Single-cell time-series analysis of metabolic rhythms in yeast

Arin Wongprommoon

Doctor of Philosophy

The University of Edinburgh

2023

To my family,

Dad and Mum, my brother Arun, and aunt Pinnapa.

Abstract

The yeast metabolic cycle (YMC) is a biological rhythm in budding yeast (*Saccharomyces cerevisiae*). It entails oscillations in the concentrations and redox states of intracellular metabolites, oscillations in transcript levels, temporal partitioning of biosynthesis, and, in chemostats, oscillations in oxygen consumption. Most studies on the YMC have been based on chemostat experiments, and it is unclear whether YMCs arise from interactions between cells or are generated independently by each cell. This thesis aims at characterising the YMC in single cells and its response to nutrient and genetic perturbations. Specifically, I use microfluidics to trap and separate yeast cells, then record the time-dependent intensity of flavin autofluorescence, which is a component of the YMC.

Single-cell microfluidics produces a large amount of time series data. Noisy and short time series produced from biological experiments restrict the computational tools that are useful for analysis. I developed a method to filter time series, a machine learning model to classify whether time series are oscillatory, and an autocorrelation method to examine the periodicity of time series data.

My experimental results show that yeast cells show oscillations in the fluorescence of flavins. Specifically, I show that in high glucose conditions, cells generate flavin oscillations asynchronously within a population, and these flavin oscillations couple with the cell division cycle. I show that cells can individually reset the

phase of their flavin oscillations in response to abrupt nutrient changes, independently of the cell division cycle. I also show that deletion strains generate flavin oscillations that exhibit different behaviour from dissolved oxygen oscillations from chemostat conditions.

Finally, I use flux balance analysis to address whether proteomic constraints in cellular metabolism mean that temporal partitioning of biosynthesis is advantageous for the yeast cell, and whether such partitioning explains the timing of the metabolic cycle. My results show that under proteomic constraints, it is advantageous for the cell to sequentially synthesise biomass components because doing so shortens the timescale of biomass synthesis. However, the degree of advantage of sequential over parallel biosynthesis is lower when both carbon and nitrogen sources are limiting.

This thesis thus confirms autonomous generation of flavin oscillations, and suggests a model in which the YMC responds to nutrient conditions and subsequently entrains the cell division cycle. It also emphasises the possibility that subpopulations in the culture explain chemostat-based observations of the YMC. Furthermore, this thesis paves the way for using computational methods to analyse large datasets of oscillatory time series, which is useful for various fields of study beyond the YMC.

Lay Summary

Living things have biological clocks that make sure that their processes happen in a sequence. An example is the circadian rhythm, which makes sure that our body's processes happen at the correct times in a day. For example, the circadian rhythm makes sure we sleep at night and are awake during daytime.

Baker's yeast has a biological clock called the yeast metabolic cycle (YMC). Because of this cycle, as the yeast cell grows, its metabolism and the concentrations of chemicals in the yeast cell changes over time. Most studies on the YMC have focused on large cultures. So, we do not know whether the cells need to talk to each other to generate the YMC, or whether each cell can generate the YMC on its own. To study the YMC, I use a microfluidics platform that can separate yeast cells. This platform allows me to monitor the changing fluorescence of a metabolite, which is part of the metabolic cycle, by taking a time-lapse of images.

My experiments showed that when the yeast cells grow in beneficial conditions, each cell generates its metabolic cycle in sync with cell division, but out-of-sync with other cells. But, when I took away the cells' nutrients, these cells started generating their metabolic cycles in sync. Additionally, I showed that individual cells behave differently from cells in large cultures.

Because my experiments produce a large amount of temporal data, I explored ways to computationally analyse this data. I was able to teach a machine learning algorithm, a common type of computer algorithm, to tell apart signals that repeated in cycles from signals that did not.

I then asked the question of: does it make sense for the yeast cell to make its building blocks in sequence while it grows, or does it make more sense for it to make all its building blocks in parallel? I solved a set of equations that describes the chemical reactions that occur in the yeast cell to determine which reactions fire the most when the cell has limited resources. The solutions of the equations suggest that if the yeast cell makes the components of its biomass in sequence, it saves time, but not when it has few nutrients in its environment.

To summarise, my thesis shows that each yeast cell generates its YMC on its own and how it does so depends on which nutrients the cell grows on. My results were different from studies based on large, bulk cultures. To explain both kinds of results, there may be sub-populations of cells in bulk culture that may generate metabolic cycles with different properties. Furthermore, my thesis shows that computational methods can be used to analyse large amounts of temporal data, and these methods are useful for studying time series in different fields.

Abstract (Thai)

วัฏจักรเมตาบอลิซึมในยีสต์ (ภาษาอังกฤษ: “yeast metabolic cycle” หรือตัวย่อ YMC) เป็นวัฏจักรทางชีวภาพในยีสต์ทำขนมปัง (*Saccharomyces cerevisiae*) ในวัฏจักรนี้ ความเข้มข้นกับสถานะรีดอกซ์ของเมตาบอไลต์และระดับกรดไรโบนิวคลีอีกภายในเซลล์มีจังหวะขึ้นลง มีการแบ่งช่วงเวลาในการสังเคราะห์สารในเซลล์ และการใช้ออกซิเจนมีจังหวะขึ้นลงที่สังเกตได้จากอุปกรณ์คีโมสแตท (chemostat)

โดยส่วนมาก การศึกษาวัฏจักรเมตาบอลิซึมในยีสต์ใช้คีโมสแตทเป็นเครื่องมือ จึงไม่เป็นที่ชัดเจนว่าวัฏจักรเมตาบอลิซึมในยีสต์เกิดจากปฏิสัมพันธ์ระหว่างเซลล์หรือถูกสร้างอย่างเป็นอิสระในแต่ละเซลล์ วิทยานิพนธ์นี้จึงมุ่งแสดงคุณลักษณะของวัฏจักรเมตาบอลิซึมในยีสต์ในเซลล์เดี่ยวและมุ่งศึกษาผลตอบสนองของวัฏจักรนี้ต่อการเปลี่ยนแปลงของสารอาหารและสารพันธุกรรม โดยจำเพาะ ข้าพเจ้าใช้เครื่องมือไมโครฟลูอิดิกส์ (microfluidics) เพื่อดักและแยกเซลล์ยีสต์ และวัดความเรืองแสงของฟลาวิน (flavin) ตามกาลเวลา เป็นการเฝ้าสังเกตวัฏจักรเมตาบอลิซึมในยีสต์

ไมโครฟลูอิดิกส์เซลล์เดี่ยวสร้างข้อมูลอนุกรมเวลาเป็นจำนวนมาก เนื่องจากอนุกรมเวลาที่สร้างโดยการทดลองเชิงชีววิทยามีลักษณะสั้นและมีสัญญาณรบกวน เครื่องมือทางคอมพิวเตอร์ที่สามารถใช้ในการวิเคราะห์จึงถูกจำกัด ดังนั้นข้าพเจ้าพัฒนาวิธีกรองอนุกรมเวลา พัฒนาโมเดลปัญญาประดิษฐ์เพื่อแยกแยะอนุกรมกรมเวลาที่มีจังหวะซ้ำ และพัฒนาวิธีสหสัมพันธ์เชิงอนุกรมเพื่อประมาณความถี่ของจังหวะในข้อมูลอนุกรมเวลา

ผลการทดลองของข้าพเจ้าแสดงให้เห็นว่าเซลล์ยีสต์แสดงความเรืองแสงของฟลาวินเป็นจังหวะ คู่ขนานกับวัฏจักรการแบ่งตัวของเซลล์ โดยในสภาพกลูโคสสูงประชากรเซลล์สร้างจังหวะนี้อย่างไม่พร้อมกัน ข้าพเจ้าแสดงให้เห็นว่าเซลล์สามารถตั้งเฟสของวัฏจักรเมตาบอลิซึมใหม่เมื่อสารอาหารเปลี่ยนแปลงอย่างกะทันหัน และอย่างเป็นอิสระจากวัฏจักรการแบ่งตัวของเซลล์ นอกจากนี้ข้าพเจ้าแสดงให้เห็นว่าพันธุ์เซลล์ที่เกิดจากการตัดยีนมีจังหวะการเรืองแสงของฟลาวินที่มีพฤติกรรมที่แตกต่างจากจังหวะของความเข้มข้นของออกซิเจนในคีโมสแตท

ข้าพเจ้าใช้ flux balance analysis เพื่อตอบคำถามว่าข้อจำกัดในโปรติโอมนำไปสู่ความได้เปรียบของการแยกเวลาการสังเคราะห์สารในเซลล์หรือไม่ และตอบคำถามว่าการแยกเวลาการสังเคราะห์สารนี้สามารถอธิบายระยะเวลาของวัฏจักรเมตาบอลิซึมได้หรือไม่ ผลการศึกษาของข้าพเจ้าแสดงให้เห้นว่าข้อจำกัดในโปรติโอมนำไปสู่การแยกเวลาการสังเคราะห์สารในเซลล์เนื่องจากการแยกเวลานี้ย่นระยะเวลาการสังเคราะห์สารในเซลล์โดยรวม แต่ทว่าความได้เปรียบนี้ลดต่ำลงหากแหล่งคาร์บอนและไนโตรเจนมีปริมาณต่ำ

กล่าวโดยสรุป วิทยานิพนธ์นี้ยืนยันว่าจังหวะการเรืองแสงของฟลาวินถูกสร้างโดยอิสระในเซลล์ และเสนอว่าวัฏจักรเมตาบอลิซึมในยีสต์ตอบสนองต่อสภาพสารอาหารแล้วกำหนดเวลาการแบ่งตัวของเซลล์ วิทยานิพนธ์นี้เน้นว่ากลุ่มเซลล์ย่อยในการเพาะเลี้ยงเซลล์อาจอธิบายผลการสังเกตวัฏจักรเมตาบอลิซึมในยีสต์ในคีโมสแตท นอกจากนี้วิทยานิพนธ์นี้ปูทางสำหรับการใช้วิธีทางคอมพิวเตอร์ในการวิเคราะห์ชุดข้อมูลอนุกรมเวลาขนาดใหญ่ ซึ่งมีประโยชน์กับสาขาวิชาอื่นนอกจากวัฏจักรเมตาบอลิซึมในยีสต์

Acknowledgements

This work would not have been possible without the support of many people and organisations.

I would like to thank the financial support provided by School of Biological Sciences, University of Edinburgh, and the Edinburgh Global Scholarship.

I would like to express my deepest gratitude towards my supervisors Dr Diego Oyarzún and Prof Peter Swain for the unique opportunity of joining two research teams, and for creating such a great environment for me to learn and work in Edinburgh. I am also very grateful your advice, patience, and mental support — especially in the very isolating days of the COVID-19 pandemic.

I am most grateful for Dr Kevin Correia for his mentorship in the early parts of my PhD: teaching me how to me ask questions, bringing in laboratory best practices from his experience, and even philosophical insights in science and life.

I am very grateful for the main software developers of the Swain lab: Dr Julian Pietsch, Dr Ivan Clark, Dr Diane Adjavon, and Dr Alán Muñoz Gonzalez. I would like to thank Julian in particular for his patience as I grasped the ropes of the old software pipeline. Thank you, Diane and Alán, for including me in your very ambitious project to re-write the pipeline — this was an immense learning experience for me. Alán — I am most grateful to have worked alongside you with *aliby*, and for your converting me to `doom emacs` too.

I am very grateful for Dr Ivan Clark and Iseabail Farquhar for imparting their wet-lab expertise and their patience during my experiments. Ish, I'd like to thank you for your patience and perseverance training me CRISPR during the chaotic times of the COVID-19 lockdown.

I would also like to thank my other colleagues at the Swain Lab: Dr Luis Fernando Montano-Gutierrez, Dr François El-Daher, Dr Julien Hurbain, Dr Nahuel Manzanaro Moreno, Dr Yu Huo, and Sofía Esteban Serna.

I would like to thank my colleagues at the Biomolecular Control Group for creating a fantastic social environment and continuously sparking ideas. I am grateful for Dr Vanessa Smer-Barreto and Dr Babita Verma for their mentorship and feedback on my research. I would like to thank Dr Mona Tonn and Denise Thiel for our weekly movie nights. I also would like to thank Evangelos-Marios Nikolados for a great collaboration opportunity and, along with Michał Kobiela, ensuring I have a good social life.

I would like to thank my friends from my undergraduate studies — Christos Nicolaou, Esther Ng, Patawee Jintana, among others — for their moral support. I would like to extend special thanks to my best friends Jessica Wang and Maetavee Asavabhokhin for their unwavering emotional support during all ups and downs. And I would also like to thank my high school friends, scattered across the globe, Attawit Chaiyaroj, Paloch Vasudhara, and Dr Pravarit Athiprayoon for their support too.

Last but not least, I would like to thank my family. I express my deepest gratitude for my parents who have taught me the wonder of science, the importance of education, and the value of working hard. None of this would have been possible without you. Aunt Pinnapa, I now have a PhD! And lastly, I thank my brother Arun Wongprommoon and cousin Pornsom Ruengpirasiri for their continuous support in this long journey.

Declaration

I declare that this thesis was composed by myself, that the work contained herein is my own except where explicitly stated otherwise in the text, and that this work has not been submitted for any other degree or professional qualification except as specified.

Arin Wongprommoon

Contents

Abstract **iii**

Lay Summary **v**

Abstract (Thai) **vii**

Acknowledgements **viii**

Declaration **xi**

Figures and Tables **xvii**

1 Introduction **1**

1.1 Biological rhythms 2

1.1.1 Biological basis of biological rhythms 2

1.1.2 Mathematical basis of biological rhythms 7

1.2 Yeast metabolic cycle 10

1.2.1 Definition and description of the yeast metabolic cycle . . 10

1.2.2 Yeast metabolic cycles under perturbations 18

1.2.3 Modelling the yeast metabolic cycle 24

1.2.4 Big picture/Hypothesis 26

1.2.5 Disputes and unresolved questions 27

1.2.6 Implications of the metabolic cycle 32

1.3 Flavins and flavoproteins 36

1.3.1 Introduction to cellular autofluorescence 36

1.3.2 Biochemical basis of flavins and flavoproteins 37
1.3.3 Flavins and flavoproteins in the yeast metabolic cycle . . . 43
1.4 Flux balance analysis 45
1.4.1 Introduction to flux balance analysis 45
1.4.2 Development of genome-scale metabolic models for budding yeast 47
1.4.3 Allocation of proteomic resources 50
1.5 Aims of thesis and research questions 53

2 Methods 54
2.1 Laboratory methods 54
2.1.1 Strains and media 54
2.1.2 Single-cell microfluidics 56
2.2 Image analysis methods 60
2.3 Time series analysis methods 63
2.3.1 Data pre-processing 63
2.3.2 Classical periodogram 63
2.3.3 Autoregressive model 65
2.3.4 Precision and recall 67
2.3.5 Cross-correlation function 68
2.3.6 Generation of synthetic oscillatory data 69
2.3.7 Signal-to-noise ratio 71
2.4 Methods related to flux balance analysis 73
2.4.1 ecYeast8 formalisms 73
2.4.2 Computing molecular weights of pseudometabolites 74

3 Microfluidics and fluorescence microscopy for cellular metabolic cycles 76
3.1 Coupled oscillations in permissive conditions 78

3.2 Decoupling between the metabolic and cell division cycles 83
3.3 Metabolic cycles in different genetic backgrounds 90
3.4 Metabolic cycles in different carbon sources 93
3.5 Metabolic cycles persist in potassium-deficient media 97
3.6 Metabolic cycles in deletion strains 103
3.7 Discussion 107
3.7.1 Summary of results 107
3.7.2 Interpretation of results 107
3.7.3 Study caveats and future directions 112

4 Analysis of oscillatory time series in the yeast metabolic cycle 115
4.1 Analysing time series in a biological context 117
4.2 Data cleaning: filtering out long-term trends 119
4.3 Visualising groups in the dataset 121
4.3.1 UMAP 122
4.3.2 Graph-based clustering 127
4.4 Detection of rhythmicity 132
4.4.1 Spectral methods 132
4.4.2 Model fitting 136
4.4.3 Machine learning 139
4.5 Period estimation using the autocorrelation function 142
4.5.1 Effect of noise parameters on the autocorrelation function of synthetic sinusoids 144
4.5.2 FitzHugh-Nagumo oscillator: effect of oscillation shape . . 149
4.5.3 Real data from fluorescence microscopy in microfluidics experiments 149
4.6 Detection of synchrony 151
4.6.1 Synthetic data 152

4.6.2 Real data from fluorescence microscopy in microfluidics experiments . . . 153
4.7 Discussion . . . 153

5 Modelling yeast biosynthesis strategies under constraints 158
5.1 Modelling temporal scheduling of biosynthesis . . . 159
5.2 The Yeast8 and ecYeast8 models and their formalisms . . . 162
5.2.1 GECKO . . . 163
5.3 Ablating pseudometabolites from the biomass reaction . . . 164
5.3.1 Definition of ablation . . . 164
5.3.2 Effect of ablation on allocation of proteome to enzymes . . . 166
5.4 Estimating timescale of biosynthesis . . . 170
5.5 Effect of restricting the enzyme pool . . . 174
5.6 Effect of carbon and nitrogen sources . . . 175
5.6.1 Saturation of exchange reactions . . . 176
5.6.2 Effect of carbon and nitrogen sources on biomass synthesis strategies . . . 179
5.6.3 Relationship between proteome allocation and resource allocation strategy . . . 185
5.7 Discussion . . . 193

6 Conclusions 197
6.1 Microfluidics and fluorescence microscopy for cellular metabolic cycles . . . 198
6.2 Analysis of oscillatory time series in the yeast metabolic cycle . . . 200
6.3 Modelling yeast biosynthesis strategies under constraints . . . 201
6.4 Summary and broader context of thesis . . . 203

Appendices

A Time series **206**

A.1 catch22 features 206

A.2 UMAP hyperparameters 206

A.3 Classification pipeline 208

A.4 Gillespie algorithm for stochastic chemical systems 209

B Flux balance analysis **212**

B.1 Modifications of chemical reaction in GECKO, general cases . . . 212

B.2 Modifications of chemical reactions in GECKO for non-simple cases 214

B.3 Computing molecular weights of pseudometabolites in ecYeast8 . 215

B.3.1 Carbohydrate, DNA, RNA, cofactor, and ion pseudometabolites 217

B.3.2 Protein pseudometabolite 217

B.3.3 Lipid pseudometabolite 219

B.4 Mathematical explanation of the effect of restricting the enzyme pool 221

Bibliography **225**

Figures and Tables

Figures

1.1 Overview of the cell division cycle 4
1.2 Glycolytic cycles 7
1.3 Phases of the yeast metabolic cycle 11
1.4 The yeast metabolic cycle has been described as spontaneous respiratory cycles 12
1.5 The yeast metabolic cycle is characterised by transcript cycling . . . 12
1.6 The yeast metabolic cycle seen as sequential scheduling of cellular processes into phases 13
1.7 The yeast metabolic cycle seen as coordinated cycling of metabolites in the cell 17
1.8 Effects of genetic perturbations on the yeast metabolic cycle 21
1.9 The relationship between the yeast metabolic cycle and the cell division cycle can be seen as a system of coupled oscillators 26
1.10 Model for how sub-populations of cells may account for oscillations of dissolved oxygen observed in the chemostat 30
1.11 Chemical structure of FMN and FAD 37
1.12 Simplified schematic of biosynthesis of flavins 38
1.13 Reference pathway for biosynthesis of riboflavin and derivatives . . . 39
1.14 Fluorescence spectrum of riboflavin 40
1.15 Flavoproteins shown by abundance 41

1.16 Timeline of development of genome-scale metabolic models of *Saccharomyces cerevisiae*. 47
1.17 Illustration of *isa* reactions in Yeast5. 49

2.1 Overview of ALCATRAS microfluidic chip design. 57
2.2 Fluid dynamic simulations of the ALCATRAS microfluic chip. . . . 58
2.3 Overview of single-cell microfluidics set-up using the ALCATRAS system 59
2.4 Example time course of images from a microfluidics experiment, processed by *aliby*. The extracted values were then saved in an array of numbers, with rows indicating cells (labelled by IDs) and columns indicating time in minutes — each row thus represents a time series. 62
2.5 Illustration of signal-to-noise ratio. 72

3.1 Engineering a fluorescent protein cassette fused to *HTB2* allows the identification of phases of the cell division cycle through monitoring changes in fluorescence of the fluorescent protein. 78
3.2 Flavin fluorescence and histone 2B levels in a single, representative FY4 HTB2::mCherry cell grown in $20\,\mathrm{g\,L^{-1}}$ glucose. 79
3.3 Mean Fourier spectrum of flavin fluorescence time series across cells. Median autocorrelation functions of flavin fluorescence and histone 2B levels time series, along with the periods of each oscillator across cells. Data are from FY4 HTB2::mCherry cells under $20\,\mathrm{g\,L^{-1}}$ glucose. . . 81
3.4 Heatmap showing the flavin fluorescence and budding events of each cell over time. Median flavin fluorescence signal across cells, aligned to first budding event. Median cross-correlation function between flavin and histone 2B signals. Data are from FY4 HTB2::mCherry cells in $20\,\mathrm{g\,L^{-1}}$ glucose. 82

3.5 Flavin fluorescence and histone 2B levels in a single, representative FY4 HTB2::mCherry cell. Heatmap showing the flavin fluorescence and budding events of each cell. Data are from FY4 and HTB2::mCherry cells, subject to glucose starvation. 84

3.6 Mean growth rate, derived from changes in the sum of estimated parent cell and bud volumes and mean cumulative number of budding events per cell of FY4 and HTB2::mCherry strains over time during the glucose-starvation experiment. 85

3.7 Relationship between the spatial position of trap and the mean raw flavin signal across all time points, and the period of flavin oscillation, for each parent cell, during glucose starvation. 87

3.8 Flavin fluorescence and histone 2B levels in two sample FY4 HTB2::mCherry cells in the glucose starvation experiment. 88

3.9 Distributions of flavin and mCherry fluorescence over time, for the glucose-starvation experiment. 89

3.10 Mean Fourier spectrum of flavin fluorescence time series across cells. Median autocorrelation function of flavin fluorescence time series, along with the periods across cells. Heatmap showing the flavin fluorescence and budding events of each cell. Median flavin fluorescence signal across cells, aligned to first budding event. Data are from BY4741 cells in $10\,\mathrm{g\,L^{-1}}$ glucose. 91

3.11 Mean Fourier spectrum of flavin fluorescence across cells. Median autocorrelation function of flavin fluorescence time series, along with the periods across cells. Median flavin fluorescence signal across cells, aligned to first budding event. Data are from CEN.PK113-7D cells in $10\,\mathrm{g\,L^{-1}}$ glucose. 92

3.12 Flavin fluorescence and histone 2B levels in a single, representative FY4 HTB2::mCherry cell grown in $20\,\mathrm{g\,L^{-1}}$ pyruvate. Median autocorrelation functions of flavin fluorescence and histone 2B levels time series, along with the periods of each oscillator across cells. Median cross-correlation function between flavin and histone 2B signals. Data are from FY4 HTB2::mCherry cells in $20\,\mathrm{g\,L^{-1}}$ pyruvate. 94

3.13 Flavin fluorescence and histone 2B levels in a single, representative FY4 HTB2::mCherry cell grown in $0.010\,\mathrm{g\,L^{-1}}$ glucose. Median autocorrelation functions of flavin fluorescence and histone 2B levels time series, along with the periods of each oscillator across cells. Data are from FY4 HTB2::mCherry cells in $0.010\,\mathrm{g\,L^{-1}}$ glucose. 95

3.14 Mean growth rate, derived from changes in the sum of estimated parent cell and bud volumes of FY4 and HTB2::mCherry strains over time during the glucose-starvation experiment, for the glucose-limiting condition and the high glucose condition. Similarly, the mean cumulative number of budding events per cell. 96

3.15 Distributions of signal-to-noise ratios of flavin signals from cells in $0.010\,\mathrm{g\,L^{-1}}$ glucose, $20\,\mathrm{g\,L^{-1}}$ pyruvate, and $20\,\mathrm{g\,L^{-1}}$ glucose. 97

3.16 Decreasing extracellular potassium (K^+) concentration shortens, then under $1\,\mathrm{mmol\,dm^{-3}}$, destroys metabolic oscillations in the chemostat. Adapted from J. S. O'Neill et al. (2020). 98

3.17 Flavin fluorescence and histone 2B levels in a single, representative FY4 HTB2::mCherry cell. Median autocorrelation functions of flavin fluorescence and histone 2B levels time series, along with the periods of each oscillator across cells. Data are from FY4 and HTB2::mCherry cells; autocorrelation functions only used time points from the potassium-deficient condition. 99

3.18 Distributions of signal-to-noise ratios of flavin signals from cells before, during, and after potassium-deficiency. 100

3.19 Mean growth rate, derived from changes in the sum of estimated parent cell and bud volumes and mean cumulative number of budding events per cell of FY4 and HTB2::mCherry strains over time during the potassium-deficient experiment. 101

3.20 Distributions of flavin and mCherry fluorescence over time, for the potassium-deficient experiment. 101

3.21 Flavin fluorescence levels in a single, representative *zwf1*Δ cell. Mean Fourier spectrum of flavin fluorescence across cells. Median autocorrelation function of flavin fluorescence time series, along with the periods across cells. Median flavin fluorescence signal across cells, aligned to first budding event. Distribution of signal-to-noise ratios of flavin signals from cells. Data are from *zwf1*Δ (BY4741) cells in $10\,\mathrm{g\,L^{-1}}$ glucose. 104

3.22 Flavin fluorescence levels in a single, representative *tsa1*Δ *tsa2*Δ cell. Overlaid Fourier spectra of flavin fluorescence across cells. Median autocorrelation function of flavin fluorescence time series, along with the periods across cells. Median flavin fluorescence signal across cells, aligned to first budding event. Distribution of signal-to-noise ratios of flavin signals from cells. Data are from *tsa1*Δ *tsa2*Δ (BY4742) cells in $10\,\mathrm{g\,L^{-1}}$ glucose. 106

4.1 Time series and Fourier spectra corresponding to a sample raw time series of flavin autofluorescence, the time series processed by a high-pass Butterworth filter, and the time series detrended using a moving average. 120

4.2 UMAP embedding of a dataset of time series featurised using *catch22*. 123

4.3 Sample non-oscillatory time series from regions of interest in UMAP embedding space 124
4.4 UMAP embedding of a dataset of the first half of time series, featurised using *catch22*. 125
4.5 Grid search of UMAP hyperparameters. 126
4.6 Process of preparing a dataset of time series for modularity clustering. 129
4.7 Pruned graph of BY4741 and *zwf1*Δ time series from the same experiment. 130
4.8 Sample time series from BY4741 and *zwf1*Δ cells. 131
4.9 Best five time series in the *zwf1*Δ dataset and their periodograms, ranked by the quality of oscillation based on the maximum power in the periodogram (Glynn et al., 2006) 133
4.10 Worst five time series in the *zwf1*Δ dataset and their periodograms, ranked by the quality of oscillation based on the maximum power in the periodogram (Glynn et al., 2006) 134
4.11 ROC curve of classifier based on Glynn et al. (2006). 135
4.12 Power spectra analytically derived from fitting an autoregressive model to time series can be divided into four types. 136
4.13 Sample time series, with a fitted autoregressive model computed according to Jia and Grima (2020). 137
4.14 138
4.15 Precision and recall from five-fold cross-validation of support vector classifiers trained using different featurisation methods. 140
4.16 Histogram of probabilities of whether a time series in the test data set is classified as oscillatory by the SVC, and as a control, with labels randomly re-assorted to time series. 141

4.17 Sample time series arranged by probability that each is oscillatory, as predicted by the support vector classifier. 141
4.18 Permutation feature importance values of *catch22* features in classifying manually-scored oscillatory and non-oscillatory time series 143
4.19 Sample sinusoids without noise, with Gaussian noise, and with Gillespie noise, along with their autocorrelation functions. 145
4.20 Illustration of the birth-death process. 146
4.21 Effect of death rate of Gillespie noise on the autocorrelation function. 147
4.22 Effect of birth rate of Gillespie noise on the autocorrelation function. 148
4.23 Sample FitzHugh-Nagumo oscillators with Gillespie noise, and its autocorrelation function. 149
4.24 Sample time series of flavin autofluorescence and its autocorrelation function. 150
4.25 Sample time series of histone 2B abundance, its autocorrelation function. 151
4.26 Sample sinusoid and FitzHugh-Nagumo oscillators without noise and with Gillespie noise, along with their cross-correlation functions. . . 152
4.27 Sample time series of flavin autofluorescence and histone 2B abundance, along with the cross-correlation function. 154

5.1 Modifications to the stoichiometric matrix of a genome-scale model that GECKO imposes 163
5.2 Growth rates from the original model and from the ablated versions of the model 165
5.3 $\log_2$(FC) of enzyme usage reaction flux in rounds of ablation 168
5.4 Estimated synthesis times of biomass components, from ablation and from scaling doubling time by mass fraction. 170

5.5 Estimated synthesis times of biomass components, for BY4741, *zwf*Δ, and *tsa2*Δ 174
5.6 Effect of the size of the proteome pool available for enzymes 176
5.7 Effect of exchange reactions on growth rate 178
5.8 Effect of glucose and ammonium exchange rates 181
5.9 Effect of pyruvate and ammonium exchange rates 184
5.10 Changes in enzyme usage reaction flux and Kendall's τ-b rank correlation coefficient for each pair, $R_{\mathrm{glc}} = 1.69\,\mathrm{mmol}\ \mathrm{g}_{\mathrm{DW}}^{-1}\ \mathrm{h}^{-1}$, $R_{\mathrm{amm}} = 1.05\,\mathrm{mmol}\ \mathrm{g}_{\mathrm{DW}}^{-1}\ \mathrm{h}^{-1}$. 186
5.11 Changes in enzyme usage reaction flux and Kendall's τ-b rank correlation coefficient for each pair, $R_{\mathrm{glc}} = 16.89\,\mathrm{mmol}\ \mathrm{g}_{\mathrm{DW}}^{-1}\ \mathrm{h}^{-1}$. 188
5.12 Changes in enzyme usage reaction flux and Kendall's τ-b rank correlation coefficient for each pair, $R_{\mathrm{pyr}} = 3.73\,\mathrm{mmol}\ \mathrm{g}_{\mathrm{DW}}^{-1}\ \mathrm{h}^{-1}$, $R_{\mathrm{amm}} = 0.90\,\mathrm{mmol}\ \mathrm{g}_{\mathrm{DW}}^{-1}\ \mathrm{h}^{-1}$. 190
5.13 Changes in enzyme usage reaction flux and Kendall's τ-b rank correlation coefficient for each pair, $R_{\mathrm{pyr}} = 8.89\,\mathrm{mmol}\ \mathrm{g}_{\mathrm{DW}}^{-1}\ \mathrm{h}^{-1}$. 191
5.14 Measures of similarity between enzyme usage fluxes in different nutrient conditions 192

Tables

1.1 Roles of the most abundant flavoproteins. 42

2.1 Strains used in this thesis. 54
2.2 Composition of base minimal medium 55
2.3 Composition of trace metal mix 55

2.4 Composition of vitamin mix 56

2.5 Supplements to minimal media for auxotrophic strains 56

2.6 Confusion matrix 67

2.7 Computed molecular weights of bulk metabolites in ecYeast8, compared to experimentally recorded biomass composition 75

3.1 Summary statistics of time series of flavin oscillations for each experiment. 108

4.1 Confusion matrix to evaluate the performance of using the autoregressive model (Jia & Grima, 2020) to detect rhythmicity. 137

A.1 *catch22* features 207

B.1 Straightforward cases of computing pseudometabolite molecular weights from pseudoreactions in ecYeast8 218

B.2 ecYeast8 reactions that generate the `fatty acid backbone` metabolite 220

B.3 f_i values for each biomass component. 221

Chapter 1

Introduction

This thesis aims to understand how an organism adapts its metabolism and cellular processes in response to external conditions, in the context of a biological rhythm. Specifically, I use the yeast metabolic cycle (YMC) as a framework for biological rhythms. My reasons are twofold: (a) biological rhythms are important for coordination of responses and are present across kingdoms, (b) there are unanswered questions about the mechanistic basis of the YMC and about reconciling evidence from two types of experimental studies.

This thesis is divided into six chapters:

1. Chapter 1 discusses the background behind the yeast metabolic cycle, using flavin autofluorescence as a way to monitor the yeast metabolic cycle, and the use of genome-scale metabolic models and flux balance analysis to investigate the metabolism of budding yeast.
2. Chapter 2 discusses the methods: single-cell microfluidics of yeast strains, an automated image analysis pipeline, and time series analysis methods.
3. Chapter 3 presents results from single-cell microfluidics and fluorescence microscopy to detect flavin-based metabolic cycles in yeast cells. I show that the metabolic cycle and cell division cycle are autonomous and synchronise in permissive conditions, while perturbations affect the relationship between these two biological oscillators.

4. Chapter 4 discusses the analysis of oscillatory time series. Given the challenges of analysing noisy low-resolution time series, this deserves discussion in its own right. This chapter explores processes to visualise groups in the dataset, detecting rhythmicity, period estimation, and detecting synchrony.
5. Chapter 5 discusses using flux balance analysis to address whether proteome constraints explain sequential scheduling of biosynthesis in the yeast metabolic cycle.
6. Finally, Chapter 6 presents a conclusion based on the previous three results chapters and suggests further avenues of study.

1.1 Biological rhythms

1.1.1 Biological basis of biological rhythms

Biological rhythms are repeating physiological or cellular processes. Biological rhythms include the circadian rhythm, the cell division cycle, the glycolytic cycle, and the yeast metabolic cycle. Genetic oscillators, biochemical oscillators, and metabolic oscillators, all linked to a cellular redox cycle, govern biological rhythms (Mellor, 2016). Biological rhythms can occur at different time scales, from seconds, to ultradian cycles (more frequent than 24 hours), to circadian rhythms (24 hours). Biological rhythms are important in temporally separating physiological processes. This separation is instrumental in responding to external conditions, including nutrient conditions, growth requirements, or the day-night cycle.

To demonstrate the definition of biological rhythm, I discuss the cell division cycle, which is well-characterised, and the glycolytic cycle, which is less well-characterised.

Cell division cycle

The cell division cycle is a series of cellular events that ensure that a parent cell divides into two progeny cells. These events include cell growth and accumulation of biomass, replication of genetic material with proofreading, and dividing the cell into two compartments. In eukaryotes, the latter event consists of karyokinesis (division of the nucleus) and cytokinesis (division of the cell). In budding yeast (*Saccharomyces cerevisiae*), the cell division cycle is divided into the G_1, S, G_2, and M phases (Fig. 1.1). In the gap phases (G_1 and G_2), the cell primarily grows and accumulates biomass. The cell replicates DNA in the S phase and in the M phase it conducts mitosis, in which chromosomes are segregated between the progeny cells.

The cell division cycle is important in unicellular organisms such as budding yeast because it is the mechanism by which the organism reproduces. Regulation of the cell division cycle is thus important (a) because the cell must only divide when necessary, (b) because the cell must have resources available for division before it divides, (c) because the cell must ensure faithful replication of DNA to prevent deleterious mutations in progeny cells, and (d) because the cell must ensure that its components are divided equally between the two progeny cells so that its progeny cells can function normally. The cell division cycle in budding yeast has checkpoints between phases to ensure that biological events from the previous phases are completed before the cell proceeds to the next phase. The most important checkpoint is START in late G_1 phase. Upon START, the cell checks whether it has the resources needed to replicate, and if requirements are met, it irreversibly commits to cell division. The cell division cycle in budding yeast is also governed by a series of gene regulatory networks that interact in a feedback loop, resulting in oscillatory expression of cyclin-CDK complexes that

Figure 1.1: Overview of the cell division cycle in budding yeast. The cell division cycle consists of G_1, S, G_2, and M phases (black arrows). The cell expresses different cyclins (coloured curves), cell division cycle regulators, as it transitions through different phases of the cell division cycle. Adapted from Adler et al. (2022).

regulate cellular events in a temporal manner (Adler et al., 2022; A. W. Murray, 2004; Orlando et al., 2008). Specifically, cyclins are proteins that sequentially accumulate and are destroyed as the cell transitions through phases of the cell division cycle. In budding yeast, these cyclins bind to and activate the cyclin-dependent kinase (CDK) Cdc28, which is constitutively expressed. The different cyclin-CDK combinations in each phase thus determines the events in the cell (Adler et al., 2022):

1. In early G_1, Cln3-Cdc28 phosphorylates Whi5, leading to activation of genes that regulate budding and DNA replication.
2. In late G_1, Cln1-Cdc28 and Cln2-Cdc28 hyperphosphorylates Sic1, leading to activation of DNA replication which marks S phase.
3. In late S phase, Clb1-Cdc28 and Clb2-Cdc28 phosphorylates Ndd1, leading to a feedback loop that activates entry into mitosis.
4. To exit mitosis, the cell activates a system to target Clb1 and Clb2 for degradation.

To coordinate cell division and metabolism with nutrient availability, budding yeast also has a system of cross-talk between nutrient signalling, growth, and the cell division cycle (Ewald, 2018). The cell integrates information from nutrient-sensing systems at START. In response to carbon deprivation, Cip1 delays START, while Msa1/2 responds to nutrient depletion and blocks START. In addition, Rim15 integrates information from the TOR and PKA pathways, which respond to nutrient depletion and other stresses, and induces cell division cycle arrest. Rim15 also responds to nutrient-poor conditions like non-fermentable carbon sources by inducing an earlier entry into START.

Furthermore, there is also evidence that sudden starvation at other phases affects cell division cycle progression, for example, through the S-phase inhibitor Sic1. The cell also coordinates internal nutrient stores with the cell division cycle. During $S/G_2/M$, Cdc28 activates the trehalase Nth1 so that the storage carbohydrate trehalose can be liquidised to provide glucose for glycolysis (Ewald et al., 2016). Additionally, during G_1/S, Cdc28 activates the lipase Tgl4 to break down storage lipids (Kurat et al., 2009). There is also evidence for an additional cyclin-dependent kinase, Pho85 (Huang et al., 2007), which inhibits the expression of genes involved in the phosphate starvation response in high levels of environmental phosphate (E. M. O'Neill et al., 1996), and also has roles in inhibiting glycogen synthase (Wilson et al., 1999).

Glycolytic cycle

The glycolytic oscillation is a biochemical oscillator in budding yeast, characterised by damped oscillations in the levels of glycolytic intermediates at a period of 40–50 seconds (Ghosh & Chance, 1964). These oscillations have been observed as a response to high-glucose conditions and in anaerobic conditions. Later studies show that levels of NADH (Lloyd, 2019; Olsen & Lunding, 2021), pH, and mitochondrial membrane potential (B. J. T. Dodd & Kralj, 2017) also oscillate at the same frequency (Fig. 1.2).

Many hypotheses have been proposed to explain the existence of glycolytic oscillations, but there is no consensus (Lloyd, 2019), though Thoke et al. (2018) proposed that these oscillations are the result of cells attempting to maintain a constant low-entropy state while remaining metabolically active. Evidence shows that glycolytic oscillations are regulated through solely biochemical means. Specifically, a high ADP/ATP ratio and the presence of fructose-1,6-bisphosphate activate the

Figure 1.2: Glycolytic cycles are characterised by oscillations in NADH (orange) and pH (blue) at a period of approximately 40–50 s, and are usually highly damped. Adapted from B. J. T. Dodd and Kralj (2017).

activity of phosphofructokinase, which then controls the flux through glycolysis, forming a negative feedback loop that causes oscillations (Ghosh & Chance, 1964; Higgins, 1964). Such a biochemical mechanism would explain how oscillations are sustained at a short timescale.

1.1.2 Mathematical basis of biological rhythms

The mathematical modelling of biological rhythms originated as simple systems of ordinary differential equations to describe negative feedback control circuits (Goodwin, 1965; Griffith, 1968). Experimental observations have then informed the development of models with finer detail. Furthermore, subsequent studies have modelled and synthesised artificial genetic circuits (Elowitz & Leibler, 2000).

The well-characterised cell division cycle has inspired models with a variety of approaches. Early models are based on a negative feedback loop of key components as identified by experimental studies. For example, Goldbeter (1991) assumed a minimal model of one cyclin, one kinase, and one protease to construct a negative

feedback loop with a delay, giving rise to stable oscillations. Such a strategy forms the basis of later models that incorporate more detail, including additional control points of the cell division cycle (K. C. Chen et al., 2004), responses to perturbations such as osmotic stress (Adrover et al., 2011), and relationship with other oscillators like the circadian rhythm (Charvin et al., 2009; Droin et al., 2019; Gérard & Goldbeter, 2012). More recent, comprehensive models include Adler et al. (2022) which is based on a system of ordinary differential equations adapted for the modelling to pheromone and osmotic shock responses, and Novak and Tyson (2022), which models the cell division cycle as a series of switches between two stable steady states whose behaviour is regulated by the CDK oscillator.

Models of glycolytic oscillations have less precision because the oscillation is less well-characterised. Most models focus on few intermediates of glycolysis that would explain the observed oscillations. Ghosh and Chance (1964) proposed a simple biochemical mechanism governed by the action of phosphofructokinase, dependent on the concentration of fructose-1,6-bisphosphate. Higgins (1964) then tested this mechanism, by describing differential equations that model six chemical reactions. Later, Termonia and Ross (1981) incorporated pyruvate kinase kinetics and levels of AMP, ADP, and ATP as part of their kinetic model based on Michaelis-Menten kinetics. Other studies focus on non-linear dynamics. Goldbeter and Lefever (1972) models the product-activated phosphofructokinase reaction, taking into account the allosteric nature of the enzyme. This model contains a single positive feedback loop as a instability-generating mechanism in a bistability model that explains oscillations. In Morán and Goldbeter (1984), the model was modified to incorporate a reaction of product recycling into substrate to explain birhythmicity, namely, the potential for oscillations of different amplitudes. Another development of the model includes a three-variable model that represents two coupled reactions each under a positive feedback loop, explaining

more complex oscillatory phenomena that could arise from pulsing of substrates (Decroly & Goldbeter, 1982). In contrast to work that models the cell division cycle, gene expression dynamics and the effects of perturbations have not been incorporated in the modelling of the glycolytic cycle.

As biological rhythms are often coupled with each other, forced and coupled oscillators have been modelled. If an oscillator is forced, it has a natural oscillation frequency, but is forced from it due to an external force applied at a regular interval. An example is the circadian clock, which is entrained to the light-dark cycle (Goldbeter & Yan, 2022). Yeast glycolytic oscillations can also be entrained via a periodic input of substrate (Boiteux et al., 1975).

Forced oscillators are closely linked to coupled oscillators, in which two oscillators are coupled to each other by certain activation or deactivation events. Two coupled oscillators tend to oscillate at a compromise frequency if the natural frequencies of each are close enough to each other. Otherwise, complex oscillations can occur: the oscillators lock to a rational ratio of frequencies — i.e. one oscillator goes through p periods while the other goes through q periods. In this case, the exact ratio depends on the ratio of the natural frequencies. Furthermore, in certain cases, chaos can occur. There is a mathematical basis in Arnold tongues (Heltberg et al., 2021). Experimental observations support this. For example, Charvin et al. (2009) showed that externally forcing cell division cycles via glucose pulsing leads to phase-locking of the cell division cycle oscillator only within a range of extrinsic periods.

1.2 Yeast metabolic cycle

1.2.1 Definition and description of the yeast metabolic cycle

The yeast metabolic cycle is an ultradian biological rhythm which has been described to entail oscillations in oxygen consumption, metabolite concentrations, transcript levels, and cellular events, at the population level. This yeast metabolic cycle is linked to the cell division cycle, but operates autonomously.

Because the yeast metabolic cycle exhibits gene-expression, biochemical, and metabolite oscillations, it can be considered as a type of biological rhythm. Specifically, the yeast metabolic cycle has gene-expression oscillators as evidenced by transcript cycling in its phases. Furthermore, it has biochemical oscillators as evidenced by changes in dissolved oxygen in the chemostat. Finally, it has metabolite oscillations as evidenced by changes in the levels of compounds that undergo redox reactions like NADH/NADPH and flavins.

History of evidence for the yeast metabolic cycle

Aspects of the YMC have been observed over decades. Nosoh and Takamiya (1962) discovered that synchronised *S. cerevisiae* cultures show oscillatory oxygen consumption. Kaspar von Meyenburg (1969) showed that gas metabolism and energy generation increase upon budding, while Mochan and Pye (1973) described a high-amplitude respiratory oscillation following a substrate shift from glucose to ethanol. Satroutdinov et al. (1992) were the first to describe the metabolic components of a 40-minute YMC for cells in continuous culture. Tu et al. (2005) first incorporated transcript cycling in the description of the YMC and defined the YMC events based on a chemostat-based investigation of growth of budding yeast on glucose-starved conditions.

Phases of the yeast metabolic cycle

Based on chemostat studies, the YMC can be divided into two major phases: an oxidative, high-oxygen consumption (OX/HOC) phase and a reductive, low-oxygen consumption (RED/LOC) phase (Fig. 1.3).

