Systems Tools for Metabolic Engineering of Parageobacillus thermoglucosidasius NCIMB 11955

(Alternative Format Thesis)

Bennett, Martyn

Award date:
2022

Awarding institution:
University of Bath

Systems Tools for Metabolic Engineering of Parageobacillus thermoglucosidasius NCIMB 11955

(Alternative Format Thesis)

Bennett, Martyn

Award date:
2022

Awarding institution:
University of Bath

Systems Tools for Metabolic Engineering of *Parageobacillus thermoglucosidasius* NCIMB 11955

Martyn Cameron Bennett

A thesis submitted for the degree of ***Doctor of Philosophy in Integrated Sustainable Chemical Technologies***

Department of Biology and Biochemistry, University of Bath, Bath, BA2 7AY, UK

Principal Investigator: Professor David Leak

Co-supervisor: Dr. Tom Arnot

Industrial Partner: Corbion (Netherlands)

International Partner: Dr. Shyam Masakapalli (Indian Institute of Technology, Mandi)

December 2021

Table of contents

Acknowledgements

Firstly, I would like to thank my supervisor Prof. David Leak for accepting me to do this PhD project and for his support throughout. I would also like to thank my Co-supervisor Prof. Tom Arnot for his advice on the construction of the microbioreactors and ultimately the CSCT for allowing me to do this PhD and the EPSRC for my funding and for granting me the COVID-19 funding extension.

At the University of Bath, I would like to thank the CSCT cohort '16 and the changing members of 4S 1.28 for all their support over the years. I would like to thank Lesley Chapman, Martin White, and the rest of the stores team South sciences stores staff for covering all the equipment and reagent orders I put in over the years. I would particularly like to thank Dave Elkins and Robin Mocock in the workshop for helping to repair the Braun control unit multiple times and for sourcing the eventual chiller unit, and Ewan Basterfield for keeping everything else in 4S running. I would also like to thank Shaun Reeksting based in Pharmacy and Pharmacology for his advice on running GC samples and fixing GCs themselves.

With regards to the third chapter of this thesis, I would like to thank Dr. Beata Lisowska for the initial development of the *P. thermoglucosidasius* genome-scale metabolic model (GSMM) which featured as a tool throughout the PhD. Next, I would like to thank my collaborators on the GSMM from the Technical University of Denmark, in particular Vivienne Mol and Ben Sánchez, who introduced me to Github and were great conversation over the lockdown periods.

At IIT Mandi I would particularly like to thank Dr. Shyam Kumar Masakapalli. He provided his unprocessed anaerobic ^{13}C xylose INST-^{13}C-MFA HPLC and GC-MS data, which are analysed here for the first time in this work, as well as invaluable advice on isotopomer network development and ^{13}C data fitting. Additionally, I would like to thank him for agreeing to have me join in his lab group at IIT Mandi for 3 months and develop my understanding of experimental ^{13}C-MFA with his lab's bioreactors and ^{13}C-labelled sugars. From the IIT Mandi metabolic systems lab, I would first like to thank the now Dr. Manu Shree Vishnoi for her reciprocal research visit to our lab and her advice on ^{13}C data processing, PhD student Jyotika Thakur for her advice on carbon transition network construction and PhD student Chanderkant Joshi for his assistance in the early 0.075 h^{-1} ^{13}C-MFA experiments, his advice on downstream data processing and for great conversations.

I also would like to thank my sister and mum for always supporting me and my Dad for always believing in me until the end. I would like to Dr. Iona Lewis-Smith my incredible partner, for being herself, for all the Sunday evening lifts to the lab to start cultures and for being my best friend through thick and thin. I love you very much.

Lastly, a quick shout-out to River T. Cat, adopted over the course of this PhD, who was a frequent source of entertainment and was always there for me when it was convenient for her.

Declaration of material from previously submitted thesis and of work done in conjunction with others

The genome scale metabolic model of *P. thermoglucosidasius* was initially developed by Dr. Beata Lisowska as the subject of her PhD thesis and is discussed in Chapter 3. The model was jointly developed to its presented state with Vivienne Mol at The Novo Nordisk Foundation Center for Biosustainability, Technical University of Denmark who also performed the experimental validation of media supplements and generated the associated figures. The work division is declared at the beginning of Chapter 3.2.
The experimental estimations of biomass components were performed by Dr. Shyam Masakapalli and Dr. Masakapalli provided the unprocessed isotopically instationary ^{13}C Xylose metabolite data for *P. thermoglucosidaisus* grown under anaerobic conditions.

Abstract
As global society transitions away from fossil fuels there is a need to produce chemical products from renewable precursors. Microbial fermentation is one method to achieve this aim which can be developed through metabolic engineering. Progressively, computational systems modelling and functional multi-omics analysis, are being applied to guide these metabolic engineering strategies. The research aimed to develop computational tools and acquire metabolic data to support this integrated system metabolic engineering approach for the thermophilic, facultative anaerobe *P. thermoglucosidasius NCIMB 11955*, a microbial chassis which offers the potential of being developed for sustainable bioconversion of renewable lignocellulosic waste to numerous products through its thermophilicity and catabolic versatility. This research built upon the foundation of an existing genome-scale metabolic model (GSMM) of *P. thermoglucosidasius NCIMB 11955* such that it could perform genome-wide analysis of *P. thermoglucosidasius* metabolism. This predicted experimentally-supported results demonstrating that a combination of thiamine, biotin and iron(III) could support anaerobic growth of *P. thermoglucosidasius* and identified potentially oxygen-dependent biochemical pathways to critical metabolites for anaerobic growth. This research also generated fluxomic data of *P. thermoglucosidasius* metabolism through a dynamic feeding *in vivo* isotopic tracer approach, known as isotopically instationary ^{13}C-Metabolic Flux Analysis (INST-^{13}C-MFA). This research presents the first INST-^{13}C-MFA data sets for a *Parageobacillus* species grown on glucose and xylose under aerobic and anaerobic conditions at a range of growth rates. A workflow for this analysis was established involving the evaluation of custom micro-bioreactors run as chemostats and a combined HPLC and GC-MS approach. However, statistically acceptable flux distributions models to represent this data remain a work in progress. Ultimately, with a more accurate, and ideally fluxomics-constrained, GSMM to act as a reference, existing strain design techniques could be used to develop *P. thermoglucosidasius* as microbial chassis for sustainable bioprocesses that could use waste lignocellulosic material as feedstocks.

Abbreviations

µL – microliter	Mal - malate
^{13}C-MFA - 13C Metabolic Flux Analysis	MCS - Multiple Cloning Sites
3PG - 3-phosphoglyceric acid	MDV - Mass Distribution Vector
6PD - 6-phosphogluconate dehydrogenase	ME - Metabolic Engineering
6PG - 6-phosphogluconate	MEETHF - methylenetetrahydrofolate,
	METHF - methyltetrahydrofolate
6PGL - 6-Phosphogluconolactonase	MIDs - Mass Isotopomer Distributions
AcCoA - Acetyl-coenzyme A	MKI - 45 mL working volume bioreactor vessel
AKG - α-ketoglutarate	MKI - 90 mL working volume bioreactor vessel
ASM - Ammonium Salts Media	MLE - Metabolically Less Efficient, with reference to pFBA
B. subtilis - *Bacillus subtilis*	MOMA - Minimization of Metabolic Adjustment
BDO - 1-4 Butanediol	MS - Mass Spectrometry
CCF - Cell culture Filtrate	MtBSTFA - N-tert-butyldimethylsilyl-N-methyltrifluoroacetamide
CCM - Central Carbon Metabolism	mV - Millivolts
CCVCs - cell culture volume changes	NCIMB - National Collection of Industrial & Marine Bacteria
Chor - Chorismate	NIST - National Institute of Standards and Technology
Cit - Citrate	NMR - Nuclear Magnetic Resonance
CN - Copy Number, relating to plasmids	non-oxPP - The non-oxidative branch of the pentose phosphate pathway
COBRA - Constraint-Based Reconstruction and Analysis	OAA - oxaloacetate
CSTR - Continuous Stirred Tank Reactor	OAA - Oxaoloacetate
DAD - Diode Array Detector	OD600 - The optical density of cell culture samples at a wavelength of 600 nm
DCW - Dry Cell Weight	ODE - Ordinary Differential Equation
dFBA - Dynamic Flux Balance Analysis	ORF - Open reading Frame
DHAP - Dihydroxyacetone phosphate	ORP - ordinary redox potential
DI-H_2O - De-ionized water	oxPPP - The oxidative branch of the pentose phosphate pathway
DO_2 - Dissolved Oxygen	*P. thermoglucosidasius* - *Parageobacillus thermoglucosidasius*

E. coli - *Escherichia coli*	PDH - pyruvate dehydrogenase complex
E.C. - enzyme commission	PEP - phosphoenolpyruvate
E4P - erythrose-4-phosphate	pFBA - Parsimonius Flux Balance Analysis
ED - Entner-Doudoroff	PFL - pyruvate formate lyase
EI - Electrospray ionisation	PID - proportional integral derivative
ELE - Enzymatically Less Efficient, with reference to pFBA	PYR - Pyruvate
EMU - Elementary Metabolic Units	R5P - ribose-5-phosphate
F6P - fructose-6-phosphate	RBS - Ribosome Binding Site
FBA - Flux Balance Analysis	rFBA - Regulatory Flux Balance Analysis
FBP - fructose-bis-phosphate	RID- Refractive Index Detector
FTHF - formyltetrahydrofolate	
Fum - fumarate	RPM - Revolutions per Minute
FVA - Flux Variability Analysis	Ru5P - ribulose-5-phosphate
G6P - glucose-6-phosphate	Ru5PE - ribulose-5-phosphate epimerase
G6PD - glucose-6-phosphate dehydrogenase	S7P - sedo-heptulose-7-phosphate
GAP - Glyceraldehyde-3-phosphate	SB - Synthetic Biology
GAPDH - Glyceraldehyde 3-phosphate dehydrogenase	SDGs - Sustainable Development Goals
GC-MS - Gas Chromatography-Mass Spectrometry	sf-GFP - super-folded Green Florescent Protein
GEM - Genome Scale Metabolic Model	SSR - Sum of the squared residuals
GPR - Gene-Protein-Reaction	Suc - Succinate
GSMM - Genome-Scale Metabolic Model	SysME - Systems Metabolic Engineering
HCl - Hydrochloric Acid	TA - Transaldolase
HPLC - High-Performance Liquid Chromatography	TBDMS - tert -Butyldimethylsilyl
HPLC-DAD-RID - High-Performance Liquid-Chromatography-Dioade Array Detector-Refractive Index Detector	*t*-BSMSCl - *tert*-butyldimethylchlorosilane
HPLC-RID-VWD - High-Performance Liquid-Chromatography-Refractive Index Detector-Variable-Wavelength Detector	TCA - tricarboxylic acid
ICDHc - cytosolic isocitratedehydrogenase	TIC - Total Ion Current
ICit - isocitrate	TK - Transketolase
IDV - Isotopomer Distribution vector	TK-C2 - Transketolase C-2
IIT Mandi - Indian Institue of Technology, Mandi	TPP - Thiamin pyrophosphate
ILEs - Isotope Labelling Experiments	TSBA - Tryptone Soya Broth Agar

INCA - Isotopomer Network Compartmental Analysis	UB - Upper Bound
INST-^{13}C-MFA - Isotpically Instationary 13C metabolic flux analysis	UN - United Nations
KIE - Kinetic Isotope Effects	VVM - Ratio of Volume of Air to Volume of media
KOH - Potassium Hydroxide	VWD - Variable-Wavelength Detector
LB - Lower Bound	Xu5P - xylulose-5-phosphate
LC-MS - liquid chromatography-mass spectrometry	
m/z – mass-to-charge ratio	

Table of Figures

Table of Tables

1. Chapter 1: Introduction

1.1 Biotechnology and Sustainable Development Biotechnological fermentation – An advantageous sustainable solution to finite resources

A pervasive issue with the modern synthetic chemical industry is that many of the fine and commodity chemicals produced are reliant on the underlying use of finite, petroleum-derived feedstocks. Such feedstocks are extracted, processed, and used in manners that are unsustainable and ultimately contribute towards greenhouse gas production and climate change[1]. As a result, there is an ongoing culturally driven industrial shift away from these finite fossil fuel resources towards more environmentally friendly production of bio-based chemicals and materials derived from renewable resources to reduce contributions towards climate change.

This has spurred research to investigate alternative, and more sustainable chemical technologies able to meet our current chemical production demands in a manner which is sustainable for the nations of global society.

1.1.1 What are the UN sustainable development goals?

In order to establish a global consensus on how to develop society in a more sustainable manner, in 2016 the United Nations published a collection of 17 Sustainable Development Goals (SDGs) and 169 targets, representing the three main pillars of sustainability of environmental protection, economic viability and social equity (Figure 1). This drive towards sustainable development underpins the purpose of the research presented in this thesis.

Figure 1. Promotional material detailing the 17 UN Sustainable Development goals.

1.1.2 Microbial biotechnological fermentation – part of a sustainable solution?

One biotechnological strategy that could be part of a sustainable solution which addresses these goals is microbial fermentation for the production of organic chemicals. This can utilise the biosynthetic capabilities of the microbial cells themselves to produce both naturally occurring and non-natural commodity and fine chemicals, polymers and alternative "renewable fuel" molecules such as biofuels and hydrogen [2].

Microbial fermentations, even when integrated as an upstream process into the existing architecture of synthetic organic manufacture, can offer several potential sustainability benefits [3]. The sustainability benefits of these microbial fermentations can include simultaneously economically and environmentally beneficial factors such as: high atom efficiencies, little reliance on external heavy metal catalysts and regular performance in water, the greenest of solvents [4]. Furthermore, microbial fermentations typically operate at more ambient temperatures, lower pressures, and milder pH values close to neutrality [5] . Several of the host microbes of these fermentations also naturally utilize carbon substrates that can be derived from renewable, inedible and inexpensive plant-based feedstocks. Some microbial hosts are even able to utilize C1 carbon sources which can operate as greenhouse gases, including: methane (CH_4) [6] , CO_2 and even CO [7], offering the promise of bioprocesses able to simultaneously sequester greenhouse gases, and produce renewable carbon products.

The use of microbes is of particular advantage when targeting large and complex natural products from secondary metabolic pathways of microbes including complex antibiotics such as erythromycin [8], polyketide products and non-ribosomal peptide products[9] . Indeed, many terpenoid natural products, which can be found naturally in microbes as well as higher organisms, can conventionally require low yield extractions from plants and animals or could otherwise require long and inefficient total chemical syntheses. As such, microbial fermentations that can uncouple the synthesis of such natural products from the environment and operate under controlled conditions can offer the additional advantage of ensuring a consistent and reliable supply of geographically distant or seasonal products, unaffected by geopolitics and climate change.

Notably, a cost-benefit analysis by Straathof and Bampouli (2017) of potential microbial production of commodity chemicals versus petrochemical production suggested that more oxidized commodity chemicals such as: 1,4-butanediol, acrylic acid and acrylate esters could be competitively produced from carbohydrate sources[10].

Indeed, they suggested that the best chemicals to target would be those that would otherwise require multiple conversion steps if produced from petrochemical sources. However, as microbial cells have usually not evolved to maximize the production of a chemical of human interest, yields are often lower than their theoretical maxima and infrequently achieve the approximately 50 g/L yields suggested by Sun and Alper (2015) to be required for industrial scale processes of some commodity chemicals [11]. These processes have therefore only seen limited application as they

are commonly less economically competitive than their petroleum-based alternatives [3].

One strategy to improve the competitiveness of these more sustainable microbial fermentation approaches is to increase the titres, yields and productivity of the 'microbial cell factories' themselves through genetic modification of the cell's metabolic pathways [12] [13].

1.1.3 Microbial Metabolic Engineering and Synthetic Biology

Metabolic engineering (ME) enables the rationally guided modification of cellular networks through genetic engineering strategies for the introduction, deletion or exchange of target pieces of DNA to redirect the flow of carbon through microbial metabolism, in order to achieve new strains of microbes with enhanced production properties that can be termed 'microbial cell factories' [14]. These properties can directly include the key performance indicators of chemical production such as the product titre, rate of formation and the product yield per unit feedstock, but can also include enhancements to overall performance through modifications to make the microbe generally more resilient to imposed culture conditions, such as enhanced tolerance to potentially toxic products [15] .

Inspired by the engineering principles of modularity and the standardization of parts, and bolstered through an increased understanding of genomics, metabolism and genetic engineering, modern ME is often applied as part of the emerging discipline of "Synthetic Biology" (SB). SB can broadly be categorized into aims and methodologies to introduce characterization, standardization and modularisation of these pieces of DNA, or 'parts', to enable more rapid metabolic pathway optimization and microbial strain development [16] .

These genetic engineering strategies can include: the targeted removal of existing native enzymes or biosynthetic pathways, the up- or down-regulation of regulatory elements, the introduction of non-native heterologous genes to construct hybrid or whole completely new created pathways with non-native functionalities [17] to create non-native or even non-natural products (such as the bio-based production of 1,4 butanediol discussed in 1.3.4). Furthermore, additional strategies can focus on larger cellular architecture, such as the channelling of target substrates through the co-localization of pathways through the application of micro-compartments and synthetic protein scaffolds [18] .

SB relies on a 'chassis', which is an engineering term applied to the choice of microbial host which serves as the cell factory upon which the genetic engineering is performed. For a bacterial chassis, a common tool for DNA transfer into a cell is using modified plasmids as vectors to introduce new genes. Plasmids are a circular, horizontally transferrable extrachromosomal form of DNA that can encode situationally advantageous genes such as enhanced resistance to antibiotics or other xenobiotics or niche catabolic abilities. Plasmids are typically described in terms of their average 'copy number' (CN), which refers to the average number of copies of a given plasmid per cell. The CN spectrum ranges from 'high' (500–700) to 'low' (5–20), depending on the mechanism of control of replication and therefore the choice of plasmid vector can be considered a rough way of tuning overall protein expression.

A SB methodology can seek to engineer modularity into plasmids for the insertion of specific genes, for example through the introduction of collection specific enzymatic restriction recognition sites known as Multiple Cloning Sites (MCSs). A simple plasmid-based DNA vector consists of one or more origins of replication, which enable the plasmid to be replicated in a specific host, a gene encoding a selection marker, such as antibiotic resistance, and a modular unit for gene transcription called an expression cassette (Figure. 2), which can be inserted into an MCS. The vector may also contain additional genes and sequences which enable transfer by conjugation between different host organisms. Alternatively, in several cases plasmids could soon be forgone entirely in favour of techniques which enable direct modification of the organism's chromosome, such as recent CRISPR/Cas systems, useful for gene knockouts and small modifications [13].

A model synthetic expression cassette contains the gene of interest as well as regulatory DNA sequences in the forms of: a promoter sequence, a sequence encoding a ribosome binding site (RBS), and a sequence encoding a transcriptional terminator [19] (Figure 2). Expression cassettes can also include additional sites called operators which can affect the ability of genes to be transcribed. Through the lens of SB these 'parts' (the gene of interest and each of these control elements) can be separately manipulated to get the desired level of expression of an individual gene or operon. Due to the falling costs of DNA synthesis, direct synthesis of libraries of these parts is now becoming increasingly possible too. Therefore, the synthetic expression cassette matters for ME approaches to design new microbial cell factories as this ability to screen combinations of parts tune protein expression can help tackle issues associated with that expression and help key performance indicators of any process.

Examples of this can be seen for each element of the standard expression cassette. The DNA of the gene of interest is a coding sequence for a protein of interest. Metabolic engineering can seek to develop attributes of the expressed protein itself, particularly if that protein is an enzyme able to carry out catalysis of key metabolic reactions. Enzyme engineering can be performed to design novel enzymes which are better suited to catalyse a reaction or use a particular substrate, or to design enzymes which are more durable, with enhanced properties such as thermostability. Furthermore, particularly in the case of heterologous genes, the DNA sequences interest can also be 'codon-harmonized' to better align the gene to the codon usage frequencies of the microbial host expressing the protein [20], and hence improve the rate of transcription.

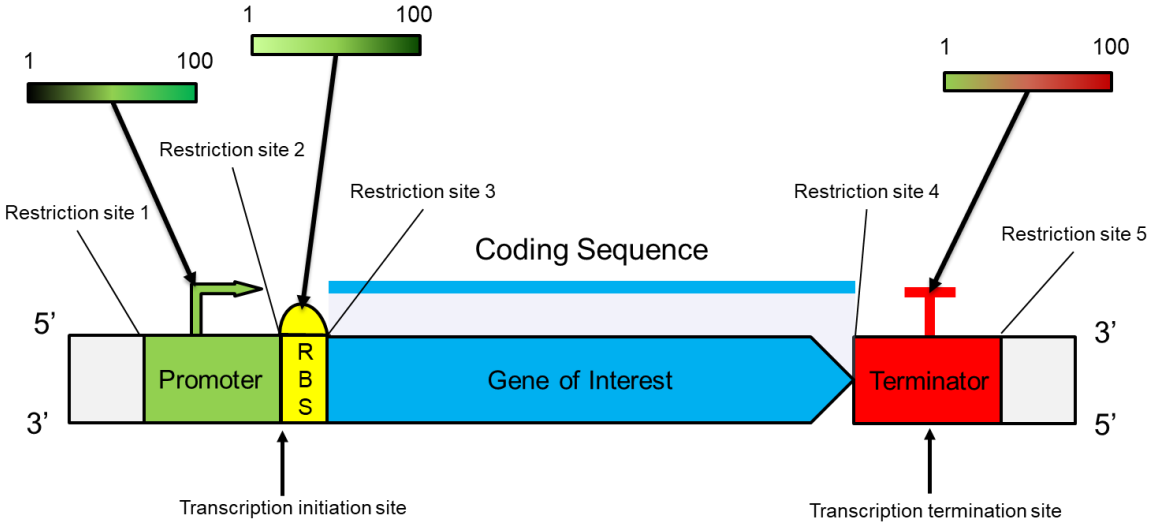


Figure 2. Schematic of a basic expression cassette within an MCS. A promoter, RBS, gene of interest and terminator are located within an MCS defined by 5 restriction sites. Bars labelled 1-100 with colour gradients represent a potential for creating libraries of different strengths for the associated control element.

The promoter sequence in an expression cassette acts as a control element for the initiation of transcription of the gene, defining where bacterial RNA polymerase, and some associated transcription factors, can bind to start transcription. They are typically classed as either constitutively expressed (i.e., always transcribing its associated gene/s) or as inducible by the presence or absence of a particular chemical signal or environmental stimulus [21].

The RBS sequence is located upstream of the start of the coding sequence of the gene of interest which, when transcribed, recruits the ribosome for the initiation of translation of the mRNA transcript and therefore acts as an mRNA-based control element of translation [22]. As a result, variations in the RBS sequence can be tuned for target translation initiation rates and, therefore, for the regulation of protein expression. In some cases, this can even be predicted from sequence alone. For example, 'the RBS calculator' algorithm developed by Howard Salis (2011) can not only predict rates of bacterial translation initiation based on start codons but can inversely generate synthetic RBS sequences designed for a particular rate of translation initiation.

Lastly, the choice of transcriptional terminator codon and terminator DNA sequence represent the end point in an operon, where transcription is terminated. These are arguably also targetable factors to influence overall rates of protein expression, depending on how they stabilize the mRNA transcript, as efficient termination of transcription will reduce the metabolic energy drain of transcription [23]. These termination strategies are classed as either intrinsic, relying on a G-C rich sequence which results in an RNA-polymerase-complex-destabilizing secondary structure architecture, or factor-dependent, relying on a circular ATP-driven protein Rho which translocates the mRNA sequence but blocks the RNA polymerase [24].

Ultimately, from a metabolic engineering perspective, the varying sequences for promoters, RBSs and terminators combined can therefore act as distinct 'tuning knobs' for the rate of protein production in microbial chasses and are often

described in terms of their relative 'strengths', that is, a relative spectrum of the capacity and effectiveness of these control element sequences when performing their specific roles. As metabolic engineering often demands fine control of protein expression, engineering-inspired research seeks also to construct libraries of RBSs and either constitutive or inducible promoters with different expression strengths, to fine tune the expression of a target proteins to pool levels desired[25]. Key examples of this are discussed for the organism central to this thesis in 1.3.4. However, despite the variety of possible approaches available to tune protein expression to solve ME problems, an ME strategy alone can miss the bigger picture of whole-cell metabolism.

1.1.4 The need for a systems view in metabolic engineering approaches

Cellular metabolism is a large, genetically encoded network (or genotype) of interconnected biochemical conversions. These conversions are predominantly catalysed by enzymes and their associated cofactors. This genotype is acted upon by environmental stimuli presented to the cells and converts nutrient substrates into the necessary macromolecules to maintain life and to produce new cells in an expressed pattern of metabolic behaviour known as a phenotype [26].

Often when aiming to developing the metabolism of an organism into a phenotype the production of a particular biochemical product, focus is understandably applied to metabolic engineering approaches which aim to divert as much carbon flux as possible towards increasing the product yield of a desired metabolite under a set of environmental growth conditions.

However, while proof of principle systems can be engineered (as discussed for *P. thermoglucosidasius* in 1.3.4), many of these rationally designed overexpression systems can be metabolically throttled and therefore unable to reach industrially acceptable product yields.

Recurrent problems can occur when these proof of principle strategies do not consider the inter-dependent nature of cellular metabolism operating as a whole network. The potential global impacts of enhancing metabolic flux towards a target compound can distort the overall metabolic homeostasis of the cell, particularly if overproduction of the target compound is in direct competition for resources needed for survival and growth. This competition for carbon, cofactors and energy can result in metabolic carbon imbalances, like the accumulation of carbon intermediates such as pyruvate [27], as well as imbalances in redox metabolism which can diminish yields [28] [29]. For example, an introduced and deliberately overexpressed biosynthetic pathway for a biochemical product could demand a particular universal co-factor such as: NADH, NADPH or FADH or draw from or distort a key pool of carbon metabolites, such as glyceraldehyde-3-phosphate, pyruvate or acetyl-CoA, that would otherwise contribute towards cell growth. If this demand was not accounted for with a compensatory change to ensure redox balance, enhanced carbon uptake and/or tie the product to cell growth, the process would not only fail to reach its theoretical potential, but could also impact upon the cell's growth rate, further reducing potential overall fermentation yields of product [30].

Underpinning an efficient ME approach therefore is the requirement of a systems-level understanding of the metabolic reaction network of the organism in order to

design a ME strategy which achieves both over-production and balanced growth. Thus, there is a need to leverage systems biology approaches [3].

1.2 Systems Metabolic Engineering for Sustainable Development

In parallel with the advances in SB, there have been concurrent developments in systems biology generating experimental genomic, transcriptomic, proteomic, metabolomic and fluxomic data which underpins the biochemical networks of cellular metabolism, as well as the construction and validation of genome-scale *in silico* models able to predict cellular behaviours and suggest modifications for particular strain design criteria [31 32].

This integration of the inter-dependent fields of traditional ME approaches with computational modelling and so-called '-omics' data analysis results in the design-build-test philosophy for the development of microbial cell factories that has been termed 'Systems Metabolic Engineering' (SysME)[33]. From the perspective of sustainable development, the SysME approach for the generation of compounds from renewable feedstocks has been directly discussed by Yang *et al.* (2017) as supporting up to 9 of the 17 UN SDGs (Figure 3) [34].

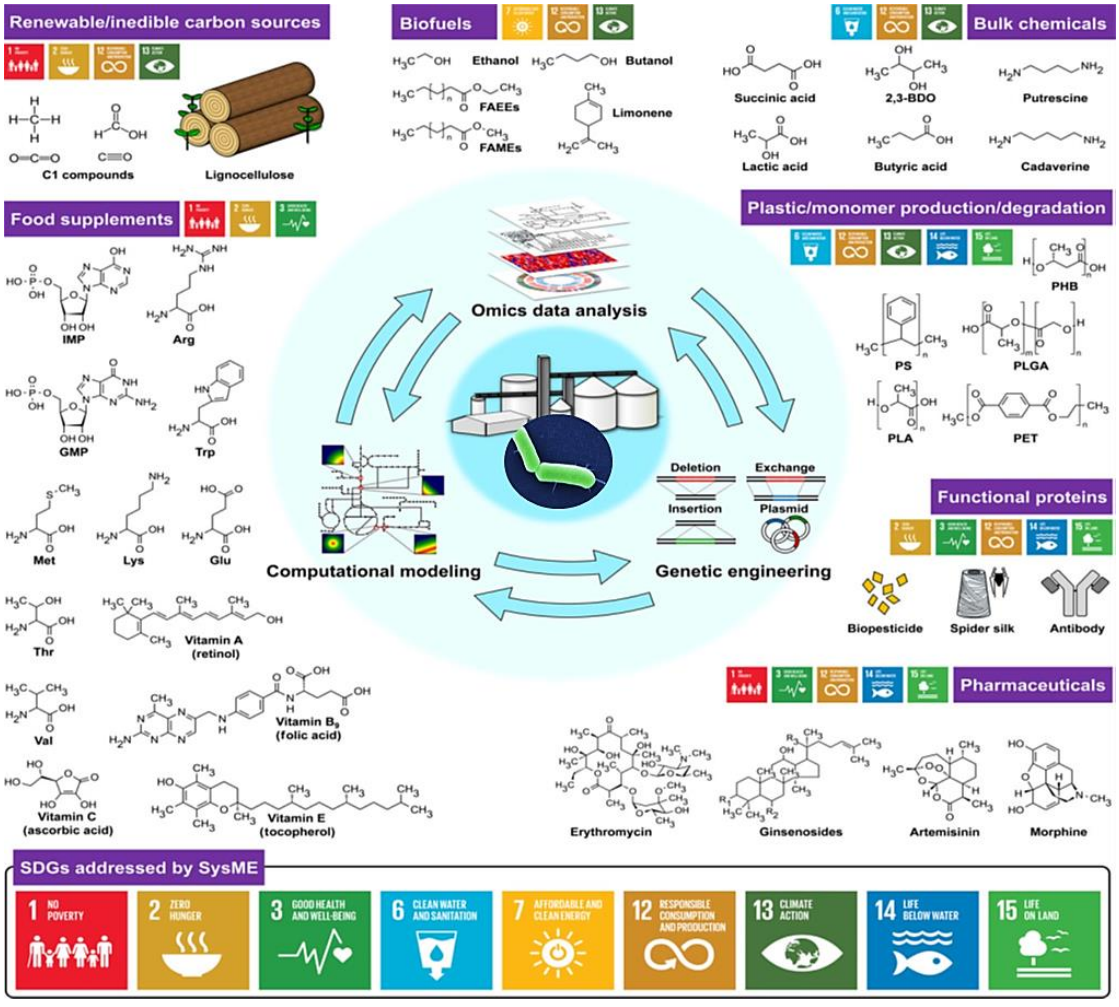


Figure 3. The interplay between SysME and the UN Sustainable Development goals. Taken from Yang, Cho et *al.* (2017)[34].

1.2.1 Computational modelling: Genome-scale metabolic modelling and Flux Balance Analysis

From the perspective of computational modelling, as the whole-genome sequencing of organisms has become cheaper, more and more genome sequences of different organisms have become available. Concurrent to this, our computational processing power has increased and computational applications of these knowledge databases of genetic information have enabled (gene-protein) predictive tools to be developed and refined using these genomes to enhance our understanding of the whole metabolic genotype–phenotype relationships of organisms we wish to develop [35]. A particular form of computational modelling focused upon in this thesis has been the emergence of the aforementioned organism-specific genome-scale metabolic model (commonly abbreviated as GSMMs or GEMs) which catalogues all known genes, metabolites and metabolic reactions of an organism [36], forming a knowledge base in which the presence of each reaction in the network is supported by an annotated gene sequence encoding a particular enzyme [35 36]. While such metabolic network

databases are valuable on their own, the accurate metabolic behaviours of organisms cannot be described without also describing the behaviour and influence of the environments which they inhabit [38]. Therefore, to gain an accurate insight into this unified field of behaviour of metabolic reactions and environment, or phenotype, these knowledge databases need to be interrogated on a systems or network level where there can be thousands of difficult to measure variables influencing cellular metabolic behaviour simultaneously.

To enable us to comprehend the number of variables involved in describing and interrogating this metabolic behaviour, genome-scale metabolic modelling has typically relied on the Constraint-Based Reconstruction and Analysis (COBRA) methodology Flux Balance Analysis (FBA) approach, under an assumption of intracellular steady-state concentrations of metabolites, to predict particular phenotypes [39 40 41]. This approach converts this knowledge database of a metabolic network into a mathematical structural model which quantitatively describes the relationships between the species or metabolites involved in the system and, through the application of constraints and objective functions, enables the prediction of optimal steady-state flux distributions for all metabolite species in all reactions in the metabolic network reconstruction.

Once the metabolic structural framework is formulated, experimental measurements such as substrate uptake rates, growth rates, growth yields and maintenance energy can be incorporated to constrain this framework. FBA can then be performed on a genome-scale metabolic model to establish the theoretical maximum and minimum flux limits of the metabolic network of an organism. This enables the rapid *in silico* simulation of specific phenotype hypotheses concerning *in vivo* maintenance requirements and growth rates of organisms: on specific substrates and media compositions, along the boundaries of theoretically feasible or infeasible cell growth and under niche or difficult to achieve experimental conditions, such as the 50% carbon monoxide atmosphere demonstrated by Mohr *et al.* (2018) to enable H_2 production in *P. thermoglucosidasius DSM 2542* [42] .

From a systems metabolic engineering perspective, FBA of a completed model allows it to be used to identify metabolic imbalances in the network that could throttle product yields, such as redox co-factor asymmetry or the accumulation of metabolic intermediates. Additionally, a completed GSMM can also be used as a platform for *in-silico* metabolic pathway design and testing to suggest system-wide genetic modifications of an organism to achieve the theoretical maximum yields of a particular metabolite [43 44], evaluate the systems-wide effects of those alterations and evaluate the impact of specific environmental conditions, such as oxygen availability and nutrient supplementation on the growth of the organism. A deeper explanation of GSMMs and FBA is provided at the start of Chapter 3.

Ultimately however, metabolic flux predictions generated by FBA are only a good as the accuracy of the underlying networks and choice of assumed constraints. Thus, not only is ensuring adequate curation a necessity but, as a huge number of assumptions underpin simulation outputs, any results also need to be validated against experimental measurements under the same simulated conditions before it can be used to inform new metabolic engineering designs [45 46].

1.2.2 The umbrella field of '-omics' data analysis

In systems biology, experimental quantifiable measurements are achieved through the different '-omics' analyses of: genomics, transcriptomics, proteomics, metabolomics and fluxomics, all of which can help inform strain development strategies [47] Each type of these analyses is dedicated to the investigation of cellular physiology and metabolism through the lenses of the different scales and stages of the central dogma of molecular biology (Figure. 4).

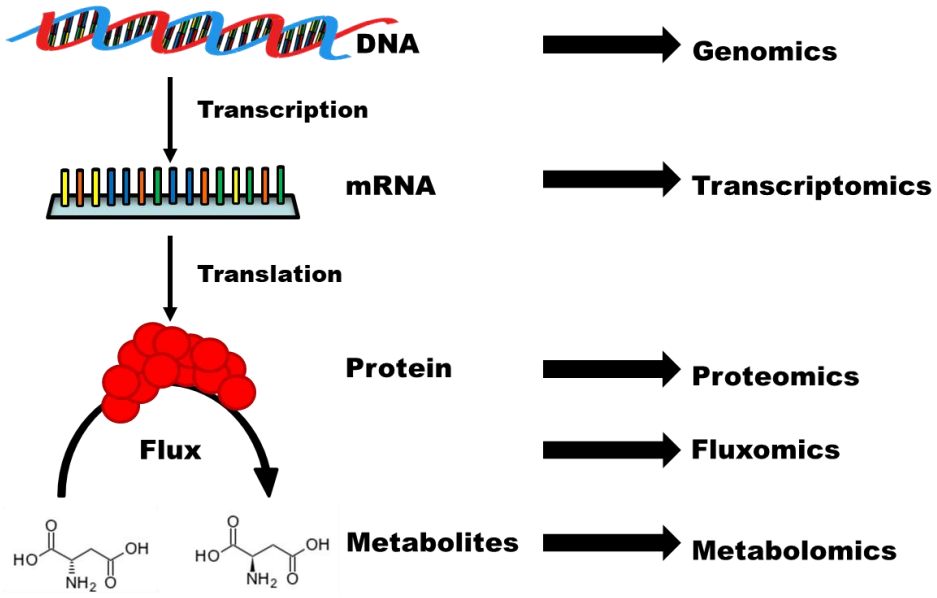


Figure 4. Schematic of the central dogma of molecular biology describing the associated -omics fields of study.

Genomics is the analysis of the structures and functions of the coding and non-coding DNA sequences which collectively constitute the genome of an organism. Genomics can therefore aid in the identification and understanding of the genetic basis of a particular metabolic behaviour and forms the basis of *the P. thermoglucosidasius* GSMM of Chapter 3. However, while FBA on a GSMM can predict pathway fluxes, without verification or applied constraints derived from the other '-omics' analyses, the predicted flux distribution may not be an accurate representation of cellular metabolic behaviour under target growth environments studies. Indeed, the presence of a gene or associated final enzymatic reaction does not guarantee its expression.

Such information can be measured and integrated by using transcriptomics, the study of expression levels of all the RNA molecules in a cell, and by using proteomics, the study of the complete protein expression profile of a cellular network under a given set of conditions [48], but this research focuses on metabolomics and fluxomics.

Metabolomics is the study of the complete profile of cellular metabolites produced by a system under particular conditions, including: amino acids, lipids,

carbohydrates, nucleotides and any target metabolite products such as lactic acid and ethanol 49. The concentrations of metabolites observed through metabolomics studies represent the final biochemical products of metabolism. Hence, these metabolites can signify the final behavioural responses of a metabolically engineering system to its environment and any genetic modifications. Fluxomics is an applied extension of metabolomics and is the study of the complete profile of the metabolic fluxes which produce the metabolome, a profile called a fluxome or flux distribution.

1.2.3 Fluxomics and ^{13}C Metabolic flux analysis

This research focuses primarily on fluxomics, in particular ^{13}C metabolic flux analysis (^{13}C-MFA) informed by ^{13}C isotope labelling experiments (ILEs). Fluxomics enables the estimation of *in-vivo* intracellular activity of not directly measurable metabolic reaction rates for both reactions and whole metabolic pathways through post-hoc analysis of cell growth experiments [50]. It is most commonly based on the feeding of growing cells with heavy stable isotope labelled tracers (including ^{2}H, ^{13}C, ^{15}N, ^{34}S, etc.) which incorporate into metabolites in fixed and predictable ways [51]. Such downstream metabolites (typically stable proteinogenic amino acids and fatty acids) can be analysed and measured by Gas Chromatography-Mass Spectrometry (GC-MS) and/or Nuclear Magnetic Resonance (NMR) to reveal their heavy isotope labelling patterns. Having developed out of the computational FBA field, the quantification of fluxes in ^{13}C-MFA is performed through defining a stoichiometric metabolic reaction network featuring, at its centre, the carbon atom transitions associated with each reaction leading from the substrate to the analyte [52]. The *in silico* metabolic flux through this carbon transition network is then constrained to match the experimentally measured rates of carbon consumption and excretion and, most importantly, the ^{13}C labelling patterns of downstream metabolites. Non-linear programming is used to minimize the sum of the squared residuals (SSR) between the (forward) simulated flux distributions derived from the metabolic network and choice of isotopic tracer, and the experimentally determined isotope labelling patterns themselves. If an acceptable fit, as determined by a certain SSR threshold, ^{13}C-MFA yields a model which describes the ^{13}C labelling pattern data of the analysed metabolites in terms of a flux distribution for the reactions for central carbon metabolism, complete with associated flux error. This therefore quantitatively describes the *in vivo* metabolic phenotype of the cell under the specific growth conditions used. These metabolic flux rates represent a function of the final catalytic rates of enzymes in pathways and for this reason are particularly valuable when knowledge of the underlying genomic to proteomic networks are not well understood, as is the case for the organism central to this thesis.

As a result, fluxomics and ^{13}C-MFA enables the most accurate systems level description of cellular metabolism among the '-omics' strategies and can help explain the phenotypic behaviours of both Wild-Type and ME organisms [48]. In addition to being directly relatable to FBA solutions of GSMMs, ^{13}C-MFA flux distribution outputs can also identify targets for further metabolic engineering. Fluxomics and ^{13}C-MFA are described in more detail in Chapter 4.1.

1.2.4 The combined SysME approach

Through the integration of genetic engineering, computational modelling and -omics analysis, SysME's design-build-test framework has been used to develop strains beyond the point which purely genetic engineering strategies could accomplish alone.

A key sustainability-centric example which brings these fields together, is the research concerning the hereteologus bio-catalytic production of 1-4-Butanediol (BDO) in *E.coli* from renewable carbohydrate feedstocks. The non-natural chemical intermediate BDO is typically synthesised through petrochemical processes and can be used in the manufacture of plastic and elastic polymers, such as those in Lyrca/Spandex. Yim *et al.* (2011), used an iterative SysME design-test-build cycle of host improvement to reach industrially viable titres of BDO [53]. Initially, Yim *et al.* (2011) constructed a synthetic operon a high-copy number plasmid of: a lactose-regulated inducible promoter PA1, the *sucCD*, *sucD* and *4hbd* genes each with individual RBSs and a T1 transcriptional terminator. This was coupled with a 4-hydroxybutyryl-CoA transferase (*cat2*) on a separate medium-copy number plasmid and then both plasmids were transformed into *E coli*. This was then followed by the application of the FBA-variant OptKnock, discussed in 3.1.8, on a highly curated GSMM of *E. coli*, which suggested gene knockouts of: lactate dehydrogenase A (Δ*ldhA*), pyruvate formate lyase (Δ*pflB*), alcohol dehydrogenase (Δ*adhE*) and malate dehydrogenase (Δ*mdhA*) to improve the anaerobic capabilities of the chassis citric acid cycle. When applied to the *E. coli* chassis, this led to a fed-batch fermentation yield of 18 g L^{-1} of BDO after 130 hours [53]. Notably from a sustainability perspective, while the highest BDO yields were obtained from growth on dextrose, BDO production was observed to be equally efficient on a range of pure and mixed C_5 and C_6 sugars. This suggests that a future process could be developed to use lignocellulosic biomass hydrolysates as a feedstock.

In a later study, the use of transcriptomics and ^{13}C fluxomics on this strain was then able to identify a metabolic bottleneck resulting in the accumulation of pyruvate and acetate, which suggested competition for flux between the heterologous BDO pathway and growth. This led to the deletion of two succinate semialdehyde dehydrogenases and a resulting elevated titre of 29 g/L of BDO [53] . A further competing route from the BDO pathway to the TCA cycle was identified and removed, enzyme engineering was performed to generate improved variants of the 4-hydroxybutyryl-CoA transferase and aldehyde dehydrogenase enzymes, and the complete operon was integrated into the *E. coli* chassis chromosome. This led to a final, commercially viable, yield of BDO of over 110 g/L [53].

Overall, this demonstrates the value of the holistic SysME approach in identifying and addressing metabolic bottlenecks for the development of high-performance microbial chassis and industrially viable sustainable bioprocesses. With the underpinning methodology of the thesis defined, the selected microbial chassis to which this strategy aims to be applied can be introduced.

1.3 What is *P. thermoglucosidasius* NCIMB 11955?

Parageobaciillus thermoglucosidasius (Occasionally referred to as *P. thermoglucosidans* [54] NCIMB 11955 is a Gram-positive, facultative anaerobic, spore-forming, thermophilic bacteria ripe for development as an industrial microbial chassis and is the organism central to this thesis [55].

1.3.1 Phylogenetic history

Parageobacillus thermoglucosidasius was initially described and classified as a thermophilic species of genus *Bacillus* in 1983[56]. However, after a full suite of 16s rRNA and morphological characterisation by Nazina *et al.* (2001), a new genus of *Geobacillus*, containing the thermophilic group 5 of the former *Bacillus* genus including the species, *thermoglucosidasius*, was defined [57].
More recently, the genus *Geobacillus* was differentiated further into two phylogenetically close, but distinct clades [54]. The first of these clades, with a G-C content range of 48.5-53.1% remained as the genus *Geobacillus*. However, the second of these clades, with a G-C content range of 42.1-44.4%, was separated into the new genus *Parageobaciillus* containing our organism of interest *Parageobacillus thermoglucosidasius* NCIMB 11955. A review of the taxonomy of the genera *Geobacillus* and *Parageobacillus* by Najar and Thakur in 2020 [58] suggests that there are currently 4-5 species of *Parageobacillus* and 13-14 species of *Geobacillus*.

Theorised to be as a result of being spore-forming extremophiles with very low buoyant density and easy to disperse spores, *Parageobaciillus* and *Geobacillus* species have also been of particular ecological interest as they have been isolated from a diverse range of environments [59] [60] . These include the isolation of: a *Geobacillus* species from the Mariana trench [61], a *Parageobacillus* species from marine sediment [62], a *Parageobacillus thermoglucosidasius* strain C56-Y93 from hot springs [63], isolation of *Geobacillus tobeii* from hay compost [64], a *P. thermoglucosidasius* strain TNO-09.020 from a dairy processing environment [65] and a strain of *Geobacillus thermodenitrificans* from a 2000 m deep oil reservoir [66].

Furthermore, observation of rapid adaptation and gene-diversity in the genus *Geobacillus* by Hirokazu Suzuki (2018) have even led the author to postulate that *Geobacillus* sp. adapt so quickly due to a combination of high horizontal-gene transfer [67] and an ability to perform stress-induced mutagenesis when faced with nutrient limitation or xenobiotic stressors [68] . For example, exposure of *G. thermoantarticus* to UV radiation, water and temperature conditions reflecting outer space by Di Donato *et al.* (2018) did not kill the organism, and it was suggested to have survived through rapid adaptation of its protein and lipid composition and additional protection of its DNA [69].

1.3.2 What makes *P. thermoglucosidasius* interesting from a sustainable industrial biotechnology perspective?

There is currently a narrow range of model bacterial chassis frequently used for industrial microbial fermentation, but these may not represent the best organisms for biosynthetic processes or target chemicals. In that respect, *P. thermoglucosidasius* has several metabolic characteristics which confer advantages when compared to such conventional model bacterial chassis including *Escherichia coli* [70] and *Bacillus subtilis* [71] [72].

Thermophilicity

The first of these traits is the moderately thermophilic (or ‘heat-loving’) nature of both *Parageobaciillus* and *Geobacillus* species, with growth temperatures which range

from 42-69°C [73]. While the molecular basis of this thermoadaptation is not yet clear, it is suggested to be connected to their high genomic G-C content and their abilities to synthesise protamine and spermidine, compounds thought to be able to stabilise nucleic acids [74]. The ability of *P. thermoglucosidasius* to grow between 50°C and 60°C is of value for large scale fermentations compared with mesophiles for several reasons.

Not only do enhanced temperatures enable higher chemical conversion rates and processes with multi-stage temperature procedures, but they also reduce the general risk of fermentation contamination with mesophilic species [75]. Notably, as biotechnological fermentations are exothermic processes, operating temperatures of 50-60°C remove the need for significant cooling of large-scale fermentations [76] . Furthermore, at their optimum growth temperatures, *Parageobaciillus* and *Geobacillus* species can achieve growth rates that rival mesophilic industry favourite *Escherichia coli*. Additionally, 50-55ºC is the optimum temperature for many commercial cellulases, making simultaneous saccharification and fermentation a realistic prospect.

Catabolic Versatility

The second key trait of *P. thermoglucosidasius* is their catabolic versatility which supports them in degrading a broad range of polymeric substrates. This could ultimately enable them to feature as more efficient whole-cell biocatalysts [77].

Lignocellulosic biomass, derived from waste plant material, represents a globally abundant potential source of renewable carbon for fermentation [78]. This complex polymeric material can contain between 65-85% carbohydrate, with hemicellulose (mainly composed of pentoses) and cellulose (hexose) being the main components, depending on its source. While this requires physico-chemical pre-treatment to separate these sugars from the aromatic polymer lignin and other non-fermentable components, hydrolysates of these sugars can be microbially fermented [79]. Thus, in the development of more sustainable bioprocesses, a microbial chassis capable of valorising as much of this carbon as possible would be advantageous.

Parageobacillus and *Geobacillus* species are known to synthesise thermostable glucanases [80], xylanases [81] , arabinanases and pectinases [82] to digest these polysaccharides. *P. thermoglucosidasius* can then directly transport and metabolise both the pentose and hexose monomers and some short-chain oligosaccharides [83] . The ability to transport short oligosaccharides, including cellobiose and branched hemicellulose-derived oligosaccharides for intracellular degradation offers considerable savings in enzyme pretreatment costs and is not found in any conventional industrial yeasts [84]. Overall, *P. thermoglucosidasius* represents a promising bacterial chassis that could be engineered into microbial cell factories for new, industrial viable sustainable bioprocesses [85].

1.3.3 What genetic engineering tools are available for *P. thermoglucosidasius* NCIMB 11955?

To facilitate the improvement of *P. thermoglucosidasius* as a biotechnological chassis, several specific tools have been developed to enable its genetic modification, resulting in it being described by Lau *et al.* (2021) as one of the few thermophiles that can currently be genetically manipulated [86] .

These include strategies for cell manipulation and DNA transfer to introduce new genes of interest, as well as the development of various genetic parts for operon and pathway engineering, tuning and validation. Transforming new genes into the 'undomesticated' gram-positive *P. thermoglucosidasius* is more challenging than with gram-negative species like *E. coli* as the gram-positive bacteria cell wall contains a thick peptidoglycan layer which can impact upon DNA uptake[87]. While initially electroporation was the preferred technique of DNA transfer, alternate DNA transformation strategies which rely on the conjugative horizontal transfer of plasmids to *P. thermoglucosidasius* from an *E. coli* donor have also been developed. In most cases, the initial plasmid vectors were derived from *Parageobacillus* species themselves. In most cases, the initial plasmid vectors were derived from *Parageobacillus* species themselves. These include the shuttle plasmid pUCG18 developed by Taylor *et al.* (2008) [88] and pTMO19 and pTMO31 developed by Cripps *et al.* (2009) [89] with the first conjugative vector, a pUCG18 derivative bearing an origin of transfer (or *oriT*) gene to enable conjugative transfer, developed by Suzuki and Yoshida (2012) [90] ; Kananavičiūtė and Čitavičius (2015) [91].

Other more direct genetic engineering strategies that introduce DNA directly into the chromosome of *P. thermoglucosidasius* have also been developed. For example, Bacon *et al.* (2017) streamlined a technique for the targeted integration or deletion of a gene of interest in the chromosome through a double-crossover, homologous recombination process which eliminated a previous negative screening step from the overall procedure [92]. Most recently, Lau *et al.* (2021) demonstrated not only deletion of the two native plasmids in *P. thermoglucosidasius,* but also the insertion of a gene encoding a fluorescent reporter through the development and testing of a CRISPR (clustered regularly interspaced short palindromic repeats)/ Cas9 genome editing system based on two Cas9 nucleases derived from *Streptococcus thermophilus,* a species of streptococcus which grows at slightly higher temperatures than *S. pyogenes,* the source of the original Cas9 [93].

At the same time, several research groups have developed additional synthetic biology tools for tuneable control over *P. thermoglucosidasius* gene expression. Two examples of this are Reeve *et al.* (2016) [94] and Pogrebynakov *et al*. (2017) [95] who took similar approaches to develop more modular genetic toolkits for engineering operons and pathways in *P. thermoglucosidasius*.

Reeve *et al.* (2016) [94] assembled two high copy number plasmids, 'PG1' and 'PG2' for *P. thermoglucosidasius* which possessed a multiple restriction cloning site and three reporter genes (including a gene for a super-folder green fluorescent protein (sf-GFP)), whereas Pogrebynakov *et al*. (2017) [95] modified the conjugative pUCG18 plasmid into the 'pIPGE' plasmid which incorporated a singular super-folded Green Florescent Protein (sf-GFP) reporter.

Both groups generated libraries of new promoters of various expression strengths to enable tuneable gene expression. Reeve *et al.* (2016) generated a library of 20 promoters through mutagenesis PCR of a constitutive high expression promoter P_{RP1} and tested their expression levels in both *P. thermoglucosidasius* and *E. coli* in order to consider both plasmid propagation in *E. coli* and the conjugative transformation process also involving *E. coli*. While only one of the new promoters had a higher expression level than the P_{RP1} template, the library generated did cover over a 100-fold range in expression levels in *P. thermoglucosidasius* between the weakest and strongest promoters. Comparably, Pogrebynakov *et al*. (2017)

generated a library of 17 constitutive promoters based on the groESL promoter. While again generating an over 100-fold range of expression strengths, the 'P1' and 'P2' promoters demonstrated higher levels of expression than the wild-type promoter. Additionally, Pogrebynakov *et al*. (2017) also isolated and validated a xylose-sensitive inducible promoter P_{xylA}, which could be of specific value to processes designed with hemicellulosic biomass in mind as particular production of any target product could be induced by xylose after initial cell growth.

Both groups also generated ribosome binding site sequence libraries using a predictive design method developed by Salis *et al.* (2009) [96]. Interestingly however, all the RBS sequences generated by both groups demonstrated transcript expression levels lower than their original parent RBS sequences, suggesting that the method, while able to generate initial expression libraries, may be less applicable to *P. thermoglucosidasius* and potentially thermophiles in general.

Overall, what these genetic engineering developments indicated is that the genetic toolkits for *P. thermoglucosidasius* are expanding and becoming increasingly accessible, suggesting that the bacterium is ripe for more diverse metabolic engineering strategies for new bioprocesses.

1.3.4 What has *P. thermoglucosidasius* been used to produce?

Using such genetic engineering tools, *P. thermoglucosidasius* has been engineered to produce a variety of biofuels and fine chemicals.
For example, the generation of the *P. thermoglucosidasius* TM242 strain by Cripps *et al.* (2009) produced a strain which able to achieve up to 95% of the theoretical yield of bioethanol [89], while the production of isobutanol in *P. thermoglucosidasius* by Lin *et al*. (2014) [97] achieved a maximum titre of 3.3 g/L isobutanol. From a sustainable fuel perspective, perhaps one of the most interesting recently discovered abilities of several *P. thermoglucosidasius* strains is the ability to produce the alternative biofuel of hydrogen from carbon monoxide via the Water-gas-shift reaction ($CO + H_2O \rightarrow CO_2 + H_2$) [98]. Further analysis indicated this reaction was catalysed by a genomically-encoded carbon monoxide dehydrogenase complex, suggesting that *P. thermoglucosidasius* could be developed further as a source of cleaner fuel from renewable lignocellulosic biomass [98 99]. Notably from the perspective of understanding the metabolism of *P. thermoglucosidasius*, transcriptomic analysis of *P. thermoglucosidasius* grown under CO and air suggested that the decrease in biomass and the start of expression of sporulation genes by *P. thermoglucosidasius* when O_2 is removed could be rescued through the introduction of CO[98]. This suggests that anaerobic growth is possible through feeding of CO which has the additional benefit of producing potentially valuable H_2 as a by-product of that growth.

Examples of commodity and fine chemicals that have been produced using *P. thermoglucosidasius* also include: the metabolic engineering of *P. thermoglucosidasius* to produce 1.03 g/L of Vitamin B2/riboflavin [100], the production of the production of 7.2 g/L 2-3-butanediol by Zhou *et al.* (2020) in the *P. thermoglucosidasius* strain DSM2542 [101] , and the recent proof of principle biosynthesis of isoprenoids from waste bread by Styles *et al.* (2020) [102].

Alternatively, when conventional mesophilic enzymes employed in industry are rendered inefficient by high temperature conditions [103], thermophilic variants of these enzymes sourced from *Parageobacillus* and *Geobacillus* species have the potential to be used instead. Examples of such thermostable enzymes include: amylases [104], carboxylesterases [105] [106], lipases [107] [108] [109], proteases [110] [111] and a DNA polymerase I from *G. stearathermophilus* GIM1.543 [112].

1.3.5 Why is there a need to understand more about *P. thermoglucosidasius*?

P. thermoglucosidasius is a bacterial chassis that is ripe for further development, offering the potential of sustainable bioconversion of renewable lignocellulosic waste to numerous products through its thermophilicity and catabolic versatility. Galvanising progress towards that goal, a burgeoning suite of metabolic engineering tools and standardized genetic parts have been developed. However, while several proof of principle fermentations for biochemical products have been performed, with even a few niche commercialized roles to produce ethanol and lactic acid, few have achieved economically viable yields.

A critical reason for this is that, despite increasing interest, *P. thermoglucosidasius*, and related *Geobacillus* species, remain understudied compared to more common bacterial chassis such as *E. coli* and *Bacillus subtilis*. As a result, a full system understanding is still lacking. Although it is understood that *P. thermoglucosidasius* cells maintain redox balance under micro-aerobic conditions through expression of genes for mixed-acid fermentation (acetic, lactic, and formic acid/formate) and ethanol fermentation, there remains a need to better understand the metabolic behaviour of *P. thermoglucosidasius* cells under micro-aerobic and truly anaerobic conditions for future metabolic engineering designs. For example, a recurring observation of anaerobic chemostat growth experiments of *P. thermoglucosidasius* cells is that washout will occur without supplementation of oxygen or yeast extract [89] [56] [113], suggesting a particular nutrient may be required for anaerobic growth. Furthermore, a transcriptomic analysis study by Loftie-Eaton *et al.* (2012) focused on the metabolic transcriptome changes *P. thermoglucosidasius* NCIMB 11955 cells as oxygen conditions were restricted [114]. In a comparison of the transcripts the authors suggested that, rather than substitute the respiratory electron transport chain for an anaerobic growth strategy when oxygen conditions were restricted, *P. thermoglucosidasius* NCIMB 11955 cells were continuing to operate in an oxygen-scavenging state and struggled under arguably their most valuable growth conditions from an industrial perspective.

1.4 Overall aims: The development of Systems Tools for Metabolic Engineering of *P. thermoglucosidasius* NCIMB 11955 for Bioprocess development

As a facultative anaerobe, *P. thermoglucosidasius* performs aerobic respiration and fermentation (production of: acetate, formate, lactate and ethanol) in response to intracellular oxygen conditions and largely influenced by environmental oxygen concentrations. Indeed, an understanding of the hierarchy of products of *P. thermoglucosidasius* fermentation metabolism led to the development of a high ethanol yielding strain through ME [89].

To aid the development of *P. thermoglucosidasius* NCIMB 11955 as a microbial chassis for sustainable biotechnological bioprocesses, there is a need to both better understand the fermentative metabolic behaviours of *P. thermoglucosidasius* NCIMB 11955 and to develop computational tools for strain design able to accurately predict cellular processes based on extracellular measurements (eg. dissolved oxygen concentration).

To help realise this overall aim, this researched has focused on the 'computational modelling' and '(flux)omics' analysis fields of the SysME approach. The objectives within these fields are as follows:

1.4.1 Computational genome-scale metabolic modelling and Flux Balance Analysis

Aim: The improvement and validation of a GSMM of *P. thermoglucosidasius* NCIMB 11955

The availability of a reliable GSMM of *P. thermoglucosidasius* would create a foundational platform for *P. thermoglucosidasius* strain design able to optimize flux through central metabolic pathways in a manner which would balance both target metabolite production and cell growth for maximum overall metabolite yields. This research developed a GSMM for *P. thermoglucosidasius* NCIMB 11955 which was initially annotated, curated and constrained by Dr. Beata Lisowska [113]. When this research began network evaluation investigations it discovered a combination of issues with the initial model including: several stoichiometrically and mass unbalanced reactions, reaction connectivity issues and a few unrealistic transport reactions. This meant that model in its original state did not yet accurately represent growth of *P. thermoglucosidasisus* NCIMB 11955.

The overall objectives of the research were therefore as follows:

1. To identify and improve existing inaccuracies and issues within the model;
2. To validate the *in-silico* model against existing experimental data; and
3. To use the model to suggest how *P. thermoglucosidasius* grows under anaerobic conditions and what could be done to support anaerobic growth.

The research towards these objectives is described in Chapter 3.

1.4.2 Fluxomics Data Analysis

Aim: INST-^{13}C-MFA of *P. thermoglucosidasius* NCIMB 11955

The use of a fluxomics technique is of particular advantage in this case as the metabolic behaviours of *P. thermoglucosidasius* are poorly understood when compared to many model chassis. This research performed ^{13}C-MFA to combine new experimentally determined ^{13}C labelling data, and previously unprocessed anaerobic ^{13}C-xylose data (Masakapalli S.K. (2014), unpublished) with the additional novel development of detailing isotopically in-stationary ^{13}C isotope labelling of *P. thermoglucosidasius* NCIMB 11955 amino acids over time (INST-^{13}C-MFA).

The overall objectives of the research were therefore as follows:

1. To establish a workflow for ^{13}C-MFA at the University of Bath.
2. To perform isotopically instationary ^{13}C labelling experiments for cultures of *P. thermoglucosidasius* NCIMB 11955 when grown:
   a) Under aerobic and anaerobic/fermentative conditions;
   b) At a range of growth rates; and
   c) On lignocellulosic biomass model substrates of either glucose or xylose as sole carbon sources.
3. To quantify the resulting amino acid labelling patterns and generate *in vivo* metabolic flux distributions of *P. thermoglucosidasius* NCIMB 11955 central carbon metabolism under the above growth conditions.

The research towards these objectives in described in Chapter 4.

2. Chapter 2: Materials and Methods

2.1 Materials

2.1.1 Bacterial strain and culture medias for storage, maintenance and bioreactor experiments

Organism used in this research

Table 1: The bacterial strain utilised in this research.

Bacterial species and strain	Commentary	Source
Parageobacillus thermoglucosidasius NCIMB 11955	Glycerol stocks maintained at University of Bath and IIT Mandi. Despite minor genome sequence differences, this strain is the same as DSM2542.	National Collection of Industrial & Marine Bacteria (Aberdeen, UK)

Glycerol stocks

Stocks of *P. thermoglucosidasius* NCIMB 11955 for long term storage were maintained as glycerol stocks. *P. thermoglucosidasius* cells were first grown in a New Brunswick Scientific Co. Innova 44 Stackable Incubator Shaker for 12-14 hours in 50 mL conical-bottom centrifuge tubes containing 10 mL 2TY medium at 60°C, 210 rotations per minute (RPM) shaking. 500 µL of this culture was combined with 500 µL of 60% w/v glycerol solution in a cryogenic vial (Thermo Scientific, Nalgene, Loughborough, UK). These mixed solutions were then then flash-frozen in liquid nitrogen and immediately stored as glycerol stocks in a -80°C freezer until required.

Media for Maintenance and Routine Growth

For short term maintenance, *P. thermoglucosidasius* NCIMB 11955 cells were grown overnight at 55 ^{o}C on Tryptone Soya Broth Agar (TSBA) (17 g/L Tryptone, 3 g/L Soytone, 5 g/L NaCl, 2.5 g/L K_2HPO_4, 15 g/L Agar) and for routine growth in liquid 2x Tryptone (2TY) (16 g/L Tryptone, 10 g/L Yeast extract, 5 g/L NaCl, pH 7) medium.

Ammonium Salts Medium (ASM) – Minimal defined media for bioreactor growth experiments

For all bioreactor experiments, *P. thermoglucosidasius* NCIMB 11955 cells were initially grown in filter-sterilized fully defined Ammonium Salts Media (ASM) (Table 2) containing 0.3% w/v naturally labelled ^{D}-(+)-glucose (99.5% pure, Merck, UK) or 0.3% w/v naturally labelled ^{D}-(+)-xylose (≥99% pure, Merck, UK) and 5 mL/L sulphate trace elements.

For the instationary ^{13}C-MFA glucose experiments, 0.3% w/v (60% 1-^{13}C and 40% U-^{13}C) ^{D}-(+)-glucose (Cambridge Isotope Laboratories, USA) was the carbon source. The chemical purity of both ^{13}C glucose isotopes was >99 % and ^{13}C-enrichment was >98%.
For the instationary ^{13}C-MFA xylose experiments, 0.3% w/v (75% 1-^{13}C and 25% naturally labelled) ^{D}-(+)-xylose (Cambridge Isotope Laboratories, USA and Merck, UK, respectively) was the carbon source. The chemical purity of the 1-^{13}C xylose was 98 % and ^{13}C-enrichment was 99%.
When used as a pre-inoculation sub-culture media, 500 mL batches of ASM were corrected to a pH of 7 before filter-sterilization. When being prepared as a reservoir for continuous culture experiments, the ASM was adjusted to a pH of 4.15 before filter-sterilization to reduce the growth of contaminants once the reservoirs were connected to the bioreactor system. Foaming was supressed for all ^{13}C-MFA experiments by inclusion of 0.0125% Antifoam 204 (Sigma-Aldrich) in the ASM reservoir.

Table 2: Composition of the minimal, defined ammonium salts media (ASM).

Chemical	Final Concentration	Concentration (gL^{-1})
$NaH_2PO_4.2H_2O$	20 mM	3.12
K_2SO_4	10 mM	1.74
Citric Acid.H_2O	8 mM	1.68
$MgSO_4.7H_2O$	5 mM	1.23
$CaCl_2$	0.08 mM	0.09
$Na_2MoO_4.2H_2O$	1.65 µM	0.0004
$(NH_4)_2SO_4$	25 mM	3.3
Glucose/xylose	0.3%	3
Biotin	12 µM	0.00292
Thiamine	12 µM	0.0036
Sulphate Trace Elements Solution*	See Below	5 mL

Table 3: Composition of Sulphate Trace Elements stock solution used in ASM.

Chemical	Final Concentration	Volume (mL)
Conc. H_2SO_4		5 mL
$ZnSO_4.7H_2O$	25 µM	1.44
$FeSO_4.7H_2O$	100 µM	5.56
$MnSO_4.H_2O$	50 µM	1.69
$CuSO_4.5H_2O$	5 µM	0.25
$CoSO_4.7H_2O$	10 µM	0.562

$NiSO_4.6H_2O$	16.85 µM	0.886
H_3BO_3		0.08
dH_2O		1000 mL

2.1.2 Bench-top bioreactors and microbial fermentation control systems

To achieve the experimental steady-state conditions, it was necessary to have Continuous Stirred Tank Reactor (CSTR) bioreactor systems which were able to operate as chemostat cultures (continuous cultures in which input rates of media and output flow rates of culture are equal).

2.1.2.1 The Custom MKI and MKII Bioreactor systems for aerobic INST-^{13}C-MFA experiments

On account of the high isotope sugar costs, all aerobic pilot experiments were performed in one of two miniature 'micro-bioreactors' of either a 45 mL (MKI) or 90 mL (MKII) working volume vessels. Based on the oxygen transfer issues discussed in chapter 4.3.3 all ^{13}C MFA experiments were performed with the 90 mL working volume MK II system.

The shared control unit and feed line setup:

The bioreactor control system for both the MKI and MKII reactors was comprised of a modified Biostat B Fermentation Cell Culture Bioreactor Control Tower (B. Braun Biotech Industries (now Sartorius), De) (Figure. 5) (hereafter referred to as the 'Braun unit'), which maintained the target temperature, pH, and air flow rate process variables of each experiment through inbuilt proportional integral derivative (PID) controllers, and a custom peristaltic-pump feed line set up for controlling media flow rates.

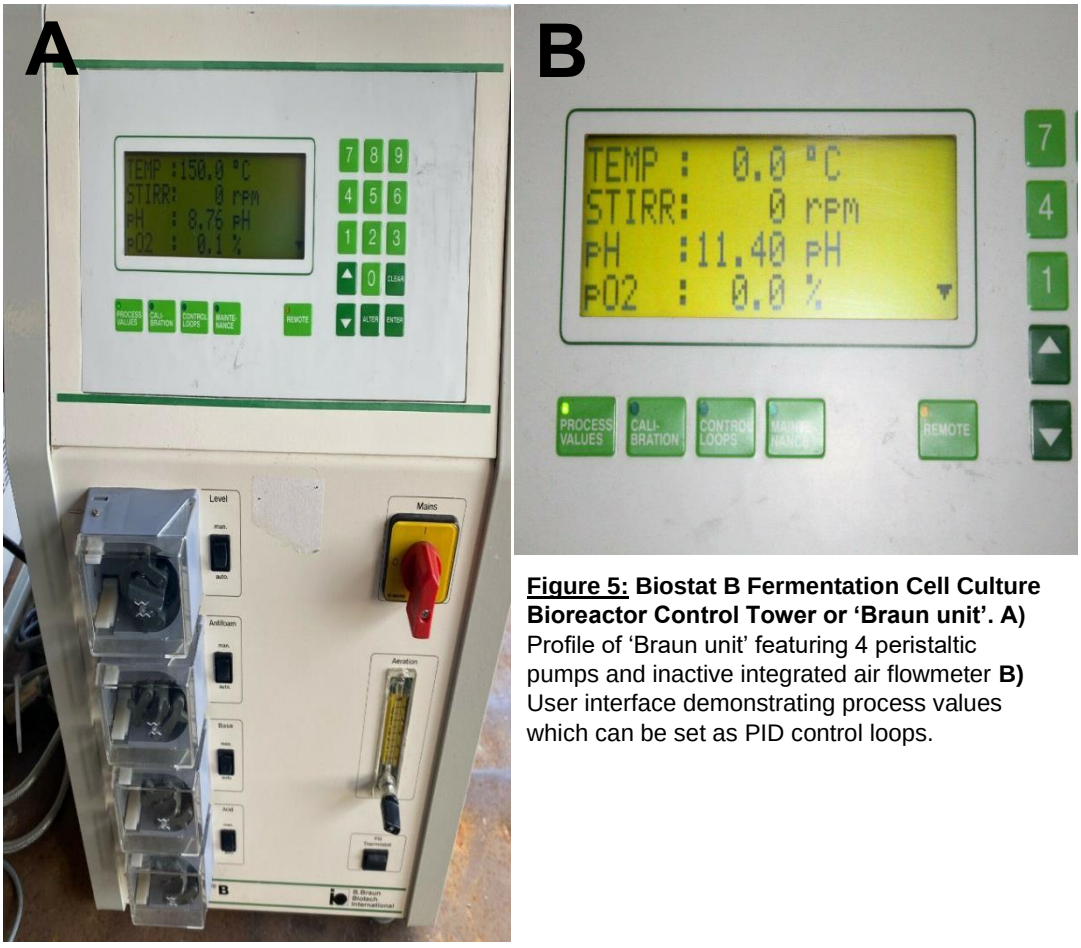


Figure 5: Biostat B Fermentation Cell Culture Bioreactor Control Tower or 'Braun unit'. A) Profile of 'Braun unit' featuring 4 peristaltic pumps and inactive integrated air flowmeter **B)** User interface demonstrating process values which can be set as PID control loops.

Temperature control

The Braun unit incorporated a Pt-100 temperature probe which fitted a thermowell in the MKI and MKII units. Maintenance of the target 60°C temperature was achieved by heating and cooling as required an internal reservoir under PID control which supplied a continuous flow (and return) of water around the water-jackets of the MKI and MKII vessels via silicon tubing.

During these experiments, the heating element of the Braun unit's temperature control suffered two separate over-runs, requiring repair. After this point, *a* Julabo FE-500 Refrigerated Circulator was sourced to pre-cool the water that entered the control unit's reservoir for temperature control.

Additionally, both the MKI and MKII vessels were fitted with a custom 10 cm reflux condenser (Chemglass Life Sciences, USA) fitted with a 0.2 µm MiniSart filter (Sartorius Stedim, Ger) to minimise evaporation and any loss of volatile compounds from the culture as well as equalise air pressure within the sealed vessel. The flow of chilled water around the condenser system was maintained from an ice-chilled water reservoir by repurposing the antifoam control pump on the Braun unit to operate in manual pumping mode.

pH control

The pH of the culture was measured using the pH probe specific to either the MKI or MKII system immersed in the culture medium. The target pH 7 (deadband ± 0.05) condition of the culture was maintained using the Braun unit's integrated pH controller and pump system. To maintain this pH, the control unit automatically pumped either 0.9M KOH or 0.2M HCl to the culture from separate 1L reservoirs via 2 mm (id) silicon tubing connected to separate 15.2 cm stainless steel needles (Thermo Scientific, UK) which were pierced through septa in entry ports on the glass top plates of the MKI and MKII vessels.

For each experiment, the PID pH controller was calibrated through the MKI- or MKII-specific pH probe (detailed below) and using known pH standards.

Air flow control

As the in-built air flow meter of the Braun unit was non-functional from the start of the research, the flow rate of compressed air was controlled through insertion of one of two air flow meters (10-150 mL/min and 0.1-1.3 L/min, Omega, UK) into the outlet gas line of the Braun unit depending on the aeration requirements (Figure. 6B). The compressed air was then pumped through silicon tubing via an in-line 0.2 µm MiniSart filter (Sartorius Stedim, Ger) and introduced to the culture at the bottom of the MKI and MKII vessels through a ring sparger (Figure. 6A). Air and liquid leakage detection was performed for all connections at the before inoculation using Snoop liquid leak detector (Swagelok, USA).

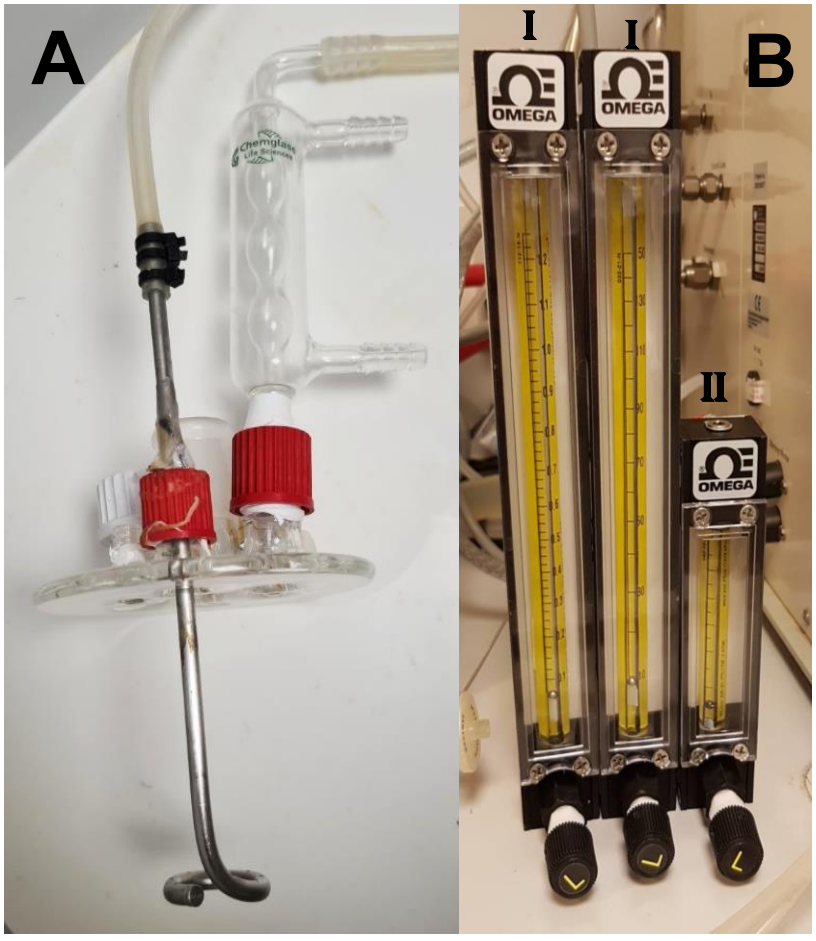


Figure 6: The complete aeration control set up. A) Top plate of MKII vessel featuring ring sparger. **B)** Auxiliary flowmeters (OMEGA, UK) for control of air flow rates.

I) Maximum air flow rate: 1.3 L/min.

II) Maximum air flow rate: 150 mL/min.

III) Maximum air flow rate: 8 mL/min.

High-Performance Liquid Chromatography (HPLC) and GC-MS data from the 1-^{13}C xylose-fed anaerobic growth of *P. thermoglucosidasius* NCIMB 11955 was generated in an older iteration of the MKI which utilized a 101U peristaltic pump (Watson-Marlow Falmouth, UK) to control air flow rate rather than the air flow meters described above.

Control of media feed rates

ASM was delivered into the vessels from the reservoirs via a 15.2 cm stainless steel needle (Thermo Scientific, UK) pieced through a septum in a port on the top plate and connected to 3.2 mm (id) silicon tubing. This system incorporated an anti-grow back tube to prevent cells growing back into the media reservoir. Continuous culture was achieved by incorporating two additional peristaltic pumps (101U Watson Marlow, Falmouth, Cornwall) on the inflow and outflow tubing and the incorporation of a 3-way water connector enabling a sterile switch from continuous pumping of ^{12}C glucose ASM to ^{13}C labelled ASM without affecting flow rate (Figure. 7).

Waste culture was removed from the vessel via a 10 mm (id) silicon tube connected to the overflow port and pumped to discourage outlet biofilm formation and build-up.

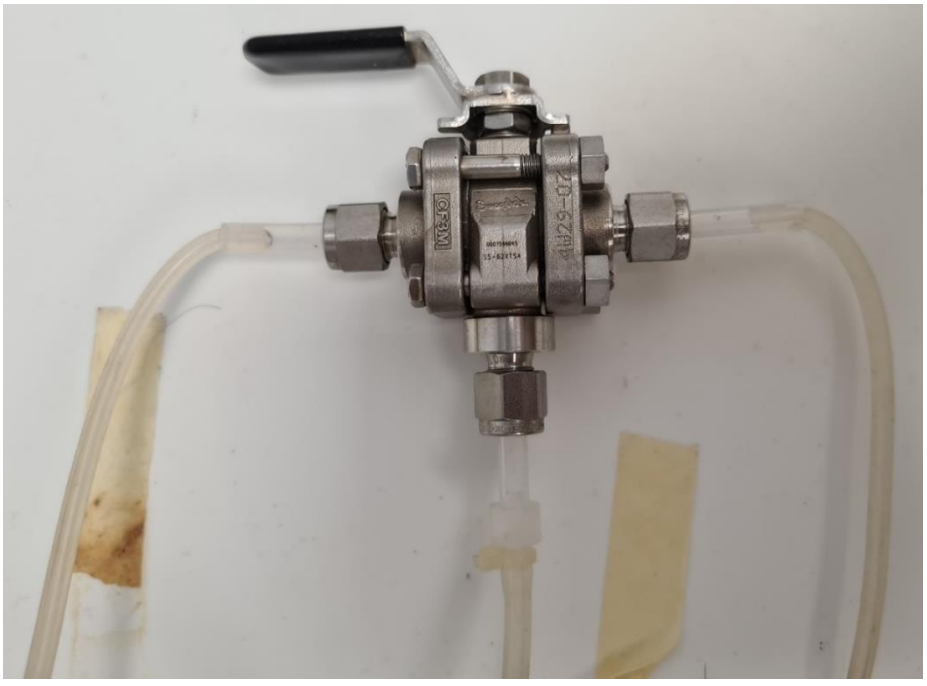


Figure 7: Three-way connector enabling a sterile switch from continuous pumping of ^{12}C glucose ASM to ^{13}C-labelled ASM.

Measurement of redox potential

The continuous measurement of the redox potential of the cultures was used an important indicator of fermentation status of *P. thermoglucosidasius* cultures. This was done using one of two ordinary redox potential pH / Redox (ORP)

probes linked to an HPT63 display console (LTH Electronics Ltd. UK). For each experiment, the system was calibrated using +220mV and +468mV standards.

The MKI vessel

The MKI 45 mL working volume bioreactor was the first to be constructed and evaluated (Figure 8). The original custom glass water-jacketed vessel and triple port top plate were designed by Jeremy Bartosiak-Jentys and manufactured by Soham Scientific (Soham, UK) based on a modification of their simulated gut reactors developed with Reading University[115].

Culture mixing in the MKI was provided by an 8 mm micro magnetic stirring bar and analogue magnetic stirrer plate (Thermo Scientific, Loughborough, UK). The measurement of culture pH and redox potential was performed for all experiments using a 120 x 12 mm autoclavable F-995 redox electrode (Broadley-James, Silsoe, UK) and an autoclavable 120 x 12 mm F-995 pH probe (Broadley-James, Silsoe, UK).

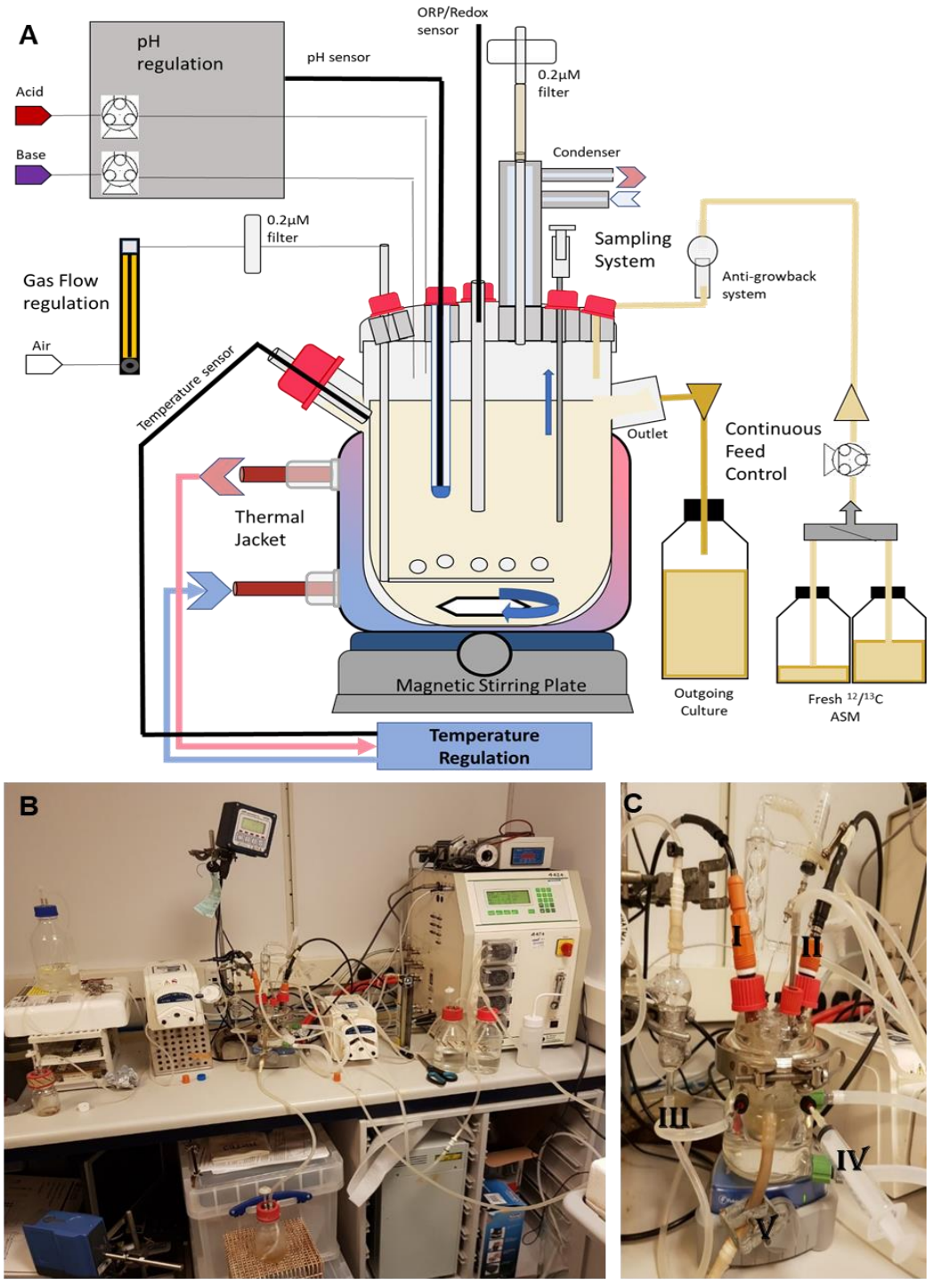


Figure 8: Set up of the 'MKI' bioreactor 45 mL culture system. **A)** Schematic of the MKI system detailing how: agitation, continuous feeding, pH, gas flow, pressure and temperature regulation were maintained when growing *P. thermoglucosidasius* cultures. Stirring controlled by magnetic flea and magnetic stirrer plate. Sampling performed with syringe through port. glass 'anti-grow back' system prevents contamination of reservoir ASM. **B)** Image of complete set up of MKI system. **C)** *P. thermoglucosidasius* cultures growing as an oxygen rich chemostat culture in the MKI microbioreactor. **I)** F-995 redox electrode (Broadley-James, Silsoe, UK). **II)** F-995 pH probe (Broadley-James, Silsoe, UK). **III)** Media inlet featuring glass 'anti-grow back' system **IV)** Sampling syringe and port **V)** Analogue magnetic stirrer plate (Thermo Scientific, Loughborough, UK).

The MKII vessel

The MKII 90 mL working volume bioreactor was the second reactor to be constructed and evaluated and was used for all aerobic ^{13}C-MFA experiments. This custom glass water-jacketed vessel was designed by Prof. David Leak based on the results of Jeremy Bartosiak-Jentys and manufactured by Soham Scientific (Soham, UK).

In addition to a having a greater working volume, the MKII system incorporated several improvements. The MKII vessel incorporated a 26° angled side inlet port to function as a thermowell for the temperature probe. The MKII top plate design was flatter instead of bevelled to increase the number of ports (Figure 9) and incorporated a port for a programmable motor-driven flat blade agitator, used for every experiment to support more reliable mixing at greater RPM values and to improve oxygen mass transfer to the culture media.

The pH of the culture was measured by an EasyFerm Plus VP pH/RX 225 pH probe (Hamilton, UK) and the culture redox potential was measured using a EasyFerm Plus PHI K8 225 S8 redox probe (Hamilton, UK).

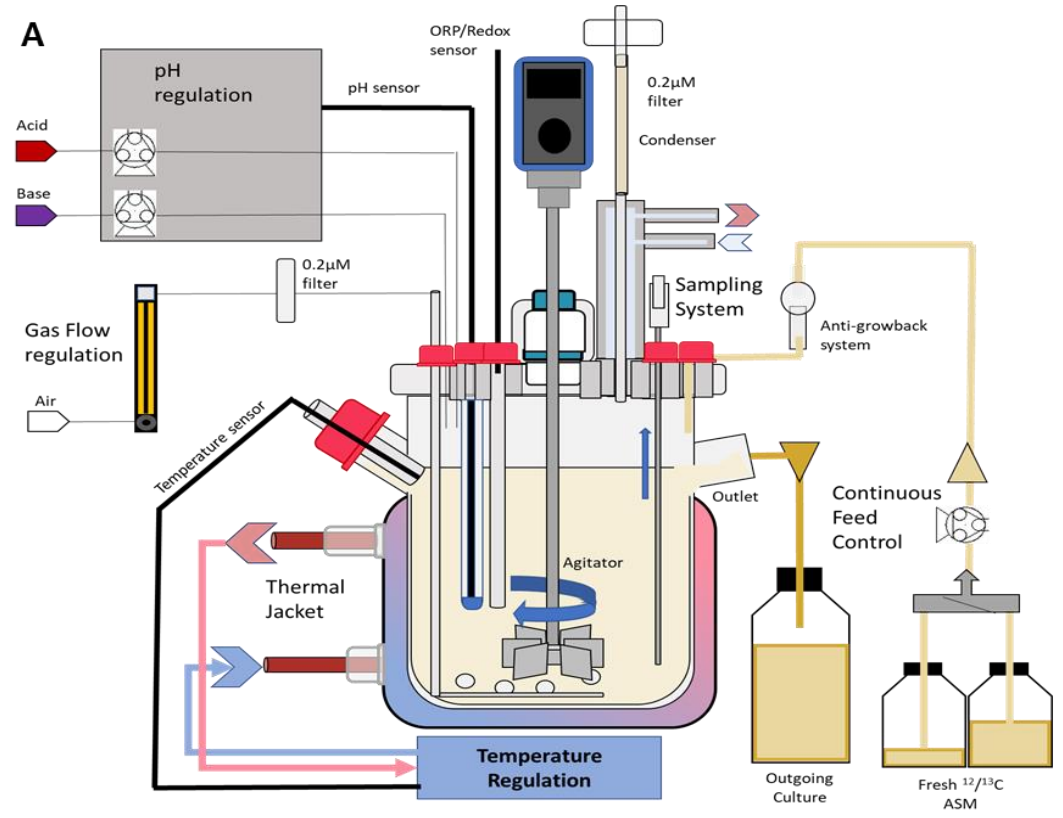


Figure 9: Set up of the 'MKII' bioreactor 90 mL culture system. **A)** Schematic of the MKII system detailing how: agitation, continuous feeding, pH, gas flow, pressure and temperature regulation were maintained when growing *P. thermoglucosidasius* cultures. Gas flow rate controlled by external flowmeter. Stirring controlled by a programmable rotor. Sampling performed with syringe through port. glass 'anti-grow back' system prevents contamination of reservoir ASM. **B)** Complete set up of the MKII system. At the time of this image, the chiller unit had not yet been installed. **C)** *P. thermoglucosidasius* cells growing as an aerobic chemostat culture in the MKII microbioreactor. **I)** Digital programmable agitator. **II)** Sampling syringe. **III)** Media inlet featuring glass 'anti-grow back' attachment. **IV)** EasyFerm Plus PHI K8 225 S8 redox probe (Hamilton, UK). **V)** EasyFerm Plus VP pH/RX 225 pH probe (Hamilton, UK).

2.1.2.2 Applikon® MiniBio2 250 system

At the Indian Institute of Technology Mandi, India, an Applikon® Biotechnology MiniBio2 250 bioreactor system (Figure 10, Applikon® Biotechnology) was used for the first aerobic glucose 0.075 h^{-1} growth rate ^{13}C-MFA experiment. The complete system consisted of a 50-250 mL working volume borosilicate glass vessel with a detachable stainless steel top plate bearing: an L-type sparger for aeration, thermowell for temperature sensor, motor-driven Rushton impeller, conductivity-based level sensor and ports for: inoculation, sampling, culture transport and two ports for pH, redox potential or dissolved oxygen probes. The measurement of culture pH was performed for all experiments using an incorporated pH probe and the measurement of dissolved oxygen (DO_2) using an AppliSens DO_2-sensor. Additionally, each system incorporated a computer-linked control unit which was used to control and maintain: a temperature of 60°C, a pH of 7, the media flow rates of the inlet and, the agitation speed of the impeller and the inlet air flow rate. Air and liquid leakage detection was performed for all connections at the before inoculation using Snoop liquid leak detector (Swagelok, USA).

Real-time monitoring and data-logging of: redox potential, temperature, air flow, pH and agitator speed were perfored for all growth experiments through the linked computer using the Lucullus Process Information Management System version 3.5.2 (Applikon® Biotechnology).

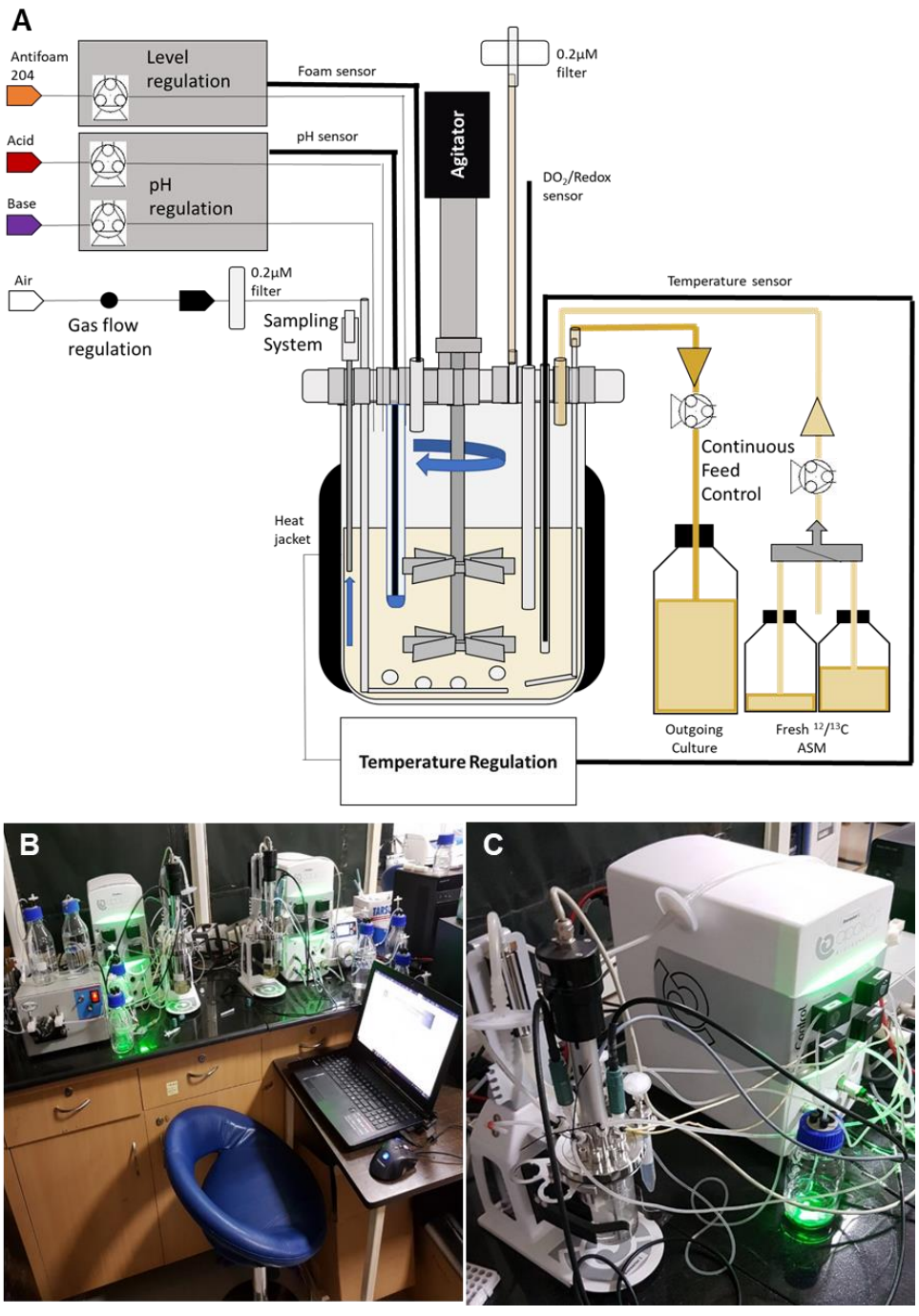


Figure 10: Set up of complete Applikon® Biotechnology MiniBio2 250 culture system. **A)** Schematic of the MiniBio2 250 system detailing how: agitation, continuous feeding, pH, foam level, gas flow, pressure and temperature regulation were maintained when growing *P. thermoglucosidasius* cultures. Temperature regulation provided by heating jacket. Stirring controlled by a programmable rotor. Sampling performed with syringe through port. ASM. **B)** Complete set up of MiniBio2 250 mL system and workstation. **C)** Close up of MiniBio2 250 system and associated control unit.

2.1.2.3 Biostat® B system for anaerobic and 'micro-aerobic' ^{13}C-MFA experiments

All anaerobic and 'micro-aerobic' pilots and INST-^{13}C-MFA experiments were performed using a Biostat® B benchtop bioreactor system (Sartorius, DE.). This consisted of the Biostat® B control tower for overview and control of process parameters and a stirred tank bioreactor (Figure 11A).

Through in-built PID controllers, the Biostat® B control tower was able to maintain pH and redox potential using a 200 mm EasyFerm Plus VP8 Pt-1000 dual pH/ORP sensor (Hamilton, CH), temperature using an incorporated 151mm Pt-1000 temperature probe and oxygen saturation levels using a 215mm OxyFerm FDA VP8 225 DO_2sensor (Hamilton, CH). The Biostat® B system also utilised inbuilt mass flow controllers to control the rates compressed air and compressed nitrogen to aid in the creation of low oxygen and low redox potential environments for anaerobic cell growth.

The stirred-tank bioreactor consisted of a 2 L water-jacketed boro-silicate glass vessel used at its minimum working volume of 350 mL with a stainless-steel top-plate. The top-plate had ports for sampling, media inlet and outlet, a stainless-steel condenser to minimize evaporation and loss of volatile products, a thermowell for the temperature probe, a ring sparger to introduce gas to the culture underneath the impeller and a motor-driven 6-blade dual disc Rushton impeller for culture agitation and mixing controlled by the Biostat® B control tower.

The complete vessel and top plate were placed on an integrated weighing scale (Sartorius, DE) connected to the Biostat® B control tower. For the ^{13}C MFA continuous culture experiments, the system was adapted to feature a 101U peristaltic pump (Watson-Marlow Pumps Group, Falmouth, Cornwall) to introduce a fixed flow rate of ASM through 3.2 mm internal diameter silicon tubing. On account of recurrent vessel pressure issues and condenser failures with the Biostat B bioreactor system, a tygon® tubing outlet line was fixed to the top of the non-functional condenser to a dreschel gas wash bottle filled with 100 mL water, submersed in a flask of ice. Under the assumption that any evaporated products would no longer condense back into the culture, a single 500 µL sample of the dreschel bottle was taken alongside each time point of the glucose anaerobic 0.075 h^{-1} ^{13}C-MFA experiment.

As with the smaller reactors, two media reservoirs were connected via a 3-way connector, enabling a sterile switch from continuous pumping of naturally labelled ASM to ^{13}C labelled ASM without affecting flow rate. Flow of culture out of the bioreactor was controlled gravimetrically by the Biostat® B tower by setting a fixed weight for the vessel and top plate determined at 350 mL working volume with all lines in and out of the bioreactor at full capacity. Thus, the dilution rate of a continuous experiment was entirely controlled via the input flow rate of media. Sampling was performed through a 3-port vacuum sampling system (Figure 11.B) which prevented contamination from sampling by uni-directional flow.

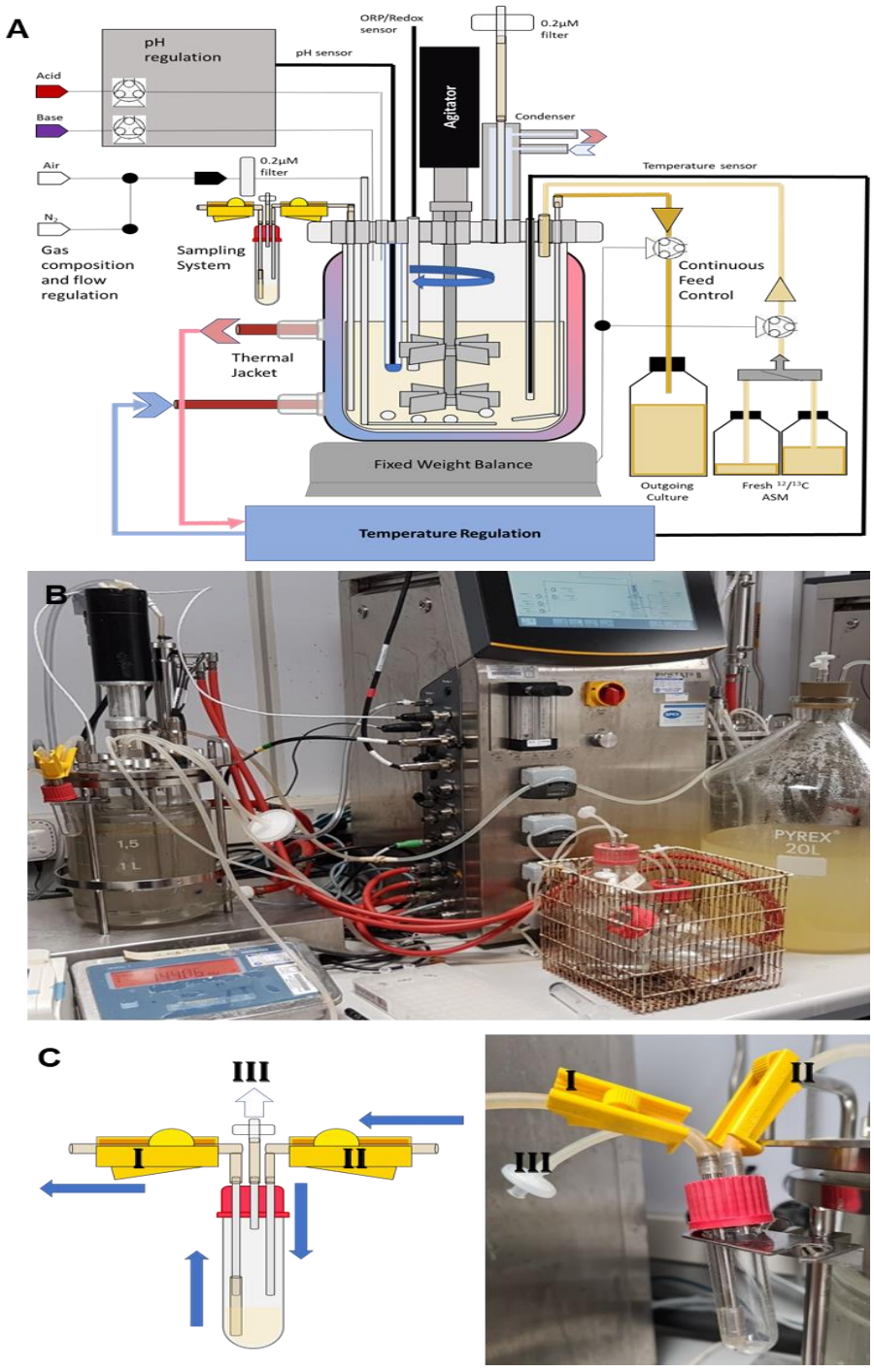


Figure 11: Set up Biostat® B bioreactor 1.5 L culture system. **A)** Schematic of the Biostat® B bioreactor system detailing how: agitation, continuous feeding, pH, gas flow, pressure and temperature regulation were maintained when growing *P. thermoglucosidasius* cultures. Stirring controlled by a programmable rotor. Continuous flow maintained by fixed input flow rate and weight balance set to maintain a target weigh by removal of culture. Sampling performed with sterile pressure sampling system (C). **B)** *P. thermoglucosidasius* cells growing under oxygen limited conditions as a chemostat culture in the Biostat® B bioreactor. **C)** Biostat® B bioreactor sterile pressure sampling system. Roller clamp **II** opened, and pressure differential created via a syringe applied to filter **III** which draws culture into the intermediary vial. Roller clamp **II** is then closed, and roller clamp **I** is opened, allowing culture samples to be drawn via syringe from the intermediary vial.

2.1.3 Computational Materials for INST-^{13}C-MFA

2.1.3.1 IsoCor v2.2

In order to obtain an accurate understanding of the metabolite mass isotopomer distributions and labelling ratios as a result of the introduced ^{13}C-labelled carbon, it was necessary to remove the influence of other anticipated naturally abundant isotopes i.e. ^{2}H, ^{13}C, ^{15}N, $^{17\text{-}18}$O, $^{29\text{-}30}$Si, and $^{33\text{-}36}$S) on the obtained ^{13}C MS data[116].

Natural isotope correction was performed using the IsoCor v2.2 tool[117] (Figure 12) using the manually generated metabolite and derivative tables (Tables 4 and 5) and yielding the corrected mass isotopomer distributions for each amino acid fragment and their relative abundances.

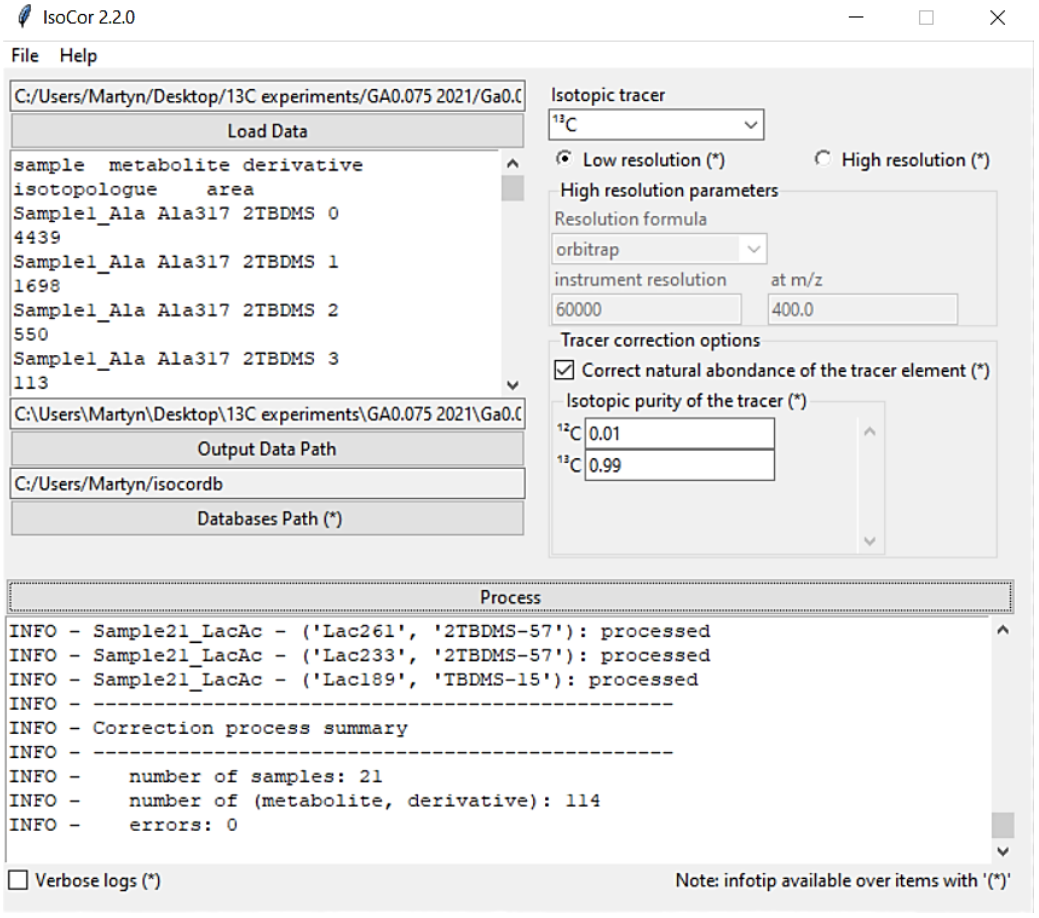


Figure 12: Graphical User-Interface of IsoCor v.2.2.0. Data is loaded as individual note files of each sample featuring the: sample, metabolite fragment formula, chemical derivative associated with each metabolite fragment and the uncorrected peak area and processed with respect to user defined metabolites, derivatives and isotopic tracer purity.

