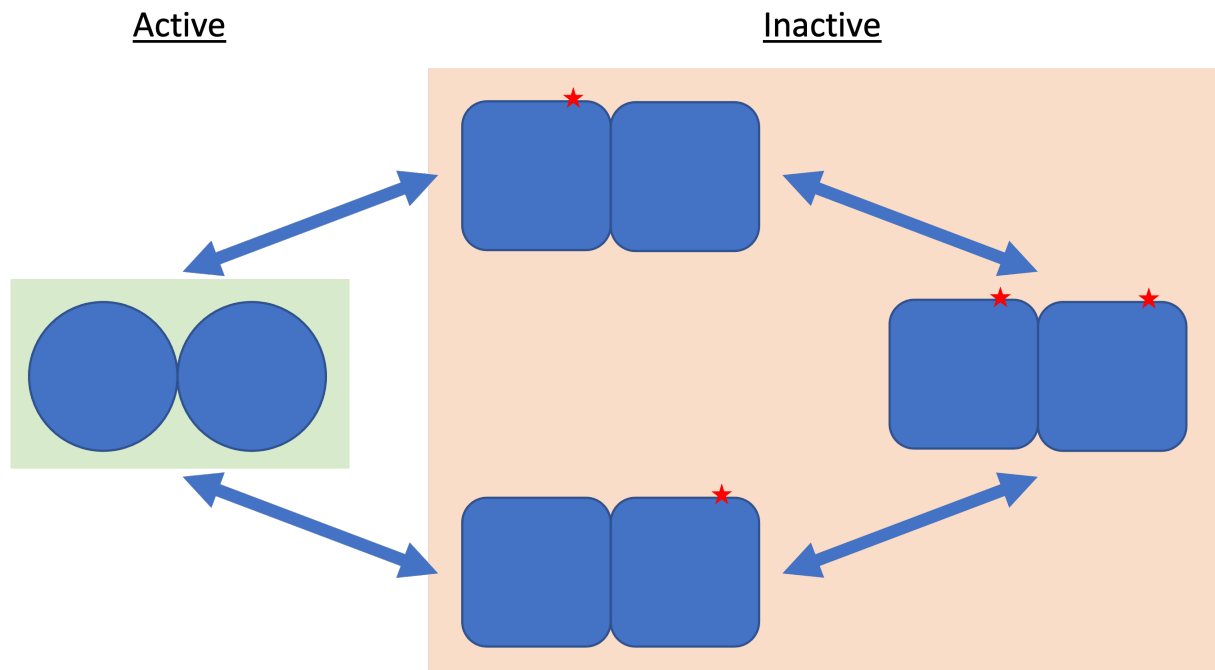
Figure 5.3: Model of combinatorial phosphorylation where the transitions between phosphorylation states is given by the blue arrows. Active and inactive enzymes are shown by the indicated background and by the shape, with square enzymes being inactive and circles being active. Phosphorylation of an individual subunit it denoted by a red star.

The generation of the schematic in Figure 5.3 represents a dimer and all of the possible active and inactive states. This model differs from the mwc representation as a single phosphorylation event is assumed to inactivate the enzyme.

5.2   Methods

There are three metrics that we used to compare these models: 1) Dynamic range, which is the change in phosphorylation fraction required to inactivate flux, and 2) Sensitivty region, this is where the enzyme is able to respond to changes in phosphohrylation fraction.

5.2.1 Dynamic Range

Measuring the dynamic range is possible either by using the maximum inflection point, or by calculating the difference in phosphorylation for a given difference in active enzyme. To give a tangible measure of the cooperativity experience we determined the change in phosphorylation fraction required to go from 90% active to 10% active.

Fortunately for every mapping $f : \psi \rightarrow X$, where ψ is the phsphorylation fraction and X is the fraction of active enzyme, there exists its inverse $f^{-1} : X \rightarrow \psi$, given by Equations 5.9-5.11. Therefore, we can analytically determine the dynamic range for each model using equation 5.12.

$$\text{Linear: } \psi_X = 1 - X \tag{5.9}$$

$$\text{MWC: } \psi_X = \frac{(\frac{1}{X} - 1)^{\frac{1}{N}}}{\frac{1}{L_0^{\frac{1}{N}}} + (\frac{1}{X} - 1)^{\frac{1}{N}}} \tag{5.10}$$

$$\text{Combinatorial: } \psi_X = 1 - X^{\frac{1}{N}} \tag{5.11}$$

$$80\% \text{ DR} = \psi_{90} - \psi_{10} \tag{5.12}$$

We then calculated the dynamic range for both the combinatorial and mwc model between 1 and 64 subunits, varying by powers of 2. By taking the difference between the dynamic ranges at increasing numbers of subunits we can inspect which model exhibits a greater change in active enzyme with a smaller change in phosphorylation fraction. We did not plot the dynamic range for the linear model as it is always 0.8.

For all dynamic range calculations the transfer constant L_0 in equation 5.10 was set to 1. This variable has an impact on the dynamic range as we demonstrate in Figure 5.6, with the dynamic range at its lowest at $L_0 = 1$ and increases as L_0 increases or decreases.

5.2.2 Sensitivity Region

To determine how the sensitivity region changed when increasing the number of enzyme subunits we did a case study by increasing the number of subunits and calculating the active fraction of enzyme as ψ varied between 0 and 0.99. Subunits were varied in powers of 2, from 1 to 16 to visualise the impact of the number of subunits has on where the dynamic range is located.

5.3 Results

Now that we have defined how we will determine these comparative metrics, we can compare the linear, mwc and combinatorial models of phosphorylation.

5.3.1 Comparison of dynamic range

To investigate the dynamic range we measured the change in phosphorylation fraction to change the active enzyme fraction from 90% to 10%. We were able to compute this analytically as shown in Section 5.2.1. We found that as the subunits increased the dynamic range of the MWC model can be up to 0.14 less than the required change in phosphorylation fraction for the combinatorial model, see Figure 5.4. The difference in sensitivity between the two models is greatest at 4 subunits and declines thereafter as both models became increasingly more sensitive to small changes in phosphorylation.
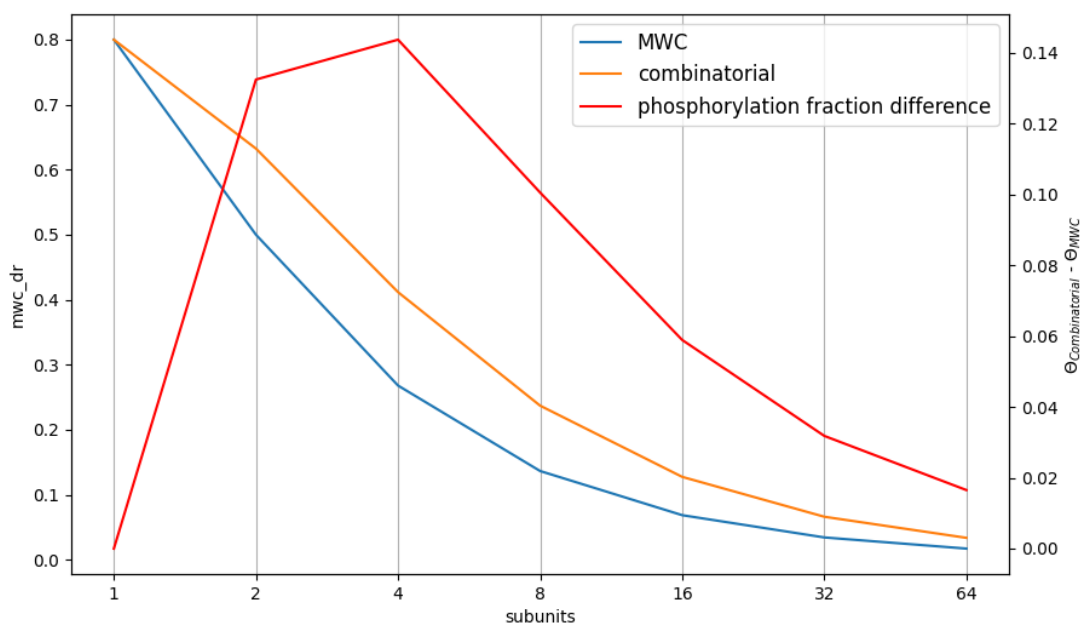
Figure 5.4: A dynamic range comparison of the combinatorial and mwc phosphorylation models comparing the fraction change required to achieve a change in flux from 90% to 10%, as calculated in Section 5.2.1. The orange and blue lines represent the combinatorial and mwc phosphorylation dynamic ranges for varying numbers of subunits. The red line presents the difference between the dynamic ranges, with a higher number indicating more phosphorylation is needed to elicit a change in flux between 90 % to 10%.

We are aiming to devise a mathematical representation of phosphorylation that exhibits cooperativity so that a small change in phosphorylated fraction results in a large change in flux. Both the combinatorial and MWC model are capable of cooperativity, whereas the linear system is not (data not presented).

This suggests that if phosphorylation were to be as sensitive as possible then the selected model would be the MWC model.

5.3.2 Increased sensitivity decreases the $\Theta_{\frac{1}{2}}$

Here performed case studies on both the combinatorial and mwc model as specified in Section 5.2.2. We found that as the number of subunits increase the area where the enzyme is sensitive to changes varies between the assumed models. In the instance of the MWC model, the region of sensitivity was stable as the number of subunits increased at 50 % phosphorylation fraction, whereas the combinatorial model shifted towards 0 %.

Due to the assumptions behind the combinatorial model, increasing the number of subunits directly decreases the number of occupied sites required to inactivate the enzyme Figure 5.5. As any single phosphorylation event inactivates the enzyme, as enzyme complexes increase the probability that an enzyme will be inactivated also increases.
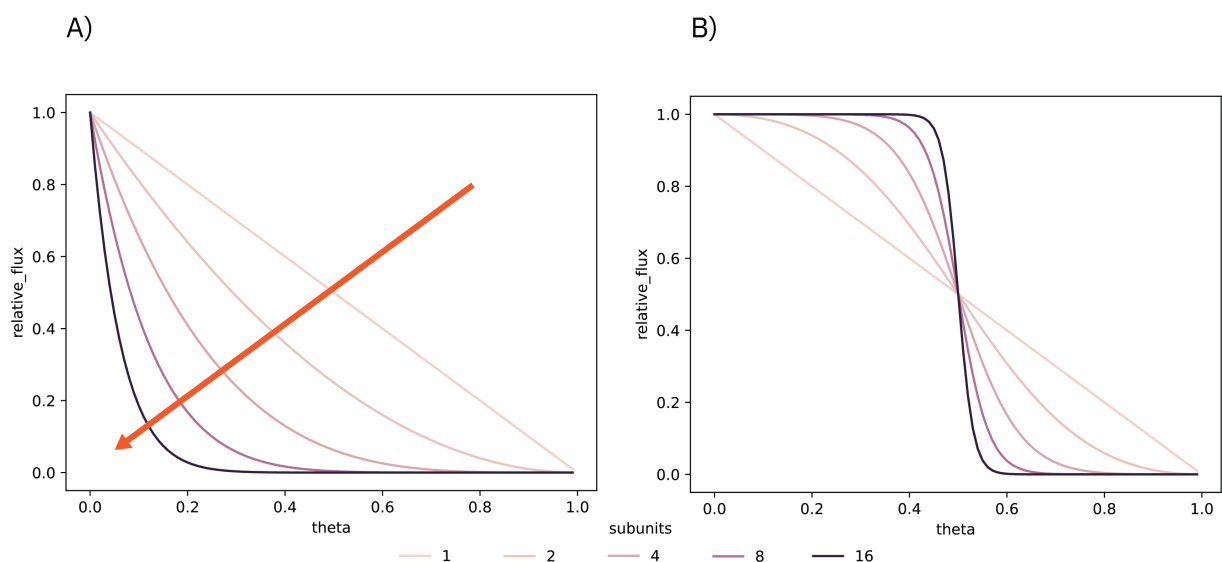
Figure 5.5: The response of the flux under increasing phosphorylation fractions. A) Presents the changing activity profiles at defined numbers of subunits for the combinatorial mechanism. And B) is the same evaluation but for the MWC mechanism. The orange arrows represents the general decrease in phosphorylation fraction where the sensitive region occurs for the combinatorial model, and this change is not observed in the MWC model

This is in contrast to the MWC model where increasing the number of subunits solely increases the sensitivity. Figure 5.6 presents the same case study conducted to investigate the sensitivity of the mwc and combinatorial models, however, rather than changing the subunits we altered the equilibrium constant L_0 between 0.01 and 100 of the mwc model. Here we show that we can change the sensitivity region by altering a single number. As a side effect, the sensitivity also increases the further the transfer constant moves from 1, see Figure 5.6. Therefore we cannot eliminate either model if the sensitivity region of phosphorylation sites predominantly occurs in low phosphorylation fractions.
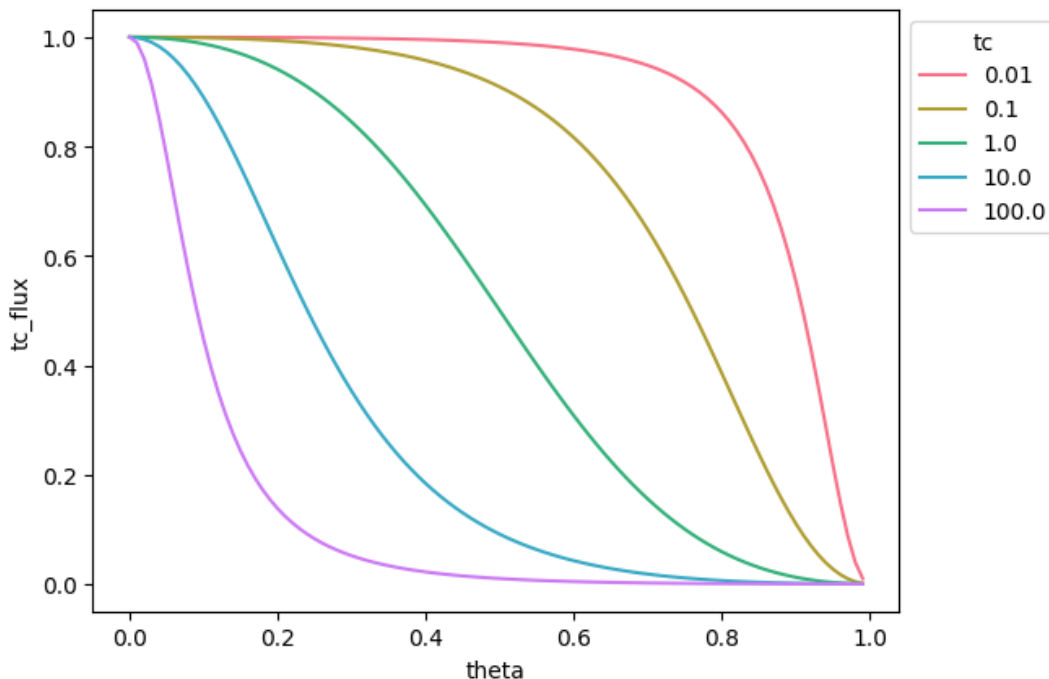
Figure 5.6: MWC model sensitivity towards the transfer constant. Increasing the equilibrium constant shifts the inflection point towards the left and vice versa for decreasing the transfer constant. Secondarily, increasing or decreasing the constant from equilibrium of 1 also increases the sensitivity of the enzyme to phosphorylation.

5.4 Discussion

The combinatorial and MWC model presented in this chapter are to our knowledge the first attempt at constructing a generalised approximation for phosphorylation in metabolic networks. At the cost of an additional parameter, the reformulated MWC model can approximate the combinatorial model and also achieve sensitivity at higher phosphorylation occupancy rates. As with the metabolic allostery model such as the standard MWC and the KNF models, no single model is necessarily correct. However, we hope to provide a framework for future studies to integrate phosphorylation into metabolic models.