Figure 1.3: Phases of the yeast metabolic cycle, with (left) high- (HOC) and low-oxygen consumption (LOC) phases defined by changes in dissolved oxygen concentration (dO_2) over time in the chemostat and (right) oxidative (OX), reductive-building (RB) and reductive-charging (RC) phases defined by cycling of transcripts. Adapted from Mellor (2016).

Many authors (Causton, 2018; D. B. Murray et al., 2011; Slavov et al., 2011) use oxygen consumption rates, evidenced by the change of dissolved oxygen concentrations over time, as a basis to refer to the YMC as a two-phase cycle (Fig. 1.4). In contrast, some authors (Machné & Murray, 2012) base their two-phase model on the clustering of transcript level patterns. Krishna and Laxman (2018) interpret the oxidative phase as a growth state, while the reductive phase is a quiescent state.

Figure 1.4: The yeast metabolic cycle has been described as spontaneous respiratory cycles of 4–5 hours, as evidenced by regular oscillations of dissolved oxygen in the chemostat, after a starvation period. Adapted from Tu et al. (2005).

In contrast to the two-phase model, some authors identify a three-phase model with a reductive-building (RB) phase and a reductive-charging (RC) phase within the reductive phase. This three-phase model is primarily based on cellular events, including clustering of transcript trajectories (Tu et al., 2005) (Fig. 1.5) and of metabolite concentration trajectories (Tu et al., 2007).

Figure 1.5: The yeast metabolic cycle is characterised by transcript cycling. Such transcripts are divided into three clusters based on their patterns and phase relationships. The peaking of these transcripts correspond to the three (OX, RB, RC) phases of the yeast metabolic cycle. Adapted from Tu et al. (2005).

Cellular processes occur with the phases of the metabolic cycle (Fig. 1.6). In the oxidative phase, cells consume oxygen at a high rate as respiration, fermentation, and energy-demanding processes like biosynthesis and gene expression occur. Biosynthesis and associated gene expression are confirmed by increased transcripts from genes encoding components of the translation machinery and amino acid biosynthesis (Tu et al., 2005). 'Redox state' metabolites, including NADH, NADPH, glutathione (Lloyd & Murray, 2005), and flavins (FMN and FAD) (D. B. Murray et al., 2011) become most oxidised in this phase. As cells transition from the oxidative to the reductive phase, 70% of metabolite concentrations peak, when NAD(P)H autofluorescence peaks, and the DNA synthesis rate is at its maximum (Lloyd & Murray, 2006).

Figure 1.6: The yeast metabolic cycle seen as sequential scheduling of cellular processes into phases. Here, in the low-oxygen consumption (LOC) phase, the cell accumulates carbohydrates, amino acids, and solutes. In contrast, in the high-oxygen consumption (HOC) phase, the reverse is true: the cell uses its accumulated resources for biosynthesis and translation. Once reserves are exhausted, the cell resumes its LOC phase. Adapted from J. S. O'Neill et al. (2020).

In the reductive phase, cells consume oxygen at a low rate. During the reductive-building phase, activities linked to mitochondrial growth occur. In the early reductive-building phase, ethanol and acetate concentrations in the medium peak, marking a transition from oxidative respiration to glycolytic metabolism (Tu et al., 2007). There is evidence to suggest that activities linked to cell proliferation — such as initiation of the cell division cycle, DNA replication, and spindle pole activity — are gated to the reductive-building phase for both the short-period and long-period YMC. Such evidence includes budding activity and the pattern of the expression of *YOX1*, which encodes a cell division cycle repressor (Tu et al., 2005).

Finally, during the reductive-charging phase, non-respiratory metabolism and degradation processes occur to prepare the cell for the oxidative phase. This non-respiratory metabolism includes glycolysis, ethanol and fatty acid metabolism, and nitrogen metabolism. With these metabolic modes, under the regulation of the transcription factors Msn2p and Msn4p (Kuang et al., 2017), acetyl CoA accumulates so ATP can be produced in the oxidative phase (Tu et al., 2005). After acetyl CoA levels reach a threshold, it promotes histone acetylation and thus induces the oxidative phase. These metabolic pathways also optimise production of NADPH — based on the induction of *GND2* — to buffer against oxidative stress in the oxidative phase. Genes associated with protein degradation, ubiquitinylation, peroxisomes, vacuoles, and the proteosome also peak in the reductive-charging phase.

It has been hypothesised that gating activities linked to cell proliferation to the reductive-building phase creates a temporal separation between oxidative biochemical processes and the cell division cycle. This temporal separation may prevent reactive oxygen species generated by oxidative process from damaging DNA. However, measuring DNA content and oxygen consumption in cells grown

at different growth rates showed that the S phase of the cell cycle may occur in the oxidative phase if the cells have a slow growth rate (Slavov & Botstein, 2011). This may be explained by the YMC gating the early and late cell cycle independently, as evidenced by periodic localisation of the anaphase-promoting complex and mitotic exit activator Cdc14 phosphatase in metaphase-arrested cells (Y. Lu & Cross, 2010) and, conversely, the persistence of NAD(P)H cycling upon arresting of the late cell cycle by depletion of Cdc14 (Papagiannakis et al., 2017), all observed in single cells. This evidence indicates that the gating between the YMC and the cell cycle is flexible. Although there is no consensus on the function of such flexible gating, gating can aid the allocation of metabolites to different temporal phases of both cellular oscillators.

Single-cell studies of the yeast metabolic cycle

Most studies of the YMC arise from chemostat experiments, but few studies have investigated the YMC in single-cell settings using microfluidic devices. Laxman et al. (2010) was an early attempt at using microfluidics to address the bulk vs single-cell issue. They cultured strains with fluorescent gene expression reporters for each phase (OX, RB, RC) of the metabolic cycle by transferring cells from a chemostat to a microfluidic device. The study showed that low glucose levels were required for the synchrony of metabolic cycles across cells. This study also shows the presence of quiescent cells, and then proposed that during OX phase, cells decide whether to commit to cell growth or enter a quiescent state, leading to a model of two sub-populations in the culture. However, it lacks quantitative time-series analysis; rather, it reports qualitative interpretations of fluorescence images. The microfluidic device did not truly physically separate each cell individually, thus it was unable to eliminate the possibility of cell-to-cell communication via a diffusible signalling chemical.

Later microfluidics studies revealed additional features of the YMC in single cells. Papagiannakis et al. (2017) revealed that YMCs are an intrinsic feature of single cells and are autonomous with respect to the cell division cycle, based on measurements of the level of ATP and of the combined level of NADH and NADPH in single cells in microfluidic devices. Specifically, they showed that such metabolic cycles synchronise with the cell division cycle at different compromise periods as the carbon source changes, and that cell division cycle events are gated to distinct phases in the metabolic cycle. Furthermore, such metabolic cycles can progress without progress of the cell division cycle progress, both spontaneously and under chemically induced G_1 arrest.

In addition, by measuring flavin fluorescence in the cell, Baumgartner et al. (2018) demonstrated that there is strong coupling between the YMC and the cell division cycle across perturbations in nitrogen sources, showing that as nitrogen sources become poorer, the coupled oscillations oscillated at longer periods and spontaneous decoupling between the YMC and the cell division cycle became more frequent. This study importantly also demonstrated that respiration is not required for the YMC; evidence includes the observation that YMCs persist in mutants deficient in oxidative phosphorylation, and that the cell division cycle inhibitor rapamycin desynchronises the YMC and the cell division cycle.

More recently, Özsezen et al. (2019) extended the experimental data reported by Papagiannakis et al. (2017) and estimated parameters of a Kuramoto model, which suggested that the metabolic cycle separately gates START and the M phases of the cell division cycle, and that the S phase oscillator follows START without input from the metabolic cycle. This mathematical model of coupled oscillators predicted that synchrony should be maintained if the period of the metabolic cycle is perturbed; this prediction was validated experimentally using nutrient shifts. Furthermore, the model predicted that perturbation of mitotic

exit should still result in synchrony of the oscillators that remain and that additional removal S-phase exit should give a system that oscillates at a compromise frequency; such predictions were validated by dynamic depletion of Cdc14 and Cdc20.

Unlike chemostat-based studies, single-cell studies do not discuss phases as the single-cell microfluidic set-up does not allow live monitoring of transcription, and oxygen consumption rate can only be measured from chemostat cultures. Such studies define the metabolic cycle as autonomous oscillations in metabolite levels in individual cells over time (Fig. 1.7). In later chapters, I adhere to this single-cell definition because I perform a single-cell study of the yeast metabolic cycle. The two- or three-phase response as reported from chemostat-based studies may result from cellular adaptation to glucose limitation in chemostat cultures, and it is unknown whether these dynamics hold true in glucose-rich conditions, which cannot be created in a chemostat (Slavov & Botstein, 2011).

Figure 1.7: The yeast metabolic cycle seen as coordinated cycling of metabolites in the cell, generated autonomously of the cell division cycle, but are linked in permissive conditions. Adapted from Zylstra and Heinemann (2022).

Contrasting the yeast metabolic cycle with the glycolytic oscillation

The yeast metabolic cycle is longer and is more robust than a similar biological oscillator, the glycolytic oscillation. The glycolytic oscillation has a period of approximately 40 seconds (Olsen et al., 2009). In contrast, the yeast metabolic cycle has been described, using various definitions, to either exhibit a 40-minute short-period cycle (C. M. Li & Klevecz, 2006; Lloyd & Murray, 2005, 2007), or a long-period cycle, which is most commonly described to be 4–5 hours (Tu et al., 2005, 2007), but also ranges between 1.4 hours to 14 hours, depending on the chemostat dilution rate (Beuse et al., 1998). Glycolytic oscillations are highly damped, but yeast metabolic oscillations are robust, lasting for weeks (Lloyd & Murray, 2007). Additionally, glycolytic oscillations have been observed in anaerobic conditions (Lloyd, 2019), but yeast metabolic cycles have been observed in aerobic conditions. Moreover, glycolytic oscillations are characterised by fluctuations in NADH fluorescence and glycolytic intermediates. However, yeast metabolic cycles consist of fluctuations in NADH fluorescence, flavin fluorescence, and ATP concentrations as well as biosynthetic intermediates in TCA cycle, amino acid, and nucleic acid metabolism (Tu et al., 2007).

1.2.2 Yeast metabolic cycles under perturbations

Perturbations in growth conditions can affect the length of the metabolic cycle and its relationship with other cellular events. The long-period cycle may vary from 1.4 to 14 hours under constant chemostat conditions with various changes in nutrient media and dilution rates (Causton, 2018), although 24-hour cycles have also been observed when temperature cycles were imposed, and immediately after (Eelderink-Chen et al., 2010). The main nutrient perturbations that have been studied are perturbations in carbon sources and in nitrogen sources.

Perturbations in growth conditions

Perturbations in carbon sources are well-documented. Lower glucose concentrations prolong the metabolic cycle, as evidenced by both chemostat studies that assess the effect of changing the dilution rate (Burnetti et al., 2016; J. S. O'Neill et al., 2020) and by single-cell studies that assess the effect of glucose concentrations in the limiting region (Papagiannakis et al., 2017). For example, increasing the dilution rate in chemostats increases the growth rate, and in turn decreases the duration of the low oxygen consumption phase while the duration of the high oxygen consumption phase holds constant (J. S. O'Neill et al., 2020; Slavov & Botstein, 2011). This effect is pronounced in the region of glucose limitation; this was evidenced by a small increase in metabolic cycle period in single cells when the glucose concentration was decreased from $10\,\mathrm{g\,L^{-1}}$ to $0.05\,\mathrm{g\,L^{-1}}$ compared to a large increase in period when glucose concentration was decreased from $0.05\,\mathrm{g\,L^{-1}}$ to $0.025\,\mathrm{g\,L^{-1}}$ (Papagiannakis et al., 2017).

Such experimental observations could by explained by models such as Jones and Kompala (1999), which suggest that as the dilution rate is decreased, metabolic oscillations acquire greater amplitudes and longer periods, but if it is low enough, metabolism becomes entirely oxidative, and metabolic oscillations disappear. However, non-fermentable carbon sources like pyruvate give long-duration metabolic cycles in single cells, comparable to cells under limiting levels of glucose (Papagiannakis et al., 2017).

In addition, bulk depletion or addition of a carbon source can reset the phase of the YMC. Adding a bulk carbon source such as acetate, ethanol, or acetaldehyde can reset the phase of the YMC (Krishna & Laxman, 2018; Kuang et al., 2017) or eliminate it (Jones & Kompala, 1999).

Perturbations in nitrogen sources are less well-studied. Baumgartner et al. (2018), based on single-cell observations, suggest that decreasing nitrogen concentration prolongs the YMC, as evidenced by longer flavin oscillations when cells are grown on lower concentrations of yeast nitrogen base (YNB) media or on urea, a non-preferred nitrogen source.

In addition, perturbations outside nutrient sources also affect the YMC. For example, externally applied hydrogen peroxide, as a source of oxidative stress, shifts the YMC to the oxidative phase (Amponsah et al., 2021). The oscillation period is insensitive to temperatures from 25 °C to 35 °C and media pH values from 2.9 to 6.0 (Lloyd & Murray, 2005) — though the period of dissolved-oxygen oscillations decrease as pH decreases to below 2.9 and the oscillations disappear when conditions are too acidic (J. S. O'Neill et al., 2020). Additionally, the dissolved-oxygen oscillations are robust to media potassium ion concentrations varying between 1 and 10 mM, but disappear when the potassium concentration falls below 1 mM (J. S. O'Neill et al., 2020).

Genetic perturbations

Although the molecular basis of the yeast metabolic cycle is not well-characterised, gene deletions shed light on it. Genes that control the cell division cycle and metabolism have been deleted in various studies.

Several deletions have been shown to remove the metabolic oscillations in chemostats: *zwf1*Δ (Tu et al., 2007), *gsy2*Δ, and *gph1*Δ (J. S. O'Neill et al., 2020) (Figs. 1.8a–1.8b).

Figure 1.8: Effects of genetic perturbations on the yeast metabolic cycle, as evidenced by the dissolved oxygen cycles of deletion strains observed in the chemostat. **(1.8a)** *zwf1*Δ shows no dissolved oxygen cycles. Adapted from Tu et al. (2007). **(1.8b)** *gsy2*Δ and *gph1*Δ show no dissolved oxygen cycles. Adapted from J. S. O'Neill et al. (2020). **(1.8c)** *rim11*Δ shows shorter dissolved oxygen cycles. Adapted from Causton et al. (2015). **(1.8d)** *swe1*Δ shows shorter dissolved oxygen cycles and a modified coupling ratio between the metabolic and cell division cycle oscillators. Adapted from Causton et al. (2015). **(1.8e)** *tsa1*Δ *tsa2*Δ shows dissolved oxygen cycles of a different shape. Adapted from Causton et al. (2015).

ZWF1 codes for glucose-6-phosphate dehydrogenase and is thus responsible for entry into the pentose phosphate pathway and subsequently a major source of NADPH generation. Therefore, deleting this gene may impair control of cellular redox. However, because of its role, this gene deletion impairs adapting to oxidative and pH stress and also causes methionine auxotrophy, so it may be difficult to draw conclusions from this deletion in particular. Furthermore, Idp2p and Ald6p catalyse reactions that generate NADPH and have shown to compensate for the loss of *ZWF1* when cells are grown on lactate plates or on liquid cultures with glucose as the carbon source (Minard & McAlister-Henn, 2005). This observation therefore raises the question of just how important *ZWF1* is to the yeast metabolic cycle, and to what extent is NADPH generation needed for control of cellular redox.

On the other hand, *GSY2* has a role in glucose storage and *GPH1* has a role in glycogen mobilisation. The absence of dissolved oxygen cycles in the associated deletions thus suggests that cycling of carbohydrate stores may be needed for the function of the metabolic cycle. However, metabolic oscillations have been observed in high-glucose conditions (Baumgartner et al., 2018; Papagiannakis et al., 2017) in which glycogen synthesis is repressed. These observations therefore suggest that glycogen cycling may play a more minor role in defining the yeast metabolic cycle and another nutrient cycling phenomenon may be more responsible.

In addition, *MSN2* and *MSN4* have been shown to regulate acetyl CoA accumulation in the reductive-charging phase, as evidenced by the lack of YMCs in deletion strains (Kuang et al., 2017). This observation suggests that genes involved in signalling pathways also play an important role in the integrity of the metabolic cycle.

Additionally, other deletions have been shown to change the frequency or shape of dissolved oxygen cycles. Causton et al. (2015) provide several examples, of which I discuss *rim11*Δ, *swe1*Δ, and *tsa1*Δ *tsa2*Δ.

Rim11p is the yeast homolog of the GSK3β serine/threonine kinase, which regulates metabolism and plays a role in setting the speed of the circadian clock. The *RIM11* deletion has been shown to give shortened periods of dissolved-oxygen metabolic cycles in the chemostat, thus pointing towards a common mechanism for both biological oscillators (Fig. 1.8c).

Swe1p is a conserved cell division cycle regulator that functions at the G2/M checkpoint and has roles in coupling the cell division cycle with the circadian rhythm. Deleting *SWE1* also resulted in shortened periods of dissolved-oxygen metabolic cycles but with the same rate of DNA replication, suggesting a dysregulation in the coupling between the yeast metabolic cycle and the cell division cycle (Fig. 1.8d).

Tsa1p and Tsa2p are paralogous cytoplasmic thioredoxin peroxidases that cooperate in the peroxiredoxin-thioredoxin system to eliminate reactive oxygen species and have been shown to be a marker for circadian rhythms. A double deletion of the two genes still results in metabolic cycles, but with an additional burst in high oxygen consumption during what would otherwise be the reductive-charging phase, showing that the peroxiredoxin-thioredoxin system is instrumental in the integrity of the yeast metabolic cycle (Fig. 1.8e). In addition, Amponsah et al. (2021), using the chemostat, showed that the oxidation of peroxiredoxin cycles during the YMC, with a corresponding cycling of hydrogen peroxide. They also confirmed that inactivating peroxiredoxins disrupts the metabolic cycle and decouples it from the cell division cycle, through inducible degradation of an additional cytosolic peroxiredoxin Ahp1p.

Taken together, these deletion studies show that regulators of other biological rhythms and of redox metabolism play a role in the regulation of the YMC. However, few genetic perturbation studies have been attempted in single-cell studies. The most significant is in Baumgartner et al. (2018), in which by deleting genes (*atp5*Δ, *cyt1*Δ) required for respiration, they showed that metabolic cycling does not require respiration.

1.2.3 Modelling the yeast metabolic cycle

Mathematical models have been developed to explain the aspects of the YMC. An early model is Jones and Kompala (1999) which uses differential equations to simulate dynamic competition between three modes of metabolism: fermentation, glucose oxidation, and ethanol oxidation. This model predicts spontaneous generation of oscillations in dissolved oxygen, cell mass, and storage carbohydrates in continuous cultures. This prediction is consistent with chemostat-based studies of the yeast metabolic cycle. Furthermore, the model predicts that, within a window of dilution rate values, if the dilution rate decreases, then the dissolved oxygen oscillations increase in amplitude and period. The increase in period agrees with experimental studies such as J. S. O'Neill et al. (2020). However, the model also predicts oscillations in the extracellular concentrations of glucose and ethanol. In theory, such oscillations can only occur if there is a nutrient-consuming prey species that is in turn consumed by two competing predators, or if there are two competitors and an inhibitor added to the chemostat that inhibits only one of the competitors (Smith & Waltman, 1995).

Krishna and Laxman (2018) use a frustrated bistability model to describe a relaxation oscillator that explains how a population of yeast cells switches between quiescent and growth states when faced with a limited amount of metabolic resources. This model assumes that the cells retain hysteresis of their current state

and posits that cells of two populations communicate through diffused acetyl-CoA to sustain population-level oscillatory behaviour. Burnetti et al. (2016) also propose that yeast cells committed to the metabolic cycle secrete metabolites that induce other cells to enter the metabolic cycle, provided that they have enough storage carbohydrates. Taken together, the models provide an attractive cell-to-cell signalling explanation for the population-level behaviour observed in the chemostat. However, such an explanation does not explain the presence of metabolic cycling in single-cell conditions in which cells are physically separated and thus signalling between cells cannot occur. Though, autonomous generation of metabolic cycles and synchrony of metabolic cycles in a population can each arise from mechanisms that are independent of each other.

Based on single-cell experimental observations, Özsezen et al. (2019) use a deterministic Kuramoto model to explain the interaction between one metabolic oscillator and three cell cycle oscillators at different stages. The study uses growth on different carbon source conditions to determine parameters that define the natural frequencies of the cell division cycle oscillators and the strength of the coupling between the four oscillators. Parameter optimisation predicts that the metabolic cycle most strongly influences the START point of the cell division cycle, and more weakly influences the M and S phases, while the three stages of the cell division cycle negligibly influence each other (Fig. 1.9). Under perturbations, the model system remains stable but shows a shift in oscillation frequency, agreeing with experimental observations, and also predicts the effects of Cdc20 and Cdc14 dynamic depletions. However, a key criticism of this model-based study is that the Kuramoto model makes simplistic assumptions about the oscillators, which may be unrealistic, given how little is known about the mechanistic basis of the metabolic oscillator.

Figure 1.9: The relationship between the yeast metabolic cycle and the cell division cycle can be seen as a system of coupled oscillators. Namely, the cell division cycle is modelled as three oscillators (START, S, M), and the yeast metabolic oscillator (MET) gates entry into START and M phases, while progression from START to S is independent of the metabolic cycle. Adapted from Özsezen et al. (2019).

Taken together, modelling approaches have been able to predict some aspects of the metabolic cycle. However, most models focus on specific aspects to the detriment of other experimental observations, and none sufficiently reconcile observations from chemostat-based and single-cell studies. Constructing more accurate models is complicated by the scant knowledge of the mechanistic basis of the yeast metabolic cycle.

1.2.4 Big picture/Hypothesis: a nutrient sensor than entrains the cell division cycle?

From existing evidence, we can create a big picture of the yeast metabolic cycle. The yeast metabolic cycle is an autonomous biological oscillator that operates at a range of frequencies in response to a range of permissive growth conditions, as evidenced by how extreme conditions impair the oscillator. Based on chemostat-based studies, such extreme conditions include poor nutrient quality, media being too acidic, and potassium ion concentration being too low (J. S. O'Neill et al., 2020). However, there is reason to believe that the metabolic oscillator can

function in some conditions previously deemed to be unfavourable. For example, single-cell studies show that yeast cells show metabolic oscillations in high-glucose conditions. Within the permissive growth conditions, different conditions affect the frequency of the metabolic cycle. For example, a low concentration of glucose or nitrogen source results in longer cycles, and bulk addition of certain compounds can reset the phase of the metabolic cycle. These observations support the idea that the metabolic oscillator includes the functionality of a nutrient sensor.

The observations suggest that the yeast metabolic cycle creates windows of opportunities for the cell to commit to START if conditions are favourable, for example, good carbohydrate or lipid stores. Thus, this oscillator acts as a timing mechanism for cellular processes, most importantly, the cell division cycle, biosynthetic processes, and redox reactions. The relationship between the metabolic cycle and the cell division cycle is governed by the mathematical basis of coupled oscillators. Specifically, there is a small window of frequencies in which both oscillators can be phase-locked, but more complicated relationships, such as multiple metabolic cycles per cell division cycle, may exist.

1.2.5 Disputes and unresolved questions with the yeast metabolic cycle

Chemostat vs single-cell studies

There is a dispute of whether the same conclusions can be drawn from chemostat-based studies and from single-cell based studies. Reconciling the two types of studies is difficult because the readouts and conditions are different: chemostat studies produce dissolved oxygen and transcript cycling readings, while single-cell experiments cannot report on dissolved oxygen and chiefly report metabolite cycling. This leads to differing definitions of the YMC. Some authors (Causton,

2018; Laxman et al., 2010) only use the term metabolic cycle to refer to synchronised cycles of dissolved oxygen concentrations observed in chemostat cultures that must have gone through a starvation phase. In contrast, single-cell studies (Baumgartner et al., 2018; Zylstra & Heinemann, 2022) define the metabolic cycle as metabolite cycling and sequences of cellular events. This is because in such settings, the cells are not synchronised by diffusible chemical signals and dissolved oxygen concentrations cannot be measured. Additionally, these studies do not include the requirement of a starvation phase as part of their definition as they show that cell exhibit metabolite cycling even without having gone through a period of starvation.

I argue that there are three caveats to chemostat-based studies: the experimenter cannot assume that the chemostat is in steady-state, the chemostat obscures contributions from sub-populations, and the chemostat imposes glucose starvation. These caveats affect the interpretation of YMC studies.

There is a wide assumption that the chemostat is in steady-state, but it may not be true. A mathematical model shows that levels of solutes change over time (Jones & Kompala, 1999). In addition, observations of the metabolic cycle in chemostats may reflect individual cells' responses to the initial starvation imposed at the start of chemostat-based studies. The subsequent response to regularly changing media conditions could explain temporal segregation of physiological processes in phases of the YMC, and may not reflect cell-autonomous behaviour. In other words, the conditions of the chemostat may force the population of cells to behave in a certain way. However, temporal segregation of physiological process has also been reported in single-cell studies (Takhaveev et al., 2023), suggesting that the cycling of solutes in the chemostat could affect some, but not all, temporal aspects of the metabolic cycle.

The chemostat obscures contribution of sub-populations of cells (Fig. 1.10). Burnetti et al. (2016) suggest that sub-populations within the yeast culture that enter the yeast metabolic cycle in a staggered manner can be responsible for the metabolic cycle and the cell division cycle appearing coupled one-to-one in the chemostat while having different periods. This proposition was informed by observing one pulse of DNA replication per YMC and observing that a fraction of cells that exhibited the YMC initiated cell division. Contributions from sub-populations of cells were further highlighted by Bagamery et al. (2020), who used a microfluidic platform to show that a group of genetically identical yeast cells divide themselves into two populations. Such a bet-hedging strategy results in some percentage of the population surviving in a glucose-starved or a glucose-rich condition, beneficial for long-term population survival. Taken together, it is possible that phenotypically different sub-populations in the chemostat culture may partially explain the observations in the chemostat so far.

In addition, there is the question of whether cells individually generate the metabolic cycle or is a diffusible chemical responsible for synchrony, as proposed by Krishna and Laxman (2018). Furthermore, Smith and Waltman (1995) showed, theoretically, if a chemostat has two competing species, and an inhibitor which inhibits one of the species is added, then the chemostat can generate oscillations. In the case of the yeast metabolic cycle, the competitors can be genetically identical sub-populations of the yeast cells that have different levels of sensitivities to an inhibitor, perhaps a metabolic by-product. Bulk culture set-ups, including chemostats, are not able to address questions about cell sub-populations and autonomy of the metabolic cycle. However, single-cell set-ups may fill in such a technical gap.

Figure 1.10: Model for how sub-populations of cells may account for oscillations of dissolved oxygen observed in the chemostat. Cells may enter the cell division cycle (pink-blue-red lines), gated by the metabolic cycle, in a staggered manner, and the combined effect of all cells may explain dissolved oxygen concentrations (black lines, top row). HOC refers to the high oxygen consumption phase, while LOC refers to the low oxygen consumption phase. Adapted from Mellor (2016).

Finally, the chemostat imposes glucose starvation, and single-cell studies with different carbon sources give a different picture in terms of metabolic requirements. Chemostat studies and related models suggest that glucose starvation and oxidative metabolism are required for oscillations in dissolved oxygen level that define YMCs. NAD(P)H oscillations have been recorded in non-fermentative conditions, such as pyruvate or low-glucose media (Papagiannakis et al., 2017). However, NAD(P)H (Özsezen et al., 2019; Papagiannakis et al., 2017) and flavin (Baumgartner et al., 2018) oscillations still occur in constant high-glucose conditions, and only within a window of periods, in contrast to the 1.4–14 hour range reported for chemostat-based studies. Furthermore, *ATP5* and *CYT1* deletions that impair oxidative respiration do not remove single-cell flavin-based metabolic oscillations (Baumgartner et al., 2018), thus giving additional evidence that oxidative metabolism is not required for the YMC.

Single-cell microfluidic studies are well-positioned to address the limitations of the chemostat, although there have been only few studies (Baumgartner et al., 2018; Laxman et al., 2010; Özsezen et al., 2019; Papagiannakis et al., 2017), discussed earlier in Section 1.2.1. These single-cell studies address a small fraction of the knowledge covered by chemostat studies, and further such studies are required.

Molecular and genetic mechanisms

There are unknowns in the molecular mechanism that drives YMCs. Genome-wide transcript cycling has two superclusters that correspond to the oxidative and reductive-building phases (Machné & Murray, 2012). However, there has been no genome-wide analysis of genes that influence cycling (Mellor, 2016), though some genes seem to have key roles. Although proteome analysis has suggested

two clusters of cycling proteins that correspond to the HOC and LOC phases (J. S. O'Neill et al., 2020), the small proportion of cycling proteins relative to the whole proteome suggests a post-translational regulation mechanism (Zylstra & Heinemann, 2022), which is yet to be elucidated.

Metabolome cycling may play a role in the metabolic cycle and can explain sequential scheduling of biosynthesis, but the evidence so far is indirect as it is based on the cell division cycle. Campbell et al. (2020) showed that lipid biosynthesis in budding yeast is periodic with the cell division cycle and peaks during S phase, as evidenced by an increase in the number of metabolites implicated in lipid metabolism in such phases, based on metabolomics analysis of prototrophic cells with synchronised cell division cycles. Ewald et al. (2016) also show that the cell division cycle machinery regulates trehalose mobilisation, showing the coupling between carbohydrate store levels and cellular oscillators. They also showed that lipid metabolism increased during S/G2/M, likely due to the synthesis of new cell membranes during bud growth, as evidenced by pathway enrichment analysis. Based on the coupling between the yeast metabolic cycle and the cell division cycle, lipid store cycling and, perhaps to a lesser extent, carbohydrate store cycling are likely instrumental to the yeast metabolic cycle. Though, investigation of how an impairment in lipid use affects the yeast metabolic cycle in single cells is needed to prove that such cycles are responsible for the metabolic cycle.

1.2.6 Implications of the metabolic cycle

The YMC shares regulatory mechanisms with the cell division cycle and the circadian rhythm, leading to the question of whether the metabolic cycle reflects a fundamental system that serves as the evolutionary or functional origin of all biological rhythms.

Similar metabolic cycles have been described in other organisms. *E. coli* shows oscillations in NAD(P)H fluorescence coupled to its cell division cycle, as evidenced by time-lapse microscopy of single cells (Zhang et al., 2018). Addition of glucose or hydrogen peroxide to the medium results in global changes in autofluorescence, reflecting a response to nutrient conditions. In addition, metabolic cycles have been observed in mammalian cells. For example, Zhu (2022) describes a 12-hour metabolic cycle in liver cells that includes sequential scheduling of metabolic processes into energy homeostasis, genetic integrity maintenance processes, immune response, and gene expression — linking the processes to the circadian rhythm and the whole cycle to a more general 12-hour mammalian ultradian clock. Importantly, this hepatic metabolic cycle operates independently from the spatial organisation of cells in the liver, reminiscent of the cell-to-cell independence of the yeast metabolic cycle. In addition, HeLa cells with synchronised cell division have been shown to exhibit both NAD(P)H and ATP oscillations throughout the cell division cycle (Ahn et al., 2017), but the literature is conflicted about these oscillations' dynamics across different mammalian cell types (Zylstra & Heinemann, 2022). There is reason to believe that a wide range of organisms exhibit biochemical phenomena similar to the yeast metabolic cycle, as the aims of controlling cell division to match environmental conditions along with the temporal coordination of biosynthesis and cellular redox state with cell division should be fundamental goals that apply to multiple domains of life.

Metabolic oscillations may be the origins of biological timekeeping mechanisms. Lloyd and Murray (2007) assert that ultradian oscillations form the basis of longer-period biological oscillators like the circadian rhythm or the cell cycle, based on temperature compensation and sensitivity of the period. Circadian rhythms can occur in cells of eukaryotes without transcription (J. S. O'Neill & Reddy, 2011; J. S. O'Neill et al., 2011), refuting the idea that gene circuits are

responsible for such rhythms. Additionally, the eukaryotic cell cycle evolved before cyclin-dependent kinases (Papagiannakis et al., 2017), so metabolic oscillations may have served to regulate the cell cycle before cyclin-dependent kinases evolved. Furthermore, YMCs share mechanisms with the circadian oscillator (Arata & Takagi, 2019; Causton et al., 2015), suggesting a common evolutionary origin. Thus, studying YMCs may shed light on the evolution of biological rhythms.

A question about operation principles that unites all biological oscillators is: why have oscillations? The answers may differ for each type of biological oscillator.

The answer is clear for circadian rhythms: circadian clocks evolved as an adaptation to the Earth's 24-hour rotation, and allow organisms to match biological processes that benefit from light or warmth to the time of day at which they occur (Millar, 2004). The selective advantage of such a system is highlighted by how circadian clocks of similar, negative-feedback gene circuit architectures evolved independently at least four times across kingdoms (A. N. Dodd et al., 2005). As a specific example, *Arabidopsis thaliana* plants with a clock period matched to the environment contain more chlorophyll, fix more carbon, grow faster, and survive better, as opposed to mutants with clock lengths that do not match environmental light-dark cycles (A. N. Dodd et al., 2005). The explanation is that light-harvesting complex proteins and chlorophyll are unstable in their unbound state. So, synthesising these components in sync with the light-dark cycle is advantageous for the plant. The plant can only synchronise synthesis with the light-dark cycle with correct anticipation of dawn and dusk, and a circadian clock that matches the environmental light-dark cycle allows this to happen.

The importance of coordinating the sequence of events in the cell division cycle was discussed in Section 1.1.1. Briefly, the cell division cycle is an energy-intensive and resource-intensive process that engages all compartments of the cell. In addition, maintaining genetic fidelity is important for ensuring that progeny cells are functional; therefore, cells need a robust system of tight control of the cell division cycle. The importance of the cell division cycle is highlighted by the conserved design of the cell division cycle across kingdoms. For example, cyclin-dependent kinases (CDKs) differ in their number in different organisms — one classical CDK in budding yeast, but at least 11 classical CDKs in humans (Malumbres et al., 2009) — but their structure and function are conserved. The importance of linking the cell division cycle with cellular resource use is highlighted by the cell's systems of coupling the cell division cycle machinery with metabolic processes (Salazar-Roa & Malumbres, 2017). In particular, oxidative phosphorylation peaks upon S- and M-phase entry in plants and glycolysis peaks upon S-phase entry in lymphocytes. In addition, the CDKs regulate cycles of mitochondrial fusion and fission to ensure that mitochondria in the parent cell are divided to progeny cells based on their volumes. The importance of the cell division cycle control system is further highlighted by the result of impairments in the genetic control of the cell division cycle: uncontrolled cell proliferation, characteristic of cancers in multicellular organisms.

The question of 'why have oscillations?' has been asked of the yeast metabolic cycle — specifically, why cycle metabolites and why cycle transcripts? It has been proposed that cycling of metabolites and metabolic activity serve to create 'just-in-time' biosynthesis of compounds when they are needed in the yeast metabolic cycle (Zylstra & Heinemann, 2022). For example, the HOC phase of the metabolic cycle is driven by the high energetic demands of protein synthesis in this phase, which is closely linked to the G_1 phase of the cell division cycle (J. S. O'Neill et al.,

2020), and the evidence includes high protein synthesis rate and high ribosomal protein abundance. The rationale for cycling transcripts is less clear, however. Earlier studies propose that transcript cycling in clusters that peak according to three metabolic cycle phases reflect the metabolic demands in each phase and the translation of such transcripts lead to changes in metabolite concentrations (Tu et al., 2005). However, Feltham et al. (2020) showed that although transcripts levels cycle, protein levels chiefly remain constant, but post-translational modifications cycle instead. This study then proposes that cyclic metabolic state changes cause post-translational modifications which then coordinate the metabolic cycle with cellular processes. The study therefore suggests that transcript cycling is effect of the metabolic cycle and have roles in the chromatin environment. Further clarification of the mechanistic basis of the yeast metabolic cycle is needed to answer the 'why have oscillations' question.

1.3 Flavins and flavoproteins

1.3.1 Introduction to cellular autofluorescence

Cellular autofluorescence is the intrinsic fluorescence of a cell without fluorescent tags. It is caused by the autofluorescence of compounds that have light emission properties (Maslanka et al., 2018). Such endogenous fluorophores include co-enzymes, vitamins, and amino acids with aromatic chemical groups, flavins being one of them. However, autofluorescence poses a difficulty in cellular microscopy because its wavelengths can overlap with other fluorophores and therefore it is difficult to draw biochemical conclusions from the fluorescence signal alone. For example, flavin autofluorescence overlaps with the spectrum of the fluorescent

glucose analogue 6-NBDG (Maslanka et al., 2018), and thus interferes with studies that use this analogue to study glucose uptake. Cellular autofluorescence can indicate the physiology and metabolism of the cell. Autofluorescence thus offers an easy way to monitor cell physiology without engineering genetic constructs.

1.3.2 Biochemical basis of flavins and flavoproteins

Flavins are a group of organic compounds that share an aromatic moiety that allows redox reactions (Fig. 1.11). Specifically, the flavin moiety can exist in the oxidised, semiquinone, or reduced states. Flavins thus function as electron carriers in the cell.

Figure 1.11: Chemical structure of FMN and FAD, with redox states of the aromatic flavin moiety shown. Adapted from Patel (2006).

In *S. cerevisiae*, flavin is present as FMN and FAD, which function as prosthetic groups in flavin-dependent proteins, or flavoproteins, whose genes account for 1.1% of the genome (Gudipati et al., 2014). FMN and FAD can be covalently bound to these proteins or be free (Mewies et al., 1998). FAD is a co-enzyme and has major roles in transferring electrons from the tricarboxylic acid cycle to the mitochondrial electron transport chain.

Flavins in *S. cerevisiae* are derived from riboflavin (Fig. 1.12). Riboflavin can be synthesised *de novo* from purine biosynthesis and the oxidative pentose phosphate pathway (Fig. 1.13). Based on the metabolism of flavins, the cell only synthesises new flavin for synthesis of FMN and FAD. Therefore, monitoring of flavins monitors the combined pool of FMN and FAD in their oxidised states.

Figure 1.12: Simplified schematic of biosynthesis of flavins and detection of the oxidation states in fluorescence microscopy.

From a technical standpoint, the redox states of flavins reflect the emission and absorption of electromagnetic radiation by the flavin moiety. The redox biochemistry of flavins give rise to fluorescence, so monitoring flavin autofluorescence monitors the redox state of the cell. Flavins, in their oxidised forms (FMN and FAD), have a peak excitation frequency of ≈460 nm and a peak emission frequency of

≈535 nm (Maslanka et al., 2018; Wagnieres et al., 1998), displayed in Fig. 1.14. Comparison of *in vivo* autofluorescence in mammalian cells and the fluorescence spectrum of riboflavin in PBS confirms this fluorescence behaviour (Aubin, 1979). In contrast, the reduced forms $FMNH_2$ and $FADH_2$ have negligible fluorescence (Masters, 1994). Both chemostat-based (D. B. Murray et al., 2011; Sasidharan et al., 2012) and single-cell microfluidic studies (Baumgartner et al., 2018) have monitored flavin autofluorescence to study the YMC.

Descriptions of key flavoproteins and their roles

Gudipati et al. (2014) describe 68 genes that code for 47 flavoproteins in budding yeast (Fig. 1.15). Of these, 35 require FAD, 15 require FMN, and 3 require both. In budding yeast, most flavins sit in the active site without covalent bonding. The biochemical and enzymatic properties of many flavoproteins are poorly characterised (Koch et al., 2017).

Figure 1.13: Reference pathway for biosynthesis of riboflavin and derivatives, KEGG pathway database (Kanehisa et al., 2023).

Figure 1.14: Fluorescence spectrum of riboflavin, (dotted line) excitation and (solid line) emission spectra shown, FPbase (Lambert, 2023).

The most abundant flavoproteins catalyse redox reactions (Table 1.1). Specifically, these reactions include reduction of reactive chemical species to respond to oxidative stress — though, not all enzymes involved in the response to reactive chemical species have flavin co-factors. Additionally, the reactions include biosynthetic reactions. Many of these reactions require NADPH or NADH to donate electrons, suggesting a link between flavins and NAD(P)H in regulating the cellular redox state. In particular, Oye2p catalyses the NADPH redox reaction, thus providing a link between flavins and NAD(P)H. One exception flavoprotein is Ilv2p, which does not catalyse a redox reaction. It has been hypothesised that an ancestral form of Ilv2p catalysed a redox reaction, but the argument is weak because it is inferred from the presence of FAD (Pang et al., 2002). To maintain the cellular redox state, it is thus reasonable to assume that the redox equilibrium of all flavoprotein-catalysed reactions are in the same direction at any point of the YMC. Supporting

Figure 1.15: Flavoproteins (blue bars) shown by abundance (Ho et al., 2018), with Rps7ap (grey bar) shown as reference. Only the 17 most abundant flavoproteins are shown.

this, Siano and Mutharasan (1989) show that NAD(P)H fluorescence and the fluorescence of lipoamide dehydrogenase, a flavoprotein, indicate simultaneous reduction in response to lowered dissolved oxygen. They further show redox equilibrium in both fluorophores in response to glucose addition.

It is important to rule out the possibility that flavin cycling is merely a function of the cell division cycle to make sure that flavin monitoring monitors the YMC. None of these flavoproteins are strictly cell division cycle proteins, but this does not exclude cycling of flavin autofluorescence linked to the cell division cycle. For example, fatty acid synthesis proteins should cycle along with the cell division cycle as cell synthesises more plasma membranes.

Protein	Name	Reaction catalysed	Reference
Fas1	beta subunit of fatty acid synthetase	acetyl−CoA + malonyl−CoA + NADPH + ATP $\longrightarrow$ palmitate	Singh et al. (2020)
Yhb1	nitric oxide oxidoreductase	$2\,NO + 2\,O_2 + NAD(P)H \longrightarrow 2\,NO_3{}^- + NAD(P)^+ + H^+$	Bonamore and Boffi (2008)
Ura1	dihydroorotate dehydrogenase	dihydroorotic acid + fumarate $\longrightarrow$ orotic acid + succinate	Zameitat et al. (2007)
Pst2	NAD(P)H-quinone oxidoreductase	$NAD(P)H + H^+ +$ quinone $\longrightarrow NAD(P)^+ +$ hydroquinone	Koch et al. (2017)
Trr1	cytoplasmic thioredoxin reductase	$H^+ + NADPH +$ thioredoxin disulfide $\longrightarrow NADP^+ +$ thioredoxin	Machado et al. (1997)
Ilv2	acetolactate synthase	2 pyruvate $\longrightarrow$ 2-acetolactate $+ CO_2$	Pang et al. (2002)
Oye2	NADPH oxidoreductase	$NADPH + H^+ +$ acceptor $\rightleftharpoons NADP^+ +$ reduced acceptor	Odat et al. (2007)
Dld3	2-hydroxyglutarate transhydrogenase	D−2-hydroxyglutarate + pyruvate $\longrightarrow$ α-ketoglutarate + lactate	Becker-Kettern et al. (2016)
Pdx3	pyridoxine phosphate oxidase	pyridoxamine 5-phosphate + $H_2O + O_2 \longrightarrow$ pyridoxal 5-phosphate + $NH_3 + H_2O$	Tsuge et al. (1979)
Erg1	squalene epoxidase	squalene $+ H^+ +$ $NADPH + O_2 \longrightarrow$ 2,3−oxidosqualene + $NADP^+ + H_2O$	Satoh et al. (1993)
Lpd1	dihydrolipoamide dehydrogenase	dihydrolipoamide + $NAD^+ \longrightarrow$ lipoamide + $NADH^+ + H^+$	Morrison (2021)

Table 1.1: Roles of the most abundant flavoproteins.

1.3.3 Flavins and flavoproteins in the yeast metabolic cycle

Flavin fluorescence can be used to monitor the metabolic cycle. The biological basis of flavins justifies this use. Flavins are linked to NAD(P)H via nitric oxide oxidoreductase (Yhb1p), as discussed in Section 1.3.2, and NAD(P)H cycles have been implicated in bulk-culture (Tu et al., 2005) and single-cell (Papagiannakis et al., 2017) studies of the YMC, as discussed in Section 1.2.1. The oxidation of flavin is at its maximum at the start of the reductive state of the YMC, as evidenced by how flavin fluorescence peaks just before dissolved oxygen concentration in the chemostat (D. B. Murray et al., 2011; Sasidharan et al., 2012). Riboflavin abundance in the cell has been shown to oscillate and peak in the oxidative state of the YMC, while FAD abundance is at its maximum in the reductive-building phase, as evidenced by metabolic profiling of extracts from chemostat cultures taken at evenly-spaced intervals (Tu et al., 2007).

Flavoproteins may have roles linked to the YMC. The most abundant is Fas1p (fatty acid synthetase). Because there is evidence that cycles of fatty acid stores are implicated in metabolic cycling in yeast (Campbell et al., 2020), it is likely that fatty acid synthetase is heavily implicated. Following this, the second most abundant is Yhb1p, which may play a major role as discussed earlier.