Tables 4 and 5: Curated libraries featuring the names and chemical formula of the Derivatives **4)** and Metabolites **5)** defined for use in isotope correction with IsoCor 2.2.

Name	Formula
TMS	C_3H9Si
TBDMS	$C_6H_{15}Si$
TBDMS-15	$C_5H_{12}Si$
TBDMS-45	C_3H_6Si
TBDMS-57	C_2H_6Si
2TBDMS	$C_{12}H_{30}Si_2$
2TBDMS-15	$C_{11}H_{27}Si_2$
2TBDMS-45	$C_9H_{21}Si_2$
2TBDMS-57	$C_8H_{21}Si_2$
3TBDMS	$C_{18}H_{45}Si_3$
3TBDMS-15	$C_{17}H_{42}Si_3$
3TBDMS-45	$C_{15}H_{36}Si_3$
3TBDMS-57	$C_{14}H_{36}Si_3$
4TBDMS	$C_{24}H_{60}Si_4$
4TBDMS-15	$C_{23}H_{57}Si_4$
4TBDMS-45	$C_{21}H_{51}Si_4$
4TBDMS-57	$C_{20}H_{51}Si_4$

Name	Formula
Ala317	$C_3H_5O_2N$
Ala302	$C_3H_5O_2N$
Ala260	$C_3H_5O_2N$
Ala232	C_2H_5ON
Ala158	C_2H_5N
Gly303	$C_2H_3NO_2$
Gly288	$C_2H_3NO_2$
Gly246	$C_2H_3NO_2$
Gly218	C_1H_2NO
Gly144	C_1H_3N
Val345	$C_5H_9NO_2$
Val330	$C_5H_9NO_2$
Val288	$C_5H_9NO_2$
Val260	C_4H_9NO
Val186	C_3H_9N
Leu359	$C_6H_{11}NO_2$
Leu344	$C_6H_{11}NO_2$

Name	Formula
Leu302	$C_6H_{11}NO_2$
Leu274	$C_5H_{11}NO_2$
Leu200	$C_5H_{11}NO$
ILe359	$C_6H_{11}NO_2$
ILe344	$C_6H_{11}NO_2$
ILe302	$C_6H_{11}NO_2$
ILe274	$C_5H_{11}NO2$
ILe200	$C_5H_{11}NO$
Pro343	$C_5H_5NO_2$
Pro328	$C_5H_5NO_2$
Pro301	$C_5H_5NO_2$
Pro286	$C_5H_5NO_2$
Pro258	$C_4H_5NO_2$
Pro184	C_4H_5NO
Met377	$C_5H_9NO_2S$
Met302	C_4H_9NOS
Met292	C_4H_9NOS

Name	Formula
Met218	C_4H_9NS
Ser447	$C_3H_4NO_3$
Ser432	$C_3H_4NO_3$
Ser390	$C_3H_4NO_3$
Ser362	$C_2H_4NO_2$
Ser288	C_2H_4NO
Thr461	$C_4H_6NO_3$
Thr446	$C_4H_6NO_3$
Thr404	$C_4H_6NO_3$
Thr376	$C_3H_6NO_2$
Thr302	C_4H_6NO
Phe393	$C_9H_9NO_2$
Phe378	$C_9H_9NO_2$
Phe336	$C_9H_9NO_2$
Phe308	C_9H_9NO
Phe234	C_9H_9N
Asn474	$C_4H_5N_2O_3$

Name	Formula
Asn459	$C_4H_5N_2O_3$
Asn417	$C_4H_5N_2O_3$
Asn302	$C_4H_5N_2O_3$
Asn389	$C_3H_5N_2O_3$
Asn315	$C_3H_5N_2O$
Arg516	$C_6H_{11}N_4O_2$
Arg501	$C_6H_{11}N_4O_2$
Arg459	$C_6H_{11}N_4O_2$
Arg402	$C_6H_{12}N_4O_2$
Arg357	$C_5H_{11}N_4$
Arg317	$C_5H_{12}N_4O$
Arg300	$C_5H_{11}N_4$
Arg288	$C_6H_{13}N_4O_2$
Arg273	$C_6H_{13}N_4O_2$
Arg231	$C_6H_{13}N_4O_2$
Gln488	$C_5H_7N_2O_3$
Gln473	$C_5H_7N_2O_3$

Name	Formula
Gln431	$C_5H_7N_2O_3$
Gln403	$C_5H_7N_2O_2$
Gln358	$C_5H_7N_2O_3$
Gln329	$C_5H_7N_2O$
Gln300	$C_5H_7N_2O_2$
Gln272	$C_5H_7N_2O$
Glu489	$C_5H_9NO_4$
Glu474	$C_5H_9NO_4$
Glu432	$C_5H_9NO_4$
Glu404	$C_5H_9NO_3$
Glu330	$C_5H_9NO_2$
Glu302	$C_5H_8NO_3$
Glu187	$C_5H_8NO_3$
His497	$C_6H_6N_3O_2$
His482	$C_6H_6N_3O_2$
His440	$C_6H_6N_3O_2$
His412	$C_5H_6N_3O$

Name	Formula
His338	$C_5H_7N_3$
His196	$C_5H_7N_3O$
Lys488	$C_6H_{11}N_2O_2$
Lys473	$C_6H_{11}N_2O_2$
Lys431	$C_6H_{11}N_2O_2$
Lys403	$C_5H_{11}N_2O$
Lys329	$C_6H_{11}N_2$
Lys300	$C_5H_{11}N_2O$
Cys406	$C_3H_4NO_2S$
Cys378	C_2H_4NOS
Cys304	C_2H_4NS
Tyr523	$C_9H_8NO_3$
Tyr508	$C_9H_8NO_3$
Tyr466	$C_9H_8NO_3$
Tyr438	$C_8H_8NO_3$
Tyr364	$C_8H_8NO_2$
Tyr221	$C_8H_9NO_3$

Name	Formula
Trp546	$C_{11}H_9N_2O_2$
Trp531	$C_{11}H_9N_2O_2$
Trp489	$C_{11}H_9N_2O_2$
Trp461	$C_{10}H_9N_2O$
Trp388	$C_{10}H_9N_2O$
Trp330	$C_{10}H_9N_2$
Trp302	$C_3H_2O_2N$
Trp244	$C_{10}H_9NO$
Asp475	$C_4H_4NO_4$
Asp460	$C_4H_4NO_4$
Asp418	$C_4H_4NO_4$
Asp390	$C_3H_4NO_3$
Asp316	$C_3H_4NO_3$
Asp302	$C_3H_5NO_3$
Asp173	$C_3H_4NO_3$
Cit634	C_4O_7
Cit591	C_4O_7

Name	Formula
Cit459	C_3O_6
Cit431	C_3O_5
Cit357	C_3O_5
Cit346	$C_4H_4O_4$
Cit331	$C_4H_4O_4$
Cit289	$C_4H_4O_4$
Cit147	$C_3H_5O_3$
Lac318	$C_3H_4O_3$
Lac303	$C_3H_4O_3$
Lac261	$C_3H_4O_3$
Lac233	$C_2H_4O_2$
Lac189	$C_3H_5O_3$
Suc	$C_4H_4O_3$
Fum	$C_4H_3O_4$
Mal	$C_4H_5O_5$
Cit	$C_6H_7O_7$
aKG	$C_5H_5O_5$

Name	Formula
OA	$C_4H_3O_5$
Aco	C_6HO_3
2/3PG	$C_3H_6O_3$
G3P	$C_3H_7O_3$
PEP	$C_3H_4O_3$
E4P	$C_4H_8O_4$
R5P	$C_5H_8O_4$
G6P	$C_6H_{10}O_5$
F1P	$C_6H_{10}O_5$
F6P	$C_6H_{10}O_5$
M6P	$C_6H_{10}O_5$
6PG	$C_6H_{10}O_6$
S7P	$C_7H_{11}O_6$
FBP	$C_6H_{11}O_8P$
AMP	$C_{10}H_{14}N_5O_4$
ADP	$C_{10}H_{14}N_5O_4$
ATP	$C_{10}H_{16}N_5O_{10}$

2.1.3.2 Isotopomer Network Compartmental Analysis (INCA)

Due to its ability to perform isotopically instationary metabolic flux analysis, the metabolic network model definition and all ^{13}C informed metabolic flux estimation of *P. thermoglucosidasius* was performed using INCA v.1.7 software package[118] in MATLAB R2017b (The MathWorks Inc, USA).

2.1.3.3 The COBRA toolbox v3.0

All flux balance, parsimonius flux balance and flux variability analysis was performed using the COBRA Toolbox v3.0 software suite[119] in either MATLAB R2017b (The MathWorks Inc, USA) or Python 3.0.

2.1.3.4 Memote

Throughout the genome-scale metabolic model refinement and improvement process, the quality of the genome scale model was evaluated with the MEMOTE tool[120] and used it to identify areas of further model refinement.

2.2 Methods

2.2.1 Instationary ^{13}C-MFA methods

2.2.1.1 Experimental Design

^{13}C isotopic labelling strategy

The ^{13}C isotope labelling strategy for the ^{13}C-MFA experiments in which glucose was the sole substrate at 0.3% w/v contained 60% 1-^{13}C and 40% U-^{13}C ^{D}-(+)-glucose. This was selected based on Dr. Charlotte Ward's observation that successful and improved resolution of some *P. thermoglucosidasius* metabolic pathways could be achieved with this mix of labelled sugars[121].

The ^{13}C isotope labelling strategy for the ^{13}C-MFA experiments in which xylose was the sole substrate at 0.3% w/v contained 75% 1-^{13}C and 25% natural labelled ^{D}-(+)-xylose, chosen in order to directly incorporate an anaerobic xylose ^{13}C-MFA experiment performed by Shyam Maskapalli in 2014 into the overall analysis which used 0.5% w/v (75% 1-^{13}C and 25% natural labelled) ^{D}-(+)-xylose as the sole substrate.

Preparation of ASM for ^{13}C-MFA experiments

To improve the accuracy of the dynamic isotope incorporation trends obtained, it was important to ensure any metabolic variation observed between the ^{13}C labelled sugar and unlabelled sugar was only down to the carbon source and not batch-to-batch variation in ASM preparation. Therefore, for every instationary ^{13}C-MFA experiment, all reservoir ASM was first prepared as a single batch and pH adjusted to 4.15 before being split, the requisite sugar added and final ASM filter-sterilized (Figure 13):

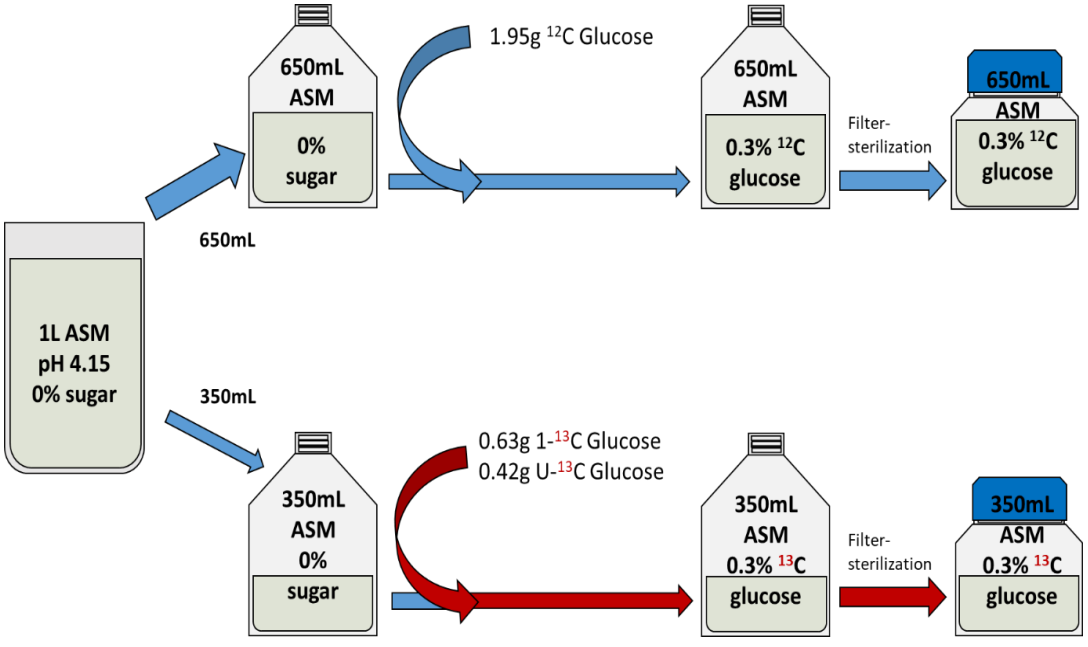


Figure 13: Depiction of the ASM preparation strategy employed to minimize batch-to-batch variation for all ^{13}C-MFA experiments, featuring the weights and volumes of an aerobic glucose experiment.

2.2.1.2 Micro-bioreactor growth experiments

Inlet feed-line flow rate calibration

A fundamental assumption of metabolic flux analysis experiments in chemostat systems is that, once the cells reach a metabolic steady state, the growth rate of the cells in the culture is equal to the dilution rate of the chemostat[122]. Therefore, for an accurate observation of cellular metabolic behavior at a target growth rate within the chemostat system, it is necessary to precisely calibrate the flow rate of media added through the feed line. In order to ensure this precise sub-100 µL/min media flow rates, peristaltic pump rotation speeds (RPM) versus media flow rate standard curves were routinely generated for all media inlet lines for the two potential media reservoirs prior to each ^{13}C-MFA experiment. For each rotation speed on the 101U peristaltic pump (Watson-Marlow Pumps Group, Falmouth, Cornwall) a flow rate was determined by weighing the amount of water that flowed into a pre-weighed measuring cylinder over a fixed time (3 minutes). Using the density of water this was converted to a volumetric flow rate (mL/min) associated with each speed setting of the pump. Lastly, dilution rate was calculated by dividing the flow rate by the working volume of the bioreactor in millilitres. It was assumed that the viscosities and densities of 0.3% glucose or xylose ASM at room temperature and pressure were equivalent to that of water.

Bioreactor assembly and sterilization

To ensure effective sterilisation by autoclaving, every bioreactor was first assembled as a complete glass vessel with: 10% total working volume of de-ionised water (DI-H_2O), integrated autoclavable pH and redox potential probes and tube feed lines connected to empty reservoir bottles. While still an open system with respect to ensuring a balance of pressure in the vessels during autoclaving, to prevent microbial contamination the only system boundaries of the bioreactor systems were silicon tube lines which allowed air flow through 0.2 µm MiniSart filters (Sartorius Stedim, Ger). For the MKII system, the flat blade agitator used to mix the culture was also wrapped in aluminium foil and autoclaved separately at the same time.

All bioreactor vessels connected place-holder reservoir bottles and foil-wrapped flat blade agitator were sterilized through autoclaving at 121 °C and 1.03 bar pressure for 40 minutes.

Once autoclaved, the media, acid, and base reservoirs (and the MKII flat blade agitator), were connected to the system in a class II microbiological safety cabinet to minimize the risk of microbial contamination.

Seed culture

For each bioreactor growth experiment, fresh working plates of *P. thermoglucosidasius* cells were prepared by streaking from a glycerol stock onto a 50°C pre-warmed solid 2TY agar plate. Cells were then incubator for 14-18 hours in a 55°C static incubator. Single colonies from those plates were used to inoculate 12.5 mL of liquid 2TY media in a 50 mL centrifuge tube. This seed culture was then grown at 55°C, 220 RPM in an Innova 44 shaking incubator (New Brunswick Scientific, UK) for 14-18 hours.

Sub-culture and inoculation

In order to acclimate the *P. thermoglucosidasius* cells as they transitioned from rich media to minimal media, 2 mL of the 2TY seed culture was sub-cultured into 10 mL of sterile, pre-warmed (60^{o}C) ASM at pH 7 in a 50 mL centrifuge tube. This was

then incubated at 60°C with shaking at 220 rpm in an Innova 44 shaking incubator (New Brunswick Scientific, UK) until the culture reached an OD_{600} of approximately 1.5.

At this point, 5 mL of culture was aseptically inoculated into either the MKI or MKII bioreactors containing ASM which had been previously equilibrated to pH 7, 60°C and the air flow rate of the specific experiment. Cells were grown in batch mode under the specific parameters of each experiment until an OD_{600} value of approximately 1 was reached. At this point, continuous culture conditions were initiated at the dilution rate of the specific experiment.

Continuous culture growth experiments

For all *P. thermoglucosidasius* growth experiments, cells were first grown under the target continuous culture conditions of the experiment in naturally labelled ASM until reaching a metabolic steady state after approximately 3.5 cell culture volume changes (CCVCs). This metabolic steady state was determined by obtaining consistent OD_{600} values and consistent fermentation product concentrations of three samples taken at different time points as indicated HPLC analysis of sample cell culture filtrate (CCF) fermentation products from three separate time points. Once metabolic steady state was achieved and target fermentation status was concluded to have been met, naturally labelled and/or pilot growth experiments were ended.

Procedure for starting the INST-^{13}C-MFA ILEs

For all *P. thermoglucosidasius* INST-^{13}C-MFA ILEs, cells were first grown in naturally labelled ASM until reaching a metabolic steady state as defined above. Once shown to be growing at a metabolic steady state, the system was aseptically switched to grow on either the 0.3% ^{13}C glucose or 0.3% ^{13}C xylose labelled ASM, starting theInstationary ^{13}C-MFA experiment.

2.2.1.3 Sampling cell culture during chemostat growth experiments

As the *P. thermoglucosidasius* continuous culture growth experiments progressed, 500 µL technical triplicate samples of cell culture were taken at experiment-specific time points for: OD_{600} readings, HPLC analysis of cell culture filtrates and GC-MS analysis of proteinogenic ^{13}C amino acids.
Under the assumption that any evaporated products in the Biostat B bioreactor system would no longer condense back into the culture due to an issue with the condenser, a single 500 µL sample of the condenser outlet dreschel bottle was taken at each time point of the glucose anaerobic 0.075 h^{-1} ^{13}C-MFA experiment. For each sample, the culture supernatant was immediately separated from the cell pellet by centrifugation at 14,000*g* for 10 minutes. For each culture sample taken, the separated supernatant was filtered through a 0.22 µM syringe filter (Phenomenex, USA) into a 1.5 mL micro-centrifuge tube to obtain a sterile, cell-free CCF. 100 µL of CCF from aerobic experiments or 200 µL of CCF from anaerobic experiments was then pipetted and mixed into either 900 or 800 µL respectively of dH_2O in a sealed glass HPLC vial for HPLC analysis. The separated cell pellets were stored at -20°C until later cell hydrolysis.

Optical density of samples at 600nm (OD_{600})

The optical density of cell culture samples at a wavelength of 600 nm (OD_{600}) was used to monitor the growth of the cells in all experiments. The OD_{600} was measured using 1:10 culture: water dilutions using a Jenway 6305 UV/Visible Spectrophotometer (VWR, UK).

Biomass Composition estimation

The biomass compositions of *P. thermoglucosidasius* NCIMB 11955 used in this research were measured under chemostat growth conditions by Dr. Shyam Masakapalli in 2014 according to the method of Durot *et al*. (2009)[123]. Biomass compositions were determined for *P. thermoglucosidasius* NCIMB 11955 cells grown at 60°C, pH 7 on 1% Glucose ASM under aerobic and anaerobic conditions (indicated by redox potential) and was repeated in biological duplicate.

2.2.1.4 High-Performance Liquid Chromatography (HPLC) analysis of extracellular cell culture metabolites and residual sugars

In order to gauge the fermentation status of *P. thermoglucosidasius* cells during continuous culture growth and to quantify the extracellular concentrations of fermentation products and any residual sugars, CCFs were analysed via either a High-Performance Liquid-Chromatography-Refractive Index Detector (RID)-Variable-Wavelength Detector (VWD) (HPLC-RID-VWD) system (University of Bath) or a HPLC-Diode Array Detector (DAD)-RID (HPLC-DAD-RID) system (IIT Mandi).

All CCF samples were then analysed for the extracellular concentrations of the metabolites: ethanol, acetate, formate, lactate and pyruvate and any residual glucose or xylose (Table 6).

Detector	Cell Culture Filtrate Compound	University of Bath Retention Time Range (mins)	IIT Mandi Retention Time Range (mins)
RID	Glucose	9.7 -> 10.4	10.5 -> 10.6
	Xylose	10.7	11.2
	Ethanol	21 -> 23	22.8 -> 23
UV/DAD	Pyruvate	9.9 -> 10	9.9 -> 10
	Lactate	13 -> 13.5	13.6 -> 13.8
	Formate	14 -> 14.5	15.7 -> 16.0
	Acetate	15 -> 15.5	14.5 -> 14.7

Table 6: Table of expected retention times (mins) on HPLC columns at the University of Bath and IIT Mandi.

HPLC-RID-UVD - The University of Bath, UK

At the University of Bath, CCF analysis was performed using an Agilent 1200 Series HPLC system (Agilent, USA) equipped with a Rezex™ RHM Monosaccharide-H-column (300 x 7.8 mm, 8 µm, Phenomenex, USA) with compound separation achieved using a mobile phase of 5 mM H_2SO_4 at a 0.6 mL/min flow rate and a column temperature of 65°C.
Ethanol, glucose and xylose were detected and quantified using the Agilent 1200 Series' integrated G1362A Infinity Refractive Index Detector (Agilent, USA).The organic acids acetate, formate, lactate and pyruvate were detected and quantitated using a G1314B VWD set to an absorption wavelength of 215 nm. The resulting chromatograms and peak areas were visualised and integrated using the ChemStation software and the resulting concentrations of metabolites were derived using standard curves for each chemical generated

over a concentration range of 2.5-100 mM. Peak areas below the range of these standard curves were assumed to be below the limit of detection of the instruments and were thus recorded as zero values for further analysis.

HPLC-DAD-RID - The Indian Institute of Technology, Mandi

At the IIT Mandi, CCF compound separation was achieved on an Agilent 1260 Infinity HPLC system (Agilent, In.) equipped with either a Rezex™ ROA-Organic Acid H+ (8%), LC Column (300 x 7.8 mm, 8 µm, Phenomenex, UK) or a SUPELCOGEL C-610H LC column (300 x 7.8 mm, Sigma-Aldrich, In). CCF compound separation was achieved with an injection volume of 10 µL at a column temperature of a 65°C using a mobile phase of 5 mM H_2SO_4 at a flow rate of 0.6 mL/min.

All CCF compounds were separated over an elution time of 30 minutes and were identified based on the retention times of authentic standards (Appendix Table 3). The resulting chromatograms and peak areas were visualised and integrated using the Agilent EZChrom Elite software (Agilent, In).

The cell culture filtrate concentrations of glucose, xylose and ethanol were quantified with the integrated Agilent 1260 Infinity RID using standard curves derived from seperate 2, 4, 6, 8, and 10 g/L glucose and xylose in water standards and 0.2, 0.4, 0.6, 0.8, and 1 g/L ethanol in water standards respectively. The cell culture filtrate concentrations of acetate, lactate, formate and pyruvate were quantified using the integrated Agilent 1260 DAD (Agilent, In.) set to an absorption wavelength of 214 nm and standard curves derived from 0.2, 0.4, 0.6, 0.8, and 1 g/L standards.

2.2.1.5 Gas chromatography mass spectrometry (GC-MS) analysis of proteinogenic amino acids and extracellular compounds

The complete proteinogenic amino acid analysis procedure was developed from a combination of: You *et al.* (2012)[124] and in particular Antoniewicz *et al.* (2019)[125].

Cell hydrolysis

In order to convert all cellular protein to its constituent proteinogenic amino acids, all biomass pellets were first re-suspended through the addition of 500 µL of 6N HCl and then transferred to acid resistant 2 mL cryogenic storage vials (Fisher scientific, UK). Each sealed cryogenic storage vial was then incubated at 100°C for 20 hours in an oven to fully hydrolyse the biomass to amino acids. A metal block was placed on top of the vials to prevent evaporation.

Hydrolysate separation and drying

After hydrolysis, the samples were transferred to separate 1.5 mL microcentrifuge tubes and centrifuged at 14,000*g* for 5 minutes to concentrate residual cellular material.

450 µL of supernatant of each sample were then transferred to another new 1.5 mL microcentrifugre tube. Samples were then dried at 60°C under an air stream for 12-14 hours in order to prepare the amino acids for the subsequent moisture-sensitive silylation.

Amino acid Derivatization with MtBSTFA + 1% t-BDMCS

Derivatization of the polar -OH, -NH_2 and -SH moieties of the amino acids in the hydrolysates is required to make the amino acids both less reactive and volatile enough such that they can be adequately separated by GC.

Firstly, 35 µL of pyridine was added to each dried sample to act as a solvent and the samples were resuspended through gentle aspiration. The sample mixtures were incubated at 60°C for 15 minutes on a static heat block to dissolve the amino acids and to equilibrate the reaction mixture. Then 50 µL of the *N-tert*-butyldimethylsilyl-*N*-methyltrifluoroacetamide (MtBSTFA) + 1% *tert*-butyldimethylchlorosilane (t-BDMSCl) derivatization agent (Merck, UK) was added to each sample and the samples were incubated at 60 °C for 35 minutes on a static heat block. (Figure. 14).

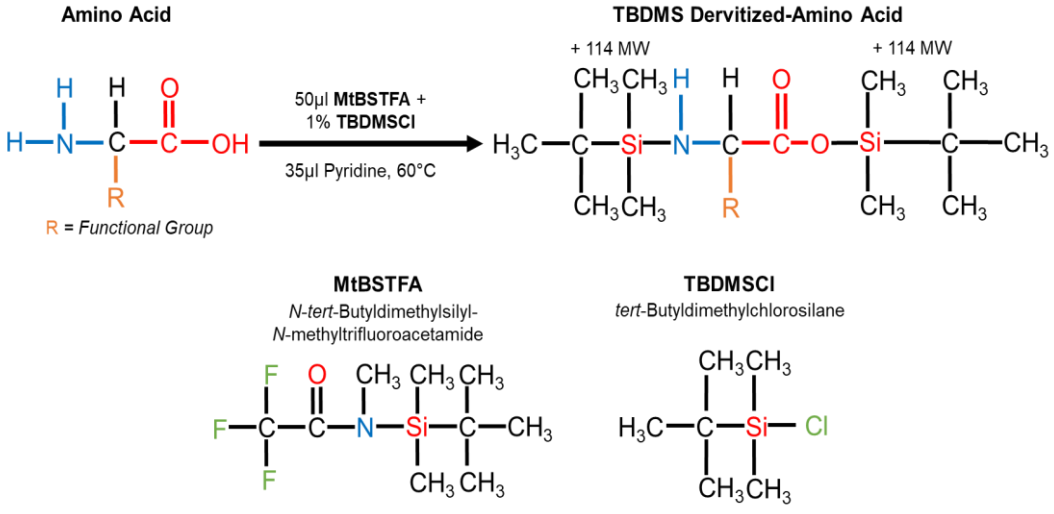


Figure 14: The amino acid chemical derivitization process featuring the structures of *N-tert*-butyldimethylsilyl-*N*-methyltrifluoroacetamide (MtBSTFA) and *tert*-butyldimethylchlorosilane (t-BDMSCl).

The active hydrogens in exposed -OH, -NH_2 and –SH moieties then undergo substitution to be replaced with a tert -Butyldimethylsilyl (TBDMS) moiety. This adds 114 per substitution to the molecular weight of the final TBDMS amino acid derivatives. Once derivitized, the samples in microcentrifuge tubes were centrifuged at 14,000*g* for 1 minute to condense any remaining cell debris and 75 µL of each derivitized sample was transferred to a labelled GC vial with 100 µL insert for GC-MS analysis.

Gas chromatographic separation and detection programmes

The University of Bath, UK

Compound separation and analysis was performed at the University of Bath using electrospray ionisation (EI) on an 8890-gas chromatography (GC, Agilent) system coupled with a 5977B MSD (MS, Agilent) GC-ALS-MS-5977B (GC-MS) system (Agilent, UK) equipped with a HP-5MS (5%-Phenyl)-methylpolysiloxane column (30 m, 0.25 mm, 0.25 µm, Agilent, UK) and quadrupole detector using He (BOC) as a carrier gas at a constant flow rate of 1.0 mL/min.

Split injections of 1 µL sample were performed, with a 10:1 split ratio (split flow of 10 mL/min), using a split inlet liner, lined with deactivated glass wool (210-4022-5 Agilent, UK). The inlet temperature was 280ºC with a 3.0 mL/min septum purge flow. The column oven was initially held at 100ºC for 1.5 minutes before a 20 ºC/min ramp to 130ºC, followed by a 10 ºC/min ramp to 220ºC with a 3-minute hold. This was then followed by a 5ºC/min ramp to 280ºC with a 4-minute hold, resulting in a total analysis time of 31 minutes. The MSD transfer line was set at 280°C, the MSD source set at 230°C, and the MSD quad temperature was set at 150°C.

After a 7-minute solvent delay, full scan mode detection was performed over a 180-550 m/z range, with a scan time of 1562 µs, and a gain factor of 15.

The Indian institute of technology, Mandi

Compound separation was performed at IIT Mandi using electrospray ionisation (EI) on a GC-ALS-5977B-MSD (GC-MS) system (Agilent, In) equipped with a HP-5MS (5%-Phenyl)-methylpolysiloxane column (30 m x 250 µm x 0.2 µm, Agilent, In) and quadropole detector using He (BOC) as a carrier gas at a constant flow rate of 1.3 mL/min. The inlet temperature was 270ºC. After 1 µL splitless injection, the column oven gradient was initially held at 120ºC with a 5-minute hold before a subsequent 4 ºC/min ramp to 270ºC where it was held for 3 minutes. This was then followed by a 20ºC/min ramp to 320ºC, and which was held for 1 minute. Baseline correction of all raw chromatograms was then performed using the metAlign 3.0 tool[126] in MATLAB version R2017b (The MathWorks Inc, Natick, USA).

At both institutions and for each GC-MS analysis run, the retention time for each TBDMS-derivitized amino acid was identified through triplicate 20, 15 10 5 1 0.5 and 0.1 ng/µL amino acid standards subjected to the same hydrolysis, drying and derivatization protocol detailed in 2.2.1.5.

2.2.2 Computational methods for instationary ^{13}C-MFA

2.2.1.2 GC-MS data processing and TBDMS-derivitized molecular mass fragment validation

GC-MS data visualization and identification of metabolites

Raw GC-MS chromatograms were visualized, and initial data interrogation was performed using Agilent MassHunter Workstation Qualitative analysis v.10.0 (Agilent, UK). Each GC peak of the amino acid standard chromatograms was interrogated, and TBDMS-derivitized amino acid fragments were identified, through spectral matching to the National Institute of Standards and Technology (NIST) 17 mass spectral database. The retention times for the peaks corresponding to the fragments were then recorded and the process was repeated for the naturally labelled samples to identify other TBDMS-derivitized compounds of interest.

Total Ion Current (TIC) chromatogram extraction

For each sample, individual total ion current (TIC) chromatograms for the 180-550 m/z range were extracted for each retention time range corresponding to the peak area of a TBDMS-derivitized amino acid or compound of interest.

Natural isotope abundance mass correction using IsoCor v2.2.0

As the obtained mass isotopomer distributions will also be affected by the presence of the naturally occurring stable heavy mass isotopes: ^{2}H, ^{13}C, ^{15}N, $^{17\text{-}18}O$, $^{29\text{-}30}Si$, and $^{33\text{-}36}S$ in both the metabolites and TBDMS derivatives, natural isotope correction is required. Natural isotope correction was performed for each extracted compound TIC through the IsoCor v2.2.0 tool[117] in Python 3.5 using the manually generated metabolite and derivative tables in described in materials 2.1.3.1 and yielding the corrected mass isotopomer distributions of each amino acid molecular weight ion.

For each technical triplicate sample, an average ^{13}C incorporation was then determined from for each mass isotopomer from the resulting corrected fractional abundances.

Defining the validity of TBDMS-derivitized molecular mass fragments for flux analysis

As the natural isotope abundance corrected TICs pertaining to each amino acid also contain background noise ion counts as well as the desired molecular fragment ion counts, it is necessary to determine which of the characteristic TBDMS-derivitized amino acid and other molecular ion mass fragments (Figure 15 and Table 7) should be used for metabolic flux analysis.

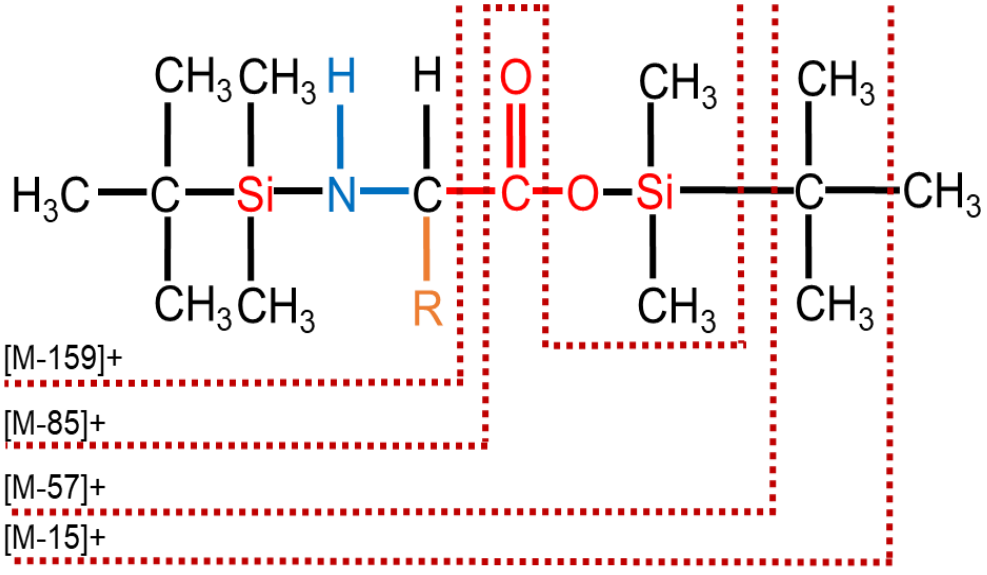


Figure 15: Depiction of the structures commonly examined molecular ion mass fragments resulting from electrospray ionization. [M-15]+ represents the loss of a -CH_3 or methyl group, [M-57]+ represents the loss of a -C_4H_9 or tert-butyl group, [M-85]+ represents the loss of a-C_4H9 tert-butyl group and a -CO group and [M-159]+ represents the loss of a -CO-O-TBDMS moiety.

Table 7: Anticipated metabolites and expected m/z ions corresponding to mass fragments.

Amino Acid	Base Molecular Weight (g/mol)	Assumed TBDMS groups post-derivitization	Full TBDMS derivitized Molecular Weight (g/mol) [M-0]+	Mass Fragment (m/z)			
				[M-15]$^+$ (-CH_3)	[M-57]$^+$ (-C_4H_9)	[M-85]$^+$ (-C_4H_9 + CO)	[M-159]$^+$ (CO-O-

							TBDMS)
Alanine	89.00	2	317	302	260	232	158
Glycine	75.07	2	303	288	246	218	144
Valine	117.15	2	345	330	288	260	186
Leucine	131.17	2	359	344	302	274	200
Isoleucine	131.17	2	359	344	302	274	200
Proline	115.13	2	343	328	286	258	184
Methionine	149.21	2	377	362	320	292	218
Serine	105.09	3	447	432	390	362	288
Threonine	119.12	3	461	446	404	376	303
Phenylalanine	165.19	2	393	378	336	308	234
Aspartate	133.11	3	475	460	418	390	316
Histidine	155.15	3	497	482	440	412	338
Glutamate	147.13	3	489	474	432	404	330
Lysine	146.19	3	488	473	431	403	329
Tyrosine	181.19	3	523	508	466	438	436

In this case, the fragments after natural isotope correction were qualified as either ‘Valid’, ‘Conditional Valid’ or ‘Invalid’. The characteristic molecular weight ions considered ‘Valid’ to represent each amino acid or compound in flux analysis estimation all had a calculated mean enrichment value for the unlabeled samples as less than or equal to the assumed natural abundance of carbon^{-13} of 1.11%. In cases where none of the characteristic molecular weight ions for a particular amino acid had an appropriate enrichment percentage, any of the characteristic molecular weight ions with natural enrichment of <2% was taken as “Conditionally Valid”. Those that did not fit those criteria were considered ‘Invalid’ for further use in flux analysis.

Complete mass isotopomer distributions of fractional abundances encompassing each isotopically instationary ^{13}C-MFA growth experiment were then created for all 'Valid' and 'Conditional Valid' molecular weight ions.

2.2.1.3 Flux estimation

Carbon Transition Network Construction

A network model of stoichiometric, carbon transition reactions representing *P. thermoglucosidasius* Central Carbon Metabolism (CCM) was defined in the Isotopomer Network Compartmental Analysis (INCA) v.2.0 software package[118] run in MATLAB R2017b (The MathWorks Inc, USA). This metabolic network and associated reaction directionality was adapted first from the carbon transition network defined by Cordova *et al.* (2016)[125] for *Geobacillus* LC300 and influenced by the metabolic networks devised by: Tang *et al.* (2009)[128] for *Parageobacillus thermoglucosidasius* M10EXG, the PhD thesis of Dr. Charlotte Ward (2014) for *P. thermoglucosidasius* NCIMB 11955[121] and the total network for the *p-thermo* GSMM presented in Chapter 3. The carbon transition network was built using the elementary metabolite unit (EMU) algorithm which combines the reaction stoichiometry, reversibility, and positional carbon information into each atom transition reaction. To act as additional constraints, reactions included the generation and usage of ATP, NADH, NADPH and one carbon metabolism metabolites methylenetetrahydrofolate (MEETHF), methyltetrahydrofolate (METHF) and formyltetrahydrofolate (FTHF). The metabolites succinate, fumarate which can act as symmetrical substrates to produce two products were defined as being symmetrical in the INCA software to prevent ^{13}C scrambling throughout the network.

The network itself represented the major pathways of upper metabolism, including glycolysis, gluconeogenesis and the pentose phosphate pathway, lower metabolism, including the Citric acid cycle, glyoxylate pathway and anaplerotic reactions, condensed amino acid synthesis pathways, fermentation efflux pathways and a biomass formation pseudo-reaction to operate as a sink for amino acids and carbon intermediates needed for growth. The proportion of protein in *P. thermoglucosidasius* NCIMB 11955 biomass, and the amino acid composition of that protein that was required for the biomass equation coefficients, had been measured previously by Shyam Masakapali and reported in Lisowska (2016)[113]. The coefficients of each biomass component were normalized to the rate of substrate uptake and input the units of millimole per gram of Dry Cell Weight per hour (mmol g^{-1} DCW h^{-1}). For pathways not included in the model, such as nucleotide, cell wall and lipid biosynthesis, the central carbon metabolite which operates as a precursor was used to represent it in the biomass equation. These stoichiometric values were taken from Codrova *et al.* (2016)[125] and assumed as constant for all ILEs. Measured yields of fermentation products were also normalized to the rate of substrate uptake and were applied to the respective upper and lower bounds of the respective efflux reactions. This effectively forces efflux of each fermentation product to its experimentally measured value (i.e., the lower bound for each amino acid sink reaction was set to a positive value).

This carbon transition network, with reactions included or omitted based on growth conditions, was used as the metabolic network framework for all the INST-^{13}C-MFA ILEs, which and is discussed in 4.4.4.1. The complete reaction network is available in Appendix 4.

2.2.1.4 Experimentally derived metabolic network constraints

The flux values for the amino acid sink pseudo-reactions were ultimately normalized to the specific sugar uptake rates of *P. thermoglucosidasius* in each INST-^{13}C-MFA experiment.

OD600nm to gDCW conversion
The conversion factor used for OD_{600} to gDCW for *P. thermoglucosidasius* NCIMB 11955 was measured by Jinghui Liang (pers. Comm.) to be 0.42 g/L DCW per 1 AU measured at OD_{600}.

Adapting metabolite concentrations into reaction constraints
For each isotope labelling experiment, the concentrations of extracellular fermentation products, overflow pyruvate and the concentrations of sugar in the culture and ^{13}C ASM reservoir were measured from CCF samples by HPLC analysis. A technical triplicate sample of media in the ^{13}C ASM reservoir was taken at the end of the experiment to determine the concentration of either glucose or xylose in the ^{13}C substrate reservoir.
The combination of biomass composition measurements, OD_{600nm} and specific rates of substrate uptake were used to calculate fluxes for amino acid coefficients in the biomass equation, and for fermentation product transport reactions, in units of mmol g^{-1} DCW h^{-1}.

For fermentation product transport reactions, the measured concentrations of product were normalized to the specific substrate uptake rate of the cells. As values were calculated for every time point, the maximum calculated value was set as the upper bound of the reaction and the minimum calculated value was set as the lower bound of the reaction. However, only one value could be applied to represent the biomass coefficient reactions in INCA, despite the dynamic nature of the experiments. Thus, for each experiment, to account for the effects of the variation in the OD_{600nm} measurements and sugar concentration in the media (which underpin the specific substrate uptake rate), an average value was calculated for each amino acid considering every sample taken during the ^{13}C portion of each isotope labelling experiment.

Metabolic model simulation using experimental data
Each experiment's fully processed and normalized mass isotopomer distributions of validated TBDMS derivitized amino acid fragments, and their respective standard errors, were input into INCA.

The positional and purity information of the ^{13}C tracer for each experiment was defined as input pools and applied to the relevant experiments. Additionally, on account of carboxylation reactions within the network, naturally labelled CO_2 (i.e., CO_2 with an enrichment of 1.13 % ^{13}C) was defined as a third potential 'tracer'. Flux estimation was performed for each experiment in INCA through a Levenberg-Marquardt (local search) algorithm for 10 independent resets to minimize the variance-weighted SSR between forward the (forward) simulated flux networks and

the experimentally determined measurements until the best fit, complete with associated individual reaction flux error values, was achieved.

2.2.1.5 Statistical analysis

Goodness-of-fit of the experimentally determined Mass Isotopomer Distributions (MIDs) and the minimum SSR fit solution for each data set was assessed through a χ^2-statistical test which calculates threshold SSR values for different confidence intervals based on the number of fitted MIDs and independent fluxes of the network as part of the inbuilt capabilities of the INCA v.1.7 software package.

This can be set to a confidence interval of choice, but flux distributions are conventionally deemed as successful fits if the resulting variance weighted SSR falls below a threshold value representing a confidence interval of 95%. If the target SSR threshold value has been surpassed, a statistically reliable metabolic flux distribution representing a best fit flux solution for metabolic phenotype under the cellular growth conditions has been obtained. In situations where the minimized SSR failed to reach the target threshold value, the metabolic flux distribution was regarded as a local flux distribution solution.

3. Chapter 3: GSMM to platform for design and discovery

3.1 Introduction:

3.1.1 Reconstructing the metabolic network topology of an organism

The first stage in generating a GSMM of a specific organism is to reconstruct the foundational topology of the metabolic network of interest from annotated whole genome sequencing data.

This strategy identifies open reading frames (ORFs) in the genomic sequence and aims to identify and annotate the functional role/s of the corresponding proteins in the metabolic network, commonly known as gene-protein-reaction (GPR) associations, by comparing their similarity to existing annotated sequences[129,130].

Several published tools can automatically carry out annotation of genomic features of draft genomes including: the RAST server[131], SEED[132], Prokka[133], BG7[134] and the NCBI's Prokaryotic Genome Annotation Pipeline[135]. Initial genome annotation was performed for the GSMM of *P. thermoglucosidasius* NCIMB 11955 by Lisowska (2016) [113] using the RAST server.

Recently, pipelines have also been developed which can perform both genome assembly and functional gene annotations such as GAAP[136] and the BRAKER2 pipeline for eukaryotes[137]. In this way, enzyme coding sequences annotated in these draft genome reconstructions can effectively be connected to known metabolic functions and reactions in the model. However, a recurring issue with these metabolic network reconstructions is that automatic annotation can result in erroneous, incomplete and/or lower confidence annotations. This can result from a combination of failure to account for potentially multi-functional proteins, failure to suggest singular metabolic function which results from multiple gene product or the underlying quality of the genome sequencing data itself[138].

Additionally, in cases of less-well annotated organisms with infrequent codon usages or specialised growth characteristics, such as the thermophilic nature of *P. thermoglucosidasius*, automatic annotation tools that rely on sequence homology may mis-annotate or fail to recognise more novel enzymes or pathways.

As a result, before meaningful predictions can be made with these models, a degree of manual assessment, verification, and curation of the suggested GPRs in the draft reconstruction is typically required[138]. Such curation can be manual or semi-automated[139] and can draw on a range of catalogued reference databases that can provide organism-specific evidence for the functions of proteins and larger metabolic reaction pathways (Table 8).

Table 8: Databases of biochemical information available online with their relevancy for reconstruction and curation of GSMMs. Table adapted from Simeonidis and Price (2015)[140]

Name	Relevant information for GSMM curation	URL	Reference
BiGG	Repository of over 70 curated GSMMs.	http://bigg.ucsd.edu/	King *et al.* (2015)[141]

BRENDA	Enzyme reaction database with organism-specific enzyme activity, specificity and localization data annotated with GPR associations and references.	https://www.brenda-enzymes.org	Chang *et al.* (2020)[142]
EC2PDB	Enzyme database of the Protein Data Bank (PDB) with curated catalytic reactions of proteins and structural information	https://www.ebi.ac.uk/thornton-srv/databases/enzymes/	Lakowski *et al.* (2017)[143]
KEGG	Knowledgebase of organism-specific genomic, reaction and metabolic pathway function information. Provides tools to map whole-genome sequencing information onto standardized metabolic pathway maps of metabolites and reactions.	https://www.genome.jp/kegg/	Kanehisa and Goto (2000)[144]
MetaCyc	Database of annotated metabolic genes, proteins, metabolites and reaction pathways for over 3000 organisms.	https://metacyc.org/	Caspi *et al.* (2013)[145]
MetaNetX	Platform for accessing GSMMs with tools for FBA, gap-filling and pathway modification	https://www.metanetx.org/	Moretti *et al.* (2020)[146]
NCBI	Sequenced gene, whole genome and protein database. Annotated gene and protein functions with taxonomic data.	https://www.ncbi.nlm.nih.gov/	NCBI Resource Coordinators [147]
Reactome	Metabolic pathway database with tools to visualise the enzymes, reactions and directionality in gene-sequencing supported metabolic networks.	https://reactome.org/	Jassal *et al.* (2019)[148]
TransportDB	Membrane transporter database with reaction equation and directionality data along with gene and protein sequence-informed taxonomic classifications.	http://www.membranetransport.org/transportDB2/index.html	Elbourne et al. (2016)[149]
UniProtKB	Protein knowledgebase with manually annotated (UniProtKB/Swiss-Prot) and automatically annotated (UniProtKB/TrEMBL) protein activity, kinetics, gene expression, metabolic pathway and sequencing data.	https://www.uniprot.org/	Bateman *et al.* (2020)[150]

In addition, several existing GSMMs have been deposited on the BiGG database[141]. In cases where organism-specific literature is scarce, the presence of reactions in

phylogenetically related organisms can help inform decisions on whether a particular reaction should be included in the overall metabolic network[138].

3.1.2 Iterative Model validation and curation using MEMOTE:

In addition to these reference compendia of biochemical and metabolic pathway knowledge, this research made particular use of the recently published MEMOTE tool developed by Lieven et al. (2020)[120]. This tool can evaluate metabolic models and highlight particular issues with the user-submitted metabolic network relating to its: Stoichiometric Consistency, Mass and Charge Balance and its Metabolite connectivity.

Once curated in this way, the accurate reconstruction can be considered a genome scale metabolic model ready for FBA.

3.1.3 Flux Balance Analysis: Metabolites, reactions and the stoichiometric matrix

A classical FBA problem begins with the conversion of all reactions in the genomic reconstruction of an organism, and their associated metabolites, to a mathematical model. This is performed through tabulation of this metabolic network in terms of a size *m*×*n* matrix formalization of interconnected linear mass balance equations known as a stoichiometric matrix (**S**)[152] :

$$\mathbf{S} = (\mathbf{S}_{i,j})_{m\times n}\{i = 1,\ldots,\boldsymbol{m} | j = 1,\ldots,\boldsymbol{n}\} \quad (1)$$

Whereby, the column vectors of the stoichiometric matrix represent the mass balance reactions ($\boldsymbol{n}_j$) in the metabolic network, including transport and exchange reactions, the rows of the stoichiometric matrix represent unique species/metabolites ($\boldsymbol{m}_i$) which operate as the reactants and/or products for those reactions and the matrix elements themselves ($\mathbf{S}_{i,j}$) denote the stoichiometric coefficients representing the relative number of moles for each metabolite *i* with respect to every reaction *j* in the network (Figure 16). Once completed, this defines the feasible space of all possible phenotypes of the organism[151].

Alongside the stoichiometric matrix, two additional vectors are defined. The first is a column vector of fluxes ($\vec{v}$) with a length of (***n***) which represents the flux rates ($\boldsymbol{v}$) through each reaction in the stoichiometric matrix and the second is a vector of metabolite concentrations (x) of length (***m***).
In the stoichiometric matrix, describing the change in concentration of a particular metabolite,x_i, in a particular reaction *j* (*i.e.* a matrix element $\mathbf{S}_{i,j}$), a positive coefficient indicates that the metabolite *i* is being generated in the reaction *j* and that the corresponding position in the flux vector ($\vec{v}$) there is a positive flux or gain of material ($\boldsymbol{v}_j$). A negative stoichiometric coefficient indicates that metabolite *i* is consumed in the reaction *j* and will have a negative flux or loss of material and a coefficient of zero indicates the metabolite *i* is not involved in the reaction *j.* Additionally, if a reaction is conditionally reversible and operates in the reverse direction, this will reverse the sign of the corresponding coefficients:

$$\frac{dx_i}{dt} = \sum_{j=1}^{n}\left(S_{i,j}^{+}\boldsymbol{v}_j\right) - \sum_{j=1}^{n}\left(S_{i,j}^{-}\boldsymbol{v}_j\right) \qquad i = 1,\ldots,\boldsymbol{m} \quad (2)$$

As every reaction in the matrix is a column vector which includes the coefficients of every species/metabolite in the metabolic network, the matrix's number of columns is also equal to the number of rows in the flux vector.
As a result, the fluxes and concentrations of metabolites for the entire metabolic reaction network can be defined as a dot product problem as follows:

$$\frac{d\vec{x}}{dt} = \mathbf{S} \cdot \vec{v} \qquad (3)$$

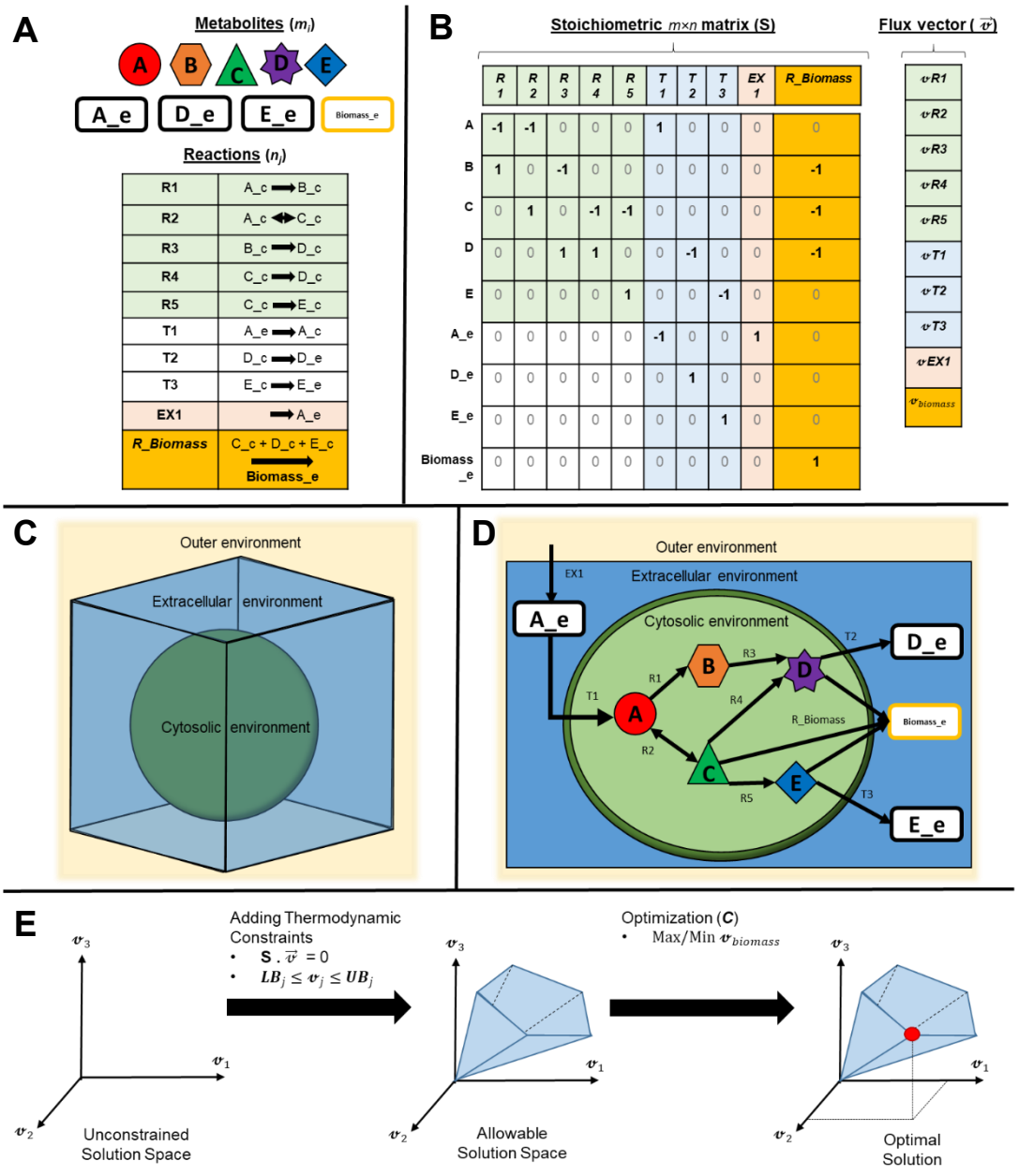


Figure 16: Mathematical and conceptual descriptions of reaction networks and FBA.
A) A reaction network for metabolites ($\boldsymbol{m_i}$) A, B, C, D and E is defined featuring: Intracellular Reactions (*j,* R1-5, Green), Transport reactions (T1-3), an Exchange reaction (EX1) and a demand reaction known as a biomass pseudo-reaction (*R_Biomass*). **B)** The five-metabolite reaction network of A formalized as an *mxn* stoichiometric matrix (**S**) with accompanying column flux vector ($\vec{v}$). Consumption of a metabolite in a particular reaction is represented by a negative coefficient, production by a positive coefficient and no involvement with a coefficient of zero. **C)** The conceptual boundaries of compartments in the GSMM. **D)** A pictorial representation of the five-metabolite reaction network featuring compartmentalisation and transport across boundaries. **E)** The applications of constraints to reach a solution in FBA. A mass balanced reaction network will yield an unconstrained solution space. The application of the steady-state assumption, reaction thermodynamic constraints ($\boldsymbol{LB}_{jMin}$ and $\boldsymbol{UB}_{jMax}$) and extracellular media constraints generates a feasible solution space. The application of an objective function, in this case maximizing the flux through a biomass reaction (Max $\boldsymbol{v}_{biomass}$), yields an optimal flux distribution solution.

3.1.3.1 Compartmentalisation, transport and exchange reactions

Each reaction is then considered in terms of the defined conceptual compartment in which it would take place (i.e., does a reaction take place inside or outside a cell?). In its simplest form, a model has two compartments (Figure 16 C+D). The first compartment is the external environment. This compartment is assumed as the environment surrounding and proximal enough to the cell such that metabolites within it can be moved into the cell by transport reactions.
The second is the cytosolic environment, compartmentalised by a systems boundary representing the cell membrane, in which most cellular metabolic reactions take place.

The compartment in which any metabolite in the model resides is defined, in this case, with a suffix _e for being present in the extracellular environment or _c for being present in the cytosol. Representing the boundary of the cell membrane, the movement of metabolite 'A' between the external environment and cytosolic compartments requires the definition of a stoichiometric exchange or transport reaction to enable the model to function:

$$A\_e <-> A\_c \quad (4)$$

Transport reactions are typically defined by supporting genomic data which suggests the presence of a key metabolite transporter. However, not only is such genomic evidence for transporters often limited[153], metabolites which can enter or leave via diffusion gradients or passive transport without inferred proteins must also be included.

While not defined as a separate compartment from the external environment, nor acting as an additional systems boundary, due to possessing a variety of transport mechanisms our model chose to include the concept of the zone outside the external extracellular environment in its the conceptual framework. Through what this research has defined as exchange reactions (separate from transport reactions) these metabolites are assumed to enter the external environment proximal to the cell from the outer zone.

With the definition of external and cytosolic metabolites, mass balanced reactions and the inclusion of transport and exchange reactions, a basic metabolic network of an organism could therefore be represented in stoichiometric matrix form (Figure 16.B).

3.1.4 Constraining metabolism: Constraint-based mathematical background

The validity and usefulness of predicted flux distributions constrained only by the mass balance stoichiometries in the stoichiometric matrix will be limited because they are not restricted by bounds of feasibility (Figure 16 E).

Thus, to interrogate this metabolic network in a meaningful way that accurately simulates metabolism, several constraints introducing thermodynamic and physicochemical principles must be applied to the metabolic network to restrict the overall solution space of flux distributions. The constraints include: the steady-state assumption, reaction thermodynamic constraints, the ionisation state of metabolites in both reactions and in free solution, extracellular media constraints and notably the objective function, an assumption of metabolic optimality, for defining a flux distribution of interest. When such constraints are applied, the restricted solution

spaces of stoichiometric matrices can be used to simulate all feasible cellular phenotypes of an organism under a set series of conditions and enables the characterisation of the functions of the metabolic network[154].

3.1.4.1 Steady-state

FBA of metabolic networks applies the physicochemical constraint of an assumption of a (pseudo) metabolic steady state. Under this constraint, it is assumed that all metabolite concentrations do not change with respect to time, i.e., the rates of generation and consumption of each metabolite are assumed to cancel each other out[122].

As a result of invoking this assumption, the dot-product problem of EQ. (3) can be simplified to:

$$\frac{d\vec{x}}{dt} = \mathbf{S} \cdot \vec{v} = 0 \quad (5)$$

$$\mathbf{S}\vec{v} = 0 \quad (6)$$

This simplification reduces the overall problem to a set of linear equations representing each reaction.

3.1.4.2 Thermodynamic Constraints: Reaction reversibility, upper and lower bounds

The thermodynamic constraints are typically applied to reactions within the network and define both the reversibility and overall directionality of each reaction, establishing a bounded range of potential flux rates for each reaction in the stoichiometric matrix with respect to the target environmental conditions of the simulation[154].

Reaction reversibility and directionality

The application of these thermodynamic constraints, such as reaction directionality and flux rate limits, are required to constrain reactions within the network to realistic levels. For example, without such constraints, not only could flux through a given reaction be infinite, but it could also operate at an infinite rate in an energetically infeasible direction.

Directionality and reversibility are typically determined by the Gibbs free energy of a reaction under the target environmental conditions of the simulation and can be further informed through *in vivo* observations or kinetic studies of related enzymes[156]. Likewise, flux rates limits can be measured experimentally and can also incorporate enzyme concentrations and kinetic constants under specific environmental conditions. These can be applied to any reaction, together with an associated error, to improve the accuracy of reactions in the metabolic network. Notably, this can also include transport reactions like substrate uptake or product excretion to constrain the edges of the solution space of the network. In situations where this flux rate information is not known, a baseline rate limit can be applied on top of the directionality constraint.
Additionally, in situations where genomic annotation has not indicated the presence of certain suspected reactions, Gibbs free energy calculations can indicate the feasibility of potentially spontaneous reactions.

Reaction flux bounds

These reaction directionality and rate constraints are codified in the form of applied upper and lower flux bounds (UBs, LBs), defining the flux rate maxima ($\boldsymbol{v}_{jMax}$) and minima ($\boldsymbol{v}_{jMin}$) for every reaction *j* in the metabolic network, narrowing the overall range of feasible fluxes in the solution space[151].

For every reaction rate in the network ($\boldsymbol{v}_j$), these bounds are set as follows:

$$\boldsymbol{v}_{jMin} \leq \boldsymbol{v}_j \leq \boldsymbol{v}_{jMax} \quad (7)$$

Where $\boldsymbol{v}_{jMin}$represents the lower bound of the reaction and $\boldsymbol{v}_{jMax}$ the upper bound. For a reaction which is only thermodynamically feasible in a single direction (as informed by $\Delta G^{0'}$) , i.e., an irreversible reaction, the bound in the infeasible direction will be set to 0 mmol g^{-1} DCW h^{-1}. In this way, simple enzyme knock-out simulations can also be introduced by setting both the UB and LB to 0 mmol g^{-1} DCW h^{-1}. Alternatively, setting non-zero values for the lower bound will force the model to carry minimal flux through the reaction if the substrates are available. Combined, these easy to apply changes enable rapid testing of genetic engineering hypotheses and different growth environments.

External Environment Constraints – an *in-silico* media

To accurately simulate the metabolic behaviour of organisms, it is important to apply as a constraint a precise definition of the growth media composition used for *in silico* cell growth that is represented in the extracellular environment of the metabolic model[152].

To change the environmental or media conditions, nutrients, carbon sources and other potentially importable metabolites must be defined as available in the *in-silico* growth media, and have a corresponding transport reaction, to be taken up and used in the model. Changing the upper and lower bounds of these transport reactions can also help represent substrate availability in the media. For example, if experimental measurements of the uptake or excretion rates of external metabolites exist, these can be applied to the metabolic model as upper and/or lower bounds constraints on the respective transport reactions[151]. If the metabolic network of the organism is dependent on any of these components for growth or particular phenotypes, their presence or absence acts as a constraint to the overall model[154]. To accurately reflect real-world growth scenarios, the composition of this *in silico* media can be defined to replicate known growth media recipes, or hypothetical substrate mixtures to evaluate what impact they may have on cell growth.

Together, these thermodynamic and environmental constraints improve the precision of any FBA solution obtained. However, even with these constraints applied, the system of equations is typically still underdetermined, with more unknown variables (reactions/fluxes) than metabolites (*j* > *i*). Therefore, rather than a unique flux vector solution to the system of equations, the stoichiometric matrix dot product instead yields a multi-dimensional solution space representing all possible cellular phenotypes under a defined set of environmental conditions[151] (Figure 16. E).

3.1.5 The objective function

This multi-dimensional flux solution space alone is of limited value for understanding metabolic flux distributions as it operates under the assumption that each reaction within the metabolic network is of equal weighting[152].
To narrow the FBA solution space to single vectors representing flux distributions of biological interest, the most important constraint to perform FBA that needs to be applied to the metabolic network is the assumption of optimality for a defined metabolic objective, such as the common targets of cell growth and efficient ATP generation or for flux through a reaction that affects target product yield[26].
This optimality is defined through an objective function, which usually involves the minimisation or maximization of flux through a single reaction in the metabolic network. This is introduced as an additional vector (***C***) to the network which defines the proportional contribution of each reaction in the network to the user-designated objective or phenotype of interest. Therefore, the vector ***C*** is usually comprised almost entirely of zeroes for every reaction but has a value of one at the position of the corresponding reaction of interest[151].

3.1.6 The Biomass Reaction

A conventional example of a FBA objective function is that of maximizing the flux through a demand reaction known as a biomass pseudo-reaction ($\mathrm{Max}\ \boldsymbol{v}_{biomass}$), which is used as a proxy for predicting maximum cell growth rate[157].

The biomass reaction is a mass balance reaction added to the stoichiometric matrix which consumes stoichiometric quantities of the constituent biomass components defined in the rows of the matrix, (including nucleic acids, amino acids, cell membrane and wall lipid components) and generates an artificial 'biomass' metabolite [26]. The stoichiometries of the biomass reaction can be derived experimentally from the dry weight composition of a cell[156]. To make it more accurate, the flux through this conversion reaction can also be scaled so that it is equal to an experimentally measured maximum cell growth rate[158].

The final maximization or minimization optimization problem for the whole metabolic field of behaviour of interest can therefore be simplified to:

$$\mathrm{Max}\ \boldsymbol{v}_{biomass} \tag{8}$$

Subject to: $\mathbf{S}\vec{v} = 0$

$$\boldsymbol{v}_{jMin} \leq \boldsymbol{v}_{j} \leq \boldsymbol{v}_{jMax} \quad \forall j$$

Where the objective function in this case is defined as the maximization of biomass generation through the solving of the stoichiometric matrix dot product, under the steady state assumption and with the applied thermodynamic and environmental constraints[127]. Through linear programming algorithms, this optimisation problem can then be solved to narrow the solution space output to a singular, vector filled with flux values ($\vec{v}$), which satisfies the objective function[157]. Therefore, this flux vector represents an optimal and feasible flux distribution for the metabolism of a modelled organism of interest which provides flux values, defined in units of mmol per gram of dry cell weight per hour (mmol g^{-1} DCW h^{-1}), for every reaction in the whole-organism metabolic network.

3.1.7 Limitations

FBA based purely on constrained stoichiometric matrices does have a few limitations which curtail the overall accuracy of this approach for modelling metabolism which are considered below.

3.1.7.1 Steady-state and Dynamic behaviour

As one of the key assumptions with classical FBA flux outputs is that the system is operating under a metabolic steady-state, actual dynamic kinetics, and concentrations of metabolites within the system cannot be accurately predicted[35], and there is a much more limited capacity to predict metabolic phenotypes which would result from any cascade responses to stimuli or structural changes to the stoichiometric matrix/metabolic network itself.

Indeed, while arguably accounted for in experimentally derived growth and non-growth associated maintenance energy calculations, classical FBA does not inherently take into account the energetic impact of: genetic regulation mechanisms, protein synthesis and potential enzyme activation by specific protein kinases[159]. As a result of not incorporating this information, seemingly more energy-efficient optimal paths may be taken through the metabolic network which do not accurately reflect their true energetic costs.

3.1.7.2 Optimisation

FBA will yield a vector predicting optimal flux distributions for a specific objective function. However, there is also a more philosophical question about the validity of having a biological 'objective', such as biomass production or ATP generation, as assumed 'goals' that cells have evolved towards in order to survive[160].

3.1.7.3 Insufficient gap-filling of the network:

Additionally, any FBA flux predictions will only be as accurate and valid as the metabolic network used to make them. Thus, poor gene annotation, insufficient gap-filling and limited reaction thermodynamic curation of the metabolic network will result in inaccurate flux prediction outputs.

3.1.8 Variations on objective functions and FBA improvements

Although maximizing the production of the 'biomass' pseudo-metabolite is one of the most common optimization targets, variation on the objective function used, and modified versions of FBA itself, can improve predicted flux distributions and help to mitigate some of the limitations of 'classical' FBA.

Alternatives to the standard biomass objective function typically revolve around the minimization and maximization theme. This can include things like minimizing substrate consumption fluxes to find the most efficient growth strategy, maximizing the yield of ATP as a direct alternative to biomass, maximizing the production of desirable metabolites, maximizing the consumption of a naturally low-priority substrate such as in bioremediation approaches or even maintaining certain ratios of ATP to NADH to support the production of a range of high energy cost metabolites [155].

From a systems metabolic engineering perspective, alternate objectives can also be used *in silico* to rapidly suggest, create and evaluate gene modifications to the metabolic network to produce metabolites of interest. Indeed, FBA-based algorithms that can better accommodate genetic engineering strategies can be more accurate predictors of modified phenotypes than standard FBA and hence more useful for evaluating strain designs. Two examples of these alternate objective strategies are the OptKnock [161] and the Minimization of Metabolic Adjustment (MOMA) [162] frameworks.

3.1.8.1 OPTKNOCK

OptKnock was developed by Burgard, Pharkya and Maranas (2003)[161] to use GSMMs more effectively to suggest and evaluate potential *in vivo* gene knockout strategies to produce metabolites of interest. The strategy embeds the assumption that cells have not evolved for the over-production of, potentially non-native, metabolites into a bi-level optimization problem. This problem introduces additional stoichiometry to reactions essential for growth, coupling these reactions to the production of the target metabolite such that the target metabolite effectively becomes a by-product of growth itself.

With this framing, OptKnock solves the bi-level optimization problem by optimizing for the maximum production of the target metabolite, subject to the maximization of the objective function. As an output, OptKnock identifies various combinations of gene knockouts, and the corresponding objective function values, that enable the maximum achievable flux towards the metabolite of interest.

3.1.8.2 The Minimization of Metabolic Adjustment (MOMA) framework

The Minimization of Metabolic Adjustment (MOMA) framework factors in an underlying assumption that new mutant strains may need adjustment to adapt to growing in a new optimal manner for their new metabolic network [162]. MOMA uses quadratic programming to effectively map the optimal wild-type flux distribution onto the new flux solution space of the modified metabolic network [161]. If *in silico* growth is feasible with the modified metabolic network, the most likely optimal flux distribution will therefore be one that can get as close as possible (in Euclidean distance terms) to the original optimal FBA solution with the minimal redistribution of fluxes across pathways.

Alternatively, modified versions of classical FBA can include strategies such as regulatory flux balance analysis (rFBA) which can introduce Boolean gene transcription and gene regulatory constraints to improve the sensitivity of a metabolic model to the specified environment and dynamic flux balance analysis (dFBA) which aims to account for metabolite concentration changes over time by performing several interative FBAs for discrete time intervals, then integrating over each time interval to yield a change in concentration [35].

Two similar variations on classical FBA performed in the validation of the *P. thermoglucosidasius* model are flux variability analysis (FVA) and parsimonious flux balance analysis (pFBA).

3.1.8.3 Flux Variability Analysis (FVA)

Although FBA should yield a single optimal solution, for a larger and underdetermined metabolic network this solution is not necessarily unique. Several different flux distributions through reactions could exist which return the same maximum value for the objective function e.g., the same growth rate or phenotype.

FVA can be used to identify alternate flux distributions which yield optimal solutions by holding the optimal objective function solution as a constraint and performing FBA to return the flux ranges for every reaction in the network[165]. FVA can therefore not only be used to identify alternate flux distributions but also the flexibility of the metabolic network and potential redundant reactions and can act as a form of sensitivity analysis of a particular FBA solution. Reactions that return a lower flux variability are of greater importance to achieving the objective function than those with higher variability[166].

Alternatively, by not setting an objective function and effectively giving every flux distribution equal weighting, FVA can also be used to establish what the maximum and minimum flux ranges of each reaction in the network could be[167].

3.1.8.4 Parsimonious FBA (pFBA)

pFBA is a variation of FBA which qualifies the influence of genes and their reactions in the network in achieving the optimal flux distribution under the specified constraints. Lewis *et al.* (2010) [168.]

Once again, pFBA takes the optimal objective function solution as a constraint, then minimises the flux for reaction in the metabolic network. This assumes that the strains which grow the fastest and most efficiently will require the minimal amount of flux to achieve the optimal solution.
With these results, classifications are then assigned to each metabolic gene/s underpinning the reactions, based on the extent to which flux through those reactions contributes to the optimal solution under the specified constraints. These genes are defined by Lewis *et al.* (2010) [168] as either:

1. *Essential genes,* which are necessary for growth,
2. *pFBA optima,* which are non-essential genes predicted to contribute towards growth,
3. *Enzymatically less efficient (ELE),* which are genes whose reactions can meet the target solution but represent a less efficient pathway than other available pathways.
4. *Metabolically less efficient (MLE),* which are genes predicted to be detrimental to growth, and
5. *pFBA no-flux,* which are genes whose associated reactions are unable to carry flux.

3.1.9 Flux Balance Analysis of *P. thermoglucosidasius*:

For the duration of the research, only two GSMMs of *P. thermoglucosidasius* were publicly available. The second of these was a GSMM of *P. thermoglucosidasius* C56-YS93 (designated iGT736) published by Ahmad *et al.* in 2017 and remains the only other formally published model [169]. As discussed in this chapter, this model could not be used for meaningful analysis as it lacked a biomass reaction, had an incomplete stoichiometric balance, and lacked several transports/exchange reactions.

This chapter however is built upon the foundation of the *P. thermoglucosidasius* (*P. thermoglucosidasius* NCIMB 11955) model reconstructed, annotated and initially evaluated by Dr. Beata Lisowska in 2016[113]. Lisowska performed the initial automatic genome annotation using the RAST server and subsequently performed extensive reaction network gap analysis and identified reaction stoichiometries, kinetics, directionality though Gibbs fee energy calculations, while cross-referencing it all to the sequenced genome of *P. thermoglucosidasius* NCIMB 11955 to shed light on it's broad substrate utilisation capabilities. Additionally, Lisowska noted that the genome of *P. thermoglucosidasius* NCIMB 11955 seemingly featured a blend of genes for enzymes in the aerobic and anaerobic biosynthetic pathways to Vitamin B12. On this basis, Lisowska proposed a theoretical novel route to Vitamin B12 synthesis which uses a combination both pathways. Lastly, analysis of the genome identified the presence of a NADP-dependent Glyceraldehyde 3-phosphate dehydrogenase (GAPDH) and quantified its activity with NADP using a linked enzyme assay.

However, network evaluation investigations as part of this research identified several stoichiometrically and mass unbalanced reactions. Notably, analysis of the provided FBA solution results of growth of *P. thermoglucosidasius* NCIMB 11955 on a rich media under aerobic conditions pointed towards missing transport reactions and an undefined growth media which could have greatly diminished the predictive potential of the GSMM in its original state. In particular, the original model lacked transport reactions for both CO_2 and oxygen. As a result, CO_2 demand seemed to be met through H_2CO_3 generated through reversal of pyruvate carbon dioxide ligase (Table 9). Due to an absence of oxygen import under nominally aerobic conditions, in the provided results nitrate was instead imported and nitrate was used both as a terminal electron acceptor and to meet the oxygen demand of the rest of metabolism by a reaction for an $NADPH_2$: oxygen oxidoreductase operating to produce O_2 from Nitrate (Table 9).

Table 9: GSMM Reactions displaying usage of CO_2 and nitrate. Flux units in mmol g^{-1} DCW h^{-1}. Table data taken from the PhD thesis of Dr. Beata Lisowska[113].

Name of reaction	Reaction Equation	Enzyme Commission Number	KEGG Reaction ID	Flux aerobic
Carbonic acid hydro---lyase	|H+| + |H2CO3| <=> |H2O|+ |CO2|	4.2.1.1	R00132	650.863
Pyruvate: Carbon---dioxide ligase (ADP---forming)	|ATP| + |Pyruvate| + |H2CO3| <=> |ADP| + |Phosphate| +|Oxaloacetate| +|H+|	6.4.1.1	R00344	-914.722
Nitrate transport in via proton symport	|H+[e]|+|Nitrate[e]|<=> |H+| + |Nitrate|	Undetermined	None	-997.48
Nitrite transport in Via proton symport	|H+[e]| + |Nitrite[e]| <=> |H+| + |Nitrite|	Undetermined	None	997.48
Nitric oxide, NADPH2:oxygen oxidoreductase	|NADPH| + (2)|O2| + (2)|NO| <=> |NADP| +|H+|+ (2)|Nitrate|	1.14.12.17	R05725	-1000
Nitric oxide, NAD(P)H2:oxygen oxidoreductase	|NADH| + (2)|O2| + (2)|NO| <=> |NAD| +|H+|+ (2)|Nitrate|	1.14.12.17	R05724	1000
Nitrate reductase (Menaquinol---8)	2)|H+| + |Nitrate| + |Menaquinol 8| <=> |H2O| + (2)|H+[e]| + |Nitrite| + |Menaquinone 8|	Undetermined	None	-997.48
cytochrome---c oxidase	(H+/e---=2) (0.5) |O2| + (6)|H+| + (2)|Cytochrome c2+| <=> |H2O| + (4)|H+[e]|+ (2)|Cytochrome c3+|	1.9.3.1,,1.9.3.1,	None	1000
Ferrocytochrome---c:oxygen oxidoreductase	|O2|+ (4)|H+| + (4)|Cytochrome c2+|<=> (2)|H2O|+ (4)|Cytochrome c3+ |	1.9.3.1	R00081	-405.687

As a result, the provided solution showed production, and where possible export, of known *P. thermoglucosidasius* fermentation products: acetate, lactate, formate and succinate (Table 10). Additionally, no transport reactions were present in the network for known excreted metabolites: acetate, pyruvate and ethanol.

Table 10: GSMM reactions displaying production and export of excreted metabolites. Flux units in mmol g^{-1} DCW h^{-1}. Table data taken from the PhD thesis of Dr. Beata Lisowska[113].

Name of reaction	Reaction Equation	Enzyme Commission Number	KEGG Reaction ID	Flux aerobic
Acetyl---CoA:formate C--acetyltransferase	|Acetyl---CoA| + |Formate| <=> |CoA| + |Pyruvate|	2.3.1.54	R00212	-1000
Acetate:CoA ligase (AMP---forming)	|ATP| + |CoA| + |Acetate| + |H+| <=> |PPi| + |AMP| + |Acetyl---CoA|	6.2.1.1	R00235	-133.974
ATP:acetate phosphotransferase	|ATP| + |Acetate| + |H+| <=> |ADP| + |Acetylphosphate|	2.7.2.1,2.7.2.15	R00315	-865.74
(S)---Lactate:NAD+ oxidoreductase	|NAD| + |L---Lactate| <=> |NADH| + |Pyruvate| + |H+|	1.1.1.27	R00703	-1000
L---lactate reversible transport Via proton symport	|H+[e]| + |L---Lactate[e]| <=> |H+| + |L---Lactate|	Undetermined	None	-811.375
Succinate transporter in/out via proton symport	|Succinate[e]| + |H+[e]| <=> |Succinate| + |H+|	TC---2.A.56,2.A.56	None	-1000
Formate transport in via proton symport	|Formate[e]| + |H+[e]| <=> |Formate| + |H+|	Undetermined	None	-950.569

Additionally, due to the lack of definition of the rich media external to the cellular compartment, the result was growth in an unconstrained media which could have every form of carbon substrate from carbohydrates to amino acids. While the potential for operation of these reactions may have genetic support, the unconstrained media likely resulted in an unrealistic optimal FBA solution involving the import of a mixture of carbon substrates, carbon intermediates and amino acids (Table 11).

Table 11: GSMM reactions displaying import of varied carbon sources. Flux units in mmol g^{-1} DCW h^{-1}. Table data taken from the PhD thesis of Dr. Beata Lisowska.[113]

Name of reaction	Reaction Equation	Enzyme Commission Number	KEGG Reaction ID	Flux aerobic
	Carbon Substrates			
Glucose---phosphotransferase (PTS) system	|D---Glucose[e]| + |Phosphoenolpyruvate| <=> |Pyruvate| + |D---glucose--6---phosphate|	Undetermined	None	770.125
Syn---Glycerol ABC transport	|H2O| + |ATP| + |Glycerol---3---phosphate[e]| => |ADP| + |Phosphate| + |H+| + |Glycerol---3---phosphate|	Undetermined	None	863.762
Mannitol transport via PEP:Pyr PTS	|Phosphoenolpyruvate| + |D-- Mannitol[e]| <=> |Pyruvate| + |D---mannitol---1---phosphate|	Undetermined	None	1000
Cellobiose transport via PEP:Pyr PTS	|Phosphoenolpyruvate| + |CELB[e]| <=> |Pyruvate| + |cellobiose 6---phoshate|	Undetermined	None	832.238
	Carbon Intermediates			
Malate transport via proton symport	(2H) (2)|H+[e]| + |L---Malate[e]| <=> (2)|H+|+ |L---Malate|	Undetermined	None	664.289
Fumarate transport via proton symport	(2H) (2) |H+[e]| + |Fumarate[e]| <=> (2) |H+| + |Fumarate|	Undetermined	None	1000
Xanthine ion---coupled transport	|H+[e]| + |XAN[e]| <=> |H+| + |XAN|	TC---2.A.40,2.A.40	None	1000
Cytosine transport in via proton symport	|H+[e]| + |Cytosine[e]| <=> |H+| + |Cytosine|	Undetermined	None	829.041
D---glucosamine transport via PEP:Pyr PTS	|Phosphoenolpyruvate| + |GLUM[e]| <=> |Pyruvate| + |D---Glucosamine phosphate|	Undetermined	None	365.679
	Amino Acids			
PHEt6	|L---Phenylalanine[e]| + |H+[e]| <=> |L---Phenylalanine| + |H+|	TC---2.A.3.1,2.A.3.1	None	27.9564
Aspartate transport via proton symport	(2H)|L---Aspartate[e]| + (2)|H+[e]| <=> |L---Aspartate| + (2)|H+|	Undetermined	None	1000
Arginine/ornithine antiporter	|L---Arginine[e]| + |Ornithine| <=> |L---Arginine| + |Ornithine[e]|	Undetermined	None	1000
L---proline transport in via proton symport	|H+[e]| + |L---Proline[e]| <=> |H+| + |L---Proline|	Undetermined	None	1000
L---lysine reversible transport via proton symport	|L---Lysine[e]| + |H+[e]| <=> |L---Lysine| + |H+|	Undetermined	None	51.3618
GLUt2	|L---Glutamate[e]| + |H+[e]| <=> |L---Glutamate| + |H+|	Undetermined	None	41.518
Na+:proline symport	|L---Proline[e]| + |Na+[e]| <=> |L---Proline| + |Na+|	Undetermined	None	1000
Branched chain amino acid:H+ symporter (Leucine)	|H+[e]| + |L---Leucine[e]| <=> |H+| + |L---Leucine|	Undetermined	None	20.516
L---valine transport in via proton symport	|H+[e]| + |L---Valine[e]|<=> |H+| + |L---Valine|	Undetermined	None	23.7121
Branched chain amino acid:H+ symporter (Isoleucine)	|H+[e]| + |L-Isoleucine[e]| <=> |H+| + |L---Isoleucine|	Undetermined	None	56.1028

Primarily, the accuracy of a GSMM and the flux predictions generated from it is of particular importance for mutant strain ME algorithms, such as the OptKnock and MOMA, which rely on an accurate simulation of the wild-type strain. If a GSMM is inaccurate, this will likely lead to incosistencies between the *in silico* generated flux distributions and experimentally obtained *in vivo* data. As a result, this research continued the development of Lisowska's GSMM for *P. thermoglucosidasius* NCIMB 11955 at the stage of manual refinement and curation of the reactions in the model, with the aim of developing it for use as a predictive model for the ME of *P. thermoglucosidasius* metabolism. The publication describing *p-thermo*, which appears as part of this chapter, represents the most complete GSMM of the genus *Parageobacillus* to date.

3.2 Genome-scale metabolic modeling of *P. thermoglucosidasius* NCIMB 11955 reveals metabolic bottlenecks in anaerobic metabolism

This declaration concerns the article entitled:			
Genome-scale metabolic modeling of P. thermoglucosidasius NCIMB 11955 reveals metabolic bottlenecks in anaerobic metabolism			
Publication status (tick one) Draft manuscript [ ] Submitted [ ] In review [ ] Accepted [ ] Published [✓]			
Publication details (reference)	Authors: Viviënne Mol*, Martyn Bennett*, Benjamín J.Sánchez, Beata K. Lisowska, Markus J.Herrgård, Alex Toftgaard Nielsen, David J.Leak, Nikolaus Sonnenschein		
Copyright status (tick the appropriate statement) I hold the copyright for this material [ ] Copyright is retained by the publisher, but I have been given permission to replicate the material here [✓]			
Candidate's contribution to the paper (provide details, and also indicate as a percentage)	Martyn Bennett (32.5%): Methodology, validation, investigation, data curation, writing. Vivienne Mol (32.5%): Methodology, validation, investigation, data curation, writing. Beata K. Lisowska (25%): Conceptualization, methodology, investigation. Benjamín J.Sánchez (5%): Methodology regarding lipid composition and supervision of the work. David Leak, Markus Herrgård (5%) contributed to the manuscript though editing and corrections, conceptualisation, ideas, funding and supervision of the work. Alex Toftgaard Nielsen, Nikolaus Sonnenschein and contributed to ideas, supervision and the manuscript though editing and corrections. All authors read and approved the final manuscript.		
Statement from Candidate	This paper reports on original research I conducted during the period of my Higher Degree by Research candidature.		
Signed	[redacted]	Date	08/07/2021

Metabolic Engineering 65 (2021) 123–134

Contents lists available at
ScienceDirect
Metabolic Engineering

Genome-scale metabolic modeling of *P. thermoglucosidasius* NCIMB 11955 reveals metabolic bottlenecks in anaerobic metabolism

Viviënne Mol [a, 1], Martyn Bennett [b, c, 1], Benjamín J. Sánchez [a, d], Beata K. Lisowska [b], Markus J. Herrgård [a, e], Alex Toftgaard Nielsen [a,***], David J. Leak [b, c,**], [2], Nikolaus Sonnenschein [d,*, 2]

[a] *The Novo Nordisk Foundation Center for Biosustainability, Technical University of Denmark, Kongens Lyngby, Denmark*

[b] *The Department of Biology & Biochemistry, University of Bath, Claverton Down, Bath, BA2 7AY, United Kingdom*

[c] *The Centre for Sustainable Chemical Technologies (CSCT), University of Bath, Claverton Down, Bath, BA2 7AY, United Kingdom*

[d] *Department of Biotechnology and Biomedicine, Technical University of Denmark, Kongens Lyngby, Denmark*

[e] *BioInnovation Institute, Copenhagen N, Denmark*

ARTICLEINFO

Keywords:

Genome-scale metabolic model
Systems metabolic engineering
Thermophile

In silico strain design
Anaerobic metabolism
Flux balance analysis

ABSTRACT

Parageobacillus thermoglucosidasius represents a thermophilic, facultative anaerobic bacterial chassis, with several desirable traits for metabolic engineering and industrial production. To further optimize strain productivity, a systems level understanding of its metabolism is needed, which can be facilitated by a genome-scale metabolic model. Here, we present *p-thermo*, the most complete, curated and validated genome-scale model (to date) of *Parageobacillus thermoglucosidasius* NCIMB 11955. It spans a total of 890 metabolites, 1175 reactions and 917 metabolic genes, forming an extensive knowledge base for *P. thermoglucosidasius* NCIMB 11955 metabolism. The model accurately predicts aerobic utilization of 22 carbon sources, and the predictive quality of internal fluxes was validated with previously published ^{13}C-fluxomics data. In an application case, *p-thermo* was used to facilitate more in-depth analysis of reported metabolic engineering efforts, giving additional insight into fermentative metabolism. Finally, *p-thermo* was used to resolve a previously uncharacterised bottleneck in anaerobic metabolism, by identifying the minimal required supplemented nutrients (thiamin, biotin and iron (III)) needed to sustain anaerobic growth. This highlights the usefulness of *p-thermo* for guiding the generation of experimental hypotheses and for facilitating data-driven metabolic engineering, expanding the use of *P. thermoglucosidasius* as a high yield production platform

Introduction

As the global transition away from petroleum-derived feedstocks continues, the need to produce commodity and fine chemicals using sustainable feedstocks has accelerated the interest in establishing microbial bioprocesses with lower environmental footprints (Steen et al., 2010; Lee and Kim, 2015; Nielsen and Keasling, 2016). The microbial 'chassis' organisms of these bioprocesses have been developed through modern metabolic engineering strategies. Such strategies have enabled the redirection of carbon flux in metabolic pathways of the corresponding microbes towards target products, in what are commonly termed 'microbial cell factories' (Hollinshead et al., 2014). Without an accurate picture of how cellular metabolism operates as a whole, metabolic engineering strategies can produce flux imbalances, resulting in the accumulation of carbon intermediates, metabolic bottlenecks and/or imbalances in the overall cellular redox ratio (Liu et al., 2017; Ma et al., 2018). As a result, there can be large upfront costs in microbial strain engineering to ensure economically viable biochemical product yields (Nielsen and Keasling, 2016; Tang et al., 2009). To bolster traditional metabolic engineering efforts and help elucidate genotype-phenotype relationships, systems metabolic engineering aims to describe a more holistic representation of cellular metabolism though the integration of stoichiometric modeling and -omics data analyses (Choi et al., 2019).

Abbreviations

Metabolites

13dpg	3-Phosphoglyceroyl phosphate
2pg	2-phosphoglycerate
3pg	3-phosphoglycerate
6pgl	6-phosphogluconolactone
ac	acetate
acald	acetaldehyde
accoa	acetyl-CoA
actp	acetyl phosphate
akg	α-ketoglutarate
asp	aspartate
cit	citrate
dhp	dihydroxyacetone phosphate
e4p	erythrose 4-phosphate
etoh	ethanol
f6p	fructose-6-phosphate
fdp	fructose 1,6-bisphosphate
for	formate
fum	fumarate
g3p	glyceraldehyde-3-phosphate
g6p	glucose-6-phosphate
glc	glucose
gly	glycine
icit	iso-citrate
lac	L-lactate
mal	malate
oaa	oxaloacetate
pep	phosphoenolpyruvate
phe	phenylalanine
pyr	pyruvate
r5p	ribose-5-phosphate
ru5p	ribulose-5-phosphate
s7p	sedoheptulose 7-phosphate
ser	serine
succ	succinate
succoa	succinyl-CoA
udpg	uridine diphosphate glucose
xu5p	xylulose-5 -phosphate

Reactions

ACKr	Acetate kinase
ACONTa	Aconitase
ALCD2x	Alcohol dehydrogenase
ASPTA	Aspartate transaminase
CS	Citrate synthase
DDPA	3-deoxy-D-arabino-heptulosonate 7-phosphate synthetase
ENO	Enolase
FBA	Fructose-bisphosphate aldolase
FUM	Fumarate hydratase
G6PDH2r	Glucose 6-phosphate dehydrogenase
GALUi	UTP-glucose-1-phosphate uridylyltransferase
GHMT	Glycine hydroxymethyltransferase
GLCtpts	Glucose phosphotransferase transporter
GLUSy	Glutamate synthase
ICL	Isocitrate lyase
LDH_L	L-lactate dehydrogenase
MDH	Malate dehydrogenase
PC	Pyruvate carboxylase
PDH	Pyruvate dehydrogenase
PFL	Pyruvate formate lyase
PGCD	Phosphoglycerate dehydrogenase
PGI	Glucose-6-phosphate isomerase
PGK	Phosphoglycerate kinase
PPCK	Phosphoenolpyruvate carboxykinase
PRPPS	Phosphoribosylpyrophosphate synthetase
PSCVT	3-phosphoshikimate 1-carboxyvinyltransferase
SUCDi	Succinate dehydrogenase
SUCOAS	Succinyl-CoA synthetase
TALA	Transaldolase
TKT1	Transketolase
TKT2	Transketolase

In particular, the advent of cheaper DNA sequencing has given rise to genome-scale metabolic models (GEMs), *in silico* reconstructions of the metabolic reaction networks of a given organism, derived from its an- notated genome sequence (Durot et al., 2009). In addition to operating as a knowledge base of metabolic information for a particular organism, GEMs can be used via constraint-based flux balance analysis to simulate carbon flux through metabolic reaction networks, enabling the rapid screening of metabolic behaviours under a range of environmental variables and biological contexts (Orth et al., 2010). Through comprehensive *in silico* predictions of metabolic phenotypes under target conditions, GEMs can also identify potential cellular redox imbalances (Zhang et al., 2016) and metabolic bottlenecks and generate hypotheses for rational, targeted genetic modifications for improved performance (Choi et al., 2019). GEMs can even guide the construction and optimization of carbon flux for either endogenous or novel heterologous microbial strain pathways towards high yields of desired products (Pharkya et al., 2004; Jensen et al., 2019).

Parageobacillus thermoglucosidasius NCIMB 11955 represents a Gram-positive, facultative anaerobic, thermophilic bacterial chassis with several advantageous traits for industrial bioprocesses when compared to many model bacterial chassis such as *Escherichia coli* and *Bacillus subtilis* (Hussein et al., 2015; Wada and Suzuki, 2018). Firstly, the thermophilicity of *Parageobacillus* spp. enables fermentations between 48 and 70°C (Zeigler, 2014; Suzuki, 2018) at growth rates surpassing other thermophilic organisms (Dahal et al., 2016) and comparable to that of *E. coli* (Panikov et al., 2003). Compared to equivalent mesophilic fermentations, these process temperatures enable a reduction in both the cooling costs of large-scale exothermic fermentations, and a reduction in the risk of contamination from mesophilic microbes (Choi et al., 2019; Krüger et al., 2018). Furthermore, for industrial bioprocesses aiming for simultaneous saccharification and fermentation (SSF), the thermophilicity of *P. thermoglucosidasius* is complemented by a catabolic versa- tility. Through extracellular secretions of thermostable amylases (Hussein et al., 2015), xylanases (Bartosiak-Jentys et al., 2013; Huang et al., 2017; Bibra et al., 2018) and other hemicellulases (Liu et al., 2012; Balazs et al., 2013; Maayer et al., 2014), *Parageobacillus* spp. are able to metabolize a wide range of C_5 and C_6 sugar monomers. Notably, they are able to transport then metabolize complex hemicellulosic (Maayer et al., 2014) and cellulosic (Hussein et al., 2015) polysaccharides derived from hydrolysates of lignocellulosic biomass, potentially reducing the reliance on externally supplied hydrolases involved in lignocellulosic pre-treatment.

A number of synthetic biology tools applicable *to P. thermoglucosidasius* have been devised including: shuttle vectors for reliable transformation (Bartosiak-Jentys et al., 2013; Taylor et al., 2008), chromosomal integration strategies (Bacon et al., 2017) promoter and RBS libraries to enable tuneable gene expression and validated reporter genes (Kananavičiūtė and Čitavičius, 2015; Reeve et al., 2016; Pog- rebnyakov et al., 2017; Drejer et al., 2018). Such tools have enabled *P. thermoglucosidasius*, and genetically similar *(Para)geobacillus* spp., to be used in the production of fuels such as bioethanol (Cripps et al., 2009; Niu et al., 2015), isobutanol (Lin et al., 2014) and hydrogen gas (Mohr et al., 2018; Aliyu et al., 2020) and also in fine chemicals including 2–3 butanediol (Kulyashov et al., 2020; Zhou et al., 2020), riboflavin (Yang et al., 2020) and isoprenoids (Styles et al., 2020). *Parageobacillus* spp. and *Geobacillus* spp. have also been the source of thermostable variants of industrially useful proteases (Zhu et al., 2007), carboxyl esterases (Ewis et al., 2004; Zhu et al., 2015; Chen et al., 2020), lipases (Zhu et al., 2015; Fotouh et al., 2016) along with a thermostable DNA polymerase I from *G. stearothermophilus* GIM1.543 (Ma et al., 2016).

In spite of these advances, (with the exception of natural end products of glycolytic metabolism, such as ethanol) none of these engineered pathways have approached their potential maximum yields. In general, they have relied on natural flux to their metabolic precursors and its inherent control. The availability of a reliable GEM would enable a systems metabolic engineering approach of *P. thermoglucosidasius* to address the optimization of flux through central metabolic pathways to balance the requirements of both production and growth. At present, only one publicly available GEM of a *P. thermoglucosidasius* exists, the related strain *P. thermoglucosidasius* C56-YS93 (denoted iGT736) (Ahmad et al., 2017). While comprising 1159 reactions and 1163 metabolites, analysis of iGT736 using the GEM assessment tool Memote developed by Lieven et al. (2020) suggests that it currently lacks some fundamental features, including a biomass equation, transport reactions and stoichiometric balance (Supplementary File 1), preventing mean- ingful application for quantitative analysis. Additionally, a few examples exist of smaller central carbon metabolism scale models derived from experimental ^{13}C isotopic tracer experiments. This includes models representing *P. thermoglucosidasius* M10EXG (Tang et al., 2009) under aerobic and anaerobic growth conditions, and similar *Geobacillus* spp. *G. icigianus* (Kulyashov et al., 2020) and *Geobacillus* LC300 (Cordova et al., 2015). However, they are less useful for illustrating the scale and complexity of whole cell metabolism.

The newly constructed genome-scale metabolic model *of P. thermo- glucosidasius* NCIMB 11955 presented herein (named hereafter as *p- thermo*) represents 917 genes and comprises of 890 metabolites and 1175 reactions across two compartments: cytosolic and extracellular space (representing the medium). After iterative cycles of manual curation, model refinement and analysis with Memote (Lieven et al., 2020), *p-thermo* exhibits a 100% stoichiometric consistency, 100% charge balance and a 99.9% mass balance. It accurately captures experimentally determined utilization of 22 carbon sources using the sole input of measured production and consumption rates (Lisowska, 2016) and is represented in the Systems Biology Markup Language (SBML) (Hucka et al., 2019) compliant format, making it compatible with commonly used constraint-based modeling software such as COBRApy (Ebrahim et al., 2013) and the COBRA Toolbox v3.0 (Heirendt et al., 2019) as well as more specialised software facilitating systems metabolic

engineering (Cardoso et al., 2018; Rocha et al., 2010). Validation of the predictive quality of *p-thermo* under aerobic, oxygen limited and anaerobic conditions is demonstrated through mapping the resulting *in silico* fluxes to experimentally determined ^{13}C-flux data ob- tained from ^{13}C-isotopic labelling experiments of the genetically and metabolically similar *P. thermoglucosidasius* M10EXG strain (Tang et al., 2009). The predictive power of *p-thermo* is further demonstrated through recapitulation of a metabolically engineered homoethanolo- genic strain of *P. thermoglucosidasius* (Cripps et al., 2009). Lastly, *p-thermo* was used to investigate the fundamental requirements and metabolic bottlenecks of *P. thermoglucosidasius* during anaerobic growth. Currently, *p-thermo* represents the most complete, curated and experimentally validated genome-scale metabolic model for a *Parageobacillus* sp, and will be a foundational platform for guiding rational metabolic engineering strategies, -omic data integration, and strain optimization to further the potential of *P. thermoglucosidasius* NCIMB 11955 to operate as microbial chassis for sustainable bioprocesses.

Results

1.1. Model reconstruction

The presented genome-scale metabolic reconstruction of *P. thermoglucosidasius* NCIMB 11955 is based on genome sequencing by ERGO™ Integrated Genomics (Overbeek et al., 2003) and Sheng et al. (2016). Genome annotation was performed through the ERGO™ Integrated Genomics suite (Overbeek et al., 2003) and the RAST annotation server (Aziz et al., 2008), followed by gap filling with Pathway Booster (Liberal et al., 2015). The reconstruction was extensively manually curated using available literature and databases (KEGG, BRENDA, MetaCyc, MetaNetX and EC2PDB), according to benchmark approaches (Thiele and Palsson, 2010). Detailed manual curation and refinement can be followed in Lisowska (Hussein et al., 2015) and in the GitHub repository. Specific attention was given to ensure nucleotide, amino acid and co-factor metabolism was accurate based on the current available knowledge. Additionally, extra care was paid to the selection and stoichiometry of the biomass precursors and their respective biosynthetic pathways, with particular focus on a more accurate representation of fatty acid and lipid biosynthesis (Sánchez et al., 2019).

This metabolic model consists of 890 metabolites, involved in a total of 1175 reactions, encoded for by 917 genes, across two compartments: cytosolic and extracellular space (representing the medium). Manual curation was critical to ensure complete consistency of the model (Supplementary report 1). Central carbon metabolism of the model re- sembles that of previously reported *(Para)geobacillus* spp (Tang et al., 2009; Cordova and Antoniewicz, 2015) (Fig. 1A). Of all reactions, 9.3% are involved in transport or exchange, highlighting the flexibility of the strain to grow on various carbon sources (Fig. 1B). Predominantly manual gap filling, based on available literature, was used to ensure correct active metabolic pathways, accounting for 20.6% of the final *p-thermo* reactions. After gap filling, 17 dead end and 22 orphan metabolites remained, resulting in 148 universally blocked reactions (12.6% of total reactions). All dead end and orphan metabolites were manually reviewed, but a lack of conclusive evidence surrounding them with respect to *Parageobacillus* sp. prevented these gaps from being filled with sufficient accuracy and so have been deliberately unmodified until further knowledge is accrued.

The model as well as scripts used in the reconstruction and manual curation are made publicly available through Github, at https://github. com/biosustain/p-thermo/releases/v1.0. The model is stored using the community standard SMBL format (Level 3, FBC Version 2) (Olivier and Bergmann, 2015) and can additionally be accessed as Supplementary File 2.

1.2. Biomass composition and growth energetics

To capture biological growth in stoichiometric models, a demand reaction referred to as a biomass pseudo-reaction, was added. An overview of how the biomass pseudo-reaction was defined is explained in Materials & Methods, with the final reaction components and associated stoichiometry given in Supplementary Table 1**.** Energetic parameters were fitted from aerobically grown chemostat experiments (Lisowska, 2016). The energy required to maintain cellular homeostasis is reflected in the non-growth associated maintenance (NGAM) and was found to be 3.141 $mmol_{ATP}/g_{DW}h^{-1}$ in *p-thermo.* The growth associated maintenance, (GAM), was estimated as 152.3 $mmol_{ATP}/g_{DW}$ and reflects the energy needed for cell replication, including macromolecule synthesis. The contribution of polymerization energy, required for macro-molecule synthesis, to the obtained GAM was estimated to be approximately 20% (Supplementary Table 2); relatively low compared to previously reported mesophiles (30–40%) (Verduyn et al., 1991; Förster et al., 2003; Monk et al., 2017; Neidhardt et al., 1996). It was previously observed that thermophilic organisms tend to require higher levels of energy for growth and homeostasis at elevated temperatures and thus have a reduced growth efficiency, shown in the high mainte- nance estimated (Dahal et al., 2016; Robb et al., 2007). This trait of thermophiles makes them valuable hosts for bioproduction as it leads to higher production rates of catabolic products compared to other organisms.

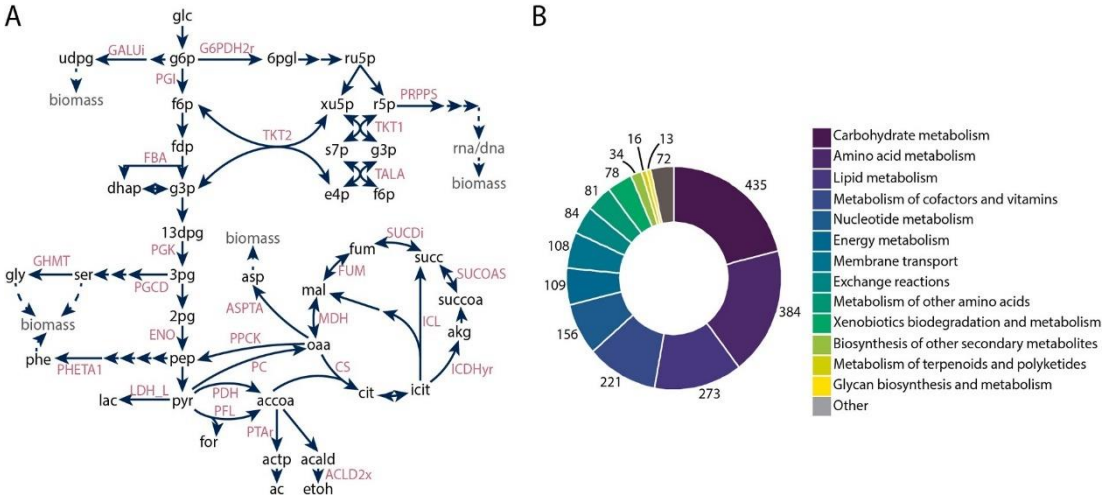


Fig. 1. A) Central carbon metabolism map, with several reaction IDs highlighted. For a more detailed overview, see Supplementary File 3. B) The number of reactions in the model for several reaction class types.