Both the MWC and combinatorial model are capable of exhibiting allostery that increases in sensitivity under an increasing number of subunits. However, there is evidence that suggests that the combinatorial model is more plausible. Studies by Sygusch and Beaudry, 1997 suggest that even a single phosphorylation event is sufficient to inactivate an entire quaternary complex. This is consistent with how phosphorylation is described by the combinatorial model. Furthermore, glycolytic enzymes in *E. coli* are believed to be regulated by phosphorylation and were observed in Schastnaya et al., 2021 to have a low phosphorylation fractions. This implies that the region of dynamic range will also occur at low phosphorylation fractions, which is consistent with the combinatorial model.

Unfortunately, this is not enough evidence to eliminate the MWC model. The assumptions that underlie the combinatorial model are more restrictive than the MWC model because of the smaller region of cooperativity. If the true region of coooperativity is outside of the one allowed by combinatorial model this could result in catastrophic model failure. The MWC model is arguably a safer choice.

Determining the true mechanism underlying phosphorylation using kinetic models is non-trivial. Given the current analytical techniques it will be difficult to devise an experiment that would differentiate between the MWC and combinatorial models unless phosphorylation regulation was observed outside of the low dynamic range region.

Choosing between the MWC and combinatorial model requires an experimental quantification of phosphorylation fraction and activity of an enzyme. This can be conducted by comparing two contrasting combinatorial and MWC model and evaluating the log-probability of each model. selecting the model that is most consistent with the observed data. Further improvements of determining the true model will be improved by measurements of the phosphorylation fraction of the target enzyme.

In the context of Chapter 4, we now have a mathematical model of phosphorylation that can be integrated for the HEX2 and PFKFB enzymes. The resulting model will have explanatory values for how phosphorylation effects metabolism and we can now use the inputs of protein concentration of the kinases and phosphatases as inputs to the model. This will improve the predictive power of the model for different states, and is directly measurable.

References

Salazar, C., & Höfer, T. (2006). Kinetic models of phosphorylation cycles: A systematic approach using the rapid-equilibrium approximation for protein–protein interactions. *Biosystems*, *83*(2), 195–206. https://doi.org/10.1016/j.biosystems.2005.05.015

Salazar, C., & Höfer, T. (2009). Multisite protein phosphorylation – from molecular mechanisms to kinetic models. *The FEBS Journal*, *276*(12), 3177–3198. https://doi.org/10.1111/j.1742-4658.2009.07027.x

Xia, J., Sánchez, B. J., Chen, Y., Campbell, K., Kasvandik, S., & Nielsen, J. (2022). Proteome allocations change linearly with the specific growth rate of Saccharomyces cerevisiae under glucose limitation. *Nature Communications*, *13*(1), 2819. https://doi.org/10.1038/s41467-022-30513-2

Oliveira, A. P., Ludwig, C., Picotti, P., Kogadeeva, M., Aebersold, R., & Sauer, U. (2012). Regulation of yeast central metabolism by enzyme phosphorylation. *Molecular Systems Biology*, *8*, 623. https://doi.org/10.1038/msb.2012.55

Madhukar, N. S., Warmoes, M. O., & Locasale, J. W. (2015). Organization of Enzyme Concentration across the Metabolic Network in Cancer Cells (V. D. Appanna, Ed.). *PLOS ONE, 10*(1), e0117131. https://doi.org/10.1371/journal.pone.0117131

Popova, S. V., & Sel'kov, E. E. (1975). Generalization of the model by Monod, Wyman and Changeux for the case of a reversible monosubtrate reaction SR,TP. *FEBS Letters*, *53*(3), 269–273. https://doi.org/10.1016/0014-5793(75)80034-2

Saa, P., & Nielsen, L. K. (2015). A general framework for thermodynamically consistent parameterization and efficient sampling of enzymatic reactions. *PLoS Computational Biology*, *11*(4), e1004195. https://doi.org/10.1371/journal.pcbi.1004195

Sygusch, J., & Beaudry, D. (1997). Allosteric communication in mammalian muscle aldolase. *The Biochemical Journal*, *327 (Pt 3)*, 717–720. Retrieved October 19, 2018, from https://www.ncbi.nlm.nih.gov/pubmed/9581547

Schastnaya, E., Raguz Nakic, Z., Gruber, C. H., Doubleday, P. F., Krishnan, A., Johns, N. I., Park, J., Wang, H. H., & Sauer, U. (2021). Extensive regulation of enzyme activity by phosphorylation in Escherichia coli. *Nature Communications*, *12*(1), 5650. https://doi.org/10.1038/s41467-021-25988-4

6   Final Remarks

In this thesis we developed a kinetic model of CHO metabolism with the objective understanding the the Warburg effect phenotype. We developed a high throughput multiomics workflow that allowed us to collect the information fundemental to understanding the metabolic space, and collected data for a diverse set of phenotypes (Chapter 2). The preliminary results suggested that the Warburg effect was not a result of a changing proteome and that metabolomics would suggest few changes between the high and low flux states. Understanding metabolism required an inference framework that could efficiently explore the posterior space and compute a mechanistic model, this necessitated the development of Maud (Chapter 3). We showed that considering the posterior space rather than approximations reduces cofidence intervals and improves model predicitons. And finally, we applied a kinetic model to the data collected in Chapter 2 to elucidate the regulatory mechanism between the high flux CHO-S cell line and the low flux CHO-ZeLa cell line that lacked the Warburg effect (Chapter 4). Rather than sampling the full posterior space, the maximal *a posteriori* was used as an estimate of model behaviour due to the computational demand. This analysis suggested that the Warburg effect requires allosteric regulation for HEX2 and PFKM and that further research is required in the TPI and GAPD reactions. The findings from this study resulted in the proposal of mechansims for phosphorylation, which are essential for understanding the regulatory network surrounding the Warburg effect (Chapter 5).

Towards the goal of answering the oustanding question, often termed, *the enigma of the Warburg effect*, we were successful in futhering understading the regulatory structure of the EMP pathway and revealed enzymes that are not fully understood. However, there are hurdles that once overcome will improve inference and advance kinetic modelling as a field, these will be discussed in the following section.

6.1   Limitations

There are several limitation to this thesis that must be addressed, and will be separated into two catagories, currently fixable, and unresolved challanges.

6.1.1   Currently Fixable

The most obvious limitation of this thesis was the lack of full HMC sampling applied to the CHO kinetic model. This is a limitation due to the computational time required to sample the full posterior space. As we showed in Chapter 3, the posterior intervals cannot be summaried with the MAP and hessian, however, the tradeoff is a considerable increase in sampling time. There are two methods for decreasing the computational time of the HMC sampler: 1) Improving the ODE solver, and 2) Parallelisation along the experimental dimension. Improving the ODE solver can either be achieved by using

a faster integrator, or by improving the initial guess such that few evaluations are required. The current integrator is developed by Sundials and has been benchmarked as one of the most efficent solvers for metabolic models Städter et al., 2021. Furthermore, there are properties of the steady state problem that can be exploited by using an ODE solver to initialise the guess for an agebraic solver Margossian, 2018. This was implemented as an experimental feature in Chapter 4 to solve the optimisation problem, however, further tuning is required to be stable for all problems. Secondly, as each leapfrog step in the HMC sampler is close to the previous solution, previous steady states can be leveraged to improve the initial guess for the ODE solver. Finally, parallelisation along the expeiemental dimension will decrease the computational time proportional to the number of experimental blocks fragmented. There are limits to this method as 4 cores with 4 fragments requires 16 cores, which is a standard limit of user laptops.

The modular rate law is a powerful tool that alleviates the need for user defined rate laws, however, there are mechanisms that are not encapsulated by the modular rate law such as ordered mechanisms and reaction mechanisms where the intermediate states are not in rapid equilibrium with the substrates. Studies that expand the mechanism space such as those conducted by Cleland, W.W., 1989, provide a more diverse set of mechanism behaviour at the expense of increased user input and increased parameter requirements Cleland, 1989; Leskovac et al., 2006. Reactions that diverge from the behaviour of the modular rate law should be evaluated on a case by case basis. In this thesis, improvements to the GAPD and phosphoglycerate kinase (PGK) enzymes are required to address substrate inhibition due to binding order.

6.1.2 Unresolved challanges

Bayesian inference produces a posterior set of points that may be highly correlated. Considering the future of biology and the emergence of biofoundries, iterative modelling will be commonplace that would greatly benefit from past modelling efforts. If the posterior space was summarised into a multivariate distribution using the summary statistics of the mean and covariance, sampling times would be not grow with the amount of biofoundry cycles. However, as we have shown with the MAP comparison, the posterior space is not well captured by summary statistics and the model requires retraining on the complete dataset.

6.2 Conclusions

This is the first kinetic model of CHO metabolism to use a contrasting Warburg effect phenotype and also collect a metabolomics, proteomics, and fluxomics. This is distinct from previous attempts that include non-Warburg effects conditions that are undergoing slow growth as the model still needs to satisfy the relatively high biosynthetic needs of CHO-ZeLa growth. Because of the diverse conditions involved with the amount of multiomics data collected, we were able to conclude that the TPI and GAPD enzymes were inadequetly described using the assumed mechanism in the model

(Chapter 4). This would not have been possible without the diverse multiomics data collected in Chapter 2.

Kinetic models are useful tools for elucidating metabolic regulation. Without consolidating all regulatory interactions into a single model identification of regulation would be impossible. We elucidated that the Warburg effect is most likely allosterically regulated at HEX2 and PFKM nodes with the other enzymes in the EMP are regulated either through an increase in thermodynamic driving force or via an increase in saturation (Chapter 4).

Point estimates summarising the posterior space are insufficient to capture the true model behaviour. Inference methods that consider information summarised at a single point cannot account for the correlation in sloppy parameter systems, and therefore, have poor predictive performance. It is recommended that HMC sampling using best practice diagnostics be standard practice. Firstly, approximations based on point estimates give bad results as shown in Chapter 3. Finaly, while variational inference may work in some cases it can produce biased results and lacks diagnostics.

Maud provides a standardised platform for Bayesian kinetic modelling. Maud is the first general kinetic modelling software that implements adaptive HMC and is based on mechanistically sound, modular, and interpretable rate laws. The presented diagnostics of convergence and effective sample size allow the user to reject bad model runs without requiring expert knowledge. Maud lowers the barrier for constructing Bayesian kinetic models by Secondly, the automated generation of the system of ODEs complemented by Stan's HMC algorithm lowers the barrier for constructing Bayesian kinetic models. And finally, the posterior sampling itself removes the need for further statistical testing that may have implicit assumptions about the data, instead, quantiles of the posterior distribution are used for statistical inference.

Hexokinase 2 is a strong mediator glycolysis. Unlike previous kinetic models of the Warburg effect, we included regulation of HEX2 that proved to be essential for the regulation of glycolytic flux. HEX2 regulation has been previously unexplored, possibly due to the lack of models of phosphorylation and/or data collected for phosphoproteomics. We can now use the knowledge that regulation of HEX2 via phosphorylation is essential to determing the root causes of the Warburg effect in an effort to connect how and why the Warburg effect occurs.

High throughput multiomics is required for constructing kinetic models. As kinetic models are a representation of metabolism, a good model should predict behaviour at multiple states. As we presented in Chpater 4, the model was unable to predict the behaviour of TPI and GAPD, futhermore, the PP pathway was inadequately represented. These failings to achieve the observed model behaviour indicate where the model requires improvement, which may not be obvsered without a sufficient training and validation dataset.

Phosphorylation can be modelled using an extended MWC model. We have demonstrated that the MWC model can display cooperativity, and change the regulation sensitivity to occur in any region of phosphorylation fraction (Chapter 5). The phosphorylation model was designed with a minimal parameterisation in mind so we can include this model in the Maud framework with minimal impact on performance.

Maud will continue to be developed to improve user experience and computational efficiency. Maud is the first kinetic modelling software that implements efficient MCMC sampling. The next stage in Maud is to optimise the ODE solvers and parallelise the sampling, this will make Maud a viable kinetic modelling tool and improve it for larger models. Additionally, the phosphorylation mechanisms (Chapter 5) will be implemented to model PFKFB3 regulation by AMPK and HEX2 mediated mitochondrial binding via Akt. Including post-translational mechansism

6.3   Recommendations

Identifying modes of action for allosteric regulation. Despite identifying the nodes that can mediate glycolytic flux, the driving forces of these changes are yet to be explored. The next step is to expand the model definition to account for phophoryation mechanisms surrounding the HEX2 and PFKFB3 nodes so the underlying mechanisms that differentiates the high and low fluxes can be understood. The original hypothesis posed by Luengo *et al.,* 2021, is that the membrane potential prevents regeneration of nad+ and therefore requires lactate production to regenerate nad+ Luengo et al., 2021. However, in this study we have shown using a comparison of the CHO-S and CHO-ZeLa cell lines, that the Warburg effect can be eliminated and additional interactions are required outside of an increase in thermodynamic driving force (Chapters 2 and 4). Therefore, we are in search of a link between a change of redox state and protein interactions that influence HEX2 and PFKFB3 activity before the Warburg effect can be fully understood.

Proteomics, absolute metabolomics, and fluxomics are essential for understanding metabolism. irrespective of the model used, we have shown in this study that each of these -omes are dependent and contribute to flux. as any modelling approach is an attempt at describing the underlying mechanism, these -omes are required to undestand metabolism. it is recommended that research focused on metabolism should measure as much of the multiomics as possible. this will deconvolude interactions between the regulatory factors and provide a more complete picture of metabolism.

Integrating isotopic labelling into kinetic models. The current workflow considers that the internal flux measurements are independent. Unfortunately, the isotoptomer measurements are correlated with exchange fluxes, resulting in larger uncertainties when estimating fluxes. The inclusion of isotopic data into a kinetic model would rely solely on the isotoptomer measurements and are uncorrelated with the exchange fluxes. Including isoptomers into kinetic models benefits more than just the

error model. A common issue with 13C-MFA is the large number of unidentified exchange fluxes, overlaying a kinetic model onto isotopic data provides a direct estimate of the exchange flux. Ideally this will both improve flux estimates and reduce uncertainties in parameter estimates.