So, for these reasons, I expect flavin autofluorescence to be oscillatory and be a useful readout of the yeast metabolic cycle. Few studies have characterised how such flavin oscillations respond to changing nutrient conditions or to gene deletions. Thus, filling in this knowledge gap is an avenue for further research.

Nevertheless, there are caveats to using flavin autofluorescence. Riboflavin fluorescence is captured too, though its intracellular abundance is two orders of magnitudes lower than that of FMN and that of FAD, and these two flavin derivatives are present at the same order of magnitude (Tu et al., 2007). In addition, different

concentrations of riboflavin influence the autofluorescence signal and influence the physiological state of the cell (Maslanka et al., 2018). However, the experimenter can eliminate the effects of riboflavin by using riboflavin-free minimal media (Verduyn et al., 1992).

Additionally, flavin fluorescence is the aggregate of many flavoprotein components. Therefore, it cannot be concluded that flavin fluorescence is the readout of one protein in particular — one can only draw conclusions about the overall redox state. Furthermore, the changes in flavin fluorescence can be because of changes in the 'flavin pool' — the amount of flavin-derived moieties in a cell across all their redox states — or due to global changes in intracellular flavin redox state, as a function of intracellular redox state. Most studies assume a constant flavin pool and see oscillations as periodic shifts in redox equilibrium. These caveats are not unique to flavin fluorescence, but are shared limitations with other autofluorescing cellular components like NAD(P)H, and the benefits of having a non-invasive method to monitor cellular metabolism outweighs the caveats.

1.4 Flux balance analysis

1.4.1 Introduction to flux balance analysis

Metabolic network reconstructions are mathematical representations of a set of metabolic pathways in a living organism. Usually, each metabolite is represented as a node, and metabolites are connected to each other through reactions that are represented as links (Palsson, 2015). This information can be represented in a two-dimensional stoichiometric matrix, in which the rows of the matrix represent the metabolites, the columns represent the reactions, and the values of each element in the matrix show the stoichiometry of the reactions in the system.

For example, if reaction R_1 is defined by:

$$1\,\mathrm{M}_1 + 2\,\mathrm{M}_2 \longrightarrow 3\,\mathrm{M}_3 \tag{1.1}$$

the elements in the stoichiometric matrix that correspond to the metabolite-reaction combinations (M_1, R_1), (M_2, R_1), and (M_3, R_1) are -1, -2, and 3, respectively.

A genome-scale metabolic model is, in simple terms, a metabolic network reconstruction that aims to cover every biochemical reaction in a living system that is catalysed by a gene-encoded enzyme. In most of the reactions represented in a metabolic network reconstruction, one or more chemical species react to create a different set of chemical species as products. However, reactions in a metabolic network reconstruction may include processes that are not chemical reactions. Such processes may include exchange of nutrients, as well as a reaction that models biomass formation.

Flux balance analysis (FBA) is a mathematical method that finds the steady-state flux of reactions through a metabolic network that is best for a given condition (Orth et al., 2010). These metabolic fluxes represent rates of chemical reactions. At its core, FBA is a method of solving the linear programming problem of finding the flux values that optimise the output value of an objective function, subject to biological constraints.

Mathematically, the linear programming problem of FBA can be expressed as:

$$\max \mathbf{c}^{\intercal}\mathbf{v} \tag{1.2}$$

subject to

$$\begin{gathered} \mathbf{S}\mathbf{v} = \mathbf{0} \\ v_{i,\min} \leq v_i \leq v_{i,\max} \end{gathered} \tag{1.3}$$

where $\mathbf{c}$ is a vector of weights such that $\mathbf{c}^{\intercal}\mathbf{v}$ defines the objective function, $\mathbf{S}$ is the stoichiometric matrix, and $\mathbf{v}$ is the vector of fluxes. The expression $v_{i,\min}$ represents the lower bound and $v_{i,\max}$ represents the upper bound for each flux v_i in $\mathbf{v}$.

The objective function is given as the task of maximising a mathematical expression that is based on a subset of fluxes (Eq. 1.2). Most commonly, the objective function is maximising the flux of the biomass reaction, thus optimising the growth rate of the cell. The constraints for FBA are, in the most basic case, imposed by two factors: the stoichiometric matrix and reaction flux bounds (Eq. 1.3). The stoichiometric matrix balances reaction inputs and outputs, while flux bounds impose upper and lower limits on the fluxes of each reaction. These constraints restrict the solution space for the FBA problem.

FBA thus offers a computationally inexpensive way to simulate metabolism in a living system, as opposed to solving a large set of differential equations that describe the kinetics of biochemical reactions. Such differential equations can be difficult to construct and parametrise.

1.4.2 Development of genome-scale metabolic models for budding yeast

Several versions of genome-scale metabolic models for *Saccharomyces cerevisiae* have been developed since the first model, iFF708 (Förster et al., 2003), reviewed in Y. Chen et al. (2022) (Fig. 1.16). Subsequent developments have included the addition of new reactions, the addition of new cellular compartments, and correcting gene-protein associations.

Figure 1.16: Timeline of development of genome-scale metabolic models of *Saccharomyces cerevisiae*. Adapted from Y. Chen et al. (2022).

Initially, new models were built based on iFF708, adhering to BiGG conventions (Norsigian et al., 2020). These models include iND750 (Duarte et al., 2004), iLL672 (Nookaew et al., 2008), and iMM904 (Mo et al., 2009) and consist of increasing numbers of reactions.

Yeast1 (Herrgård et al., 2008) became the first consensus genome-scale metabolic model of budding yeast, combining knowledge from previous models in standardised terminology; however, it was not until Yeast4 (Dobson et al., 2010) when gaps in the network were filled and constraint-based simulations were possible.

More recent developments of consensus genome-scale metabolic models of budding yeast created formalisms that are relevant to this thesis. Yeast5 (Heavner et al., 2012) introduced '*isa*' reactions and pseudometabolites to group specific lipid species within a more general lipid class (Fig. 1.17). This formalism has two aims. First, *isa* reactions account for generic chemical compounds, as opposed to specific chemical species, represented in KEGG pathways of sphingolipid metabolism. Such generic compounds include 'complex sphingolipid', 'fatty acid', and 'acyl-CoA' rather than specific chemical species. Second, *isa* reactions create *OR* logic in what constitutes biomass — in other words, users of the model can choose whether one set of lipid species or another set forms part of biomass. This *OR* logic accounts for different lipid species in aerobic and anaerobic growth. Yeast5 provides two biomass equations, one for aerobic growth and the other for anaerobic growth, that have different stoichiometric coefficients. The formalisms in Yeast5 resemble that of iIN800 (Nookaew et al., 2008), which has two alternative biomass reactions for carbon-limited and nitrogen-limited conditions.

Later, Yeast6 (Heavner et al., 2013) eliminated the separate biomass equations for aerobic and anaerobic growth conditions. Subsequently, Yeast7 (Aung et al., 2013) revised representation of fatty acid metabolism to correct for errors and expanded the variety of glycerolipid and glycerophospholipid composition covered by *isa* reactions.

Figure 1.17: Illustration of *isa* reactions in Yeast5. **(A)** 'Complex sphingolipid' is a pseudometabolite that is comprised of two chemical species, MIPC and $M(IP)_2C$, and one pseudometabolite, IPC. To implement these relationships in Yeast5, *isa* reactions are defined — e.g. 'MIPC is a complex sphingolipid'. **(B)** Another pseudoreaction is defined in order to define lipids as an a collective of other compounds, complex sphingolipids among them. Here, 'lipid' is a pseudometabolite. **(C)** The biomass reaction is defined with chemical species and pseudometabolites as reactants and a 'biomass' pseudometabolite as the product. Adapted from Heavner et al. (2012).

The development of the most recent consensus model, Yeast8 (H. Lu et al., 2019), included additional pseudometabolites, incorporating SLIMEr (Sánchez et al., 2019), and improving the biomass reaction. First, in addition to lipids as in previous consensus models, Yeast8 includes pseudometabolites that represent proteins, carbohydrates, DNA, RNA, cofactors, and ions, making it easy to study each component of biomass separately. Second, SLIMEr (Splits Lipids Into Measurable Entities) solves the problem of representing many possible lipid species in genome-scale metabolic models through the addition of two types of pseudoreactions. Specifically, it creates SLIME reactions to split lipids into their basic components (e.g. glycerol and fatty acid chains for triglycerides) and it creates lipid pseudoreactions to impose constraints on lipid classes and acyl chain distribution to make the ratio between each class match measured abundances. Third, the biomass reaction in Yeast8 is improved by the addition of trace metal ions and cofactors, along with re-scaling the stoichiometric coefficients in biomass composition and re-fitting the growth-associated maintenance energy requirements, to preserve mass balance.

Consequently, Yeast8 is the most comprehensive genome-scale metabolic model of budding yeast metabolism, with improved metabolic scope compared to earlier models, closer to the real metabolic network.

1.4.3 Allocation of proteomic resources

Metabolism in microbes are subject to limited resources: metabolic building blocks, energy currency, biosynthesis machinery, and spatial compartments (Elsemman et al., 2022). An important aspect of this resource limitation is a limited proteome, implying that the cell must allocate its proteome to its proteins, as evidenced by observations on how RNA-to-protein ratios affect growth rate (Scott & Hwa, 2011).

To account for limited proteomic resources, genome-scale metabolic models can be extended. Here, I review two main approaches: minimising protein cost and explicit addition of optimisation of resources to models (Adjavon, 2022).

Some adaptations of FBA include methods to ensure that the protein cost is minimised during optimisation. These methods include three strategies: adding an additional objective function, imposing additional constraints, and imposing bounds on fluxes. Among studies that use additional objective functions, Schuetz et al. (2012) tested setting reactions other than the biomass reaction as the objective and showed that combining maximisation of ATP yield and of biomass yield together with minimising the sum of absolute fluxes resulted in highest optimality. Additionally, Wang et al. (2019) aimed to understand diauxie by defining a new variable, enzyme utilisation efficiency, to be optimised. Enzyme utilisation efficiency was defined based on carbon flux and an enzyme cost variable derived from enzyme concentrations and molecular weights.

Alternatively, additional constraints can be imposed for the linear programming problem in flux balance analysis. For example, Mori et al. (2016) proposed a Constrained Allocation Flux Balance Analysis method, which divided the proteome into four sectors and imposed the constraint that sum of fluxes in each sector must be below a threshold value. Furthermore, bounds can be set directly on fluxes. This approach was implemented in GECKO (Domenzain et al., 2022; Sánchez et al., 2017), in which pseudoreactions were defined to model allocation of an enzyme-available proteome pool to individual enzymes, and the size of the proteome pool was set setting an upper bound on the flux of the corresponding pseudoreaction.

Rather than adding relatively small formalisms to FBA problems, the optimisation of resources can be explicitly added to models. This strategy includes two deviations from ordinary flux balance analysis: resource balance analysis (Goelzer et al., 2011) and DynamicME (Yang et al., 2019). Resource balance analysis predicts the distribution of resources from the nutrient medium among cellular subsystems by modelling the living system as an interconnection of subsystems that compete for common resources. Thus, growth rate depends on resource allocation among subsystems. Unlike ordinary FBA, resource balance analysis predicts concentrations of metabolites in addition to fluxes, but still gives a steady-state picture of metabolism. In contrast, DynamicME simulates changes in metabolism and protein expression over time. To create dynamic simulations, DynamicME expresses the stoichiometric matrix and flux bounds as functions of growth rate, and to simulate optimisation of resources, it imposes macromolecular crowding constraints.

In this thesis, GECKO is used to simulate allocation of proteomic resources because there is an easily available derivative of Yeast8 with the GECKO formalism already applied. Furthermore, GECKO provides an easy way to control the size of the enzyme-available proteome, thus making it easy to test the hypothesis of whether the size of the proteome affects metabolic strategies in the budding yeast cell.

1.5 Aims of thesis and research questions

This thesis aims to address these research questions:

1. Do metabolic oscillations arise from communication between cells or from external synchronisation?
2. If metabolic cycles are persistent across nutrient and genetic perturbations, is it because metabolic cycles reflect sequential synthesis of biomass components?

To test whether metabolic oscillations in chemostat-based studies arise from communication between cells or from external synchronisation, I use a single-cell microfluidics platform to monitor flavin-based metabolic cycles. I chose this platform because it physically separates cells and has rapid media flow, and both eliminate communication between cells that may occur in the chemostat. Single-cell microfluidics alone cannot assess the role of communication between cells without comparing with bulk-culture set-ups such as the chemostat or the turbidostat. However, using such bulk-culture equipment is beyond the scope of this thesis, and this thesis aims to address the first research question by comparing single-cell results with corresponding chemostat-based results in the literature.

The observation that metabolic cycles are persistent across nutrient and genetic perturbations gave rise to the second research question. To test whether this persistence reflects sequential synthesis of biomass components, I use flux balance analysis with an enzyme-constrained genome-scale metabolic model of budding yeast. Specifically, I assess whether imposing a constraint on the proteome or changing nutrient conditions leads to sequential synthesis of biomass components taking less time than parallel synthesis of biomass components. Answering this latter research question could lead to an understanding of the yeast metabolic cycle as a fundamental metabolic adaptation to physiological constraints.

Chapter 2

Methods

2.1 Laboratory methods

2.1.1 Strains and media

The *Saccharomyces cerevisiae* strains used in this thesis are described in Table 2.1.

Name	Background	Genotype	Origin	Notes
FY4	FY4	—	EUROSCARF	Winston et al. (1995)
htb2::mCherry	FY4	HTB2::mCherry	In-house, CRISPR	—
BY4741	BY4741	*MAT*a *his3*Δ*1 leu2*Δ*0 met15*Δ*0 ura3*Δ*0*	EUROSCARF	Brachmann et al. (1998)
*zwf1*Δ	BY4741	*zwf1*Δ*::KAN*	Edinburgh Genome Foundry	Yeast deletion collection
BY4742	BY4742	*MAT*α *his3*Δ*1 leu2*Δ*0 lys2*Δ*0 ura3*Δ*0*	Bruce Morgan	Calabrese et al. (2019)
*tsa1*Δ *tsa2*Δ	BY4742	*tsa1*Δ*::natNT2 tsa2*Δ*::kanMX4*	Bruce Morgan	Calabrese et al. (2019)
CEN.PK113-7D	CEN.PK113-7D	—	Peter Kötter	Nijkamp et al. (2012)

Table 2.1: Strains used in this thesis.

The minimal medium described by Verduyn et al. (1992) was used unless otherwise stated. This minimal medium does not contain riboflavin, thus minimising its effect on flavin autofluorescence imaging, and its composition is known and easily controlled. Specifically, the composition of the carbon source-limiting medium is described in Tables 2.2–2.4, and the media pH was adjusted to 6.0 using potassium hydroxide, or sodium hydroxide for potassium-free media.

Reagent	Concentration	Remarks
KH_2PO_4	$3\,g\,L^{-1}$	
$MgSO_4 \cdot 7\,H_2O$	$0.5\,g\,L^{-1}$	
$(NH_4)_2SO_4$	$5\,g\,L^{-1}$	
Trace metals	$1\,mL\,L^{-1}$	See Table 2.3
Vitamins	$1\,mL\,L^{-1}$	See Table 2.4. Add upon use.
Carbon source	variable	Add upon use.

Table 2.2: Composition of base minimal medium. For potassium-free media, replace KH_2PO_4 with $2.65\,g\,L^{-1}$ NaH_2PO_4, which gives the same molarity.

Reagent	Formula	Concentration $[g\,L^{-1}]$
EDTA	$C_{10}H_{14}N_2Na_2O_8 \cdot 2\,H_2O$	15.00
Zinc sulfate	$ZnSO_4 \cdot 7\,H_2O$	4.50
Manganese (II) chloride	$MnCl_2 \cdot 2\,H_2O$	0.84
Cobalt (II) chloride	$CoCl_2 \cdot 6\,H_2O$	0.30
Copper (II) sulfate	$CuSO_4 \cdot 5\,H_2O$	0.30
Sodium molybdate	$Na_2MoO_4 \cdot 2\,H_2O$	0.40
Calcium chloride	$CaCl_2 \cdot 2\,H_2O$	4.50
Iron (II) sulfate	$FeSO_4 \cdot 7\,H_2O$	3.00
Boric acid	H_3BO_3	1.00
Potassium iodide	KI	0.10

Table 2.3: Composition of trace metal mix for minimal media described in Table 2.2.

For auxotrophic strains, supplements were added according to Table 2.5. Then, a carbon source is added as appropriate to create the growth medium.

Reagent	Formula	Concentration [$g\,L^{-1}$]
D-(+)-biotin	$C_{10}H_{16}N_2O_3S$	0.05
D-panthothenic acid calcium salt	$Ca(C_9H_{16}NO_5)_2$	1.00
Nicotinic acid	$C_6H_5NO_2$	1.00
myo-Inositol	$C_6H_{12}O_6$	25.00
Thiamine chloride hydrochloride	$C_{12}H_{15}ClN_4OS \cdot HCl$	1.00
Pyridoxal hydrochloride	$C_8H_{12}ClNO_3$	1.00
4-aminobenzoic acid	$C_7H_7NO_2$	0.20

Table 2.4: Composition of vitamin mix for minimal media described in Table 2.2.

Reagent	Concentration [$mg\,L^{-1}$]
histidine	125
leucine	500
tryptophan	75
methionine	100
uracil	150

Table 2.5: Supplements to minimal media for BY4741-background auxotrophic strains, compositions derived from Pronk (2002). For BY4742-background strains, replace methionine with $100\,mg\,L^{-1}$ lysine-HCl.

2.1.2 Single-cell microfluidics

Cells were grown from colonies on solid agar in a liquid culture composed of minimal media formulation appropriate for the experiment, supplements appropriate for the strain's auxotrophy, and a carbon source (glucose or pyruvate) appropriate for the experiment (see Section 2.1.1). The cells were incubated at 30 °C for 14 h (overnight) if the carbon source is glucose or 48 h if the carbon source is pyruvate. Subsequently, the cells were diluted so that the resulting culture had an OD_{600} of 0.10–0.20, and were then incubated for a further 4 h.

To monitor metabolic cycles in single cells, ALCATRAS microfluidics (Crane et al., 2014) devices were prepared and connected to syringes containing nutrient media (Fig. 2.1). ALCATRAS microfluidic devices were made from polydimethylsiloxane (PDMS), with a 10:1 ratio between Sylgard 184 (Dow Corning) and curing agent. The devices consist of chambers that contain of a dense array of cell traps to trap parent cells, allowing approximately 100 traps to be imaged at once using a 40 × objective lens magnification.

Figure 2.1: Overview of ALCATRAS microfluidic chip design. **(A)** Sample brightfield image of cell traps in the chip, overlaid with fluorescence image of Doa1p-GFP, to indicate cells (green). **(B)** Switching between media sources (yellow to pink) occurs within 6 s. **(C)** Microfluidic chip design. **(D)** Switching rate was assessed by imaging the fluorescence in the chamber whilst the input medium was switched between a non-fluorescent medium and medium containing 0.1% fluorescein. **(E)** Trap dimensions, in micrometres. Adapted from Crane et al. (2014).

The pressure on the syringes were controlled automatically to achieve a constant $4\,\mu L\,min^{-1}$ media flow, creating laminar flow. The shape of the traps allows parent cells to be trapped in place, while progeny cells leave the trap after they bud from parent cells as nutrient medium flows across the device. Specifically, when a cell falls into a trap, the local pressure gradient changes to create a low-energy pocket that keeps the cell trapped, as evidenced by a simulation of fluid dynamics given the device design (Fig. 2.2).

Figure 2.2: Fluid dynamic simulations of the ALCATRAS microfluic chip. **(A)** Velocity of fluid (in units of $\mu m\,s^{-1}$) as a function of spatial position in the device. **(B)** Pressure of fluid (in units of Pa) as a function of spatial position in the device. Adapted from Crane et al. (2014).

PDMS is known to absorb small, hydrophobic molecules (Toepke & Beebe, 2006), and molecules with such properties are also known to adsorb onto the surface of PDMS (N. Li et al., 2009). These physical interactions decrease the effective concentration of such molecules in the nutrient media. To decrease adsorption, a thin layer of silane was evaporated onto the wafer before moulding of PDMS devices (Crane et al., 2014), and during experimental set-up, bovine serum albumin was added to create an additional coating on the internal surfaces of the microfluidic device.

To prepare for an experiment, a device's multiple chambers were filled with growth media supplemented with 0.05% w/v bovine serum albumin, which was added to further decrease adsorption (Toepke & Beebe, 2006). Cells were then loaded into the ALCATRAS chambers — different chambers can house cells from different strains (Fig. 2.3). There are two syringes that can contain different media, and in experiments that require different nutrient conditions at different times, the ALCATRAS system is programmed to switch between the two syringes. The cells and ALCATRAS chambers were located in an incubation chamber (Oko-labs) that was maintained at 30 °C.

Figure 2.3: Overview of single-cell microfluidics set-up using the ALCATRAS system. Cells are loaded into chambers within devices, where they are trapped and separated (centre). The media composition the cells experience are controlled with syringe pumps (left). Brightfield and fluorescent images are taken at regular intervals, and then processed using *aliby* (Muñoz González, 2023) to obtain time series of fluorescence intensity changes for each parent cell.

Microscopy was performed using a 40 × 1.4 NA oil immersion objective (Nikon), and the Nikon Perfect Focus System was used to ensure consistent focus. X-Y spatial positions were defined for each chamber to maximise spatial coverage of the chamber while ensuring that the microscope takes less time to move positions

and capture images than the interval period. Images were taken every 5 min, and the duration of image acquisition varied for each experiment. Brightfield and flavin images were captured in all strains and mCherry images were additionally captured for the HTB2::mCherry strain. Five z-slices were taken for brightfield images, with a spacing of 0.6 µm between slices. Fluorescence imaging was performed with an OptoLED light source (Cairn Research), and LED voltage was optimised for maximum signal intensity without LED cut-off prior to experiments. For flavin imaging, the excitation filter was set to 430/24 (418 nm to 442 nm), the emission filter was set to 535/30 (520 nm to 550 nm), and the exposure time was 60 ms. One z-slice was taken for each flavin image in each position. For mCherry imaging, the excitation filter was set to (555 nm to 590 nm), the emission filter was to set to 632/60 (602 nm to 682 nm) and the exposure time was 100 ms. Five z-slices were taken for mCherry images, with a spacing of 0.6 µm between slices.

2.2 Image analysis methods

I used *aliby* (Muñoz González, 2023), an end-to-end Python-based software package developed for time-lapse microscopy, to process the microscope images in order to obtain flavin and mCherry time series for further analysis.

aliby tracks tiles that correspond to a trap across time-lapse images to account for expected spatial drifting in the microscope. It then uses *BABY* (Pietsch et al., 2023) to segment the images of traps to identify the outlines of cells and to track cells from one time point to another, creating a lineage of cells (Fig. 2.4). *aliby* then overlays the cell outlines onto the fluorescence (flavin and mCherry) images

to extract fluorescence intensity, and assigns a fluorescence value to each cell at each time point based on the mean intensity of pixels within the cell's outline. The background fluorescence is also computed, based on the pixel intensity outside cell outlines, and is then subtracted from the cell fluorescence.

Flavin fluorescence thus represents the oxidation of flavins throughout the yeast metabolic cycle (see Section 1.3), and mCherry fluorescence thus represents the amount of histone proteins as a proxy for cell division cycle progression (Garmendia-Torres et al., 2018).

(a)

				0	5	10	15	20	25	30	35	40
group	**position**	**trap**	**cell_label**									
by4742morgan	**by4742morgan_009**	**0**	**1**	24.772610	23.852300	25.205446	27.832461	26.400000	27.875989	27.985330	27.255981	28.903981
		2	**1**	27.925208	26.661202	26.717678	29.116022	28.820937	26.954424	28.685567	27.431981	28.188482
		13	**1**	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
		14	**1**	27.315353	26.448544	26.237911	28.489109	28.425490	26.700193	27.048148	25.534843	25.229452
		16	**1**	32.257895	30.159292	30.147826	31.741652	28.512238	29.964093	30.235993	28.889078	30.412587
...	**...**	**...**	**...**	...	...	...	...	...	...	...	...	...
tsa1tsa2morgan	**tsa1tsa2morgan_024**	**33**	**7**	10.657143	11.175824	10.653846	11.621951	11.454545	8.912698	8.629032	10.739130	11.523810
			8	5.811765	6.186047	4.436782	5.211765	5.786408	6.239316	6.535088	5.672727	5.641026
			9	8.750000	9.797297	9.183673	9.457831	9.117647	8.701149	NaN	8.276316	9.481481
			10	10.123967	10.579832	9.638298	9.260870	9.751825	10.245902	10.202899	10.692982	11.006803
		36	**12**	9.271028	9.326087	10.680851	9.896907	10.606742	9.872727	9.075000	9.132530	9.932203

1108 rows × 240 columns

(b)

Figure 2.4: (2.4a) Example time course of images from a microfluidics experiment, processed by *aliby* (Muñoz González, 2023). Top row shows brightfield images taken over time; only the first three time points are shown for illustrative purposes. Here, *BABY* (Pietsch et al., 2023) segments brightfield images to obtain cell outlines (blue and brown). Bottom row shows images from a sample fluorescence channel taken over time. The cell outlines from image segmentation were overlaid onto the fluorescence channel images so that the fluorescence intensity within each cell can be extracted. **(2.4b)** The extracted values were then saved in an array of numbers, with rows indicating cells (labelled by IDs) and columns indicating time in minutes — each row thus represents a time series.

2.3 Time series analysis methods

2.3.1 Data pre-processing

To filter out long-term trends that may confound analysis of oscillatory time series, a high-pass Butterworth filter, with a critical frequency of $2.86 \times 10^{-3}\,\mathrm{min}^{-1}$, corresponding to a period of 350 min, was applied to fluorescence time series generated from *aliby*. Specifically, Figs. 3.2, 3.3–3.5, 3.10–3.13, 3.15, 3.17–3.18, and 3.21–3.22 in Chapter 3 and all figures from Chapter 4 except for 4.1a are derived from filtered time series. Section 4.2 further discusses the rationale for this filtering method.

However, the long-term shifts of fluorescence intensities are important for interpreting some results based on imposing temporary environmental perturbations on yeast cells. In such cases, such long-term shifts would be filtered away by the high-pass Butterworth filter, and thus the filter was not applied. Specifically, Figs. 3.8–3.9, and 3.20 are derived from unfiltered time series.

2.3.2 Classical periodogram

To detect rhythmicity in time series, I computed the periodogram, based on the classical (original) definition of the periodogram (Scargle, 1982; Schuster, 1898), and by using the fast Fourier transform, then used a statistical test for rhythmicity based on Glynn et al. (2006), described as follows:

1. Let the data have $\mathcal{G}$ cells. Let cell $g = 1, \ldots, \mathcal{G}$ have a time series with N_g time points. The time series is thus denoted $Y_g(t) = y_g(t_1), \ldots, y_g(t_{N_g})$.
2. For each time series, I define a range of test frequencies linearly from $\frac{1}{N_g}$ to the Nyquist limit (i.e. half the rate of image acquisition).

With this definition, I compute the classical periodogram for each time series, via the Fourier transform:

$$P_g(\omega) = \frac{N_g}{2\sigma^2} \left| \int_{-\infty}^{\infty} Y_g e^{-2\pi i t} dt \right|, \tag{2.1}$$

where σ^2 is the sample variance of Y_g. In this equation, the periodogram is normalised by the coefficient $N_g/2\sigma^2$ so that the area under the periodogram is constant across all time series. The Lomb-Scargle periodogram is equivalent to the classical periodogram if the time points are equally spaced (Lomb, 1976), as is the case for the vast majority of my data.

3. For each cell g, I denote the peak $h_g = \max_j P_g(\omega)$. The peak of the normalised classical periodogram of each time series was used as a proxy for the quality of oscillation.

4. I define an effective number of independent frequencies $M = f_{max} N_g$ for each time series, where f_{max} is the Nyquist limit (VanderPlas, 2018). I then calculate the p-value of testing the null hypothesis that such a peak is due to chance:

$$p_g = 1 - (1 - e^{-h_g})^M \tag{2.2}$$

This formula is based on the exponential distribution of the power at a given frequency in the periodogram (Scargle, 1982).

5. I order the cells by p-values: $p_1 \leq p_2 \leq \cdots \leq p_{\mathcal{G}}$. This order thus ranks the cells by oscillation quality.

6. To control the false discovery rate (Benjamini & Hochberg, 1995), I find $\hat{k}$ according to:

$$\hat{k} = \arg \max_{1 \leq k \leq \mathcal{G}} \{k : p_k \leq qk/\mathcal{G}\} \tag{2.3}$$

where q is a defined false discovery rate.

7. Cells whose p-values correspond to $p_1, p_2, \ldots, p_{\hat{k}}$ are thus denoted to have statistically significant oscillatory behaviour for the false discovery rate q.

2.3.3 Autoregressive model

As an alternative method to detect rhythmicity in time series, I fitted an autoregressive model and used its parameters to compute an analytically-defined periodogram, based on Jia and Grima (2020), as follows:

1. The algorithm relies on fitting a single time series $n(0), n(1), \ldots, n(M-1)$ with an autoregressive model AR(P) with order P:

$$\phi_0 n_t + \phi_1 n_{t-1} + \phi_2 n_{t-2} + \cdots + \phi_P n_{t-P} = \theta_0 \epsilon_t \tag{2.4}$$

   where ϵ_t is a white noise satisfying $\langle \epsilon_t \rangle = 0$, $\phi_0 = 1$, and $\phi_1, \ldots, \phi_P$ are real numbers such that the complex zeros of the polynomial $\Phi(z) = \sum_{k=0}^{P} \phi_k z^k$ lie outside the unit circle.

2. The sample mean of the time series is estimated by:

$$\langle n \rangle = \frac{1}{M} \sum_{k=0}^{M-1} n(k) \tag{2.5}$$

3. The sample autocorrelation function is estimated as:

$$R_i = \frac{1}{M} \sum_{k=0}^{M-1-i} (n(k) - \langle n \rangle)(n(k+i) - \langle n \rangle) \tag{2.6}$$

4. The coefficients $\phi_1, \ldots, \phi_P$ are estimated by solving the Yule-Walker equation:

$$\begin{bmatrix} R_0 & R_1 & \cdots & R_{P-1} \\ R_1 & R_0 & \cdots & R_{P-2} \\ \vdots & \vdots & \ddots & \vdots \\ R_{P-1} & R_{P-2} & \cdots & R_0 \end{bmatrix} \begin{bmatrix} \phi_1 \\ \phi_2 \\ \vdots \\ \phi_P \end{bmatrix} = \begin{bmatrix} R_1 \\ R_2 \\ \vdots \\ R_P \end{bmatrix} \tag{2.7}$$

5. The parameter θ_0 is estimated as:

$$\theta_0^2 = R_0 - \sum_{k=1}^{P} \phi_k R_k \tag{2.8}$$

6. The order P is determined by minimising the Akaike information criterion:

$$\mathrm{AIC}(P) = \log \theta_0^2(P) + 2\frac{P}{M} \tag{2.9}$$

where $\theta_0(P)$ is the estimated θ_0 (Eq. 2.8) for a specific P. In this step, P is varied with $1 \leq P \leq 3\sqrt{M}$, and the optimum order (P) is the one that gives the smallest value of $\mathrm{AIC}(P)$

7. The power spectrum is thus estimated analytically using the parameters found in earlier steps by:

$$G(\xi) = \frac{1}{2\pi} \cdot \frac{\theta_0^2}{|\sum_{k=0}^{P} \phi_k \mathrm{e}^{-ik\xi}|^2}, -\pi \leq \xi \leq \pi \tag{2.10}$$

where ξ represents frequency.

2.3.4 Precision and recall

To evaluate the performance of rhythmicity detection methods, precision and recall are computed. The rhythmicity detection methods are seen as a binary classifier that predicts a 'positive' or 'negative' label to each observation, and these predicted labels are compared against true labels — in the context of this thesis, manually-defined labels of whether a time series is oscillatory or not.

Given the confusion matrix:

		Predicted labels	
		Positive	Negative
True labels	Positive	True positives (TP)	False negatives (FN)
	Negative	False positives (FP)	True negatives (TN)

Table 2.6: Confusion matrix

Precision is defined as:

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{2.11}$$

And recall is defined as:

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{2.12}$$

Precision and recall can be used with datasets with class imbalance, i.e. if the two labels are not split evenly across observations.

2.3.5 Cross-correlation function

To estimate the periodicity of signals and to detect the synchrony between two types of signals, I adapted the cross-correlation function as used in Pietsch et al. (2023), as follows:

1. Let the data have M cells. Each cell i in the population of M cells has a time series $x_1^{(i)}, \ldots, x_j^{(i)}, \ldots, x_N^{(i)}$ of quantity x and a time series $y_1^{(i)}, \ldots, y_j^{(i)}, \ldots, y_N^{(i)}$ of quantity y. Let both time series have a sampling interval of Δt.
2. The deviation from the population mean for each time series is computed. This population mean is calculated over replicates at each time point. The caveat of this calculation is that the signals must be out-of-phase, and I ensure this by generating synthetic signals with a random phase. Otherwise, the underlying signal will be subtracted from all time series and the cross-correlation of noise will be computed — this is undesired.

$$\delta x_t^{(i)} = x_t^{(i)} - \frac{1}{M} \sum_j x_t^{(j)} \tag{2.13}$$

$$\delta y_t^{(i)} = y_t^{(i)} - \frac{1}{M} \sum_j y_t^{(j)} \tag{2.14}$$

3. Based on Kiviet et al. (2014), the cross-covariance of the two time series x and y at a time lag of $r\Delta t$, is thus given by:

$$C_{xy}^{(i)}(r\Delta t) = \begin{cases} \frac{1}{N-r} \sum_{t=1}^{N-r} \delta x_t^{(i)} \cdot \delta y_{t+r}^{(i)} & \text{if } r \geq 0 \\ C_{yx}^{(i)}(-r\Delta t) & \text{if } r < 0 \end{cases} \tag{2.15}$$

4. $C_{xx}^{(i)}(0)$ and $C_{yy}^{(i)}(0)$ thus give the variances of x and y. The cross-correlation is thus given, with normalising by the standard deviation, by:

$$R_{xy}^{(i)}(r\Delta t) = \frac{C_{xy}^{(i)}(r\Delta t)}{\sqrt{C_{xx}^{(i)}(0)C_{yy}^{(i)}(0)}} \tag{2.16}$$

The autocorrelation of a time series x is thus the cross-correlation of the time series with itself, i.e. $R_{xx}^{(i)}(r\Delta t)$.

2.3.6 Generation of synthetic oscillatory data

To understand the effect of the shape and noise properties present in biological time series on the cross-correlation function, I emulated such time series by generating synthetic time series using the harmonic (sinusoid) and FitzHugh-Nagumo oscillators (FitzHugh, 1961).

The harmonic oscillator $y(t)$ is defined as the solution of:

$$\frac{\mathrm{d}^2 y}{\mathrm{d}t^2} = -\omega^2 y \tag{2.17}$$

where the sole parameter ω represents the angular frequency.

The FitzHugh-Nagumo oscillator $v(t)$ is defined as a solution of the system:

$$\begin{aligned} \frac{\mathrm{d}\,v}{\mathrm{d}t} &= v - \frac{v^3}{3} - w + RI_{\text{ext}} \\ \tau\frac{\mathrm{d}\,w}{\mathrm{d}t} &= v + a - bw \end{aligned} \tag{2.18}$$

where RI_{ext}, τ, a, and b are constant parameters to be determined.

In addition, I emulated noise by generating Gaussian and Gillespie noise.

Gaussian noise was generated by randomly drawing samples from the normal distribution $\mathcal{N}(0, \sigma^2)$. Here, σ denotes the standard deviation of the distribution and thus controls the size of the noise.

Gillespie noise emulates noise from biological systems. Gillespie noise was generated using the direct method of the Gillespie algorithm (Gillespie, 1977) on the birth-death process model (Appendix A.4).

The birth-death process is a simple stochastic model used for the modelling of gene expression. The model describes a species that is produced at a linear birth rate k_0 and destroyed at a linear death rate d_0, and is defined by the system of equations Eq. 2.19:

$$\begin{aligned} R_1 &: \varnothing \xrightarrow{\mathrm{k_0}} \mathrm{P} \\ R_2 &: \mathrm{P} \xrightarrow{\mathrm{d_0}} \varnothing \end{aligned} \tag{2.19}$$

To produce Gillespie noise, a stochastic simulation employing the direct method of the Gillespie algorithm was performed on the birth-death process model with defined k_0 and d_0 parameters. The final time was defined in such a way that allows the trajectory of the amount of P over time to reach a steady state (Fig. 4.20b). This time varied depending on the k_0 and d_0 values, but the final time of 1500 was chosen as it was long enough to have the trajectory reach steady state for the k_0 and d_0 values used in this study.

The latter half of the trajectory was taken and then put on a grid with 1000 regularly-spaced time points, equal to the number of time points for the synthetic oscillators (harmonic and FitzHugh-Nagumo). The time series was then normalised by subtracting the mean (k_0/d_0) and then dividing by $\sqrt{1/d_0}$ to create a time

series representing Gillespie noise with mean 0 and standard deviation $\sqrt{k_0}$. This Gillespie noise thus has a standard deviation of noise amplitude $A = \sqrt{k_0/d_0}$ and noise timescale $\tau = 1/d_0$ — in other words, the rate parameters of the birth-death process control the noise properties of this Gillespie noise.

2.3.7 Signal-to-noise ratio

To evaluate the quality of oscillatory time series in a dataset and to indirectly measure the amplitude of the oscillations, I computed a signal-to-noise ratio for each time series. Assuming a constant noise introduced by the combination of intrinsic noise (stochasticity in biochemical processes) and extrinsic noise (variations introduced by measurement instruments), a low signal-to-noise ratio suggests a low oscillation amplitude, and the reverse is true for a high signal-to-noise ratio.

The signal-to-noise ratio can be defined as follows:

Given a time series $x(t) = x(t_1), x(t_2), \ldots, x(t_N)$, the normalised classical periodogram (Fourier spectrum) is given by

$$P(\omega) = \frac{N}{2\sigma^2} \left| \int_{-\infty}^{\infty} x(t) e^{-2\pi i t} dt \right|, \tag{2.20}$$

where N is the number of time points in x and σ^2 is the sample variance of x. The periodogram P is defined as a function of the angular frequency ω.

A critical frequency ω_c is then defined to divide signal and noise — that is, very high-frequency components of the periodogram correspond to noise and lower-frequency components correspond to the meaningful oscillations.

The signal to noise ratio $r_{s/n}$ (Fig. 2.5) is thus defined as

$$r_{s/n} = \frac{s}{n} \tag{2.21}$$

where the signal s is defined as

$$s = \int_0^{\omega_c} P(\omega) d\omega \tag{2.22}$$

and the noise n is defined as

$$n = \int_{\omega_c}^{\infty} P(\omega) d\omega \tag{2.23}$$

(a) **(b)**

Figure 2.5: (2.5a) Illustration of signal-to-noise ratio. The signal-to-noise ratio is defined (Eqs. 2.20–2.23) as the area under the periodogram below a cut-off frequency (yellow) divided by the area under the periodogram above a cut-off frequency (pink). **(2.5b)** Histogram of signal-to-noise ratios from a sample experiment ($n = 403$).

In this thesis, I define $\omega_c = 0.018\,\mathrm{min}^{-1}$, corresponding to a period of 55 min, because this period is smaller than the periods of all oscillations of flavin autofluorescence I observed in my experiments.

2.4 Methods related to flux balance analysis

To evaluate biosynthesis strategies in the *S. cerevisiae* cell, I used the ecYeast8 (H. Lu et al., 2019) genome-scale metabolic model and performed flux balance analysis.

2.4.1 ecYeast8 formalisms

The formalisms used in ecYeast8 differ from the usual formalisms used in genome-scale metabolic models. ecYeast8 contains four formalisms relevant to this thesis:

1. The biomass reaction is defined by having 'pseudometabolites' as reactants and a biomass species as a product. These pseudometabolites include lipids, proteins, carbohydrates, RNA, DNA, ions, and cofactors. This is in contrast to BiGG genome-scale models (Norsigian et al., 2020). In these models, biomass reactions have chemical species as reactants, and each has a stoichiometric coefficient that is equal to the species' abundance in units of mmol g_{DW}^{-1}.
2. The pseudometabolites are defined by '*isa*' reactions, which group specific chemical species into classes of metabolites (Heavner et al., 2012). These reactions account for how some KEGG definitions of reactions require generic compounds and these reactions also allow flexibility of biomass definition in different growth conditions. The *isa* reactions have chemical species as reactants, each with a stoichiometric coefficient representing the species' abundance in units of mmol g_{DW}^{-1}, and a pseudometabolite as a product. In effect, the species abundance information is shifted one reaction away from the biomass reaction.

3. The models implements SLIMEr (Sánchez et al., 2019), which Splits Lipids Into Measurable Entities, and adds constraints on lipid classes and acyl chain distribution. This formalism is needed because species of lipid backbones and acyl chain can combine to form lipids in more than a thousand ways, and the resulting lipid species are difficult to all be represented in a genome-scale metabolic model. SLIMEr thus introduces reactions that split lipids into their basic components and lipid pseudoreactions to preserve the distribution of acyl chains. As a result, the definitions of lipids are flexible.
4. GECKO (Sánchez et al., 2017) was applied to Yeast8 to produce ecYeast8. GECKO modifies the stoichiometric matrix of a genome-scale metabolic model to account for enzyme abundances and kinetics. Specifically, it adds to enzyme-catalysed reactions enzyme species with a stoichiometric coefficient derived from its k_{cat} value. The formalism also adds reactions to model drawing enzymes from a pool. GECKO simulates an upper limit of amino acids available for enzyme production.

2.4.2 Computing molecular weights of pseudometabolites

To compute the synthesis time for each biomass component, the mass fractions of each biomass component must be known. These quantities are typically stored as molecular weights in a genome-scale model. As biomass components in ecYeast8 are pseudometabolites without specified molecular weights, I computed the mass fractions of each biomass component based on their *isa* reactions (Appendix B.3). Rather than using mass fractions based on experimental studies, I used mass fractions based on the model because the mass fraction of each biomass component varies according to strain and growth rate (Elsemman et al., 2022; Nilsson & Nielsen, 2016).

Table 2.7 summarises the molecular weight of the pseudometabolites. The ratio between the molecular weights are similar to the ratio between the mass of each class of macromolecule in the yeast cell dry weight shown by Canelas et al. (2011), thus validating my calculations.

Metabolite	Computed molecular weight ($\mathrm{g\,mol^{-1}}$)	Biomass composition at growth rate $0.375\,\mathrm{h^{-1}}$ ($\mathrm{g\,kg_{DW}^{-1}}$)
Protein	504.37	505
Carbohydrate	350.37	237
RNA	64.04	105
Lipid	31.57	57
Cofactors	4.83	
DNA	3.90	5
Ions	2.48	
Total	961.57	

Table 2.7: Computed molecular weights of bulk metabolites in ecYeast8, compared to experimentally recorded biomass composition by Canelas et al. (2011).

Adding together molecular weights of pseudometabolites gave a molecular weight of the biomass pseudometabolite of $961.57\,\mathrm{g\,mol^{-1}}$, close to the $966\,\mathrm{g\,mol^{-1}}$ computed by Takhaveev et al. (2023) from a different genome-scale model. In theory, this number should be $1000\,\mathrm{g\,mol^{-1}}$ because the stoichiometric coefficients of the species that form biomass components are expressed in terms of mmol $\mathrm{g_{DW}^{-1}}$ (Palsson, 2015; Thiele & Palsson, 2010), but the deviation from 1000 could be explained by the SLIMEr formalism. In addition, the sum of stoichiometric coefficients is not always verified in genome-scale models (Chan et al., 2017).

Chapter 3

Microfluidics and fluorescence microscopy for cellular metabolic cycles

Few published studies have investigated the YMC in cells isolated from each other (Baumgartner et al., 2018; Özsezen et al., 2019; Papagiannakis et al., 2017; Silverman et al., 2010). Most studies instead have examined the YMC from a population of cells in a chemostat (Amponsah et al., 2021; Causton et al., 2015; Mellor, 2016; D. B. Murray et al., 2011; J. S. O'Neill et al., 2020; Tu et al., 2005, 2007). However, such chemostat-based studies are unable to account for cell-to-cell heterogeneity. Furthermore, in such conditions, cell density and environmental conditions are far removed from the natural habitat of budding yeast.

To reconcile the evidence about the characteristics of the yeast metabolic cycle (YMC) from single-cell and chemostat experiments, I sought to use a single-cell experimental platform to address whether cellular metabolic cycles confirm chemostat-based studies.

In this chapter, I use single-cell microfluidics to physically separate budding yeast cells. In these experiments, I use fluorescence microscopy to monitor the yeast metabolic cycle and the cell division cycle.

