This thesis has been submitted in fulfilment of the requirements for a postgraduate degree (e. g. PhD, MPhil, DClinPsychol) at the University of Edinburgh. Please note the following terms and conditions of use:

- This work is protected by copyright and other intellectual property rights, which are retained by the thesis author, unless otherwise stated.
- A copy can be downloaded for personal non-commercial research or study, without prior permission or charge.
- This thesis cannot be reproduced or quoted extensively from without first obtaining permission in writing from the author.
- The content must not be changed in any way or sold commercially in any format or medium without the formal permission of the author.
- When referring to this work, full bibliographic details including the author, title, awarding institution and date of the thesis must be given.
The development of high throughput metabolomics to aid the synthetic biology ‘design-build-test-learn’ cycle

Georgie Barrett

Doctor of Philosophy

Institute of Quantitative Biology, Biochemistry and Biotechnology

The University of Edinburgh

2023
“Let us choose for ourselves our path in life, and let us try to strew that path with flowers.”

— Émilie Du Châtelet
Abstract

Metabolomics offers a comprehensive insight to metabolite profiles within biological systems. It often uses mass spectrometry as an analytical technique, enabling detection with high sensitivity and selectivity. Metabolomics has been successfully utilized to study a range of fields, including synthetic biology. The basis of synthetic biology is the design and construction of new biological systems, and the re-design of existing natural systems for a specific purpose. Due to its complexity, synthetic biology often requires several iterations of the ‘design-build-test-learn’ cycle to optimise a bioprocess. Despite major advances in strain engineering throughput, there are few publications demonstrating the same improvement in strain analysis. Therefore, due to lengthy analytical techniques, the ‘test’ phase remains a bottleneck. To explore this under-developed area, a combination of high throughput metabolomics and low-cost automation were used for strain optimisation.

High throughput metabolomics was implemented using a targeted method on a triple quadrupole instrument using flow injection. This method was optimised using multiple reaction monitoring, allowing 36 metabolites in the pathways of interest to be identified. This method has a short run time of 2 minutes per sample, enabling a full 96 well plate to be run in less than 4 hours. A liquid handling robot, built from LEGO® and programmed using MINDSTORMS® software, was implemented and illustrates that low-cost automation is a viable option to increase throughput.

As an example of strain optimisation, the metabolism of *Escherichia coli* NST74 was examined. This strain is capable of overproducing L-phenylalanine due to feedback de-regulation. Therefore, eight inhibitors were used to probe the metabolism of this strain, with the aim to increase the yield of L-phenylalanine and provide potential targets for further genetic modification. One inhibitor, L-tryptophan - a feedback inhibitor of anthranilate synthase - increased the yield of extracellular L-phenylalanine by five-fold compared to the control, whilst another, L-tyrosine, increased the yield by seven-fold, also through feedback inhibition. This illustrates successful yield optimisation.
In summary, this research aimed to reduce the bottleneck of the synthetic biology ‘design-build-test-learn’ cycle. Automation and high throughput mass spectrometry methods were implemented and decreased the time needed for sample preparation and data acquisition. Optimisation of the strain resulted in a seven-fold increase in the desired product. The combination of high throughput metabolomics and low-cost automation were used for strain optimisation successfully, representing a starting point for a high throughput platform for strain optimisation, and rapid testing using metabolomics.
Synthetic biology redesigns organisms to have useful purposes, by engineering them to have new abilities. This involves genetic engineering of bacteria, yeast and mammalian cells, and can help solve problems in many fields, such as medicine, agriculture and the textile industry. However, due to the complexity of synthetic biology, the process from start to finish can be a lengthy one. The work conducted here aims to make the process more streamlined and increase throughput. This was achieved using a strain of bacteria, *E. coli*, as an example organism.

One analytical technique that can be used is mass spectrometry, which allows the measurement of hundreds of molecules. An example of molecules that can be analysed are metabolites. These are small and dynamic, meaning they respond to their environment, and therefore analysis provides a ‘snapshot’ of what is happening at a particular moment in time. Another platform that helps is automation, as this can increase the speed and reproducibility of experiments. Liquid handling robots are usually used, which can be expensive.

In this thesis, high throughput mass spectrometry methods were developed and used on samples from *E. coli* NST74. This strain of *E. coli* has been genetically engineered to overproduce L-phenylalanine, an amino acid. Parts of the metabolism were restricted using inhibitors, to see if this would lead to an increase of the desired product. Ultimately, it was possible to increase the yield of L-phenylalanine by seven-fold; therefore, this work provides targets for further genetic modifications.

As automation is typically expensive, low-cost automation was explored to see if this could be a useful addition to a synthetic biology lab. This was done using a liquid handling robot, built from LEGO® pieces. This robot was capable of running useful experiments, and therefore low-cost automation could increase accessibility to liquid handling robots in the lab. Overall, this work demonstrates a platform consisting of a high throughput mass spectrometry method, low-cost automation, and inhibitor panel, employed to increase the speed of progress within the field of synthetic biology. Further, this platform shows proof of concept of strain optimisation with *E. coli* NST74 and the increase of yield in L-phenylalanine by seven-fold.
I would firstly like to thank my supervisor, Dr Karl Burgess, for all his help and understanding during my academic journey. Karl, you helped grow my love of mass spectrometry and I will forever be grateful that I stumbled across your lab and ended up doing a Masters project and then a PhD with you. Next, to the members of my PhD committee, Professor Susan Rosser, Professor Maurice Gallagher, and Professor Malcolm Walkinshaw; thank you for your guidance throughout this project. A special thank you to Professor Stephen Wallace and his group, particularly Annemette Kjeldsen, for all the help with the styrene work.

My heartfelt gratitude to the past and present members of the Burgess Group, and EdinOmics. In no particular order: Martina, Joan, Luke, Ricardo, Dasha, Bart, Aya, Tessa, Lisa, Andy, Fraser, Yumali and Jessica. Your knowledge and enthusiasm made my PhD journey easier, and a lot more enjoyable. A special thanks to Jess Tallis, for always answering my questions, about anything and everything!

This would not have been possible without the support of all the people mentioned below. Thank you all for listening to me talk constantly about mass spectrometry and metabolomics when you had no idea what I was talking about.

To my Mum and Dad, thank you for inspiring my curiosity and giving me the courage to try. To my brother Charlie, I am so proud of you and your strength. To Anna, thank you for the countless cups of tea, and endless patience. And to Amy, for helping me believe that I could do this. To my other friends, thank you for all the laughs and support.

To Blade, Beau, Bramble and Bumble, the four-legged friends who kept me smiling and active during the stresses of a PhD.

Finally, I dedicate this thesis to my best friend, my Nan, Sylvia Lindsay. You always told me to just try my best, well Nan, I did.
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.

Georgie Barrett

28th September 2023
## Contents

Abstract ........................................................................................................................................ iii
Lay Summary ...................................................................................................................................... v
Acknowledgements ........................................................................................................................ vi
Declaration ......................................................................................................................................... vii
Contents ........................................................................................................................................... viii
Figures and Tables ............................................................................................................................ xv
   Figures ........................................................................................................................................... xv
   Tables ............................................................................................................................................. xx
List of Abbreviations ........................................................................................................................... xxiii
Chapter 1. ............................................................................................................................................... 1
Introduction ......................................................................................................................................... 1
   1.1 Metabolomics .............................................................................................................................. 1
   1.2 Separation techniques ................................................................................................................ 5
      1.2.1 Liquid chromatography ...................................................................................................... 5
      1.2.2 Gas Chromatography ......................................................................................................... 6
   1.3 Alternatives to prior separation .................................................................................................. 7
      1.3.1 Flow injection ................................................................................................................... 7
      1.3.2 Direct infusion/injection ..................................................................................................... 7
   1.4 Mass spectrometry ...................................................................................................................... 8
      1.4.1 Ion sources ....................................................................................................................... 8
      Electron Ionisation ..................................................................................................................... 8
      MALDI ....................................................................................................................................... 8
      Atmospheric pressure ionisation methods .................................................................................... 9
      Electrospray ionisation ............................................................................................................... 9
   1.4.2 Mass analysers ...................................................................................................................... 11
      Quadrupoles ............................................................................................................................. 11
      Time of Flight ............................................................................................................................ 12
      Ion trap .................................................................................................................................... 13
      Orbitrap ..................................................................................................................................... 14
   1.4.3 Hybrid instruments ............................................................................................................... 15
      Triple quadrupole ...................................................................................................................... 15
      Ion mobility quadrupole-time of flight ...................................................................................... 16
      Other ......................................................................................................................................... 17
   1.5 Synthetic biology and the ‘design-build-test-learn’ cycle ......................................................... 17
   1.6 Inhibitors ................................................................................................................................... 20
1.6.1 Inhibitors as product enhancers ............................................... 22
1.7 Aims ......................................................................................... 23
Chapter 2. ................................................................................... 24
Materials and Methods ................................................................. 24
2.1 Reagents ................................................................................... 24
2.2 Bacterial strains ....................................................................... 24
2.3 Growth media .......................................................................... 25
2.4 Culture conditions .................................................................... 25
2.5 OD_{600} measurement ............................................................... 26
2.5.1 Flask cultures ................................................................. 26
2.5.2 96 well plate cultures ....................................................... 26
2.6 Metabolite extraction ............................................................... 26
2.6.1 Flask cultures ................................................................. 26
Intracellular metabolite extraction .............................................. 26
Extracellular metabolite extraction .............................................. 27
2.6.2 96 well plate cultures ....................................................... 27
Intracellular metabolite extraction .............................................. 28
Extracellular metabolite extraction .............................................. 28
2.7 Targeted mass spectrometry .................................................... 28
2.7.1 Flow injection .................................................................. 28
2.7.2 Instrument parameters ...................................................... 29
2.7.3 Data processing and analysis ............................................. 29
2.8 Untargeted mass spectrometry ................................................ 31
2.8.1 Data acquisition ............................................................. 31
2.8.2 Data processing .............................................................. 31
2.8.3 Data analysis .................................................................. 32
Chapter 3. ................................................................................... 33
High throughput mass spectrometry using flow injection on a triple quadrupole mass spectrometer ......................................................... 33
3.1 Introduction ........................................................................... 33
3.2 Aims ....................................................................................... 38
3.3 Methods ................................................................................ 38
3.3.1 Initial flow injection mass spectrometry set up ................... 38
3.3.2 Optimisation .................................................................... 39
Solvent system selection ............................................................ 39
Extraction solvents .................................................................... 40
MRMs ......................................................................................... 40
Low-cost LEGO® automation to aid the synthetic biology ‘design-build-test-learn’ cycle .......................................................... 55

4.1 Introduction .................................................................................................................. 55
  4.1.1 Background ......................................................................................................... 55
  4.1.2 Frugal science .................................................................................................. 58
  4.1.3 LEGO® robotics in STEM .............................................................................. 59
  4.1.4 Programming robots .................................................................................... 60

4.2 Aims ............................................................................................................................ 63

4.3 Method ........................................................................................................................ 63
  4.3.1 Building .............................................................................................................. 63
  4.3.2 Programming .................................................................................................... 68
  Serial dilution ............................................................................................................. 68
  Different cell culture conditions .............................................................................. 68
  Metabolite extraction .............................................................................................. 69

4.4 Results ........................................................................................................................ 70
  4.4.1 Uses .................................................................................................................... 70
  L-phenylalanine dilution curve ............................................................................... 70
  Different cell culture conditions and metabolite extractions .................................. 73

4.5 Discussion .................................................................................................................... 73
  4.5.1 Build .................................................................................................................. 75
  Comparison to Gerber et al. robot .......................................................................... 76
  Comparison to other open-source robots .................................................................. 76
  4.5.2 Electronics and Programming ........................................................................ 80
Chapter 5. Synthetic biology for styrene production

5.1 Introduction

5.1.1 Styrene background

5.1.2 Styrene production

5.1.3 Bioplastics

5.1.4 Biological styrene production

5.1.5 E. coli as a chassis system

5.1.6 E. coli NST74 with a styrene producing plasmid

5.1.7 Shikimate pathway

E. coli NST74

Styrene producing plasmid

5.2 Aims

5.3 Methods

5.3.1 No induction metabolism investigation

5.3.2 Stable isotope labelling

5.3.3 Fast and slow induction

5.3.4 Data processing and data analysis

5.3.5 Culturing

5.3.6 Flask culture metabolite extraction

5.3.7 IM-q-TOF acquisition, data processing and analysis

4.6 Conclusion
Using an inhibitor to alter the flux of L-phenylalanine in an engineered E. coli strain

6.1 Introduction.................................................................................................................. 124
   6.1.1 E. coli NST74 ........................................................................................................ 124
   6.1.2 Importance of L-phenylalanine............................................................................. 130
   6.1.3 Using inhibitors to alter metabolism ................................................................. 130
   6.1.4 Targets to inhibit ............................................................................................... 131
6.2 Aims .................................................................................................................................. 132
6.3 Method ............................................................................................................................ 132
   6.3.1 Preliminary experiment ...................................................................................... 132
   6.3.2 Addition of L-tryptophan to cell culture in 96 well plates ............................... 133
   6.3.3 Addition of L-tryptophan with time delay ......................................................... 135
6.4 Results and Discussion ................................................................................................. 135
   6.4.1 Preliminary experiment ...................................................................................... 135
   6.4.2 Culturing with added L-tryptophan – Targeted MS ....................................... 138
      Growth curves .......................................................................................................... 139

Culturing ............................................................................................................................. 102
Gas Chromatography Quadrupole Time-Of-Flight Mass Spectrometry .................................. 103
Data analysis ........................................................................................................................ 104
5.3.4 24 and 48-hour induction ...................................................................................... 104
Culturing ............................................................................................................................. 104
Gas Chromatography Quadrupole Time-Of-Flight Mass Spectrometry .................................. 105
Data analysis ........................................................................................................................ 105
5.4 Results ............................................................................................................................ 105
   5.4.1 No induction metabolism investigation ......................................................... 105
   5.4.2 Stable isotope labelling .................................................................................... 112
   5.4.3 Fast and slow induction .................................................................................... 114
   5.4.4 24 and 48-hour induction ................................................................................ 115
5.5 Discussion ....................................................................................................................... 118
   5.5.1 IM-Q-TOF results ............................................................................................. 118
      Culturing three strains, no induction ................................................................. 118
      Stable isotope labelling ....................................................................................... 119
   5.5.2 GC-MS results .................................................................................................. 120
      Fast and Slow Induction ...................................................................................... 120
      24 and 48-hour induction .................................................................................... 120
5.6 Conclusion ...................................................................................................................... 121

Chapter 6 .............................................................................................................................. 124
Exploring a panel of inhibitors on an engineered *E. coli* strain ................................................................. 181

7.1 Introduction ................................................................................................................................. 181
  7.1.1 Anthranilate synthase ............................................................................................................. 182
  7.1.2 Isochorismate synthase ........................................................................................................ 182
  7.1.3 Chorismate pyruvate lyase ................................................................................................... 183
  7.1.4 Aminodeoxychorismate synthase ......................................................................................... 183
  7.1.5 Prephenate dehydrogenase ................................................................................................ 184

7.2 Aims ........................................................................................................................................... 185

7.3 Method ..................................................................................................................................... 185
  7.3.1 Cell culture ............................................................................................................................ 185
  7.3.2 Addition of inhibitors ............................................................................................................ 185
  7.3.3 Metabolite extraction ............................................................................................................. 186
  7.3.4 Targeted Mass spectrometry acquisition, data processing and analysis ....................... 186

7.4 Results and Discussion ............................................................................................................. 186
  7.4.1 Enzyme target: Anthranilate synthase, Inhibitor: Anthranilic acid ........................................ 186
  Growth curves ............................................................................................................................... 187
  L-phenylalanine ............................................................................................................................ 188
Figures

All figures were created using Biorender.com or GraphPad Prism version 8.4.2.

Chapter 1. Introduction

Figure 1. 1. The central dogma of molecular biology. ............................................. 1
Figure 1. 2. The basic components of a mass spectrometer. ........................................ 4
Figure 1. 3. Gas chromatography mass spectrometry. ................................................. 6
Figure 1. 4. Electrospray ionisation model. ................................................................. 10
Figure 1. 5. Quardupole mass analyser ................................................................. 12
Figure 1. 6. Time of Flight mass analyser. ................................................................. 13
Figure 1. 7. Ion trap mass analyser. ................................................................. 14
Figure 1. 8. Orbitrap mass analyser. ................................................................. 15
Figure 1. 9. Triple quadrupole instrument ................................................................. 16
Figure 1. 10. Schematic of the Agilent IM-q-TOF mass spectrometer. ....................... 17
Figure 1. 11. ‘Design-Build-Test-Learn’ cycle .................................................. 19
Figure 1. 12. Vmax and Km ................................................................................. 21

Chapter 2. Materials and Methods

Figure 2. 1. Targeted mass spectrometry workflow ................................................. 30
Chapter 3. High throughput mass spectrometry using flow injection on a triple quadrupole mass spectrometer

Figure 3. 1. Flow injection mass spectrometry ................................................................. 34
Figure 3. 2. MRM scanning on a triple quadrupole instrument .................................. 36
Figure 3. 3. Factors for analytical platform choice ......................................................... 37
Figure 3. 4. Short tubing total ion chromatograms ....................................................... 42
Figure 3. 5. Long tubing total ion chromatograms ......................................................... 43
Figure 3. 6. Mobile phase solvents ............................................................................... 45
Figure 3. 7. Reproducibility of the flow injection method on six metabolites .......... 51

Chapter 4. Low-cost LEGO® automation to aid the synthetic biology cycle

Figure 4. 1. Programming for serial dilutions on different software ......................... 62
Figure 4. 2. The LEGO® 2D robot ................................................................................ 64
Figure 4. 3. Set up of 96 well plate for different cell culture conditions ................... 69
Figure 4. 4. Programming for serial dilution in 96 well plate ................................... 71
Figure 4. 5. Serial dilutions - LEGO® robot and manual pipetting ......................... 72
Figure 4. 6. Individual replicates for LEGO® robot serial dilutions ......................... 73

Chapter 5. Synthetic biology for styrene production

Figure 5. 1. The structure of styrene ............................................................................. 89
Figure 5. 2. Industrial styrene production ..................................................................... 91
Figure 5. 3. Reactions taking place during styrene production ................................. 92
Figure 5. 4. *E. coli* NST74 with the pTpal-fdc plasmid................................. 97
Figure 5. 5. The biosynthetic pathway of styrene............................................. 99
Figure 5. 6. Metabolite extraction using 100% ACN.......................................... 107
Figure 5. 7. Metabolite extraction using 100% MeOH........................................ 108
Figure 5. 8. L-phenylalanine area peak for negative mode IM-q-TOF data............ 110
Figure 5. 9. L-phenylalanine area peak for positive mode IM-q-TOF data............. 111
Figure 5. 10. $^{12}$C and $^{13}$C glucose in *E. coli* NST74 samples ....................... 113
Figure 5. 11. Styrene peak with fast and slow induction methods......................... 115
Figure 5. 12. Styrene standard curve.............................................................. 116
Figure 5. 13. Styrene area peak after 24 and 48 hours...................................... 117
Figure 5. 14. Biobased plastics............................................................................ 123

Chapter 6. Using an inhibitor to alter the flux of L-phenylalanine in an engineered *E. coli* strain

Figure 6. 1. L-phenylalanine pathway.................................................................... 129
Figure 6. 2. L-phenylalanine in *E. coli* NST74 grown with added L-tryptophan..... 136
Figure 6. 3. Growth curves of *E. coli* NST74 with added L-tryptophan............... 139
Figure 6. 4. Extracellular L-phenylalanine with added L-tryptophan..................... 141
Figure 6. 5. Intracellular L-phenylalanine with added L-tryptophan...................... 143
Figure 6. 6. Extracellular metabolites of *E. coli* NST74 with added L-tryptophan... 144
Figure 6. 7. Extracellular and intracellular L-tryptophan peak area...................... 145
Figure 6. 8. IM-q-TOF extracellular negative metabolites...................................... 151
Figure 6. 9. PCA graphs for extracellular negative metabolites............................ 153
Figure 6. 10. IM-q-TOF extracellular positive metabolites.................................... 154
Figure 6.11. IM-q-TOF extracellular positive metabolites (continued) ...................... 155
Figure 6.12. PCA graphs for extracellular positive metabolites ....................... 157
Figure 6.13. IM-q-TOF intracellular negative metabolites ............................. 161
Figure 6.14. IM-q-TOF intracellular negative metabolites (continued) ............. 162
Figure 6.15. PCA graphs for intracellular negative metabolites ...................... 163
Figure 6.16. IM-q-TOF intracellular positive metabolites ............................... 164
Figure 6.17. IM-q-TOF intracellular positive metabolites (continued) ............. 165
Figure 6.18. PCA graphs for intracellular positive metabolites ...................... 166
Figure 6.19. L-phenylalanine IM-q-TOF data ............................................. 168
Figure 6.20. Untargeted metabolic pathways shown to be effected by the addition of the L-tryptophan ........................................................................................................ 175
Figure 6.21. L-phenylalanine with L-tryptophan added at stationary phase .. .... 177

Chapter 7. Exploring a panel of inhibitors on an engineered E. coli strain

Figure 7.1. Growth curves and L-phenylalanine for E. coli NST74 with added anthranilic acid ...................................................................................................................... 187
Figure 7.2. Extracellular metabolites in E. coli NST74 with added anthranilic acid 190
Figure 7.3. A pathway diagram for L-phenylalanine, L-tyrosine, and L-glutamate. 191
Figure 7.4. Growth curves and L-phenylalanine for E. coli NST74 with added methyl anthranilate ........................................................................................................ 193
Figure 7.5. Growth curves for E. coli NST74 with different concentrations of added NEM ....................................................................................................................... 197
Figure 7.6. L-phenylalanine in E. coli NST74 with added NEM ....................... 198
Figure 7.7. Growth curves for E. coli NST74 with added vanillic acid .............. 200
Figure 7. 8. L-phenylalanine in E. coli NST74 with added vanillic acid.............. 201
Figure 7. 9. Growth curves for E. coli NST74 with added nickel sulphate. ........... 203
Figure 7. 10. L-phenylalanine in E. coli NST74 with added nickel sulphate. ....... 204
Figure 7. 11. Growth curves for E. coli NST74 with added L-tyrosine ................. 206
Figure 7. 12. L-phenylalanine in E. coli NST74 with added L-tyrosine, all samples including media controls................................................................. 207
Figure 7. 13. Growth curves for media only controls for E. coli NST74 with added L-tyrosine........................................................................................................ 208
Figure 7. 14. Extracellular L-phenylalanine in E. coli NST74 with added L-tyrosine ......................................................................................................................... 209
Figure 7. 15. Intracellular L-phenylalanine in E. coli NST74 with added L-tyrosine.211
Figure 7. 16. Growth curves for E. coli NST74 with added diethyl malonate........ 213
Figure 7. 17. L-phenylalanine in E. coli NST74 with added diethyl malonate........ 214
Figure 7. 18. Extracellular L-phenylalanine in E. coli NST74 with both L-tryptophan and L-tyrosine added......................................................................................................................... 216
Figure 7. 19. Diagram of L-phenylalanine synthesis, the pathways targeted in this thesis and the effect these inhibitors had on L-phenylalanine yield................................. 220

Chapter 8. General discussions and conclusions

Figure 8. 1. DBTL cycle with the platforms developed in this work. .................... 226
Tables

Chapter 3. High throughput mass spectrometry using flow injection on a triple quadrupole mass spectrometer

Table 3. 1. MRM list for all experiments performed on the TSQ Quantiva™ triple quadrupole mass spectrometer. ........................................................................................................... 48

Table 3. 2. Reproducibility of the mass spectrometry methods on six metabolites .. 52

Chapter 4. Low-cost LEGO® automation to aid the synthetic biology cycle

Table 4. 1. Modifications in the LEGO® robot. ................................................................. 65

Table 4. 2. Comparison of open-source robots. ................................................................. 77

Chapter 5. Synthetic biology for styrene production

Table 5. 1. The genetic modifications in E. coli NST74. .................................................. 96

Table 5. 2. The strain and plasmids used........................................................................ 97

Table 5. 3. 24 and 48-hour induction experiment......................................................... 105

Table 5. 4. Significantly different metabolites from the 100% ACN extraction........ 109

Table 5. 5. Average peak area for $^{12}$C and $^{13}$C glucose for extracellular samples in negative mode........................................................................................................... 114
Chapter 6. Using an inhibitor to alter the flux of L-phenylalanine in an engineered *E. coli* strain

Table 6.1. The different concentration of L-tryptophan added to media. .............. 134
Table 6.2. Preliminary data of extracellular L-phenylalanine and the fold change. 137
Table 6.3. Average optical density of *E. coli* NST74 with added L-tryptophan. .... 140
Table 6.4. Mean peak area of extracellular L-phenylalanine and the fold change when grown with L-tryptophan. ................................................................. 142
Table 6.5. Statistics showing the variability within the 10mM added L-tryptophan samples. ........................................................................................................... 142
Table 6.6. IM-q-TOF extracellular metabolites. ...................................................... 150
Table 6.7. IM-q-TOF intracellular metabolites ....................................................... 159
Table 6.8. Mean peak area and fold change of extracellular L-phenylalanine in *E. coli* NST74 with L-tryptophan added at stationary phase ........................................ 177

Chapter 7. Exploring a panel of inhibitors on an engineered *E. coli* strain

Table 7.1. Overview of the enzyme and inhibitors used in this chapter .............. 181
Table 7.2. Overview of the inhibitors used and the different concentrations that were added ........................................................................................................... 186
Table 7.3. Average peak area and overall fold change for *E. coli* NST74 with different concentrations of added anthranilic acid. ......................................................... 189
Table 7.4. The average peak area and overall fold change of samples of *E. coli* NST74 with different concentrations of added L-tyrosine ....................................... 210
Table 7.5. Average peak area and overall fold change for extracellular E. coli NST74 samples grown with different concentrations of added L-tryptophan and L-tyrosine.

Table 7.6. Summary of inhibitors used and the fold increase this had on the yield of extracellular L-phenylalanine.
List of Abbreviations

5-HTP 5-hydroxytryptophan
ACN acetonitrile
AmpC ampicillin
APCI atmospheric pressure chemical ionisation
APPI atmospheric pressure photoionisation
CAD computer aided design
CCS collision cross section
CMW chloroform: methanol: water
CRM charged residue model
DAHP 3-Deoxy-D-arabinohexulosonate 7-phosphate
DBTL design-build-test-learn
DC direct current
DHB 2,3-dihydroxybenzoate
DIY do it yourself
EI electron ionisation
ESI electrospray ionisation
FDC1 ferulic acid decarboxylase
FDR false detection rate
FWHM full width at half maximum
GC gas chromatography
HILIC hydrophilic interaction liquid chromatography
HPLC/UHPLC high/ultra high-performance liquid chromatography
ICIS Interactive Chemical Information System
IEM ion evaporation model
IM-q-TOF ion mobility quadrupole-time of flight
IPTG isopropyl β-D-1-thiogalactopyranoside
LC liquid chromatography
MALDI matrix assisted laser desorption ionisation
MeOH methanol
MRM multiple reaction monitoring
MS mass spectrometry
m/z mass to charge
NEM N-ethylmaleimide
NMR nuclear magnetic resonance
OD$_{600}$ optical density at 600 nm
PAL2 phenylalanine ammonia-lyase
PEEK polyetheretherketone
PET polyethylene terephthalate
PHA polyhydroxyalkanoates
PHB poly-3-hydroxybutyrate
PLA polylactic acid
QQQ triple quadrupole
Q-TOF quadrupole-time of flight
RP reversed phase
SRM single reaction monitoring
TIC total ion chromatogram
TOF time of flight

WCW wet cell weight
Chapter 1.