References

Städter, P., Schälte, Y., Schmiester, L., Hasenauer, J., & Stapor, P. L. (2021). Benchmarking of numerical integration methods for ODE models of biological systems. *Scientific Reports*, *11*(1), 2696. https://doi.org/10.1038/s41598-021-82196-2

Margossian, C. (2018). Computing Steady States With Stan'S Nonlinear Algebraic Solver. https://doi.org/10.5281/ZENODO.1284375

Cleland, W. W. (1989). The kinetics of enzyme-catalyzed reactions with two or more substrates or products. I. Nomenclature and rate equations. 1963. *Biochimica et Biophysica Acta*, *1000*, 213–220. Retrieved October 12, 2018, from https://www.ncbi.nlm.nih.gov/pubmed/2673369

Leskovac, V., Trivić, S., Peričin, D., & Kandrač, J. (2006). Deriving the rate equations for product inhibition patterns in bisubstrate enzyme reactions. *Journal of Enzyme Inhibition and Medicinal Chemistry*, *21*(6), 617–634. https://doi.org/10.1080/14756360600829381

Luengo, A., Li, Z., Gui, D. Y., Sullivan, L. B., Zagorulya, M., Do, B. T., Ferreira, R., Naamati, A., Ali, A., Lewis, C. A., Thomas, C. J., Spranger, S., Matheson, N. J., & Vander Heiden, M. G. (2021). Increased demand for NAD+ relative to ATP drives aerobic glycolysis. *Molecular Cell*, *81*(4), 691–707.e6. https://doi.org/10.1016/j.molcel.2020.12.012

A   Hierarchical model stan code

```
1   data {
2       int<lower=0> I; // number of fluxes
3       real flux[I]; // measured fluxes
4       real<lower=0> sigma[I]; // std of measured fluxes
5   }
6   transformed data {
7       real largest_measured_distance = max(flux) - min(flux);
8       real p10_max_flux = max(flux)*0.1;
9   }
10  parameters {
11      real mu0; // Biological Mean
12      vector[I] mu; // individual measured flux
13      real<lower=0> eta0_z; // Standardized Biological Standard Deviation
14  }
15  transformed parameters{
16      real eta0 = eta0_z*p10_max_flux + largest_measured_distance/2;
17  }
18  model {
19      mu ~ normal(mu0, eta0);
20      flux ~ normal(mu, sigma);
21      eta0_z ~ std_normal(); // eta0 ~ normal(largest_measured_distance/2,
            p10_max_flux)
22  }
```

B   Methionine cycle parameters

parameter name	prior quantile 1%	prior quantile 99%	true value	prior Z-score of true value
$e_R^{CBS1,amet_c}$	3.430000e-06	0.002480	9.300000e-05	0.004
$e_R^{GNMT1,amet_c}$	3.000000e-05	0.002000	2.000000e-05	-2.787
$e_R^{MAT3,amet_c}$	1.000000e-04	0.001000	3.170000e-04	0.003
$e_R^{MAT3,met-L_c}$	4.500000e-04	0.000800	6.000000e-04	0.000
$e_R^{MTHFR1,ahcys_c}$	1.120000e-07	0.000081	2.000000e-06	-0.101
$e_T^{GNMT1,mlthf_c}$	1.120000e-05	0.008050	2.290000e-04	-0.136
$e_T^{MTHFR1,amet_c}$	1.120000e-07	0.000081	1.500000e-05	0.791
k_{cat}^{AHC1}	1.200000e+02	400.000000	2.340000e+02	0.259
k_{cat}^{BHMT1}	6.000000e+00	35.000000	1.380000e+01	-0.135
k_{cat}^{CBS1}	1.000000e+01	188.000000	7.020000e+00	-2.887
k_{cat}^{GNMT1}	7.000000e-01	60.000000	1.050000e+01	0.507
k_{cat}^{MAT1}	8.200000e-02	59.100000	7.900000e+00	0.639
k_{cat}^{MAT3}	5.890000e-01	424.000000	1.990000e+01	0.116
$k_{cat}^{METH-Gen}$	4.840000e-01	349.000000	1.160000e+00	-1.209
k_{cat}^{MS1}	1.000000e+00	3.300000	1.770000e+00	-0.091
k_{cat}^{MTHFR1}	1.300000e+00	4.200000	3.170000e+00	1.204
k_{cat}^{PROT1}	1.590000e-01	0.222000	2.650000e-01	3.423
$K_I^{GNMT1,ahcys_c}$	2.000000e-06	0.001400	5.300000e-05	0.010
$K_I^{MAT1,amet_c}$	3.000000e-04	0.000400	3.470000e-04	0.014
$K_I^{METH-Gen,ahcys_c}$	1.000000e-06	0.000030	6.000000e-06	0.021
$K_M^{AHC1,ahcys_c}$	5.220000e-05	0.037600	2.320000e-05	-2.050
K_M^{AHC1,adn_c}	1.670000e-07	0.000120	5.660000e-06	0.118
$K_M^{AHC1,hcys-L_c}$	1.580000e-07	0.000114	1.060000e-05	0.458
$K_M^{BHMT1,hcys-L_c}$	1.200000e-05	0.000032	1.980000e-05	0.049
$K_M^{BHMT1,glyb_c}$	4.720000e-05	0.034000	8.460000e-03	0.949
$K_M^{CBS1,hcys-L_c}$	1.000000e-06	0.000025	4.240000e-05	3.090
$K_M^{CBS1,ser-L_c}$	2.000000e-06	0.000004	2.830000e-06	0.004
$K_M^{GNMT1,amet_c}$	1.300000e-05	0.009400	5.200000e-04	0.198
$K_M^{GNMT1,ahcys_c}$	4.100000e-07	0.000295	1.100000e-05	0.000
K_M^{GNMT1,gly_c}	5.480000e-05	0.039500	2.540000e-03	0.273
$K_M^{GNMT1,sarcs_c}$	3.730000e-09	0.000003	1.000000e-07	0.000
$K_M^{MAT1,met-L_c}$	1.400000e-05	0.000720	1.070000e-04	0.074
K_M^{MAT1,atp_c}	5.270000e-05	0.038000	2.030000e-03	0.181
$K_M^{MAT3,met-L_c}$	4.470000e-05	0.032200	1.130000e-03	-0.029
K_M^{MAT3,atp_c}	5.270000e-05	0.038000	2.370000e-03	0.258
$K_M^{METH-Gen,amet_c}$	7.000000e-06	0.000013	9.370000e-06	-0.135
$K_M^{MS1,5mthf_c}$	3.320000e-06	0.002390	6.940000e-05	-0.124
$K_M^{MS1,hcys-L_c}$	1.000000e-06	0.000003	1.710000e-06	-0.054
$K_M^{MTHFR1,mlthf_c}$	7.500000e-05	0.000088	8.080000e-05	-0.158
$K_M^{MTHFR1,nadph_c}$	1.600000e-05	0.000028	2.090000e-05	-0.105
$K_M^{PROT1,met-L_c}$	4.500000e-05	0.000085	4.390000e-05	-2.507
L_0^{CBS1}	3.730000e-02	26.800000	1.030000e+00	0.017
L_0^{GNMT1}	3.730000e-02	26.800000	1.310000e+02	2.438
L_0^{MAT3}	3.730000e-03	2.680000	1.080000e-01	0.037
L_0^{MTHFR1}	1.120000e-01	80.500000	3.920000e-01	-1.018

Table B.1: True parameter values and marginal prior distributions used for the methionine cycle case study

C CHO kinetic model definition

Hexokinase

Hexokinase 1
reaction: $glc - D + atp \leftrightarrow g6p + adp$
EC number: 2.7.1.1/2.7.1.2 Characterisation of this enzyme was described by White and Wilson, 1989 where the hexokinase isomer was isolated from rat brain. Rates were calculated using spectrophotometric glucose-6-phosphate dehydrogenase (G6PDH) assays. The enzyme exists as a dimer with a catalytically active C-terminus and non active N-terminus. In the study conducted by White and Wilson, 1989, the g6p analog 1,5-AnG6P was used to evaluate regulation of hexokinase. This was selected for as it allows for quantification using the G6PDH assay. Table C.1 shows the kinetic parameter priors used.

Table C.1: Hexokinase 1 kinetic parameters

parameter	substrate	value	units	regulation	references
k_{cat}		93.5	s^{-1}		White and Wilson, 1989
K_m	glc-D	$6e - 5$	M		White and Wilson, 1989
K_m	atp	$8.8e - 4$	M		White and Wilson, 1989
K_i	g6p	$5.5e - 5$	M	competitive	White and Wilson, 1989

Hexokinase 2
reaction: $glc - D + atp \leftrightarrow g6p + adp$
EC number: 2.7.1.1/2.7.1.2 Similar to hexokinase 1, hexokinase 2 is a dimer that is competitively inhibited by g6p. In contrast, HEX2 has two active dimers, which are described as allosteric in Mulukutla et al., 2014.

Table C.2: Hexokinase 2 kinetic parameters

parameter	substrate	value	units	regulation	references
k_{cat}		739 ± 0.14	s^{-1}		Ahn et al., 2009
k_{cat}		180	s^{-1}		Mulquiney and Kuchel, 1999
K_m	glc-D	$3.7e - 4 \pm 1.2e - 4$	M		Ahn et al., 2009
K_m	atp	$8.1e - 4 \pm 1.2e - 4$	M		Ahn et al., 2009
K_i	g6p	$2.4e - 4 \pm 6e - 5$	M	competitive	Ahn et al., 2009

Glucose-6-phosphate isomerase

Glucose-6-phosphate isomerase (GPI) is often modelled as a reversible Michaelis Menten process with no regulation Mulukutla et al., 2014. There are instances of competitive inhibition by 6-phosphogluconate (6pgc), however, this was only observed at concentrations of 6pgc in the mM range Parr, 1956, whereas CHO cells typically have concentrations on the order of μM. Hence, 6pgc inhibition is not considered in this model. The following binding constants are used to construct the priors.

Table C.3: Glucose-6-phosphate isomerase kinetic parameters

parameter	substrate	value	units	regulation	references
k_{cat}		3300	s^{-1}		Tsuboi et al., 1971
K_m	g6p	$2.7e-4$	M		Gracy and Tilley, 1975
K_m	g6p	$1.05e-4$	M		Kahana et al., 1960
K_m	f6p	$3.0e-5$	M		Kahana et al., 1960

Table C.4: Phosphofructokinase L kinetic parameters

parameter	substrate	value	units	regulation	references
k_{cat}		127	s^{-1}		Kagimoto and Uyeda, 1979
K_m	f6p	$7e-5$	M		Durante et al., 1995
K_m	f6p	$8e-5$	M		Brand and Söling, 1974
K_m	atp	$1e-4$	M		Brand and Söling, 1974
K_m	atp	$7e-5$	M		Durante et al., 1995
K_m	adp	$1.4e-4$	M		Brand and Söling, 1974
K_m	fdp	$4.3e-4$	M		Brand and Söling, 1974
L_0		$8-18.5$			Moreno-Sánchez et al., 201
e_R	f6p	same as Km		allosteric activation	
e_R	fdp	$2-20e-3$	M	allosteric activation	Durante et al., 1995
e_T	atp	$1-20e-6$	M	allosteric inhibition	Durante et al., 1995

Phosphofructokinase

reaction: $f6p + atp \leftrightarrow f16bp + adp$

EC number: 2.7.1.11

Phosphofructokinase L

Phosphofructokinase L (PFKL) is a tetramer that is allosterically regulated by a variety of molecules, most notably, atp and f16bp. Durante et al., 1995 identified that the L isoenzyme was insensitive to citrate concentrations even up to 10 mM, however, was regulated by cooperative binding of f6p and fdp. Additionally, atp was found to be an allosteric inhibitor of PFKL. The following binding constants were used as priors for the model. The mechanisms of PFKL is described as an ordered Bi Bi reaction using the notation defined by Cleland, 1973 and was identified Brand and Söling, 1974. However, in Maud mechanisms are defined using the modular rate law, and therefore lacks the generative property of substrate inhibition.

Phosphofructokinase M

Phosphofructokinase M (PFKM) is a tetramer that is allosterically regulated and is hypothesised to be the primary active form when exhibiting the Warburg effect. Tetramers are possible with the L isoform of PFK Brüser et al., 2012 however, this model will consider the isoforms as separate tetramers for simplicity. Similar to PFKL atp is an allosteric inhibitor and f6p is an allosteric activator. Unlike PFKL, f26bp is an allosteric activator of PFKM Zancan et al., 2007. The authors in Durante et al., 1995 identified that both PFK L and M isoenzymes were sensitive to fdp allosteric activation. Table C.5 defines the binding and rate constants used to define the priors in the model

Table C.5: Phosphofructokinase L kinetic parameters

parameter	substrate	value	units	regulation	references
k_{cat}		822	s^{-1}		Mulquiney and Kuchel, 1999
K_m	f6p	$7e-5$	M		Durante et al., 1995
K_m	atp	$3-90e-6$	M		Durante et al., 1995
K_m	adp	$1.4-2.2e-3$	M		Etiemble et al., 1977
K_m	fdp	$3.3-13.2e-3$	M		Etiemble et al., 1977
L_0		$18-25$			Moreno-Sánchez et al., 2011
e_R	f6p	same as Km		allosteric activation	
e_R	fdp	$2-20e-3$	M	allosteric activation	Durante et al., 1995
e_T	atp	$5e-3$	M	allosteric inhibition	Zancan et al., 2007
e_T	f26bp	$1-20e-6$	M	allosteric activation	Zancan et al., 2007

Fructosebisphosphate aldolase

reaction: *f16bp → g3p + dhap*
EC number: 4.1.2.13

Fructosebisphosphate aldolase (FBP) splits fructose 1,6-bisphosphate into glyceraldehyde 3-phosphate and dihydroxyacetone phosphate. The mechanism is defined as an ordered Uni Bi mechanism Cleland, 1973 by Sygusch and Beaudry, 1997. But will be defined as a modular rate law for the purposes of this model. The following binding and rate constants are used to define the priors in the model.

Table C.6: Glucose-6-phosphate isomerase kinetic parameters

parameter	substrate	value	units	regulation	references
k_{cat}		60	s^{-1}		Kusakabe et al., 1994
K_m	fdp	$5.2e-5$	M		Kusakabe et al., 1994
K_m	g3p	$1.9e-4$	M		Mulukutla et al., 2014
K_m	dhap	$3.5e-5$	M		Mulukutla et al., 2014

Triosephosphate isomerase

reaction: *g3p → dhap*
EC number: 2.7.2.3 Triosephosphate isomerase (TPI) is an enzyme that catalyses the reversible reaction between glyceraldehyde 3-phosphate and dihydroxyacetone phosphate.

Table C.7: Triosephosphate isomerase kinetic parameters

parameter	substrate	value	units	regulation	references
k_{cat}		3800	s^{-1}		Eber and Krietsch, 1980
K_m	g3p	$1.3e-5$	M		Eber and Krietsch, 1980
K_m	dhap	$1.8e-4$	M		Eber and Krietsch, 1980

Glyceraldehyde 3-phosphate dehydrogenase

reaction: *g3p + nad + pi → 13bpg + nadh*
EC number: 1.2.1.12

Glyceraldehyde 3-phosphate dehydrogenase (GAPDH) is a kinase that phosphorylates g3p and generates nadh to be used in the TCA cycle. The mechanism has previously been defined as an orderedTerBi mechanism in Wang and Alaupovic, 1980 in human erythrocytes. But, will be defined using the modular rate law for the purposes of this model. To account for the competitive substrate inhibition as a result of the mechanism, we will define g3p as a competitive inhibitor. The priors were defined in table C.8.

Table C.8: Glyceraldehyde 3-phosphate dehydrogenase kinetic parameters

parameter	substrate	value	units	regulation	references
k_{cat}		161	s^{-1}		Heinz and Freimüller, 1982
K_m	g3p	$9.5e-5$	M		Wang and Alaupovic, 1980
K_m	nad	$4.5e-5$	M		Wang and Alaupovic, 1980
K_m	pi	$7.8e-5$	M		Wang and Alaupovic, 1980
K_m	nadh	$3.3e-6$	M		Furfine and Velick, 1965
K_m	13bpg	$8e-7$	M		Furfine and Velick, 1965
K_i	13bpg	$3e-7$	M		Furfine and Velick, 1965
K_i	nadh	$1e-6$	M		Furfine and Velick, 1965
K_i	g3p	$3.1e-4$	M		Wang and Alaupovic, 1980

Phosphoglycerate kinase

reaction: *13dpg + adp → 3pg + atp*

EC number: 2.7.2.3

Phosphoglycerate kinase (PGK) is a reversible enzyme that catalyses the conversion between 13dpg and 3pg. The mechanism is proposed to be a random order mechanism with product binding as defined in Lee and O'Sullivan, 1975. No other inhibition was observed. The binding constants area defined in table C.9 the values were calculated using an enzyme located in rabbit muscle Krietsch and Bucher, 1970.

Table C.9: Phosphoglycerate kinase kinetic parameters

parameter	substrate	value	units	regulation	references
k_{cat}		430	s^{-1}		Lee and O'Sullivan, 1975
K_m	atp	$4.2e-4$	M		Lee and O'Sullivan, 1975
K_m	adp	$1.5e-4$	M		Krietsch and Bucher, 1970
K_m	3pg	$1.32e-3$	M		Krietsch and Bucher, 1970
K_m	13bpg	$1.9e-6$	M		Ali and Brownstone, 1976

Phosphoglycerate mutase

reaction: *3pg → 2pg*

EC number: 5.4.2.1 Phosphoglycerate mutase (PGM) is a reversible enzyme that catalyses the conversion between 2pg and 3pg. The mechanism was determined to be a ping-pong mechanism that requires binding from 2,3-bpg Grisolia and Cleland, 1968. To account for activation we will use the MWC formalism with 2,3-bpg as an activator and consider the interconversion as a uni-uni reaction. The priors for the model are

defined below.