Specifically, I aim to evaluate these hypotheses:

1. Yeast cells independently generate yeast metabolic cycles. Each cell generates the metabolic cycle autonomously of other cellular oscillators, but the metabolic cycle can phase-lock the cell division cycle.
2. The yeast metabolic cycle is retained in different nutrient and genetic perturbations, but characteristics of the cycle change in response.
3. Flavin autofluorescence of single yeast cells recapitulate oscillations in dissolved oxygen in the chemostat. If there are discrepancies between these two manifestations of the yeast metabolic cycle, they may be explained by individual cells continuing to generate the metabolic cycle but without synchrony between cells.

In this chapter, I show that metabolic cycles are generated autonomously and are coupled to the cell division cycle in permissive conditions, confirming previous single-cell studies (Özsezen et al., 2019; Papagiannakis et al., 2017). To decouple the metabolic and cell division oscillators, confirming that the metabolic oscillator is independently and autonomously generated, I used fast nutrient switching to induce starvation. I further show that the metabolic cycle is robust across nutrient and genetic perturbations, but with changes to its oscillatory parameters. Finally, to address whether single-cell metabolic cycles confirm findings from chemostat-based studies, I emulated situations known to affect dissolved-oxygen traces. These situations include potassium deficiency and deletions of genes with roles in metabolism and biological timekeeping.

3.1 Coupled oscillations in permissive conditions

To show that metabolic cycles are generated autonomously and are coupled to the cell division cycle, I replicated the single-cell flavin oscillations observed by Baumgartner et al. (2018), using prototrophic FY4 cells engineered with an HTB2::mCherry insertion. Replicating results from a previous flavin-based microfluidics study is important to confirm that my use of ALCATRAS (Crane et al., 2014) monitored the yeast metabolic cycle, especially given that the microfluidics set-up differed from previous studies (Baumgartner et al., 2018; Papagiannakis et al., 2017).

The HTB2::mCherry insertion allows monitoring phases of the cell division cycle through quantifying the intensity of the fluorescence of the inserted protein over time (Garmendia-Torres et al., 2018) (Fig. 3.1), while also allowing monitoring flavin fluorescence by avoiding the overlap of flavin and GFP emission spectra.

Figure 3.1: (A) Engineering a fluorescent protein cassette fused to *HTB2* **(B)** allows the identification of phases of the cell division cycle through monitoring changes in fluorescence of the fluorescent protein. Adapted from Garmendia-Torres et al. (2018).

Fig. 3.2 shows that oscillations in flavin fluorescence peak when a bud forms shortly before G_2/M, as evidenced by prototrophic FY4 HTB2::mCherry cells grown in minimal medium supplemented with $20\,\mathrm{g\,L^{-1}}$ glucose. This figure also shows that in some cases, a metabolic oscillation occurred without cell division cycle progression or bud formation. Such cases, also revealed by Papagiannakis et al. (2017) via cycles of NAD(P)H fluorescence, confirmed that the metabolic cycle is generated autonomously from the cell division cycle.

Figure 3.2: Flavin fluorescence (blue, solid lines) and histone 2B (orange, dotted lines) levels in a single, representative FY4 HTB2::mCherry cell grown in $20\,\mathrm{g\,L^{-1}}$ glucose. Vertical lines (black, dashed) indicate budding events. Star ($\star$) indicates a flavin oscillation without a corresponding cell division cycle.

As observed, oxidation of flavin upon budding was expected for these reasons:

1. Flavin fluorescence peaks (becomes most oxidised) and NAD(P)H fluorescence peaks (becomes most reduced) at the same time in chemostat cultures (D. B. Murray et al., 2011).
2. NAD(P)H is in the reduced form when buds form (Papagiannakis et al., 2017).
3. The flavoprotein lipoamide dehydrogenase, coded by the gene *LPD1*, is in redox equilibrium with NAD(P)H (Siano & Mutharasan, 1989).

To quantify the period of the oscillators, I combined time series analysis methods. Fig. 3.3 shows that flavin fluorescence oscillated at a period of approximately 90 min, based on the mean Fourier spectrum and median autocorrelation function. Figs. 3.3b and 3.3d additionally show that the cell division cycle proceeded at the same period, as evidenced by the autocorrelation function of mCherry. The duration of the cell division cycles agrees with previously reported values (Herskowitz, 1988).

To visualise the relationship between the metabolic cycle and the cell division cycle, Fig. 3.4 shows that budding events synchronise with peaks in fluorescence and that the cell division cycle varies between cells, with most just under 2 h, agreeing with Fig. 3.3b (inset). The oscillatory shape of the median flavin fluorescence time series when aligned to the first budding event (Fig. 3.4b) further confirms the synchrony between the metabolic cycle and budding events.

Finally, to quantify the relationship between the metabolic cycle and the cell division cycle, I computed the cross-correlation function between the flavin and mCherry signals across the population (Fig. 3.4c). The flavin signal peaks, on average, precede the mCherry signal peaks by 5 min, as evidenced by the location of the peak of the cross-correlation function that is closest to the vertical axis. The cross-correlation function thus demonstrates the coincidence between peaks of flavin oscillations and mitosis.

Figure 3.3: (3.3a, 3.3c) Mean Fourier spectrum of flavin fluorescence time series across cells. **(3.3b, 3.3d)** Median autocorrelation functions of flavin fluorescence (blue) and histone 2B levels (orange) time series, along with ***(insets)*** the periods of each oscillator across cells as determined by the frequency with the greatest power in each signal's Fourier spectrum. Data are from FY4 HTB2::mCherry cells under 20 g L^{-1} glucose. Two experiment repeats shown: (3.3a)–(3.3b) are from the first repeat, and (3.3c)–(3.3d) are from the second repeat.

Figure 3.4: (3.4a) Heatmap showing the flavin fluorescence (pixels on a red-blue scale) and budding events (black pixels) of each cell over time. Signals are aligned by the first budding event. **(3.4b)** Median flavin fluorescence signal across cells, aligned to first budding event (two repeats: $n_1 = 361$, $n_2 = 144$). **(3.4c)** Median cross-correlation function between flavin and histone 2B signals (two repeats: $n_1 = 392$, $n_2 = 170$). Data are from FY4 HTB2::mCherry cells in $20\,\mathrm{g\,L^{-1}}$ glucose.

3.2 Decoupling between the metabolic and cell division cycles

To provide additional evidence that cells generate metabolic oscillations autonomously of the cell division cycle, I created a condition in which cells did not undergo cell division. Specifically, I did so by inducing starvation: I cultured FY4 and HTB2::mCherry cells in $7.5\,\mathrm{g\,L^{-1}}$ glucose for 7 h, switching them to $0\,\mathrm{g\,L^{-1}}$ glucose for 8 h, and then resumed $7.5\,\mathrm{g\,L^{-1}}$ glucose for 7 h. This abrupt induction of starvation is similar to experiments described by Bagamery et al. (2020), which showed that a population of genetically identical budding yeast cells, upon glucose starvation, formed two sub-populations that had different cellular physiology.

Fig. 3.5 shows that when cells were in high glucose, metabolic oscillations were asynchronous, consistent with Section 3.1, Papagiannakis et al. (2017), and Baumgartner et al. (2018). When cells grown in high glucose were abruptly starved of glucose, the proportion of rhythmic cells dropped (87% to 12%, see Table 3.1); among the rhythmic cells, flavin oscillations reset their phase and exhibited lower amplitudes.

Fig. 3.6 further shows that during starvation, growth rate dropped and budding events were sparse. The partial, rather than full, recovery of growth rate after starvation may be explained by accumulated phototoxicity over the course of the experiment generated by fluorescence imaging, with similar patterns across all experiments. Alternatively, the partial recovery of growth rate could also be explained by a memory of glucose starvation. In addition, the lower cumulative number of budding events in the HTB2::mCherry strain, consistent across all experiments, may be explained by the increased phototoxicity generated by mCherry imaging in five z-slices (Methods Section 2.1.2).

(a)

(b)

Figure 3.5: (3.5a) Flavin fluorescence (blue, solid lines) and histone 2B (orange, dotted lines) levels in a single, representative FY4 HTB2::mCherry cell. Vertical lines (black, dashed) indicate budding events. Shading (yellow) indicates glucose starvation. **(3.5b)** Heatmap showing the flavin fluorescence (pixels on a red-blue scale) and budding events (black pixels) of each cell. Data are from FY4 and HTB2::mCherry cells, subject to $7.5\,\mathrm{g\,L^{-1}}$ glucose for 7 h before being abruptly switched to $0\,\mathrm{g\,L^{-1}}$ glucose for 8 h and then resumed to $7.5\,\mathrm{g\,L^{-1}}$ glucose for 7 h.

(a) (b)

Figure 3.6: (3.6a) Mean growth rate, derived from changes in the sum of estimated parent cell and bud volumes (shading: 95% confidence intervals) and **(3.6b)** mean cumulative number of budding events per cell (shading: confidence intervals from bootstrapping, $n = 30$) of FY4 (blue) and HTB2::mCherry (orange) strains over time during the glucose-starvation experiment. Vertical lines (red) show changes in the nutrient medium.

The results show that metabolic oscillations can be generated when the cell division cycle is halted, providing strong evidence that the metabolic cycle is generated autonomously and independently of the cell division cycle. In addition, the results show that each cell can individually reset the phase of its metabolic cycle in response to abrupt changes in environmental conditions. Similar phenomena have been observed upon bulk addition of carbon sources (Krishna & Laxman, 2018; Kuang et al., 2017).

Importantly, the results also suggest that diffusion of signalling chemicals between cells is not required for generation of metabolic cycles. To prove this assertion, a bulk-culture experiment, such as from a turbidostat or chemostat, should be performed as a control. However, interpreting my single-cell microfluidics results in conjunction with reports from analogous chemostat-based studies (Tu et al., 2005) should still give strong evidence in support of the lack of requirement for cell-to-cell signalling.

The PDMS that forms the microfluidic devices may additionally affect results because small, hydrophobic molecules are known to be absorbed by (Toepke & Beebe, 2006) and adsorb onto the surface of (N. Li et al., 2009) PDMS. Though, the PDMS in my experiments has been treated to reduce absorption and adsorption (Crane et al., 2014). Additionally, Krishna and Laxman (2018) propose that acetyl-CoA and NADPH may serve as signalling chemicals for the metabolic cycles. Therefore, if absorption and adsorption still occur in my microfluidic devices, these processes would remove such signalling chemicals, further eliminating the possibility that signalling chemicals are required for generation of metabolic cycles.

To show that my microfluidics set-up creates fluid flow in chambers that would flush any signalling chemicals away from cells, making cell-to-cell communication impossible, Fig. 3.7 shows that there is no gradient of response along the direction of fluid flow. Specifically, there is no significant correlation between spatial position and the mean intensity of the raw flavin signal (x-axis: $\rho = -0.11$, $p = 0.10$, y-axis: $\rho = -0.0047$, $p = 0.95$). Additionally, there is also no significant correlation between spatial position and the period of flavin oscillations (x-axis: $\rho = -0.063$, $p = 0.44$, y-axis: $\rho = 0.016$, $p = 0.85$).

To further demonstrate that fluid flow in my microfluidic devices eliminates cell-to-cell signalling, further experiments can collect 'spent medium', i.e. nutrient medium that has passed through a chamber that contains cells. This spent medium can then be filtered and used as input nutrient media for a repeat experiment. Such an experiment would show that no signalling molecule that can synchronise metabolic cycles are in the spent medium.

(a)

(b)

Figure 3.7: Relationship between the spatial (X, Y) position of trap and **(3.7b)** the mean raw flavin signal across all time points, and **(3.7b)** the period of flavin oscillation (first peak of autocorrelation function), for each parent cell, during glucose starvation. Media flow is along the X axis.

Combined with results from the previous section, my data suggest that the metabolic cycle responds to external conditions and allows the cell division cycle to be initiated during a specific phase of the metabolic cycle, if conditions are favourable for growth.

If the metabolic cycle gates initiation of the early and late cell division cycle (Özsezen et al., 2019), then a condition that creates unfavourable conditions for cell division should lead to each cell progressing through its cell division cycle until its next gap phase, before halting its cell division cycle. Depending on which phase in the cell division cycle the cell happens to be in upon starvation, this next gap phase can either be G_1 or G_2/M. To test this implication, Fig. 3.8 shows that cells may remain in G_1 during glucose starvation (Fig. 3.8a), as evidenced by low mCherry intensity, or in G_2/M (Fig. 3.8b), as evidenced by high mCherry intensity.

(a)

(b)

Figure 3.8: Flavin fluorescence (blue, solid lines) and histone 2B (orange, dotted lines) levels in two sample FY4 HTB2::mCherry cells in the glucose starvation experiment. Vertical lines (black, dashed) indicates budding events. Shading (yellow) indicate glucose starvation. **(3.8a)** is an example of a cell with a low intensity of mCherry during starvation, while **(3.8b)** is an example with a high intensity of mCherry. The flavin and mCherry time series were normalised to give a mean of 0 and standard deviation of 1 so that they can be plotted on the same vertical axes, but the high-pass Butterworth filter was not applied.

Extending this investigation across a population of cells, Fig. 3.9b shows that the distribution of mCherry intensity becomes broader during starvation before resuming to a distribution resembling the initial condition upon restoration of glucose. This observation can be explained by a larger proportion of cells in G_2/M, giving high mCherry intensity, in contrast to the usually short time cells spend in G_2/M relative to the rest of the cell division cycle (Fig. 3.1). In contrast, Fig. 3.9a suggests that during starvation, the distribution of flavin intensity became narrower. In addition, the figure suggests that cells retain memory of starvation — in other words, whether each cell has encountered starvation before influences the cell's metabolic cycle — as evidenced by how the flavin signals after starvation were, overall, higher than before starvation.

Figure 3.9: Distributions of **(3.9a)** flavin and **(3.9b)** mCherry fluorescence over time, for the glucose-starvation experiment. Vertical lines (red, dashed), indicate times of medium changes. As a control, distributions of **(3.9c)** flavin and **(3.9d)** mCherry fluorescence over time for the high glucose ($20\,\mathrm{g\,L^{-1}}$) experiment are also shown. Raw time series were used to calculate the distributions.

3.3 Metabolic cycles in different genetic backgrounds

To show that the metabolic cycle is robust, I monitored flavin autofluorescence signals from the auxotrophic BY4741 strain. Cells of this strain were grown in minimal medium supplemented with uracil and the amino acids required for this strain to grow, in addition to $10\,\mathrm{g\,L^{-1}}$ glucose as the carbon source. Showing that metabolic cycles occur in an auxotroph is important because it shows that many cellular aspects must be impaired for the cycle to disappear, thus suggesting that the metabolic cycle is an intrinsic property of budding yeast. Similar to FY4 HTB2::mCherry cells, BY4741 cells showed robust, consistent oscillations in flavin fluorescence that peak upon budding (Fig. 3.10), although metabolic cycles have a period of $\approx$75 min in this case. The shorter period may be explained by a lack of burden caused by a lack of an mCherry insertion, and by nutritional supplements.

FY4 and BY4741 both derive from the S288c background strain. To show that the metabolic cycle is generated from a budding yeast strain other than S288c background strains, I performed a similar experiment with the prototrophic CEN.PK strain grown in minimal medium. Showing that metabolic cycles additionally occur in a different genetic background is important to emphasise that the cycles are intrinsic to budding yeast. CEN.PK is an important background to consider because it harbours genetic differences relative to S288c that results in physiological differences, including biotin prototrophy and malate metabolism (Nijkamp et al., 2012). Furthermore, CEN.PK has greater mitochondrial stability and a better gene regulatory response to levels of oxygen, owing to the insertion mutation that

Figure 3.10: (3.10a) Mean Fourier spectrum of flavin fluorescence time series across cells. **(3.10b)** Median autocorrelation function of flavin fluorescence time series, along with ***(inset)*** the periods across cells as determined by the frequency with the greatest power in each signal's Fourier spectrum. **(3.10c)** Heatmap showing the flavin fluorescence (pixels on a red-blue scale) and budding events (black pixels) of each cell. Signals are aligned by the first budding event. **(3.10d)** Median flavin fluorescence signal across cells, aligned to first budding event. Data are from BY4741 cells in $10\,\mathrm{g\,L^{-1}}$ glucose.

deactivates the *HAP1* gene in S288c (Gaisne et al., 1999). Fig. 3.11 suggests that CEN.PK113-7D cells exhibited ≈90 min flavin oscillations that were synchronised with budding events, similar to FY4 cells, thus further confirming the robustness of the metabolic cycle across genetic backgrounds.

Figure 3.11: (3.11a) Mean Fourier spectrum of flavin fluorescence across cells. **(3.11b)** Median autocorrelation function of flavin fluorescence time series, along with ***(inset)*** the periods across cells as determined by the frequency with the greatest power in each signal's Fourier spectrum. **(3.11c)** Median flavin fluorescence signal across cells, aligned to first budding event. Data are from CEN.PK113-7D cells in 10 g L^{-1} glucose.

3.4 Metabolic cycles in different carbon sources

To show that the metabolic cycle responds to nutrient conditions and, accordingly, adjusts the cell's metabolism and cell division cycle, I cultured cells in pyruvate and in a growth-limiting glucose concentration. These experiments are important as they confirm conclusions about varying nutrient conditions made by Papagiannakis et al. (2017), but using flavin autofluorescence. Specifically, pyruvate provided an example of a non-fermentable carbon source to test whether the switch from fermentative to respiratory metabolism affected the metabolic cycle. Additionally, a growth-limiting glucose concentration emulated low-glucose concentrations in a chemostat and was thus used to test whether long YMCs observed in such conditions can be replicated in a microfluidics platform.

Fig. 3.12 shows that FY4 HTB2::mCherry cells had longer metabolic cycles and cell division cycles (approximately 4 h) when grown in minimal media supplemented with $20\,\mathrm{g\,L^{-1}}$ pyruvate, compared to growth in high glucose. In addition, fewer (42%, see Table 3.1) cells exhibited metabolic cycles than in high-glucose conditions (87%). Furthermore, the synchrony between the metabolic cycle and cell division cycle remained, but with a longer lag (60 min) of the mCherry signal peak with respect to the flavin signal peak (Fig. 3.12c). Fig. 3.12a shows that the longer cell division cycles were because of longer G_1 phases but unchanged S/M phases, as evidenced by the longer flat regions of the mCherry signal.

(a)

(b)

(c)

Figure 3.12: (3.12a) Flavin fluorescence (blue, solid lines) and histone 2B (orange, dotted lines) levels in a single, representative FY4 HTB2::mCherry cell grown in $20\,\mathrm{g\,L^{-1}}$ pyruvate. Vertical lines (black, dashed) indicate budding events. **(3.12b)** Median autocorrelation functions of flavin fluorescence (blue) and histone 2B levels (orange) time series, along with ***(inset)*** the periods of each oscillator across cells as determined by the frequency with the greatest power in each signal's Fourier spectrum. **(3.12c)** Median cross-correlation function between flavin and histone 2B signals. Data are from FY4 HTB2::mCherry cells in $20\,\mathrm{g\,L^{-1}}$ pyruvate.

Fig. 3.13 shows that FY4 HTB2::mCherry cells had longer metabolic cycles when grown in minimal media supplemented with 0.010 g L^{-1} glucose.

(a)

(b)

Figure 3.13: (3.13a) Flavin fluorescence (blue, solid lines) and histone 2B (orange, dotted lines) levels in a single, representative FY4 HTB2::mCherry cell grown in 0.010 g L^{-1} glucose. Vertical lines (black, dashed) indicate budding events. **(3.13b)** Median autocorrelation functions of flavin fluorescence (blue) and histone 2B levels (orange) time series, along with ***(inset)*** the periods of each oscillator across cells as determined by the frequency with the greatest power in each signal's Fourier spectrum. Data are from FY4 HTB2::mCherry cells in 0.010 g L^{-1} glucose.

Additionally, Fig. 3.14 shows that the growth rate and the rate of bud formation of cells on limiting glucose was lower than on high glucose ($20\,\mathrm{g\,L^{-1}}$).

Figure 3.14: Mean growth rate, derived from changes in the sum of estimated parent cell and bud volumes (shading: 95% confidence intervals) of FY4 (blue) and HTB2::mCherry (orange) strains over time, for **(3.14a)** the glucose-limiting condition ($0.010\,\mathrm{g\,L^{-1}}$) and **(3.14b)** the high glucose condition ($20\,\mathrm{g\,L^{-1}}$). Similarly, the mean cumulative number of budding events per cell (shading: confidence intervals from bootstrapping, $n = 30$), of the same strains for **(3.14c)** the glucose-limiting condition and **(3.14d)** the high glucose condition.

Furthermore, Fig. 3.15 shows that the amplitude of the flavin oscillations in this glucose-limiting condition was low relative to other conditions, as evidenced by the lower signal-to-noise ratios (two-sided Kolmogorov-Smirnov test: low-glucose vs pyruvate $p = 1.6 \times 10^{-82}$, low-glucose vs high-glucose $p = 1.9 \times 10^{-171}$). The low quality of signals was further supported by the low proportion (12%, see Table 3.1) of cells manually scored as rhythmic.

Figure 3.15: Distributions of signal-to-noise ratios of flavin signals from cells in (yellow) 0.010 g L^{-1} glucose, (blue) 20 g L^{-1} pyruvate, and (orange) 20 g L^{-1} glucose. Vertical axis shows probability density, computed by dividing the number of cells in each bin by the total number of cells and then by the bin width.

Finally, Fig. 3.13b suggests that measurements of the metabolic cycle and the cell division cycle lost synchrony in limiting glucose. This was evidenced by an approximately 2.5 h to 3 h average metabolic cycle, though not robust, but an absence of consistent oscillations in mCherry intensity. This decoupling can be explained by a lack of cell division cycle events.

3.5 Metabolic cycles persist in potassium-deficient media

To address whether single-cell flavin traces from microfluidic experiments recapitulate dissolved-oxygen yeast metabolic cycles in chemostats, I replicated conditions of chemostat-based studies in which nutrient or genetic perturbations severely affected the metabolic cycle. These nutrient conditions included potassium deficiency and deletion strains included *zwf1*Δ and *tsa1*Δ *tsa2*Δ. Replicating conditions of chemostat-based studies is important in showing that the

single-cell metabolic cycle and the chemostat metabolic cycle are the same cycle, or to prove otherwise. Chemostat experiments obscure the behaviour of individual cells, and single-cell microfluidics experiments can provide a bottom-up explanation of high-level observations of the metabolic cycle in the chemostat. Such single-cell experiments could address, for example, whether the cellular behaviour of the yeast metabolic cycle explains the changes in dissolved-oxygen oscillations.

To test whether potassium deficiency eliminates metabolic cycles, I treated cells with potassium-deficient minimal medium, replacing monopotassium phosphate (KH_2PO_4) in the minimal medium with an equivalent molarity of monosodium phosphate (NaH_2PO_4), as described in Table 2.2, with both media supplemented with $20\,\mathrm{g\,L^{-1}}$ glucose as a carbon source. J. S. O'Neill et al. (2020) suggested that, in chemostat cultures, as potassium in the nutrient medium is gradually replaced with sodium, the period and amplitude of dissolved-oxygen oscillations decrease until the oscillations disappear (Fig. 3.16).

Figure 3.16: Decreasing extracellular potassium (K^+) concentration shortens, then under $1\,\mathrm{mmol\,dm^{-3}}$, destroys metabolic oscillations in the chemostat. Adapted from J. S. O'Neill et al. (2020).

Fig. 3.17a shows that FY4 HTB2::mCherry cells retained synchronised metabolic cycles and cell division cycles when cells were abruptly switched from potassium-containing to potassium-deficient minimal medium. Such cycles were longer and were generated less reliably (40% manually scored as oscillatory, see Table 3.1) as in the normal growth medium (Fig. 3.17b).

(a)

(b)

Figure 3.17: (3.17a) Flavin fluorescence (blue, solid lines) and histone 2B (orange, dotted lines) levels in a single, representative FY4 HTB2::mCherry cell. Vertical lines (black, dashed) indicate budding events. Shading (yellow) indicates the potassium-deficient period. **(3.17b)** Median autocorrelation functions of flavin fluorescence (blue) and histone 2B levels (orange) time series, along with ***(inset)*** the periods of each oscillator across cells as determined by the frequency with the greatest power in each signal's Fourier spectrum. Data are from FY4 and HTB2::mCherry cells; autocorrelation functions only used time points from the potassium-deficient condition (6 h to 16 h).

In addition, the distribution of signal-to-noise ratios of time series before potassium-deficiency and during potassium-deficiency were identical (Fig. 3.18; two-sided Kolmogorov-Smirnov test, $p = 0.10$), indicating an unchanged amplitude of flavin signal; however, the same was not true when comparing time series during potassium-deficiency and after potassium-deficiency (two-sided Kolmogorov-Smirnov test, $p = 4.09 \times 10^{-7}$).

Figure 3.18: Distributions of signal-to-noise ratios of flavin signals from cells before, during, and after potassium-deficiency.

In addition, to test whether potassium deficiency affected cell growth and division, Fig. 3.19a shows that growth rates recovered soon after a sharp decrease upon the abrupt switch to the potassium-deficient medium. This global response across cells is suggestive of an osmotic response as a result of intracellular potassium leaking out the cell, though the change and recovery of growth rate is slower than the change and recovery of cell volume in response to osmotic stress reported by Granados et al. (2017). Fig. 3.19b further shows that the rate of budding was unaffected during potassium-deficiency, in contrast to a pause under glucose starvation (Fig. 3.6b).

(a) (b)

Figure 3.19: (3.19a) Mean growth rate, derived from changes in the sum of estimated parent cell and bud volumes (shading: 95% confidence intervals) and **(3.19b)** mean cumulative number of budding events per cell (shading: confidence intervals from bootstrapping, $n = 30$) of FY4 (blue) and HTB2::mCherry (orange) strains over time during the potassium-deficient experiment. Vertical lines (red) show changes in the nutrient medium.

Finally, in contrast to glucose starvation, Fig. 3.20 suggests that potassium deficiency did not affect the time each cell spent in each phase of the metabolic and cell division cycles as they progressed through growth.

(a) (b)

Figure 3.20: Distributions of **(3.20a)** flavin and **(3.20b)** mCherry fluorescence over time, for the potassium-deficient experiment. Vertical lines (red, dashed), indicate times of medium changes. Raw time series were used to calculate the distributions.

In summary, my results show that upon switching from potassium-containing to potassium-deficient, sodium-containing media, cells' growth rates briefly dropped, but soon after, some cells resumed generating metabolic cycles and dividing in synchrony. My observations also indicate that during potassium-deficiency, the amplitude of metabolic oscillations remained constant, as evidenced by a lack of statistically significant changes in the signal-to-noise ratio distributions of flavin signals. These observations contrast J. S. O'Neill et al. (2020), which showed that in chemostats, as the potassium concentration in the medium is decreased, the amplitude of dissolved oxygen cycles becomes smaller until the cycles disappear. The differences between my observations and J. S. O'Neill et al. (2020) may be explained by the different experimental set-ups: if a population of cells generates metabolic cycles less reliably and over a broader range of periods, as revealed by the single-cell microfluidic set-up, then the combined effect of these cells in the chemostat can be the loss of dissolved oxygen cycles that arise from synchronous metabolic cycles. In addition, J. S. O'Neill et al. (2020) proposed that intracellular carbohydrate or amino acid stores are responsible for metabolic cycles. This leads to an alternative explanation for experimental discrepancies: these stores can be sufficient to drive metabolic oscillations during a relatively short period of potassium-deficiency in my experiment. Furthermore, the significance of the potassium-deficient, sodium-containing nutrient condition is highlighted by the toxicity of sodium ions for budding yeast (Ariño et al., 2010; Casey et al., 2013; Watcharawipas et al., 2018).

3.6 Metabolic cycles in deletion strains

To continue the investigation of whether single-cell flavin-based metabolic cycles recapitulate dissolved-oxygen metabolic cycles, I investigated the *zwf1*Δ and *tsa1*Δ *tsa2*Δ deletion strains. The investigation of deletion strains is important as they can lead to mechanistic explanations of the YMC.

To investigate whether the *zwf1*Δ strain shows abolition of the metabolic cycle in single-cell microfluidics, I used a *zwf1*Δ strain with the BY4741 background. Chemostat-based studies have suggested that in the *zwf1*Δ strain, metabolic cycles are abolished but with little change in growth rate (Tu et al., 2007). Cells were pre-cultured in $20\,\mathrm{g\,L^{-1}}$ pyruvate over 48 h and then cultured in $10\,\mathrm{g\,L^{-1}}$ glucose in the microfluidic device because higher glucose concentrations disfavour growth in this strain. As the strain had an auxotrophic background, the required nutrient supplements were also added. Fig. 3.21 shows that the *zwf1*Δ cells showed oscillations of approximately 3 h, but with low robustness (56% manually scored as oscillatory, see Table 3.1) and a wide distribution of signal-to-noise ratios, while the reference BY4741 strain showed robust flavin oscillations of approximately 1.5 h. These results conflict with the results from the chemostat-based study (Tu et al., 2007) that suggested that metabolic cycles are abolished in this strain.

To investigate whether the *tsa1*Δ *tsa2*Δ strain shows metabolic oscillations of a different waveform in single-cell microfluidics, I used a *tsa1*Δ *tsa2*Δ strain with the BY4742 background. Chemostat-based studies suggest that in the *tsa1*Δ *tsa2*Δ strain, metabolic cycles are shorter and exhibit an M-shaped dissolved oscillation trace due to an additional dip of oxygen consumption in the reductive-charging phase (Causton et al., 2015). To be consistent with *zwf1*Δ, cells were pre-cultured and cultured in the same conditions, but with the appropriate supplements for the auxotrophy of BY4742.

Figure 3.21: (3.21a) Flavin fluorescence (blue, solid lines) levels in a single, representative *zwf1*Δ cell. Vertical lines (black, dashed) indicate budding events. **(3.21b)** Mean Fourier spectrum of flavin fluorescence across cells. **(3.21c)** Median autocorrelation function of flavin fluorescence time series, along with ***(inset)*** the periods across cells as determined by the frequency with the greatest power in each signal's Fourier spectrum. **(3.21d)** Median flavin fluorescence signal across cells, aligned to first budding event. **(3.21e)** Distribution of signal-to-noise ratios of flavin signals from cells. Data are from *zwf1*Δ (BY4741) cells in 10 g L^{-1} glucose.

Fig. 3.22 suggests that the metabolic cycles generated from the auxotrophic *tsa1*Δ *tsa2*Δ strain were not consistent. Supporting this is the fact that the fraction of rhythmic cells fell from 83% in the parent strain to 32% in the *tsa1*Δ *tsa2*Δ strain (Table 3.1), though rhythmic cells exhibited oscillations with low noise, as evidenced by relatively high signal-to-noise ratios (median = 2.2). Additionally, the Fourier spectra (Fig. 3.22b) suggest that oscillations of around 2.5 h and 4 h were prominent in the population. These two prominent periods were obscured by the median fluorescence signal aligned by first budding event (Fig. 3.22d), which merely suggests that a 3 h oscillation was prominent in the population, and by the median autocorrelation function (Fig. 3.22c), which suggests that the oscillations were not at a consistent frequency across the population, in contrast to the BY4742 wild-type that showed robust 1.5 h oscillations.

Taken together, the flavin-based metabolic cycles from single-cell experiments and the dissolved-oxygen cycles from chemostat experiments with *zwf1*Δ and *tsa1*Δ *tsa2*Δ exhibited discrepancies. These discrepancies may be explained by the physical separation between cells in the microfluidic device and the lack of a glucose starvation treatment that is usually performed in chemostat-based experiments. For *zwf1*Δ, the population of cells in the single-cell microfluidic device shows a heterogeneity of flavin signals, unlike the more homogeneous parent BY4741 strain. Therefore, the absence of dissolved oxygen cycles observed in the chemostat from this strain, as reported by Tu et al. (2007), may be the combined effect of metabolic cycles of varies periods that are asynchronous due to the loss of ability to reset phase in response to the glucose starvation at the beginning of the chemostat experiment. For *tsa1*Δ *tsa2*Δ, the population of cells in the single-cell microfluidic device exhibits flavin signals that mostly fall under two main periods. Therefore, the M-shaped oxygen cycles observed in the chemostat from

Figure 3.22: (3.22a) Flavin fluorescence (blue, solid lines) levels in a single, representative *tsa1*Δ *tsa2*Δ cell. Vertical lines (black, dashed) indicate budding events. **(3.22b)** Overlaid Fourier spectra of flavin fluorescence across cells ($n = 212$). **(3.22c)** Median autocorrelation function of flavin fluorescence time series, along with ***(inset)*** the periods across cells as determined by the frequency with the greatest power in each signal's Fourier spectrum. **(3.22d)** Median flavin fluorescence signal across cells, aligned to first budding event. **(3.22e)** Distribution of signal-to-noise ratios of flavin signals from cells. Data are from *tsa1*Δ *tsa2*Δ (BY4742) cells in $10\,\mathrm{g\,L^{-1}}$ glucose.

this strain, as reported by Causton et al. (2015), may result from the additive effect of two sub-populations of cells that each have a different natural metabolic cycle period, and still retain the ability to reset phase in response to the glucose starvation at the beginning of the chemostat experiment.

3.7 Discussion

3.7.1 Summary of results

To summarise the results in this chapter, Table 3.1 shows summary statistics pertaining to the period of metabolic oscillations for each experiment discussed.

3.7.2 Interpretation of results

This chapter confirms the presence of flavin-based single-cell metabolic cycles, and further confirms that they are autonomous and may gate the cell division cycle across nutrient and genetic perturbations.

My work differs from Baumgartner et al. (2018), a similar single-cell microfluidics study of flavin-based metabolic cycles, on several aspects. In Baumgartner et al. (2018), a row of several cells can be physically touching within a cell trap; my work uses a trap design that can separate single cells with more physical distance and allow more media flow around the cells (Crane et al., 2014), and this design better minimises cell-to-cell communication. Furthermore, the two cell division cycle reporters that Baumgartner et al. (2018) used must be in separate cell strains and can each only indicate single events — Whi5p localisation indicating START and Nhp6a localisation indicating separation between parent and progeny cell nuclei. In contrast, the Htb2p reporter in this chapter can alone indicate three cell division cycle events: the G_1-S transition, the S-G_2/M transition, and

Experiment & section referenced	Number of time series	Number of oscillatory time series	Mean estimated period & standard error of the mean [min]	Notes
High glucose, repeat 1 (3.1)	190	166 (87.4%)	115.7 ± 3.6	
High glucose, repeat 2 (3.1)	52	42 (80.8%)	111.9 ± 7.9 †	
Glucose starvation (3.2)	212	26 (12.3%)	117.3 ± 10.2 †	Only time points during starvation used.
BY4741 (3.3)	298	246 (82.6%)	92.4 ± 2.6 ∗	
CEN.PK (3.3)	29	6 (20.7%)	118.3 ± 24.7 †	
Pyruvate (3.4)	360	152 (42.2%)	247.4 ± 6.2 ∗	
Low glucose (3.4)	1520	183 (12.0%)	193.6 ± 7.6 ∗	
Potassium deficient (3.5)	511	206 (40.3%)	132.8 ± 2.8 ∗	Only time points during potassium deficiency used.
*zwf1*Δ (3.6)	112	63 (56.3%)	192.8 ± 7.7 ∗	
*tsa1*Δ *tsa2*Δ (3.6)	212	67 (31.6%)	157.7 ± 7.7 ∗	Single value for mean period may not be representative because population likely contains two dominant periods.

Table 3.1: Summary statistics of time series of flavin oscillations for each experiment. Oscillatory and non-oscillatory time series were human-labelled, and the percentage of total time series that were labelled as oscillatory is shown. Periods were estimated by finding the location of the first peak of the autocorrelation function of each oscillatory time series. Two-sample independent t-tests for the means were performed between the 'high glucose, repeat 1' experiment and other experiments — ∗ indicates statistically significant differences, while † indicates statistically insignificant differences ($p = 0.05$).

anaphase/cytokinesis, thus giving a more informative view of the cell division cycle (Garmendia-Torres et al., 2018) In addition, Baumgartner et al. (2018) and this chapter investigated different perturbations of the metabolic cycle. The former investigated the effect of nitrogen sources and respiration-deficiency, while my work investigated the effect of carbon sources, the effect of potassium, and

deficiencies in redox regulation. Though, the presence of metabolic cycles in *atp5*Δ and *cyt1*Δ mutants (Baumgartner et al., 2018) and in high-glucose conditions in this chapter support the assertion that respiration is not required for the metabolic cycle.

My results suggest that yeast cells independently generate the metabolic cycle which locks the cell division cycle in-phase. This conclusion was supported by the observation that flavin cycles were asynchronous between cells and peaks coincide with bud formation. These observations were consistent with Papagiannakis et al. (2017) and Baumgartner et al. (2018).

Results in pyruvate additionally reveal that as the metabolic cycle lengthens, G_1 lengthens but S/M stays the same length, suggesting a model in which a specific phase of the metabolic cycle gates entry into the cell division cycle. Importantly, metabolic cycles still occurred even when cells did not divide. This holds true for one-off skipping of cell division and conditions in which cells pause cell division for long periods of time.

My results additionally show that the metabolic cycle and the cell division cycle can be decoupled, reinforcing the idea that the metabolic cycle is autonomous from other cellular oscillators. In particular, I observed that single-cell flavin oscillations could synchronise and reset phase in response to abrupt starvation, while the cell division cycle was paused. This observation also suggests that the metabolic cycle is individually generated across cells without the need of a diffusible metabolite as proposed by Krishna and Laxman (2018). The conclusion that the metabolic cycle is generated independently across cells is evidenced by the continued presence of metabolic cycles during starvation even though the cells were physically separated, with nutrient media perfused across them.

The distributions of flavin and mCherry signals during starvation provided some mechanistic basis for the coupling or decoupling between the two oscillators. This observation suggests that if starvation occurs before START, the cell remains in the G_1 phase. Otherwise, the cell proceeds through the cell division cycle until it pauses in the G_2 phase. Such observations strengthen the idea that the metabolic cycle independently gates the early (START) and the late (mitotic exit) phases of the cell division cycle (Özsezen et al., 2019). However, the biochemical mechanism by which the cell uses to reset the phase of its metabolic cycle remains unclear.

My observations confirm that cells adapt their metabolic cycle to nutrient conditions: the metabolic cycle lengthened when cells were grown on pyruvate, while the cycle exhibited a lengthened period and a lower amplitude when cells were grown on low glucose. A possible explanation is that nutrient conditions that favour respiration over fermentation — and thus slower growth rate — leads to slower YMCs.

My results suggest discrepancies between chemostat and single-cell studies of the metabolic cycle, in particular, with regards to potassium-deficient conditions and the *zwf1*Δ and *tsa1*Δ *tsa2*Δ deletion strains. Such discrepancies warrant models to explain the observations. Alternatively, these discrepancies may represent additional cases in which the metabolic cycle is observed in single cells in conditions previously thought to be unfavourable based on chemostat observations, as was the case for metabolic cycles in high-glucose conditions.

I observed that metabolic cycles persist in potassium-deficient conditions, in contrast to J. S. O'Neill et al. (2020) which suggested that the oscillations disappear. The disappearance of dissolved oxygen cycles can alternatively be explained by a loss of synchrony in the population. However, while J. S. O'Neill et al. (2020) gradually replaced potassium with sodium over a longer period of time, my

experiments included an abrupt potassium depletion but over a shorter period of time. Therefore, the carbohydrate or amino acid stores that J. S. O'Neill et al. (2020) proposed were responsible for metabolic oscillations may still have been sufficient to drive metabolic oscillations during potassium-deficiency in my experiment.

I also observed that *zwf1*Δ exhibited metabolic cycles, though with varying amplitudes. This contrasts with Tu et al. (2007), which suggested that metabolic cycles in this strain were abolished. To reconcile findings, a potential explanation is that cells lose synchrony or the ability to reset phase, while growth is still maintained. *ZWF1* codes for glucose-6-phosphate dehydrogenase, which catalyses the first step of the pentose phosphate pathway involving the reduction of $NADP^+$ to produce NADPH, a key metabolite in the YMC (Nogae & Johnston, 1990). Thus, it is expected that the *zwf1*Δ deletion should affect a broad range of metabolic processes, including flavin oscillations, owing to the role of NAD(P)H redox in the function of the most abundant flavoproteins (Gudipati et al., 2014). However, some of the deleterious effects of *zwf1*Δ deletion may be compensated by *ALD6* and *IDP2* as they also catalyse reactions that produce NADPH (Minard & McAlister-Henn, 2005), therefore, explaining why growth is retained in the *zwf1*Δ strain.

My results additionally suggest that *tsa1*Δ *tsa2*Δ exhibits a range of metabolic cycle frequencies. This observation can be reconciled with the M-shaped dissolved oxygen cycles described by Causton et al. (2015) through a potential explanation: there are at least two substantial cell populations that each produce a different frequencies of metabolic oscillations, and the M-shaped waveform is the sum of the effect of individual cells. *TSA1* and *TSA2* are paralogous genes that are involved in redox metabolism. Specifically, these genes code for peroxiredoxins as part of the peroxiredoxin-thioredoxin system, which functions to reduce intracellular

reactive oxygen species (Wong et al., 2002). In addition, these genes have been suggested to be linked to the circadian rhythm, as evidenced by an approximately 24 h oscillation in oxidation patterns (Edgar et al., 2012). Therefore, deletion of these genes may lead to loss of regulation of timekeeping, leading to the different oscillation frequencies.

Taken together, the discrepancies between chemostat and single-cell studies highlight the role of sub-populations that cannot be captured in the chemostat, but possibly in single-cell studies.

3.7.3 Study caveats and future directions

Characteristics of the single-cell metabolic cycle

Time series of NAD(P)H oscillations, especially if recorded alongside flavin in the same cells, would strengthen the evidence that flavin autofluorescence oscillations are equivalent to the single-cell metabolic oscillations described by previous microfluidics studies. Such data would also provide a novelty: two fluorophores that act as read-outs of the metabolic cycle have, to my knowledge, never been recorded from the same cell.

To explore the link between the components of flavin autofluorescence and cycling of storage lipids as a proposed biochemical mechanism of the metabolic cycle, the *fas1*Δ strain may be studied. *FAS1* codes for the most abundant flavoprotein, the beta subunit of fatty acid synthetase, which has a role in lipid metabolism (Gudipati et al., 2014). The investigation may be strengthened with a rescue experiment using lipid sources such as glycerol trihexanoate or glycerol trioctanoate. This avenue of exploration may lead to additional insight on the biochemical basis of the yeast metabolic cycle, which is still poorly characterised.

To explore the conditions that make budding yeast cells reset their metabolic cycle phases, future experiments may include adding carbon sources in bulk, using the media-switching system of ALCATRAS. Such experiments could include acetate, acetaldehyde, or ethanol (Krishna & Laxman, 2018; Kuang et al., 2017). Insights from such experiments may lead to a broader understanding of the control of the sequence of events in the metabolic cycle.

Chemostat vs single-cell

Additional nutrient conditions can be used to address the discrepancies between the chemostat and single-cell microfluidics, namely, low glucose conditions and feast-and-famine conditions.

Low glucose conditions emulate conditions in the chemostat and may lead to long metabolic cycles, thus explaining why cycles with periods up to 14 h have been observed in chemostats. However, the glucose concentrations in chemostats are below the tolerance of measurement with current technologies; therefore, I propose that the limiting concentration of $0.010\,\mathrm{g\,L^{-1}}$ is used. Experiments with deletion strains can then be performed under these low-glucose conditions to investigate whether such conditions lead to a closer equivalence between chemostat and single-cell studies.

Feast-and-famine conditions have been modelled in chemostat cultivation of yeast (Jones & Kompala, 1999). Further experiments can therefore use rapid media-switching in ALCATRAS to produce regular glucose pulsing. In this type of experiment, cells are fed with a glucose-limited medium for an amount of time, then switched on to a glucose-rich medium for a short period of time (approximately 10 min), and the cycle repeats. The interval between glucose pulses can be varied to investigate the effect of an external entraining mechanism on the system of coupled oscillators that defines the yeast metabolic cycle. This design would be

similar to Charvin et al. (2009), which investigated the effect of intervals of glucose pulsing on the cell division cycle in budding yeast. A glucose pulsing experiment can thus also lead to a mathematical model of coupled oscillations in the yeast metabolic cycle.

Chapter 4

Analysis of oscillatory time series in the yeast metabolic cycle

Short and noisy oscillatory time series are challenging to analyse to give usable characteristics such as period and phase, especially because of the limited information they encode. Microfluidics experiments capture up to 5–10 periods of metabolic cycles, so a Fourier spectrum gives period estimates at low resolution, unless additional steps are taken, such as zero-padding the time series.