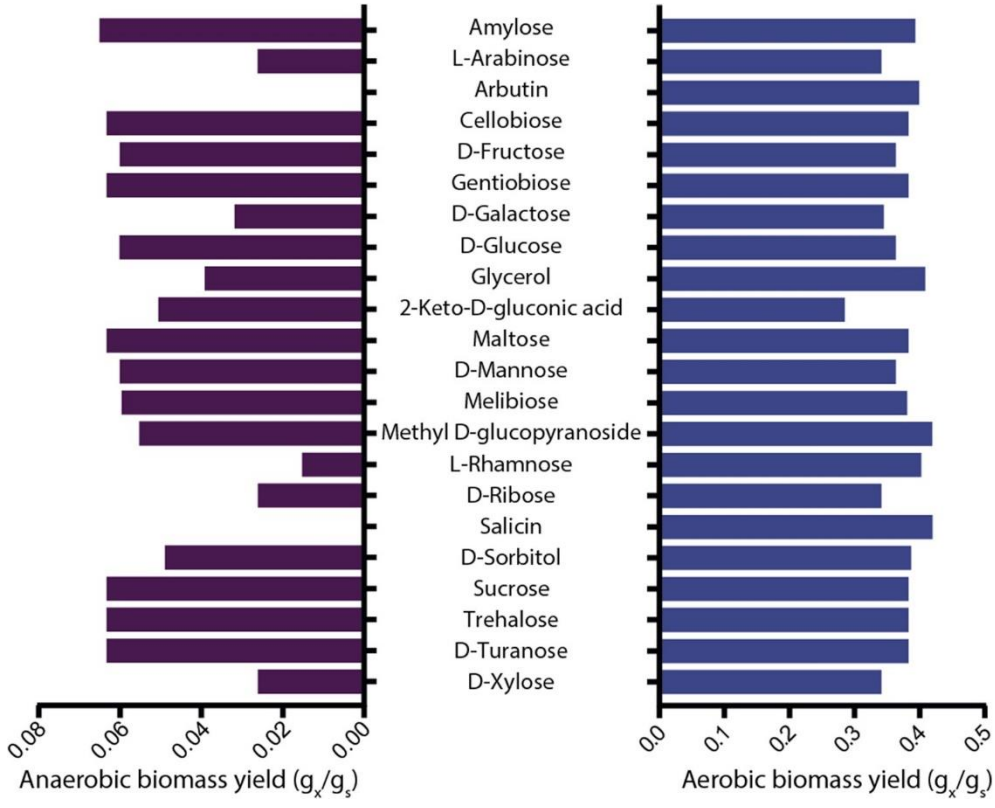


Fig. 2. Anaerobic (left) and aerobic (right) predicted biomass yields for 22 different carbon sources, for which aerobic growth has been experimentally confirmed (Lisowska, 2016). Carbon substrates were all supplied in the model at 30°C mol/gDWh to account for differences in composition between the carbon sources.

1.3. Overview of Metabolism

To provide a comprehensive overview, two pathway maps of the model were drawn using Escher (King et al. Escher, 2015) corresponding to central carbon and amino acid metabolism (Supplementary Files 3 and 4) and deposited in the GitHub repository at p-thermo/maps. Traits specific to *Geobacillus* spp. and *P.*

thermoglucosidasius NCIMB 11955, presented in the literature, were used to validate the model's metabolism. Detailed step-by-step decisions that were made, can be followed in the GitHub repository at "p-thermo/notebooks". As an example, in central carbon metabolism research has shown that *Geobacillius* spp., unlike many mesophilic *Bacillus* species, lack genes for a 6-phosphoglu- conolactonase (6PGL), responsible for part of the oxidative pentose phosphate pathway (PPP) (Hussein et al., 2015). Instead, the reaction can occur spontaneously and, at thermophilic temperatures, has been hypothesized to be sufficiently rapid to maintain the requisite PPP flux (Miclet et al., 2001). The absence of 6PGL was captured in the model, but to reflect the active PPP pathway, a pseudo-reaction was added to allow the complete oxidative PPP to function. *(Para)geobacillus* species are known to be capable of growth on a wide range of carbohydrates, and have been shown to secrete various polysaccharide degrading enzymes such as xylanases and other hemi- cellulose degrading enzymes (Hussein et al., 2015; Liu et al., 2012; Balazs et al., 2013; Maayer et al., 2014). To assess the metabolic capacity of the model, growth on various carbon sources was simulated (Fig. 2). The choice of carbon sources was made based on previous qualitative growth experiments which demonstrate a range of sole carbon sources which allow aerobic growth of *P. thermoglucosidasius* NCIMB 11955 (Lisowska, 2016). Additionally, anaerobic growth on these substrates was computationally predicted. In both cases, carbon supply was normalized to 30 $Cmol_s/g_{DW}h$, to accommodate different polymeric substrate forms being present in the data set. Initially, the model showed no aerobic growth on arbutin, salicin and rhamnose, due to dead-end metabolites being formed as side products in the first steps of their break down. Available literature was used to fill the gaps in the catabolic pathways, which enabled aerobic growth on all three carbon sources. Anaerobically, *in silico* growth on arbutin and salicin was unfeasible, as current knowledge suggests that their catabolism is oxygen dependent. Both arbutin and salicin are non-conventional carbon sources and are glycosides, composed of either a hydroquinone or salicyl alcohol func- tional group attached to glucose, respectively. It is known that meta- bolism of these glycosides occurs through splitting of the glycosidic bond, with the two functional groups being catabolized individually. With currently available knowledge, the further breakdown of the salicyl alcohol and hydroquinone functional groups is dependent on oxygen, deeming the *in-silico* prediction of anaerobic growth unfeasible. As there is little knowledge about microbial catabolism of these carbon substrates, this hypothesis would warrant experimental validation.

1.4. Assessment of predictive power through ^{13}C-flux fitting

Prior to using a genome-scale model for metabolic analyses or *ab initio* predictions, it is critical to validate its predictive power based on previously attained experimental data. This was done by analysis of how well the simulated fluxes of *p-thermo* match published flux distributions of *P. thermoglucosidasius*. ^{13}C-isotopic labelling is a standard tool used to elucidate intracellular fluxes in central carbon metabolism, through extensive experimental work and data analysis. Flux variability analysis (FVA) is an *in silico* approach that can allow *ab initio* analysis of metabolism without the need for laborious experimental data (Mahadevan and Schilling, 2003). Comparing the two data types can give insights into metabolism and allow the generation of hypotheses for metabolic engineering purposes.

To make this comparison, ^{13}C-flux data from *P. thermoglucosidasius* M10EXG subject to varying oxygen conditions was used to qualitatively assess the predictive quality of *p-thermo* (Tang et al., 2009). Whole proteome analysis (on a sequence basis) of the *P. thermoglucosidaius* M10EXG and NCIMB 11955 strains shows that the ORFs between the two strains are highly similar (Supplementary Fig. 1, Supplementary Table 3). Importantly, considering the metabolic genes that would be captured as reactions in a metabolic model, there are only 11 and 12 unique reactions in *P. thermoglucosidasius* NCIMB 11955 and *P. thermoglucosidasius* M10EXG respectively (Supplementary Tables 3, 4 and 5). Therefore, based on the overall metabolic similarity between the two strains, we assume that the ^{13}C-flux data from *P. thermoglucosidasius* M10EXG can be utilized for a qualitative assessment of *p-thermo*.

In order to test if *p-thermo* can predict intracellular fluxes close to the ^{13}C-flux data, the measured production and consumption rates were fixed in the model as exchange rates, and internal fluxes were predicted in a sensitivity analysis with FVA. Parsimonious enzyme usage flux balance analysis (pFBA), which has previously been shown to predict fluxes that correlate with experimental measurements, was also performed (Lewis et al., 2010). The *in-silico* fluxes and pFBA results were mapped to experimentally determined fluxes in aerobic, oxygen limited and anaerobic conditions (Fig. 3A, B and C respectively). pFBA showed good correlation to the measured data for each condition (Supplementary Fig. 2), and together with FVA showed accurate predictions of the internal central carbon fluxes (Fig. 3A, B, and C). Biomass yields (Fig. 3D) and oxygen consumption rates (Supplementary Fig. 3) were adequately predicted as well. These analyses validate the predictive quality of the created model and highlight the power of using metabolic models for understanding intracellular fluxes when only extracellular consumption or production rates are available.

1.5. Recapitulating & interpreting knockout physiology

To evaluate the utility of *p-thermo* for metabolic engineering applications, we recreated previously reported homoethanologenic mutants of *P. thermoglucosidasius* NCIMB 11955 *in silico.* Cripps et al. (2009) engineered

lactate dehydrogenase (*ldh*) and pyruvate formate lyase (*pfl*) knockouts *in vivo*, and supplemented their ethanol yields with an upre- gulation of pyruvate dehydrogenase expression (PDH_{up}). Using *p-thermo*, the wild type (WT), Δ*ldh* and Δ*ldh*Δ*pfl*(PDH_{up}) strains were recreated, as stoichiometric modeling cannot distinguish between upregulated expression levels (i.e. between Δ*ldh*Δ*pfl* and Δ*ldh*Δ*pfl* PDH_{up}). Exchange rates of the main fermentation metabolites were predicted using *p-thermo* and their accuracy evaluated based on measured data (Fig. 4). In performing the analysis, two distinct thresh- olds for Flux Variability Analysis (FVA) were selected, 95% and 99% of optimum biomass production (Mahadevan and Schilling, 2003), to assess the flexibility of exchange rates to the simulated conditions.

The performed simulations show a substantial discrepancy between predicted and measured yields in the WT and Δ*ldh* strains, whereas simulations tightly match the measured yields in the Δ*ldh*Δ*pfl* PDH_{up} strain. Still, the mismatch between the experimental and *in silico* data, *p-thermo* can be used to understand metabolic branch points. The main discrepancy observed lies in the lactate and formate yields for the WT and Δ*ldh* strains, which can be traced to the cellular decision of what to do with the synthesized pyruvate (Fig. 4). In this regard, there are three options: 1) conversion into lactate by lactate dehydrogenase (LDH), 2) anaerobic conversion into acetyl-CoA by pyruvate formate lyase (PFL) or 3) aerobic conversion into acetyl-CoA by pyruvate dehydrogenase (PDH). In both the WT and Δ*ldh* strain, *p-thermo* showed flux from pyruvate to acetyl-CoA to be exclusively carried through PFL, fitting with experimental expectations under anaerobic conditions due to high [NADH] (Kim et al., 2008). Additionally, the conversion of pyruvate into acetate and ethanol results in one additional ATP per glucose, compared to converting pyruvate into lactate (Wang et al., 2010). Therefore, from a stoichiometric perspective, *p-thermo* predicts this to be the most optimal pathway for growth, explaining the high concentrations of formate, ethanol and acetate predicted in the simulation.

However, this was not observed in the experimental yields, presumably because of subtle differences in dissolved oxygen availability in the experimental setup that influence multiple levels of regulation *in vivo*, intrinsically not accurately captured by stoichiometric models. In the experimental dataset, undefined oxygen limited conditions were used in which a gradual decline in available dissolved oxygen concentration would have occurred during growth, whereas simulations were performed under anaerobic conditions. Under oxygen-limited conditions, PDH is expressed in the wild type *P. thermoglucosidasius* (Cripps et al., 2009), where PFL is typically only active under completely anaerobic conditions (Sawers and Bock, 1988). The transition of physiological states in response to decreasing oxygen availability results in excess NADH and creates a redox imbalance in the cell which is alleviated through production of lactate as the production of formate by PFL is restricted. This could explain the discrepancy between the experimentally measured low formate and high lactate production in the WT strain and the prediction by *p-thermo*. In the LDH knockout at low dissolved oxygen conditions, which prevents PFL activity, PDH instead predominantly carries flux to acetyl-CoA. In this instance, in order to maintain cellular redox balance, the Δ*ldh* cells increase the produced ethanol/acetate ratio. This picture highlights the complexity of cellular and enzymatic regulation that is poorly captured in stoichiometric models, as well as the difficulty in simulating uncontrolled environments accurately.

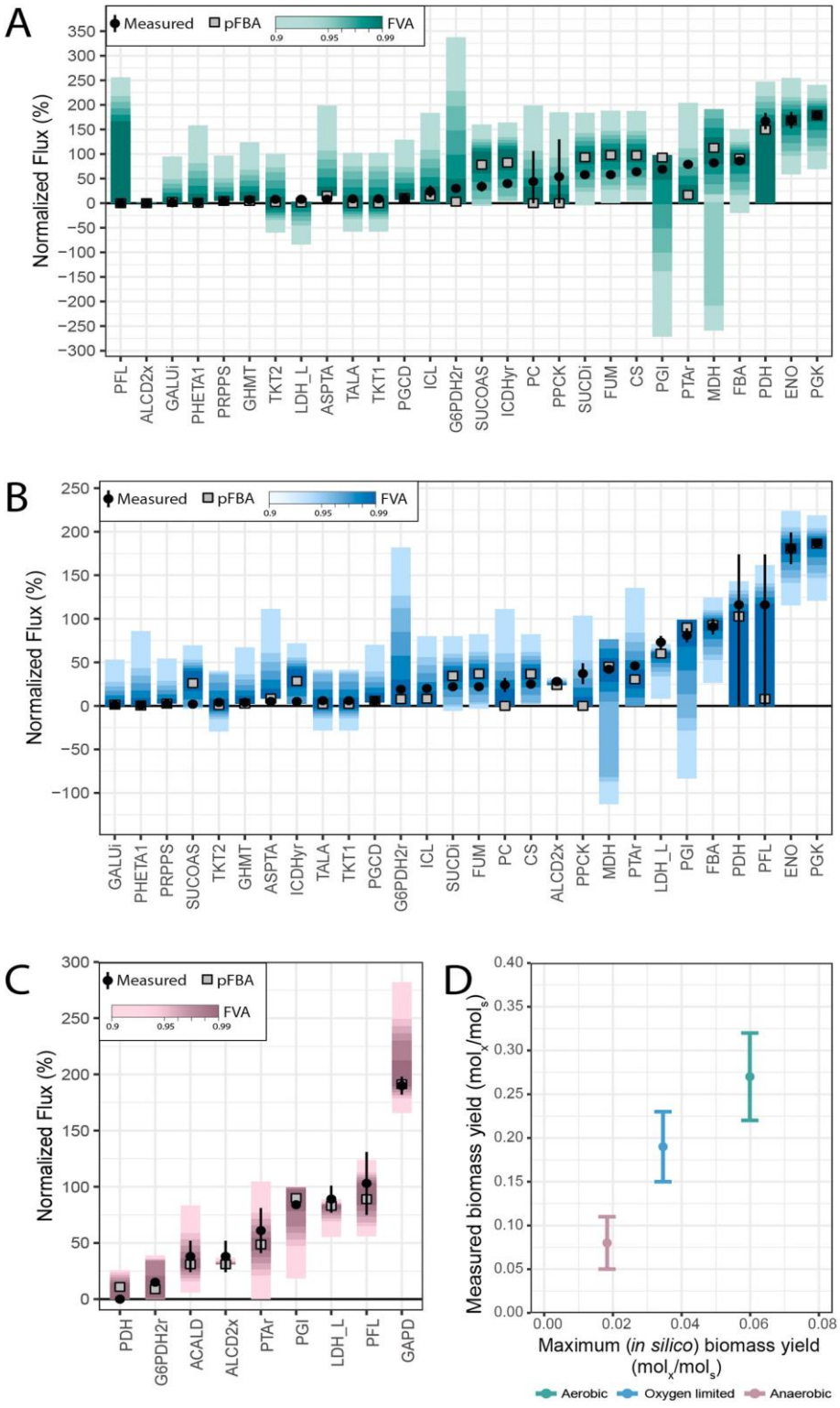


Fig. 3. Results of fixing experimentally measured exchange rates and predicting intracellular flux distributions (Tang et al., 2009) in aerobic (A), oxygen limited (B) and anaerobic (C) conditions, normalized to the glucose uptake rate. Fig. 1A shows the stoichiometry of all the reactions shown on the x-axis. FVA sensitivity analysis is shown in line ranges. Predicted and measured maximum biomass yields, for a FVA threshold set at 99% of optimum biomass, are shown in (D).

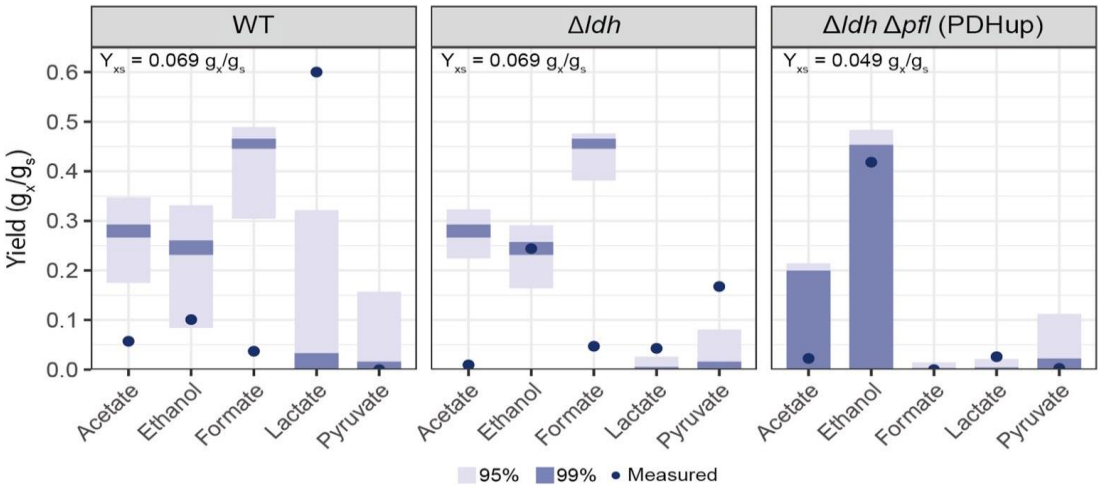


Fig. 4. Comparison of *in silico* predictions of fermentation product yields in three engi- neered strains with experimentally deter- mined data from Cripps et al. (Cripps et al., 2009), when solely the carbon uptake rate and knockouts were fixed in the model. Yield (g_x/g_s) is shown for predicted and measured exchange rates. Each panel highlights a different strain: wild type (WT), Δ*ldh* and Δ*ldh*Δ*pfl* (PDHup). Two varying thresholds for FVA were run: 95% and 99% of the optimum biomass production. The *in silico* predicted biomass yield (Y_{xs}) for 99% of the optimum biomass production is shown for each condition.

However, the performed simulations can still be used to visualize and understand the burden that lactate production can have on cellular growth. The inability to induce PFL at moderate levels of oxygen limi- tation, puts a larger reliance on fermentative metabolism to lactate, providing less energy. We used *p-thermo* to investigate the possible impact this has. First, all measured exchange rates for the three strains were fitted to the model and used in subsequent determination of pre- dicted biomass yields. This showed that the model is physiologically capable of capturing the measured data, albeit with a lower predicted biomass yield than was experimentally measured, suggesting that stoi- chiometrically sub-optimal fermentation pathways were active *in vivo* (Supplementary Fig. 4A). Finally, the effect of increasing lactate pro- duction on biomass yield was computed, showing the energetic loss that occurs from lactate production (Supplementary Fig. 4B). Overall, this highlights the importance of complex regulation in dictating *in vivo* metabolism, over pure stoichiometric optima per se.

1.1. *Genome-scale metabolic modeling allows the elucidation of metabolic bottlenecks*

The availability of a comprehensive GEM can also facilitate the elucidation of metabolic bottlenecks and identification and optimization of chemically defined growth media (Branco dos Santos et al., 2017). Thus, *p-thermo* was used to help resolve known issues of the anaerobic metabolic physiology of *P. thermoglucosidasius*. Although *P. thermoglucosidasius* is clearly capable of classical mixed acid fermentation and shows elements of a regulated aerobic-anaerobic switch as revealed by transcriptomic analysis (Loftie-Eaton et al., 2013) (although oxygen-scavenging state under fermentative conditions), it has long been known that growth under anaerobic conditions on existing mini- mal defined growth medias requires additional supplements in comparison to growth under aerobic conditions. Typically, this was resolved by supplementation with a small amount of oxygen or yeast extract (Hussein et al., 2015; Cripps et al., 2009; Lisowska, 2016). Therefore, here we used simulations of *p-thermo* to find a minimal set of defined nutrients that can achieve anaerobic growth of *P. thermoglucosidasius*. As a first observation, when fed true minimal, anaerobic medium, the model predicted no growth, in accordance with experimental observa- tions. However, fermentative energy generation was observed, which highlights that oxygen requirement comes from critical secondary me- tabolites or cofactors that cannot be synthesized anaerobically, which is corroborated by previous observations (Hussein et al., 2015). By mini- mizing the oxygen uptake in the model, a critical reaction set requiring oxygen was generated (Table 1). This analysis highlighted a complex combination of components that cannot be synthesized anaerobically: thiamine, biotin, folate, vitamin B12, spermine, spermidine and hemin. Additionally, iron(III) must be available in the medium to allow porphyrin biosynthesis.

I*n silico* supplementation of these components rescued anaerobic growth, providing a combination of candidates for experi- mental validation. *P. thermoglucosidasius* DSM 2542 was obtained from the DSM stock center (Leibniz Institute.-G, 2021). While DSM 2542 carries several mutations (11 SNPs and 2 indels) compared to the NCIMB 11955 sequence (Sheng et al., 2016), the only differences found in coding sequences are annotated as hypothetical proteins. Thus, it is expected that they would not impact the predictive outcome of the model between

±

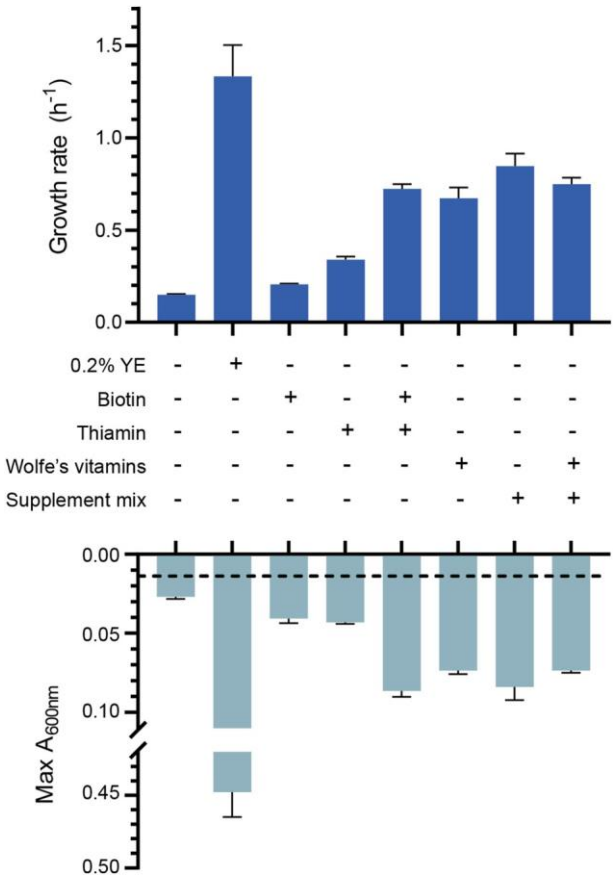


Fig. 5. Experimental growth rates calculated and maximum observed absorbance values when *P. thermoglucosidasius* NCIMB 11955 was grown anaerobi- cally in a microtiter plate reader in TMM base medium, supplemented with various nutrients, as indicated. Dashed line indicates inoculation absorbance when an inoculation optical density of 0.05 was used.

the two strains. The simulated essential components were experimentally added together in trace amounts to form a supplementation mix (see Materials and Methods); to assess if it would allow anaerobic growth. This was compared to Wolfe's vitamin solution, a commonly used mix of vitamins in base thermophilic minimal medium (TMM) (Pogrebnyakov et al., 2017; Fong et al., 2006). It should be noted that Wolfe's vitamin solution contains thiamin, biotin, folate and vitamin B12, amongst other nutrients, and that TMM contains trace amounts of iron (III). To uncover the minimal sets of components needed to rescue anaerobic growth, eight different conditions were tested, all composed of base TMM with 10 g/L glucose: 1) no added nutrients, 2) 0.2% yeast extract, 3) biotin, 4) thiamin, 5) biotin and thiamin, 6) Wolfe's vitamins, 7) Supplementation mix and 8) Wolfe's vitamins plus the unique components of the supplementation mix (spermine, spermidine and heme) (Fig. 5, Supplementary Fig. 5).

Experimental observations suggested that a combination of thiamin, biotin and iron(III) were the minimal required supplementation set needed to sustain anaerobic growth, as no difference was observed when additional defined supplementation was added (Fig. 5). Yeast extract also contains significant amounts of amino acids and other components and so provides an additional growth advantage, as expected. However, this highlights a discrepancy with the model predictions, as a larger minimal supplementation set was originally predicted (Table 1). Finally, as expected, base TMM can support aerobic growth, at a maximum rate of 0.267 0.021 h^{-1} (Supplementary Fig. 5A), confirming the synthesis of the critical components in the presence of oxygen.

There are several reasons that can explain the differences between the *in silico* and experimentally determined minimal supplementation set. The incomplete understanding of thermophilic life introduces additional levels of complexity that are typically not captured by auto- matic annotation pipelines dependent on predominantly mesophilic datasets (Aziz et al., 2008; Seemann, 2014) leading to errors in the annotation of thermophilic traits (Mendoza et al., 2019). For example, genomes of thermophilic organisms show a correlation with higher G/C content, less intergenic regions and a higher

Table 1

Overview of critical reactions that require oxygen to allow growth in p-thermo.

Reaction ID	EC-code	Reaction	*In silico* requirement
CAT	1.11.1.6 & 1.11.1.21	Catalase-peroxidase	Glyoxylate and dicarboxylate metabolism
PMES	1.14.14.46	Pimeloyl-[acyl-carrier	Biotin biosynthesis
BLUB	1.13.11.79	protein] synthase 5,6-dimethylbenzimidazole synthase	Vitamin B12 biosynthesis
GLYCTO1	1.1.3.15	(S)-2-hydroxy-acid oxidase	Folate biosynthesis
PPPGO_1	1.3.3.4	Protoporphyrinogen	Porphyrin (heme)
FERO	1.16.3.1	oxidase	biosynthesis
		Ferroxidase	Porphyrin (Fe) metabolism
ACDO	1.13.11.54	Acireductone dioxygenase	Aliphatic polyamine biosynthesis
GLYHOR	1.4.3.19	Glycine oxidase	Thiamine biosynthesis

functional stability (re- flected by the lower ratio of non-synonymous to synonymous sub- stitutions over time) (Sabath et al., 2013; Wang et al., 2015; Lusk, 2019). Additionally, thermostable proteins can have significantly altered structure compared to their mesophilic counterparts performing the same reaction, confounding homology-based annotation (Huang et al., 2020). Through the observed discrepancies, we can unveil additional insights of anaerobic metabolism of *P. thermoglucosidasius.*

In the first place, the *in-silico* dependence on vitamin B12 highlights the inaccuracy of annotation pipelines. Vitamin B12 synthesis is classi- cally divided into two routes: canonical (aerobic) and non-canonical (anaerobic) (Fang et al., 2017). Although the genome annotation of *P. thermoglucosidasius* NCIMB 11955 reveals parts of either pathway, neither is complete. It has been proposed that possibly a novel, blended pathway may be present; however, this may arise from incorrect an- notations based on lacking knowledge of thermophilic vitamin B12 biosynthesis genes (Lisowska, 2016; Raux et al., 2000; Gajcy, 1973). The possibility to grow without vitamin B12 supplementation does highlight both an aerobic and anaerobic functional pathway in *P. thermoglucosidasius* NCIMB 11955. To further understand the *de novo* biosynthesis of vitamin B12, experimental validation would be required. In *p-thermo*, the *in silico* oxygen requirement for spermine and sper midine biosynthesis comes from the downstream recycling of a biosynthetic by-product: 5′-methylthioadenosine (5-MTA). 5-MTA recycling is also important in a novel, oxygen independent MTA- isoprenoid shunt, involved in the methionine salvage pathway (North et al., 2016). This pathway has been characterized in *Rhodospirillum rubrum* and orthology analysis highlights the possible presence of parts of this pathway in various facultative anaerobic *Bacillus* spp (North et al., 2017). This presents the possibility of an alternate 5-MTA recy- cling pathway, explaining the independence of anaerobic growth to spermine or spermidine addition.

Similarly, to spermine and spermidine, the oxygen requirement in the *in-silico* folate biosynthesis pathway stems from the formation of gly- colaldehyde as a side product, which is further oxidized to glyoxylate. The *in silico* oxidase responsible for glyoxylate formation requires oxy- gen (EC 1.1.3.15). However, reports show that *Moorella thermoacetica*, a thermophilic obligate anaerobe, can grow on glycolate through the formation of glyoxylate, highlighting the possibility for a (to date) un- known alternate electron acceptor (Seifritz et al., 1999; Sakai et al., 2008).

Finally, heme is suggested to be synthesized in *P. thermoglucosidasius* from glycine using a 5-aminolevulinic acid synthase (Lisowska, 2016) and notably using an oxygen-dependent protoporphyrinogen oxidase. Both *E. coli* and *B. subtilis* have an oxygen independent coproporphyrinogen-III oxidase (hemN), known to be responsible for anaerobic heme biosynthesis, using other electron acceptors such as fumarate, or nitrate over oxygen (Jacobs and Jacobs, 1976; Hippler et al., 1997; Layer et al., 2002; Möbius et al., 2010). While the current genome annotation of *P. thermoglucosidasius* NCIMB 11955 (Sheng et al., 2016) suggests that only the oxygen dependent path is present, the data presented herein suggest that supplementation with hemin is not required for growth (Fig. 5). One possible explanation for this discrep- ancy can be found when performing a tBLASTn with the *B. subtilis* hemN (NCBI accession CAB61616) against the *P. thermoglucosidasius* NCIMB 11955 genome. This highlighted a significant hit (CP016622 region 3448674, 3449762, 52% identity, E-value: 10^{-115}). This suggests the possibility that some form of this oxygen independent heme biosynthesis route could also be present highlighting the need to better understand the metabolism of non-model organism chassis.

This identification of the minimal, defined anaerobic medium high- lights how GEMs can be used to facilitate experimental hypotheses, where previous hypotheses have failed. The result, a defined minimalanaerobic medium is valuable for further investigation into anaerobic metabolism through ^{13}C-characterization studies, where defined media are critical. Additionally, this identification of a series of components which support anaerobic growth of *P. thermoglucosidasius* at a minimal medium level can further help inform the development of industrial growth media for other microbial chassis used in anaerobic bioprocesses improving growth and chemical product yields.

Discussion

Parageobacillus spp. represent valuable microbial chassis for meta- bolic engineering and fermentative bioproduction. Many advantages derive from their thermophilic character, with additional advantages coming from species specific traits. However, to further develop *Parageobacillus* spp. into fully optimized microbial cell factories, additional in-depth and systems level understanding of metabolism is required, for which omic analyses and genome scale metabolic models are critical. Currently, various automatic pipelines exist for generating metabolic models on the sole basis of a genome sequence. Yet, for thermophilic organisms, significant faults resulting from automatic annotation pipe- lines are evident, as these are based on predominantly mesophilic datasets. Thermophilic genomes show different characteristics, the ef- fect of which on metabolism is still poorly understood, making trans- lation into a predicted function difficult. Thus, significant manual curation is needed in the generation of GEMs for thermophilic organisms, which is limited by the availability of knowledge on thermophilic metabolism. This is reflected by the relatively high need for manual gap-filling, and a high resulting percentage of blocked reactions in *p-thermo*. To increase the understanding of genotype-phenotype relationships in thermophilic hosts, the availability of a GEM acts as a considerable step facilitating systems level studies. As a result, with more knowledge arising, iterative rounds of model improvement are possible.

Therefore, in this work, we developed *p-thermo*, to date the most complete, curated and validated genome-scale metabolic model for a facultative anaerobic *Parageobacillus* sp. In it, genomic and biochemical knowledge were combined into a single powerful knowledge base, providing a critical tool for data-driven metabolic engineering, -omic data integration, process design and optimization. The model accurately captured substrate usage *in silico*, showing the metabolic flexibility of the strain for production with alternative carbon sources (Lisowska, 2016). Furthermore, ^{13}C-isotopic data verified the quality of *p-thermo* for predicting central internal fluxes of the model, when solely production and consumption rates are measured, a common practice when evalu- ating metabolic engineering designs.

Going beyond validation, *p-thermo* was used to provide more in- depth analysis of previously reported metabolic engineering ap- proaches (Cripps et al., 2009). Initial *p-thermo* simulations did not completely match experimental data, as stoichiometric models are incapable of capturing complex levels of regulation that play a dominant role in *in vivo* metabolism. Nonetheless, *p-thermo* was used to investigate the pyruvate branch point in central metabolism and allowed additional insights into metabolic flux distributions in various genetic back- grounds. With *p-thermo* further insights into metabolism can be gained, allowing improved targeted metabolic engineering in subsequent designs.

Finally, we used *p-thermo* to generate hypothesis driven experiments to alleviate a bottleneck in anaerobic metabolism, where previous experimental design was unsuccessful (Hussein et al., 2015). Doing so revealed fundamental insights into the metabolism of *P. thermoglucosidasius*, and also demonstrated that significant knowledge gaps still exist. This analysis, in combination with the ^{13}C-based verifi- cation, highlights an additional obstacle in working with thermophilic GEMs, where annotation pipelines are less precise: information on central carbon metabolism can be inferred with relative accuracy, whereas peripheral metabolic pathways are not significantly understood and require further systems-level investigation. Overall, *p-thermo*, together with other systems level and omics-based approaches, act as a tool to improve our understanding of genotype- phenotype relationships. Genome-scale metabolic models are in this way a critical part of an iterative cycle and are essential to the use and efficacy of thermophilic hosts for metabolic engineering and industrial bioproduction.

Materials & Methods

1.2. Model construction & curation

Genome sequencing of *Parageobacillus thermoglucosidasius* NCIMB 11955 was initially performed by ERGOTM Integrated Genomics (Overbeek et al., 2003) (funded by TMO Renewables Ltd) and subse- quently updated using the published *P. thermoglucosidasius* NCIMB 11955 genome sequence (Sheng et al., 2016) (NCBI accession CP016622 [chromosome], CP016623[pNCI001], and CP016624[pNCI002]).
Genome annotation was performed through the ERGOTM Integrated Genomics suite (Overbeek et al., 2003) and the RAST server (Brettin et al., 2015). Pathway Booster was used for gap filling, resolving gaps through comparisons with evolutionarily-related genomes (Liberal et al., 2015). Upon base construction of the model, further manual curation was done following standard procedures (Thiele and Palsson, 2010). To do so, missing information was primarily obtained from literature and using various databases: BRENDA, EC2PDB, KEGG, MetaNetX or MetaCyc (Lisowska, 2016). Whenever information on *P. thermoglucosidasius* was lacking, available references from other (*Para)geobacillus* spp or *Bacillus* spp were added. All further manual curation and refinement can be found in the GitHub repository. Model improvement was ensured by running Memote (Lieven et al., 2020) after each modification.

1.3. Biomass composition and growth energetics

To model growth, a biomass pseudo-reaction was added to the model. The reaction pools metabolites needed for growth into a biomass metabolite. Base biomass composition was previously determined experimentally according to reported practices (Durot et al., 2009; Lisowska, 2016). Lipid composition was obtained from previous reports (Tang et al., 2009), and was incorporated into the model according to a restrictive approach (Sánchez et al., 2019), in which a determined acyl chain length is assumed for all lipid species. Further fine-tuning of biomass composition was performed based on available enzymatic and metabolic requirements of the strain, with case-by-case justification given in the GitHub repository. Critical metabolites known to be required for catabolic functions of essential enzymes, such as heme, were added at trace stoichiometries based on knowledge from related organisms and scaled to ensure that all biomass components added up to 1 g/g_{DW} (Chan et al., 2017). Growth energetics (ATP cost of growth-associated maintenance and ATP requirement for non-growth associated maintenance) were estimated by minimizing the prediction error of the specific substrate consumption rate and the specific growth rate of glucose fed, aerobic chemostats (Lisowska, 2016; Thiele and Palsson, 2010). The P/O ratios were obtained from the given data for *B. subtilis* (Oh et al., 2007). The contribution of polymerization of each metabolite type to the total growth associated maintenance was esti- mated based on previously reported polymerization energies (Verduyn et al., 1991).

1.4. Transport reactions

The model has two compartments: extracellular and intracellular. Transport reactions were inferred from genome annotations and omology to known transporters. Additionally, knowledge about growth on various substrates was used to validate the presence of the corre sponding transporters.

1.5. Stoichiometric modeling & applied constraints

In traditional flux balance analysis (Orth et al., 2010), reaction stoichiometries are converted into a stoichiometric matrix (**S**), with *m* x *n* dimensions, where *m* represents the various metabolites and *n* repre- sents the number of reactions. Coefficients in the matrix are either positive or negative, reflecting production and consumption, respec- tively. Stoichiometric modeling works under the assumption of a pseudo-steady-state, represented as:

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

Where the vector **v** contains the fluxes of all reactions, given in units of $mmol/g_{DW}h$. As there are more metabolites than reactions, to solve this underdetermined system, linear programming is used by formulating an objective function (*z*), per default set as biomass accumulation. Reversibility of reactions is set based on thermodynamic prediction and a default medium is defined; in the case of *p-thermo*, minimal medium, with D-glucose as default carbon source is used.
Quantification of metabolic fluxes was performed using flux vari- ability analysis (FVA) (Mahadevan and Schilling, 2003). When running FVA, a threshold below the optimum is used to represent the metabolic freedom that is given to a model. In this study, a sensitivity analysis was run with FVA thresholds from 90 to 99%, to evaluate at what sensitivity level the model better matches the experimental data. Additionally, parsimonious flux balance analysis (pFBA) was run, by conducting a bilevel linear programming optimization that computes the optimum (growth) solution of the network, whilst minimizing the sum of all fluxes (Lewis et al., 2010). In doing so, this optimization predicts the most stoichiometrically efficient pathway set, and captures the maximum biomass per unit flux objective that has previously been described to be well supported by proteomic and transcriptomic data (Lewis et al., 2010; Schuetz et al., 2007).

1.6. Genome comparison

For an unbiased genome comparison, the *P. thermoglucosidasius* NCIMB 11955 genome was obtained

from NCBI (Accession: CP016622), and the *P. thermoglucosidasius* M10EXG genome was obtained from the Integrated Microbial Genome database (ID 2501416905). Genome annotation was performed using RASTk (Brettin et al., 2015), after which the two proteomes were compared through blast bi-directional best hits to create a homology matrix between the strains, based on a published pipeline (Norsigian et al., 2020). To filter for metabolic genes, any ORF associated to a predicted EC code was considered metabolic. The exact workflow can be followed in the GitHub repository.

1.7. Experimental procedures

The *P. thermoglucosidasius* NCIMB 11955 (DSM2542) strain was ob- tained from DSMZ (Leibniz Institute.-G, 2021). The strain was grown in either 2SPY medium or base thermophile minimal medium (TMM), modified from Fong et al. (2006). 2SPY was used for a first preculture, and contains per liter, 16 g soy peptone, 10 g yeast extract and 5 g NaCl, adjusted to pH 6.8. Base TMM contains, per liter: 930 ml Six salts so- lution (SSS), 40 ml 1 M MOPS (pH 8.2), 10 ml 1 mM $FeSO_4$ in 0.4 M tricine, 10 ml 0.132 M K_2HPO_4, 10 ml 0.953 M NH_4Cl, 0.5 ml 1 M $CaCl_2$ and trace element solution, adjusted to a final pH of 6.8. SSS contains, per 930 ml: 4.6 g NaCl, 1.35 g Na_2SO_4, 0.23 g KCl, 0.037 g KBr, 1.72 g $MgCl_2{\cdot}6\ H_2O$ and 0.83 g $NaNO_3$. The trace element solution contained, per liter, 1 g $FeCl_3{\cdot}6\ H_2O$, 0.18 g $ZnSO_4{\cdot}7\ H_2O$, 0.12 g $CuCl_2\ {\cdot}2\ H_2O$, 0.12 g $MnSO_4{\cdot}H_2O$ and 0.18 g $CoCl_2{\cdot}6\ H_2O$. D-glucose to a final concentration of 10 g/L was added to the base TMM.

When indicated, the base TMM was supplemented with one of the following, to the indicated final concentrations: 0.2% (w/v) yeast extract, 2 µg/L biotin, 5 µg/L thiamine-HCl, 1x Wolfe's vitamins, or 1x supplementation mix. 1000x Wolfe's vitamins consist of, per liter, 10 mg pyridoxine hydrochloride, 5.0 mg thiamine-HCl, 5.0 mg riboflavin, 5.0 mg nicotinic acid, 5.0 mg calcium D-( )-pantothenate, 5.0 mg p-ami- nobenzoic acid, 5.0 mg thioctic acid, 2.0 mg biotin, 2.0 mg folic acid and 0.1 mg vitamin B12. The 1000x supplementation mix contained, per liter, 2.0 mg biotin, 5.0 mg thiamine-HCl, 2.0 mg folic acid, 0.1 mg vitamin B12, 127 µg/L spermidine, 174 µg/L spermine tetrahydro- chloride and 0.7 mg hemin.

The aerobic cultures were inoculated to a starting OD_{600} of around 0.05, after an overnight culture on the base TMM medium. Growth was monitored through OD600 measurements, during growth at 60°, 200 RPM in baffled shake flasks. Anaerobic medium was prepared similarly to aerobic medium, but 1 µg/L resazurin was added to ensure complete anaerobic conditions. The medium was flushed with nitrogen gas prior to use. All anaerobic work was performed in an anaerobic chamber. Overnight cultures were run in anaerobic serum flasks at 60 °C, 200 rpm and used to inoculate a microtiter plate to a final OD of 0.05 in 200 µL volume. After sealing, the OD_{600} was measured every 15 min for 10 h in a Biotek Epoch2 microplate spectrophotometer, placed in an anaerobic chamber (run at 60 °C, with linear shaking).

Availability of data and materials

The metabolic model, scripts and corresponding datasets generated during the study are all freely available, under an Apache 2.0 license, at the GitHub repository: https://github.com/biosustain/p-thermo/rele ases/v1.0.

Acknowledgements

We thank Dr. Shyam Maskapalli for help with biomass composition analysis.

+

References

Ahmad, A., et al., 2017. A Genome Scale Model of Geobacillus thermoglucosidasius (C56- YS93) reveals its biotechnological potential on rice straw hydrolysate. J. Biotechnol.251, 30–37.

Aliyu, H., Mohr, T., Cowan, D., Maayer, P. De, 2020. Time-course transcriptome of Parageobacillus thermoglucosidasius DSM 6285 grown in the presence of carbon monoxide and air. Int. J. Mol. Sci. 21.

Aziz, R.K., et al., 2008. The RAST Server: rapid annotations using subsystems technology. BMC Genom. 9, 1–15.

Bacon, L.F., Hamley-Bennett, C., Danson, M.J., Leak, D.J., 2017. Development of an efficient technique for gene deletion and allelic exchange in Geobacillus spp. Microb. Cell Factories 16, 1–8.

Balazs, Y.S., et al., 2013. Identifying critical unrecognized sugar – protein interactions in GH10 xylanases from Geobacillus stearothermophilus using STD NMR. FEBS J. 280, 4652–4665.

Bartosiak-Jentys, J., Hussein, A.H., Lewis, C.J., Leak, D.J., 2013. Modular system for assessment of glycosyl hydrolase secretion in Geobacillus thermoglucosidasius. Microbiology 159, 1267–1275.

Bibra, M., Kunreddy, V.R., Sani, R.K., 2018. Thermostable xylanase production by geobacillus sp . Strain DUSELR13 , and its application in ethanol production with lignocellulosic biomass. Microorganisms 93, 1–25.

Branco dos Santos, F., et al., 2017. Probing the genome-scale metabolic landscape of Bordetella pertussis, the causative agent of whooping cough. Appl. Environ. Microbiol. 83, 1–19.

Brettin, T., et al., 2015. RASTtk: a modular and extensible implementation of the RAST algorithm for building custom annotation pipelines and annotating batches of genomes. Sci. Rep. 5.

Cardoso, J.G.R., et al., 2018. Cameo: a Python library for computer aided metabolic engineering and optimization of cell factories. ACS Synth. Biol. 7, 1163–1166.

Chan, S.H.J., Cai, J., Wang, L., Simons-Senftle, M.N., Maranas, C.D., 2017. Standardizing biomass reactions and ensuring complete mass balance in genome-scale metabolic models. Syst. Biol. (Stevenage) 33, 3603–3609.

Chen, P., Liu, C., Chen, Y., Hsu, F., Shaw, 2020. J. Isolation, expression and characterization of the thermophilic recombinant esterase from Geobacillus thermodenitrificans PS01. Appl. Biochem. Biotechnol. 191, 112–124.

Choi, K.R., et al., 2019. Systems metabolic engineering strategies: integrating systems and synthetic biology with metabolic engineering. Trends Biotechnol. 37, 817–837.

Cordova, L.T., Antoniewicz, M.R., 2015. 13C metabolic flux analysis of the extremely thermophilic, fast growing, xylose-utilizing Geobacillus strain LC300. Metab. Eng. 33, 148–157.

Cordova, L.T., Long, C.P., Venkataramanan, K.P., Antoniewicz, M.R., 2015. Complete genome sequence, metabolic model construction and phenotypic characterization of Geobacillus LC300, an extremely thermophilic, fast growing, xylose-utilizing bacterium. Metab. Eng. 32, 74–81.

Cripps, R.E., et al., 2009. Metabolic engineering of Geobacillus thermoglucosidasius for high yield ethanol production. Metab. Eng. 11, 398–408.

Dahal, S., Poudel, S., Thompson, R.A., 2016. Genome-scale modeling of thermophilic microorganisms. Adv. Biochem. Eng. Biotechnol. https://doi.org/10.1007/10.

Drejer, E.B., Hakvåg, S., Irla, M., Brautaset, T., 2018. Genetic tools and techniques for recombinant expression in thermophilic bacillaceae. Microorganisms 6, 1–19.

Durot, M., Bourguignon, P., Schachter, V., 2009. Genome-scale models of bacterial metabolism: reconstruction and applications. FEMS Microbiol. Rev. 33, 164–190. Ebrahim, A., Lerman, J.A., Palsson, B.O., Hyduke, D.R., 2013. COBRApy: COnstraints-based reconstruction and analysis for Python. BMC Syst. Biol. 7, 1–6.

Ewis, H.E., Abdelal, A.T., Lu, C., 2004. Molecular cloning and characterization of two thermostable carboxyl esterases from Geobacillus stearothermophilus. Gene 329,187–195.

Fang, H., Kang, J., Zhang, D., 2017. Microbial production of vitamin B 12: a review and future perspectives. Microb. Cell Factories 16, 1–14.

Fong, J.C.N., et al., 2006. Isolation and characterization of two novel ethanol-tolerant facultative-anaerobic thermophilic bacteria strains from waste compost. Extremophiles 10, 363–372.

Förster, J., Famili, I., Fu, P., Palsson, B.Ø., Nielsen, J., 2003. Genome-scale reconstruction of the Saccharomyces cerevisiae metabolic network. Genome Res. 13, 244–253.

Fotouh, D.M.A., Bayoumi, R.A., Hassan, M.A., 2016. Production of thermoalkaliphilic lipase from Geobacillus thermoleovorans DA2 and application in leather industry. Enzym. Res. 1–9.

Gajcy, H., 1973. Biosynthesis of Vitamin B12 by mesophilic and thermophilic strains of Bacillus megaterium. Pol. J. Microbiol. 5, 119–127.

Heirendt, L., et al., 2019. Creation and analysis of biochemical constraint-based models: the COBRA Toolbox v3.0. Nat. Protoc. 14, 639–702.

Hippler, B., et al., 1997. Characterization of Bacillus subtilis hemN. J. Bacteriol. 179,7181–7185.

Hollinshead, W., He, L., Tang, Y.J., 2014. Biofuel production: an odyssey from metabolic engineering to fermentation scale-up. Front. Microbiol. 5, 1–8.

Huang, D., et al., 2017. Synergistic hydrolysis of xylan using novel xylanases , β-xylosidases , and an α - L -arabinofuranosidase from Geobacillus thermodenitrificans NG80-2. Appl. Microbiol. Biotechnol. 101, 6023–6037.

Huang, P., et al., 2020. Evaluating protein engineering thermostability prediction tools using an independently generated dataset. ACS Omega 6487–6493. https://doi.org/ 10.1021/acsomega.9b04105.

Hucka, M., et al., 2019. The systems biology Markup Language (SBML): language specification for level 3 version 2 core release 2. J. Integr. Bioinf. 16.

Hussein, A.H., Lisowska, B.K., Leak, D.J., 2015. The genus Geobacillus and their biotechnological potential. In: Advances in Applied Microbiology, vol. 92. Elsevier.

Jacobs, N.J., Jacobs, J.M., 1976. Nitrate, Fumarate and oxygen as electron acceptors for a late step in microbial heme synthesis. Biochim. Biophys. 449, 1–9. Jensen, K., Broeken, V., Hansen, A.S.L., Sonnenschein, N., Herrgård, M.J., 2019.

OptCouple: joint simulation of gene knockouts, insertions and medium modifications for prediction of growth-coupled strain designs. Metab. Eng. Commun. 8.

Kananavičiūtė, R., Čtavičius, D., 2015. Genetic engineering of Geobacillus spp. J. Microbiol. Methods 111, 31–39.

Kim, Y., Ingram, L.O., Shanmugam, K.T., 2008. Dihydrolipoamide dehydrogenase mutation alters the NADH sensitivity of pyruvate dehydrogenase complex of Escherichia coli K-12. J. Bacteriol. 190, 3851–3858.

King, Z.A., et al., 2015. A web application for building, sharing, and embedding data-rich visualizations of biological pathways. PLoS Comput. Biol. 1–13. https://doi.org/ 10.1371/journal.pcbi.1004321.

Krüger, A., Schäfers, C., Schröder, C., Antranikian, G., 2018. Towards a sustainable biobased industry – highlighting the impact of extremophiles. N. Biotech. 40, 144–153.

Kulyashov, M., Peltek, S.E., Akberdin, I.R., 2020. A genome-scale metabolic model of 2,3-butanediol production by thermophilic bacteria Geobacillus icigianus. Microorganisms 8. Layer, G., Verfürth, K., Mahlitz, E., Jahn, D., 2002. Oxygen-independent coproporphyrinogen-III oxidase HemN from Escherichia coli. J. Biol. Chem. 277, 34136–34142.

Lee, S.Y., Kim, H.U., 2015. Systems strategies for developing industrial microbial strains. Nat. Biotechnol. 33, 1061–1072.

Leibniz Institute, 2021. DSMZ-German Collection of Microorganisms and Cell Cultures GmbH. https://www.dsmz.de/.

Lewis, N.E., et al., 2010. Omic data from evolved E . coli are consistent with computed optimal growth from genome-scale models. Mol. Syst. Biol. https://doi.org/ 10.1038/msb.2010.47.

Liberal, R., Lisowska, B.K., Leak, D.J., Pinney, J. W. PathwayBooster, 2015. A tool to support the curation of metabolic pathways. BMC Bioinf. 16, 4–9.

Lieven, C., et al., 2020. MEMOTE for standardized genome-scale metabolic model testing. Nat. Biotechnol. 38, 272–276.

Lin, P.P., et al., 2014. Isobutanol production at elevated temperatures in thermophilic Geobacillus thermoglucosidasius. Metab. Eng. 24, 1–8.

Lisowska, B.K., 2016. Genomic Analysis and Metabolic Modelling of Geobacillus Thermoglucosidasius NCIMB 11955. University of Bath.

Liu, B., et al., 2012. Characterization of a recombinant thermostable xylanase from hot spring thermophilic Geobacillus sp. TC-W7. J. Microbiol. Biotechnol. 22, 1388–1394.

Liu, Y., Li, J., Du, G., Chen, J., Liu, L., 2017. Metabolic engineering of Bacillus subtilis fueled by systems biology: recent advances and future directions. Biotechnol. Adv.35, 20–30.

Loftie-Eaton, W., et al., 2013. Balancing redox cofactor generation and ATP synthesis: key microaerobic responses in thermophilic fermentations. Biotechnol. Bioeng. 110, 1057–1065.

Lusk, B.G., 2019. Thermophiles; or, the Modern Prometheus: the importance of extreme microorganisms for understanding and applying extracellular electron transfer. Front. Microbiol. 10, 1–10.

Ma, Y., et al., 2016. Enhancement of polymerase activity of the large fragment in DNA Polymerase I from Geobacillus stearothermophilus by site-directed mutagenesis at the active site. BioMed Res. Int. 1–8.

Ma, W., et al., 2018. Metabolic engineering of carbon over flow metabolism of Bacillus subtilis for improved N-acetyl-glucosamine production. Bioresour. 250, 642–649.

Maayer, P. De, Brumm, P.J., Mead, D.A., Cowan, D.A., 2014. Comparative analysis of the Geobacillus hemicellulose utilization locus reveals a highly variable target for improved hemicellulolysis. BMC Genom. 15, 1–17.

Mahadevan, R., Schilling, C.H., 2003. The effects of alternate optimal solutions in constraint-based genome-scale metabolic models. Metab. Eng. 5, 264–276.

Mendoza, S.N., Olivier, B.G., Molenaar, D., Teusink, B., 2019. A systematic assessment of current genome-scale metabolic reconstruction tools. Genome Biol. 20, 1–20.

Miclet, E., Michels, P.A.M., Opperdoes, F.R., Lallemand, J., Duffieux, F., 2001. NMR spectroscopic analysis of the first two steps of the pentose-phosphate pathway elucidates the role of 6-phosphogluconolactonase. J. Biol. Chem. 276, 34840–34846.

Mo̧bius, K., et al., 2010. Heme biosynthesis is coupled to electron transport chains for energy generation. Proc. Natl. Acad. Sci. Unit. States Am. 107, 10436–10441.

Mohr, T., et al., 2018. CO-dependent hydrogen production by the facultative anaerobe Parageobacillus thermoglucosidasius. Microb. Cell Factories 17, 1–12.

Monk, J.M., et al., 2017. iML1515, a knowledgebase that computes Escherichia coli traits. Nat. Biotechnol. 35, 8–12.

Neidhardt, F.C., Ingraham, J.L., Schaechter, M., 1996. Chapter 3: Escherichia coli and Salmonella: cellular and molecular biology. Physiology of the Bacterial Cell: a Molecular Approach. American Society of Microbiology (ASM) Press.

Nielsen, J., Keasling, J.D., 2016. Engineering cellular metabolism. Cell 164, 1185–1197.

Niu, H., Leak, D., Shah, N., Kontoravdi, C., 2015. Metabolic characterization and modeling of fermentation process of an engineered Geobacillus thermoglucosidasius strain for bioethanol production with gas stripping. Chem. Eng. Sci. 122, 138–149.

Norsigian, C.J., Fang, X., Seif, Y., Monk, J.M., Palsson, B.O., 2020. A workflow for generating multi-strain genome-scale metabolic models of prokaryotes. Nat. Protoc.15, 1–14.

North, J.A., et al., 2016. Metabolic regulation as a consequence of anaerobic 5-methylthioadenosine recycling in Rhodospirillum rubrum. mBio 7, 1–12.

North, J.A., Miller, A.R., Wildenthal, J.A., Young, S.J., Tabita, F.R., 2017. Microbial pathway for anaerobic 5'methylthioadenosine metabolism coupled to ethylene formation. Proc. Natl. Acad. Sci. Unit. States Am. 10455–10464. https://doi.org/10.1073/pnas.1711625114.

Oh, Y.K., Palsson, B.O., Park, S.M., Schilling, C.H., Mahadevan, R., 2007. Genome-scale reconstruction of metabolic network in Bacillus subtilis based on high-throughput phenotyping and gene essentiality data. J. Biol. Chem. 282, 28791–28799.

Olivier, B.G., Bergmann, F.T., 2015. The systems biology Markup Language (SBML) level 3 package: flux balance constraints. J. Integr. Bioinf. 12.

Orth, J.D., Thiele, I., Palsson, B.Ø., 2010. What is flux balance analysis? Nat. Biotechnol. 28, 245–248.

Overbeek, R., et al., 2003. The ErgoTM genome analysis and discovery system. Nucleic Acids Res. 31, 164–171.

Panikov, N.S., Popova, N.A., Dorofeev, A.G., Nikolaev, Y.A., Verkhovtseva, N.V., 2003. Growth of the thermophilic bacterium geobacillus uralicus as a function of temperature and pH: an SCM-based kinetic analysis. Microbiology 72, 320–327.

Pharkya, P., Burgard, A.P., Maranas, C. D. OptStrain, 2004. A computational framework for redesign of microbial process. Genome Res. 14, 2367–2376.

Pogrebnyakov, I., Jendresen, C.B., Nielsen, A.T., 2017. Genetic toolbox for controlled expression of functional proteins in Geobacillus spp. PloS One 12.

Raux, E., Schubert, H.L., Warren, M.J., 2000. Biosynthesis of cobalamin (vitamin B12): a bacterial conundrum. Cell. Mol. Life Sci. 57, 1880–1893.

Reeve, B., Martinez-Klimova, E., Jonghe, J. de, Leak, D.J., Ellis, T., 2016. The geobacillus plasmid set: a modular toolkit for thermophile engineering. ACS Synth. Biol. 5, 1342–1347.

Robb, F., Antranikian, G., Grogan, D., Driessen, A., 2007. Thermophiles: Biology and Technology at High Temperatures. CRC Press.

Rocha, I., et al., 2010. OptFlux: an open-source software platform for in silico metabolic engineering. BMC Syst. Biol. 4, 1–12.

Sabath, N., Ferrada, E., Barve, A., Wagner, A., 2013. Growth temperature and genome size in bacteria are negatively correlated, suggesting genomic streamlining during thermal adaptation. Genome Biol. Evol. 5, 966–977.

Sakai, S., Inokuma, K., Nakashimada, Y., Nishio, N., 2008. Degradation of glyoxylate and glycolate with ATP Synthesis by a thermophilic anaerobic bacterium, Moorella sp. strain HUC22-1. Appl. Environ. Microbiol. 74, 1447–1452.

Sánchez, B.J., Li, F., Kerkhoven, E.J., Nielsen, J., 2019. SLIMEr: probing flexibility of lipid metabolism in yeast with an improved constraint-based modeling framework. BMC Syst. Biol. 13, 1–9.

Sawers, G., Bock, A., 1988. Anaerobic regulation of pyruvate formate-lyase from escherichia coli K-12. J. Bacteriol. 170, 5330–5336.

Schuetz, R., Kuepfer, L., Sauer, U., 2007. Systematic evaluation of objective functions for predicting intracellular fluxes in Escherichia coli. Mol. Syst. Biol. https://doi.org/ 10.1038/msb4100162.

Seemann, T. Prokka, 2014. Rapid prokaryotic genome annotation. Bioinformatics 30, 2068–2069.

Seifritz, C., Frȍstl, J.M., Drake, H.L., Daniel, S.L., 1999. Glycolate as a metabolic substrate for the acetogen Moorella thermoacetica. FEMS Microbiol. Lett. 170, 399–405.

Sheng, L., Zhang, Y., Minton, N.P., 2016. Complete genome sequence of Geobacillus thermoglucosidasius NCIMB 11955, the progenitor of a bioethanol production strain. Genome Announc. 4, 4–5. Steen, E.J., et al., 2010. Microbial production of fatty-acid-derived fuels and chemicals from plant biomass. Nature 463, 559–563.

Styles, M.Q., et al., 2020. The heterologous production of terpenes by the thermophile Parageobacillus thermoglucosidasius in a consolidated bioprocess using waste bread. Metab. Eng. https://doi.org/10.1016/j.ymben.2020.11.005.

Suzuki, H., 2018. Peculiarities and biotechnological potential of environmental adaptation by Geobacillus species. Appl. Microbiol. Biotechnol. 10425–10437.

Tang, Y.J., et al., 2009. Analysis of metabolic pathways and fluxes in a newly discovered thermophilic and ethanol-tolerant geobacillus strain. Biotechnol. Bioeng. 102, 1377–1386.

Taylor, M.P., Esteban, C.D., Leak, D.J., 2008. Development of a versatile shuttle vector for gene expression in Geobacillus spp. Plasmid 60, 45–52.

Thiele, I., Palsson, B.Ø., 2010. A protocol for generating a high-quality genome-scale metabolic reconstruction. Nat. Protoc. 5, 93–121.

Verduyn, C., Stouthamer, A.H., Scheffers, W.A., van Dijken, J.P., 1991. A theoretical evaluation of growth yields of yeasts. Antonie Leeuwenhoek 59, 49–63.

Wada, K., Suzuki, H., 2018. Biotechnological platforms of the moderate thermophiles , Geobacillus species: notable properties and genetic tools. *Physiological and Biotechnological Aspects of Extremophiles* 195–218. INC. https://doi.org/10.1016/ B978-0-12-818322-9.00015-0.

Wang, Q., et al., 2010. Metabolic flux control at the pyruvate node in an anaerobic escherichia coli strain with an active pyruvate dehydrogenase*. Appl. Environ.Microbiol. 76, 2107–2114.

Wang, Q., Cen, Z., Zhao, J., 2015. The survival mechanisms of thermophiles at high temperatures: an angle of omics. Physiology 97–106. https://doi.org/10.1152/ physiol.00066.2013.

Yang, Z., et al., 2020. Engineering thermophilic Geobacillus thermoglucosidasius for riboflavin production. Microb. Biotechnol. 1–11, 0.

Zeigler, D.R., 2014. The Geobacillus paradox: why is a thermophilic bacterial genus so prevalent on a mesophilic planet ? Microbiology 1–11. https://doi.org/10.1099/ mic.0.071696-0.

Zhang, J., et al., 2016. Engineering an NADPH/NADP redox biosensor in yeast. ACS Synth. Biol. 5, 1546–1556.

Zhou, J., Lian, J., Rao, C.V., 2020. Metabolic engineering of Parageobacillus thermoglucosidasius for the efficient production of (2R,3R)-butanediol. Appl. Microbiol. Biotechnol. 104, 4303–4311.

Zhu, W., Cha, D., Cheng, G., Peng, Q., Shen, P., 2007. Purification and characterization of a thermostable protease from a newly isolated Geobacillus sp . YMTC 1049. Enzym. Microb. Technol. 40, 1592–1597.

Zhu, Y., et al., 2015. Molecular cloning and characterization of a thermostable lipase from deep-sea thermophile Geobacillus sp. EPT9. World J. Microbiol. Biotechnol. 31, 295–306

3.3 Further GSMM discussion

3.3.1 Overall improvements compared to original model

Assessment of the original model with the Memote tool developed by Lieven *et al.* (2020)[120] identified a range of inter-connected reaction, metabolite and flux consistency issues, summarised and scored into the five areas of: Stoichiometric Consistency (70.8%), Mass Balance (43.3%), Charge Balance (52.3%), Metabolite Connectivity (37.8%) and Unbounded Flux in Default Medium (46.9%) (Figure. 17A).

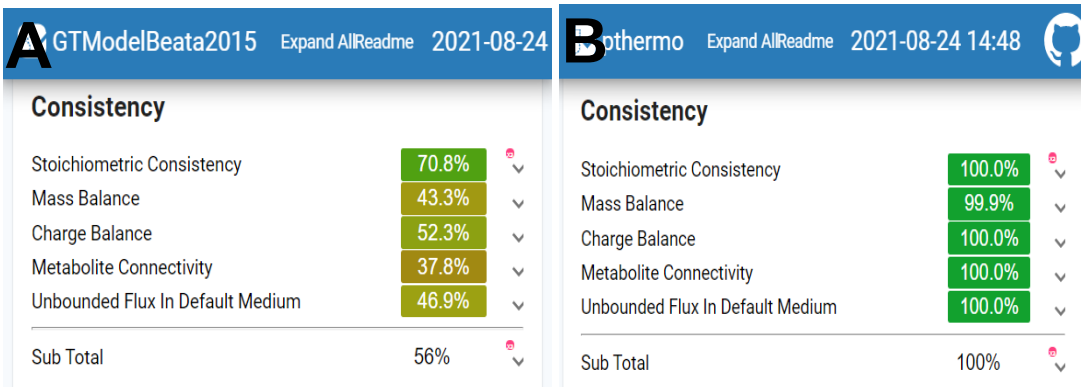


Figure 17: Memote analysis of original *P. thermoglucosidasius* model versus published *p-thermo* model. A) Original model (GTModelBeata2015). **B)** Current model (*p-thermo*). Analysis with Memote performed on 2021-08-24.

The stoichiometric consistency, mass balance and charge balance together constitute the conditions of the reactions in the network. Issues with these three factors in a steady-state model point to reactions that do not adhere to the law of conservation of mass, i.e., the net masses and/or changes of both sides of a particular reaction are unequal. This can have a substantial impact on the accuracy of any flux predictions, as unbalanced cycles can result in the system's ability to create mass or change from nothing or, alternatively, destroy mass in the system. 'Metabolite connectivity' evaluates the extent to which the metabolites in the model are involved in reactions. In this case, Memote suggested that there were 1418 metabolites with no connections to reactions in the model (62.22%). These could have been artefacts of the initial reconstruction, but we were mindful that they could also signify legitimate missing reactions in the metabolic network. Lastly, Memote's FVA-mediated evaluation of the "Unbounded Flux in the Default Medium" identified that 288 reactions in the model were able to carry unlimited flux (53.14%), reactions which in many cases could result in thermodynamically unrealistic flux rates through certain pathways.

Memote was then used to perform cycles of analysis and manual curation of the issues of the original model until evaluation the Memote tool output suggested the current *P. thermoglucosidasius* NCIMB 11955 model (hereafter referred to as *p-thermo*) demonstrated a 100% stoichiometric consistency, metabolite connectivity, charge balance and unbounded flux using the default medium, as well as a 99.9% mass balance (Fig. 17B).

3.3.2 The oxPPP and 6PGL: How a missing enzyme points to areas for improvement

Despite the improvements made to the *p-thermo* model, one notable and persistent limitation of the CCM of the base model still to be resolved was flux through the oxPPP. There is disagreement between the optimal flux distribution solution of the

p-thermo model, which suggests no flux through the oxPPP even under aerobic conditions and ^{13}C-MFA studies into the CCM of *P. thermoglucosidasius* and closely related *Geobacillus* species (Discussed in detail in Chapter 4) which suggest that the oxPPP still carries flux[121,127,128]. This disagreement in a key CCM reaction could help explain why pFBA of *p-thermo* alongside *Geobacillus* M10EXG measurements, suggested that it accurately represented growth *of P. thermoglucosidasius* under aerobic conditions (R^2 = 0.98) but was a less accurate representation under oxygen-limited (R^2 = 0.8) and anaerobic conditions (R^2 = 0.79) (Appendix Paper Supplementary Figure S3) as this pathway has a lower observed flux under those conditions.

3.3.2.1 Is 6PGL supposed to be missing?

While the genome of *P. thermoglucosidaisus* does encode both a glucose-6-phosphate dehydrogenase (G6PD) and 6-phosphogluconate dehydrogenase (6PD) from the oxidative branch of the pentose phosphate pathway (oxPPP), it does not contain a gene which encodes a 6-Phosphogluconolactonase (6PGL) [E.C. 3.1.1.31][55]. The lack of a gene for 6PGL can be explained through a combination of two factors. The first is a thermodynamic explanation that indicates this reaction can proceed spontaneously at higher temperatures.

As a result, while some thermophilic species such as Thermobacillus composti and Rhodothermus marinus do have a 6PGL enzyme, high growth temperatures are cited as the assumed reason for the common absence of this gene in annotated genome sequences of thermophiles such as *P. thermoglucosidasius*, genetically similar *Geobacillus* species *G. stearothermophilus* and *G. kaustophilus* and a wider range of thermophilic organisms such as *Anoxybacillus flavithermus* and *Thermus thermophilus* [170]. This is suggested by Wang, Cen and Zhao (2015) to have led to the hypothesis that selective gene loss which reduces genome complexity is an energetic cost minimizing mechanism that supports growth at high temperatures [171].

3.3.2.2 What did this mean for *p-thermo*?

Ultimately, this research assumed that the spontaneous hydrolysis of the D-glucono-1,5-lactone-6P sufficiently fast enough in *P. thermoglucosidasisus* to support the PPP flux required for growth. To account for this spontaneous hydrolysis in the GSMM, a pseudo-reaction (reaction ID "R_PGL" in *p-thermo*) was introduced into the *p-thermo* model to represent 6PGL:

D-Glucono-1,5-lactone 6-phosphate + H_2O => 6-Phospho-D-gluconate

However, a current issue of the *p-thermo* model is that when FVA is performed on the optimal biomass distribution solution, it suggests that the R_PGL pseudo-reaction is unable to carry flux under optimal growth (Figure. 18A). Through the application of a positive lower bound for the reaction, the pathway could still carry flux instead of the encoded transketolase [2.2.1.1] if forced to do so (Figure. 18B). However, this was accompanied by a decrease in biomass yield from 0.699809 to 0.692448 mmol g^{-1} DCW h^{-1}. This slight decrease implies that the non oxidative branch of the pentose phosphate pathway (non-oxPPP) is a more energy efficient path in *p-thermo* to take for growth.

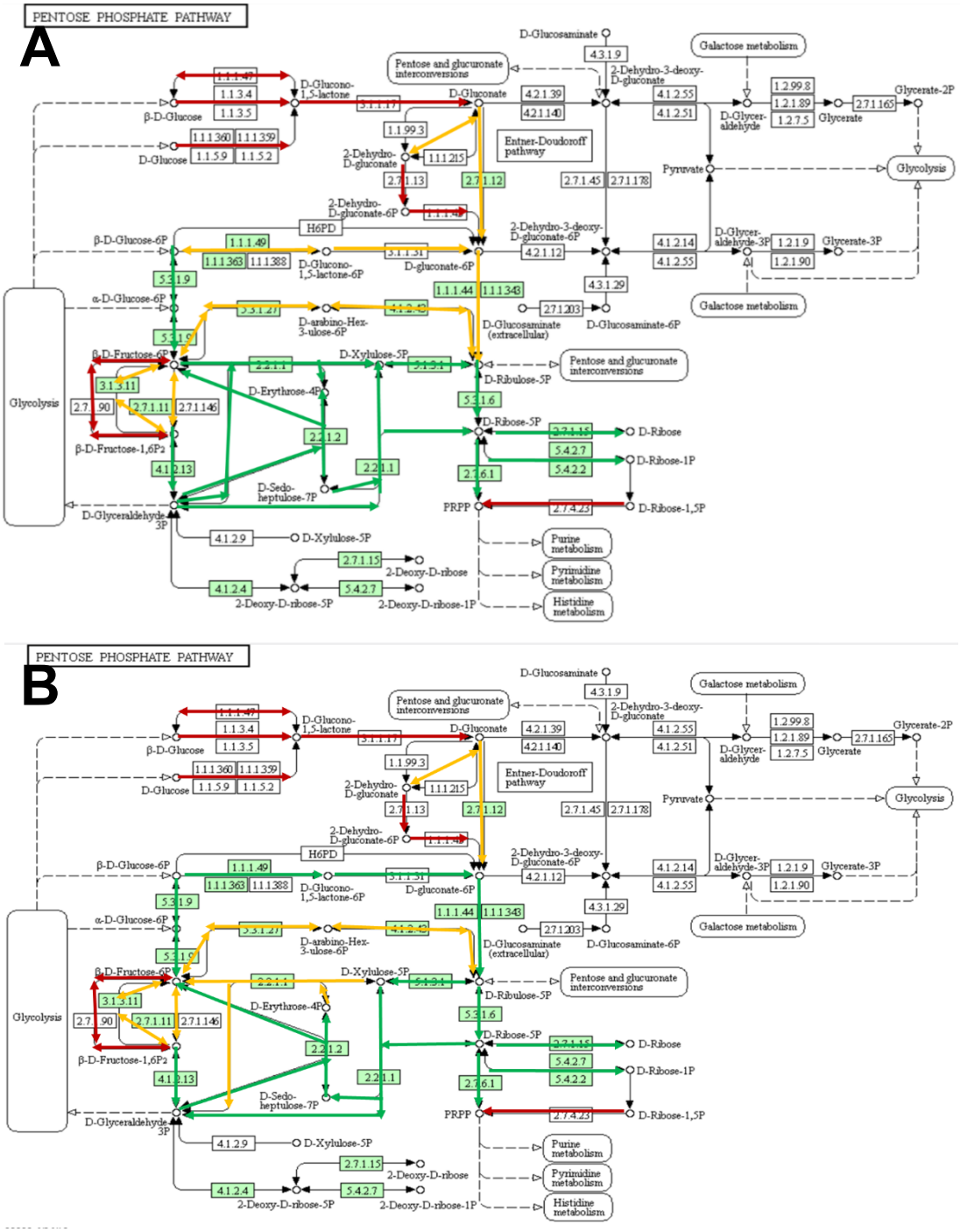


Figure. 18. FVA of *p-thermo* focused on the pentose phosphate pathway. A) *p-thermo* demonstrating no flux through the oxPPP. **B)** *p-thermo* demonstrating flux through the oxPPP when forced by constraints. Arrows indicate the net direction of flux. Red arrows indicate that the associated reaction is not present in the p-thermo model. Yellow arrows indicate that the associated reaction is present in the p-thermo model but does not carry any flux under the best fit FBA solution. Green arrows indicate that the associated reaction is present in the *p-thermo* model and carries flux under the best fit FBA solution.

The particular situation of Figure 18A, where the non-oxPPP is thermodynamically preferable, is termed riboneogenesis and reflects a feasible situation where the demand for ribose 5-phosphate is greater than the demand for NADPH[172].

Issues with the oxPPP are not uncommon in GSMMs and can point towards an overall imbalance in the NADPH/NADH cofactor ratios. Indeed, A meta-analysis evaluating GSMMs for eukaryote *Saccharomyces cerevisiae* by Pereira, Nielsen and Rocha (2016) [173] identified that several of the published models investigated by the authors predicted erroneous flux in the oxPPP when compared to experimental flux data. One of the key findings of this work was that flux and NADPH generation could be restored to the oxPPP with the deactivation of a cytosolic isocitratedehydrogenase (ICDHc), which in several cases was found to be supplying the entire requirement of NADPH for growth. However, there is only the single NADP linked isocitrate dehydrogenase noted in P thermoglucosidasius genome (reaction ID "R_ICDHyr" in *p-thermo*) which is critical for operation of the citric acid cycle and growth. What this points to however is that flux through another NADPH generating reaction may be responsible and that future work should investigate and compare the NADPH and NADH generating reactions of the model to potentially find a culprit for this imbalance [173].

Alternatively, these observations combined with the existing data surrounding the spontaneity of the hydrolysis of the ^{D}-glucono-1,5-lactone-6P substrate, suggest that the oxPPP may not be favoured by the *p-thermo* model as the stoichiometric reaction does not consider a potentially lower energy catalytic route of reaction provided by high temperature. Given the importance of the oxPPP in the production of histidine and nucleic acids required for cell growth, future ^{13}C-MFA research surrounding *P. thermoglucosidasius* could investigate whether flux through the oxPPP changes under different temperature conditions. As suggested by Crown, Long and Antoniewicz (2016), 1-^{13}C glucose operating as the only tracer would be the best option to resolve the oxPPP [174]. Using this tracer and comparing oxPPP fluxes of *P. thermoglucosidasius* cells grown at different temperatures could help to quantify the specific impact of flux through this pathway on cell growth and indicate whether this hydrolysis could operate as a rate-limiting step at lower growth temperatures.

Overall, this unresolved issue highlights how valuable the integration of fluxomics data to a GSSM can be to validate the predictions of a model and gain a more accurate systems level understanding of an organism, and therefore support the higher goal of designing a realistic ME strategy which achieves both over-production of a target metabolite and balanced growth.

3.3.3 Roles of thio-cofactors

Perhaps the most interesting results demonstrating the value of *p-thermo* were the observations that minimal media supplementation with a combination of biotin, thiamine and iron (III) was able to support *P. thermoglucosidasius* cell growth under anaerobic conditions. This is important as it is this condition which is most valuable from the perspective of using *P. thermoglucosidasius* as a microbial chassis for industrial fermentation due to its use in producing fermentation products lactic acid and ethanol. Notably, these supplements and (other suggested supplements) fall under the umbrella of term thio-cofactors which includes metabolites such as: thiamine, biotin, molybdopterins and iron-sulphur clusters [175]. This suggests that part of the reason *P. thermoglucosidasius* may struggle under anaerobic conditions to generate/regenerate the production and/or oxidation state of these sulphur containing co-factors.

Arguments for this hypothesis can be seen when examining the roles of thiamine and iron (III).

3.3.3.1 Thiamine

Thiamine, and it's phosphprylated form Thiamin pyrophosphate (TPP), is an essential cofactor in carbohydrate metabolism, often directly integrated into enzymes. Notably for the CCM of *P. thermoglucosidasius*, enzymes dependent on TPP as a co-factor are involved in catalyzing both the aerobic and anaerobic conversions of pyruvate to Acetyl-CoA as well as the oxidation of citric acid cycle intermediate α-ketoglutarate to succinyl CoA.

Under aerobic conditions the TPP-dependent pyruvate dehydrogenase complex (PDH) catalyses the production of acetyl-CoA and CO_2 from pyruvate [176]. However, under conditions where oxygen (or another electron acceptor) is absent, *P. thermoglucosidasius* will predominantly utilise TPP-dependent pyruvate formate lyase (PFL)[56] instead, which uses pyruvate as an electron acceptor and produces one molecule of acetyl CoA and one molecule of formate to help maintain overall redox balance in the cell.

Given the importance of PFL-dependent production of Acetyl-CoA to oxygen-limited growth of *P. thermoglucosidasius*, it is therefore logical that experimental supplementation of thiamine would support continued growth under low oxygen conditions. However, while there does not seem to be transport machinery for direct import of complete thiamine in *P. thermoglucosidasius*, it is also recognised that thiamine forms hydrolysis products in water. If these smaller products can be imported, it is likely they can be integrated into the thiamine biosynthesis pathway an energetically favourable manner by the thiamine salvage pathway to form TPP [176].

Additionally, acetolactate synthases, which catalyse the decarboxylation of pyruvate to acetolactate, are also TPP-dependent enzymes [178]. As acetolactate represents a precursor for proteinogenic amino acids L-valine and L-leucine, increased metabolic pools of thiamine likely also support this synthesis.
In addition to the potential direct import of thiamine or its component hydrolysis products in these experiments, the pool of available thiamine would conventionally depend on its ability to be synthesized intracellularly. Notably, the set of critical reactions to enable anaerobic growth in *p-thermo* which required oxygen included a glycine oxidase (Paper Table 1) (defined by reaction ID "R_ GLYHOR" in *p-thermo*) catalysing the reaction:

L-Tyrosine + S-Adenosyl-L-methionine + Reduced acceptor <=> Iminoglycine + 4-Cresol + 5'-Deoxyadenosine + L-Methionine + Acceptor

This enzyme has been long recognized in *Bacillus subtilis* to be critical in the oxygen-dependent formation of the thiazole moiety precursor iminoglycine required to synthesise thiamine pyrophosphate [179].If activity of this enzyme is limited under low oxygen conditions, this could therefore impact the ability of *P. thermoglucosidasius* to synthesise the cofactor thiamine pyrophosphate for PDH and reliant growth.

Notably however, while not encoded in the *P. thermoglucosidasius* genome, an oxygen independent 2-iminoacetate synthase [EC:4.1.99.19] exists that can also catalyse iminoglycine formation from tyrosine through the reaction:

L-Tyrosine + S-Adenosyl-L-methionine + Reduced acceptor <=> Iminoglycine + 4-Cresol + 5'-Deoxyadenosine + L-Methionine + Acceptor

Future work could therefore investigate whether the heterologous expression of a thermophilic 2-iminoacetate synthase in *P. thermoglucosidasius* supports growth under low oxygen conditions.

This result could also help explain the finding of Dr. Charlotte Ward in her PhD thesis that supplementation of serine to a minimal growth media for *P. thermoglucosidisus* helped support growth of *P. thermoglucosidasius* under micro-aerobic condtions[121]. Rather than simply providing an addition carbon source, the fact that serine is a direct precursor to glycine may have meant that adding serine increased the pool of available glycine substrate for an oxygen-limited glycine oxidase and helped support overall thiamine/thiamine pyrophosphate production for growth.

3.3.3.2 Iron

Two findings of the *p-thermo* associated growth experiments were that: 1) supplementation with iron (III) to the medium supported anaerobic growth of *P. thermoglucosidasius* cells *in vivo* and that 2) hemin, a porphyrin containing Fe (iron) and Cl (chloride), could not be produced under anaerobic conditions in *p-thermo*.

Heme biosynthesis

The importance of glycine was also implicated in the oxygen-dependent biosynthesis of heme and vitamin B12. This was due to a native *P. thermoglucosidiasius* plasmid-conferred 5-aminolevulinic acid synthase, (defined by reaction ID "R_ ALASm" in *p-thermo*, which synthesised the 5-aminolevulinic acid from a glycine precursor rather than a more common L-glutamate precursor. While a tBLASTn search performed as part of the published research suggested a possible unannotated homolog to oxygen-idependent *B. subtilis* HemN in the *P. thermoglucosidasius* NCIMB 11955, further examination of the complete heme biosynthesis pathway in the genome of *P. thermoglucosidasius* suggested that additional enzymes in the pathway could also explain why heme biosynthesis could be oxygen-dependent (Figure 19).

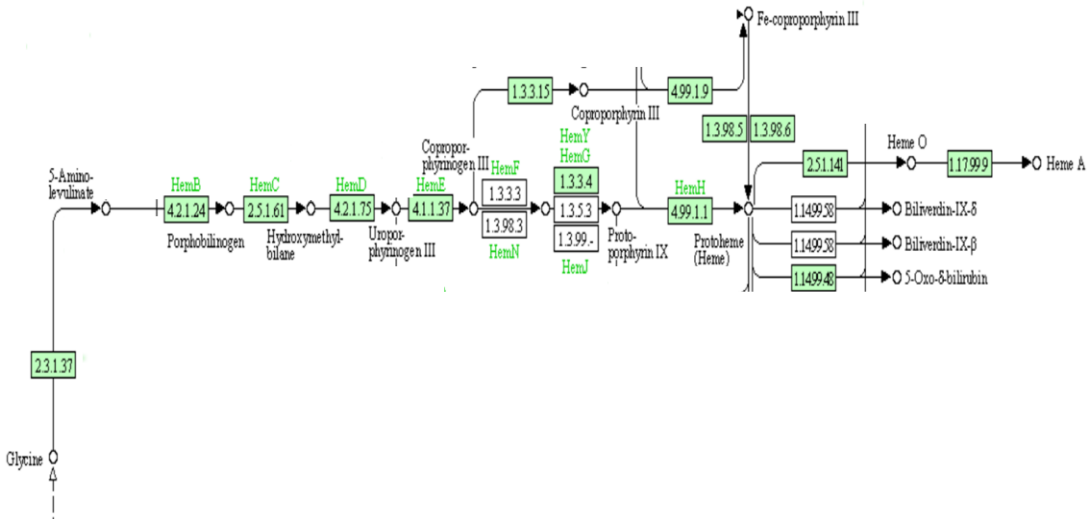


Figure 19: KEGG biosynthesis pathway from glycine to heme featuring *P. thermoglucosidasius* genes. Green filled boxes represent pathway enzymes with corresponding genes in the *P. thermoglucosidaisus* genome, white boxes represent pathway enzymes with no associated genes.

While genes HemB,C,D,E, and H are encoded for enzymes which are not oxygen dependent, a protoporphyrinogen/coproporphyrinogen III oxidase encoded by a single gene can perform the roles of EC numbers [1.3.3.4] (HemY) and [1.3.3.15] and is dependent on oxygen. This therefore suggests that the current route to heme from 5-aminolevulinic acid is oxygen dependent due to using this oxygen dependent protoporphyrinogen. An oxygen independent route could therefore involve both the aforementioned HemN and a menaquinone-dependent protoporphyrinogen oxidase [EC:1.3.5.3] termed HemG. Future work could therefore investigate whether the heterologous expression of genes for oxygen-independent HemN (oxygen-independent coproporphyrinogen III oxidase [EC:1.3.98.3] and HemG (menaquinone-dependent protoporphyrinogen oxidase [EC:1.3.5.3] in *P. thermoglucosidasius* could further support growth under low oxygen conditions.

Ferredoxin-linked enzymes

A final key thio-cofactor that could have been supported through addition of iron (III) are iron-sulphur clusters, which can feature in anaerobic metabolism as oxygen-sensitve ferredoxins.

Ferredoxins are [4Fe-4S] and [2Fe-2S] clusters that are involved with electron transfer in many metabolic reactions. While reactions containing ferredoxins are unsual for many lactic acid producing Bacilli, enzymes involving ferredoxins are present within the *P. thermoglucosidasius* genome [180]. While initially appearing unusual, it is important to remember that *P. thermoglucosidasius* were first extracted from the notably iron-rich sulphourous hot-springs in Yellowstone National Park [181]. This suggests that an evolutionary advantage of ferredoxin-containing proteins in *P. thermoglucosisisus* history and could explain why supplementation with a supplement mixture containing iron was determined to be advantageous to anaerobic growth of *P. thermoglucosidaisus* when compared to the control (Paper Fig. 5). Indeed, supplementation of iron could be advantageous both for supporting several ferredoxins reactions present in the *P. thermoglucosidasius* genome and *p-thermo* that could be operating under anaerobic conditions including a reduced ferredoxin:H+ oxidoreductase (reaction ID "R_HYDA" in *p-thermo*), catalysing the model reaction of: $2Fd_{Rd} + 2H^+ <-> 2Fd_{Ox} + H_2$, an Ammonia:ferredoxin oxidoreductase (reaction ID "R_NO2R" in *p-thermo*) catalysing the model reaction of: $NH_4 + 6Fd_{Ox} <-> 6Fd_{Rd} + 8H^+ + NO_2$ and a 4-hydroxy-3-methylbut-2-en-1-yl-diphosphate:oxidized ferredoxin oxidoreductase (reaction ID "R_MECDPOR" in *p-thermo*), an enzyme involved with isoprenoid biosynthesis.

Perhaps the most interesting of the reactions encoded in the *P. thermoglucosisisus* genome and in *p-thermo* is the existence of an oxoglutarate:ferredoxin oxidoreductase (reaction ID "R_OOR3r" in *p-thermo*). This catalyses the unidirectional model reaction of: $CoA + 2Fd_{Ox} + AKG -> CO_2 + 2Fd_{Rd} + SucCoA + H+$ in the citric acid cycle suggesting that this may support continued citric acid cycle turning under low oxygen conditions. Indeed, this could help explain the transcriptomic observations of Loftie-Eaton *et al.* (2012) that *P. thermoglucosidasius* NCIMB 11955 cells growing under anaerobic conditions continued to operate in an oxygen-scavening state [114]. Additionally, the recently discovered hydrogen producing capabilities of *P. thermoglucosidasisus* (captured in *p-thermo* through reaction ID "R_COOR") when grown on carbon monoxide was identified by Mohr *et al.* (2018)[7] to require Ni-Fe to form part of a metallocenter for the discovered protein complex of CO dehydrogenase (CODH) and energy-converting hydrogenase (ECH) which performs the hydrogenogenesis. From this perspective, supplementation with iron is also therefore likely to support future research into the optimal growth conditions of *P. thermoglucosidasius* under CO atmosphere condtions for the production of hydrogen.

4. Chapter 4: INST-^{13}C-MFA

4.1 Introduction

4.1.1 Isotope labelling experiments, isotopically stationary ^{13}C Metabolic flux Analysis and isotopically instationary ^{13}C Metabolic flux Analysis

In the context of metabolism, flux is the absolute rate of metabolite conversion in reaction, or a pathway made up of multiple reactions [182]. The physiological phenotype expressed by an organism is determined by the distributions of flux through its complete metabolic reaction network [183]. The aim of fluxomics therefore is to describe the distributions of fluxes in a given metabolic network, subject to sets of environmental constraints, in order to better understand the systems-level functional behaviours of an organism. Indeed, such *in vivo* metabolic flux distributions inherently incorporate the interplay of the all the regulatory constraints of metabolism which constitute the observed phenotype [47] [184].

4.1.1.1 ^{13}C Isotope labelling experiments (ILEs): The bow-tie conceptual framework

As it is not possible to directly measure fluxes in vivo, conventional fluxomics indirectly measures the metabolic fluxes of *in vivo* reactions through stable ^{13}C isotope-tracer labelling experiments [183].

A ^{13}C ILE revolves around culturing cells at a fixed growth rate on a ^{13}C-enriched carbon tracer in a minimal defined media. Under the ‘bow-tie’ conceptual framework of metabolism, the ^{13}C carbon tracer will be catabolized and its constitutive carbons will be rearranged through CCM into metabolite pools of universal carbon intermediates [185]. These ^{13}C-labelled carbon intermediates will then funnel back out of CCM and incorporate into a measurable downstream product of CCM (typically the amino acids of cell proteins, which can be analysed after hydrolysis, in predictable manners (Figure. 20).

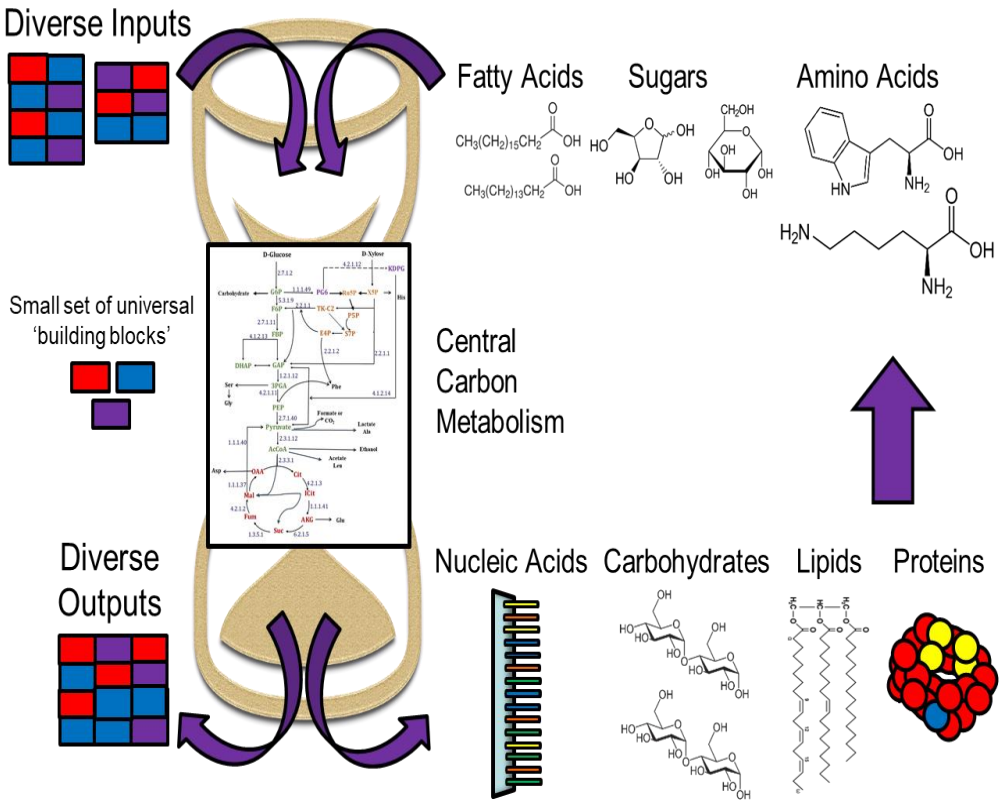


Figure. 20. Simplified schematic of the 'bow-tie' conceptual framework of metabolism. Diverse substrate inputs get broken down into metabolites of central carbon metabolism. These metabolites intermediates then get integrated into metabolic products required for cell growth. When required, these products can also be catabolized back to substrates (represented by the single straight arrow).

4.1.1.2 Isotopomer networks, distributions, and forward estimates

Isotopomers

As a catabolized ^{13}C atom travels through CCM, it can get incorporated into a downstream metabolite in a range of ways depending on the atom's original position in the ^{13}C tracer substrate, the percentage of catabolized substrate with a ^{13}C isotope and the different metabolic pathways the ^{13}C atom travelled through. As a result, isotope labelling experiments can generate metabolites with specific isotope isomers known as 'isotopomers'[118]. These metabolite isotopomers can be further classified as either positional isotopomers (essentially mass regioisomers) which incorporate ^{13}C atoms in different locations in the same chemical metabolite and define which atoms are labelled in a molecule, and mass isotopomers (or isotopologues) which represent groups of one or more positional isotopomers which have the same cumulative atomic mass and define how many atoms are ^{13}C labelled in a molecule (Figure. 21).

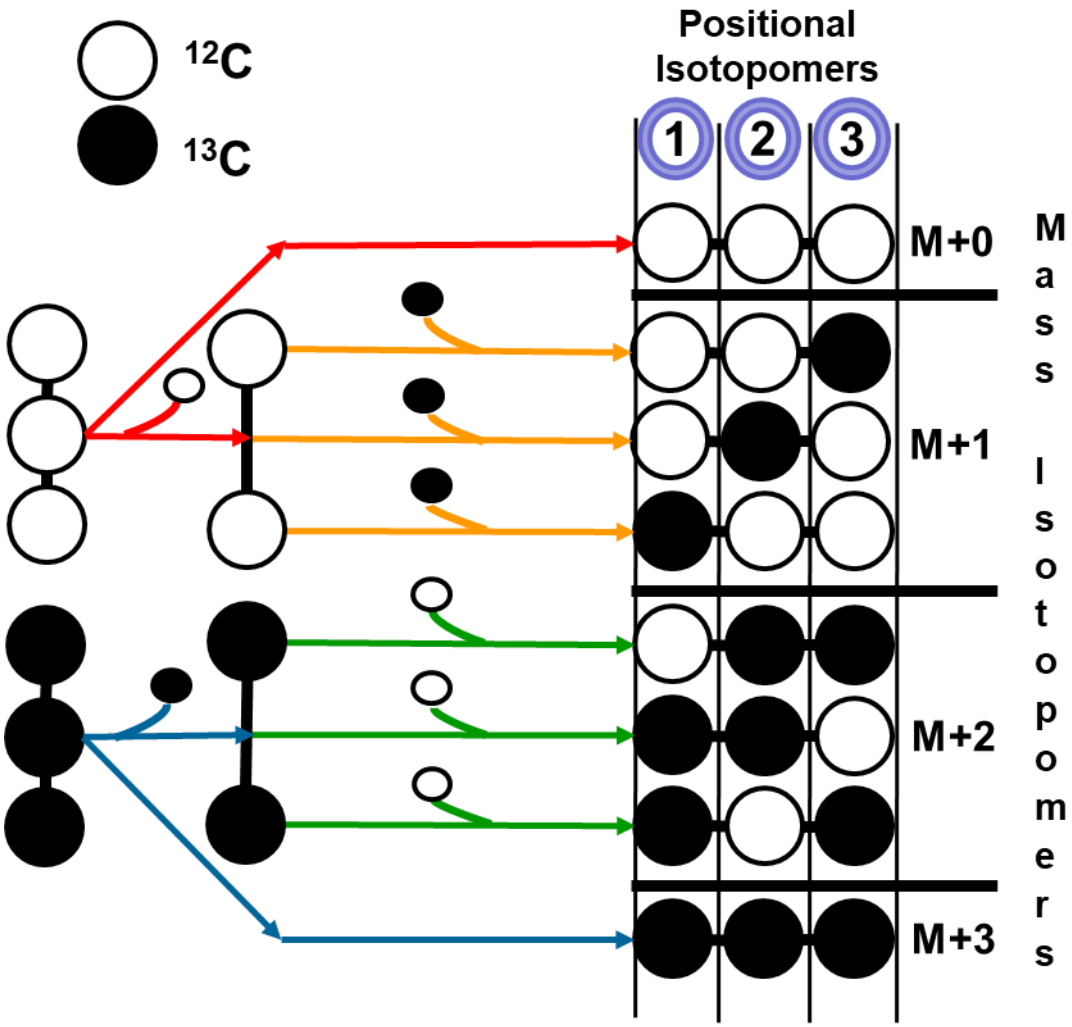


Figure. 21. Pictoral explanation of positional and mass isotopomers for a 3-carbon compound. The position in a carbon metabolite in which ^{13}C-labelled or naturally labelled carbons are eliminated or integrated gives rise to their possible positional isotopomers (1-3) or mass isotopomers (defined here in terms of their mass shift from M+0 to M+3). Adapted from Toya *et al.* (2011)[186] .

For analysis, this information is typically expressed in vector form. The complete collection of all positional isotopomers in vector form is commonly termed the isotopomer distribution vector (IDV), while the complete collection of mass isotopomers of a metabolite (i.e., M+1, M+2, M+3, … M+*n*) relative to its original molecular mass (defined as M+0) (Figure. 21) is termed the mass isotopomer distribution (MID) vector. When combined with measured ^{13}C isotope incorporation data, these MIDs can ultimately be expressed as Mass Distribution Vectors (MDVs) containing varying fractional abundances of each mass isotopomer[118].

Isotopomer Transition Networks and Isotopomer Distribution Vectors

The biochemical reactions which comprise the CCM of an organism can be expressed as carbon transition networks. The essential reaction stoichiometry and atom mapping information of these networks are typically validated using experimentally derived genomic and biochemical data from the organism in question [182].

Carbon transition networks codify carbon-containing metabolites as IDVs. Metabolic reactions in the network therefore become non-linear isotopomer balance equations describing the carbon transitions and rearrangements between IDVs as part of an isotopomer transition network[187]. These isotopomer transition networks therefore provide a structural framework detailing how every possible labelling state of every metabolite pool in the network is connected to one another.
As a result, an isotopomer transition network allows forward *in-silico* predictions of potential *in-vivo* IDVs using only a given input of a ^{13}C labelled tracer molecule [186]. In contrast to GSMMs, the metabolic network topology of an isotopomer transition network is intentionally simplified to be the minimum number of key metabolic reactions required to predict IDVs. In ^{13}C-MFA, it is these IDVs to which experimental ^{13}C labelling data is compared.

Simplifying the isotopomer balance equation - EMUs

Due to the vast number of possible combinations of metabolic positional and mass isotopomers, quantitative estimation of metabolic fluxes based on ^{13}C isotopomer transition networks composed of non-linear isotopomer balance equations becomes computationally expensive. To tackle this challenge, several modelling approaches have been developed which decompose the supplied isotopomer transition network of non-linear equations into simpler, linear equations allowing for more efficient computation[188]. These modelling approaches introduced the: 'Cumomer'[189] , 'Bondomer'[190] and Elementary Metabolite Units (EMUs)[191,192], with the EMU framework employed in this research. These approaches are reviewed in detail by Weitzel, Wiechert and Nöh (2007) [187, 189].

The EMU framework condenses the complete collection of possible isotopomers into groups of only the essential positional and mass isotopomers required to both satisfy the connectivity constraints imposed by the choice of atom transition network and describe the measured MIDs of labelled fragment ions [193]. In many cases, if no carbon transitions of importance to the overall MIDs take place in a particular metabolic pathway, the EMU framework can effectively reduce a pathway of consecutive enzymatic reactions into a single carbon transition reaction. As a result, the EMU framework significantly decreases the number of balance equations that need to be solved and converts the non-linear system of isotopomer equations into a system of linear equations featuring EMUs[187]. Indeed, Antoniewicz *et al.* (2007) demonstrated that, the application of the EMU framework, a 26 amino acid fragment labelling experiment of a simple *E. coli* metabolic model could condense a non-linear 4612 isotopomer-based problem to a linear 223 EMU-based problem [191].

Physiological Steady-states and Chemostat continuous culture systems

Like FBA, a typical ^{13}C-MFA assumes that harvested cells had reached a metabolic steady-state (i.e., a stable growth rate in which the sum of intracellular influxes and effluxes are zero). Additionally, isotopically steady-state ^{13}C-MFA also assumes that a similar isotopic equilibrium had been reached where the distributions of ^{13}C atoms also did not change with respect to time. Theoretically, these conditions can be achieved experimentally by harvesting cells at an isotopic and metabolic quasi-steady-state during mid-exponential batch growth on a ^{13}C carbon media, however in reality the transitory changes in substrate and metabolite products, even during a short sample window, will affect the accuracy of any flux determination [194].

In order to more accurately achieve the metabolic steady-state conditions required for INST-^{13}C-MFA experiments, this research grew *P. thermoglucosidasius* cells in 3 different CSTR configurations operated as chemostats. Under the chemostat configuration, sterile feed media continually enters the controlled environment of the CSTR vessel and is balanced by a concurrent removal of the equivalent volume of

cell culture [195]. As a result, when a population of cells reaches a steady-state, the maximum specific growth rate of those cells is equal to the dilution rate [194]. Therefore, by varying the media addition rate, different phenotypic behaviours of cells at different fixed growth rates can be examined by sampling at any time from this continuous culture. Ultimately, the chemostat growth method enables the direct control of cellular growth rates for ^{13}C-MFA experiments at a range of programmable growth rates and on different concentrations of media components.

Experimental determination of ^{13}C labelling patterns

Experimental determination of the labelling of metabolite isotopomers from cell culture samples is typically performed using a GC-MS or liquid chromatography-mass spectrometry (LC-MS), which can provide predominantly mass isotopomer [196], and/or NMR data [197].
While NMR can be used to distinguish positional isotopomers from low analyte concentrations, as well as mass isotopomers, this research focuses on the use of more sensitive GC-MS to elucidate MDVs for TBDMS-derivitized amino acids which were originally isolated from hydrolyzed samples of *P. thermoglucosidasius* biomass.

Natural abundance correction

Once initial MDVs have been determined for metabolites of interest in an ILE, the measured abundances of each mass isotopomer must have the contribution of any naturally present stable isotopes of: hydrogen, carbon, nitrogen, oxygen, silicon and sulphur removed. This natural isotope abundance correction must be performed to ensure only the mass isotopomers generated as a result of the fed ^{13}C tracer are considered in further analysis. The importance, history and best strategies for natural isotope correction are reviewed in detail by Midani *et al.* (2017) [116].

Qualitative and Quantitative mass isotopomer distribution analysis

Once the amino acid MDVs have been corrected, analysis of the degrees and patterns of incorporation of ^{13}C can be performed to both qualitatively and quantitatively to infer the influence of metabolic pathways on the resulting isotopomer patterns and to describe carbon flux distributions between different pathways of CCM[199].

4.1.2.1 Qualitative ^{13}C fingerprinting

In qualitative mass isotopomer distribution analysis, known as ^{13}C-fingerprinting, fully corrected MIDs enable qualitative observations to be made of the metabolic network structure and connectivity of a particular organism using knowledge of the overall biochemical network of an organism's metabolism, rather than attempting to fit the data to an assumed isotopomer distribution network [200].

Such qualitative observations and interpretations can be performed by considering the total ^{13}C abundance and the ^{13}C-labelling of MDVs. This is because the resulting labelling patterns of specific amino acid products will be derived from specific branching nodes in CCM which will reflect the available carbon sources and environmental influences[174] (Figure. 22).