Table C.10: Phosphoglycerate mutase kinetic parameters

parameter	substrate	value	units	regulation	references
k_{cat}		795	s^{-1}		Mulquiney and Kuchel, 1999
K_m	3pg	$(0.6 - 4.6)e - 3$	M		Krietsch and Bucher, 1970
K_m	3pg	$0.168e - 3$	M		Mulquiney and Kuchel, 1999
K_m	2pg	$14 - 25e - 6$	M		Mulquiney and Kuchel, 1999
L_0		1			default prior
e_R	2,3-bpg	$(1.4 - 42)e - 6$			Grisolia and Cleland, 1968

Enolase

reaction: $2pg \rightarrow h2o + pep$

EC number: 4.2.1.11 There are three isoenzymes of enolase, however, as shown in Rider and Taylor, 1973 there are minimal kinetic differences between enzymes and hence, will be considered as a single enolase for the purpose of this model. No regulation was noted for this enzyme. The priors are defined in Table C.11

Table C.11: Triosephosphate isomerase kinetic parameters

parameter	substrate	value	units	regulation	references
k_{cat}		57	s^{-1}		Kornblatt, 2005
K_m	2pg	$1.3e - 4$	M		Rider and Taylor, 1973
K_m	pep	$0.34e - 4$	M		Rider and Taylor, 1973

Pyruvate kinase

reaction: $pep + adp \rightarrow atp + pyr$

EC number: 2.7.1.40 Pyruvate has two isoforms expressed across the analysed cell lines, pyruvate kinase M1 (PKM1) and pyruvate kinase M2 (PKM2). Based on proteomics conducted throughout this thesis the LR isoform was not expressed at a significant level.

Pyruvate kinase M1

pyruvate kinase M1 (PKM1) is the most abundant isoform of pyruvate kinase in CHO cells as indicated by proteomics. The kinetic properties of PKM1 do not exhibit the allostery found in the other pyruvate kinases towards fdp or pep Ikeda et al., 2000.

Table C.12: Pyruvate kinase M1 kinetic parameters

parameter	substrate	value	units	regulation	references
k_{cat}		627	s^{-1}		Ikeda et al., 2000
K_m	pep	$5.8e - 5$	M		Ikeda et al., 2000
K_m	adp	$5.6e - 4$	M		Ikeda et al., 2000

Pyruvate kinase M2

PKM2 is an allosteric enzyme active in its tetrameric state Gao et al., 2012. and is primarily activated by fdp. There is strong inhibition by phenalanine and moderate inhibition by serine Macpherson et al., 2018. However, at physiological concentrations

of fdp, serine inhibition is almost completely negated by fdp activation. Macpherson et al., 2018.

Table C.13: Pyruvate Kinase M2 kinetic parameters

parameter	substrate	value	units	regulation	references
k_{cat}		350	s^{-1}		Macpherson et al., 2018
k_{cat}		570	s^{-1}		Ikeda and Noguchi, 1998
K_m	pep	$2.3e-4$	M		Macpherson et al., 2018
K_m	pep	$5e-5$	M		Ikeda and Noguchi, 1998
K_m	adp	$3.2e-4$	M		Ikeda and Noguchi, 1998
L_0		0.66			Macpherson et al., 2018
e_T	phe	$1.91e-4$			Macpherson et al., 2018
e_R	fdp	$25e-8$			Macpherson et al., 2018
e_R	pep	$2e-4$	M		Ikeda and Noguchi, 1998

Lactate dehydrogenase A

reaction: *pyr + nadh → nad + lac*

EC number: 1.1.1.27 Lactate dehydrogenase A (LDHA) converts the reversible reaction of pyruvate to lactate. It is known to have a tetrameric quaternary structure but exhibits no allostry Valvona et al., 2016. The enzyme is inhibited by pyruvate in a competitive

Table C.14: Laactate dehydrogenase A kinetic parameters

parameter	substrate	value	units	regulation	references
k_{cat}		277	s^{-1}		Jackson et al., 2016
K_m	pyr	$2.1e-4$	M		Zewe and Fromm, 1965
K_m	pyr	$7.8e-4$	M		Talaiezadeh et al., 2015
K_m	nadh	$7.4e-6$	M		Zewe and Fromm, 1965
K_i	nadh	$3e-4$	M		Talaiezadeh et al., 2015
K_m	nad	$1.7e-4$	M		Zewe and Fromm, 1965
K_m	nad	$5e-4$	M		Talaiezadeh et al., 2015
K_m	lac	$1.09e-2$	M		Zewe and Fromm, 1965
K_m	lac	$1.07e-2$	M		Talaiezadeh et al., 2015
K_i	pyr	$3.3e-3$	M	competitive	Jackson et al., 2016

Glucose-6-phosphate dehydrogenase

reaction: *g6p + nadp → nadph + 6pgl*

EC number: 1.1.1.49

Glucose-6-phosphate dehydrogenase (G6PD) is a tetrameric enzyme that represents the first step in the oxidative Pentose Phosphate Pathway (oxPPP). branch. Studies investigating the structure of G6PD in erythrocytes found the enzyme to exist as a monomer, dimer, and tetramer configuration with both the tetramer and dimer being active forms Bonsignore et al., 1971. As the tetramer is a dimer of dimers, allostery will be applied assuming a 2 subunit system that is activated by nadp Bonsignore et al., 1971. The mechanism was determined by Kanji et al., 1976 to be an ordered mecha-

nism with the cofactor binding first and released last. As a consequence of Maud considering only random binding mechanisms the saturation behaviour of G6PD cannot capture the substrate inhibition from g6p. Furthermore, atp and adp were also found to be inhibitors of G6PD Kanji et al., 1976.

Table C.15: g6pdh kinetic parameters

parameter	substrate	value	units	regulation	references
k_{cat}			s^{-1}		
K_m	g6p	$3.6e-5$	M		Kanji et al., 1976
K_m	nadp	$(4.8-9.0)e-6$	M		Kanji et al., 1976
L_0		1			default prior
e_R	nadp	$1.5e-4$			Bonsignore et al., 1971
K_i	adp	$1.9e-3$	M	competitive	Kanji et al., 1976
K_i	atp	$(4.5-7.2)e-3$	M	competitive	Kanji et al., 1976
K_i	g6p	$4.4e-3$	M	competitive	Kanji et al., 1976

6-phosphogluconate lactonase

reaction: *6pgl + h2o → 6pgc + h*

EC number: 3.1.1.31

6-phosphogluconate lactonase (6PGl) catalyses the irreversible hydrolysis of 6pgl to 6pgc. This enzyme has been found to exhibit no regulation and will be considered as a simple Michaelis-Menten enzyme Bauer et al., 1983. The parameters used as priors for this model are defined in the table C.16.

Table C.16: 6pgl kinetic parameters

parameter	substrate	value	units	regulation	references
k_{cat}		28.52	s^{-1}		Bauer et al., 1983
K_m	6pgl	$8e-5$	M		Schofield and Sols, 1976

Phosphogluconate dehydrogenase

reaction: *6pgc + nadp → ru5p-D + nadph + co2*

EC number: 1.1.1.44

Phosphogluconate dehydrogenase (GND) catalyses the reversible conversion to ru5p-D and generates the reduced cofactor napdh. GND is a dimer that exhibits no allosteric interactions, however is inhibited by nadph competitively Topham et al., 1986.

Ribose-5-phosphate isomerase

reaction: *ru5p-D → r5p*

EC number: 5.3.1.6

Ribose-5-phosphate isomerase (RPI) catalyses the reversible conversion of ru5p-D to r5p. This enzyme is a dimer that exhibits no allosteric interactions Jung et al., 2000. The parameters used as priors for this model are defined in table C.18.

A kinetic investigation of the Warburg effect in CHO cells

Table C.17: gnd kinetic parameters

parameter	substrate	value	units	regulation	references
k_{cat}		16.7	s^{-1}		Topham et al., 1986
k_{ca}		33.3	s^{-1}		Mcintyre et al., 1989
K_m	6pgc	$2.27e-5$	M		Topham et al., 1986
K_m	nadp	$2.9e-6$	M		Topham et al., 1986
K_m	nadph	$3e-7$	M	competitive	Villet and Dalziel, 1969
K_m	co2	$1.5e-2$	M	competitive	Villet and Dalziel, 1969
K_m	ru5p-D	$3e-5$	M	competitive	Villet and Dalziel, 1969
K_i	nadph	$1.1e-6$	M	competitive	Topham et al., 1986

Table C.18: rpi kinetic parameters

parameter	substrate	value	units	regulation	references
k_{cat}		37.5	s^{-1}		Urivetzky and Tsuboi, 1963
k_{cat}		3440	s^{-1}		Jung et al., 2000
K_m	r5p	$2.2e-3$	M		Urivetzky and Tsuboi, 1963
K_m	r5p	$6.3e-4$	M		Jung et al., 2000
K_m	ru5p-D	$(2.6-10)e-3$	M		Kiely et al., 1973

Ribulose-5-phosphate 3-epimerase

reaction: $ru5p\text{-}D \rightarrow xu5p\text{-}D$
EC number: 5.1.3.1 Ribulose-5-phosphate 3-epimerase (RPE) is part of the non-oxidative pentose phosphate pathway catalysing the interconversion of ru5p-D and r5p. RPE is a homodimer that exhibits cooperativity with respect to ru5p-D Karmali et al., 1983.

Table C.19: rpe kinetic parameters

parameter	substrate	value	units	regulation	references
k_{cat}		120	s^{-1}		Wood, 1974
K_m	ru5p-D	$1.9e-4$	M		Wood, 1979
K_m	ru5p-D	$2.9e-3$	M		Wood, 1974
K_m	xu5p-D	$1.5e-4$			Wood, 1974
L_0		1			default prior
e_R	ru5p-D	$1.9e-4$			default prior

Transketolase

Transketolase (TKT1)

reaction: $r5p + xu5p\text{-}D \rightarrow g3p + s7p$
EC number: 5.1.3.1

Transketolase 1 (TKT1) is part of the non-oxidative PPP and acts promiscuously with Transketolase 2 (TKT2). Promiscuity is not currently available in Maud and therefore, each reaction will be modelled separately using the measured enzyme concentration. Apart from promiscuity there is no additional regulation between these enzymes. The parameters used as priors for this model are defined in table C.20.

Table C.20: tkt1 kinetic parameters

parameter	substrate	value	units	regulation	references
k_{cat}		2.8	s^{-1}		Paoletti, 1983
k_{cat}		$13 - 18.5$	s^{-1}		Schenk et al., 1998a
K_m	r5p	$6.6e - 5$	M		Paoletti, 1983
K_m	r5p	$3.0e - 4$	M		Warnock and Prudhomme, 1982
K_m	xu5p-D	$2.5e - 5$	M		Paoletti, 1983
K_m	xu5p-D	$1.8e - 4$			Warnock and Prudhomme, 1982
K_m	s7p	$4e - 3$	M		Sprenger et al., 1995
K_m	g3p	$(2.1 - 4.3)e - 3$	M		Schenk et al., 1998b

Transketolase (TKT2)
reaction: *e4p + xu5p-D→ g3p + f6p*
EC number: 5.1.3.1

Transketolase 2 (TKT2) is part of the non-oxidative PPP and acts promiscuously with Transketolase 2 (TKT1). Promiscuity is not currently available in Maud and therefore, each reaction will be modelled separately using the measured enzyme concentration. The parameters used as priors for this model are defined in table C.21.

Table C.21: tkt2 kinetic parameters

parameter	substrate	value	units	regulation	references
k_{cat}		69	s^{-1}		Nilsson et al., 1998
K_m	e4p	$3.6e - 4$	M		Schenk et al., 1998b
K_m	f6p	$3.40e - 4$	M		Meshalkina et al., 2011
K_m	xu5p-D	$2.5e - 5$	M		Paoletti, 1983
K_m	xu5p-D	$1.8e - 4$			Warnock and Prudhomme, 1982
K_m	g3p	$(2.1 - 4.3)e - 3$	M		Schenk et al., 1998b

Transaldolase
reaction: *f6p + e4p → g3p + s7p*
EC number: 2.2.1.2

Table C.22: rpe kinetic parameters

parameter	substrate	value	units	regulation	references
k_{cat}		18	s^{-1}		Schneider et al., 2008
K_m	f6p	$2.3e - 3$			Schneider et al., 2008
K_m	e4p				
K_m	g3p				
K_m	s7p				

References

White, T. K., & Wilson, J. E. (1989). Isolation and characterization of the discrete N- and C-terminal halves of rat brain hexokinase: Retention of full catalytic activity in the isolated C-terminal half. *Archives of Biochemistry and Biophysics*, *274*(2), 375–393. Retrieved February 14, 2019, from https://www.ncbi.nlm.nih.gov/pubmed/2802617

Mulukutla, B. C., Yongky, A., Daoutidis, P., & Hu, W.-S. (2014). Bistability in Glycolysis Pathway as a Physiological Switch in Energy Metabolism (P. Dzeja, Ed.). *PLoS ONE*, *9*(6), e98756. https://doi.org/10.1371/journal.pone.0098756

Ahn, K. J., Kim, J., Yun, M., Park, J. H., & Lee, J. D. (2009). Enzymatic properties of the N- and C-terminal halves of human hexokinase II. *BMB Reports*, *42*(6), 350–355. Retrieved February 15, 2019, from https://www.ncbi.nlm.nih.gov/pubmed/19558793

Mulquiney, P. J., & Kuchel, P. W. (1999). Model of 2,3-bisphosphoglycerate metabolism in the human erythrocyte based on detailed enzyme kinetic equations: Equations and parameter refinement. *The Biochemical Journal*, *342 Pt 3*, 581–596.

Parr, C. W. (1956). Inhibition of phosphoglucose isomerase. *Nature*, *178*(4547), 1401–1401. https://doi.org/10.1038/1781401a0

Tsuboi, K. K., Fukunaga, K., & Chervenka, C. H. (1971). Phosphoglucose Isomerase from Human Erythrocyte. *Journal of Biological Chemistry*, *246*(24), 7586–7594. https://doi.org/10.1016/S0021-9258(19)45817-1

Gracy, R. W., & Tilley, B. E. (1975). [84] Phosphoglucose isomerase of human erythrocytes and cardiac tissue. In *Methods in Enzymology* (pp. 392–400). Elsevier. https://doi.org/10.1016/S0076-6879(75)41086-2

Kahana, S. E., Lowry, O. H., Schulz, D. W., Passonneau, J. V., & Crawford, E. J. (1960). The kinetics of phosphoglucoisomerase. *The Journal of Biological Chemistry*, *235*, 2178–2184.

Durante, A., Raleigh, X., Gomez, M., Campos, G., & Ryder, E. (1995). Isozyme Analysis of Human Normal Polymorphonuclear Leukocyte Phosphofructokinase. *Biochemical and Biophysical Research Communications*, *216*(3), 898–905. https://doi.org/10.1006/bbrc.1995.2706

Cleland, W. W. (1973). Derivation of rate equations for multisite ping-pong mechanisms with ping-pong reactions at one or more sites. *The Journal of Biological Chemistry*, *248*(24), 8353–8355. Retrieved January 19, 2019, from https://www.ncbi.nlm.nih.gov/pubmed/4762915

Brand, I. A., & Söling, H. D. (1974). Rat liver phosphofructokinase. Purification and characterization of its reaction mechanism. *The Journal of Biological Chemistry*, *249*(24), 7824–7831.