In addition, there is no standard method of analysing oscillatory time series such as those that arise from the yeast metabolic cycle. This is evidenced by the variety of methods previously used to analyse time series from single-cell metabolic oscillations. Specifically, Papagiannakis et al. (2017) used a smoothing spline to remove trends, the autocorrelation function and trough-to-trough intervals to estimate periods, and a Cox proportional hazard model (Cox, 1972) to characterise features of time series. In contrast, Baumgartner et al. (2018) used the Savitzky-Golay filter to filter time series, used the autocorrelation function along with both peak-to-peak intervals and trough-to-trough intervals to estimate periods, and simple differences to quantify the synchrony between two types of time series. Finally, Özsezen et al. (2019) used LOWESS spline fitting to remove trends in time series.

To characterise the properties of large sets of time series generated by microfluidics experiments, I sought to develop a pipeline of time series analysis methods.

Here, I propose steps for the analysis of datasets of 100–1000 time series related to the yeast metabolic cycle from single cells, using my experimental data as an example case.

Specifically, this chapter focuses on:

1. Data cleaning: choosing data and filtering out long-term trends that may confound analysis.
2. Visualising groups in a dataset: identifying groups within a population of time series based on their similarities. Such a relationship may include groupings or structures within the population of time series.
3. Detection of rhythmicity: determining whether a time series exhibits oscillations.
4. Period estimation: identifying the period of a single time series.
5. Detection of synchrony: identifying whether two types of signal from the same cell are synchronous, and to what extent.

In this chapter, I show that a high-pass Butterworth filter gives control over frequencies when filtering out long-term trends. To discover structure within a dataset, I show that UMAP, a dimension-reduction technique, and modularity clustering, a community detection technique, led to similar groupings. Subsequently, I compared three approaches to rhythmicity detection: a statistical test based on a spectral method, model-fitting and analytically computing a periodogram, and a simple machine learning model. To estimate the period and noise parameters of time series, I then explored the effect of noise on the autocorrelation function of synthetic time series. Finally, I used the cross-correlation function to detect synchrony and to quantify the relationship between two types of oscillators.

4.1 Analysing time series in a biological context

Previous studies have described computational pipelines that included mathematical methods to analyse biological time series. Zieliński et al. (2022) described a software pipeline (BioDare/BioDare2), catered to circadian rhythm studies, to estimate the period and detect rhythmicity in time series. This pipeline includes choices of methods to detrend and normalise time series, followed by a choice of methods to estimate the period, phase, and amplitude of the time series. Furthermore, the pipeline also includes statistical tests for the presence of an oscillation in a time series: an implementation of the JTK_CYCLE test (Hughes et al., 2010) along with an empirical derivative (Hutchison et al., 2015).

The software BioDare2 builds upon Zieliński et al. (2014), which compared and contrasted a set of period-estimation methods — FFT-NLLS, mFourFit, MESA, the Enright periodogram, the Lomb-Scargle periodogram, and spectrum resampling — to conclude with recommendations on time series analysis.

Studies of biological rhythms have used the BioDare pipeline to quantify features of oscillations of fluorescence. These included using FFT-NLLS to calculate the period and amplitude error of fluorescence in mouse brain sections to determine the mechanistic basis of the synchronisation of the suprachiasmatic nucleus (Hamnett et al., 2019). Another study used linear detrending of data followed by FFT-NLLS and spectral resampling to estimate the period and amplitude error of delayed fluorescence of chloroplasts in *Kalanchoe fedtschenkoi* leaves to determine whether phosphorylation of phosphoenolpyruvate affects robustness of circadian rhythms (Boxall et al., 2017).

In addition, Fulcher and Jones (2017) described a software pipeline, termed *hctsa*, that computes over 7700 time series features for input time series. The resulting feature matrix — a row for each time series and a column for each feature — could then be used to identify sets of features that are useful to discriminate between sets of time series or to identify clusters of time series based on their properties. The publication then showed that *hctsa* could be used to distinguish five *Caenorhabditis elegans* strains based on their movement patterns, and to identify clusters in the feature space of time series of *Drosophila melanogaster* movement patterns which correspond well to experimental groups. To reduce computation time, Lubba et al. (2019) identified 22 features, termed *catch22*, of *hctsa* that performed well in time series classification tasks based on 93 datasets (Table A.1 in Appendix A.1).

Taken together, BioDare and *hctsa* serve as two examples of approaches to analyse time series. However, these approaches overlap: the *catch22* feature set, itself a subset of *hctsa*, includes Fourier coefficients used in some BioDare methods. Additionally, *catch22* includes and a periodicity measure based on the autocorrelation function (Wang et al., 2007), similar to the period estimation method that will be discussed in Section 4.5. Such overlaps in measures may mean that the use of *catch22* may reduce to a few measures equivalent to analysis using BioDare — this implication is explored further in Section 4.4.3.

4.2 Data cleaning: filtering out long-term trends

Biological time series often have long-term trends. In my study of the yeast metabolic cycle, such trends include slow, global changes in flavin autofluorescence, which must be removed to uncover the periodic behaviour of flavin autofluorescence that is a component of the metabolic cycle. To determine the detrending method that is most appropriate for my data, I compared a frequency filtering method with a sliding-window detrending method.

To demonstrate the use of a method that modifies the frequency profile of time series to remove trends, Figs. 4.1a–4.1b show how a time series and its Fourier spectrum changes after the application of a high-pass Butterworth filter with a critical frequency of $2.86 \times 10^{-3}\,\mathrm{min}^{-1}$, corresponding to a period of 350 min. Defining a signal filter offers direct control over frequencies. The critical frequency was chosen as a reasonable upper limit of periods of the yeast metabolic cycle, based on my observations in single-cell microfluidics experiments. Defining the critical frequency in this way excludes the possibility of metabolic cycles that have very long periods in favour of emphasising metabolic oscillations of an expected frequency.

To show how sliding-window methods may adversely affect the frequency profile of time series when used for detrending, I computed the Fourier spectrum of time series detrended using the moving average method. Sliding-window methods are common in detrending biological time series. For example, Cuny et al. (2022) used a moving average: a constant, defined sliding window to smooth time series of yeast cell mass during growth. Fig. 4.1c shows that the moving average method introduced an artefact in the frequency spectrum near the reciprocal of the window size and decreases the number of time points.

Figure 4.1: (Left panels) Time series and (right panels) Fourier spectra corresponding to **(4.1a)** a sample raw time series of flavin autofluorescence, **(4.1b)** the time series processed by a high-pass Butterworth filter with a critical frequency $2.86 \times 10^{-3}\,\mathrm{min}^{-1}$, and **(4.1c)** the time series detrended using a moving average (window size 30 time points). Arrow (↘) indicates artefact.

4.3 Visualising groups in the dataset

To identify structures in datasets of time series, I implemented UMAP, a dimension-reduction method, and modularity clustering, a graph-based clustering method. Such data visualisation methods are important because the structures they show may identify differences between groups that are biologically relevant — for example, sub-populations of oscillations with similar properties. Previous efforts in using computational methods to identify groups in a set of biological time series include using k-means clustering to identify clusters of transcript cycling patterns that correspond to phases of the YMC (Tu et al., 2005), development of a method to cluster featurised multivariate time series based of videos of human motion (Wang et al., 2007), and using signal entropy to featurise fMRI signals followed by modularity clustering to partition the signals into brain regions (Shafiei et al., 2019).

To demonstrate the data visualisation methods, I used time series of flavin auto-fluorescence oscillations from one experiment with both the wild-type BY4741 strain ($n = 206$) and the mutant *zwf1*Δ strain ($n = 425$). These time series had time points sampled every 5 min in the experiment, for a total of 163 time points. I manually labelled the time series to indicate whether they were oscillatory or not, with 142 of the 206 BY4741 time series classed as oscillatory and 224 of the 425 *zwf1*Δ classed as oscillatory.

4.3.1 UMAP

UMAP (McInnes et al., 2020) is an unsupervised dimension reduction method that can be used to visualise structure in a dataset. Specifically, UMAP aims to find a manifold structure of the input observations and compute a low-dimensional embedding that preserves the topological structure of the manifold. This embedding thus serves as coordinates to plot the data onto a low-dimensional space.

To evaluate whether UMAP was able to discover a structure within the BY4741 & *zwf1*Δ dataset that corresponded to meaningful divisions, I featurised the time series with *catch22*, then used UMAP to compute two-dimensional embeddings.

Fig. 4.2a demonstrates that UMAP suggested a small group of non-oscillatory time series that differed markedly from the rest ($*$ in figure), and a larger group that was more similar to oscillatory time series ($**$ in figure). However, the figure also suggested that a third group of non-oscillatory time series occupied the same region in embedding space as oscillatory time series. In addition, Fig. 4.2b demonstrates that UMAP suggested that the BY4741 time series were more similar to each other than *zwf*Δ were to each other. This was evidenced by how *zwf1*Δ occupied larger regions of the embedding space than BY4741. These embeddings agreed with my observation that time series from the *zwf1*Δ strain had a larger variety of shapes and oscillation quality than the BY4741 strain.

The overlaps between oscillatory and non-oscillatory time series and between BY4741 and *zwf1*Δ time series in embedding space can be explained by two factors: the characteristics of the time series themselves, or the UMAP algorithm. To demonstrate that the characteristics of the time series contribute to their groupings within the embedding space, Fig. 4.3 shows non-oscillatory time series that correspond to the two regions described by Fig. 4.2a along with those that occupy the same region of embedding space as oscillatory time series. This figure

Figure 4.2: UMAP embedding ($n = 5$, min_dist $= 0.5$, $d = 2$, Euclidean distance as the metric) of a dataset of time series featurised using *catch22*. Each node represents a time series, coloured either by **(4.2a)** whether each is oscillatory or not, human-labelled ($*$ and $**$ indicating two groups of non-oscillatory nodes of interest), or by **(4.2b)** strain ('BY4741' or '*zwf1*Δ').

suggests that the groups of interest were defined by the characteristics of the time series. Specifically, Region 1 corresponds to non-oscillatory time series with very high signal-to-noise ratios, Region 2 corresponds to non-oscillatory time series with high signal-to-noise ratios, and Region 3 corresponds to time series with inconsistent oscillatory behaviour. Such time series in Region 3 were labelled non-oscillatory due to their inconsistent behaviour, but likely occupy the same region in embedding space with oscillatory time series because they share similar noise characteristics.

To further emphasise the role of the characteristics of the time series, Fig. 4.4 repeats the UMAP analysis, but on the first half of the time series. As the first half of each time series is likely to have similar time series characteristics as the whole time series, this figure provides further evidence to suggest that the characteristics of each time series are the major determinant of each node's position in the

Figure 4.3: Sample non-oscillatory time series from regions of interest in UMAP embedding space shown in Fig. 4.2. **(4.3a)** Non-oscillatory group (small group, $*$ in Fig. 4.2a). **(4.3b)** Non-oscillatory group (large group, $**$ in Fig. 4.2a). **(4.3c)** Non-oscillatory time series that occupy the same region of embedding space as oscillatory time series.

(a) **(b)**

Figure 4.4: UMAP embedding ($n = 5$, min_dist $= 0.5$, $d = 2$, Euclidean distance as the metric) of a dataset of the first half of time series used in Fig. 4.2, featurised using *catch22*. Each node represents a time series, coloured either by **(4.4a)** whether each is oscillatory or not, human-labelled, or by **(4.4b)** strain ('BY4741' or '*zwf1*Δ').

embedding space. This conclusion is further evidenced by the continued presence of a small group of non-oscillatory time series that differed substantially from the rest and a larger group that was more similar to oscillatory time series (Fig. 4.4a), similar to what is shown in Fig. 4.2a.

To improve the visualisation in Fig. 4.2, I performed a grid search of the n and min_dist UMAP hyperparameters (Appendix A.2) to find the best combination that separates the groups of time series. Fig. 4.5 suggests that $50 \leq n \leq 150$ and $0.25 \leq$ min_dist ≤ 1 resulted in a good separation between the BY4741 and *zwf1*Δ nodes. However, in this region, the two groups of non-oscillatory time series marked by $*$ and $**$ in 4.2 were merged. Altogether, these non-oscillatory time series were consistently displayed into groups separate from the rest as the hyperparameters were varied.

Figure 4.5: Grid search of UMAP hyperparameters: number of neighbours along the horizontal axis and minimum distance along the vertical axis. Data points are coloured according to category: grey indicates non-oscillatory time series, blue indicates oscillatory time series from *zwf1*Δ cells, and orange indicates oscillatory time series from BY4741 cells.

Varying parameter values to obtain better separation between groups of time series decreases the contribution of the UMAP algorithm to the overlap between groups. However, some overlap between non-oscillatory and oscillatory time series and between BY4741 and *zwf1*Δ time series remain, leading to the conclusion that the characteristics of the time series are chiefly responsible for their positions in embedding space. Thus, UMAP may have potential to visualise groups in a dataset of time series that may not correspond to biologically relevant groups, and this may lead to finding heterogeneity in time series among cells of the same strain.

4.3.2 Graph-based clustering

Modularity clustering is a mathematical method that partitions a graph into groups to optimise a 'modularity' value, defined so that the method finds a trade-off between maximising the connections within a cluster and minimising the connections between clusters (Newman, 2006). This optimisation problem is computationally difficult, so approximations such as the Louvain algorithm are needed for large networks (Blondel et al., 2008), with the Leiden algorithm (Traag et al., 2019) subsequently developed to ensure that communities are well-connected and to provide an optimum number of communities.

Furthermore, the intrinsic scale of modularity scales with the square root of the number of connections in the network. Therefore, if the network is large, there is a large resolution limit, preventing a modularity clustering algorithm from detecting small-scale structures (Fortunato & Barthélemy, 2007; Traag et al., 2011). To remedy this, algorithms that implement a resolution parameter (γ), were devised; the value of this parameter thus controls the scale at which communities are detected (Kumpula et al., 2007; Reichardt & Bornholdt, 2004).

To assess the performance of a graph-based clustering method in identifying clusters in time series data, I represented a dataset of time series as a graph before using modularity clustering to identify clusters. Fig. 4.6 illustrates this process, specifically:

1. *Constructing a graph representation:* Each time series was represented as a vector of features in n-dimensional space, where n is the length of the vector. Here, I represented each time series with a vector of 22 features using *catch22*. The cosine distances between each pair of vector was computed, and became the edge weights of a complete graph with each time series as a node.
2. *Pruning:* The complete graph was pruned by deleting edges, so that each node was connected to at least the k nearest neighbours. I used $k = 10$.
3. *Modularity clustering:* Modularity clustering was performed on the graph to partition the pruned graph into communities. In this step, I used the constant Potts model (Traag et al., 2011) and the Leiden algorithm (Traag et al., 2019).

Fig. 4.7a shows that construction of a pruned graph based on similarities between time series highlights two groups of non-oscillatory time series. Subsequently, Figs. 4.7b–4.7d show that the resolution parameter (γ) controls the number of communities detected. Non-oscillatory time series were assigned to a separate group when $\gamma = 0.01$, and the two sub-groups of non-oscillatory time series were separated when $\gamma = 0.02$; however, when $\gamma = 0.03$, the divisions changed. The Leiden algorithm (Traag et al., 2019) suggested 10 as the optimal number of communities, which was realised by $\gamma = 0.02$. For all γ values, modularity clustering was able to further show communities among the oscillatory time series.

Figure 4.6: Process of preparing a dataset of time series for modularity clustering. **(4.6a)** Constructing a graph representation: each time series was featurised and the cosine distances in feature space became edge weights. **(4.6b)** Pruning the complete graph so that each node had at least the k nearest neighbours. **(4.6c)** Modularity clustering, using the Leiden algorithm (Traag et al., 2019), to identify communities. Fig. 4.6c is adapted from Newman (2006). In all subfigures, data are synthetic and only serve to illustrate the process.

Figure 4.7: Pruned graph of BY4741 and *zwf1*Δ time series from the same experiment. **(4.7a)** Nodes coloured by group: grey, non-oscillatory; blue, oscillatory *zwf1*Δ; orange, oscillatory BY4741. Thickness of edges represent edge weights, scaled by similarity found by cosine distances. Additionally, nodes coloured by community as found by the constant Potts model (Traag et al., 2011) as the resolution parameter (γ) was varied: **(4.7b)** $\gamma = 0.01$ (4 communities), **(4.7c)** $\gamma = 0.02$ (10 communities), and **(4.7d)** $\gamma = 0.03$ (12 communities).

However, such communities did not divide along the division between BY4741 and *zwf1*Δ cells, suggesting that time series features alone were not able to divide these two strains. This agreed with my observation that some oscillatory *zwf1*Δ time series resembled BY4741 time series (Fig. 4.8).

Figure 4.8: Sample time series from BY4741 and *zwf1*Δ cells. *zwf1*Δ sample 1 does not resemble the BY4741 sample, while *zwf1*Δ sample 2 does.

In sum, the general agreement between UMAP and modularity clustering shows that the BY4741 & *zwf1*Δ dataset had internal structure defined by rhythmicity of time series. However, it was not clear from UMAP whether there were sub-populations among the BY4741 and *zwf1*Δ time series whose members exhibit similar types of oscillations. In contrast, while modularity clustering suggests that such sub-populations can be found based on connectivity between nodes, the number of sub-populations depends on the value of a resolution parameter.

4.4 Detection of rhythmicity

To identify metabolic cycles from flavin autofluorescence signals, it is important to have a systematic method to determine whether a time series is oscillatory. To determine the time series classification method that is most appropriate for my data, I compared a spectral method, a model-fitting method, and a machine learning method.

4.4.1 Spectral methods

In order to classify oscillatory and non-oscillatory time series, I modified a classifier based on a spectral method that included a statistical test (Methods Section 2.3.2). This classifier was based on Glynn et al. (2006), which described a method that employed the peak power from the Lomb-Scargle periodogram (Lomb, 1976) to rank time series by the quality of oscillation and to perform a statistical test to determine whether a time series is oscillatory or non-oscillatory (Scargle, 1982), as shown by Eqs. 2.2–2.3.

Figs. 4.9–4.10 suggest that the best- and worst-ranked time series according to the quality of their oscillatory signals conformed to subjective judgements of quality. Highest-ranked time series resembled sinusoids and therefore led to periodograms with a strong power corresponding to the frequency of the sinusoid that would model the time series. Conversely, lowest-ranked time series resembled white noise and led to periodograms with power equally spread across all frequencies, thus bringing down the height of the highest peak. However, some time series with irregular oscillations based on visual inspection (Fig. 4.9; ranks 2, 3) were given higher ranks than those with more regular oscillations based on visual inspection, thus calling into question the reliability of the ranking method.

Figure 4.9: (4.9a) Best five time series in the *zwf1*Δ dataset and **(4.9b)** their periodograms, ranked by the quality of oscillation based on the maximum power in the periodogram (Glynn et al., 2006).

Figure 4.10: (4.10a) Worst five time series in the *zwf1*Δ dataset and **(4.10b)** their periodograms, ranked by the quality of oscillation based on the maximum power in the periodogram (Glynn et al., 2006).

To detect rhythmicity in one time series, Glynn et al. (2006) calculates the probability of the null hypothesis that a peak in the periodogram occurs due to chance. When extended across a population of time series, rhythmicity detection by this method thus becomes a task of testing multiple hypotheses. Glynn et al. (2006) thus proposed controlling the false discovery rate, defined as the proportion of cases in which the null hypothesis is true among all hypotheses in which the test is declared significant (see Methods Section 2.3.2, specifically Eq. 2.3). Controlling the false discovery rate thus controls the expected proportion of oscillations that are classified as oscillatory.

To assess the performance of this method as a classifier for rhythmicity detection, Fig. 4.11 shows the receiver operating characteristic (ROC) curve, created as the false discovery rate was varied. The area under the ROC curve (0.762) suggests that the classifier performed modestly well, especially for a large ($n = 425$) dataset of time series with a large variety of quality of oscillations.

Figure 4.11: ROC curve of classifier based on Glynn et al. (2006) as the false discovery rate was varied. The true positive rate and false positive rates were computed based on the manual labels of the *zwf1*Δ dataset.

4.4.2 Model fitting

To assess the performance of a time series classification method based on the autoregressive model, I implemented the method described by Jia and Grima (2020), used to characterise synthetic time series of stochastic, oscillatory gene expression in a dividing cell. In the implementation of the autoregressive model used by Jia and Grima (2020), each data point was expressed as a linear combination of a number of data points that precede it, and model parameters led to an analytical solution for the periodogram, thus giving an advantage over the low-resolution Fourier spectrum (Methods Section 2.3.3). The resulting power spectra fell into four categories, one of which (Type I) corresponded to a lack of oscillations, characterised by an absence of a local maximum in the power spectrum (Fig. 4.12). This method thus allows computing the frequency of the oscillation from the location of the peak in the periodogram and quality of the oscillation from the height of the peak.

Figure 4.12: Power spectra (a) analytically derived from fitting an autoregressive model to time series (b) can be divided into four types. Type I lacks a local maximum and is denoted as lacking oscillations. Figure adapted from Jia and Grima (2020).

(a) **(b)**

Figure 4.13: (4.13a) Sample time series (orange), with a fitted autoregressive model (blue) of order 18 computed according to Jia and Grima (2020). **(4.13b)** Periodogram defined based on parameters of the autoregressive model.

Fig. 4.14 shows that the autoregressive model was able to correctly identify a time series as oscillatory at a period of 75.2 min, as evidenced by the location of a peak in the periodogram that the model predicted.

To assess the performance of the use of the autoregressive model for rhythmicity detection across a dataset, I extended this method across the *zwf1*Δ dataset, treating Type I power spectra (Fig. 4.12) as non-oscillatory. The confusion matrix (Table 4.1) suggests that the method leads to poor performance (precision = 0.532, recall = 0.629, no-skill classifier: precision = recall = 0.527, see Methods Section 2.3.4 for definitions).

		Predicted by AR model		
		Positive	Negative	Total
Human-defined labels	Positive	141	83	224
	Negative	124	77	201
	Total	265	160	425

Table 4.1: Confusion matrix to evaluate the performance of using the autoregressive model (Jia & Grima, 2020) to detect rhythmicity in the *zwf1*Δ dataset.

To find an explanation for the poor performance, I consider the false negatives (83 time series) and false positives (124 time series). To address the false negatives, Fig. 4.14a shows that Type I power spectra have lower optimal orders of the autoregressive model compared to Type II and Type III power spectra. With lower orders, less information is used to inform the power spectrum, and therefore it is more likely that the resulting power spectrum is of Type I.

A potential explanation for the false positives is that non-oscillatory time series correspond to power spectra with multiple local maxima, as opposed to a single local maximum that corresponds to the most likely period of an oscillatory time series. Such power spectra with multiple local maxima were thus classified as Types II or III. However, Fig. 4.14b shows that there is no relationship between the number of local minima and whether a time series is oscillatory (contingency table $\chi^2 = 2.90$, 3 degrees of freedom, $p = 0.41$). Thus, a better explanation is needed to explain the false positives.

Figure 4.14: (4.14a) Optimal order of autoregressive function fitted to time series, categorised by whether they are non-oscillatory or oscillatory time series (human-labelled) and the type of oscillation based on the analytical periodogram derived from the autoregressive function (Jia & Grima, 2020). No oscillatory time series have Type IV power spectra, thus not shown. **(4.14b)** Proportions of non-oscillatory and oscillatory time series that have 0 to 3 local maxima in their analytical periodograms.

4.4.3 Machine learning

As an alternative to the mathematical methods previously discussed, I trained a support vector classifier to classify oscillatory and non-oscillatory time series from the *zwf1*Δ cells (Appendix A.3).

To ensure that the dynamic ranges of the fluorescence signals do not affect rhythmicity detection, I normalised each time series $x_i(t_1), \ldots, x_i(t_j), \ldots, x_i(t_N)$ to produce a processed time series $z_i(t_1), \ldots, z_i(t_j), \ldots, z_i(t_N)$ as follows:

$$z_i(t_j) = \frac{x_i(t_j) - \mu_i}{\sigma_i} \tag{4.1}$$

where μ_i is the mean value of x_i computed across all time points, and σ_i is the standard deviation of x_i computed across all time points. As a result, each normalised time series z_i has a mean of 0 and a standard deviation of 1. From this input data, 75% of the time series formed the training set.

To determine the most effective way to featurise the data, I computed the precision and recall (defined in Methods Section 2.3.4) of support vector classifiers trained on data featurised using different methods. All support vector classifiers were trained using a radial bias kernel, a kernel coefficient $\gamma = 1/N$, where N is the number of features, and a regularisation parameter $C = 10$. Fig. 4.15 suggests that featurisation using *catch22* and the Fourier spectrum gave comparably high performances, as evidenced by high precision and recall, and with a low degree of overfitting, as evidenced by a small variation of both metrics across the rounds of cross-validation.

Figure 4.15: (Left) Precision and (right) recall from five-fold cross-validation of support vector classifiers trained using different featurisation methods: using the time points as features, using the power values in the Fourier spectrum as features, and using *catch22* features. As a control, the oscillatory and non-oscillatory labels were randomly reassigned to the time series and the time points were used as features.

To predict the probability that each time series was oscillatory, I used Platt scaling (Platt, 1999) with the support vector classifier, as implemented by the `predict_proba` method in the Python package `scikit_learn`. Platt scaling involves finding parameters for a logistic sigmoid function which approximates the posterior class probability for a binary classifier. Fig. 4.16 suggests that the classifier performed well in discriminating between the two classes, as evidenced by the U-shaped histogram of probabilities (Fig. 4.16a), in contrast to the control (Fig. 4.16b).

(a) **(b)**

Figure 4.16: (4.16a) Histogram of probabilities of whether a time series in the test data set is classified as oscillatory by the SVC (featurisation with *catch22*, $\gamma = 1/22$, $C = 10$), and as a control **(4.16b)** with labels randomly reassigned.

In addition, Fig. 4.17 demonstrates that the probabilities can serve as a good score to rank time series by oscillation quality.

Figure 4.17: Sample time series arranged by probability that each is oscillatory, as predicted by the support vector classifier.

To assess whether a specific mathematical measure of rhythmicity within the *catch22* feature set is responsible for classification of time series, Fig. 4.18 evaluates the *catch22* features that were most influential for the classifier. To do so, I used a Random Forest architecture with the same train-test split and featurisation as the best-performing SVC model, so that I can compute the mean decrease in accuracy as the value of each feature is randomly shuffled (Breiman, 2001). Fig. 4.18 suggests that the importance of each feature in rhythmicity detection was small and that there was no set of features that had significantly greater importance than others. In other words, classification of rhythmicity did not depend only on a few explicit mathematical measures of rhythmicity and adequate classification was only possible when all features were combined. This suggests that the use of *catch22* does not reduce to a few features equivalent to analysis using tools like BioDare (Zieliński et al., 2014).

4.5 Period estimation using the autocorrelation function

Estimating the period of oscillatory time series is important as it provides a quantitative measure of how yeast metabolic cycles respond to genetic nutrient perturbations, as was previously used by Papagiannakis et al. (2017). To show that the autocorrelation function can be used to estimate the period and noise properties of both symmetric and asymmetric oscillations, I adapted the autocorrelation function as used by Pietsch et al. (2023) (Methods Section 2.3.5). To calibrate the method, I generated synthetic oscillations — sinusoids and the FitzHugh-Nagumo oscillator (FitzHugh, 1961) — to investigate the effect of their

properties on the autocorrelation function. Subsequently, I applied the autocorrelation function to characterise experimentally-recorded time series. Details on how the sinusoids and the FitzHugh-Nagumo oscillators were defined can be found in Methods Section 2.3.6.

Figure 4.18: Permutation feature importance values of *catch22* features in classifying manually-scored oscillatory and non-oscillatory time series from the *zwf1*Δ dataset, arranged by the mean decrease in accuracy when the feature of interest is randomly shuffled (Breiman, 2001). To compute feature importance, a random forest classifier was fitted to the training dataset. Error bars show inter-trees variability.

4.5.1 Effect of noise parameters on the autocorrelation function of synthetic sinusoids

To compare the effect of Gaussian noise and Gillespie noise on the autocorrelation function, I computed autocorrelation functions from a population of sinusoids with either type of noise added via element-wise sums. Fig. 4.19b shows that the autocorrelation function computed from a population of out-of-phase sinusoids could be modelled by a cosine with the same period as the sinusoids. Following this, Fig. 4.19d shows that the addition of Gaussian noise preserved the point $(0, 1)$, but the amplitude of the cosine that models the autocorrelation function was decreased. Furthermore, the variation of the autocorrelation function among time series at long lags was increased, as evidenced by the interquartile range, because less data was used to compute the autocorrelation function at longer lags.

Gillespie noise is based on the birth-death process (Fig. 4.20), and its two parameters control noise parameters (Methods Section 2.3.6). Specifically, given a birth rate k_0 and a death rate d_0, the noise has a standard deviation of noise amplitude $A = \sqrt{k_0/d_0}$ and noise timescale $\tau = 1/d_0$.

Fig. 4.19f shows that when Gillespie noise was added to the sinusoids, the median autocorrelation followed the exponential decay function $y = \mathrm{e}^{-2d_0T}$, where T represents lag. In addition, the locations of the peaks of the autocorrelation function were preserved. The observation thus suggests that the death rate d_0 parameter of Gillespie noise controlled the shape of the autocorrelation function.

Figure 4.19: **(4.19a)** Sample sinusoids without noise, and **(4.19b)** its autocorrelation function. **(4.19c)** Sample sinusoids with Gaussian noise defined by drawing samples from $\mathcal{N}(0, \sigma^2 = 3)$, and **(4.19d)** its autocorrelation function. **(4.19e)** Sample sinusoids of with Gillespie noise ($k_0 = 5$ and $d_0 = 0.05$), and **(4.19f)** its autocorrelation function. Red line is defined by $y = \mathrm{e}^{-2d_0T}$, where T represents the lag in units of period of the sinusoids. For each case, the frequency of the sinusoids was 0.03, and there were 100 repeats, randomly out-of-phase.

(a)

(b)

Figure 4.20: (4.20a) Illustration of the birth-death process. The birth-death process is defined by two reactions: birth ($R_1 : \varnothing \longrightarrow \mathrm{P}$), with rate k_0, and death ($R_2 : \mathrm{P} \longrightarrow \varnothing$) with rate d_0, both of which affect the number N of species P. Random firing of the birth and death reactions, simulated by the Gillespie algorithm (Gillespie, 2007), defines the trajectory of N over time, which in turn defines Gillespie noise. **(4.20b)** Sample trajectory of a substrate created and destroyed by the birth-death process, simulated by the Gillespie algorithm ($k_0 = 5$, $d_0 = 0.05$, $t_{\max} = 1500$).

Figure 4.21: (4.21a) Sample sinusoids with Gillespie noise ($k_0 = 5$ and $d_0 = 0.5$), and **(4.21b)** its autocorrelation function. **(4.21c)** Sample sinusoids with Gillespie noise ($k_0 = 5$ and $d_0 = 0.005$), and **(4.21d)** its autocorrelation function. Red lines are defined by $y = \mathrm{e}^{-2d_0T}$, where T represents the lag in units of period of the sinusoids. For each case, the frequency of the sinusoids was 0.03, and there were 100 repeats, randomly out-of-phase.

To quantify the effect of the noise timescale on the shape of the autocorrelation function, I varied the death rate parameter d_0 when generating Gillespie noise. Fig. 4.21 shows that a higher death rate decreased the decay timescale of the autocorrelation function (Fig. 4.21b), while a lower death rate introduced long-term trends in the simulated signals (Fig. 4.21c). A lower death rate also increased the variation between autocorrelation functions between replicates (Fig. 4.21d).

Figure 4.22: (4.22a) Sample sinusoids with Gillespie noise ($k_0 = 25$ and $d_0 = 0.05$), and **(4.22b)** its autocorrelation function. **(4.22c)** Sample sinusoids with Gillespie noise ($k_0 = 1$ and $d_0 = 0.05$), and **(4.22d)** its autocorrelation function. Red lines are defined by $y = \mathrm{e}^{-2d_0T}$, where T represents the lag in units of period of the sinusoids. For each case, the frequency of the sinusoids was 0.03, and there were 100 repeats, randomly out-of-phase.

To quantify the effect of the noise amplitude on the autocorrelation function, I varied the birth rate parameter k_0 when generating Gillespie noise. Fig. 4.22 shows that a higher birth rate increased the amplitude of noise (Fig. 4.22a) and increased the variation between replicate autocorrelation functions (Fig. 4.22b), while the opposite was true for a higher birth rate (Figs. 4.22c–4.22d).

4.5.2 FitzHugh-Nagumo oscillator: effect of oscillation shape

To test whether Gillespie noise parameters can be estimated from the autocorrelation function computed from an asymmetric oscillation, I added Gillespie noise with varying d_0 and k_0 to FitzHugh-Nagumo oscillators as defined in the Methods Section 2.3.6. Fig. 4.23b shows that when the oscillator had a different shape, the waves in the autocorrelation function changed shape, becoming more pointed. Additionally, the effect of noise parameters on the autocorrelation function is preserved when the oscillators switch from sinusoid to FitzHugh-Nagumo oscillators.

(a) **(b)**

Figure 4.23: (4.23a) Sample FitzHugh-Nagumo oscillators ($RI_{\text{ext}} = 0.4$, $\tau = 12.5$, $a = 0.7$, $b = 0.82$) with Gillespie noise ($k_0 = 5$ and $d_0 = 0.05$), and **(4.23b)** its autocorrelation function. Red line is defined by $y = \mathrm{e}^{-2d_0T}$, where T represents the lag in units of period of the sinusoids. There were 100 repeats, randomly out-of-phase.

4.5.3 Real data from fluorescence microscopy in microfluidics experiments

To deduce the period and noise parameters of a experimentally-recorded sinusoid-like signal (Fig. 4.24a), I computed the autocorrelation functions of a population of flavin autofluorescence time series (Fig. 4.24b). The autocorrelation function suggests an average period of 19 time points, corresponding to 95 min, as expec-

ted from the nutrient conditions. However, estimation of noise parameters was complicated by the damping in the autocorrelation function, giving a different shape compared to the synthetic data and fewer peaks and troughs for fitting. Nevertheless, relating the shape of the autocorrelation function to the effect of noise parameters on synthetic sinusoids suggested a noise timescale of 7.35 and a noise amplitude of 110.45.

(a) **(b)**

Figure 4.24: (4.24a) Sample time series of flavin autofluorescence. **(4.24b)** Autocorrelation function across a population of time series of flavin autofluorescence.

In addition to flavin autofluorescence, I also recorded time series of histone 2B abundance as an indicator of the phases of the cell division cycle (Garmendia-Torres et al., 2018), to investigate whether the flavin autofluorescence oscillations and the cell division cycle synchronised. The abundance of histone 2B follow an asymmetric oscillation (Fig. 4.25a).

Similar to the previous section, to deduce the period and noise parameters of this signal, I computed the autocorrelation functions of a population of histone 2B abundance time series (Fig. 4.25b). The autocorrelation function also suggests an average period of 19 time points, corresponding to 95 min. As was the case for the flavin autofluorescence time series, the damping in the autocorrelation function complicated estimation of noise parameters, but the function suggested a noise

(a) **(b)**

Figure 4.25: (4.25a) Sample time series of histone 2B abundance. **(4.25b)** Autocorrelation function across a population of time series of histone 2B abundance.

timescale of 9.35 and a noise amplitude of 168.12. The differences of these noise parameters relative to the flavin autofluorescence time series suggest different noise properties, which can be explained by the different fluorescence channels and exposure times used to generate each type of signal.

4.6 Detection of synchrony

To test a method to detect the synchrony and quantify the temporal lag between two types of oscillations, I computed the cross-correlation functions of a population of sinusoid and FitzHugh-Nagumo oscillators. Cross-correlation has been used to investigate the relationship between the expression levels of two genes in a model feed-forward loop (Dunlop et al., 2008), and to investigate the relationship between instantaneous growth rate and the expression of *lac* genes of enzymes in central metabolism across a population of *E. coli* cells (Kiviet et al., 2014).

4.6.1 Synthetic data

Fig. 4.26b shows that the cross-correlation function identifies that the sinusoids, on average, peaked 20 time points before the FitzHugh-Nagumo oscillators, close to the actual value of 20.75 time points. This shift was evidenced by the position of the peak of the cross-correlation function closest to the vertical axis. The cross-correlation function further showed that synchrony between the two oscillators was maintained along the entire time series, across all time series. Furthermore, Fig. 4.26d suggests that even with strong Gillespie noise, the lag between the two oscillators could still be deduced from the cross-correlation function.

Figure 4.26: (4.26a) (Blue) Sample sinusoid ($f = 0.0235$) and (orange) FitzHugh-Nagumo oscillator ($RI_{\text{ext}} = 0.4$, $\tau = 12.5$, $a = 0.7$, $b = 0.82$) of the same frequency and without noise, and **(4.26b)** the cross-correlation function of the FitzHugh-Nagumo oscillators with respect to the sinusoids. **(4.26c)** (Blue) Sample sinusoid and (orange) FitzHugh-Nagumo oscillator with same parameters as 4.26a, but with Gillespie noise ($d_0 = 0.05$, $k_0 = 5$), and **(4.26d)** the cross-correlation function of the FitzHugh-Nagumo oscillators with respect to the sinusoids. For each case, there were 400 repeats, randomly out-of-phase.

4.6.2 Real data from fluorescence microscopy in microfluidics experiments

To show how the cross-correlation function can be used to quantify the synchrony between flavin autofluorescence oscillations and HTB2::mCherry levels in a population of cells, Fig. 4.27 displays a sample pair of time series (Fig. 4.27a) and the cross-correlation function from the population of cells (Fig. 4.27b). The cross-correlation function suggests that the histone 2B oscillations peaked after the flavin autofluorescence oscillations by an average of 5 min.

4.7 Discussion

This chapter discusses methods to filter long-term trends in time series data, to visualise structures within a dataset of time series, to detect rhythmicity in a time series, to estimate period and noise parameters, and to detect the synchrony between two time series.

The methods used in this chapter are more sophisticated than existing methods that have been used for analysing single-cell time series from the yeast metabolic cycle. The filtering methods are frequency-based, rather than based on fitting a model to the raw time series (Baumgartner et al., 2018; Özsezen et al., 2019). Detecting sub-groups in a dataset of metabolic cycle time series is also a novel approach. In addition, this chapter proposes using feature engineering and machine learning to extend existing mathematical approaches for rhythmicity detection. Furthermore, cross-correlation offers a way to robustly quantify the time lag between two types of time series which accounts for both noise and different

(a)

(b)

Figure 4.27: (4.27a) Sample time series of flavin autofluorescence (blue) and histone 2B abundance (orange). **(4.27b)** Cross-correlation function between the flavin autofluorescence time series and the histone 2B abundance time series.

shapes of oscillators, which is more robust than using simple differences as used by Baumgartner et al. (2018). Taken together, the methods in this chapter can give more information about single-cell time series of metabolic cycles and can more easily be automated than previous methods.

My results suggest that using a high-pass Butterworth filter to filter out long-term trends in time series data gives better control over the frequency profile of the time series than moving-average methods, which is often used to detrend time series from biological oscillators. Such results highlight that a degree of caution is needed to choose methods for such a crucial step in data analysis.

My exploration of UMAP and modularity clustering suggests that both methods were useful in discovering structure within a dataset of time series, particularly in discriminating between oscillatory and non-oscillatory time series. These methods further indicated sub-groups of time series that may have similar properties, such as shape or oscillation quality, which may correspond to sub-populations of metabolic cycle-producing cells in a culture. The consistency between the two methods strongly suggest that such groups in the dataset are meaningful.

Subsequently, my exploration of three approaches to detect rhythmicity — deriving a statistical test of a power spectrum, deriving a periodogram from an autoregressive model, and a binary classifier — highlights the difficulty of rhythmicity detection in noisy biological time series. The spectral method described by Glynn et al. (2006) had a modest performance. The autoregressive model was able to identify the most likely period in some time series, but otherwise classified most time series as non-oscillatory, and lacked a tuning parameter. Finally, the support vector classifier suggests that a simple machine learning model could be adapted for rhythmicity detection, subject to a good feature set and training data.

Changes in featurisation or classifier architecture may improve performance in discovering structure in a dataset of time series or rhythmicity detection. Other time series feature sets such as *tsfresh* (794 features from 63 methods) (Christ et al., 2018) and *FATS* (45 features developed for astronomical light curve data) (Nun et al., 2015) may be used in place of *catch22*. In addition, other machine learning architectures may be used. Such architectures include deep learning models more suited for time series data, such as Long Short-Term Memory, which learns how to remember and forget information (Hochreiter & Schmidhuber, 1997), and temporal convolutional networks, shown to outperform recurrent networks on tasks utilising time- and sequence-dependent data (Bai et al., 2018).

Ultimately, rhythmicity detection requires supplying a threshold in some form, be it a range of frequencies in which oscillations are expected (Zieliński et al., 2014), a parameter that controls the proportion of time series detected as oscillatory, or training labels. This is because there is no way to objectively specify a failure rate for a rhythmicity detection method as there is no independent method to estimate rhythmicity (Zieliński et al., 2014).

My observations concerning the autocorrelation function confirms its use for estimating the period of an oscillatory time series, as used previously by Papagiannakis et al. (2017). As periodicity-estimation methods have a limited ability to estimate the period of short, noisy time series owing to little input data, combining several such methods can be useful to produce a picture of the periodicity of oscillatory time series. For example, Potvin-Trottier et al. (2016) combines the autocorrelation function and the Fourier transform to study the changes in the periodicity of a modified model of the repressilator.

Furthermore, I showed that the autocorrelation function may be used to estimate parameters to describe noise, assuming that the noise can be modelled by the birth-death process. However, further work, such as synthetic time series generated from a wider variety of parameters or additional estimation methods are likely required for adequate estimation of noise parameters from real data. In addition, it is possible that other types of noise better describe the noise from real data, and such types of noise may lead to different effects on the autocorrelation function.

Finally, my results show that the cross-correlation function, as used by Dunlop et al. (2008), Kiviet et al. (2014), and Pietsch et al. (2023), can be used to detect synchrony between two sets of time series and to quantify the temporal relationship between the time series, even if the time series are very noisy.

Taken together, the analysis methods discussed in this chapter can form the basis of a powerful data analysis pipeline to analyse large datasets of oscillatory biological time series.

Chapter 5

Modelling yeast biosynthesis strategies under constraints

To better understand the mechanistic basis of the YMC, I sought to build a model of how the cyclic sequence of cellular events responds to extracellular nutrient conditions. However, it is challenging to develop a fine-grained model for the aspects of the YMC, especially if the detailed molecular mechanisms are unclear. Further complicating the development of a fine-grained model is the fact that the main read-outs from single-cell studies — NAD(P)H and flavin autofluorescence — are aggregate signals from several biochemical phenomena.

Thus, it is more feasible to construct coarse-grained models to answer biological questions about the YMC. One such question is whether a finite proteome is responsible for the sequence of events in the yeast metabolic cycle, especially considering the high energetic and resource requirements of protein synthesis (J. S. O'Neill et al., 2020; Zylstra & Heinemann, 2022).

Here, I use a genome-scale metabolic model and flux balance analysis (FBA) to address whether cellular events in the metabolic cycle reflect a strategy in resource allocation that optimises growth.

Specifically, I aim to evaluate these hypotheses:

1. A finite proteome pool gives rise to sequential scheduling of the synthesis of biomass components during growth: lipids, carbohydrates, amino acids, and nucleic acids. In other words, as an adaption, the cell synthesises biomass components in sequence rather than in parallel. This would explain the timing of biosynthetic events in the phases of the yeast metabolic cycle.
2. This resource allocation strategy remains advantageous in many nutrient conditions and deletion strains, which would explain the robustness of the yeast metabolic cycle.
3. Synthesis of biomass components in parallel, as opposed to in sequence, may be advantageous if the processes of synthesising each biomass component share similar enzyme levels across biomass components.

5.1 Modelling temporal scheduling of biosynthesis

Previous studies have attempted to use FBA to model how each phase of the YMC has different metabolic requirements. Takhaveev et al. (2023) showed that in different stages of the cell division cycle, the cell synthesises different components of its biomass at different levels. In the study, they blocked synthesis of each class of macromolecule and recorded the changes in single-cell NAD(P)H cycles, representing the YMC, to quantify the level of each class of macromolecule that the cell synthesises at each time point within a cell division cycle. Then, they used these activities as coefficients for a modified thermodynamic-stoichiometric metabolic model at each time point and used FBA to deduce biomass production rates.

Additionally, Cesur et al. (2022) constructed a different FBA model for each YMC phase based on transcriptomic and epigenetic data. They did so using the GIMME algorithm (Becker & Palsson, 2008), which excludes reactions that correspond to genes that are not expressed at certain time points. Both studies model the metabolic state of the cell at each phase of the YMC, rather than predicting the time the cell takes to replicate or to synthesise biomass components.

An attempt in extending FBA to solve a time-dependent resource allocation problem was Reimers et al. (2017). This study extended a genome-scale model of the cyanobacterium *Synechococcus elongatus* PCC7942 to find the temporal order of intracellular synthesis reactions which optimises the growth rate of the cell, under resource constraints. This model imposes time dependence by implementing a light-dark cycle as an external oscillator.