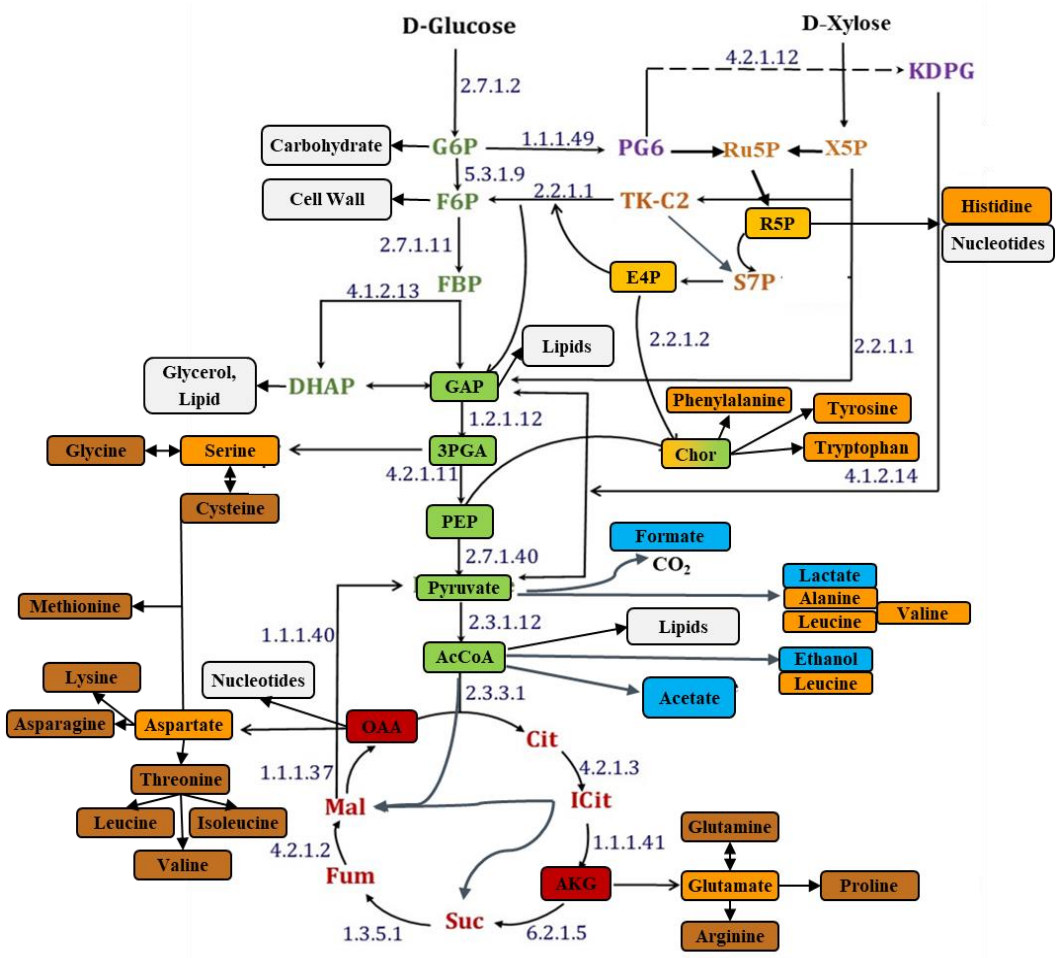


Figure 22. Simplified schematic of *P. thermoglucosidasius* aerobic and anaerobic central carbon metabolism fed with glucose and xylose. Boxed nodes highlight precursors involved with amino acid metabolism. Boxed amino acids highlighted in orange represent amino acids derived from a central carbon metabolite. The remaining amino acids derived from other amino acids are highlighted in brown. Central carbon metabolites are highlighted in green (glycolysis), yellow (pentose phosphate pathway), purple (Entner-Doudoroff pathway) and red (Citric Acid cycle). The fermentation metabolites measured in this research are highlighted blue. Available metabolic paths informed from genome sequence of *P. thermoglucosidasius* DSM 2542, complete with enzyme commission (EC) numbers (purple text). AcCoA (acetyl-coenzyme A), AKG (α-ketoglutarate), Cit (citrate), Chor (Chorismate), DHAP (Dihydroxyacetone phosphate), E4P (erythrose-4-phosphate), FBP (fructose-bis-phosphate), Fum (fumarate), GAP (Glyceraldehyde-3-phosphate), G6P (glucose-6-phosphate), 6PG (6-phosphogluconate), ICit (isocitrate), Mal (malate), OAA (oxaloacetate), F6P (fructose-6-phosphate), PEP (phosphoenolpyruvate), 3PGA (3-phosphoglyceric acid), R5P (ribose-5-phosphate), Ru5P (ribulose-5-phosphate), S7P (sedo-heptulose-7-phosphate), Suc (Succinate), TK-C2 (Transketolase C-2) and Xu5P (xylulose-5-phosphate). Adapted from Shree *et al.* (2018)[201], Tang *et al.* (2009)[128] and Cordova and Antoniewicz (2016)[127].

As a result, this analysis can help to describe metabolic pathway activity and help to infer the relative contributions of different metabolic pathways to the observed phenotype of an organism [202].

For example, this qualitative approach was use by Jyoti *et al.* (2020) to suggest an unusual mode of aerobic metabolism in the phytopathogenic bacterium *Ralstonia solanaceaum* [203]. Through feeding *R. solanaceaum* with a range of glucose tracers, including 1-^{13}C glucose, the authors were able to suggest that the Entner-Doudoroff (ED) and the non-oxidative branch of the PPP (non-oxPP) were used under aerobic conditions instead of glycolysis. This was inferred from observations that the glycolysis path 3PGA-and-PEP-node derived serine and PEP-node derived Phenylalanine and Tyrosine were unlabelled while ^{13}C labelling was observed for pyruvate-node-derived alanine (19.5%), OAA node-derived valine (11.5%) and R5P node-derived histidine (4%).

4.1.2.2 Quantitative ^{13}C MFA Stationary ^{13}C fluxomics

Quantitative fluxomics analysis can be performed to infer absolute fluxes through metabolic pathways, complete with accompanying statistical data, in what is commonly defined as ^{13}C-Metabolic Flux Analysis (^{13}C-MFA) [200].

Fundamentally, this involves back-tracing the experimentally measured ^{13}C-labelled metabolite profiles to their *in silico* simulated equivalents. The quantification of the unknown fluxes through metabolic pathways are determined through solving the non-linear least squares problem, which ultimately seeks to minimize the SSR between the assumed IDV and metabolite profiles (simulated from the isotopomer transition network, specific ^{13}C labelled tracers and any growth, substrate import or product export rates and the experimentally observed MDVs fit to them[187].

4.1.2.3 ^{13}C-isotopic tracer MFA for non-steady states

While ILEs involving rapid reaction quenching techniques for more metabolomics focused ^{13}C-MFA experiments have been developed to allow for direct quantification of some transient metabolites, the current research focuses on determining flux distributions of *P. thermoglucosidasius* from dynamic ^{13}C labelling experiments, in what is termed isotopically instationary (occasionally termed non-stationary) ^{13}C-MFA (INST-^{13}C-MFA) [199 204 205].

Under this dynamic system framework, it is still assumed that a metabolic steady state (with consistent fluxes and metabolite pool-sizes) has been achieved by a population of cells [202]. However, as ^{13}C-labelled substrate is introduced to the continuous cell culture it slowly replaces the unlabelled substrates. As a result, in an INST-^{13}C-MFA framework the ^{13}C enrichment and resulting isotopomer distributions of cellular metabolites, in this case proteinogenic amino acids, is assumed to be able to change over time (Figure. 23) [199 206]

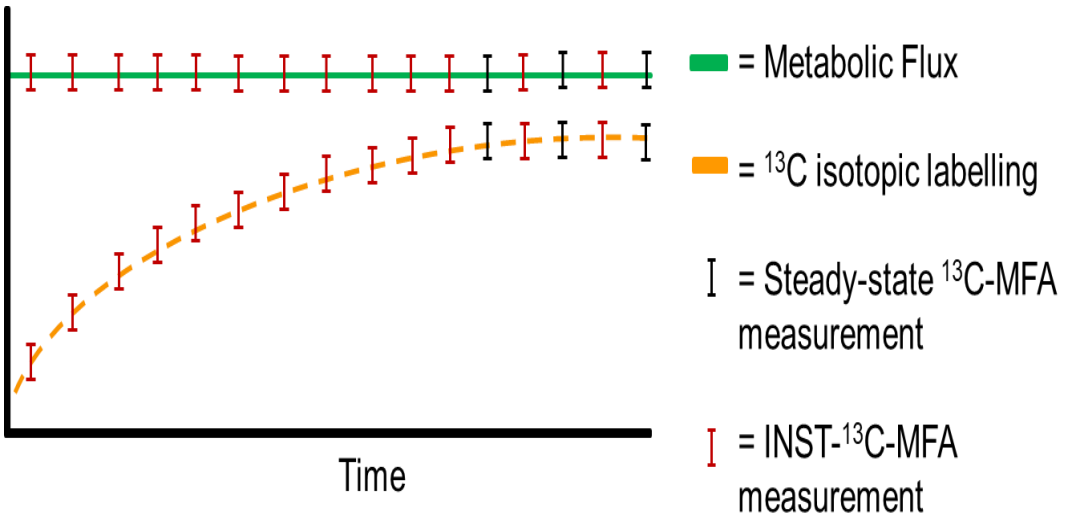


Figure 23. Graphical comparison of isotopically steady-state versus instationary ^{13}C-MFA. Both ^{13}C-MFA methodologies assume an unchanging metabolic flux (green line) over time. Sampling in isotopically steady-state ^{13}C-MFA (black bars) is performed once ^{13}C-labelling (orange dashed line) has incorporated over time to a steady state, whereas INST-^{13}C-MFA sampling (red bars) is performed throughout the ^{13}C-label incorporation process.

To measure the rates of ^{13}C isotope incorporation over time, this research measured isotopomer distributions of ^{13}C-fed *P. thermoglucosidasius* cell culture samplesat various time points from the introduction of the labelled tracer to an isotopic steady-state, thereby enabling both isotopically instationary and stationary ^{13}C-MFA.

Compared to isotopically steady-state ^{13}C-MFA, INST-^{13}C-MFA comes with both advantages and disadvantages. The main disadvantages of INST-^{13}C-MFA are experimental and financial practicalities. Experimentally, to capture the dynamic rates of ^{13}C incorporation, more samples representing numerous time points are required. This increases the size of the overall sample and data processing task. Additionally, INST-^{13}C-MFA experiments are longer as they first require cells to have reached a metabolic steady state before the introduction of a ^{13}C tracer. Financially, a major limitation of all INST-^{13}C-MFA experiments performed in this research was the cost of the tracers themselves [184]. While Crown, Long and Antoniewicz (2016) [174] and Cordova and Antoniewicz (2016) [127] have demonstrated that different ^{13}C labelling positions in glucose tracers in *E. coli* and xylose tracers in *Geobacillus* LC 300 respectively can improve flux resolutions for different metabolic pathways, the costs for many of these single tracers, and tracer combinations, can be high (Table 12).

Table 12. Commercially available ^{13}C positionally labelled glucose and xylose tracers. Tracers shown only those readily in stock from suppliers Cambridge Isotope Labs, USA (CIL) and Merck, UK (Merck) without further requested quotes. No price without inquiry is designated by --. *Shown is the lowest listed price per gram on 12/11/2021

^{13}C Tracer	List price (£/g*)
[U-^{13}C] Glucose	£168 (CIL), £220 (Merck),
[1-^{13}C] Glucose	£189 (CIL), £247 (Merck),
[2-^{13}C] Glucose	£460 (CIL), £632 (Merck),
[3-^{13}C] Glucose	£1,132 (CIL), £1,750 (Merck)
[4-^{13}C] Glucose	£1,486 (CIL)
[5-^{13}C] Glucose	£1,592 (CIL)
[6-^{13}C] Glucose	£689 (CIL)
[1,2-^{13}C] Glucose	£664 (CIL), £871 (Merck)
[1,6-^{13}C] Glucose	£1,422 (CIL), £552 (Merck)
[1,2,3-^{13}C] Glucose	£3,143 (CIL)
[U-^{13}C] Xylose	£808 (CIL), -- (Merck)
[1-^{13}C] Xylose	£598 (CIL), -- (Merck)
[2-^{13}C] Xylose	-- (CIL)
[3-^{13}C] Xylose	-- (CIL)
[4-^{13}C] Xylose	-- (CIL)
[5-^{13}C] Xylose	£1,540 (CIL)
[6-^{13}C] Xylose	-- (CIL)
[1,2-^{13}C] Xylose	£952 (CIL)

However, the key advantages of INST-^{13}C-MFA were demonstrated through a comparative ^{13}C-MFA vs. INST-^{13}C-MFA study of 1-^{13}C and U-^{13}C-glucose fed cultures of *Corynebacterium glutamicum* performed by Noack *et al.* (2011) [207]. The determined absolute fluxes differed between the two approaches and, when the resulting data was fit to a corresponding GSMM of *C. glutamicum*, it was found that only the estimated flux distribution from the INST-^{13}C-MFA approach resulted in viable biomass production (i.e., cell growth). This was believed to have resulted from the greater number of measurements of the INST-^{13}C-MFA approach, with the concurrent increase in the accuracy and precision of the overall measurements of the ^{13}C labelled isotopomers translating to the resulting estimated flux distributions [208]. This heightened level of precision could also help to test hypotheses about unusual metabolic network topologies of under-studied organisms with potentially missing or inaccurate reaction annotations, as was identified for *P. thermoglucosidasius* during development of the GSMM. The complete INST-^{13}C-MFA workflow featured in this research is depicted in Figure. 24.

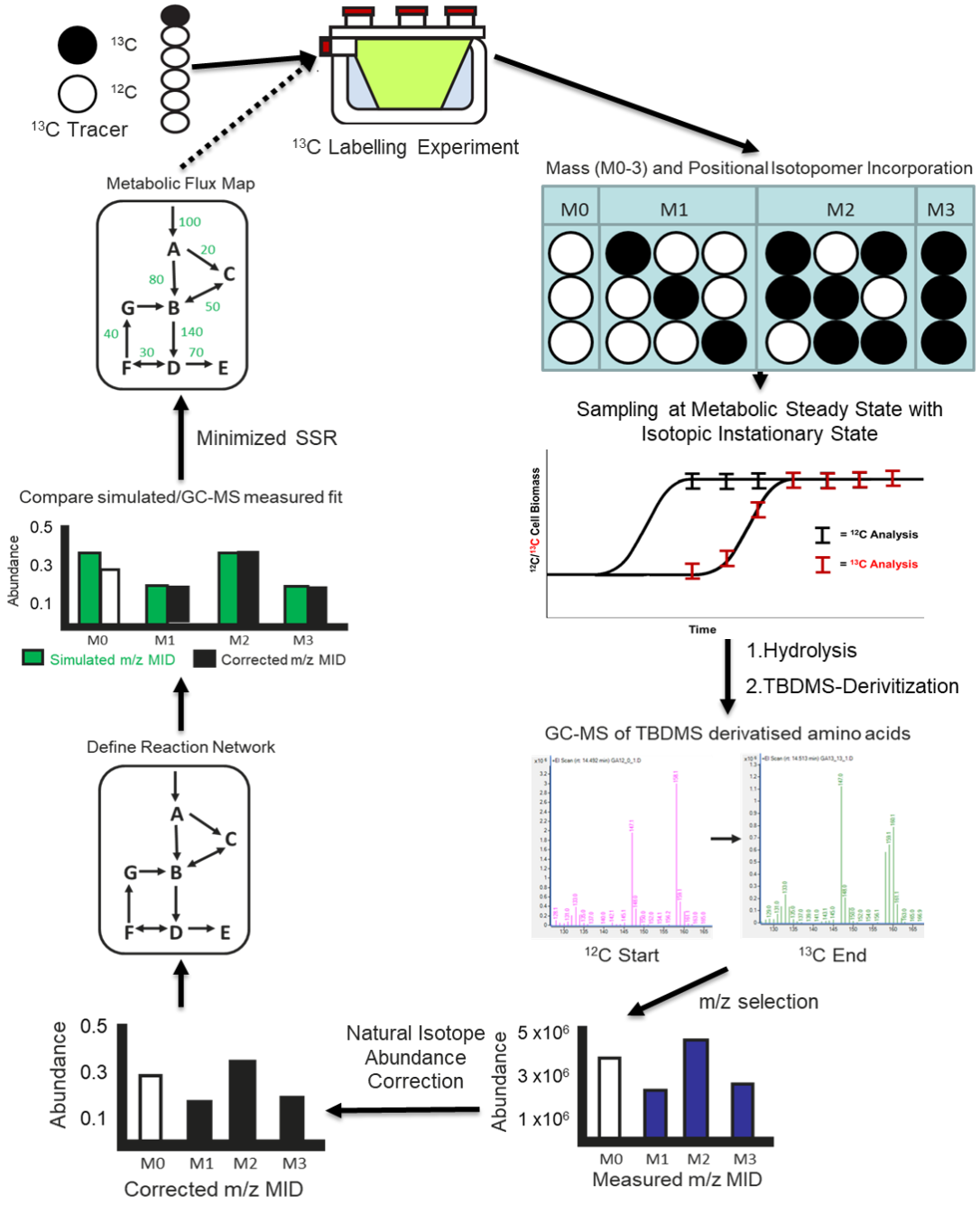


Figure. 24. Schematic of the complete INST-^{13}C-MFA flux determination workflow featured in this research. Minimal, defined media with ^{13}C labelled tracers as sole carbon substrates are continuously fed to a chemostat culture of growing *P. thermoglucosidasius* cells at a fixed dilution rate. The ^{13}C atoms incorporate into metabolic products generating specific mass and positional isotopomers. Regular sampling is then performed to capture the rate of ^{13}C labelling over time. Biomass samples are hydrolysed, and the resulting amino acids are derivitized to TBDMS-derivatives. The TBDMS-derivatised amino acids are then separated and abundances analysed using GC-MS. MIDs are determined for key fragments and natural isotope correction is performed to yield corrected MIDs. For flux analysis a reaction network for the organism must be defined. Minimization of the SSR between the forward simulated and experimentally determined MIDs is then performed to quantify the central metabolic pathways yield a metabolic flux map. These fluxes can then be depicted as metabolic flux maps which can then be used to describe cellular physiology and inform future experiments (represented by the dashed arrow).

4.1.3 Other examples of Geobacillus ^{13}C-MFA derived models

The resulting flux maps from ^{13}C-MFA seek to describe the original fluxes and metabolite balances of cellular metabolism under different phenotypic states. To the author's knowledge there are four separate examples of ^{13}C-MFA isotope labelling experiments describing *Parageobacillus* or *Geobacillus* species, none of which utilised an isotopically instationary form of analysis.

Table 13. Existing ^{13}C-MFA studies of *P. thermoglucosidaisus* and *Geobacillus* species.

ORGANISM	^{13}C CARBON SOURCE/S	^{13}C-MFA OR INST-^{13}C-MFA?	OXYGEN CONDITIONS	REFERENCE
***P. THERMOGLUCOSIDASIUS* M10EXG**	**0.3% D-Glucose [1-^{13}C] or [2-^{13}C]**	^{13}C-MFA	Aerobic, Micro-Aerobic and Anaerobic	Tang *et al.,* (2009)[128]
***P. THERMOGLUCOSIDASIUS* NCIMB 11955**	**0.5% D-Glucose [U-^{13}C] 20:80 [U-^{12}C] and 0.5% D-Glucose [U-^{13}C] 60:40 [U-^{12}C]**	^{13}C-MFA	Aerobic, Micro-Aerobic and Anaerobic	Dr. Charlotte Ward Thesis[121]
***G.* LC300**	**Six parallel labelling experiments of 0.2% singly labelled D-Xylose (i.e. [1-^{13}C], [2-^{13}C], [3-^{13}C], [4-^{13}C] and [5-^{13}C])**	^{13}C-MFA and 'COMPLETE-MFA' (Leighty and Antoniewicz, 2013)	Aerobic	Cordova *et al*., (2016)[127]
	Six parallel labelling experiments of 0.32% singly labelled D-Glucose (i.e. [1-^{13}C], [2-^{13}C], [3-^{13}C], [4-^{13}C], [5-^{13}C] and [6-^{13}C])		Aerobic	Cordova *et al.,* (2017)[170]

This includes models representing the: glucose metabolism of *P. thermoglucosidasius* M10EXG[128] under aerobic, micro-aerobic and anaerobic growth conditions, glucose metabolism of *P. thermoglucosidasius* NCIMB 11955 grown under aerobic and anaerobic conditions[121] and the glucose and xylose metabolism of *Geobacillus* LC300 grown under aerobic conditions[127 170].

4.1.3.1 *P. thermoglucosidasius* M10EXG

The first example of a form of ^{13}C-MFA applied to a *Parageobacillus* species was Tang *et al.* (2009)'s characterisation of *P. thermoglucosidasius* M10EXG[128]. As discussed in Chapter 3, M10EXG was previously determined to have only 29 unique ORFs which could encode metabolic enzymes M10EXG, making this species a relevant comparison.

Tang *et al.* (2009) grew M10EXG cells in minimal growth media at 60°C with shaking at 200 RPM containing either [1-^{13}C] or [2-^{13}C] labelled glucose as a sole substrate. Metabolic fluxes were determined using a combination of ^{13}C ILEs analysed by GC-MS and enzyme assays of activities of specific CCM enzymes. The ^{13}C ILEs assumed a pseudo-metabolic steady state and harvested cells during mid-

exponential growth under aerobic, micro-aerobic and anaerobic conditions. ^{13}C-MFA fitting was performed using the $[M-57]^+$ and $[M-159]^+$ mass fragments from only 7 amino acids (Alanine, Glycine, Leucine, Serine, Phenylalanine, Aspartate and Glutamate) each assumed to represent a particular carbon intermediate node in CCM. Estimations of the confidence intervals of the calculated fluxes was performed using a Monte Carlo method.

From the resulting glucose uptake normalized flux distributions, Tang *et al.* (2009) suggested that aerobic glucose metabolism in M10EXG cells involved a mixture of flux through glycolysis (69) and the oxidative PPP (30 ± 2) in upper metabolism, with high TCA cycle activity (64 ± 3) and high flux through the anapleurotic pyruvate shunt (58 ± 38) and through PEP carboxylase (44 ± 31), with high error values indicating it was difficult to resolve the two pathways.
Micro-aerobic glucose metabolism demonstrated a slight shift towards glycolysis (81) from the oxPPP (19 ± 2) in upper metabolism, but the most pronounced changes were seen in lower metabolism with reduced TCA cycle activity (25 ± 2), comparably higher use of the glyoxylate shunt (20 ± 4), continuing high flux through the pyruvate shunt (37 ± 12) and PEP carboxylase (24 ± 8) and the increased production of fermentation products: lactate (67), formate (13) and ethanol (28). Lastly, these trends continued to anaerobic glucose metabolism with increased flux through glycolysis (84 ± 2) away from the oxidative PPP (15 ± 2), a purely biosynthetic TCA cycle and the increase in flux to fermentation products: acetate (61 ± 10), lactate (89 ± 2), formate (103 ± 14) and ethanol (38 ± 7). Additonally, the authors also did not detect activity for the Entner-Doudoroff pathway, nor any transhydrogenase activity detected by assay from cells under any of the growth conditions.

Additionally, the authors suggested that the, potentially renewable, ethanol production of the cells could be improved through deletion of their acetate and lactate producing capabilities; a strategy demonstrated to be successful by Cripps *et al.* (2009)[89].

A specific limitation of this study is that the isotopomer transition network is both derived from *G. kaustophilus* and the biochemical reactions of that network were set to unidirectional determined from thermodynamic reaction data taken from the reactions of mesophilic *B. subtilis*. While helping to prevent underdetermination of the system and to reduce degrees of freedom for the resulting isotopomer balance equations, this introduces the possibility of ignoring reversible reactions which could be a more accurate description of M10EXG metabolism.

4.1.3.2 *P. thermoglucosidasius* DL33, DL44 and DL66

The most recent example of isotopically steady-state ^{13}C-MFA applied to a *Parageobacillus* species was performed as the focus of a thesis by Dr. Charlotte Ward in 2014 on *P. thermoglucosidasius* NCIMB 11955 termed DL33, to analyse the metabolic differences in CCM flux between this wild-type and two mutants termed DL44 and DL66 when grown on glucose as a sole substrate[121]. The DL33 mutant possessed a lactate dehydrogenase deletion (Δldh) while the DL66 mutant possessed deletions of lactate dehydrogenase and pyruvate formate-lyase, alongside upregulated expression of pyruvate dehydrogenase (Δldh Δpfl↑pdh). Ward grew cells of all strains of *P. thermoglucosidasius* under three different aeration conditions at 66°C in continuous culture mode set to a dilution rate of 0.1 h^{-1} for 3 CCVCs. These aeration conditions were defined as: 0.5 VVM (22.5 mL/min) for fermentative growth, 1.5 VVM (67.5 mL/min) for microaerobic growth and 12 VVM (540 mL/min) for aerobic growth.

While initial attempts to resolve flux distributions using a 20:80 [U-^{13}C] glucose to naturally labelled [U-^{12}C] glucose feed ratio could only be resolved to a statistically acceptable 95% confidence interval with the PPP removed, repeat experiments using a 60:40 [1-^{13}C] glucose to [U-^{13}C] glucose feed ratio combined with a Monte Carlo simulation approach to minimize SSR, enabled Ward to achieve statistically acceptable metabolic flux distribution solutions for the full CCM of the three studied strains. In these successful instances, fitting was achieved for all strains using the same set of 15 different amino acid fragments and the [M-57]+ fragment of malate.

The key focus of the 60:40 [1-^{13}C] glucose to [U-^{13}C] glucose feed ratio results was the contrasting growth under fermentative conditions. The resulting flux distributions for the three strains showed similar activity in upper metabolism with the majority of flux travelling through the glycolytic path for DL33 (196 ± 1), DL44 (196 ± 1) and DL66 (195 ± 80) with correspondingly low flux through the oxidative PPP for DL33 (15 ± 2), DL44 (15 ± 3) and DL66 (16 ± 2). However, the low flux through the PPP for all three strains could have been a result of the reactions of the typically reversible non-oxidative branch of the PPP being constrained to unidirectional flow in the forward direction and represents one possible limitation of this network.

Much greater variation was observed in the calculated fluxes for fermentation pathways, the TCA cycle and the anaplerotic reactions between the strains. Under fermentative conditions, the wild-type DL33 strain produced fermentation products acetate (29 ± 0), lactate (98 ± 0) and ethanol (32 ± 3) and a TCA cycle flux of 34 ± 3. Because of the deletion of the ldh gene, DL44 produced more acetate (44 ± 0) and ethanol (89 ± 2), but no lactate and with a lower TCA cycle flux of (18 ± 1.2). With the additional deletion of the pfl gene and upregulation of expression of the pdh gene, DL66 produced no lactate and slightly less acetate (39 ± 0), but even more ethanol (123 ± 4) with a TCA cycle flux slightly higher than DL33 (21 ± 2).

Notably in terms of fermentation products, Ward also detected concentrations of external alanine for all three *P. thermoglucosidasius* strain cultures grown under micro-aerobic and fermentative conditions. This alanine export was hypothesised to regenerate pools of NAD+ using an NADH dependent dehydrogenase in the synthesis of alanine. Under fermentative conditions this was observed as 6 ± 1 for DL33, 34 ± 0 for DL44 and 15 ± 4 for DL66. A further notable difference was observed between the strains for the single reaction representing the anaplerotic reactions between the two combined pools of phosphenolpyruvate-pyruvate and malate-oxaloacetate. In the wild-type DL33 strain flux went towards the phosphenolpyruvate-pyruvate pool from the TCA cycle (8 ± 2) whereas in the DL44

and DL66 strains this directionality reversed towards the malate-oxaloacetate pool (3 ± 1, 21 ± 2). However, given the simplifying of these metabolic pools and reactions, it is difficult to tell to what extent the change is a result of changes in flux of both enzymes or the changes in fluxes of single enzyme which determines the net direction of the singular reaction.

Additionally, based on the genome sequencing and flux analysis results observed by Ward, the reaction for a reversible NADP+ dependent glyceraldehyde-3-phosphate dehydrogenase (GAP (abc) <-> 3PG (abc) + ATP + NADPH) was included in this atom transition network for all conditions and excretion of alanine as a fermentation product was considered for the microaerobic and anaerobic maps of this research.

Despite the unparalleled detail of steady-state CCM fluxes of *P. thermoglucosidasius* NCIMB 11955 and two mutants under a broad range of conditions, this research did have 2 specific limitations. The first of these was inclusion of supplementary: L-glutamate (953 μM), L-serine (428 μM) and L-threonine (336 μM) in the growth media.
Although Ward determined that only removal of serine impacted upon the production of biomass and ethanol under fermentative conditions, the inclusion of these amino acids may have affected the overall homeostasis of the network. Indeed, the network itself lacks a reaction for glutamate synthesis as it was assumed glutamate would be provided solely from uptake.
Secondly, while a potential typographic error not reflected in the results, the communicated version of the atom transition network includes a reaction (BM_oaa7) which creates an additional carbon from the catabolism of isoleucine to succinate: “lLE (#ABCDEF) + CX (#a) -> SUC (#**A**BCDa) + CO2 (#**A**) + AcCoA (#EF)”. If accurate, this would affect the validity of the estimated flux values.

4.1.3.3 *Geobacillus* spp. LC300

The metabolic behaviours of *Geobacillus* spp*.* LC300 on xylose[127] and glucose [170] as sole substrates were identified in successive publications through a simultaneous parallel tracer ^{13}C-MFA strategy termed ‘^{13}C-MFA COMPLETE-MFA’.

In the determination of the metabolic fluxes of *Geobacillus* spp*.* LC300 on xylose, Cordova and Antoniewicz (2016) grew LC300 cells in batch mode at 72°C with 11.3 mL/min air in 5 parallel growth experiments with each experiment using 2 g/L (0.2%) of one singly positionally labelled xylose tracer of [1-^{13}C], [2-^{13}C], [3-^{13}C], [4-^{13}C], and [5-^{13}C] as the sole substrate[127]. Later, as part of a broader investigation into the metabolic similarities and differences of thermophilic metabolisms, the determination of the metabolic fluxes of Geobacillus *spp.* LC300 was performed by Cordova *et al*., (2017)[170] under the same growth conditions. However, cells were instead grown on 3.2 g/L (0.32%) of one singly positionally labelled glucose tracer of [1-^{13}C], [2-^{13}C], [3-^{13}C], [4-^{13}C], [5-^{13}C], and [6-^{13}C] as the sole substrate.

In both cases, the ^{13}C xylose and glucose ILEs assumed a pseudo-metabolic steady state the cells were harvested during what was determined to be mid-exponential growth. The metabolic flux distributions were estimated for each of the individual ^{13}C-labelled xylose and glucose experiments, and the combined xylose and glucose data sets, through a minimization of the SSR between the forward simulated and experimentally GC-MS measured MIDs of 24 (xylose) and 25 (glucose) different amino acid fragments and CO_2 using an in-line gas analyser. For the combined data set ‘COMPLETE-MFA’ approach, all experimentally determined xylose and glucose tracer data sets were simultaneously fit to single separately forward simulated MID

models. Statistically acceptable SSR values (95% confidence interval) were obtained for all xylose and glucose data sets, under the assumption of universal 0.4 mol. % measurement errors.

For both the xylose and glucose flux maps of *Geobacillus* sp. LC300, flux values were normalised to the measured xylose or glucose uptake rate set to a value of 100.
The flux distribution of LC300 grown on xylose demonstrated high non-oxidative PPP flux (71 ± 0.5) from xylose entering the network as xylulose-5-phosphate, limited oxidative PPP flux (16 ± 0.5), high flux through glycolysis (136 ± 0.4), and the TCA cycle (69 ± 1), with moderate contributions from the anaplerotic reaction of pyruvate to oxaloacetate (40 ± 2) and very low flux through the glyoxylate (3 ± 1) and ED pathways.
Alternatively, the flux distribution of LC300 grown on glucose demonstrated moderate oxidative PPP flux (37 ± 0) and limited non-oxidative PPP flux (17 ± 0) due to glucose entering the network as glucose-6-phosphate. While a similarly high glycolytic flux (154 ± 1) and anaplerotic flux from pyruvate to oxaloacetate (32 ± 2) was observed, the observed TCA cycle flux was lower (31 ± 1) due to production of acetate (38 ± 2). Similarly, low flux was also determined for the glyoxylate (1 ± 1) and ED (0 ± 0) pathways.

Both papers make informed suggestions about LC300 metabolism which may translate to *P. thermoglucosidasius*. In the paper concerning LC300 growth on xylose, the authors conclude based on their ^{13}C labelled CO_2 data that carbon loss in upper metabolism is due to low flux (16 ± 0.5) through the oxidative PPP (which loses the C1 carbon of G6P as CO_2) rather than the ribulose monophosphate RUMP pathway (which loses the same C1 carbon of F6P as formaldehyde) which had a negligible flux of (1 ± 1). In the paper concerning LC300 growth on glucose, the authors suggest that the lack of observed flux from Threonine to Glycine and from pyruvate and Acetyl-CoA to Isoleucine is further evidence to support the lack of a respective Threonine Aldolase or Citramalate Synthase, genes for which are also not found in the LC300 genome. Furthermore, oxidative PPP activity in both models suggested that the lack of a gene for 6-phosphogluconolactonase did not prevent flux from travelling down the oxidative PPP, with the authors citing the frequently held hypothesis that the reaction is spontaneous at the high growth temperatures of LC300. Additionally, the authors noted that there was no single superior xylose tracer that could resolve the entire LC300 metabolic network, but that certain tracers were optimal for particular metabolic pathways. For example, they found that [2-^{13}C] xylose was the best tracer for resolving the fluxes of the glycolytic and pentose phosphate pathways whereas [5-^{13}C] xylose was the best tracer for resolving the TCA cycle and lower metabolism.

Despite the level of precision, both LC300 studies share a specific and notable limitation. In both cases, the ^{13}C labelled media included 0.05 g/L of yeast extract, supposedly to overcome an initial sub-culturing lag phase which could presumably affect the pseudo-metabolic steady state assumption. Yeast extract represents a variable source of amino acids, nitrogen, vitamins and additional carbon substrates which would introduce an unknown amount of naturally labelled carbon to the culture. This could introduce disagreement between the forward simulated isotopomer network and the experimentally measured MIDs and potentially affect the overall validity of the otherwise accurately and reliably obtained flux distributions.

4.1.4 General limitations

While addressed to some extent in the research by Ward through a focus on fermentation products, a key limitation of these studies is that of the somewhat

arbitrarily defined growth conditions of the experiments themselves. ‘Aerobic’ was defined by Tang *et al.* (2009) simply as growth at 60°C in a shake flask at 220 RPM, ‘micro-aerobic’ was defined as a stoppered flask with a 1:1 ait-to-liquid ratio and ‘anaerobic’ a stoppered shake flask with a head-space of argon. Alternatively, aerobic was defined by Cordova *et al.* (2016 and 2017) for LC300 as growth at 72°C with 11.3 mL/min air flow.

Under these ill-defined conditions it is difficult to distinguish to what extent the observed flux values are a result of actual metabolic behaviour or the growth conditions. Furthermore, by not accurately defining the growth conditions, it is unclear how similar they actually are in terms of oxygen concentration and therefore obfuscates the validity of comparisons between them. For example, a notable yield of acetate was detected in the aerobic exponential growth experiments of *P. thermoglucosidasius* M10EXG (0.64 ± 0.12) and G. sp. LC300 on glucose (0.38 ± 0.03 mol of acetate produced per mol of glucose consumed) but not on xylose. While this could be explained as a by-product response to rapid growth on an excess of substrate, acetate is also an initial fermentation product seen for micro-aerobic and anaerobic M10EXG (40 ± 0.5 and 61 ± 10 respectively) and micro-aerobic and fermentative NCIMB 11955 too (59 ± 0 and 29 ± 0 respectively). Ultimately, the lack of aerobic distinction affects the accuracy of the overall observations.

4.2 INST-^{13}C-MFA specific aims and objectives:

With the publishing of the *p-thermo* GSMM, we now have a more accurate computational model of *P. thermoglucosidasius* NCIMB 11955 able to generate qualitative, but not quantitative, flux distribution predictions about its aerobic and fermentative metabolic behaviours through FBA. However, a key limitation of such *in silico* GSMMs is that their reaction-level stoichiometric matrix foundations do not typically account for the gene and flux regulation that control the abundances, rates and stabilities of catalytic enzymes for particular reactions [184]. While this information could be encoded using various assumptions from more model organisms, to understand how the global changes in gene expression can control metabolic flux redistributions in the non-model *P. thermoglucosidasius* in response to changing conditions, there is a need to quantitatively determine these metabolic fluxes to experimentally validate the *p-thermo* GSMM [198]. Notably, a commonly cited prerequisite to using strain engineering approaches, such as MOMA, to suggest ME designs is the experimental validation of any GSMM generated flux maps [44].

Influenced by the experimental designs of the existing ^{13}C-MFA research into *Parageobacillus* and *Geobacillus* species, the aim of the fluxomics research was to be the first to use an INST-^{13}C-MFA strategy to describe the metabolic fluxes through the central carbon metabolism of *P. thermoglucosidasius* NCIMB 11955 at the level of the phenotype for target specific growth rates, degrees of aerobic and anaerobic behaviour and on glucose- or xylose-limited chemostat cultures. This was done to contribute to the greater overall goal of further validating and improving the *p-thermo* model as a tool for SysME strain design and provide metabolic insights that could also be used to inform future metabolic engineering strategies.

The specific objectives of the fluxomics research can be found as follows:

- Chapter 4.3 details the validation of chemostat systems to establish the INST-^{13}C-MFA workflow at the University of Bath and the method development involved in obtaining metabolic steady-state conditions which accurately reflected *P. thermoglucosidasius* phenotypes under defined redox potential conditions.

- Chapter 4.4.1 details how the INST-^{13}C-MFA workflow was used to observe the resulting ^{13}C labelling patterns of amino acid fragments and make qualitative observations about central carbon metabolism fluxes of *P. thermoglucosidasius* NCIMB 11955.

- Chapter 4.4.3 details how the INST-^{13}C-MFA workflow was used to improve quantitation of *in vivo*, central carbon metabolism fluxes of *P. thermoglucosidasius* NCIMB 11955.

4.3 Method development: Establishing the microbioreactor systems and validating the conditions for continuous growth under the target aerobic and anaerobic conditions

Before the INST-^{13}C-MFA experiments could be achieved, the workflow to perform accurate INST-^{13}C-MFA experiments in the University of Bath research group needed to be established.

As discussed in 4.1.1.2.4, stable continuous cultures of *P. thermoglucosidasius* operating as chemostats are one method to achieve the metabolic steady-state for ^{13}C-MFA experiments that avoids the need to assume a pseudo-metabolic steady state from a batch culture. The first stage in this research therefore was to construct and evaluate these bioreactor systems for their ability to support continuous cultures of *P. thermoglucosidasius*. The second of these stages was to define sets of parameters that would enable the continuous culture of *P. thermoglucosidasius* under different metabolic steady states. Both stages relied on the use of HPLC-RI-UV to quantify the concentrations of sugars in sampled cell culture filtrates, indicating whether cellular growth is substrate limited and thus growing at a rate equivalent to the set dilution rate, and any fermentation or overflow metabolites which are indicative of aerobic or anaerobic metabolic behaviours.

Chapter 4.3 details the overall method development process, and initial ^{13}C labelled sugar continuous culture runs, which led to the final and complete INST-^{13}C-MFA experiments.

4.3.1 Defining aerobic, anaerobic, and micro-aerobic growth conditions through redox potential and HPLC analysis

While the cellular growth rates and the choice of carbon substrate and concentration were easily quantifiable, to directly address some of the limitations around potentially inaccurate definitions of oxygen environments in previous research the definitions of the ‘aerobic’ and ‘anaerobic’ growth conditions needed to be qualified against the metabolic behaviours of *P. thermoglucosidasius* that were of most interest.

Fundamentally, an aerobic culture will contain dissolved oxygen, while an anaerobic culture will not contain any dissolved oxygen. How aerobic or anaerobic a culture can be dictated by the concentration of the dissolved oxygen in the media and what concentration of that dissolved oxygen is available to an individual cell in the culture. When growing cultures using a bioreactor system, the concentration of dissolved oxygen is dictated by four main factors. The first is the flow rate of oxygen or air itself, which controls the total possible concentration of oxygen. The second of these is the sparger. Different models of sparger allow a tuning of the bubble size of the pumped oxygen which control the surface area ratio of gas to water on which oxygen can dissolve. The third of these is the programmable agitator. Agitators can have varying blade sizes, shapes, heights and spacing within a culture and total numbers of blades. These agitiators can also be set to different rotation speeds and agitation modes. Combined, this enables control over the rate of gas bubble mixing and hence rate of oxygen dissolving. The fourth and final factor are particular chemical components in the media itself which can affect the ability of oxygen to dissolve.

The concentration of dissolved oxygen in a culture is typically measured by a DO_2 tension probe [209]. However, due to the detection limits of the probe, the percentage oxygenation (pO_2) measurements may only indicate a relative concentration of dissolved oxygen and thus can fail to translate to a description of the metabolic behaviour of a single cell in a culture of cells. This is particularly relevant for the low

oxygen concentration environments that are of importance to potentially understanding *P. thermoglucosidasius* metabolism under truly anoxic conditions.

A potentially better descriptor of an individual cell's metabolic behaviours under these low oxygen conditions can be the net ORP of the extracellular culture, measured by an ORP electrode in millivolts (mV). While separated from the intracellular redox state of a cell, the extracellular redox potential does correlate to metabolic activity. For example, during respiration cells producing electron-donating NADH from both intracellular anabolic and catabolic processes can consume dissolved oxygen in the media to act as a terminal electron acceptor. As a result, the measured extracellular ORP of the media is an inter-dependent relationship between the concentration of dissolved oxygen and the intracellular metabolic activity of the overall cell population [210]. This will ultimately stabilise to a fixed ORP value under chemostat conditions. High overall metabolic activity and/or low dissolved oxygen concentrations will therefore result in a low ORP value, whereas a high dissolved oxygen concentration and/or low overall metabolic activity will result in a more positive ORP value.

In terms of defining 'anaerobic' growth of *P. thermoglucosidasius*, the metabolic behaviour of interest was the production and secretion of the fermentation metabolite ethanol, produced by alcohol dehydrogenase from Acetyl-CoA, even if this meant supplying small amount of oxygen to the media. This had been determined previously to start being detectable in *P. thermoglucosidasius* cultures at a redox potential of ~ -260 to -280 mV (Jinghui Liang, 2021 pers comm.). 'Micro-aerobic' growth was defined as the metabolic behaviour of *P. thermoglucosidasius* whereby it produces and exports the industrially relevant overflow metabolite of lactic acid, detectable in cultures at a redox potential of ~ -200 mV.

Under aerobic conditions, it is assumed the only growth-limiting factor in a chemostat is the carbon substrate concentration in the media. Therefore, if use of HPLC analysis of the cell culture detects that the concentration of the carbon substrate in the cell culture filtrate is zero, it can be assumed that the growth rate of the organism in the culture is equal to the set dilution rate of the media. HPLC detection of overflow metabolites or typical fermentation products in the CCF therefore suggests that the cells are also limited in the amount of oxygen they can take up. Thus, 'aerobic' growth of *P. thermoglucosidasius* was defined as a metabolic behaviour which does not produce detectable concentration of overflow carbon metabolites, such as pyruvate or acetate, which could be indicative of suboptimal throughput of the citric acid cycle in several common microbial chassis [211].

In order to achieve metabolic steady-states for these different aeration conditions, the choice of growth vessel for the INST-^{13}C-MFA experiments was small volume CSTRs. These bioreactors can support comparatively higher cell densities than shake-flasks, due to additional pH and dissolved oxygen tension control, while critically also minimizing the volume of media and hence use of costly ^{13}C-labelled sugar required to achieve isotopic steady state [212].

4.3.2 Initial Vessel Pilots: MKI vs MKII Micro-Bioreactor construction and *P. thermoglucosidasius* pilot growth experiments Pilot screening experiments

The earliest pilot screening experiments performed sought to broadly evaluate the two custom CSTRs, the 40 mL working volume 'MKI' and the 90 mL working volume 'MKII', which were assembled to function as chemostats from available components. The detailed description and general operation of these bioreactor systems is provided in 2.1.2.1

The MKI and MKII systems were evaluated for their abilities to achieve the target aerobic and micro-aerobic *P. thermoglucosidasius* growth conditions through the monitoring of the redox potential of cultures and quantification CCF metabolite and glucose concentrations by HPLC.

4.3.2.1 MKI

The 40 mL working volume MKI microbioreactor system was the first to be assembled (Figure. 25B) and evaluated. *P. thermoglucosidasius* NCIMB 11955 cells were grown on 1% glucose ASM, at 60°C, pH 7 and with a constant 3.33 VVM air flow rate. The magnetic flea was set to a maximum stable arbitrary rotation rate of '8.0' on the magnetic stirrer for two separate experiments fed at media dilution rates of 0.04 h^{-1} and 0.06 h^{-1}. Each culture was sampled after 3 CCVCs, which was assumed to be approaching or equal to metabolic steady state according to Dr. Ward[121], and the compound concentrations in the CCFs were analysed by HPLC-RID-UVD (Figure. 25A). The near absence of glucose detected in CCFs of both experiments suggests that the cells were growing under carbon-limited chemostat conditions and thus it can be assumed that the respective media dilution rates were equivalent to the cell growth rates.

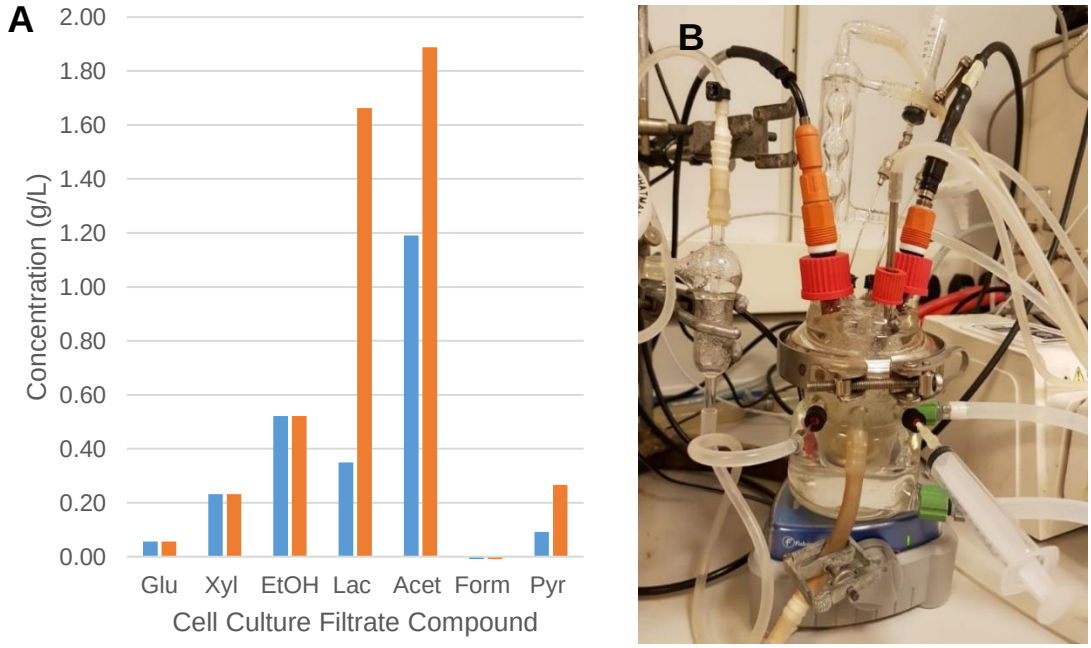


Figure 25: MKI microbioreactor *P. thermoglucosidasius* chemostat growth condition evaluation. A) HPLC-RI-UVD analysed products at 0.04 h^{-1} (blue) and 0.06 h^{-1} (orange) media dilution rates. n = 1 biological and technical replicate. **B)** *P. thermoglucosidasius* cells growing as a chemostat culture in the MKI microbioreactor.

Dilution rate (h^{-1})	0.04	0.06
O.D.600	2.52	2.91
Redox Potential (mV)	-237	-226

Table 14: Accompanying $O.D_{600}$ and redox potential readings for MKI *P. thermoglucosidasius* chemostat growth experiments at 0.04h^{-1} and 0.06h^{-1} media dilution rates.

In the CCF samples, HPLC-UVD at 214nm suggested the presence of 0.52 g/L ethanol, 0.35 g/L and 1.66 g/L lactic acid and 1.19 g/L and 1.89 g/L acetate in the respective 0.04 h^{-1} and 0.06 h^{-1} dilution rate growth experiments (Figure 25A). Furthermore, the culture redox potential values for the experiments were read to have achieved -237 mV at a media dilution rate of 0.04 h^{-1} and -226 mV at a dilution rate of 0.06 h^{-1} (Table 14). Combined, these values suggested that the MKI system could be effective for future micro-aerobic growth experiments.

4.3.2.2 MKII

As the maximum stirring speed could not achieve aerobic growth conditions in the MKI reactor could not be achieved, the MKII 90 mL working volume reactor was constructed which incorporated a digital programmable rotor (Figure. 26B).

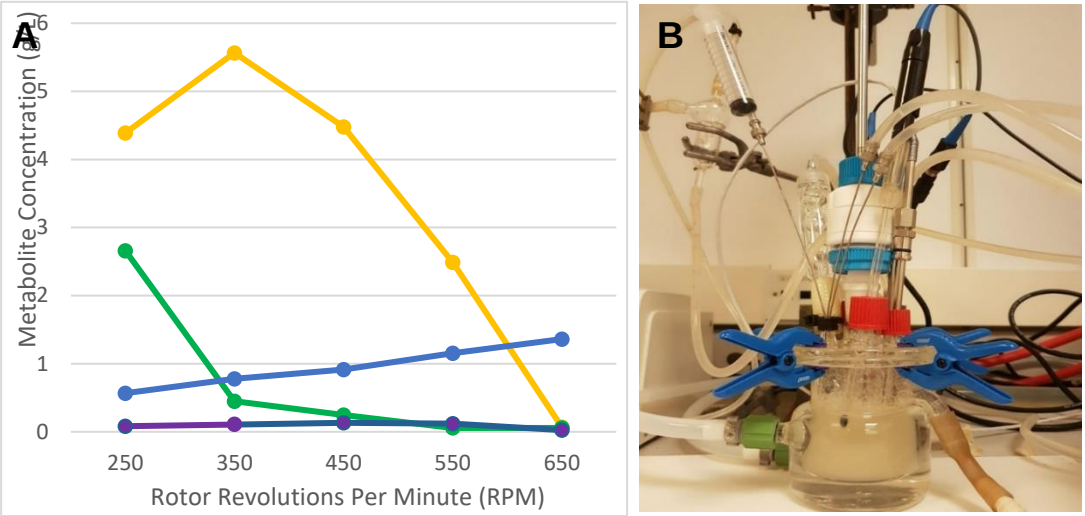


Figure 26. MKII microbioreactor *P. thermoglucosidasius* chemostat growth condition evaluation. (**A**) HPLC-RI-UVD analysed CCF products at 250-650 rotor RPM featuring Lactate (yellow), Glucose (green), Acetate (blue) and Pyruvate (purple). n = 1 biological and technical replicate. (**B**) *P. thermoglucosidasius* cells growing as a chemostat culture in the MKII microbioreactor.

P. thermoglucosidasius NCIMB 11955 cells were grown at 60°C, pH 7, 1 VVM air on 1% glucose ASM and at a dilution rate of 0.1 h^{-1}. In this pilot screening experiment, the continuous cultures of *P. thermoglucosidasius* had the agitator speed for the programmed rotor varied from 250-650 RPM, allowing one cell culture volume change (10 hours) before OD_{600} measurements and cell culture samples were taken for HPLC analysis. At this point the agitator RPM was changed to the next in sequence.

Over the 40-hour experiment, the OD_{600} of the culture rose from 1.86 at an agitator speed of 250 RPM to 3.93 at an agitator speed of 650 RPM while the redox potential of the culture increased from -211 mV at 250 RPM to -189 mV at 650 RPM (Table 15).

Table 15. Accompanying $O.D_{600}$ and redox potential readings for MKII *P. thermoglucosidasius* chemostat growth experiments grown at agitator RPM values of 250-650.

Agitator RPM	250	350	450	550	650
O.D.600	1.86	1.89	2.19	3.34	3.93
Redox Potential (mV)	-211	-208	-194	-200	-186

Concurrently, despite an initial increase in lactate from 4.39 g/L to 5.57 g/L between agitator speeds of 250 and 350 RPM, the concentration of lactate in the CCFs then dropped from 5.57 g/L at 350 rotor RPM to below the limit of detection at an agitator speed of 650 RPM (Figure 26A).

Discussion

The MKI pilot experiments demonstrated that the CSTR system had promise as a basis for micro-aerobic and anaerobic experiments through achieving measured redox potential values of -237 mV and ethanol concentrations of 0.52 g/L. However, the observation of apparently ethanologenic conditions were at a high air flow rate of 3.33 VVM, suggested that continuous aerobic growth conditions would not be feasible in the MKI system, which then prompted evaluation of the MKII system. Indeed, attempts to increase magnetic stirrer bar RPM to enhance mass transfer of oxygen to the media hit a limit. Beyond a value of 8.2 on the analogue dial of the magnetic stirrer plate, the stir bar would become erratic, inconsistent, and prone to failure. This was further exacerbated by the ring-type sparger used, which restricted available stirring space on the base of the vessel.

The MKI was considered for anaerobic growth experiments, but the stirring controlled by magnetic flea was considered too unreliable. To develop the MKI for more reliable anaerobic experiments, a 'MKIII' top-plate design combining the fully seal-able, bevel/concave with an additional port for a programmable agitator as in the MKII was designed. Designs were discussed with the manufacturer of the MKI and MKII. However, due to costs, lead time and then the collapse of the company, this was disregarded as an anaerobic growth strategy. Future fluxomics experiments in general should consider the design of a CSTR vessel able to run in continuous mode and use the ≤45 mL working volume of the MKI, to minimize the costs associated with the heavy isotope tracers, but with a form of programmable agitation alongside temperature, pH and aeration control. Further reliability of the results obtained could also be improved if such as a system was able to operate multiple growth vessels simultaneously.
As a result, and due to time constraints, the anaerobic growth experiments were instead eventually performed in the 350 mL minimum working volume Biostat B bioreactor system, which had a proven internal record of achieving anaerobic ethanologenic cultures of *P. thermoglucosidasius*.

For the MKII, the combination of the *P. thermoglucosidasius* culture redox potential value of -186 mV and an $O.D._{600}$ value of 3.93, under a HPLC-RID suggested glucose limitation on 1% glucose ASM, suggested that the parameters of 1 VVM air flow rate, an agitator speed of 650 RPM, 60°C and pH 7 could support 0.1 h^{-1} carbon-limited micro-aerobic growth ^{13}C-MFA experiments and could be iterated upon to generate aerobic growth conditions. In retrospect it is likely that the formative vessel pilot experiments focused on defining growth conditions did not accurately represent *P. thermoglucosidasius* cells at a metabolic steady state, or potentially under their target aeration conditions, at any agitator RPM as only a single culture volume change occurred between each RPM condition. However, based on being able to achieve lactic acid free microaerobic conditions with reliable mixing due to the programmable rotor, the MKII was selected for the development of aerobic ^{13}C MFA experiments at the University of Bath. Future work could investigate whether different agitator stirring speeds could be used as an alternate form of redox control besides air flow rate.

4.3.3 The development of aerobic and anaerobic growth conditions and initial glucose ^{13}C-MFA isotope labelling experiments (ILEs)

With CSTRs for stable continuous cultures of *P. thermoglucosidasius* established at the University of Bath, attention then turned to the ^{13}C-MFA workflow itself. As there were no remaining researchers with experience of MFA experiments, attention turned to the international partner of the project, Dr. Shyam Masakapalli. A former research group member with ^{13}C-MFA experience, he suggested some initial

research on this project be performed in his research group at IIT Mandi, which had a fully established ^{13}C-MFA workflow. This workflow formed the basis of the INST-^{13}C-MFA methodology detailed in 2.2.1.

Notably, in order to better achieve aerobic conditions and reduce experimental costs, the concentration of ^{13}C D-glucose in the ASM was reduced from 1% w/v to 0.5% and the VVM was increased to 3. Combined these factors increase the concentration of dissolved oxygen available to an individual cell, both by increasing the concentration of flowing oxygen itself through agitation-driven gas-liquid mixing, but also by limiting the potential size of the cell population. At IIT Mandi, rather than use redox potential, the concentration of oxygen was measured by a DO_2 probe.

4.3.3.1 Initial aerobic *P. thermoglucosidasisus* NCIMB 11955 INST-^{13}C-MFA ILE

The first glucose-based aerobic INST-^{13}C-MFA ILE was performed at IIT Mandi to experience the full ^{13}C-MFA workflow. At IIT Mandi, a Applikon® Biotechnology MiniBio2 250 bioreactor system modified to run in continuous culture (detailed in 2.1.2.3), was used to grow *P. thermoglucosidasius* NCIMB 11955 cells on 0.5% D-glucose (60% 1-^{13}C, 40% U-^{13}C) ASM, at 60°C, pH 7, a constant 3 VVM air flow rate, an agitator speed of 1000 RPM, and a media dilution rate of 0.075 h^{-1}.

Over the measured 65 hours of the time-course experiment, the OD_{600} and dissolved oxygen values remained consistent with respective mean values of 2.78 (± 0.16) and 1.35 % (± 0.21%).

Table 16: OD_{600} values and DO_2 percentages for the initial 65-hour GA0.075 glucose *P. thermoglucosidasius* INST-^{13}C-MFA experiment identifying the time point the switch was made to ^{13}C labelled ASM.

^{12}C/^{13}C Glucose	^{12}C	^{13}C													
Time (Hours)	-16.1	0	0.5	1.5	2.5	4.5	6.5	10.5	22.5	26.5	30.5	34.5	45.5	46.5	48.6
OD_{600}	2.84	2.91	2.98	2.95	2.38	2.36	2.73	2.78	2.68	2.84	2.53	2.64	3.22	2.66	2.67
DO_2 (%)	1.3	1.4	1.4	1.6	1.2	1.2	1.1	1.8	1.1	1.2	1.2	1.4	1.6	1.4	1.2

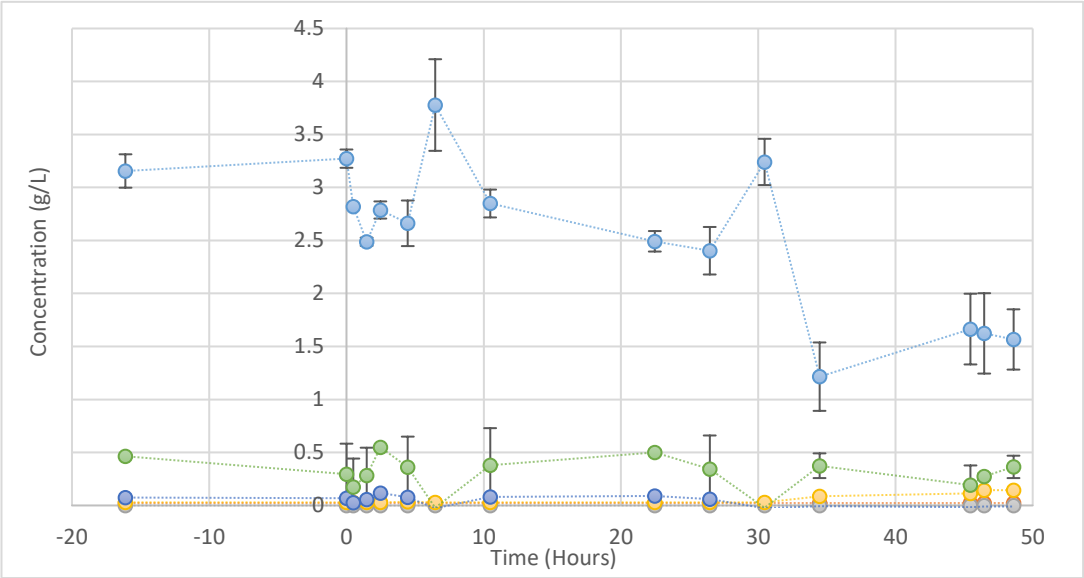


Figure 27: HPLC-RID-DAD analysed CCF products for the initial *P. thermoglucosidasius* GA0.075 INST-^{13}C-MFA experiment. Glucose (light blue), Lactate (orange), Ethanol (grey), Pyruvate (yellow), Formate (dark blue) and Acetate (green). Error is SD for technical triplicate.

Notably however, HPLC-RID analysis of the CCFs identified a decreasing presence of residual glucose, starting at 3.27 g/L (±0.09, n = 3) at the introductory point of the ^{13}C-labelled media and ending at 1.57 g/L (± 0.28, n = 3) by the final time point (Figure. 27). Concurrently, HPLC-DAD analysis of the CCFs identified a low but consistent, average concentration of acetate across the ^{13}C time points of the experiment of 0.30 g/L (± 0.15, n = 13), as well as a gradual increase in the concentration of pyruvate between hours 34.5 to 48.6 of 0.09 g/L (± 0.05, n = 3) to 0.14 g/L (± 0.02, n = 3).

Discussion

Although the OD_{600} values suggest a relatively stable cell population, the presence of excess residual glucose, low dissolved oxygen concentration readings and detectable fermentation productions of *P. thermoglucosidasius* suggests that the cells were not carbon-limited and may have been under oxygen limited conditions. This therefore suggested that the metabolic behaviours pf the cells may not accurately reflect a fully aerobic metabolism and would need future repetition. However, in addition to experience of a downstream ^{13}C-MFA workflow, this experiment suggested future attempts to develop aerobic *P. thermoglucosidasius* culturing conditions should reduce the working sugar concentration of the ASM to 0.5% w/v and increase the air flow to the cultures.

4.3.3.2 Initial *P. thermoglucosidasisus* NCIMB 11955 INST-^{13}C-MFA GA0.15 GA0.225 ILEs

Upon returning to the MKII system at the University of Bath, the practice of 0.5% w/v glucose ASM was implemented, and focus was placed on increasing air flow rate to the culture. A new external air flow meter was installed to the Braun unit enabling measurable air flow rates up to 150 mL/min. Employing a 120 mL/min or 1.3˙VVM air flow rate, *P. thermoglucosidasius* cells were cultured in two separate experiments at 60°C and a pH 7 on 0.5% w/v D-glucose (60% 1-^{13}C, 40% U-^{13}C) ASM with an agitator RPM of 650. In one experiment, the dilution rate was 0.15 h^{-1} and in the second it was 0.225 h^{-1}.

GA0.15 – Glucose ASM Aerobic 0.15 h^{-1} dilution rate ILE

For the first aerobic ^{13}C glucose ASM, 0.15 h^{-1} dilution rate ILE (GA0.15), the redox potential values of the culture varied between -44 mV and -62 mV, with an average value of -51 mV (± 3.90 mV). The OD_{600} values of the culture declined from 4.64 across the duration of the ^{13}C portion of the continuous culture experiment to 3.01, representing a 35.13% decrease. (Table 17).

Table 17: OD_{600}, DCW and redox values (mV) for the initial GA0.15 aerobic glucose *P. thermoglucosidasius* INST-^{13}C-MFA experiment identifying the time point the switch was made to ^{13}C labelled ASM.

Sample	12C 8	12C 9	12C 10	13C 1	13C 2	13C 3	13C 4	13C 5	13C 6	13C 7	13C 8	13C 9	13C 10	13C 11	13C 12	13C 13	13C 14	13C 15	13C 16	13C 17	13C 18	13C 19
Time (hours)	-3.97	-2.97	-0.97	0.00	1.25	1.58	2.05	2.68	4.35	5.02	6.07	7.02	8.02	13.02	16.02	17.68	19.18	20.85	22.52	24.52	25.52	25.52
Cell culture volume changes (CCVCs)	-0.60	-0.45	-0.15	0.00	0.19	0.24	0.31	0.41	0.66	0.76	0.92	1.06	1.21	1.97	2.42	2.67	2.90	3.15	3.40	3.70	3.86	3.86
OD.600	5.30	5.31	4.64	4.51	4.54	4.07	4.36	4.25	4.11	3.95	3.65	3.66	3.63	3.46	3.42	3.46	3.81	4.01	3.54	3.20	3.01	3.01
DCW (g/L)	2.25	2.25	1.97	1.91	1.92	1.73	1.85	1.80	1.74	1.67	1.55	1.55	1.54	1.47	1.45	1.47	1.62	1.70	1.50	1.36	1.28	1.28
Redox (miliVolts)	-52	-52	-54	-52	-53	-56	-49	-51	-53	-51	-55	-54	-52	-52	-52	-47	-48	-47	-44	-43	-52	-62

HPLC-RI analysis of all CCF samples determined near zero average residual glucose levels of 0.08 g/L (± 0.03, n = 21) while HPLC-RI-UV suggested no detectable concentrations of fermentation products in any chromatogram (Figure 28).

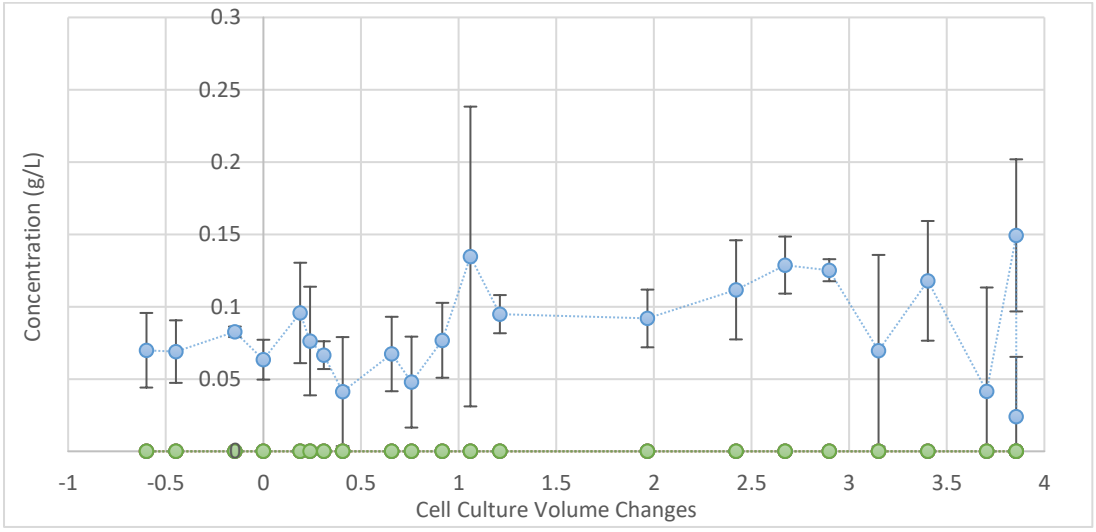


Figure 28: HPLC-RID-UVD analysed CCF products for the initial *P. thermoglucosidasius* GA0.15 INST-^{13}C-MFA experiment. Glucose (light blue), Lactate (orange), Ethanol (grey), Pyruvate (yellow), Formate (dark blue) and Acetate (green). Error is SD for technical triplicate.

GA0.225 – Glucose ASM Aerobic 0.225 h^{-1} dilution rate ILE

For the first ^{13}C ASM, 0.225 h^{-1} dilution rate experiment, this decline in OD_{600} values of the culture was even more prominent. The OD_{600} values declined from a stable average of 5.12 (± 0.01) for the first 5 samples of the ^{13}C portion of the continuous culture experiment to 1.58, representing a 69.14% decrease (Table 18). The redox potential values of the culture generally increased to reflect this falling cell population, increasing from -58 mV to -34 mV before declining again to -59 mV by the end of the experiment.

Table 18: OD_{600}, DCW and redox values (mV) for the initial GA0.225 aerobic glucose *P. thermoglucosidasius* INST-^{13}C-MFA experiment identifying the time point the switch was made to ^{13}C-labelled ASM.

Sample	12C 8	12C 9	12C 10	13C 1	13C 2	13C 3	13C 4	13C 5	13C 6	13C 7	13C 8	13C 9	13C 10	13C 11	13C 12	13C 13	13C 14	13C 15	13C 16	13C 17
Time (hours)	-2.72	-1.72	-0.72	0.00	0.18	0.43	0.75	1.25	2.50	3.00	3.75	4.50	5.25	7.50	9.00	9.77	11.25	12.43	13.68	16.32
Cell culture volume changes (CCVCs)	-0.61	-0.39	-0.16	0.00	0.04	0.10	0.17	0.28	0.56	0.67	0.84	1.01	1.18	1.68	2.02	2.19	2.53	2.79	3.07	3.66
OD.600	5.54	5.56	5.12	5.10	5.12	5.12	5.11	5.14	4.98	4.72	4.35	3.81	3.32	2.96	2.82	2.76	1.42	1.41	1.41	1.58
DCW (g/L)	2.35	2.36	2.17	2.16	2.17	2.17	2.17	2.18	2.11	2.00	1.84	1.62	1.41	1.26	1.20	1.17	0.60	0.60	0.60	0.67
Redox (miliVolts)	-56	-57	-69	-58	-62	-56	-54	-54	-69	-62	-55	-46	-42	-36	-34	-34	-36	-38	-44	-59

HPLC-UV analysis of the CCF samples of the 0.225 h^{-1} dilution rate experiment detected the presence of a sharp spike in acetate in ^{13}C samples 8 (0.37 g/L (± 0.45)) and 9 (1.72 g/L (± 0.30)) but was not able to detect the presence of any other potential fermentation products in in any other chromatogram (Figure. 29). HPLC-RI analysis of the CCFs demonstrated that, at the same time as the detected spike in acetate, concentrations of glucose began a continual increase from 0.03 g/L (± 0.06) to 3.24 g/L (± 0.22).

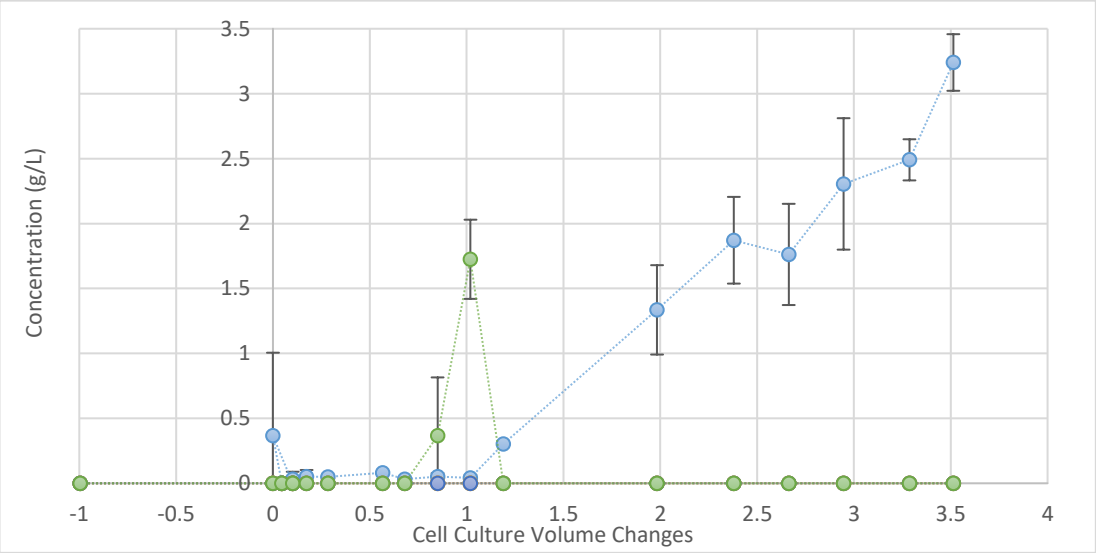


Figure 29: HPLC-RID-UVD analysed CCF products for the initial *P. thermoglucosidasius* GA0.225 INST-^{13}C-MFA experiment. Glucose (light blue), Lactate (orange), Ethanol (grey), Pyruvate (yellow), Formate (dark blue) and Acetate (green). Error is SD for technical triplicate.

Discussion

Combined, these experiments suggested that the process variables of a 0.5% w/v glucose concentration and an air flow rate of 120 mL/min could result in continuous cultures of *P. thermoglucosidasius* displaying aerobic metabolic behaviours at near carbon-limitation. However, the respective 35.13% and 69.14% declines in

observed cell concentration across the 0.15 h^{-1} and 0.225 h^{-1} experiments suggested that these populations were far from the metabolically steady states as is assumed by ^{13}C-MFA. As a result, these experiments did not proceed further in the ^{13}C MFA workflow.

There were three potential factors which could provide explanations for this repeated decline in observed concentration after introduction of the ^{13}C glucose ASM. The first of these explanations was that the sugar itself was the issue. Due to cost constraints, the initially available ^{13}C tracers had been in storage for several years prior to the start of this research. If the drop in cell density was as a result of the concentration and purity of ^{13}C tracer deviating, this could seriously affect any attempt at mapping resulting ^{13}C labelling data to forward simulated MIDs based on defined tracer enrichments and purities. However, a simultaneous 24 well plate batch growth experiment comparing *P. thermoglucosidasius* cells grown in 6x 1 mL 0.5% ^{12}C glucose ASM versus 6x 1 mL 0.5% ^{13}C glucose ASM suggested a negligible difference in growth rate or final OD_{600} readings, indicating this was not an issue.

The second factor could have been the response of the *P. thermoglucosidasius* cells themselves to the ^{13}C labelled substrate. Although a key assumption of ^{13}C-MFA is a lack of any kinetic isotope effects (KIE)s on the overall network as a result of the tracer, distribution of carbon^{-13} through biochemical reactions known to display KIEs *in vitro* could theoretically depress overall cell growth rate and hence the cumulative observed OD_{600} values *in vivo*. As a result, if this were the case it would affect the validity of any biological insights derived from the INST-^{13}C-MFA ILEs. However, despite a theoretical basis, particular investigations into KIEs in metabolic models by Millard, Portais and Mendes (2015) [213] and Sandberg *et al.* (2016) [213] of *E. coli* cells did not support this hypothesis.

Millard, Portais and Mendes (2015) developed a framework for integrating possible KIEs into isotopomer models of *E. coli* and evaluated the KIE impact that different positionally ^{13}C labelled glucose tracers would have on CCM reaction fluxes, ^{13}C-labelling patterns and metabolite pool concentrations. Notably, under U-^{13}C-glucose feeding the difference between simulations with and without KIEs in the individual CCM reactions: of G6PD, 6PD, ribulose-5-phosphate epimerase (Ru5PE), Transaldolase (TA), and transketolases TK1 and TK2 would be limited to a maximum relative flux change of approximately -2%. Despite these individual flux differences, for a system-wide analysis the addition of KIEs only contributed an additional error of approximately <0.001% for most metabolites. The highest error reported was for downstream isotopologue labelling patterns derived from the metabolite node of pyruvate. However, this was still a percentage of only 0.0047%, which was lower than the precision and accuracy which could be obtained by MS, which was stated to be 1%. Overall, Millard, Portais and Mendes (2015) demonstrated that the impact of KIEs was significantly lower at the systems level of a metabolic model than anticipated from measurements of individual enzymes and hypothesised that this global robustness was down to the bidirectionality of many reactions in the global network.

Alternatively, Sandberg *et al.* (2016) performed an Adaptive Laboratory Evolution (ALE) experiment comparing of *E. coli* cells serially cultured for 40 days on 6x M9 minimal media cultures featuring U-^{13}C-glucose and naturally labelled (^{12}C) glucose. Phenotypic characterisation of the cells across this duration suggested that, although there appeared to be a minor benefit to growth rate growing on ^{12}C vs ^{13}C glucose (3 ± 2%), this was not statistically significant and also did not impact the genetic modifications observed in both cell populations across the 40-day ALE. Additionally, the authors also observed that neither the wild-type nor ^{13}C evolved

cells demonstrated any preference for either naturally labelled or ^{13}C-labelled glucose, suggesting a lack of any selection pressure for or against a heavy isotope substrate. Overall, Sandberg *et al.* (2016) also concluded that there were no significant impacts of growing *E. coli* on U-^{13}C-glucose compared to naturally labelled glucose. Ultimately, this literature suggested that any KIEs were not likely to have a notable impact in whole cell systems and was thus unlikely to be the cause of such a substantial observed drop in $O.D._{600}$ in for *P. thermoglucosidasius* in these ILE results.

The third factor was a more general issue of culture foaming observed in these two experiments, and previous aerobic experiments, which was considered the key issue. Foaming during fermentations is the production of gaseous bubbles which occur as because of mechanical agitation combined with the introduction of gas. The liquid lamellas of these bubbles bond with others creating the foam, which is further stabilised as a result of cell proteins, typically from cell lysis[214]. Typically, culture foaming in bioreactor systems is managed either through integrated mechanical foam breakers or through incorporated sensors which instruct the bioreactor control unit to dispense a form of antifoam agent from a reservoir when an electrical circuit is created by rising foam touching the sensor. The MKI and MKII top plates both lacked an available port for an additional sensor and the Braun control unit lacked the foam sensor itself. Additionally, in comparison to the design of many CSTR setups which draw effluent from the base of the vessel, the design of the MKI and MKII had an efflux port on the wall of the vessel, essentially dictating the maximum working volume Figure. 30. Excessive foaming in these systems can therefore resulting in only foam being removed from the reactor instead of culture, leading to a more fed-batch type of growth model. This subsequently impacts the accuracy of any ^{13}C-MFA experiment assuming steady-state as well as the calculation of the specific substrate uptake rate under the set conditions. Ultimately, if this behaviour occurred during the 12C portion of the ILE, but dissipated for the ^{13}C portion of the experiment, the change in growth modes from quasi fed-batch back to continuous would explain the substantial observed drop in OD_{600}.

Figure 30: 'MKII' bioreactor system featuring excessive foaming. Due to the wall mounted efflux port, excessive foaming could result in only foam being removed from the culture, leading to a more fed-batch style of growth.

As a result, two changes were made to the composition of the ASM for further experiments. The first of these was the inclusion of 0.0125% Antifoam 204 in the ASM itself. While this prevented the formation of foam, it was important to keep the

added concentration to a minimum as high concentrations of antifoam can influence the oxygen transfer rate of the culture system [214].

Secondly, to further reduce foaming as the result of high cell concentrations and to reduce the overall cost of experiments, the concentration of ^{13}C substrate used for all experiments was reduced from 0.5% w/v to 0.3% w/v. Now that culture foaming was controlled, to help counteract any potential impact of the 0.0125% antifoam in the ASM on the concentrations of dissolved oxygen, the air flow rate for subsequent experiments was also increased to 140 mL/min or 1.5 VVM.

Overall, through these initial pilot experiments, a set of operating conditions was established for meeting the key assessment criteria of carbon-limited continuous cultures of *P. thermoglucosidasius* NCIMB 11955. Through the reduction of sugar concentrations from 1 to 0.5 to 0.3 % w/v, increase of air flow rate to 1.5 VVM and the inclusion of 0.0125% Antifoam 204 to the media, reliable aerobic metabolic behaviours of *P. thermoglucosidasius* for ^{13}C-MFA could be achieved. These final ASM and operating conditions were maintained throughout all the future aerobic glucose and xylose INST-^{13}C-MFA experiments.

4.3.3.3 Anaerobic growth development by HPLC analysis of cell culture filtrates

Anaerobic growth experiments of *P. thermoglucosidasius* were instead performed in the 350 mL working volume Biostat B bioreactor system, which had an internal record of achieving anaerobic ethanologenic cultures of *P. thermoglucosidasius* when grown on glucose and xylose ASM. A detailed description of this modified system, including the outlet line fixed to the top of the condenser, and its operation are present in materials and methods 2.1.2.3. The greatest constraints that impacted the glucose anaerobic method development were the minimum working limits of the Biostat B bioreactor system itself. In addition to a minimum working volume of 350 mL, it had a minimum active air flow rate of 0.05 L/min or no air flow rate. Unlike the MKI and MKII system, the Biostat B bioreactor system also had an independent nitrogen gas line, with flow rates of 0.05 L/min, to aid in driving air from the culture and headspace. With these minimum limits and a goal of the most anoxic culture able to sustain *P. thermoglucosidasius* growth, focus was given to rotor RPM to control the concentration of dissolved oxygen in the culture.

The variation of agitator speed and the impact on cell populations and culture redox potential.

A series of growth pilots were performed in the Biostat B bioreactor system which aimed to establish the impact of reducing speed of the programmable agitator on the growth and growth conditions of *P. thermoglucosidasius* cells under oxygen-limited continuous culture. Agitator speeds of 200, 100 and 50 RPM were evaluated for their ability to induce production of fermentation products and/or ethanologenic growth conditions.

Results

The $O.D._{600}$ and redox potential values decreased across the three continuous culture experiments (Table 19). Looking at the last 4 samples of each experiment (most likely to represent metabolic steady state), the average OD_{600} values of 200, 100 and 50 RPM experiments were: 0.81 (± 0.04, n = 4), 0.58 (± 0.03, n = 4) and 0.58 (± 0.02, n = 4) respectively. Similarly, as rotor RPM decreased, the average redox potential values of the last 4 samples of each experiment also decreased, from -188 mV (± 12.51, n = 4) for the 200 RPM experiment, to -210 mV (± 3.30, n = 4) for the 100 RPM experiment to -252 mV (± 3.50, n = 4) for the 50 RPM experiment.

Table 19: OD_{600}, DCW and redox values (mV) for for the *P. thermoglucosidasius* variation of agitator RPM experiments at 200, 100, 50 RPM.

	Sample	200 RPM 12C 1	200 RPM 12C 2	200 RPM 12C 3	200 RPM 12C 4	200 RPM 12C 5	200 RPM 12C 6	200 RPM 12C 7	200 RPM 12C 8	200 RPM 12C 9	200 RPM 12C 10
200 RPM	Time (hours of run)	0	3.333	6.667	10	13.667	24	25.667	29	30.667	49.167
	Cell Culture volume Changes	0	0.25	0.5	0.75	1.025	1.8	1.925	2.175	2.3	3.688
	OD.600	0.54	0.65	0.71	0.69	0.75	0.88	0.80	0.80	0.88	0.77
	DCW (g/L)	0.231	0.276	0.301	0.292	0.320	0.373	0.338	0.338	0.373	0.325

	Redox (mV)	-210	-218	-207	-217	-224	-194	-190	-197	-198	-167
100 RPM	**Sample**	**100 RPM 12C 1**	**100 RPM 12C 2**	**100RPM 12C 3**	**100 RPM 12C 4**	**100 RPM 12C 5**	**100 RPM 12C 6**	**100 RPM 12C 7**			
	Time (hours)	0	4.333	7.667	24.667	30.833	32.5	49.5			
	Cell Culture volume Changes	0	0.325	0.575	1.85	2.313	2.438	3.713			
	OD.600	0.71	0.67	0.702	0.544	0.615	0.597	0.564			
	DCW g/L	0.301	0.284	0.298	0.231	0.261	0.253	0.239			
	Redox (miliVolts)	-194	-263	-240	-212	-212	-210	-204			
50 RPM	**Sample**	**50 RPM 12C 1**	**50 RPM 12C 2**	**50 RPM 12C 3**	**50 RPM 12C 4**	**50 RPM 12C 5**	**50 RPM 12C 6**	**50 RPM 12C 7**			
	Time (hours)	0	3.167	6.167	24.333	27.333	30.333	48			
	Cell Culture volume Changes	0	0.238	0.463	1.825	2.05	2.275	3.6			
	OD.600	0.81	0.961	0.787	0.614	0.568	0.562	0.585			
	DCW g/L	0.343	0.407	0.334	0.260	0.241	0.238	0.248			
	Redox (miliVolts)	32	-268	-265	-255	-253	-254	-246			

HPLC-RI-UV analysis was again used to evaluate the chemical concentrations present in the CCFs of samples over the course of the three continuous culture experiments. HPLC-RI analysis of all sample CCFs demonstrated differing changes in concentration of glucose across the experiment (Figure 31), depending on what point in the growth cycle continuous culture was initiated. Once again looking at the last 4 time points of each experiment, the CCFs of the last 4 samples of the 200 PM experiment demonstrated an average glucose concentration of 0.06 g/L (± 0.03, n = 4), while the last 4 samples of the 100 PM experimented demonstrated a near zero average concentration of glucose of 0.01 g/L (± 0.03, n = 4). However, for the 50 RPM continuous culture experiment the final 4 sample CCFs demonstrated an average glucose concentration of 0.13 g/L (± 0.02, n = 4) and did not reach a steady state. HPLC-RI analysis of all sample CCFs did not detect the presence of ethanol in either the 200 or 100 RPM experiments. Notably however, the presence of 0.07 g/L (± 1.27) ethanol was observed in the final CCF of the 50 RPM experiment.

HPLC-UV analysis of the CCFs of each experiment demonstrated the presence of the *P. thermoglucosidasius* fermentation metabolites of: lactate, formate and acetate which varied with respect to the agitator RPM (Figure 31). No pyruvate was detected in any sample's CCFs implying that fermentation pathways were handling any potential accumulation of pyruvate. Looking at the final 4 time points of each experiment, the average concentrations of lactate decrease as the rotor RPM decreases, averaging at 1.30 g/L (± 0.21, n = 4) for the 200 RPM experiment, 0.70 g/L (± 0.11, n = 4) for the 100 RPM experiment and below the limit of detection for the 50 RPM experiment. The average concentrations of formate and acetate in the final 4 time points of each experiment show an inverse correlation to that of lactate, both increasing in concentration as the rotor RPM decreases. For formate, these average concentrations were 0.13 g/L (± 0.04, n = 4) for the 200 RPM experiment 0.50 g/L (± 0.03, n = 4) for the 100 RPM experiment and an average concentration of at 0.72 g/L (± 0.16, n = 4) for the 50 RPM experiment. For acetate, no

concentration could be detected in the CCFs of the 200 PM experiment, an average of 0.54 g/L (± 0.16, n = 4) was detected for the 100 RPM experiment and an average concentration of 1.04 g/L (± 0.12, n = 4) was detected for the 50 RPM experiment.

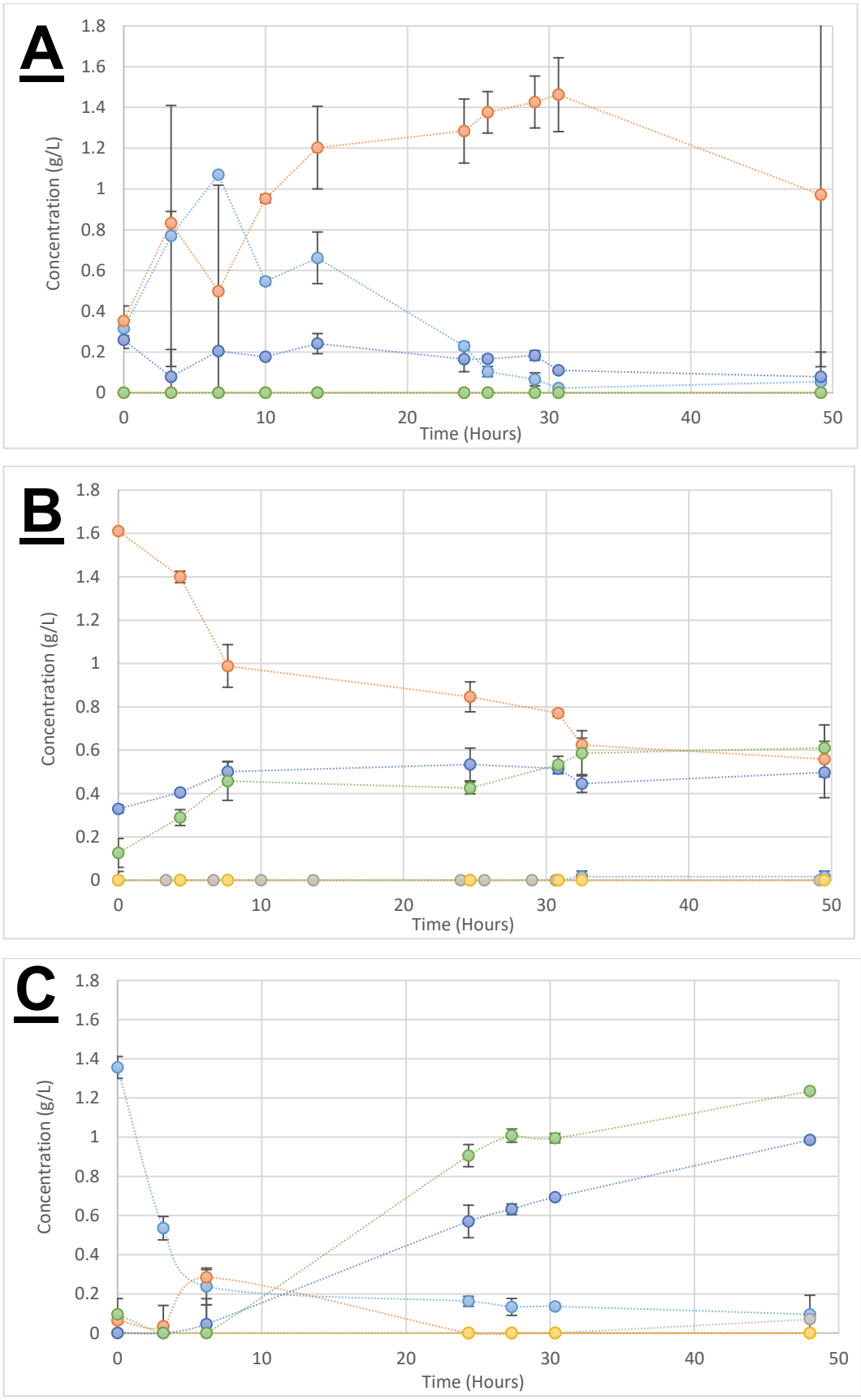


Figure 31: HPLC-RID-UVD analysed CCF products for the *P. thermoglucosidasius* variation of agitator RPM experiment. A) 200 RPM **B)** 100 RPM **C)** 50 RPM. Glucose (light blue), Lactate (orange), Ethanol (grey), Pyruvate (yellow), Formate (dark blue) and Acetate (green). Error is SD for technical triplicate.

Discussion

Despite the observations of 0.13 g/L (± 0.02, n = 4) glucose in the 50 RPM agitator speed experiment which did not reach a steady state, these pilot experiment results do suggest that the chosen operating conditions could sustain near carbon limited growth of *P. thermoglucosidasius* cells in continuous culture under low oxygen concentration conditions. As a result of being able to observe a small amount of ethanol, an agitator speed of 50 RPM was chosen for future ILE runs. While the agitator could be programmed to stop, to ensure culture mixing was continued to provide adequate suspension and dispersion of cells as well as the prevention of biofilm formation.

A reduction in agitator speed RPM reducing the rate of gas-liquid mixing of the culture and hence the maximum potential dissolved oxygen concentration. In turn, this reduces the redox potential of the culture and hence the metabolic behaviours of *P. thermoglucosidasius* cells. Thus, as these fermentation products are a result of the oxygen limitation of the culture, it is unsurprising that there is a relationship between the average concentrations of these fermentation products and the rotor RPM which ultimately affects the concentration of oxygen available to an individual cell. However, these trends do help to exemplify the metabolic transitions of *P. thermoglucosidasius* from predominant production of lactate by lactate dehydrogenase, to acetate and formate by PFL and even down to early ethanol production by alcohol dehydrogenase. Indeed, these results suggest that agitator RPM could be an important variable for precisely tuning the redox potential of a continuous culture for the study of the metabolic behaviours of *P. thermoglucosidasius* at precise redox potential values, such as the production of ethanol and could be used in future work to investigate the suggested observed metabolic switch from lactate to formate and acetate production.

Supplementation with $FeCl_3$

While the agitator RPM pilots represented progress towards establishing operating conditions ethanologenic anaerobic conditions, the HPLC-RI detected concentration of 0.07 g/L (± 1.27) ethanol was ethanol in the final sample of the 50 RPM experiment was low and had a high value of standard error. This was not deemed enough to enable a conclusion that the targeted ethanologenic ‘anaerobic’ growth conditions had been achieved. As the continuous culture runs were constrained by the cost of the ^{13}C sugar and the already minimum operating conditions of Biostat B bioreactor system, focus turned to the *p-thermo* GSMM experimental results obtained suggesting that addition of thiamine, biotin and Fe(III) could support anaerobic growth of *P. thermoglucosidasius*. Notably, as thiamine and biotin were already components of ASM, this suggested that Fe(III) was missing in the media in order to allow porphyrin biosynthesis. The previous ASM composition already included a source of iron in the form of dissolved $FeSO_4.7H_2O$, which represents iron in its +2-oxidation state. If there was a particular growth advantage of the presence of iron in its +3-oxidation state (as discussed in 3.3.3), it is reasonable to assume that under oxygenic conditions oxidation from Fe(II) to Fe(III) either in water or by a potentially genomically encoded Fe2+:oxygen oxidoreductase (*p-thermo* reaction ID ‘R_FERO’) could occur and meet this demand. However, under oxygen-limited conditions it is also reasonable to believe that this oxidation would be less likely. Therefore a stable yet soluble form of Fe(III) was sought to add to the ASM. The *p-thermo* GSMM experimental results suggested that *P. thermoglucosidasius* was unable to synthesize hemin under anaerobic conditions, which contains an Fe^{3+} - Cl bound to a porphyrin group.

Early research by Amartey, Leak and Hartley (1991)[215] identified that an optimized 1% sucrose defined medium containing 0.07 g/L $FeCl_3.6H_2O$ was able to support

aerobic growth of a *G. stearothermophilus* mutant. It was however noted that the optimized defined media could not support anaerobic growth on its own. Given this precedent however, an equivalent concentration of 0.042 g/L anhydrous $FeCl_3$ was added to the ASM for the final anaerobic glucose experiment.

While only representing a single biological replicate, this $FeCl_3$ supplemented growth experiment resulted in the lowest observed steady state value for redox potential yet, maintaining a fixed redox potential of -259 mV for the final 3 hours 46 minutes of the ^{12}C ASM growth experiment. However, given the remaining time constraints of the project, this continuous culture experiment proceeded to be the final anaerobic glucose, 0.075h^{-1} dilution rate INST-^{13}C-MFA experiment.

4.4 The isotopically instationary ^{13}C-ILEs and metabolic flux analysis

With the experimental conditions defined, the final INST-^{13}C-MFA ILEs for *P. thermoglucosidasius* NCIMB 11955 were performed as described in the Materials and Methods chapter 2. This research collates and discusses different INST-^{13}C-MFA experiments for *P. thermoglucosidasius* NCIMB 11955 with an overall goal to provide data sets to further develop the utility of the GSMM. A complete summary of the target aeration conditions, ^{13}C carbon sources and dilution/growth rates of the experiments performed and analyzed as part of this thesis provided in Table 20.

Table 20: Table detailing the different INST-^{13}C-MFA ILEs performed for *P. thermoglucosidasius* NCIMB 11955, including the composition of the ^{13}C source used, who performed the experiment and who analysed the resulting data.

Organism	Aeration conditions	^{13}C Carbon Source	Dilution rate (h^{-1})	^{13}C-MFA experiment performed	^{13}C-MFA experiment data analysed
***Parageobacillus thermoglucosidasius* NCIMB 11955**	Aerobic	0.3% D-Glucose (60% 1-^{13}C, 40% U-^{13}C) ASM	0.075	**This Study (Martyn Bennett)**	**This Study (Martyn Bennett)**
			0.15		**This Study (Martyn Bennett)**
			0.225		**This Study (Martyn Bennett)**
			0.32		
		0.3% D-Xylose (75% 1-^{13}C, 25% ^{12}C) ASM	0.075	**This Study (Martyn Bennett)**	**This Study (Martyn Bennett)**
			0.15	**This Study (Martyn Bennett)**	**This Study (Martyn Bennett)**
	'Micro-aerobic'	0.3% D-Glucose (60% 1-^{13}C, 40% U-^{13}C) ASM	0.075	**This Study (Martyn Bennett)**	**This Study (Martyn Bennett)**
	Anaerobic	0.3% D-Glucose (60% 1-^{13}C, 40% U-^{13}C) ASM + Fe(III)	0.075	**This Study (Martyn Bennett)**	**This Study (Martyn Bennett)**
		0.5% D-Xylose (75% 1-^{13}C, 25% ^{12}C) ASM	0.075	2014 (Dr. Masakapalli, UoB/IIT Mandi)	**This Study (Martyn Bennett)**

4.4.1 The final INST-^{13}C-MFA ILEs

To enable more direct comparisons of ^{13}C incorporation across the growth experiments, sampling time points for all INST-^{13}C-MFA experiments were chosen with respect to the dilution rate dependent CCVCs. The target CCVCs and respective time points in hours for each ILE are presented in Table 21.

Table 21: Table of the sampling strategy employed for all INST-^{13}C-MFA experiments emphasising how the sampling time point in hours was determined with respect the fraction of a single cell culture volume change (CCVC).

^{13}C Sample	CCVC	Dilution rate (h^{-1})			
		0.075	0.15	0.225	0.3
		Sampling Time Point (hours)			
1	0	0	0	0	0
2	0.0375	0.5	0.25	0.167	0.125
3	0.075	1	0.5	0.333	0.25
4	0.15	2	1	0.667	0.5
5	0.25	3.333	1.667	1.111	0.833
6	0.5	6.667	3.333	2.222	1.667
7	0.6	8	4	2.667	2
8	0.75	10	5	3.333	2.5
9	0.9	12	6	4	3
10	1.05	14	7	4.667	3.5
11	1.5	20	10	6.667	5
12	1.8	24	12	8	6
13	1.95	26	13	8.667	6.5
14	2.25	30	15	10	7.5
15	2.5	33.333	16.667	11.111	8.333
16	2.75	36.667	18.333	12.222	9.167
17	3	40	20	13.333	10
18	3.25	43.333	21.667	14.444	10.833
19	3.5	46.667	23.333	15.556	11.667
20	3.75	50	25	16.667	12.5

4.4.1.1 The glucose ASM aerobic INST-^{13}C-MFA GA0.075 GA0.15 GA0.225 and GA0.32 ILEs

Using the 0.3% glucose, 0.0125% antifoam 204 ASM and the growth operation conditions detailed in Table 20, the final aerobic *P. thermoglucosidasius* NCIMB 11955 INST-^{13}C-MFA ILEs were carried out as described in (Materials and methods 2.2.1) at final media dilution rates of 0.075 h^{-1}, 0.15 h^{-1}, 0.225 h^{-1} and 0.32 h^{-1} (Table 22) (Figures 32, 33, 34). While a dilution rate of 0.3 h^{-1} was initially planned, the 0.32 h^{-1} dilution rate was the closest possible dilution rate that could be achieved by the precision of the 101U peristaltic pump employed.

No foaming was observed for any experiment. For dilution rates 0.075 h^{-1}, 0.15 h^{-1}, 0.225 h^{-1} and 0.32 h^{-1} the average respective $O.D._{600}$ and redox potential values were 2.25 ± 0.12 and -46.78 mV (± 4.81, n = 18), 4.40 ± 0.44 and -39.62 mV (± 3.55, n = 21), 3.2 ± 0.40) and -82 (± 8.17, n = 19) and 2.45 ± 0.11 and -78.15 mV (± 1.65, n = 20) (Tables 22 and 23). This error as a proportion of average O.D.600 for each experiment ranged from 4.66% (0.32 h^{-1}) to 12.46% (0.225 h^{-1}) and for redox potential ranged from 2.11% (0.32 h^{-1}) to 9.96% (0.225 h^{-1}). nMeasured redox potential decreased from the start of the ^{13}C portion of the experiment from -78 mV, reached a trough at -100 mV and increased steadily into the -70mV range before increasing rapidly from -71mV to -41mV in the last hour of the experiment.

Table 22: OD_{600}, DCW and redox values (mV) for the final GA0.075 and GA0.15 aerobic glucose *P. thermoglucosidasius* INST-^{13}C-MFA ILEs.

GA0.075	Sample	12C 8	12C 9	12C 10	13C 1	13C 2	13C 3	13C 4	13C 5	13C 6	13C 7	13C 8	13C 9	13C 10	13C 11	13C 12	13C 13	13C 14	13C 15			
	Time (hours)	-4.97	-3.5	-2	0	1.15	2.02	3.32	12	13.98	19.98	23.98	25.98	36.65	40.02	43.35	46.68	50.02	51.52			
	Cell Culture volume Changes	-0.37	-0.26	-0.15	0	0.09	0.15	0.245	0.9	1.05	1.50	1.80	1.95	2.75	3.00	3.25	3.50	3.75	3.86			
	OD600	2.02	2.02	2.07	2.14	2.18	2.18	2.18	2.31	2.33	2.36	2.33	2.32	2.33	2.33	2.35	2.32	2.34	2.34			
	DCW (g/L)	0.86	0.86	0.88	0.91	0.92	0.92	0.92	0.98	0.99	1.0	0.99	0.99	0.99	0.99	1.00	0.98	0.99	0.99			
	Redox (miliVolts)	-43	-41	-42	-48	-48	-47	-48	-46	-47	-48	-48	-48	-48	-49	-48	-47	-48	-48			
GA0.15	Sample	12C 3	12C 4	12C 5	13C 1	13C 2	13C 3	13C 4	13C 5	13C 6	13C 7	13C 8	13C 9	13C 10	13C 11	13C 12	13C 13	13C 14	13C 15	13C 16	13C 17	13C 18
	Time (hours)	-3.97	-2.97	-0.97	0	1.25	1.58	2.05	2.68	4.35	5.02	6.07	7.02	8.02	13.02	16.02	17.68	19.18	20.85	22.52	24.52	25.52
	Cell culture volume changes CCVCs	-0.60	-0.45	-0.15	0	0.19	0.24	0.31	0.41	0.66	0.76	0.92	1.06	1.21	1.97	2.42	2.67	2.90	3.15	3.40	3.70	3.86
	OD600	4.77	4.47	4.83	4.88	4.78	4.79	4.84	4.78	4.68	4.67	4.72	4.63	4.61	4.26	4.01	3.87	3.78	3.77	3.68	3.77	3.83
	DCW g/L	2.02	1.90	2.05	2.07	2.03	2.03	2.05	2.03	1.98	1.98	2.00	1.96	1.95	1.80	1.70	1.64	1.60	1.60	1.56	1.60	1.62
	Redox (miliVolts)	-33	-32	-34	-33	-38	-42	-43	-41	-41	-46	-42	-42	-41	-40	-40	-40	-41	-40	-42	-41	-40

Table 23: **OD_{600}, DCW and redox values (mV) for the final GA0.225 and GA0.32 aerobic glucose *P. thermoglucosidasius* INST-^{13}C-MFA ILEs.**

GA0.225	Sample	12C 5	12C 6	12C 7	13C 1	13C 2	13C 3	13C 4	13C 5	13C 6	13C 7	13C 8	13C 9	13C 10	13C 11	13C 12	13C 13	13C 14	13C 15	13C 16
	Time (hours)	-4.67	-3.67	-2.67	0.00	0.20	0.45	0.77	1.25	2.50	3.00	3.75	4.50	5.25	8.75	10.50	11.75	13.00	14.50	15.50
	Cell Culture volume Changes	-1.00	0.05	0.10	0.00	0.05	0.10	0.17	0.28	0.57	0.68	0.85	1.02	1.19	1.98	2.38	2.66	2.95	3.29	3.51
	OD 600	2.30	2.30	2.33	3.10	3.22	3.22	3.29	3.27	3.38	3.42	3.46	3.46	3.52	3.44	3.44	3.40	3.52	3.38	3.28
	DCW g/L	0.98	0.98	0.99	1.31	1.37	1.37	1.39	1.39	1.43	1.45	1.47	1.47	1.49	1.46	1.46	1.44	1.49	1.43	1.39
	Redox (miliVolts)	-84	-84	-84	-78	-79	-85	-87	-88	-94	-100	-94	-95	-91	-80	-78	-74	-71	-71	-41
GA0.32	Sample	12C 8	12C 9	12C 10	13C 1	13C 2	13C 3	13C 4	13C 5	13C 6	13C 7	13C 8	13C 9	13C 10	13C 11	13C 12	13C 13	13C 14	13C 15	13C 16	13C 17
	Time (hours)	-2.80	-1.80	-0.80	0.00	0.30	0.83	1.18	1.50	2.00	2.53	3.03	4.35	5.37	5.87	6.87	7.70	8.55	9.55	10.05	10.92
	Cell culture volume changes	-0.88	-0.57	-0.25	0.00	0.09	0.26	0.37	0.47	0.63	0.80	0.96	1.37	1.69	1.85	2.17	2.43	2.70	3.01	3.17	3.44
	OD600	2.17	2.30	2.33	2.36	2.41	2.42	2.37	2.40	2.44	2.40	2.43	2.51	2.46	2.41	2.53	2.59	2.57	2.60	2.62	2.61
	DCW g/L	0.92	1.00	1.02	1.03	1.00	1.02	1.03	1.02	1.03	1.06	1.04	1.02	1.07	1.10	1.09	1.10	1.11	1.11	0.00	0.00
	Redox (miliVolts)	-80	-76	-79	-79	-77	-82	-79	-78	-80	-77	-79	-77	-76	-77	-75	-77	-78	-80	-78	-79

GA0.075

HPLC-RI analysis of the CCF samples of the 0.075 h^{-1} dilution rate experiment demonstrate an initial concentration of glucose which rises from 0.13 g/L (± 0.06) in ^{12}C 8, peaks at an average between ^{13}C samples 1-3 of 16.5 g/L (± 0.02, n = 3), then steadily declines to below detectable levels by ^{13}C 8, remaining at near zero levels up to the end of the experiment. HPLC-UV analysis of the CCFs suggests that, as glucose reaches zero, there is a single peak of 0.21 g/L (± 0.19) pyruvate detected. However, no other fermentation products were detected.

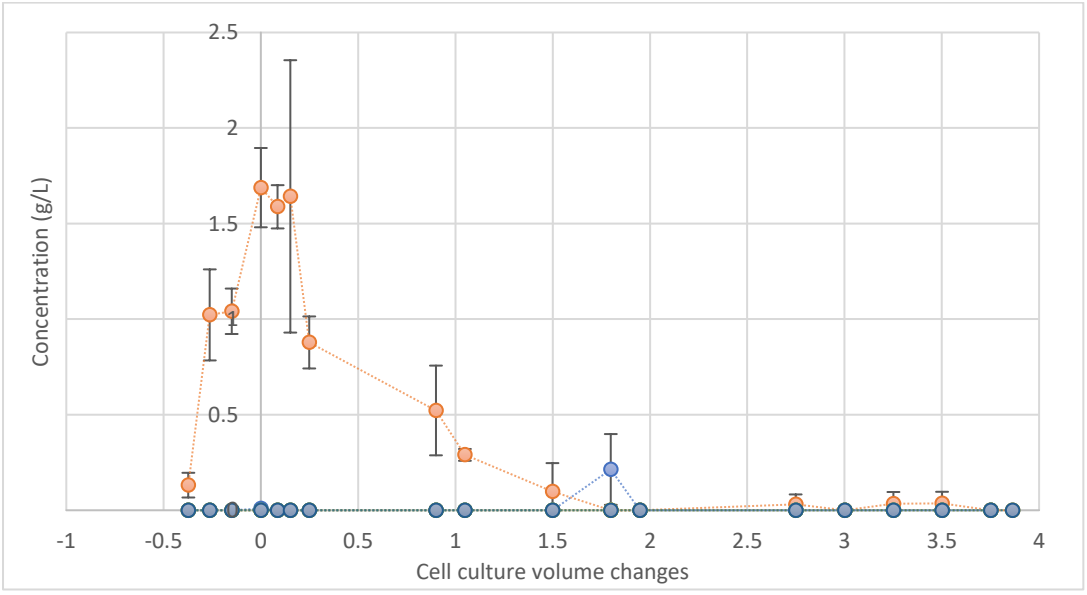


Figure 32: **HPLC-RID-UVD analysed CCF products for the final INST-^{13}C-MFA GA0.075 *P. thermoglucosidasius* ILE.** Glucose (orange), Lactate (grey), Ethanol (yellow, Pyruvate (light blue), Formate (green) and Acetate (dark blue). Error is SD for technical triplicate.