Kagimoto, T., & Uyeda, K. (1979). Hormone-stimulated phosphorylation of liver phosphofructokinase in vivo. *Journal of Biological Chemistry*, *254*(13), 5584–5587. https://doi.org/10.1016/S0021-9258(18)50449-X

Moreno-Sánchez, R., Marín-Hernández, A., Gallardo-Pérez, J. C., Quezada, H., Encalada, R., Rodríguez-Enríquez, S., & Saavedra, E. (2011). Phosphofructokinase type 1 kinetics, isoform expression and gene polymorphisms in cancer cells. *Journal of Cellular Biochemistry*, n/a–n/a. https://doi.org/10.1002/jcb.24039

Brüser, A., Kirchberger, J., Kloos, M., Sträter, N., & Schöneberg, T. (2012). Functional linkage of adenine nucleotide binding sites in mammalian muscle 6-phosphofructokinase. *The Journal of Biological Chemistry*, *287*(21), 17546–17553. https://doi.org/10.1074/jbc.M112.347153

Zancan, P., Almeida, F. V. R., Faber-Barata, J., Dellias, J. M., & Sola-Penna, M. (2007). Fructose-2,6-bisphosphate counteracts guanidinium chloride-, thermal-, and ATP-induced dissociation of skeletal muscle key glycolytic enzyme 6-phosphofructo-1-kinase: A structural mechanism for PFK allosteric regulation. *Archives of Biochemistry and Biophysics*, *467*(2), 275–282. https://doi.org/10.1016/j.abb.2007.08.032

Etiemble, J., Picat, C., & Boivin, P. (1977). Reaction mechanism of erythrocyte phosphofructokinase. *Biochimie*, *59*(8-9), 673–678. https://doi.org/10.1016/S0300-9084(77)80245-9

Sygusch, J., & Beaudry, D. (1997). Subunit interaction in mammalian aldolases. *The Biochemical Journal*, *323 (Pt 3)*, 671–676. Retrieved October 19, 2018, from https://www.ncbi.nlm.nih.gov/pubmed/9169599

Kusakabe, T., Motoki, K., & Hori, K. (1994). Human Aldolase C: Characterization of the Recombinant Enzyme Expressed in Escherichia coli1. *The Journal of Biochemistry*, *115*(6), 1172–1177. https://doi.org/10.1093/oxfordjournals.jbchem.a124475

Eber, S. W., & Krietsch, W. K. (1980). The isolation and characterization of the multiple forms of human skeletal muscle triosephosphate isomerase. *Biochimica et Biophysica Acta (BBA) - Enzymology*, *614*(1), 173–184. https://doi.org/10.1016/0005-2744(80)90178-3

Wang, C. S., & Alaupovic, P. (1980). Glyceraldehyde-3-phosphate dehydrogenase from human erythrocyte membranes. Kinetic mechanism and competitive substrate inhibition by glyceraldehyde 3-phosphate. *Archives of Biochemistry and Biophysics*, *205*(1), 136–145. Retrieved April 4, 2019, from https://www.ncbi.nlm.nih.gov/pubmed/7447472

Heinz, F., & Freimüller, B. (1982). [52] Glyceraldehyde-3-phosphate dehydrogenase from human tissues. In *Methods in Enzymology* (pp. 301–305). Elsevier. https://doi.org/10.1016/S0076-6879(82)89054-X

Furfine, C. S., & Velick, S. F. (1965). The Acyl-enzyme Intermediate and the Kinetic Mechanism of the Glyceraldehyde 3-Phosphate Dehydrogenase Reaction. *Journal of Biological Chemistry*, *240*(2), 844–855. https://doi.org/10.1016/S0021-9258(17)45252-5

Lee, C. S., & O'Sullivan, W. J. (1975). Properties and mechanism of human erythrocyte phosphoglycerate kinase. *Journal of Biological Chemistry, 250*(4), 1275–1281. https://doi.org/10.1016/S0021-9258(19)41810-3

Krietsch, W. K. G., & Bucher, T. (1970). 3-Phosphoglycerate Kinase from Rabbit Sceletal Muscle and Yeast. *European Journal of Biochemistry, 17*(3), 568–580. https://doi.org/10.1111/j.1432-1033.1970.tb01202.x

Ali, M., & Brownstone, Y. S. (1976). A study of phosphoglycerate kinase in human erythrocytes. II. Kinetic properties. *Biochimica et Biophysica Acta, 445*(1), 89–103. Retrieved February 17, 2019, from https://www.ncbi.nlm.nih.gov/pubmed/953034

Grisolia, S., & Cleland, W. W. (1968). Influence of Salt, Substrate, and Cofactor Concentrations on the Kinetic and Mechanistic Behavior of Phosphoglycerate Mutase. *Biochemistry, 7*(3), 1115–1121.

Rider, C. C., & Taylor, C. B. (1973). Isoenzymes of enolase in rat tissues. *Biochemical Society Transactions, 1*(5), 1127–1128. https://doi.org/10.1042/bst0011127

Kornblatt, M. J. (2005). Changing the metal ion selectivity of rabbit muscle enolase by mutagenesis: Effects of the G37A and G41A mutations. *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics, 1748*(1), 20–25. https://doi.org/10.1016/j.bbapap.2004.11.020

Ikeda, Y., Taniguchi, N., & Noguchi, T. (2000). Dominant Negative Role of the Glutamic Acid Residue Conserved in the Pyruvate Kinase M1 Isozyme in the Heterotropic Allosteric Effect Involving Fructose-1,6-bisphosphate. *Journal of Biological Chemistry, 275*(13), 9150–9156. https://doi.org/10.1074/jbc.275.13.9150

Gao, X., Wang, H., Yang, J. J., Liu, X., & Liu, Z.-R. (2012). Pyruvate kinase M2 regulates gene transcription by acting as a protein kinase. *Molecular Cell, 45*(5), 598–609. https://doi.org/10.1016/j.molcel.2012.01.001

Macpherson, J. A., Theisen, A., Masino, L., Fets, L., Driscoll, P. C., Encheva, V., Snijders, A. P., Martin, S. R., Kleinjung, J., Barran, P. E., Fraternali, F., & Anastasiou, D. (2018). Functional cross-talk between allosteric effects of activating and inhibiting ligands underlies PKM2 regulation. *BioRxiv*. https://doi.org/10.1101/378133

Ikeda, Y., & Noguchi, T. (1998). Allosteric Regulation of Pyruvate Kinase M2 Isozyme Involves a Cysteine Residue in the Intersubunit Contact. *Journal of Biological Chemistry, 273*(20), 12227–12233. https://doi.org/10.1074/jbc.273.20.12227

Valvona, C. J., Fillmore, H. L., Nunn, P. B., & Pilkington, G. J. (2016). The regulation and function of lactate dehydrogenase A: Therapeutic potential in brain tumor. *Brain Pathology, 26*(1), 3–17. https://doi.org/10.1111/bpa.12299

Jackson, E., López-Gallego, F., Guisan, J., & Betancor, L. (2016). Enhanced stability of l -lactate dehydrogenase through immobilization engineering. *Process Biochemistry, 51*(9), 1248–1255. https://doi.org/10.1016/j.procbio.2016.06.001

Zewe, V., & Fromm, H. J. (1965). Kinetic Studies of Rabbit Muscle Lactate Dehydroge-
nase. II. Mechanism of the Reaction [*]. *Biochemistry*, *4*(4), 782–792. https://doi.
org/10.1021/bi00880a024

Talaiezadeh, A., Shahriari, A., Tabandeh, M. R., Fathizadeh, P., & Mansouri, S. (2015).
Kinetic characterization of lactate dehydrogenase in normal and malignant hu-
man breast tissues. *Cancer Cell International*, *15*(1), 19. https://doi.org/10.1186/
s12935-015-0171-7

Bonsignore, A., Cancedda, R., Nicolini, A., Damiani, G., & De Flora, A. (1971). Metabolism
of human erythrocyte glucose-6-phosphate dehydrogenase. *Archives of Biochem-
istry and Biophysics*, *147*(2), 493–501. https://doi.org/10.1016/0003-9861(71)
90406-1

Kanji, M. I., Toews, M. L., & Carper, W. R. (1976). A kinetic study of glucose-6-phosphate
dehydrogenase. *The Journal of Biological Chemistry*, *251*(8), 2258–2262.

Bauer, H. P., Srihari, T., Jochims, J. C., & Hofer, H. W. (1983). 6-phosphogluconolactonase.
Purification, properties and activities in various tissues. *European Journal of Bio-
chemistry / FEBS*, *133*(1), 163–168. https://doi.org/10.1111/j.1432-1033.1983.
tb07442.x

Schofield, P., & Sols, A. (1976). Rat liver 6-phosphogluconolactonase: A low Km en-
zyme. *Biochemical and Biophysical Research Communications*, *71*(4), 1313–1318.
https://doi.org/10.1016/0006-291X(76)90798-1

Topham, C. M., Matthews, B., & Dalziel, K. (1986). Kinetic studies of 6-phosphogluconate
dehydrogenase from sheep liver. *European Journal of Biochemistry / FEBS*, *156*(3),
555–567. Retrieved March 6, 2019, from https://www.ncbi.nlm.nih.gov/
pubmed/3699023

Mcintyre, L. M., Thorburn, D. R., Bubb, W. A., & Kuchel, P. W. (1989). Comparison of
computer simulations of the F-type and L-type non-oxidative hexose monophos-
phate shunts with 31P-NMR experimental data from human erythrocytes. *Eu-
ropean Journal of Biochemistry*, *180*(2), 399–420. https://doi.org/10.1111/j.1432-
1033.1989.tb14662.x

Villet, R. H., & Dalziel, K. (1969). The nature of carbon dioxide substrate and equilib-
rium constant of the 6-phosphogluconate dehydrogenase reaction. *Biochemical
Journal*, *115*(4), 633–638. https://doi.org/10.1042/bj1150633

Jung, C. H., Hartman, F. C., Lu, T. Y., & Larimer, F. W. (2000). D-ribose-5-phosphate
isomerase from spinach: Heterologous overexpression, purification, character-
ization, and site-directed mutagenesis of the recombinant enzyme. *Archives of
Biochemistry and Biophysics*, *373*(2), 409–417. https://doi.org/10.1006/abbi.
1999.1554

Urivetzky, M., & Tsuboi, K. K. (1963). Enzymes of the human erythrocyte. V. Pentose
phosphate isomerase, purification and properties. *Archives of Biochemistry and
Biophysics*, *103*(1), 1–8. https://doi.org/10.1016/0003-9861(63)90002-X

Kiely, M. E., Stuart, A. L., & Wood, T. (1973). Partial purification and kinetic properties of ribose-5-phosphate ketol-isomerase and ribulose-5-phosphate 3-epimerase from various sources. *Biochimica et Biophysica Acta (BBA) - Enzymology*, *293*(2), 534–541. https://doi.org/10.1016/0005-2744(73)90360-4

Karmali, A., Drake, A. F., & Spencer, N. (1983). Purification, properties and assay of D-ribulose 5-phosphate 3-epimerase from human erythrocytes. *The Biochemical Journal*, *211*(3), 617–623. https://doi.org/10.1042/bj2110617

Wood, T. (1974). A spectrophotometric procedure for following the d-ribulose 5-phosphate 3-epimerase reaction in the direction of ribulose 5-phosphate formation. *Analytical Biochemistry*, *62*(2), 509–513. https://doi.org/10.1016/0003-2697(74)90183-3

Wood, T. (1979). Purification and properties of d-ribulose-5-phosphate 3-epimerase from calf liver. *Biochimica et Biophysica Acta (BBA) - Enzymology*, *570*(2), 352–362. https://doi.org/10.1016/0005-2744(79)90155-4

Paoletti, F. (1983). Purification and properties of transketolase from fresh rat liver. *Archives of Biochemistry and Biophysics*, *222*(2), 489–496. https://doi.org/10.1016/0003-9861(83)90547-7

Schenk, G., Duggleby, R. G., & Nixon, P. F. (1998a). Heterologous expression of human transketolase. *The International Journal of Biochemistry & Cell Biology*, *30*(3), 369–378. https://doi.org/10.1016/S1357-2725(97)00154-4

Warnock, L., & Prudhomme, C. (1982). The isolation and preliminary characterization of apotransketolase from human erythrocytes. *Biochemical and Biophysical Research Communications*, *106*(3), 719–723. https://doi.org/10.1016/0006-291X(82)91770-3

Sprenger, G. A., Schorken, U., Sprenger, G., & Sahm, H. (1995). Transketolase a of Escherichia coli K12. Purification and Properties of the Enzyme from Recombinant Strains. *European Journal of Biochemistry*, *230*(2), 525–532. https://doi.org/10.1111/j.1432-1033.1995.0525h.x

Schenk, G., Duggleby, R. G., & Nixon, P. F. (1998b). Properties and functions of the thiamin diphosphate dependent enzyme transketolase. *The International Journal of Biochemistry & Cell Biology*, *30*(12), 1297–1318. https://doi.org/10.1016/S1357-2725(98)00095-8

Nilsson, U., Hecquet, L., Gefflaut, T., Guerard, C., & Schneider, G. (1998). Asp [477] is a determinant of the enantioselectivity in yeast transketolase. *FEBS Letters*, *424*(1-2), 49–52. https://doi.org/10.1016/S0014-5793(98)00136-7

Meshalkina, L. E., Solovjeva, O. N., & Kochetov, G. A. (2011). Interaction of transketolase from human tissues with substrates. *Biochemistry (Moscow)*, *76*(9), 1061–1064. https://doi.org/10.1134/S0006297911090112

Schneider, S., Sandalova, T., Schneider, G., Sprenger, G. A., & Samland, A. K. (2008). Replacement of a Phenylalanine by a Tyrosine in the Active Site Confers Fructose-6-phosphate Aldolase Activity to the Transaldolase of Escherichia coli and Hu-

man Origin. *Journal of Biological Chemistry*, *283*(44), 30064–30072. https://doi.org/10.1074/jbc.M803184200

D   13C-MFA input

Table D.1: Measured intracellular mass isotopomer distributions for defined metabolites

experiment	replicate	met_id	$\mu\, m_0$	$\sigma\, m_0$	$\mu\, m_1$	$\sigma\, m_1$	$\mu\, m_2$	$\sigma\, m_2$	$\mu\, m_3$	$\sigma\, m_3$
CHO-S wt	A	3pg	0.781	0.020	0.054	0.020	0.160	0.021		
		6pgc	0.446	0.036	0.157	0.020	0.360	0.033	0.037	0.020
		dhap	0.617	0.020	0.060	0.020	0.317	0.020		
		f6p	0.425	0.026	0.079	0.020	0.452	0.043	0.044	0.020
		fdp	0.488	0.020	0.085	0.020	0.388	0.020	0.040	0.020
		g3p	0.665	0.020	0.064	0.020	0.264	0.020		
		g6p	0.444	0.040	0.081	0.020	0.440	0.055	0.035	0.020
		pep	0.764	0.049	0.046	0.020	0.182	0.043		
		ru5p-D	0.541	0.020	0.191	0.020	0.235	0.020	0.033	0.020
		s7p	0.409	0.083	0.177	0.020	0.316	0.054	0.098	0.020
	B	3pg	0.693	0.020	0.047	0.020	0.254	0.034		
		6pgc	0.376	0.030	0.145	0.020	0.439	0.041	0.040	0.020
		dhap	0.567	0.020	0.042	0.020	0.383	0.020		
		f6p	0.363	0.022	0.062	0.020	0.531	0.051	0.044	0.020
		fdp	0.428	0.020	0.071	0.020	0.456	0.020	0.045	0.020
		g3p	0.570	0.020	0.046	0.020	0.375	0.027		
		g6p	0.379	0.034	0.062	0.020	0.522	0.065	0.038	0.020
		pep	0.679	0.043	0.039	0.020	0.274	0.064		
		ru5p-D	0.443	0.020	0.256	0.020	0.247	0.020	0.055	0.020
		s7p	0.297	0.061	0.170	0.020	0.412	0.070	0.121	0.020
	C	3pg	0.783	0.020	0.062	0.020	0.149	0.020		
		6pgc	0.414	0.033	0.152	0.020	0.391	0.036	0.042	0.020
		dhap	0.646	0.020	0.064	0.020	0.279	0.020		
		f6p	0.397	0.025	0.084	0.020	0.472	0.045	0.046	0.020
		fdp	0.459	0.020	0.089	0.020	0.407	0.020	0.044	0.020
		g3p	0.624	0.020	0.072	0.020	0.296	0.021		
		g6p	0.414	0.037	0.085	0.020	0.458	0.057	0.043	0.020
		pep	0.755	0.048	0.051	0.020	0.186	0.044		
		ru5p-D	0.465	0.020	0.230	0.020	0.263	0.020	0.042	0.020
		s7p	0.313	0.064	0.194	0.020	0.376	0.064	0.117	0.020
	D	3pg	0.791	0.021	0.058	0.020	0.146	0.020		
		6pgc	0.461	0.037	0.143	0.020	0.357	0.033	0.039	0.020
		dhap	0.684	0.020	0.064	0.020	0.245	0.020		
		f6p	0.428	0.026	0.082	0.020	0.451	0.043	0.039	0.020
		fdp	0.485	0.020	0.082	0.020	0.392	0.020	0.040	0.020