Introduction

1.1 Metabolomics

The central dogma of molecular biology states that information flows directionally, from genomic DNA through mRNA transcripts; these are then translated into proteins. The protein products, such as enzymes, structural proteins, and transport proteins, then affect the concentrations of substrates and products. These are part of tightly controlled metabolic pathways (figure 1.1). It is now understood that this information is omnidirectional, and often relies on feedback loops (Roberts et al., 2012). The ‘omics’ are named as they are the study of the ‘-ome’, which refers to the entire complement of a class of biomolecules, for example, genomics is the study of the genome. Thus metabolomics, also known as metabonomics and metabolic profiling, is the comprehensive analysis of all metabolites and low molecular weight molecules in a biological specimen (Oliver, 1998).

Figure 1. 1. The central dogma of molecular biology.

Whilst the term ‘metabolome’ was coined by Oliver (1998), its history predates this, with metabolic profiling studies being performed as early as the 1960’s with Dalgleish
et al. (1966) paper on gas-liquid-chromatography to separate metabolites in urine and tissue extracts.

Metabolites are small molecules, typically with a mass of less than 1500Da (Qui et al., 2023). Metabolites are dynamic and respond to the cellular environment, and therefore metabolomics provides a ‘snapshot’ of the phenotype at that moment in time. Metabolomics can enable further study of biosynthetic and biodegradation pathways (Villas-Boas and Bruheim, 2007). It can also allow the analysis of metabolic flux, which is described as “the passage of a metabolite through a reaction system over time” (Burgess et al., 2014). Metabolites consist of a diverse range of chemical classes, such as lipids, sugars, and organic acids, with a range of biological functions. Due to this diversity, no single methodology will be able to analyse and detect every small molecule, and therefore there is always some degree of bias (Wang, 2010). An example of this bias would be the use of different extraction solvents. For example, polar metabolites may be more efficiently extracted by polar solvents such as water, and therefore the choice of solvent can bias the detection of metabolites (Kirkwood et al., 2013).

Metabolomics can be split into two general methodologies: untargeted and targeted. Untargeted metabolomics is the analysis of all measurable small molecules within a sample. In this type of analysis, the coverage of the metabolome is only restricted by sample preparation and the limitations of the chosen analytical technique, and therefore untargeted metabolomics can enable novel discovery, for example biomarkers (Schrimpe-Rutledge et al., 2016). Due to its complexity, this produces extensive data sets and requires advanced data processing and analysis (Gertsman and Barshop, 2018).

In contrast, targeted metabolomics is the measurement of a defined group of metabolites. This can be carried out in a quantitative manner with the use of standards, and usually involves knowledge of established biochemical pathways. When looking at targeted metabolites, sample preparation can be optimised based on the physiochemical properties of these metabolites. Data sets are often easier to analyse (Roberts et al., 2012).

The typical analytical techniques used in metabolomics are nuclear magnetic resonance (NMR) and mass spectrometry (MS).
NMR spectroscopy is a physiochemical technique used to obtain structural information of molecules. The technology is founded on the excitation of atomic nuclei when exposed to a magnetic field. NMR has been successfully implemented in metabolomics on numerous occasions (Gowda and Raftery, 2021). NMR is non-destructive, easy to quantify and requires little sample preparation. Additionally, NMR is highly automated and reproducible, making it ideal for large-scale metabolomic studies (Wishart et al., 2022). Typically, less metabolites can be detected with NMR than in other techniques, such as MS. Further, NMR is 10 to 100 times less sensitive than MS, and although each technique has advantages and limitations, MS accounts for ~80% of all published metabolomic studies (Emwas et al., 2019).

Mass spectrometry is an analytical technique that separates ions based on their mass (m) to charge (z) ratio. The history of MS goes back to the early 1900s; J. J. Thomson was already notable for his discovery of the electron in 1897, and with the help of his student, Francis Aston, they then designed the first mass spectrometer. The instrument used gas discharge tubes to disperse ions by mass. The ions passed through electrical and magnetic fields before being deflected and detected on a photographic plate (Griffiths, 2008). Advancements continued and by the 1930s, MS was a confirmed technique for the separation of atomic ions by mass. By the 1940s, mass spectrometers were commercially available, and MS was firmly established (Nier, 1998).

Modern day MS has evolved into a versatile field, with a choice of ionisation sources and mass analysers, creating a wide array of choices and possibilities. Each type of ionisation source and mass analyser has its advantages and limitations, and therefore it is important to consider these before selecting an instrument (Cooks and Yan, 2018).

One advantage of MS is it has a wide detection range, capable of analysing small metabolites to entire proteins. Another advantage of MS is its high throughput capabilities and analytical speed. It also has the ability to detect hundreds of compounds, ranging from low concentrations to high. The basic structure of a MS is ion source -> mass analyser -> ion detector -> computer system (figure 1.2).
MS analysis generally involves prior separation through chromatography. This is liquid chromatography (LC) for compounds in the liquid phase, and gas chromatography (GC) for molecules in the gas phase (Coskun, 2016). Despite this, ions are always detected in the gas phase. The basic fundamentals of MS are prior separation, ionisation, and detection. As mentioned, separation can occur through LC or GC, but is not a necessary component. Ionisation happens in the ion source and involves the formation of gas phase ions. The mass analyser plays the role of separating the ions, through various methods that will be discussed later in this chapter. The detector detects and records the filtered ions. The final result is a spectra which requires interpretation and analysis. Two important terms within MS are resolution and mass accuracy. Resolution refers to the ability to distinguish two ions of similar mass to charge ratio. The most common method to measure resolution is using the full width at half maximum (FWHM) of the peak (Marshall et al., 2013).

\[
Resolution = \frac{m/z \text{ measured}}{FWHM}
\]

Mass accuracy is defined as the difference between the theoretical and measured m/z of a molecule (Brenton and Godfrey, 2010).

\[
Mass \text{ accuracy (ppm)} = \frac{m/z \text{ theoretical} - m/z \text{ measured}}{m/z \text{ theoretical}} \times 10^6
\]
1.2 Separation techniques

1.2.1 Liquid chromatography

Liquid chromatography (LC) was first used in 1903, by the botanist Mikhail Tswett, who used LC to separate plant pigments (Tswett, 1905; Hesse and Weil, 1954; Ali et al., 2010). Whilst the technology has changed in that time, the concept remains the same. The sample passes through a column, containing a packing material, known as the stationary phase. The sample then interacts with the stationary phase, this leads to elution at different times based upon each components affinity to the packing material. The liquid components are known as the mobile phase, which moves and separates the sample across the stationary phase (Coskun, 2016).

High Performance Liquid Chromatography (HPLC) and Ultra High-Performance Liquid Chromatography (UHPLC) are often used in metabolomics and provide efficient chromatographic separation. When HPLC or UHPLC is coupled to a MS this results in enhanced sensitivity and metabolome coverage (Gika et al., 2014). As the stationary phase separates molecules based on different physiochemical properties, obtaining comprehensive metabolomic coverage is difficult. However, it is possible by using different column chemistries, for instance using Normal Phase (NP), Reversed Phase (RP) and Hydrophilic Interaction Liquid Chromatography (HILIC). In NP, the stationary phase is polar, and the mobile phase is non-polar, this causes more polar compounds to interact more strongly with the stationary phase, resulting in later elution. In RP chromatography, the stationary phase is non-polar, and the mobile phase is polar. It separates analytes based on their hydrophobicity, with more hydrophobic compounds interacting more strongly with the stationary phase and therefore eluting later. HILIC has a stationary phase which is polar, and the mobile phase is typically non-polar or low in polarity. This increases the retention times of polar molecules, however unlike NP the mobile phase can be highly organic, which can increase sensitivity when coupled to an MS with electrospray ionisation (Moldoveanu and David, 2022). Typically, untargeted RP-UHPLC is designed to profile medium and non-polar metabolites, whereas HILIC is more suitable for polar metabolites. This usually occurs separately, but there are other options such as
column switching, coupled column separations, and simultaneous parallel separations using two columns (Xu et al., 2021).

1.2.2 Gas Chromatography

Gas chromatography (GC) is the separation technique of choice for smaller volatile and semi-volatile organic molecules such as hydrocarbons, alcohols and aromatics. It is also used for pesticides, steroids, fatty acids, and hormones (Choi and Chung, 2014).

In GC, the sample is vaporised into gas phase and separated into various ions by a capillary column coated with a stationary phase. The compounds are pushed along by an inert carrier gas such as helium. Typically, a GCMS column is made from fused silica with varying diameters and lengths (Hopfer, 2022). Each ion interacts with the column stationary phase differently based upon their chemistry, resulting in separation. As the components are separated, they elute at different times, and therefore hit the detector at different times (figure 1.3) (Gawale et al., 2022).

![Gas Chromatography Mass Spectrometry](image)

*Figure 1.3. Gas chromatography mass spectrometry.*
1.3 Alternatives to prior separation

Separation prior to analysis with an MS is not always necessary.

1.3.1 Flow injection

Flow injection is not a separation technique as there is no prior separation, rather it is a method of sample presentation. This method is described in more detail in chapter 3, but briefly, flow injection involves the injection of a sample into a liquid flow that then undergoes ionisation and analysis in the MS. This has limitations, such as reducing the sensitivity of the instrument through two main mechanisms. The first is ion suppression at the ionisation source, and the second is ion competition in the detection system (Sarvin et al., 2020).

1.3.2 Direct infusion/injection

As with flow injection, direct infusion does not involve prior separation. Instead, it involves continuous ionisation of a static sample. This analytical technique has been used in untargeted electrospray (ESI)-MS studies including plant, food, environmental, animal and human tissue sample analysis. As with flow injection, the advantages of direct infusion MS are the speed and throughput capabilities (Kozlova et al., 2022).

However, its limitations include strong matrix effects and ion suppression caused by the ions entering the MS at the same time. These limitations can be minimized through method optimisation, for example using nano ESI to deliver extremely low flow rates and increase ionisation efficiency. This is due to smaller droplets than conventional ESI, which requires less sample desolvation and results in more ions successfully entering the MS (Chekmeneva et al., 2017).
1.4 Mass spectrometry

1.4.1 Ion sources

The purpose of the ion source is to ionise the molecules from the sample, causing them to be in the gas phase. There are different techniques for ionisation; electron ionisation (EI), matrix assisted laser desorption (MALDI) and atmospheric pressure ionisation methods. The atmospheric methods can be broken down into electrospray ionisation (ESI), atmospheric pressure chemical ionisation (APCI) and atmospheric pressure photo ionisation (APPI) (Bhardwaj and Hanley, 2014). Out of these, ESI is the most commonly used.

Electron Ionisation

Electron ionisation (EI) is typically used with GC-MS and occurs in a vacuum. The parent molecule, in gas phase, passes through a beam of high energy electrons. When the parent molecule collides with the electron beam, it can cause the parent molecule to eject one or more electrons from the molecule, creating an ion. This is then accelerated towards the detector (Amirav et al., 2020).

MALDI

Matrix assisted laser desorption ionisation (MALDI) involves taking the analyte and placing it on a conductive plate and mixing it with a matrix, usually an organic acid, such as 2,5-dihydroxybenzoic acid, α-cyano-4-hydroxy-trans-cinnamic acid and sinapinic acid. This is then dried out, and the crystallised mixture is hit with a laser. The matrix molecules absorb the energy, and this creates a plume of molecules in the gas phase. The protons from the organic acid transfer to the analyte, which is then pulsed towards the MS (Li et al., 2022).
Atmospheric pressure ionisation methods

Atmospheric pressure ionisation methods are known as ‘soft’ ionisation methods. This is because they lead to less in-source fragmentation than other techniques. There are three types: electrospray ionisation (ESI), atmospheric pressure chemical ionisation (APCI), and atmospheric pressure photo ionisation (APPI). APCI is most suitable for non-polar molecules such as lipids and steroid hormones, whereas APPI is most suitable for aromatic hydrocarbons that do not ionise easily (Thomas, 2019).

Electrospray ionisation

Electrospray was first theorised by John Strutt, Lord Rayleigh, in 1882. He claimed that the evaporation of a charged droplet of solvent will lead to a point where the repulsion of charges overcomes the surface tension of the droplet, causing the droplet to disperse. This charge-based dispersion of droplets is the basis of electrospray (Banerjee and Mazumdar, 2012). In 1968, Malcolm Dole first described electrospray ionisation, linking it with MS. In 2002, John Fenn received a Nobel Prize for his work with ESI, mainly the possibility to transfer large molecules such as proteins into gas phase without fragmentation. This work led to ESI becoming the most common ionisation technique in MS (Wilm, 2011).

In ESI, the sample is sprayed from a needle with a voltage applied. This causes a potential difference between the capillary and the cone of the MS. The action of this voltage causes a Taylor cone to form. Droplets spray from the Taylor cone and are covered in charges. Solvent is removed, partly due to the heated auxiliary gas, and partly due to the volatile nature of the solvents. This causes the droplets to shrink and increases the charge density in the droplet. When the repulsive forces inside the droplet become higher than the droplet surface tension (Rayleigh instability limit), the droplet explodes in a process known as Coulomb fission, and forms charged progeny droplets. The solvent then evaporates from these, leaving naked charged analytes. The potential difference between the needle and ion transfer tube is used to attract the naked charged analyte towards and into the MS (figure 1.4) (Ho et al., 2003).
It is important to note there are two models for ESI: the ion evaporation model (IEM) and charge residue model (CRM). The IEM theorises that gas-phase ions are produced from highly charged droplets, and ions are ejected from the droplet surface. The CRM theorises that for high molecular weight molecules the excess charges from the droplet are transferred to, and remain, on the molecule enclosed within a droplet after solvent evaporation (Sunner et al., 2006).

There are different factors that affect the ionisation performance of ESI. These are the applied potential difference, the flow rate, solvent choice, sheath and auxiliary gas rate, and temperature (Liyanage et al., 2018). The applied voltage can improve ionisation when increased to a certain voltage, however the signal can become unstable, and the signal/noise can increase if the voltage is too high (Gardner and Brodbelt, 2009). The flow rate has an effect, as a lower flow rate requires less solvent evaporation, and therefore this helps ion formation, however in LC slow flow rates can lead to longer acquisition times and less than ideal chromatography. The choice of mobile phase can have an impact, as solvents with low surface tension produce stable and reproducible electrospray (Liyanage et al., 2018). This includes solvents such as methanol, isopropanol, and acetonitrile. The auxiliary gas flow rate and temperature can also have an effect, as this can help with de-solvation of the sample and can be optimised to reduce solvent droplet size. This then leads to an increase in ion formation and analyte signal (Gardner and Brodbelt, 2009).

![Electrospray ionisation model, based on the diagram by Banerjee and Mazumdar (2012).](image-url)
This method is a soft ionisation technique, this means it has reduced in-source fragmentation and therefore the structure of a molecule can enter the MS instrument intact.

1.4.2 Mass analysers

As well as different ionisation methods, there are several different types of mass analysers within MS. These are responsible for separating ions based on their m/z values. The choice of mass analyser depends on numerous different factors, including:

- The desired m/z range
- The required resolution
- The ion source needed
- The limit of detection

There is no single mass analyser that is suitable for all applications, and therefore before choosing an instrument it is important to consider these factors. Here quadrupoles and time of flight analysers are discussed, as these were used within the experimental section of this thesis, however other mass analysers are also discussed in order to give an overview.

Quadrupoles

Quadrupole mass analysers are one of the most common types of MS. They are often bench top mass spectrometers, typically low cost, and known for their durability and reliability. They are often used in tandem, either with other quadrupoles in a triple quadrupole instrument or with other analysers such as Time of Flight or Orbitraps (Li et al., 2021).

The quadrupole acts as a mass filter. It consists of four cylindrical or hyperbolic rods, with opposite rods connected together. A radio frequency (Rf) is applied, as is a direct current (DC). The combination of Rf and DC causes ions to oscillate as they
pass through the quadrupole. Only ions of a certain m/z will have a stable trajectory and are able to pass through, others will hit the rods and be filtered out. The m/z of the ion can be selected for by changing the Rf and DC potential (figure 1.5) (Miller and Denton, 1986).

![Quadrupole mass analyser](image)

**Figure 1.5. Quadrupole mass analyser showing the alternating Rf and DC on opposite rods.**

Quadrupoles are robust instruments and require little maintenance. They are suitable for both GC and LC due to their need for continuous beam of ions. However, this makes them less suitable for ion sources such as MALDI where the ions are pulsed in. Quadrupoles also have limited mass ranges and poor resolution (Li et al., 2021). Quadrupoles are commonly used in targeted metabolomics.

**Time of Flight**

Another mass analyser is Time of Flight (TOF). In TOF instruments, ions are accelerated by a standard kinetic energy from a voltage grid and fly down a vacuum tube with a detector at the end. As all the ions have the same kinetic energy applied to them, the small ions will fly faster and hit the detector first. The ions hit the
detector plate at different times, and by measuring the time taken for the ion to hit the detector, the m/z of the ion can be determined (figure 1.6) (Boesl, 2016).

Figure 1.6. Time of Flight mass analyser.

TOF analysers have the highest mass range of all analysers, as well as high resolution and high mass accuracy, however as the m/z value increases the resolution decreases. This is overcome by the use of reflectron TOFs, which reflects the ion path back in the direction of the ion source before being detected. TOF analysers work well with both ESI and MALDI ion sources (Singhal et al., 2015). Within the field of metabolomics, untargeted work is usually performed on instruments with a TOF mass analyser.

**Ion trap**

The ion trap mass analyser is a modification of the quadrupole. The ion trap consists of two hyperbolic electrodes facing each other and a hyperbolic ring electrode placed between them. As with quadrupoles, an oscillating Rf and DC potential are applied causing ions to become trapped between the electrodes. The ion trap is filled with a
dampening gas, typically helium, which collides with the ions and reduces their kinetic energy. Ions of different m/z are then selectively removed from the ion trap by the varying Rf potential. Once removed from the trap, the ions then hit the detector (figure 1.7) (Nair and Clarke, 2017).

Figure 1.7. Ion trap mass analyser.

**Orbitrap**

The Orbitrap mass analyser is composed of three main parts, an inner spindle electrode covered by two hollow outer electrodes facing each other. The ions are collected in the C-trap, and an ion packet is pulsed into the Orbitrap (figure 1.8). The ion packets then rotate around the central electrode, forming rings and oscillating laterally. This causes the ions to separate based on their m/z. The outer electrodes also act as receiver plates that detect the back-and-forth axial motion of the ions. The signal produced by the ion oscillation then undergoes Fourier transformation to provide mass spectra (Zubarev and Makarov, 2013). One of the main advantages of the Orbitrap analyser is its high resolving power. Nonetheless, a limiting factor is that
they are prone to space charge effects if the Orbitrap becomes overfilled (Perry et al., 2008).

**Figure 1.8.** Orbitrap mass analyser.

### 1.4.3 Hybrid instruments

Often multiple mass analysers can be combined in one single MS instrument. This includes the Q-TOF (quadrupole-time of flight) and QQQ (triple quadrupole), along with other examples. Here the Q-TOF and triple quadrupole instrument will be discussed, as they were used in this thesis.

**Triple quadrupole**

Triple quadrupole mass spectrometry technology was invented by Christie G. Enke and Richard A. Yost in the late 1970s (Yost and Enke, 1978). Triple quadrupole
instruments consist of three sets of quadrupole analysers used in sequence. The first quadrupole is referred to as Q1, and can either scan across a range of m/z values or can selectively filter specific ions based on their m/z. The ions then pass through to the second quadrupole, Q2, this is used exclusively as a collision cell to fragment the ions from Q1 and consists of neutral gas atoms, such as nitrogen, to enable collision induced dissociation. The product ions formed by this fragmentation are scanned in the third quadrupole, Q3 (figure 1.9).

**Figure 1.9. Triple quadrupole instrument.**

**Ion mobility quadrupole-time of flight**

An ion mobility quadrupole-time of flight (IM-q-TOF) instrument consists of two mass analysers: a quadrupole and a TOF analyser. Ion mobility is a technique where ions are generated by an ion source and then enter a gas-filled drift tube (figure 1.10). In the drift tube, ions are accelerated with an applied electric field whilst a buffer gas opposes the ion motion. Ions with different sizes and charges will have different ion mobilities, and subsequently different drift times. For example, ions with higher charge states will experience higher electric force leading to increased drift velocity. Furthermore, ions with larger collision cross-sections will experience more interactions with the buffer gas, and therefore move slower. Therefore, ion mobility MS provides a new dimension of separation and shape information, because separation is based on the conformation of a molecule in addition to its mass (Kanu et al., 2008). It provides the ability to assess the drift times and collision cross-
section values, therefore alongside the other information gained, e.g. retention time and m/z value, enables more confidence in identifications (Dodds and Baker, 2019).

Other hybrid instruments include Tribrids consisting of three mass analysers, for example the Thermo Scientific™ Orbitrap™ Ascend Tribrid™ with a quadrupole, ion trap and Orbitrap. Another example is a Q-trap, such as the SCIEX QTRAP® 6500 with a quadrupole and ion trap. Hybrid instruments combine the capabilities of the different mass analysers, with increased versatility and flexibility than single mass analyser instruments.

1.5 Synthetic biology and the ‘design-build-test-learn’ cycle

Metabolomics and mass spectrometry are employed in a variety of fields, such as clinical research, disease treatment, drug characterisation, and agricultural research (Yang et al., 2019). In recent years, metabolomics and MS has started to play a crucial role in the field of synthetic biology. An example of this includes identifying
enzymatic bottlenecks with alcohol dehydrogenase, leading to genetic modifications, resulting in increased yield of 1-butanol in *E. coli* (Novak et al., 2020). Therefore, in this thesis, metabolomics and mass spectrometry were applied to a synthetic biology application.

Synthetic biology is a multidisciplinary field of science that applies engineering principles to biological systems, resulting in the modification of organisms and the redesign of existing systems. The concept of synthetic biology was first proposed in 1914 by Stephane Le Duc (Le Duc, 1914). However, it was not until 1980 that the term ‘synthetic biology’ first appeared in literature, to describe bacteria that had been genetically engineered using recombinant DNA technology (Benner, 2004).

Synthetic biology has the potential to address a variety of significant issues, which pose predominant problems to society, such as reducing the spread of disease and pests, enhancing food production, and restoring biodiversity in its natural environments (Hobman et al., 2022). The range of uses include medical applications, such as engineered kidneys cells for diabetic treatments (Xie et al., 2016); textile applications, such as engineered silkworms to produce indigoidine, allowing the production of blue coloured silk (Jia et al., 2023). Another example is food production, including the use of an engineered plant-based burger, Impossible™ Foods. Here, yeast is engineered to produce soy leghaemoglobin to improve meaty flavours and aromas. It was also found that compared to a beef patty, the Impossible™ Burger requires 96% less land and produces 89% fewer greenhouse gases, leading to a positive environmental effect too (Voigt, 2020).

To make the process more streamlined, the ‘design-build-test-learn’ (DBTL) cycle has been implemented (figure 1.11). This aims to increase the throughput and success of each project, however, bottlenecks still exist (Chao et al., 2016). This is illustrated by 1,3-propanediol which took over 10 years and $100 million to develop a biosynthetic process for production (Committee on Industrialization of Biology, 2015).

In the past decade, the “design” and “build” stages have been propelled by massive improvements in DNA sequencing and synthesising technologies, leading to significant reductions in cost and turnaround time. Currently, the bottleneck is within the “test” region, as data acquisition and data analysis can be time consuming and complicated. The “learning” stage has seen rapid advancement due to machine
learning which can cope with the large amounts of data that is generated (Opgenorth et al., 2019).

In the “design” phase, the construction of the biological system is planned, for example selecting and designing genetic constructs. The “build” phase involves constructing these genetic modifications. This process can often be automated to increase speed and throughput with robotic liquid handlers and DNA synthesis machines. Automation enables multiple actions to be performed simultaneously and around the clock. In principle, automated robots could enable replicate numbers beyond the capabilities of human researchers. In the “test” phase, the engineered strains are evaluated for their performance and functionality. The data generated allows for the identification of successful modifications. This is the current bottleneck due to lengthy data acquisition techniques and long data analysis times (Delépine et al., 2016). Recent advances in high throughput analytical techniques have increased

Figure 1. 11. ‘Design-Build-Test-Learn’ cycle. The bottleneck is in the ‘test’ phase due to the lengthy analytical techniques used.
the speed, however due to complexity of the analysis (e.g. proteomics or metabolomics) the process is still time consuming. An example of a high throughput system is the Agilent RapidFire platform, which couples solid-phase extraction with HT-ESI-MS for rapid and automated MS analysis (Woodruff et al., 2016). The “learn” phase evaluates the data and uses it to refine the design of the biological system. The cycle then repeats until the product is optimised.

Targeted proteomics has been successfully applied using the DBTL cycle to identify bottlenecks in the mevalonate pathway, enabling an improvement in the production of a sesquiterpene (Redding-Johanson et al., 2011). Studies have utilised proteomics and metabolomics in combination for analysing engineered strains, resulting in improved production of target molecules (Brunk et al., 2016; George et al., 2014). Based on these successes, it was decided to use metabolomics and MS, alongside the synthetic biology ‘design-build-test-learn’ cycle for product enhancement. This was through the use of inhibitors to see if they could influence the production of the target molecule, whilst high throughput platforms were developed to decrease the bottleneck in the DBTL cycle.

1.6 Inhibitors

Inhibitors are essential tools in both biological research and medicine (Pérez de la Lastra et al., 2021). They can be used to study the functions of enzymes and receptors and are also used as therapeutic agents to treat various medical conditions (Kong et al., 2023). For example, enzyme inhibitors can be used to slow down or block specific metabolic pathways in the treatment of diseases like cancer or hypertension (El-Kenawi and El-Remessy, 2013). One other potential use of inhibitors is to direct the flux of metabolic pathways towards a high value product and provide targets for genetic editing in synthetic biology (Du et al., 2014).