GA0.15 and GA0.32

HPLC-RI-UV analysis of the CCFs of the second aerobic 0.15 h^{-1} and first 0.32 h^{-1} ^{13}C glucose ILEs show similarity in the respective ^{13}C portions of the experiments. From the 2nd ^{13}C sample analysed from each experiment, the chromatograms of both experiments demonstrate low average residual concentrations of glucose of 0.06 g/L (± 0.03, n = 15) for 0.15 h^{-1} and of 0.06 g/L (± 0.01, n = 16) for 0.32 h^{-1}. From the 2nd ^{13}C sample, neither sets of ^{13}C chromatograms demonstrated a presence of any overflow or fermentation metabolites (Figures. 33 & 34). However, HPLC-UV of the 8th, 9th and 10th time point samples of the ^{12}C media and 1st sample from the ^{13}C media show unexpectedly high average concentrations of acetate on the chromatographs of 3.18 g/L ± 1.08 g/L (n = 4).

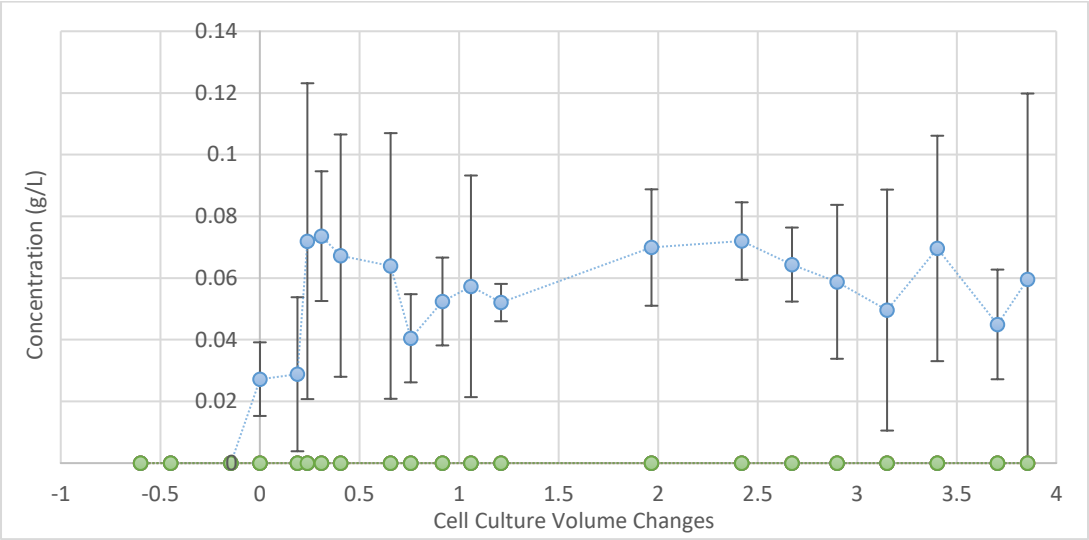


Figure 33: HPLC-RID-UVD analysed CCF products for the final INST-^{13}C-MFA GA0.15 *P. thermoglucosidasius* ILE. Glucose (light blue), Lactate (orange), Ethanol (grey), Pyruvate (yellow), Formate (dark blue) and Acetate (green). Error is SD for technical triplicate.

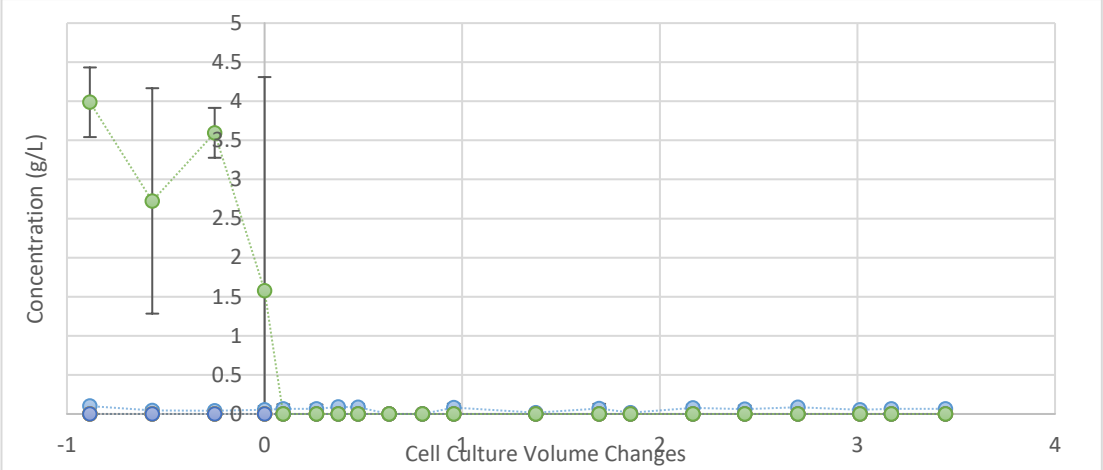


Figure 34: HPLC-RID-UVD analysed CCF products for the final INST-^{13}C-MFA GA0.32 *P. thermoglucosidasius* ILE. Glucose (light blue), Lactate (orange), Ethanol (grey), Pyruvate (yellow), Formate (dark blue) and Acetate (green). Error is SD for technical triplicate.

GA0.225

HPLC-RI analysis of the CCF samples of the second aerobic, 0.225 h^{-1} 13C glucose ILE (GA0.225) demonstrate a similarly low average concentration of glucose across the experiment of 0.09 g/L (± 0.07, n = 19). HPLC-UV analysis of the CCFs suggests that lactate was possible detected at a concentration of 0.392 g/L (± 0.64), in the 3rd ^{13}C sample CCF (Figure 35). Otherwise, no lactate or pyruvate were detected for the duration of the experiment. However, the CCFs do demonstrate a low presence of fermentation products formate and acetate. Presence of formate is initially observed once at a concentration of 0.16 g/L (± 0.06) in the CCF of the 3rd ^{13}C sample. Formate is then observed near continuously in the CCFs of ^{13}C samples 7-16, at an average concentration of 0.10 g/L (± 0.05, n = 10). Acetate is only observed in the CCFs of ^{13}C samples 11-16, where it increases from 0.06 g/L (± 0.64) to 0.44 g/L (± 0.06).

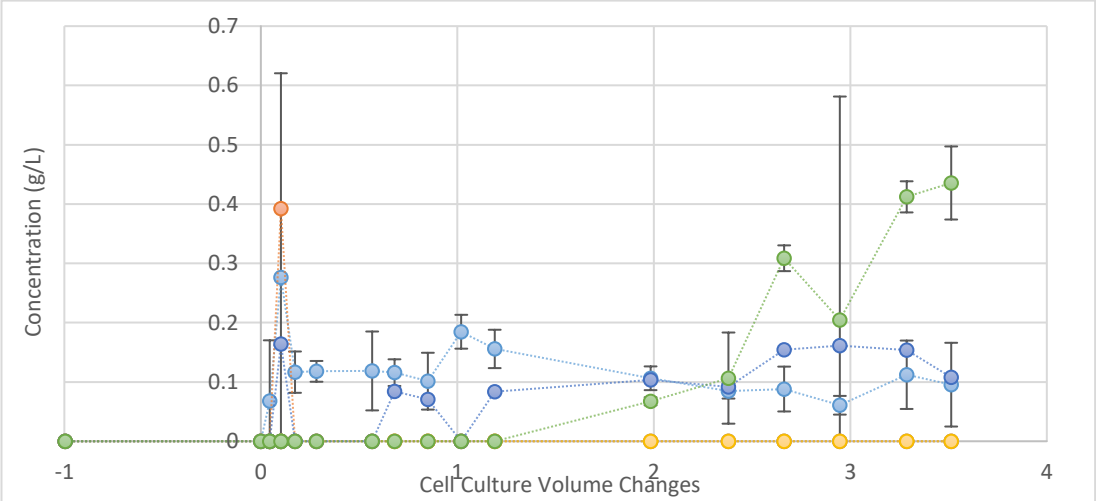


Figure 35: HPLC-RID-UVD analysed CCF products for the final INST-^{13}C-MFA GA0.225 *P. thermoglucosidasius* ILE. Glucose (light blue), Lactate (orange), Ethanol (grey), Pyruvate (yellow), Formate (dark blue) and Acetate (green). Error is SD for technical triplicate.

Discussion

Combined, the OD_{600} and redox potential values of the different dilution rate ILEs suggest that an aerobic metabolic steady state for *P. thermoglucosidasius* had been

achieved throughout the key ^{13}C isotope labelling portion of the GA0.075, GA0.15, and GA0.32 experiments and that the carbon-limited growth rates of the cells reflect the dilution rates of the experiments.

However, for the GA0.075 ILE, the spike in glucose likely results from excess ^{13}C media being added in the changeover point between ^{12}C and ^{13}C-glucose ASM, perturbing the metabolic steady state. These results suggest that the *P. thermoglucosidasius* cells are likely to have reached an acceptably stable metabolic equilibrium both before the perturbation but may not accurately represent metabolic steady state or constant rates of isotope incorporation at the target conditions by the end of the experiment.

For the GA0.225 ILE, the relatively high variability of the OD_{600} and redox potential values of the 0.025 h^{-1} dilution rate experiment suggests that perturbation of the *P. thermoglucosidasius* cells population across the experiment, indicating that metabolic steady state may not have been achieved across the experiment. HPLC-RI suggests the cells were near carbon-limited and therefore likely growing at a rate near the 0.225 h^{-1} dilution rate. However, the HPLC-UV detected presence of formate and acetate in the CCFs, while much lower concentrations than detected for glucose microaerobic and anaerobic experiments on a DCW basis, does suggest that the cells were experiencing a degree of oxygen limitation compared to the other aerobic experiments and PFL may have been active.

The presence of 4.73 g/L acetate in the ^{12}C portion of the GA0.32 ILE experiment (Figure 34) was unexpected. However, as the concentration of acetate seems to completely disappear in the 32 minutes between the 1st and 3rd ^{13}C sample, its presence (in light of the stable 78.15 ± 2.11 mV (n = 3) redox potential of the experiment) suggests that this is unlikely to be an accurate observation of *P. thermoglucosidasius* metabolism and more likely to be an artefact of cross-contamination of other samples in the queue of the HPLC instrument. This argument is supported by the observations that the 1st of the technical triplicate CCF samples of ^{13}C sample 1 shows a 4.73 g/L concentration while the 2nd and 3rd samples do not register any acetate above the limit of detection of the instrument.

4.4.1.2 The glucose ASM anaerobic and micro-aerobic INST-^{13}C-MFA GAn0.075 and GMA0.075 ILEs

GAn0.075

For the anaerobic glucose 0.075h^{-1} dilution rate INST-^{13}C MFA experiment (GAn0.075), *P. thermoglucosidasius* cells were cultured as described in materials and methods 2.21 at the 50 RPM agitator speed informed by the anaerobic pilots, 60°C, pH 7, 0.05 L/min air (0.14 VVM) and 1L/min N_2 (2.86 VVM) on 0.3% w/v ^{D}-Glucose (60% 1-^{13}C, 40% U-^{13}C) ASM (supplemented with 0.042 g/L $FeCl_3$).The average OD_{600} and redox potential values observed were 0.51 ± 0.05 and -259 mV ± 2.25 (n = 15) (Table 24).

Table 24: OD_{600}, DCW and redox values (mV) for the final GAn0.075 anaerobic glucose *P. thermoglucosidasius* INST-^{13}C-MFA ILEs.

	Sample	12C 13	12C 14	12C 15	13C 1	13C 2	13C 3	13C 4	13C 5	13C 6	13C 7	13C 8	13C 9	13C 10	13C 11	13C 12	13C 13	13C 14	13C 15
GAn0.075	Time (hours)	-3.22	-1.97	-0.72	0.00	0.83	1.67	3.00	6.33	7.67	9.42	19.67	21.73	25.67	29.67	33.00	43.00	47.50	50.75
	Cell Culture volume Changes	-0.24	-0.15	-0.05	0.00	0.06	0.13	0.23	0.48	0.58	0.71	1.48	1.63	1.93	2.23	2.48	3.23	3.56	3.81
	OD_{600}	0.46	0.46	0.41	0.47	0.46	0.46	0.45	0.48	0.47	0.51	0.51	0.54	0.59	0.56	0.58	0.56	0.56	0.59
	DCW g/L	0.19	0.20	0.18	0.20	0.20	0.19	0.19	0.20	0.20	0.21	0.21	0.23	0.25	0.24	0.25	0.24	0.24	0.25
	Redox (miliVolts)	-259	-258	-259	-258	-258	-261	-261	-256	-256	-256	-258	-258	-265	-260	-260	-261	-260	-262

HPLC-UV of the sample CCFs demonstrated the presence of the typical *P. thermoglucosidasius* 11955 overflow and fermentation products when grown on glucose of pyruvate, lactate, acetate and formate (Figure. 36). Pyruvate was detected at near zero g/L concentrations for the duration of the experiment, averaging at 0.04 (± 0.03, n = 14). Lactic acid concentrations demonstrated a near exponential decrease in concertation over the time course of the ^{13}C ILE (R^2 = 0.98), from 1.67 g/L (± 0.07) at the ^{13}C zero mark to 0.257 g/L (± 0.01) at the end of the experiment.

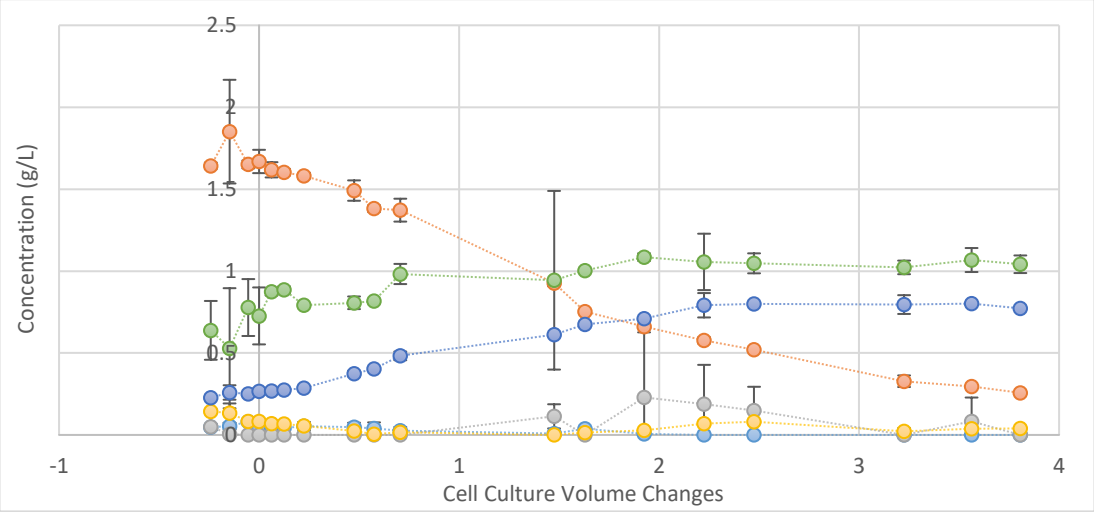


Figure 36: HPLC-RID-UVD analysed CCF products for the final INST-^{13}C-MFA GAn0.075 *P. thermoglucosidasius* ILE. Glucose (light blue), Lactate (orange), Ethanol (grey), Pyruvate (yellow), Formate (dark blue) and Acetate (green). Error is SD for technical triplicate.

The HPLC-UV detected concentrations of formate and acetate gradually increase over the first half of the ^{13}C experiment before plateauing. From the first ^{13}C time point until the 10th, formate concentrations were observed to increase from 0.27 g/L (± 0.02) to 0.71 g/L (n = 10), while acetate concentrations increased from 0.726 g/L (± 0.174) to 1.09 g/L (± 0.02, n = 10). Both product concentrations then plateau and stabilize for the remainder of the experiment, formate to an average concentration of 0.78 g/L (± 0.08, n = 6) and acetate to an average concentration of 1.05 g/L (± 0.02, n = 6).

HPLC-RI analysis of the CCFs demonstrates near zero g/L concentrations of glucose, averaging 0.03 g/L (± 0.02), indicating that the *P. thermoglucosidasius* cells were likely carbon-limited. HPLC-RI analysis of the ethanol concentrations of the CCF and outlet condensate concentrations varied (Figure 37). The average HPLC-RI detected concentration of ethanol in the CCFs was 0.05 g/L (± 0.07, n = 18) whereas the average concentration of ethanol detected in the outlet condensate filtrates was higher at 0.53 g/L (± 0.08, n = 18). For the first 10 hours of the ^{13}C ASM portion of the experiment, ethanol concentrations were near the limit of detection, averaging 0.01 g/L (± 0.01, n = 10) in the ^{13}C sample CCFs and 0.02 g/L (± 0.04, n = 10) in the outlet condensate filtrate samples. However, from the 20-hour mark until the end of the experiment, the detected concentrations of ethanol were higher, averaging 0.1 g/L (± 0.08, n = 8) in the ^{13}C CCFs and 1.16 g/L (± 0.10, n = 8) in the outlet condensate filtrate. For sampled time points where ethanol was detected in both the CCF and outlet condensate filtrate, the concentration in the outlet condensate filtrate was on average 8.48 (± 3.63, n = 5).

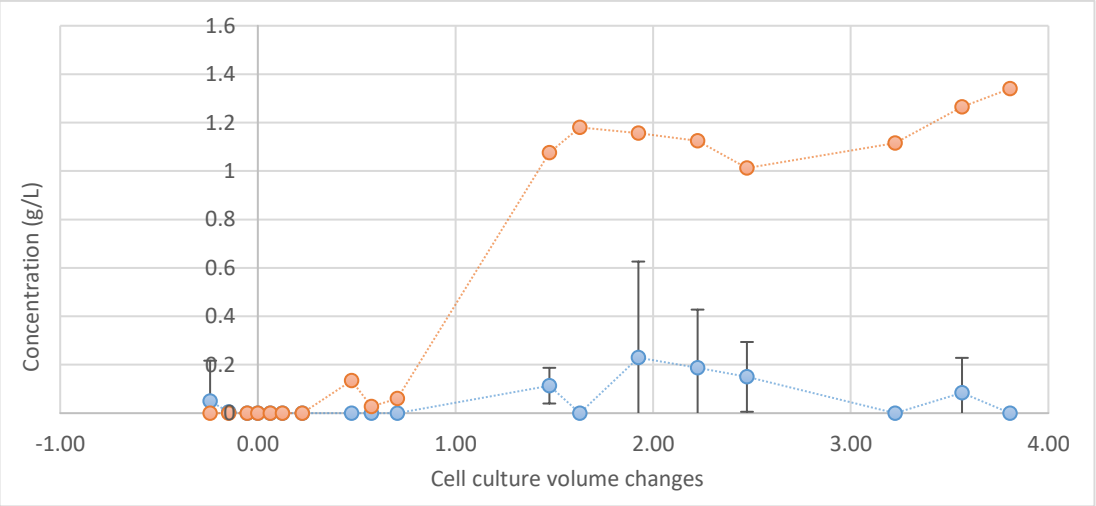


Figure 37: HPLC-RID analysed ethanol products for the final INST-^{13}C-MFA GAn0.075 *P. thermoglucosidasius* ILE. Ethanol measured in CCFs (light blue, n = 3), Ethanol measured in outlet condensate filtrate samples (orange, n= 1).

GMA0.075

The last ^{13}C glucose ILE to be performed was GMA0.075. After a combination of the agitator RPM pilots and the final anaerobic glucose 0.075 h^{-1} INST ^{13}C-MFA experiment, a limited logarithmic relationship was observed between the observed

culture redox potential and the agitator rotor RPM values of the final 4 samples of each experiment (R^2 = 0.911) (Figure 38).

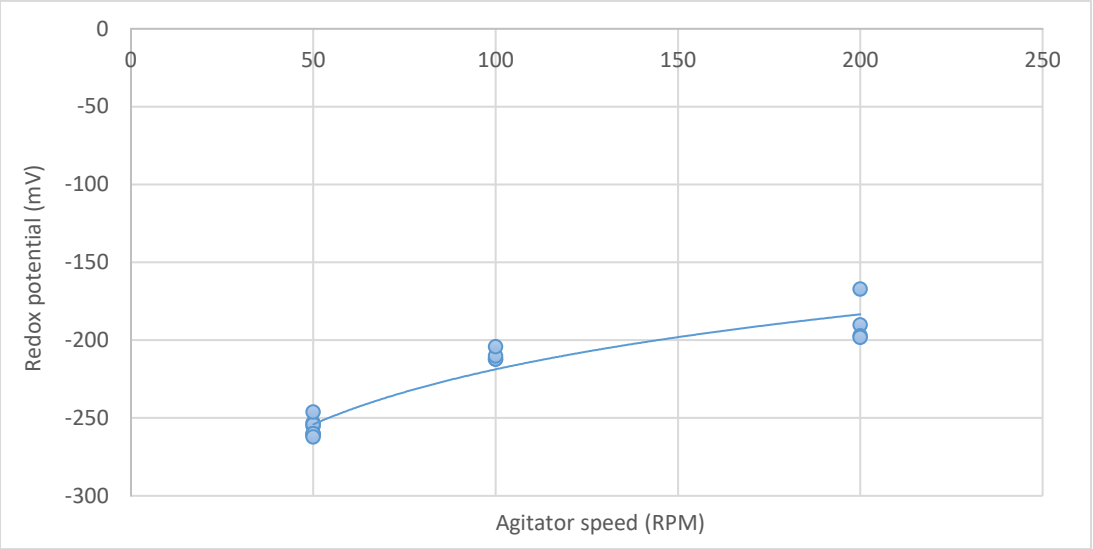


Figure 38: The relationship between agitator RPM and measured culture redox potential from the agitator speed anaerobic pilot growth experiments. The final 4 measurements of culture redox potential were used from each experiment as they were assumed to represent metabolic steady state.

On this basis, it was hypothesised than an agitator rotor RPM value of 75 would result in an average redox potential at metabolic steady state of approximately –239 mV and a metabolic behaviour of *P. thermoglucosidasius* representing mixed acid fermentation, but not producing any ethanol.

For the micro-aerobic glucose 0.075 h^{-1} dilution rate INST-^{13}C MFA experiment, *P. thermoglucosidasius* cells were cultured as described in materials and methods 2.2 1 at an agitator speed of 75 RPM, 60°C, pH 7, 0.05 L/min air (0.14 VVM) and 1L/min N_2 (2.86 VVM) on 0.3% w/v D-Glucose (60% 1-^{13}C, 40% U-^{13}C) ASM (supplemented with 0.042 g/L $FeCl_3$). Over the course of the experiment the average $O.D._{600}$ and redox potential values were a respective 0.585 (± 0.02, n = 19) and -208 mV (± 4.06, n = 19) (Table 25).

Table 25: OD_{600}, DCW and redox values (mV) for the final GMA0.075 micro-aerobic glucose *P. thermoglucosidasius* INST-^{13}C-MFA ILE.

	Sample	12C 6	12C 7	12C 8	13C 1	13C 2	13C 3	13C 4	13C 5	13C 6	13C 7	13C 8	13C 9	13C 10	13C 11	13C 12	13C 13	13C 14	13C 15
GMA0.075	Time (hours)	-4.97	-3.50	-2.00	0.00	1.20	2.17	3.40	6.67	8.00	10.00	20.08	24.00	26.00	30.00	33.33	43.00	46.33	49.67
	Cell Culture volume Changes	-0.37	-0.26	-0.15	0.00	0.09	0.16	0.26	0.50	0.60	0.75	1.51	1.80	1.95	2.25	2.50	3.23	3.48	3.73
	OD_{600}	0.57	0.58	0.59	0.59	0.62	0.62	0.62	0.60	0.58	0.56	0.56	0.61	0.59	0.58	0.58	0.56	0.58	0.56
	DCW g/L	0.24	0.25	0.25	0.25	0.26	0.26	0.26	0.26	0.25	0.24	0.24	0.26	0.25	0.24	0.24	0.24	0.25	0.24
	Redox (mV)	-211	-209	-210	-210	-215	-215	-212	-207	-209	-208	-213	-209	-208	-204	-204	-203	-203	-202

HPLC-RI analysis of the CCFs of the samples taken identified a low average residual concentration of glucose of 0.03 g/L (± 0.02, n = 19) and did not detect the presence of any ethanol (Figure 39).

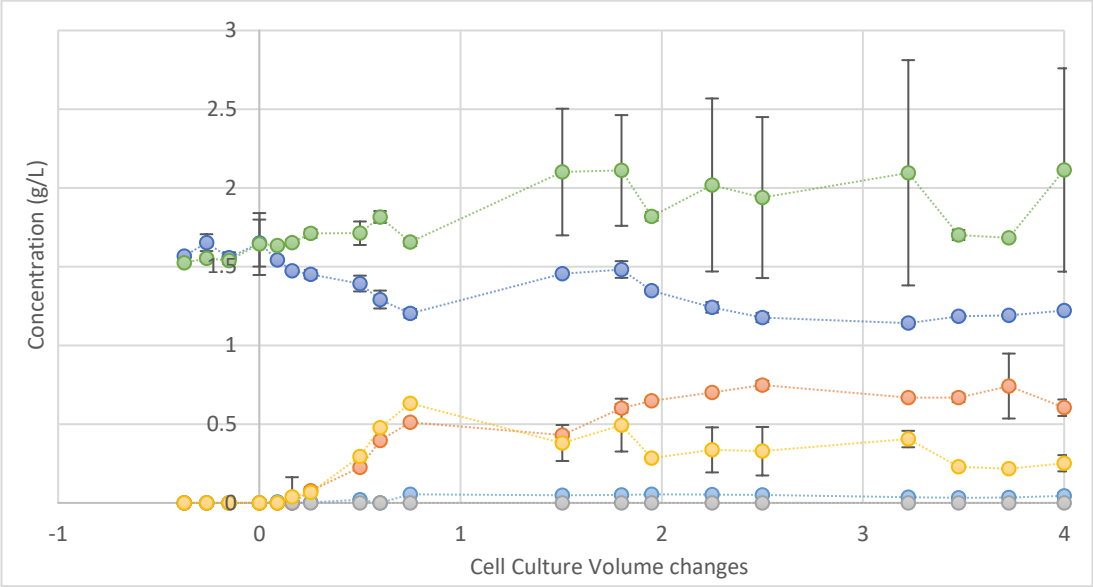


Figure 39: HPLC-RID-UVD analysed CCF products for the final INST-^{13}C-MFA GMA0.075 *P. thermoglucosidasius* ILE. Glucose (light blue), Lactate (orange), Ethanol (grey), Pyruvate (yellow), Formate (dark blue) and Acetate (green). Error is SD for technical triplicate.

HPLC-UV analysis did however demonstrate the presence of the typical *P. thermoglucosidasius* fermentation products of lactate, formate and acetate (Figure 39).
HPLC-UV analysis of the sample CCFs did not detect the presence of pyruvate until the 3rd ^{13}C time point, at a concentration of 0.04 g/L (± 0.12). The detected concentration of pyruvate then rapidly rose over nearly 8 hours to a concentration of 0.63 g/L (± 0.01) before steadily declining over the next 18 hours to a concentration of 0.28 g/L (± 0.01) at the 10th 13C time point. The concentration of pyruvate then remains somewhat constant for the next 17 hours of the experiment, with and average concentration of 0.34 g/L (± 0.06, n = 4), before declining from the 14th to 16th ^{13}C time point to a stable average concentration of 0.23 g/L (± 0.01, n = 3).

The HPLC-UV detected concentrations of acetate in the sample CCFs display two different patterns. Between samples ^{12}C 6 and ^{13}C 7, the average concentration of acetate detected was 1.64 g/L (± 0.20, n = 10), with a low average standard error for each triplicate sample of ± 0.04 g/L (n = 10). However, between ^{13}C samples 8 and 16, the average concentration of acetate detected was 1.95 g/L (± 0.17, n = 9), with a higher average standard error for each triplicate sample of ± 0.36 g/L (n = 9).

The HPLC-UV detected concentrations of lactate and formate demonstrate a roughly inverse correlation to each other. Between samples ^{12}C 6 and ^{13}C 7, the detected concentration of formate declined from 1.57 g/L (± 0.01) to 1.20 g/L (± 0.03), but lactate concentrations increased from below detectable levels to 0.51 g/L (± 0.01). Over the following 10 hours of the experiment, the formate concentration increased to 1.46 g/L (± 0.05), while the lactate concentration slightly decreased to 0.43 g/L (± 0.01). Then concentration of formate then peaks at 1.48 g/L (± 0.05) in the 9th ^{13}C sample, before declining to a steady average concentration for the final 5 ^{13}C CCF samples of 1.18 g/L (± 0.03, n = 5). Concurrently, the concentrations of lactate increase from 0.51 g/L (± 0.01) at 13C 7, until a stable, average

concentration of is observed for the final 5 ^{13}C CCF samples of 0.6 g/L (± 0.05, n = 5).

Discussion

Overall, the stable $O.D._{600}$, redox potential values and the HPLC-RI-UV analysis of the two experiments suggests that both experiments were carbon-limited, and thus growing at a rate near the dilution rate of 0.075 h^{-1}. The observed OD_{600} values for the GAn0.075 ILE are all lower than GA0.075 and GMA0.075, indicating oxygen limitation was observed in growth.

For the GAn0.075 ^{13}C ILE, despite the stable average redox potential value range of -259 mV ± 2.25 (n = 15), the near exponential decrease in lactic acid concentrations across the experiment (R^2 = 0.98) alongside the possible production of ethanol and increases in the OD_{600} values and formate and acetate concentrations suggest that the available oxygen to the increasing *P. thermoglucosidasius* population was decreasing. What that likely translated to on a metabolic level was increasing activity of PFL and alcohol dehydrogenase and a potential deceasing in activity of lactate dehydrogenase. An interesting observation for the GAn0.075 ILE can be made when comparing the: average redox potential values of all samples where ethanol was detected -260 mV (± 2.20, n = 7), all samples where ethanol was detected, adjacent ethanol-lacking samples -259 mV (± 2.21, n = 12) and all samples where ethanol was not detected -258.7 mV (± 2.14, n = 11). This suggests that ethanologenesis and activation of alcohol dehydrogenase in this experiment initiated at a redox potential value of -260 mV and that the INST-^{13}C-MFA experiment captures the metabolic behaviour of *P. thermoglucosidasius* at the point of activation of the production of potential bioethanol. However, as a result, this experiment may not accurately represent *P. thermoglucosidasius* metabolic behaviour under the high ethanol producing anaerobic conditions initially sought.

For the GMA0.075 ^{13}C ILE, the stable observed redox potential value of -208.11 mV combined with stably observed CCF concentrations of fermentation products lactate, acetate and formate suggests that lactate dehydrogenase, phosphotransacetylase/acetate kinase and PFL were active and that the *P. thermoglucosidasius* cells represented the mixed acid fermentation metabolic behaviour sought for micro-aerobic conditions.

4.4.1.3 The xylose ASM aerobic and anaerobic INST-^{13}C-MFA XA0.075 XA0.15 and XAn0.075 ILEs

XA0.075 and XA0.15

For the aerobic xylose 0.075 h^{-1} and 0.015 h^{-1} dilution rate ^{13}C-MFA experiments (XA0.075 and XA0.015, *P. thermoglucosidasius* cells were cultured under the same 60°C, pH 7, 650 rotor RPM and 1.5 VVM air as the glucose experiments. The ASM also contained 0.0125% antifoam 204, but instead used 0.3% ^{D}-Xylose (75% 1-^{13}C, 25% ^{12}C) as the sole carbon substrate. The OD_{600} and redox potential values of the two experiments both display low variation and are similar (Table 26). The average OD_{600} and redox potential values were 2.48 (± 0.08) and -38 mV (± 0.63, n = 15) for the 0.075 h^{-1} experiment and 2.39 (± 0.12) and -37 mV (± 3.54, n = 18) for the 0.15 h^{-1}.

Table 26: OD_{600}, DCW and redox values (mV) for the final XA0.075 and XA0.15 xylose aerobic *P. thermoglucosidasius* INST-^{13}C-MFA ILEs.

XA0.075	Sample	12C 5	12C 6	12C 7	13C 1	13C 2	13C 3	13C 4	13C 5	13C 6	13C 7	13C 8	13C 9	13C 10	13C 11	13C 12			

	Time (hours)	-6.00	-4.42	-2.75	0.00	1.17	2.00	3.33	6.67	14.00	20.00	24.00	27.33	30.67	37.33	40.67				
	Cell Culture volume Changes	-0.49	-0.36	-0.23	0.00	0.10	0.16	0.27	0.55	1.15	1.64	1.97	2.24	2.51	3.06	3.33				
	OD.600	2.37	2.45	2.39	2.39	2.33	2.42	2.48	2.47	2.61	2.58	2.57	2.57	2.52	2.54	2.48				
	DCW g/L	1.00	1.04	1.01	1.01	0.99	1.03	1.05	1.05	1.11	1.09	1.09	1.09	1.07	1.08	1.05				
	Redox (miliVolts)	-6.00	-4.42	-2.75	0.00	1.17	2.00	3.33	6.67	14.00	20.00	24.00	27.33	30.67	37.33	40.67				
XA0.15	Sample	12C 7	12C 8	12C 9	13C 1	13C 2	13C 3	13C 4	13C 5	13C 6	13C 7	13C 8	13C 9	13C 10	13C 11	13C 12	13C 13	13C 14	13C 15	
	Time (hours)	-3.18	-2.25	-1.23	0.00	0.50	1.00	1.67	3.33	4.00	5.05	6.00	7.05	10.00	16.67	18.67	20.33	21.83	23.50	
	Cell culture volume changes	-0.48	-0.34	-0.19	0.00	0.08	0.15	0.25	0.50	0.60	0.76	0.90	1.06	1.50	2.50	2.80	3.05	3.28	3.53	
	OD600	2.33	2.31	2.25	2.32	2.28	2.25	2.32	2.32	2.36	2.30	2.32	2.35	2.35	2.40	2.60	2.48	2.60	2.61	
	DCW g/L	0.99	0.98	0.95	0.98	0.97	0.95	0.98	0.98	1.00	0.98	0.98	1.00	1.00	1.02	1.10	1.05	1.10	1.11	
	Redox (miliVolts)	-39	-40	-39	-39	-39	-38	-37	-37	-33	-32	-32	-31	-31	-39	-34	-39	-40	-42	

HPLC-RI-UV analysis of all CCFs from the 0.075 h^{-1} and 0.15 h^{-1} experiments did not detect the presence of any fermentation products (Figures 40 and 41).

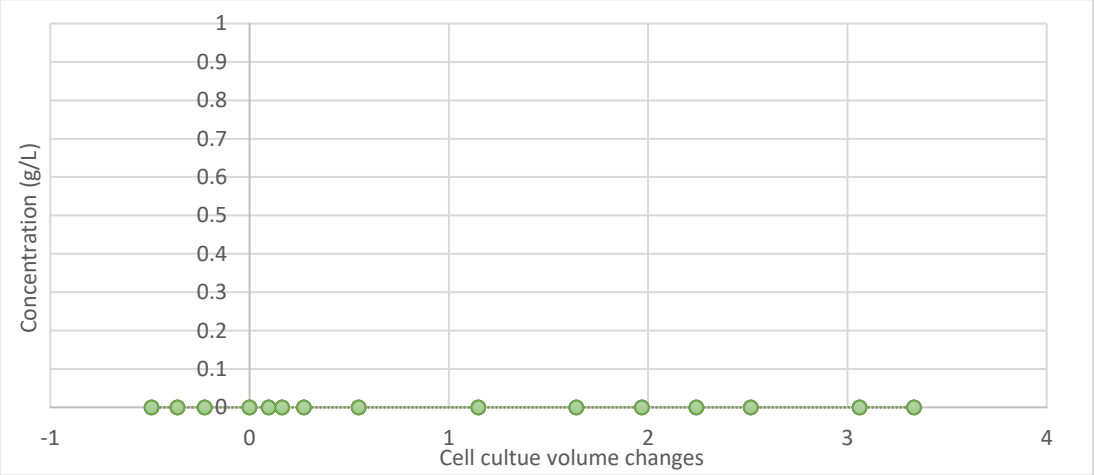


Figure 40: HPLC-RID-UVD analysed CCF products for the final INST-^{13}C-MFA XA0.075 *P. thermoglucosidasius* ILE. Xylose (light blue), Lactate (orange), Ethanol (grey), Pyruvate (yellow), Formate (dark blue) and Acetate (green). Error is SD for technical triplicate.

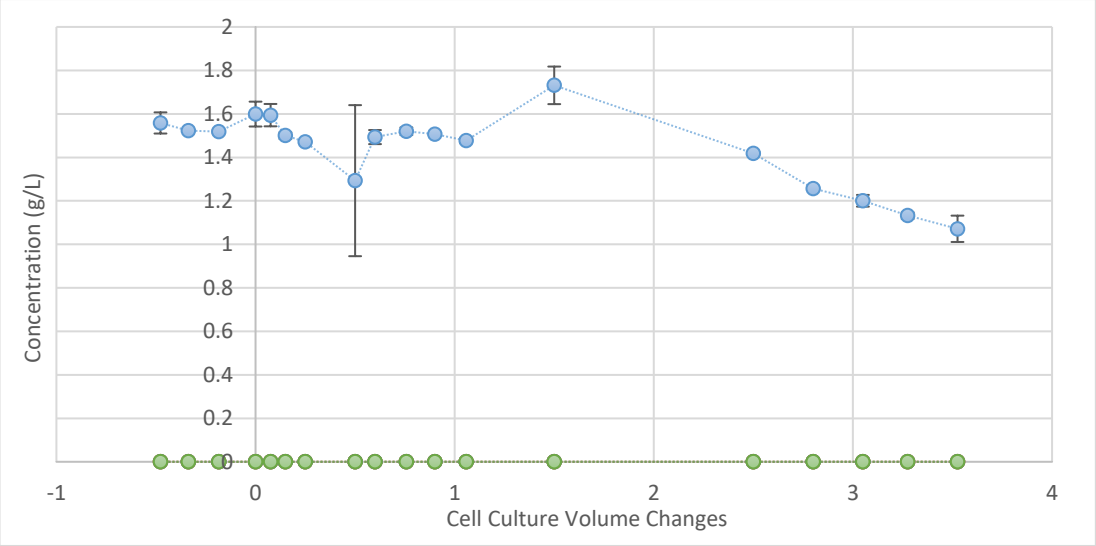


Figure 41: HPLC-RID-UVD analysed CCF products for the final INST-^{13}C-MFA XA0.15 *P. thermoglucosidasius* ILE. Xylose (light blue), Lactate (orange), Ethanol (grey), Pyruvate (yellow), Formate (dark blue) and Acetate (green). Error is SD for technical triplicate.

While HPLC-RI analysis of the CCFs of the 0.075 h^{-1} experiment did not detect the presence of any xylose for the entire experiment, analysis of the CCFs of the 0.15 h^{-1} experiment indicating the presence of fluctuating concentrations of xylose across the experiment. The concentration of xylose remains a roughly average concentration of 1.50 g/L (± 0.10, n = 9) for the first 7.05 hours of the ^{13}C portion of the experiment before peaking at 1.73 g/L (± 0.09). The concentration then undergoes a linear decline (R^2 = 0.90) over the next 16.45 hours to 1.07g/L (± 0.06) and does not reach a stable equilibrium by the end of the experiment.

XAn0.075

In 2014 at the University of Bath, international partner Dr. Shyam Masakapalli used an older iteration of the MKI 45 mL working volume micro bioreactor to grow *P. thermoglucosidasius* NCIMB 11955 cells at 60°C, pH 7 on 0.5% ^{D}-Xylose (75% 1-^{13}C, 25% ^{12}C) ASM, at 0.031 VVM air, 50 rotor RPM and a dilution rate of 0.075 h^{-1}. While this INST-^{13}C-MFA experiment was performed in 2014, the resulting HPLC and GC-MS data had not been analysed until it was provided by Dr. Masakapalli to this research in 2018.

Across the 47-hour duration of the experiment, the OD600 of the cell cultures remained in a range of 0.49-0.52, redox potential of the cultures remained between -251 and -258 mV (Table 27). The concentration of xylose in the cell culture filtrate, as measured by HPLC-RI, remained between 3.92 g/L and 3.70 g/L. HPLC-UV also demonstrated similarly consistent cell culture filtrate concentrations of acetate (0.75-0.87 g/L) formate (0.56-0.66 g/L) and pyruvate (0.33-0.55 g/L) (Figure 42).

Table 27: **OD_{600}, DCW and redox values (mV) for the final XAn0.075 xylose anaerobic *P. thermoglucosidasius* INST-^{13}C-MFA ILE**.

^{12} or ^{13}C Xylose	12	13												
Time (Hours)	**0**	**0.5**	**1**	**2**	**3**	**6**	**8**	**12**	**24**	**28**	**32**	**36**	**47**	**48**
OD600	0.52	0.52	0.52	0.52	0.5	0.52	0.51	0.52	0.5	0.49	0.52	0.52	0.52	0.52
Redox Potential (mV)	-255	-253	-251	-248	-256	-251	-252	-253	-254	-255	-256	-257	-258	-254

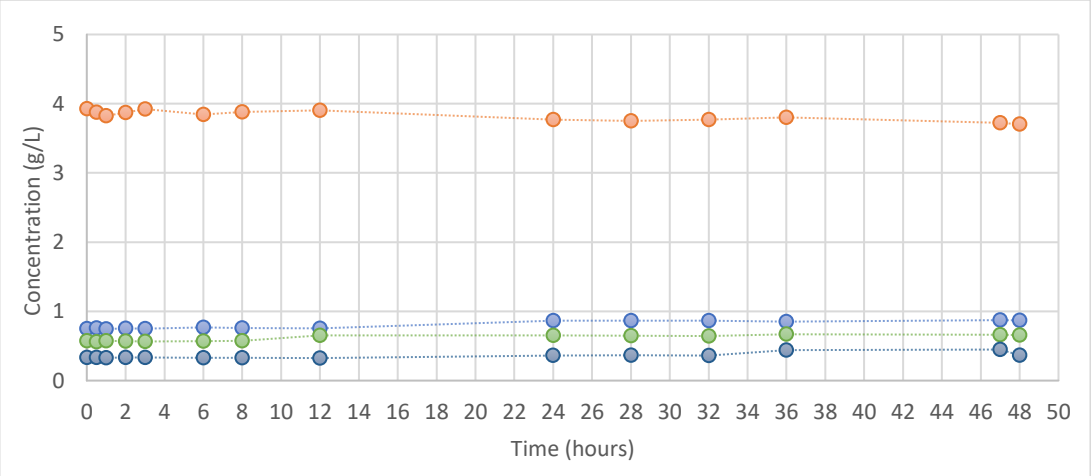


Figure 42: **HPLC-RID-UVD analysed CCF products for the final INST-^{13}C-MFA XAn0.075 *P. thermoglucosidasius* ILE.** Xylose (green), Acetate (blue), Formate (orange) and Pyruvate (purple). n = 1 technical replicate.

Discussion

For the XA0.075 and XA0.15 INST-^{13}C-MFA experiments, all analysed CCFs displayed no presence of fermentation products by HPLC-RI-UV analysis, suggesting that the *P. thermoglucosidasius* cells were growing under aerobic conditions. However, for the XA0.075 INST-^{13}C-MFA experiment the average detected concentration of 1.50 g/L (± 0.10, n = 9) suggested that the cells were not growing under carbon-limited conditions and that the cells may not have achieved a metabolic steady state.

For the XAn0.075 ILE, the stable observed values for OD_{600}, redox potential and HPLC-detected metabolites suggests that anaerobic metabolic steady state was constant throughout the isotope labelling experiment. However, the residual 3.92 g/L and 3.70 g/L (representing in this case a value out of 5 g/L and a final w/v percentage of 0.5% xylose), suggest that it was not fully carbon-limited growth. Notably, the typical anaerobic *P. thermoglucosidasius* NCIMB 11955 fermentation products from glucose feeding of lactic acid and ethanol were not identified in any anaerobic xylose chromatogram suggesting that they may not have been produced by the cells grown on xylose.

4.4.2 Qualitative INST-^{13}C-MFA Results

To describe the ^{13}C-labeling distributions and rates of ^{13}C isotope incorporation across CCM in the 8 *P. thermoglucosidasius* NCIMB 11955 ILEs, TBDMS-derivatized cell protein hydrolysate, consisting of up to 16 TBDMS-derivitized proteinogenic amino acids, was analysed by GC-MS for each technical triplicate sample in each experiment. Chapter 4.4.2. covers the presentation and initial qualitative metabolomic analysis of the resulting MIDs of TBDMS-derivatized amino acid mass fragments.

4.4.2.1 GC separation of TBDMS-derivatized amino acids

Representative chromatograms for each growth condition performed and analysed as part of this research are presented in (Figure 43) and demonstrate low noise baselines with no suggestions of baseline drift or column bleed to impact downstream quantitation:

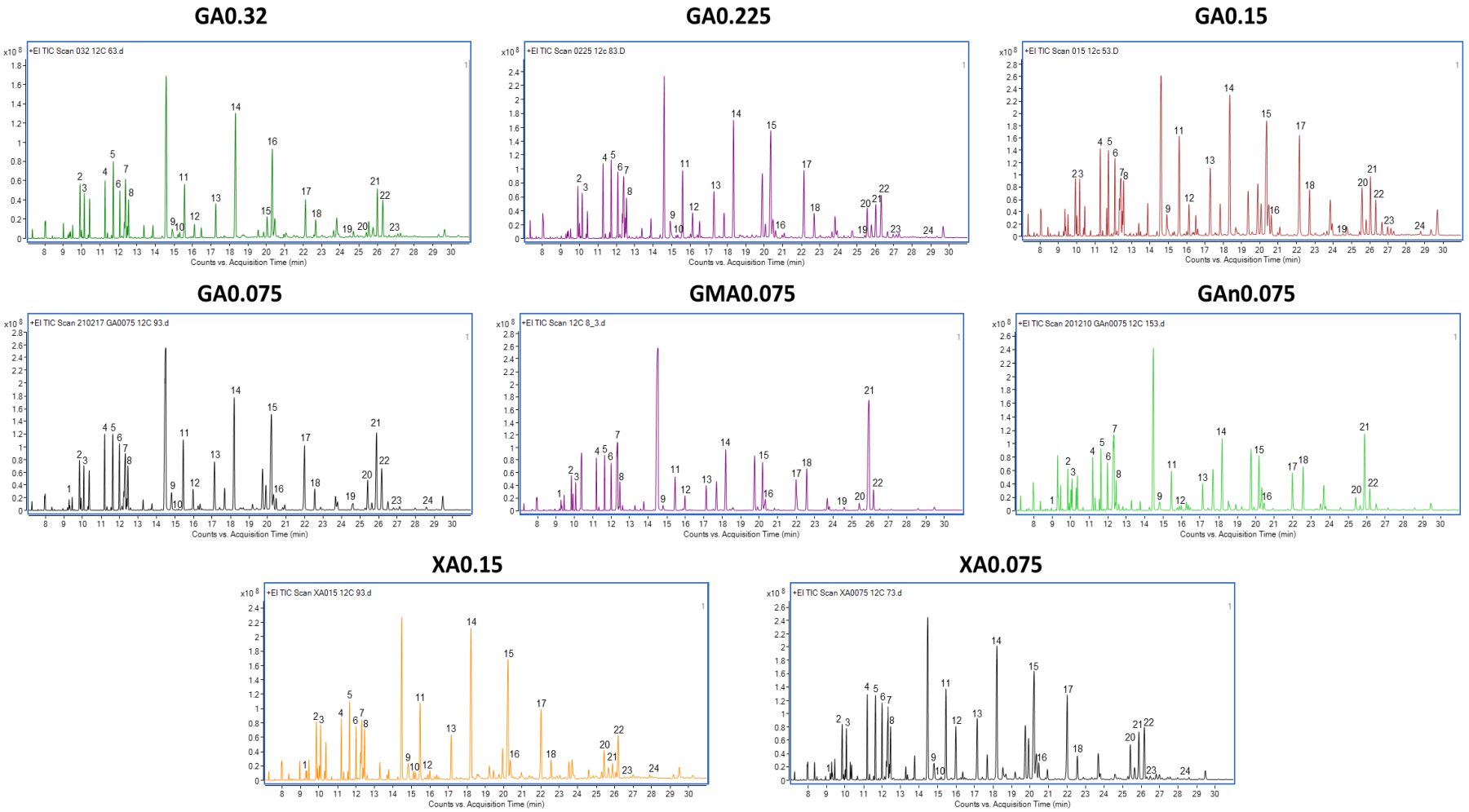


Figure 43: Representative GC chromatograms for each of the INST-13C-MFA ILEs performed in this research. These chromatograms represent the final sample taken before the introduction of the ^{13}C tracer. Identities of numbered peaks are listed in Table 28.

The retention times corresponding to the mass spectrum of fragmented ions associated with each TBDMS-derivitized amino acid, or additionally detected metabolite, are presented in Table 28.

Table 28: Metabolites identified from characteristic mass spectra for each ILE with experiment specific GC retention times.

Metabolite		Experiment-Specific GC Retention Times (mins)					
GC peak	NIST Identity	GA0.075	GA0.15 GA0.225 GA0.32	GAM0.075	GAn0.075	XA0.075	XA0.15
1	Lactic Acid 2TBDMS	9.23:9.31	--	9.23:9.31	9.26:9.31	9.26:9.31	9.25:9.31
2	Alanine 2TBDMS	9.79:9.87	9.87:9.95	9.79:9.87	9.79:9.87	9.79:9.88	9.81:9.89
3	Glycine 2TBDMS	10.02:10.10	10.10:10.18	10.02:10.10	10.04:10.11	10.04:10.11	10.04:10.13
4	Valine 2TBDMS	11.14:11.22	11.23:11.34	11.14:11.22	11.15:11.24	11.15:11.24	11.16:11.24
5	Leucine 2TBDMS	11.59:11.67	11.67:11.76	11.59:11.67	11.59:11.67	11.59:11.69	11.60:11.69
6	Isoleucine 2TBDMS	11.90:12.03	12.03:12.13	11.95:12.03	11.95:12.03	11.95:12.04	11.96:12.05
7	Threonine 2TBDMS	12.18:12.25	12.28:12.37	12.18:12.27	12.20:12.27	12.20:12.29	12.20:12.34
8	Proline 2TBDMS	12.36:12.49	12.50:12.58	12.39:12.49	12.40:12.50	12.40:12.50	12.42:12.50
9	Pyroglutamic Acid 2TBDMS	14.76:14.86	14.89:15.03	14.76:14.86	14.76:14.86	14.76:14.86	14.79:14.90

10	Methionine 2TBDMS	15.04: 15.15	15.17: 15.30	--	--	15.03: 15.17	15.06:1 5.18
11	Serine 3TBDMS	15.36: 15.50	15.49: 15.65	15.36:1 5.50	15.37:1 5.53	15.37: 15.53	15.39:1 5.53
12	Threonine 3TBDMS	15.91: 16.03	16.07: 16.17	15.91:1 6.02	15.91:1 6.04	15.93: 16.04	15.93:1 6.05
13	Phenylalanine 2TBDMS	17.07: 17.18	17.20: 17.34	17.06:1 7.20	17.07:1 7.21	17.07: 17.21	17.07:1 7.21
14	Aspartate 3TBDMS	18.09: 18.24	18.25: 18.38	18.09:1 8.24	18.10:1 8.28	18.10: 18.28	18.12:1 8.26
15	Glutamate 3TBDMS	20.09: 20.24	20.24: 20.38	20.09:2 0.24	20.08:2 0.26	20.08: 20.26	20.11:2 0.26
16	Asparagine 3TBDMS	20.65: 20.78	20.80: 20.94	20.65:2 0.78	20.64:2 0.81	20.64: 20.81	20.67:2 0.81
17	Lysine 3TBDMS	21.91: 22.05	22.04: 22.20	21.91:2 2.08	21.91:2 2.05	21.91: 22.08	21.93:2 2.08
18	Glutamine 3TBDMS	22.63: 22.79	22.78: 22.94	22.82:2 2.92	22.64:2 2.82	22.64: 22.84	22.67:2 2.82
19	Biotin 3TBDMS	24.50: 24.59	25.48: 25.67	24.50:2 4.59	--	--	--
20	Histidine 3TBDMS	25.33: 25.48	25.33: 25.56	25.33:2 5.56	25.33:2 5.49	25.33: 25.53	25.36:2 5.53
21	Citric Acid 4TBDMS	25.79: 25.96	25.96: 26.09	25.80:2 5.98	25.79:2 5.96	25.82: 25.97	25.82:2 5.96

22	Tyrosine 3TBDMS	26.08: 26.23	26.23: 26.39	26.08:26.23	26.08:26.23	26.09: 26.26	26.08:26.23
23	Tryptophan 2TBDMS	26.69: 26.88	26.91: 27.04	--	--	26.74: 26.90	26.69:26.88
24	Cysteine 4TBDMS	28.24: 28.64	28.64: 28.98	--	--	28.23: 28.64	28.23:28.64

The inability to detect arginine for all experiments is as expected from the choice of method as this oxidizes during the drying of the hydrolysate under air. Unexpectedly, mass spectra consistent with TBDMS-derivitized cysteine and tryptophan, commonly reported as lost during the cell hydrolysis process when under oxygen, were detected for several experiments (Table 28). In the case of cysteine, mass spectra indicative 4-TBDMS derivitized cysteine was observed for experiments GA0.075-GA0.32 and XA0.075-XA0.15 at combined retention times of 28.23-28.98 minutes.

In the case of tryptophan, mass spectra indicative of 2-TBDMS derivitized tryptophan was observed for experiments GA0.075-GA0.32, GAM0.075 and XA0.075-XA0.15 at a combined retention time of 26.91-27.05 minutes.

Individual TIC chromatograms were then extracted for each retention time range. Each TIC was then baseline corrected and corrected for natural abundances of heavy isotopes as described in methods 2.2.2. This yielded corrected MIDs for each mass ion fragment.

4.4.2.2 The validity of TBDMS-derivitized molecular mass fragments for further flux analysis

To determine which MIDs were valid for further analysis, all-natural isotope abundance corrected MIDs of all TBDMS-derivitized amino acid mass ion fragments were subjected to a validity check. Each amino acid fragment was qualified as either 'Valid' (possessing an average unlabelled carbon^{-13} enrichment of 0-1.11%), 'Conditionally Valid' (possessing an average unlabelled carbon^{-13} enrichment of ≤ 2.11%) or 'Invalid' (possessing an average unlabelled carbon^{-13} enrichment of >2%) for further analysis as described in methods 2.2.1.5. In cases where no amino acid fragment was qualified as valid, a single conditionally valid fragment was used. The average ^{13}C abundances and resulting validity of 31 amino acid fragments (and 3 mass fragments of TBDMS-derivitized lactic acid) used for further analysis for each experiment are presented in Tables 29, 30, 31, 32. The complete validity tables for all observed mass ion fragments are presented in Appendix 2. None of the observed mass isotopomer fragments representing Asn, Gln, or Trp were found to be valid or conditionally valid in any of the experiments. Mass isotopomer fragments representing Met were only found valid for the XA0.075 and XA0.15 experiments and fragments representing Thr were only found valid for XA0.075.

Table 29: TBDMS-derivtized amino acid mass fragments for the aerobic glucose ILEs and their calculated percentage ^{13}C incorporation after naturally abundant ^{13}C isotope correction. Mass fragments are determined as either 'Valid', 'Conditionally Valid' or 'Invalid'.

GA0.32			GA0.225			GA0.15			GA0.075		
Amino acid fragment		S.D. n = 9	Amino acid fragment		S.D. n = 9	Amino acid fragment		S.D. n =9	Amino acid fragment		S.D. n=3
Ala302	0.20%	0.06%	**Ala302**	0.19%	0.06%	**Ala302**	0.25%	0.05%	**Ala302**	0.15%	0.08%
Gly288	0.22%	0.18%	**Gly288**	0.46%	0.16%	**Gly288**	0.35%	0.16%	**Gly288**	0.32%	0.07%
Val330	0.59%	0.14%	**Val330**	0.67%	0.33%	**Val330**	0.46%	0.11%	**Val330**	0.20%	22.20%
Pro328	3.43%	0.83%	**Pro328**	4.23%	2.70%	**Pro328**	3.04%	0.98%	**Pro328**	3.43%	0.66%
Pro258	0.88%	0.78%	**Pro258**	1.92%	1.37%	**Pro258**	4.58%	1.29%	**Pro258**	3.06%	1.99%
Met377	12.80%	8.78%	**Met377**	30.67%	6.65%	**Met377**	27.13%	5.80%	**Met377**	49.92%	0.14%
Phe378	15.68%	12.81%	**Phe378**	2.39%	0.44%	**Phe378**	3.81%	0.50%	**Phe378**	3.05%	0.32%
Phe336	10.68%	14.02%	**Phe336**	0.91%	0.38%	**Phe336**	1.52%	0.46%	**Phe336**	0.89%	0.44%
Phe308	10.30%	12.54%	**Phe308**	1.18%	0.46%	**Phe308**	1.67%	0.59%	**Phe308**	0.89%	0.39%
Phe234	9.59%	12.65%	**Phe234**	1.13%	0.52%	**Phe234**	1.93%	0.47%	**Phe234**	1.16%	0.59%
Ser432	0.05%	0.06%	**Ser432**	0.19%	0.13%	**Ser432**	0.14%	0.07%	**Ser432**	0.12%	0.06%
Thr446	1.39%	3.92%	**Thr446**	1.39%	3.92%	**Thr446**	1.39%	3.92%	**Thr446**	16.91%	14.04%
Thr404	2.54%	7.18%	**Thr404**	2.54%	7.18%	**Thr404**	2.54%	7.18%	**Thr404**	21.91%	2.07%
Lys488	0.13%	0.03%	**Lys488**	0.19%	0.07%	**Lys488**	0.20%	0.06%	**Lys488**	0.20%	0.06%
Lys473	1.24%	0.40%	**Lys473**	0.94%	0.31%	**Lys473**	0.81%	0.15%	**Lys473**	0.63%	0.60%
Lys431	0.24%	0.05%	**Lys431**	0.30%	0.10%	**Lys431**	0.40%	0.12%	**Lys431**	0.29%	0.19%
Lys329	0.94%	0.89%	**Lys329**	0.75%	0.17%	**Lys329**	0.70%	0.20%	**Lys329**	0.46%	0.16%
Cys304	17.58%	9.50%	**Cys304**	0.31%	0.40%	**Cys304**	0.00%	0.00%	**Cys304**	0.00%	10.05%
Tyr508	9.34%	3.43%	**Tyr508**	1.65%	0.13%	**Tyr508**	2.30%	0.47%	**Tyr508**	1.84%	0.10%
Tyr466	1.26%	0.32%	**Tyr466**	0.79%	0.13%	**Tyr466**	1.04%	0.27%	**Tyr466**	0.43%	0.18%
Tyr438	2.23%	0.38%	**Tyr438**	1.82%	0.29%	**Tyr438**	2.33%	0.54%	**Tyr438**	1.61%	0.21%
Tyr364	2.02%	0.44%	**Tyr364**	0.87%	0.06%	**Tyr364**	1.24%	0.22%	**Tyr364**	0.78%	0.04%
His482	6.37%	2.07%	**His482**	2.88%	0.59%	**His482**	1.60%	0.12%	**His482**	1.75%	0.22%
His440	0.71%	0.09%	**His440**	0.95%	0.57%	**His440**	1.25%	0.75%	**His440**	0.83%	1.63%
Glu489	0.32%	0.09%	**Glu489**	0.27%	0.11%	**Glu489**	0.23%	0.06%	**Glu489**	0.24%	0.11%
Glu474	0.22%	0.05%	**Glu474**	0.23%	0.07%	**Glu474**	0.27%	0.04%	**Glu474**	0.16%	0.07%
Glu404	0.08%	0.02%	**Glu404**	0.32%	0.18%	**Glu404**	0.35%	0.16%	**Glu404**	0.12%	0.02%
Asp460	0.22%	0.07%	**Asp460**	0.19%	0.07%	**Asp460**	0.17%	0.07%	**Asp460**	0.16%	0.09%
Asp390	1.95%	0.95%	**Asp390**	3.76%	1.97%	**Asp390**	6.40%	1.59%	**Asp390**	4.00%	0.28%
Leu344	0.28%	0.05%	**Leu344**	0.49%	0.20%	**Leu344**	0.44%	0.18%	**Leu344**	0.17%	0.11%
ILe344	0.54%	0.17%	**ILe344**	0.94%	0.51%	**ILe344**	0.41%	0.13%	**ILe344**	0.19%	0.11%
Lac303	20.39%	9.92%	**Lac303**	30.65%	15.14%	**Lac303**	34.92%	14.36%	**Lac303**	1.38%	3.81%
Lac261	9.09%	4.07%	**Lac261**	12.44%	5.54%	**Lac261**	19.87%	9.62%	**Lac261**	0.58%	1.38%
Lac233	7.70%	2.62%	**Lac233**	8.56%	2.94%	**Lac233**	11.58%	4.98%	**Lac233**	0.24%	1.18%

Table 30: TBDMS-derivtized amino acid mass fragments for the GA0.075, GMA0.075 and GAn0.075 ILEs and their calculated percentage ^{13}C incorporation after naturally abundant ^{13}C isotope correction. Mass fragments are determined as either 'Valid', 'Conditionally Valid' or 'Invalid'.

GA0.0752			GMA0.075			GAn0.075		
Amino acid fragment		S.D. n=3	Amino acid fragment		S.D. n=9	Amino acid fragment		S.D. n=9
Ala302	0.15%	0.08%	**Ala302**	0.15%	0.08%	**Ala302**	1.10%	0.15%
Gly288	0.32%	0.07%	**Gly288**	0.41%	0.21%	**Gly288**	1.34%	3.13%
Val330	0.20%	22.20%	**Val330**	0.65%	0.19%	**Val330**	1.50%	7.67%

Pro328	3.43%	0.66%	**Pro328**	8.11%	1.28%	**Pro328**	7.98%	2.62%
Pro258	3.06%	1.99%	**Pro258**	0.77%	0.81%	**Pro258**	3.20%	4.25%
Met377	49.92%	0.14%	**Met377**	16.81%	7.54%	**Met377**	37.45%	3.21%
Phe378	3.05%	0.32%	**Phe378**	5.10%	1.12%	**Phe378**	7.22%	14.03%
Phe336	0.89%	0.44%	**Phe336**	0.79%	0.10%	**Phe336**	1.79%	0.40%
Phe308	0.89%	0.39%	**Phe308**	0.71%	0.11%	**Phe308**	2.28%	1.45%
Phe234	1.16%	0.59%	**Phe234**	0.44%	0.30%	**Phe234**	1.66%	0.14%
Ser432	0.12%	0.06%	**Ser432**	0.12%	0.08%	**Ser432**	1.10%	0.36%
Thr446	16.91%	14.04%	**Thr446**	1.87%	3.78%	**Thr446**	0.00%	1.39%
Thr404	21.91%	2.07%	**Thr404**	2.74%	7.11%	**Thr404**	0.00%	2.54%
Lys488	0.20%	0.06%	**Lys488**	0.16%	0.05%	**Lys488**	1.11%	0.73%
Lys473	0.63%	0.60%	**Lys473**	2.53%	1.13%	**Lys473**	1.91%	0.62%
Lys431	0.29%	0.19%	**Lys431**	0.23%	0.06%	**Lys431**	1.20%	0.66%
Lys329	0.46%	0.16%	**Lys329**	0.78%	0.37%	**Lys329**	1.67%	2.18%
Cys304	0.00%	10.05%	**Cys304**	0.00%	0.00%	**Cys304**	0.63%	22.09%
Tyr508	1.84%	0.10%	**Tyr508**	0.87%	0.32%	**Tyr508**	3.56%	6.26%
Tyr466	0.43%	0.18%	**Tyr466**	0.40%	0.05%	**Tyr466**	1.84%	0.85%
Tyr438	1.61%	0.21%	**Tyr438**	0.88%	0.19%	**Tyr438**	3.04%	0.59%
Tyr364	0.78%	0.04%	**Tyr364**	1.04%	0.10%	**Tyr364**	2.30%	0.70%
His482	1.75%	0.22%	**His482**	4.25%	1.40%	**His482**	3.58%	1.33%
His440	0.83%	1.63%	**His440**	0.54%	0.19%	**His440**	1.49%	2.98%
Glu489	0.24%	0.11%	**Glu489**	0.49%	0.18%	**Glu489**	1.64%	0.59%
Glu474	0.16%	0.07%	**Glu474**	0.42%	0.13%	**Glu474**	1.31%	0.40%
Glu404	0.12%	0.02%	**Glu404**	0.11%	0.01%	**Glu404**	1.36%	1.49%
Asp460	0.16%	0.09%	**Asp460**	0.20%	0.09%	**Asp460**	1.24%	0.03%
Asp390	4.00%	0.28%	**Asp390**	0.70%	1.35%	**Asp390**	2.06%	2.35%
Leu344	0.17%	0.11%	**Leu344**	0.37%	0.09%	**Leu344**	1.35%	1.70%
ILe344	0.19%	0.11%	**ILe344**	0.46%	0.09%	**Ile344**	1.47%	0.52%
Lac303	1.38%	3.81%	**Lac303**	1.46%	0.45%	**Lac303**	1.68%	0.17%
Lac261	0.58%	1.38%	**Lac261**	0.19%	0.17%	**Lac261**	12.18%	0.69%
Lac233	0.24%	1.18%	**Lac233**	0.12%	0.08%	**Lac233**	11.23%	0.87%

For all glucose experiments, GMA0.075 had the greatest number of valid fragments for further analysis (23) while GAn0.075 had the fewest valid (4) and most conditionally valid fragments (16). This comparably poor result for the GAn0.075 ^{13}C ILE was likely down to smaller total biomass concentrations in each 500 µL sample taken forward for further analysis. As a result, these values are potentially a reflection of a worse signal to noise ratio resulting lower masses of derivitized amino acid fragments detected compared to the noise baseline. While quantity of biomass is perhaps less important for determining the MID ^{13}C-labelling ratios needed for further flux analysis, this result suggests that future work in this area should use the measured OD_{600} readings to standardize the amount of cellular material per sample taken forward for further analysis to improve the obtained mass ion signal.

Table 31: TBDMS-derivtized amino acid mass fragments for the XA0.075 and XA0.15 ILEs and their calculated percentage ^{13}C incorporation after naturally abundant ^{13}C isotope correction. Mass fragments are determined as either 'Valid', 'Conditionally Valid' or 'Invalid'.

XA0.075	XA0.15

Amino acid fragments		S.D	Amino acid fragments		S.D
Ala302	0.78%	1.80%	**Ala302**	0.21%	0.12%
Gly288	0.33%	0.13%	**Gly288**	0.36%	0.25%
Val330	0.08%	0.20%	**Val330**	0.36%	0.21%
Pro328	0.33%	0.33%	**Pro328**	0.67%	0.80%
Pro258	1.50%	1.50%	**Pro258**	1.23%	2.01%
Met377	0.13%	0.13%	**Met377**	0.15%	0.18%
Phe378	0.74%	0.31%	**Phe378**	0.35%	0.46%
Phe336	0.77%	0.46%	**Phe336**	0.98%	0.51%
Phe308	0.61%	0.43%	**Phe308**	1.28%	0.44%
Phe234	1.00%	0.58%	**Phe234**	1.03%	0.67%
Ser432	0.10%	0.12%	**Ser432**	0.04%	0.06%
Thr446	0.23%	0.20%	**Thr446**	5.05%	9.41%
Thr404	0.09%	0.07%	**Thr404**	3.42%	6.38%
Lys488	0.07%	0.02%	**Lys488**	0.26%	0.22%
Lys473	0.95%	0.24%	**Lys473**	0.85%	0.03%
Lys431	0.13%	0.04%	**Lys431**	0.31%	0.25%
Lys329	0.52%	0.14%	**Lys329**	0.94%	0.45%
Cys304	7.47%	10.72%	**Cys304**	3.23%	9.12%
Tyr508	0.26%	0.13%	**Tyr508**	0.06%	0.08%
Tyr466	0.43%	0.14%	**Tyr466**	0.14%	0.17%
Tyr438	1.05%	0.77%	**Tyr438**	0.16%	0.19%
Tyr364	0.42%	0.06%	**Tyr364**	0.21%	0.26%
His482	0.52%	0.31%	**His482**	0.31%	0.41%
His440	0.87%	0.30%	**His440**	1.04%	0.16%
Glu489	0.25%	0.07%	**Glu489**	0.33%	0.26%
Glu474	0.16%	0.05%	**Glu474**	0.25%	0.16%
Glu404	0.01%	0.01%	**Glu404**	0.33%	0.32%
Asp460	0.14%	0.09%	**Asp460**	0.22%	0.19%
Asp390	1.17%	1.07%	**Asp390**	1.17%	2.32%
Leu344	0.18%	0.09%	**Leu344**	0.24%	0.17%
ILe344	0.17%	0.05%	**ILe344**	0.51%	0.44%
Lac303	0.00%	0.00%	**Lac303**	0.00%	2.06%
Lac261	0.47%	0.47%	**Lac261**	0.31%	0.31%
Lac233	0.16%	0.16%	**Lac233**	0.17%	0.17%

For the XA0.075 and XA0.15 ^{13}C ILEs performed as part of this research in the MKII bioreactor system, XA0.075 had 28 amino acid fragments deemed valid with only 2 fragments deemed conditionally valid, whereas XA0.15 had 26 amino acid fragments deemed valid and only 2 fragments deemed conditionally valid. Notably however, for the data set provided by Shyam Maskapalli from his anaerobic xylose ILE in 2014, a different range of amino acid fragments were qualified as part of this research as valid, conditionally valid, and invalid Table 32.

Table 32: TBDMS-derivtized amino acid mass fragments for the XAn0.075 ILE and their calculated percentage ^{13}C incorporation after naturally abundant ^{13}C isotope correction. Mass fragments are determined as either 'Valid', 'Conditionally Valid' or 'Invalid'.

XAn0.075	
Amino acid fragments	
Ala302	23.65%
Ala232	0.71%
Gly288	3.66%
Val330	1.95%
Val288	1.14%
Pro328	1.25%
Pro258	2.02%
Met377	No data.
Phe378	0.59%
Phe336	0.99%
Phe308	1.16%
Phe234	1.55%
Ser432	8.65%
Ser362	1.07%
Thr446	0.00%
Thr404	3.19%
Lys488	3.56%
Lys473	0.00%
Lys431	2.06%
Lys329	5.21%
Cys304	No data.
Tyr508	No data.
Tyr466	3.69%
Tyr438	3.36%
Tyr364	4.53%
Tyr302	1.47%
His482	No data.
His440	No data.
Glu489	4.29%
Glu474	3.23%
Glu404	1.48%
Glu330	0.97%
Asp460	6.37%
Asp418	1.32%
Asp390	2.90%
Leu344	5.64%
Leu200	3.91%
Ile344	3.95%
Ile200	3.62%
Lac303	No data.
Lac261	No data.
Lac233	No data.

Additionally, 2-TBDMS derivitized lactic acid fragments Lac261 and Lac233 were found to be valid for experiments GA0.075, GMA0.075, XA0.075 and XA0.15. While this suggests that trace cell culture carried over with the cell pellet, these observations suggest that ^{13}C labelling patterns of TBDMS derivitized lactic acid can be inferred alongside amino acids.

4.4.3 Isotopically instationary ^{13}C-labelling and MIDs across ILEs

To compare the MIDs and overall extents of ^{13}C incorporation for the proteinogenic amino acid mass fragments across the different dilution rate experiments, each data set was normalized to the CCVCs of the different dilution rates (as displayed in Table 21). Furthermore, to broadly compare different labelling patterns in different pathways of CCM, the isotopically instationary ^{13}C labelling data for all experiments is presented in terms of assumed CCM node (Figure 44). Across the different dilution rate, aeration and sugar substrate experiments, all potentially proteinogenic TBDMS-derivitized amino acid mass fragments demonstrate an observable increase in ^{13}C isotopic labelling over increasing CCVCs. Furthermore, these CCVC normalised rates do differ between experiments as do the mass isotopomers (M+0, M+1…M+$_n$) which predominate for each observed amino acid.

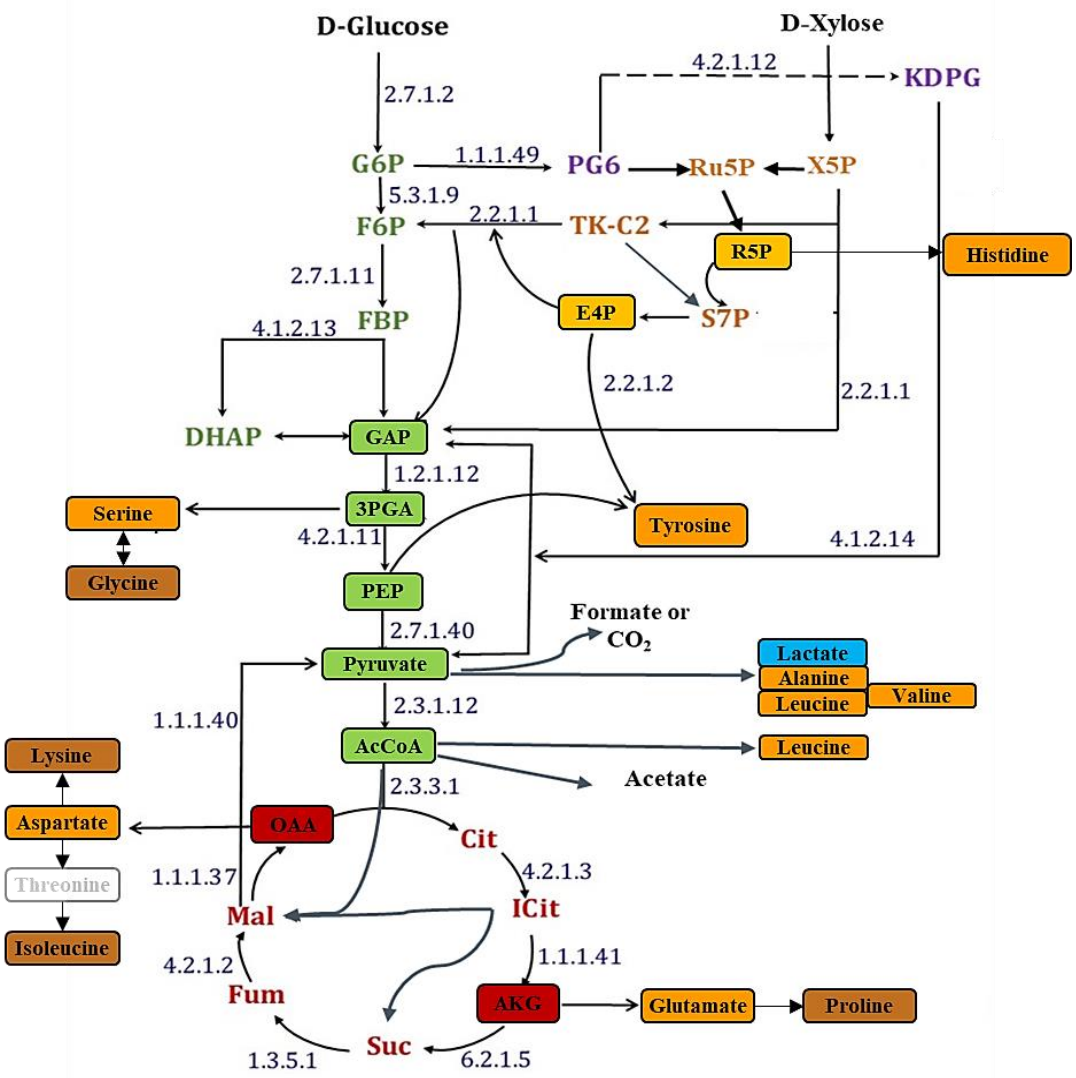


Figure 44: Simplified schematic of assumed *P. thermoglucosidasius* central carbon metabolism framing the presented amino acids and nodes. Boxed intermediate metabolites highlight precursors involved with amino acid metabolism discussed below. Boxed amino acids highlighted in orange represent discussed amino acids which function as intermediates for other amino acids. Threonine phased out in white represents its involvement in the biosynthetic route from aspartate to isoleucine, but that it is not discussed here. The remaining amino acids discussed are highlighted in brown. Central carbon metabolites are highlighted in green (glycolysis), yellow (pentose phosphate pathway), purple (Entner-Doudoroff pathway) and red (TCA cycle). The fermentation metabolites discussed in the following analysis are highlighted blue. Available metabolic paths informed from genome sequence of *P. thermoglucosidasius* DSM 2542, complete with enzyme commission (EC) numbers (purple text). AcCoA (acetyl-coenzyme A), AKG (α-ketoglutarate), Cit (citrate), Chor (chorismate), DHAP (dihydroxyacetone phosphate), E4P (erythrose-4-phosphate), FBP (fructose-bis-phosphate), Fum (fumarate), GAP (glyceraldehyde-3-phosphate), G6P (glucose-6-phosphate), 6PG (6-phosphogluconate), ICit (isocitrate), Mal (malate), OAA (oxaloacetate), F6P (fructose-6-phosphate), PEP (phosphoenolpyruvate), 3PGA (3-phosphoglycerate), R5P (ribose-5-phosphate), Ru5P (ribulose-5-phosphate), S7P (sedo-heptulose-7-phosphate), Suc (succinate), TK-C2 (transketolase C-2) and Xu5P (xylulose-5-phosphate). Adapted from Shree *et al.* (2018) 201, Tang *et al.* (2009)[128] and Cordova and Antoniewicz (2016)[127].

4.4.3.1 Nodes: R5P – His440 E4P/PEP Tyr466, Phe 336

Beginning at the pentose phosphate pathway, three observed TBDMS-derivitized amino acids can receive direct contributions from flux through this pathway. The first is six-carbon histidine (derived from R5P and represented here by mass fragment His440), the second is nine-carbon tyrosine (represented by mass fragment Tyr466) and the third is nine-carbon phenylalanine (represented by mass fragment Phe336). Both tyrosine and phenylalanine are derived from a combination of PPP metabolite E4P and glycolysis intermediate PEP.

Comparing GA0.075 GA0.15 GA0.225 and GA0.32

The first set of experiments which can be directly compared are the aerobic, 0.075, 0.15, 0.225 and 0.32 h^{-1} dilution rate ILEs where 0.3% w/v (60% 1-^{13}C and 40% U-^{13}C) ^{D}-glucose was the sole substrate (GA0.075, GA0.15, GA0.225 and GA0.32).

For the GA0.15, GA0.225 and GA0.32 ILEs, an isotopic steady state is observable for the MIDs of the His440, Tyr466 and Phe366 mass fragment populations by the 3rd CCVC but is not obviously achieved for GA0.075 (Figure. 45). The respective final totals of ^{13}C labelling for all three amino acids observed for GA0.32 were: 90.7% for His440, 93.4% for Tyr466 and 93.4% for Phe 336 (Fig. 45, graph column 1), observed for GA0.225 were: 94.6% for His440, 93.9% for Tyr466 and 93.9% for Phe 336 (Fig.45, graph column 2) and observed for GA0.15 were: 92.3% for His440, 93.2% for Tyr466 and 93.5% for Phe 336 (Fig. 45, graph column 3). GA0.075 demonstrated lower observed final total ^{13}C labelling values of: 68.7% for His440, 69.7% for Tyr466 and 69.6% for Phe 336 (Fig.45, graph column 4).

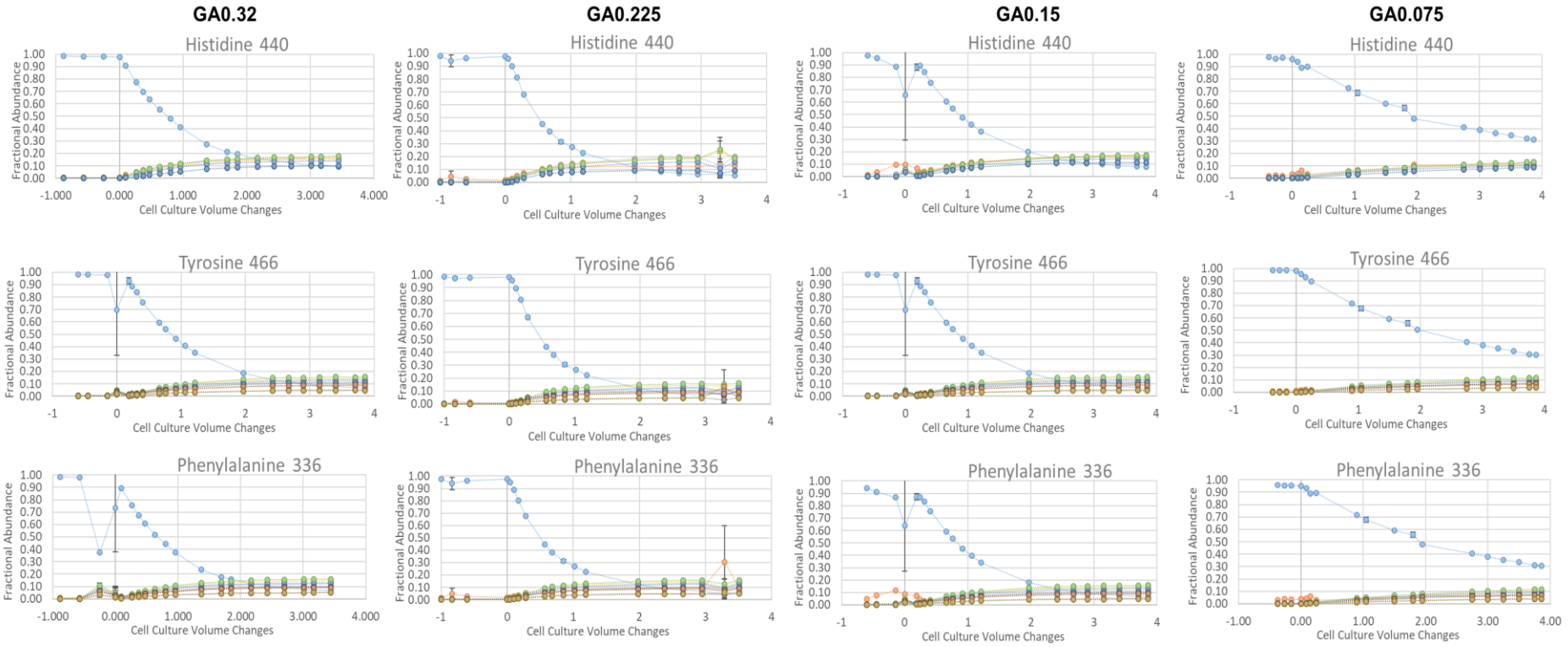


Figure 45: PPP derived TBDMS-derivitized amino acid MIDs for mass fragments His440, Tyr466 and Phe336 for the GA0.32, GA0,225, GA0.15 and GA0.075 of the INST-13C-MFA ILEs performed in this research. M+0 (Light blue), M+1 (Orange), M+2 (Light Grey), M+3 (Yellow), M+4 (Blue), M+5 (Green), M+6 (Dark Blue), M+7 (Dark Brown), M+8 (Dark Grey), M+9 (Light Brown). Reported values are means of technical triplicate samples with error bars representing SD.

Across the His440, Tyr466 and Phe336 mass fragments of the GA0.15, GA0.225 and GA0.32 ILEs, the mass isotopomer with the highest abundance for 6 carbon histidine was either the M+2 or M+5 mass isotopomer and for 9 carbon tyrosine and phenylalanine was the M+5 mass isotopomer. This resulted in steady state fractional abundances of: 0.173 (± 0.001) for His440 M+2, 0.158 (± 0.002) for Tyr466 M+5 and 0.158 (± 0.001) for Phe336 M+5 in GA0.32, steady state fractional abundances of 0.199 (± 0.00) for His440 M+2, 0.159 (± 0.002) for Tyr466 M+5 and 0.158 (± 0.001) for Phe336 M+5 in GA0.225, steady state fractional abundances of 0.313 (± 0.001) for His440 M+5, 0.157 (± 0.001) for Tyr466 M+5 and 0.158 (± 0.001) for Phe336 M+5 observed in GA0.15. For the GA0.075 ILE there were observed final fractional abundances of 0.313 (± 0.001) for the His440 M+0 mass fragment, 0.303 (± 0.002) for the Tyr466 M+0 mass fragment and 0.304 (± 0.003) for the Phe336 M+5 mass fragment observed in GA0.075.

Comparing GA0.075 GMA0.075 GAn0.075

The second set of experiments which can be directly compared are the aerobic 0.075 h^{-1}, micro-aerobic 0.075 h^{-1} and anaerobic 0.075 h^{-1} dilution rate ILEs where 0.3% w/v (60% 1-^{13}C and 40% U-^{13}C) D-glucose was the sole substrate (GA0.075, GAM0.075 and GAn0.075). In addition to the aforementioned GA0.075 ILE, for the GAM0.075 ILE it appears that an isotopic steady state may have been reached for the MIDs of the His440, Tyr466 and Phe366 mass fragment populations by the final 4 CVCC time point, however this is inconclusive. Alternatively, for the GAn0.075 ILEs, an isotopic steady state is observable from the 3.56 CCVC time point (Figure. 46). The respective final totals of ^{13}C labelling for all three amino acids observed for GAM0.075 were: 84.4% for His440, 90.5% for Tyr466 and 90.4% for Phe 336 (Figure. 46, graph column 2) and for GAn0.075 were: 87.4% for His440, 92.9% for Tyr466 and 92.6% for Phe 336 (Figure. 46, graph column 3).

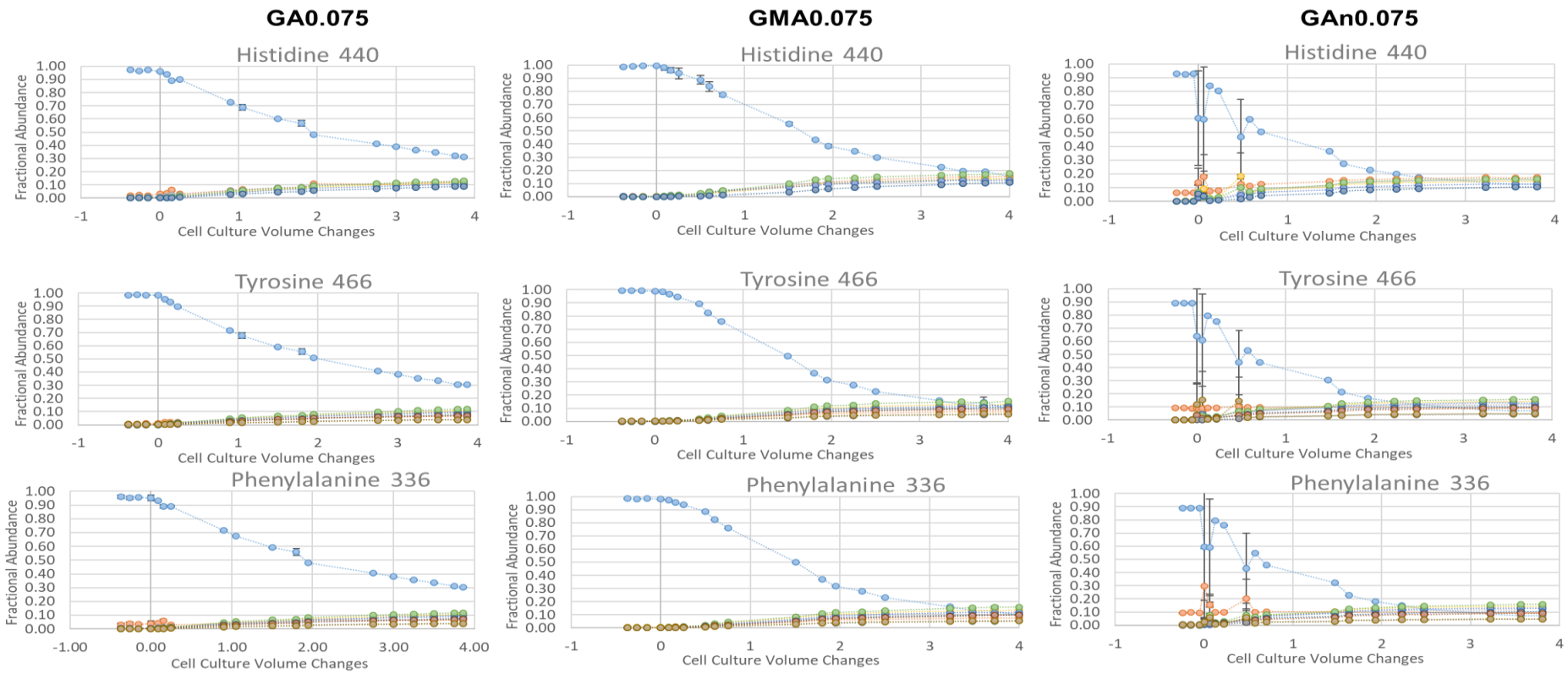


Figure 46: PPP derived TBDMS-derivitized amino acid MIDs for mass fragments His440, Tyr466 and Phe336 for the GA0.075, GMA0,075, GAn0.075 of the INST-^{13}C-MFA ILEs performed in this research. M+0 (Light blue), M+1 (Orange), M+2 (Light Grey), M+3 (Yellow), M+4 (Blue), M+5 (Green), M+6 (Dark Blue), M+7 (Dark Brown), M+8 (Dark Grey), M+9 (Light Brown). Reported values are means of technical triplicate samples with error bars representing SD.

Across the His440, Tyr466 and Phe336 mass fragments of the GAM0.075 and GAn0.075 ILEs, the mass isotopomer with the highest abundance for 6 carbon histidine varied, but for 9 carbon tyrosine and phenylalanine was again the M+5 mass isotopomer. In addition to the aforementioned GA0.075 results, this resulted steady state fractional abundances of 0.176 (± 0.002) for His440 M+2, 0.155 (± 0.003) for Tyr466 M+5 and 0.157 (± 0.001) for Phe336 M+5 observed in GAM0.075. For GAn0.075, the highest fractional abundances for histidine were close between the His440 M+1 and M+5 for mass fragments with 0.173 (± 0.002) for His440 M+1 vs 0.165 (± 0.002) for His440 M+5. For tyrosine and phenylalanine, the M+5 mass fragment predominated with observed abundances of 0.176 (± 0.004) for Tyr466 M+5 and 0.157 (± 0.002) for Phe336 M+5.

Comparing XA0.075 XA0.15 XAn0.075

The third set of experiments which can be directly compared are the xylose aerobic 0.075 h^{-1} and xylose aerobic 0.15 h^{-1} dilution rate ILEs performed in this research (XA0.075, XA0.15) and the xylose anaerobic 0.075 h^{-1} dilution rate ILE (XAn0.075). In the case of XA0.075 and XA0.15, 0.3% w/v (75% 1-^{13}C and 25% natural labelled) ^{D}-xylose was the sole substrate and in the case of XAn0.075, 0.5% w/v (75% 1-^{13}C and 25% natural labelled) ^{D}-xylose was the sole substrate. For the XA0.075 an XA0.15 ILEs, an isotopic steady state is observable for the MIDs of the His440, Tyr466 and Phe336 mass fragment populations again by the 3rd CCVC (Figure. 47). However, for the XAn0.075 ILE, while an isotpic steady state is achieved for Tyr466 and Phe336, no mass fragments of TBDMS-derivatized histidine were observed.

As a result of the 75% 1-^{13}C substrate labelling pattern, the respective final totals of ^{13}C labelling for all three amino acids were far lower than the glucose experiments, reaching final percentages observed for XA0.075 of: 79.6% for His440, 58.2% for Tyr466 and 58.5% for Phe 336 (Figure. 47, graph column 1), for XA0.15 of: 81.3% for His440, 60.0% for Tyr466 and 59.7% for Phe 336 (Figure. 47, graph column 2) and for XAn0.075 of: 58.2% for Tyr466 and 54.6% for Phe336 (Figure. 47, graph column 3).

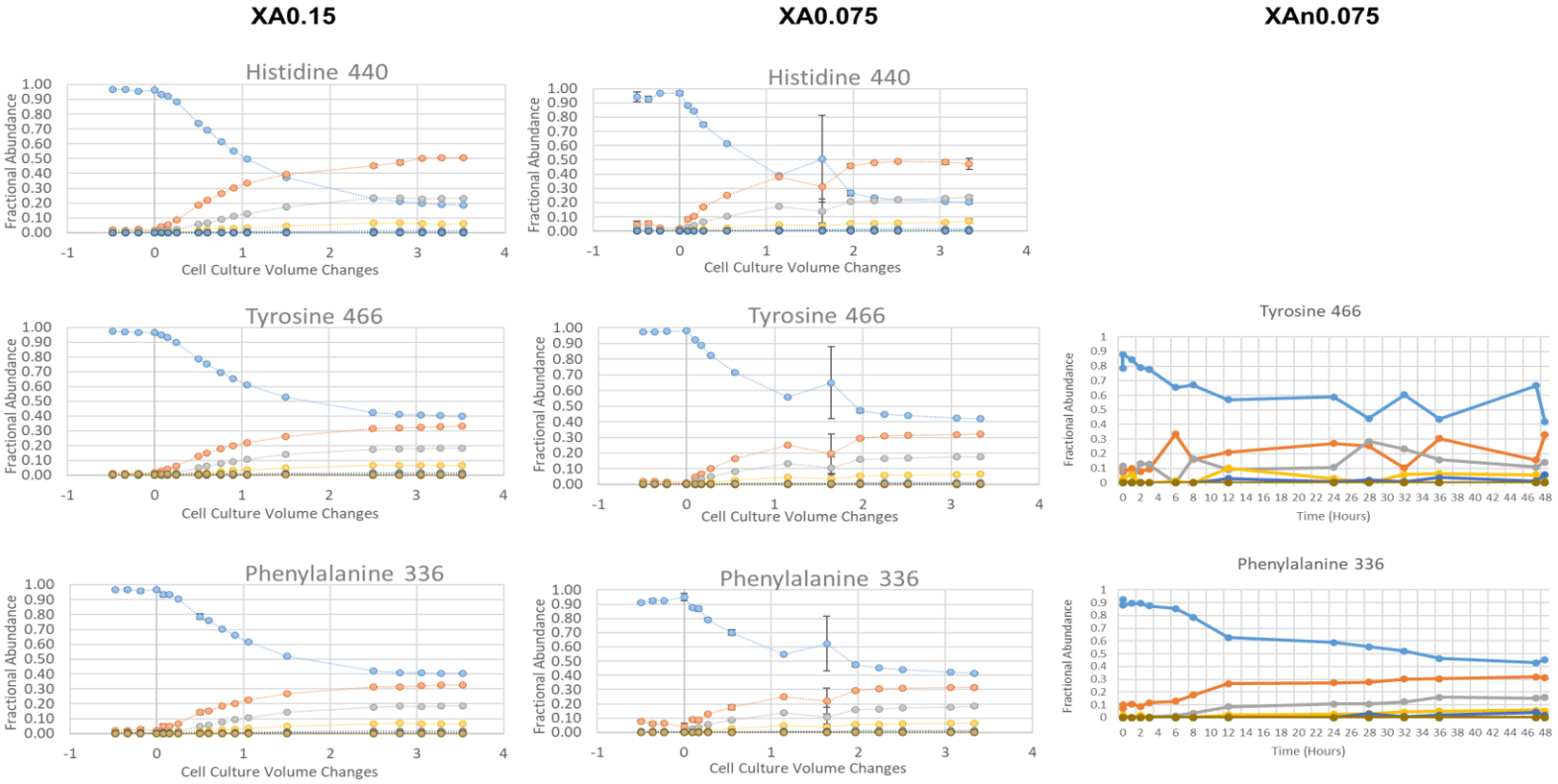


Figure 47: PPP derived TBDMS-derivitized amino acid MIDs for mass fragments His440, Tyr466 and Phe336 for the XA0.15, XA0,075, XAn0.075 of the INST-13C-MFA ILEs performed in this research. His440 was not detected for the XAn0.075. M+0 (Light blue), M+1 (Orange), M+2 (Light Grey), M+3 (Yellow), M+4 (Blue), M+5 (Green), M+6 (Dark Blue), M+7 (Dark Brown), M+8 (Dark Grey), M+9 (Light Brown). Reported values are means of technical triplicate samples with error bars representing SD.

As a further result of the 75% 1-^{13}C substrate labelling pattern, across the observed His440, Tyr466 and Phe336 mass fragments, the mass isotopomer with the highest abundance for 6 carbon histidine, when observed, was the M+1 mass isotopomer. For 9 carbon tyrosine and phenylalanine it was predominantly the M+0 unlabelled mass isotopomer. For each experiment, this resulted in steady state fractional abundances of 0.472 (± 0.039) for His440 M+1, 0.418 (± 0.002) for Tyr466 M+0 and 0.415 (± 0.003) for Phe336 M+0 in XA0.075, steady state fractional abundances of 0.506 (± 0.004) for His440 M+1, 0.400 (± 0.001) for Tyr466 M+0 and 0.403 (± 0.001) for Phe336 M+1 in XA0.15 and steady state fractional abundances of 0.418 for Tyr466 M+0 and 0.454 for Phe336 M+0 in XAn0.075.

4.4.3.2 Node: 3PG - Ser432, Gly288

'Descending' from upper CCM, the next metabolic intermediate node of interest is 3-phosphoglycerate (3PG), of which 3 carbon serine and two carbon serine-derivative glycine are direct products. The representative valid TBDMS- derivitized amino acid mass fragments of these amino acids were Ser432 and Gly288.

Comparing GA0.075, GA0.15, GA0.225 and GA0.32

For the GA0.15, GA0.225 and GA0.32 ILEs, an isotopic steady state is observable for the MIDs of the Ser432, and Gly288 mass fragment populations by the 3rd CCVC, but again is not obviously achieved for GA0.075 (Figure. 48).

The respective final totals of ^{13}C labelling for all three amino acids for GA0.32 were: 69.8% for Ser432 and 42.8% for Gly288 (Figure.48, graph column 1), for GA0.225 were: 72.0% for Ser432 and 43.2% for Gly288 (Figure. 48, graph column 2) and GA0.15 were: 69.3% for Ser432 and 44.8% for Gly288 (Figure. 48, graph column 3). GA0.075 demonstrated lower final enrichment values of: 54.5% for Ser432 and 35.9% for Gly288 (Figure. 48 graph column 4).

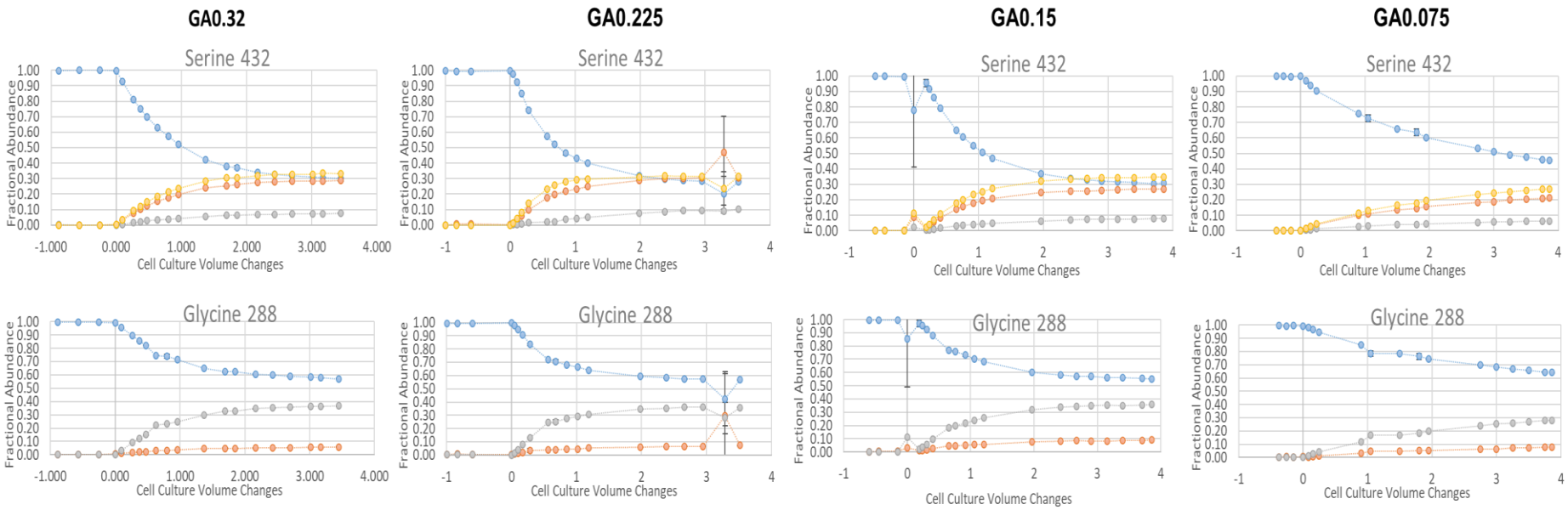


Figure 48: 3PG derived TBDMS-derivitized amino acid MIDs for mass fragments Ser432 anf Gly288 for the GA0.32, Ga0.225, GA0.15 and GA0.075 of the INST-13C-MFA ILEs performed in this research. M+0 (Light blue), M+1 (Orange), M+2 (Light Grey), M+3 (Yellow), Reported values are means of technical triplicate samples with error bars representing SD.

Across the Ser432 and Gly288 fragments, the mass isotopomer with the highest abundance for 3 carbon serine was predominantly the M+3 mass isotopomer and for 2 carbon glycine was the M+0 mass isotopomer. For each experiment, this resulted in final fractional abundances of 0.333 (± 0.002) for Ser432 M+3 and 0.572 (± 0.002) for Gly288 M+0 in GA0.32, steady state fractional abundances of 0.314 (± 0.003) for Ser432 M+3 and 0.568 (± 0.001) for Gly288 M+0 in GA0.225, steady state fractional abundances of 0.307 (± 0.002) for Ser432 M+0 and 0.552 (± 0.003) for Gly288 M+0 in GA0.15 and steady state fractional abundances of 0.455 (± 0.003) for Ser432 M+0 and 0.641 (± 0.002) for Gly288 M+0 in in GA0.075.

Comparing GA0.075, GMA0.075 and GAn0.075

For the GMA0.075 ILE, in terms of isotopic steady state for the MIDs of the Ser432 and Gly288 mass fragment populations, it appears that an isotopic steady state may have been reached by the final 4 CVCC time point, however it is inconclusive. Alternatively, for the GAn0.075 ILE, an isotopic steady state is observable from the 3.56 CCVC time point (Figure. 49). The respective final totals of ^{13}C labelling for all three amino acids observed for GAM0.075 were: 68.1% for Ser432 and 43.4% for Gly288 (Figure. 49, graph column 2) and for GAn0.075 were: 68.7% for Ser432 and 43.5% for Gly288 (Figure. 49, graph column 3).