experiment	replicate	met_id	$\mu\,m_0$	$\sigma\,m_0$	$\mu\,m_1$	$\sigma\,m_1$	$\mu\,m_2$	$\sigma\,m_2$	$\mu\,m_3$	$\sigma\,m_3$
		g3p	0.716	0.020	0.052	0.020	0.225	0.020		
		g6p	0.459	0.041	0.074	0.020	0.433	0.054	0.034	0.020
		pep	0.786	0.050	0.056	0.020	0.151	0.035		
		ru5p-D	0.434	0.020	0.231	0.020	0.298	0.020	0.037	0.020
		s7p	0.374	0.076	0.177	0.020	0.353	0.060	0.096	0.020
	A	3pg	0.763	0.020	0.049	0.020	0.185	0.025		
		6pgc	0.527	0.042	0.157	0.020	0.285	0.026	0.030	0.020
		dhap	0.759	0.020	0.051	0.020	0.186	0.020		
		f6p	0.525	0.032	0.069	0.020	0.378	0.036	0.027	0.020
		fdp	0.409	0.020	0.150	0.020	0.382	0.020	0.059	0.020
		g3p	0.773	0.020	0.048	0.020	0.175	0.020		
		g6p	0.523	0.047	0.073	0.020	0.378	0.047	0.026	0.020
		pep	0.817	0.052	0.044	0.020	0.137	0.032		
		ru5p-D	0.515	0.020	0.137	0.020	0.320	0.020	0.029	0.020
		s7p	0.423	0.086	0.181	0.020	0.316	0.053	0.081	0.020
	B	3pg	0.783	0.020	0.057	0.020	0.156	0.021		
		6pgc	0.523	0.042	0.118	0.020	0.323	0.030	0.036	0.020
		dhap	0.758	0.020	0.059	0.020	0.179	0.020		
		f6p	0.489	0.030	0.075	0.020	0.405	0.039	0.031	0.020
		fdp	0.402	0.020	0.166	0.020	0.368	0.020	0.064	0.020
		g3p	0.778	0.020	0.052	0.020	0.166	0.020		
		g6p	0.496	0.044	0.081	0.020	0.390	0.049	0.033	0.020
		pep	0.817	0.052	0.048	0.020	0.131	0.031		
CHO-ZeLa		ru5p-D	0.539	0.020	0.129	0.020	0.299	0.020	0.033	0.020
		s7p	0.412	0.084	0.150	0.020	0.353	0.060	0.085	0.020
	C	3pg	0.780	0.020	0.049	0.020	0.167	0.022		
		6pgc	0.556	0.044	0.124	0.020	0.291	0.027	0.028	0.020
		dhap	0.772	0.020	0.053	0.020	0.171	0.020		
		f6p	0.422	0.026	0.065	0.020	0.458	0.044	0.054	0.020
		fdp	0.279	0.020	0.290	0.020	0.314	0.020	0.118	0.020
		g3p	0.769	0.020	0.055	0.020	0.172	0.020		
		g6p	0.513	0.046	0.078	0.020	0.380	0.048	0.029	0.020
		pep	0.810	0.052	0.049	0.020	0.137	0.032		
		ru5p-D	0.580	0.020	0.092	0.020	0.296	0.020	0.032	0.020
		s7p	0.425	0.087	0.152	0.020	0.353	0.060	0.070	0.020
	D	3pg	0.806	0.021	0.049	0.020	0.142	0.020		
		6pgc	0.529	0.042	0.169	0.020	0.275	0.026	0.028	0.020
		dhap	0.761	0.020	0.050	0.020	0.184	0.020		
		f6p	0.537	0.033	0.070	0.020	0.366	0.035	0.027	0.020
		fdp	0.416	0.020	0.142	0.020	0.384	0.020	0.057	0.020
		g3p	0.782	0.020	0.049	0.020	0.165	0.020		
		g6p	0.514	0.046	0.078	0.020	0.378	0.047	0.030	0.020

A kinetic investigation of the Warburg effect in CHO cells

experiment	replicate	met_id	$\mu\,m_0$	$\sigma\,m_0$	$\mu\,m_1$	$\sigma\,m_1$	$\mu\,m_2$	$\sigma\,m_2$	$\mu\,m_3$	$\sigma\,m_3$
		pep	0.812	0.052	0.049	0.020	0.136	0.032		
		ru5p-D	0.539	0.020	0.167	0.020	0.257	0.020	0.037	0.020
		s7p	0.512	0.104	0.159	0.020	0.265	0.045	0.064	0.020
	A	3pg	0.788	0.021	0.044	0.020	0.165	0.022		
		6pgc	0.539	0.043	0.076	0.020	0.355	0.033	0.030	0.020
		dhap	0.803	0.020	0.042	0.020	0.154	0.020		
		f6p	0.523	0.032	0.055	0.020	0.402	0.038	0.021	0.020
		fdp	0.598	0.020	0.082	0.020	0.295	0.020	0.025	0.020
		g3p	0.814	0.020	0.028	0.020	0.158	0.020		
		g6p	0.521	0.046	0.054	0.020	0.406	0.051	0.019	0.020
		pep	0.820	0.052	0.039	0.020	0.141	0.033		
		ru5p-D	0.508	0.020	0.335	0.020	0.136	0.020	0.020	0.020
		s7p	0.475	0.097	0.198	0.020	0.271	0.046	0.056	0.020
	B	3pg	0.791	0.021	0.043	0.020	0.163	0.022		
		6pgc	0.562	0.045	0.087	0.020	0.323	0.030	0.028	0.020
		dhap	0.799	0.020	0.044	0.020	0.157	0.020		
		f6p	0.539	0.033	0.059	0.020	0.380	0.036	0.021	0.020
		fdp	0.617	0.020	0.081	0.020	0.279	0.020	0.024	0.020
		g3p	0.792	0.020	0.041	0.020	0.167	0.020		
		g6p	0.533	0.047	0.059	0.020	0.386	0.048	0.021	0.020
		pep	0.805	0.051	0.039	0.020	0.153	0.036		
CHO-S 11 KO		ru5p-D	0.589	0.020	0.242	0.020	0.142	0.020	0.027	0.020
		s7p	0.462	0.094	0.222	0.020	0.259	0.044	0.057	0.020
	C	3pg	0.786	0.021	0.050	0.020	0.160	0.021		
		6pgc	0.543	0.043	0.089	0.020	0.340	0.032	0.029	0.020
		dhap	0.769	0.020	0.047	0.020	0.184	0.020		
		f6p	0.522	0.032	0.058	0.020	0.394	0.037	0.026	0.020
		fdp	0.592	0.020	0.086	0.020	0.297	0.020	0.025	0.020
		g3p	0.785	0.020	0.043	0.020	0.172	0.020		
		g6p	0.511	0.045	0.065	0.020	0.399	0.050	0.025	0.020
		pep	0.802	0.051	0.037	0.020	0.161	0.038		
		ru5p-D	0.537	0.020	0.216	0.020	0.208	0.020	0.039	0.020
		s7p	0.472	0.096	0.177	0.020	0.293	0.050	0.058	0.020
	D	3pg	0.811	0.021	0.050	0.020	0.139	0.020		
		6pgc	0.579	0.046	0.096	0.020	0.306	0.028	0.019	0.020
		dhap	0.778	0.020	0.049	0.020	0.173	0.020		
		f6p	0.554	0.034	0.069	0.020	0.354	0.034	0.024	0.020
		fdp	0.649	0.020	0.085	0.020	0.242	0.020	0.024	0.020
		g3p	0.761	0.020	0.049	0.020	0.190	0.020		
		g6p	0.552	0.049	0.069	0.020	0.356	0.044	0.023	0.020
		pep	0.845	0.054	0.041	0.020	0.114	0.027		

experiment	replicate	met_id	$\mu\, m_0$	$\sigma\, m_0$	$\mu\, m_1$	$\sigma\, m_1$	$\mu\, m_2$	$\sigma\, m_2$	$\mu\, m_3$	$\sigma\, m_3$
		ru5p-D	0.622	0.020	0.194	0.020	0.168	0.020	0.016	0.020
		s7p	0.500	0.102	0.171	0.020	0.275	0.046	0.054	0.020
	A	3pg	0.798	0.021	0.054	0.020	0.146	0.020		
		6pgc	0.542	0.043	0.095	0.020	0.340	0.032	0.023	0.020
		dhap	0.800	0.020	0.044	0.020	0.154	0.020		
		f6p	0.503	0.031	0.061	0.020	0.408	0.039	0.028	0.020
		fdp	0.608	0.020	0.074	0.020	0.293	0.020	0.024	0.020
		g3p	0.794	0.020	0.044	0.020	0.160	0.020		
		g6p	0.552	0.049	0.065	0.020	0.356	0.045	0.027	0.020
		pep	0.829	0.053	0.042	0.020	0.126	0.030		
		ru5p-D	0.639	0.020	0.174	0.020	0.166	0.020	0.021	0.020
		s7p	0.477	0.097	0.152	0.020	0.311	0.053	0.061	0.020
	B	3pg	0.782	0.020	0.052	0.020	0.163	0.022		
		6pgc	0.528	0.042	0.112	0.020	0.332	0.031	0.028	0.020
		dhap	0.752	0.020	0.048	0.020	0.197	0.020		
		f6p	0.539	0.033	0.065	0.020	0.371	0.035	0.025	0.020
		fdp	0.587	0.020	0.069	0.020	0.320	0.020	0.025	0.020
		g3p	0.756	0.020	0.047	0.020	0.193	0.020		
		g6p	0.543	0.048	0.065	0.020	0.366	0.046	0.027	0.020
		pep	0.806	0.051	0.043	0.020	0.147	0.034		
CHO-ZenZeLa		ru5p-D	0.543	0.020	0.175	0.020	0.256	0.020	0.027	0.020
		s7p	0.460	0.094	0.151	0.020	0.328	0.056	0.061	0.020
	C	3pg	0.798	0.021	0.049	0.020	0.149	0.020		
		6pgc	0.553	0.044	0.103	0.020	0.322	0.030	0.022	0.020
		dhap	0.763	0.020	0.048	0.020	0.186	0.020		
		f6p	0.543	0.034	0.065	0.020	0.366	0.035	0.026	0.020
		fdp	0.621	0.020	0.071	0.020	0.286	0.020	0.022	0.020
		g3p	0.777	0.020	0.044	0.020	0.176	0.020		
		g6p	0.523	0.047	0.063	0.020	0.389	0.049	0.026	0.020
		pep	0.820	0.052	0.045	0.020	0.132	0.031		
		ru5p-D	0.529	0.020	0.204	0.020	0.243	0.020	0.023	0.020
		s7p	0.471	0.096	0.168	0.020	0.300	0.051	0.060	0.020
	D	3pg	0.747	0.020	0.051	0.020	0.198	0.027		
		6pgc	0.515	0.041	0.074	0.020	0.384	0.036	0.027	0.020
		dhap	0.725	0.020	0.043	0.020	0.228	0.020		
		f6p	0.521	0.032	0.063	0.020	0.388	0.037	0.028	0.020
		fdp	0.573	0.020	0.066	0.020	0.337	0.020	0.025	0.020
		g3p	0.724	0.020	0.046	0.020	0.226	0.020		
		g6p	0.515	0.046	0.066	0.020	0.392	0.049	0.027	0.020
		pep	0.783	0.050	0.041	0.020	0.172	0.040		
		ru5p-D	0.554	0.020	0.211	0.020	0.202	0.020	0.033	0.020

A kinetic investigation of the Warburg effect in CHO cells

experiment	replicate	met_id	$\mu\, m_0$	$\sigma\, m_0$	$\mu\, m_1$	$\sigma\, m_1$	$\mu\, m_2$	$\sigma\, m_2$	$\mu\, m_3$	$\sigma\, m_3$
		s7p	0.434	0.088	0.166	0.020	0.330	0.056	0.070	0.020
		3pg	0.761	0.020	0.057	0.020	0.175	0.023		
		6pgc	0.495	0.039	0.221	0.020	0.252	0.023	0.033	0.020
		dhap	0.706	0.020	0.056	0.020	0.231	0.020		
		f6p	0.482	0.030	0.076	0.020	0.406	0.039	0.036	0.020
		fdp	0.563	0.020	0.085	0.020	0.321	0.020	0.031	0.020
	A	g3p	0.707	0.020	0.055	0.020	0.232	0.020		
		g6p	0.481	0.043	0.081	0.020	0.404	0.050	0.034	0.020
		pep	0.755	0.048	0.044	0.020	0.201	0.047		
		ru5p-D	0.549	0.020	0.225	0.020	0.193	0.020	0.032	0.020
		s7p	0.403	0.082	0.202	0.020	0.309	0.052	0.085	0.020
		3pg	0.752	0.020	0.075	0.020	0.167	0.022		
		6pgc	0.509	0.041	0.159	0.020	0.295	0.027	0.036	0.020
		dhap	0.723	0.020	0.066	0.020	0.204	0.020		
		f6p	0.500	0.031	0.099	0.020	0.363	0.035	0.038	0.020
		fdp	0.572	0.020	0.104	0.020	0.291	0.020	0.033	0.020
	B	g3p	0.701	0.020	0.070	0.020	0.222	0.020		
		g6p	0.495	0.044	0.106	0.020	0.364	0.045	0.036	0.020
		pep	0.806	0.051	0.071	0.020	0.123	0.029		
CHO-Zen		ru5p-D	0.476	0.020	0.342	0.020	0.152	0.020	0.030	0.020
		s7p	0.422	0.086	0.222	0.020	0.278	0.047	0.078	0.020
		3pg	0.754	0.020	0.067	0.020	0.173	0.023		
		6pgc	0.490	0.039	0.155	0.020	0.322	0.030	0.033	0.020
		dhap	0.707	0.020	0.064	0.020	0.222	0.020		
		f6p	0.483	0.030	0.097	0.020	0.380	0.036	0.041	0.020
	C	fdp	0.584	0.020	0.099	0.020	0.287	0.020	0.030	0.020
		g3p	0.692	0.020	0.070	0.020	0.231	0.020		
		g6p	0.477	0.042	0.102	0.020	0.383	0.048	0.038	0.020
		pep	0.770	0.049	0.065	0.020	0.166	0.039		
		ru5p-D	0.527	0.020	0.233	0.020	0.200	0.020	0.039	0.020
		s7p	0.403	0.082	0.213	0.020	0.299	0.051	0.085	0.020
		3pg	0.773	0.020	0.047	0.020	0.180	0.024		
		6pgc	0.476	0.038	0.112	0.020	0.374	0.035	0.039	0.020
		dhap	0.742	0.020	0.052	0.020	0.199	0.020		
		f6p	0.473	0.029	0.075	0.020	0.415	0.039	0.037	0.020
	D	fdp	0.609	0.020	0.083	0.020	0.280	0.020	0.027	0.020
		g3p	0.731	0.020	0.054	0.020	0.215	0.020		
		g6p	0.477	0.042	0.077	0.020	0.409	0.051	0.037	0.020
		pep	0.828	0.053	0.044	0.020	0.128	0.030		
		ru5p-D	0.489	0.020	0.238	0.020	0.233	0.020	0.039	0.020
		s7p	0.393	0.080	0.175	0.020	0.346	0.059	0.086	0.020
		3pg	0.817	0.021	0.053	0.020	0.127	0.020		