Inhibitors can be classified as either reversible or irreversible. An irreversible inhibitor forms a stable complex with the enzyme, leaving the enzyme permanently inactivated, although in some cases the enzyme can be reactivated over a lengthy period (Strelow et al., 2012).
Enzyme kinetics are important and illustrate the binding between the enzyme and inhibitor, and the effect on the substrate. The mechanism of inhibition is described through calculations called the Michaelis-Menten equation which demonstrates Vmax and Km (Nomenclature Committee of the International Union of Biochemistry, 1983).

\[
v = \frac{V_{\text{max}} [S]}{K_m + [S]}
\]

[S] is the concentration of the substrate. Vmax is the maximum rate of the reaction, when all the enzymes active sites are saturated with substrate. The relationship between the rate of reaction and concentration of the substrate depends on the affinity of the enzyme for its substrate. Km is the substrate concentration at which the reaction rate is 50% of the Vmax. An enzyme with a high Km has a low affinity for its substrate and requires a greater concentration of substrate to achieve Vmax (Pant, 2022). However, this equation is only true when there is one substrate and one product and therefore is an oversimplification of what occurs in biological processes.

![Graph showing Vmax and Km for different types of inhibitors: competitive and non-competitive.](image)

**Figure 1.12.** Vmax and Km for different types of inhibitors; competitive and non-competitive.

There are three main types of inhibition: competitive, non-competitive, and uncompetitive. A competitive inhibitor binds to the active site in place of the
substrate, which will increase the Km value for the substrate and there will be no change in Vmax value (figure 1.12). When the inhibitor occupies the active site, the substrate cannot bind, and the enzymes catalytic activity is reduced. This inhibition can be overcome by increasing the substrate concentration (Ramsay and Tipton, 2017).

A non-competitive inhibitor is a type of allosteric regulation. The inhibitor binds to an allosteric site which results in a conformational change so the substrate can no longer bind. This also means the inhibitor can bind to the enzyme regardless of whether the substrate is bound to the enzyme. Non-competitive inhibitors lower Vmax and there is no effect on Km for the substrate (figure 1.12) (Delaune and Alsayouri, 2022).

Uncompetitive inhibition takes place when an enzyme inhibitor binds to the enzyme substrate complex, resulting in the complex becoming inactive. The Vmax and Km both decrease (Strelow et al., 2012).

1.6.1 Inhibitors as product enhancers

In this thesis, inhibitors were used to see if they could inhibit metabolic pathways, and direct flux towards the production of a high value product. Similar work was successful in agriculture, such as Wang et al. (2021) who tested nitrification inhibitors to increase crop yield in wheat and Khan et al. (2014) who used inhibitors to increase the yield of potatoes.

In a more biological setting, Jutras et al. (2016) exhibited the use of a protease inhibitor to increase the yield of a monoclonal antibody produced in Nicotiana benthaminana leaves. Inhibitors of the cell cycle have also been used to selectively target CDK4/6, resulting in optimisation of cell growth, product per cell and glycosylation profile in antibody production in Chinese hamster ovary (CHO) cell cultures (Du et al., 2014). However, this field seems under explored, particularly within synthetic biology and therefore it was important to investigate.
1.7 Aims

The hypothesis of this project was: is it possible to reduce the bottleneck in the synthetic biology DBTL cycle using high throughput metabolomics and mass spectrometry? Also, is it possible to use inhibitors to increase the yield of a high value product?

To do this a combination of low-cost automation, high throughput flow injection MS methods and inhibitor panel were tested, and they represent a starting point for a high throughput platform for strain optimisation and rapid testing using metabolomics.

These high throughput platforms were used on a biological example; *E. coli* NST74, a highly modified strain. Two examples of high value products were explored; firstly styrene and secondly L-phenylalanine. The yield of L-phenylalanine was then optimised through the use of inhibitors, and targets selected for future synthetic biology work.

The objectives were as follows:

1. Develop a high throughput targeted mass spectrometry method. This was done through flow injection mass spectrometry using a targeted MRM on a triple quadrupole mass spectrometer (Chapter 3).
2. Investigate a low-cost automation platform to increase throughput. This involved the use of a LEGO® liquid handling robot (Chapter 4).
3. Increase the yield of a high value product. Initially this was investigated with a styrene producing plasmid, with styrene as the high value product. Due to incompatibility with the analytical platform this was changed to L-phenylalanine (Chapters 5-7).
4. Provide targets for future synthetic biology work. This was achieved using inhibitors to block alternative pathways in order to analyse the effect on the yield of the target product, L-phenylalanine (Chapters 6 and 7).
Chapter 2.

Materials and Methods

This chapter describes the general materials and methods used throughout this thesis. Any experiment deviating from what is described here is specifically mentioned in the corresponding chapter.

2.1 Reagents

All reagents used in this work were purchased from Sigma-Aldrich (Merck Life Science UK Limited, Gillingham, Dorset, United Kingdom) with the following exceptions: LC-MS solvents were OPTIMA® LC-MS Grade from Fisher Scientific (Loughborough, Leicestershire, United Kingdom) except formic acid which was from Merck KGaA (Darmstadt, Germany). Vanillic acid, diethyl malonate, nickel (II) sulphate hexahydrate and methyl anthranilate were from Thermo Scientific Chemicals (Massachusetts, USA). Axygen® Axymat, Axygen® PCR plate and CoStar 96 well plates were used from Corning (New York, USA).

2.2 Bacterial strains

The bacterial strains used in this thesis were *Escherichia coli* K12, obtained from the teaching laboratory at the School of Biological Science, University of Edinburgh, and *E. coli* NST74 (ATCC 31884). Both *E. coli* NST74 and the pTpal-fdc plasmid were obtained from Professor Stephen Wallace at the University of Edinburgh.

*E. coli* NST74 was originally produced by Tribe (1987) and is genetically modified to overproduce L-phenylalanine, and therefore has the following modifications: aroH367, tyrR366, tna-2, lacY5, aroF394(fbr), malT384, pheA101(fbr), pheO352, aroG397(fbr). Each of these modifications is discussed in more detail in chapter 6.
For the styrene work, a pTpal-fdc plasmid was added to the *E. coli* NST74 strain. This plasmid was based on the Ptrc99A plasmid and is ampicillin resistant. The plasmid consisted of phenylalanine ammonia-lyase 2 (PAL2) from *Arabidopsis thaliana* inserted into the NCol and Xbal site, and ferulic acid decarboxylase (FDC1) of *Saccharomyces cerevisiae* inserted into the SbfI and HindIII sites. This work was conducted by McKenna et al. (2013), and enabled L-phenylalanine to be converted to styrene.

### 2.3 Growth media

Different growth mediums were used at different points in this thesis. The composition of these is as follows:

- **Luria-Bertani (LB) media:** Tryptone (10g/L); Yeast Extract (5g/L); NaCl(10g/L); 1000 mL Purified Water (MilliQ)

- **M9 media:** M9 salt solution (10X) (Na$_2$HPO$_4$ 33.7mM; KH$_2$PO$_4$ 22.0mM; NaCl 8.55mM; NH$_4$Cl 9.35mM), Glucose 0.4%, MgSO$_4$ 1 mM, CaCl$_2$ 0.3 mM, Biotin 1 µg, Thiamine 1 µg, purified water

- **MM1 media:** glucose (15g/L); MgSO$_4$·7H$_2$O (0.5 g/L); (NH$_4$)$_2$SO$_4$ (4.0g/L); MOPS (24.7 g/L); KH$_2$PO$_4$ (0.3 g/L); K$_2$HPO$_4$ (0.7 g/L); and 5mL/L ATCC Trace Mineral Supplement (Catalogue No. MD-TMS) (EDTA(0.5g/L), MgSO$_4$·7H$_2$O (3g/L), MnSO$_4$·7H$_2$O (0.5g/L), NaCl (1g/L), FeSO$_4$·7H$_2$O (0.1g/L), Co(NO$_3$)$_2$·6H$_2$O (0.1g/L), CaCl$_2$ (0.1 g/L), ZnSO$_4$·7H$_2$O (0.1g/L), CuSO$_4$·5H$_2$O (0.01g/L), AlK(SO$_4$)$_2$ (0.01 g/L), H$_3$BO$_3$ (0.01 g/L), Na$_2$MoO$_4$·2H$_2$O (0.01g/L), Na$_2$SeO$_3$ (0.001 g/L), Na$_2$WO$_4$·2H$_2$O (0.10 /L), and NiCl$_2$·6H$_2$O (0.02g/L)).

### 2.4 Culture conditions

Culture conditions varied depending on the individual experiment, therefore they have been described in their respective chapters.
2.5 OD₆₀₀ measurement

2.5.1 Flask cultures

For bacteria cultured in flasks, 200µL of the culture was added to 1800µL media. 1mL of this was then transferred to a cuvette, and the OD₆₀₀ reading was taken in a cell density meter (Ultrospec™ 10, Biochrom), this was then used to calculate the OD₆₀₀ of the culture.

2.5.2 96 well plate cultures

For bacteria cultured in 96 well plates, the OD₆₀₀ was diluted to 0.1, growth readings were then taken for the whole growth period. The 96 well plate was placed in a Tecan Infinite® 200Pro plate reader with a target temperature of 37°C, and a shaking (linear) duration of 550s. This meant an optical density reading was taken every 10 minutes.

2.6 Metabolite extraction

2.6.1 Flask cultures

For each sample, 1mL was transferred to a chilled Eppendorf tube. The samples were then centrifuged at 4 °C and 13,000 x g for 10 minutes. The supernatant and cell pellet were collected as extracellular and intracellular fractions respectively and stored at -80 °C until further extraction for LC-MS analysis.

Intracellular metabolite extraction

For intracellular metabolite extraction, 200µL of solvent was added for every 5 mg of wet cell weight (WCW) of the pellet. The WCW was determined by spinning down
1mL of the sample for 5 minutes at 13,000 x g in a pre-weighed Eppendorf tube, removing the supernatant and weighing the remaining cell pellet. The pellet was then re-suspended by pipetting. Samples were then mixed in a chilled microtube mixer or a rotary shaker for 1 hour at 4°C. They were then centrifuged for 3 minutes at 13,000 x g and 4 °C. At this point, 50µL of supernatant was transferred to MS vials and 120µL of supernatant was transferred into a fresh Eppendorf tube and stored at -80 °C, in case of further MS analysis. During handling, the extraction solvent and the samples were kept on ice.

**Extracellular metabolite extraction**

Extracellular extractions were prepared by diluting 10µL of sample into 390µL of solvent. The samples were then vortexed for 5 seconds and placed in a chilled microtube mixer or a rotary shaker for 5 minutes at 4 °C, and then centrifuged for 3 minutes at 13,000 x g and 4 °C. At this point, 50µL was transferred to MS vials, and 180µL of supernatant was transferred into a fresh Eppendorf tube and stored at -80 °C in case of further MS analysis. During handling, the extraction solvent and the samples were kept on ice.

For flask cultures different extraction solvents were used. The extraction solvent is stated in each respective section.

**2.6.2 96 well plate cultures**

The plate was removed from the plate reader and then centrifuged at 1000 x g for 5 minutes. The supernatant was then transferred to a fresh 96 well plate, resulting in one plate with a cell pellet and another with supernatant.

For these extractions the solvent was 50:50 acetonitrile: water. During handling, the solvent and samples were kept on ice.
Intracellular metabolite extraction

Samples were normalised by adding the amount of solvent in proportion to the OD$_{600}$. The highest OD$_{600}$ had 200µL of solvent added, whilst the other samples received a proportional amount. The cell pellet was resuspended by pipetting up and down. The 96 well plate was then placed on a plate rocker at 4°C for 1 hour. It was then centrifuged at 1000 x g for 5 minutes. 75µL of the supernatant was then transferred to an Axygen™ 96 well PCR plate and an Axygen Axymat™ silicone sealing mat was placed on top to seal it. The remaining sample plate was sealed and kept at -80°C in case of further analysis.

Extracellular metabolite extraction

For the extracellular metabolite extraction samples, 190µL was added to a fresh plate and 10µL of supernatant was added. The solvent was 50:50 acetonitrile: water at 4°C. This was then placed on the plate rocker at 4°C for 10 minutes and then centrifuged at 1000 x g for 5 minutes. 75µL of the extraction was then transferred to an Axygen™ 96 well PCR plate and an Axygen Axymat™ silicone sealing mat was placed on top, whilst awaiting MS analysis. The remaining sample plate was sealed and kept at -80°C in case of further analysis.

2.7 Targeted mass spectrometry

2.7.1 Flow injection

Flow injection took place using the autosampler from an Ultimate™ 3000 LC system. The mobile phase solvent was 50:50 methanol: water + 0.1% formic acid. Polyetheretherketone (PEEK) capillary tubing was used, it was 2220mm long, with an internal diameter of 0.005” (0.13 mm). The flow rate was 200µL and a wash step of 50µL was used between samples. The optimisation required for this method is discussed in chapter 3.
2.7.2 Instrument parameters

The instrument used was a Thermo Scientific™ TSQ Quantiva™ triple quadrupole mass spectrometer. The instrument has a maximum mass resolution of 0.2 Da peak width (FWHM) in Q1 and Q3. Ions were generated using an Ion Max NG source (Thermo Scientific™) with a heated electrospray ionisation (HESI) II probe. The mass spectrometer was operated using multiple reaction monitoring (MRM) method in polarity switching mode.

The parameters were as follows:

Spray voltage ± 3.5kV
Sheath gas (Arb) 25
Aux gas (Arb) 5
Sweep gas (Arb) 0
Ion transfer tube (°C) 325
Vaporiser temperature (°C) 75

Q1 resolution was set to 0.7 FWHM, Q3 resolution was set to 1.2 FWHM and the collision induced dissociation in Q2 occurred under a gas pressure of 1.5 mTorr.

Cycle time was 0.8 seconds and data was collected in centroid.

The MRM table is comprised of reference standards, previous work by Garcia et al., (2023) and Wheeler (2016). This can be found in chapter 3, table 3.1.

2.7.3 Data processing and analysis

Raw data was processed using Xcalibur™ software (version 4.4.16.14). Xcalibur Qual browser was used to evaluate chromatograms, and Xcalibur Quan Browser was used for peak detection.

ICIS™ peak detection was selected with the following parameters:
Smoothing points: 1
Baseline window: 50
Area noise factor: 5
Peak noise factor: 10

Advanced parameters were all left as default. Noise method was INCOS noise. RMS was not selected. Peak parameters were as follows:

Minimum peak width: 3
Multiple resolution: 10
Area tail extension: 5
Area scan window: 0

Peak detection was checked manually. This involved scrolling through the list of targeted metabolites to confirm optimal peak detection. On occasion the algorithm did not fully integrate a peak, meaning a peak was either missed or half identified. Therefore, manual integration was needed, this involved adding a line across the base line of the peak to ensure optimal integration.

After processing, the metabolite features were extracted as a .csv file and further analysed using GraphPad version 8.4.2.

Figure 2. 1. Targeted mass spectrometry workflow.
2.8 Untargeted mass spectrometry

2.8.1 Data acquisition

Liquid chromatography occurred on an Agilent 1290 Infinity II series UHPLC system (Agilent Technologies) using a Poroshell 120 HILIC-Z, 2.1 mm x 50 mm, 2.7 μm coupled to a InfinityLab Poroshell 120 HILIC-Z, 3.0 mm x 2.7 μm UHPLC guard column. For positive ion mode, mobile phase A was 10 mM ammonium formate in water with 0.1% formic acid, pH 3 and mobile phase B was 10 mM ammonium formate in water/acetonitrile (1:9) with 0.1% formic acid, pH 3. For negative ion mode, mobile phase A was 10 mM ammonium acetate in water, pH 9, and mobile phase B was 10 mM ammonium acetate in water/acetonitrile (1:9), pH 9. Detailed chromatography parameters can be found in Pičmanová et al. (2022). Samples were eluted in a linear gradient over 3.5 minutes at a flow rate of 200 μL/min. This was coupled to an Agilent 6560 IM-q-TOF (Agilent Technologies) with a Dual Agilent Jet Stream Electron Ionization source. The parameters of the mass spectrometry method were used as in Pičmanová et al. (2022).

2.8.2 Data processing

Data processing took place as described in Pičmanová et al. (2022). Initial data processing involved demultiplexing the files on PNNL PreProcessor. AgTofReprocessUI was used to perform accurate mass recalibration, next IM-MS browser 10.0 was used to recalibrate drift time. Molecular features were extracted in Mass Profiler 10.0 with a retention time tolerance of ± 0.3 minutes, drift time tolerance of ± 1.5% and accurate mass tolerance of ± (5 ppm + 2 mDa). The raw multiplexed data, the reconstructed demultiplexed data and Mass Profiler feature lists (.cef files) were then utilised in the High Resolution Demultiplexer (HRdm) 1.0 beta v41 for further peak deconvolution. Features were re-extracted from HRdm files using Mass Profiler 10.0 and annotated based on accurate mass and collision cross section (CCS) values using McLean CCS Compendium PCDL (Picache et al., 2019).
2.8.3 Data analysis

Data analysis took place using Metaboanalyst 5.0 (Pang et al., 2021) using .csv files generated from Mass Hunter Qualitative Analysis 10.0. Annotations were performed against the McLean Unified CCS Library (Picache et al., 2019) with a tolerance of 5ppm +/- 2mDa for m/z and +/-1% for CCS. These identifications are level 2, according to the Metabolite Standards Initiative as the identification is based upon a library spectrum match rather than a reference standard, however as orthogonal data was acquired (in the form of CCS values) this adds to the confidence level (Sumner et al., 2007; Schymanski et al., 2014). These results were converted to .csv files and analysed using Metaboanalyst 5.0. Here the data was log-transformed and scaled using pareto scaling.
Chapter 3.

High throughput mass spectrometry using flow injection on a triple quadrupole mass spectrometer

3.1 Introduction

Mass spectrometry is usually coupled with liquid or gas chromatography with the aim of separating metabolites before detection. This separation enables isobaric compounds to be detected and increases the sensitivity of the instrument by minimising the effects of ion suppression (Metz et al., 2009). However, this approach is time consuming, with chromatographic separation usually taking 20-60 minutes per sample. From a high throughput perspective, this is far from ideal. Ultra short columns are available, with shorter chromatographic separation times, but these are often limited in their ability to separate complex mixtures (Sarvin et al., 2020). The use of columns also results in an added cost, with it not being uncommon for a column to cost £1,000 or more. Optimisation of liquid chromatography methods can be complex, and involves the use of different buffers, columns, gradients, and flow rates.

In contrast, flow injection does not require any lengthy separation techniques, allowing for a fast analysis time, typically of 1 minute (figure 3.1). This makes it perfect for high throughput. There are many types of flow injection MS, but the most common involves autosamplers from existing LC systems and by-passing the column compartment, allowing direct injection into the ion source to take place (Nanita and Kaldon, 2015).
Autosamplers from LC systems have been optimised to be suitable for LC purposes, this benefits flow injection MS as the autosamplers can deliver injections of less than 1µL repeatedly and reliably. This results in increasing accuracy and precision for flow injection methods (Nanita and Kaldon, 2015). However, there are still many aspects that require optimisation.

One example of this is mobile phase solvent efficiency. This has an advantage over LC as non-LC friendly solvents and reagents can be used. Any liquid that is compatible with MS can be used as a carrier liquid in flow injection MS. Consequently, an array of solvents and additives can be tested in order to find the most appropriate for a specific set of samples (Rappold, 2022).

Another example of optimisation is the rate of diffusion, which occurs as the sample travels from the injector to the ion source. This can be altered by adjusting the liquid flow rate and the capillary volume. The analyte peak shape is dictated by the system configuration (capillary length and volume) and system conditions (injection volume and flow rate) (Nanita and Kaldon, 2015). This is important to optimise as the analyte signal intensity will increase as the inner capillary diameter decreases, however, there must be enough data points collected (minimum of five) across the peak for accurate quantitation. Therefore, optimisation is necessary to find a balance for analyte sensitivity and maintaining accurate peak width for data collection.
Flow injection analysis has been demonstrated on a wide range of analytes, including blood (Johnson, 2010), plasma (Sander et al., 2015), and urine (Niesser et al., 2012). Pesticides (Nanita et al., 2009), vitamins (Bhandari et al., 2013), and pharmaceuticals (Michel et al., 2015) have also been positively identified using flow injection analysis, showing its versatility. Flow injection has been used previously in biotechnology in the 1990s (Keay and Wang, 1997; Lüdi et al., 1990), however in recent years it has become under-utilised. Due to its ease of use, and its successful previous applications, this technique was investigated to develop as a high throughput technique to help aid the synthetic biology ‘design-build-test-learn’ cycle for *E. coli* NST74 and metabolites within the L-phenylalanine biosynthetic pathway.

Rapid MS analysis was required as part of the ultimate aim of reducing the bottleneck, but so was easy data processing and analysis. This was important when it came to selecting which MS to use. Each MS instrument has its strengths and limitations, and therefore before selecting an instrument it is important to evaluate these. For the purpose of this research, an instrument was needed which would be capable of fast analysis of samples and quick but simple data analysis. Based on these criteria, a triple quadrupole instrument was chosen.

The triple quadrupole instrument is capable of performing various different scan modes, one of these is selected reaction monitoring (SRM) or multiple reaction monitoring (MRM) (Gross, 2017). In MRM, the first quadrupole (Q1) is used to select for a specific precursor ion, filtering out ions of other m/z values. The second quadrupole, Q2, acts as a collision cell. In the collision cell, the precursor ion fragments with nitrogen (an inert gas) with a collision energy, resulting in product ions. These product ions are then selected for in the third quadrupole, Q3 (figure 3.2). The conversion from precursor/parent ion to product/daughter ions is referred to as an ion transition. Multiple transitions can be used for the same analyte in order to increase the reliability of the identification. Only compounds that meet both these criteria, i.e. specific parent ion and specific daughter ions, corresponding to the mass of the molecule of interest are isolated within the mass spectrometer (Pisitkun et al., 2007).
Figure 3.2. MRM scanning on a triple quadrupole instrument.

This is a highly specific and sensitive mass spectrometry technique that can selectively quantify compounds within complex mixtures. MRM sensitivity depends critically on the tuning of instrument parameters for the generation of maximal product ion signal (Sherwood et al., 2009), however, whilst the design of MRMs can aid selectivity it does not completely remove the possibility of misidentifications. MRMs allow targeted analysis, which enables easier data handling than untargeted. The data from targeted metabolomics is less complex as only a specific set of metabolites are being analysed, and therefore data interpretation is simpler. As data processing and analysis can often become another bottleneck, MRMs were chosen to avoid this.
Figure 3.3. Data quality includes sensitivity, accuracy, precision of quantification and number of metabolite identifications. No single methodology is sufficient to achieve full coverage of the complete metabolome. Different levels of analysis have different strengths and limitations. Single metabolite analysis usually provides the most sensitive and precise analysis. Targeted analysis can provide high quality analysis using optimised methods for a single class of compounds. Metabolite profiling can analyse a wider range of chemical classes but does not provide the same level of data quality. Metabolite fingerprinting offers the widest coverage, but data analysis is more complex and offers lower confidence in quantitation. Diagram based on Fernie et al. (2004).

Limitations of the triple quadrupole MS include a reduced mass range (m/z 10-1800) compared to other MS instruments, e.g. m/z 40 – 8000 for Thermo Scientific™ Orbitrap Exploris™ 240 Mass Spectrometer. However, for this research this does not matter, as only small molecules are being analysed. Triple quadrupole instruments also have lower resolution than other MS instruments, and therefore only a nominal mass can be obtained.
With all factors considered, a targeted method using flow injection analysis on a triple quadrupole instrument, using MRMs, was chosen to reduce the bottleneck of the synthetic biology ‘design-build-test-learn’ cycle.

Off target effects would be analysed on an IM-q-TOF using the method described in Pičmanová et al. (2022). In this research, off target effects can be defined as unintended effects on metabolism. The Pičmanová et al. (2022) method provided fast, untargeted data acquisition using ZIC-HILIC liquid chromatography, with chromatographic separation of a selection of metabolites with a variety of physiochemical properties, maintaining reproducibility. It took 3.5 minutes per sample and provided a fast and reproducible high-resolution platform for untargeted metabolomics. One limitation of this untargeted method, however, was its complex data processing and analysis as the data multiple stages of processing, such as demultiplexing.

### 3.2 Aims

The main aim of this section was to develop a targeted method for mass spectrometry analysis on a triple quadrupole instrument, for metabolites in the L-phenylalanine biosynthetic pathway. The method had to be fast, reproducible, and suitable for metabolites in the shikimate and L-phenylalanine pathway.

This chapter involved optimisation, both with the instrument method and processing method, in order to achieve the most ideal platform for analysing these samples.

### 3.3 Methods

#### 3.3.1 Initial flow injection mass spectrometry set up

The initial mass spectrometry method took place using an Ultimate 3000 autosampler and a Thermo Scientific™ TSQ Quantiva™ triple quadrupole instrument. The MS was set up using the following parameters:
Method Duration (min) = 1, Ion Source Type = H-ESI, Spray Voltage: Positive Ion (V) = 3500, Spray Voltage: Negative Ion (V) = 3500, Sheath Gas (Arb) = 35, Aux Gas (Arb) = 7, Sweep Gas (Arb) = 0, Ion Transfer Tube Temp (°C) = 325, Vaporizer Temp (°C) = 275, Experiment Type: SRM, Use Cycle Time: True, Use Chromatographic Filter: True, Cycle Time (sec): 0.800, Chromatographic Peak Width (sec): 6.0, Data Mode: Centroid, Collision Gas Pressure (mTorr): 1.5, Q1 Resolution (FWHM): 0.7, Q3 Resolution (FWHM): 1.2

This was tested on stock solutions of amino acids at different concentrations in MS vials, and optimisation began to take place. Initially a flow rate of 100µL was used. The PEEK tubing lengths were 650mm and 2220mm. The internal diameter was 130µM.

Optimisation occurred and some MS parameters were changed. These were the auxiliary gas which was changed to 5 arb, and the vaporizer temperature was reduced to 75°C.

For the flow injection methods, there were several changes made in controlling the autosampler; these were increasing the sampler wash speed, waste speed, draw speed, needle wash volume, and decreasing the sampler draw delay. The longer tubing length was selected, and the flow rate was also increased to 200µL/min.

3.3.2 Optimisation

Solvent system selection

Solvents of 60:40 isopropanol: water + 0.1% ammonium hydroxide and 50:50 methanol: water + 0.1% formic acid were tested and compared at a flow rate of 200µL/min (Cajka et al., 2023). Intracellular and extracellular metabolite extraction samples of E. coli NST74 were used, and some were spiked with additional L-phenylalanine. The comparison was made based on the total ion chromatogram (TICs) acquired and signal intensity and signal consistency of L-phenylalanine.
Extraction solvents

Different extraction solvents were tested on the *E. coli* NST74 cells, 100% ACN, 50:50 ACN: water, 100% MeOH, 50:50 MeOH: water and 40:20:40 MeOH: water: ACN (Fu et al., 2022). These were performed using the method described in chapter 2, Section 2.6.1.

The extraction solvent chosen was based on TIC traces and signal intensity of L-phenylalanine in both extracellular and intracellular samples.