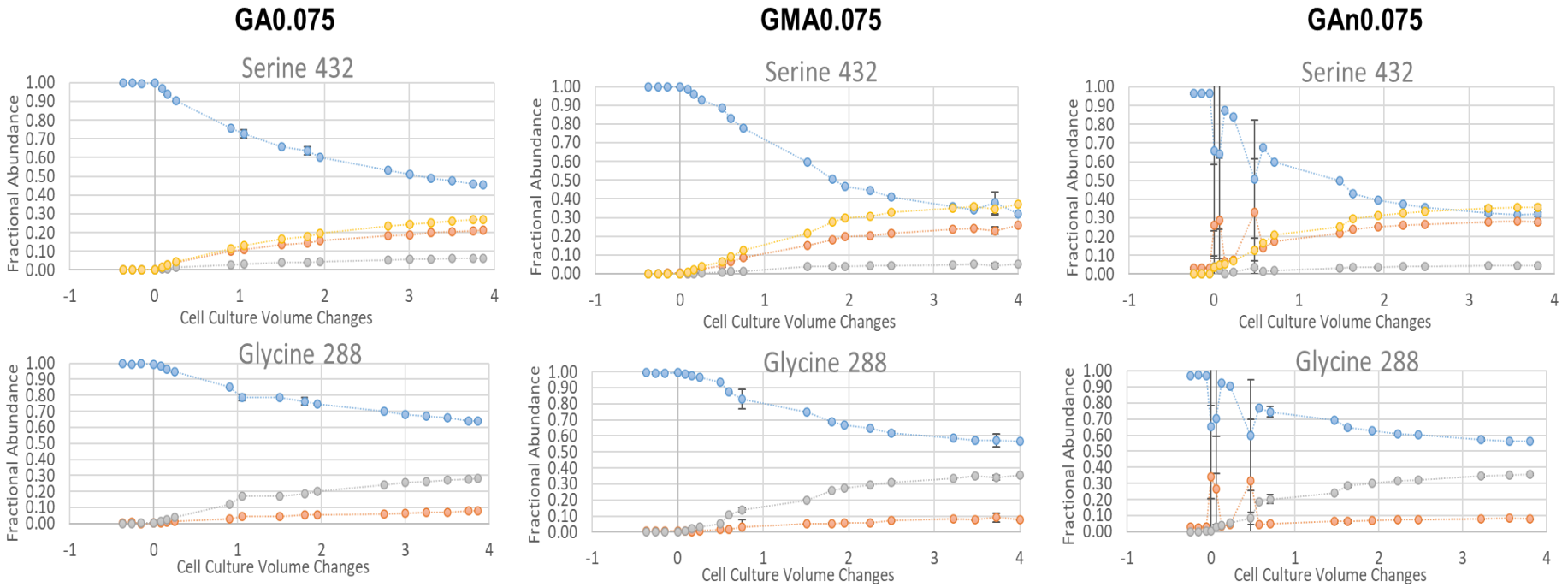


Figure 49: 3PG derived TBDMS-derivitized amino acid MIDs for mass fragments Ser432 anf Gly288 for the GA0.075, GMA0075, GAn0.075 of the INST-13C-MFA ILEs performed in this research. M+0 (Light blue), M+1 (Orange), M+2 (Light Grey), M+3 (Yellow), Reported values are means of technical triplicate samples with error bars representing SD.

For the Ser432 and Gly288 mass fragments of the GMA0.075 and GAn0.075 ILEs, the mass isotopomer with the highest abundance for 3 carbon serine was the M+3 mass isotopomer but for 2 carbon glycine was the M+0 unlabelled mass isotopomer. This resulted in steady state fractional abundances of 0.373 (± 0.004) for Ser432 M+3 and 0.566 (± 0.003) for Gly288 M+0 in GAM0.075 and steady state fractional abundances of 0.357 (± 0.010) for Ser432 M+3 and 0.565 (± 0.005) for Gly288 M+0 observed in GAn0.075

Comparing XA0.075, XA0.15 and XAn0.075

For the XA0.075 an XA0.15 ILEs, an isotopic steady state is observable for the MIDs of the Ser432 and Gly288 mass fragment populations again by the 3rd CCVC (Figure. 50), but this is complicated for the XAn0.075 ILE by the high variation in measured MID values for Ser432 and the final time point of Gly288.

For the three ^{13}C Xylose ILEs, the respective final totals of ^{13}C labelling for all three amino acids were again lower than the glucose experiments, reaching final percentages observed for XA0.075 of: 33.9% for Ser432 and 13.8% for Gly288 (Figure. 50, graph column 1), for XA0.15 of: 32.7% for Ser432 and 16.0% for Gly288 (Figure. 50, graph column 2) and observed for XAn0.075 of: 31.6% for Ser432, and 12.8% for Gly288 (Figure. 50, graph column 3).

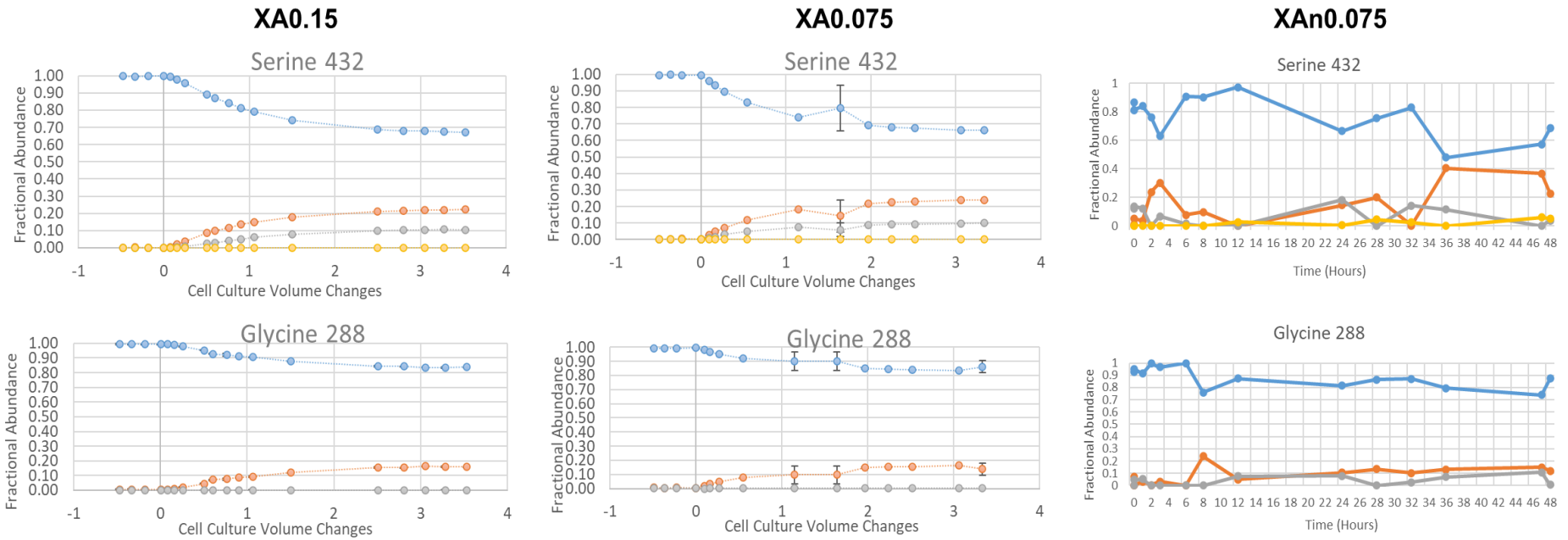


Figure 50: 3PG derivitized amino acid MIDs for mass fragments Ser432 anf Gly288 for the XA0.15, XA0.075, XAn0.075 of the INST-13C-MFA ILEs performed in this research. M+0 (Light blue), M+1 (Orange), M+2 (Light Grey), M+3 (Yellow), Reported values for XA0.15 and XA0.075 are means of technical triplicate samples with error bars representing SD

Across the Ser432 and Gly288 mass fragments for all the xylose ILEs, the mass isotopomer with the highest abundance for both 3 carbon serine and 2 carbon glycine the M+0 mass isotopomer. For each experiment, this resulted in final fractional abundances of 0.661 (± 0.002) for Ser432 M+0 and 0.862 (± 0.042) for Gly288 M+0 in XA0.075, steady state fractional abundances of 0.673 (± 0.001) for Ser432 M+0 and 0.840 (± 0.001) for Gly288 M+0 in in XA0.15 and steady state fractional abundances of 0.684 for Ser432 M+0 and 0.872 for Gly288 M+0 in in XAn0.075.

4.4.3.3 Nodes: Pyr/AcCoA – Ala302, Val330 and Leu344

The next CCM nodes of interest are those of pyruvate (PYR) and acetyl-CoA (AcCoA), represented separately below by TBDMS-derivitized amino acid mass fragments for alanine (Ala302) and valine (Val330), and collectively by the leucine mass fragment Leu344. Biosynthesis of five-carbon valine begins with two molecules of PYR which eliminates a single carbon as CO_2 from the second PYR to bind to the acetolactate synthase enzyme https://chemistry-europe.onlinelibrary.wiley.com/doi/10.1002/slct.201702128
Biosynthesis of six-carbon Leucine represents a combination of the PYR and AcCoA nodes, effectively replacing the 5th carbon from the valine biosynthesis intermediate 3-methyl-2-oxobutanoate with AcCoA.

Comparing GA0.075, GA0.15, GA0.225 and GA0.32
For the GA0.15, GA0.225 and GA0.32 ILEs, an isotopic steady state is again observable for the MIDs of the Ala302, Leu344 and Val330 mass fragment populations by the 3rd CCVC. For GA0.075, isotopic steady state could have been achieved for Ala302 and Val330 by the end of the experiment but was unlikely to have been achieved by Leu344 (Figure.51).

The respective final totals of ^{13}C-labelling for all three amino acids for GA0.32 were: 73.1% for Ala302, 93.1% for Leu344 and 87.5% for Val330 (Figure. 51 graph column 1), for GA0.225 were: 73.8% for Ala302, 95.3% for Leu344 and 88.6% for Val330 (Figure. 51, graph column 2) and GA0.15 were: 74.2% for Ala302, 95.1% for Leu344 and 88.2% for Val330 (Figure. 51, graph column 3). GA0.075 demonstrated lower final enrichment values of 54.1% for Ala302, 71.8% for Leu344 and 67.0% for Val330 (Figure 51, graph column 4).

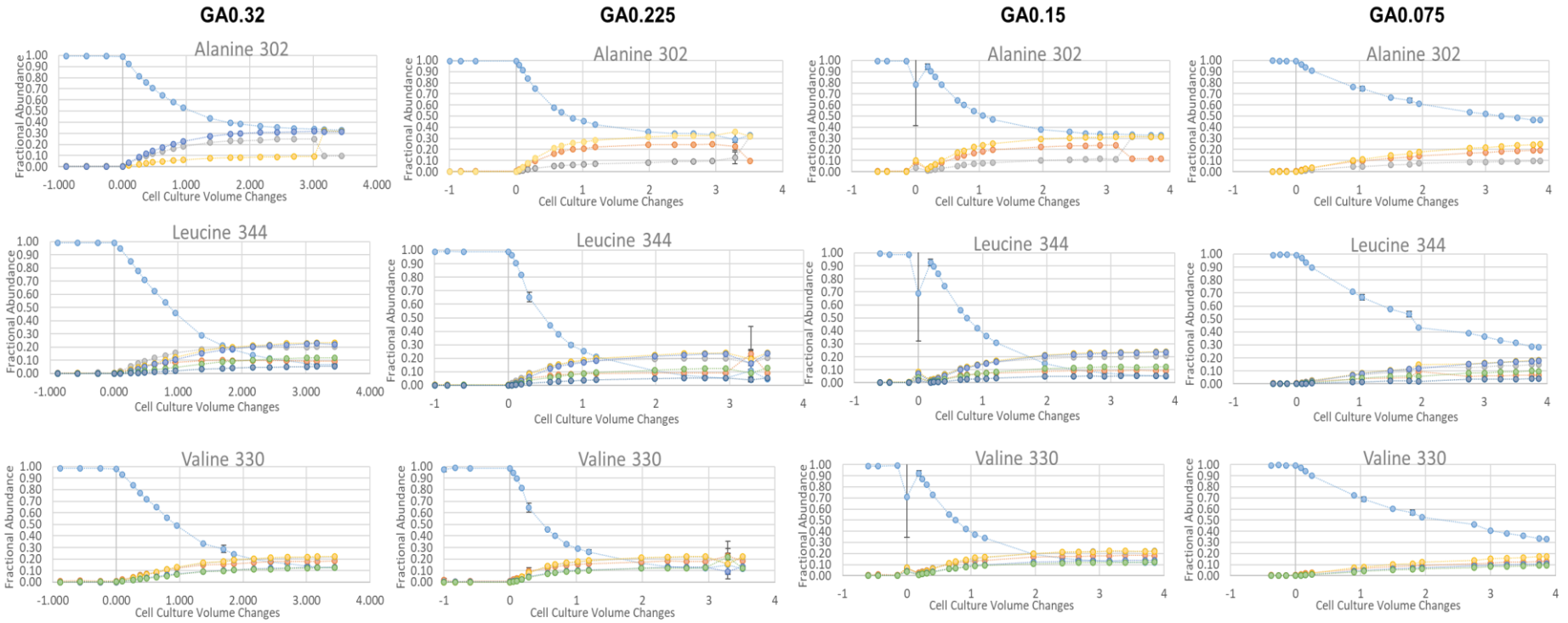


Figure 51: Pyruvate and acetate derived TBDMS-derivitized amino acid MIDs for mass fragments Ala302, Leu344 and Val330 for the GA0.32, GA0,225, GA0.15 and GA0.075 of the INST-13C-MFA ILEs performed in this research. M+0 (Light blue), M+1 (Orange), M+2 (Light Grey), M+3 (Yellow), M+4 (Blue), M+5 (Green), M+6 (Dark Blue). Reported values are means of technical triplicate samples with error bars representing SD.

Across the Ala302, Leu344 and Val330 fragments, the mass isotopomer with the highest fractional abundance for 3 carbon alanine was a close competition between the M+0 and M+3 mass isotopomers for GA0.32 (0.331 (± 0.002) M+0 vs. 0.320 (± 0.003) M+3), GA0.225 (0.331 (± 0.001) M+0 vs. 0.320 (± 0.001) M+3) and GA0.15 (0.330 (± 0.002) M+0 vs. 0.313 (± 0.002) M+3), but was clearly the M+0 mass isotopomer for GA0.075 0.541 (± 0.004). For 6 carbon leucine the predominant mass isotopomer for GMA0.075 was the M+3 isotopomer, with a steady state fractional abundance of 0.232 (± 0.004) for Leu344 M+3 for GA0.32, 0.242 (± 0.004) for Leu344 M+3 for GA0.225 and 0.239 (± 0.002) Leu344 M+3 for GA0.15. For GA0.075 the predominant fragment was the M+0 isotopomer with a final fractional abundance of 0.282 (± 0.002) Leu344 M+0. For 5 carbon valine however the M+3 mass isotopomer predominated for GA0.32 0.221 (± 0.004) Val330 M+3, GA0.225 0.224 (± 0.002) Val330 M+3 and GA0.15 0.224 (± 0.003) Val330 M+2. Once again, the M0 mass isotopomer predominated for GA0.075 0.330 (± 0.002).

Comparing GA0.075, GMA0.075 and GAn0.075

For the GMA0.075 and GAn0.075 ILEs, an isotopic steady state is observable from the 3.225 CCVC time point for the MIDs of the Ala302, Val330 and Leu344 mass fragment populations (Figure. 52).

The respective final totals of ^{13}C labelling for all three amino acids observed for GAM0.075 were: 73.1% for Ala302, 90.5% for Leu344 and 85.3% for Val330 (Figure.52, graph column 2) and for GAn0.075 were: 64.6% for Ala302, 90.7% for Leu344 and 90.7% for Val330 (Figure.52, graph column 3).

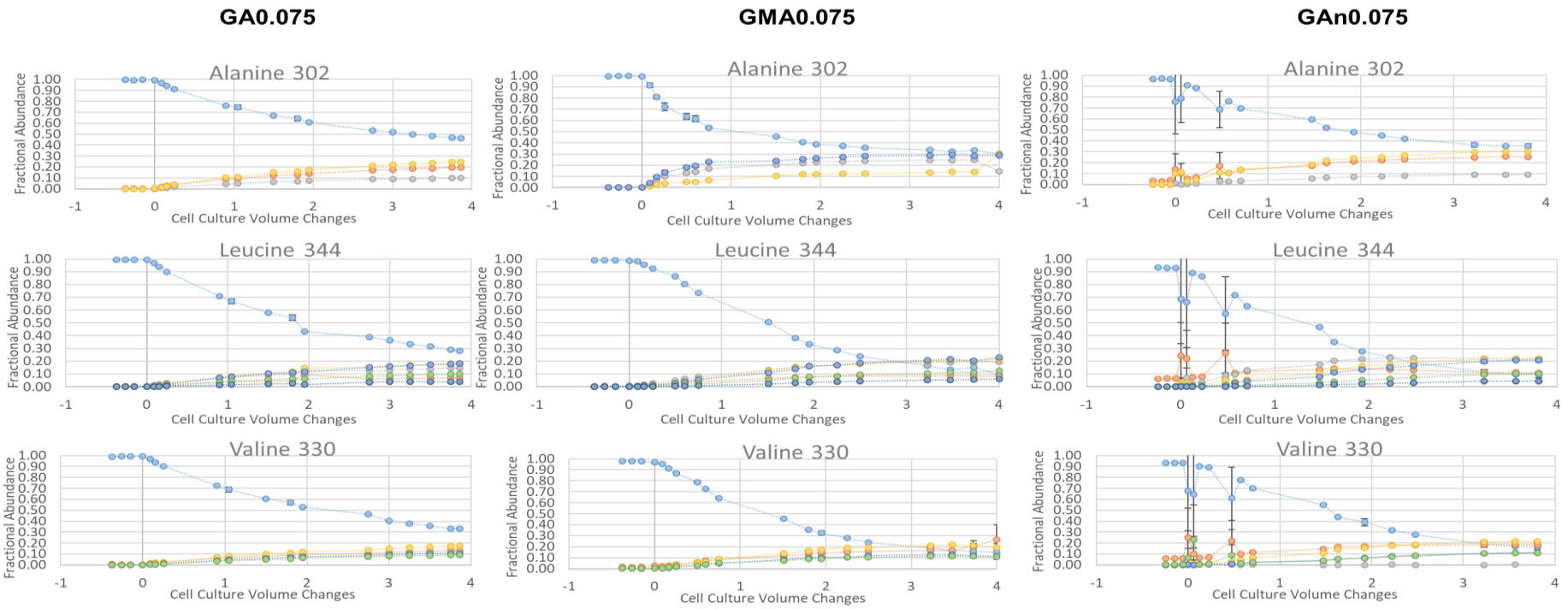


Figure 52: Pyruvate and acetate derived TBDMS-derivitized amino acid MIDs for mass fragments Ala302, Leu344 and Val330 for the GMA0.075, GM0.075 and GAn0.075 of the INST-^{13}C-MFA ILEs performed in this research. M+0 (Light blue), M+1 (Orange), M+2 (Light Grey), M+3 (Yellow), M+4 (Blue), M+5 (Green), M+6 (Dark Blue). Reported values are means of technical triplicate samples with error bars representing SD.

For the Ala302, Leu344 and Val330 mass fragments of the GMA0.075 and GAn0.075 ILEs, the mass isotopomer with the highest abundance for 3 carbon alanine was the M+0 mass isotopomer for GMA0.075 and GAn0.075 with respective fractional abundances of 0.304 (± 0.002) Ala302 M+0 and 0.354 (± 0.017) Ala302 M+0. For 6 carbon leucine the predominant mass isotopomer for GMA0.075 was the M+4 isotopomer, with a steady state fractional abundance of 0.228 (± 0.003) for Leu344 M+4, but for GAn0.075 was a close contest between the M+2 and M+3 isotopomers with steady state fractional abundance of 0.218 (± 0.002) Leu344 M+2 and 0.222 (± 0.008) Leu344 M+3. Lastly, for 5 carbon valine, for both the GMA0.075 and GAn0.075 ILEs, the mass isotopomers with the highest abundances varied due to close final values and measurement error. This resulted in steady state fractional abundances of 0.266 (± 0.136) for Val330 M+1 and 0.196 (± 0.048) for Val330 M+3 in GAM0.075 and steady state fractional abundances of 0.212 (± 0.005) for Val330 M+2 and 0.215 (± 0.006) for Val330 M+3 in GAn0.075.

Comparing XA0.075, XA0.15 and XAn0.075

For the XA0.075 and XA0.15 ILEs, an isotopic steady state is observable for the MIDs of the Ala302, Val330 and Leu344 mass fragment populations again by 2.5 CCVCs for XA0.075 and XA0.15 (Figure. 53). Once again, this is unclear for the XAn0.075 due to the high variation in measured MIDs.

For the three ^{13}C xylose ILEs, the respective final totals of ^{13}C labelling for all three amino acids were again lower than the glucose experiments, reaching final percentages observed for XA0.075 of: 32.8% for Ala302, 50.9% for Val330 and 62.2% for Leu344 (Figure 53, graph column 1), for XA0.15 of: 32.1% for Ala302, 50.5% for Val330 and 62.8% for Leu344 (Figure 53 graph column 2) and observed for XAn0.075 of: 39.4% for Ala302, 41.4% for Val330 and 50.1% for Leu344 (Figure 53, graph column 3).

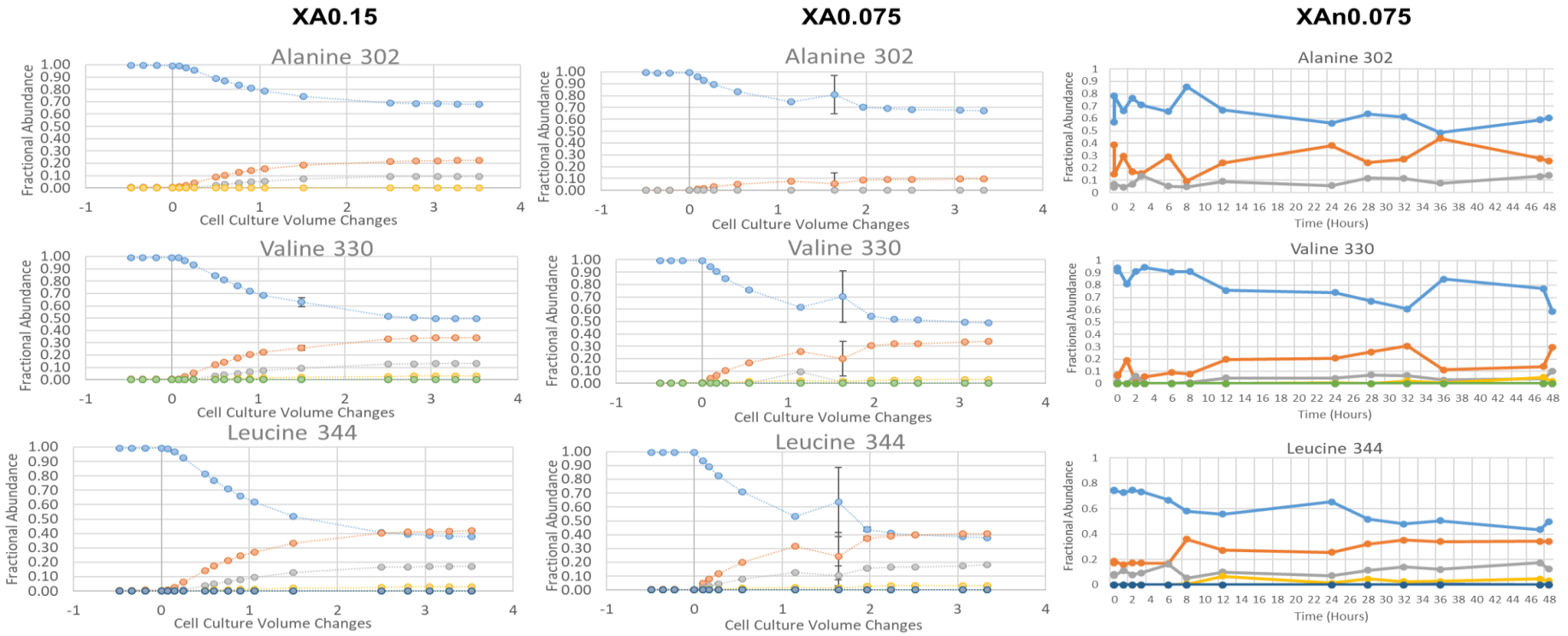


Figure 53: Pyruvate and acetate derived TBDMS-derivitized amino acid MIDs for mass fragments Ala302, Leu344 and Val330 for the XA015, XA0.075 and XAn0.075 of the INST-^{13}C-MFA ILEs performed in this research. M+0 (Light blue), M+1 (Orange), M+2 (Light Grey), M+3 (Yellow), M+4 (Blue), M+5 (Green), M+6 (Dark Blue). Reported values for XA0.15 and XA0.075 are means of technical triplicate samples with error bars representing SD.

Across the Ala302, Val330 and Leu344 mass fragments of the XA0.075, XA0.15 and XAn0.075 ILEs, the mass isotopomer with the highest abundance for 3 carbon alanine was the M+0 mass isotopomer. This final fractional abundances for each ILE were 0.672 (± 0.002) for Ala302 M+0 in XA0.075, 0.679 (± 0.002) for Ala302 M+0 in XA0.15 and 0.606 for Ala302 M+0 in XAn0.075. For each xylose ILE, the M+0 fragment predominated for 5 carbon valine, with final fractional abundances for XA0.075 of 0.491 (± 0.002) for Val330 M+0, for XA0.15 of 0.495 (± 0.001) for Val330 M+0 and for XAn0.075 of 0.586 for Val330 M+0. Lastly, for 6 carbon leucine, the predominant mass isotopomer was the M+1 isotopomer for both the XA0.075 and XA0.15 ILEs, with steady state fractional abundances of 0.408 (± 0.002) for Leu344 M+1 in XA0.075 and steady state fractional abundances of 0.421 (± 0.003) for Leu344 M+1 in XA0.15. For the XAn0.075 ILE the M+0 mass fragment predominated with a final abundance of 0.499 for Leu344 M+0.

4.4.3.4 Node: AKG – Glu474, Pro258

The results of the final α-ketoglutarate (AKG) and oxaloacetate (OAA) nodes represent the citric acid cycle.
Following the conventional clockwise directionality of the citric acid cycle, the penultimate node of interest is AKG, represented here by TBDMS-derivitized mass fragments for 5 carbon glutamate (Glu474) and 5 carbon proline. In this instance proline is represented by 4 carbon mass fragment Pro258.

Comparing GA0.075, GA0.15, GA0.225 and GA0.32

For the GA0.15, GA0.225 and GA0.32 ILEs, an isotopic steady state is observable for the MIDs of the Glu474, and Pro258 mass fragment populations by 2.6 CCVCs. It not obviously achieved for GA0.075 by the end of the ILE (Figure. 54).

The respective final totals of ^{13}C labelling for all three amino acids for GA0.32 were: 87.2% for Glu474 and 90.7% for Pro258 (Figure 54, graph column 1), for GA0.225 were: 93.0% for Glu474 and 92.2% for Pro258 (Figure 54, graph column 2) and GA0.15 were: 89.0% for Glu474 and 91.5% for Pro258 (Figure 54, graph column 3). GA0.075 demonstrated lower final enrichment values of: 76.5% for Glu474 and 68.3% for Pro258 (Figure 54 graph column 4).

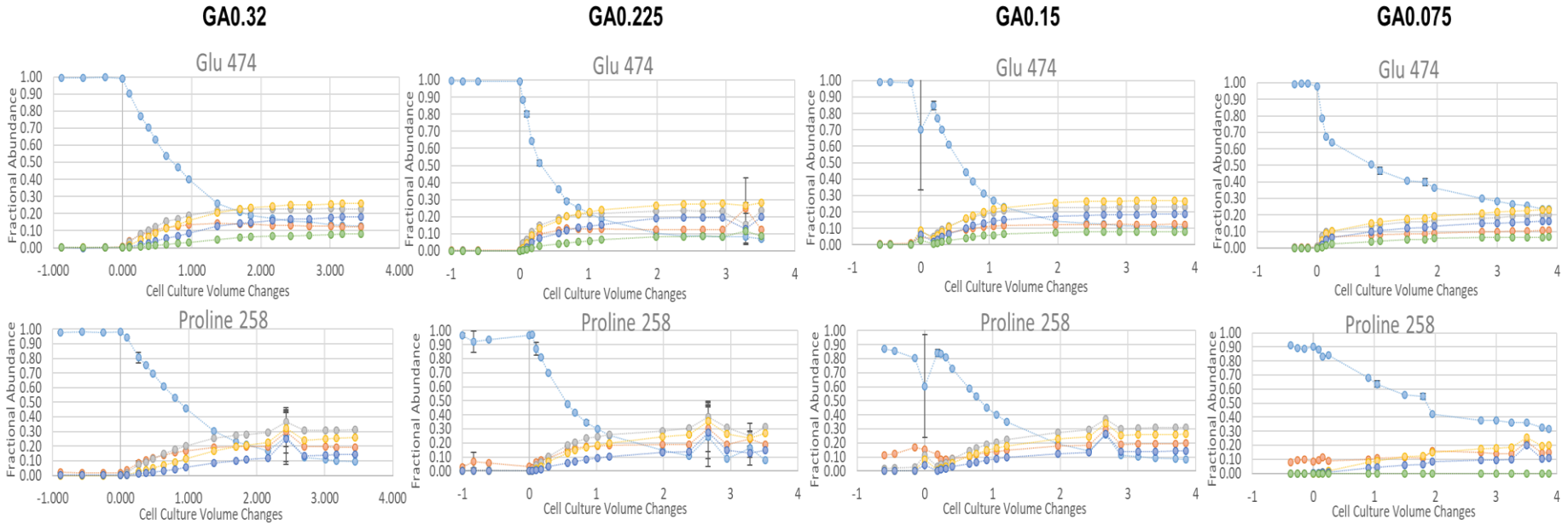
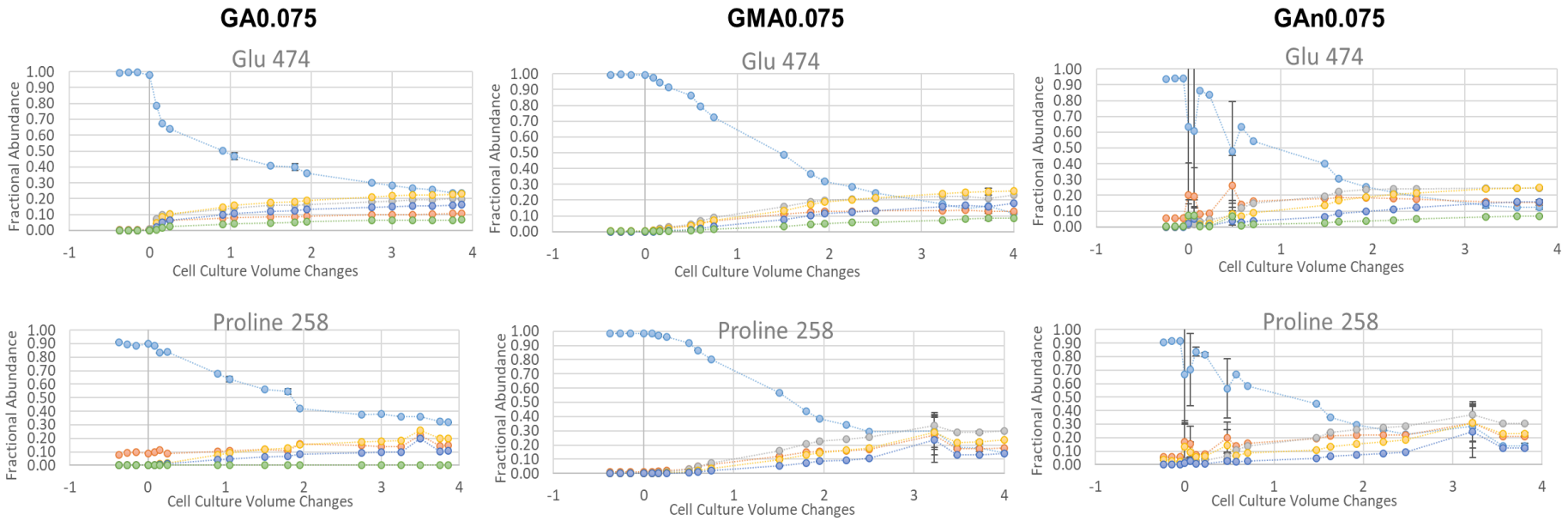


Figure 54: AKG derived TBDMS-derivitized amino acid MIDs for mass fragments Glu474 and Pro258 for the GA0.32, GA0,225, GA0.15 and GA0.075 of the INST-^{13}C-MFA ILEs performed in this research. M+0 (Light blue), M+1 (Orange), M+2 (Light Grey), M+3 (Yellow), M+4 (Blue), M+5 (Green). Reported values are means of technical triplicate samples with error bars representing SD.

Across the Glu474 and Pro258 fragments, the mass isotopomer with the highest abundance for 5 carbon glutamate was the M+3 mass isotopomer and for 5 carbon proline the M+2 mass isotopomer. For each experiment, this resulted in final fractional abundances of 0.176 (± 0.002) for Glu474 M+3 and 0.261 (± 0.001) for Pro258 M+2 in GA0.32, steady state fractional abundances of 0.282 (± 0.002) for Glu474 M+3 and 0.313 (± 0.001) for Pro258 M+3 in GA0.225, steady state fractional abundances of 0.267 (± 0.003) for Glu474 M+3 and 0.312 (± 0.001) for Pro258 M+2 in GA0.15 and final fractional abundances of 0.235 (± 0.003) for Glu474 M+0 and 0.317 (± 0.001) for Pro258 M+0 in in GA0.075.

Comparing GA0.075, GMA0.075 and GAn0.075

For the GMA0.075 ILE, in terms of isotopic steady state for the MIDs of the Glu474 and Pro258 mass fragment populations, it appears that an isotopic steady state may have been reached by the final 4 CVCC time point, however it is inconclusive. Alternatively, for the GAn0.075 ILE, an isotopic steady state is observable from the 3.56 CCVC time point (Figure. 55). The respective final totals of ^{13}C labelling for all three amino acids observed for GAM0.075 were: 88.5% for Glu474 and 85.4% for Pro258 (Figure. 55, graph column 2) and for GAn0.075 were: 87.6% for Glu474 and 86.2% for Pro258 (Figure. 55, graph column 3).



<u>Figure 55:</u> AKG derived TBDMS-derivitized amino acid MIDs for mass fragments Glu474 and Pro258 for the GA0.075, GMA0.075 and GAn0.075 INST-^{13}C-MFA ILEs performed in this research. M+0 (Light blue), M+1 (Orange), M+2 (Light Grey), M+3 (Yellow), M+4 (Blue), M+5 (Green), M+6 (Dark Blue). Reported values for XA0.15 and XA0.075 are means of technical triplicate samples with error bars representing SD.

For the Glu474 and Pro258 mass fragments of the GMA0.075 and GAn0.075 ILEs, the mass isotopomer with the highest abundance for 5 carbon glutamate was the M+3 mass isotopomer and for 5 carbon proline the M+2 mass isotopomer. This resulted steady state fractional abundances of 0.257 (± 0.001) for Glu474 M+3 and 0.300 (± 0.001) for Pro258 M+2 in GAM0.075 and steady state fractional abundances of 0.249 (± 0.005) for Glu474 M+3 and 0.306 (± 0.005) for Pro258 M+2 observed in GAn0.075.

Comparing XA0.075, XA0.15 and XAn0.075

For the XA0.075, XA0.15 and XAn0.075 ILEs, an isotopic steady state is observable for the MIDs of the Glu474 and Pro258 mass fragment populations again by the 3rd CCVC (Figure. 56).

For the three ^{13}C xylose ILEs, the respective final totals of ^{13}C labelling for all three amino acids were again lower than the glucose experiments, reaching final percentages observed for XA0.075 of: 59.0% for Glu474 and 56.4% for Pro258 (Figure. 56, graph column 1), for XA0.15 of: 54.9% for Glu474 and 54.4% for Pro258 (Figure. 56, graph column 2) and observed for XAn0.075 of: 45.7% for Glu474 and 41.5% for Pro258 (Figure. 56, graph column 3).

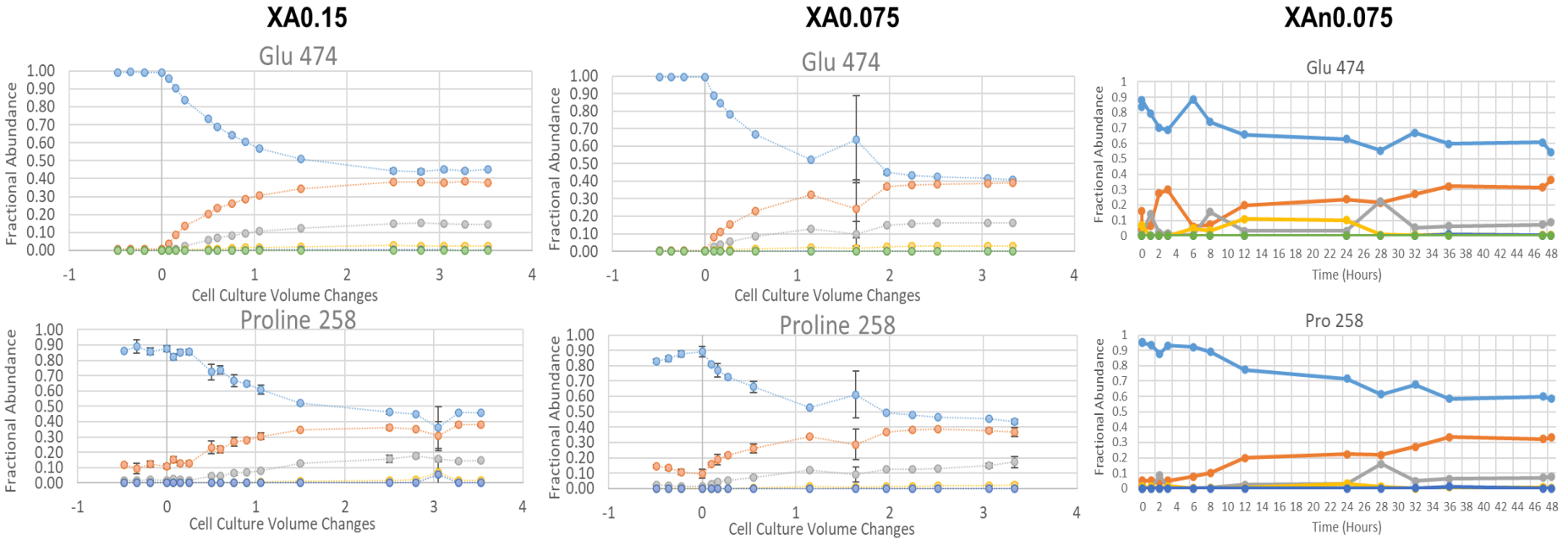


Figure 56: AKG derived TBDMS-derivitized amino acid MIDs for mass fragments Glu474 and Pro258 for the XA0.15, XA0.075 and XAn0.075 of the INST-^{13}C-MFA ILEs performed in this research. M+0 (Light blue), M+1 (Orange), M+2 (Light Grey), M+3 (Yellow), M+4 (Blue), M+5 (Green), M+6 (dark blue). Reported values for XA0.15 and XA0.075 are means of technical triplicate samples with error bars representing SD.

Once again, across the Glu474 and Pro258 mass fragments, the mass isotopomer with the highest abundance for both 5 carbon glutamate and 5 carbon proline was the M+0 mass isotopomer. For each experiment, this resulted in final fractional abundances of 0.410 (± 0.004) for Glu474 M+0 and 0.436 (± 0.016) for Pro258 M+0 in XA0.075, steady state fractional abundances of 0.451 (± 0.002) for Glu474 M+0 and 0.456 (± 0.002) for Pro258 M+0 in in XA0.15 and steady state fractional abundances of 0.543 for Glu474 M+0 and 0.585 for Pro258 M+0 in in XAn0.075.

4.4.3.5 Node: OAA – Asp460, Lys473, Ile344

The final node of interest was OAA, represented by the direct transamination product of aspartate (Asp460) and the derivative products lysine (Lys473) and isoleucine (Ile344). While routes to the production of lysine and isoleucine exist, which receive contributions from alternative nodes, the genome of *P. thermoglucosidasius* suggest that these should be absent. In particular, the *P. thermoglucosidaisius* genome appears to be missing a gene for a citramalate synthase enzyme which would enable direct contributions from PYR to pools of leucine and an LL-diaminopimelate aminotransferase which could create two routes in the biosynthetic pathway of isoleucine which would affect labelling. Absences for both genes are also reflected in *p-thermo*.

Comparing GA0.075, GA0.15, GA0.225 and GA0.32
For the GA0.15, GA0.225 and GA0.32 ILEs, an isotopic steady state is again observable for the MIDs of the Asp460, Lys473 and Ile344 mass fragment populations by 2.5 CCVCs. For GA0.075, isotopic steady state could have been achieved for Asp460, Lys473 and Ile344 by the end of the experiment (Figure 57), but this is inconclusive.

The respective final totals of ^{13}C labelling for all three amino acids for GA0.32 were: 87.5% for Asp460, 93.6% for Lys473 and 93.8% for Ile344 (Figure. 57, graph column 1), for GA0.225 were: 90.0% for Asp460, 94.8% for Lys473 and 95.3% for Ile344 (Figure. 57, graph column 2) and GA0.15 were: 89.8% for Asp460, 94.4% for Lys473 and 94.7% for Ile344 (Figure 57, graph column 3). GA0.075 demonstrated lower final enrichment values of: 69.9% for Asp460, 69.9% for Lys473 and 70.7% for Ile344 (Figure 57, graph column 4).

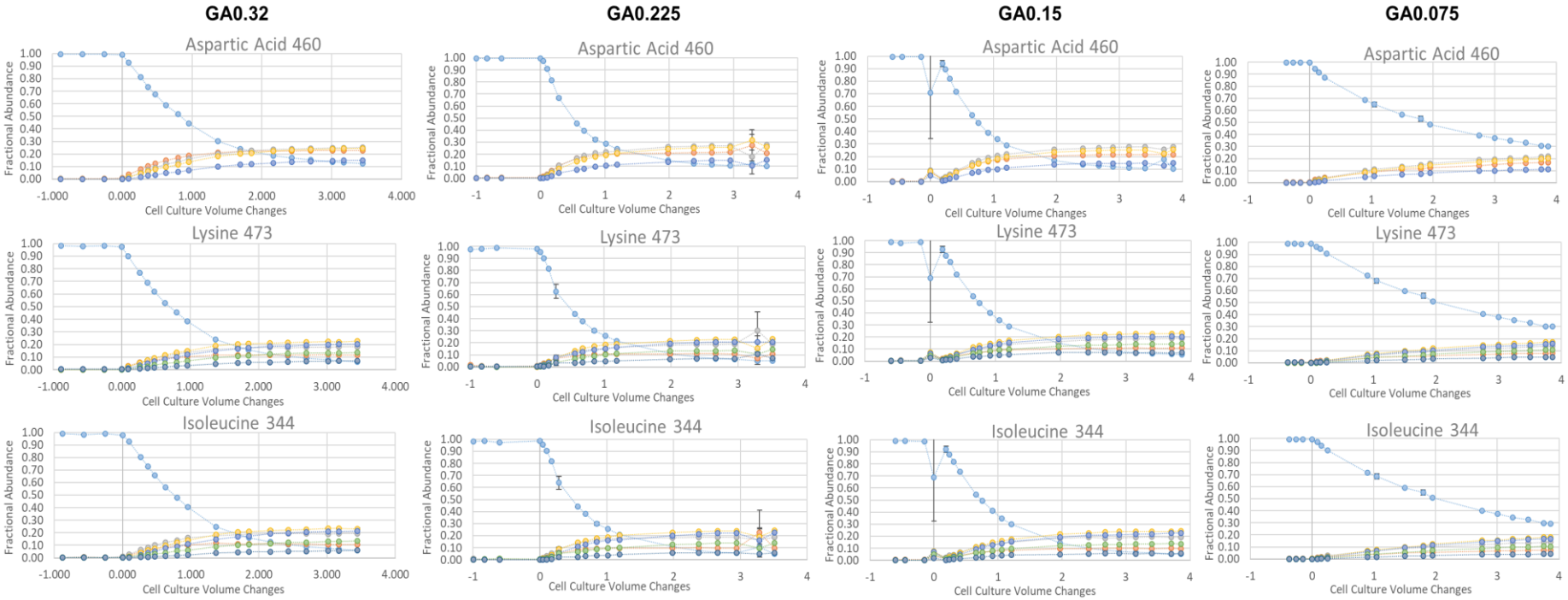


Figure 57: OAA derived TBDMS-derivitized amino acid MIDs for mass fragments Asp460, Lys473 and Ile344 for the GA0.32, GA0,225, GA0.15 and GA0.075 INST-^{13}C-MFA ILEs performed in this research. M+0 (Light blue), M+1 (Orange), M+2 (Light Grey), M+3 (Yellow), M+4 (Blue), M+5 (Green), M+6 (Dark Blue). Reported values are means of technical triplicate samples with error bars representing SD.

Across the Asp460, Lys473 and Ile344 fragments of the GA0.15, GA0.225 and GA0.32 ILEs, the mass isotopomer with the highest fractional abundance for 4 carbon aspartic acid the M+2 mass isotopomer predominated for GA0.32 (0.253 (± 0.001) Asp460 M+2, GA0.225 (0.277 (± 0.00) Asp460 M+2 and GA0.15 0.278 (± 0.001) Asp460 M+2. For 6 carbon lysine however the M+3 mass isotopomer predominated for GA0.32 0.228 (± 0.003) Lys473 M+3, GA0.225 0.231 (± 0.003) and GA0.15 0.230 (± 0.002). For 6 carbon isoleucine the M+3 mass isotopomer predominated for GA0.32 0.232 (± 0.005) Ile344 M+3, GA0.225 0.245 (± 0.004) Ile344 M+3 and GA0.15 0.241 (± 0.003). Once again, for the GA0.075 ILE the M0 mass isotopomer with final abundances of 0.301 (± 0.003) Asp460 M+0, 0.301 (± 0.003) Lys473 M+0, 0.293 (± 0.003) Ile344 M+0.

Comparing GA0.075, GMA0.075 and GAn0.075

For the GMA0.075 and GAn0.075 ILEs, an isotopic steady state is observable from the 3.225 CCVC time point for the MIDs of the Asp460, Lys473 and Ile344 mass fragment populations (Figure. 58). The respective final totals of ^{13}C labelling for all three amino acids observed for GAM0.075 were: 82.5% for Asp460, 76.8% for Lys473 and 89.9% for Ile344 (Figure. 58, graph column 2) and for GAn0.075 were: 82.6% for Asp460, 89.2% for Lys473 and 90.5% for Ile344 (Figure 58, graph column 3).

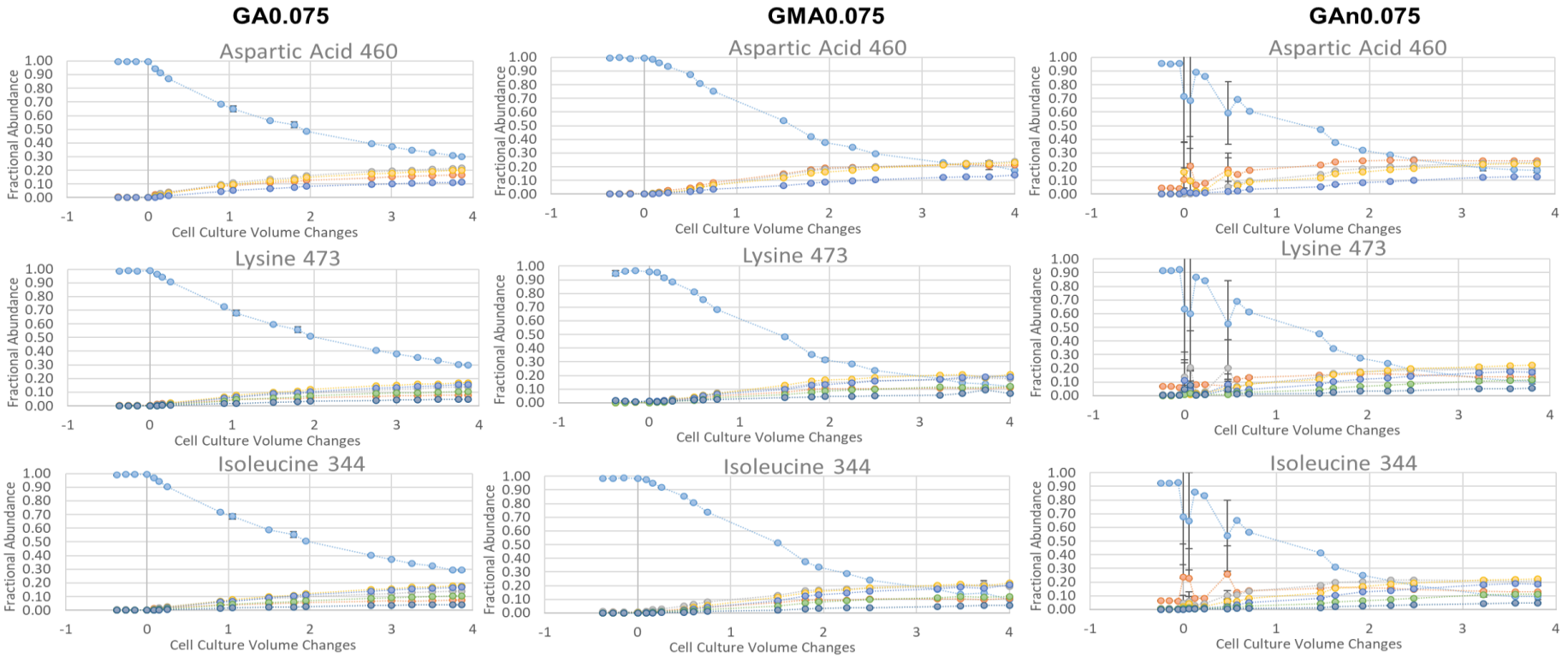


Figure 58: OAA derived TBDMS-derivitized amino acid MIDs for mass fragments Asp460, Lys473 and Ile344 for the GA0.075, GMA075, and GAn0.075 INST-^{13}C-MFA ILEs performed in this research. M+0 (Light blue), M+1 (Orange), M+2 (Light Grey), M+3 (Yellow), M+4 (Blue), M+5 (Green), M+6 (Dark Blue). Reported values are means of technical triplicate samples with error bars representing SD.

For the Asp460, Lys473 and Ile344 mass fragments of the GMA0.075 and GAn0.075 ILEs, the mass isotopomer with the highest abundance for 4 carbon aspartate was the M+2 mass isotopomer for GMA0.075 with a fractional abundance of 0.245 (± 0.01) Asp460 M+2 and was the M+1 mass isotopomer for GAn0.075 with a fractional abundance of 0.245 (± 0.002). For 6 carbon leucine the predominant mass isotopomer for both GMA0.075 and GAn0.075 was the M+3 isotopomer, with respective steady state fractional abundances of 0.208 (± 0.005) for Lys473 M+3 and 0.223 (± 0.001) Lys473 M+3. Lastly, for 6 carbon isoleucine, for both the GMA0.075 and GAn0.075 ILEs, the mass isotopomer with the highest abundance for isoleucine was the M+3 mass isotopomer. This resulted in observed steady state fractional abundances of 0.220 (± 0.001) for Ile344 M+3 in GAM0.075 and steady state fractional abundances of 0.224 (± 0.005) for Ile344 M+3 in GAn0.075

Comparing XA0.075, XA0.15 and XAn0.075

For the XA0.075 and XA0.15 ILEs, an isotopic steady state is observable for the MIDs of the Asp460, Lys473 and Ile344 mass fragment populations again by 2.5 CCVCs for XA0.075 and XA0.15 (Figure 59). However, for the XAn0.075 ILE but this is unclear. It may have been achieved for the Asp460 and Ile344 mass fragments by the end of the experiment but was not likely achieved by lysine 473.

For the three ^{13}C xylose ILEs, the respective final totals of ^{13}C labelling for all three amino acids were again lower than the glucose experiments, reaching final percentages observed for XA0.075 of: 52.2% for Asp460, and 63.6% for Lys473 and 64.4% for Ile344 (Figure. 59, graph column 1), for XA0.15 of: 50.4% for Asp460, and 62.9% for Lys473 and 63.3% for Ile344 (Figure. 59, graph column 2) and observed for XAn0.075 of: 51.0% for Asp460, and 63.8% for Lys473 and 50.1% for Ile344 (Figure 59, graph column 3).

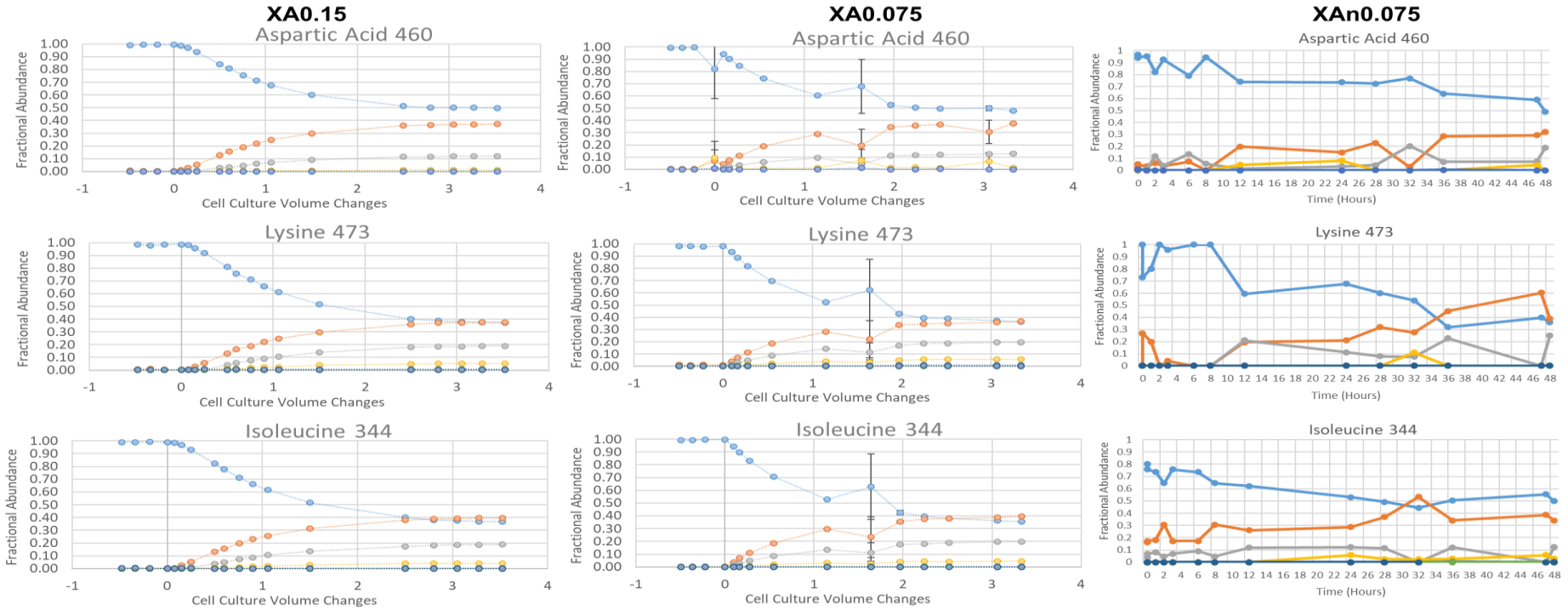


Figure 59: OAA derived TBDMS-derivitized amino acid MIDs for mass fragments Asp460, Lys473 and for the XA0.15, XA0.075 and XAn0.075 INST-^{13}C-MFA ILEs performed in this research. M+0 (Light blue), M+1 (Orange), M+2 (Light Grey), M+3 (Yellow), M+4 (Blue), M+5 (Green), M+6 (dark blue). Reported values for XA0.15 and XA0.075 are means of technical triplicate samples with error bars representing SD.

Across the Asp460, Lys473 and Ile344 and mass fragments of the XA0.075 and XA0.15, the mass isotopomer with the highest abundance for 4 carbon aspartate was the M+0 mass isotopomer for XA0.075 with a fractional abundance of 0.478 (± 0.001) M+0 Asp460, for with a fractional abundance of 0.496 (± 0.001) M+0 Asp460 XA0.15 and a fractional abundance of 0.490 M+0 Asp460 XAn0.075. For 6 carbon lysine the predominant mass isotopomer for XA0.075 was the M+1 isotopomer. For each experiment, this resulted in steady state fractional abundances of 0.368 (± 0.002) for Lys473 M+1 in XA0.075, 0.378 (± 0.002) for Lys473 M+1 in XA0.15 and 0.389 for Lys473 M+1 in XAn0.075. Lastly, for 6 carbon isoleucine, for both the XA0.075 and XA0.15 ILEs, the mass isotopomer with predominant abundance for isoleucine was the M+1 mass isotopomer. This resulted in observed steady state fractional abundance of 0.396 (± 0.003) for Ile344 M+1 in XA0.075 and steady state fractional abundance of 0.396 (± 0.004) for Ile344 M+1 in XA0.15, but a final fractional abundance of 0.499 for Ile344 M+0 in XAn0.075.

4.4.3.6 Discussion: Qualitative observations on glycolysis and the PPP, Serine vs Tyrosine and Alanine

In terms of general observations of the MDV data, high SD values were observed for several time points in different experiments for all amino acids. In terms of corresponding CCVCs, these time points consisted of point 0 ^{13}C CCVCs for the GA0.15 ILE, point 3.29 ^{13}C CCVCs for the GA0.225 ILE, points 0, 0.0625 and 0.475 CCVCs for the GAn0.075 ILE and point 1.64 CCVCs for the XA0.075 ILE. No error could be calculated for XAn0.075 as each data point represented a single sample per time point rather than technical triplicate. Additionally, observation of the ^{13}C incorporation trends for GA0.075 suggest that this experiment did not reach an isotopic steady state by the end of the ILE.

A peculiar observation can also be seen for all Pro258 trends across all experiments but XA0.075 and XAn0.075 in that single data point in each trend generates high error. This occurs at CCVCs of 2.43 (GA0.32), 2.66 (GA0.225), 2.67 (GA0.15), 3.5 (GA0.075), 3.225 (GMA0.075 and GAn0.075) and 3.05 (XA0.25) which in each case corresponds to about the point that the ^{13}C-labelling trends converge at an isotopic steady state of proline. However, given this is observation is not repeated for any other amino acid fragment and not observed for proline's precursor glutamate Glu474, this was most likely some form of recurring processing error than an accurate observation of metabolic behaviour at near isotopic steady state.

Under the assumption that these high error time-points were not accurate observations of *P. thermoglucosidasius* metabolic behaviour, the MDVs at all of these points were excluded from final quantitative analysis.

Rather than presenting the INST-^{13}C-MFA MDV data sets in isolation, as a result of comparable values for total ^{13}C incorporation, it was decided that they would be presented in terms of their assumed carbon intermediate node in CCM such that qualitative observations can be made about how pathways from these nodes contribute to the respect metabolic phenotypes of *P. thermoglucosidasius*. As a result, the similarities and differences in total ^{13}C incorporation and predominant mass isotopomer of each amino acid fragment can be compared to qualitatively indicate if specific biosynthetic paths were operating differently under different conditions.

The ^{13}C-labelling pattern of 3 carbon serine directly represents the labelling of the 3PG precursor. Across all the glucose ILEs demonstrating that, at isotopic steady state on 60% 1-^{13}C and 40% U-^{13}C ^{D}-glucose as a sole substrate, an average of 33.0% (± 5.7%) of 3PG was unlabelled (M+0), 33.2% (± 3.3%) of 3PG was completely ^{13}C-labelled (M+3), and only 6.9% (± 1.8%) of 3PG had 2 of 3 carbons labelled (Table 33).

Table 33: MID comparisons for Serine (Ser432) for all experiments. Average and SD values were calculated based on all Glucose ILEs and all Xylose INST-^{13}C-MFA ILEs.

INST-^{13}C-MFA ILE	Ser432			
	M+0	M+1	M+2	M+3
GA0.32	0.302	0.288	0.077	0.333
GA0.225	0.280	0.305	0.101	0.314

GA0.15	0.307	0.269	0.077	0.347
GA0.075	0.455	0.212	0.062	0.270
GMA0.075	0.319	0.257	0.052	0.373
GAn0.075	0.320	0.277	0.046	0.357
Average	0.330	0.268	0.069	0.332
SD	0.057	0.029	0.018	0.033
XA0.075	0.661	0.237	0.101	0.000
XA0.15	0.673	0.222	0.104	0.001
XAn0.075	0.684	0.226	0.038	0.052
Average	0.673	0.228	0.081	0.018
SD	0.009	0.006	0.030	0.024

For the xylose ILEs, grown on 75% 1-^{13}C, 25% ^{12}C xylose as a sole substrate, an average of 167.0% (± 0.9%) of 3PG was therefore unlabelled (M+0) and 22.8% (± 0.6%) of 3PG had 2 of 3 carbons labelled.

Furthermore, as the ^{13}C-labelling patterns of serine and glycine are indicative of 3PG, it can be interesting to compare the total ^{13}C incorporation and ^{13}C-labelling patterns of serine (Ser432) to those of tyrosine (Tyr466) the latter of which is derived from the subsequent glycolytic node of PEP and of E4P in the PPP. Differences between them can indicate the degree of contribution from the PPP and E4P node. Comparing the total incorporation of the glucose and xylose MIDs demonstrates this phenomenon (Table 34).

Table 34: Comparison of percentage final ^{13}C incorporation values for Serine (Ser432) and Tyrosine (Tyr466) for all INST-^{13}C-MFA ILEs.

INST-^{13}C-MFA ILE	Final ^{13}C incorporation (%)		
	Ser432	Tyr466	Difference
GA0.32	69.8	93.4	23.6
GA0.225	72	93.9	21.9
GA0.15	69.3	93.9	24.6
GA0.075	54.5	69.7	15.2
GMA0.075	68.1	90.5	22.4
GAn0.075	68.7	92.9	24.2
XA0.075	33.9	58.2	24.3
XA0.15	32.7	60	27.3
XAn0.075	31.6	58.2	26.6

The average final observed ^{13}C incorporation into Ser432 for the GA0.32, GA0.225 and GA0.15 ILEs is 70.4% (± 1.2%), whereas in Tyr466 it is an average incorporation of 93.7% (± 0.2%), constituting an average difference of 23.4%. A similar difference of 23.3% is seen between the final ^{13}C incorporation values of Ser432 in GMA0.075 (68.1) and GMA0.075 (68.7) and those in Tyr466 in GMA0.075 (68.1) and GMA0.075 (68.7). Combined this suggests that, when grown on glucose, synthesis of chorismate-derived tyrosine and phenylalanine in *P. thermoglucosidasius* receives the same contribution from glycolysis and the PPP under both aerobic and anaerobic conditions and at different aerobic growth rates.

When grown on xylose under aerobic and anaerobic conditions on 75% 1-^{13}C and 25% naturally labelled xylose, this difference was also a similar 26.1%. The average final ^{13}C incorporation into Ser432 for the XA0.075, XA0.15 and Xan0.075 ILEs is 32.7% (± 0.9%), whereas in Tyr466 it is an average incorporation of 58.8% (± 0.8%). This result in accordance with the ^{13}C-MFA flux map generated for *Geobacillus* LC300 grown on xylose as a sole substrate which also demonstrated PPP activity[127].

A similar form of qualitative analysis can also be performed between serine and alanine (Ala302), derived from pyruvate. Comparison of the M+0 unlabeled mass isotopomer ID predominant pathway in the *P. thermoglucosidasius* metabolic network, or whether the metabolite pool of pyruvate has been partially derived from the ED pathway or anaplerotic reactions Tang *et al.* (2009)[128].

Table 35: Comparison of percentage final ^{13}C incorporation values for the Serine Ser432 M+0 mass fragment with the Alanine Ala302 M+0 mass fragment for all INST-^{13}C-MFA ILEs.

INST-^{13}C-MFA ILE	Final ^{13}C incorporation	
	Ser432 M+0	Ala302 M+0
GA0.32	0.30	0.33
GA0.225	0.28	0.33
GA0.15	0.31	0.33
GA0.075	0.46	0.54
GMA0.075	0.32	0.30
GAn0.075	0.32	0.35
Average	0.33	0.37
SD	0.06	0.08
XA0.075	0.66	0.67
XA0.15	0.67	0.68
XAn0.075	0.67	0.61
Average	0.67	0.46
SD	0.01	0.22

The Ser432 and Ala302 amino acid M+0 MID values for each ILE demonstrate similar levels of ^{13}C incorporation, respective averages of 0.33 (± 0.06) versus 0.37 (± 0.08) for the glucose ILEs and respective averages of 0.67 (± 0.01) versus 0.37 (± 0.08) for the glucose ILEs. This suggests that for both substrates and under all oxygen conditions that glycolysis was still the main metabolic pathway for *P. thermoglucosidasius* and that there was minimal, if any, impact of the ED pathway or anaplerotic reactions on the pools of pyruvate.

Overall, the dynamic MIDs described in 4.4.3 were determined as suitable for further quantitative ^{13}C-MFA to generate flux models in chapter 4.4.4.

4.4.4 Quantitative INST-^{13}C-MFA

While several qualitative inferences can be made about *P. thermoglucosidasius* metabolism by comparing the MIDs of different nodes, quantitative ^{13}C-MFA is required to represent the above MID data in terms of calculated metabolic reaction fluxes. In order to perform the MID data fitting, an isotopomer transition network featuring the necessary carbon transitions was required.

Given its public availability (and published success in representing growth of *Geobacillus* on xylose as a sole substrate), the existing carbon transition network from the *Geobacillus* LC300 ^{13}C-MFA experiments of Cordova *et al.* (2016)[127] were used as a foundation. Notably, unlike the atom transition networks of Tang *et al.* (2009)[128] and C. Ward (2016)[121]. these reactions included the production of cofactors of: NADH, NADPH, ATP and the one carbon metabolism metabolites of: MEETHF, METHF and FTHF.

In the following work, flux estimations were performed in the INCA v.2.0 software using the amino acid mass fragments deemed valid for each experiment to minimize the SSR between the forward predicted MID solutions and the experimentally measured MID data. Importantly, for each fitting procedure flux estimation was restarted 10 times with randomized initial guesses for unconstrained flux and metabolite pool sizes to offer the best possibility of finding the global optimal flux solution. Each reported solution represents the solution with the lowest SSR as a result of this method.

The MID data set from the GA0.15 ILE was selected as the test case on account of having a broad range of valid fragments while also demonstrating low measurement error across the amino acids for the ^{13}C portion of the ILE. However, initial attempts at flux estimation did not yield an optimal solution with a statistically acceptable SSR. Indeed, the SSR value was several 10s of million higher than the maximum statistically allowed SSR. As a result, focus turned first to the evaluation of the carbon transition reactions and reaction stoichiometries themselves.

4.4.4.1 Development of the Carbon Transition Network

All reactions present in initial carbon transition network were represented in *p-thermo* and the *P. thermoglucosidasius* NCIMB 11955 genome, implying that the *Geobacillus* LC300 transition network[127] was a valid representation of *P. thermoglucosidasius*.

However, several potentially relevant reactions present in the *P. thermoglucosidasius* NCIMB 11955 genome were not included in the initial network. As the network was developed for aerobic growth on xylose, the first of these modifications was the straightforward introduction of 3 reactions to import glucose as a substrate:

Gluc.ext (abcdef) + ATP -> Gluc (abcdef)

Gluc (abcdef) + ATP -> G6P (abcdef)

Gluc.ext (abcdef) + PEP (ghi) -> G6P (abcdef) + Pyr (ghi)

These first two reactions were introduced based on the ATP-dependent glucose transporters in the genome and the third was introduced for anaerobic conditions to represent the present phosphotransferase system in *P. thermoglucosidasius*.

The second set of reactions added to the model concerned reactions associated with fermentative behaviours of *P. thermoglucosidasius*. While production and export of acetate and lactate were already in the network, a reaction representing PFL and reactions for the production and export of pyruvate, formate and ethanol were also added:

Pyr (abc) -> Pyr_eff (abc)

Pyr (abc) -> AcCoA (bc) + For_eff (a)

AcCoA (ab) -> Eth_eff (ab)

Additionally, based on the findings discussed in chapter 3.3.3.2 concerning ferredoxin linked enzymes, a reaction was introduced for the *p-thermo* and *P. thermoglucosisisus* genome encoded oxoglutarate:ferredoxin oxidoreductase and a pseudo-reaction allowing it's regeneration:

2*FeOx + AKG (abcde) -> 2*FeRed + SucCoA (abcd) + CO2 (e)

2*FeRed <-> 2*FeOx

Inclusion of these reactions and their respective efflux constraints in the carbon transition network for each fitting procedure was based on their detection HPLC in the CCFs of each ILE.

One of the key findings in the thesis of B. Lisowska was the presence of an NADP-dependent GAPDH in the genome of *P. thermoglucosidaisus* NCIMB 11955, the activity of which with NADP was validated through enzyme assay[113]. As a result, an NAPH variant of the pre-existing NADH-linked GAPDH was added to the reaction network:

GAP (abc) > 3PG (abc) + ATP + NADH

GAP (abc) -> 3PG (abc) + ATP + NADPH

Both reactions were also set to being unidirectional to avoid transhydrogenase-like cycling
Inclusion of this resulted in a 66% decrease in SSR and was thus maintained for all future networks.
Lastly, one reaction encoded in the *P. thermoglucosidasius* NCIMB 11955 genome but not in *p-thermo* or the carbon transition network was succinyl-CoA:acetate CoA-transferase [E.C. 2.8.3.18] which represented the reaction:

SucCoA (abcd) + Ac (ef) <-> Suc (abcd) + AcCoA (ef)

This introduced the possibility of an ATP and NADPH-free route to both regenerate acetate to AcCoA and support the citiric acid cycle. Given acetate was encoded to be produced as a result of cysteine biosynthesis, this could potentially impact both aerobic and anaerobc flux estimations. However, introduction of this reaction into the network increased the minimized SSR value by 13.2%, so it was not included in further analysis.

Once these reactions had been introduced, focus turned to reducing the size of the overall network by condensing connecting reactions that didn't involve carbon rearrangements.
The reactions involved in synthesis of lysine:

Asp (abcd) + Pyr (efg) + Glu (hijkl) + SucCoA (mnop) + ATP + 2 NADPH →

LL-DAP (½ abcdgfe + ½ efgdcba) + AKG (hijkl) + Suc (½ mnop + ½ ponm)

LL-DAP (½ abcdefg + ½ gfedcba) → Lys (abcdef) + CO_2 (g)

Were condensed to single reaction of:

Asp (abcd) + Pyr (efg) + Glu (hijkl) + ATP + 2*NADPH + SucCoA (mnop) -> Lys (efgdcb) + CO2 (a) + AKG (hijkl) + Suc (½ mnop + ½ ponm)

Lastly, reactions involving Ru5P were condensed to involve a single pool of P5P to reduce complexity and degrees of freedom in the PPP.

Constraints

Similarly to FBA, the constraints imposed on the directions of reactions, and their subsequent flux in a carbon transition network, shape the overall topology of the solution space and the number of degrees of freedom of the system.

The first of these constraints applied to the network were thermodynamic constraints of each reaction which determined whether individual reactions were considered unidirectional or reversible. The initial thermodynamic constraints of the underlying *Geobacillus* LC300 network on which this network is based were mostly unchanged. However, where reactions demonstrated unrealistic exchange fluxes and high standard error, the directionality of each reaction was deferred to the unidirectional form of the reaction utilised by Tang *et al.* (2009) for *P. thermoglucosidasius* M10EXG[128].

Biomass efflux constraints

The second of these constraints were net constraints which determined the rate of amino acid efflux to biomass, based on the experimentally determined *P. thermoglucosidasius* OD_{600}:DCW relationship of 1 OD to 0.424 and the amino acid compositions of *P. thermoglucosidaisus* DCW protein, measured ODs and substrate uptake rates (Table 36).

Table 36: Average glucose or xylose (substrate) uptake rates for all INST-^{13}C-MFA *P. thermoglucosidasius* ILEs.

ILE	Number of measurements considered (n)	Average Substrate uptake rate (gDCW/L/h)	Standard deviation	Average specifc rate of substrate uptake (qs)	Standard deviation
GA0.075	45	0.202	0.083	0.206	0.084
GA0.15	54	0.251	0.027	0.139	0.030
GA0.225	48	0.542	0.026	0.381	0.031
GA0.32	51	1.065	0.041	1.016	0.073
GMA0.075	48	0.895	0.027	3.609	0.224
GAn0.075	45	0.945	0.087	4.385	0.809
XA0.075	36	0.209	0.009	0.197	0.014
XA0.15	45	0.303	0.022	0.299	0.022
XAn0.075	42	0.404	0.027	1.853	0.416

Two different cell protein compositions of dry mass, performed by Shyam Maskapalli in 2014 (Appendix 3), were assumed and used to calculate the amino acid requirements for biomass production for each ILE in units mmol/gDCW.

Calculated concentrations of metabolic intermediates representing cellular macro molecules drained to biomass were taken from literature measurements of *Geobacillus* LC300 and assumed to be independent of the growth conditions of the ILE. These values were supplied to a biomass equation in further flux analysis (Table 37).

Table 37: Calculated concentrations for amino acid net constraints, and assumed macromolecule composition net constraints, supplied to the biomass equation for all INST-^{13}C-MFA ILEs. All in units of mmol/gDCW.

Metabolite	GA0.075	GA0.15	GA0.225	GA0.32	GAn 0.075	GMA 0.075	XA0.075	XA0.15	XAn 0.075
Glycine	0.411	0.850	0.613	0.456	0.053	0.056	0.455	0.453	0.048
L-Alanine	0.519	1.074	0.774	0.576	0.067	0.071	0.574	0.572	0.060
L-Valine	0.380	0.785	0.566	0.422	0.049	0.052	0.420	0.418	0.044
L-Leucine	0.351	0.725	0.523	0.389	0.045	0.048	0.388	0.386	0.041
L-Isoleucine	0.275	0.569	0.410	0.305	0.036	0.038	0.304	0.303	0.032
L-Serine	0.183	0.378	0.272	0.203	0.024	0.025	0.202	0.201	0.021
L-Threonine	0.271	0.560	0.404	0.301	0.035	0.037	0.300	0.299	0.031
L-Phenylalanine	0.146	0.302	0.218	0.162	0.019	0.02	0.162	0.161	0.017
L-Tyrosine	0.105	0.216	0.156	0.116	0.013	0.014	0.116	0.115	0.012
L-Tryptophan	0.104	0.214	0.155	0.115	0.013	0.014	0.115	0.114	0.012
L-Cysteine	0.109	0.225	0.162	0.121	0.027	0.015	0.120	0.120	0.013
L-Methionine	0.086	0.177	0.128	0.095	0.021	0.012	0.095	0.094	0.010
L-Lysine	0.320	0.662	0.477	0.355	0.041	0.044	0.354	0.352	0.037
L-Arginine	0.185	0.382	0.276	0.205	0.024	0.025	0.204	0.204	0.021
L-Histidine	0.078	0.162	0.117	0.087	0.010	0.011	0.089	0.086	0.090
L-Aspartate	0.224	0.463	0.334	0.249	0.058	0.030	0.248	0.247	0.026
L-Glutamate	0.325	0.672	0.485	0.361	0.084	0.044	0.359	0.358	0.038
L-Asparagine	0.224	0.463	0.334	0.249	0.058	0.030	0.248	0.247	0.026
L-Glutamine	0.325	0.672	0.485	0.361	0.084	0.044	0.359	0.358	0.038
L-Proline	0.168	0.348	0.251	0.187	0.022	0.023	0.186	0.185	0.194
P5P (RNA/DNA)	0.949	0.949	0.949	0.949	0.949	0.949	0.949	0.949	0.949
GAP (Lipid)	0.298	0.298	0.298	0.298	0.298	0.298	0.298	0.298	0.298
3PG	0.66	0.66	0.66	0.66	0.66	0.66	0.66	0.66	0.66
PEP	0.074	0.074	0.074	0.074	0.074	0.074	0.074	0.074	0.074
PYR	0.257	0.257	0.257	0.257	0.257	0.257	0.257	0.257	0.257
AcCoA (Lipids)	1.31	1.31	1.31	1.31	1.31	1.31	1.31	1.31	1.31
AKG	0.074	0.074	0.074	0.074	0.074	0.074	0.074	0.074	0.074
OAA (RNA/dNA)	0.48	0.48	0.48	0.48	0.48	0.48	0.48	0.48	0.48
G6P (Carbohydrate)	0.074	0.074	0.074	0.074	0.074	0.074	0.074	0.074	0.074
F6P (Cell Wall)	0.479	0.479	0.479	0.479	0.479	0.479	0.479	0.479	0.479

Additionally, measured concentrations of efflux products lactate, ethanol, pyruvate, formate and acetate were also calculated in terms of mmol g^{-1} DCW h^{-1} and as efflux with respect to 100 molecules of substrate for each oxygen-limited ILE (Table

38). These were added as constraints to the respective networks as UBs and LBs representing ±10% of the stated value.

Table 38: Average calculated net efflux constraints for all anerobic INST-^{13}C-MFA ILEs in absolute values and normalized to 100 molecules of feed substrate.

Metabolite	GAn0.075		GMA0.075		XAn0.075	
	Average concentration (mmol g^{-1} DCW h^{-1})	With respect to 100 glucose molecules	Average concentration (mmol g^{-1} DCW h^{-1})	With respect to 100 glucose molecules	Average concentration (mmol g^{-1} DCW h^{-1})	With respect to 100 xylose molecules
Lactate	4.737	10.042	2.708	13.430	0	0
Ethanol	0.192	0.417	0	0	0	0
Pyruvate	0.172	0.370	1.590	7.886	2.913	7.798
Formate	2.124	4.252	6.921	34.320	4.967	13.299
Acetate	3.780	7.424	8.979	44.525	6.565	17.576

Variation of the network P/O ratio

Lastly, the global energy consumption of the network in terms of ATP can influence the outcome of the phenotypic flux predictions[217]. As the atom transition network featured reactions for the oxidative phosphorylation of NADH and $FADH_2$, one constraint strategy to reduce the SRR was to tune the (AT)P:O ratio (or the number of molecules of phosphorylated per atoms of oxygen reduced to water).

The base oxidative phosphorylation reactions of the transition network were:

$$2*NADH + O_2 \rightarrow 4\ ATP$$

$$2*FADH2 + O_2 \rightarrow 2\ ATP$$

Two rounds of tuning the P/O ratio were performed for the network using the same GA0.15 data set at different points in the network development (Table 39).

Table 39: The impact of tuning the stoichiometries of the P/O ratio for oxidative phosphorylation reactions of NADH and $FADH_2$.

Variation 1			Variation 2		
NADH	**$FADH_2$**	**Change in SSR (%)**	**NADH**	**$FADH_2$**	**Change in SSR (%)**
3	2	-7.726	3	2	0
2.5	1.5	-6.907	3.1	2.1	-29.563
2.33	1.4	-3.111	3.2	2.1	-54.041
2	1	0	3.2	2.2	-58.417
1.85	1	3.152	3.3	2.2	-58.016
1.5	1	2.908	3.2	2.3	-45.610
1.33	0.45	0.834	3.4	2.3	-32.551
0.66	0.87	15.009	3.5	2.3	-7.877

Ultimately, this demonstrated optimum values of 3.2 ATP per NADH and 2.2 ATP per $FADH_2$ (Table 39). As a result, these stoichiometries were input for the two oxidative phosphorylation reactions for all ILE fitting attempts as additional constraints. Future work in this area would first apply this strategy for each independent ILE data set to evaluate how the optimal P/O ratio changes across the studied growth conditions.

Assumptions about measurement error

The calculated measurement errors themselves represent constraints on the overall system and directly influence the calculation of SSR. For the initial stages of the ^{13}C-data fitting process, focus had been placed on developing the carbon transition network itself to reduce the SSR. However, in a 2018 guide for ^{13}C-MFA for cancer biology Maciek Antoniewicz suggests that typical measurement errors for MS derived MID data could be up to ± 0.01[218]. On this basis, all measurement error in each amino acid mass fragment MID was either increased to 0.01 or left as its original value if greater than 0.01.

The effect on the obtained SSR value when applied to the fitting of the GA0.15 ILE data set was immediate, taking the obtained SSR value from 1,023,493 to 448,831, a 56% decrease and improvement. As a result, this to increase to ≥ ± 0.01 measurement error was applied to all data sets before further ^{13}C-fitting attempts.

4.4.4.2 Data fitting strategies: MID groupings

For initial flux estimates, two groupings of MID data sets were considered to infer metabolic understanding of *P. thermoglucosidasius* under the different growth conditions explored.

The first grouping strategy “All valid” involved simultaneous fitting of all MDVs for all amino acid fragments determined to be valid for each experiment (Tables 29-32 in chapter 4.4.2.2). This was also the approach taken for the final flux estimations.

The second group strategy “Nodes” was based on the work done by Tang *et al.* (2009)[126] who performed ^{13}C-MFA fitting using mass fragments from only 7 amino acids (Alanine, Glycine, Phenylalanine, Serine, Glutamate, Aspartate and Leucine), each assumed to represent a particular carbon intermediate node in CCM. Following the strategy, of a similar group of 9 amino acids were also chosen to represent CCM intermediate nodes (Alanine, Valine, Phenylalanine, Serine, Tyrosine, Histidine, Glutamate, Aspartate and Leucine) (Table 40).

Table 40: Amino acids featured in the “Nodes” MID data set strategy with their respective assumed CCM intermediate.

Amino Acid	CCM node/s represented
Alanine	Pyruvate
Valine	Pyruvate
Phenylalanine	E4P and PEP
Serine	3PG
Tyrosine	E4P
Histidine	P5P
Glutamate	AKG
Aspartate	OAA
Leucine	Pyruvate and AcCoA

Compared to the original strategy, this included Histidine (derived from P5P), Valine (as a second amino acid derived from pyruvate), Tyrosine (as a second amino acid derived from PEP and E4P) and used Serine as a single representative of intermediate 3PG, rather than Serine and Glycine. While less likely to represent a valid or accurate solution, this second fitting was much faster computationally and performed to identify reactions and amino acids of issue in the overall carbon transition network.

The intiial flux distributions comparing these approaches using for the aerobic glucose and xylose ILEs are presented in chapter 4.4.2.4.

4.4.4.3 Initial flux analysis for the aerobic Glucose and Xylose INST-^{13}C-MFA ILEs: SSR values and Goodness-of-fit analysis

Initial flux estimations were only performed for the aerobic glucose and xylose INST-^{13}C-MFA ILE data sets to understand what the key areas of the simpler aerobic transition network were that required further improvement. For each ILE, both the "All valid" and "Nodes" data set approaches were use for ^{13}C-data fitting. Each reported solution was the solution with the lowest SSR as a result of 10 restarts with random initial guesses for unconstrained flux and metabolite pool sizes. In each case, the obtained SSR value was greater than the maximum statistically acceptable SSR at a confidence interval of 95%, suggesting that the fits were not accepted for the aerobic glucose (Table 41) or xylose (Table 42) "All valid" or "Nodes" data set

Table 41: Goodness-of-fit analysis for analysis of the “all valid” and “node” aerobic glucose INST-^{13}C-MFA amino acid fragment data sets using the developing carbon transition network featuring the final obtained SSR values and the individual squared residuals (SRES) for each amino acid mass fragment.

Metabolite	GA0.075				GA0.15				GA0.225				GA0.32			
	All valid		Node		All valid		Node		All valid		Nodes		All valid		Nodes	
	SRES	Contribution	SRES	Contribution	SRES	Contribution	SRES	Contribution	SRES	Contribution	SRES	Contribution	SRES	Contribution	SRES	Contribution
Ala 302	76.34	25.81	30.44	23.55	1944.00	15.03	76.73	26.24	799.66	22.65	764.95	25.42	1369.50	24.09	1291.80	30.08
Gly 288	77.49	19.60			382940.00	18.73			243.89	21.70			409.57	21.29		
Val 330	40.68	17.16	64.21	18.38	439.49	18.03	79.72	22.63	81.50	17.82	65.31	21.01	103.96	20.71	130.54	22.64
Pro 328																
Pro 258	813.44	21.29											478.92	20.35		
Phe 336			99.16	17.80	301.95	17.01	93.66	21.49	63.62	19.29	37.69	18.02			92.05	21.65
Phe 234									627.35	20.95						
Ser 432	148.13	24.05	50.80	22.50	45236.00	26.70	46.94	40.23	404.20	22.08	536.56	35.74	404.70	21.90	498.25	37.48
Lys 488	1019.00	15.58			967.50	17.24			1103.20	18.41			646.71	21.04		
Lys 473	1668.00	14.86			3046.90	17.33			2992.80	19.07						
Lys 431	1045.80	14.58			990.51	17.24			909.74	17.04			642.79	21.04		
Lys 329	1067.10	15.55			1122.60	17.25			1481.80	16.79			1127.00	20.73		
Tyr 508																
Tyr 466	52.42	16.65	42.47	18.81	313.99	17.46	111.29	21.35	69.54	19.45	41.93	18.05	89.97	25.87	83.90	21.66
Tyr 438		21.00														
Tyr 364	58.88	35.28							88.37	23.39						
His 440	1775.30	23.43	1200.00	34.64	6957.60	18.02	6327.30	17.58	5158.30	27.51	4740.00	45.49	4156.60	39.39	3624.80	30.52
Glu 489	1017.80	23.42			744.59	17.05			1591.50	18.05	191.07	42.88	168.94	21.46		
Glu 474	1044.80	23.44			734.37	17.05	38.69	53.99	1677.20	18.05			200.26	21.46	207.95	28.55
Glu 404	778.76	14.00	614.51	29.33	1194.00	17.03			4774.10	25.07			652.10	21.96		
Glu 330																
Asp 460	854.80	20.17	398.13	25.03	414.17	19.49	319.35	22.56	721.35	28.58	396.44	24.49	1038.70	22.72	643.92	27.06
Leu 344	50.75	22.11	239.96	30.96	168.50	18.03	99.89	25.94	40.79	20.60	73.55	24.90	252.78	24.89	276.78	26.36
Ile 344	1716.80	17.13			1317.00	24.55			770.47	24.50			824.20	22.08		
	SSR	Expected SSR range	SSR	Expected SSR range	SSR	Expected SSR range	SSR	Expected SSR range	SSR	Expected SSR range	SSR	Expected SSR range	SSR	Expected SSR range	SSR	Expected SSR range
	13,282	1429-1778	2,740	576-717.3	448,831	1337-1548	7193.6	690.2 -843.5	23,599	1428-1646	6,848	613.4-758	12,886	1191.4-1390.4	6,850	708-865
	WRES mean	WRES SD	WRES mean	WRES SD	WRES mean	WRES SD	WRES mean	WRES SD	WRES mean	WRES SD	WRES mean	WRES SD	WRES mean	WRES SD	WRES mean	WRES SD
	-0.58	± 2.61	-0.41	± 1.76	-2.50	± 43.79	-0.49	± 2.67	-0.41	± 3.51	-0.37	± 2.76	-0.41	± 2.75	-0.43	± 2.58

Table 42: Goodness-of-fit analysis for analysis of the "all valid" and "node" aerobic xylose INST-^{13}C-MFA amino acid fragment data sets using the developing carbon transition network. This featued the final obtained SSR values, the individual squared residuals (SRES) for each amino acid mass fragment and the overall weighted residuals (WRES) of the complete data set.

Metabolite	XA0.075				XA0.15			
	All valid		Node		All valid		Node	
	SRES	Contribution	SRES	Contribution	SRES	Contribution	SRES	Contribution
Ala 302	399.95	12.07	479.64	14.76	504.91	16.04	495.83	15.87
Gly 288	23.52	18.58			1713.40	15.00		
Val 330	234.90	12.26	383.77	17.18	474.36	16.01	725.64	16.02
Pro 328	1468.00	23.05			74909.00	16.52		
Pro 258	583.13	19.13	602.01	16.06	1095.20	15.68	929.81	16.72
Phe 336	635.06	21.45			1119.80	14.90		
Phe 234	492.13	18.83	758.93	14.75	1431.70	14.87		
Ser 432	582.19	14.69			1053.30	15.57	679.07	79.92
Lys 488	568.94	14.69			1129.90	15.15		
Lys 473	804.62	15.21			2083.90	15.23		
Lys 431	306.15	18.34			1187.50	15.15		
Lys 329	310.88	18.47			1787.90	15.10		
Tyr 508	118.54	17.97			1923.40	15.47		
Tyr 466	53.48	18.38			988.75	15.66		
Tyr 438	9738.00	19.01			691.11	15.61		
Tyr 364	239.50	16.70	109.77	15.49	621.87	15.65	580.38	16.97
His 440	309.34	16.64	6360.00	50.24	13591.00	15.38	10404.00	37.25
Glu 489	852.60	24.19	441.99	41.69	1822.40	15.74		
Glu 474	368.66	15.01			1811.20	15.74		
Glu 404	131.37	12.78			1421.70	15.41	1899.50	18.47
Glu 330	399.95	12.07	470.16	28.10	11226.00	14.12		
Asp 460	23.52	18.58	300.11	18.73	612.54	16.70	1372.60	18.86
Leu 344	234.90	12.26	479.64	14.76	419.44	103.37	1091.30	15.91
Ile 344	1468.00	23.05			1235.40	17.95	495.83	15.87
	SSR	Expected SSR range	SSR	Expected SSR range	SSR	Expected SSR range	SSR	Expected SSR range
	18,554	1284-1490	9,906	663.7-814.11	124,856	1862.4-2109.3	18,178	841-1010
	WRES mean	WRES SD	WRES mean	WRES SD	WRES mean	WRES SD	WRES mean	WRES SD
	-0.85	± 3.19	-0.91	± 3.51	-1.57	± 7.03	-1.62	± 9.02

Although neither of these data fitting strategies yielded statistically acceptable SSR values, these initial fitting results indicated the amino acids contributing most the obtained SSR values and what needed further improvement in the fitting process.

While the outcome of the natural isotope abundance correction for the His440 histidine mass fragment was "conditionally valid" for the GA0.15 data set, for all other data sets evaluated in this analysis His440 was deemed valid. However, the most notable observation for the "Node" data sets across all the flux estimations is the impact on the goodness of fit of His440, which provided the largest contribution to the final SSR value (Tables 40 and 41). As the data was valid and the direct biosynthesis reaction of histidine itself was verified against the *P. thermoglucosidasius* genome, this suggested that proximal factors in the biosynthesis of histidine were incorrect in the underlying network. Indeed, Cordova *et al.* (2016) were only able to resolve the oxidative PPP through the integration of ^{13}C labelled CO_2 data which suggested a negligible flux (1 ± 1) through the ribulose monophosphate pathway[127], suggesting that this could be applied as a further constraint.

Parallel to this idea, these led to the consideration of new sources of potentially unlabelled carbon into the network as proxy reactions to support the ^{13}C fitting procedure.

The first of these ideas was more general and allowed for potential import of ^{12}C glucose into the network to account for the potential of pre-exisitng pools of naturally labelled carbon and would aid the fitting in situations where the predicted ^{13}C labelling pattern of mass fragments was higher than experimentally measured This was set as a free flux with an UB of 10 to represent up to 10% of the normalized glucose uptake rate.

The second of these ideas was more specific to the synthesis of histidine, which required as a reactant the one carbon metabolism metabolite FTHF. As a result, a free flux for import and export of central one carbon-metabolism precursor MEETHF was added to the network to account for pre-exisitng or variable metabolite pools of MEETHF. A similar strategy was also applied by Ward, who introduced a single carbon 'C_x' pool which could freely import and export.

The final modification added to the model was tangential to histidine biosynthesis and the PPP and introduced a reaction for the synthesis of OAA-derived RNA monomer Uridine monophosphate (UMP) from L-aspartate to the network (Asp (abcd) + CO_2 (e) -> UMP (abcd) + CO_2 (e)) and added it as a net constraint to the biomass equation with an assumed constant value of 0.119 mmol g^{-1} DCW h^{-1} derived from the biomass equation of *p-thermo*.

The impact of these additions for the ^{13}C fitting procedure was evaluated individually against the GA0.15 "nodes" data set (Table 43).

Table 43: Goodness-of-fit analysis for analysis of "nodes" GA0.15 INST-^{13}C-MFA amino acid fragment data set comparing the individual impact of new network modifications. This featued the final obtained SSR values, the individual squared residuals (SRES) for each amino acid mass fragment and the overall weighted residuals (WRES) of the complete data set.

Metabolite	GA0.15 Nodes data set							
	Original		UMP in Biomass Eq.		MEETHF		Pre-existing ^{12}C	
	SRES	Contribution	SRES	Contribution	SRES	Contribution	SRES	Contribution
Ala 302	76.73	26.24	55.58	22.18	74.78	27.28	92.06	21.71
Val 330	79.72	22.63	46.08	27.24	55.34	42.98	69.09	24.32
Phe 336	93.66	21.49	87.96	18.76	99.65	19.02	111.56	25.92
Ser 432	46.94	40.23	325.37	32.42	309.72	31.14	324.05	28.61
Tyr 466	111.29	21.35	101.14	18.60	113.54	18.90	126.96	26.10
His 440	6327.30	17.58	4141.90	24.64	4142.00	26.06	4237.30	45.34
Glu 474	38.69	53.99	175.21	35.55	187.65	33.70	187.34	46.92
Asp 460	319.35	22.56	368.79	49.82	395.62	30.74	428.25	26.25
Leu 344	99.89	25.94	53.42	36.80	72.90	32.17	95.63	25.83
	SSR	Expected SSR range	SSR	Expected SSR range	SSR	Expected SSR range	SSR	Expected SSR range
	7193.6	690.2 - 843.5	5355.4	690.2 - 843.5	5451.2	690.2 - 843.5	5672.24	688.3 - 841.4
	WRES mean	WRES SD	WRES mean	WRES SD	WRES mean	WRES SD	WRES mean	WRES SD
	-0.49	± 2.67	-0.26	± 2.32	-0.28	± 2.34	-0.24	± 2.35

As a result, all three of these modifications were added to the underlying network for the glucose experiments and the UMP and MEETHF modifications for the xylose networks with a respective ^{12}C-xylose import reaction. Due to time constraints, these were the final additions to the carbon transition network.

4.4.4.4 Final flux analysis for the aerobic and anaerobic Glucose and Xylose INST-^{13}C-MFA ILEs: SSR values and Goodness-of-fit analysis

Initial flux estimations were only performed for the aerobic glucose and xylose INST-^{13}C-MFA ILE data sets to understand what the key areas of the simpler aerobic transition network were that required further improvement. For each ILE, the “All valid” data set approaches were use for ^{13}C-data fitting (Tables 44-46). Each reported solution was the solution with the lowest SSR as a result of 10 restarts with random initial guesses for unconstrained flux and metabolite pool sizes.

Table 44: Goodness-of-fit analysis for analysis of "All Valid" INST-^{13}C-MFA amino acid fragment data sets for all aerobic glucose ILEs comparing the collective impact of new network modifications. This featued the final obtained SSR values, the individual squared residuals (SRES) for each amino acid mass fragment and the overall weighted residuals (WRES) of the complete data set.

Metabolite	GA0.075 All Valid Data Set				GA0.15 All Valid Data Set				GA0.225 All Valid Data Set				GA0.32 All Valid Data Set			
	Original Network		Final Network		Original Network		Final Network		Original Network		Final Network		Original Network		Final Network	
	SRES	Contribution	SRES	Contribution	SRES	Contribution	SRES	Contribution	SRES	Contribution	SRES	Contribution	SRES	Contribution	SRES	Contribution
Ala 302	76.34	25.81	1336.30	23.88	1944.00	15.03	101.29	27.09	799.66	22.65	706.63	28.28	1369.50	24.09	1348.3	22.7148
Gly 288	77.49	19.60	409.17	27.15	382940.00	18.73	1054.8	30.3103	243.89	21.70	401.63	23.67	409.57	21.29	420.4639	25.3039
Val 330	40.68	17.16	106.29	20.71	439.49	18.03	95.10	20.71	81.50	17.82	69.13	20.69	103.96	20.71	106.1639	25.7292
Pro 328																
Pro 258	813.44	21.29	499.03	19.81									478.92	20.35	501.0137	21.4101
Phe 336					301.95	17.01	96.89	18.98	63.62	19.29	42.62	16.40				
Phe 234									627.35	20.95	705.66	16.46				
Ser 432	148.13	24.05	439.47	24.32	45236.00	26.70	53.78	20.59	404.20	22.08	537.90	23.06	404.70	21.90	440.9362	22.6676
Lys 488	1019.00	15.58	640.50	20.06	967.50	17.24	851.96	18.9275	1103.20	18.41	1079.80	18.61	646.71	21.04	637.148	24.9721
Lys 473	1668.00	14.86	635.92	20.05	3046.90	17.33	2847.1	20.0566	2992.80	19.07	2942.00	18.77				
Lys 431	1045.80	14.58	1132.90	20.34	990.51	17.24	862.63	18.93	909.74	17.04	904.60	17.08	642.79	21.04	633.5634	24.9684
Lys 329	1067.10	15.55	439.47	24.32	1122.60	17.25	1432	19.51	1481.80	16.79	1526.30	17.33	1127.00	20.73	1130.7	19.6214
Tyr 508																
Tyr 466	52.42	16.65	104.11	27.47	313.99	17.46	110.8201	18.8572	69.54	19.45	46.29	16.43	89.97	25.87	105.0775	20.2355
Tyr 438																
Tyr 364	58.88	35.28	147.15	26.79					88.37	23.39	100.85	18.46				
His 440	1775.30	23.43	3973.40	40.61	6957.60	18.02	5364	39.7365	5158.30	27.51	5104.6	25.88	4156.60	39.39	4014.5	28.2215
Glu 489	1017.80	23.42	152.31	20.79	744.59	17.05	32.04	23.42	1591.50	18.05	1837	21.05	168.94	21.46	164.4851	27.9601
Glu 474	1044.80	23.44	181.64	20.79	734.37	17.05	31.98	23.40	1677.20	18.05	1941	21.03	200.26	21.46	193.7961	28.004
Glu 404	778.76	14.00	686.78	21.18	1194.00	17.03	445.68	21.8343	4774.10	25.07	4857	28.99	652.10	21.96	693.7675	28.862
Glu 330																
Asp 460	854.80	20.17	958.63	20.56	414.17	19.49	696.51	22.48	721.35	28.58	1260.7	22.7	1038.70	22.72	982.3842	21.957
Leu 344	50.75	22.11	300.12	47.46	168.50	18.03	78.55	19.92	40.79	20.60	45.27	22.76	252.78	24.89	274.4637	25.7261
Ile 344	1716.80	17.13	785.73	22.83	1317.00	24.55	819.34	22.23	770.47	24.50	932.34	20.35	824.20	22.08	796.29	20.6463
	SSR	Expected SSR range	SSR	Expected SSR range	SSR	Expected SSR range	SSR	Expected SSR range	SSR	Expected SSR range	SSR	Expected SSR range	SSR	Expected SSR range	SSR	Expected SSR range
	13,282	1429-1778	12,489	1429-1778	448,831	1337-1548	14,974	1337-1548	23,599	1428-1646	25041	1428-1646	12,886	1191.4-1390.4	12,443	1191.4-1390.4
	WRES mean	WRES SD	WRES mean	WRES SD	WRES mean	WRES SD	WRES mean	WRES SD	WRES mean	WRES SD	WRES mean	WRES SD	WRES mean	WRES SD	WRES mean	WRES SD
	-0.58	± 2.61	-0.42	2.72	-2.50	± 43.79	-2.5	± 43.79	-0.41	± 3.51	-0.46	± 3.61	-0.41	± 2.75	-0.4	± 2.74

Table 45: Goodness-of-fit analysis for analysis of "All Valid" INST-^{13}C-MFA amino acid fragment data sets for all aerobic xylose ILEs comparing the collective impact of new network modifications. This featued the final obtained SSR values, the individual squared residuals (SRES) for each amino acid mass fragment and the overall weighted residuals (WRES) of the complete data set.

Metabolite	XA0.075 All Valid Data Set				XA0.15 All Valid Data Set			
	Original Network		Final Network		Original Network		Final Network	
	SRES	Contribution	SRES	Contribution	SRES	Contribution	SRES	Contribution
Ala 302	399.95	12.07	421.14	13.19	504.91	16.04	415.60	16.0
Gly 288	23.52	18.58	277.38	15.09	1713.40	15.00	138.11	22.41
Val 330	234.90	12.26	400.28	13.93	474.36	16.01	180.09	16.62
Pro 328	1468.00	23.05	2048.30	18.29	74909.00	16.52	45972	26.64
Pro 258								
Phe 336	583.13	19.13	768.05	15.15	1095.20	15.68	601.02	15.55
Phe 234	635.06	21.45	986.54	14.00	1431.70	14.87	726.50	16.05
Ser 432	492.13	18.83	768.40	23.60	1053.3	15.5733	712.22	22.39
Lys 488	582.19	14.69	984.06	14.48	1129.90	15.15	656.97	16.19
Lys 473					2083.90	15.23	2170.9	16.48
Lys 431	568.94	14.69	960.03	14.48	1187.50	15.15	702.6401	16.17
Lys 329	804.62	15.21	1331.90	14.10	1787.90	15.10	1077.7	50.79
Tyr 508	306.15	18.34	411.67	13.40	1923.40	15.47	1493.2	15.23
Tyr 466	310.88	18.47	422.430	13.42	988.75	15.66	503.33	15.29
Tyr 438	118.54	19.01	157.016	13.16	691.11	15.61	369.36	15.24
Tyr 364	53.48	16.70	99.91	13.24	621.87	15.65	276.27	15.26
His 440	9738	16.64	7092	32.75	13591.00	15.38	16,944	16.72
Glu 489	239.50	24.19	635.35	20.50	1822.40	15.74	202.20	17.27
Glu 474	309.34	15.01	764.39	20.43	1811.20	15.74	195.037	17.26
Glu 404	852.60	12.78	1479.8	18.46	1421.70	15.41	464.67	16.77
Glu 330					11226.00	14.12	10827	14.40
Asp 460	368.66	18.58	962.72	30.88	612.54	16.70	279.90	17.42
Leu 344	131.37	12.26	460.47	14.53	419.44	103.37	153.43	16.50
Ile 344	389.61	23.05	713.12	23.69	1235.40	17.95	2598.10	42.09
	SSR	Expected SSR range	SSR	Expected SSR range	SSR	Expected SSR range	SSR	Expected SSR range
	18,610	1284-1490	22,145	1373-1586	124,856	1862.4-2109.3	88,160	1860-2107
	WRES mean	WRES SD	WRES mean	WRES SD	WRES mean	WRES SD	WRES mean	WRES SD
	-0.85	± 3.19	-0.85	± 3.19	-1.57	± 7.03	-1.22	± 5.93

Table 46: Goodness-of-fit analysis for analysis of "All Valid" INST-^{13}C-MFA amino acid fragment data sets for all anaerobic ILEs. This featued the final obtained SSR values, the individual squared residuals (SRES) for each amino acid mass fragment and the overall weighted residuals (WRES) of the complete data set.

Metabolite	GMA0.075 All Valid Data Set		GAn0.075 All Valid Data Set	
	Final Network		Final Network	
	SRES	Contribution	SRES	Contribution
Ala 302	591.43	22.94	861.56	33.53
Gly 288	214.82	20.87	206.53	16.48
Val 330	134.52	17.89	239.08	19.32
Phe 336	265.78	19.15		
Phe 308	381.58	18.34		
Phe 234	327.02	18.36	697.97	22.71
Ser 432	404.26	20.75	729.82	35.29
Lys 488	696.52	23.45	451.15	27.42
Lys 431	688.28	23.41		
Lys 329	1008.00	18.25		
Tyr 508	267.62	19.03		
Tyr 466	265.09	19.04	508.42	22.47
Tyr 438	367.11	18.76	1920.20	38.68
Tyr 364	385.87	18.71		
His 440	1710.40	31.29		
Glu 489	119.20	17.69		
Glu 474	149.03	17.69	279.11	20.25
Glu 404	650.61	17.27		
Asp 460	842.36	22.24	736.31	24.16
Asp 390	5234.70	18.23		
Leu 344	185.16	17.90	284.67	21.22
Ile 344	830.08	21.43		
Lac 303	835.31	22.29		
Lac 261	609.91	24.72		
	SSR	Expected SSR range	SSR	Expected SSR range
	17,375	2004-2260	7488	802-965
	WRES mean	WRES SD	WRES mean	WRES SD
	-0.35	± 2.56	-0.63	± 2.48

Metabolite	XAn0.075 All Valid Data Set	
	Final Network	
	SRES	Contribution
Ala 232	9494.00	12
Val 330	6518.00	11
Pro 328	12281.00	12
Met 320	7300.40	11
Phe 336	5635.80	12
Ser 362	10601.00	12
Lys 488	6766.90	12
Tyr 466	5701.40	12
Glu 404	9746.40	11
Asp 418	7959.00	12
	SSR	Expected SSR range
	82,004	462-588
	WRES mean	WRES SD
	-6.77	± 8.30

Once again however, the obtained SSR value was greater than the maximum statistically acceptable SSR at a confidence interval of 95% for all data sets, suggesting that the fits were not accepted for any of the "All valid" data sets with the updated model (Table 45). Comparison against the pre-existiing "All valid" data set fits suggested these modifications had different impacts of the fitting of the complete amino acid mass fragment data sets for the different ILEs. Reductions in obtained SSRs were observed for the GA0.075, GA0.15, GA0.32 and XA0.15 "all valid" data sets representing respective fit improvements of: 5.97%, 96.66%, 3.51% and 29.39%, whereas increases were observed for the "all valid" data sets of GA0.225 and XA0.075 with respective fit impacts of -6.11% and -20%. Frustratingly, this analysis suggested that the Gly288 MID data of the benchmark GA0.15 data set was a likely contributor to the overall changes in SSR observed. These results suggest these modfications need to be investigated on a case-by-case basis and further examined on whether such changes should be permanent. Additionally, future work should consider whether introducing additional pseudo-reactions to account for potentially pre-existing pools of individual amino acid or other metabolite pools would aid the ^{13}C-fitting.