Traditional genome-scale models assume that the uptake rate of carbon source limits production. However, levels of each enzyme also restrict reaction fluxes, leading to the development of enzyme-constrained models. An enzyme-constrained model fits the assumption that there is a fixed number of amino acids the cell has to distribute (Weiße et al., 2015). Models like Sánchez et al. (2017) and Elsemman et al. (2022) constrain the total sum of fluxes based on a defined total amount of enzyme. Elsemman et al. (2022) additionally imposed a ribosome capacity constraint and compartment constraints.

In this chapter, I used an enzyme-constrained genome-scale model of *Saccharomyces cerevisiae* and performed FBA to simulate two strategies of scheduling of the synthesis of biomass components: in sequence and in parallel. In contrast to Takhaveev et al. (2023), which modelled the levels of biosynthesis for each biomass

component at each phase of the cell division cycle based on the aggregate single-cell response observed in experimental data, without predicting timescales, this chapter predicts the timescale for biosynthesis for different proteome pool sizes and different nutrient conditions.

To simulate the cell prioritising the synthesis of each biomass component in sequence, I ablated metabolites from the biomass reaction so that one biomass component remained, then optimised the model. To assess the advantage of sequential synthesis over parallel synthesis, I estimated the synthesis time for each biomass component based on ablation, then compared these times predicted by the unmodified model. To show that sequential synthesis is an adaptation for a finite proteome pool, I varied the size of the enzyme-available proteome pool to observe its effect on the cell's preferred scheduling strategy. Finally, to show how nutrient conditions affect the cell's scheduling strategy, I modelled changes in carbon and nitrogen source concentrations, then observed how the changes affect allocation of the proteome to metabolic enzymes.

FBA was chosen as a proof-of-concept method to simulate sequential scheduling of the synthesis of biomass components for its ease of use, and may serve as a basis for more fine-grained modelling to evaluate the hypotheses in this chapter in future investigations. As a limitation, FBA only examines the steady state of metabolism in a biological system. However, FBA is useful when combined with ablating the biomass reaction, because ablation introduces a series of different steady states — as each biomass component is prioritised — that can occur in sequence. This thus reflects, to some extent, how the metabolic state of the cell changes over time.

An additional limitation is that FBA simulates an average behaviour of yeast cells. To simulate a population of cells, extensions to FBA can be applied. For example, Tourigny et al. (2021) developed a stochastic simulation algorithm with FBA embedded, which models fluctuations in enzyme and metabolite copy numbers, for *Mycobacterium genitalium*. Applying this algorithm to ecYeast8 must account for the greater genome complexity and variability of lipid composition (Sánchez et al., 2019) in budding yeast. In addition, Labhsetwar et al. (2017) developed a population FBA approach, which was based on sampling enzyme copy numbers from an experimental distribution. Applying this idea to ecYeast8 may mean a stochastic size of the enzyme-available proteome pool. Both approaches are challenging and are beyond the scope of this thesis, but may be used in extensions to this work.

5.2 The Yeast8 and ecYeast8 models and their formalisms

To impose proteome constraints, I used the enzyme-constrained Yeast8 (ecYeast8) model (H. Lu et al., 2019) and performed FBA. I used this model because it is recent, offers a good coverage of reactions, and is continuously updated in a well-characterised and well-documented software repository. Here, I used the model ecYeast8.6.0, the latest version for which both original and enzyme-constrained variants are available. ecYeast8 uses the GECKO formalism (Sánchez et al., 2017) — specifically, GECKO 2 (Domenzain et al., 2022), the latest published version. GECKO applies an enzyme constraint by modifying the stoichiometric matrix of

a genome-scale metabolic model. In addition, the model has 'pseudometabolites' defined by reactions that group specific chemical species in general classes. These formalisms allow studying each class of biomass component (e.g. lipid, protein, carbohydrate) individually (Methods Section 2.4.1).

5.2.1 GECKO

In a conventional genome-scale model, metabolic fluxes through reactions are constrained by lower and upper bounds. This constraint narrows the solution space when the objective function is optimised. GECKO imposes an additional constraint on the metabolic fluxes based on the concentration of the enzyme that catalyses the reaction (Fig. 5.1).

Figure 5.1: Modifications to the stoichiometric matrix of a genome-scale model that GECKO imposes. The original stoichiometric matrix **S** includes metabolites $M_1 \dots M_m$ and reactions $v_1 \dots v_n$. GECKO extends this matrix to include enzymes $E_1 \dots E_p$ and enzyme usage reactions $e_1 \dots e_p$. The new stoichiometric matrix can be seen as four submatrices concatenated together: the upper left submatrix is the original **S**, the upper right submatrix is **0**, the lower left submatrix encodes kinetic information, and the lower right submatrix is **I**. Figure adapted from Sánchez et al. (2017).

GECKO modifies the linear programming of FBA as defined by Eq. 1.2 and 1.3 so that enzymes are expressed as metabolites that take part in reactions. This can be expressed as:

$$\max \mathbf{c}^{\intercal}\mathbf{v} \tag{5.1}$$

subject to

$$\begin{gathered} \mathbf{Sv} = \mathbf{0} \\ v_{j,\min} \leq v_j \leq v_{j,\max} \\ v_j \leq k_{\text{cat}}^{ij} \cdot [E_i] \end{gathered} \tag{5.2}$$

where v_j is the flux of each reaction R_j catalysed by enzyme E_i, k_{cat}^{ij} is the catalytic constant of the enzyme E_i for reaction R_j, and $[E_i]$ represents the concentration of the enzyme E_i. These constraints ensure that each v_j does not exceed the $v_{\max}$ of E_i. Each enzyme E_i may catalyse one or more reactions R_j. Appendices B.1–B.2 illustrate these modifications by considering how they modify chemical reactions in the model.

5.3 Ablating pseudometabolites from the biomass reaction

5.3.1 Definition of ablation

To simulate producing each class of biomass component in turn, I took advantage of pseudometabolites in ecYeast8 to remove them in turn from the biomass reaction — in other words, ablating pseudometabolites from the biomass reaction.

Consider the objective function, the biomass reaction:

```
 47.5883 atp_c + 47.5883 h2o_c + lipid_c + protein_c + carbohydrate_c
+ dna_c + rna_c + cofactor + ion
-> 47.5883 adp_c + biomass_c + 47.5883 h_c + 47.5883 pi_c
```

There are seven pseudometabolites: lipid, protein, carbohydrate, DNA, RNA, cofactor, and ion.

To simulate the cell prioritising biosynthesis of lipids, I set the stoichiometric coefficients of all pseudometabolites except for lipids to zero in the above equation, giving:

```
 47.5883 atp_c + 47.5883 h2o_c + lipid_c
-> 47.5883 adp_c + biomass_c + 47.5883 h_c + 47.5883 pi_c
```

Using this modified reaction as the objective function, the model was optimised using FBA, and this process was repeated for the other pseudometabolites, resulting in different growth rates for each round of ablation (Fig. 5.2).

Figure 5.2: Growth rates from the original model (leftmost bar) and from the ablated versions of the model (other bars)

5.3.2 Effect of ablation on allocation of proteome to enzymes

To assess whether ablation leads to metabolic changes that reflect cellular biochemistry, I quantified how proteome allocation to enzymes changes across rounds of ablation. When the model prioritised a different biomass component in each round of ablation, the model partitioned the limited proteome available for enzyme production differently. In other words, in each round of ablation, the vector of fluxes carried by each enzyme usage pseudoreaction (defined in Eq. B.5) was different from each other and from the non-ablated case, in which all biomass components were synthesised in parallel.

To quantify changes in proteome allocation across rounds of ablation, I computed the base-2 logarithm of fold changes of fluxes relative to the non-ablated, parallel case. This computation is defined as:

$$\log_2(\mathrm{FC}_{i,j}) = \log_2\left(\frac{e'_{i,j}}{e'_{i,\mathrm{par}}}\right) \tag{5.3}$$

where, to ensure that $\log_2(\mathrm{FC}_{i,j})$ can be defined for all i and j,

$$e' = \begin{cases} \epsilon, & \text{if } |e| < \epsilon \\ e, & \text{otherwise} \end{cases} \tag{5.4}$$

where e' is either $e'_{i,j}$ or $e'_{i,\mathrm{par}}$, e is the corresponding $e_{i,j}$ or $e_{i,\mathrm{par}}$, and ϵ is the minimum flux of $1\,\mathrm{molecule\,cell}^{-1} = 1.11 \times 10^{-10}\,\mathrm{mmol\ g}_{\mathrm{DW}}^{-1}$, computed assuming a cell dry weight of 15 pg dry weight per cell (Sherman, 2002), and with units consistent with enzyme usage reactions as described by Sánchez et al. (2017).

Here, $e_{i,\text{par}}$ represents the flux of an enzyme usage reaction associated with each enzyme E_i in the model (defined in Eq. B.3) in the non-ablated, parallel case, and $e_{i,j}$ represents the flux of the enzyme usage reaction when biomass component j is prioritised in a round of ablation.

If $\log_2(\text{FC}_{i,j}) > 0$, the cell allocates more of its proteome to produce enzyme E_i when biomass component j is prioritised; the reverse is true if $\log_2(\text{FC}_{i,j}) < 0$. In addition, if enzyme expression switches on ($e_{i,\text{par}} = 0$), $\log_2(\text{FC}_{i,j}) \ll 0$. Conversely, if enzyme expression switches off ($e_{i,j} = 0$), $\log_2(\text{FC}_{i,j}) \gg 0$.

To show that the re-allocation of the proteome in each round of ablation reflects metabolic pathways that are relevant for prioritising the synthesis of the respective biomass component, Fig. 5.3 categorises fold changes of enzyme-catalysed reactions by subsystem.

Specifically, the figure shows:

1. When any biomass component was prioritised, the cell de-allocated its proteome to most of its enzymes, as $\log_2(\text{FC}_{i,j}) < 0$ for most values of i, j. However, there were cases with strong increases in allocation ($\log_2(\text{FC}_{i,j}) > 0$). For example, when cofactors were prioritised, there were strong increases in biosynthesis of unsaturated fatty acids and in riboflavin metabolism. Additionally, there were strong increases in oxidative phosphorylation when carbohydrate, DNA, RNA, cofactor, or ion was prioritised.
2. When I modelled a cell that prioritises lipid biosynthesis, the model showed the least change relative to the parallel (non-ablated case), compared to other biomass components. The changes included increases of fluxes in the subsystems of fatty acid biosynthesis, glycerolipid metabolism, glycerophospholipid metabolism, inositol phosphate metabolism, steroid biosynthesis, and terpenoid backbone. Given that these subsystems are directly related

Figure 5.3: $\log_2(\text{FC})$ of enzyme usage reaction flux (Eq. B.5) in rounds of ablation, according to biomass component prioritised. Each column shows the component that remains in each round of pseudometabolite ablation (labels on top). Each row represents an enzyme, and rows are grouped by subsystem (labels on left). Colours represent $\log_2(\text{FC})$ (defined in Eq. 5.3) showing how enzyme usage fluxes change in rounds of ablation: green shows an increase, while pink shows a decrease. Rows in which $|\log_2(\text{FC})| < 11$ for all biomass components are not shown to restrict the number of reactions ($n = 3897$) for visualisation.

to lipid metabolism, such changes were expected. In addition, during lipid biosynthesis, the model showed decreases of fluxes in oxidative phosphorylation, the TCA cycle, and in amino acid metabolism, with fluxes varying depending on the amino acid.

3. When I modelled a cell that prioritises protein biosynthesis, the model showed small increases in fluxes associated with amino acid metabolism, tRNA metabolism, and oxidative phosphorylation. Such increases were expected as these led to production of substrates that are required for translation. Conversely, when other biomass components (carbohydrate, ion, DNA, RNA, and cofactor) were prioritised, there were decreases in fluxes in glycine, serine, and threonine metabolism.
4. When I modelled a cell that prioritises carbohydrate biosynthesis, the model showed increases in fructose metabolism, mannose metabolism, N-glycan biosynthesis, and starch and sucrose metabolism. This was expected given that these reactions relate directly to pathways for synthesis of carbohydrates.
5. When I modelled a cell that prioritises RNA biosynthesis, the model showed increases in the purine metabolism and pyrimidine metabolism subsystems, along with a mixed picture in the pentose phosphate pathway. The increases were expected given the presence of purines and pyrimidines in RNA and the role of the pentose phosphate pathway in generating the precursors for these compounds. However, these subsystems showed weak decreases in flux when DNA is prioritised, though the enzyme usage flux profile overall is similar to when RNA is prioritised.

Because the changes of fluxes reflect cellular biochemistry, results thus validate ablation of biomass components as a method to simulate sequential synthesis of biomass components by the cell.

5.4 Estimating timescale of biosynthesis

To evaluate whether sequential or parallel synthesis of biomass components offers a time advantage during cell growth, I estimated the timescale of biosynthesis for either resource allocation strategy (Fig. 5.4).

Figure 5.4: Estimated synthesis times of biomass components, from ablation (blue bars) and from scaling doubling time by mass fraction (red bars). Bars under 'all biomass' indicate (blue) T_{seq} (defined in Eq. 5.9) and (red) the doubling time t_0 (defined in Eq. 5.5). Other bars indicate (blue) $t_{\mathrm{seq},i}$ (defined in Eq. 5.6) and (red) $t_{\mathrm{par},i}$ (defined in Eq. 5.7); T_{par} is the greatest $t_{\mathrm{par},i}$.

Based on the objective function of the unmodified model, the doubling time was computed as follows:

$$t_0 = \frac{\ln 2}{\lambda_0} \tag{5.5}$$

where t_0 is the doubling time and λ_0 is the growth rate, equivalent to the optimised flux of the biomass reaction.

Based on the growth rates computed in rounds of ablation (Fig. 5.2), the synthesis time of each biomass component was computed as follows:

$$t_{\mathrm{seq},i} = f_i \cdot \frac{\ln 2}{\lambda_{\mathrm{seq},i}} \tag{5.6}$$

where i represents each of the biomass components (lipids, proteins, carbohydrates, DNA, RNA, cofactors, and ions), $t_{\mathrm{seq},i}$ is the predicted time for synthesis of each biomass component, f_i is the mass fraction of each biomass component (Table B.3), and $\lambda_{\mathrm{seq},i}$ is the optimal flux of the ablated biomass reaction. In Eq. 5.6, f_i acts as a scaling factor so that the synthesis time of a biomass component is proportional to the proportion of the biomass component in dry cell mass.

For comparison, I computed estimates of the time for each biomass component, assuming that it is proportional to the mass fraction:

$$t_{\mathrm{par},i} = f_i \cdot t_0 \tag{5.7}$$

where t_0 is the doubling time found in Eq. 5.5.

Eq. 5.7 assumes that different biomass components require the same synthesis time per unit of mass. Challenging this assumption is the fact that protein synthesis is a metabolically expensive process, with translation requiring many precursors and a series of catalytic processes, compared to the polymerisation of other macromolecules. Such complex translation machinery creates multiple rate-limiting steps for protein synthesis (Shah et al., 2013), which also vary depending on the gene each protein is expressed because of regulatory elements (Lahtvee et al., 2017) and because of amino acid composition (Y. Chen & Nielsen, 2022). So, per unit of mass, protein synthesis may require more time than the synthesis of other biomass components.

To evaluate the assumption that synthesis time is proportional to mass fraction, the inferred synthesis times can be compared to findings from previous studies. However, model estimates give a large range of protein synthesis rates — 1900 to 19 500 proteins per cell per second (Seeger et al., 2023) — thus making such a comparison difficult. In addition, in simulating parallel biosynthesis, it is not possible to simulate different synthesis rates for different biomass components as only one objective function is optimised. In the absence of precise time estimates for the synthesis of biomass components, the assumption that synthesis time is proportional to mass fraction may be the best assumption for the parallel case, which is one of two hypothetical scenarios, alongside the sequential case.

To determine whether sequential biosynthesis of biomass components or parallel biosynthesis of biomass components is advantageous, I defined a ratio $\tau_{\text{seq/par}}$ that represents the ratio between the total time predicted by ablation and the biomass component that is predicted to take the most time:

$$\tau_{\text{seq/par}} = \frac{T_{\text{seq}}}{T_{\text{par}}} \tag{5.8}$$

where T_{seq} represents the predicted growth time assuming sequential biosynthesis and T_{par} represents the limiting biomass synthesis time assuming parallel biosynthesis.

The quantities T_{seq} and T_{par} are defined:

$$T_{\text{seq}} = \sum_i t_{\text{seq},i} \tag{5.9}$$

$$T_{\text{par}} = \operatorname*{argmax}_i t_{\text{par},i} \tag{5.10}$$

Therefore,

$$
\begin{aligned}
\tau_{\text{seq/par}} &= \frac{T_{\text{seq}}}{T_{\text{par}}} \\
&= \frac{\left(\sum_i t_{\text{seq},i}\right)}{\left(\text{argmax}_i\, t_{\text{par},i}\right)} \\
&= \frac{\left(\sum_i f_i \cdot \frac{\ln 2}{\lambda_{\text{seq},i}}\right)}{\left(f_{\text{protein}} \cdot \frac{\ln 2}{\lambda_0}\right)} \\
&= \left(\frac{f_{\text{lipid}}}{\lambda_{\text{seq,lipid}}} + \frac{f_{\text{protein}}}{\lambda_{\text{seq,protein}}} + \cdots + \frac{f_{\text{ion}}}{\lambda_{\text{seq,ion}}}\right) \cdot \frac{\lambda_0}{f_{\text{protein}}}
\end{aligned}
\tag{5.11}
$$

The expression in Eq. 5.11 means that the definition of the $\tau_{\text{seq/par}}$ ratio does not reduce to a trivial expression and depends on the λ_0 and the $\lambda_{\text{seq},i}$ values, which are independent of each other. A $\tau_{\text{seq/par}} < 1$ means that synthesising biomass components in sequence saves more time, and sequential biosynthesis is favoured. Conversely, $\tau_{\text{seq/par}} > 1$ indicates that parallel synthesis of biomass components is favoured as synthesising biomass components in sequence does not save time.

To confirm that the advantage of the sequential biosynthesis strategy is retained in other genetic backgrounds, I extended the computation $\tau_{\text{seq/par}}$ and related quantities to models in which genes were deleted. To corroborate results from Chapter 3, the BY4741 *zwf1*Δ and BY4742 *tsa2*Δ strains were simulated — the latter was substituted for *tsa1*Δ *tsa2*Δ because the ecYeast8 model does not include reactions that correspond to *TSA1*. The deletions were made by restricting to zero the reaction fluxes that are associated with the deleted genes in the model. For BY4741-background strains, supplements were simulated by allowing uptake of histidine, leucine, tryptophan, methionine and uracil. The same applied to BY4742-background strains, but lysine uptake replaced methionine uptake.

Fig. 5.5 shows that $\tau_{\text{seq/par}} < 1$ still held for auxotrophs and deletion strains, suggesting that the sequential biosynthesis strategy remained advantageous. Assuming that temporal scheduling of the synthesis of biomass components during growth explains the timing of biosynthetic events in the yeast metabolic cycle, these observations supports results in Chapter 3 which shows that auxotrophs and deletion strains have YMCs.

Figure 5.5: Estimated synthesis times of biomass components: from ablation and T_{par}, for wild-type (unmodified model), BY4741, *zwf*Δ in the BY4741 background, and *tsa2*Δ in the BY4742 background

5.5 Effect of restricting the enzyme pool

The yeast cell has a finite enzyme-available proteome pool, so it must decide which enzymes to allocate the greatest proportions of the pool to. To study this effect using the ecYeast8 model, I imposed a constraint on the enzyme-available proteome pool by varying the value of the upper limit of the flux e_{pool} of the enzyme pool pseudoreaction (Eq. B.3), taking advantage of a GECKO formalism

that is easy to modify and interpret. With a smaller e_{pool}, the sum of fluxes of enzyme usage pseudoreactions must decrease, and the model must decide which enzyme usage pseudoreactions to allocate a higher flux to, modelling the biological response to a restricted enzyme pool.

Fig. 5.6 shows that constraining the proteome pool available for enzymes leads to a greater advantage of sequential biosynthesis of biomass components over parallel biosynthesis. Within the range of e'_{pool} that gives realistic growth rates, this is evidenced by a decreasing $\tau_{\text{seq/par}}$ ratio if e'_{pool} decreases (Fig. 5.6a). Concurrently, as e'_{pool} decreases, the wild type growth rate (λ_0) decreases linearly to zero and ablated growth rates ($\lambda_{\text{seq},i}$) decrease in linear segments independently of each other and of the growth rate. These observations can be explained by considering Eq. 5.11 and modelling the changes of λ_0 and $\lambda_{\text{seq},i}$ with respect to $e'_{\text{pool}}/e_{\text{pool}}$ as linear equations (Appendix B.4).

5.6 Effect of carbon and nitrogen sources

To explore whether sequential synthesis of biomass components remains advantageous across nutrient conditions, I investigated how changes in the concentrations of nitrogen (ammonium) and carbon sources (glucose and pyruvate) affected the resource allocation strategies. Ammonium is the form of nitrogen in minimal growth media. Glucose is the preferred carbon source for budding yeast, and as pyruvate is a non-fermentable carbon source, using it as a carbon source would show how substantial changes to central carbon metabolism affected resource allocation strategies.

Figure 5.6: Effect of the size of the proteome pool available for enzymes (e'_{pool}) on **(5.6a, 5.6b)** $\tau_{\text{seq/par}}$, the growth rate, and **(5.6c, 5.6d)** the optimal flux of the ablated biomass reaction $\lambda_{\text{seq},i}$ in each round of ablation. **(5.6a)** and **(5.6c)** show $e'_{\text{pool}}/e_{\text{pool}} \leq 2$, while **(5.6b)** and **(5.6d)** show $e'_{\text{pool}}/e_{\text{pool}} \leq 20$.

5.6.1 Saturation of exchange reactions

Investigating the effect of carbon and nitrogen sources requires finding ranges of concentrations of each source to be used in an FBA problem that leads to biologically informative results. The saturation point of the nutrient, defined as the concentration of the nutrient at which the growth rate reaches its maximum, thus serves as a biologically relevant reference point. Although genome-scale metabolic models have nutrient exchange reactions that simulate the presence or absence of nutrients, saturation points from experimental studies cannot be directly used as flux bounds for these reactions because FBA does not account

for substrate concentrations. Instead, I constrained the flux of exchange reactions so that the optimised growth rate matched experimental observations, as was performed in previous FBA-based studies (Elsemman et al., 2022; Famili et al., 2003).

To model the effect of nutrient concentrations on growth rate, I created saturation curves that show how the effect of nutrient exchange flux on the objective function. Figs. 5.7a and 5.7b show that the saturation curves of glucose and pyruvate had different shapes. In addition, the maximum growth rate on glucose ($0.38\,\mathrm{h}^{-1}$) was greater than the maximum growth rate on pyruvate ($0.25\,\mathrm{h}^{-1}$). Furthermore, Fig. 5.7c shows that the maximum growth rate on the carbon source sets the saturation point for ammonium.

These growth saturation curves agree with similar studies. Specifically, the growth saturation curve for glucose (Fig. 5.7a) is similar to that simulated by Elsemman et al. (2022) using another derivative of the Yeast8 model. The maximum growth rate on glucose agrees with Domenzain et al. (2022), which used GECKO 2 to create the ecYeast7 model to predict maximum growth rates on various carbon sources. Domenzain et al. (2022) did not simulate growth on pyruvate, but a lower maximum growth rate on pyruvate is consistent with my experimental observations.

(a) **(b)**

(c)

Figure 5.7: Effect of glucose **(5.7a)**, pyruvate **(5.7b)**, and ammonium **(5.7c)** exchange reactions on growth rate. The saturation point of exchange reactions is defined as the flux of an exchange reaction above which the objective function reaches its maximum.

5.6.2 Effect of carbon and nitrogen sources on biomass synthesis strategies

Glucose and ammonium

To assess the effect of glucose and ammonium concentration on biomass synthesis strategies, I ablated components in the biomass reaction in different nutrient conditions set by glucose and ammonium exchange fluxes to obtain $\tau_{\text{seq/par}}$, λ_0, $t_{\text{seq,carbohydrate}}$, and $t_{\text{seq,protein}}$. Additionally, to determine whether the carbon or nitrogen source is limiting for each of these quantities, I defined the sensitivity at each nutrient condition $(R_{\text{glc}}, R_{\text{amm}})$, with respect to each axis i as:

$$s_i(R_{\text{glc}}, R_{\text{amm}}) = \frac{R_i}{y(R_{\text{glc}}, R_{\text{amm}})} \cdot \frac{\partial y(R_{\text{glc}}, R_{\text{amm}})}{\partial R_i} \tag{5.12}$$

where i indicates glucose (glc) or ammonium (amm), R_i indicates the exchange rate, glucose (R_{glc}) or ammonium (R_{amm}), an $(R_{\text{glc}}, R_{\text{amm}})$ pair defines a nutrient condition, and $y(R_{\text{glc}}, R_{\text{amm}})$ represents the quantity of interest at each nutrient condition. For a specific nutrient condition $(R_{\text{glc}}, R_{\text{amm}})$, if $s_{\text{glc}}(R_{\text{glc}}, R_{\text{amm}}) > s_{\text{amm}}(R_{\text{glc}}, R_{\text{amm}})$ for a quantity of interest y, then the quantity is more sensitive to glucose exchange in this condition. Conversely, if $s_{\text{glc}}(R_{\text{glc}}, R_{\text{amm}}) < s_{\text{amm}}(R_{\text{glc}}, R_{\text{amm}})$, the quantity is more sensitive to ammonium exchange.

Fig. 5.8b divides the nutrient conditions into three regions according to nutrient limitation: a glucose-limiting region, an ammonium-limiting region, and a region where neither glucose nor ammonium is limiting. These regions are determined by $(s_{\text{glc}}, s_{\text{amm}})$ vectors. Fig. 5.8a shows that parallel biosynthesis was advantageous $(\tau_{\text{seq/par}} > 1)$ in two circumstances: when both glucose and ammonium limited the growth rate, and when ammonium exchange was at saturation while glucose

exchange is greater than saturation. In contrast, when the growth rate was near its maximum, where neither glucose nor ammonium was limiting, when glucose or ammonium exchange increased, sequential biosynthesis became more advantageous ($\tau_{\text{seq/par}} < 1$).

To link the changes in the values of $t_{\text{seq,carbohydrate}}$ and $t_{\text{seq,protein}}$ with cellular biochemistry, Fig. 5.8c and 5.8d show how both values are affected by R_{glc} and R_{amm}. Fig. 5.8c shows that glucose exchange limited carbohydrate synthesis time in all nutrient conditions, and this observation could be explained by glucose uptake being directly linked to carbohydrate synthesis. In contrast, Fig. 5.8d shows that ammonium exchange limited protein synthesis time for most nutrient conditions, except for conditions in which glucose exchange was very low. This observation can be explained by the fact that both carbon and nitrogen sources contribute to protein biosynthesis.

To determine whether changes in the synthesis time of biomass components are responsible for switching of scheduling strategies between sequential and parallel biosynthesis, I investigated how R_{glc} and R_{amm} affected $\frac{t_{\text{seq,carbohydrate}}}{t_{\text{seq,protein}}}$. Synthesis times $t_{\text{seq},i}$ were computed to assess whether the ratio between the synthesis times differed in different conditions. The ratio $\frac{t_{\text{seq,carbohydrate}}}{t_{\text{seq,protein}}}$ serves as the principal measure to quantify how $t_{\text{seq},i}$ values change as nutrient conditions change. This ratio is informative because:

1. Of all biomass components, predicted synthesis times of these biomass components varied the most as the glucose and ammonium exchange rates were varied.
2. Each accounts for a large proportion of biomass: protein accounts for 52.5% and carbohydrate accounts for 36.4% (Table B.3).

Figure 5.8: Effect of glucose and ammonium exchange rates on **(5.8a)** $\tau_{\text{seq/par}}$, **(5.8b)** growth rate based on unmodified biomass reaction, λ_0, **(5.8c)** predicted carbohydrate synthesis time, $t_{\text{seq,carbohydrate}}$ **(5.8d)** predicted protein synthesis time, $t_{\text{seq,protein}}$, and **(5.8e)** ratio of carbohydrate synthesis time to protein synthesis time, $t_{\text{seq,carbohydrate}}/t_{\text{seq,protein}}$. Exchange rates are expressed in percentages of growth saturation from Fig. 5.7, with black straight lines indicate 100% of saturation. Contours show regions in which $\tau_{\text{seq/par}} > 1$. Arrows overlaid on heatmaps indicate sensitivity of the quantity as a vector $(s_{\text{glc}}, s_{\text{amm}})$ (Eq. 5.12).

3. Carbohydrate synthesis has a clear biochemical relationship with the level of a carbon source. Additionally, protein synthesis has a clear biochemical relationship with the level of ammonium as amino acids contain amino groups.

Fig. 5.8e suggests that carbon and nitrogen limitation affect the ratios of synthesis times among biomass components, as shown by the change in the value of $\frac{t_{\text{seq,carbohydrate}}}{t_{\text{seq,protein}}}$ as R_{glc} and R_{amm} changed. Specifically, within the glucose-limiting region, $\frac{t_{\text{seq,carbohydrate}}}{t_{\text{seq,protein}}}$ was at a high and constant value. In contrast, within the ammonium-limiting region, as R_{amm} increases, $\frac{t_{\text{seq,carbohydrate}}}{t_{\text{seq,protein}}}$ increased. In addition, along the boundary of both regions, where glucose and ammonium were both near-limiting and $\tau_{\text{seq/par}}$ was high, $\frac{t_{\text{seq,carbohydrate}}}{t_{\text{seq,protein}}}$ remained roughly constant. These observations link synthesis times of biomass components with scheduling strategy. Specifically, the observations suggest that the ratios of synthesis times alone did not explain the choice of scheduling strategy, and that total biosynthesis time was an additional factor.

Pyruvate and ammonium

To investigate the effect of a non-fermentable carbon source on biomass synthesis strategies, I repeated the investigation using pyruvate as the carbon source.

Results show that the saturation curve of pyruvate controlled the effect of exchange rates on scheduling strategies. Fig. 5.9 shows that when pyruvate was the carbon source, the region of carbon source limitation was larger than when glucose was the carbon source. This observation reflected the different shapes of the pyruvate and glucose saturation curves (Fig. 5.7): below growth rate saturation, the relationship between pyruvate exchange rate and growth rate is linear, while the relationship between glucose exchange rate and growth rate is not linear. In other words, on pyruvate, the growth rate reached half of its maximum value

when the pyruvate exchange rate was half of the value that corresponds to the maximum growth rate. In contrast, on glucose, when the growth rate reached half of its maximum value, the glucose exchange rate was well below half of the value that corresponds to the maximum growth rate.

Fig. 5.9 shows that the relationships between exchange rates and the quantities $\tau_{\text{seq/par}}$, λ_0, $t_{\text{seq,carbohydrate}}$, and $t_{\text{seq,protein}}$ were similar to the glucose-ammonium case. In other words, when pyruvate was the carbon source, the relationship between carbon and nitrogen source uptake and the synthesis of carbohydrates and proteins still determined how scheduling strategies changed according to nutrient conditions. Parallel biosynthesis was advantageous in two conditions: when both pyruvate and ammonium limited the growth rate, and when ammonium exchange was at saturation while pyruvate exchange was greater than saturation (Fig. 5.9a). In contrast, sequential biosynthesis was advantageous when neither pyruvate nor ammonium was limiting. $\frac{t_{\text{seq,carbohydrate}}}{t_{\text{seq,protein}}}$ remained at high values when pyruvate is limiting or when pyruvate and ammonium were both limiting (Fig. 5.9e). Furthermore, the behaviour of $\frac{t_{\text{seq,carbohydrate}}}{t_{\text{seq,protein}}}$ as nutrient conditions varied could also be explained by pyruvate exchange limiting carbohydrate synthesis time (Fig. 5.9c) and by both pyruvate and ammonium exchange limiting protein synthesis time (Fig. 5.9d).

Figure 5.9: Effect of pyruvate and ammonium exchange rates on **(5.9a)** $\tau_{\text{seq/par}}$, **(5.9b)** growth rate based on unmodified biomass reaction, λ_0, **(5.9c)** predicted carbohydrate synthesis time, $t_{\text{seq,carbohydrate}}$ **(5.9d)** predicted protein synthesis time, $t_{\text{seq,protein}}$, and **(5.9e)** ratio of carbohydrate synthesis time to protein synthesis time, $t_{\text{seq,carbohydrate}}/t_{\text{seq,protein}}$. Exchange rates are expressed in percentages of growth saturation from Fig. 5.7, with black straight lines indicate 100% of saturation. Contours show regions in which $\tau_{\text{seq/par}} > 1$. Arrows overlaid on heatmaps indicate sensitivity of the quantity as a vector $(s_{\text{pyr}}, s_{\text{amm}})$ (Eq. 5.12).

5.6.3 Relationship between proteome allocation and resource allocation strategy

Glucose and ammonium

The investigation of the effect of carbon and nitrogen sources on scheduling strategies led to a new hypothesis: the cell favours parallel biosynthesis of biomass components when the carbon source and the nitrogen source are both limiting, especially if the associated biomass components share metabolic pathways and if the conditions dictate similar levels of enzymes. This hypothesis arose from the observation that the highest $\tau_{\text{seq/par}}$ occurred at the boundary at which both the carbon source and the nitrogen source were limiting.

To evaluate this hypothesis, I compared the enzyme usage reaction flux vectors in rounds of ablation between a high $\tau_{\text{seq/par}}$ condition and a low $\tau_{\text{seq/par}}$ condition created by combinations of glucose and ammonium exchange rates. Enzyme usage reaction flux vectors show how the cell allocates its finite proteome to enzymes when each biomass component is prioritised (Section 5.4).

To test whether similarities in proteome allocation to enzymes explain parallel biosynthesis, Fig. 5.10a shows that when both glucose and ammonium were limiting, the allocation patterns were similar across rounds of ablation, thus explaining the advantage of parallel biosynthesis of biomass components. This similarity is quantified in Fig. 5.10b, which shows that parallel biosynthesis exhibited relatively similar proteome allocation with the synthesis of individual biomass components, as evidenced by relatively high pairwise Kendall's τ-b rank correlation coefficients (mean = 0.38).

Figure 5.10: For the high $\tau_{\text{seq/par}}$ condition ($R_{\text{glc}} = 1.69\,\text{mmol g}_{\text{DW}}^{-1}\ \text{h}^{-1}$, $R_{\text{amm}} = 1.05\,\text{mmol g}_{\text{DW}}^{-1}\ \text{h}^{-1}$): **(5.10a)** changes in enzyme usage reaction flux in rounds of ablation. Columns show the biomass component prioritised. In each column, rows represent enzyme usage reactions, arranged in descending order of flux. Colours identify the reactions, with white indicating reactions that carry zero flux in the parallel case. **(5.10b)** Pairwise Kendall's τ-b rank correlation coefficient (Kendall, 1945) for each pair of enzyme usage flux profiles.

In contrast, Fig. 5.11a shows that when neither glucose nor ammonium limited growth, biomass components could be divided into two groups based on enzyme usage reaction flux vectors. These groups suggest an explanation for sequential biosynthesis: the cell may synthesise lipids and proteins together in one stage of its growth cycle, while synthesising carbohydrates, DNA, RNA, cofactors, and ions together in another stage. The bias towards sequential biosynthesis is further emphasised by Fig. 5.11b, which shows that parallel biosynthesis exhibited different proteome allocation compared to the synthesis of individual biomass components, as evidenced by relatively low pairwise Kendall's τ-b rank correlation coefficients compared to Fig. 5.10b (mean = 0.28).

Figure 5.11: For the low $\tau_{\text{seq/par}}$ condition ($R_{\text{glc}} = 16.89\,\text{mmol}\ \text{g}_{\text{DW}}^{-1}\ \text{h}^{-1}$): **(5.11a)** changes in enzyme usage reaction flux in rounds of ablation. Columns show the biomass component prioritised. In each column, rows represent enzyme usage reactions, arranged in descending order of flux. Colours identify the reactions, with white indicating reactions that carry zero flux in the parallel case. **(5.11b)** Pairwise Kendall's τ-b rank correlation coefficient (Kendall, 1945) for each pair of enzyme usage flux profiles.

Pyruvate and ammonium

To evaluate, in a different carbon source, whether parallel biosynthesis is also favoured when biomass components share metabolic pathways, I extended the comparison of enzyme usage reaction flux vectors in rounds of ablation to nutrient conditions created by pyruvate and ammonium exchange rates.

Similar to the glucose-ammonium investigation, in the high $\tau_{\text{seq/par}}$ condition where pyruvate and ammonium were limiting, Figs. 5.12a and 5.12b show that proteome allocation were similar across rounds of ablation, supporting parallel biosynthesis as a scheduling strategy. However, the results for the low $\tau_{\text{seq/par}}$ case suggest that grouping of biomass components explains sequential biosynthesis for only very low $\tau_{\text{seq/par}}$.

Fig. 5.13a shows that when pyruvate was the carbon source, the contrast of proteome allocation between biomass components in the low $\tau_{\text{seq/par}}$ condition was lessened compared to when glucose was the carbon source. This may be explained by the higher $\tau_{\text{seq/par}}$ in this condition ($\tau_{\text{seq/par}} = 0.9$), compared to the $\tau_{\text{seq/par}}$ in the glucose-ammonium investigation ($\tau_{\text{seq/par}} = 0.7$).

The use of enzyme usage vectors to explain biomass synthesis scheduling strategies is called into question by the multiplicity of solutions in FBA. Fig. 5.14 shows that the relationship between exchange rates and measures of similarity between enzyme usage fluxes when carbohydrate was prioritised and when protein was prioritised was not continuous.

Figure 5.12: For the high $\tau_{\text{seq/par}}$ condition ($R_{\text{pyr}} = 3.73\,\text{mmol g}_{\text{DW}}^{-1}\ \text{h}^{-1}$, $R_{\text{amm}} = 0.90\,\text{mmol g}_{\text{DW}}^{-1}\ \text{h}^{-1}$): **(5.12a)** changes in enzyme usage reaction flux in rounds of ablation. Columns show the biomass component prioritised. In each column, rows represent enzyme usage reactions, arranged in descending order of flux. Colours identify the reactions, with white indicating reactions that carry zero flux in the parallel case. **(5.12b)** Pairwise Kendall's τ-b rank correlation coefficient (Kendall, 1945) for each pair of enzyme usage flux profiles.

(a) **(b)**

Figure 5.13: For the low $\tau_{\text{seq/par}}$ condition ($R_{\text{pyr}} = 8.89\,\text{mmol g}_{\text{DW}}^{-1}\,\text{h}^{-1}$): **(5.13a)** changes in enzyme usage reaction flux in rounds of ablation. Columns show the biomass component prioritised. In each column, rows represent enzyme usage reactions, arranged in descending order of flux. Colours identify the reactions, with white indicating reactions that carry zero flux in the parallel case. **(5.13b)** Pairwise Kendall's τ-b rank correlation coefficient (Kendall, 1945) for each pair of enzyme usage flux profiles.

Figure 5.14: Measures of similarity between enzyme usage fluxes in different nutrient conditions. Cosine distances between the vector of enzyme usage fluxes when protein is prioritised and the equivalent vector when carbohydrate is prioritised are shown **(5.14a)** for glucose-ammonium conditions and **(5.14b)** pyruvate-ammonium conditions; smaller distances indicate a greater similarity between the two conditions. Means of the set of Kendall's τ-b rank correlation coefficients between the parallel case and each biomass-prioritised case are shown **(5.14c)** for glucose-ammonium conditions and **(5.14d)** pyruvate-ammonium conditions; larger mean values indicate a greater similarity between parallel biosynthesis and individual biomass component synthesis.

This observation is likely explained by the multiplicity of solutions in FBA. Namely, while FBA finds the optimal value of the objective function, there is no guarantee that the flux vector $\mathbf{v}$ (Eq. 1.2) is equivalent across linear programming solvers. In sum, although choosing representative nutrient conditions led to an attractive picture that may confirm the hypothesis that the cell favours parallel biosynthesis in nutrient limiting conditions if biomass components share metabolic pathways, the computational limitations of FBA may cast doubts on conclusions.

5.7 Discussion

This chapter uses a genome-scale model of budding yeast and flux balance analysis to test whether sequential synthesis of biomass components reflects an adaptation to limited cellular resources.

My results suggest that ablation of components of the biomass reaction was a viable method to simulate sequential synthesis of biomass components, as evidenced by how ablation predicted biologically relevant changes in proteome allocation to enzymes. In addition, ablation led to a way to estimate the time of synthesis of biomass components. These times suggest that sequential scheduling of biosynthesis saves time during growth and remains advantageous across deletion strains.

My results further show that within realistic growth rates, a smaller proteome pool led to a greater advantage of sequential biosynthesis over parallel biosynthesis. However, parallel scheduling of biosynthesis becomes advantageous in some nutrient conditions, such as when both carbon and nitrogen sources are limiting. Further simulations suggest an explanation for this advantage of parallel biosynthesis: as each biomass component is synthesised, the cell allocates its proteome pool to enzymes in similar patterns.

The advantage of sequential scheduling of biosynthesis over parallel scheduling of biosynthesis may explain why yeast cells exhibit metabolic cycles, in which there is a sequence of synthesising biomass components (Mellor, 2016). Conversely, the advantage of parallel scheduling of biosynthesis over sequential scheduling of biosynthesis in nutrient-limiting conditions may explain the metabolic cycle being less reliably generated and of low quality in some conditions in my observations. For example, the model predicts an advantage of parallel scheduling of

biosynthesis when there is low (10-20% saturation) glucose exchange and when ammonium exchange lies below saturation. This may correspond to low-glucose conditions in Section 3.4, from which most flavin signals are non-oscillatory. However, such correspondence is only true when the glucose exchange rate in the model corresponds to the glucose concentration in the experiment, and experiments to calibrate the two quantities must be performed to verify this correspondence.

Furthermore, it is possible that implementing conditions not tested by this chapter, such as highly acidic conditions (J. S. O'Neill et al., 2020), in flux balance analysis simulations may predict on advantage of parallel scheduling of biosynthesis, thus explaining the absence of metabolic cycles in such conditions. The metabolic cycle may thus be an adaptation to a limited proteome pool and having to carry out the metabolically expensive process of protein synthesis during cell growth (J. S. O'Neill et al., 2020). Clustering of the synthesis of lipids and proteins in one group may explain cycling of storage lipids (Campbell et al., 2020), while clustering of carbohydrate, DNA, and RNA synthesis in another group to group together processes that require oxidative phosphorylation may explain respiratory cycles in the YMC (Tu et al., 2005).

However, the limitations of using FBA to investigate resource allocation strategies in this chapter include multiplicity of solutions and a lack of ability to preform concentration-dependent, time-dependent, and compartment-dependent simulations. The multiplicity of solutions in FBA makes it challenging to draw conclusions about how quantities derived from flux values apart from growth. These values in the GECKO model include enzyme usage reaction fluxes, which would inform proteome allocation strategies. In addition, FBA gives a steady-state picture of metabolism. To describe changes in fluxes over time, derivatives such as dynamic FBA (Mahadevan et al., 2002) could be used. Alternatively, to introduce time-dependence of resource allocation, a phenomenological description of the

metabolic cycle can be included; such an approach would be equivalent to the addition of a light-dark cycle as an external oscillator as performed by Reimers et al. (2017). To further increase the precision of the model and thus its ability to make predictions, compartmentalisation can be considered (Elsemman et al., 2022), or resource balance analysis (Goelzer et al., 2011) can be used.

To validate the predictions from flux balance analysis in this chapter with experimental results from Chapter 3, conditions that correspond to an advantage of parallel scheduling of biosynthesis over sequential scheduling of biosynthesis, and vice versa, can be applied to yeast cells in a single-cell microfluidics experiment. Performing such experiments will necessitate calibrating carbon and nitrogen source concentrations in the nutrient media with exchange rates in the ecYeast8 model. In addition, other nutrient perturbations and genetic perturbations in this thesis — namely, potassium-deficiency and the *zwf1*Δ and *tsa1*Δ *tsa2*Δ strains — can be simulated using the modelling approach in this chapter, and can lead to further cross-validation between the single-cell microfluidics and modelling approaches used in this thesis. As the precision of the modelling approach is increased by using derivations of FBA, the results from Chapter 3 can be used to calibrate such computational methods, and a more precise model can then be used to predict new nutrient perturbations or genetic perturbations that impair the metabolic cycle, forming a design-build-test-learn cycle.

The synthesis times predicted by this chapter may inform the period length to be used to simulate time series as discussed in Chapter 4. In addition, if stochastic extensions to FBA (Labhsetwar et al., 2017; Tourigny et al., 2021) are used in an extension to this study, this can lead to a population of time series with different periods or noise parameters. In such a case, methods to visualise groups in a

dataset (Section 4.3) or to estimate the period in a population of time series (Section 4.5) can be used to evaluate whether the population of simulated time series in this case approximates recorded time series from cells as presented in Chapter 3.