MRMs

Several reiterations of MRM optimisation took place. This included the addition of metabolites of interest, and removal of metabolites that were not detected, in order to optimise dwell time. Different polarities were also tested; positive, negative, and positive/negative switching. The number of transitions for each metabolite was also optimised. The final table can be seen below, table 3.1.

Processing methods

The processing method was optimised alongside the MRM optimisation, this mostly consisted of MRM and transition selection, as well as optimising the peak detection algorithm. The ICIS™ peak detection algorithm was chosen, with the following parameters:

- Smoothing points: 1
- Baseline window: 50
- Area noise factor: 5
- Peak noise factor: 10

Advanced parameters were left as default and are available to see in chapter 2.
Other optimisation

Wash steps

To avoid carry-over of metabolites, and contamination, a wash step of 100% ACN was introduced between each injection. This was optimised by using different volumes in the wash step and analysing any carry over. The volumes of wash step tested were 10µL, 50µL and 300µL.

Injection volume

Injection volumes of 1µL and 5µL were compared. This was performed on *E. coli* NST74 with added L-tryptophan as an inhibitor (chapter 6). Extracellular metabolite samples of *E. coli* NST74 were tested.

3.3.3 Reproducibility test

To test reproducibility, the same sample of extracellular metabolite extraction of *E. coli* NST74 was injected ten times using the final method developed. The TICs and signal intensity for L-phenylalanine and other metabolites were then analysed.

3.4 Results and Discussion

3.4.1 Initial flow injection mass spectrometry set up

The initial flow injection set up required some immediate optimisation. It became clear that the tubing used was too short, resulting in the sample reaching the mass
spectrometer too quickly, and therefore the sample is only partially analysed. This can be seen in figure 3.4, where both the TICs and chromatogram for L-phenylalanine show approximately half of the sample being analysed. However, increasing the tubing length enabled the whole peak to be measured (figure 3.5). More optimisation occurred including changing the flow rate from 100µL/min to 200µL/min and changes to MS parameters such as the auxiliary gas and vaporizer temperature were made.

**Figure 3.4.** A) Total ion chromatograms of amino acid standards mixture using the shorter (650mm) tubing with an internal diameter of 130µM. Eight different injections are shown. B) chromatogram for the peak of L-phenylalanine. Four different injections are shown. Note that when the MS analysis starts the injection band has already started.
Figure 3. 5. A) Total ion chromatograms of amino acid standards mixture using the longer (2220mm) tubing with an internal diameter of 130µM. Eight different injections are shown. B) chromatogram for the peak of L-phenylalanine. Eight different injections are shown. Note the full injection band is able to be analysed.

As mentioned previously in this chapter, the length and diameter of the tubing is an important factor in flow injection analysis. This is due to injection band diffusion, as the sample travels from the needle to the mass spectrometer. The analyte peak shape is determined by the capillary dimensions and conditions (e.g. injection volume and flow rate) (Nanita and Kaldon, 2015). The capillary of 2220mm long and 130µM diameter, alongside an injection volume of 1µL and flow rate of 200µL/min provided good analyte peak shapes for the targeted metabolites in this work.

One MS parameter which was altered was the auxiliary gas, this was decreased from 7 to 5 arb. The purpose of the auxiliary gas is to help aid the ions from the probe to the ion transfer tube. For these samples, this parameter worked best when less auxiliary gas was flowing. Another parameter change was to the vaporizer temperature. The sheath gas and auxiliary gas also serve the purpose of sample desolvation, which is completed within the heated capillary. The auxiliary gas heats
as it passes through a vaporizer. The vaporizer is thermally insulated from the sample tube to prevent direct heating of the sample. The desolvation process is important in ESI and occurs when the heated auxiliary gas is applied to the charged droplet, leading to solvent evaporation. This causes the droplet size to decrease, and the charged droplet is now at the Rayleigh limit, resulting in charged progeny droplets which dry out and become individual gas phase analyte ions. These are then guided through the ion optics into the mass analyser (Daub and Cann, 2011). In this method, the vaporiser temperature was reduced in order to decrease the speed of desolvation and ensure optimum ion transfer into the mass spectrometer.

All the changes conducted on the autosampler were to decrease the time taken for an injection, whilst ensuring the injection needle was kept clean, to avoid contamination. This includes increasing the draw speed and decreasing the draw delay.

### 3.4.2 Optimisation

**Solvent system selection**

These solvents were chosen as both methanol and isopropanol have low surface tension and produce reproducible electrospray (Liyanage et al., 2018; Ho et al., 2003). Out of the tested solvents, the mobile phase chosen was 50:50 methanol: water + 0.1% formic acid. This was chosen because there was a difference in retention times of the main peaks as shown in the TIC examples in figure 3.6. For the 60:40 isopropanol: water + 0.1% ammonium hydroxide the peak is at the beginning, which could lead to some metabolites being undetected (figure 3.6A) whereas the 50:50 methanol: water + 0.1% formic acid peak appears approximately 20 seconds later (figure 3.6B). The traces for L-phenylalanine in the isopropanol solvent were also slightly less consistent than the methanol solvent.
In flow injection, the entire sample injection band enters the mass spectrometer very soon after the instrumental analysis run has started, as there is no prior separation. This can lead to severe analyte signal suppression (Nanita and Kaldon, 2015). Therefore, it is important that the entire injection band is able to be analysed by the MS. As the retention time of the 60:40 isopropanol: water solvent is faster, it means part of the band is missed. One reason for this difference in retention time could be the polarity. Increasing the polarity of the mobile phase leads to longer retention times and methanol has higher dipole moment than isopropanol, and therefore is more polar. Other solvent systems could have been tested such as acetonitrile, which has a higher ionisation efficiency of ESI than methanol (Kageyama Kaneshima et al., 2019), however the chosen solvent system worked well so there was no need for further optimisation.

Figure 3.6. Total ion chromatograms for the same three extracellular E. coli NST74 samples using different mobile phase solvents A) 60:40 isopropanol: water + 0.1% ammonium hydroxide, B) 50:50 methanol: water + 0.1% formic acid.
Extraction solvents

Metabolite extraction solvent optimisation was also conducted, and 50:50 ACN: water at 4°C was chosen. This was because it showed the most consistent results across multiple metabolite peaks both intracellularly and extracellularly, with the lowest standard deviations in metabolite peak area.

The choice of solvent is very important, there is a wide range of metabolites, so it is not possible to extract them with one single solvent (Saw et al., 2021). For metabolomics, an example of an extraction solvent would be 1:3:1 ratio of chloroform: methanol: water, in order for metabolites in a range of polar groups to be covered (Creek et al., 2011). However, as these extractions occurred in 96 well plates, where the chloroform caused plastic leaching and plate damage, alternatives had to be explored, and 50:50 acetonitrile: water was chosen.

Acetonitrile is an aprotic polar organic solvent with a polarity index of 5.8 and water is also a polar solvent, with a polarity index of 10.2 (Jordan et al., 2022). Therefore, this solvent extraction solution is favoured towards polar metabolites, and introduces some bias into the extraction. However, as this is a targeted analysis bias already exists as there is a certain set of metabolites that are the focus. Therefore, whilst 50:50 acetonitrile: water might not be the optimal extraction solvent for non-polar metabolites, it is still able to extract non-polar metabolites, such as L-phenylalanine, and therefore was used successfully for work conducted in this thesis.

MRM

Several iterations of MRM optimisation took place in order to achieve maximum results. Factors that were changed include the number of metabolites, number of transition per metabolite, and the polarities (positive only, negative only, and positive/negative switching). This was repeated several times to develop a finalised MRM list seen below (table 3.1).

For the purpose of this thesis, the focus was on the biosynthesis of L-phenylalanine and inhibition of other chorismate-driven pathways, and that is why these
metabolites were included. Some metabolites were excluded if they were not found to be present across a range of intra- and extracellular samples. This was to increase dwell time and increase signal to noise. The dwell time refers to the amount of time the instrument is collecting data for a specific MRM mass transition (in milliseconds). Increasing the dwell time will increase the sensitivity and precision of the method, but it will reduce the number of mass transitions that can be monitored (Cook-Botelho et al., 2017). The final method consisted of a dwell time of approximately 10 milliseconds per transition.

Some metabolites also had some transitions removed, as it became clear that they could be identified from the precursor and one product ion, and therefore this was another way to increase dwell time. The MRM table below (table 3.1) is comprised of reference standards, previous work by Garcia et al., (2023) and Wheeler (2016).
Table 3. 1. MRM list for all experiments performed on the TSQ Quantiva™ triple quadrupole mass spectrometer.

<table>
<thead>
<tr>
<th>Compound</th>
<th>Polarity</th>
<th>Precursor (m/z)</th>
<th>Product (m/z)</th>
<th>Collision Energy (V)</th>
</tr>
</thead>
<tbody>
<tr>
<td>PABA</td>
<td>Positive</td>
<td>138.24</td>
<td>120</td>
<td>17</td>
</tr>
<tr>
<td>L-Glutamine</td>
<td>Positive</td>
<td>147.076</td>
<td>84</td>
<td>17.55</td>
</tr>
<tr>
<td></td>
<td>Positive</td>
<td>147.076</td>
<td>101.214</td>
<td>11.91</td>
</tr>
<tr>
<td></td>
<td>Positive</td>
<td>147.076</td>
<td>129.929</td>
<td>9.63</td>
</tr>
<tr>
<td>L-Phenylalanine</td>
<td>Positive</td>
<td>166.086</td>
<td>77</td>
<td>39.84</td>
</tr>
<tr>
<td></td>
<td>Positive</td>
<td>166.086</td>
<td>120.054</td>
<td>14.06</td>
</tr>
<tr>
<td>3-Dehydroshikimate</td>
<td>Positive</td>
<td>171.215</td>
<td>127</td>
<td>30</td>
</tr>
<tr>
<td>Shikimate</td>
<td>Positive</td>
<td>173.235</td>
<td>73</td>
<td>19.8</td>
</tr>
<tr>
<td></td>
<td>Positive</td>
<td>173.235</td>
<td>93</td>
<td>26.2</td>
</tr>
<tr>
<td>Nicotinic acid</td>
<td>Positive</td>
<td>180</td>
<td>105.9</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td>Positive</td>
<td>180</td>
<td>135.9</td>
<td>14</td>
</tr>
<tr>
<td>L-Tyrosine</td>
<td>Positive</td>
<td>182.081</td>
<td>136.089</td>
<td>13.93</td>
</tr>
<tr>
<td></td>
<td>Positive</td>
<td>182.081</td>
<td>165.071</td>
<td>10.27</td>
</tr>
<tr>
<td>L-Tryptophan</td>
<td>Positive</td>
<td>203.315</td>
<td>74</td>
<td>22</td>
</tr>
<tr>
<td></td>
<td>Positive</td>
<td>203.315</td>
<td>116</td>
<td>20</td>
</tr>
<tr>
<td>Shikimate</td>
<td>Positive</td>
<td>204</td>
<td>189</td>
<td>8</td>
</tr>
<tr>
<td>Salicylate</td>
<td>Positive</td>
<td>209</td>
<td>91</td>
<td>8</td>
</tr>
<tr>
<td>Chorismate</td>
<td>Positive</td>
<td>225.265</td>
<td>91</td>
<td>19.96</td>
</tr>
<tr>
<td>Prephenate</td>
<td>Positive</td>
<td>225.265</td>
<td>181</td>
<td>7</td>
</tr>
<tr>
<td>Maleate</td>
<td>Positive</td>
<td>245</td>
<td>216.7</td>
<td>4</td>
</tr>
<tr>
<td>Shikimate-3-Phosphate</td>
<td>Positive</td>
<td>253.21</td>
<td>97</td>
<td>35</td>
</tr>
<tr>
<td>Shikimate</td>
<td>Positive</td>
<td>255</td>
<td>239</td>
<td>4</td>
</tr>
<tr>
<td>Isocitrate</td>
<td>Positive</td>
<td>257</td>
<td>200.7</td>
<td>4</td>
</tr>
<tr>
<td>Itaconate</td>
<td>Positive</td>
<td>259</td>
<td>130.8</td>
<td>20</td>
</tr>
<tr>
<td>D-Glucose 6-phosphate</td>
<td>Positive</td>
<td>259.22</td>
<td>199</td>
<td>15.7</td>
</tr>
<tr>
<td>Salicylate</td>
<td>Positive</td>
<td>267</td>
<td>209</td>
<td>8</td>
</tr>
<tr>
<td>p-coumaroyl shikimate</td>
<td>Positive</td>
<td>319.2</td>
<td>163.1</td>
<td>17</td>
</tr>
<tr>
<td>Pyruvate</td>
<td>Negative</td>
<td>87.008</td>
<td>43.083</td>
<td>8.79</td>
</tr>
<tr>
<td></td>
<td>Negative</td>
<td>87.008</td>
<td>68.929</td>
<td>5.25</td>
</tr>
<tr>
<td>Fumarate</td>
<td>Negative</td>
<td>115.003</td>
<td>69.929</td>
<td>13.34</td>
</tr>
<tr>
<td></td>
<td>Negative</td>
<td>115.003</td>
<td>71.03</td>
<td>8.16</td>
</tr>
<tr>
<td>Succinate</td>
<td>Negative</td>
<td>117.019</td>
<td>73.113</td>
<td>11.74</td>
</tr>
<tr>
<td></td>
<td>Negative</td>
<td>117.019</td>
<td>98.827</td>
<td>7.86</td>
</tr>
<tr>
<td>L-Aspartate</td>
<td>Negative</td>
<td>132.03</td>
<td>88.143</td>
<td>11.91</td>
</tr>
<tr>
<td></td>
<td>Negative</td>
<td>132.03</td>
<td>115.071</td>
<td>9.8</td>
</tr>
<tr>
<td>(S)-Malate</td>
<td>Negative</td>
<td>133.014</td>
<td>70.988</td>
<td>13.34</td>
</tr>
<tr>
<td></td>
<td>Negative</td>
<td>133.014</td>
<td>114.929</td>
<td>11.15</td>
</tr>
</tbody>
</table>
Further optimisation could take place by obtaining reference standards for all metabolites of interest and creating an MRM based on these, however the MRM created worked well for the samples obtained in this thesis, in chapters 6 and 7.

### Processing method setup

The processing method was optimised to ensure that the MRMs in the instrument method were being detected based upon their transitions. Peak detection aims to
pull apart peaks generated by real metabolites from those generated by random electrical and chemical noise (Zhang et al., 2009). The peak detection algorithm chosen was Interactive Chemical Information System (ICIS™). The ICIS™ detection algorithm has been designed for MS data and is said to have superior peak detection efficiency at low MS signal levels. As one of the limitations of flow injection is that matrix effects may reduce ionisation and thus signal intensities, this algorithm was chosen (Farke et al., 2023).

Other optimised steps

A washing volume of 50µL was chosen, this provided a compromise of avoiding carry over resulting incorrect analysis, and also avoiding a large time increase of data acquisition. The syringe wash was 100% acetonitrile.

A 1µL injection for each sample was chosen for this method. This enables this method to be transferred to experiments in which low quantity of the sample may be available. It also tested the sensitivity of the instrument. A 1µL injection provided enough sample for the targeted metabolites to be detected. Whilst the intensity for the 5µL injection was higher, it ultimately did not add to the analysis, and increased the risk of carry over and therefore provided no benefit for this data set.

The plate layout was designed randomly each time to prevent batch effects during the whole process. Batch effects are technical, non-biological factors that affect variation in the data; therefore, it is extremely important to avoid them (Leek et al., 2010). During the extraction process some empty wells were filled with solvent, these then acted as a blank during the MS analysis to check for carry over. These minor optimisations all contributed to produce a high throughput flow injection method, targeted at metabolites of interest.
3.4.3 Reproducibility

Ten injections of the same sample were performed with five other samples injected in between. This showed reproducible peaks for a range of metabolites, as shown in figure 3.7 and table 3.2.

Figure 3. 7. Average peak area for six metabolites and their standard deviations, illustrating the reproducibility of the data acquired using the flow injection method on the triple quadrupole instrument. Error bars indicate standard deviation.
As the figure above shows there is a small standard deviation, indicating that there is little difference in peak area across the injections. The instrument method and peak detection algorithm were optimised for this method and were reproducible across a range of metabolites. Both phosphoenolpyruvate and citrate were detected at much lower levels than the other metabolites and have larger variation, however the data is still reproducible. This is also indicated in the table below (table 3.2).

Table 3.2 Mean peak area, standard deviation, standard error mean and relative standard deviations for 6 metabolites within the MRM showing the reproducibility achieved.

<table>
<thead>
<tr>
<th>Metabolite</th>
<th>Mean peak area</th>
<th>Standard deviation</th>
<th>Standard error mean</th>
<th>Relative standard deviations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Phosphoenolpyruvate</td>
<td>3044.43</td>
<td>686.95</td>
<td>217.23</td>
<td>0.226</td>
</tr>
<tr>
<td>Citrate</td>
<td>1106.71</td>
<td>400.33</td>
<td>151.3091</td>
<td>0.391</td>
</tr>
<tr>
<td>L-Phenylalanine</td>
<td>50231.00</td>
<td>1435.42</td>
<td>453.92</td>
<td>0.028</td>
</tr>
<tr>
<td>L-Proline</td>
<td>64863.86</td>
<td>4068.68</td>
<td>1286.63</td>
<td>0.063</td>
</tr>
<tr>
<td>L-Tryptophan</td>
<td>36382.57</td>
<td>1838.11</td>
<td>581.26</td>
<td>0.051</td>
</tr>
<tr>
<td>L-Alanine</td>
<td>138431.40</td>
<td>6030.37</td>
<td>1906.97</td>
<td>0.044</td>
</tr>
</tbody>
</table>

3.5 Conclusion

Flow injection MS has been previously used for high throughput bacterial identification (Vaidyanathan et al., 2002). This research used a flow rate of 50µL/min using 50% aqueous acetonitrile containing 10mM formic acid and was analysed on a Q-TOF mass spectrometer. Vaidyanthan et al. (2002) were able to identify five different bacterial species successfully, with reproducible and stable spectra. Another successful implementation of flow injection MS was by Ramzan et al. (2021) who developed a method to identify the antimicrobial potential of compounds. This was achieved with a flow rate of 200µL/min using different solvents, on a Q-TOF
instrument. This method was successfully implemented to screen anti-bacterial drugs.

The work conducted in this chapter illustrates that flow injection using targeted MRM$s on a triple quadrupole instrument is successful for the evaluation of metabolites in the shikimate and L-phenylalanine pathway. It was used to run a number of inhibitor samples, shown in chapters 6 and 7. Whilst not all the metabolites in these pathways were present, this was due to a lack of reference standards, and therefore could be optimised for further use. The previous method used in this lab was a HILIC LC-MS method taking 40 minutes per sample. This new flow injection method was able to analyse samples against MRM$s for 34 metabolites in a method of 1-minute length. The total length between raw files generated was 2.5 minutes. This means a full 96 well plate could be analysed in 4 hours. This method can also be implemented in any lab that has an LC and triple quadrupole mass spectrometer as it does not require any additional instrumentation, for example a NanoMate®. This completes the aim of generating a high throughput flow injection method to reduce the bottleneck in the synthetic biology DBTL cycle.

As well as fast analysis, this method also has quick data processing. The processing method means that the data can be exported to an excel sheet almost immediately, as the optimisation of the peak detection algorithm resulted in little manual input being required. If manual input was needed it involved scrolling through the list of metabolites and checking the algorithm had correctly detected the peaks. In most cases it was successful but occasionally the peak would have to be manually integrated by using the software to draw a line across the base of the peak. This process became routine and was fast and easy to complete. The excel sheet generated from the processing method provided lots of information, but for the purposes of this research the key findings was the peak area. This is directly related to the concentration of the metabolite, and therefore when analysing the effect of inhibitors on L-phenylalanine was the most important.

For future work, it would be ideal to include dilutions of the metabolites of interest, at known concentrations, to create a calibration curve for quantitation. Triple quadrupole instruments are often used for quantitation, and it is one of their advantages over other MS instruments. It would also be ideal to expand the MRM to
include more metabolites of interest, but it is important to keep an optimum dwell
time to maintain sensitivity of the instrument.

Another factor to optimise could be to decrease sampler speed more and perhaps
even alter the instrument parameters further, in order to gain higher sensitivity.
Ultimately this method worked well for the samples it was running; it was quick,
reproducible, and easy to analyse, fulfilling the initial criteria with ease.
Chapter 4.

Low-cost LEGO® automation to aid the synthetic biology ‘design-build-test-learn’ cycle

4.1 Introduction

4.1.1 Background

Synthetic biology is an interdisciplinary field, which combines biology and engineering to design and construct new biological systems, as well as modifying existing ones. This is achieved through the use of genetic engineering, DNA sequencing, and computer modelling to modify the characteristics of an organism (Benner and Sismour, 2005). The goal of synthetic biology is to create biological functions and systems, which do not exist in nature, or to enhance existing biological processes beyond their natural capabilities. This includes, but is not limited to:

- medical applications, for example: CAR-T therapies directed against the CD19 protein for the treatment of acute lymphoblastic leukaemia and large B-cell lymphoma (Yan et al., 2023).
- environmental purposes, for example: an engineered E. coli strain was produced to convert phenol to a carbon source, in order to degrade phenol in contaminated waste water (Wang et al., 2019).
- agricultural applications, for example: engineered proteobacterium (KV137) for nitrogen fixing in corn roots (Voigt, 2020).

Synthetic biology is predicted to impact most human activities, such as biofuel (Qiao et al., 2015), food (e.g. biosynthesis of carotenoids (Lv et al., 2021)), and medicine (e.g. yeast engineered for malaria treatment (Kung et al., 2018)). However, despite the growth and progress of synthetic biology, its full potential has yet to be reached.

Automation is defined as “the technique of making an apparatus, process, or a system operate automatically” (Merriam Webster Dictionary). In biology, this can
refer to a range of processes and workflows - such as laboratory tasks through to data analysis. In the laboratory, automation can be used to improve accuracy and efficiency in tasks such as pipetting, sample preparation, and high-throughput screening (Medina et al., 2023). Automated systems can also be used for the imaging and analysis of biological samples, such as microscopy and flow cytometry (Hayashi et al., 2011). In the field of clinical medicine, automation is becoming increasingly important for tasks such as electronic medical record management, patient monitoring, and diagnostic testing (Davenport and Kalakota, 2019). Overall, automation in biology has the potential to improve the speed, accuracy, and reproducibility of research and clinical applications, ultimately leading to a better understanding and treatment of biological processes and diseases.

Automation has become a critical requirement for synthetic biology, enabling researchers to design, build, test, and optimise genetic systems quickly and efficiently. This can include processes such as DNA synthesis, high-throughput screening, genome editing, and data analysis (Szymański et al., 2011). This results in reduced time and cost, increased efficiency, as well as helping researchers identify patterns and predict outcomes. A large benefit of automation within synthetic biology is the ability to rapidly generate genetic constructs, allowing large libraries to be created. Automation can also assemble DNA sequences in high throughput, allowing researchers to study a large number of organisms simultaneously (Walsh et al., 2019). This approach can help identify the most promising results for further study and optimisation. Another advantage of automation in synthetic biology is the ability to control experimental variables more precisely.

Automated platforms can precisely dispense reagents and perform measurements with high accuracy and reproducibility, thereby reducing the potential for experimental errors and variability. As synthetic biology is an iterative process, propelled along through the ‘design-build-test-learn’ cycle, automation allows this process to become streamlined and can reduce errors, as tedious and repetitive tasks are prone to human failings (Yeow et al., 2014). Automation has been consistently and positively implemented in large scientific companies, such as Amyris, Zymergen, Ginkgo, and Genomatica for the above-mentioned reasons (Carbonell et al., 2019). Overall, automation has significantly accelerated progress in synthetic biology, and is likely to continue to play a crucial role in the future.
One subcategory of automation is liquid handling robots. These are automated devices designed to transfer and dispense multiple liquid volumes accurately and precisely. These robots use advanced technology and software to manipulate a robotic arm to dispense various types of liquids, such as water, solvents, reagents, and samples. Liquid handling robots have seen incredible advances in recent years, such as increased throughput, increased accuracy and precision. They usually consist of modular systems enabling more customised workflows, user-friendly interfaces (Tegally et al., 2020), and integration with other technologies such as lateral flow assays (Anderson et al., 2022). Liquid handling robots offer practical and reliable ways of reducing workload in the laboratory, by offering consistency, reliability, as well as utilising workflows to become more efficient and cost-effective (Kong et al., 2012). Sample preparation methodologies can often be time consuming and repetitive and are often a frequent cause of the bottleneck in the ‘design-build-test-learn’ cycle. These robots are widely used in life science research, drug discovery, clinical diagnostics, and biotechnology in order to reduce these bottlenecks (Tegally et al., 2020).

Liquid handling robots come in different sizes, shapes, and configurations, depending on their intended use and application. Some robots can handle a few microlitres of liquid, whilst others can process hundreds of litres. Common features of these robots include liquid handling arms or needles, dispensing heads, pipetting tips, microplate holders, automated liquid dispensing systems, integrated sensors, and sometimes touch-screen interfaces (Tegally et al., 2020).

Overall, liquid handling robots play a crucial role in laboratory workflows, enabling researchers to carry out complex experiments with minimal errors and maximum efficiency. Therefore, it was an important part of this project to see if automation could be investigated and optimised for maximum benefits. Several low-cost liquid handling robots were reviewed for their suitability. This includes the Andrew+ robot by Waters™, capable of manipulating Gilson™ pipettes (sizes P2 to P1000), which is suitable for a lab benchtop with a low weight of 16kg. However, with an approximate cost of $24,000 this was not feasible (Semac et al., 2013). Another, lower cost, alternative was the Opentrons® robot (OT-2) which starts at a value of $9,000 but requires additional add-ons at extra cost in order to increase its usability (Villanueva-Cañas et al., 2021).
However, despite these options being considered “low-cost” they remain expensive and inaccessible to many labs, therefore, the search for a true low-cost automation platform that could be deployed in any lab continued. Therefore, low-cost open-source robots were investigated; the term “open source” refers to something individuals can modify and share as its design is publicly accessible, and therefore open source liquid handling robots can be modified to their users’ needs. A paper on the use of LEGO® robotics within STEM in education, by Gerber et al. (2017) came to light, and seemed like an interesting concept to pursue. This paper designed two liquid handling robots using a commercially available LEGO® kit and used it to perform simple experiments to capture the interest of schoolchildren studying science. This robot cost approximately £325 per kit and was built and modified easily from standard LEGO® parts, which are cheap and accessible worldwide.