A

experiment	replicate	met_id	$\mu\,m_0$	$\sigma\,m_0$	$\mu\,m_1$	$\sigma\,m_1$	$\mu\,m_2$	$\sigma\,m_2$	$\mu\,m_3$	$\sigma\,m_3$
		6pgc	0.591	0.047	0.117	0.020	0.262	0.024	0.031	0.020
		dhap	0.782	0.020	0.065	0.020	0.149	0.020		
		f6p	0.572	0.035	0.085	0.020	0.317	0.030	0.026	0.020
		fdp	0.695	0.020	0.109	0.020	0.177	0.020	0.019	0.020
		g3p	0.774	0.020	0.062	0.020	0.159	0.020		
		g6p	0.560	0.050	0.082	0.020	0.333	0.042	0.025	0.020
		pep	0.853	0.054	0.042	0.020	0.103	0.024		
		ru5p-D	0.611	0.020	0.233	0.020	0.134	0.020	0.022	0.020
		s7p	0.559	0.114	0.230	0.020	0.170	0.029	0.041	0.020
	B	3pg	0.746	0.020	0.060	0.020	0.188	0.025		
		6pgc	0.515	0.041	0.136	0.020	0.315	0.029	0.035	0.020
		dhap	0.692	0.020	0.063	0.020	0.239	0.020		
		f6p	0.474	0.029	0.077	0.020	0.415	0.040	0.034	0.020
		fdp	0.592	0.020	0.096	0.020	0.283	0.020	0.030	0.020
		g3p	0.702	0.020	0.067	0.020	0.225	0.020		
		g6p	0.492	0.044	0.053	0.020	0.422	0.053	0.033	0.020
		pep	0.782	0.050	0.051	0.020	0.163	0.038		
		ru5p-D	0.573	0.020	0.227	0.020	0.167	0.020	0.033	0.020
		s7p	0.425	0.086	0.218	0.020	0.282	0.048	0.075	0.020
	C	3pg	0.807	0.021	0.059	0.020	0.130	0.020		
		6pgc	0.593	0.047	0.132	0.020	0.250	0.023	0.025	0.020
		dhap	0.779	0.020	0.058	0.020	0.159	0.020		
		f6p	0.568	0.035	0.081	0.020	0.325	0.031	0.026	0.020
		fdp	0.674	0.020	0.101	0.020	0.205	0.020	0.021	0.020
		g3p	0.769	0.020	0.053	0.020	0.173	0.020		
		g6p	0.554	0.049	0.080	0.020	0.341	0.043	0.025	0.020
		pep	0.854	0.054	0.044	0.020	0.102	0.024		
		ru5p-D	0.633	0.020	0.208	0.020	0.130	0.020	0.029	0.020
		s7p	0.525	0.107	0.218	0.020	0.208	0.035	0.049	0.020
	D	3pg	0.771	0.020	0.050	0.020	0.175	0.024		
		6pgc	0.557	0.044	0.117	0.020	0.299	0.028	0.028	0.020
		dhap	0.734	0.020	0.047	0.020	0.215	0.020		
		f6p	0.488	0.030	0.091	0.020	0.384	0.037	0.037	0.020
		fdp	0.598	0.020	0.087	0.020	0.290	0.020	0.025	0.020
		g3p	0.734	0.020	0.053	0.020	0.209	0.020		
		g6p	0.592	0.053	0.058	0.020	0.322	0.040	0.028	0.020
		pep	0.806	0.051	0.041	0.020	0.149	0.035		
		ru5p-D	0.517	0.020	0.221	0.020	0.224	0.020	0.038	0.020
		s7p	0.469	0.096	0.212	0.020	0.255	0.043	0.063	0.020

E   CHO fitted model parameters

E.1   Fitted parameter values and their prior distributions

Table E.1: Fitted parameter values and their prior distributions. Z-scores represent the number of standard deviations the fitted values are from their priors.

parameter	value	prior mean	prior sd (log-normal)	Z-score	parameter	value	prior mean	prior sd (log-normal)	Z-score
k_{cat}^{HEX1}	3.4E+05	3.4E+05	5.0E-01	-4.4E-02	$K_m^{G6PDH,g6p_c}$	3.6E-05	3.6E-05	5.0E-01	-6.4E-03
k_{cat}^{HEX2}	2.2E+06	7.4E+05	1.0E+00	-1.1E+00	$K_m^{G6PDH,nadp_c}$	6.6E-06	6.6E-06	1.4E-01	-5.6E-03
k_{cat}^{GPI}	1.4E+07	1.2E+07	1.0E+00	-1.8E-01	$K_m^{PGL,6pgl_c}$	8.0E-05	8.0E-05	5.0E-01	0.0E+00
k_{cat}^{PFKL}	4.1E+05	4.6E+05	1.0E+00	1.2E-01	$K_m^{GND,6pgc_c}$	2.2E-05	2.3E-05	5.0E-01	7.0E-02
k_{cat}^{PFKM}	5.4E+06	3.0E+06	1.0E+00	-6.0E-01	$K_m^{GND,nadp_c}$	2.8E-06	2.9E-06	5.0E-01	6.7E-02
k_{cat}^{FBA}	2.0E+05	2.2E+05	1.0E+00	7.8E-02	$K_m^{GND,nadph_c}$	3.1E-07	3.0E-07	5.0E-01	-6.5E-02
k_{cat}^{TPI}	1.4E+07	1.4E+07	1.0E+00	-1.1E-03	$K_m^{GND,ru5p-D_c}$	3.1E-05	3.0E-05	5.0E-01	-5.9E-02
k_{cat}^{GAPD}	3.0E+04	5.8E+05	1.0E+00	3.0E+00	$K_m^{GND,co2_c}$	1.5E-02	1.5E-02	1.0E-01	-9.3E-04
k_{cat}^{PGK}	1.5E+06	1.5E+06	1.0E+00	7.8E-05	$K_m^{RPI,r5p_c}$	1.1E-03	1.2E-03	2.7E-01	3.1E-01
k_{cat}^{PGM}	2.9E+06	2.9E+06	1.0E+00	1.7E-05	$K_m^{RPI,ru5p-D_c}$	5.0E-03	5.1E-03	2.9E-01	6.8E-02
k_{cat}^{ENO}	2.1E+05	2.1E+05	1.0E+00	4.9E-05	$K_m^{RPE,ru5p-D_c}$	5.6E-04	7.4E-04	5.9E-01	4.7E-01
k_{cat}^{PYK1}	3.0E+04	2.3E+06	1.0E+00	4.3E+00	$K_m^{RPE,xu5p-D_c}$	1.5E-04	1.5E-04	5.0E-01	6.1E-02
k_{cat}^{PYK2}	1.6E+06	1.6E+06	1.0E-01	2.6E-01	$K_m^{TKT1,r5p_c}$	1.4E-04	1.4E-04	3.3E-01	1.7E-02
k_{cat}^{G6PDH}	2.9E+05	3.6E+05	2.0E+00	1.0E-01	$K_m^{TKT1,xu5p-D_c}$	6.6E-05	6.7E-05	4.2E-01	3.1E-02
k_{cat}^{PGL}	1.0E+05	1.0E+05	2.0E+00	-5.6E-04	$K_m^{TKT1,s7p_c}$	4.0E-03	4.0E-03	1.0E+00	-6.3E-03
k_{cat}^{GND}	1.1E+05	1.1E+05	2.6E-01	-4.5E-02	$K_m^{TKT1,g3p_c}$	3.0E-03	3.0E-03	1.5E-01	7.7E-05
k_{cat}^{RPI}	1.5E+06	1.3E+06	9.7E-01	-1.7E-01	$K_m^{TKT2,e4p_c}$	3.4E-04	3.6E-04	5.0E-01	1.2E-01
k_{cat}^{RPE}	2.6E+06	4.3E+05	1.5E+00	-1.2E+00	$K_m^{TKT2,f6p_c}$	3.5E-03	3.4E-03	5.0E-01	-3.9E-02
k_{cat}^{TKT1}	2.7E+04	2.6E+04	4.1E-01	-6.2E-02	$K_m^{TKT2,xu5p-D_c}$	6.5E-05	6.7E-05	4.2E-01	5.7E-02
k_{cat}^{TKT2}	3.2E+05	2.5E+05	1.0E+00	-2.6E-01	$K_m^{TKT2,g3p_c}$	3.0E-03	3.0E-03	1.5E-01	-2.3E-04
k_{cat}^{TALA}	1.2E+05	6.5E+04	1.0E+00	-6.4E-01	$K_m^{TALA,f6p_c}$	2.1E-03	2.3E-03	5.0E-01	1.4E-01

parameter	value	prior mean	prior sd (log-normal)	Z-score	parameter	value	prior mean	prior sd (log-normal)	Z-score
$K_m^{HEX1,glc-D_c}$	6.0E-05	6.0E-05	5.0E-01	1.3E-04	$K_m^{TALA,g3p_c}$	1.0E-03	1.0E-03	2.0E+00	-1.1E-03
K_m^{HEX1,atp_c}	8.8E-04	8.8E-04	5.0E-01	1.4E-03	$K_m^{TALA,s7p_c}$	1.8E-03	1.0E-03	2.0E+00	-2.8E-01
$K_m^{HEX2,glc-D_c}$	3.7E-04	3.7E-04	5.0E-01	1.7E-02	$K_m^{TALA,e4p_c}$	9.8E-05	1.0E-03	2.0E+00	1.2E+00
K_m^{HEX2,atp_c}	8.0E-04	8.1E-04	5.0E-01	2.0E-02	$K_i^{HEX2,g6p_c}$	2.4E-04	2.4E-04	5.0E-01	-6.9E-03
$K_m^{GPI,g6p_c}$	1.7E-04	1.7E-04	2.0E-01	2.7E-02	K_i^{HEX2,gdp_c}	3.2E-03	3.2E-03	1.5E+00	-1.0E-03
$K_m^{GPI,f6p_c}$	3.1E-05	3.0E-05	5.0E-01	-6.6E-02	K_i^{TPI,pep_c}	3.2E-05	3.2E-05	1.5E+00	1.2E-03
$K_m^{PFKL,f6p_c}$	7.1E-05	7.0E-05	5.0E-01	-1.9E-02	$K_i^{GAPD,g3p_c}$	3.1E-04	3.1E-04	5.0E-01	6.5E-06
K_m^{PFKL,atp_c}	4.5E-05	4.5E-05	6.4E-01	-2.5E-02	$K_i^{GAPD,nadh_c}$	1.0E-06	1.0E-06	5.0E-01	2.0E-05
K_m^{PFKL,adp_c}	1.4E-04	1.4E-04	5.0E-01	1.9E-02	$K_i^{GAPD,13dpg_c}$	3.0E-07	3.0E-07	5.0E-01	-3.3E-05
K_m^{PFKL,fdp_c}	4.3E-04	4.3E-04	5.0E-01	1.9E-02	$K_i^{G6PDH,g6p_c}$	4.4E-03	4.4E-03	5.0E-01	1.5E-04
$K_m^{PFKM,f6p_c}$	7.0E-05	7.0E-05	5.0E-01	7.7E-03	K_i^{G6PDH,adp_c}	1.9E-03	1.9E-03	5.0E-01	1.2E-03
K_m^{PFKM,atp_c}	1.6E-05	1.6E-05	7.3E-01	1.9E-04	K_i^{G6PDH,atp_c}	5.7E-03	5.7E-03	1.0E-01	3.1E-04
K_m^{PFKM,adp_c}	1.8E-03	1.8E-03	9.7E-02	1.7E-05	$K_i^{GND,nadph_c}$	1.1E-06	1.1E-06	5.0E-01	-1.6E-03
K_m^{PFKM,fdp_c}	6.6E-03	6.6E-03	3.0E-01	1.0E-04	L_0^{HEX1}	1.0E+00	1.0E+00	1.0E-01	8.7E-03
K_m^{FBA,fdp_c}	5.4E-05	5.2E-05	5.0E-01	-5.8E-02	L_0^{HEX2}	6.4E+00	2.0E+01	1.5E+00	7.6E-01
$K_m^{FBA,dhap_c}$	3.4E-05	3.5E-05	5.0E-01	5.2E-02	L_0^{PFKL}	1.0E+01	1.0E+01	5.3E-01	-1.6E-03
$K_m^{FBA,g3p_c}$	1.9E-04	1.9E-04	5.0E-01	-2.9E-02	L_0^{PFKM}	1.6E+01	1.6E+01	4.9E-01	-2.9E-02
$K_m^{TPI,dhap_c}$	1.8E-04	1.8E-04	5.0E-01	6.7E-05	L_0^{PYK2}	6.6E-01	6.6E-01	2.0E+00	0.0E+00
$K_m^{TPI,g3p_c}$	1.3E-05	1.3E-05	5.0E-01	-6.6E-04	L_0^{G6PDH}	1.0E+01	1.5E+00	2.0E+00	-9.7E-01
$K_m^{GAPD,g3p_c}$	1.0E-04	9.5E-05	5.0E-01	-1.1E-01	L_0^{RPE}	3.5E-01	1.0E+00	2.0E+00	5.3E-01
K_m^{GAPD,pi_c}	8.1E-05	7.8E-05	5.0E-01	-7.6E-02	$e_T^{HEX1,g6p_c}$	1.5E-03	1.0E-03	2.0E+00	-2.1E-01
K_m^{GAPD,nad_c}	4.5E-05	4.5E-05	5.0E-01	-5.9E-04	e_T^{HEX1,gdp_c}	3.5E-03	3.2E-03	1.5E+00	-7.7E-02
$K_m^{GAPD,nadh_c}$	3.3E-06	3.3E-06	5.0E-01	1.2E-05	e_R^{HEX2,mit_c}	1.0E+00	1.0E+00	1.0E-01	4.4E-02

parameter	value	prior mean	prior sd (log-normal)	Z-score	parameter	value	prior mean	prior sd (log-normal)	Z-score
$K_m^{GAPD,13dpg_c}$	8.0E-07	8.0E-07	5.0E-01	-5.0E-06	$e_R^{PFKL,f6p_c}$	6.3E-05	6.3E-05	1.5E+00	1.6E-03
$K_m^{PGK,13dpg_c}$	1.9E-06	1.9E-06	5.0E-01	-3.2E-05	e_R^{PFKL,fdp_c}	6.3E-03	6.3E-03	4.9E-01	-7.2E-03
K_m^{PGK,adp_c}	1.5E-04	1.5E-04	5.0E-01	-2.7E-05	e_T^{PFKL,atp_c}	1.2E-03	1.4E-03	2.8E+00	4.6E-02
K_m^{PGK,atp_c}	4.2E-04	4.2E-04	5.0E-01	1.9E-05	$e_R^{PFKM,f6p_c}$	5.5E-05	6.3E-05	1.5E+00	9.2E-02
$K_m^{PGK,3pg_c}$	1.3E-03	1.3E-03	5.0E-01	0.0E+00	e_R^{PFKM,fdp_c}	6.4E-03	6.3E-03	4.9E-01	-2.6E-02
$K_m^{PGM,3pg_c}$	8.8E-04	8.8E-04	7.1E-01	-1.2E-05	e_T^{PFKM,atp_c}	9.2E-03	4.5E-05	3.3E+00	-1.6E+00
$K_m^{PGM,2pg_c}$	1.9E-05	1.9E-05	1.2E-01	-5.6E-06	e_R^{PFKM,amp_c}	3.2E-04	3.0E-04	1.0E+00	-7.6E-02
$K_m^{ENO,2pg_c}$	1.3E-04	1.3E-04	5.0E-01	-3.1E-05	$e_R^{PFKM,f26bp_c}$	1.0E+00	1.0E+00	1.0E-02	0.0E+00
K_m^{ENO,pep_c}	3.4E-05	3.4E-05	5.0E-01	1.2E-05	e_T^{PFKM,cit_c}	1.0E+00	1.0E+00	1.0E-02	7.2E-03
K_m^{PYK1,pep_c}	1.2E-04	5.8E-05	5.0E-01	-1.5E+00	e_R^{PYK2,pep_c}	2.0E-04	2.0E-04	1.0E+00	0.0E+00
K_m^{PYK1,adp_c}	5.9E-04	5.6E-04	5.0E-01	-1.2E-01	e_R^{PYK2,fdp_c}	2.5E-07	2.5E-07	1.0E+00	0.0E+00
K_m^{PYK2,pep_c}	1.3E-04	1.1E-04	3.3E-01	-6.6E-01	$e_R^{G6PDH,nadp_c}$	1.5E-04	1.5E-04	1.0E+00	-2.8E-02
K_m^{PYK2,adp_c}	3.3E-04	3.2E-04	5.0E-01	-4.1E-02	$e_R^{RPE,ru5p-D_c}$	1.7E-04	1.9E-04	1.0E+00	8.4E-02

E.2   allostery results

The values in Table E.2 are the calculated regulatory coefficients for the allosteric contribution of each enzyme.