4.4.4.5 Initial flux maps of *P. thermoglucosidasius* NCIMB 11955 on glucose and xylose and under aerobic and anaerobic conditions

While the resulting flux distributions suggest that a statiscally acceptable fit of INST-^{13}C-MFA data to model was ultimately not achieved in this research cannot represent a final statement on *P. thermoglucosidasius* metabolic behaviour for these ILEs, comparing the best fit flux distribution for GA0.075, GAn0.075, XA0.075 and XAn0.075 do suggest limited similarities with existing research and do aslo help point towards areas of future network refinement (Figures 60-63).

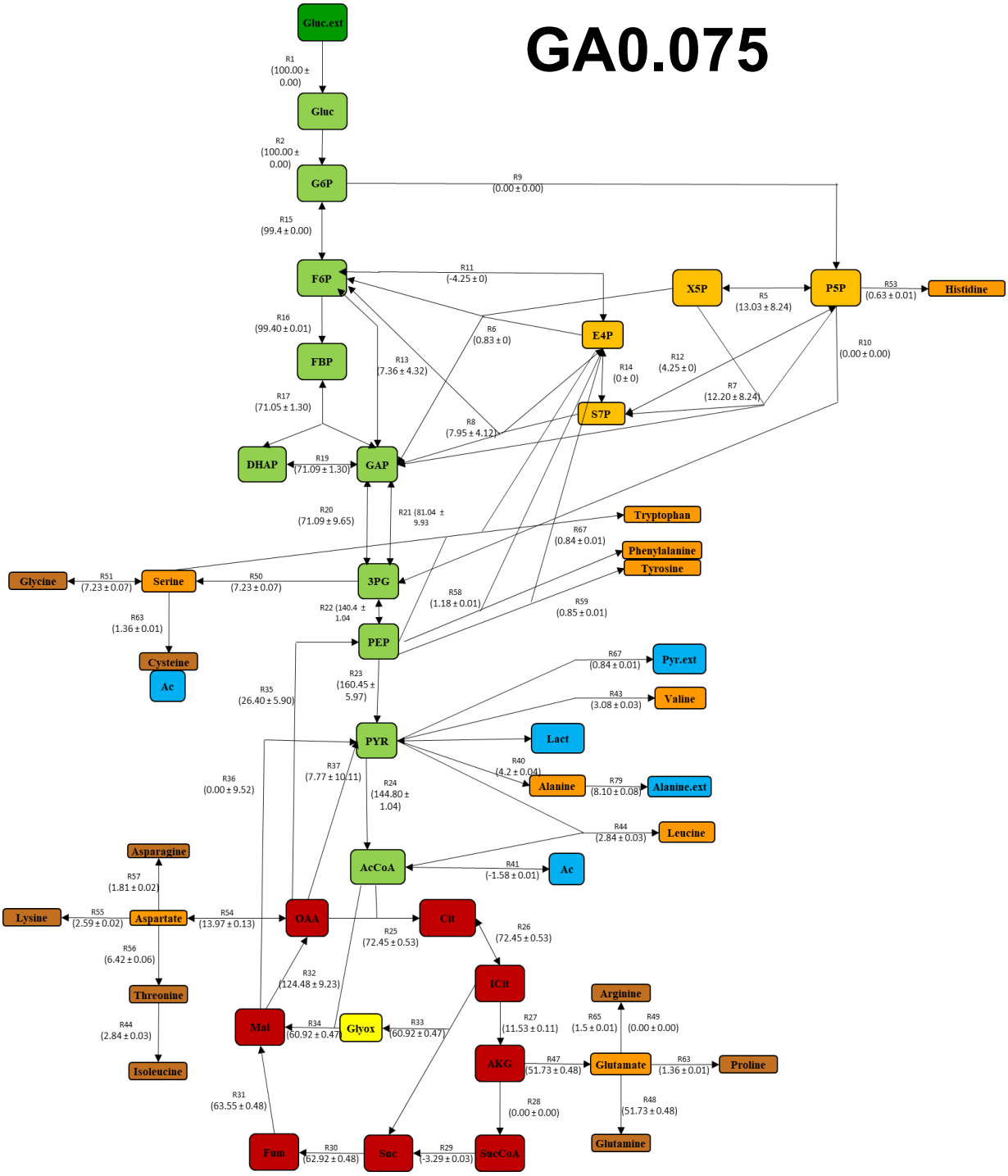


Figure 60: Flux map of *P. thermoglucocidasius* NCIMB 11955 grown on glucose as a sole substrate under aerobic conditions at a growth rate of 0.075h^{-1}.

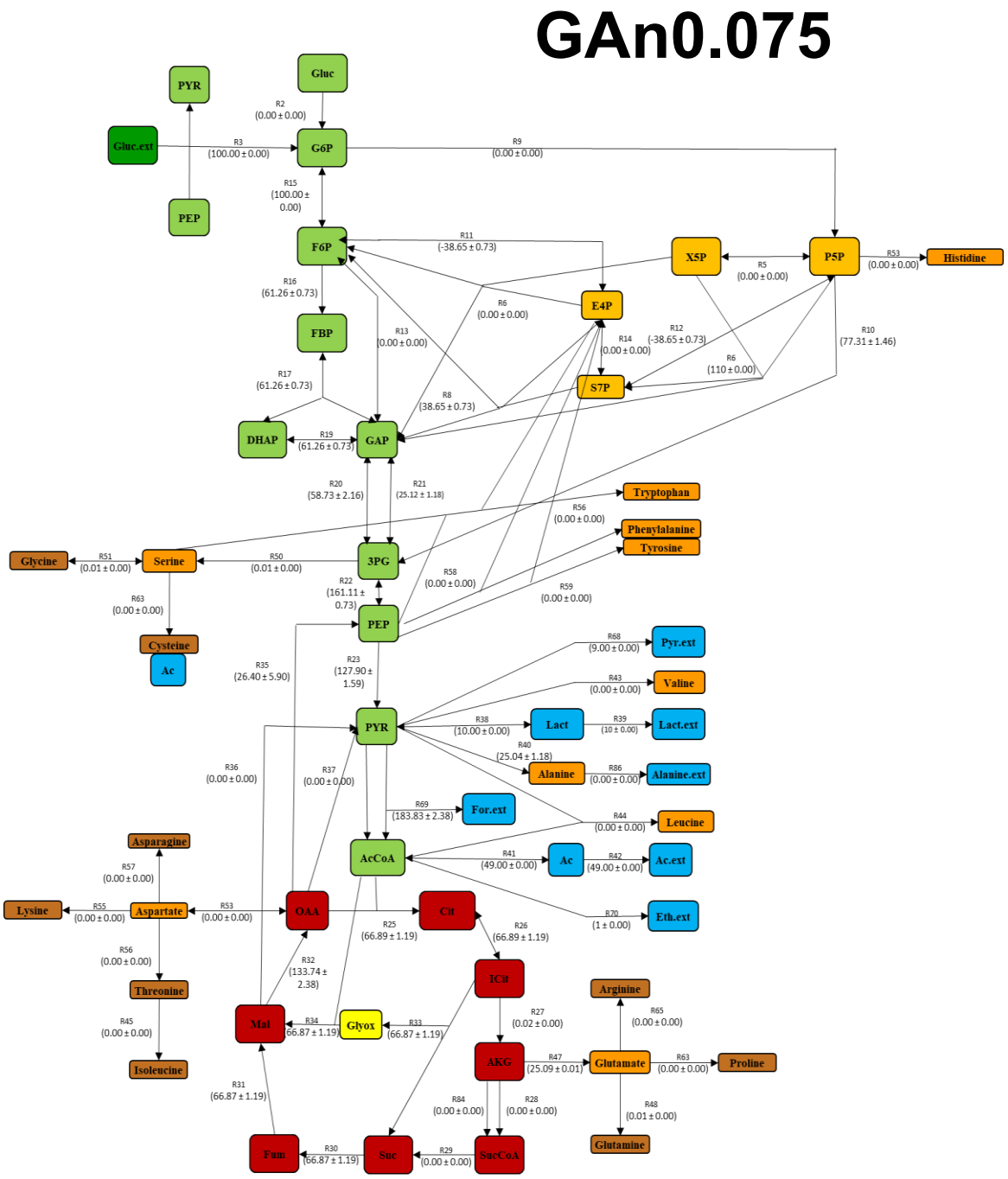


Figure 61: Flux map *of P. thermoglucocidasius* NCIMB 11955 grown on glucose as a sole substrate under anaerobic conditions at a growth rate of 0.075h^{-1}.

The initial *in vivo* flux distributions of *P. thermoglucosidasius* grown under aerobic and anaerobic conditions on gluose as a sole substrate do identify potential areas for improvement. In both cases there were a preference for the glycolytic pathway with little flux seen for the PPP. This absence of oxPPP flux (R9) clashes with statistically accepted flux distributions generated for *P. thermoglucosidasius* NCIMB 11955 as generated by Ward (15 (± 2) anaerobic oxPPP flux)[121], M10EXG by Tang *et al.* (2009) (aerobic oxPPP flux (30 ± 2), anaerobic oxPPP flux 15 (± 2))[128], and *Geobacillus* LC300 as generated by Cordova *et al.* (2017) (aerobic oxPP flux 37 (± 0)[170]. Combined this suggests that the glycolytic pathway has been represented correctly but provides further evidence that the PPP needs readdressing.

Additionally, both flux maps demonstrate a clear dependency on the glyoxylate shunt in the TCA cycle (R33 + R34) and negible flux through alpha-ketoglutarate dehydrogenase (R28). While potentially expected under low oxygen conditions, its mirrored flux under aerobic conditions suggests this reaction should also be evaluated against the glyoxylate cycle to ensure co-factor demands are adequately met.

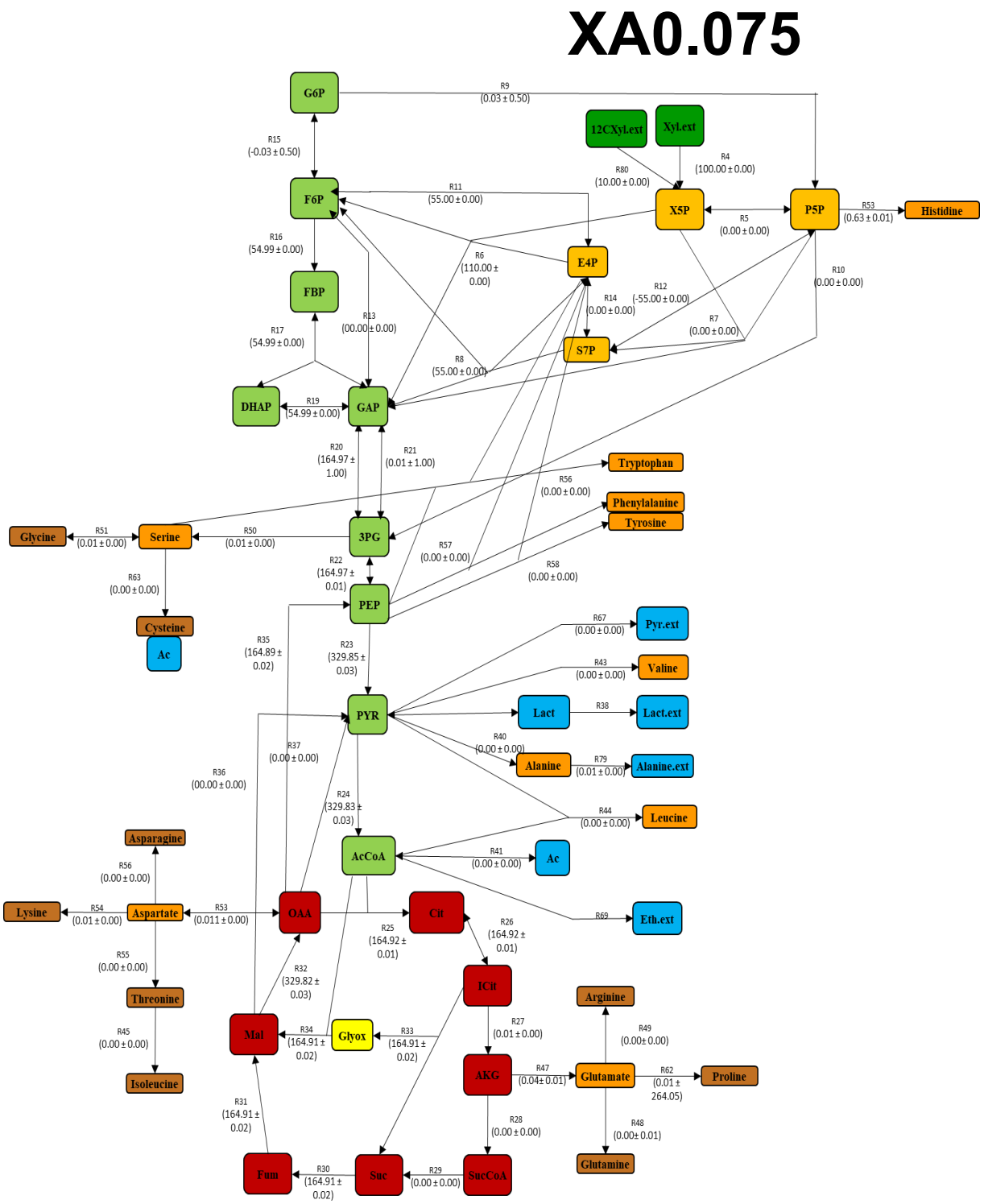


Figure 62: Flux map *of P. thermoglucocidasius* NCIMB 11955 grown on xylose as a sole substrate under aerobic conditions at a growth rate of 0.075h^{-1}.

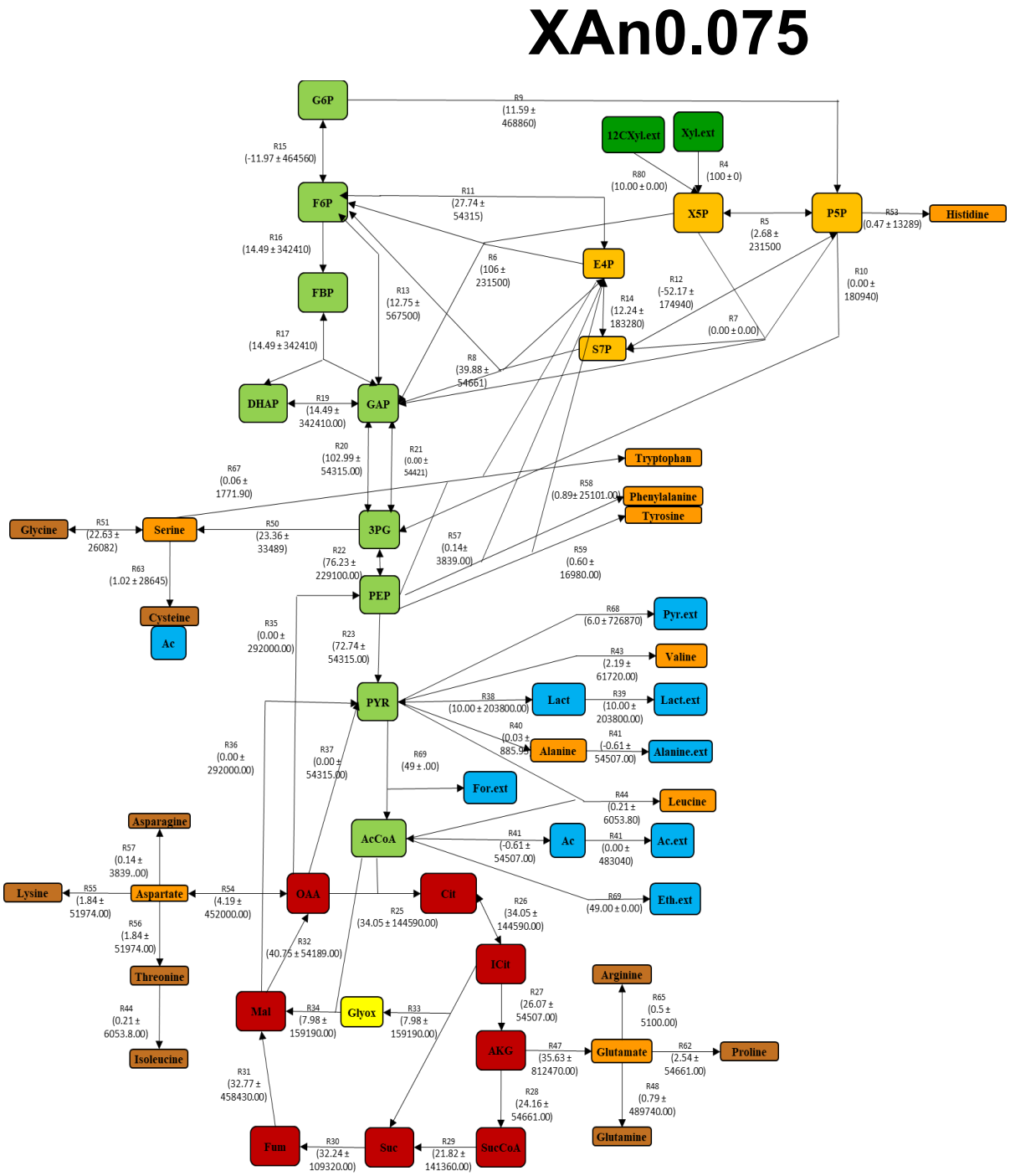


Figure 63: Flux map *of P. thermoglucocidasius* NCIMB 11955 grown on glucose as a sole substrate under anaerobic conditions at a growth rate of 0.075h^{-1}.

The initial *in vivo* flux distributions of *P. thermoglucosidasius* grown under aerobic and on xylose as a sole substrate also identify potential areas for improvement. Once again, there is no flux through the oxPPP (R9) which was observed for Geobacillus LC300 grown on xylose at a flux value of 16 (± 0.5)[127], suggesting once again that there are issues with the PPP. Notably, the biomass output of the best fit flux distribution in this case was incredibly low compared to that of glucose and is reflected in the neglible fluxes towards amino acid biosynthesis. Far less can be observed from the initial flux distribution of anaerobic growth on xylose. In

accordance with it's high SSR value, the standard deviation for all fluxes in the network was far too high to draw meaningful conclusions from this flux distribution.

Overall, these results demonstrate that, while the amino acid mass fragment data sets for each ILE themselves are likely accurate, the combination of statistically unacceptable SSR values suggests further network refinement is required to adequately fit the data to a flux map for further analysis and potential integration alongside p-thermo. Indeed, given the determined validity and low SD associated with the His440 mass fragments in the experiments where it was observed, this suggests that future work should focus on refining the network reactions which constitute the PPP to achieve more accurate flux distributions.

5. Chapter 5: General Discussion and future work

The motivation behind this research was to contribute to the field of biotechnology in a way which could help meet some of the global sustainability challenges encompassed within the UN SDGs. Whether at the scale of the global environment or a single bacterial cell, taking a systems-level view of a challenge which tries to account for the many underlying variables which influence the observed behavior can help us to both understand problems and indicate potential solutions.

Progressively, applications of systems biology (including computational systems modelling and functional multi-omics analysis) are being applied to metabolic engineering strategies.This form of a 'Design-Build-Test' approach is neatly described through the umbrella term: SysME. The overall aim of this research was to apply this integrated SysME approach to the thermophilic facultative anaerobe *P. thermoglucosidasius*, a microbial chassis which offers the potential of being developed for sustainable bioconversion of renewable lignocellulosic waste to numerous products through its thermophilicity and catabolic versatility.

While the concurrent development of a cumulative suite of genetic engineering tools and standardized genetic parts had been developed specifically for *P. thermoglucosidasius*, the metabolic behaviours of *P. thermoglucosidasius*, particularly under micro-aerobic and anaerobic conditions, were understudied compared to many more industrially utilised microbial species. Thus, to aid the development of *P. thermoglucosidasius* as a microbial chassis for more environmentally friendly chemical production, this research focused on the 'computational modelling' and '(flux)omics' analysis aspects of the SysMe approach to develop computational tools and acquire metabolic data of *P. thermoglucosidasius* NCIMB 11955 to help accurately understand and predict its cellular metabolism.

5.1 Computational genome-scale metabolic modelling and Flux Balance Analysis

The aim of generating a computational modelling tool in this research was built upon previous PhD research to generate a GSMM representing *P. thermoglucosidasius* NCIMB 11955 metabolism. However, several structural inaccuracies within the model meant that it was limited in its ability to accurately predict *P. thermoglucosidasius* NCIMB 11955 metabolism.

Ultimately, in partnership with the other authors and building upon the groundwork of the model which came before it, this research was able to improve the original GSMM of *P. thermoglucosidasius*. Through iterative cycles of analysis and manual curation with the Memote tool this research improved the original model such that *p-thermo* demonstrates a 100% stoichiometric consistency, metabolite connectivity, charge balance and unbounded flux in our default growth medium, as well as a 99.9% mass balance. As a result, this novel model of *P. thermoglucosidasius* NCIMB 11955 could finally be published for the first time and currently represents the most accurate GSMM of *P. thermoglucosidasius* available.

The most valuable aspect of this research for future research into *P. thermoglucosidasius* is that *p-thermo* can now perform genome-wide analysis of *P. thermoglucosidasius* metabolism and predict flux distributions throughout the metabolic network. In particular, it can do so with an increased degree of accuracy which reflected experimentally measured fluxes of *Geobacillus* M10EXG[128] and existing *P. thermoglucosidasius* mutants under the low-oxygen conditions which are

most relevant to the *P. thermoglucosidasius* fermentation products of lactic acid and ethanol.

In terms of future computational modelling research towards sustainable bioprocesses, a key first thing that could be done while further developing the accuracy of the *p-thermo* model is to simulate the growth of *P. thermoglucosidasius* on more complex carbon substrates or mixtures of them. This could be achieved through two steps. First new sugar compositions and media could be composed which reflect experimentally defined, real-world lignocelliulosic biomass hydrolysate sugars after pre-treatment. Second, p-thermo could then be used to simulate how *P. thermoglucosidasius* could grow on them and how they could produce various bio-products. Additionally, *p-thermo* could be further validated through the generation of additional *P. thermoglucosidasius* NCIMB 11955 mutant models as generated experimentally by Sheng et al. (2017)[1], including a similar lactate dehydrogenase double mutant also lacking a uracil metabolism enzyme pyrE, LS001 (ΔpyrEΔldh). As the PyrE mutation was used as a selection system, it would be interesting to see if there was an impact on metabolism of this mutation. Alternatively, new reactions could be added to the *p-thermo* model to replicate the experimental production of isobutanol in *P. thermoglucosidasius* by Lin *et al.* (2014)[97] or terpenoid production by Styles *et al.* (2020)[102].

A key aim for the GSMM research was to use the model to suggest what could be done to support anaerobic growth of *P. thermoglucosidasius*. Perhaps the most interesting results of the validation of *p-thermo* therefore were the *in-silico* predicted and experimentally-supported results that showed that a combination of thiamine, biotin and iron(III) could support anaerobic growth of *P. thermoglucosidasius* The requirement for thiamine was suggested to be a result of both the oxygen-dependent synthesis of TPP and the TPP dependency of key anaerobic enzyme PFL (PFL), whereas supplementation of iron(III) was postulated to be supportive to several genome-annotated ferredoxin-dependent metabolic reactions of *P. thermoglucosidasius*, which originated from iron and sulphur rich hot springs. Genomic evidence also suggested that production of heme was restricted under anaerobic conditions due to the presence of an oxygen-dependent protoporphyrinogen/coproporphyrinogen III oxidase (HemY).

From the perspective of SB, future work could therefore investigate whether the heterologous expression of a thermophilic 2-iminoacetate synthase, oxygen-independent HemN and oxygen-independent HemG in *P. thermoglucosidasius* supports growth under low oxygen conditions. Additionally, if suggested to confer a growth advantage under anaerobic conditions when compared to the wild-type, plasmids bearing these genes could theoretically form the basis of an antibiotic-free selection system for *P. thermoglucosidasius*. Additionally, on account of the predicted oxygen sensitivity of the native glycine oxidase involved in thiamine biosynthesis, future work could also investigate whether introduction of a thermophilic 2-iminoacetate synthase in *P. thermoglucosidasius* could support growth under low oxygen conditions.

5.2 INST-^{13}C-MFA of *P. thermoglucosidasius* NCIMB 11955

In tandem to the development of *p-thermo*, the second key aim of this research was to generate fluxomic data of *P. thermoglucosidasius* to further validate, and potentially incorporate into, the GSMM to enhance its capability to reflect growth under a range of growth conditions.
To the author's knowledge, this represents the first INST-^{13}C-glucose and ^{13}C D-xylose feeding ^{13}C-MFA ILEs for a *Parageobacillus* or *Geobacillus* species in fixed dilution/growth rate chemostat systems.

Focus was given during the initial experimental design phase to try and improve on the existing fluxomic analysis of *Parageobacillus* and *Geobacillus* species. This was done by using a defined minimal media without inclusion of yeast extract or additional amino acid carbon sources, by screening for carbon-limitation of cultures a different growth rates through the monitoring of OD_{600} and substrate concentrations, by giving particular focus to ensuring that each experiment was an accurate representation of *P. thermoglucosidasius* metabolism under the target oxygen growth conditions through the measurement of redox potential and by performing ^{13}C-MFA with the additional novel development of detailing isotopically in-stationary ^{13}C isotope labelling of *P. thermoglucosidasius* NCIMB 11955 amino acids over time.

To perform INST-^{13}C-MFA, it was necessary to establish a workflow for the analysis at the University of Bath. Chapter 4.3 collates a collection of key experiments performed over the course of the research involved with establishing the small volume MKI and MKII CSTR growth systems. While only key early experiments have been mentioned, these experiments represented a critical part of the overall aim of developing a viable INST-^{13}C-MFA workflow and was a revival of the research group's ability to perform chemostat growth experiments and ^{13}C-MFA experiments. The approach of using HPLC-RI-UV to monitor the concentrations of metabolic products and residual sugar allowed a direct quantification of whether each chemostat growth experiment had achieved the desired metabolic steady state under the target aeration conditions.

Initial setbacks in performing growth experiments were ultimately solved through the inclusion of antifoam 204 in the ASM. Through a variation of the: ASM sugar concentrations, air flow rates, agitator speeds and additional media supplements, the results of Chapter 4.3 demonstrate that aerobic and microaerobic growth on minimal, defined media could be achieved for *P. thermoglucosidaisus* cells grown under chemostat conditions. Through the additional supplement of $FeCl_3$ as suggested by the *p-thermo* model, these results suggest that threshold conditions of ethanolgenesis may have been reached for *P. thermoglucosidasius* in the GAn0.075 ILE, but that future work can improve upon these conditions and investigate whether stable populations of *P. thermoglucosidasius* can be maintained at even lower culture redox potentials, which would likely represent a greater degree of ethanologenesis.

One of the most confounding issues with the overall experimental design resulted from the prohibitive cost of the ^{13}C sugars themselves, particularly with respect to the anerobic growth expriments. For the anerobic growth experiments it was necessary to work within a series of very narrow constraints: a minimum working volume of 350 mL, a minimum air flow rate of 50 mL/min or 0 and a ^{13}C sugar cost which, particularly given the continuous culture strategy, demanded a low % sugar in the feed media. Inversely however, more carbon could increase the size of the cell population, reducing the available concentration of O_2 per cell and could potentially encourage sustainable anaerobic growth conditions. Indeed, *P. thermoglucosidasius* has the capacity to be grown fermentatively on 1% ASM and at a higher working volume (which enables a lower concentration of dissolved oxygen at the 50 mL/min minimum operating flow rate of the equipment). Future work with this set up should therefore aim to use a greater concentration of ^{13}C tracer to achieve more anerobic conditions and greater degrees of ethanologenesis. An initial attempt was made to grow *P. thermoglucosidasius* cells aerobically on ^{12}C xylose at a dilution rate of 0.225 h^{-1} to mirror the glucose experiments. However, this experiment proved to be the final performance for the Braun unit, which could not be repaired after a final heater failure. Future work could therefore aim to perform this

growth experiment of *P. thermoglucosidasius* using these conditions. Overall however, these conditions ultimately defined the final INST-^{13}C-MFA ILEs performed and analysed in Chapter 4.4 and focus from this point therefore turned to data processing and analysis.

The qualitative results of chapter 4.4.2-3 suggest that the overall INST-^{13}C-MFA workflow developed over the course of this research was successful in its aim to be the first to perform INST-^{13}C-MFA ILEs of *P. thermoglucosidasius* grown on glucose and xylose under aerobic and anaerobic conditions at a range of growth rates and successfully observe the resulting dynamic rates of ^{13}C labelling over time for its proteinogenic amino acids. Indeed, the results of these experiments provide a series of potentially valuable isotopically in-stationary data sets for future research into *P. thermoglucosidasius*. These results also suggest that the node-by-node comparison approach was valid way to discuss the data and that biosynthesis of amino acids in *P. thermoglucosidasius* grown at different aerobic growth rates and under different oxygen conditions primarily uses glycolysis, with little qualitative suggestions of impact of the ED pathway or considerable flux through the PPP.

The most critical limitation of these INST-^{13}C-MFA ILE results themselves however is that they, combined with their respective CCF analyses, represent a single biological replicate and therefore the reliability of these results as descriptions of metabolic behaviours of *P. thermoglucosidasius* NCIMB 11955 is limited. Indeed, this is most notable for the results of the GA0.075 ILE. A recurring observation in these results was that the total ^{13}C-labelling towards the final samples of the GA0.075 ILE suggest that an isotopic steady state had not been reached. This in turn may affect attempts to model this experimental data set in terms of CCM flux, even when simulations are set to account for non-steady state labelling. On the other hand, the GA0.075 ILE data set suggests that, even if a researcher's intent is to only perform isotopically steady-state ^{13}C-MFA using a chemostat approach, there is value in processing and analysing several time-points towards the assumed end of the ILE to ensure downstream analysis is validly assuming isotopic steady state.

The quantitative results of chapter 4.4.2-3 however suggest, that, although improvements were made to the underlying carbon transition network, futher network variation, constraint and refinement is required to yield flux distribution models which are statistically acceptable representations of these experimental data sets.

One reason for this could be that the ^{13}C-MFA fitting was performed using only the mass fragments of amino acids determined to be valid after natural isotope abundance correction. However, Long and Antoniewicz (2019)[123] suggest that ideally the [M-57]+ and [M-159]+ mass fragments of each amino acid should be used for flux fitting. Future work could therefore investigate whether the use of these fragment MDVs, regardless of their resulting values after natural isotope abundance correction, can aid in achieving SSR accepted flux maps with more accurate flux distributions. Future work would also aim to achieve statistically accepted SSR fits by looking at reactions in the network, in particular those relating to the PPP and key SSR contributor His440, and by exploring whether further addition of pseudo-reactions to the network could potentially improve fitting of the ^{13}C-data.

Time constraints were a considerable factor and were largely a result of underestimation of how long the data processing steps, including TIC extraction, natural isotope abundance correction and data entry to INCA, would take for the technical triplicate samples of each experiment.

5.3 General future work: FBA of a ^{13}C-constrained GSMM

Ultimately however, future work involving this research should work towards the application and integration of statistically acceptable flux distributions of the INST-^{13}C-MFA ILEs with *p-thermo*. Two methods for achieving this have been developed by Martín *et al.* (2015)[219] and Gopalakrishnan and Maranas (2015)[220]. Both methods rely on first performing the ^{13}C data fitting with ILE derived amino acid mass fragments to a model of CCM, then using the output fluxes as constraints for the larger GSMM network. In this way, there is also less of a need to assume an evolutionary optimization principle, such as the maximization of flux through a biomass equation. Such-large scale metabolic analysis would both further improve the predictive accuracy of the model and enable specific characterization of *P. thermoglucosidasius* phenotypes under the aerobic, micro-aerobic and anaerobic conditions as sought in this research.

Such an ideally fluxomics-constrained *p-thermo* would also create a foundational platform for strain design. Indeed, with a more accurate foundational GSMM to act as a reference, techniques for strain design including OptKnock and MOMA could be applied to *p-thermo* to inform strain engineering approaches for *P. thermoglucosidasius* to develop it as a microbial chassis for potential future thermophilic bioprocesses maximize production of existing metabolic products of interest such as: lactic acid, ethanol, and hydrogen. Alternatively, the base GSMM could be modified with new reactions, allowing *in silico* predictions to suggest genetic modifications maximize production of non-native bioproducts such as for the proof of principle isobutanol and terpenoid systems, or to meet broader demands for platform chemicals in a more sustainable manner.

6. References

1. Sheng, L., Kovács, K., Winzer, K., Zhang, Y. and Minton, N. (2017). Development and implementation of rapid metabolic engineering tools for chemical and fuel production in Geobacillus thermoglucosidasius NCIMB 11955. *Biotechnology for Biofuels*, 10(1).
2. Biz, A., Proulx, S., Xu, Z., Siddartha, K., Mulet Indrayanti, A. and Mahadevan, R., 2019. Systems biology based metabolic engineering for non-natural chemicals. *Biotechnology Advances*, [online] 37(6), p.107379. Available at: <https://www.sciencedirect.com/science/article/abs/pii/S0734975019300606>.
3. Nielsen, J. and Keasling, J., 2016. Engineering Cellular Metabolism. *Cell*, [online] 164(6), pp.1185-1197. Available at: <https://www.sciencedirect.com/science/article/pii/S0092867416300708> [Accessed 1 January 2022].
4. Byrne, F., Jin, S., Paggiola, G., Petchey, T., Clark, J., Farmer, T., Hunt, A., Robert McElroy, C. and Sherwood, J., 2016. Tools and techniques for solvent selection: green solvent selection guides. *Sustainable Chemical Processes*, [online] 4(1). Available at: <https://sustainablechemicalprocesses.springeropen.com/articles/10.1186/s40508-016-0051-z#citeas> [Accessed 1 January 2022].
5. Soetaert, W. and Vandamme, E., 2006. The impact of industrial biotechnology. *Biotechnology Journal*, [online] 1(7-8), pp.756-769. Available at: <https://pubmed.ncbi.nlm.nih.gov/16897819/> [Accessed 1 January 2022].
6. Kalyuzhnaya, M., Puri, A. and Lidstrom, M., 2015. Metabolic engineering in methanotrophic bacteria. *Metabolic Engineering*, [online] 29, pp.142-152. Available at: <https://www.sciencedirect.com/science/article/abs/pii/S1096717615000361> [Accessed 1 January 2022].
7. Mohr, T., Aliyu, H., Küchlin, R., Zwick, M., Cowan, D., Neumann, A. and de Maayer, P., 2018. Comparative genomic analysis of Parageobacillus thermoglucosidasius strains with distinct hydrogenogenic capacities. *BMC Genomics*, [online] 19(1). Available at: <https://bmcgenomics.biomedcentral.com/articles/10.1186/s12864-018-5302-9#citeas> [Accessed 1 January 2022].
8. Zhang, H., Wang, Y., Wu, J., Skalina, K. and Pfeifer, B., 2010. Complete Biosynthesis of Erythromycin A and Designed Analogs Using E. coli as a Heterologous Host. *Chemistry & Biology*, [online] 17(11), pp.1232-1240. Available at: <https://pubmed.ncbi.nlm.nih.gov/21095573/> [Accessed 1 January 2022].
9. Tippelt, A. and Nett, M., 2021. Saccharomyces cerevisiae as host for the recombinant production of polyketides and nonribosomal peptides. *Microbial Cell Factories*, [online] 20(1). Available at: <https://microbialcellfactories.biomedcentral.com/articles/10.1186/s12934-021-01650-y#citeas> [Accessed 1 January 2022].
10. Straathof, A. and Bampouli, A., 2017. Potential of commodity chemicals to become bio-based according to maximum yields and petrochemical prices. *Biofuels, Bioproducts and Biorefining*, [online] 11(5), pp.798-810. Available

at: <https://onlinelibrary.wiley.com/doi/10.1002/bbb.1786> [Accessed 1 January 2022].
11. Sun, J. and Alper, H., 2015. Metabolic engineering of strains: from industrial-scale to lab-scale chemical production. *Journal of Industrial Microbiology and Biotechnology*, [online] 42(3), pp.423-436. Available at: <https://academic.oup.com/jimb/article/42/3/423/5995523> [Accessed 1 January 2022].
12. Hollinshead, W., He, L. and Tang, Y., 2014. Biofuel production: an odyssey from metabolic engineering to fermentation scale-up. *Frontiers in Microbiology*, [online] 5. Available at: <https://www.frontiersin.org/articles/10.3389/fmicb.2014.00344/full> [Accessed 1 January 2022].
13. Lau, M., Sheng, L., Zhang, Y. and Minton, N., 2021. Development of a Suite of Tools for Genome Editing in Parageobacillus thermoglucosidasius and Their Use to Identify the Potential of a Native Plasmid in the Generation of Stable Engineered Strains. *ACS Synthetic Biology*, [online] 10(7), pp.1739-1749. Available at: <https://pubs.acs.org/doi/10.1021/acssynbio.1c00138> [Accessed 1 January 2022].
14. Vilaça, P., Maia, P., Giesteira, H., Rocha, I. and Rocha, M., 2017. Analyzing and Designing Cell Factories with OptFlux. *Methods in Molecular Biology*, [online] pp.37-76. Available at: <https://link.springer.com/protocol/10.1007%2F978-1-4939-7528-0_2#citeas> [Accessed 1 January 2022].
15. Mukhopadhyay, A., 2015. Tolerance engineering in bacteria for the production of advanced biofuels and chemicals. *Trends in Microbiology*, [online] 23(8), pp.498-508. Available at: <https://www.sciencedirect.com/science/article/pii/S0966842X1500102X> [Accessed 1 January 2022].
16. Kelwick, R., MacDonald, J., Webb, A. and Freemont, P., 2014. Developments in the Tools and Methodologies of Synthetic Biology. *Frontiers in Bioengineering and Biotechnology*, [online] 2. Available at: <https://www.frontiersin.org/articles/10.3389/fbioe.2014.00060/full> [Accessed 1 January 2022].
17. Gu, Y., Lv, X., Liu, Y., Li, J., Du, G., Chen, J., Rodrigo, L. and Liu, L., 2019. Synthetic redesign of central carbon and redox metabolism for high yield production of N-acetylglucosamine in Bacillus subtilis. *Metabolic Engineering*, [online] 51, pp.59-69. Available at: <https://www.sciencedirect.com/science/article/abs/pii/S1096717618303306> [Accessed 1 January 2022].
18. Myhrvold, C., Polka, J. and Silver, P., 2016. Synthetic Lipid-Containing Scaffolds Enhance Production by Colocalizing Enzymes. *ACS Synthetic Biology*, [online] 5(12), pp.1396-1403. Available at: <https://pubs.acs.org/doi/10.1021/acssynbio.6b00141> [Accessed 1 January 2022].
19. Maki, M., Leung, K. and Qin, W., 2009. The prospects of cellulase-producing bacteria for the bioconversion of lignocellulosic biomass. *International Journal of Biological Sciences*, [online] pp.500-516. Available at: <https://pubmed.ncbi.nlm.nih.gov/19680472/> [Accessed 1 January 2022].
20. Punde, N., Kooken, J., Leary, D., Legler, P. and Angov, E., 2019. Codon harmonization reduces amino acid misincorporation in bacterially expressed P. falciparum proteins and improves their immunogenicity. *AMB Express*,

[online] 9(1). Available at: <https://amb-express.springeropen.com/articles/10.1186/s13568-019-0890-6#citeas> [Accessed 1 January 2022].
21. Nature. 2014. *Definition: Promoter*. [online] Available at: <https://www.nature.com/scitable/definition/promoter-259/> [Accessed 1 January 2022].
22. Eriksen, M., Sneppen, K., Pedersen, S. and Mitarai, N., 2017. Occlusion of the Ribosome Binding Site Connects the Translational Initiation Frequency, mRNA Stability and Premature Transcription Termination. *Frontiers in Microbiology*, [online] 8. Available at: <https://www.frontiersin.org/articles/10.3389/fmicb.2017.00362/full> [Accessed 1 January 2022].
23. Joseph, B., Pichaimuthu, S. and Srimeenakshi, S., 2015. An Overview of the Parameters for Recombinant Protein Expression in Escherichia coli. *Journal of Cell Science & Therapy*, [online] 06(05). Available at: <https://www.longdom.org/open-access/an-overview-of-the-parameters-for-recombinant-protein-expression-in-escherichia-coli-2157-7013-1000221.pdf> [Accessed 1 January 2022].
24. Peters, J., Vangeloff, A. and Landick, R., 2011. Bacterial Transcription Terminators: The RNA 3′-End Chronicles. *Journal of Molecular Biology*, [online] 412(5), pp.793-813. Available at: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3622210/> [Accessed 1 January 2022].
25. Dou, W., Zhu, Q., Zhang, M., Jia, Z. and Guan, W., 2021. Screening and evaluation of the strong endogenous promoters in Pichia pastoris. *Microbial Cell Factories*, [online] 20(1). Available at: <https://microbialcellfactories.biomedcentral.com/articles/10.1186/s12934-021-01648-6#citeas> [Accessed 1 January 2022].
26. Maarleveld, T., Khandelwal, R., Olivier, B., Teusink, B. and Bruggeman, F., 2013. Basic concepts and principles of stoichiometric modeling of metabolic networks. *Biotechnology Journal*, [online] 8(9), pp.997-1008. Available at: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4671265/> [Accessed 1 January 2022].
27. Weber, J., Li, Z. and Rinas, U., 2021. Recombinant protein production provoked accumulation of ATP, fructose-1,6-bisphosphate and pyruvate in E. coli K12 strain TG1. *Microbial Cell Factories*, [online] 20(1). Available at: <https://microbialcellfactories.biomedcentral.com/articles/10.1186/s12934-021-01661-9#citeas> [Accessed 1 January 2022].
28. Liu, Y., Li, J., Du, G., Chen, J. and Liu, L., 2017. Metabolic engineering of Bacillus subtilis fueled by systems biology: Recent advances and future directions. *Biotechnology Advances*, [online] 35(1), pp.20-30. Available at: <https://www.sciencedirect.com/science/article/abs/pii/S0734975016301458> [Accessed 1 January 2022].
29. Ma, W., Liu, Y., Shin, H., Li, J., Chen, J., Du, G. and Liu, L., 2018. Metabolic engineering of carbon overflow metabolism of Bacillus subtilis for improved N-acetyl-glucosamine production. *Bioresource Technology*, [online] 250, pp.642-649. Available at: <https://www.sciencedirect.com/science/article/abs/pii/S0960852417318242> [Accessed 1 January 2022].
30. Dugar, D. and Stephanopoulos, G., 2011. Relative potential of biosynthetic pathways for biofuels and bio-based products. *Nature Biotechnology*,

[online] 29(12), pp.1074-1078. Available at: <https://pubmed.ncbi.nlm.nih.gov/22158355/> [Accessed 1 January 2022].
31. Andreozzi, S., Chakrabarti, A., Soh, K., Burgard, A., Yang, T., Van Dien, S., Miskovic, L. and Hatzimanikatis, V., 2016. Identification of metabolic engineering targets for the enhancement of 1,4-butanediol production in recombinant E. coli using large-scale kinetic models. *Metabolic Engineering*, [online] 35, pp.148-159. Available at: <https://pubmed.ncbi.nlm.nih.gov/26855240/> [Accessed 1 January 2022].
32. Vilaça, P., Maia, P., Giesteira, H., Rocha, I. and Rocha, M., 2017. Analyzing and Designing Cell Factories with OptFlux. *Methods in Molecular Biology*, [online] pp.37-76. Available at: <https://link.springer.com/protocol/10.1007%2F978-1-4939-7528-0_2#citeas> [Accessed 1 January 2022].
33. Lee, J., Na, D., Park, J., Lee, J., Choi, S. and Lee, S., 2012. Systems metabolic engineering of microorganisms for natural and non-natural chemicals. *Nature Chemical Biology*, [online] 8(6), pp.536-546. Available at: <https://www.nature.com/articles/nchembio.970#citeas> [Accessed 1 January 2022].
34. Yang, D., Cho, J., Choi, K., Kim, H. and Lee, S., 2017. Systems metabolic engineering as an enabling technology in accomplishing sustainable development goals. *Microbial Biotechnology*, [online] 10(5), pp.1254-1258. Available at: <https://pubmed.ncbi.nlm.nih.gov/28696000/> [Accessed 1 January 2022].
35. Gianchandani, E., Chavali, A. and Papin, J., 2010. The application of flux balance analysis in systems biology. *Wiley Interdisciplinary Reviews: Systems Biology and Medicine*, [online] 2(3), pp.372-382. Available at: <https://wires.onlinelibrary.wiley.com/doi/abs/10.1002/wsbm.60> [Accessed 1 January 2022].
36. Kelley, J., Maor, S., Kim, M., Lane, A. and Lun, D., 2017. MOST-visualization: software for producing automated textbook-style maps of genome-scale metabolic networks. *Bioinformatics*, [online] 33(16), pp.2596-2597. Available at: <https://academic.oup.com/bioinformatics/article/33/16/2596/3738797?login=true> [Accessed 1 January 2022].
37. Orth, J., Thiele, I. and Palsson, B., 2010. What is flux balance analysis?. *Nature Biotechnology*, [online] 28(3), pp.245-248. Available at: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3108565/> [Accessed 1 January 2022].
38. Dewey, J. and Bentley, A., 1949. *Knowing and the known*. Westport, Conn.: Greenwood Pr., p.VI.
39. Orth, J., Thiele, I. and Palsson, B., 2010. What is flux balance analysis?. *Nature Biotechnology*, [online] 28(3), pp.245-248. Available at: <https://www.nature.com/articles/nbt.1614#citeas> [Accessed 1 January 2022].
40. O'Brien, E., Monk, J. and Palsson, B., 2015. Using Genome-scale Models to Predict Biological Capabilities. *Cell*, [online] 161(5), pp.971-987. Available at: <https://www.sciencedirect.com/science/article/pii/S0092867415005681> [Accessed 1 January 2022].
41. Gianchandani, E., Chavali, A. and Papin, J., 2010. The application of flux balance analysis in systems biology. *Wiley Interdisciplinary Reviews:*

Systems Biology and Medicine, [online] 2(3), pp.372-382. Available at: <https://pubmed.ncbi.nlm.nih.gov/20836035/> [Accessed 1 January 2022].
42. Mohr, T., Aliyu, H., Küchlin, R., Polliack, S., Zwick, M., Neumann, A., Cowan, D. and de Maayer, P., 2018. CO-dependent hydrogen production by the facultative anaerobe Parageobacillus thermoglucosidasius. *Microbial Cell Factories*, [online] 17(1). Available at: <https://microbialcellfactories.biomedcentral.com/articles/10.1186/s12934-018-0954-3#citeas> [Accessed 1 January 2022].
43. Long, C. and Antoniewicz, M., 2019. High-resolution 13C metabolic flux analysis. *Nature Protocols*, [online] 14(10), pp.2856-2877. Available at: <https://www.nature.com/articles/s41596-019-0204-0#citeas> [Accessed 1 January 2022].
44. Gonzalez, J., Long, C. and Antoniewicz, M., 2017. Comprehensive analysis of glucose and xylose metabolism in Escherichia coli under aerobic and anaerobic conditions by 13C metabolic flux analysis. *Metabolic Engineering*, [online] 39, pp.9-18. Available at: <https://pubmed.ncbi.nlm.nih.gov/27840237/> [Accessed 1 January 2022].
45. Orth, J., Thiele, I. and Palsson, B., 2010. What is flux balance analysis?. *Nature Biotechnology*, [online] 28(3), pp.245-248. Available at: <https://www.nature.com/articles/nbt.1614#citeas> [Accessed 1 January 2022].
46. Ko, Y., Kim, J., Lee, J., Han, T., Kim, G., Park, J. and Lee, S., 2020. Tools and strategies of systems metabolic engineering for the development of microbial cell factories for chemical production. *Chemical Society Reviews*, [online] 49(14), pp.4615-4636. Available at: <https://pubs.rsc.org/en/content/articlelanding/2020/CS/D0CS00155D> [Accessed 1 January 2022].
47. Choi, K., Jang, W., Yang, D., Cho, J., Park, D. and Lee, S., 2019. Systems Metabolic Engineering Strategies: Integrating Systems and Synthetic Biology with Metabolic Engineering. *Trends in Biotechnology*, [online] 37(8), pp.817-837. Available at: <https://www.sciencedirect.com/science/article/abs/pii/S0167779919300034> [Accessed 1 January 2022].
48. Gahlaut, A., Vikas, Dahiya, M., Gothwal, A., Kulharia, M., Chhillar, A., Hooda, V. and Dabur, R., 2013. Proteomics & metabolomics: Mapping biochemical regulations. *Drug Invention Today*, [online] 5(4), pp.321-326. Available at: <https://www.sciencedirect.com/science/article/pii/S0975761913000744> [Accessed 1 January 2022].
49. Cortassa, S., Caceres, V., Bell, L., O'Rourke, B., Paolocci, N. and Aon, M., 2015. From Metabolomics to Fluxomics: A Computational Procedure to Translate Metabolite Profiles into Metabolic Fluxes. *Biophysical Journal*, [online] 108(1), pp.163-172. Available at: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4286601/> [Accessed 1 January 2022].
50. *Frontiers Research Topics*, 2019. Fluxomics and Metabolic Analysis in Systems Microbiology. [online] Available at: <https://www.frontiersin.org/research-topics/7614/fluxomics-and-metabolic-flux-analysis-in-systems-microbiology> [Accessed 1 January 2022].
51. Noack, S., Nöh, K., Moch, M., Oldiges, M. and Wiechert, W., 2011. Stationary versus non-stationary 13C-MFA: A comparison using a consistent

dataset. *Journal of Biotechnology*, [online] 154(2-3), pp.179-190. Available at: <https://pubmed.ncbi.nlm.nih.gov/20638432/> [Accessed 1 January 2022].
52. Burgard, A., Burk, M., Osterhout, R., Van Dien, S. and Yim, H., 2016. Development of a commercial scale process for production of 1,4-butanediol from sugar. *Current Opinion in Biotechnology*, [online] 42, pp.118-125. Available at: <https://pubmed.ncbi.nlm.nih.gov/27132123/> [Accessed 1 January 2022].
53. Yim, H., Haselbeck, R., Niu, W., Pujol-Baxley, C., Burgard, A., Boldt, J., Khandurina, J., Trawick, J., Osterhout, R., Stephen, R., Estadilla, J., Teisan, S., Schreyer, H., Andrae, S., Yang, T., Lee, S., Burk, M. and Van Dien, S., 2011. Metabolic engineering of Escherichia coli for direct production of 1,4-butanediol. *Nature Chemical Biology*, [online] 7(7), pp.445-452. Available at: <https://www.nature.com/articles/nchembio.580#citeas> [Accessed 1 January 2022].
54. Aliyu, H., Lebre, P., Blom, J., Cowan, D. and De Maayer, P., 2016. Phylogenomic re-assessment of the thermophilic genus Geobacillus. *Systematic and Applied Microbiology*, [online] 39(8), pp.527-533. Available at: <https://www.sciencedirect.com/science/article/abs/pii/S0723202016300959?via%3Dihub> [Accessed 1 January 2022].
55. Hussein, A., Lisowska, B. and Leak, D., 2015. The Genus Geobacillus and Their Biotechnological Potential. *Advances in Applied Microbiology*, [online] pp.1-48. Available at: <https://www.sciencedirect.com/science/article/abs/pii/S0065216415000064> [Accessed 1 January 2022].
56. Suzuki, Y., Kishigami, T., Inoue, K., Mizoguchi, Y., Eto, N., Takagi, M. and Abe, S., 1983. Bacillus thermoglucosidasius sp. nov., a New Species of Obligately Thermophilic Bacilli. *Systematic and Applied Microbiology*, [online] 4(4), pp.487-495. Available at: <https://pubmed.ncbi.nlm.nih.gov/23194806/> [Accessed 1 January 2022].
57. Nazina, T., Tourova, T., Poltaraus, A., Novikova, E., Grigoryan, A., Ivanova, A., Lysenko, A., Petrunyaka, V., Osipov, G., Belyaev, S. and Ivanov, M., 2001. Taxonomic study of aerobic thermophilic bacilli: descriptions of Geobacillus subterraneus gen. nov., sp. nov. and Geobacillus uzenensis sp. nov. from petroleum reservoirs and transfer of Bacillus stearothermophilus, Bacillus thermocatenulatus, Bacillus thermoleovorans, Bacillus kaustophilus, Bacillus thermodenitrificans to Geobacillus as the new combinations G. stearothermophilus, G. th. *International Journal of Systematic and Evolutionary Microbiology*, 51(2), pp.433-446.
58. Najar, I. and Thakur, N., 2020. A systematic review of the genera Geobacillus and Parageobacillus: their evolution, current taxonomic status and major applications. *Microbiology*, [online] 166(9), pp.800-816. Available at: <https://pubmed.ncbi.nlm.nih.gov/32744496/> [Accessed 1 January 2022].
59. Zeigler, D., 2014. The Geobacillus paradox: why is a thermophilic bacterial genus so prevalent on a mesophilic planet?. *Microbiology*, [online] 160(1), pp.1-11. Available at: <https://www.microbiologyresearch.org/content/journal/micro/10.1099/mic.0.071696-0> [Accessed 1 January 2022].

60. Studholme, D., 2014. Some (bacilli) like it hot: genomics ofGeobacillusspecies. *Microbial Biotechnology*, [online] 8(1), pp.40-48. Available at: <https://ore.exeter.ac.uk/repository/handle/10871/16493> [Accessed 1 January 2022].
61. Takami, H., Nishi, S., Lu, J., Shimamura, S. and Takaki, Y., 2004. Genomic characterization of thermophilic Geobacillus species isolated from the deepest sea mud of the Mariana Trench. *Extremophiles*, [online] 8(5), pp.351-356. Available at: <https://pubmed.ncbi.nlm.nih.gov/15168170/> [Accessed 1 January 2022].
62. Inoue, M., Tanimura, A., Ogami, Y., Hino, T., Okunishi, S., Maeda, H., Yoshida, T. and Sako, Y., 2019. Draft Genome Sequence of Parageobacillus thermoglucosidasius Strain TG4, a Hydrogenogenic Carboxydotrophic Bacterium Isolated from a Marine Sediment. *Microbiology Resource Announcements*, [online] 8(5). Available at: <https://pubmed.ncbi.nlm.nih.gov/30714041/> [Accessed 1 January 2022].
63. Brumm, P., Land, M. and Mead, D., 2015. Complete genome sequence of Geobacillus thermoglucosidasius C56-YS93, a novel biomass degrader isolated from obsidian hot spring in Yellowstone National Park. *Standards in Genomic Sciences*, [online] 10(1). Available at: <https://pubmed.ncbi.nlm.nih.gov/26442136/> [Accessed 1 January 2022].
64. Sung, M., Kim, H., Bae, J., Rhee, S., Jeon, C., Kim, K., Kim, J., Hong, S., Lee, S., Yoon, J., Park, Y. and Baek, D., 2002. Geobacillus toebii sp. nov., a novel thermophilic bacterium isolated from hay compost. *International Journal of Systematic and Evolutionary Microbiology*, [online] 52(6), pp.2251-2255. Available at: <https://pubmed.ncbi.nlm.nih.gov/12508894/> [Accessed 1 January 2022].
65. Zhao, Y., Caspers, M., Abee, T., Siezen, R. and Kort, R., 2012. Complete Genome Sequence of Geobacillus thermoglucosidans TNO-09.020, a Thermophilic Sporeformer Associated with a Dairy-Processing Environment. *Journal of Bacteriology*, [online] 194(15), pp.4118-4118. Available at: <https://pubmed.ncbi.nlm.nih.gov/22815439/> [Accessed 1 January 2022].
66. Feng L., Wang W., Cheng J., Ren Y., Zhao G., Gao C., Tang Y., Liu- X., Han W., Peng X., Liu R., Wang L., 2007. Genome and proteome of long-chain alkane degrading Geobacillus thermodenitrificans NG80-2 isolated from a deep-subsurface oil reservoir. *Proceedings of the National Academy of Sciences of the United States of America*. 2007 Mar; 5602-5607. [online] 104(13), pp.5602-5607.
67. Suzuki, H., Taketani, T., Tanabiki, M., Ohara, M., Kobayashi, J. and Ohshiro, T., 2021. Frequent Transposition of Multiple Insertion Sequences in Geobacillus kaustophilus HTA426. *Frontiers in Microbiology*, [online] 12. Available at: <https://pubmed.ncbi.nlm.nih.gov/33841375/> [Accessed 1 January 2022].
68. Suzuki, H., 2018. Peculiarities and biotechnological potential of environmental adaptation by Geobacillus species. *Applied Microbiology and Biotechnology*, [online] 102(24), pp.10425-10437. Available at: <https://pubmed.ncbi.nlm.nih.gov/30310966/> [Accessed 1 January 2022].
69. Di Donato, P., Romano, I., Mastascusa, V., Poli, A., Orlando, P., Pugliese, M. and Nicolaus, B., 2018. Survival and Adaptation of the Thermophilic Species Geobacillus thermantarcticus in Simulated Spatial Conditions. *Origins of Life and Evolution of Biospheres*, [online] 48(1), pp.141-158.

Available at: <https://link.springer.com/article/10.1007/s11084-017-9540-7#citeas> [Accessed 1 January 2022].
70. Pontrelli, S., Chiu, T., Lan, E., Chen, F., Chang, P. and Liao, J., 2018. Escherichia coli as a host for metabolic engineering. *Metabolic Engineering*, [online] 50, pp.16-46. Available at: <https://www.sciencedirect.com/science/article/abs/pii/S1096717618300740?via%3Dihub> [Accessed 1 January 2022].
71. Hussein, A., Lisowska, B. and Leak, D., 2015. The Genus Geobacillus and Their Biotechnological Potential. *Advances in Applied Microbiology*, [online] pp.1-48. Available at: <https://www.sciencedirect.com/science/article/abs/pii/S0065216415000064?via%3Dihub> [Accessed 1 January 2022].
72. Wada, K. and Suzuki, H., 2020. Biotechnological platforms of the moderate thermophiles, Geobacillus species: notable properties and genetic tools. *Physiological and Biotechnological Aspects of Extremophiles*, [online] pp.195-218. Available at: <https://www.sciencedirect.com/science/article/pii/B9780128183229000150?via%3Dihub> [Accessed 1 January 2022].
73. Najar, I. and Thakur, N., 2020. A systematic review of the genera Geobacillus and Parageobacillus: their evolution, current taxonomic status and major applications. *Microbiology*, [online] 166(9), pp.800-816. Available at: <https://pubmed.ncbi.nlm.nih.gov/32744496/> [Accessed 1 January 2022].
74. Takami, H., Takaki, Y., Chee, G., Nishi, S., Shimamura, S., Suzuki, H., Matsui, S. and Uchiyama, I., 2004. Thermoadaptation trait revealed by the genome sequence of thermophilic Geobacillus kaustophilus. *Nucleic Acids Research*, [online] 32(21), pp.6292-6303. Available at: <https://academic.oup.com/nar/article/32/21/6292/1101682> [Accessed 1 January 2022].
75. Krüger, A., Schäfers, C., Schröder, C. and Antranikian, G., 2018. Towards a sustainable biobased industry – Highlighting the impact of extremophiles. *New Biotechnology*, [online] 40, pp.144-153. Available at: <https://pubmed.ncbi.nlm.nih.gov/28512003/> [Accessed 1 January 2022].
76. Lin, B. and Tao, Y., 2017. Whole-cell biocatalysts by design. *Microbial Cell Factories*, [online] 16(1). Available at: <https://microbialcellfactories.biomedcentral.com/articles/10.1186/s12934-017-0724-7#citeas> [Accessed 1 January 2022].
77. Peralta-Yahya, P., Zhang, F., del Cardayre, S. and Keasling, J., 2012. Microbial engineering for the production of advanced biofuels. *Nature*, [online] 488(7411), pp.320-328. Available at: <https://pubmed.ncbi.nlm.nih.gov/22895337/> [Accessed 1 January 2022].
78. Gonzalez, J., Long, C. and Antoniewicz, M., 2017. Comprehensive analysis of glucose and xylose metabolism in Escherichia coli under aerobic and anaerobic conditions by 13C metabolic flux analysis. *Metabolic Engineering*, [online] 39, pp.9-18. Available at: <https://pubmed.ncbi.nlm.nih.gov/27840237/> [Accessed 1 January 2022].
79. Rastogi, G., Bhalla, A., Adhikari, A., Bischoff, K., Hughes, S., Christopher, L. and Sani, R., 2010. Characterization of thermostable cellulases produced by Bacillus and Geobacillus strains. *Bioresource Technology*, [online] 101(22), pp.8798-8806. Available at: <https://pubmed.ncbi.nlm.nih.gov/20599378/> [Accessed 1 January 2022].

80. Brumm, P., De Maayer, P., Mead, D. and Cowan, D., 2015. Genomic analysis of six new Geobacillus strains reveals highly conserved carbohydrate degradation architectures and strategies. *Frontiers in Microbiology*, [online] 6. Available at: <https://www.frontiersin.org/articles/10.3389/fmicb.2015.00430/full> [Accessed 1 January 2022].
81. Demir, N., Nadaroglu, H., Tasgin, E., Adiguzel, A. and Gulluce, M., 2011. Purification and characterization of a pectin lyase produced by Geobacillus stearothermophilus Ah22 and its application in fruit juice production. *Annals of Microbiology*, [online] 61(4), pp.939-946. Available at: <https://annalsmicrobiology.biomedcentral.com/articles/10.1007/s13213-011-0217-6#citeas> [Accessed 1 January 2022].
82. De Maayer, P., Brumm, P., Mead, D. and Cowan, D., 2014. Comparative analysis of the Geobacillus hemicellulose utilization locus reveals a highly variable target for improved hemicellulolysis. *BMC Genomics*, [online] 15(1), p.836. Available at: <https://bmcgenomics.biomedcentral.com/articles/10.1186/1471-2164-15-836#citeas> [Accessed 1 January 2022].
83. Hussein, A., Lisowska, B. and Leak, D., 2015. The Genus Geobacillus and Their Biotechnological Potential. *Advances in Applied Microbiology*, [online] pp.1-48. Available at: <https://www.sciencedirect.com/science/article/abs/pii/S0065216415000064> [Accessed 1 January 2022].
84. Olofsson, K., Bertilsson, M. and Lidén, G., 2008. A short review on SSF – an interesting process option for ethanol production from lignocellulosic feedstocks. *Biotechnology for Biofuels*, [online] 1(1). Available at: <https://biotechnologyforbiofuels.biomedcentral.com/articles/10.1186/1754-6834-1-7#citeas> [Accessed 1 January 2022].
85. Olson, D., McBride, J., Joe Shaw, A. and Lynd, L., 2012. Recent progress in consolidated bioprocessing. *Current Opinion in Biotechnology*, [online] 23(3), pp.396-405. Available at: <https://pubmed.ncbi.nlm.nih.gov/22176748/> [Accessed 1 January 2022].
86. Lau, M., Sheng, L., Zhang, Y. and Minton, N., 2021. Development of a Suite of Tools for Genome Editing in Parageobacillus thermoglucosidasius and Their Use to Identify the Potential of a Native Plasmid in the Generation of Stable Engineered Strains. *ACS Synthetic Biology*, [online] 10(7), pp.1739-1749.
87. Nijland, R., Burgess, J., Errington, J. and Veening, J., 2010. Transformation of Environmental Bacillus subtilis Isolates by Transiently Inducing Genetic Competence. *PLoS ONE*, 5(3), p.e9724.
88. Taylor, M., Esteban, C. and Leak, D., 2008. Development of a versatile shuttle vector for gene expression in Geobacillus spp. *Plasmid*, [online] 60(1), pp.45-52.
89. Cripps, R., Eley, K., Leak, D., Rudd, B., Taylor, M., Todd, M., Boakes, S., Martin, S. and Atkinson, T., 2009. Metabolic engineering of Geobacillus thermoglucosidasius for high yield ethanol production. *Metabolic Engineering*, [online] 11(6), pp.398-408. Available at: <https://www.sciencedirect.com/science/article/abs/pii/S1096717609000688> [Accessed 1 January 2022].
90. Suzuki, H., 2012. Genetic Transformation of Geobacillus kaustophilus HTA426 by Conjugative Transfer of Host-Mimicking Plasmids. *Journal of*

Microbiology and Biotechnology, [online] 22(9), pp.1279-1287. Available at: <https://pubmed.ncbi.nlm.nih.gov/22814504/> [Accessed 1 January 2022].
91. Kananavičiūtė, R. and Čitavičius, D., 2015. Genetic engineering of Geobacillus spp. *Journal of Microbiological Methods*, [online] 111, pp.31-39. Available at: <https://www.sciencedirect.com/science/article/abs/pii/S0167701215000536?via%3Dihub> [Accessed 1 January 2022].
92. Bacon, L., Hamley-Bennett, C., Danson, M. and Leak, D., 2017. Development of an efficient technique for gene deletion and allelic exchange in Geobacillus spp. *Microbial Cell Factories*, [online] 16(1). Available at: <https://microbialcellfactories.biomedcentral.com/articles/10.1186/s12934-017-0670-4#citeas> [Accessed 1 January 2022].
93. Lau, M., Sheng, L., Zhang, Y. and Minton, N., 2021. Development of a Suite of Tools for Genome Editing in Parageobacillus thermoglucosidasius and Their Use to Identify the Potential of a Native Plasmid in the Generation of Stable Engineered Strains. *ACS Synthetic Biology*, [online] 10(7), pp.1739-1749. Available at: <https://pubs.acs.org/doi/10.1021/acssynbio.1c00138> [Accessed 1 January 2022].
94. Reeve, B., Martinez-Klimova, E., de Jonghe, J., Leak, D. and Ellis, T., 2016. The Geobacillus Plasmid Set: A Modular Toolkit for Thermophile Engineering. *ACS Synthetic Biology*, [online] 5(12), pp.1342-1347. Available at: <https://pubs.acs.org/doi/full/10.1021/acssynbio.5b00298> [Accessed 1 January 2022].
95. Pogrebnyakov, I., Jendresen, C. and Nielsen, A., 2017. Genetic toolbox for controlled expression of functional proteins in Geobacillus spp. *PLOS ONE*, [online] 12(2), p.e0171313. Available at: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5289569/> [Accessed 1 January 2022].
96. Salis, H., Mirsky, E. and Voigt, C., 2009. Automated design of synthetic ribosome binding sites to control protein expression. *Nature Biotechnology*, [online] 27(10), pp.946-950. Available at: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2782888/> [Accessed 1 January 2022].
97. Lin, P., Rabe, K., Takasumi, J., Kadisch, M., Arnold, F. and Liao, J., 2014. Isobutanol production at elevated temperatures in thermophilic Geobacillus thermoglucosidasius. *Metabolic Engineering*, [online] 24, pp.1-8. Available at: <https://pubmed.ncbi.nlm.nih.gov/24721011/> [Accessed 1 January 2022].
98. Mohr, T., Aliyu, H., Biebinger, L., Gödert, R., Hornberger, A., Cowan, D., de Maayer, P. and Neumann, A., 2019. Effects of different operating parameters on hydrogen production by Parageobacillus thermoglucosidasius DSM 6285. *AMB Express*, [online] 9(1). Available at: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6928187/> [Accessed 1 January 2022].
99. Chou, K., Xiong, W., Magnusson, L., Seibert, M. and Maness, P., 2021. Electrons and Protons | Renewable Hydrogen From Biomass Fermentation. *Encyclopedia of Biological Chemistry III*, [online] pp.551-559. Available at: <https://www.sciencedirect.com/science/article/pii/B9780128194607000177> [Accessed 1 January 2022].
100. Yang, Z., Sun, Q., Tan, G., Zhang, Q., Wang, Z., Li, C., Qi, F., Wang, W., Zhang, L. and Li, Z., 2020. Engineering thermophilic Geobacillus

thermoglucosidasius for riboflavin production. *Microbial Biotechnology*, [online] 14(2), pp.363-373. Available at: <https://sfamjournals.onlinelibrary.wiley.com/doi/10.1111/1751-7915.13543> [Accessed 1 January 2022].

101. Zhou, J., Lian, J. and Rao, C., 2020. Metabolic engineering of Parageobacillus thermoglucosidasius for the efficient production of (2R, 3R)-butanediol. *Applied Microbiology and Biotechnology*, [online] 104(10), pp.4303-4311. Available at: <https://pubmed.ncbi.nlm.nih.gov/32221689/> [Accessed 1 January 2022].
102. Styles, M., Nesbitt, E., Hoffmann, T., Queen, J., Ortenzi, M. and Leak, D., 2021. The heterologous production of terpenes by the thermophile Parageobacillus thermoglucosidasius in a consolidated bioprocess using waste bread. *Metabolic Engineering*, [online] 65, pp.146-155. Available at: <https://www.sciencedirect.com/science/article/abs/pii/S1096717620301750?via%3Dihub> [Accessed 1 January 2022].
103. Zeldes, B., Keller, M., Loder, A., Straub, C., Adams, M. and Kelly, R., 2015. Extremely thermophilic microorganisms as metabolic engineering platforms for production of fuels and industrial chemicals. *Frontiers in Microbiology*, [online] 6. Available at: <https://www.frontiersin.org/articles/10.3389/fmicb.2015.01209/full> [Accessed 1 January 2022].
104. Jiang, T., Cai, M., Huang, M., He, H., Lu, J., Zhou, X. and Zhang, Y., 2015. Characterization of a thermostable raw-starch hydrolyzing α-amylase from deep-sea thermophile Geobacillus sp. *Protein Expression and Purification*, [online] 114, pp.15-22. Available at: <https://pubmed.ncbi.nlm.nih.gov/26073094/> [Accessed 1 January 2022].
105. Zhu, Y., Li, H., Ni, H., Xiao, A., Li, L. and Cai, H., 2014. Molecular cloning and characterization of a thermostable lipase from deep-sea thermophile Geobacillus sp. EPT9. *World Journal of Microbiology and Biotechnology*, [online] 31(2), pp.295-306. Available at: <https://pubmed.ncbi.nlm.nih.gov/25388475/> [Accessed 1 January 2022].
106. Chen, P., Liu, C., Chen, Y., Hsu, F. and Shaw, J., 2020. Isolation, Expression and Characterization of the Thermophilic Recombinant Esterase from Geobacillus thermodenitrificans PS01. *Applied Biochemistry and Biotechnology*, [online] 191(1), pp.112-124. Available at: <https://pubmed.ncbi.nlm.nih.gov/31956956/> [Accessed 1 January 2022].
107. Quintana-Castro, R., Díaz, P., Valerio-Alfaro, G., García, H. and Oliart-Ros, R., 2008. Gene Cloning, Expression, and Characterization of the Geobacillus Thermoleovorans CCR11 Thermoalkaliphilic Lipase. *Molecular Biotechnology*, 42(1), pp.75-83.
108. Ebrahimpour, A., Rahman, R., Basri, M. and Salleh, A., 2011. High level expression and characterization of a novel thermostable, organic solvent tolerant, 1,3-regioselective lipase from Geobacillus sp. strain ARM. *Bioresource Technology*, [online] 102(13), pp.6972-6981. Available at: <https://pubmed.ncbi.nlm.nih.gov/21531550/> [Accessed 1 January 2022].
109. Abol Fotouh, D., Bayoumi, R. and Hassan, M., 2016. Production of Thermoalkaliphilic Lipase from Geobacillus thermoleovorans DA2 and Application in Leather Industry. *Enzyme Research*, [online] 2016, pp.1-9. Available at: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4735910/> [Accessed 1 January 2022].

110. Iqbal, I., Aftab, M., Afzal, M., Ur-Rehman, A., Aftab, S., Zafar, A., Ud-Din, Z., Khuharo, A., Iqbal, J. and Ul-Haq, I., 2014. Purification and characterization of cloned alkaline protease gene ofGeobacillus stearothermophilus. *Journal of Basic Microbiology*, [online] 55(2), pp.160-171. Available at: <https://pubmed.ncbi.nlm.nih.gov/25224381/> [Accessed 1 January 2022].
111. Chang, C., Gong, S., Liu, Z., Yan, Q. and Jiang, Z., 2021. High level expression and biochemical characterization of an alkaline serine protease from Geobacillus stearothermophilus to prepare antihypertensive whey protein hydrolysate. *BMC Biotechnology*, [online] 21(1). Available at: <https://bmcbiotechnol.biomedcentral.com/articles/10.1186/s12896-021-00678-7#citeas> [Accessed 1 January 2022].
112. Ma, Y., Zhang, B., Wang, M., Ou, Y., Wang, J. and Li, S., 2016. Enhancement of Polymerase Activity of the Large Fragment in DNA Polymerase I from Geobacillus stearothermophilus by Site-Directed Mutagenesis at the Active Site. *BioMed Research International*, [online] 2016, pp.1-8. Available at: <https://www.hindawi.com/journals/bmri/2016/2906484/> [Accessed 1 January 2022].
113. Genomic analysis and metabolic modelling of *Geobacillus thermoglucosidasius*NCIMB 11955, Lisowska, B. (Author). 29 Jun 2016
114. Loftie-Eaton, W., Taylor, M., Horne, K., Tuffin, M., Burton, S. and Cowan, D., 2012. Balancing redox cofactor generation and ATP synthesis: Key microaerobic responses in thermophilic fermentations. *Biotechnology and Bioengineering*, [online] 110(4), pp.1057-1065. Available at: <https://onlinelibrary.wiley.com/doi/abs/10.1002/bit.24774> [Accessed 1 January 2022].
115. Metabolic engineering and metabolic flux analysis of thermophilic, ethanogenic *Geobacillus* spp. Bartosiak-Jentys, J. (Author). 29 Jun 2011
116. Midani, F., Wynn, M. and Schnell, S., 2017. The importance of accurately correcting for the natural abundance of stable isotopes. *Analytical Biochemistry*, [online] 520, pp.27-43. Available at: <https://www.sciencedirect.com/science/article/abs/pii/S0003269716304250?via%3Dihub> [Accessed 1 January 2022].
117. Millard, P., Delépine, B., Guionnet, M., Heuillet, M., Bellvert, F. and Létisse, F., 2019. IsoCor: isotope correction for high-resolution MS labeling experiments. *Bioinformatics*, [online] 35(21), pp.4484-4487. Available at: <https://academic.oup.com/bioinformatics/article/35/21/4484/5418798> [Accessed 1 January 2022].
118. Young, J., 2014. INCA: a computational platform for isotopically non-stationary metabolic flux analysis. *Bioinformatics*, [online] 30(9), pp.1333-1335. Available at: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3998137/> [Accessed 1 January 2022].
119. Heirendt, L., Arreckx, S., Pfau, T., Mendoza, S., Richelle, A., Heinken, A., Haraldsdóttir, H., Wachowiak, J., Keating, S., Vlasov, V., Magnusdóttir, S., Ng, C., Preciat, G., Žagare, A., Chan, S., Aurich, M., Clancy, C., Modamio, J., Sauls, J., Noronha, A., Bordbar, A., Cousins, B., El Assal, D., Valcarcel, L., Apaolaza, I., Ghaderi, S., Ahookhosh, M., Ben Guebila, M., Kostromins, A., Sompairac, N., Le, H., Ma, D., Sun, Y., Wang, L., Yurkovich, J., Oliveira, M., Vuong, P., El Assal, L., Kuperstein, I., Zinovyev, A., Hinton, H., Bryant, W., Aragón Artacho, F., Planes, F., Stalidzans, E., Maass, A., Vempala, S., Hucka, M., Saunders, M., Maranas,

C., Lewis, N., Sauter, T., Palsson, B., Thiele, I. and Fleming, R., 2019. Creation and analysis of biochemical constraint-based models using the COBRA Toolbox v.3.0. *Nature Protocols*, [online] 14(3), pp.639-702. Available at: <https://www.nature.com/articles/s41596-018-0098-2#citeas> [Accessed 1 January 2022].

120. Lieven, C., Beber, M., Olivier, B., Bergmann, F., Ataman, M., Babaei, P., Bartell, J., Blank, L., Chauhan, S., Correia, K., Diener, C., Dräger, A., Ebert, B., Edirisinghe, J., Faria, J., Feist, A., Fengos, G., Fleming, R., García-Jiménez, B., Hatzimanikatis, V., van Helvoirt, W., Henry, C., Hermjakob, H., Herrgård, M., Kaafarani, A., Kim, H., King, Z., Klamt, S., Klipp, E., Koehorst, J., König, M., Lakshmanan, M., Lee, D., Lee, S., Lee, S., Lewis, N., Liu, F., Ma, H., Machado, D., Mahadevan, R., Maia, P., Mardinoglu, A., Medlock, G., Monk, J., Nielsen, J., Nielsen, L., Nogales, J., Nookaew, I., Palsson, B., Papin, J., Patil, K., Poolman, M., Price, N., Resendis-Antonio, O., Richelle, A., Rocha, I., Sánchez, B., Schaap, P., Malik Sheriff, R., Shoaie, S., Sonnenschein, N., Teusink, B., Vilaça, P., Vik, J., Wodke, J., Xavier, J., Yuan, Q., Zakhartsev, M. and Zhang, C., 2020. MEMOTE for standardized genome-scale metabolic model testing. *Nature Biotechnology*, [online] 38(3), pp.272-276. Available at: <https://www.nature.com/articles/s41587-020-0446-y#citeas> [Accessed 1 January 2022].

121. Application of metabolic flux and transcript analyses to understanding the physiology of engineered *Geobacillus thermoglucosidasius*. Ward, J. (Author). 2014. Available at: < https://ethos.bl.uk/OrderDetails.do?uin=uk.bl.ethos.659523>

122. Ziv, N., Brandt, N. and Gresham, D., 2013. The Use of Chemostats in Microbial Systems Biology. *Journal of Visualized Experiments*, [online] (80). Available at: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3940325/> [Accessed 1 January 2022].

123. Durot, M., Bourguignon, P. and Schachter, V., 2009. Genome-scale models of bacterial metabolism: reconstruction and applications. *FEMS Microbiology Reviews*, [online] 33(1), pp.164-190. Available at: <https://pubmed.ncbi.nlm.nih.gov/19067749/> [Accessed 1 January 2022].

124. You, L., Page, L., Feng, X., Berla, B., Pakrasi, H. and Tang, Y., 2012. Metabolic Pathway Confirmation and Discovery Through ¹³C-labeling of Proteinogenic Amino Acids. *Journal of Visualized Experiments*, [online] (59). Available at: <https://pubmed.ncbi.nlm.nih.gov/22314852/> [Accessed 1 January 2022].

125. Long, C. and Antoniewicz, M., 2019. High-resolution 13C metabolic flux analysis. *Nature Protocols*, [online] 14(10), pp.2856-2877. Available at: <https://www.nature.com/articles/s41596-019-0204-0#citeas> [Accessed 1 January 2022].

126. Lommen, A. and Kools, H., 2011. MetAlign 3.0: performance enhancement by efficient use of advances in computer hardware. *Metabolomics*, 8(4), pp.719-726.

127. Cordova, L. and Antoniewicz, M., 2016. 13C metabolic flux analysis of the extremely thermophilic, fast growing, xylose-utilizing Geobacillus strain LC300. *Metabolic Engineering*, 33, pp.148-157.

128. Tang, Y., Sapra, R., Joyner, D., Hazen, T., Myers, S., Reichmuth, D., Blanch, H. and Keasling, J., 2009. Analysis of metabolic pathways and fluxes in a newly discovered thermophilic and ethanol-tolerantGeobacillusstrain. *Biotechnology and Bioengineering*, 102(5), pp.1377-1386.

129. Cuevas, D., Edirisinghe, J., Henry, C., Overbeek, R., O'Connell, T. and Edwards, R., 2016. From DNA to FBA: How to Build Your Own Genome-Scale Metabolic Model. *Frontiers in Microbiology*, 7.
130. Bordbar, A., Monk, J., King, Z. and Palsson, B., 2014. Constraint-based models predict metabolic and associated cellular functions. *Nature Reviews Genetics*, 15(2), pp.107-120.
131. Aziz, R., Bartels, D., Best, A., DeJongh, M., Disz, T., Edwards, R., Formsma, K., Gerdes, S., Glass, E., Kubal, M., Meyer, F., Olsen, G., Olson, R., Osterman, A., Overbeek, R., McNeil, L., Paarmann, D., Paczian, T., Parrello, B., Pusch, G., Reich, C., Stevens, R., Vassieva, O., Vonstein, V., Wilke, A. and Zagnitko, O., 2008. The RAST Server: Rapid Annotations using Subsystems Technology. *BMC Genomics*, 9(1).
132. Overbeek, R., Olson, R., Pusch, G., Olsen, G., Davis, J., Disz, T., Edwards, R., Gerdes, S., Parrello, B., Shukla, M., Vonstein, V., Wattam, A., Xia, F. and Stevens, R., 2013. The SEED and the Rapid Annotation of microbial genomes using Subsystems Technology (RAST). *Nucleic Acids Research*, 42(D1), pp.D206-D214.
133. Seemann, T., 2014. Prokka: rapid prokaryotic genome annotation. *Bioinformatics*, 30(14), pp.2068-2069.
134. Pareja-Tobes, P., Manrique, M., Pareja-Tobes, E., Pareja, E. and Tobes, R., 2012. BG7: A New Approach for Bacterial Genome Annotation Designed for Next Generation Sequencing Data. *PLoS ONE*, 7(11), p.e49239.
135. Tatusova, T., DiCuccio, M., Badretdin, A., Chetvernin, V., Nawrocki, E., Zaslavsky, L., Lomsadze, A., Pruitt, K., Borodovsky, M. and Ostell, J., 2016. NCBI prokaryotic genome annotation pipeline. *Nucleic Acids Research*, 44(14), pp.6614-6624.
136. Kong, J., Huh, S., Won, J., Yoon, J., Kim, B. and Kim, K., 2019. GAAP: A Genome Assembly + Annotation Pipeline. *BioMed Research International*, 2019, pp.1-12.
137. Brůna, T., Hoff, K., Lomsadze, A., Stanke, M. and Borodovsky, M., 2021. BRAKER2: automatic eukaryotic genome annotation with GeneMark-EP+ and AUGUSTUS supported by a protein database. *NAR Genomics and Bioinformatics*, 3(1).
138. Thiele, I. and Palsson, B., 2010. A protocol for generating a high-quality genome-scale metabolic reconstruction. *Nature Protocols*, 5(1), pp.93-121.
139. Liberal, R., Lisowska, B., Leak, D. and Pinney, J., 2015. PathwayBooster: a tool to support the curation of metabolic pathways. *BMC Bioinformatics*, 16(1).
140. Simeonidis, E. and Price, N., 2015. Genome-scale modeling for metabolic engineering. *Journal of Industrial Microbiology and Biotechnology*, 42(3), pp.327-338.
141. King, Z., Lu, J., Dräger, A., Miller, P., Federowicz, S., Lerman, J., Ebrahim, A., Palsson, B. and Lewis, N., 2015. BiGG Models: A platform for integrating, standardizing and sharing genome-scale models. *Nucleic Acids Research*, 44(D1), pp.D515-D522.
142. Chang, A., Jeske, L., Ulbrich, S., Hofmann, J., Koblitz, J., Schomburg, I., Neumann-Schaal, M., Jahn, D. and Schomburg, D., 2020. BRENDA, the ELIXIR core data resource in 2021: new developments and updates. *Nucleic Acids Research*, 49(D1), pp.D498-D508.
143. Laskowski, R., Jabłońska, J., Pravda, L., Vařeková, R. and Thornton, J., 2017. PDBsum: Structural summaries of PDB entries. *Protein Science*, 27(1), pp.129-134.

144. Kanehisa, M., 2000. KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Research*, 28(1), pp.27-30.
145. Caspi, R., Altman, T., Billington, R., Dreher, K., Foerster, H., Fulcher, C., Holland, T., Keseler, I., Kothari, A., Kubo, A., Krummenacker, M., Latendresse, M., Mueller, L., Ong, Q., Paley, S., Subhraveti, P., Weaver, D., Weerasinghe, D., Zhang, P. and Karp, P., 2013. The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of Pathway/Genome Databases. *Nucleic Acids Research*, 42(D1), pp.D459-D471.
146. Moretti, S., Tran, V., Mehl, F., Ibberson, M. and Pagni, M., 2020. MetaNetX/MNXref: unified namespace for metabolites and biochemical reactions in the context of metabolic models. *Nucleic Acids Research*, 49(D1), pp.D570-D574.
147. *Nucleic Acids Research*, 2015. Database resources of the National Center for Biotechnology Information. 44(D1), pp.D7-D19.
148. Jassal, B., Matthews, L., Viteri, G., Gong, C., Lorente, P., Fabregat, A., Sidiropoulos, K., Cook, J., Gillespie, M., Haw, R., Loney, F., May, B., Milacic, M., Rothfels, K., Sevilla, C., Shamovsky, V., Shorser, S., Varusai, T., Weiser, J., Wu, G., Stein, L., Hermjakob, H. and D'Eustachio, P., 2019. The reactome pathway knowledgebase. *Nucleic Acids Research*,
149. Elbourne, L., Tetu, S., Hassan, K. and Paulsen, I., 2016. TransportDB 2.0: a database for exploring membrane transporters in sequenced genomes from all domains of life. *Nucleic Acids Research*, 45(D1), pp.D320-D324.
150. Bateman, A., Martin, M., Orchard, S., Magrane, M., Agivetova, R., Ahmad, S., Alpi, E., Bowler-Barnett, E., Britto, R., Bursteinas, B., Bye-A-Jee, H., Coetzee, R., Cukura, A., Da Silva, A., Denny, P., Dogan, T., Ebenezer, T., Fan, J., Castro, L., Garmiri, P., Georghiou, G., Gonzales, L., Hatton-Ellis, E., Hussein, A., Ignatchenko, A., Insana, G., Ishtiaq, R., Jokinen, P., Joshi, V., Jyothi, D., Lock, A., Lopez, R., Luciani, A., Luo, J., Lussi, Y., MacDougall, A., Madeira, F., Mahmoudy, M., Menchi, M., Mishra, A., Moulang, K., Nightingale, A., Oliveira, C., Pundir, S., Qi, G., Raj, S., Rice, D., Lopez, M., Saidi, R., Sampson, J., Sawford, T., Speretta, E., Turner, E., Tyagi, N., Vasudev, P., Volynkin, V., Warner, K., Watkins, X., Zaru, R., Zellner, H., Bridge, A., Poux, S., Redaschi, N., Aimo, L., Argoud-Puy, G., Auchincloss, A., Axelsen, K., Bansal, P., Baratin, D., Blatter, M., Bolleman, J., Boutet, E., Breuza, L., Casals-Casas, C., de Castro, E., Echioukh, K., Coudert, E., Cuche, B., Doche, M., Dornevil, D., Estreicher, A., Famiglietti, M., Feuermann, M., Gasteiger, E., Gehant, S., Gerritsen, V., Gos, A., Gruaz-Gumowski, N., Hinz, U., Hulo, C., Hyka-Nouspikel, N., Jungo, F., Keller, G., Kerhornou, A., Lara, V., Le Mercier, P., Lieberherr, D., Lombardot, T., Martin, X., Masson, P., Morgat, A., Neto, T., Paesano, S., Pedruzzi, I., Pilbout, S., Pourcel, L., Pozzato, M., Pruess, M., Rivoire, C., Sigrist, C., Sonesson, K., Stutz, A., Sundaram, S., Tognolli, M., Verbregue, L., Wu, C., Arighi, C., Arminski, L., Chen, C., Chen, Y., Garavelli, J., Huang, H., Laiho, K., McGarvey, P., Natale, D., Ross, K., Vinayaka, C., Wang, Q., Wang, Y., Yeh, L., Zhang, J., Ruch, P. and Teodoro, D., 2020. UniProt: the universal protein knowledgebase in 2021. *Nucleic Acids Research*, 49(D1), pp.D480-D489.
151. Orth, J., Thiele, I. and Palsson, B., 2010. What is flux balance analysis?. *Nature Biotechnology*, 28(3), pp.245-248.
152. Yasemi, M. and Jolicoeur, M., 2021. Modelling Cell Metabolism: A Review on Constraint-Based Steady-State and Kinetic Approaches. *Processes*, 9(2), p.322.

153. Lee, J., 2006. Flux balance analysis in the era of metabolomics. *Briefings in Bioinformatics*, 7(2), pp.140-150.
154. Price, N., Reed, J. and Palsson, B., 2004. Genome-scale models of microbial cells: evaluating the consequences of constraints. *Nature Reviews Microbiology*, 2(11), pp.886-897.
155. Chavali, A., D'Auria, K., Hewlett, E., Pearson, R. and Papin, J., 2012. A metabolic network approach for the identification and prioritization of antimicrobial drug targets. *Trends in Microbiology*, 20(3), pp.113-123.
156. Bernhard O. Palsson. Systems Biology: Simulation of Dynamic Network States. Cambridge University Press, 2011
157. Feist, A. and Palsson, B., 2010. The biomass objective function. *Current Opinion in Microbiology*, 13(3), pp.344-349.
158. Chan, S., Cai, J., Wang, L., Simons-Senftle, M. and Maranas, C., 2017. Standardizing biomass reactions and ensuring complete mass balance in genome-scale metabolic models. *Bioinformatics*, 33(22), pp.3603-3609.
159. COVERT, M., SCHILLING, C. and PALSSON, B., 2001. Regulation of Gene Expression in Flux Balance Models of Metabolism. *Journal of Theoretical Biology*, 213(1), pp.73-88.
160. Edwards, J. and Palsson, B., 2000. The Escherichia coli MG1655 in silico metabolic genotype: Its definition, characteristics, and capabilities. *Proceedings of the National Academy of Sciences*, 97(10), pp.5528-5533.
161. Burgard, A., Pharkya, P. and Maranas, C., 2003. Optknock: A bilevel programming framework for identifying gene knockout strategies for microbial strain optimization. *Biotechnology and Bioengineering*, 84(6), pp.647-657.
162. Segre, D., Vitkup, D. and Church, G., 2002. Analysis of optimality in natural and perturbed metabolic networks. *Proceedings of the National Academy of Sciences*, 99(23), pp.15112-15117.
163. Burgard, A., Pharkya, P. and Maranas, C., 2003. Optknock: A bilevel programming framework for identifying gene knockout strategies for microbial strain optimization. *Biotechnology and Bioengineering*, 84(6), pp.647-657.
164. Hartmann, A., Vila-Santa, A., Kallscheuer, N., Vogt, M., Julien-Laferrière, A., Sagot, M., Marienhagen, J. and Vinga, S., 2017. OptPipe - a pipeline for optimizing metabolic engineering targets. *BMC Systems Biology*, 11(1).
165. Mahadevan, R. and Schilling, C., 2003. The effects of alternate optimal solutions in constraint-based genome-scale metabolic models. *Metabolic Engineering*, 5(4), pp.264-276.
166. Gudmundsson, S. and Thiele, I., 2010. Computationally efficient flux variability analysis. *BMC Bioinformatics*, 11(1).
167. Leighty, R. and Antoniewicz, M., 2011. Dynamic metabolic flux analysis (DMFA): A framework for determining fluxes at metabolic non-steady state. *Metabolic Engineering*, [online] 13(6), pp.745-755. Available at: <https://www.sciencedirect.com/science/article/abs/pii/S1096717611001005?via%3Dihub> [Accessed 2 January 2022].
168. Lewis, N., Hixson, K., Conrad, T., Lerman, J., Charusanti, P., Polpitiya, A., Adkins, J., Schramm, G., Purvine, S., Lopez-Ferrer, D., Weitz, K., Eils, R., König, R., Smith, R. and Palsson, B., 2010. Omic data from evolved E. coli are consistent with computed optimal growth from genome-scale models. *Molecular Systems Biology*, 6(1), p.390.
169. Ahmad, A., Hartman, H., Krishnakumar, S., Fell, D., Poolman, M. and Srivastava, S., 2017. A Genome Scale Model of Geobacillus

thermoglucosidasius (C56-YS93) reveals its biotechnological potential on rice straw hydrolysate. *Journal of Biotechnology*, 251, pp.30-37.
170. Cordova, L., Cipolla, R., Swarup, A., Long, C. and Antoniewicz, M., 2017. 13C metabolic flux analysis of three divergent extremely thermophilic bacteria: Geobacillus sp. LC300, Thermus thermophilus HB8, and Rhodothermus marinus DSM 4252. *Metabolic Engineering*, 44, pp.182-190.
171. Wang, Q., Cen, Z. and Zhao, J., 2015. The Survival Mechanisms of Thermophiles at High Temperatures: An Angle of Omics. *Physiology*, 30(2), pp.97-106.
172. Stincone, A., Prigione, A., Cramer, T., Wamelink, M., Campbell, K., Cheung, E., Olin-Sandoval, V., Grüning, N., Krüger, A., Tauqeer Alam, M., Keller, M., Breitenbach, M., Brindle, K., Rabinowitz, J. and Ralser, M., 2014. The return of metabolism: biochemistry and physiology of the pentose phosphate pathway. *Biological Reviews*, 90(3), pp.927-963.
173. Pereira, R., Nielsen, J. and Rocha, I., 2016. Improving the flux distributions simulated with genome-scale metabolic models of Saccharomyces cerevisiae. *Metabolic Engineering Communications*, 3, pp.153-163.
174. Crown, S., Long, C. and Antoniewicz, M., 2016. Optimal tracers for parallel labeling experiments and 13C metabolic flux analysis: A new precision and synergy scoring system. *Metabolic Engineering*, 38, pp.10-18.
175. Das, M., Dewan, A., Shee, S. and Singh, A., 2021. The Multifaceted Bacterial Cysteine Desulfurases: From Metabolism to Pathogenesis. *Antioxidants*, 10(7), p.997.
176. Marsden, S., McMillan, D. and Hanefeld, U., 2020. Assessing the Thiamine Diphosphate Dependent Pyruvate Dehydrogenase E1 Subunit for Carboligation Reactions with Aliphatic Ketoacids. *International Journal of Molecular Sciences*, 21(22), p.8641.
177. Jenkins, A., Schyns, G., Potot, S., Sun, G. and Begley, T., 2007. A new thiamin salvage pathway. *Nature Chemical Biology*, 3(8), pp.492-497.
178. Eram, M. and Ma, K., 2013. Decarboxylation of Pyruvate to Acetaldehyde for Ethanol Production by Hyperthermophiles. *Biomolecules*, 3(4), pp.578-596.
179. Park, J., Dorrestein, P., Zhai, H., Kinsland, C., McLafferty, F. and Begley, T., 2003. Biosynthesis of the Thiazole Moiety of Thiamin Pyrophosphate (Vitamin B1). *Biochemistry*, 42(42), pp.12430-12438.
180. Campbell, I., Bennett, G. and Silberg, J., 2019. Evolutionary Relationships Between Low Potential Ferredoxin and Flavodoxin Electron Carriers. *Frontiers in Energy Research*, 7.
181. Teske, A., 2018. Aerobic Archaea in iron-rich springs. *Nature Microbiology*, 3(6), pp.646-647.
182. Dauner, M., 2010. From fluxes and isotope labeling patterns towards in silico cells. *Current Opinion in Biotechnology*, 21(1), pp.55-62.
183. Sauer, U., 2006. Metabolic networks in motion: 13 C-based flux analysis. *Molecular Systems Biology*, 2(1), p.62.
184. Winter, G. and Krömer, J., 2013. Fluxomics - connecting 'omics analysis and phenotypes. *Environmental Microbiology*, 15(7), pp.1901-1916.
185. Friedlander, T., Mayo, A., Tlusty, T. and Alon, U., 2015. Evolution of Bow-Tie Architectures in Biology. *PLOS Computational Biology*, 11(3), p.e1004055.
186. Toya, Y., Kono, N., Arakawa, K. and Tomita, M., 2011. Metabolic Flux Analysis and Visualization. *Journal of Proteome Research*, 10(8), pp.3313-3323.

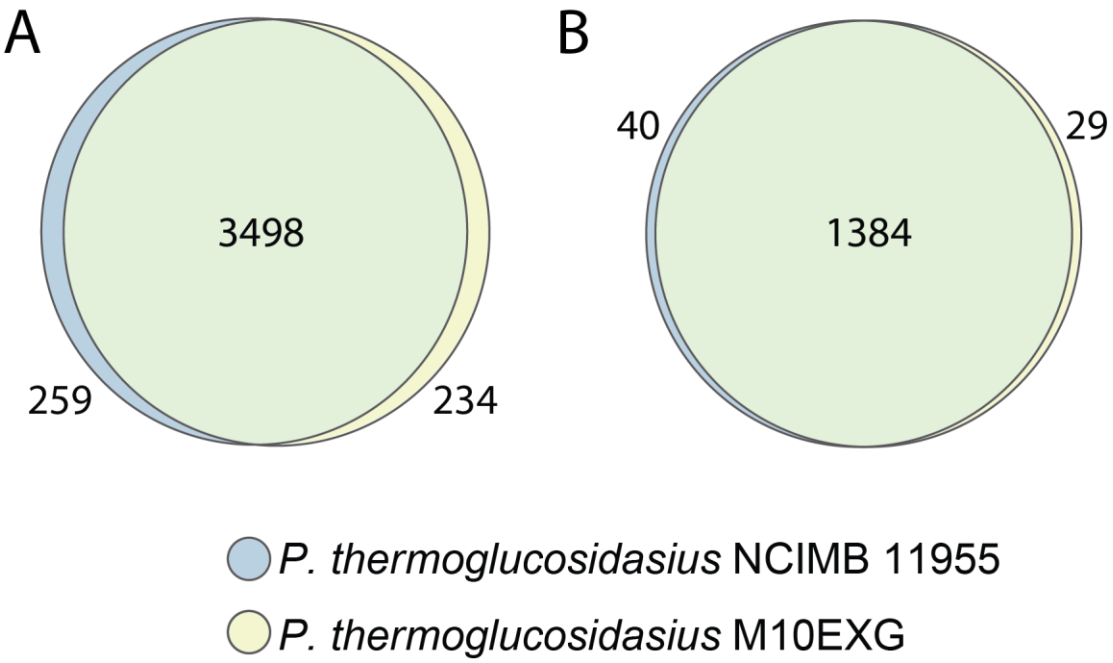
187. Weitzel, M., Wiechert, W. and Nöh, K., 2007. The topology of metabolic isotope labeling networks. *BMC Bioinformatics*, 8(1).
188. Dai, Z. and Locasale, J., 2017. Understanding metabolism with flux analysis: From theory to application. *Metabolic Engineering*, 43, pp.94-102.
189. Wiechert, W., Mallney, M., Isermann, N., Wurzel, M. and de Graaf, A., 1999. Bidirectional reaction steps in metabolic networks: III. Explicit solution and analysis of isotopomer labeling systems. *Biotechnology and Bioengineering*, 66(2), pp.69-85.
190. Sriram, G., 2004. Improvements in metabolic flux analysis using carbon bond labeling experiments: bondomer balancing and Boolean function mapping. *Metabolic Engineering*, 6(2), pp.116-132.
191. Antoniewicz, M., Kelleher, J. and Stephanopoulos, G., 2007. Elementary metabolite units (EMU): A novel framework for modeling isotopic distributions. *Metabolic Engineering*, 9(1), pp.68-86.
192. Choi, J. and Antoniewicz, M., 2019. Tandem Mass Spectrometry for 13C Metabolic Flux Analysis: Methods and Algorithms Based on EMU Framework. *Frontiers in Microbiology*, 10.
193. Jazmin, L., O'Grady, J., Ma, F., Allen, D., Morgan, J. and Young, J., 2013. Isotopically Nonstationary MFA (INST-MFA) of Autotrophic Metabolism. *Plant Metabolic Flux Analysis*, pp.181-210.
194. Adamberg, K., Valgepea, K. and Vilu, R., 2015. Advanced continuous cultivation methods for systems microbiology. *Microbiology*, 161(9), pp.1707-1719.
195. Yasid, N., Rolfe, M., Green, J. and Williamson, M., 2016. Homeostasis of metabolites in Escherichia coli on transition from anaerobic to aerobic conditions and the transient secretion of pyruvate. *Royal Society Open Science*, 3(8), p.160187.
196. Heuillet, M., Bellvert, F., Cahoreau, E., Letisse, F., Millard, P. and Portais, J., 2018. Methodology for the Validation of Isotopic Analyses by Mass Spectrometry in Stable-Isotope Labeling Experiments. *Analytical Chemistry*, 90(3), pp.1852-1860.
197. Nargund, S., Joffe, M., Tran, D., Tugarinov, V. and Sriram, G., 2013. Nuclear Magnetic Resonance Methods for Metabolic Fluxomics. *Methods in Molecular Biology*, pp.335-351.
198. Antoniewicz, M., 2015. Parallel labeling experiments for pathway elucidation and 13C metabolic flux analysis. *Current Opinion in Biotechnology*, 36, pp.91-97.
199. Cheah, Y. and Young, J., 2018. Isotopically nonstationary metabolic flux analysis (INST-MFA): putting theory into practice. *Current Opinion in Biotechnology*, 54, pp.80-87.
200. Hollinshead, W., He, L. and Tang, Y., 2019. 13C-Fingerprinting and Metabolic Flux Analysis of Bacterial Metabolisms. *Methods in Molecular Biology*, pp.215-230.
201. Shree, M. and K. Masakapalli, S., 2018. Intracellular Fate of Universally Labelled 13C Isotopic Tracers of Glucose and Xylose in Central Metabolic Pathways of Xanthomonas oryzae. *Metabolites*, 8(4), p.66.
202. Jang, C., Chen, L. and Rabinowitz, J., 2018. Metabolomics and Isotope Tracing. *Cell*, 173(4), pp.822-837.
203. Jyoti, P., Shree, M., Joshi, C., Prakash, T., Ray, S., Satapathy, S. and Masakapalli, S., 2020. The Entner-Doudoroff and Nonoxidative Pentose Phosphate Pathways Bypass Glycolysis and the Oxidative Pentose Phosphate Pathway in Ralstonia solanacearum. *mSystems*, 5(2).
204. Nöh, K. and Wiechert, W., 2011. The benefits of being transient: isotope-based metabolic flux analysis at the short time scale. *Applied Microbiology and Biotechnology*, 91(5), pp.1247-1265.

205. Wiechert, W. and Nöh, K., 2013. Isotopically non-stationary metabolic flux analysis: complex yet highly informative. *Current Opinion in Biotechnology*, 24(6), pp.979-986.
206. Antoniewicz, M., 2021. A guide to metabolic flux analysis in metabolic engineering: Methods, tools and applications. *Metabolic Engineering*, 63, pp.2-12.
207. Noack, S., Nöh, K., Moch, M., Oldiges, M. and Wiechert, W., 2011. Stationary versus non-stationary 13C-MFA: A comparison using a consistent dataset. *Journal of Biotechnology*, 154(2-3), pp.179-190.
208. Antoniewicz, M., 2015. Methods and advances in metabolic flux analysis: a mini-review. *Journal of Industrial Microbiology and Biotechnology*, 42(3), pp.317-325.
209. Place, T., Domann, F. and Case, A., 2017. Limitations of oxygen delivery to cells in culture: An underappreciated problem in basic and translational research. *Free Radical Biology and Medicine*, 113, pp.311-322.
210. Liu, C., Qin, J. and Lin, Y., 2017. Fermentation and Redox Potential. *Fermentation Processes*,.
211. Papagianni, M., 2012. Recent advances in engineering the central carbon metabolism of industrially important bacteria. *Microbial Cell Factories*, 11(1), p.50.
212. Borja, G., Meza Mora, E., Barrón, B., Gosset, G., Ramírez, O. and Lara, A., 2012. Engineering Escherichia coli to increase plasmid DNA production in high cell-density cultivations in batch mode. *Microbial Cell Factories*, 11(1).
213. Millard, P., Portais, J. and Mendes, P., 2015. Impact of kinetic isotope effects in isotopic studies of metabolic systems. *BMC Systems Biology*, 9(1).
214. Routledge, S., 2012. BEYOND DE-FOAMING: THE EFFECTS OF ANTIFOAMS ON BIOPROCESS PRODUCTIVITY. *Computational and Structural Biotechnology Journal*, 3(4), p.e201210001.
215. Amartey, S., Leak, D. and Hartley, B., 1991. Effects of temperature and medium composition on the ethanol tolerance ofBacillus stearothermophilus LLD-15. *Biotechnology Letters*, 13(9), pp.627-632.
216. Lonhienne, T., Garcia, M., Noble, C., Harmer, J., Fraser, J., Williams, C. and Guddat, L., 2017. High Resolution Crystal Structures of the Acetohydroxyacid Synthase-Pyruvate Complex Provide New Insights into Its Catalytic Mechanism. *ChemistrySelect*, 2(36), pp.11981-11988.
217. Durot, M., Bourguignon, P. and Schachter, V., 2009. Genome-scale models of bacterial metabolism: reconstruction and applications. *FEMS Microbiology Reviews*, 33(1), pp.164-190.
218. Antoniewicz, M., 2018. A guide to 13C metabolic flux analysis for the cancer biologist. *Experimental & Molecular Medicine*, 50(4), pp.1-13.
219. García Martín, H., Kumar, V., Weaver, D., Ghosh, A., Chubukov, V., Mukhopadhyay, A., Arkin, A. and Keasling, J., 2015. A Method to Constrain Genome-Scale Models with 13C Labeling Data. *PLOS Computational Biology*, 11(9), p.e1004363.
220. Gopalakrishnan, S. and Maranas, C., 2015. ^{13}C metabolic flux analysis at a genome-scale. *Metabolic Engineering*, 32, pp.12-22.