4.1.2 Frugal science

Frugal science refers to scientific research that is conducted using low-cost and accessible materials and equipment. This movement was sparked by Professor Manu Prakash at the University of Stanford and highlights the demand for cheap and accessible equipment to make science open to everyone, regardless of socioeconomic status or geographic location (Newby, 2017). This emphasises creativity, resourcefulness, and simplicity, as well as community collaboration and interdisciplinary approaches. This is particularly important where access to expensive laboratory equipment and resources are limited. Examples of frugal science include electroporation using the ElectroPen (Byagathvalli et al., 2020) and the FoldScope, a microscope made from paper (Cybulski et al., 2014). Frugal science has the potential to be applied to many areas of research, including health, agriculture, energy, and education. Open-source automation research fits into this timely surge in frugal science and can enable a liquid handling robot to be built and developed at a low cost.
4.1.3 LEGO® robotics in STEM

LEGO® automation refers to the use of LEGO® building blocks to create automated machines or devices. It can be designed and built using various LEGO® pieces, motors and sensors, and can be programmed to perform tasks using software such as LEGO® Mindstorms®. The possibilities for LEGO® automation are extensive, with options to create simple machines to complex robots that can move, detect obstacles, and even solve Rubik’s Cube puzzles (Milstein, 2001). Many educational institutions also use LEGO® automation to teach students about engineering, programming, and robotics in a fun and interactive way.

As mentioned above, the search continued for a low-cost automation solution that could be used in any lab, and LEGO® robotics fit the criteria. The paper, by Gerber et al. (2017), on the use of LEGO® robotics within STEM in education, had successfully built two liquid handling robots – a 1D and a 2D robot. The 1D robot was capable of simple pipetting experiments whereas the 2D one was based on a linear rail system, and therefore could complete more complex experiments. The 1D robot was built from the parts in the standard Education EV3 LEGO® 45544 Mindstorms® kit and standard lab apparatus, such as 1mL plastic syringe to uptake and dispense liquids, and cuvettes. For the 2D robot, there was a more advanced design to enable better liquid handling, as the syringe was driven by a linear rail system. This required additional pieces, acquired separately from the EV3 kit. The 2D robot seemed worth exploring as Gerber et al. (2017) was able to achieve delivery of droplets of various sizes down to 2.5μL (25% precision, 8% accuracy) using a 1mL syringe, and down to 0.15μL (15% precision, 8% accuracy) using a 25-μL syringe, which is better than results from a P2 pipette (Gerber et al., 2017). In comparison, the TECAN Freedom EVO liquid handler shows precision of 10% and accuracy of 6% when pipetting 3μL water using a 1mL syringe (Bessemans et al., 2016).

This was chosen as the basis for the robot in this project, as it was based on commercially available parts, used open-source software, and had a low build cost of £325 per robot.
4.1.4 Programming robots

In general, programming for robotics consists of writing software, which can be uploaded to the robot as firmware, instructing it to perform a specific task or function. This task or function can vary from simple actions to complex processes and is dependent on the robots physical capabilities and the coding used, for example, the code may include instructions for controlling the robots movements through motors and sensors.

The most used programming languages for robots include C++, Python, Java, and MATLAB. A strength of these coding languages within automation is the precision of movement that can be obtained; however, it can be difficult to learn these coding languages, and therefore the complexity of the code could be a limitation (Abbasimoshaei et al., 2022). After the code has been written, it must be tested to ensure that it works as required. This may involve running the code on a simulated robot platform or on the actual robot, and making adjustments as necessary to ensure that the robot performs the desired tasks. It can then be transferred to the robot; this may involve uploading the code to the robot's computer or connecting to the robot remotely to run the code. The languages used for the LEGO® robot were LabVIEW and Scratch, both are visual programming languages known for their user-friendly interfaces.

Initially, programming for the 1D robot took place using LEGO® Mindstorms® EV3 Home Edition, based on LabVIEW software. LabVIEW (short for Laboratory Virtual Instrument Engineering Workbench) is a visual programming language, developed by National Instruments in 1986, to enable engineers and scientists to create virtual instruments that can be easily customised and programmed (Elliot et al., 2007). LabVIEW provides a graphical user interface for designing the front panel of virtual instruments, and a block diagram for programming the functionality of the instrument. The front panel contains the user interface elements, such as icons of each motor and sensor, while the block diagram contains the code for processing the information received from these elements (Figure 4.1A).

From June 2021, LEGO® retired Home Edition and replaced it with LEGO® MINDSTORMS® EV3 Home app. This uses Scratch as its programming language,
which was developed by the Lifelong Kindergarten Group at the Massachusetts Institute of Technology (MIT) and is designed to teach young people basic programming concepts and computational thinking in a fun and interactive way. This is done similarly to LabVIEW, as it is a visual programming platform where users create programs by dragging and dropping blocks, which represent code (Figure 4.1B). Like LabVIEW, Scratch is also open source, however it does not use icons.
Figure 4. 1. A) Serial dilution in cuvettes as performed on the 1D robot using LabVIEW based software where icons are dragged and dropped in the sequence required B) Serial dilution in 96 well plate using the Scratch based software where blocks are dragged and dropped in the correct order.
4.2 Aims

The aim of this project was to build and program a low-cost accessible liquid handling robot, by improving upon the Gerber et al. (2017) 2D robot, to make it suitable for use in biotechnology. There were three objectives:

- Build a liquid handling robot capable of addressing a 96 well plate.
- Develop programs for dilution curves, metabolite extraction, and different cell culture conditions.
- Evaluate robot pipetting performance and accuracy.

4.3 Method

The methods for the LEGO® liquid handling robot is split into two sections: the build and the programming. After the initial build, the modifications to the physical robot occurred alongside the programming for optimisation purposes.

4.3.1 Building

The initial build of the LEGO® liquid handling robot was based upon the 2D robot and was constructed as per Gerber et al. (2017) by following computer aided design (CAD) files. This can be pictured below, figure 4.2. For this robot to be deployed in a laboratory environment, several modifications were needed, these are shown in Table 4.1.
Figure 4.2. The LEGO® 2D robot built and programmed in this thesis, based on the robot by Gerber et al. (2017).
Table 4. The components in the LEGO® liquid handling robot that were altered from the original paper by Gerber et al. (2017). This table describes the component and the effect of the change.

<table>
<thead>
<tr>
<th>Component</th>
<th>Description</th>
<th>Effect</th>
</tr>
</thead>
<tbody>
<tr>
<td>LEGO® brick</td>
<td>CAD files showed the LEGO® brick directly attached to the cart, however this needed to be structurally secure to withstand the weight and movement of the syringe system. The extra weight from the brick was hampering free movement and caused inconsistencies with the volume of liquid dispensed. Therefore, the decision was made to remove the brick from the cart and instead to use cables to connect it. The cables provided with the kit were too short and again caused the cart to be easily dislodged from its intended position. Specialist cables were obtained from MindSensors.com to connect the brick and allow it to be placed further away.</td>
<td>The longer cables eliminated the issue with the cart; however, it caused another problem. The wires were now getting tangled within the moving parts of the robot. This was solved by connecting the wires using cable ties and ensuring there was sufficient movement for the robot to freely move, ensuring the wires could not get trapped.</td>
</tr>
<tr>
<td>Syringe</td>
<td>Another requirement was changing the syringe from a plastic syringe with a pipette tip glued onto it, to a glass syringe, which would be more compatible with standard laboratory use. Whilst the Gerber et al. (2017) paper shows that the swapping of syringes is possible, it does not give any instructions on how this was implemented.</td>
<td>This ensures stability but also enables freedom of movement of the raising mechanism, whilst keeping access to the removable needle clear. This allowed the robot to be used for Hamilton syringes of 1mL and 500µL. This was also done as these syringes have removable needles to help mitigate any contamination issues that may occur.</td>
</tr>
</tbody>
</table>
To maximise the effectiveness of the robot it was essential to make it possible to use a range of different syringe sizes. Therefore, this part of the robot was modified, which involved widening the area where the syringe sits. This involved changing the LEGO® pieces to longer ones and using a different method to secure the syringe within the syringe system. This is done by slotting the syringe tip into a separate LEGO® piece.

If working with a particularly viscous liquid, then this would require further calibration to ensure all liquid from the syringe was ejected. Similarly, if a liquid has low viscosity, then a way round this would be to get the syringe to draw up the liquid, lift up out of the well and then have syringe pull up some air, this should ensure the liquid stays within the syringe until it has moved to the well where it is then expelled. To overcome viscosity issues, accurate calibration for each liquid is required.

<table>
<thead>
<tr>
<th>Syringe injection</th>
<th>The issue of liquid viscosity affected how much liquid was ejected from the syringe. This was solved by intensive calibration and ensuring the motor was set to push the plunger all the way to the bottom of the syringe, in a set time.</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>96 well plate</th>
<th>To keep the 96 well plate stable, LEGO® pieces were added to the frame acting as a vice grip to keep it secure and stable.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Reservoir</strong></td>
<td>There was a requirement for a liquid reservoir to enable liquid transfer. To keep this in a fixed position LEGO® pieces were used to create a holder.</td>
</tr>
<tr>
<td><strong>Cart and frame</strong></td>
<td>Other LEGO® pieces were added to assist with the consistent loading of the robot, such as extra pieces on the stage to ensure it was always reset to the same point. This was mainly to help with any calibration issues but also to prevent any confusion if the robot was to be used by a different user.</td>
</tr>
</tbody>
</table>
4.3.2 Programming

Initially programming took place using Mindstorms® EV3 Home Edition based on LabVIEW. The programming involves controlling each motor and sensor individually and in the correct sequence. This was done by dragging and dropping icons of selected motor or sensor and deciding the exact movement needed, for example motor A rotate clockwise 60°.

Serial dilution

Firstly, a program was developed to perform serial dilutions in 24 well plates. Once this was working well, it was modified to enable serial dilutions on 96 well plates. A program to re-set the robot to its original start position was also developed.

The programming software then changed to Mindstorms® app using Scratch. Therefore, the program had to be repeated on the newer software. To check the programming worked for this a serial dilution of L-phenylalanine stock solution was completed. This was replicated 4 times, alongside a serial dilution with manual pipetting for comparison. The plate was then analysed on a Thermo Scientific™ TSQ Quantiva™ triple quadrupole mass spectrometer using the flow injection MS method and data analysis described in Chapter 2 – Section 2.7.

Different cell culture conditions

Other programmes were then developed for more specific lab use to support high throughput metabolomics, and to support the remainder of this project.

The first of these was the creation of cell culture or inhibitor assays, enabling each well to have a different media composition for growth experiments. This was programmed to work by transporting minimal media from a reservoir to all the growth wells and transporting reagents such as glucose/glycerol/inhibitors from a reservoir to empty wells. The next step is mixing the reagent, for example glucose, with the
minimal media in one well in a ratio of 1:1. A serial dilution is then performed, reducing the concentration of the reagent (figure 4.3). Once the robot has set up the plate, bacterial inoculation must occur manually, and the plate can then be put in a plate reader where growth data can be obtained. The programming for this was tested using minimal media and water as the reagent and liquid was transferred between wells as intended.

Figure 4.3. The intended set up of a 96 well plate for cell culture assays, resulting in each well having a different media composition for growth assays.

**Metabolite extraction**

Another program that was developed was for metabolite extractions. In metabolomics, one example of an extraction solvent is chloroform: methanol: water (CMW) in a 1:3:1 ratio (Creek et al., 2011), however as the robot uses well plates made from plastic, this would not be a suitable extraction solvent, as plastic leaching would cause contamination of the samples. When solvents were added to the 96 well plates, visible damage was observed using CMW, and no visible damage
occurred using 100% ice-cold acetonitrile or 100% ice-cold methanol. Therefore, it was decided to use either of these solvents for future use with the robot.

This methodology was split into two sections, extracellular metabolite extractions and intracellular metabolite extractions. The extracellular metabolite extraction method was programmed to transfer 20µL of supernatant from a used well plate to a fresh plate and then add 180µL of solvent from a reservoir. In contrast, the intracellular metabolite extraction method was programmed to transfer 200µL of solvent from a reservoir to a well, and then pipette up and down to re-suspend the cell pellet in the solvent. This then involved a wash step in another reservoir to avoid contamination, before repeating for another well.

4.4 Results

4.4.1 Uses

L-phenylalanine dilution curve

As mentioned, the first stage was to re-write the program for serial dilutions in 24 well plates. This was then transferred to 96 well plates, although this was more difficult due to the accuracy required to move the syringe to the correct location for each well. The modification of changing to the Hamilton syringe helped solve this, as there was a smaller needlepoint, and therefore it was easier to move the needle above the intended well.

To test the serial dilution programming on a 96 well plate, a serial dilution of 1M L-phenylalanine stock solution was performed, both by the robot (programming in figure 4.4), and manual pipetting. This was replicated four times in each condition. The plate was then analysed on a Thermo Scientific™ TSQ Quantiva™ triple quadrupole mass spectrometer, using the flow injection method mentioned in chapter 3, and the data analysed in Xcalibur™. Overall, the results from this serial dilution show both conditions follow the same pattern, but the robot has more variation than manual pipetting (figure 4.5).
Figure 4. Programming for serial dilution in 96 well plate using the Scratch based software.
Upon breaking down these results into their individual replicates (figure 4.6), it is clear that replicate 1 and 2 show almost no variability between conditions. Replicate 3, however, shows huge variability between the robot and manual pipetting, particularly in well 3 and 4. It is theorised that this variability was caused by the serial dilution being conducted in the wrong order or misplacement of the syringe. This suggests that there is an error, either with how the robot is physically built or how its programmed, which could cause this inconsistency between replicates. Replicate 4 shows some recovery to a more consistent level but is not as good as the first two replicates, with well 3 and 4 still being slightly variable. This variability in replicate 4 could be due to carry over. Therefore, further optimisation would be required.
Different cell culture conditions and metabolite extractions

The programming for cell culture assays and metabolite extractions was tested using media and water. It has been demonstrated to work using the LEGO® liquid handling robot, verifying the programmes function as expected. However, before implementing these programmes on biological samples, further work would be required to evaluate these. This should include testing for reproducibility and useability. It would also be important to evaluate the wash step for carry over and potentially add more wash steps if necessary.

4.5 Discussion

The LEGO® liquid handling robot will be evaluated against other DIY and open-source robots, as these have similarities in cost, support, accessibility, and quality.
As previously mentioned, the term “open-source” refers to something individuals can modify and share because its design is publicly accessible, and therefore open-source liquid handling robots can be modified to their users need. The primary advantages of open-source robots are cost, customisation, and ease of use.

Commercialised robots are provided with support from the manufacturers, which includes both the physical robot and programming. They are also highly reliable and capable of high throughput (Gu et al., 2020). These factors command a large price tag, and therefore, open-source automation is on the rise. Whilst open-source robots may not have the same quality and quantity of support as their commercial competitors, they can provide a much lower cost alternative, and can be optimised for a particular task. Open-source robots are a great way to prototype and experiment for general science education, whereas their commercial counterparts are often too expensive to be used for this (Councill et al., 2021). There are several different types of open-source liquid handling robots available, with varying degrees of cost, each with its own strengths and limitations.

Several examples of open-source liquid handling robots will be briefly discussed. One such example is OTTO liquid handler, which was designed in the wake of the COVID-19 pandemic. This is a modular robot, which is capable of automatically preparing samples for qPCR, flow cytometry, and other biological assays that rely on accurate liquid dispensing. OTTO consists of a linear rail system and is made from a combination of 3D printed components and off-the-shelf parts, with a total build cost of approximately $1500 (Florian et al., 2020).

Another is the Evobot, which is also open source with a modular design, constructed with a combination of standard off-the-shelf components, 3D printing, and laser cut parts (Faiña et al., 2020). The Evobot has various uses, one of which is culturing and maintaining Microbial Fuel Cells (MFCs). It has been programmed to create and gather data from MFCs and react on the set thresholds based on a feedback loop (Theodosiou et al., 2017). It has an approximate cost of $500 and each syringe module costs approximately $100 (Faiña et al., 2020).

Pipetting Helper Imaging Lid (PHIL) is another instance of an inexpensive, small, open-source liquid handling robot. This is also constructed from a combination of commercial and 3D printed components. PHIL has many uses, such as tissue
immunostainings, and can be integrated onto a microscope increasing its usability (Dettinger et al., 2022). PHIL costs approximately 550€ to build, pricing itself similarly to the EvoBot.

Another example is the OpenLH, a liquid handler built from a robotic arm, modified pipette and 3D printed parts. This robot costs approximately $1000 to build and is aimed at encouraging bio-enthusiasts to use liquid handling robots (Gome et al., 2019).

A final open-source robot that will be discussed is the open-source DIY Arduino-controlled liquid handling robot, which is made from 3D printed and laser cut parts. This robot has been successfully used for serial dilutions, and if the user has access to a laser cutter can cost as little as $150 to make (Li et al., 2022).

Having given a brief overview of the above open-source robots, they will be compared to the LEGO® liquid handling robot in terms of build, programming, software, and finally, uses.

### 4.5.1 Build

Overall, the build of the LEGO® robot was relatively simple, however there were some issues. The first difficulty arose in the form of the CAD files from the Gerber et al. (2017) paper. These were unable to be opened in LEGO® Digital Designer as intended, so they were instead opened in Studio 2.0. This resulted in software incompatibilities, which led to particular pieces, mostly motors and sensors, being misplaced. However, this was easily overcome by spending time modifying the CAD files and comparing them to the Gerber et al. (2017) robot. The robot is modular, so each individual module was built (syringe, lift, cap, frame and cart), and then assembled together. The modular build allows each individual section to undergo further modification if required, and as LEGO® pieces are readily available and easily connectible, the modifications are easy to implement.
Comparison to Gerber et al. robot

As shown in Table 4.1, the robot underwent several modifications from the original, developed by Gerber et al. (2017). Some modifications, such as relocating the brick, were found to increase the stability of the robot, which had a positive effect on multiple aspects, such as preventing spillages, increased precision, and smoother liquid dispensing.

Other changes, such as the LEGO® holder for the reservoir and 96 well plates, were useful in increasing the function of the robot. For instance, the use of a reservoir was the quickest and easiest way to dispense the liquid, and creating a holder kept the reservoir stable and consistently in the same spot. This reduced spillages and calibration times.

Comparison to other open-source robots

Other open-source robots were constructed from various components, although 3D printing was used as a principal part of their construction (table 4.2).
<table>
<thead>
<tr>
<th>Robot name</th>
<th>Components</th>
<th>Size of x/y platform</th>
<th>Cost of build</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>LEGO® 2D robot</td>
<td>LEGO® 45544 Mindstorms® Education EV3 set Additional LEGO® pieces, Hamilton syringe, Cables from Mindstorms®</td>
<td>260 mm x 180 mm</td>
<td>£325 (~$390)</td>
<td>Gerber et al., (2017)</td>
</tr>
<tr>
<td>OTTO</td>
<td>combination of 3D printed and off the shelf components</td>
<td>500 mm x 600 mm</td>
<td>$1500</td>
<td>Florian et al., (2020)</td>
</tr>
<tr>
<td>EvoBOT</td>
<td>Off-the-shelf components, 3D printing and laser cutting</td>
<td>600 mm x 400 mm</td>
<td>$600</td>
<td>Theodosiou et al., (2017)</td>
</tr>
<tr>
<td>PHIL</td>
<td>commercial and 3D printed components</td>
<td>200 mm x 300 mm</td>
<td>€550</td>
<td>Dettinger et al., (2022)</td>
</tr>
<tr>
<td>open-source DIY</td>
<td>laser-cut acrylic pieces, hand-cut delrin rods, few small 3D printed ABS plastic parts</td>
<td>140 mm x 130 mm</td>
<td>$150 or $400</td>
<td>Li et al., (2022)</td>
</tr>
<tr>
<td>Arduino-controlled liquid handling robot</td>
<td>3D printed parts, a modified pipette and commercially available parts</td>
<td>250 mm x 250 mm</td>
<td>$1000</td>
<td>Gome et al., (2019)</td>
</tr>
</tbody>
</table>
As shown in Table 4.2, 3D printed parts are a consistent component in all the open-source robots discussed in this chapter, except for the LEGO® liquid handler. This is important to recognise; 3D printing allows customised pieces to be produced at a lower cost than commercialised or shop-bought components, however, this can still be more expensive than LEGO®. It can also be inaccessible, because not everyone has access to a 3D printer and the required filament, and those who do have access may not have the knowledge needed to customise a module of a liquid handling robot. This is an important factor to consider, as Professor Manu Prakash illustrates with his focus on accessibility within science (Newby, 2017). In contrast, LEGO® parts are low-cost and accessible worldwide. Individual parts can be bought online and shipped globally, and therefore not only is the initial build cheap, but any modifications or fixes required would also be at a low cost. LEGO® is also able to withstand regular use as a child’s toy, and therefore is potentially more robust than 3D printed pieces. LEGO® is also a fundamentally familiar concept, so the user is already at an advantage, as they understand how to connect LEGO® pieces to construct something.

Another factor to consider is the size, weight, and movability of these robots, as workspace can often be limited in a lab environment, and the ability to transport the robot provides additional flexibility. The size varies considerably, with the DIY Arduino-controlled liquid handling robot being the smallest, with an x/y range of 140mm x 130mm (Li et al., 2022), and the largest being the OTTO robot with an x/y range of 500mm x 600mm (Florian et al., 2020). The LEGO® liquid handler is the second smallest, providing useable space for four well plates at once, whilst not taking up additional space. The weight of these robots is not readily available, but the weight of the LEGO® robot is 1kg. Without seeing each of these robots up-close, it is difficult to state how they would respond to being frequently moved, but from information provided it would appear the OTTO and EvoBot would be better located in one place. The OpenLH, PHIL and DIY Arduino-controlled liquid handling robot all seem easily moveable, which also corresponds to their size.

The LEGO® liquid handling robot is very moveable, it is sturdy and lightweight and can be transported in a straightforward manner. As the LEGO® robot can be AA-battery operated, or charged using a battery pack, this also adds to its flexibility, as it does not need to be located near a power source. During this work the battery pack
was found to be long lasting, approximately 20 hours per charge and took approximately 3.5 hours to charge. It also does not need to be connected to a computer, as the programmes can be pre-loaded onto the LEGO® brick. These factors enable the LEGO® liquid handling robot to be used in various locations/environments, such as different labs, inside a fume hood, or anaerobic chamber. One limitation, however, is that coloured LEGO® blocks are not able to be autoclaved, and so the robot would have to be sterilised with 70% ethanol or UV light. Clear LEGO® blocks can be autoclaved, so this robot could be rebuilt using clear bricks in the future. The manoeuvrability of the LEGO® robot is a big advantage, as it allows use in sterile environments, an issue that even commercial robots can struggle to overcome (Torres-Acosta et al., 2022).

Comparing the robots discussed in this chapter, the LEGO® liquid handling robot is the cheapest to build, modify, and maintain. This makes it the most cost-effective open-source liquid handling robot, as the initial build is low cost, it is quick to build, the modifications are easy, and any breakages would require replacing LEGO® pieces. This is in comparison to replacing custom made or purpose-built parts as with the other open-source robots. However, it is important to state that the LEGO® robot is built from pieces from the standard Educational EV3 LEGO® kit (45544), which is discontinued and now no longer commercially available. However, it did require additional pieces, and it would still be possible to build this robot, as these parts are available online, it would just be more time consuming than the initial build, as pieces would have to be ordered individually rather than provided in one kit.

If comparing how easy it is to build each of these robots, then the LEGO® liquid handler has the upper hand. It can be built by school age children and has been replicated many times with both Gerber et al. (2017) and Tarrés-Puertas et al. (2023), illustrating its reproducibility by school age children. It requires no 3D printing, laser cutting or soldering to build, just connecting LEGO® pieces. Any additional modifications made involved removing or adding additional LEGO® parts, much easier than customising 3D printed or laser cut parts.

The OpenLH only took 1 hour to build, this is due to the main component being an existing robotic arm (Gome et al., 2019). In contrast, EvoBot took 2 weeks to be built and tested, this was done using a research assistant hired for this task (Theodosiou
et al., 2017). PHIL took 7 days for all the 3D pieces to be printed, and then 2.5 – 6 hours for full assembly of the robot by PhD students (Dettinger et al., 2022). The DIY Ardino took 1-3 days for production and assembly time (Li et al., 2022). However, the times noted here are the times recorded by the people who designed the robot, and it can be assumed that external people will take additional time. Replicating robots is not easy as many issues crop up. This was experienced in the recreation of the OTTO robot with a PhD student spending 20+ hours of troubleshooting just to get the correct wiring, with the help of the original designer (Wiseman, 2021). This illustrates the complexities of reproducing these robots, and therefore the time and ease of build should be a major factor when deciding which to pursue.

However, it is also worth noting that there are several improvements that could be made on the LEGO® robot. This could include the addition of customised 3D modules and parts to add extra complexities to the build, for example, the plate holder and reservoir could be replaced by 3D printed parts. Some of the other robots can operate unattended, such as EvoBot and OTTO, and therefore the addition of a camera could aid this and allow work to be carried out remotely (Florian et al., 2020; Faiña et al., 2020). Robots, such as the EvoBot, can work using feedback loops (Faiña et al., 2020), which could be a great addition to the LEGO® liquid handling robot, and could be possible with the integration of sensors.

4.5.2 Electronics and Programming

Electronics

From June 2021 LEGO® retired Home Edition and replaced it with LEGO® MINDSTORMS® EV3 Home app. This features a coding language based on Scratch, and therefore the existing programmes were not compatible. The existing LabVIEW programmes were available to use as they had been uploaded on the Mindstorms® EV3 brick, however they could no longer be modified and calibrated, meaning they were no longer fit for purpose. This resulted in all the programmes needing to be rewritten on new software, using a different coding language.
Each open-source robot needs electronics, and the robots discussed here (except the LEGO® liquid handling robot) all use Arduino boards of differing sizes and processing powers. Arduino is an open-source electronics platform, which is easy to use, for both hardware and software. Arduino boards are able to read inputs and turn them into an output. They can have code uploaded to them, so they do not need to always be plugged into a computer, but they do still need a power source.

Comparatively, for the LEGO® robot, the electronics come in the form of the Mindstorms® EV3 brick. The code can be uploaded to the brick, either by Bluetooth connection or USB-C cable, and once this has occurred the robot can run from anywhere, providing the brick has sufficient charge (from AA batteries or a battery pack).

Overall, the Arduino has more processing power than the Mindstorms® brick (the smallest Arduino board is 8-bit, the same size as the Mindstorms® brick). The Arduino also enables more inputs and outputs, whereas the EV3 brick is limited to four inputs and four outputs. For this particular LEGO® robot this works well, but could perhaps be a limiting factor for future work. The Arduino boards have more versatility as there are many different types and they are also extremely cheap.

It is possible to combine Mindstorms® EV3 and Arduino boards so this could be an option to explore in the future to be able to maximise output, especially if changes to the build were to occur. For example, if optimising the build so that feedback loops would be possible, extra sensors would be required. This would require additional inputs and outputs, but would greatly improve the usability of the robot. This would not be possible with the use of the Mindstorms® EV3 brick alone but could be achieved with the addition of an Arduino board, and would increase the possibilities.