Chapter 6

Conclusions

Although biological rhythms are crucial for living organisms to control their physiological processes in response to external conditions, not all biological rhythms are well-characterised. In contrast to the circadian rhythm and the cell division cycle, our knowledge of the biochemical basis of the yeast metabolic cycle is incomplete. It is unclear what mechanism drives cell division-independent cycling of biosynthesis observed in the yeast metabolic cycle, and it is also unclear what mechanism is responsible for the adaptation of the yeast metabolic cycle to changing demands (Zylstra & Heinemann, 2022). Additionally, chemostat-based and single-cell experiments led to conflicting conclusions about the yeast metabolic cycle because each type of experiment creates different culture conditions and have different types of measurements.

The primary goal of this thesis was thus to develop an explanation to reconcile chemostat and single-cell studies on the yeast metabolic cycle. Specifically, I developed such explanations through testing whether specific characteristics of the yeast metabolic cycle as observed in the chemostat could be recapitulated in single-cell microfluidics. In addition, this thesis aimed to show whether proteomic constraints could explain why the yeast cell temporally segregates biosynthetic events as it progresses through the yeast metabolic cycle. This secondary goal provided a coarse-grained explanation of a model of the yeast metabolic cycle as a fundamental metabolic adaptation to physiological constraints.

6.1 Microfluidics and fluorescence microscopy for cellular metabolic cycles

In Chapter 3, I used the ALCATRAS (Crane et al., 2014) single-cell microfluidics platform to physically separate budding yeast cells and fluorescence microscopy to monitor the yeast metabolic cycle and the cell division cycle. I showed that yeast cells independently generated flavin-based single-cell metabolic cycles. In addition, a specific phase of such cycles likely gated the cell division cycle, as evidenced by decoupling between the metabolic and cell division cycles during starvation. I further showed that the metabolic cycle was retained in nutrient perturbations and in deletion strains. In particular, I showed that cells generated such cycles in potassium-deficient conditions, contrary to J. S. O'Neill et al. (2020), although at a longer period than standard, high-glucose, single-cell conditions. I also showed that *zwf1*Δ and *tsa1*Δ *tsa2*Δ cells generated flavin cycles whose waveforms differed from cycles of dissolved oxygen previously observed in the chemostat. Specifically, I observed that *zwf1*Δ generated longer flavin cycles than wild-type cells, contrary to the absence of dissolved oxygen cycles observed by Tu et al. (2007). Additionally, I observed that *tsa1*Δ *tsa2*Δ cells generated flavin cycles of two main period lengths, contrary to the M-shaped dissolved oxygen cycles observed by Causton et al. (2015).

My results suggest that the yeast metabolic cycle is likely an intrinsic cycle in budding yeast that oscillates within a range of natural frequencies, but the cell is able to adjust this frequency to respond to nutrient conditions. If conditions are permissive, the metabolic cycle provides windows of opportunities for the cell division cycle to be initiated. Otherwise, if conditions are not permissive, the metabolic cycle continues while the cell division cycle is halted at a gap phase (G_1 or G_2/M).

My results further suggest that the presence of sub-populations in the yeast culture could explain the discrepancy between single-cell and chemostat observations. The idea of sub-populations that stagger their entry into the yeast metabolic cycle has been suggested before to explain chemostat observations (Burnetti et al., 2016), and the presence of genetically identical sub-populations that respond differently to the same nutrient perturbation has later been shown (Bagamery et al., 2020). Additionally, genetically identical sub-populations that have different levels of sensitivities to an inhibitor may also explain chemostat oscillations (Smith & Waltman, 1995).

To provide more clarity to the role of nutrient storage in the yeast metabolic cycle, future work may include experiments with strains deficient in lipid synthesis. Additionally, a feast-and-famine experimental set-up which better emulates chemostat conditions could lead to a clearer explanation of previous chemostat-based studies. The glucose pulses imposed by this set-up may lead to a mathematical model of coupled oscillations that links the intrinsic yeast metabolic cycle to extrinsically-imposed oscillations.

6.2 Analysis of oscillatory time series in the yeast metabolic cycle

Because the ALCATRAS platform produces large datasets of time series, in Chapter 4, I developed a series of time series analysis methods. These methods clean data, visualise groups in a dataset, detect rhythmicity, estimate periodicity of signals, and detect synchrony between two types of signals. I showed that a high-pass filter offered good control over the frequency domain of time series. Subsequently, I showed that dimension-reduction (UMAP) and clustering (modularity clustering) methods agreed on a division between oscillatory and non-oscillatory time series in a dataset. Following this, I demonstrated that a statistical method based on the power spectrum and a support vector classifier offer modest performances in rhythmicity detection. Additionally, I showed that the autocorrelation function could be used to estimate periodicity and noise parameters from synthetic data. However, my current implementation of the autocorrelation function has limited ability in characterising noise parameters from real data. Finally, I showed that the cross-correlation function could be used to quantify the shift of one type of time series relative to another, across a population of paired time series

Rhythmicity detection is complicated by its different definitions depending on the approach — reflected in the variety of rhythmicity detection methods compared in Chapter 4. From a signal processing perspective, it can be defined as finding a strong signal within a range of expected frequencies (Zieliński et al., 2014). However, from a data science perspective, rhythmicity detection can be seen as identifying the values of a set of time series features that best discriminate between non-oscillatory and oscillatory time series, which I defined visually.

To improve the usefulness of the time series analysis methods, further refinement is needed. To make the clustering methods and the support vector classifier generalisable, we require a large enough dataset of signals that includes a variety of oscillation types and shapes, and hyperparameter tuning. Furthermore, to improve the ability of the autocorrelation function to infer noise properties of real data, a broader range of noise parameters should be simulated. Such simulations would provide additional information that leads to a more precise relationship between noise parameters and the shape of the autocorrelation function. A precise way to detect of noise parameters can then be useful to compare the noise from different environmental conditions and from imaging methods.

With the improvements in place, the methods developed in Chapter 4 can form a powerful time series analysis pipeline for oscillatory signals from any natural phenomenon. For example, the methods can detect rhythmicity in the time series presented in Chapter 3. In addition, to test whether there are sub-populations of cells that have different metabolic cycling behaviours among a genetically homogeneous culture, methods to visualise groups in a dataset of time series (Section 4.3) can be used if there are natural divisions within an experiment that suggest biologically relevant sub-populations.

6.3 Modelling yeast biosynthesis strategies under constraints

Finally, in Chapter 5, I used an enzyme-constrained genome-scale model of budding yeast and flux balance analysis to address whether a limited proteome pool leads to a preference of sequential biosynthesis over parallel biosynthesis. In this chapter, I used the novel approach of ablating components of the biomass reaction

to simulate temporal segregation of biosynthesis, and devised a time ratio that indicates whether sequential or parallel biosynthesis was more advantageous. I showed that sequential scheduling of biosynthesis was advantageous across deletion strains, and became more advantageous if the proteome pool was smaller. This prediction agrees with the presence of single-cell metabolic cycles in the *zwf1*Δ and *tsa1*Δ *tsa2*Δ strains as reported in Section 3.6.

However, I also showed that parallel scheduling of biosynthesis became advantageous when both carbon and nitrogen sources were limiting. This observation may be explained by the synthesis pathways across different biomass components sharing enzymes. This observation also agrees with experimental observations; specifically, the low-glucose conditions discussed in Section 3.4 fall in a region of parallel scheduling of biosynthesis being advantageous and thus may explain the lack of robustness of metabolic cycles in such conditions.

The advantage of sequential biosynthesis may explain why the yeast cell sequentially schedules biosynthesis of biomass components across phases of the yeast metabolic cycle, even when such sequential scheduling is not needed to coordinate events of the cell division cycle — e.g. when the metabolic cycle proceeds without cell division during starvation. Furthermore, the advantage of parallel biosynthesis in some conditions suggests that the metabolic cycle may cease to occur if nutrient conditions are too harsh, in concordance with studies that suggest disappearance of the metabolic cycle in extreme conditions (J. S. O'Neill et al., 2020). To improve model predictability, this study could be extended by using derivations of flux balance analysis that account for compartmentalisation or temporality, such as dynamic flux balance analysis. In addition, derivations that account for stochastic

variation in metabolism can be used to estimate periods and noise parameters of metabolic oscillations, which can then be used to generate synthetic time series that can in turn be characterised using time series analysis methods discussed in Chapter 4.

6.4 Summary and broader context of thesis

Put together, single-cell analysis of flavin-based yeast metabolic cycles (Chapter 3) and modelling of the metabolism of budding yeast (Chapter 5) may provide a mechanistic explanation for such an under-characterised biological rhythm. Such insights may lead to a biochemical mechanism for the autonomous generation of the yeast metabolic cycle and for its response to nutrient conditions. The biochemistry of the yeast metabolic cycle could then be modelled using techniques such as flux balance analysis. In addition, robust time series analysis methods (Chapter 4) would be able to discover classes of oscillations within a microfluidics experiment that could correspond to sub-populations in the culture. Identification of such sub-populations could then potentially reconcile results of single-cell and chemostat experiments.

Biological rhythms are an important physiological adaptation of all living organisms. This thesis, in sum, shows the robustness of the yeast metabolic cycle and relates it to resource allocation strategies, thus potentially shedding light on what could be a fundamental biological process.

The presence of the yeast metabolic cycle leads to a question: what benefits of biological oscillations justify their existence? The benefits of the circadian rhythm and the cell division cycle are clear: circadian rhythms synchronise physiological processes to the light-dark cycle, while the cell division cycle coordinates

resource-intensive processes with the presence of nutrient stores and maintains genetic fidelity. In the yeast metabolic cycle, it has been proposed that sequential scheduling of biosynthesis optimises the use of the limited proteome for cell growth (J. S. O'Neill et al., 2020; Takhaveev et al., 2023) so that biomass components are synthesised 'just-in-time' for when they are required for phase of the cell division cycle (Zylstra & Heinemann, 2022). However, the continued presence of the yeast metabolic cycle when the cell division cycle is halted implies that the metabolic cycle has functions other than control of the cell division cycle. Alternatively, the metabolic cycle may function as a environment-sensitive background rhythm to ensure that just-in-time biosynthesis for the cell division cycle can be switched on rapidly, as supported by my observation of rapid responses of cells to restoration of glucose after starvation.

Furthermore, the common control features of the yeast metabolic cycle and other biological rhythms may reflect a common evolutionary or functional origin. The yeast metabolic cycle may be mediated via post-translational modifications, as evidenced by how the proteome shows far less cyclic variation than the transcriptome (Feltham et al., 2020) or the metabolome (J. S. O'Neill et al., 2020). Specifically, studies have highlighted the role of cycles of peroxiredoxin oxidation in the integrity of the yeast metabolic cycle in chemostats (Amponsah et al., 2021; Causton et al., 2015), while other studies highlight the importance of chromatin remodelling in the metabolic cycle (Gowans et al., 2018; Nocetti & Whitehouse, 2016).

Continued cycles of peroxiredoxin oxidation have also been observed in circadian rhythms of transcription-disabled *Ostreococcus tauri* cells (J. S. O'Neill et al., 2011) and DNA-lacking human red blood cells (J. S. O'Neill & Reddy, 2011). Furthermore, human red blood cells also exhibit NADH/NADPH and ATP cycles that are independent of the glycolytic cycle (J. S. O'Neill & Reddy, 2011),

similar to the yeast metabolic cycle (Papagiannakis et al., 2017). In addition, the glycolytic cycle is regulated through solely biochemical means (Ghosh & Chance, 1964; Higgins, 1964), so this cycle may represent a basal mechanism of regulation of biological rhythms. Such similarities, across a range of biological kingdoms, thus strongly suggest a non-genetic common origin of biological rhythms.

Appendix A

Time series

A.1 catch22 features

Table A.1 lists *catch22* features.

A.2 UMAP hyperparameters

UMAP (McInnes et al., 2020) is a dimension reduction method which aims to find a global manifold structure of the input observations and computes a low-dimension embedding that preserves the structure of the manifold.

UMAP has several hyperparameters, of which four have major effects on the embedding:

1. The *number of neighbours* (n) to consider when approximating the local metric controls how the method balances local and global structure in the data. If this parameter takes low values, the algorithm concentrates on very local structure, potentially to the detriment of the big picture. As the value increases, the algorithm 'glues' more nodes together to form clusters.
2. The *largest embedding dimension* (d) controls the number of dimensions the data is reduced to. In other words, it controls whether the resulting map is one-dimensional, two-dimensional, three-dimensional, or of higher dimensions.

Feature name	Description
DN_HistogramMode_5	Mode of z-scored distribution (5-bin histogram)
DN_HistogramMode_10	Mode of z-scored distribution (10-bin histogram)
SB_BinaryStats_mean_longstretch1	Longest period of consecutive values above the mean
DN_OutlierInclude_p_001_mdrmd	Time intervals between successive extreme events above the mean
DN_OutlierInclude_n_001_mdrmd	Time intervals between successive extreme events below the mean
first_1e_ac	First 1/e crossing of autocorrelation function
firstMin_acf	First minimum of autocorrelation function
SP_Summaries_welch_rect_area_5_1	Total power in lowest fifth of frequencies in the Fourier power spectrum
SP_Summaries_welch_rect_centroid	Centroid of the Fourier power spectrum
FC_LocalSimple_mean3_stderr	Mean error from a rolling 3-sample mean forecasting
CO_trev_1_num	Time-reversibility statistic, $\langle (x_{t+1} - x_t)^3 \rangle_t$
CO_HistogramAMI_even_2_5	Automutual information, $m = 2, \tau = 5$
IN_AutoMutualInfoStats_40_-gaussian_fmmi	First minimum of the automutual information function
MD_hrv_classic_pnn40	Proportion of successive differences exceeding 0.04σ
SB_BinaryStats_diff_longstretch0	Longest period of successive incremental decreases
SB_MotifThree_quantile_hh	Shannon entropy of two successive letters in equiprobable 3-letter symbolization
FC_LocalSimple_mean1_tauresrat	Change in correlation length after iterative differencing
CO_Embed2_Dist_tau_d_expfit_-meandiff	Exponential fit to successive distances in 2-d embedding space
SC_FluctAnal_2_dfa_50_1_2_logi_-prop_r1	Proportion of slower timescale fluctuations that scale with DFA (50% sampling)
SC_FluctAnal_2_rsrangefit_50_1_-logi_prop_r1	Proportion of slower timescale fluctuations that scale with linearly rescaled range fits
SB_TransitionMatrix_3ac_-sumdiagcov	Trace of covariance of transition matrix between symbols in 3-letter alphabet
PD_PeriodicityWang_th0_01	Periodicity, based on first peak of autocorrelation function, satisfying certain conditions (Wang et al., 2007)

Table A.1: *catch22* features, adapted from Lubba et al. (2019).

3. The *minimal distance* (min_dist) controls the desired separation between close points in the embedding space. Specifically, this parameter controls how tightly the algorithm is allowed to pack points together. With low values, the visualisation forms 'clumps'.
4. The previous hyperparameters are numerical, but the *metric* hyperparameter instead specifies the distance metric that is used to compute distances in the ambient space of the input data. For example, this metric can be the Euclidean distance, the cosine distance, or other metrics used to compute the distances between two vectors of numerical data.

A.3 Classification pipeline

In machine learning, classification is defined as the process of identifying a category that a piece of input data belongs to. In this section, the classification task is identifying whether a time series (input data) is oscillatory (belongs to one category of two) or non-oscillatory (belongs to the other category of two).

A typical classification pipeline can be described by the following steps:

1. *Pre-processing of data:* Input data are cleaned or normalised. For example, to classify oscillatory time series, the input time series may be normalised to give similar dynamic ranges.
2. *Labelling:* Each piece of input data has a label assigned to it to denote which category it belongs to. For example, to classify oscillatory time series, a human can subjective assign the label '0' for non-oscillatory time series and '1' for oscillatory time series, for a total of two categories.

3. *Featurisation:* Input data are converted to feature vectors in the process of featurisation. This process uses domain knowledge related to the type or origin of the data to define characteristics of the data that may be useful for classification.
4. *Train-test split:* The input data are then randomly divided into a training data set and a test data set.
5. *Training of model:* The machine learning model is then fit on the (featurised) training data set and its labels to fit parameters in the model.
6. *Evaluation of model on test dataset:* The model, trained on the training dataset, is used to predict the labels of data in the (featurised) test data set. The performance of the model is then evaluated on the test data set. This evaluation is based on computing quantities that express how well the model assigns labels to data, compared to the labels defined earlier.

A.4 Gillespie algorithm for stochastic chemical systems

To define the Gillespie algorithm (Gillespie, 2007), consider such a system with M reactions $R_1, \ldots, R_j, \ldots R_M$ involving N species $S_1, \ldots, S_i, \ldots S_N$ in a fixed volume V at thermal equilibrium. Let $X_i(t)$ represent the number of molecules of S_i at time t, and the state vector

$$\mathbf{X}(t) = [X_1(t), \ldots, X_N(t)] \tag{A.1}$$

thus gives the state of the system at any given time t.

Each reaction R_j is described by two quantities:

1. A state-change vector $\mathbf{v}_j = [v_{1,j}, \ldots, v_{N,j}]$ which defines how the stoichiometry of the system changes if the reaction occurs. $v_{i,j}$ represents the change in the stoichiometry of S_i when R_j occurs.
2. A propensity function $a_j(\mathbf{x})$, which gives the probability, that one R_j reaction occurs in the volume V within the following short time interval $[t, t{+}dt)$, given the state $\mathbf{X}(t) = \mathbf{x}$.

The definition of the propensity function $a_j(\mathbf{x})$ differs depending on the reaction. For a unimolecular reaction ($\mathrm{S}_1 \longrightarrow \mathrm{P}$),

$$a_j(\mathbf{x}) = k_j X_1 \tag{A.2}$$

For a bimolecular reaction with two different species ($\mathrm{S}_1 + \mathrm{S}_2 \longrightarrow \mathrm{P}$),

$$a_j(\mathbf{x}) = \frac{k_j}{V} X_1 X_2 \tag{A.3}$$

And finally, for a bimolecular reaction with two molecules of the same species ($\mathrm{S}_1 + \mathrm{S}_1 \longrightarrow \mathrm{P}$),

$$a_j(\mathbf{x}) = \frac{k_j}{V} X_1 (X_1 - 1) \tag{A.4}$$

Where, in all cases, k_j is the rate constant of reaction R_j.

The Gillespie algorithm aims to predict the temporal evolution of the state vector given the initial state $\mathbf{X}(t_0) = \mathbf{x}_0$. It does so by iteratively choosing the next reaction that occurs, based on its probability, and then choosing its firing time based on a propensity function. Combining these simulations gives a trajectory of state vectors across the time course of interest. In detail, the direct Gillespie algorithm can be defined as stated in Algorithm 1:

Algorithm 1: Direct method of the Gillespie algorithm

Input: Stochastic model (with species $S_1, \ldots, S_i, \ldots S_N$ and reactions $R_1, \ldots, R_j, \ldots R_M$, along with a state-change vector $\mathbf{v_j}$ and a rate constant k_i for each reaction R_j); initial time t_0; and initial model state $\mathbf{X}(t_0) = \mathbf{x}_0$

Output: Trajectory of state vectors $\mathbf{X}(t)$, with t taking discrete values in $[t_0, t_{\max}]$

while $t < t_{\max}$ **do**

- Calculate the propensities $a_j(\mathbf{x})$ based on the current state $\mathbf{x}$;
- Calculate the combined propensity $a_0(\mathbf{x}) = \sum_j a_j(\mathbf{x})$;
- Generate two random numbers r_1 and r_2, both from the uniform distribution $U(0, 1)$;
- Choose the next reaction R_j, with j given by the smallest integer that satisfies $\sum_{j'}^{j} a_{j'(\mathbf{x})} > r_1 a_0(\mathbf{x})$;
- Calculate the time to the next reaction $\tau = \frac{1}{a_0(\mathbf{x})} \ln(\frac{1}{r_2})$;
- Simulate the next reaction by updating the state vector $\mathbf{x} \leftarrow \mathbf{x} + \mathbf{v_j}$ and store the new vector in $\mathbf{X}(t)$;
- Update the time by $t \leftarrow t + \tau$ and store the new time;

end

return Trajectory of state vectors $\mathbf{X}(t)$ for a vector of times t;

Appendix B

Flux balance analysis

B.1 Modifications of chemical reaction in GECKO, general cases

To illustrate how GECKO modifies chemical reactions, consider a simple example, a reaction R_j catalysed by enzyme E_i:

$$\mathrm{A} + \mathrm{B} \xrightarrow{\mathrm{E}_i} \mathrm{C} + \mathrm{D} \tag{B.1}$$

To apply the constraints described by Eq. 5.2, the chemical reaction in Eq. B.1 is modified by adding a term:

$$\mathrm{n_{ij}E}_i + \mathrm{A} + \mathrm{B} \longrightarrow \mathrm{C} + \mathrm{D} \tag{B.2}$$

with the stoichiometric coefficient $n_{ij} = 1/k_{\mathrm{cat}}^{ij}$. GECKO takes k_{cat} values from BRENDA (Chang et al., 2021).

This transformation, adding the enzyme as a pseudometabolite, is based on the intuition that the system uses some amount of enzyme at a specific time to catalyse the flux through the reaction. Slightly different formalisms are applied to reversible reactions, isozymes, promiscuous enzymes, and enzyme complexes (Appendix B.2).

To constrain overall enzyme levels in the model, GECKO defines a pseudoreaction

$$ER_{\text{pool}} : \varnothing \longrightarrow \text{E}_{\text{pool}} \tag{B.3}$$

where E_{pool} is a pseudometabolite that represents the proteome pool available for enzymes. This pseudoreaction has a flux

$$e_{\text{pool}} \leq (P_{\text{total}} - P_{\text{measured}}) \cdot f \cdot \sigma \tag{B.4}$$

in units of $\text{g}\,\text{g}_{\text{DW}}^{-1}$, where P_{total} is the total protein fraction with respect to the dry weight of the cell, P_{measured} is the protein fraction of proteins whose weight are accounted for in the model, based on proteomic data, f represents the fraction of proteins that are enzymes, and σ is a parameter that represents the average saturation of enzymes. ecYeast8.6.0 assumes parameter values of $f = 0.5$, $P_{\text{total}} = 0.5$, and $\sigma = 0.5$. If no proteomic data is used, as is the case in this thesis, $P_{\text{measured}} = 0$.

Defining such parameters is a judgement call, especially when the protein fraction varies across growth rates (Elsemman et al., 2022), but $f = 0.5$ is close to the protein mass fraction of ecYeast8.6.0. Subsequently, GECKO changes the carbohydrate composition based on the assumption that a change in the amino acid composition is offset by the reverse change in the carbohydrate composition; experimental data justifies this assumption (Nissen et al., 1997).

Then, for each enzyme E_i, GECKO defines enzyme usage pseudoreactions of the form

$$ER_i : \text{MW}_i\,\text{E}_{\text{pool}} \longrightarrow \text{E}_i \tag{B.5}$$

where MW_i represents the molecular weight of the enzyme in units of $g\,mmol^{-1}$. The flux of enzyme usage pseudoreactions are defined in units of $mmol\,g_{DW}^{-1}$. GECKO takes enzyme data from SWISSPROT (The UniProt Consortium, 2023) and KEGG (Kanehisa et al., 2023), including molecular weight of proteins and associated pathways.

Taken together, the modelled cell thus has an enzyme pool in terms of a mass fraction of the cell's dry weight, and the modelled cell allocates certain fractions of this mass to the synthesis of each enzyme at steady-state. The mass of each enzyme in the cell determines the amount (in moles) of each enzyme and therefore its catalytic activity.

B.2 Modifications of chemical reactions in GECKO for non-simple cases

GECKO modifies chemical reactions in a genome-scale model so that enzymes are expressed as metabolites that take part in reactions. Section 5.2.1 describes how this was performed for the simple case of a singular chemical reaction and one enzyme having a one-to-one association. Slightly different formalisms are applied to reversible reactions, isozymes, promiscuous enzymes, and enzyme complexes. Namely:

- Reversible reactions are modelled as the forward and reverse reactions separately.
- For isozymes, the original reaction is copied multiple times corresponding to the number of reactions that the isozyme catalyses. Each has an isozyme catalysing the reaction. In addition, there is an 'arm' reaction to act as an intermediate between the substrate and the products.

- No actions are needed for promiscuous enzymes.
- Enzyme complexes are modelled as one reaction that uses all subunit proteins that all share the same k_{cat} value.

B.3 Computing molecular weights of pseudometabolites in ecYeast8

In genome-scale metabolic models, mass fractions are represented in the biomass reaction, set as the objective function, which is defined as:

$$\mathrm{f_1M_1 + f_2M_2 + \cdots + f_nM_n \longrightarrow B} \tag{B.6}$$

where $M_1 \ldots M_n$ represent the chemical species that make up the cell's biomass, the stoichiometric coefficients $f_1 \ldots f_n$ represent the mass fraction of each species in units of $\mathrm{g\,g_{DW}^{-1}}$, and B represents biomass. If a chemical species M_i has a mass fraction f_i, then 1 g of cell dry weight has f_i g of chemical species M_i.

In the ecYeast8 model, the mass fraction of biomass components cannot simply be obtained by taking the stoichiometric coefficients (f_i in Eq. B.6) from the objective function. This is because the objective function of Yeast8 does not conform to this format, and instead contains pseudometabolites. This formalism can be expressed as:

$$\begin{aligned}
\mathrm{f_1M_{B,1} + \cdots + f_nM_{B,n} + f_{n+1}P_1 + \cdots + f_{n+k}P_k} &\longrightarrow \mathrm{B} \\
\mathrm{f_{P_1,1}M_{P_1,1} + \cdots + f_{P_1,n}M_{P_1,n}} &\longrightarrow \mathrm{P_1} \\
&\cdots \\
\mathrm{f_{P_k,1}M_{P_k,1} + \cdots + f_{P_k,n}M_{P_k,n}} &\longrightarrow \mathrm{P}_k
\end{aligned} \tag{B.7}$$

where each M is a chemical species with a defined molecular weight, each P is a pseudometabolite, and each f is a stoichiometric coefficient. The objective function remains the reaction that produces B, but some chemical species $M_{B,1} \dots M_{B,n}$ are retained in the objective function, while other chemical species are replaced by pseudometabolites $P_1 \dots P_k$. The reactions that produce $P_1 \dots P_k$ are *isa* reactions. *isa* reactions define pseudometabolites by having chemical species with known molecular weights as reactants, with their stoichiometric coefficients representing abundance in mmol $\mathrm{g}_{\mathrm{DW}}^{-1}$. In Yeast8, the objective function is defined as:

```
 47.5883 atp_c + 47.5883 h2o_c + lipid_c + protein_c + carbohydrate_c
+ dna_c + rna_c + cofactor + ion
-> 47.5883 adp_c + biomass_c + 47.5883 h_c + 47.5883 pi_c
```

Here, there are seven pseudometabolites: lipid, protein, carbohydrate, DNA, RNA, cofactor, and ion.

As the ecYeast8 model does not specify the molecular weights of these pseudometabolites, in order to obtain the mass fraction of each biomass component represented by the pseudometabolites, I treated each pseudometabolite as a chemical species and calculated its molecular weight by assuming mass balance (Chan et al., 2017; Dinh et al., 2022; Takhaveev et al., 2023). Namely, I assumed that in reactions that produce the pseudometabolites, there is conservation of mass, and therefore:

$$\sum_{r=j}^{n_r} m_r c_r = \sum_{p=i}^{n_p} m_p c_p \tag{B.8}$$

where $s = 1, \ldots n_s$ represents substrates of the reaction in question, $p = 1, \ldots n_p$ represents products. m_r represents molar mass of reactant r, m_p represents molar mass of product p, c_r represents stoichiometric coefficient of reactant r, and c_p represents stoichiometric coefficient of product p.

The resulting molecular weight will thus represent the mass fraction of each biomass component in units of $\mathrm{g\,g_{DW}^{-1}}$.

B.3.1 Carbohydrate, DNA, RNA, cofactor, and ion pseudometabolites

Computing the molecular weights of the carbohydrate, DNA, RNA, cofactor, and ion pseudometabolites is straightforward. This is because the equations similarly have reactants with molecular weights specified in the model and only the pseudometabolite, the sole product, does not have a molecular weight specified. In such cases, Eq. B.8 can be applied directly, i.e. the molecular weight of the pseudometabolite is equal to $\sum_{r=j}^{n_r} m_r c_r$, where m_r values are taken directly from the molecular weights specified in the model. The results for these pseudometabolites are shown in Table B.1.

B.3.2 Protein pseudometabolite

Other metabolites were less straightforward. To compute the molecular weight of the protein pseudometabolite, I inspected reaction `r_4047`:

```
 0.57284 Ala-tRNA(Ala) + 0.200644 Arg-tRNA(Arg) + 0.126979 Asn-tRNA(Asn)
+ ... + 0.330369 Val-tRNA(Val)
-> 0.57284 tRNA(Ala) + 0.200644 tRNA(Arg) + 0.126979 tRNA(Asn)
+ ... + 0.330369 tRNA(Val) + protein
```

ID	Reaction	Computed molecular weight ($g\,mol^{-1}$)
r_4048	0.684535 (1->3)-beta-D-glucan + 0.228715 (1->6)-beta-D-glucan + 0.330522 glycogen + 0.650171 mannan + 0.126456 trehalose -> carbohydrate	350.37
r_4050	0.0036 dAMP + 0.0024 dCMP + 0.0024 dGMP + 0.0036 dTMP -> DNA	3.90
r_4049	0.0445348 AMP + 0.0432762 CMP + 0.0445348 GMP + 0.0579921 UMP -> RNA	64.04
r_4598	0.00019 coenzyme A + 1e-05 FAD + 0.00265 NAD + 0.00015 NADH + 0.00057 NADP(+) + 0.0027 NADPH + 0.00099 riboflavin + 1.2e-06 TDP + 6.34e-05 THF + 1e-06 heme a -> cofactor	4.83
r_4599	3.04e-05 iron(2+) + 0.00363 potassium + 0.00397 sodium + 0.02 sulphate + 0.00129 chloride + 0.00273 Mn(2+) + 0.000748 Zn(2+) + 0.000217 Ca(2+) + 0.00124254 Mg(2+) + 0.000659 Cu(2+) -> ion	2.48

Table B.1: Straightforward cases of computing pseudometabolite molecular weights from pseudoreactions in ecYeast8

In Yeast8, aminoacyl-tRNA and tRNA species do not have molecular weights specified in the model. This is because their chemical formulas are incompletely specified in the model as the model uses `R` to represent the tRNA. For example, `Ala-tRNA(Ala)`, alanyl-tRNA, is represented as `C3H7NOR`, and `tRNA(Ala)` is represented as `RH`. As a consequence, the molecular weights of these species cannot be directly computed from the chemical formula. Because tRNAs are unmodified during translation, `R` can be ignored. In other words, I treated `R` as a chemical

element of atomic mass 0 when computing m_r for each reactant and m_p for each product, leaving only m_p for `protein` undefined. This m_p can then be found by rearranging Eq. B.8, thus giving the molecular mass of the protein pseudometabolite.

B.3.3 Lipid pseudometabolite

Finally, the lipid pseudometabolite is the least straightforward because the model does not specify the molecular weights of some of the reactants of the lipid pseudoreaction. The lipid pseudoreaction is represented in reaction `r_2108`:

```
lipid backbone + lipid chain -> lipid
```

And both `lipid backbone` and `lipid chain` have no molecular weight specified.

Reaction `r_4065` specifies a lipid chain pseudoreaction, in which `lipid chain` is generated:

```
0.0073947 C16:0 chain + 0.0217019 C16:1 chain + 0.0020726 C18:0 chain
+ 0.000796243 C18:1 chain
-> lipid chain
```

As all reactants have molecular weights defined in the model, the molecular weight of `lipid chain` can be computed from the mass balance of this reaction.

Reaction `r_4063` specifies a lipid backbone pseudoreaction, in which `lipid backbone` is generated:

```
0.000631964 1-phosphatidyl-1D-myo-inositol backbone
+ 0.00243107 ergosterol + 0.000622407 ergosterol ester backbone
+ 0.000135359 fatty acid backbone + ...
-> lipid backbone
```

ID	Reaction	Computed molecular weight ($g\,mol^{-1}$)
r_3975	palmitate -> 0.255421 fatty acid backbone 0.256429 C16:0 chain	742.54
r_3976	palmitoleate -> 0.253405 fatty acid backbone 0.254413 C16:1 chain	744.56
r_3977	stearate -> 0.283475 fatty acid backbone 0.284483 C18:0 chain	714.49
r_3978	oleate -> 0.281459 fatty acid backbone 0.282467 C18:1 chain	716.51

Table B.2: ecYeast8 reactions that generate the `fatty acid backbone` metabolite

The model specifies molecular weights for all species in the reaction that generates `lipid backbone`, expect for `fatty acid backbone`. To compute the molecular weight of `fatty acid backbone`, the reactions that produce this species must be used. Because of SLIMEr, four reactions in the model produce `fatty acid backbone` (Table B.2). The model specifies molecular weights for all species in these four reactions, except for `fatty acid backbone`. Therefore, using each chemical equation, the molecular weight of `fatty acid backbone` can be solved for by rearranging Eq. B.8 with parameters defined to match each reaction. However, the molecular weights computed from each equation are different. Because the differences are slight, I took the mean of the four weights to give $729.53\,g\,mol^{-1}$. Subsequently, the molecular weight of `lipid backbone` was computed from this mean value and the molecular weights of other species involved in the production of `lipid backbone`, giving $21.31\,g\,mol^{-1}$. With the molecular weights of `lipid backbone` and `lipid chain` defined, the molecular weight of `lipid` is thus the sum of the two.

A summary of molecular weights can be found in Table 2.7.

Subsequently, the mass fraction of each biomass component is computed by dividing the molecular weight of the corresponding pseudometabolite by the molecular weight of biomass (Table B.3).

Metabolite	f_i
Protein	0.524 53
Carbohydrate	0.364 38
RNA	0.066 60
Lipid	0.032 83
Cofactors	0.005 02
DNA	0.004 06
Ions	0.002 58

Table B.3: f_i values for each biomass component.

B.4 Mathematical explanation of the effect of restricting the enzyme pool

To explain how $\tau_{\text{seq/par}}$ increases as the proteome pool shrinks, let $\tau_{\text{seq/par}}$, given by Eq. 5.11, depend on x:

$$\tau_{\text{seq/par}}(x) = \left(\sum_i \frac{f_i}{\lambda_{\text{seq},i}(x)} \right) \cdot \frac{\lambda_0(x)}{f_{\text{protein}}} \tag{B.9}$$

where $x = e'_{\text{pool}}/e_{\text{pool}}$. The expression in Eq. B.9 takes into account how λ_0 and $\lambda_{\text{seq},i}$ values vary with x, and how f_i values are constants.

We thus obtain:

$$\begin{aligned}
\frac{\mathrm{d}\,\tau_{\text{seq/par}}(x)}{\mathrm{d}x} &= \frac{1}{f_{\text{protein}}}\frac{\mathrm{d}}{\mathrm{d}x}\left[\left(\sum_i \frac{f_i}{\lambda_{\text{seq},i}(x)}\right)\cdot\lambda_0(x)\right] \\
&= \frac{1}{f_{\text{protein}}}\left[\left(\sum_i \frac{f_i}{\lambda_{\text{seq},i}(x)}\right)\cdot\frac{\mathrm{d}\,\lambda_0(x)}{\mathrm{d}x} + \lambda_0(x)\frac{\mathrm{d}}{\mathrm{d}x}\left(\sum_i \frac{f_i}{\lambda_{\text{seq},i}(x)}\right)\right] \\
&= \frac{1}{f_{\text{protein}}}\left[\left(\sum_i \frac{f_i}{\lambda_{\text{seq},i}(x)}\right)\cdot\frac{\mathrm{d}\,\lambda_0(x)}{\mathrm{d}x} - \lambda_0(x)\sum_i\left(\frac{f_i}{\lambda_{\text{seq},i}(x)^2}\cdot\frac{\mathrm{d}\,\lambda_{\text{seq},i}(x)}{\mathrm{d}x}\right)\right]
\end{aligned} \tag{B.10}$$

To explain the increase in $\tau_{\text{seq/par}}$ as e'_{pool} increases, I consider the behaviour of λ_0 and $\lambda_{\text{seq},i}$ values with respect to e'_{pool} in intervals.

With reference to Fig. 5.6, consider $0 \leq x \leq 0.5$. In this region of x, based on the observations in the figure, we model $\lambda_0 = k_0 x$ and $\lambda_{\text{seq},i} = k_i x$, where constants $k_0, k_i > 0$. This models how these values initially increase linearly in Fig. 5.6. Thus, Eq. B.10 becomes:

$$\begin{aligned}
\frac{\mathrm{d}\,\tau_{\text{seq/par}}(x)}{\mathrm{d}x} &= \frac{1}{f_{\text{protein}}}\left[\left(\sum_i \frac{f_i}{k_i x}\right)\cdot k_0 - k_0 x\sum_i\left(\frac{f_i}{(k_i x)^2}\cdot k_i\right)\right] \\
&= \frac{1}{f_{\text{protein}}}\left[\frac{k_0}{x}\sum_i\frac{f_i}{k_i} - k_0 x\left(\sum_i\frac{f_i}{k_i x^2}\right)\right] \\
&= \frac{1}{f_{\text{protein}}}\left[\frac{k_0}{x}\sum_i\frac{f_i}{k_i} - \frac{k_0}{x}\sum_i\frac{f_i}{k_i}\right] \\
&= 0
\end{aligned} \tag{B.11}$$

Eq. B.11 thus explains the constant $\tau_{\text{seq/par}}$ in this region.

Now, consider $0.5 < x \leq 9$. In this region, the trajectories of $\lambda_{\text{seq},i}$ with respect to time remain linear, but some with changes in slope. In other words, in a sub-region where the slopes of all $\lambda_{\text{seq},i}$ are constant, we can let: $\lambda_0 = k_0 x$ and $\lambda_{\text{seq},i} = m_i x + c_i$, where $k_0, m_i, c_i > 0$. Thus, Eq. B.10 becomes:

$$\begin{aligned}
\frac{\mathrm{d}\,\tau_{\text{seq/par}}(x)}{\mathrm{d}x} &= \frac{1}{f_{\text{protein}}}\left[\left(\sum_i \frac{f_i}{m_i x + c_i}\right)\cdot k_0 - k_0 x \sum_i \left(\frac{f_i}{(m_i x + c_i)^2}\cdot m_i\right)\right] \\
&= \frac{k_0}{f_{\text{protein}}}\left[\left(\sum_i \frac{f_i}{m_i x + c_i}\right) - x\left(\sum_i \frac{f_i m_i}{(m_i x + c_i)^2}\right)\right] \\
&= \frac{k_0}{f_{\text{protein}}}\sum_i \left[\frac{f_i}{m_i x + c_i} - \frac{x f_i m_i}{(m_i x + c_i)^2}\right] \\
&= \frac{k_0}{f_{\text{protein}}}\sum_i \left[\frac{f_i c_i}{(m_i x + c_i)^2}\right]
\end{aligned} \tag{B.12}$$

As $f_i, c_i, m_i > 0$ for all biomass components i, and $k_0 > 0$, we get $\frac{\mathrm{d}\,\tau_{\text{seq/par}}(x)}{\mathrm{d}x} > 0$ regardless of the values that these constants take. Because k_0 does not change over the region of x considered, m_i, c_i, and x values thus determine the magnitude of $\frac{\mathrm{d}\,\tau_{\text{seq/par}}(x)}{\mathrm{d}x}$. If within a region of x, m_i and c_i values remain constant for all i, then as x increases, $\frac{\mathrm{d}\,\tau_{\text{seq/par}}(x)}{\mathrm{d}x}$ should decrease — this was the case, as observed in Fig. 5.6.

Lastly, consider $x > 9$. In this region, λ_0 becomes constant, thus we let $\lambda_0 = k_0$. We keep $\lambda_{\text{seq},i} = m_i x + c_i$, and as before, $k_0, m_i, c_i > 0$. Thus, Eq. B.10 becomes:

$$\begin{aligned}
\frac{\mathrm{d}\,\tau_{\text{seq/par}}(x)}{\mathrm{d}x} &= \frac{1}{f_{\text{protein}}}\left[0 - k_0 \sum_i \left(\frac{f_i}{(m_i x + c_i)^2}\cdot m_i\right)\right] \\
&= -\frac{k_0}{f_{\text{protein}}}\sum_i \left[\frac{f_i m_i}{(m_i x + c_i)^2}\right]
\end{aligned} \tag{B.13}$$

This predicts *decreasing* $\tau_{\text{seq/par}}$ as x increases in this region. Because k_0 is constant in this region, the rate of this decrease is thus controlled by m_i and c_i values. As each $\lambda_{\text{seq},i}$ trajectory becomes flat as x increases, each $\frac{f_i m_i}{(m_i x + c_i)^2}$ term becomes zero, thus shrinking the magnitude of $\frac{\mathrm{d}\,\tau_{\text{seq/par}}(x)}{\mathrm{d}x}$. Finally, as all $\lambda_{\text{seq},i}$ trajectories become flat at $x > 15$, $\frac{\mathrm{d}\,\tau_{\text{seq/par}}(x)}{\mathrm{d}x} = 0$.

Bibliography

Adjavon, D.-Y. (2022). *Computational analysis of single-cell dynamics: Protein localisation, cell cycle, and metabolic adaptation* [Doctoral dissertation, University of Edinburgh].

Adler, S. O., Spiesser, T. W., Uschner, F., Münzner, U., Hahn, J., Krantz, M., & Klipp, E. (2022). A yeast cell cycle model integrating stress, signaling, and physiology. *FEMS Yeast Research*, foac026.

Adrover, M. À., Zi, Z., Duch, A., Schaber, J., González-Novo, A., Jimenez, J., Nadal-Ribelles, M., Clotet, J., Klipp, E., & Posas, F. (2011). Time-Dependent Quantitative Multicomponent Control of the G1-S Network by the Stress-Activated Protein Kinase Hog1 upon Osmostress. *Science Signaling*, *4*(192), ra63–ra63.

Ahn, E., Kumar, P., Mukha, D., Tzur, A., & Shlomi, T. (2017). Temporal flux-omics reveals oscillations in TCA cycle flux throughout the mammalian cell cycle. *Molecular Systems Biology*, *13*(11), 953.

Amponsah, P. S., Yahya, G., Zimmermann, J., Mai, M., Mergel, S., Mühlhaus, T., Storchova, Z., & Morgan, B. (2021). Peroxiredoxins couple metabolism and cell division in an ultradian cycle. *Nature Chemical Biology*, *17*(4), 477–484.

Arata, Y., & Takagi, H. (2019). Quantitative Studies for Cell-Division Cycle Control. *Frontiers in Physiology*, *10*.

Ariño, J., Ramos, J., & Sychrová, H. (2010). Alkali Metal Cation Transport and Homeostasis in Yeasts. *Microbiology and Molecular Biology Reviews*, *74*(1), 95–120.

Aubin, J. E. (1979). Autofluorescence of viable cultured mammalian cells. *Journal of Histochemistry & Cytochemistry*, *27*(1), 36–43.

Aung, H. W., Henry, S. A., & Walker, L. P. (2013). Revising the Representation of Fatty Acid, Glycerolipid, and Glycerophospholipid Metabolism in the Consensus Model of Yeast Metabolism. *Industrial Biotechnology*, *9*(4), 215–228.

Bagamery, L. E., Justman, Q. A., Garner, E. C., & Murray, A. W. (2020). A Putative Bet-Hedging Strategy Buffers Budding Yeast against Environmental Instability. *Current Biology*.

Bai, S., Kolter, J. Z., & Koltun, V. (2018). An Empirical Evaluation of Generic Convolutional and Recurrent Networks for Sequence Modeling.

Baumgartner, B. L., O'Laughlin, R., Jin, M., Tsimring, L. S., Hao, N., & Hasty, J. (2018). Flavin-based metabolic cycles are integral features of growth and division in single yeast cells. *Scientific Reports*, *8*(1), 1–10.

Becker, S. A., & Palsson, B. O. (2008). Context-Specific Metabolic Networks Are Consistent with Experiments. *PLOS Computational Biology*, *4*(5), e1000082.

Becker-Kettern, J., Paczia, N., Conrotte, J.-F., Kay, D. P., Guignard, C., Jung, P. P., & Linster, C. L. (2016). Saccharomyces cerevisiae Forms d-2-Hydroxyglutarate and Couples Its Degradation to d-Lactate Formation via a Cytosolic Transhydrogenase. *Journal of Biological Chemistry*, *291*(12), 6036–6058.

Benjamini, Y., & Hochberg, Y. (1995). Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society. Series B (Methodological)*, *57*(1), 289–300.

Beuse, M., Bartling, R., Kopmann, A., Diekmann, H., & Thoma, M. (1998). Effect of the dilution rate on the mode of oscillation in continuous cultures of Saccharomyces cerevisiae. *Journal of Biotechnology*, *61*(1), 15–31.

Blondel, V. D., Guillaume, J.-L., Lambiotte, R., & Lefebvre, E. (2008). Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment*, *2008*(10), P10008.