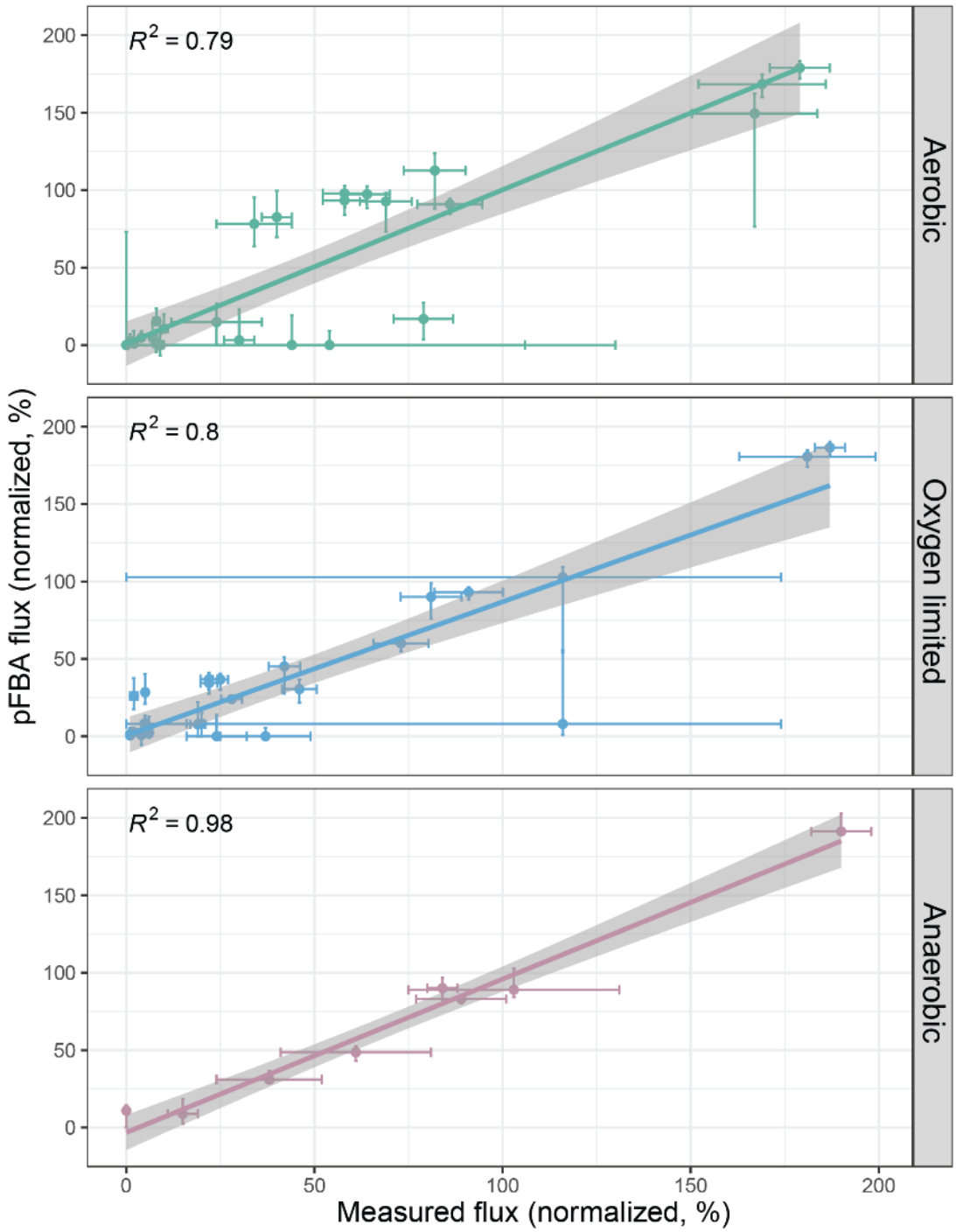
7. Appendices

7.1 Appendix 1: Supplementary Figures to ‘Genome-scale metabolic modeling of *P. thermoglucosidasius* NCIMB 11955 reveals metabolic bottlenecks in anaerobic metabolism’.

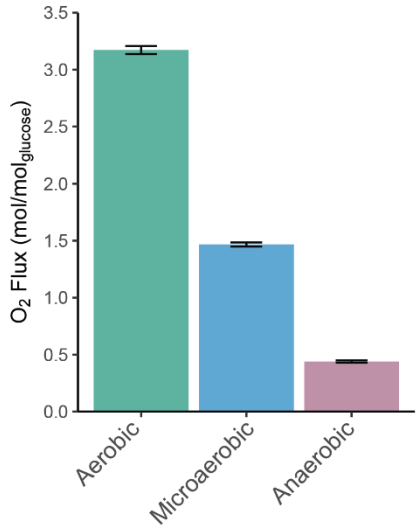
Supplementary figure 1: Whole proteome comparison between *P. thermoglucosidasius* NCIMB 11955 and *P. thermoglucosidaius M10EXG*, for all ORFs (A) and when filtered for metabolic genes (B). Supplementary tables 4 and 5 list the unique metabolic ORFs between the two strains.



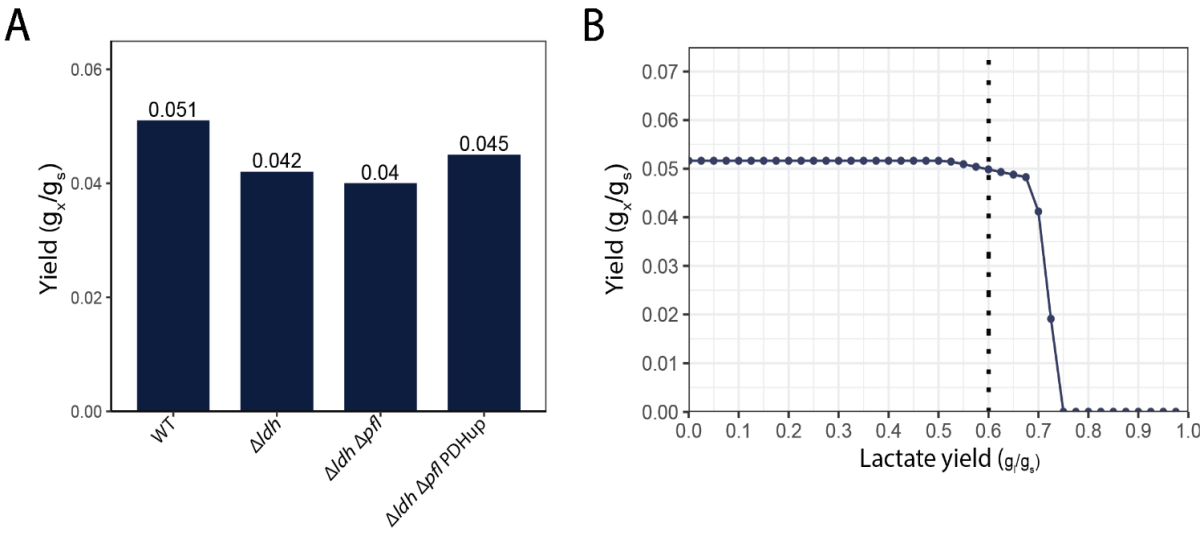
Supplementary figure 2: Correlation between pFBA analysis of the model and experimentally derived data[7], normalized to glucose uptake rate, in aerobic, oxygen limited and anaerobic conditions. Errors for measured fluxes and variability in pFBA fluxes are shown. A linear fit has been applied to assay correlation, with the R value indicated per condition.



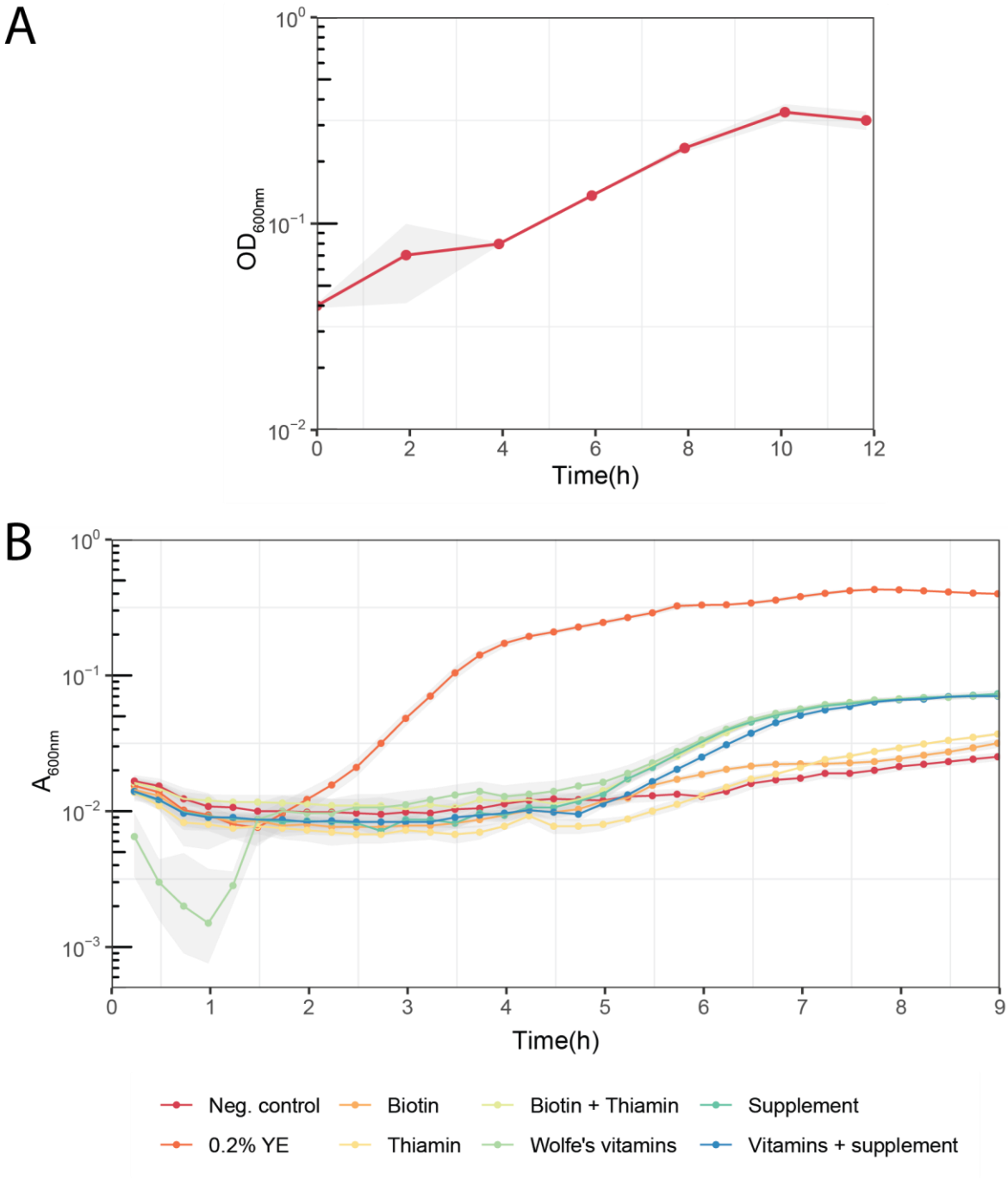
Supplementary figure 3: Predicted oxygen consumption rates for the three conditions, when measured exchange rates of fermentation products were fit to the model.



Supplementary figure 4: A) result of computing predicted *in silico* biomass yield, when measured exchange rates, carbon uptake rates and genetic manipulations (i.e. knockouts) have been introduced. B) Effect of lactate production on biomass yield when all other measured exchange rates are fixed for the WT strain. The dotted line indicates the measured lactate production in these conditions.



Supplementary figure 5: A) Aerobic shake flask experiment of *P. thermoglucosidasius* NCIMB 11955 on TMM base medium. Shaded area shows standard devaition between three biological replicates. B) Anaerobic growth curves of *P. thermoglucosidasius* NCIMB 11955 grown on base TMM supplemented with various nutrients. Experiment was performed in a microtiter plate reader, and shaded area represents standard deviation of measurements over quadruplicates. (YE = yeast extract)



Supplementary Tables

Supplementary Table 1: Stoichiometry of the biomass reaction in *p-thermo*.

	Compound	mmol/gDCW
Proteins		
	Ala	0.5142
	Arg	0.1831
	Asp	0.2219
	Asn	0.2219
	Cys	0.1079
	Glu	0.322
	Gln	0.322
	Gly	0.4077
	His	0.0778
	Ile	0.2728
	Leu	0.3475
	Lys	0.3172
	Met	0.0832
	Phe	0.1451
	Pro	0.1665
	Ser	0.1811
	Thr	0.2688
	Trp	0.1026
	Tyr	0.1036
	Val	0.3761
Nucleic Acids	*Ribonucleic acid (RNA)*	
	AMP	0.1193
	CMP	0.0915
	GMP	0.0915
	UMP	0.1193
	Deoxyribonucleic acid (DNA)	
	dAMP	0.0077
	dCMP	0.0059
	dGMP	0.0059
	dTMP	0.0077
Lipids	*Phospholipids*	
	PE	0.0290
	PG	0.0277
	CLPN	0.0296
Carbohydrates		
	D-Fructose	0.1048
	GDP-Mannose	0.0057
	UDP-D-Galactose	0.1895
	UDP-D-Xylose	0.1825
	UDP-D-Glucose	0.0096
	UDP-L-Arabinose	0.0407

	Compound	mmol/gDCW
Salts		
	Phosphorous	0.0420
	Calcium	0.0028
	Potassium	0.6323
	Magnesium	0.0875
	Iron	0.0304
Vitamins & cofactors		
	Thiamin B1	0.0002
	Riboflavin B2	0.0002
	Vitamin B12	0.0002
	Vitamin B6	0.0002
	10-Formyltetrahydrofolate	0.0004
	Biotin	1.79E-06
	Bacillithiol	6.23E-05
	Bacillithiol disulfide	1.56E-07
	Chorismate	0.0002
	FAD	0.0002
	FMN	0.0002
	Heme	0.0002
	Proto-heme	0.0002
	Siroheme	0.0002
	Menaquinol	0.0003
	NAD+	0.0145
	NADH	0.0267
	NADP	0.0042
	NADPH	0.0027
	S-Adenosyl-L-methionine	0.0002
	Spermidine	0.0011
	Spermine	0.0025
	Putrescine	0.0054
	CoA	0.0002
Intracellular Metabolites		
	Acetyl-CoA	0.0002
	Succinyl-CoA	8.75E-05
Energy Requirement		
	ADP	- 152.28
	Pi	-152.27
	Ppi	0.0011
	H	- 152.28
	H2O	152.28
	ATP	152.31

Supplementary table 2: Estimation of polymerization energy needed to form biomass from the different monomer classes present in the biomass reaction. This energy fraction constitutes part of the growth associated maintenance that is present in the biomass reaction. Polymerization energy per molecule was obtained from literature[65].

	Cellular content (w/w)	Polymerization energy	
		Per molecule (mmol ATP/g polymer)	Total (mmol ATP/gcell)
Protein	0.52	37.7	19.604
Carbohydrates	0.1	12.8	1.28
RNA	0.16	26	4.16
DNA	0.1	26	2.6
Lipids	0.09	25.6	2.304
Sum			29.948

Supplementary table 3: Overview of an analyses of filtering the unique ORFs detected in the genome analyses between two *(Para)geobacillus* strains, when various levels of filtering are applied to elucidate how many reactions would be unique in models made of each strain, and finally which would be connected to any pre-existing metabolites in the network. Supplementary table 4 and 5 highlight the unique metabolic ORFs identified.

	***P. thermoglucosidasius* NCIMB 11955**	***P. thermoglucosiodasius* M10EXG**
Unique ORFs	259	234
Unique metabolic ORFs[1]	40	29
Unique ORFs as reactions[2]	18	13
Unique reactions[3]	11	12
Connected unique reactions[4]	8	6

[1] Unique metabolic ORFs are ORFs with an E.C. code associated to them
[2] Subset of unique metabolic ORFs that would actually be captured as a reaction in the model
[3] Subset of unique ORFs that would cause new reactions to the model
[4] Subset of unique reactions that are connected by a main metabolite to one or more pre-existing metabolites in the model. Note, all these reactions are still blocked (i.e. no two main metabolites found in model).

Supplementary table 4: Metabolic ORFs unique to *P. thermoglucosidasius* NCIMB 11955, detected in the genome comparison.

Gene annotation	EC	Annotated Kegg Ontology
abfA	3.2.1.55	Alpha-L-arabinofuranosidase
adk	2.7.4.3	adenylate kinase
araA	5.3.1.4	L-arabinose isomerase
cocE	3.1.1.84	Carboxylesterase
ecfA2	3.6.3.-	Hydrolase, involved in transmembrane movement
glf	5.4.99.9	UDP-galactopyranose mutase
gltX	6.1.1.17	Glutamyl-tRNA synthetase
hdl IVa	3.8.1.2	2-haloacid dehalogenase
helD_2	3.6.4.12	DNA helicase
ispD	2.7.7.60	2-C-methyl-D-erythritol 4-phosphate cytidylyltransferase
ispF	4.6.1.12	2-C-methyl-D-erythritol 2,4-cyclodiphosphate synthase
kpsU	2.7.7.38	3-deoxy-manno-octulosonate cytidylyltransferase (CMP-KDO synthetase)
lytC_3	3.5.1.28	N-acetylmuramoyl-L-alanine amidase
mazF	3.1.27.-	Esterase (component of type II toxin-antitoxin system)
mcsB	2.7.14.1	Protein arginine kinase
mngB	3.2.1.170	Mannosylglycerate hydrolase
mrnC	3.1.26.-	Esterase (ribonuclease)
mtlD	1.1.1.17	Mannitol-1-phosphate 5-dehydrogenase
ppk	2.7.4.1	Polyphosphate kinase
ppx	3.6.1.11	Exopolyphosphatase
pseG	3.6.1.57	UDP-2,4-diacetamido-2,4,6-trideoxy-beta-L-altropyranose hydrolase
rapA_1	3.6.4.-	RNA Polymerase associated hydrolase
rbsD	5.4.99.62	D-ribose pyranase
rhpA	3.6.4.13	ATP-dependent helicase
rpoA	2.7.7.6	DNA-directed RNA polymerase
rpoB	2.7.7.6	DNA-directed RNA polymerase
rpoC	2.7.7.6	DNA-directed RNA polymerase
srlB	2.7.1.198	Glucitol/sorbitol PTS system EIIA component
srlE	2.7.1.198	Glucitol/sorbitol PTS system EIIA component
tagD	2.7.7.39	Glycerol-3-phosphate cytidylyltransferase
trmB	2.1.1.33	tRNA (guanine-N7-)-methyltransferase
truA	5.4.99.12	tRNA pseudouridine synthase
wecA	2.7.8.40	UDP-GalNAc:undecaprenyl-phosphate GalNAc-1-phosphate transferase
xylA_1	3.2.1.37	Xylan 1,4-beta-xylosidase
xynA	3.2.1.8	Endo-1,4-beta-xylanase
xynB	3.2.1.37	Xylan 1,4-beta-xylosidase
YDAf_2	2.3.1.-	Acyltransferase
yeeF	3.1.-.-	Ribonuclease
yjjG	3.1.3.5	5'-nucleotidase
yobL	3.1.-.-	Ribonuclease

Supplementary table 5: Metabolic ORFs unique to *P. thermoglucosidasius* M10EXG, detected in the genome comparison.

Gene annotation	EC	Annotated Kegg Ontology
cocE	3.1.1.84	Carboxylesterase
cwlK	3.4.-.-	Peptidase
cynS	4.2.1.104	Cyanate lyase
derK	2.7.1.210	D-erythrulose 4-kinase
dhaL	2.7.-.-	Phosphotransferase
fdtA	5.3.2.3	TDP-4-oxo-6-deoxy-alpha-D-glucose-3,4-oxoisomerase
hsdM_1	2.1.1.72	Site-specific DNA-methyltransferase (adenine-specific)
hsdM_2	2.1.1.72	Site-specific DNA-methyltransferase (adenine-specific)
hsdR_1	3.1.21.3	Type I restriction enzyme
hsdR_2	3.1.21.3	Type I restriction enzyme
iolB	5.3.1.30	5-deoxy-glucuronate isomerase
iolE	4.2.1.44	Inosose dehydratase
iolI	5.3.99.11	2-keto-myo-inositol isomerase
iolX	1.1.1.370	Scyllo-inositol 2-dehydrogenase (NAD+)
lhgO	1.1.3.15	Glycolate oxidase
lipA	2.8.1.8	Lipoyl synthase
lsrF	2.3.1.245	3-hydroxy-5-phosphonooxypentane-2,4-dione thiolase
mutT4	3.6.1.-	Esterase
NA	2.4.1.161	Oligosaccharide 4-alpha-D-glucosyltransferase
radD	3.6.4.12	DNA repair helicase
recD	3.1.11.5	Exodeoxyribonuclease V
rfbC	5.1.3.13	dTDP-4-dehydrorhamnose 3,5-epimerase
sqhC	4.2.1.137	Sporulenol synthase
sunS	2.4.1.-	Glycosyltransferase
tatD	3.1.21.-	Endodeoxyribonucleases
uvrD1	3.6.4.12	ATP-dependent DNA helicase
wapA_1	3.1.-.-	tRNA(Glu)-specific nuclease
wapA_2	3.1.-.-	tRNA(Glu)-specific nuclease
wapA_3	3.1.-.-	tRNA(Glu)-specific nuclease

7.2 Appdendix 2 - Complete validity tables

GA0.075

Table S1: All TBDMS-derivtized amino acid mass fragments for the GA0.075 ILE and their calculated percentage ^{13}C incorporation after naturally abundant ^{13}C isotope correction. Mass fragments are determined as either 'Valid', 'Conditionally Valid' or 'Invalid'.

GA0.0752					
metabolite	AA standa	Standard d	12C Meta-	Type of fragment accuracy	
Ala317	4.62%	0.65%	4.09%	Invalid	Invalid
Ala302	0.15%	0.08%	0.18%	**Valid**	Conditonal valid
Ala260	14.49%	3.76%	14.39%	Invalid	Invalid
Ala232	22.89%	5.65%	22.79%	Invalid	Invalid
Gly303	9.59%	1.55%	7.74%	Invalid	Invalid
Gly288	0.32%	0.07%	0.27%	**Valid**	Conditonal valid
Gly246	17.25%	4.55%	17.09%	Invalid	Invalid
Gly218	25.83%	6.22%	25.40%	Invalid	Invalid
Val345	6.98%	16.86%	6.88%	Invalid	Conditonal valid
Val330	0.20%	22.20%	0.25%	**Valid**	Conditonal valid
Val288	12.39%	1.22%	13.32%	Invalid	Invalid
Val260	8.56%	9.41%	8.44%	Invalid	Conditonal valid
Val186	18.35%	19.88%	18.27%	Invalid	Conditonal valid
Pro343	14.90%	3.45%	17.52%	Invalid	Invalid
Pro328	3.43%	0.66%	3.88%	Invalid	Invalid
Pro301	31.83%	3.91%	33.20%	Invalid	Invalid
Pro286	4.38%	1.59%	4.86%	Invalid	Invalid
Pro258	3.06%	1.99%	2.93%	Invalid	Conditonal valid
Pro184	10.73%	2.38%	10.57%	Invalid	Invalid
Met377	49.92%	0.14%	42.70%	Invalid	Invalid
Met302	11.31%	0.27%	12.15%	Invalid	Invalid
Met292	32.88%	3.24%	23.47%	Invalid	Invalid
Met218	25.08%	3.11%	17.30%	Invalid	Invalid
Phe393	32.83%	6.75%	33.87%	Invalid	Invalid
Phe378	3.05%	0.32%	3.30%	Invalid	Invalid
Phe336	0.89%	0.44%	0.90%	**Valid**	Conditonal valid
Phe308	0.89%	0.39%	0.85%	**Valid**	Conditonal valid
Phe234	1.16%	0.59%	1.15%	**Valid**	Conditonal valid
Ser447	2.76%	1.31%	3.26%	Invalid	Invalid
Ser432	0.12%	0.06%	0.12%	**Valid**	Conditonal valid
Ser390	8.46%	2.96%	8.17%	Invalid	Invalid
Ser362	10.16%	4.07%	9.71%	Invalid	Invalid
Ser288	11.29%	3.88%	10.89%	Invalid	Invalid
Thr461	17.38%	12.06%	17.59%	Invalid	Invalid
Thr446	16.91%	14.04%	18.32%	Invalid	Invalid
Thr404	21.91%	2.07%	21.92%	Invalid	Invalid
Thr376	80.11%	8.03%	84.41%	Invalid	Invalid
Thr302	8.53%	0.66%	9.13%	Invalid	Invalid
Lys488	0.20%	0.06%	0.21%	**Valid**	Conditonal valid
Lys473	0.63%	0.60%	0.57%	**Valid**	Conditonal valid
Lys473	0.63%	0.60%	0.57%	**Valid**	Conditonal valid
Lys431	0.29%	0.19%	0.29%	**Valid**	Conditonal valid
Lys403	4.88%	4.20%	5.32%	Invalid	Conditonal valid
Lys329	0.46%	0.16%	0.48%	**Valid**	Conditonal valid
Lys300	6.07%	0.25%	6.18%	Invalid	Invalid
Cys406	21.39%	4.44%	21.07%	Invalid	Invalid
Cys378	36.03%	22.75%	33.71%	Invalid	Invalid
Cys304	0.00%	10.05%	0.00%	Invalid	Conditonal valid
Tyr523	33.98%	2.44%	34.28%	Invalid	Invalid
Tyr508	1.84%	0.10%	1.80%	Invalid	Invalid
Tyr466	0.43%	0.18%	0.45%	**Valid**	Conditonal valid
Tyr438	1.61%	0.21%	1.47%	Invalid	Invalid
Tyr364	0.78%	0.04%	0.79%	**Valid**	Conditonal valid
Tyr221	4.50%	0.15%	4.58%	Invalid	Invalid
Trp546	7.63%	3.42%	7.07%	Invalid	Invalid
Trp531	52.06%	2.06%	51.49%	Invalid	Invalid
Trp531	52.06%	2.06%	51.49%	Invalid	Invalid
Trp489	35.08%	1.79%	35.51%	Invalid	Invalid
Trp461	56.33%	2.92%	55.89%	Invalid	Invalid
Trp388	41.33%	1.36%	40.98%	Invalid	Invalid
Trp330	31.68%	8.52%	31.22%	Invalid	Invalid
Trp302	11.45%	2.89%	12.86%	Invalid	Invalid
Trp244	51.52%	1.12%	52.98%	Invalid	Invalid
His497	10.42%	0.66%	10.20%	Invalid	Invalid
His482	1.75%	0.22%	1.75%	Invalid	Invalid
His440	0.83%	1.63%	0.72%	**Valid**	Conditonal valid
His412	3.10%	0.88%	3.04%	Invalid	Invalid
His338	1.78%	0.46%	1.68%	Invalid	Invalid
Glu489	0.24%	0.11%	0.25%	**Valid**	Conditonal valid
Glu474	0.16%	0.07%	0.18%	**Valid**	Conditonal valid
Glu432	11.80%	2.04%	11.71%	Invalid	Invalid
Glu404	0.12%	0.02%	0.11%	**Valid**	Conditonal valid
Glu330	7.37%	1.25%	7.28%	Invalid	Invalid
Glu302	1.88%	0.20%	1.99%	Invalid	Invalid
Glu187	46.06%	0.22%	46.00%	Invalid	Invalid
Asp475	2.05%	0.88%	2.23%	Invalid	Conditonal valid
Asp460	0.16%	0.09%	0.15%	**Valid**	Conditonal valid
Asp418	9.95%	2.07%	9.87%	Invalid	Invalid
Asp390	4.00%	0.28%	3.98%	Invalid	Invalid
Asp316	7.51%	1.07%	7.45%	Invalid	Invalid
Asp302	12.97%	2.58%	12.91%	Invalid	Invalid
Leu359	3.39%	2.33%	3.20%	Invalid	Conditonal valid
Leu344	0.17%	0.11%	0.16%	**Valid**	Conditonal valid
Leu302	6.35%	1.47%	6.28%	Invalid	Invalid
Leu274	7.80%	1.84%	7.76%	Invalid	Invalid
Leu200	11.70%	1.91%	11.60%	Invalid	Invalid
ILe359	5.27%	1.55%	5.12%	Invalid	Invalid
ILe344	0.19%	0.11%	0.24%	**Valid**	Conditonal valid
ILe302	6.23%	2.12%	6.16%	Invalid	Invalid
ILe274	5.72%	2.32%	5.65%	Invalid	Invalid
ILe200	9.01%	2.30%	8.94%	Invalid	Invalid
Arg516	0.00%	0.00%	0.00%	Invalid	Conditonal valid
Arg501	0.00%	0.00%	0.00%	Invalid	Conditonal valid
Arg459	0.00%	0.00%	0.00%	Invalid	Conditonal valid
Arg402	0.00%	0.00%	0.00%	Invalid	Conditonal valid
Arg357	0.00%	0.00%	0.00%	Invalid	Conditonal valid
Arg317	0.00%	0.00%	0.00%	Invalid	Conditonal valid
Arg300	0.00%	0.00%	0.00%	Invalid	Conditonal valid
Arg288	0.00%	0.00%	0.00%	Invalid	Conditonal valid
Arg273	0.00%	0.00%	0.00%	Invalid	Conditonal valid
Arg231	0.00%	0.00%	0.00%	Invalid	Conditonal valid
Gln488	38.16%	0.06%	37.42%	Invalid	Invalid
Gln473	41.88%	0.05%	44.98%	Invalid	Invalid
Gln431	16.50%	1.90%	16.26%	Invalid	Invalid
Gln403	28.73%	3.22%	28.54%	Invalid	Invalid
Gln358	31.81%	0.13%	30.37%	Invalid	Invalid
Gln329	38.61%	0.07%	38.57%	Invalid	Invalid
Gln300	52.10%	0.46%	57.39%	Invalid	Invalid
Gln272	31.33%	0.11%	31.22%	Invalid	Invalid
Suc303	0.00%	0.59%	0.00%	Invalid	Conditonal valid
Suc261	0.00%	12.15%	0.00%	Invalid	Conditonal valid
Lac318	69.99%	2.67%	66.69%	Invalid	Invalid
Lac303	1.38%	3.81%	1.42%	Invalid	Conditonal valid
Lac261	0.58%	1.38%	0.53%	**Valid**	Conditonal valid
Lac233	0.24%	1.18%	0.32%	**Valid**	Conditonal valid
Lac189	6.10%	3.45%	5.88%	Invalid	Invalid

GA0.15

Table S2: All TBDMS-derivtized amino acid mass fragments for the GA0.15 ILE and their calculated percentage ^{13}C incorporation after naturally abundant ^{13}C isotope correction. Mass fragments are determined as either 'Valid', 'Conditionally Valid' or 'Invalid'.

GA0.15

metabolite	AA standar	Standard d	12C Meta-	AA standar
Ala317	4.00%	0.65%	4.00%	Invalid
Ala302	0.23%	0.09%	0.25%	**Valid**
Ala260	15.60%	2.19%	16.55%	Invalid
Ala232	25.03%	3.38%	26.45%	Invalid
Gly303	9.22%	1.72%	8.57%	Invalid
Gly288	0.30%	0.18%	0.35%	**Valid**
Gly246	21.57%	5.01%	22.31%	Invalid
Gly218	30.58%	5.32%	31.78%	Invalid
Val345	17.00%	3.30%	15.11%	Invalid
Val330	0.59%	0.13%	0.46%	**Valid**
Val288	6.51%	2.67%	11.95%	Invalid
Val260	8.62%	2.18%	10.06%	Invalid
Val186	18.62%	2.69%	20.39%	Invalid
Pro343	4.21%	0.98%	11.45%	Invalid
Pro328	1.04%	0.37%	3.04%	**Valid**
Pro301	7.43%	1.90%	22.77%	Invalid
Pro286	5.13%	1.94%	5.14%	Invalid
Pro258	7.95%	2.57%	4.58%	Invalid
Pro184	14.77%	2.18%	12.38%	Invalid
Met377	0.25%	0.09%	27.13%	**Valid**
Met302	18.72%	0.47%	9.25%	Invalid
Met292	11.29%	2.96%	36.97%	Invalid
Met218	11.55%	2.62%	12.84%	Invalid
Phe393	28.77%	4.01%	37.04%	Invalid
Phe378	0.93%	0.19%	3.81%	**Valid**
Phe336	1.95%	0.82%	1.52%	Invalid
Phe308	1.85%	0.80%	1.67%	Invalid
Phe234	2.43%	0.74%	1.93%	Invalid
Ser447	4.04%	1.02%	3.50%	Invalid
Ser432	0.28%	0.15%	0.14%	**Valid**
Ser390	17.49%	2.67%	10.59%	Invalid
Ser362	24.37%	4.33%	13.89%	Invalid
Ser288	23.76%	3.69%	14.93%	Invalid
Thr461	0.00%	0.00%	3.07%	Invalid
Thr446	0.00%	0.00%	1.39%	Invalid
Thr404	0.00%	0.00%	2.54%	Invalid
Thr376	0.00%	0.00%	10.05%	Invalid
Thr302	0.00%	0.00%	1.07%	Invalid
Lys488	0.12%	0.06%	0.20%	**Valid**
Lys473	1.24%	0.36%	0.81%	Invalid
Lys473	1.24%	0.36%	0.81%	Invalid
Lys431	0.23%	0.10%	0.40%	**Valid**
Lys403	10.93%	2.59%	6.70%	Invalid
Lys329	1.14%	0.35%	0.70%	**Valid**
Lys300	5.52%	1.61%	7.46%	Invalid
Cys406	12.21%	5.01%	14.92%	Invalid
Cys378	32.38%	24.27%	26.72%	Invalid
Cys304	10.76%	10.26%	0.00%	Invalid
Tyr523	12.39%	1.25%	41.23%	Invalid
Tyr508	0.20%	0.06%	2.30%	**Valid**
Tyr466	2.18%	0.83%	1.04%	Invalid
Tyr438	1.16%	0.52%	2.33%	**Valid**
Tyr364	1.14%	0.50%	1.24%	**Valid**
Tyr221	3.63%	0.21%	6.56%	Invalid
Trp546	8.83%	2.68%	7.79%	Invalid
Trp531	52.94%	1.86%	50.53%	Invalid
Trp531	52.94%	1.86%	50.53%	Invalid
Trp489	37.61%	10.08%	30.31%	Invalid
Trp461	49.04%	1.57%	57.38%	Invalid
Trp388	16.15%	2.79%	40.05%	Invalid
Trp330	48.60%	15.58%	25.05%	Invalid
Trp302	9.99%	3.92%	15.02%	Invalid
Trp244	16.25%	1.19%	57.72%	Invalid
His497	5.32%	3.21%	10.61%	Invalid
His482	0.59%	0.24%	1.60%	**Valid**
His440	4.34%	2.49%	1.25%	Invalid
His412	3.84%	1.06%	4.00%	Invalid
His338	16.15%	1.38%	2.16%	Invalid
Glu489	0.22%	0.11%	0.23%	**Valid**
Glu474	0.18%	0.04%	0.27%	**Valid**
Glu432	9.38%	1.74%	12.69%	Invalid
Glu404	0.08%	0.07%	0.35%	**Valid**
Glu330	5.45%	1.52%	8.56%	Invalid
Glu302	1.80%	0.14%	1.98%	Invalid
Glu187	46.21%	0.18%	46.09%	Invalid
Asp475	2.43%	0.94%	2.68%	Invalid
Asp460	0.40%	0.31%	0.17%	**Valid**
Asp418	7.42%	3.45%	11.85%	Invalid
Asp390	4.22%	1.82%	6.40%	Invalid
Asp316	7.31%	1.84%	9.95%	Invalid
Asp302	8.85%	5.62%	15.51%	Invalid
Leu359	3.80%	1.12%	3.25%	Invalid
Leu344	0.27%	0.14%	0.44%	**Valid**
Leu302	5.70%	1.24%	7.12%	Invalid
Leu274	7.26%	1.48%	8.94%	Invalid
Leu200	11.25%	1.19%	12.67%	Invalid
ILe359	3.54%	0.85%	4.13%	Invalid
ILe344	0.35%	0.14%	0.41%	**Valid**
ILe302	6.72%	1.15%	6.90%	Invalid
ILe274	6.71%	1.36%	6.89%	Invalid
ILe200	9.91%	1.08%	10.04%	Invalid
Arg516	0.00%	0.00%	0.00%	Invalid
Arg501	0.00%	0.00%	0.00%	Invalid
Arg459	0.00%	0.00%	0.00%	Invalid
Arg402	0.00%	0.00%	0.00%	Invalid
Arg357	0.00%	0.00%	0.00%	Invalid
Arg317	0.00%	0.00%	0.00%	Invalid
Arg300	0.00%	0.00%	0.00%	Invalid
Arg288	0.00%	0.00%	0.00%	Invalid
Arg273	0.00%	0.00%	0.00%	Invalid
Arg231	0.00%	0.00%	0.00%	Invalid
Gln488	0.41%	0.32%	37.47%	**Valid**
Gln473	0.65%	0.43%	30.57%	**Valid**
Gln431	5.60%	3.75%	19.85%	Invalid
Gln403	9.29%	4.43%	29.08%	Invalid
Gln358	3.22%	1.17%	27.66%	Invalid
Gln329	1.24%	0.23%	37.79%	Invalid
Gln300	22.35%	0.53%	36.23%	Invalid
Gln272	4.50%	0.54%	33.51%	Invalid
Suc303	0.00%	0.00%	1.62%	Invalid
Suc261	0.00%	0.00%	24.84%	Invalid
Lac318	0.00%	0.00%	25.79%	Invalid
Lac303	0.00%	0.00%	34.92%	Invalid
Lac261	0.00%	0.00%	19.87%	Invalid
Lac233	0.00%	0.00%	11.58%	Invalid
Lac189	0.00%	0.00%	10.91%	Invalid

GA0.225

Table S3: All TBDMS-derivtized amino acid mass fragments for the GA0.225 ILE and their calculated percentage ^{13}C incorporation after naturally abundant ^{13}C isotope correction. Mass fragments are determined as either 'Valid', 'Conditionally Valid' or 'Invalid'.

GA0.225				
metabolite	AA standar	Standard d	12C Meta-	AA standar
Ala317	4.00%	0.65%	3.68%	Invalid
Ala302	0.23%	0.09%	0.19%	**Valid**
Ala260	15.60%	2.19%	13.47%	Invalid
Ala232	25.03%	3.38%	21.73%	Invalid
Gly303	9.22%	1.72%	8.17%	Invalid
Gly288	0.30%	0.18%	0.46%	**Valid**
Gly246	21.57%	5.01%	16.23%	Invalid
Gly218	30.58%	5.32%	24.41%	Invalid
Val345	17.00%	3.30%	17.39%	Invalid
Val330	0.59%	0.13%	0.67%	**Valid**
Val288	6.51%	2.67%	11.76%	Invalid
Val260	8.62%	2.18%	7.70%	Invalid
Val186	18.62%	2.69%	17.57%	Invalid
Pro343	4.21%	0.98%	13.61%	Invalid
Pro328	1.04%	0.37%	4.23%	**Valid**
Pro301	7.43%	1.90%	21.87%	Invalid
Pro286	5.13%	1.94%	3.59%	Invalid
Pro258	7.95%	2.57%	1.92%	Invalid
Pro184	14.77%	2.18%	8.81%	Invalid
Met377	0.25%	0.09%	30.67%	**Valid**
Met302	18.72%	0.47%	9.83%	Invalid
Met292	11.29%	2.96%	12.43%	Invalid
Met218	11.55%	2.62%	8.46%	Invalid
Phe393	28.77%	4.01%	34.90%	Invalid
Phe378	0.93%	0.19%	2.39%	**Valid**
Phe336	1.95%	0.82%	0.91%	Invalid
Phe308	1.85%	0.80%	1.18%	Invalid
Phe234	2.43%	0.74%	1.13%	Invalid
Ser447	4.04%	1.02%	3.20%	Invalid
Ser432	0.28%	0.15%	0.19%	**Valid**
Ser390	17.49%	2.67%	6.49%	Invalid
Ser362	24.37%	4.33%	7.83%	Invalid
Ser288	23.76%	3.69%	9.92%	Invalid
Thr461	0.00%	0.00%	3.07%	Invalid
Thr446	0.00%	0.00%	1.39%	Invalid
Thr404	0.00%	0.00%	2.54%	Invalid
Thr376	0.00%	0.00%	10.05%	Invalid
Thr302	0.00%	0.00%	1.07%	Invalid
Lys488	0.12%	0.06%	0.19%	**Valid**
Lys473	1.24%	0.36%	0.94%	Invalid
Lys473	1.24%	0.36%	0.94%	Invalid
Lys431	0.23%	0.10%	0.30%	**Valid**
Lys403	10.93%	2.59%	8.39%	Invalid
Lys329	1.14%	0.35%	0.75%	**Valid**
Lys300	5.52%	1.61%	6.00%	Invalid
Cys406	12.21%	5.01%	19.51%	Invalid
Cys378	32.38%	24.27%	25.39%	Invalid
Cys304	10.76%	10.26%	0.31%	Invalid
Tyr523	12.39%	1.25%	35.46%	Invalid
Tyr508	0.20%	0.06%	1.65%	**Valid**
Tyr466	2.18%	0.83%	0.79%	Invalid
Tyr438	1.16%	0.52%	1.82%	**Valid**
Tyr364	1.14%	0.50%	0.87%	**Valid**
Tyr221	3.63%	0.21%	5.34%	Invalid
Trp546	8.83%	2.68%	9.32%	Invalid
Trp531	52.94%	1.86%	46.98%	Invalid
Trp531	52.94%	1.86%	46.98%	Invalid
Trp489	37.61%	10.08%	38.20%	Invalid
Trp461	49.04%	1.57%	59.43%	Invalid
Trp388	16.15%	2.79%	39.42%	Invalid
Trp330	48.60%	15.58%	28.32%	Invalid
Trp302	9.99%	3.92%	17.02%	Invalid
Trp244	16.25%	1.19%	49.94%	Invalid
His497	5.32%	3.21%	13.80%	Invalid
His482	0.59%	0.24%	2.88%	**Valid**
His440	4.34%	2.49%	0.95%	Invalid
His412	3.84%	1.06%	4.80%	Invalid
His338	16.15%	1.38%	2.13%	Invalid
Glu489	0.22%	0.11%	0.27%	**Valid**
Glu474	0.18%	0.04%	0.23%	**Valid**
Glu432	9.38%	1.74%	11.46%	Invalid
Glu404	0.08%	0.07%	0.32%	**Valid**
Glu330	5.45%	1.52%	7.39%	Invalid
Glu302	1.80%	0.14%	2.10%	Invalid
Glu187	46.21%	0.18%	46.03%	Invalid
Asp475	2.43%	0.94%	2.71%	Invalid
Asp460	0.40%	0.31%	0.19%	**Valid**
Asp418	7.42%	3.45%	9.22%	Invalid
Asp390	4.22%	1.82%	3.76%	Invalid
Asp316	7.31%	1.84%	7.18%	Invalid
Asp302	8.85%	5.62%	12.40%	Invalid
Leu359	3.80%	1.12%	4.26%	Invalid
Leu344	0.27%	0.14%	0.49%	**Valid**
Leu302	5.70%	1.24%	5.78%	Invalid
Leu274	7.26%	1.48%	7.41%	Invalid
Leu200	11.25%	1.19%	11.31%	Invalid
ILe359	3.54%	0.85%	5.25%	Invalid
ILe344	0.35%	0.14%	0.94%	**Valid**
ILe302	6.72%	1.15%	5.42%	Invalid
ILe274	6.71%	1.36%	5.12%	Invalid
ILe200	9.91%	1.08%	8.69%	Invalid
Arg516	0.00%	0.00%	0.00%	Invalid
Arg501	0.00%	0.00%	0.00%	Invalid
Arg459	0.00%	0.00%	0.00%	Invalid
Arg402	0.00%	0.00%	0.00%	Invalid
Arg357	0.00%	0.00%	0.00%	Invalid
Arg317	0.00%	0.00%	0.00%	Invalid
Arg300	0.00%	0.00%	0.00%	Invalid
Arg288	0.00%	0.00%	0.00%	Invalid
Arg273	0.00%	0.00%	0.00%	Invalid
Arg231	0.00%	0.00%	0.00%	Invalid
Gln488	0.41%	0.32%	38.68%	**Valid**
Gln473	0.65%	0.43%	27.54%	**Valid**
Gln431	5.60%	3.75%	20.23%	Invalid
Gln403	9.29%	4.43%	29.82%	Invalid
Gln358	3.22%	1.17%	24.03%	Invalid
Gln329	1.24%	0.23%	39.16%	Invalid
Gln300	22.35%	0.53%	39.84%	Invalid
Gln272	4.50%	0.54%	36.08%	Invalid
Suc303	0.00%	0.00%	2.68%	Invalid
Suc261	0.00%	0.00%	24.57%	Invalid
Lac318	0.00%	0.00%	19.37%	Invalid
Lac303	0.00%	0.00%	30.65%	Invalid
Lac261	0.00%	0.00%	12.44%	Invalid
Lac233	0.00%	0.00%	8.56%	Invalid
Lac189	0.00%	0.00%	5.05%	Invalid

GA0.32

Table S4: All TBDMS-derivtized amino acid mass fragments for the GA0.32 ILE and their calculated percentage ^{13}C incorporation after naturally abundant ^{13}C isotope correction. Mass fragments are determined as either 'Valid', 'Conditionally Valid' or 'Invalid'.

GA0.32				
metabolite	AA standar	Standard d	12C Meta-	AA standar
Ala317	3.65%	0.75%	3.81%	Invalid
Ala302	0.20%	0.09%	0.20%	**Valid**
Ala260	12.96%	3.40%	9.85%	Invalid
Ala232	20.87%	5.22%	16.04%	Invalid
Gly303	8.07%	1.37%	7.37%	Invalid
Gly288	0.22%	0.17%	0.22%	**Valid**
Gly246	16.87%	5.38%	10.22%	Invalid
Gly218	24.89%	6.86%	15.57%	Invalid
Val345	14.74%	5.01%	12.90%	Invalid
Val330	0.50%	0.17%	0.59%	**Valid**
Val288	7.21%	2.12%	16.27%	Invalid
Val260	6.59%	3.90%	2.62%	Invalid
Val186	15.73%	5.52%	10.39%	Invalid
Pro343	4.58%	1.01%	12.08%	Invalid
Pro328	1.18%	0.41%	3.43%	**Valid**
Pro301	8.91%	3.05%	19.21%	Invalid
Pro286	3.57%	2.13%	2.68%	Invalid
Pro258	5.42%	3.32%	0.88%	Invalid
Pro184	12.30%	3.42%	6.04%	Invalid
Met377	0.25%	0.08%	12.80%	**Valid**
Met302	18.82%	0.18%	15.49%	Invalid
Met292	8.48%	3.48%	2.19%	Invalid
Met218	9.01%	3.18%	2.31%	Invalid
Phe393	32.66%	8.57%	44.64%	Invalid
Phe378	1.08%	0.40%	15.68%	**Valid**
Phe336	1.32%	0.85%	10.68%	Invalid
Phe308	1.60%	0.64%	10.30%	Invalid
Phe234	1.62%	1.02%	9.59%	Invalid
Ser447	3.86%	1.86%	4.31%	Invalid
Ser432	0.20%	0.17%	0.05%	**Valid**
Ser390	11.99%	7.63%	2.28%	Invalid
Ser362	16.39%	11.08%	2.64%	Invalid
Ser288	16.48%	10.19%	3.77%	Invalid
Thr461	1.97%	5.58%	3.07%	Invalid
Thr446	0.17%	0.49%	1.39%	**Valid**
Thr404	0.01%	0.02%	2.54%	**Valid**
Thr376	10.60%	29.98%	10.05%	Invalid
Thr302	0.00%	0.00%	1.07%	Invalid
Lys488	0.10%	0.05%	0.13%	**Valid**
Lys473	1.05%	0.44%	1.24%	**Valid**
Lys473	1.05%	0.44%	1.24%	**Valid**
Lys431	0.20%	0.09%	0.24%	**Valid**
Lys403	9.56%	3.40%	8.53%	Invalid
Lys329	0.86%	0.31%	0.89%	**Valid**
Lys300	4.89%	1.34%	4.33%	Invalid
Cys406	11.87%	5.41%	24.36%	Invalid
Cys378	43.43%	17.24%	37.92%	Invalid
Cys304	5.44%	9.17%	17.58%	Invalid
Tyr523	15.66%	4.57%	44.41%	Invalid
Tyr508	0.34%	0.18%	9.34%	**Valid**
Tyr466	1.76%	0.99%	1.26%	Invalid
Tyr438	1.97%	1.05%	2.23%	Invalid
Tyr364	1.03%	0.47%	2.02%	**Valid**
Tyr221	4.26%	0.74%	6.40%	Invalid
Trp546	9.58%	3.05%	6.93%	Invalid
Trp531	53.56%	2.80%	47.42%	Invalid
Trp531	53.56%	2.80%	47.42%	Invalid
Trp489	41.82%	7.25%	49.75%	Invalid
Trp461	50.39%	3.26%	59.30%	Invalid
Trp388	16.61%	2.44%	52.68%	Invalid
Trp330	57.78%	8.29%	35.53%	Invalid
Trp302	6.31%	3.69%	17.50%	Invalid
Trp244	17.06%	2.14%	55.60%	Invalid
His497	3.73%	1.35%	33.20%	Invalid
His482	0.55%	0.13%	6.37%	**Valid**
His440	4.06%	2.80%	0.71%	Invalid
His412	4.43%	1.85%	11.02%	Invalid
His338	16.61%	1.46%	3.04%	Invalid
Glu489	0.24%	0.09%	0.32%	**Valid**
Glu474	0.16%	0.05%	0.22%	**Valid**
Glu432	7.25%	3.10%	7.59%	Invalid
Glu404	0.05%	0.07%	0.08%	**Valid**
Glu330	3.88%	2.30%	3.86%	Invalid
Glu302	1.84%	0.13%	2.42%	Invalid
Glu187	46.19%	0.18%	45.83%	Invalid
Asp475	3.21%	1.81%	2.40%	Invalid
Asp460	0.37%	0.28%	0.22%	**Valid**
Asp418	4.96%	4.04%	7.12%	Invalid
Asp390	2.94%	2.40%	1.95%	Invalid
Asp316	5.24%	3.16%	4.89%	Invalid
Asp302	6.06%	5.53%	9.75%	Invalid
Leu359	3.77%	1.26%	4.30%	Invalid
Leu344	0.27%	0.14%	0.28%	**Valid**
Leu302	4.01%	2.16%	3.15%	Invalid
Leu274	5.14%	2.70%	4.20%	Invalid
Leu200	9.27%	2.63%	8.50%	Invalid
ILe359	3.31%	0.94%	6.29%	Invalid
ILe344	0.33%	0.14%	0.54%	**Valid**
ILe302	5.39%	2.50%	1.53%	Invalid
ILe274	5.28%	2.66%	1.25%	Invalid
ILe200	8.49%	2.60%	4.24%	Invalid
Arg516	0.00%	0.00%	0.00%	Invalid
Arg501	0.00%	0.00%	0.00%	Invalid
Arg459	0.00%	0.00%	0.00%	Invalid
Arg402	0.00%	0.00%	0.00%	Invalid
Arg357	0.00%	0.00%	0.00%	Invalid
Arg317	0.00%	0.00%	0.00%	Invalid
Arg300	0.00%	0.00%	0.00%	Invalid
Arg288	0.00%	0.00%	0.00%	Invalid
Arg273	0.00%	0.00%	0.00%	Invalid
Arg231	0.00%	0.00%	0.00%	Invalid
Gln488	0.39%	0.33%	37.33%	**Valid**
Gln473	0.57%	0.50%	34.51%	**Valid**
Gln431	4.54%	3.10%	15.31%	Invalid
Gln403	8.30%	5.20%	25.53%	Invalid
Gln358	3.17%	1.20%	28.53%	Invalid
Gln329	1.07%	0.13%	38.67%	**Valid**
Gln300	22.38%	0.52%	39.32%	Invalid
Gln272	4.48%	0.56%	32.05%	Invalid
Suc303	0.00%	0.00%	2.43%	Invalid
Suc261	0.00%	0.00%	24.70%	Invalid
Lac318	0.00%	0.00%	39.47%	Invalid
Lac303	0.00%	0.00%	20.39%	Invalid
Lac261	0.00%	0.00%	9.09%	Invalid
Lac233	0.00%	0.00%	7.70%	Invalid
Lac189	0.00%	0.00%	7.11%	Invalid

GMA0.075

Table S5: All TBDMS-derivtized amino acid mass fragments for the GMA0.075 ILE and their calculated percentage ^{13}C incorporation after naturally abundant ^{13}C isotope correction. Mass fragments are determined as either 'Valid', 'Conditionally Valid' or 'Invalid'.

GMA0.075						
AA standa	Standard d	12C Meta	AA standard fragment accur		Standard-	12C fragm
0.00%	0.00%	2.82%	Invalid	Conditonal valid	0.00%	Invalid
0.00%	0.00%	0.15%	Invalid	Conditonal valid	0.00%	**Valid**
0.00%	0.00%	6.11%	Invalid	Conditonal valid	0.00%	Invalid
0.00%	0.00%	10.09%	Invalid	Conditonal valid	0.00%	Invalid
0.00%	0.00%	7.09%	Invalid	Conditonal valid	0.00%	Invalid
0.00%	0.00%	0.41%	Invalid	Conditonal valid	0.00%	**Valid**
0.00%	0.00%	5.20%	Invalid	Conditonal valid	0.00%	Invalid
0.00%	0.00%	7.85%	Invalid	Conditonal valid	0.00%	Invalid
0.00%	0.00%	10.74%	Invalid	Conditonal valid	0.00%	Invalid
0.00%	0.00%	0.65%	Invalid	Conditonal valid	0.00%	**Valid**
0.00%	0.00%	1.51%	Invalid	Conditonal valid	0.00%	Invalid
0.00%	0.00%	2.90%	Invalid	Conditonal valid	0.00%	Invalid
0.00%	0.00%	11.35%	Invalid	Conditonal valid	0.00%	Invalid
0.00%	0.00%	23.19%	Invalid	Conditonal valid	0.00%	Invalid
0.00%	0.00%	8.11%	Invalid	Conditonal valid	0.00%	Invalid
0.00%	0.00%	33.44%	Invalid	Conditonal valid	0.00%	Invalid
0.00%	0.00%	5.43%	Invalid	Conditonal valid	0.00%	Invalid
0.00%	0.00%	0.77%	Invalid	Conditonal valid	0.00%	**Valid**
0.00%	0.00%	5.23%	Invalid	Conditonal valid	0.00%	Invalid
0.00%	0.00%	16.81%	Invalid	Conditonal valid	0.00%	Invalid
0.00%	0.00%	16.31%	Invalid	Conditonal valid	0.00%	Invalid
0.00%	0.00%	36.79%	Invalid	Conditonal valid	0.00%	Invalid
0.00%	0.00%	26.11%	Invalid	Conditonal valid	0.00%	Invalid
0.00%	0.00%	46.80%	Invalid	Conditonal valid	0.00%	Invalid
0.00%	0.00%	5.10%	Invalid	Conditonal valid	0.00%	Invalid
0.00%	0.00%	0.79%	Invalid	Conditonal valid	0.00%	**Valid**
0.00%	0.00%	0.71%	Invalid	Conditonal valid	0.00%	**Valid**
0.00%	0.00%	0.44%	Invalid	Conditonal valid	0.00%	**Valid**
0.00%	0.00%	3.59%	Invalid	Conditonal valid	0.00%	Invalid
0.00%	0.00%	0.12%	Invalid	Conditonal valid	0.00%	**Valid**
0.00%	0.00%	1.25%	Invalid	Conditonal valid	0.00%	Invalid
0.00%	0.00%	1.42%	Invalid	Conditonal valid	0.00%	Invalid
0.00%	0.00%	1.89%	Invalid	Conditonal valid	0.00%	Invalid
0.00%	0.00%	23.37%	Invalid	Conditonal valid	0.00%	Invalid
0.00%	0.00%	1.87%	Invalid	Conditonal valid	0.00%	Invalid
0.00%	0.00%	2.74%	Invalid	Conditonal valid	0.00%	Invalid
0.00%	0.00%	89.76%	Invalid	Conditonal valid	0.00%	Invalid
0.00%	0.00%	1.07%	Invalid	Conditonal valid	0.00%	**Valid**
0.00%	0.00%	0.16%	Invalid	Conditonal valid	0.00%	**Valid**
0.00%	0.00%	2.53%	Invalid	Conditonal valid	0.00%	Invalid
0.00%	0.00%	2.53%	Invalid	Conditonal valid	0.00%	Invalid
0.00%	0.00%	0.23%	Invalid	Conditonal valid	0.00%	**Valid**
0.00%	0.00%	12.27%	Invalid	Conditonal valid	0.00%	Invalid
0.00%	0.00%	0.78%	Invalid	Conditonal valid	0.00%	**Valid**
0.00%	0.00%	4.66%	Invalid	Conditonal valid	0.00%	Invalid
0.00%	0.00%	2.22%	Invalid	Conditonal valid	0.00%	Invalid
0.00%	0.00%	3.32%	Invalid	Conditonal valid	0.00%	Invalid
0.00%	0.00%	0.00%	Invalid	Conditonal valid	0.00%	Invalid
0.00%	0.00%	37.86%	Invalid	Conditonal valid	0.00%	Invalid
0.00%	0.00%	0.87%	Invalid	Conditonal valid	0.00%	**Valid**
0.00%	0.00%	0.40%	Invalid	Conditonal valid	0.00%	**Valid**
0.00%	0.00%	0.88%	Invalid	Conditonal valid	0.00%	**Valid**
0.00%	0.00%	1.04%	Invalid	Conditonal valid	0.00%	**Valid**
0.00%	0.00%	4.57%	Invalid	Conditonal valid	0.00%	Invalid
0.00%	0.00%	0.56%	Invalid	Conditonal valid	0.00%	**Valid**
0.00%	0.00%	5.48%	Invalid	Conditonal valid	0.00%	Invalid
0.00%	0.00%	5.48%	Invalid	Conditonal valid	0.00%	Invalid
0.00%	0.00%	3.91%	Invalid	Conditonal valid	0.00%	Invalid
0.00%	0.00%	6.26%	Invalid	Conditonal valid	0.00%	Invalid
0.00%	0.00%	4.55%	Invalid	Conditonal valid	0.00%	Invalid
0.00%	0.00%	3.01%	Invalid	Conditonal valid	0.00%	Invalid
0.00%	0.00%	1.53%	Invalid	Conditonal valid	0.00%	Invalid
0.00%	0.00%	5.97%	Invalid	Conditonal valid	0.00%	Invalid
0.00%	0.00%	35.12%	Invalid	Conditonal valid	0.00%	Invalid
0.00%	0.00%	4.25%	Invalid	Conditonal valid	0.00%	Invalid
0.00%	0.00%	0.54%	Invalid	Conditonal valid	0.00%	**Valid**
0.00%	0.00%	21.99%	Invalid	Conditonal valid	0.00%	Invalid
0.00%	0.00%	10.46%	Invalid	Conditonal valid	0.00%	Invalid
0.00%	0.00%	0.49%	Invalid	Conditonal valid	0.00%	**Valid**
0.00%	0.00%	0.42%	Invalid	Conditonal valid	0.00%	**Valid**
0.00%	0.00%	3.91%	Invalid	Conditonal valid	0.00%	Invalid
0.00%	0.00%	0.11%	Invalid	Conditonal valid	0.00%	**Valid**
0.00%	0.00%	1.40%	Invalid	Conditonal valid	0.00%	Invalid
0.00%	0.00%	3.00%	Invalid	Conditonal valid	0.00%	Invalid
0.00%	0.00%	45.41%	Invalid	Conditonal valid	0.00%	Invalid
0.00%	0.00%	7.66%	Invalid	Conditonal valid	0.00%	Invalid
0.00%	0.00%	0.20%	Invalid	Conditonal valid	0.00%	**Valid**
0.00%	0.00%	2.48%	Invalid	Conditonal valid	0.00%	Invalid
0.00%	0.00%	0.70%	Invalid	Conditonal valid	0.00%	**Valid**
0.00%	0.00%	1.91%	Invalid	Conditonal valid	0.00%	Invalid
0.00%	0.00%	4.34%	Invalid	Conditonal valid	0.00%	Invalid
0.00%	0.00%	3.50%	Invalid	Conditonal valid	0.00%	Invalid
0.00%	0.00%	0.37%	Invalid	Conditonal valid	0.00%	**Valid**
0.00%	0.00%	2.27%	Invalid	Conditonal valid	0.00%	Invalid
0.00%	0.00%	2.94%	Invalid	Conditonal valid	0.00%	Invalid
0.00%	0.00%	7.54%	Invalid	Conditonal valid	0.00%	Invalid
0.00%	0.00%	3.63%	Invalid	Conditonal valid	0.00%	Invalid
0.00%	0.00%	0.46%	Invalid	Conditonal valid	0.00%	**Valid**
0.00%	0.00%	2.20%	Invalid	Conditonal valid	0.00%	Invalid
0.00%	0.00%	1.55%	Invalid	Conditonal valid	0.00%	Invalid
0.00%	0.00%	5.28%	Invalid	Conditonal valid	0.00%	Invalid
0.00%	0.00%	0.00%	Invalid	Conditonal valid	0.00%	Invalid
0.00%	0.00%	0.00%	Invalid	Conditonal valid	0.00%	Invalid
0.00%	0.00%	0.00%	Invalid	Conditonal valid	0.00%	Invalid
0.00%	0.00%	0.00%	Invalid	Conditonal valid	0.00%	Invalid
0.00%	0.00%	0.00%	Invalid	Conditonal valid	0.00%	Invalid
0.00%	0.00%	0.00%	Invalid	Conditonal valid	0.00%	Invalid
0.00%	0.00%	0.00%	Invalid	Conditonal valid	0.00%	Invalid
0.00%	0.00%	0.00%	Invalid	Conditonal valid	0.00%	Invalid
0.00%	0.00%	0.00%	Invalid	Conditonal valid	0.00%	Invalid
0.00%	0.00%	44.64%	Invalid	Conditonal valid	0.00%	Invalid
0.00%	0.00%	53.93%	Invalid	Conditonal valid	0.00%	Invalid
0.00%	0.00%	18.85%	Invalid	Conditonal valid	0.00%	Invalid
0.00%	0.00%	27.79%	Invalid	Conditonal valid	0.00%	Invalid
0.00%	0.00%	25.38%	Invalid	Conditonal valid	0.00%	Invalid
0.00%	0.00%	33.58%	Invalid	Conditonal valid	0.00%	Invalid
0.00%	0.00%	38.93%	Invalid	Conditonal valid	0.00%	Invalid
0.00%	0.00%	36.79%	Invalid	Conditonal valid	0.00%	Invalid
0.00%	0.00%	8.78%	Invalid	Conditonal valid	0.00%	Invalid
0.00%	0.00%	21.27%	Invalid	Conditonal valid	0.00%	Invalid
0.00%	0.00%	40.80%	Invalid	Conditonal valid	0.00%	Invalid
0.00%	0.00%	1.46%	Invalid	Conditonal valid	0.00%	Invalid
0.00%	0.00%	0.19%	Invalid	Conditonal valid	0.00%	**Valid**
0.00%	0.00%	0.12%	Invalid	Conditonal valid	0.00%	**Valid**
0.00%	0.00%	4.69%	Invalid	Conditonal valid	0.00%	Invalid

GAn0.075

Table S6: All TBDMS-derivtized amino acid mass fragments for the GAn0.075 ILE and their calculated percentage ^{13}C incorporation after naturally abundant ^{13}C isotope correction. Mass fragments are determined as either 'Valid', 'Conditionally Valid' or 'Invalid'.

GAn0.075					
metabolite	AA standa	Standard d	12C Meta-	AA standard fragment accura	
Ala317	4.56%	2.21%	4.60%	Invalid	Invalid
Ala302	1.33%	1.18%	1.10%	Invalid	Conditonal valid
Ala260	12.62%	7.42%	12.32%	Invalid	Invalid
Ala232	19.97%	10.63%	19.53%	Invalid	Invalid
Gly303	5.37%	6.23%	6.87%	Invalid	Conditonal valid
Gly288	1.37%	3.13%	1.34%	Invalid	Conditonal valid
Gly246	12.30%	14.08%	11.32%	Invalid	Conditonal valid
Gly218	17.64%	12.42%	16.48%	Invalid	Invalid
Val345	18.39%	8.48%	17.27%	Invalid	Invalid
Val330	1.66%	7.67%	1.50%	Invalid	Conditonal valid
Val288	20.34%	3.32%	20.22%	Invalid	Invalid
Val260	6.61%	4.77%	6.22%	Invalid	Invalid
Val186	15.68%	8.74%	15.27%	Invalid	Invalid
Pro343	23.07%	3.87%	23.93%	Invalid	Invalid
Pro328	7.98%	2.62%	9.78%	Invalid	Invalid
Pro301	35.72%	13.84%	39.67%	Invalid	Invalid
Pro286	6.94%	4.15%	9.26%	Invalid	Invalid
Pro258	3.20%	4.25%	3.73%	Invalid	Conditonal valid
Pro184	7.72%	6.86%	7.43%	Invalid	Conditonal valid
Met377	37.45%	3.21%	38.74%	Invalid	Invalid
Met302	22.54%	7.87%	20.40%	Invalid	Invalid
Met292	39.07%	5.99%	39.87%	Invalid	Invalid
Met218	35.56%	4.71%	35.33%	Invalid	Invalid
Phe393	43.99%	15.52%	44.33%	Invalid	Invalid
Phe378	7.22%	14.03%	6.71%	Invalid	Conditonal valid
Phe336	1.88%	1.48%	1.79%	Invalid	Conditonal valid
Phe308	2.28%	1.45%	2.36%	Invalid	Conditonal valid
Phe234	1.71%	1.57%	1.66%	Invalid	Conditonal valid
Ser447	5.30%	1.97%	4.91%	Invalid	Invalid
Ser432	1.20%	0.84%	1.10%	**Valid**	Conditonal valid
Ser390	3.75%	9.02%	3.41%	Invalid	Conditonal valid
Ser362	3.51%	11.92%	3.39%	Invalid	Conditonal valid
Ser288	5.49%	11.29%	5.13%	Invalid	Conditonal valid
Thr461	0.00%	4.00%	3.07%	Invalid	Conditonal valid
Thr446	0.00%	0.00%	1.39%	Invalid	Conditonal valid
Thr404	0.00%	0.00%	2.54%	Invalid	Conditonal valid
Thr376	0.00%	0.00%	10.05%	Invalid	Conditonal valid
Thr302	0.00%	0.00%	1.07%	**Valid**	Conditonal valid
Lys488	1.25%	0.52%	1.11%	**Valid**	Conditonal valid
Lys473	1.91%	0.62%	1.95%	Invalid	Invalid
Lys473	1.91%	0.62%	1.95%	Invalid	Invalid
Lys431	1.35%	0.69%	1.20%	**Valid**	Conditonal valid
Lys403	12.30%	3.47%	13.91%	Invalid	Invalid
Lys329	1.67%	2.18%	1.62%	Invalid	Conditonal valid
Lys300	5.36%	3.79%	5.28%	Invalid	Invalid
Cys406	25.43%	5.62%	23.90%	Invalid	Invalid
Cys378	34.65%	27.05%	32.84%	Invalid	Invalid
Cys304	0.63%	22.09%	1.36%	**Valid**	Conditonal valid
Tyr523	44.75%	6.48%	43.95%	Invalid	Invalid
Tyr508	3.56%	6.26%	3.51%	Invalid	Conditonal valid
Tyr466	1.97%	1.12%	1.84%	Invalid	Conditonal valid
Tyr438	3.04%	0.59%	2.95%	Invalid	Invalid
Tyr364	2.30%	0.70%	2.23%	Invalid	Invalid
Tyr221	7.49%	1.68%	7.35%	Invalid	Invalid
Trp546	9.14%	4.17%	9.00%	Invalid	Invalid
Trp531	52.12%	20.82%	50.88%	Invalid	Invalid
Trp531	52.12%	17.45%	50.88%	Invalid	Invalid
Trp489	39.05%	15.48%	38.94%	Invalid	Invalid
Trp461	56.98%	17.50%	55.53%	Invalid	Invalid
Trp388	50.71%	14.94%	49.93%	Invalid	Invalid
Trp330	31.23%	13.67%	31.16%	Invalid	Invalid
Trp302	21.27%	10.55%	20.96%	Invalid	Invalid
Trp244	55.57%	6.41%	55.80%	Invalid	Invalid
His497	17.38%	5.56%	18.99%	Invalid	Invalid
His482	3.58%	1.33%	3.79%	Invalid	Invalid
His440	1.49%	2.98%	1.45%	Invalid	Conditonal valid
His412	8.64%	1.84%	9.52%	Invalid	Invalid
His338	3.21%	2.07%	3.36%	Invalid	Conditonal valid
Glu489	1.64%	0.59%	1.46%	Invalid	Conditonal valid
Glu474	1.31%	0.40%	1.26%	Invalid	Conditonal valid
Glu432	8.38%	4.91%	8.15%	Invalid	Invalid
Glu404	1.36%	1.49%	1.27%	Invalid	Conditonal valid
Glu330	4.60%	3.41%	4.41%	Invalid	Conditonal valid
Glu302	3.92%	0.97%	3.78%	Invalid	Invalid
Glu187	46.53%	18.71%	46.34%	Invalid	Invalid
Asp475	4.28%	13.69%	4.41%	Invalid	Conditonal valid
Asp460	1.45%	1.41%	1.24%	Invalid	Conditonal valid
Asp418	7.30%	4.29%	6.79%	Invalid	Invalid
Asp390	2.06%	2.35%	2.16%	Invalid	Conditonal valid
Asp316	4.98%	3.51%	4.60%	Invalid	Invalid
Asp302	9.52%	5.50%	8.95%	Invalid	Invalid
Leu359	7.61%	1.81%	6.88%	Invalid	Invalid
Leu344	1.35%	1.70%	1.21%	Invalid	Conditonal valid
Leu302	5.19%	2.92%	4.95%	Invalid	Invalid
Leu274	6.74%	3.55%	6.34%	Invalid	Invalid
Leu200	10.48%	4.87%	10.21%	Invalid	Invalid
ILe359	11.99%	1.77%	11.51%	Invalid	Invalid
ILe344	1.61%	1.09%	1.47%	Invalid	Conditonal valid
ILe302	4.91%	4.09%	4.65%	Invalid	Conditonal valid
ILe274	4.34%	3.96%	4.06%	Invalid	Conditonal valid
ILe200	8.22%	5.06%	7.86%	Invalid	Invalid
Arg516	0.00%	3.01%	0.00%	Invalid	Conditonal valid
Arg501	0.00%	0.00%	0.00%	Invalid	Conditonal valid
Arg459	0.00%	0.00%	0.00%	Invalid	Conditonal valid
Arg402	0.00%	0.00%	0.00%	Invalid	Conditonal valid
Arg357	0.00%	0.00%	0.00%	Invalid	Conditonal valid
Arg317	0.00%	0.00%	0.00%	Invalid	Conditonal valid
Arg300	0.00%	0.00%	0.00%	Invalid	Conditonal valid
Arg288	0.00%	0.00%	0.00%	Invalid	Conditonal valid
Arg273	0.00%	0.00%	0.00%	Invalid	Conditonal valid
Arg231	0.00%	0.00%	0.00%	Invalid	Conditonal valid
Gln488	39.69%	0.39%	40.15%	Invalid	Invalid
Gln473	46.44%	0.39%	47.35%	Invalid	Invalid
Gln431	22.52%	4.47%	21.49%	Invalid	Invalid
Gln403	32.33%	3.24%	31.26%	Invalid	Invalid
Gln358	35.98%	2.29%	34.53%	Invalid	Invalid
Gln329	43.57%	1.25%	41.97%	Invalid	Invalid
Gln300	38.74%	9.18%	42.01%	Invalid	Invalid
Gln272	38.64%	6.01%	36.77%	Invalid	Invalid
Suc303	22.00%	1.91%	21.91%	Invalid	Invalid
Suc261	15.57%	0.00%	14.86%	Invalid	Invalid
Lac318	33.63%	0.00%	36.05%	Invalid	Invalid
Lac303	1.37%	0.00%	1.68%	Invalid	Invalid
Lac261	14.06%	0.00%	12.18%	Invalid	Invalid
Lac233	11.57%	0.00%	11.23%	Invalid	Invalid
Lac189	13.56%	0.00%	13.25%	Invalid	Invalid

XA0.075

Table S6: All TBDMS-derivtized amino acid mass fragments for the XA0.075 ILE and their calculated percentage ^{13}C incorporation after naturally abundant ^{13}C isotope correction. Mass fragments are determined as either 'Valid', 'Conditionally Valid' or 'Invalid'.

XA0.075						
metabolite	AA standa	Standard d	12C Meta	AA standard fragment accura		Standard-
Ala317	3.40%	0.51%	2.57%	Invalid	Invalid	
Ala302	0.14%	0.08%	0.21%	**Valid**	Conditonal valid	**0.06%**
Ala260	7.60%	2.43%	11.69%	Invalid	Invalid	
Ala232	12.93%	3.77%	18.46%	Invalid	Invalid	
Gly303	6.31%	1.90%	6.22%	Invalid	Invalid	
Gly288	0.78%	1.80%	0.28%	**Valid**	Conditonal valid	-1.02%
Gly246	14.35%	6.36%	14.34%	Invalid	Invalid	
Gly218	26.07%	4.49%	21.11%	Invalid	Invalid	
Val345	7.64%	2.52%	5.14%	Invalid	Invalid	
Val330	0.22%	0.08%	0.17%	**Valid**	Conditonal valid	**0.14%**
Val288	5.68%	1.15%	10.86%	Invalid	Invalid	
Val260	4.92%	1.95%	7.16%	Invalid	Invalid	
Val186	13.47%	2.61%	14.81%	Invalid	Invalid	
Pro343	5.61%	1.34%	7.99%	Invalid	Invalid	
Pro328	1.25%	0.33%	2.12%	Invalid	Conditonal valid	**0.92%**
Pro301	10.44%	1.76%	19.42%	Invalid	Invalid	
Pro286	1.66%	0.82%	3.49%	Invalid	Conditonal valid	**0.84%**
Pro258	2.74%	1.50%	3.15%	Invalid	Invalid	
Pro184	9.62%	1.90%	8.76%	Invalid	Invalid	
Met377	0.33%	0.13%	28.72%	**Valid**	Conditonal valid	**0.20%**
Met302	19.09%	0.10%	12.27%	Invalid	Invalid	
Met292	2.90%	1.94%	9.19%	Invalid	Conditonal valid	**0.95%**
Met218	3.33%	1.85%	7.45%	Invalid	Invalid	
Phe393	27.47%	4.20%	24.65%	Invalid	Invalid	
Phe378	0.74%	0.31%	2.03%	**Valid**	Conditonal valid	**0.43%**
Phe336	0.77%	0.46%	0.96%	**Valid**	Conditonal valid	**0.31%**
Phe308	0.61%	0.43%	0.99%	**Valid**	Conditonal valid	**0.18%**
Phe234	1.00%	0.58%	1.14%	**Valid**	Conditonal valid	**0.43%**
Ser447	3.42%	1.03%	2.45%	Invalid	Invalid	
Ser432	0.10%	0.08%	0.12%	**Valid**	Conditonal valid	**0.02%**
Ser390	9.58%	3.02%	7.49%	Invalid	Invalid	
Ser362	12.08%	4.55%	9.37%	Invalid	Invalid	
Ser288	12.12%	4.01%	9.88%	Invalid	Invalid	
Thr461	6.70%	3.51%	7.86%	Invalid	Invalid	
Thr446	0.23%	0.20%	1.50%	**Valid**	Conditonal valid	**0.03%**
Thr404	0.09%	0.07%	3.09%	**Valid**	Conditonal valid	**0.02%**
Thr376	94.20%	3.04%	65.36%	Invalid	Invalid	
Thr302	14.38%	0.05%	11.63%	Invalid	Invalid	
Lys488	0.07%	0.02%	0.18%	**Valid**	Conditonal valid	**0.05%**
Lys473	0.95%	0.24%	0.54%	**Valid**	Conditonal valid	**0.71%**
Lys473	0.95%	0.24%	0.54%	**Valid**	Conditonal valid	**0.71%**
Lys431	0.13%	0.04%	0.28%	**Valid**	Conditonal valid	**0.09%**
Lys403	8.56%	3.33%	4.39%	Invalid	Invalid	
Lys329	0.52%	0.14%	0.62%	**Valid**	Conditonal valid	**0.38%**
Lys300	4.36%	0.43%	5.40%	Invalid	Invalid	
Cys406	8.03%	4.12%	15.46%	Invalid	Invalid	
Cys378	42.62%	23.50%	24.38%	Invalid	Invalid	
Cys304	7.47%	10.72%	0.00%	Invalid	Conditonal valid	-3.26%
Tyr523	13.24%	2.62%	23.34%	Invalid	Invalid	
Tyr508	0.26%	0.13%	1.66%	**Valid**	Conditonal valid	**0.13%**
Tyr466	0.43%	0.14%	0.46%	**Valid**	Conditonal valid	**0.29%**
Tyr438	1.05%	0.77%	1.53%	**Valid**	Conditonal valid	**0.28%**
Tyr364	0.42%	0.06%	0.61%	**Valid**	Conditonal valid	**0.37%**
Tyr221	3.70%	0.61%	3.61%	Invalid	Invalid	
Trp546	8.29%	3.64%	4.33%	Invalid	Invalid	
Trp531	59.04%	2.87%	36.82%	Invalid	Invalid	
Trp531	59.04%	2.87%	36.82%	Invalid	Invalid	
Trp489	33.46%	13.41%	29.92%	Invalid	Invalid	
Trp461	55.93%	1.54%	45.36%	Invalid	Invalid	
Trp388	13.23%	0.95%	35.14%	Invalid	Invalid	
Trp330	49.05%	16.22%	29.00%	Invalid	Invalid	
Trp302	7.83%	2.85%	12.24%	Invalid	Invalid	
Trp244	16.29%	0.89%	40.99%	Invalid	Invalid	
His497	3.30%	1.56%	10.65%	Invalid	Invalid	
His482	0.52%	0.31%	1.56%	**Valid**	Conditonal valid	**0.22%**
His440	1.98%	1.68%	0.87%	Invalid	Conditonal valid	**0.30%**
His412	3.72%	1.98%	3.47%	Invalid	Invalid	
His338	13.23%	0.55%	1.66%	Invalid	Invalid	
Glu489	0.25%	0.07%	0.16%	**Valid**	Conditonal valid	**0.17%**
Glu474	0.16%	0.05%	0.17%	**Valid**	Conditonal valid	**0.11%**
Glu432	5.55%	1.85%	9.22%	Invalid	Invalid	
Glu404	0.01%	0.01%	0.14%	**Valid**	Conditonal valid	**0.00%**
Glu330	1.98%	1.24%	5.71%	Invalid	Conditonal valid	**0.74%**
Glu302	1.96%	0.13%	1.63%	Invalid	Invalid	
Glu187	46.41%	1.16%	35.95%	Invalid	Invalid	
Asp475	2.99%	0.90%	1.54%	Invalid	Invalid	
Asp460	0.14%	0.09%	0.17%	**Valid**	Conditonal valid	**0.05%**
Asp418	5.43%	2.41%	8.46%	Invalid	Invalid	
Asp390	1.17%	1.07%	4.07%	**Valid**	Conditonal valid	**0.11%**
Asp316	3.28%	1.78%	6.52%	Invalid	Invalid	
Asp302	7.36%	2.91%	10.82%	Invalid	Invalid	
Leu359	3.17%	1.49%	2.42%	Invalid	Invalid	
Leu344	0.18%	0.09%	0.15%	**Valid**	Conditonal valid	**0.10%**
Leu302	3.24%	1.54%	5.34%	Invalid	Invalid	
Leu274	4.20%	1.87%	6.63%	Invalid	Invalid	
Leu200	8.38%	1.74%	9.47%	Invalid	Invalid	
ILe359	2.14%	0.40%	2.42%	Invalid	Invalid	
ILe344	0.17%	0.05%	0.16%	**Valid**	Conditonal valid	**0.12%**
ILe302	4.66%	1.47%	5.20%	Invalid	Invalid	
ILe274	4.18%	1.63%	5.05%	Invalid	Invalid	
ILe200	7.59%	1.47%	7.46%	Invalid	Invalid	
Arg516	0.00%	0.00%	0.00%	Invalid	Conditonal valid	0.00%
Arg501	0.00%	0.00%	0.00%	Invalid	Conditonal valid	0.00%
Arg459	0.00%	0.00%	0.00%	Invalid	Conditonal valid	0.00%
Arg402	0.00%	0.00%	0.00%	Invalid	Conditonal valid	0.00%
Arg357	0.00%	0.00%	0.00%	Invalid	Conditonal valid	0.00%
Arg317	0.00%	0.00%	0.00%	Invalid	Conditonal valid	0.00%
Arg300	0.00%	0.00%	0.00%	Invalid	Conditonal valid	0.00%
Arg288	0.00%	0.00%	0.00%	Invalid	Conditonal valid	0.00%
Arg273	0.00%	0.00%	0.00%	Invalid	Conditonal valid	0.00%
Arg231	0.00%	0.00%	0.00%	Invalid	Conditonal valid	0.00%
Gln488	0.24%	0.08%	27.55%	**Valid**	Conditonal valid	**0.16%**
Gln473	0.21%	0.06%	26.31%	**Valid**	Conditonal valid	**0.15%**
Gln431	6.55%	1.77%	15.27%	Invalid	Invalid	
Gln403	5.28%	1.02%	21.02%	Invalid	Invalid	
Gln358	2.46%	0.07%	23.07%	Invalid	Invalid	
Gln329	0.97%	0.11%	34.82%	**Valid**	Conditonal valid	**0.86%**
Gln300	19.49%	4.56%	33.66%	Invalid	Invalid	
Gln272	3.85%	0.85%	26.75%	Invalid	Invalid	
Suc303	0.00%	0.00%	6.83%	Invalid	Conditonal valid	0.00%
Suc261	0.00%	0.00%	14.66%	Invalid	Conditonal valid	0.00%
Lac318	0.00%	0.00%	51.30%	Invalid	Conditonal valid	0.00%
Lac303	0.00%	0.00%	1.22%	Invalid	Conditonal valid	0.00%
Lac261	0.00%	0.00%	0.47%	Invalid	Conditonal valid	0.00%
Lac233	0.00%	0.00%	0.16%	Invalid	Conditonal valid	0.00%
Lac189	0.00%	0.00%	3.89%	Invalid	Conditonal valid	0.00%

XA0.15

Table S7: All TBDMS-derivtized amino acid mass fragments for the XA0.15 ILE and their calculated percentage ^{13}C incorporation after naturally abundant ^{13}C isotope correction. Mass fragments are determined as either 'Valid', 'Conditionally Valid' or 'Invalid'.

XA0.15				
metabolite	AA standa	Standard d	12C Meta-	AA standa
Ala317	1.37%	1.62%	4.18%	Invalid
Ala302	0.07%	0.09%	0.21%	**Valid**
Ala260	5.68%	6.74%	15.55%	Invalid
Ala232	9.12%	10.73%	24.60%	Invalid
Gly303	2.91%	3.44%	7.27%	Invalid
Gly288	0.09%	0.12%	0.36%	**Valid**
Gly246	7.92%	9.59%	20.23%	Invalid
Gly218	11.38%	13.49%	29.28%	Invalid
Val345	2.45%	2.92%	10.34%	Invalid
Val330	0.12%	0.16%	0.36%	**Valid**
Val288	2.54%	2.93%	18.92%	Invalid
Val260	2.25%	3.08%	5.53%	Invalid
Val186	6.20%	7.30%	14.86%	Invalid
Pro343	3.03%	3.59%	9.07%	Invalid
Pro328	0.67%	0.80%	2.66%	**Valid**
Pro301	5.02%	5.87%	19.44%	Invalid
Pro286	0.86%	1.24%	3.46%	**Valid**
Pro258	1.23%	2.01%	3.59%	Invalid
Pro184	4.26%	5.10%	11.04%	Invalid
Met377	0.15%	0.18%	10.51%	**Valid**
Met302	8.46%	9.46%	18.22%	Invalid
Met292	2.24%	3.37%	1.86%	Invalid
Met218	2.54%	3.52%	1.73%	Invalid
Phe393	13.48%	15.80%	37.98%	Invalid
Phe378	0.35%	0.46%	5.53%	**Valid**
Phe336	0.33%	0.51%	0.98%	**Valid**
Phe308	0.25%	0.44%	1.28%	**Valid**
Phe234	0.41%	0.67%	1.03%	**Valid**
Ser447	1.52%	1.85%	4.44%	Invalid
Ser432	0.04%	0.06%	0.09%	**Valid**
Ser390	4.45%	5.74%	8.65%	Invalid
Ser362	5.70%	7.66%	10.77%	Invalid
Ser288	5.94%	7.64%	12.02%	Invalid
Thr461	13.10%	20.95%	21.11%	Invalid
Thr446	5.05%	9.41%	3.96%	Invalid
Thr404	3.42%	6.38%	5.91%	Invalid
Thr376	39.41%	44.36%	65.66%	Invalid
Thr302	0.00%	0.00%	1.07%	Invalid
Lys488	0.03%	0.04%	0.26%	**Valid**
Lys473	0.60%	0.82%	0.85%	**Valid**
Lys473	0.60%	0.82%	0.85%	**Valid**
Lys431	0.05%	0.06%	0.31%	**Valid**
Lys403	5.09%	6.17%	6.32%	Invalid
Lys329	0.42%	0.49%	0.94%	**Valid**
Lys300	1.93%	2.16%	6.30%	Invalid
Cys406	2.34%	2.69%	27.21%	Invalid
Cys378	16.48%	23.65%	40.61%	Invalid
Cys304	3.23%	9.12%	23.76%	Invalid
Tyr523	6.28%	7.23%	42.88%	Invalid
Tyr508	0.06%	0.08%	9.52%	**Valid**
Tyr466	0.14%	0.17%	1.28%	**Valid**
Tyr438	0.16%	0.19%	2.44%	**Valid**
Tyr364	0.21%	0.26%	2.46%	**Valid**
Tyr221	1.60%	1.79%	6.75%	Invalid
Trp546	2.97%	3.50%	6.32%	Invalid
Trp531	25.80%	29.32%	47.02%	Invalid
Trp531	25.80%	29.32%	47.02%	Invalid
Trp489	13.55%	15.48%	53.49%	Invalid
Trp461	23.65%	26.48%	60.97%	Invalid
Trp388	5.87%	6.61%	55.34%	Invalid
Trp330	13.66%	15.34%	42.96%	Invalid
Trp302	3.43%	4.81%	17.42%	Invalid
Trp244	9.09%	10.44%	57.71%	Invalid
His497	3.99%	5.50%	24.24%	Invalid
His482	0.31%	0.41%	6.26%	**Valid**
His440	0.40%	0.88%	1.04%	**Valid**
His412	2.77%	3.39%	8.45%	Invalid
His338	5.87%	1.08%	2.91%	Invalid
Glu489	0.06%	0.07%	0.33%	**Valid**
Glu474	0.07%	0.09%	0.25%	**Valid**
Glu432	3.15%	3.95%	12.90%	Invalid
Glu404	0.00%	0.00%	0.33%	**Valid**
Glu330	1.51%	2.19%	8.75%	Invalid
Glu302	0.80%	0.90%	2.10%	**Valid**
Glu187	20.82%	23.27%	46.23%	Invalid
Asp475	1.31%	1.65%	2.26%	Invalid
Asp460	0.03%	0.03%	0.22%	**Valid**
Asp418	3.53%	4.57%	12.68%	Invalid
Asp390	1.40%	2.32%	7.28%	Invalid
Asp316	2.60%	3.66%	10.62%	Invalid
Asp302	4.74%	5.97%	16.16%	Invalid
Leu359	0.81%	0.93%	5.55%	**Valid**
Leu344	0.07%	0.08%	0.24%	**Valid**
Leu302	1.81%	2.46%	5.84%	Invalid
Leu274	2.31%	3.09%	7.29%	Invalid
Leu200	4.19%	4.93%	11.13%	Invalid
Ile359	0.82%	0.92%	8.08%	**Valid**
Ile344	0.06%	0.07%	0.51%	**Valid**
Ile302	2.47%	3.06%	3.75%	Invalid
Ile274	2.36%	3.03%	3.14%	Invalid
Ile200	3.86%	4.50%	6.91%	Invalid
Arg516	0.00%	0.00%	0.00%	Invalid
Arg501	0.00%	0.00%	0.00%	Invalid
Arg459	0.00%	0.00%	0.00%	Invalid
Arg402	0.00%	0.00%	0.00%	Invalid
Arg357	0.00%	0.00%	0.00%	Invalid
Arg317	0.00%	0.00%	0.00%	Invalid
Arg300	0.00%	0.00%	0.00%	Invalid
Arg288	0.00%	0.00%	0.00%	Invalid
Arg273	0.00%	0.00%	0.00%	Invalid
Arg231	0.00%	0.00%	0.00%	Invalid
Gln488	0.15%	0.22%	30.39%	**Valid**
Gln473	0.37%	0.58%	27.59%	**Valid**
Gln431	3.15%	5.60%	7.48%	Invalid
Gln403	3.73%	6.14%	29.10%	Invalid
Gln358	1.20%	1.37%	33.21%	**Valid**
Gln329	1.29%	1.95%	35.49%	Invalid
Gln300	9.82%	10.99%	40.33%	Invalid
Gln272	2.42%	3.04%	31.91%	Invalid
Suc303	0.00%	0.00%	0.00%	Invalid
Suc261	0.00%	0.00%	0.00%	Invalid
Lac318	0.00%	0.00%	27.43%	Invalid
Lac303	0.00%	0.00%	2.06%	Invalid
Lac261	0.00%	0.00%	0.31%	Invalid
Lac233	0.00%	0.00%	0.17%	Invalid
Lac189	0.00%	0.00%	4.21%	Invalid

Xan0.075

Table S8: All TBDMS-derivtized amino acid mass fragments for the XAn0.075 ILE and their calculated percentage ^{13}C incorporation after naturally abundant ^{13}C isotope correction. Mass fragments are determined as either 'Valid', 'Conditionally Valid' or 'Invalid'.

Anaerobic Xylose		
Metabolite	Mean enri	Type of fra
Ala317	No data.	Invalid
Ala302	23.65%	Invalid
Ala260	2.00%	Invalid
Ala232	0.71%	Valid
Ala158	5.66%	Invalid
Gly303	No data.	Invalid
Gly288	3.66%	Valid
Gly246	6.51%	Invalid
Gly218	8.12%	Invalid
Gly144	34.68%	Invalid
Val345	No data.	Invalid
Val330	1.95%	Invalid
Val288	1.14%	Valid
Val260	2.20%	Invalid
Val186	1.73%	Invalid
Val302	5.85%	Invalid
Pro343	No data.	Invalid
Pro328	1.25%	Valid
Pro286	1.35%	Invalid
Pro258	2.02%	Invalid
Pro184	1.37%	Invalid
Pro301	No data.	Invalid
Met377	No data.	Invalid
Met362	No data.	Invalid
Met320	No data.	Invalid
Met292	No data.	Invalid
Met218	No data.	Invalid
Met302	No data.	Invalid
Phe393	No data.	Invalid
Phe378	0.59%	Valid
Phe336	0.99%	Valid
Phe308	1.16%	Invalid
Phe234	1.55%	Invalid
Phe302	1.63%	Invalid
Ser447	No data.	Invalid
Ser432	8.65%	Invalid
Ser390	9.71%	Invalid
Ser362	1.07%	Valid
Ser288	1.86%	Invalid
Ser302	10.21%	Invalid
Thr461	No data.	Invalid
Thr446	0.00%	Valid
Thr404	3.19%	Valid
Thr376	No data.	Invalid
Thr302	38.62%	Invalid
Thr302	29.93%	Invalid
Lys488	3.56%	Invalid
Lys473	0.00%	Valid
Lys431	2.06%	Invalid
Lys403	No data.	Invalid
Lys329	5.21%	Invalid
Lys302	6.73%	Invalid
Tyr523	No data.	Invalid
Tyr508	No data.	Invalid
Tyr466	3.69%	Invalid
Tyr438	3.36%	Invalid
Tyr364	4.53%	Invalid
Tyr302	1.47%	Valid
Tyr221	3.18%	Invalid
His497	No data.	Invalid
His482	No data.	Invalid
His440	No data.	Invalid
His412	No data.	Invalid
His338	No data.	Invalid
His302	No data.	Invalid
His195	No data.	Invalid
Glu489	4.29%	Invalid
Glu474	3.23%	Invalid
Glu432	No data.	Invalid
Glu404	1.48%	Invalid
Glu330	0.97%	Valid
Glu302	13.88%	Invalid
Glu187	67.07%	Invalid
Asp475	0.00%	Unchangin
Asp460	6.37%	Invalid
Asp418	1.32%	Valid
Asp390	2.90%	Invalid
Asp316	1.41%	Invalid
Asp302	1.47%	Invalid
Asp173	39.94%	Invalid
Leu359	No data.	Invalid
Leu344	5.64%	Invalid
Leu302	13.84%	Invalid
Leu274	5.90%	Invalid
Leu200	3.91%	Invalid
Leu302	15.60%	Invalid
ILe359	No data.	Invalid
ILe344	3.95%	Invalid
ILe302	13.92%	Invalid
ILe274	5.54%	Invalid
ILe200	3.62%	Valid
ArgR74	No data.	Invalid
Arg175	No data.	Invalid
Cit217	No data.	Invalid
Cit291	No data.	Invalid

7.3 Appendix 3 – Biomass compositions of *P. thermoglucosidasius*

Table S9: The protein compositions of *P. thermoglucosidasius* NCIMB 11955 used in this research were measured under chemostat growth conditions by Dr. Shyam Masakapalli in 2014 according to the method of Durot et al. (2009)[123]. Biomass compositions were determined *for P. thermoglucosidasius* NCIMB 11955 cells grown at 60°C, pH 7 on 1% Glucose ASM under aerobic and anaerobic conditions (indicated by redox potential) and was repeated in biological duplicate. R1A1-2 represents aerobic growth in rich media, R1A1-2 represents anaerobic growth in rich media

	% proportion	% of dry mass	umol/g CDW			
			R1A1	R1A2	R1An1	R1An2
Asp/Asn	9.822901778	5.143271371	470.9439185	456.9380975	543.2945417	445.2021817
Thr	5.223629803	2.735092565	285.1531998	276.6727746	328.9609885	269.5667609
Ser	3.032535947	1.587835822	192.1825435	186.4670555	221.7073473	181.6778693
Glu/Gln	15.9866549	8.370612505	683.2743563	662.9538512	788.2450833	645.9266639
Gly	4.471915145	2.34149477	432.5595756	419.6953008	499.0132522	408.9159223
Ala	7.02518988	3.678389421	545.5404408	529.3161273	629.3512499	515.7212673
Val	7.165843244	3.752035523	399.0794279	387.210849	460.3895843	377.2657954
Met	2.142098805	1.121602935	90.15616617	87.47493157	104.006764	85.2282412
Ile	5.931861476	3.105922669	289.4274972	280.8199548	333.8919418	273.6074257
Leu	7.555727299	3.956178814	368.6591891	357.6953048	425.295916	348.5083231
Tyr	3.249366635	1.70136837	109.9102876	106.6415676	126.7956905	103.9026048
Phe	4.104885016	2.149317794	153.9610604	149.3822752	177.6139376	145.5455677
His	2.051700609	1.074270439	82.56968965	80.11407604	95.25478498	78.05644055
Lys	7.811917709	4.090320113	336.4921384	326.4848987	388.1870748	318.0995195
Arg	5.496038973	2.877726006	194.2461579	188.4692982	224.0879926	183.6286867
Pro	3.10850815	1.627614867	176.6881844	171.433497	203.832606	167.0304299
Trp	3.677125826	1.925343082	108.8819892	105.6438506	125.6094156	102.930513
Cys	2.142098805	1.121602935	114.4424202	111.0389151	132.0240899	108.1870116

7.4 Appendix 4 – Complete Carbon Transition Network and notes

Table S10: The complete carbon atom transition network used in the flux estimation procedure. Notes indicate whether particular reaction equations were considered active or inactive for specific INST-13C-MFA MID data sets.

ID	Reaction Equation	Notes
R1	Gluc.ext (abcdef) + ATP -> Gluc (abcdef)	Glucose only
R2	Gluc (abcdef) + ATP -> G6P (abcdef)	
R3	Gluc.ext (abcdef) + PEP (ghi) -> G6P (abcdef) + Pyr (ghi)	Anaerobic Glucose Only
R4	Xyl.ext (abcde) + ATP -> X5P (abcde)	Xylose only
R5	P5P (abcde) -> X5P (abcde)	
R6	X5P (abcde) + E4P (fghi) -> GAP (cde) + F6P (abfghi)	
R7	X5P (abcde) + P5P (fghij) -> S7P (abfghij) + GAP (cde)	
R8	GAP (abc) + S7P (defghij) -> E4P (ghij) + F6P (defabc)	
R9	G6P (abcdef) -> P5P (bcdef) + CO2 (a) + 2*NADPH	

R10	P5P (abcde) -> 3PG (cde) + EC2 (ab)	
R11	F6P (abcdef) <-> E4P (cdef) + EC2 (ab)	
R12	S7P (abcdefg) <-> P5P (cdefg) + EC2 (ab)	
R13	F6P (abcdef) -> GAP (def) + EC3 (abc)	
R14	S7P (abcdefg) -> E4P (defg) + EC3 (abc)	
R15	G6P (abcdef) -> F6P (abcdef)	
R16	F6P (abcdef) + ATP -> FBP (abcdef)	
R17	FBP (abcdef) -> DHAP (cba) + GAP (def)	
R18	F6P (abcdef) -> P5P (bcdef) + Formald (a)	
R19	DHAP (abc) -> GAP (abc)	
R20	GAP (abc) -> 3PG (abc) + ATP + NADH	
R21	GAP (abc) -> 3PG (abc) + ATP + NADPH	NAPH-dependent GAPDH
R22	3PG (abc) <-> PEP (abc)	
R23	PEP (abc) -> Pyr (abc) + ATP	
R24	Pyr (abc) -> AcCoA (bc) + CO2 (a) + NADH	
R25	OAA (abcd) + AcCoA (ef) -> Cit (dcbefa)	
R26	Cit (abcdef) -> ICit (abcdef)	
R27	ICit (abcdef) -> AKG (abcde) + CO2 (f) + NADPH	
R28	AKG (abcde) -> SucCoA (bcde) + CO2 (a) + NADH	
R29	SucCoA (abcd) <-> Suc (abcd) + ATP	
R30	Suc (abcd) <-> Fum (abcd) + FADH2	
R31	Fum (abcd) <-> Mal (abcd)	
R32	Mal (abcd) -> OAA (abcd) + NADH	
R33	ICit (abcdef) -> Glyox (ab) + Suc (dcef)	
R34	Glyox (ab) + AcCoA (cd) -> Mal (abcd)	
R35	OAA (abcd) + ATP -> PEP (abc) + CO2 (d)	
R36	Mal (abcd) -> Pyr (abc) + CO2 (d) + NADPH	
R37	OAA (abcd) -> Pyr (abc) + CO2 (d) + ATP	
R38	Pyr (abc) + NADH <-> Lact (abc)	Inactive unless indicated by HPLC
R39	Lact (abc) -> Lact_eff (abc)	Inactive unless indicated by HPLC
R40	Pyr (abc) + Glu (defgh) -> Ala (abc) + AKG (defgh)	
R41	AcCoA (ab) <-> Ac (ab) + ATP	
R42	Ac (ab) -> Ac_eff (ab)	Inactive unless indicated by HPLC
R43	Pyr (abc) + Pyr (def) + Glu (ghijk) + NADPH -> Val (abefc) + CO2 (d) + AKG (ghijk)	
R44	AcCoA (ab) + Pyr (cde) + Pyr (fgh) + Glu (ijklm) + NADPH -> Leu (abdghe) + CO2 (c) + CO2 (f) + AKG (ijklm) + NADH	
R45	Thr (abcd) + Pyr (efg) + Glu (hijkl) + NADPH -> Ile (abfcdg) + CO2 (e) + AKG (hijkl) + NH3	
R46	Ile (abcdef) + METHF (g) + O2 -> Suc (bcdg) + CO2 (a) + AcCoA (ef)	
R47	AKG (abcde) + NADPH + NH3 -> Glu (abcde)	

R48	Glu (abcde) + ATP + NH3 -> Gln (abcde)	
R49	Glu (abcde) + CO2 (f) + Gln (ghijk) + Asp (lmno) + AcCoA (pq) + 5*ATP + NADPH -> Arg (abcdef) + AKG (ghijk) + Fum (lmno) + Ac (pq)	
R50	3PG (abc) + Glu (defgh) -> Ser (abc) + AKG (defgh) + NADH	
R51	Ser (abc) <-> Gly (ab) + MEETHF (c)	
R52	Gly (ab) <-> CO2 (a) + MEETHF (b) + NADH + NH3	
R53	P5P (abcde) + FTHF (f) + Gln (ghijk) + Asp (lmno) + 5*ATP -> His (edcbaf) + AKG (ghijk) + Fum (lmno) + 2*NADH	
R54	OAA (abcd) + Glu (efghi) <-> Asp (abcd) + AKG (efghi)	
R55	Asp (abcd) + Pyr (efg) + Glu (hijkl) + ATP + 2*NADPH + SucCoA (mnop) -> Lys (efgdcb) + CO2 (a) + AKG (hijkl) + Suc (mnop)	
R56	Asp (abcd) + 2*ATP + 2*NADPH -> Thr (abcd)	
R57	Asp (abcd) + 2*ATP + NH3 -> Asn (abcd)	
R58	PEP (abc) + PEP (def) + E4P (ghij) + Glu (klmno) + ATP + NADPH -> Phe (abcefghij) + CO2 (d) + AKG (klmno) + NADH	
R59	PEP (abc) + PEP (def) + E4P (ghij) + Glu (klmno) + ATP + NADPH -> Tyr (abcefghij) + CO2 (d) + AKG (klmno) + NADH	
R60	O2.ext -> O2	
R61	CO2 (a) -> CO2_eff (a)	
R62	CO2.ext (a) -> CO2 (a)	
R63	Glu (abcde) + ATP + 2*NADPH -> Pro (abcde)	
R64	Ser (abc) + AcCoA (de) + 3*ATP + 4*NADPH + SO4 -> Cys (abc) + Ac (de)	
R65	Glu (abcde) + CO2 (f) -> Arg (abcdef)	
R66	Asp (abcd) + METHF (e) + Cys (fgh) + SucCoA (ijkl) + ATP + 2*NADPH -> Met (abcde) + Pyr (fgh) + Suc (ijkl) + NH3	
R67	PEP (abc) + PEP (def) + E4P (ghij) + P5P (klmno) + Ser (pqr) + Gln (stuvw) + 3*ATP + NADPH -> 3PG (mno) + Pyr (abc) + CO2 (d) + Trp (pqrlkghijef) + Glu (stuvw)	
R68	Pyr (abc) -> Pyr_eff (abc)	Inactive unless indicated by HPLC
R69	Pyr (abc) -> AcCoA (bc) + For_eff (a)	Anaerobic only, inactive unless indicated by HPLC
R70	AcCoA (ab) -> Eth_eff (ab)	Anaerobic only, inactive unless indicated by HPLC
R71	2*NADH + O2 -> 6.4*ATP	
R72	2*FADH2 + O2 -> 4.2*ATP	
R73	ATP -> ATP.ext	
R74	NH3.ext -> NH3	
R75	SO4.ext -> SO4	
R76	MEETHF (a) + NADH -> METHF (a)	
R77	MEETHF (a) -> FTHF (a) + NADPH	
R78	Mal (abcd) -> OAA (abcd) + FADH2	
R79	-> 39.67*Biomass	Stoichiometrric composition as indicated in Table 36

R80	Asp (abcd) + CO2 (e) -> UMP (abcd) + CO2 (e)	
R81	MEETHF.ext (a) -> MEETHF (a)	
R82	MEETHF (a) -> MEETHF_eff (a)	
R83	12Gluc.ext (abcdef) + ATP -> Gluc (abcdef)	Glucose only, UB = 10
R84	2*FeOX + AKG (abcde) -> 2*FeRed + SucCoA (bcde) + CO2 (a)	Anaerobic only
R85	2*FeOX <-> 2*FeRed	Anaerobic only
R86	12xyl.ext (abcde) + ATP -> X5P (abcde)	Xylose only, UB = 10