Programming – comparison of LabVIEW to Scratch

As mentioned previously, the initial programming for the LEGO® liquid handling robot was done using Mindstorms® based on LabVIEW software, and then changed to Home app, based on Scratch. Both LabVIEW and Scratch are visual programming languages, which are user friendly, and therefore ideal for a beginner. Both
LabVIEW and Scratch work with a “drag and drop” system, however LabVIEW is icon based, with an illustrated example of which component is being modified. In contrast, Scratch is mostly text boxes. One limitation of both LabVIEW and Scratch software is that the motors are controlled using degree turns. The use of motors was improved compared to the initial robot by Gerber et al. (2017), which relied on sensors as a stop cue; however, it then became vital that the robot was reset to the same starting position each time to ensure correct calibration.

Coding for the EV3 brick can also take place using other coding platforms such as MakeCode and Python, but these options were not explored. This could be investigated further to see if these coding languages would be more useful. Again, this shows the versatility of the LEGO® robot as coding could take place in different coding languages.

**Programming – comparison to more to complex coding languages**

Other open-source robots use a range of different programming languages. This varies from Python (OTTO and EvoBot) to MatLab (PHIL), to combinations such as Python with a block-based interface (DIY LH and Open LH). All these programming languages are adequate and are largely down to personal preference and experience. Proficient coders would most likely prefer to work with Python, but those new to programming would most likely prefer a visual coding language. Ultimately, the coding language or the graphical user interface (GUI) does not impact the robot but does impact its usability, so if the robot would be used by multiple people with varying coding background, then the easier the better.

**4.5.3 Uses**

**Serial dilutions**

As previously mentioned, the LEGO® robot has space for four well plates at once, enabling it to work at a high capacity. It has been programmed successfully for serial
dilutions in both 24 and 96 well plates. As the results in section 4.4.1 show, there are issues with consistency with the serial dilution in 96 well plates, however this is something that can be overcome with further optimisation.

The next step for this experiment would be to perform a serial dilution, using the whole 96 well plate to see what the results would be for the next replicates (rows 5-12). It would be interesting to see whether they follow the pattern of replicate number 4 and continue to decrease in their variability, or whether they follow the trend of replicate 3 and remain variable. Ideally, the problem within the programming could be identified and then modified to prevent this being a continual issue. Another factor to consider is that over time (more than 4 replicates) human error will also occur with pipetting, and the human results may not be as reproducible as they were.

Cell culture assay

One of the other programs developed was the cell culture assays. Whilst the programming for this works well and the LEGO® robot operated as required, this program was not able to be implemented. This was due to biological reasons, as it is not possible to distinguish between the effects of the condition on growth and batch effects, which introduces bias to the results. Batch effects are when non-biological factors in an experiment cause change to the data, therefore it is necessary to distribute samples randomly in well plates to limit biases and reduce technical variability (Čuklina et al., 2020). To overcome batch effects, samples are placed in random block designs. However, the LEGO® robot is unable to do this, as it cannot move to a specific well. Instead, it sets up the cell culture plate by serial dilutions meaning each condition would be in the same row, and therefore undergo batch effects. Therefore, despite the robots programming working well it is not appropriate for scientific use.
Metabolite extractions

Another programme that the LEGO® liquid handling robot has been programmed to complete is metabolite extraction. For both intracellular and extracellular metabolite extraction, there is a potential for contamination between samples. This should be tested to see if the wash step in between samples is sufficient to avoid contamination. Other open-source robots, such as PHIL and the DIY LH, would experience the same issues as they also use the same pipette and syringe multiple times (Dettinger et al., 2022; Li et al, 2022). PHIL uses a single pipette and tested their wash step (pipette washed with target media in an unused well before entering the target well). The measured carry over was 1µL without a wash step, 1 wash step 1:1000, 2 wash steps 1:100,000 and 4 wash steps >1:100,000 (Dettinger et al., 2022). Therefore, it is fair to assume that adding extra wash steps to the LEGO® robot protocol would reduce contamination to an acceptable amount. In contrast, the EvoBot, which also has one syringe, overcomes the issue of contamination by using external pumps to provide each reagent. They also dispense liquid without the needle touching the sample (Faiña et al., 2020), which could also be implemented in the LEGO® robot. However, this would not work for intracellular metabolite extraction, which requires the cell pellet to be resuspended.

As with the cell culture assays, this protocol has not been sufficiently tested, and therefore it is impossible to validate whether there are any issues with inconsistencies.

Compared to the original robot built by Gerber et al. (2017), the robot built in this research does not have the same accuracy and precision. It is presumed that this is because the Gerber et al. (2017) robot would have undergone considerable optimisations prior to publication, and therefore this robot should be capable of the same precision and accuracy at a later date with additional work.
Comparison to other open-source robots

The capabilities of open-source robots are wide, and each robot has been optimised for their own purposes. OTTO, EvoBot, and PHIL have been successfully implemented for biological uses such as qPCRs and flow cytometry (Florian et al., 2020), the monitoring and culturing of microbial fuel cells (Faiña et al., 2020) and tissue immunostainings (Dettinger et al., 2022). Whereas the LEGO® liquid handling robot, OpenLH, and DIY Arduino have mostly been employed as educational robots.

However, as discussed, the LEGO® robot does have the ability to be useful in the right setting. The Gerber et al. (2017) paper proves this type of robot is capable of high precision and accuracy, with delivery of droplets of various sizes down to 2.5μl (25% precision, 8% accuracy) using a 1-mL syringe and down to 0.15μl (15% precision, 8% accuracy) using a 25-μl syringe, (Gerber et al, 2017).

It should also be explored, if changing programming languages could increase the uses of the LEGO® robot. This was also experienced by the EvoBot lab group, as they said, “One major issue is that most chemists or biologists do not have programming skills. We have provided a lot of examples and some templates to perform the experiments requested by our partners. Nevertheless, they are only a small subset of the possible experiments” (Faiña et al., 2020). DIY LH also said, “This interface however was found to be limiting and/or cumbersome to more advanced programmers, which is why we also developed the Python API” (Li et al., 2022). Therefore, this is clearly an issue experienced by many users of open-source robots, and exploring different programming languages could unlock more potential in the LEGO® robot.

4.5.4 Frugal science

As mentioned above, frugal science is about making science accessible to all, and this LEGO® liquid handling robot achieves this aim. LEGO® is easily and readily available worldwide. It is robust and designed to withstand constant use. Modifications are easy to implement as they mostly involved building LEGO® pieces.
Each part of the robot is modular, allowing for further customisation and expansion. The brick can be AA battery operated, or powered using a LEGO® battery pack, which is charged via USB-C cable. This means that the robot can be used in remote areas without power, or a plug socket nearby. There are potential issues with contamination between samples, as with many other liquid handling robots, however this robot can be used in a laminar flow hood, and if future work proves that the wash step is adequate to avoid contamination, then this would prove even more useful.

4.6 Conclusion

The overall evaluation of this LEGO® robot is that in some areas, such as ease of build and ease of programming, it excels. In other areas, other open-source liquid handling robots perform better. It is important to say that this robot still requires further optimisation and development, which provides exciting potential.

When comparing these robots, it is important to consider what factors take priority, for this project the attributes were: cost, size, uses, time taken, and ease of use. These factors have been discussed throughout this chapter.

Overall, the OTTO robot is the best. It has more uses and it is functional and practical in a lab setting. However, it is also the most expensive open-source robot discussed here, and it took the longest time to build. It also has proven hard to replicate as discussed above (Wiseman, 2021). The EvoBot is also a very useful and versatile robot that again has proven use; however, this requires additional modules to make it more useable. This robot also took a long time to build in comparison to others (Theodosiou et al., 2017). PHIL is also good; this is a similar price to the EvoBot, and its small size offers versatility and added uses such as being attached to a microscope (Dettinger et al., 2022). The OpenLH is restricted and is limited in terms of potential modifications to increase its usability. It succeeds at what it set out to achieve, inspiring the field of DIYBio to investigate automation, however in terms of lab uses it may not be the most practical (Gome et al., 2019). Similarly, the DIY Arduino is a useful educational robot but suffers from a difficulty to build, and limited uses (Li et al., 2022).
The LEGO® robot sits middle of the field; it is not advanced as OTTO, EvoBot or PHIL, but offers something unique. It is extremely easy to build and program, and has been replicated several times (Gerber et al., 2017; Tarrés-Puertas et al., 2023), whereas other robots have struggled with replications due to their complexity. Limitations of the LEGO® robot, such as potential contamination was also experienced by other open-source robots that do not have the ability to use disposable pipette tips. It is limited in its uses due to biology (batch effects and contamination), rather than technical failures and therefore for future work finding suitable uses is of upmost importance. It would be ideal to get an experienced coder to program the LEGO® liquid handling robot, to see if they could eliminate the issues with consistency - alternatively all future work could be done with 24 well plates, where this issue was not prevalent. Once these changes have been made, it would be beneficial to re-evaluate the LEGO® liquid handling robot in terms of its competitors.
Chapter 5.

Synthetic biology for styrene production

5.1 Introduction

5.1.1 Styrene background

Styrene is an organic compound, with the chemical formula $\text{C}_6\text{H}_5\text{CH} = \text{CH}_2$. It is a derivative of benzene, and is also known as styrol, phenylethylene, vinylbenzene or cinnamene. It is an unsaturated aromatic monomer (figure 5.1), which occurs naturally in some plants and foods, such as cinnamon, coffee, and peanuts (Steele et al., 1994). It is a colourless, flammable liquid, although when aged can appear yellow. Styrene is volatile and evaporates at 20°C. It has become an important feedstock due to its ability to be polymerised, and is used for the production of various polymers, including polystyrene, styrene-butadiene rubber, and styrene-acrylonitrile resins (Miller et al., 1994).

Approximately 65% of styrene is used for the production of polystyrene. This can be used to create a variety of products, such as packaging, insulation, and consumer goods (James and Castor, 2005). As mentioned above, other important uses of styrene include the production of styrene-butadiene rubber, for products such as car tyres, industrial hoses, and footwear. Another use is the production of styrene-acrylonitrile resins, and unsaturated polyester resins. These polymers are used in a variety of products such as insulation, packaging materials, appliances, and toys (Anastas and Hammond., 2016). Additionally, styrene can be used to form styrene-butadiene latexes, which are used in the production of synthetic rubber, latex paints and coatings (Miller et al., 1994). These are some examples to show the versatility of styrene and explain its importance and significance.
Figure 5.1. The structure of styrene, an aromatic monomer with a benzene ring.

Styrene is one of the top ten bulk petrochemicals in the world, with an expected global market of $62.81 billion in 2026 (Styrene Global Market Report, 2023). However, commercially available styrene is derived from finite petroleum sources. This has huge implications on an environmental scale, as using fossil fuels contributes to greenhouse gas emissions and global warming, therefore there is a growing pressure for environmentally friendly alternatives.

5.1.2 Styrene production

Typical industrial method

More than 85% of styrene is produced, firstly by the alkylation of benzene to ethylbenzene, followed by the direct dehydrogenation of ethylbenzene (Vasudevan et al., 2009). In the alkylation process, benzene is reacted with ethylene in the presence of an acid catalyst, such as hydrochloric acid or sulphuric acid (figure 5.2). This reaction produces ethylbenzene, which is a colourless liquid with a sweet odour. The next step is the dehydrogenation of ethylbenzene. This reaction is endothermic and reversible, which means high temperatures and low pressures are favoured. Styrene is produced by the vapour-phase adiabatic dehydrogenation of ethylbenzene in two reactors using a metal catalyst (iron, cobalt, or nickel) (Woodle, 2006). Steam is fed into the reactor to increase conversion by increasing temperature and lowering the partial pressure of ethylbenzene resulting in a shift of equilibrium. There are several unwanted side reactions, producing by-products such as benzene, toluene, ethylene, and carbon dioxide (figure 5.3). Temperature control
of the reactor is crucial to minimise these side reactions, and therefore the typical operating range is 600–655 °C. The reactor contents are cooled, before being sent to a three-phase separator. Here, the organic layer is sent to distillation columns to separate styrene from the other components. This method usually obtains a yield of approximately 60% (Luyben, 2011).
Figure 5.2. Industrial styrene production from ethylene and benzene. 1 = dehydrogenation step, 2 = dehydrogenated mixture, 3 = ethylbenzene/styrene monomer splitter, 4 = styrene finishing column, 5 = ethylbenzene recovery column, 6 = benzene/toluene splitter. Diagram based on Pérez-Sánchez et al. (2017).
Styrene Reaction:

1. \( C_3H_7CH_2CH_3 \rightleftharpoons C_2H_5CHCH_2 + H_2 \)

Other Reactions:

Benzene/Ethylene Reaction:

2. \( C_3H_7CH_2CH_3 \rightarrow C_6H_5 + C_2H_4 \)

Toluene/Methane Reaction:

3. \( C_6H_5CH_2CH_3 + H_2 \rightarrow C_6H_5CH_3 + CH_4 \)

Carbon Monoxide Reactions:

4. \( 2H_2O + C_2H_4 \rightarrow 2CO + 4H_2 \)
5. \( H_2O + CH_4 \rightarrow CO + 3H_2 \)

Carbon Dioxide Reaction:

6. \( H_2O + CO \rightarrow CO_2 + H_2 \)

Figure 5. 3. Reactions taking place during large-scale industrial styrene production from fossil fuels.

Other methods

Styrene can also be produced by the pyrolysis of ethylene and benzene, or by the oxidation of toluene. However, these methods are not as common as the alkylation and dehydrogenation methods, due to a lower yield of the desired product (Webb and Corson, 1947).

As mentioned above, the production of styrene is a major global industry, and the largest producers of styrene include companies such as Dow Chemical, INEOS, BASF, and Shell Chemical. Millions of tons are produced each year – with an estimated 41 million tons in 2020, representing a >$28 billion market in the US alone (Liu et al., 2018). The huge energy requirements to produce styrene make it one of the most energy intensive chemical production routes. The production of styrene can have significant environmental impacts, including the emission of greenhouse gases and other air pollutants. As a result, there is growing interest in developing more sustainable and environmentally friendly methods of styrene production, including the use of renewable feedstocks, and the development of new catalytic processes,
that reduce the energy requirements and emissions associated with styrene production. One avenue being explored, is bacteria capable of synthesising styrene from renewable sources, as a more sustainable and greener source of styrene and styrene-derived polymers (Liang et al., 2020).

5.1.3 Bioplastics

As previously mentioned, there are considerable environmental impacts from producing plastics, including styrene, from petroleum and fossil fuels. Plastic production has sharply increased over the last 70 years; in 1950, the world produced just two million tonnes. By 2019 this had increased to over 450 million tonnes, producing over 1.6 billion tonnes of greenhouse gases in plastic production and conversion (Ritchie et al., 2023). Whilst there has been an increase in recycling since the 1980s, 79% of all plastic produced has been dumped into landfill sites, 12% has been incinerated and only 9% recycled. The lack of recycling is largely due to limitations of mechanical recycling which converts plastics into new shapes through mechanical force and heat (Rosenboom et al., 2022). Therefore, there is a huge drive for biobased plastics to be developed. Plastics are referred to as biobased if they are made (at least partly) from biomass, often corn, sugarcane or cellulose (Ashter, 2016). These can then be polymerised to produce a replacement for existing polymers, such as polyethylene, or novel polymers, such as polyhydroxyalkanoates (PHAs) (Fernandez-Bunster and Pavez, 2022). These polymers can then go on to produce a variety of consumer goods in the same way petroleum-based ones can. One major advantage of biobased plastics over traditional plastics, is that they reduce reliance on fossil fuels and reduce the environmental effects of plastic production (Rosenboom et al., 2022).

Another approach to biobased plastics is using bacteria to produce plastic. Some bacteria can use organic compounds, such as sugars and starch to produce biopolymers such as PHAs. These can be produced by a variety of bacterial species, such as Streptomyces, Bacillus and Burkholderia with a variety of different carbon sources, such as glucose, lactate, and glycerol (Vicente et al., 2023). A more in-
depth review of plastic produced from bacteria can be found in the paper published by Vicente et al. (2023). Most bacterial producers of PHA show improved yield under nutrient limitation. PHAs can be divided into three subgroups depending on the length of its chain: short, medium, and long. They can also be homopolymers or heteropolymers. Poly-3-hydroxybutyrate (PHB), probably the most well studied PHA is a homopolymer. Materials made from PHB are usually stiff and brittle, with low thermal stability (McAdam et al., 2020).

As well as using bacteria to produce plastic, one other avenue being explored is microbial biodegradation, a process where microorganisms secrete extracellular enzymes that depolymerises the plastic polymer. These monomers can then be metabolised to carbon dioxide and water or upcycled to valuable products (Haider et al., 2019). An example of this a strain of engineered *E. coli* converting PET to vanillin, which can be used in food and cosmetic industry (Sadler et al., 2021).

### 5.1.4 Biological styrene production

As mentioned earlier, styrene can be found naturally in some plants and trees. One example of this is American Styrax (*L. styraciflua*) and Asian Styrax (*L. orientalis*), where styrene was found to be a major constituent (Fernandez et al., 2005). Styrene has also been detected in fungi; *Penicillium camemberti* was found to be able to produce styrene from excess L-phenylalanine in a study by Pagot et al. (2007), who used the presence of $^{13}$C styrene to confirm the conversion of phenylalanine to styrene. Further, styrene can also be produced in yeast, as *Saccharomyces cerevisiae* has been genetically modified to produce styrene (McKenna et al., 2014).

### 5.1.5 *E. coli* as a chassis system

As *E. coli* is easy to genetically manipulate (Blount, 2015), it has been used to produce many different synthetic compounds. This includes producing vanillin from plastic feedstock (Sadler and Wallace, 2021), hyaluronic acid from glucose and
galactose (Woo et al., 2019), and colorants including carotenoids, indigo, anthocyanins, and violacein (Yang et al., 2021).

These examples show the versatility of *E. coli* and based on the growing need for a sustainable source of styrene, McKenna and Nielsen (2011), explored the use of genetically manipulated *E. coli* as a styrene producer.

### 5.1.6 *E. coli* NST74 with a styrene producing plasmid

**Shikimate pathway**

In *E. coli*, the production of aromatic amino acids (phenylalanine, tryptophan, and tyrosine) is tightly regulated (Rodríguez et al., 2014). The shikimate pathway begins with phosphoenolpyruvate (from glycolysis) and erythrose-4-phosphate (from the pentose phosphate pathway) creating 3-deoxy-D-arabino-heptulosonate-7-phosphate (DAHP) through a condensation reaction. DAHP is the first metabolite within this pathway, and occurs through the DAHP synthase enzymes aroF, aroG and aroH, which are feedback sensitive to tyrosine, phenylalanine and tryptophan respectively. Through a series of reactions, DAHP is converted to chorismate (figure 5.5). Chorismate is a branching point for many further biosynthetic pathways, such as synthesis of amino acids (Gibson et al., 1962), folic acid, menaquinone, and ubiquinone (Dosselaere and Vanderletden, 2001). The enzymes of this pathway and their corresponding genes are known and well-studied, especially in *E. coli* (Pittard, 1996).

As a result of its versatility, the shikimate pathway has been engineered to produce a number of high value products in *E. coli*. This ranges from salicylic acid (Lin et al., 2014), alkaloids, flavonoids, and coumarins for use as pharmaceutical precursors (Jiang and Zhang., 2016).
**E. coli NST74**

*E. coli* NST74 is a feedback deregulated phenylalanine overproducer (Tribe, 1987). See chapter 6 for an in-depth review, however an outline is provided in table 5.1. As mentioned, above the shikimate pathway has been heavily modified many times to produce a wide range of high value products, *E. coli* NST74 is one example of this as it has been modified to overproduce L-phenylalanine.

**Table 5.1. The genetic modifications in E. coli NST74 and the effect this has on metabolism (Tribe, 1987). See chapter 6 for a more in-depth review.**

<table>
<thead>
<tr>
<th>Modification in E. coli NST74</th>
<th>Effect</th>
</tr>
</thead>
<tbody>
<tr>
<td>aroH367</td>
<td>DAHP synthase feedback regulation by tryptophan has been reduced</td>
</tr>
<tr>
<td>tyrR366</td>
<td>Regulation of aromatic amino acid biosynthesis and transport is reduced</td>
</tr>
<tr>
<td>tna-2</td>
<td>The breakdown of tryptophan to indole and pyruvate is reduced</td>
</tr>
<tr>
<td>lacY5</td>
<td>Galactosides are no longer transported into the cell</td>
</tr>
<tr>
<td>aroF394(fbr)</td>
<td>DAHP synthase is no longer feedback regulated by tyrosine</td>
</tr>
<tr>
<td>malT384</td>
<td>Malto-oligosaccharides uptake regulation is disturbed</td>
</tr>
<tr>
<td>pheA101(fbr)</td>
<td>The chorismate mutase/prephenate dehydratase enzyme complex is feedback resistant to L-phenylalanine</td>
</tr>
<tr>
<td>pheO352</td>
<td>The chorismate mutase/prephenate dehydratase enzyme complex production is reduced</td>
</tr>
<tr>
<td>aroG397(fbr)</td>
<td>DAHP synthase is no longer feedback regulated by L-phenylalanine</td>
</tr>
</tbody>
</table>

**Styrene producing plasmid**

Work by McKenna and Nielsen (2011), was conducted to introduce genes that would convert phenylalanine to styrene. This was done in several ways, one being the addition of a pTpal-fdc plasmid, containing two genes coding for two enzymes (McKenna et al., 2013) (figure 5.4). The first of these enzymes is phenylalanine ammonia-lyase 2 (PAL2), from *Arabidopsis thaliana*. This means L-phenylalanine is
converted to trans-cinnamic acid as catalysed by PAL2. The second is ferulic acid decarboxylase (FDC1), from *S. cerevisiae* which converts the trans-cinnamic acid to styrene (table 5.2). The addition of these enzymes led to the accumulation of up to 260mg/L of styrene in shake flask cultures (McKenna and Nielsen, 2011).

Table 5.2. The strain and plasmids used in this work and the modifications involved.

<table>
<thead>
<tr>
<th>Strain</th>
<th>Modification</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Plasmid</td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Ptrc</em>99A</td>
<td><em>Ptrc</em>, pBR322 ori, <em>lacIq</em>, <em>AmpR</em></td>
<td></td>
</tr>
<tr>
<td><em>pTpal</em></td>
<td>PAL2 of <em>A. thaliana</em> inserted into the NCol and XbaI site of <em>Ptrc</em>99A.</td>
<td>McKenna et al., 2013</td>
</tr>
<tr>
<td><em>pTpal-fdc</em></td>
<td>FDC1 of <em>S. cerevisiae</em> inserted into the SbfI and HindIII sites of <em>pTpal</em></td>
<td>McKenna et al., 2013</td>
</tr>
</tbody>
</table>
This strain, *E. coli* NST74 and the plasmid, pTpal-fdc, were chosen for this project to investigate further. This was due to the value and impact of styrene production in bacteria, as well as the promising results produced by McKenna and Nielsen (2011) and McKenna et al. (2013).
Figure 5.5. The biosynthetic pathway of styrene. The shikimate pathway is shown in the green box. L-phenylalanine pathway is shown in the blue box. Conversion of L-phenylalanine to styrene is shown in the red box. Pathway genes shown: aroF, aroG, aroH = DAHP synthase; aroB = DHQ synthase; aroD = DHQ dehydratase; aroE = shikimate dehydrogenase; aroK, aroL = shikimate kinase; aroA = 3-phosphoshikimate-1-carboxyvinyltransferase; aroC = chorismate synthase; pheA and tyrA subunits of chorismite mutase; ilvE = subunit of the branched amino acid aminotransferase; aspC = subunit of aspartate aminotransferase; tyrB = tyrosine aminotransferase; PAL2 = phenylalanine ammonia lyase2; FDC1 = ferulic acid decarboxylase. Adapted from EcoCyc; Keseler et al., (2013).
5.2 Aims

The aim of this work was to investigate the metabolism of *E. coli* NST74 without the induction of styrene to get a view of usual metabolism. The next step was to induce styrene and measure the amount produced, changing culture conditions to increase the yield. The final aim was to investigate the metabolism of *E. coli* NST74 when producing styrene.

5.3 Methods

5.3.1 No induction metabolism investigation

**Plasmid integration**

*E. coli* NST74 was co-transformed with a pTpal-fdc plasmid. *E. coli* NST74 competent cells were taken out of the -80°C freezer and thawed on ice. LB agar plates with 50mg/mL ampicillin (AmpC) were incubated in a 37°C static incubator. 5µL of the pTpal-fdc plasmid was added to 50µL of competent cells in an Eppendorf tube. This was then placed on ice for 20 minutes. The Eppendorf tube was then placed in a water bath at 50°C for 30 seconds, and then placed on ice for 5 minutes. 10µL of this mixture was pipetted onto the LB with AmpC agar plate and plated. The agar plate was incubated overnight at 37°C in a static incubator.

**Culturing**

Strains of *E. coli* K12 and NST74 were cultured on LB agar plates for 24 hours at 37°C in a stationary incubator. NST74_pTpal-fdc was cultured on LB with 50mg/mL AmpC in the same conditions. Pre-culture suspensions were then grown; three
colonies from the agar plates were used for each biological replicate, and each condition had 6 replicates. These were grown in 5mL LB media (LB with AmpC for NST74_pTpal-fdc) in 10mL falcon tubes, at 37°C, 180 rpm in a shaking incubator for 18 hours. The OD$_{600}$ for each replicate was then diluted to 0.2 in 10mL LB suspension, and then placed back in the shaking incubator for 6 hours. From each culture, 1.9mL was transferred into two Eppendorf tubes and centrifuged at 1000 x g for 10 minutes at 4°C, leaving a cell pellet and supernatant.

**Flask culture metabolite extraction**

Performed as in Chapter 2. Methods and Materials - Section 2.6.1. Extraction solvents were 100% acetonitrile at 4°C and 100% methanol at 4°C. Samples were normalised by adding the amount of solvent in proportion to the OD$_{600}$.

**IM-q-TOF acquisition, data processing and analysis**


**5.3.2 Stable isotope labelling**

**Culturing**

_E. coli_ NST74 was cultured on LB plates, and _E. coli_ NST74_pTpal-fdc was cultured on LB plate with 50mg/mL AmpC. Pre-culture suspensions were then grown in 5 mL M9 minimal media at 37°C and 160 rpm in a shaking incubator, with 3 colonies from the agar plate in each replicate. Each condition had 4 replicates. After 16 hours the cultures were diluted to an OD$_{600}$ of 0.1 in 30mL of M9 minimal media. Glucose was added equal to the amount present in the media. For the $^{12}$C condition, 100% $^{12}$C glucose was added, and for the $^{13}$C condition a 50:50 mix of $^{12}$C and $^{13}$C glucose
was added. The samples were then placed back in the shaking incubator at 37 °C and 160 rpm for 6 hours.

Flask culture metabolite extraction

Performed as in Chapter 2. Methods and Materials - Section 2.6.1. A 1:3:1 ratio of chloroform: methanol: water was used as the extraction solvent. Samples were normalised by adding the amount of solvent in proportion to the OD$_{600}$.