Boiteux, A., Goldbeter, A., & Hess, B. (1975). Control of oscillating glycolysis of yeast by stochastic, periodic, and steady source of substrate: A model and experimental study. *Proceedings of the National Academy of Sciences*, *72*(10), 3829–3833.

Bonamore, A., & Boffi, A. (2008). Flavohemoglobin: Structure and reactivity. *IUBMB Life*, *60*(1), 19–28.

Boxall, S. F., Dever, L. V., Kneřová, J., Gould, P. D., & Hartwell, J. (2017). Phosphorylation of Phosphoenolpyruvate Carboxylase Is Essential for Maximal and Sustained Dark CO2 Fixation and Core Circadian Clock Operation in the Obligate Crassulacean Acid Metabolism Species Kalanchoë fedtschenkoi. *The Plant Cell*, *29*(10), 2519–2536.

Brachmann, C. B., Davies, A., Cost, G. J., Caputo, E., Li, J., Hieter, P., & Boeke, J. D. (1998). Designer deletion strains derived from Saccharomyces cerevisiae S288C: A useful set of strains and plasmids for PCR-mediated gene disruption and other applications. *Yeast*, *14*(2), 115–132.

Breiman, L. (2001). Random Forests. *Machine Learning*, *45*(1), 5–32.

Burnetti, A. J., Aydin, M., & Buchler, N. E. (2016). Cell cycle Start is coupled to entry into the yeast metabolic cycle across diverse strains and growth rates. *Molecular Biology of the Cell*, *27*(1), 64–74.

Calabrese, G., Peker, E., Amponsah, P. S., Hoehne, M. N., Riemer, T., Mai, M., Bienert, G. P., Deponte, M., Morgan, B., & Riemer, J. (2019). Hyperoxidation of mitochondrial peroxiredoxin limits H2O2-induced cell death in yeast. *The EMBO Journal*, *38*(18), e101552.

Campbell, K., Westholm, J., Kasvandik, S., Bartolomeo, F. D., Mormino, M., & Nielsen, J. (2020). Building blocks are synthesized on demand during the yeast cell cycle. *Proceedings of the National Academy of Sciences*, *117*(14), 7575–7583.

Canelas, A. B., Ras, C., ten Pierick, A., van Gulik, W. M., & Heijnen, J. J. (2011). An in vivo data-driven framework for classification and quantification of enzyme kinetics and determination of apparent thermodynamic data. *Metabolic Engineering*, *13*(3), 294–306.

Casey, E., Mosier, N. S., Adamec, J., Stockdale, Z., Ho, N., & Sedlak, M. (2013). Effect of salts on the Co-fermentation of glucose and xylose by a genetically engineered strain of Saccharomyces cerevisiae. *Biotechnology for Biofuels*, *6*(1), 83.

Causton, H. C. (2018). Metabolic rhythms: A framework for coordinating cellular function. *The European Journal of Neuroscience*.

Causton, H. C., Feeney, K. A., Ziegler, C. A., & O'Neill, J. S. (2015). Metabolic Cycles in Yeast Share Features Conserved among Circadian Rhythms. *Current Biology*, *25*(8), 1056–1062.

Cesur, M. F., Çakır, T., & Pir, P. (2022). Genome-Wide Analysis of Yeast Metabolic Cycle through Metabolic Network Models Reveals Superiority of Integrated ATAC-seq Data over RNA-seq Data. *mSystems*, *0*(0), e01347–21.

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. *Bioinformatics*, *33*(22), 3603–3609.

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

Charvin, G., Cross, F. R., & Siggia, E. D. (2009). Forced periodic expression of G1 cyclins phase-locks the budding yeast cell cycle. *Proceedings of the National Academy of Sciences*, *106*(16), 6632–6637.

Chen, K. C., Calzone, L., Csikasz-Nagy, A., Cross, F. R., Novak, B., & Tyson, J. J. (2004). Integrative Analysis of Cell Cycle Control in Budding Yeast. *Molecular Biology of the Cell*, *15*(8), 3841–3862.

Chen, Y., Li, F., & Nielsen, J. (2022). Genome-scale modeling of yeast metabolism: Retrospectives and perspectives. *FEMS Yeast Research*, foac003.

Chen, Y., & Nielsen, J. (2022). Yeast has evolved to minimize protein resource cost for synthesizing amino acids. *Proceedings of the National Academy of Sciences*, *119*(4), e2114622119.

Christ, M., Braun, N., Neuffer, J., & Kempa-Liehr, A. W. (2018). Time Series FeatuRe Extraction on basis of Scalable Hypothesis tests (tsfresh – A Python package). *Neurocomputing*, *307*, 72–77.

Cox, D. R. (1972). Regression Models and Life-Tables. *Journal of the Royal Statistical Society: Series B (Methodological)*, *34*(2), 187–202.

Crane, M. M., Clark, I. B. N., Bakker, E., Smith, S., & Swain, P. S. (2014). A Microfluidic System for Studying Ageing and Dynamic Single-Cell Responses in Budding Yeast. *PLOS ONE*, *9*(6), e100042.

Cuny, A. P., Tanuj Sapra, K., Martinez-Martin, D., Fläschner, G., Adams, J. D., Martin, S., Gerber, C., Rudolf, F., & Müller, D. J. (2022). High-resolution mass measurements of single budding yeast reveal linear growth segments. *Nature Communications*, *13*(1), 3483.

Decroly, O., & Goldbeter, A. (1982). Birhythmicity, chaos, and other patterns of temporal self-organization in a multiply regulated biochemical system. *Proceedings of the National Academy of Sciences*, *79*(22), 6917–6921.

Dinh, H. V., Sarkar, D., & Maranas, C. D. (2022). Quantifying the propagation of parametric uncertainty on flux balance analysis. *Metabolic Engineering*, *69*, 26–39.

Dobson, P. D., Smallbone, K., Jameson, D., Simeonidis, E., Lanthaler, K., Pir, P., Lu, C., Swainston, N., Dunn, W. B., Fisher, P., Hull, D., Brown, M., Oshota, O., Stanford, N. J., Kell, D. B., King, R. D., Oliver, S. G., Stevens, R. D., & Mendes, P. (2010). Further developments towards a genome-scale metabolic model of yeast. *BMC Systems Biology*, *4*(1), 145.

Dodd, A. N., Salathia, N., Hall, A., Kévei, E., Tóth, R., Nagy, F., Hibberd, J. M., Millar, A. J., & Webb, A. A. R. (2005). Plant Circadian Clocks Increase Photosynthesis, Growth, Survival, and Competitive Advantage. *Science*, *309*(5734), 630–633.

Dodd, B. J. T., & Kralj, J. M. (2017). Live Cell Imaging Reveals pH Oscillations in Saccharomyces cerevisiae During Metabolic Transitions. *Scientific Reports*, *7*(1), 13922.

Domenzain, I., Sánchez, B., Anton, M., Kerkhoven, E. J., Millán-Oropeza, A., Henry, C., Siewers, V., Morrissey, J. P., Sonnenschein, N., & Nielsen, J. (2022). Reconstruction of a catalogue of genome-scale metabolic models with enzymatic constraints using GECKO 2.0. *Nature Communications*, *13*(1), 3766.

Droin, C., Paquet, E. R., & Naef, F. (2019). Low-dimensional dynamics of two coupled biological oscillators. *Nature Physics*, *15*(10), 1086–1094.

Duarte, N. C., Herrgård, M. J., & Palsson, B. Ø. (2004). Reconstruction and Validation of Saccharomyces cerevisiae iND750, a Fully Compartmentalized Genome-Scale Metabolic Model. *Genome Research*, *14*(7), 1298–1309.

Dunlop, M. J., Cox, R. S., Levine, J. H., Murray, R. M., & Elowitz, M. B. (2008). Regulatory activity revealed by dynamic correlations in gene expression noise. *Nature Genetics*, *40*(12), 1493–1498.

Edgar, R. S., Green, E. W., Zhao, Y., van Ooijen, G., Olmedo, M., Qin, X., Xu, Y., Pan, M., Valekunja, U. K., Feeney, K. A., Maywood, E. S., Hastings, M. H., Baliga, N. S., Merrow, M., Millar, A. J., Johnson, C. H., Kyriacou, C. P., O'Neill, J. S., & Reddy, A. B. (2012). Peroxiredoxins are conserved markers of circadian rhythms. *Nature*, *485*(7399), 459–464.

Eelderink-Chen, Z., Mazzotta, G., Sturre, M., Bosman, J., Roenneberg, T., & Merrow, M. (2010). A circadian clock in Saccharomyces cerevisiae. *Proceedings of the National Academy of Sciences*, *107*(5), 2043–2047.

Elowitz, M. B., & Leibler, S. (2000). A synthetic oscillatory network of transcriptional regulators. *Nature*, *403*(6767), 335–338.

Elsemman, I. E., Rodriguez Prado, A., Grigaitis, P., Garcia Albornoz, M., Harman, V., Holman, S. W., van Heerden, J., Bruggeman, F. J., Bisschops, M. M. M., Sonnenschein, N., Hubbard, S., Beynon, R., Daran-Lapujade, P., Nielsen, J., & Teusink, B. (2022). Whole-cell modeling in yeast predicts compartment-specific proteome constraints that drive metabolic strategies. *Nature Communications*, *13*(1), 801.

Ewald, J. C. (2018). How yeast coordinates metabolism, growth and division. *Current Opinion in Microbiology*, *45*, 1–7.

Ewald, J. C., Kuehne, A., Zamboni, N., & Skotheim, J. M. (2016). The Yeast Cyclin-Dependent Kinase Routes Carbon Fluxes to Fuel Cell Cycle Progression. *Molecular Cell*, *62*(4), 532–545.

Famili, I., Förster, J., Nielsen, J., & Palsson, B. O. (2003). Saccharomyces cerevisiae phenotypes can be predicted by using constraint-based analysis of a genome-scale reconstructed metabolic network. *Proceedings of the National Academy of Sciences*, *100*(23), 13134–13139.

Feltham, J. E., Xi, S., Murray, S. C., Wouters, M., Urdiain-Arraiza, J., Heilig, R., George, C., Townley, A. F., Roberts, E., Kessler, B. M., Liberatori, S., Charles, P. D., Angel, A., Fischer, R., & Mellor, J. (2020). Transcriptional changes are regulated by metabolic pathway dynamics but decoupled from protein levels. *bioRxiv*, 833921.

FitzHugh, R. (1961). Impulses and Physiological States in Theoretical Models of Nerve Membrane. *Biophysical Journal*, *1*(6), 445–466.

Förster, J., Famili, I., Fu, P., Palsson, B. Ø., & Nielsen, J. (2003). Genome-Scale Reconstruction of the Saccharomyces cerevisiae Metabolic Network. *Genome Research*, *13*(2), 244–253.

Fortunato, S., & Barthélemy, M. (2007). Resolution limit in community detection. *Proceedings of the National Academy of Sciences*, *104*(1), 36–41.

Fulcher, B. D., & Jones, N. S. (2017). Hctsa: A Computational Framework for Automated Time-Series Phenotyping Using Massive Feature Extraction. *Cell Systems*, *5*(5), 527–531.e3.

Gaisne, M., Bécam, A.-M., Verdière, J., & Herbert, C. J. (1999). A 'natural' mutation in Saccharomyces cerevisiae strains derived from S288c affects the complex regulatory gene HAP1 (CYP1). *Current Genetics*, *36*(4), 195–200.

Garmendia-Torres, C., Tassy, O., Matifas, A., Molina, N., & Charvin, G. (2018). Multiple inputs ensure yeast cell size homeostasis during cell cycle progression (A. Musacchio, Ed.). *eLife*, *7*, e34025.

Gérard, C., & Goldbeter, A. (2012). Entrainment of the Mammalian Cell Cycle by the Circadian Clock: Modeling Two Coupled Cellular Rhythms. *PLOS Computational Biology*, *8*(5), e1002516.

Ghosh, A., & Chance, B. (1964). Oscillations of glycolytic intermediates in yeast cells. *Biochemical and Biophysical Research Communications*, *16*(2), 174–181.

Gillespie, D. T. (1977). Exact stochastic simulation of coupled chemical reactions. *The Journal of Physical Chemistry*, *81*(25), 2340–2361.

Gillespie, D. T. (2007). Stochastic Simulation of Chemical Kinetics. *Annual Review of Physical Chemistry*, *58*(1), 35–55.

Glynn, E. F., Chen, J., & Mushegian, A. R. (2006). Detecting periodic patterns in unevenly spaced gene expression time series using Lomb–Scargle periodograms. *Bioinformatics*, *22*(3), 310–316.

Goelzer, A., Fromion, V., & Scorletti, G. (2011). Cell design in bacteria as a convex optimization problem. *Automatica*, *47*(6), 1210–1218.

Goldbeter, A. (1991). A minimal cascade model for the mitotic oscillator involving cyclin and cdc2 kinase. *Proceedings of the National Academy of Sciences*, *88*(20), 9107–9111.

Goldbeter, A., & Lefever, R. (1972). Dissipative Structures for an Allosteric Model. *Biophysical Journal*, *12*(10), 1302–1315.

Goldbeter, A., & Yan, J. (2022). Multi-synchronization and other patterns of multi-rhythmicity in oscillatory biological systems. *Interface Focus*, *12*(3), 20210089.

Goodwin, B. C. (1965). Oscillatory behavior in enzymatic control processes. *Advances in Enzyme Regulation*, *3*, 425–437.

Gowans, G. J., Schep, A. N., Wong, K. M., King, D. A., Greenleaf, W. J., & Morrison, A. J. (2018). INO80 Chromatin Remodeling Coordinates Metabolic Homeostasis with Cell Division. *Cell Reports*, *22*(3), 611–623.

Granados, A. A., Crane, M. M., Montano-Gutierrez, L. F., Tanaka, R. J., Voliotis, M., & Swain, P. S. (2017). Distributing tasks via multiple input pathways increases cellular survival in stress (N. Barkai, Ed.). *eLife*, *6*, e21415.

Griffith, J. S. (1968). Mathematics of cellular control processes I. Negative feedback to one gene. *Journal of Theoretical Biology*, *20*(2), 202–208.

Gudipati, V., Koch, K., Lienhart, W.-D., & Macheroux, P. (2014). The flavoproteome of the yeast Saccharomyces cerevisiae. *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics*, *1844*(3), 535–544.

Hamnett, R., Crosby, P., Chesham, J. E., & Hastings, M. H. (2019). Vasoactive intestinal peptide controls the suprachiasmatic circadian clock network via ERK1/2 and DUSP4 signalling. *Nature Communications*, *10*(1), 542.

Heavner, B. D., Smallbone, K., Barker, B., Mendes, P., & Walker, L. P. (2012). Yeast 5 – an expanded reconstruction of the Saccharomyces cerevisiae metabolic network. *BMC Systems Biology*, *6*(1), 55.

Heavner, B. D., Smallbone, K., Price, N. D., & Walker, L. P. (2013). Version 6 of the consensus yeast metabolic network refines biochemical coverage and improves model performance. *Database*, *2013*, bat059.

Heltberg, M. L., Krishna, S., Kadanoff, L. P., & Jensen, M. H. (2021). A tale of two rhythms: Locked clocks and chaos in biology. *Cell Systems*, *12*(4), 291–303.

Herrgård, M. J., Swainston, N., Dobson, P., Dunn, W. B., Arga, K. Y., Arvas, M., Blüthgen, N., Borger, S., Costenoble, R., Heinemann, M., Hucka, M., Le Novère, N., Li, P., Liebermeister, W., Mo, M. L., Oliveira, A. P., Petranovic, D., Pettifer, S., Simeonidis, E., ... Kell, D. B. (2008). A consensus yeast metabolic network reconstruction obtained from a community approach to systems biology. *Nature Biotechnology*, *26*(10), 1155–1160.

Herskowitz, I. (1988). Life cycle of the budding yeast Saccharomyces cerevisiae. *Microbiological Reviews*, *52*(4), 536–553.

Higgins, J. (1964). A chemical mechanism for oscillation of glycolytic intermediates in yeast cells. *Proceedings of the National Academy of Sciences*, *51*(6), 989–994.

Ho, B., Baryshnikova, A., & Brown, G. W. (2018). Unification of Protein Abundance Datasets Yields a Quantitative Saccharomyces cerevisiae Proteome. *Cell Systems*, *6*(2), 192–205.e3.

Hochreiter, S., & Schmidhuber, J. (1997). Long Short-Term Memory. *Neural Computation*, *9*(8), 1735–1780.

Huang, D., Friesen, H., & Andrews, B. (2007). Pho85, a multifunctional cyclin-dependent protein kinase in budding yeast. *Molecular Microbiology*, *66*(2), 303–314.

Hughes, M. E., Hogenesch, J. B., & Kornacker, K. (2010). JTK_CYCLE: An Efficient Nonparametric Algorithm for Detecting Rhythmic Components in Genome-Scale Data Sets. *Journal of Biological Rhythms*, *25*(5), 372–380.

Hutchison, A. L., Maienschein-Cline, M., Chiang, A. H., Tabei, S. M. A., Gudjonson, H., Bahroos, N., Allada, R., & Dinner, A. R. (2015). Improved Statistical Methods Enable Greater Sensitivity in Rhythm Detection for Genome-Wide Data. *PLOS Computational Biology*, *11*(3), e1004094.

Jia, C., & Grima, R. (2020). Frequency domain analysis of fluctuations of mRNA and protein copy numbers within a cell lineage: Theory and experimental validation. *bioRxiv*, 2020.09.23.309724.

Jones, K. D., & Kompala, D. S. (1999). Cybernetic model of the growth dynamics of Saccharomyces cerevisiae in batch and continuous cultures. *Journal of Biotechnology*, *71*(1), 105–131.

Kanehisa, M., Furumichi, M., Sato, Y., Kawashima, M., & Ishiguro-Watanabe, M. (2023). KEGG for taxonomy-based analysis of pathways and genomes. *Nucleic Acids Research*, *51*(D1), D587–D592.

Kaspar von Meyenburg, H. (1969). Energetics of the budding cycle of Saccharomyces cerevisiae during glucose limited aerobic growth. *Archiv für Mikrobiologie*, *66*(4), 289–303.

Kendall, M. G. (1945). The Treatment of Ties in Ranking Problems. *Biometrika*, *33*(3), 239–251.

Kiviet, D. J., Nghe, P., Walker, N., Boulineau, S., Sunderlikova, V., & Tans, S. J. (2014). Stochasticity of metabolism and growth at the single-cell level. *Nature*, *514*(7522), 376–379.

Koch, K., Hromic, A., Sorokina, M., Strandback, E., Reisinger, M., Gruber, K., & Macheroux, P. (2017). Structure, biochemical and kinetic properties of recombinant Pst2p from Saccharomyces cerevisiae, a FMN-dependent NAD(P)H:quinone oxidoreductase. *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics*, *1865*(8), 1046–1056.

Krishna, S., & Laxman, S. (2018). A minimal “push–pull” bistability model explains oscillations between quiescent and proliferative cell states. *Molecular Biology of the Cell*, *29*(19), 2243–2258.

Kuang, Z., Pinglay, S., Ji, H., & Boeke, J. D. (2017). Msn2/4 regulate expression of glycolytic enzymes and control transition from quiescence to growth (N. Barkai, Ed.). *eLife*, *6*, e29938.

Kumpula, J. M., Saramäki, J., Kaski, K., & Kertész, J. (2007). Limited resolution in complex network community detection with Potts model approach. *The European Physical Journal B*, *56*(1), 41–45.

Kurat, C. F., Wolinski, H., Petschnigg, J., Kaluarachchi, S., Andrews, B., Natter, K., & Kohlwein, S. D. (2009). Cdk1/Cdc28-Dependent Activation of the Major Triacylglycerol Lipase Tgl4 in Yeast Links Lipolysis to Cell-Cycle Progression. *Molecular Cell*, *33*(1), 53–63.

Labhsetwar, P., Melo, M. C. R., Cole, J. A., & Luthey-Schulten, Z. (2017). Population FBA predicts metabolic phenotypes in yeast. *PLOS Computational Biology*, *13*(9), e1005728.

Lahtvee, P.-J., Sánchez, B. J., Smialowska, A., Kasvandik, S., Elsemman, I. E., Gatto, F., & Nielsen, J. (2017). Absolute Quantification of Protein and mRNA Abundances Demonstrate Variability in Gene-Specific Translation Efficiency in Yeast. *Cell Systems*, *4*(5), 495–504.e5.

Lambert, T. J. (2023). Using FPbase: The Fluorescent Protein Database. In M. Sharma (Ed.), *Fluorescent Proteins: Methods and Protocols* (pp. 1–45). Springer US.

Laxman, S., Sutter, B. M., & Tu, B. P. (2010). Behavior of a Metabolic Cycling Population at the Single Cell Level as Visualized by Fluorescent Gene Expression Reporters. *PLOS ONE*, *5*(9), e12595.

Li, C. M., & Klevecz, R. R. (2006). A rapid genome-scale response of the transcriptional oscillator to perturbation reveals a period-doubling path to phenotypic change. *Proceedings of the National Academy of Sciences*, *103*(44), 16254–16259.

Li, N., Schwartz, M., & Ionescu-Zanetti, C. (2009). PDMS Compound Adsorption in Context. *SLAS Discovery*, *14*(2), 194–202.

Lloyd, D. (2019). Saccharomyces cerevisiae: Oscillatory Orchestration of Growth. In T. Satyanarayana, S. K. Deshmukh & M. V. Deshpande (Eds.), *Advancing Frontiers in Mycology & Mycotechnology: Basic and Applied Aspects of Fungi* (pp. 181–214). Springer Singapore.

Lloyd, D., & Murray, D. B. (2005). Ultradian metronome: Timekeeper for orchestration of cellular coherence. *Trends in Biochemical Sciences*, *30*(7), 373–377.

Lloyd, D., & Murray, D. B. (2006). The temporal architecture of eukaryotic growth. *FEBS Letters*, *580*(12), 2830–2835.

Lloyd, D., & Murray, D. B. (2007). Redox rhythmicity: Clocks at the core of temporal coherence. *BioEssays*, *29*(5), 465–473.

Lomb, N. R. (1976). Least-squares frequency analysis of unequally spaced data. *Astrophysics and Space Science*, *39*(2), 447–462.

Lu, H., Li, F., Sánchez, B. J., Zhu, Z., Li, G., Domenzain, I., Marcišauskas, S., Anton, P. M., Lappa, D., Lieven, C., Beber, M. E., Sonnenschein, N., Kerkhoven, E. J., & Nielsen, J. (2019). A consensus S. cerevisiae metabolic model Yeast8 and its ecosystem for comprehensively probing cellular metabolism. *Nature Communications*, *10*(1), 3586.

Lu, Y., & Cross, F. R. (2010). Periodic Cyclin-Cdk Activity Entrains an Autonomous Cdc14 Release Oscillator. *Cell*, *141*(2), 268–279.

Lubba, C. H., Sethi, S. S., Knaute, P., Schultz, S. R., Fulcher, B. D., & Jones, N. S. (2019). Catch22: CAnonical Time-series CHaracteristics. *Data Mining and Knowledge Discovery*, *33*(6), 1821–1852.

Machado, A. K., Morgan, B. A., & Merrill, G. F. (1997). Thioredoxin Reductase-dependent Inhibition of MCB Cell Cycle Box Activity in Saccharomyces cerevisiae. *Journal of Biological Chemistry*, *272*(27), 17045–17054.

Machné, R., & Murray, D. B. (2012). The Yin and Yang of Yeast Transcription: Elements of a Global Feedback System between Metabolism and Chromatin. *PLOS ONE*, *7*(6), e37906.

Mahadevan, R., Edwards, J. S., & Doyle, F. J. (2002). Dynamic Flux Balance Analysis of Diauxic Growth in Escherichia coli. *Biophysical Journal*, *83*(3), 1331–1340.

Malumbres, M., Harlow, E., Hunt, T., Hunter, T., Lahti, J. M., Manning, G., Morgan, D. O., Tsai, L.-H., & Wolgemuth, D. J. (2009). Cyclin-dependent kinases: A family portrait. *Nature Cell Biology*, *11*(11), 1275–1276.

Maslanka, R., Kwolek-Mirek, M., & Zadrag-Tecza, R. (2018). Autofluorescence of yeast Saccharomyces cerevisiae cells caused by glucose metabolism products and its methodological implications. *Journal of Microbiological Methods*, *146*, 55–60.

Masters, B. R. (1994). Confocal Redox Imaging of Cells. In E. E. Bittar & L. Bo Chen (Eds.), *Advances in Molecular and Cell Biology* (pp. 1–19, Vol. 8). Elsevier.

McInnes, L., Healy, J., & Melville, J. (2020). UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction. *arXiv:1802.03426 [cs, stat]*.

Mellor, J. (2016). The molecular basis of metabolic cycles and their relationship to circadian rhythms. *Nature Structural & Molecular Biology*, *23*(12), 1035–1044.

Mewies, M., McIntire, W. S., & Scrutton, N. S. (1998). Covalent attachment of flavin adenine dinucleotide (FAD) and flavin mononucleotide (FMN) to enzymes: The current state of affairs. *Protein Science*, *7*(1), 7–20.

Millar, A. J. (2004). Input signals to the plant circadian clock. *Journal of Experimental Botany*, *55*(395), 277–283.

Minard, K. I., & McAlister-Henn, L. (2005). Sources of NADPH in Yeast Vary with Carbon Source. *Journal of Biological Chemistry*, *280*(48), 39890–39896.

Mo, M. L., Palsson, B. Ø., & Herrgård, M. J. (2009). Connecting extracellular metabolomic measurements to intracellular flux states in yeast. *BMC Systems Biology*, *3*(1), 37.

Mochan, E., & Pye, E. K. (1973). Respiratory Oscillations in Adapting Yeast Cultures. *Nature New Biology*, *242*(119), 177–179.

Morán, F., & Goldbeter, A. (1984). Onset of birhythmicity in a regulated biochemical system. *Biophysical Chemistry*, *20*(1), 149–156.

Mori, M., Hwa, T., Martin, O. C., Martino, A. D., & Marinari, E. (2016). Constrained Allocation Flux Balance Analysis. *PLOS Computational Biology*, *12*(6), e1004913.

Morrison, H. (2021). Chapter 14 - Dihydrolipoamide dehydrogenase. In H. Morrison (Ed.), *Enzyme Active Sites and their Reaction Mechanisms* (pp. 71–77). Academic Press.

Muñoz González, A. F. (2023). *Phenotyping single cells of Saccharomyces cerevisiae using an end-to-end analysis of high-content time-lapse microscopy* [Doctoral dissertation, University of Edinburgh].

Murray, A. W. (2004). Recycling the Cell Cycle: Cyclins Revisited. *Cell*, *116*(2), 221–234.

Murray, D. B., Haynes, K., & Tomita, M. (2011). Redox regulation in respiring Saccharomyces cerevisiae. *Biochimica et Biophysica Acta (BBA) - General Subjects*, *1810*(10), 945–958.

Newman, M. E. J. (2006). Modularity and community structure in networks. *Proceedings of the National Academy of Sciences*, *103*(23), 8577–8582.

Nijkamp, J. F., van den Broek, M., Datema, E., de Kok, S., Bosman, L., Luttik, M. A., Daran-Lapujade, P., Vongsangnak, W., Nielsen, J., Heijne, W. H., Klaassen, P., Paddon, C. J., Platt, D., Kötter, P., van Ham, R. C., Reinders, M. J., Pronk, J. T., de Ridder, D., & Daran, J.-M. (2012). De novo sequencing, assembly and analysis of the genome of the laboratory strain Saccharomyces cerevisiae CEN.PK113-7D, a model for modern industrial biotechnology. *Microbial Cell Factories*, *11*(1), 36.

Nilsson, A., & Nielsen, J. (2016). Metabolic Trade-offs in Yeast are Caused by F1F0-ATP synthase. *Scientific Reports*, *6*(1), 22264.

Nissen, T. L., Schulze, U., Nielsen, J., & Villadsen, J. (1997). Flux Distributions in Anaerobic, Glucose-Limited Continuous Cultures of Saccharomyces Cerevisiae. *Microbiology*, *143*(1), 203–218.

Nocetti, N., & Whitehouse, I. (2016). Nucleosome repositioning underlies dynamic gene expression. *Genes & Development*, *30*(6), 660–672.

Nogae, I., & Johnston, M. (1990). Isolation and characterization of the ZWF1 gene of Saccharomyces cerevisiae, encoding glucose-6-phosphate dehydrogenase. *Gene*, *96*(2), 161–169.

Nookaew, I., Jewett, M. C., Meechai, A., Thammarongtham, C., Laoteng, K., Cheevadhanarak, S., Nielsen, J., & Bhumiratana, S. (2008). The genome-scale metabolic model iIN800 of Saccharomyces cerevisiae and its validation: A scaffold to query lipid metabolism. *BMC Systems Biology*, *2*(1), 71.

Norsigian, C. J., Pusarla, N., McConn, J. L., Yurkovich, J. T., Dräger, A., Palsson, B. O., & King, Z. (2020). BiGG Models 2020: Multi-strain genome-scale models and expansion across the phylogenetic tree. *Nucleic Acids Research*, *48*(D1), D402–D406.

Nosoh, Y., & Takamiya, A. (1962). Synchronization of Budding Cycle in Yeast Cells, and Effect of Carbon Monoxide and Nitrogen-Deficiency on the Synchrony. *Plant and Cell Physiology*, *3*(1), 53–66.

Novak, B., & Tyson, J. J. (2022). Mitotic kinase oscillation governs the latching of cell cycle switches. *Current Biology*.

Nun, I., Protopapas, P., Sim, B., Zhu, M., Dave, R., Castro, N., & Pichara, K. (2015). FATS: Feature Analysis for Time Series.

Odat, O., Matta, S., Khalil, H., Kampranis, S. C., Pfau, R., Tsichlis, P. N., & Makris, A. M. (2007). Old Yellow Enzymes, Highly Homologous FMN Oxidoreductases with Modulating Roles in Oxidative Stress and Programmed Cell Death in Yeast. *Journal of Biological Chemistry*, *282*(49), 36010–36023.

Olsen, L. F., Andersen, A. Z., Lunding, A., Brasen, J. C., & Poulsen, A. K. (2009). Regulation of Glycolytic Oscillations by Mitochondrial and Plasma Membrane H+-ATPases. *Biophysical Journal*, *96*(9), 3850–3861.

Olsen, L. F., & Lunding, A. (2021). Oscillations in Yeast Glycolysis. In A. Stefanovska & P. V. E. McClintock (Eds.), *Physics of Biological Oscillators: New Insights into Non-Equilibrium and Non-Autonomous Systems* (pp. 211–224). Springer International Publishing.

O'Neill, E. M., Kaffman, A., Jolly, E. R., & O'Shea, E. K. (1996). Regulation of PHO4 Nuclear Localization by the PHO80-PHO85 Cyclin-CDK Complex. *Science*, *271*(5246), 209–212.

O'Neill, J. S., Hoyle, N. P., Robertson, J. B., Edgar, R. S., Beale, A. D., Peak-Chew, S. Y., Day, J., Costa, A. S. H., Frezza, C., & Causton, H. C. (2020). Eukaryotic cell biology is temporally coordinated to support the energetic demands of protein homeostasis. *Nature Communications*, *11*(1), 4706.

O'Neill, J. S., & Reddy, A. B. (2011). Circadian clocks in human red blood cells. *Nature*, *469*(7331), 498–503.

O'Neill, J. S., van Ooijen, G., Dixon, L. E., Troein, C., Corellou, F., Bouget, F.-Y., Reddy, A. B., & Millar, A. J. (2011). Circadian rhythms persist without transcription in a eukaryote. *Nature*, *469*(7331), 554–558.

Orlando, D. A., Lin, C. Y., Bernard, A., Wang, J. Y., Socolar, J. E. S., Iversen, E. S., Hartemink, A. J., & Haase, S. B. (2008). Global control of cell-cycle transcription by coupled CDK and network oscillators. *Nature*, *453*(7197), 944–947.

Orth, J. D., Thiele, I., & Palsson, B. Ø. (2010). What is flux balance analysis? *Nature Biotechnology*, *28*(3), 245–248.

Özsezen, S., Papagiannakis, A., Chen, H., Niebel, B., Milias-Argeitis, A., & Heinemann, M. (2019). Inference of the High-Level Interaction Topology between the Metabolic and Cell-Cycle Oscillators from Single-Cell Dynamics. *Cell Systems*.

Palsson, B. Ø. (2015). *Systems Biology: Constraint-based Reconstruction and Analysis*. Cambridge University Press.

Pang, S. S., Duggleby, R. G., & Guddat, L. W. (2002). Crystal structure of yeast acetohydroxyacid synthase: A target for herbicidal inhibitors. *Journal of Molecular Biology*, *317*(2), 249–262.

Papagiannakis, A., Niebel, B., Wit, E. C., & Heinemann, M. (2017). Autonomous Metabolic Oscillations Robustly Gate the Early and Late Cell Cycle. *Molecular Cell*, *65*(2), 285–295.

Patel, R. N. (Ed.). (2006). *Flavin-Containing Oxidative Biocatalysts*. CRC Press.

Pietsch, J. M., Munoz, A. F., Adjavon, D.-Y. A., Farquhar, I., Clark, I. B., & Swain, P. S. (2023). Determining growth rates from bright-field images of budding cells through identifying overlaps (A. M. Moses, Ed.). *eLife*, *12*, e79812.

Platt, J. (1999). Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods. *Advances in large margin classifiers*, *10*(3), 61–74.

Potvin-Trottier, L., Lord, N. D., Vinnicombe, G., & Paulsson, J. (2016). Synchronous long-term oscillations in a synthetic gene circuit. *Nature*, *538*(7626), 514–517.

Pronk, J. T. (2002). Auxotrophic Yeast Strains in Fundamental and Applied Research. *Applied and Environmental Microbiology*, *68*(5), 2095–2100.

Reichardt, J., & Bornholdt, S. (2004). Detecting Fuzzy Community Structures in Complex Networks with a Potts Model. *Physical Review Letters*, *93*(21), 218701.

Reimers, A.-M., Knoop, H., Bockmayr, A., & Steuer, R. (2017). Cellular trade-offs and optimal resource allocation during cyanobacterial diurnal growth. *Proceedings of the National Academy of Sciences*, *114*(31), E6457–E6465.

Salazar-Roa, M., & Malumbres, M. (2017). Fueling the Cell Division Cycle. *Trends in Cell Biology*, *27*(1), 69–81.

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 Systems Biology*, *13*(1), 4.

Sánchez, B. J., Zhang, C., Nilsson, A., Lahtvee, P.-J., Kerkhoven, E. J., & Nielsen, J. (2017). Improving the phenotype predictions of a yeast genome-scale metabolic model by incorporating enzymatic constraints. *Molecular Systems Biology*, *13*(8), 935.

Sasidharan, K., Tomita, M., Aon, M., Lloyd, D., & Murray, D. B. (2012). Time-Structure of the Yeast Metabolism In vivo. In I. I. Goryanin & A. B. Goryachev (Eds.), *Advances in Systems Biology* (pp. 359–379). Springer.

Satoh, T., Horie, M., Watanabe, H., Tsuchiya, Y., & Kamei, T. (1993). Enzymatic Properties of Squalene Epoxidase from Saccharomyces cerevisiae. *Biological & Pharmaceutical Bulletin*, *16*(4), 349–352.

Satroutdinov, A. D., Kuriyama, H., & Kobayashi, H. (1992). Oscillatory metabolism of Saccharomyces cerevisiae in continuous culture. *FEMS Microbiology Letters*, *98*(1-3), 261–267.

Scargle, J. D. (1982). Studies in astronomical time series analysis. II - Statistical aspects of spectral analysis of unevenly spaced data. *The Astrophysical Journal*, *263*, 835–853.

Schuetz, R., Zamboni, N., Zampieri, M., Heinemann, M., & Sauer, U. (2012). Multidimensional Optimality of Microbial Metabolism. *Science*, *336*(6081), 601–604.

Schuster, A. (1898). On the investigation of hidden periodicities with application to a supposed 26 day period of meteorological phenomena. *Terrestrial Magnetism*, *3*(1), 13–41.

Scott, M., & Hwa, T. (2011). Bacterial growth laws and their applications. *Current Opinion in Biotechnology*, *22*(4), 559–565.

Seeger, M., Flöttmann, M., & Klipp, E. (2023). A dynamical stochastic model of yeast translation across the cell cycle. *Heliyon*, *9*(2), e13101.

Shafiei, G., Zeighami, Y., Clark, C. A., Coull, J. T., Nagano-Saito, A., Leyton, M., Dagher, A., & Mišić, B. (2019). Dopamine Signaling Modulates the Stability and Integration of Intrinsic Brain Networks. *Cerebral Cortex*, *29*(1), 397–409.

Shah, P., Ding, Y., Niemczyk, M., Kudla, G., & Plotkin, J. B. (2013). Rate-Limiting Steps in Yeast Protein Translation. *Cell*, *153*(7), 1589–1601.

Sherman, F. (2002). Getting started with yeast. In C. Guthrie & G. R. Fink (Eds.), *Methods in Enzymology* (pp. 3–41, Vol. 350). Academic Press.

Siano, S. A., & Mutharasan, R. (1989). NADH and flavin fluorescence responses of starved yeast cultures to substrate additions. *Biotechnology and Bioengineering*, *34*(5), 660–670.

Silverman, S. J., Petti, A. A., Slavov, N., Parsons, L., Briehof, R., Thiberge, S. Y., Zenklusen, D., Gandhi, S. J., Larson, D. R., Singer, R. H., & Botstein, D. (2010). Metabolic cycling in single yeast cells from unsynchronized steady-state populations limited on glucose or phosphate. *Proceedings of the National Academy of Sciences*, *107*(15), 6946–6951.

Singh, K., Graf, B., Linden, A., Sautner, V., Urlaub, H., Tittmann, K., Stark, H., & Chari, A. (2020). Discovery of a Regulatory Subunit of the Yeast Fatty Acid Synthase. *Cell*, *180*(6), 1130–1143.e20.

Slavov, N., & Botstein, D. (2011). Coupling among growth rate response, metabolic cycle, and cell division cycle in yeast. *Molecular Biology of the Cell*, *22*(12), 1997–2009.

Slavov, N., Macinskas, J., Caudy, A., & Botstein, D. (2011). Metabolic cycling without cell division cycling in respiring yeast. *Proceedings of the National Academy of Sciences*, *108*(47), 19090–19095.

Smith, H. L., & Waltman, P. (1995). *The Theory of the Chemostat: Dynamics of Microbial Competition*. Cambridge University Press.

Takhaveev, V., Özsezen, S., Smith, E. N., Zylstra, A., Chaillet, M. L., Chen, H., Papagiannakis, A., Milias-Argeitis, A., & Heinemann, M. (2023). Temporal segregation of biosynthetic processes is responsible for metabolic oscillations during the budding yeast cell cycle. *Nature Metabolism*, *5*(2), 294–313.

Termonia, Y., & Ross, J. (1981). Oscillations and control features in glycolysis: Numerical analysis of a comprehensive model. *Proceedings of the National Academy of Sciences*, *78*(5), 2952–2956.

The UniProt Consortium. (2023). UniProt: The Universal Protein Knowledgebase in 2023. *Nucleic Acids Research*, *51*(D1), D523–D531.

Thiele, I., & Palsson, B. Ø. (2010). A protocol for generating a high-quality genome-scale metabolic reconstruction. *Nature Protocols*, *5*(1), 93–121.

Thoke, H. S., Olsen, L. F., Duelund, L., Stock, R. P., Heimburg, T., & Bagatolli, L. A. (2018). Is a constant low-entropy process at the root of glycolytic oscillations? *Journal of Biological Physics*, *44*(3), 419–431.

Toepke, M. W., & Beebe, D. J. (2006). PDMS absorption of small molecules and consequences in microfluidic applications. *Lab on a Chip*, *6*(12), 1484–1486.

Tourigny, D. S., Goldberg, A. P., & Karr, J. R. (2021). Simulating single-cell metabolism using a stochastic flux-balance analysis algorithm. *Biophysical Journal*, *120*(23), 5231–5242.

Traag, V. A., Van Dooren, P., & Nesterov, Y. (2011). Narrow scope for resolution-limit-free community detection. *Physical Review E*, *84*(1), 016114.

Traag, V. A., Waltman, L., & van Eck, N. J. (2019). From Louvain to Leiden: Guaranteeing well-connected communities. *Scientific Reports*, *9*(1), 5233.

Tsuge, H., Ozeki, K., Sen-maru, K., & Ohashi, K. (1979). Purification and Properties of Pyridoxamine (Pyridoxine) 5′-Phosphate Oxidase from Baker's Yeast. *Agricultural and Biological Chemistry*, *43*(9), 1801–1807.

Tu, B. P., Kudlicki, A., Rowicka, M., & McKnight, S. L. (2005). Logic of the Yeast Metabolic Cycle: Temporal Compartmentalization of Cellular Processes. *Science*, *310*(5751), 1152–1158.

Tu, B. P., Mohler, R. E., Liu, J. C., Dombek, K. M., Young, E. T., Synovec, R. E., & McKnight, S. L. (2007). Cyclic changes in metabolic state during the life of a yeast cell. *Proceedings of the National Academy of Sciences*, *104*(43), 16886–16891.

VanderPlas, J. T. (2018). Understanding the Lomb-Scargle Periodogram. *The Astrophysical Journal Supplement Series*, *236*(1), 16.

Verduyn, C., Postma, E., Scheffers, W. A., & Dijken, J. P. V. (1992). Effect of benzoic acid on metabolic fluxes in yeasts: A continuous-culture study on the regulation of respiration and alcoholic fermentation. *Yeast*, *8*(7), 501–517.

Wagnieres, G. A., Star, W. M., & Wilson, B. C. (1998). In Vivo Fluorescence Spectroscopy and Imaging for Oncological Applications. *Photochemistry and Photobiology*, *68*(5), 603–632.

Wang, X., Wirth, A., & Wang, L. (2007). Structure-Based Statistical Features and Multivariate Time Series Clustering. *Seventh IEEE International Conference on Data Mining (ICDM 2007)*, 351–360.

Wang, X., Xia, K., Yang, X., & Tang, C. (2019). Growth strategy of microbes on mixed carbon sources. *Nature Communications*, *10*(1), 1279.

Watcharawipas, A., Watanabe, D., & Takagi, H. (2018). Sodium Acetate Responses in Saccharomyces cerevisiae and the Ubiquitin Ligase Rsp5. *Frontiers in Microbiology*, *9*.

Weiße, A. Y., Oyarzún, D. A., Danos, V., & Swain, P. S. (2015). Mechanistic links between cellular trade-offs, gene expression, and growth. *Proceedings of the National Academy of Sciences*, *112*(9), E1038–E1047.

Wilson, W. A., Mahrenholz, A. M., & Roach, P. J. (1999). Substrate Targeting of the Yeast Cyclin-Dependent Kinase Pho85p by the Cyclin Pcl10p. *Molecular and Cellular Biology*, *19*(10), 7020–7030.

Winston, F., Dollard, C., & Ricupero-Hovasse, S. L. (1995). Construction of a set of convenient saccharomyces cerevisiae strains that are isogenic to S288C. *Yeast*, *11*(1), 53–55.

Wong, C.-M., Zhou, Y., Ng, R. W. M., Kung, H.-f., & Jin, D.-Y. (2002). Cooperation of Yeast Peroxiredoxins Tsa1p and Tsa2p in the Cellular Defense against Oxidative and Nitrosative Stress. *Journal of Biological Chemistry*, *277*(7), 5385–5394.

Yang, L., Ebrahim, A., Lloyd, C. J., Saunders, M. A., & Palsson, B. O. (2019). DynamicME: Dynamic simulation and refinement of integrated models of metabolism and protein expression. *BMC Systems Biology*, *13*(1), 2.

Zameitat, E., Pierik, A. J., Zocher, K., & Löffler, M. (2007). Dihydroorotate dehydrogenase from Saccharomyces cerevisiae: Spectroscopic investigations with the recombinant enzyme throw light on catalytic properties and metabolism of fumarate analogues. *FEMS Yeast Research*, *7*(6), 897–904.

Zhang, Z., Milias-Argeitis, A., & Heinemann, M. (2018). Dynamic single-cell NAD(P)H measurement reveals oscillatory metabolism throughout the E. coli cell division cycle. *Scientific Reports*, *8*(1), 1–10.

Zhu, B. (2022). Logic of the temporal compartmentalization of the hepatic metabolic cycle. *Physiology*.

Zieliński, T., Hay, J., & Millar, A. J. (2022). Period Estimation and Rhythm Detection in Timeseries Data Using BioDare2, the Free, Online, Community Resource. In D. Staiger, S. Davis & A. M. Davis (Eds.), *Plant Circadian Networks: Methods and Protocols* (pp. 15–32). Springer US.

Zieliński, T., Moore, A. M., Troup, E., Halliday, K. J., & Millar, A. J. (2014). Strengths and Limitations of Period Estimation Methods for Circadian Data. *PLOS ONE*, *9*(5), e96462.

Zylstra, A., & Heinemann, M. (2022). Metabolic dynamics during the cell cycle. *Current Opinion in Systems Biology*, 100415.