Table E.2: Estimated allosteric contributions from the fitted CHO model.

edges	CHO-S wt	CHO-ZeLa	CHO-ZenZeLa	CHO-Zen
r5p-drain	1.000	1.000	1.000	1.000
HEX1_HEX	0.023	0.018	0.041	0.014
HEX2_HEX	0.605	0.339	0.215	0.394
GPI_GPI	1.000	1.000	1.000	1.000
PFKL_PFK	0.668	0.856	0.147	0.793
PFKM_PFK	0.463	0.237	0.286	0.167
FBA_FBA	1.000	1.000	1.000	1.000
TPI_TPI	1.000	1.000	1.000	1.000
GAPD_GAPD	1.000	1.000	1.000	1.000
PGK_PGK	1.000	1.000	1.000	1.000
PGM_PGM	1.000	1.000	1.000	1.000
ENO_ENO	1.000	1.000	1.000	1.000
PYK1_PYK	1.000	1.000	1.000	1.000
PYK2_PYK	1.000	1.000	1.000	1.000
G6PDH_G6PDH	0.355	0.366	0.241	0.360
PGL_PGL	1.000	1.000	1.000	1.000
GND_GND	1.000	1.000	1.000	1.000
RPI_RPI	1.000	1.000	1.000	1.000
RPE_RPE	0.861	0.891	0.833	0.860
TKT1_TKT1	1.000	1.000	1.000	1.000
TKT2_TKT2	1.000	1.000	1.000	1.000
TALA_TALA	1.000	1.000	1.000	1.000

E.3   saturation results

The values in Table E.3 are the calculated regulatory coefficients for the saturation contribution of each enzyme.

Table E.3: Estimated saturation contributions from the fitted CHO model.

edges	CHO-S wt	CHO-ZeLa	CHO-ZenZeLa	CHO-Zen
r5p-drain	1	1	1	1
HEX1_HEX	0.970	0.972	0.981	0.977
HEX2_HEX	0.946	0.947	0.961	0.950
GPI_GPI	0.251	0.249	0.249	0.249

edges	CHO-S wt	CHO-ZeLa	CHO-ZenZeLa	CHO-Zen
PFKL_PFK	0.963	0.729	0.957	0.981
PFKM_PFK	0.980	0.982	0.977	0.987
FBA_FBA	0.534	0.593	0.499	0.378
TPI_TPI	0.273	0.394	0.347	0.302
GAPD_GAPD	0.420	0.838	0.549	0.365
PGK_PGK	0.266	0.170	0.131	0.137
PGM_PGM	0.149	0.117	0.104	0.114
ENO_ENO	0.203	0.147	0.113	0.136
PYK1_PYK	0.504	0.387	0.346	0.318
PYK2_PYK	0.489	0.372	0.330	0.301
G6PDH_G6PDH	0.985	0.986	0.973	0.987
PGL_PGL	0.893	0.656	0.415	0.826
GND_GND	0.199	0.145	0.071	0.196
RPI_RPI	0.026	0.029	0.017	0.024
RPE_RPE	0.156	0.177	0.112	0.145
TKT1_TKT1	0.523	0.704	0.502	0.567
TKT2_TKT2	0.041	0.207	0.059	0.042
TALA_TALA	0.118	0.359	0.165	0.142

E.4 reversibility results

The values in Table E.4 are the calculated regulatory coefficients for the saturation contribution of each enzyme.

Table E.4: Estimated reversibility contributions from the fitted CHO model.

edges	CHO-S wt	CHO-ZeLa	CHO-ZenZeLa	CHO-Zen
r5p-drain	1	1	1	1
r5p-drain	1.000	1.000	1.000	1.000
HEX1_HEX	1.000	1.000	1.000	1.000
HEX2_HEX	1.000	1.000	1.000	1.000
GPI_GPI	0.018	0.008	0.008	0.008
PFKL_PFK	1.000	0.993	1.000	1.000
PFKM_PFK	1.000	0.993	1.000	1.000
FBA_FBA	0.690	0.304	0.480	0.514
TPI_TPI	0.017	0.006	0.006	0.008
GAPD_GAPD	1.000	1.000	1.000	1.000
PGK_PGK	0.238	0.247	0.265	0.307
PGM_PGM	0.313	0.184	0.158	0.219
ENO_ENO	0.529	0.380	0.215	0.394
PYK1_PYK	1.000	1.000	1.000	1.000

edges	CHO-S wt	CHO-ZeLa	CHO-ZenZeLa	CHO-Zen
PYK2_PYK	1.000	1.000	1.000	1.000
G6PDH_G6PDH	1.000	1.000	1.000	1.000
PGL_PGL	1.000	1.000	1.000	1.000
GND_GND	0.970	0.952	0.907	0.969
RPI_RPI	0.642	0.449	0.417	0.531
RPE_RPE	0.847	0.831	0.793	0.823
TKT1_TKT1	0.130	0.066	0.052	0.128
TKT2_TKT2	0.138	0.019	0.037	0.141
TALA_TALA	-0.668	-0.158	-0.168	-0.411

E.5   Fitted multiomics data

Tables E.5, E.6 and E.7 were calculated by solving the kinetic model defined in Chapter 4 using the data stored in

Table E.5: Estimated metabolite concentrations in $\frac{mol}{L_{cell}}$ from the fitted CHO model.

metabolites	CHO-S wt	CHO-ZeLa	CHO-ZenZeLa	CHO-Zen
glc-D_c	2.00E-02	2.00E-02	2.00E-02	2.00E-02
atp_c	3.19E-02	3.44E-02	5.24E-02	4.24E-02
adp_c	1.35E-02	1.13E-02	1.48E-02	1.81E-02
g6p_c	6.39E-03	7.56E-03	5.49E-03	9.89E-03
f6p_c	3.50E-03	4.18E-03	3.04E-03	5.48E-03
fdp_c	2.02E-03	8.19E-02	3.60E-03	7.98E-04
dhap_c	7.63E-04	7.25E-03	1.31E-03	5.98E-04
g3p_c	7.90E-05	7.58E-04	1.37E-04	6.25E-05
pi_c	1.61E-03	1.52E-03	1.56E-03	1.57E-03
nad_c	1.06E-01	1.00E-01	1.48E-01	1.22E-01
nadh_c	2.19E-08	4.79E-06	2.19E-07	2.19E-07
13dpg_c	1.95E-06	1.02E-06	7.55E-07	6.43E-07
3pg_c	1.50E-03	5.95E-04	3.72E-04	4.52E-04
2pg_c	1.63E-04	7.69E-05	4.97E-05	5.59E-05
f26bp_c	1.08E+00	9.84E-01	9.75E-01	9.85E-01
pep_c	1.34E-04	8.28E-05	6.77E-05	5.88E-05
pyr_c	1.00E-03	1.00E-03	1.00E-03	1.00E-03
lac-L_c	4.15E-02	1.58E-03	4.80E-02	8.43E-02
nadp_c	7.34E-04	7.75E-04	3.58E-04	7.52E-04
nadph_c	2.31E-04	1.99E-04	4.08E-04	2.30E-04
6pgl_c	6.67E-04	1.53E-04	5.68E-05	3.80E-04
6pgc_c	1.37E-04	1.13E-04	1.02E-04	1.28E-04
ru5p-D_c	2.02E-04	3.25E-04	1.28E-04	1.97E-04

metabolites	CHO-S wt	CHO-ZeLa	CHO-ZenZeLa	CHO-Zen
r5p_c	5.31E-04	1.31E-03	5.47E-04	6.79E-04
co2_c	1.20E-03	1.20E-03	1.20E-03	1.20E-03
xu5p-D_c	1.37E-04	2.43E-04	1.17E-04	1.55E-04
e4p_c	2.92E-05	1.66E-04	4.62E-05	3.21E-05
s7p_c	9.98E-04	4.90E-04	5.52E-04	1.83E-03
gdp_c	4.73E-03	5.32E-03	9.10E-04	3.86E-03
amp_c	5.21E-04	1.43E-03	1.82E-03	4.01E-03
mit_c	8.78E+00	2.28E+00	7.55E-01	3.15E+00
cit_c	4.09E-01	1.43E+01	3.17E+00	2.97E+01

Table E.6: Estimated fluxes in $\frac{mol}{L_{cell}} \cdot s$ from the fitted CHO model.

reaction	CHO-S wt	CHO-ZeLa	CHO-ZenZeLa	CHO-Zen
reactions CHO-S wt CHO-ZeLa CHO-ZenZeLa CHO-Zen HEX	1.87E-01	1.34E-01	1.55E-01	1.52E-01
GPI	1.63E-01	1.12E-01	1.39E-01	1.14E-01
PFK	1.77E-01	1.25E-01	1.49E-01	1.38E-01
FBA	1.77E-01	1.25E-01	1.49E-01	1.38E-01
TPI	1.77E-01	1.25E-01	1.49E-01	1.38E-01
GAPD	3.62E-01	2.57E-01	3.02E-01	2.88E-01
PGK	3.62E-01	2.57E-01	3.02E-01	2.88E-01
PGM	3.62E-01	2.57E-01	3.02E-01	2.88E-01
ENO	3.62E-01	2.57E-01	3.02E-01	2.88E-01
PYK	3.62E-01	2.57E-01	3.02E-01	2.88E-01
G6PDH	2.40E-02	2.23E-02	1.64E-02	3.72E-02
PGL	2.40E-02	2.23E-02	1.64E-02	3.72E-02
GND	2.40E-02	2.23E-02	1.64E-02	3.72E-02
RPI	9.30E-03	8.60E-03	6.69E-03	1.35E-02
RPE	1.47E-02	1.37E-02	9.71E-03	2.37E-02
TKT1	7.36E-03	6.85E-03	4.86E-03	1.18E-02
TKT2	7.36E-03	6.85E-03	4.86E-03	1.18E-02
TALA	-7.36E-03	-6.85E-03	-4.86E-03	-1.18E-02
r5p-drain	1.93E-03	1.75E-03	1.84E-03	1.70E-03

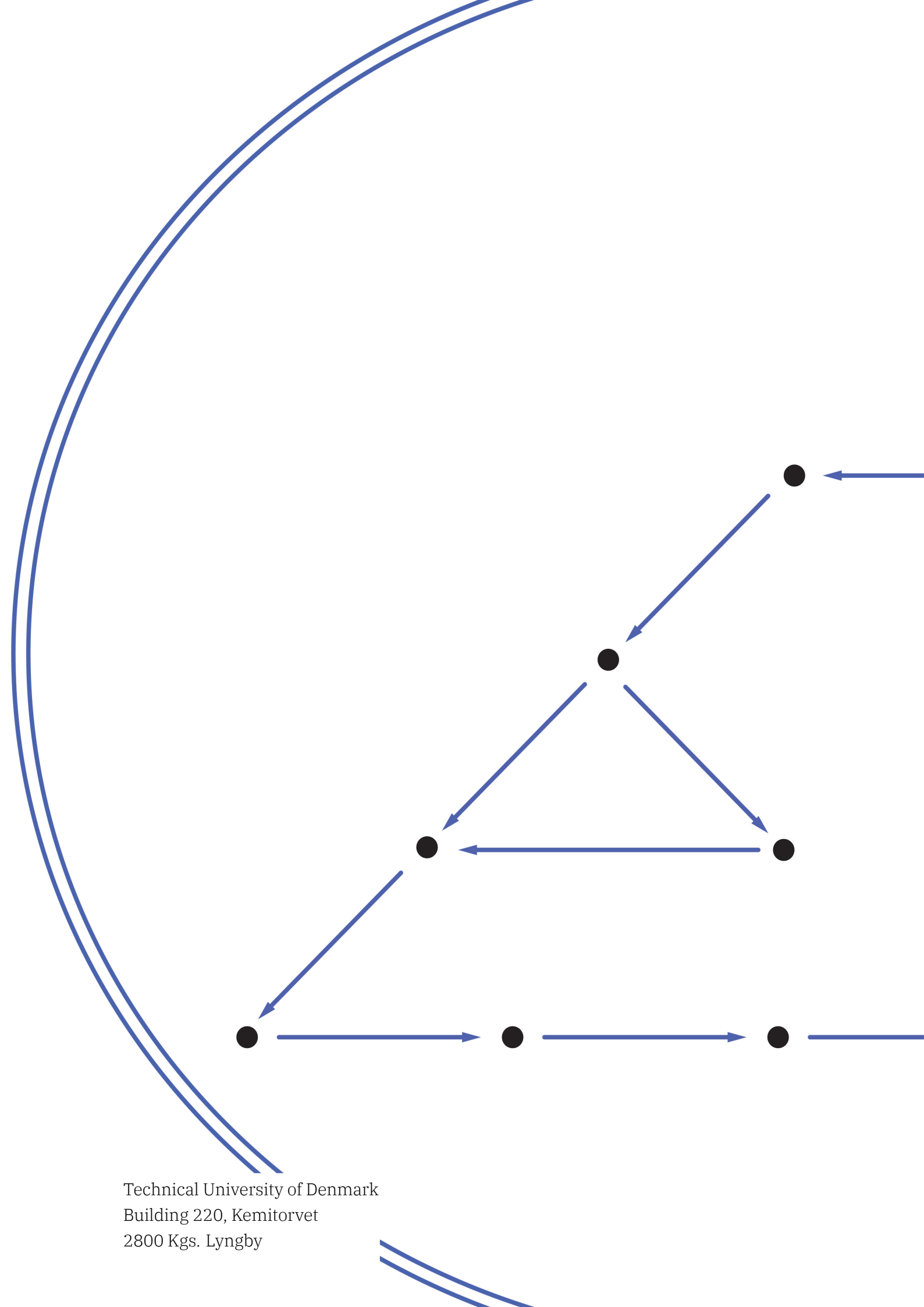Table E.7: Estimated enzyme concentrations in $\frac{mol}{L_{cell}}$ from the fitted CHO model.

enzyme	CHO-S wt	CHO-ZeLa	CHO-ZenZeLa	CHO-Zen
HEX1	2.73E-06	3.58E-06	4.97E-06	3.69E-06
HEX2	1.32E-07	1.60E-07	1.91E-07	1.65E-07

enzyme	CHO-S wt	CHO-ZeLa	CHO-ZenZeLa	CHO-Zen
GPI	2.51E-06	3.86E-06	5.15E-06	4.16E-06
PFKL	8.64E-08	1.16E-07	1.94E-07	1.34E-07
PFKM	6.29E-08	7.67E-08	9.10E-08	1.07E-07
FBA	2.41E-06	3.48E-06	3.11E-06	3.56E-06
TPI	2.84E-06	4.03E-06	5.31E-06	4.31E-06
GAPD	2.91E-05	1.04E-05	1.86E-05	2.66E-05
PGK	3.70E-06	3.96E-06	5.65E-06	4.44E-06
PGM	2.70E-06	4.17E-06	6.44E-06	4.03E-06
ENO	1.64E-05	2.25E-05	6.06E-05	2.62E-05
PYK1	1.04E-05	1.21E-05	1.73E-05	1.41E-05
PYK2	2.67E-07	1.98E-07	2.36E-07	3.24E-07
G6PDH	2.33E-07	2.10E-07	2.37E-07	3.56E-07
PGL	2.62E-07	3.31E-07	3.85E-07	4.39E-07
GND	1.12E-06	1.46E-06	2.30E-06	1.76E-06
RPI	3.58E-07	4.39E-07	6.28E-07	7.03E-07
RPE	5.05E-08	4.06E-08	5.13E-08	9.00E-08
TKT1	4.07E-06	5.52E-06	6.97E-06	6.15E-06
TKT2	4.06E-06	5.52E-06	6.97E-06	6.18E-06
TALA	7.56E-07	9.78E-07	1.43E-06	1.65E-06

Technical University of Denmark

Building 220, Kemitorvet

2800 Kgs. Lyngby