IM-q-TOF acquisition


Data processing and data analysis

In order to reduce the complexity of data analysis, the ion mobility data was not used, and the data was analysed as if it had been run on an LC-q-TOF system. This meant ignoring the CCS values and therefore identifications are based on m/z library matches. Data analysis took place using Agilent MassHunter Workstation Qualitative 10.0 and MS-Dial version 4.9.221818.

5.3.3 Fast and slow induction

Culturing

*E. coli* NST74_pTpal-fdc was cultured on LB agar plate with 50mg/mL AmpC for 24 hours in a stationary incubator at 37°C. 3 colonies for each replicate were then transferred to M9 minimal media with 50mg/mL AmpC and grown in a shaking incubator at 160 rpm for 18 hours in 50mL falcon tubes. There were 4 replicates for
each condition. The OD$_{600}$ was then taken and diluted to 0.1 in a total volume of 15mL. Two different concentrations of Isopropyl $\beta$-D-1-thiogalactopyranoside (IPTG) were used; 0.2mM and 1mM. Two methods of induction were also tested, a fast and slow protocol.

For the fast protocol, IPTG was pre-warmed to 37°C for 30 minutes, it was then added to the suspension to create final concentrations of 0.2mM IPTG and 1mM IPTG. The final volume was 15mL. This was then placed in a shaking incubator at 37°C, 180rpm for 4 hours. After 4 hours 1mL of dodecane was added to each sample and shaken. 100μL of the dodecane was transferred to GC vials for further analysis.

For the slow protocol, IPTG was pre-warmed to 20°C for 30 minutes. It was then added to the suspension to create final concentrations of 0.2mM IPTG and 1mM IPTG. The final volume was 15mL. This was then placed in a shaking incubator at 25°C, 180rpm for 16 hours. After 16 hours 1mL of dodecane was added to each sample and shaken. 100μL of the dodecane was transferred to GC vials for further analysis.

**Gas Chromatography Quadrupole Time-Of-Flight Mass Spectrometry**

GC/QTOF-MS was conducted using an Agilent 7200B GC/Q-TOF mass spectrometer with Gerstel Multi-Purpose Sampler (MPS) robotics (Anatune). Underivatized samples of 1μL were injected with a 1:10 split ratio into a DB-5ms 30m x 250μm x 0.25μm GC column (Agilent Technologies). Helium was used as the carrier gas at a flow rate of 1 mL/min. The inlet was set to 250 °C, and the GC oven was programmed to hold at 60 °C for 1 min, then increase to 325 °C (at 10 °C/min), where it was held for 15 min. The ion source was set to 230 °C, with 35 μA filament current and 70 eV electron energy. A mass range of 35-600 m/z was scanned at an acquisition rate of 5 spectra/s with a solvent delay of 4 minutes.
Data analysis

Data was normalised a final growth reading. Data analysis took place using Agilent MassHunter Workstation Qualitative 10.0, to identify the styrene peak and acquire the peak area. Data was normalised from a final growth reading. Statistical analysis took place using GraphPad Prism 8.4.2.

5.3.4 24 and 48-hour induction

Culturing

*E. coli* NST74_pTp-al-fdc was cultured on LB agar plates with 50mg/mL AmpC and cultured overnight in MM1 media.

Cultures were grown in 10mL MM1 in 50mL falcon tubes at 32°C in a shaking incubator at 180rpm for 18hours. This was the then diluted to an OD$_{600}$ of 0.1 and grown in 10mL MM1 with AmpC in a 50mL borosilicate glass flask with lid, baffled to help with aeration, at 32°C, 220 rpm which was tightly sealed to prevent styrene escaping. After 6 hours, IPTG was added at a final concentration of 0.2mM. Culturing continued for 48 hours, with samples taken at 24h and 48h. During this time pH was monitored every 6 hours and adjusted with 1M NaOH to keep it at pH7. Dodecane was added at various points, in order to capture the styrene, as in Lee et al. (2019). These are shown in table 5.3.
Table 5.3. Samples taken from the 24- and 48h induction experiment.

<table>
<thead>
<tr>
<th>Induction status</th>
<th>Dodecane status</th>
<th>Replicates at 24h</th>
<th>Replicates at 48h</th>
</tr>
</thead>
<tbody>
<tr>
<td>Uninduced</td>
<td>Dodecane added before freezing</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>Uninduced</td>
<td>Dodecane added after freezing</td>
<td>4</td>
<td>N/A</td>
</tr>
<tr>
<td>Induced</td>
<td>Dodecane added before freezer</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>Induced</td>
<td>Dodecane added after freezing</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>Induced</td>
<td>Grown with dodecane layer</td>
<td>4</td>
<td>4</td>
</tr>
</tbody>
</table>

Gas Chromatography Quadrupole Time-Of-Flight Mass Spectrometry

As explained earlier in this chapter – Section 5.3.3.

Data analysis

As explained earlier in this chapter – Section 5.3.3.

5.4 Results

5.4.1 No induction metabolism investigation

This experiment was conducted to compare the metabolic pathways between *E. coli* strains K12, NST74, and NST74 with an uninduced pTpal-fdc plasmid and to investigate any pathway differences.
This experiment also took place using two different solvents, in order to find which would be the most appropriate for future experiments. Two extraction solvents, 100% acetonitrile and 100% methanol, both at 4°C were tested. These solvents were chosen as they were most compatible with the automation platform that was hoped to be used.
Figure 5.6. Using the solvent 100% acetonitrile, some metabolite differences between the three conditions in pathways linked to L-phenylalanine and therefore styrene. These were significantly different when an ANOVA was performed (p<0.05).
Figure 5. Using the solvent 100% methanol, some metabolite differences between the three conditions in pathways linked to L-phenylalanine and therefore styrene. These were significantly different when an ANOVA was performed ($p<0.05$).
Both figures 5.6 and 5.7 show that significant changes occurred in the pathways of interest, the shikimate pathways and its offshoots. To choose the most appropriate solvent, chromatograms were compared, and it was decided that 100% acetonitrile would be used for further studies. This was due to metabolites of interest being successfully extracted and the data being more reproducible. Therefore, the data for the acetonitrile extraction was analysed in more detail, shown in table 5.4 below. Statistical analysis was performed using ANOVA and post hoc Fishers Least Significant Difference (LSD). Fishers LSD is a statistical method used alongside ANOVA to determine which means are significantly different from each other (Meier, 2006).

Table 5.4. Significantly different metabolites from the 100% acetonitrile extraction, between E. coli K12, NST74 and NST74 with the pTpal_fdc plasmid. Statistical analysis was performed on Metaboanalyst 5.0 using ANOVA with post hoc Fishers LSD (p<0.005).

<table>
<thead>
<tr>
<th>Metabolite</th>
<th>P Value</th>
<th>FDR</th>
<th>Fisher’s LSD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Extracellular</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cytosine</td>
<td>5.38E-05</td>
<td>0.002342</td>
<td>K12 - NST; K12 - PT; NST - PT</td>
</tr>
<tr>
<td>Deoxyadenosine</td>
<td>0.000188</td>
<td>0.005455</td>
<td>NST - K12; NST - PT</td>
</tr>
<tr>
<td>L-2,4-Diaminobutanoate</td>
<td>0.000657</td>
<td>0.011423</td>
<td>K12 – PT; NST - PT</td>
</tr>
<tr>
<td>Hypoxanthine</td>
<td>0.00209</td>
<td>0.03031</td>
<td>K12 - NST; K12 - PT</td>
</tr>
<tr>
<td>Intracellular</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Phenylethanol</td>
<td>2.23E-07</td>
<td>4.80E-06</td>
<td>PT - K12; PT - NST</td>
</tr>
<tr>
<td>Phenylpyruvate</td>
<td>4.40E-05</td>
<td>3.79E-04</td>
<td>NST - K12; PT - K12; PT - NST</td>
</tr>
<tr>
<td>Aminobenzoate</td>
<td>8.36E-04</td>
<td>0.005994</td>
<td>K12 - NST; K12 - PT</td>
</tr>
<tr>
<td>2-Phenylacetic acid</td>
<td>0.0032</td>
<td>0.019659</td>
<td>K12 - NST; K12 - PT</td>
</tr>
<tr>
<td>Anthranilate</td>
<td>0.007816</td>
<td>0.042011</td>
<td>K12 - NST; K12 - PT</td>
</tr>
<tr>
<td>Xanthine</td>
<td>1.36E-06</td>
<td>5.99E-05</td>
<td>NST - K12; PT - K12</td>
</tr>
<tr>
<td>Thymidine</td>
<td>1.93E-05</td>
<td>4.24E-04</td>
<td>NST - K12; PT - K12</td>
</tr>
<tr>
<td>Uridine</td>
<td>2.63E-04</td>
<td>0.00386</td>
<td>NST - K12; PT - K12</td>
</tr>
<tr>
<td>Cytosine</td>
<td>2.76E-15</td>
<td>2.15E-13</td>
<td>K12 - NST; K12 - PT</td>
</tr>
</tbody>
</table>
As can be seen in table 5.4, there were several metabolic differences between *E. Coli* K12, *E. coli* NST74, and *E. coli* NST74 with an uninduced pTpal-fdc plasmid. It was clear that many of these metabolites were close to the shikimate and phenylalanine pathway. This is expected as these pathways have been genetically modified in the NST74 strain.

As the main metabolite of interest, L-phenylalanine was explored, and the results show no significant difference.

Negative polarity mode

![L-phenylalanine](image)

*Figure 5.8. L-phenylalanine area peak for negative mode IM-q-TOF data. A) Extracellular data. B) Intracellular data. Error bars indicate standard deviation.*
Positive polarity mode

![Graph showing L-phenylalanine area peak for positive mode IM-q-TOF data. A) Extracellular data. B) Intracellular data. Error bars indicate standard deviation.](image)

Figure 5.9. L-phenylalanine area peak for positive mode IM-q-TOF data. A) Extracellular data. B) Intracellular data. Error bars indicate standard deviation.

Whilst these results are not statistically significant, this is most likely due to the large variation in each condition. However, both positive and negative mode data show a similar trend for intracellular phenylalanine (figure 5.8 and figure 5.9). There may be an increase in intracellular L-phenylalanine from *E. coli* K12 to *E. coli* NST74 and again from *E. coli* NST74 to NST74 with the pTpal-fdc plasmid.

This work was preliminary work to confirm changes between the three strains and identified targets for future works.
5.4.2 Stable isotope labelling

As shown in figure 5.10, it is clear that the ideal labelling of 50:50 mix was not achieved. This is corroborated in table 5.5, which suggests that the total labelling achieved was approximately 19%.
Figure 5.10. Examples of the peak for glucose in E. coli NST74 samples. A = unlabelled control grown with $^{12}\text{C}$ glucose, B = grown in 50:50 labelled $^{13}\text{C}$ and $^{12}\text{C}$ glucose, as you can see there is different amounts of $^{13}\text{C}$ glucose in these samples and it is not a 50:50 mix between $^{12}\text{C}$ and $^{13}\text{C}$. Therefore, this made tracking further propagation difficult.
Table 5.5. Average peak area for $^{12}$C and $^{13}$C glucose for extracellular samples in negative mode

<table>
<thead>
<tr>
<th>Extracellular Negative Mode Data</th>
<th>Average $^{12}$C Glucose Peak Area [M-H]-</th>
<th>Average $^{13}$C Glucose Peak Area [M+6-H]-</th>
</tr>
</thead>
<tbody>
<tr>
<td>$^{12}$C NST74</td>
<td>53945.4375</td>
<td>0</td>
</tr>
<tr>
<td>$^{12}$C NST74_pTpal-fdc</td>
<td>56330.5925</td>
<td>0</td>
</tr>
<tr>
<td>50:50 $^{12}$C:$^{13}$C NST74</td>
<td>26633.05</td>
<td>5078.4525</td>
</tr>
<tr>
<td>50:50 $^{12}$C:$^{13}$C NST74_pTpal-fdc</td>
<td>27404.4275</td>
<td>5419.7375</td>
</tr>
</tbody>
</table>

As mentioned in section 5.3.2, the ion mobility data, such as CCS values, was not used in the analysis of these samples. This is because the orthogonal data added extra complexity to the data processing. As a result of not using the ion mobility data, the confidence in identifications is decreased.

5.4.3 Fast and slow induction

The results for this experiment indicate that the induction of the pTpal-fdc plasmid in *E. coli* NST74 did not work, as styrene was found to be higher in the controls (n=1), than in the induced samples (n=4) (figure 5.11).
An ANOVA was performed on the data collected, and no significant differences between conditions were found.

A styrene standard was run on the GC-MS at the same time, to check that styrene could be detected through this method of analysis. This was positive, and showed that styrene could be detected through GC-MS. However, this also shows that the induction of the pTpal-fdc plasmid in *E. coli* NST74 resulted in no significant increase in styrene production. This is shown as the styrene area peak is lower in the induced samples than the control.

### 5.4.4 24 and 48-hour induction

Based on an assessment of the induction process used in the previous experiment, many changes were made to the experimental design, such as the media (changing from M9 media to MM1 media), the flask (changing from plastic to borosilicate glass
flasks, baffled to help with aeration and with a glass lid to trap the styrene), and the length of induction time (minimum 24 hours).

A styrene standard curve was conducted (figure 5.12), with dilutions in dodecane. This was done to ensure that styrene could be detected when mixed with dodecane. Spiked samples were also included to ensure detection was possible.

![Styrene standard curve](image)

*Figure 5. 12. A styrene standard curve, this does not follow the expected linear curve. This could be due to saturation in the mass analyser.*

After 24 hours of induction with 0.2mM IPTG, samples were taken. These samples show no significant difference in the amount of styrene present between the different conditions after a one-way ANOVA was conducted as statistical analysis (figure 5.13A).
Figure 5. 13. A) Styrene area peak after 24 hours in different culture conditions, no significant differences were found when an ANOVA test was performed. B) Styrene area peak after 48 hours in different culture conditions, no significant differences were found when an ANOVA test was performed.

After 48 hours, additional samples were taken. As with the 24-hour samples, a one-way ANOVA was conducted on this data, and no significant differences between conditions were found (figure 5.14B). As with the previous experiment styrene was found to be present in all samples, but in small amounts. In the 48-hour samples, there was more styrene present in the uninduced samples than in the samples that have been induced with IPTG. This data was normalised to the final optical density of the flasks.
5.5 Discussion

5.5.1 IM-Q-TOF results

**Culturing three strains, no induction**

This experiment showed that there are significant changes between the three strains, mostly in the pathways of interest, demonstrating that the modifications in the NST74 strain effect metabolism around the shikimate pathway. As this was a preliminary experiment, its intention was to check there were differences, and provide some targets for future work.

Referring back to figure 5.6, it is clear that there is a decrease in the metabolite, 4-aminobenzoate, in *E. coli* NST74 pTpal-fdc. This metabolite is involved in the pABA complex in *E. coli* and is involved in tetrafolate synthesis (Roux and Walsh, 1992). As this is decreased in the NST74-pTpal-fdc and NST74 strain, this could indicate that the flux of this pathway is reduced, and instead directed towards increased L-phenylalanine production, as would be expected.

This is also similar to the decrease seen in anthranilate. Again, this suggests that the flux is reduced in this pathway in order to increase flux towards L-phenylalanine production.

The increase in phenylpyruvate from *E. coli* K12 to NST74 and NST74_pTpal-fdc is as expected as these strains have been genetically engineered to favour this pathway, with the modification in PheA and PheO genes (table 5.1). This also indicates that phenylpyruvate is not a bottleneck for the production of L-phenylalanine in the NST74 strain.

Despite there being no significant differences, there is an increasing trend of intracellular L-phenylalanine as shown in figures 5.8 and 5.9. Between *E. coli* K12 and NST74 this is expected, as *E. coli* NST74 has numerous genetic manipulations to increase the amount of phenylalanine produced. Between *E. coli* NST74 and NST74 with the plasmid there also appears to be an increase in intracellular L-phenylalanine. This is slightly more surprising; however this may suggest that the
addition of the plasmid leads to increase flux of the L-phenylalanine biosynthetic pathway. As the plasmid is uninduced, there should be no conversion of phenylalanine to trans-cinnamic acid, and subsequently styrene, however there may actually be some conversion due to ‘leaky’ expression. ‘Leaky’ gene expression, where expression of a regulated gene is observed in its uninduced state, remains a persistent problem in synthetic biology (Ho et al., 2021).

In this data trans-cinnamic acid could not be detected, and, as this experiment was analysed on an IM-q-TOF, styrene could not be detected, so it is not possible to confirm if this is the case.

**Stable isotope labelling**

Stable isotope labelling is useful to study metabolic fluxes and determine metabolic pathways. The major benefit of stable isotope labelling is its ability to track and trace the biosynthesis, conversion, and degradation of metabolites within a living system (Triebl and Wenk., 2018). It can also enable the study of metabolic fluxes and enable metabolic pathways to be determined and followed (Chokkathukalam et al., 2014).

In the case of *E. coli* NST74 it would be particularly interesting to see how the genetic changes alter carbon flux in relation to the production of aromatic amino acids, and how the flux is altered.

The stable isotope experiment did not demonstrate propagation of the labelled carbon into the shikimate pathway – the pathway of interest. The aim was 50% $^{13}$C labelling, and less than 20% was achieved after 6 hours of culturing (table 5.5). This makes it much harder to track carbon flux through the pathways as the amount present is so low. However, there is evidence of labelling in glycolysis (figure 5.10) and future work – with the correct amount of labelled glucose – would most likely provide interesting data on how carbon flux through the shikimate and phenylalanine pathways work.

Crown et al. (2015) used $^{13}$C glucose for metabolite flux analysis in *E. coli* K-12 MG1655 and found labelled carbon in amino acids four to six hours after the addition of labelled glucose. However, other studies cultured cells for a longer time-period;
Wolfsberg et al. (2018) used 100% $^{13}$C glucose with *E. coli* BW25113 and *E. coli* ΔackA cultured for 18 hours. This allowed the tracking of carbon flux to amino acids, such as valine and glutamate. Therefore, for future work with *E. coli* NST74, a longer culture time should be tested, alongside a higher percentage of labelled glucose.

5.5.2 GC-MS results

**Fast and Slow Induction**

The uninduced control had more styrene present than the induced samples, this is unexpected and suggests styrene contamination, or leaky expression of the plasmid in the controls. In this work the fast protocol samples were incubated for 4 hours, whilst the slow protocol samples were incubated for 16 hours. McKenna and Nielsen (2011) stated that there was not enough accumulation of L-phenylalanine for conversion to styrene until culturing had occurred for 17 hours. Therefore, several changes were implemented to increase the likelihood of detecting styrene in the next experiment. These changes included culturing for longer time-periods and changing the media from M9 media to MM1 media to provide optimal growth media. Borosilicate flasks with lids were also used to prevent styrene evaporating. The flasks were baffled to enable easier aeration.

**24 and 48-hour induction**

The standard curve based upon the styrene standard was not linear (figure 5.12), the presumed reason for this is that at the high concentration, the GC-MS was over saturated, and so the level of detection may not be representative.

Despite all the changes in the culturing method, the results from the GC data showed that the different culture conditions and time of the induction, had no significant effect on styrene production (figure 5.13). Therefore, it was concluded that
controlled induction of the plasmid was not taking place. The pTpal-fdc plasmid is high-copy and whilst these are useful for producing large quantities of plasmid DNA, they are not sufficient at tightly regulated gene expression, this can lead to “leaky” expression (Gruber et al., 2008). Based on the results produced in this study, the uninduced sample is exhibiting leaky expression, hence the presence of styrene in this sample.

IPTG is a molecular analog of allolactose and removes a repression from the lac operon of the plasmid, leading to gene expression (Myung et al., 2020). The work by McKenna and Nielsen (2011) demonstrates that styrene could be produced from the induction of the plasmid through IPTG, and could produce styrene at titres of 260mg/L. However, other research on E. coli strain JM109(DE3) with a plasmid, found that the addition of IPTG had a negative effect on growth and cellular viability (Gomes et al., 2020). Therefore, it is possible that the addition of the IPTG was having a detrimental effect on the E. coli NST74.

5.6 Conclusion

As mentioned earlier in this chapter, there is a drive to increase the number of biobased plastics, and the modifications by Nielson and McKenna (2011) to convert phenylalanine to styrene definitely fulfilled this aim. They have created many different plasmids, and have achieved a yield of 260mg/L, (McKenna et al., 2011) with similar research by Lee et al. (2019) providing yields of 5.3g/L in a fed batch cultivation and 1.7g/L in a shake flask. This shows microbial production of styrene is possible, and with further optimisation could be commercialised.

Nonetheless, there are limitations to microbial styrene production. Firstly, it seems very difficult to culture styrene from these plasmids. Many different factors were changed in order to successfully produce styrene from this strain, and it was mostly unsuccessful. This could be for a variety of reasons, such as experimental error and less than ideal culture conditions, but ultimately the nature of styrene also adds to its difficulty. Styrene is volatile and evaporates easily. It is typically detected by GC-MS or flame ionisation detection (Agency for Toxic Substances and Disease Registry,
1992) and is not detectable using LC-MS which was a limiting factor in terms of analysis and in the design of experiments.

Throughout the experiments discussed in this chapter, it became important to evaluate styrene as a product, and whether it was the ideal metabolite to be studying. It soon became clear that biobased plastics are a controversial issue.

Whilst styrene, derived from *E. coli* NST74 and a plasmid would be classed as a biobased plastic, and offer a lower environmental effect than styrene produced from petroleum, it would still have a negative environmental impact. This is because it is not biodegradable.

Biodegradable in this context refers to plastics that can be broken down through microbial or enzymatic decomposition processes, resulting in natural by-products such as gases (CO₂, N₂), water, and inorganic salts. Factors such as temperature, duration, presence of microorganisms, oxygen, moisture and nutrients all affect biodegradation of plastics (Deconinck and Wilde, 2013; van den Oever et al., 2017). However, fossil fuel-based plastics such as polybutylene adipate terephthalate and polycaprolactone are biodegradable, adding to the complexity of the issue (Folino et al., 2020). An example of this is cyanobacteria species, *Phormidium lucidum* and *Oscillatoria subbrevis*, were found to be capable of degrading low density polyethylene sheets (Sarmah et al., 2018). However, as cyanobacteria needs light to survive it is also possible that degradation of these plastics is restricted to surface-accessible particles (Barone et al., 2020).

When analysing the sustainability of plastic production, it is important to consider each stage, from creation to degradation (figure 5.14). Both bio-based and biodegradable plastics are more environmentally friendly than conventional plastics, but it is also important to consider the end life cycle too. Whilst some biobased plastics are biodegradable, like PHA (and PHB) and poly lactic acid (PLA), others such as polytrimethylene terephthalate (PTT) are not (Hiraga et al., 2019).
Due to the above-mentioned factors, it was decided to cease exploring styrene as a high value product and switch to L-phenylalanine instead, as *E. coli* NST74 is a genetically engineered feedback deregulated overproducer (Tribe, 1987). Unlike styrene, L-phenylalanine can easily be detected on MS instruments, making analysis much more straightforward.
Chapter 6.

Using an inhibitor to alter the flux of L-phenylalanine in an engineered *E. coli* strain

6.1 Introduction

6.1.1 *E. coli* NST74

*E. coli* NST74 is a highly modified strain of *E. coli*, patented in 1987 by David Tribe. In wild type *E. coli*, the organism synthesises L-phenylalanine as required, resulting in low titres of L-phenylalanine (Gu et al., 2023). Therefore, Tribe introduced modifications to enable *E. coli* to produce L-phenylalanine in excess, subsequently increasing the levels extracellularly (Tribe, 1987).

Typically, phenylalanine is produced from intermediates from glycolysis and the pentose phosphate pathway, combining to make 3-deoxy-D-arabinoheptulosonate 7-phosphate (DAHP), which then goes through the shikimate pathway resulting in chorismate (Liu et al., 2019). Chorismate is then converted to phenylalanine through a series of reactions (figure 6.1). L-phenylalanine is an amino acid with a market of $660 million in 2021 (Future Market Insights, 2023). The demand for this comes from the food and beverage industry, where it is used in the synthesis of aspartame (Sprenger, 2007). It is also used in the synthesis of pharmaceuticals, as a treatment for vitiligo (Antoniou et al., 1989) and in HIV protease inhibitor anti-inflammatory drugs and rennin inhibitors (Bongaerts et al., 2001).

The patent by Tribe (1987) explains how *E. coli* NST74 was modified to become a feedback deregulated L-phenylalanine overproducer. This is achieved through a series of modifications, with the primary alteration being the inhibition of DAHP synthase from L-phenylalanine. The genotype for *E. coli* NST74 is: aroH367, tyrR366, tna2, lacY5, aroF394 (fbr), malT384, pheA101 (fbr), pheO352, aroG397 (fbr) (Tribe, 1987). This is a heavily modified strain, and therefore each of these modifications will be discussed below.
As mentioned above, DAHP is the first step of the shikimate pathway, and is formed by combining phosphoenolpyruvate and d-erythrose 4-phosphate. This occurs through a reaction involving an enzyme called DAHP synthase, and through subsequent reactions, leads to the biosynthesis of aromatic amino acids (Srinivasan and Sprinson, 1959). There are three isozymes of DAHP synthase, each specifically feedback regulated by tyrosine, phenylalanine, or tryptophan (Smith et al., 1962; Zurawski et al., 1981). AroH is the DAHP synthase isoform, which is sensitive to tryptophan, and therefore mutations in aroH lead to tryptophan-insensitive DAHP synthase (Pittard et al., 1969; Ray et al., 1988).

**TyrR**

TyrR gets its name from "tyrosine repressor". It is a dual transcriptional regulator of the tyrR regulon and involves genes that are essential for aromatic acid biosynthesis and transport (Wallace and Pittard, 1969; Brown and Somerville, 1971; Pittard et al., 2005). TyrR can act both as a repressor and as an activator of transcription (Pittard and Davison, 1991). As the NST74 strain has a mutation in this gene, it will be devoid of this regulation, and therefore will continue to produce aromatic acids, including L-phenylalanine.

**Tna-2**

The tna-2 gene codes for the enzyme tryptophanase 2. This breaks down tryptophan to indole and pyruvate (Gish and Yanofsky, 1993). In *E. coli* NST74, there is a mutation in this gene, which aims to prevent this reaction from occurring, and therefore prevents the breakdown of tryptophan.
LacY5

LacY is a lactose/proton symporter, responsible for the uptake of lactose and other galactosides. It is a well-characterised secondary transporter and was the first transporter gene to be cloned and sequenced (Teather et al., 1978; Büchel et al., 1980). A mutation here would reduce the uptake of these galactosides, presumably to ensure glucose or a similar sugar would be used as an alternative, and this would start the chain of shikimate pathway reactions, eventually leading to L-phenylalanine production.

aroF394 (fbr)

As mentioned above, DAHP synthase is vital in the synthesis of aromatic amino acids. AroF is one of the isoenzymes, and is feedback regulated by tyrosine. Mutations in aroF lead to tyrosine insensitive DAHP synthase. Transcriptional control of aroF is mediated by the tyrR repressor with the end products of the pathway acting as co-repressors (Wallace and Pittard, 1969; Davies and Davison, 1982). In E. coli NST74, this mutation means DAHP synthase is no longer feedback regulated, and instead is feedback resistant to tyrosine, and will continue to produce tyrosine in excess.

malT384

malT is a transcriptional regulator that positively regulates the transcription of the maltose regulon required for uptake and catabolism of malto-oligosaccharides (Boos and Shuman, 1998). The mutation here disrupts this mechanism.
PheA101 (fbr)

The bifunctional chorismate mutase / prephenate dehydratase complex carries out the first step in the biosynthetic pathways for the aromatic amino acids' tyrosine and phenylalanine. This involves chorismate becoming prephenate through claisen rearrangement, catalysed by chorismate mutase. Prephenate can then undergo either decarboxylation/dehydration to phenylpyruvate, catalysed by prephenate dehydratase, or prephenate can undergo decarboxylation/dehydrogenation to p-hydroxyphenylpyruvate catalysed by prephenate dehydrogenase (Zhang et al., 1998).

This mechanism is usually feedback regulated by L-phenylalanine by an allosteric mechanism (Dopheide et al., 1972). In wild type E. coli, there is almost complete inhibition of prephenate dehydratase at concentrations of L-phenylalanine greater than 1mM (Nelms et al., 1992). Therefore, to stop this becoming a bottleneck, the gene was modified to prevent feedback inhibition of L-phenylalanine, allowing L-phenylalanine production to continue.

pheO352

pheO is an operator constitutive mutation close to the structural gene pheA; therefore, a pheO mutant would reduce the amount of the chorismate mutase / prephenate dehydratase complex to be present (Im and Pittard, 1971).

aroG397 (fbr)

This is the gene for the DAHP synthase isoform for phenylalanine. DAHP synthase is active early in the pathway. The mutation here makes this strain feedback resistant to phenylalanine, and therefore the reactions can continue through the shikimate pathway and result in increased yield of L-phenylalanine.
Not only is *E. coli* NST74 able to overproduce phenylalanine, it is also a prototrophic strain and therefore does not need expensive additives to media to enable cell growth. This, along with its ability to produce phenylalanine at a concentration of 4g/L in a continuous process (Tribe, 1987), made it an ideal model organism.

Due to the above-mentioned modifications, the metabolism of *E. coli* NST74 was already of interest, and therefore it was decided to see if this metabolism could be influenced with the use of inhibitors, to block certain pathways to increase the yield of L-phenylalanine further.
**Figure 6.1. L-phenylalanine pathway. Mutations in E. coli NST74 are shown in red boxes.**

G6P = glucose 6-phosphate, F6P = fructose 6-phosphate, F1-6P = fructose 1,6 bisphosphate, G3P = glyceraldehyde-3-phosphate, PEP = phosphoenolpyruvate, PYR = pyruvate, ACoA = acetyl coenzyme A, TCA = Tricarboxylic acid, OAA = oxaloacetic acid, 6PGL = 6-phosphoglucono-lactone, 6PG = 6-phosphogluconate, RU5P = ribulose-5-phosphate, 2D3D6PG = 2-dehydro-3-deoxygluconate-6-phosphate, X5P = xylulose-5-phosphate, R5P = ribose-5-phosphate, S7P = sedoheptulose-7-phosphate, E4P = erythrose 4-phosphate, DAHP = 3-deoxy-D-arabino-heptulosonate 7-phosphate, DHQ = 3-dehydroquinate, DHS = 3-dehydroshikimate, SHIK = shikimate, S3P = shikimate-3-phosphate, EPSP = 5-enolpyruvyl-shikimate 3-phosphate, CHA = chorismate, PPA = prephenate, PPY = phenylpyruvate, HPP = 4-hydroxyphenylpyruvate, ANTA = Anthranilate, PRAA = n-(5-phospho-b-D-nbosylanthranilate), I3GP = indole 3-glycerolphosphate
6.1.2 Importance of L-phenylalanine

L-phenylalanine is an aromatic amino acid with a formula C\textsubscript{9}H\textsubscript{11}NO\textsubscript{2}. It is used as a nutritional supplement and as a precursor for the production of various catecholamines, and the artificial sweetener aspartame, commonly found in fizzy drinks (Czarnecka et al., 2021). It can also be used in pharmaceuticals, dietary supplements, nutraceuticals, and as an ingredient in cosmetics (Ding et al., 2016).

L-phenylalanine world production exceeds 15,000 tons per year, with main production capacities in China (Sprenger, 2007; Rui and Yonghua, 2011). Currently, L-phenylalanine is obtained by chemical, enzymatic, or microbial processes. Recently there has been increased interest in producing L-phenylalanine through microbial fermentation, especially by engineered strains of \textit{E. coli}, which often have controlled characteristics and high growth rates (Liu et al., 2019).

In \textit{E. coli}, the active transport of phenylalanine can be performed by two different systems, aroP and pheP. AroP is the aromatic amino acid transfer system (Brown, 1970; Piperno and Oxender, 1968), and pheP is the phenylalanine specific transport system (Brown, 1970; Whipp et al., 1980). AroP, the common transport system, has relatively broad specificity, but shows a high affinity for phenylalanine, tyrosine, and tryptophan. PheP is a phenylalanine transporter that is a member of the Amino Acid-Polyamine-Organocation (APC) superfamily of transporters (Cosgriff et al., 2000). Therefore, with two transport mechanisms L-phenylalanine is easily exported extracellularly, which is desired for ease of harvesting.

6.1.3 Using inhibitors to alter metabolism

Previous work, conducted by Rana et al. (2022), showed inhibitors could be used to increase the yield of wheat. This was echoed by Khan et al. (2014), who used inhibitors to increase the yield of potatoes. In a more biological setting, Jutras et al. (2016) exhibited the use of a protease inhibitor to increase the yield of a monoclonal antibody produced in \textit{Nicotiana benthaminana} leaves. Inhibitors of the cell cycle have also been used to selectively target CDK4/6, resulting in optimisation of cell...
growth, product per cell and glycosylation profile in antibody production in Chinese hamster ovary (CHO) cell cultures (Du et al., 2014). Despite these, the use of inhibitors to increase the yield of a high value product seems under-utilised. In this work, inhibitors were used to target different enzymes to see if this would drive metabolic flux towards increase L-phenylalanine production.

As explained in chapter 5, the shikimate pathway has many different branches separate from L-phenylalanine production. Each of the branch point enzymes must be carefully regulated to distribute chorismate properly to various downstream intermediates. In this highly modified strain, the purpose of the mutations is to increase the yield of L-phenylalanine, and therefore an alternative way to do this would be to influence the other branch points. This would increase the amount of chorismate available for conversion to L-phenylalanine.

Target enzymes were chosen based on their proximity to the chorismate pathway. These enzymes were: anthranilate synthase, isochorismate synthase, chorismate pyruvate lyase, 2-amino-4-deoxychorismate synthase, and prephenate dehydrogenase.

6.1.4 Targets to inhibit

One such target was anthranilate synthase. The anthranilate synthase enzyme consists of two subunits: α and β and forms a α2β2 tetramer. The α-subunit, also known as TrpE possesses a catalytic domain. This catalyses the conversion of chorismate and glutamine to anthranilate, glutamate and pyruvate. The β-subunit, also known as TrpGD, catalyses the transfer of the phosphoribosyl group of 5-phosphorylribose-l-pyrophosphate to anthranilate, forming N-phosphoribosylanthranilate (Kanno et al., 2005; Balderas-Hernández et al., 2009).

The end product of the pathway, tryptophan, can act as a feedback inhibitor, as it is able to bind to an allosteric site in TrpE, reducing the enzyme’s catalytic activity. This feedback inhibition mechanism helps to maintain the balance between the synthesis of tryptophan, and the availability of its precursor, chorismate (Castro-López et al., 2022; Bhagat et al., 2022).
As already mentioned, anthranilate synthase is regulated by feedback inhibition of tryptophan, therefore it was decided to supplement the media with L-tryptophan. This would test whether feedback inhibition was a good inhibition mechanism to increase the production of L-phenylalanine in *E. coli* NST74.

### 6.2 Aims

The aim of this work was to evaluate the metabolism of *E. coli* NST74, when grown with an inhibitor of anthranilate synthase, L-tryptophan, and see what effect this could have on the yield of L-phenylalanine. This was first done in 50mL flasks before transferring to 96 well plates to increase throughput. Off target effects were also investigated through untargeted metabolomics on the IM-q-TOF instrument. The addition of L-tryptophan at different time points was also investigated.

### 6.3 Method

#### 6.3.1 Preliminary experiment

**Culturing**

*E. coli* NST74 was taken from glycerol stocks and revived on a LB agar plate, at 37°C for 24 hours in a stationary incubator. Three colonies were then taken from the agar plate and placed in 10mL MM1 media for 18 hours, at 37°C, 180rpm. This was one replicate. After 18 hours, the OD\(_{600}\) was diluted to 0.1 in 10mL MM1. pH was monitored and adjusted with 1M NaOH to keep it at pH 7. The desired concentration of L-tryptophan (control – no tryptophan, 10µM, 100µM and 1mM added L-tryptophan) was added to the *E. coli* NST74 in MM1 media to volume of 10mL in 50mL falcon tubes. This was then incubated at 37°C, 180rpm for 6 hours in a shaking incubator at 180rpm.
Flask culture metabolite extraction

As in Chapter 2 Methods and Materials - Section 2.6.1. 100% acetonitrile at 4°C was used as the extraction solvent.

Targeted mass spectrometry data acquisition and analysis

As in Chapter 2 Methods and Materials – Section 2.7.

6.3.2 Addition of L-tryptophan to cell culture in 96 well plates

Culturing

E. coli NST74 was taken from glycerol stocks and revived on LB agar at 37°C for 24 hours in a stationary incubator. Three colonies were then taken from the agar plate and placed in 10mL MM1 suspension for 6 hours, this was one replicate. After 6 hours, the OD$_{600}$ of the samples was taken using 1mL of the sample in a cell density meter. The 96 well plate (Costar 96 Flat Bottom Transparent Polystyrene) was then set up and contained a total volume of 200µL in each well plate. This consisted of MM1 media, the sample at an OD$_{600}$ of 0.1 and the inhibitor at the desired concentration.

The 96 well plate was then placed in a Tecan infinite® 200Pro plate reader with a target temperature of 37°C, shaking (linear) duration of 550s prior to taking a reading, resulting in an OD reading every 10 minutes until stationary growth phase was reached.

The plate was then removed from the incubator and centrifuged at 1000 x g for 5 minutes. The supernatant was then transferred to a fresh 96 well plate, resulting in one plate with a cell pellet and another with supernatant.
Table 6.1. The different concentration of L-tryptophan added to media.

<table>
<thead>
<tr>
<th>Inhibitor</th>
<th>Concentrations added</th>
</tr>
</thead>
<tbody>
<tr>
<td>L-tryptophan</td>
<td>0, 10µM, 100µM, 1mM, 10mM</td>
</tr>
</tbody>
</table>

96-well plate metabolite extraction

For these extractions the solvent was 50:50 acetonitrile: water. During handling, the solvent and samples were kept on ice.

Intracellular metabolite extraction

Samples were normalised by adding the amount of solvent in proportion to the OD_{600}. The highest OD_{600} had 200µL of solvent added, whilst the other samples received a proportionate amount. The cell pellet was resuspended by pipetting up and down. The 96 well plate was then placed on a plate rocker at 4°C for 1 hour. It was then centrifuged at 1000 x g for 5 minutes. 75µL of the supernatant was then transferred to an Axygen™ 96 well PCR plate and an Axygen Axymat™ silicone sealing mat was placed on top to seal it. The remaining sample plate was sealed and kept at -80°C in case of further analysis.

Extracellular metabolite extraction

For the extracellular metabolite extraction samples, 190µL was added to a fresh plate and 10µL of supernatant was added. The solvent was 50:50 acetonitrile: water at 4°C. This was then placed on the plate rocker at 4°C for 10 minutes and then centrifuged at 1000 x g for 5 minutes. 75µL of the extraction was then transferred to an Axygen™ 96 well PCR plate and an Axygen™ Axymat™ silicone sealing mat was
placed on top, awaiting MS analysis. The remaining sample plate was sealed and kept at -80°C in case of further analysis.

**Targeted mass spectrometry data acquisition and data analysis**

As in Chapter 2 Methods and Materials – Section 2.7.

**Untargeted mass spectrometry data acquisition and data analysis**

As in Chapter 2 Methods and Materials – Section 2.8.

**6.3.3 Addition of L-tryptophan with time delay**

All methods were the same as above, but the L-tryptophan was added when the *E. coli* NST74 had reached stationary phase, at 10 hours of growth. Metabolite extraction and targeted mass spectrometry methods were performed as in Chapter 2 - Sections 2.6.2 and 2.7 respectively.

**6.4 Results and Discussion**

**6.4.1 Preliminary experiment**

A preliminary experiment was performed using *E. coli* NST74, grown in MM1 media with three different added concentrations of tryptophan. The results from this indicate that extracellular L-phenylalanine increased with higher concentrations of L-tryptophan. The conditions with 100µM added L-tryptophan and 1mM L-tryptophan proved to be significantly different to the control, with p values of 0.0291 and 0.0170 respectively (figure 6.2A). Despite not being significantly different, the third lowest amount of added tryptophan (10µM, p= 0.3557) was still part of the trend; increasing L-tryptophan concentration leads to an increase in L-phenylalanine concentration.
Figure 6.2. The average peak area for L-phenylalanine in samples of E. coli NST74 grown with added L-tryptophan. Error bars indicate standard deviation. A) Extracellular L-phenylalanine. Two of these results are significant with the results of a t-test compared to the
control sample, *E. coli* NST74 grown with no added L-tryptophan. B) Intracellular L-phenylalanine. These results show no significant differences between conditions.

Table 6.2. Mean peak area of extracellular L-phenylalanine in each condition and the fold change compared to the control.

<table>
<thead>
<tr>
<th>Condition</th>
<th>Mean peak area</th>
<th>Fold change</th>
</tr>
</thead>
<tbody>
<tr>
<td>NST74 + MM1</td>
<td>647848</td>
<td>-</td>
</tr>
<tr>
<td>NST74 + 10µM</td>
<td>944435</td>
<td>1.46</td>
</tr>
<tr>
<td>NST74 + 100µM</td>
<td>1240856</td>
<td>1.92</td>
</tr>
<tr>
<td>NST74 + 1mM</td>
<td>2017775</td>
<td>3.1</td>
</tr>
</tbody>
</table>

The preliminary data for the extracellular samples suggest that the added tryptophan of 1mM can cause a 3.1-fold increase in the level of L-phenylalanine (table 6.2).

Intracellular data indicated no significant differences in the peak area of L-phenylalanine between conditions (figure 6.2B).

Discussion

The increase of L-phenylalanine found in extracellular samples is large, and therefore the addition of L-tryptophan could prove beneficial in regard to increasing the yield of L-phenylalanine in *E. coli* NST74. As intracellular L-phenylalanine did not show any statistically significant changes, this could suggest that as excess L-phenylalanine was produced it was transported out of the cell, in order to maintain homeostatic balance. If this is correct, then it also proves beneficial as it makes harvesting the product easier. However, the intracellular data is highly variable within each condition. As *E. coli* NST74 appears to easily transport L-phenylalanine out of the cell, it also has potential to be grown in large bioreactors in continuous and feedback cultures.

However, there are limitations to this preliminary data. There are three replicates for the control, 10µM addition and 1mM addition and only two for the 100µM addition
due to one sample failing to grow. This means the data is not particularly robust and more replicates are required.

In addition, whilst the data suggests that adding L-tryptophan does increase the concentration of L-phenylalanine, it is not clear through which mechanism this increase takes place. For instance, it could occur through tryptophan inhibiting anthranilate synthase through feedback inhibition, and therefore increase metabolic flux towards other branches of the pathway. One other possibility is the addition of L-tryptophan aids the growth of *E. coli* NST74, and this is the cause of the increase in L-phenylalanine. This idea could be overcome by performing a metabolite extraction normalised by growth rate for future work. Another hypothesis is that as *E. coli* NST74 no longer needs to produce tryptophan as it is present in the media it can direct flux towards the synthesis of other amino acids. Ultimately, however, the goal is to increase the product and have no negative off target effect or effects on growth. This led to a more comprehensive experiment being designed and completed (Chapter 6, Section 6.4.2).

### 6.4.2 Culturing with added L-tryptophan – Targeted MS

This experiment followed the preliminary experiment. The cell growth occurred in 96 well plates so that growth data could be obtained. Different concentrations of L-tryptophan were added, 0µM, 10µM, 100µM, 1mM, and 10mM (table 6.1). During metabolite extraction, the samples were normalised by optical density.
Growth curves

Figure 6. Growth curves from a 96 well plate containing *E. coli* NST74 with a control and 4 different concentrations of added L-tryptophan.

The culture in the 96 well plate was grown in a plate reader, with readings taken every 10 minutes, to enable thorough understanding of the growth. All conditions reached stationary phase of growth at approximately 10 hours, indicating that the added L-tryptophan has no effect on the time taken to reach stationary phase. However, the growth rate is not the same, the two lowest concentrations of added tryptophan, 10µM and 100µM, grew at the same rate as the control, whilst the two
higher concentrations showed increased growth (figure 6.3). The average OD$_{600}$ after 13 hours is shown in table 6.3, where this increase can be illustrated further.

**Table 6.3.** Average optical density of E. coli NST74 with added L-tryptophan.

<table>
<thead>
<tr>
<th>Condition</th>
<th>Average optical density after 13 hours (n=4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>NST74 (control)</td>
<td>0.428</td>
</tr>
<tr>
<td>NST74 + 10µM</td>
<td>0.429</td>
</tr>
<tr>
<td>NST74 + 100µM</td>
<td>0.437</td>
</tr>
<tr>
<td>NST74 + 1mM</td>
<td>0.457</td>
</tr>
<tr>
<td>NST74 + 10mM</td>
<td>0.515</td>
</tr>
</tbody>
</table>

During metabolite extraction, the samples were normalised by optical density to account for these differences in growth as mentioned in the methods above.

**Extracellular L-phenylalanine**

The data for extracellular samples showed there was a significant difference in L-phenylalanine peak area, correlating to higher concentrations of L-tryptophan. In two of the four conditions, the 1mM and 10mM added L-tryptophan, there was a significant difference using a t-test when compared to the control (p<0.0001 and p<0.0006 respectively). This is illustrated in figure 6.4A. The other two conditions also showed an increase in L-phenylalanine, although it was not significant (figure 6.4B).
Figure 6.4. A) Differences in L-phenylalanine peak area in extracellular samples of E. coli NST74 with added tryptophan. Two conditions, 1mM and 10mM added L-tryptophan are significantly different to the control. Error bars indicate standard deviation. B) Heat map for the L-phenylalanine extracellular samples.
Table 6.4. Mean peak area of extracellular L-phenylalanine in each condition and the fold change compared to the control.

<table>
<thead>
<tr>
<th>Condition</th>
<th>Average Peak Area for extracellular L-phenylalanine</th>
<th>Fold change</th>
</tr>
</thead>
<tbody>
<tr>
<td>NST74 + MM1</td>
<td>24212.76</td>
<td>-</td>
</tr>
<tr>
<td>NST74 + 10μM</td>
<td>28237.94</td>
<td>1.17</td>
</tr>
<tr>
<td>NST74 + 100μM</td>
<td>27245.91</td>
<td>1.13</td>
</tr>
<tr>
<td>NST74 + 1mM</td>
<td>55281.09</td>
<td>2.28</td>
</tr>
<tr>
<td>NST74 + 10mM</td>
<td>111728.1</td>
<td>4.61</td>
</tr>
</tbody>
</table>

With 10mM added L-tryptophan, there is a 4.61-fold increase in L-phenylalanine when compared to the control (table 6.4).

**Intracellular L-phenylalanine**

The intracellular L-phenylalanine remains relatively consistent throughout the conditions, which corroborates the findings of the preliminary results. However, there is a large increase in the peak area of intracellular L-phenylalanine with 10mM added L-tryptophan (figure 6.5). This is significantly different when a t-test was performed (p<0.0428), however this condition does have a large standard deviation and standard error mean (SEM), seen in the table below (table 6.5).

Table 6.5. Statistics showing the variability within the 10mM added L-tryptophan samples.

<table>
<thead>
<tr>
<th>NST74 + 10mM added L-tryptophan intracellular samples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean of L-phenylalanine peak</td>
</tr>
<tr>
<td>Std. deviation of L-phenylalanine peak</td>
</tr>
<tr>
<td>Std. error of mean of L-phenylalanine peak</td>
</tr>
</tbody>
</table>
Figure 6.5. Intracellular L-phenylalanine peak area of E. coli NST74 with added L-tryptophan. Error bars indicate standard deviation.

Other metabolites

From the other metabolites targeted within the mass spectrometry method, only three were found to have significant differences between conditions.

The first of these is pyruvate, shown in figure 6.6A. This appears to be consistent between the first 4 conditions but is significantly increased in NST74 + 10mM L-tryptophan (p<0.0001).
Another extracellular metabolite that showed significant differences was succinate (figure 6.6B). The significant difference here was between the control NST74 and the NST74 grown with 10mM added L-tryptophan (p<0.0354), however there does seem to be a general increasing trend of the levels of succinate across all the conditions.

The other extracellular metabolites found to be present across most samples showed no significant differences between conditions, this includes cis-aconitate, citrate, glucose, fumarate, glyceraldehyde-3-phosphate, isocitrate, L-aspartate, L-glutamate, L-glutamine, L-tyrosine, and S-malate. These metabolites were all found to be consistent throughout all conditions, indicating that the added L-tryptophan does not have off-target effects.

Intracellularly there were no targeted metabolites, other than L-phenylalanine (figure 6.5), that showed any significant differences, however as untargeted analysis did not take place it is not possible to look at the changes in the global metabolome.
L-tryptophan itself was not found to be significantly altered in both extracellular and intracellular samples. This is shown in figure 6.7.

**Figure 6.7.** Extracellular and intracellular L-tryptophan peak area. There is no significant difference between conditions in both extracellular and intracellular samples. It is important to note that detection of L-tryptophan is low and therefore MS parameters may need further optimisation. Error bars indicate standard deviation.

**Discussion**

Higher concentrations of L-tryptophan lead to increased concentration of L-phenylalanine in extracellular samples. As the growth rate was accounted for through normalisation during the metabolite extraction, the increase in growth is not the cause for the increase in L-phenylalanine.

Therefore, it limits the reasoning to two different reasons, either L-tryptophan inhibits anthranilate synthase, which leads to an increase in other metabolic pathways, or as *E. coli* NST74 no longer needs to produce L-tryptophan the metabolism alters to account for this, and this leads to the increase. Either way there is a large increase in the desired product - extracellular L-phenylalanine.
When looking at L-tryptophan as a potential inhibitor it was important to consider what effect it could have on the cells. L-tryptophan is an amino acid usually produced by *E. coli*, and therefore it was unlikely to have any negative effects. In terms of the L-phenylalanine pathway, *E. coli* NST74 has a mutation in the aroH gene (Tribe, 1987). This gene codes for the isoenzyme of DAHP synthase which is specifically feedback regulated by tryptophan. This means that even with the added L-tryptophan, PEP and E4P are able to produce DAHP, which is the first step of the shikimate pathway. The addition of L-tryptophan does not cause feedback inhibition at this enzyme due to the mutation in aroH, and therefore the metabolism of the shikimate pathway is unaffected.

L-tryptophan was added as it is a known feedback inhibitor of anthranilate synthase, and the addition of it caused an artificial feedback loop to step into action. Consequently, it was hoped that the flux would move from the anthranilate pathway towards increased L-phenylalanine.

In tryptophan biosynthesis the feedback inhibition of anthranilate synthase is often seen as the rate limiting step, and there are many researchers who have tried to overcome this through various modifications such as increasing the amount of TrpE present (Castro-López et al., 2022) and other genetic modifications (Gu et al., 2012; Du et al., 2019).

In the case of *E. coli* NST74, inhibiting anthranilate synthase appears to be beneficial, as the strain is prototrophic and does not need L-tryptophan to survive (Tribe, 1987). Therefore, there are no negative effects to its inhibition, and it appears to increase the concentration of extracellular L-phenylalanine, whilst having minimal effects on other targeted metabolites. There was also a positive effect on growth with the higher concentrations of added L-tryptophan. However, to observe the overall effect on metabolism then untargeted metabolomics should also take place, this was conducted later and is discussed in 6.4.3.

Intracellularly, the amount of L-phenylalanine was consistent, except in the highest concentration of added L-tryptophan. This could be that the *E. coli* NST74 was struggling to export all the L-phenylalanine being produced. This could be a target for future modification. In *E. coli*, the protein YddG is responsible for the export of aromatic amino acids. Doroshenko et al. (2007) reported that *E. coli* strains...
overexpressing *yddG* accumulated less L-phenylalanine within the cell and exported L-phenylalanine threefold faster than the control. Therefore, this could be implemented in *E. coli* NST74 to see if the yield of extracellular L-phenylalanine increased further. However, the build-up of intracellular L-phenylalanine in *E. coli* NST74 in the 10mM added L-tryptophan condition, does not appear to have a detrimental effect as it did not impact growth. This might be different when scaled up in a bioreactor, and therefore the option for modification of the L-phenylalanine transport system is beneficial.

### Other metabolites

Intracellularly there were no significant differences of any detected metabolite between the conditions. This indicates that the addition of L-tryptophan does not largely affect the metabolism of *E. coli* NST74 within this targeted workflow. This is a positive observation, as the only detected change is the desired increase in L-phenylalanine, and otherwise the metabolism remains unaffected. As with all targeted metabolomics, to observe the full effect on metabolism, untargeted experiments should also take place.

Extracellular samples showed an increase in pyruvate in this highest concentration of added tryptophan (p<0.0001) (figure 6.6A). One possibility for this increase is there is elevated flux of glycolysis, also supported by the increase in growth (figure 6.3). Another possibility may be that pyruvate can be produced by the breakdown of tryptophan to indole and pyruvate by tryptophanase A. This would be unexpected as *E. coli* NST74 has a mutation in the tna-2 gene, which has the purpose of stopping the breakdown of tryptophan. However, in wild type *E. coli*, the activity of unmodified tna-2 depends directly on the amount of exogenous tryptophan (Li and Young, 2013), and therefore it is possible that at such extreme concentrations of tryptophan, the enzyme was able to have some effect and started to break down the excess tryptophan to indole and pyruvate (Li and Young, 2015). A similar case occurred with Ward and Yudkin (1976), where tna knockouts were able to grow on minimal medium containing indole and 5-methyltryptophan, meaning it must be able to make tryptophanase. However, as the mutations in the structural genes for adenyl cyclase
and the CGA protein were present, it was able to make tryptophanase only by acquiring a means of expressing tna that does not depend on the c-AMP-CGA system. Therefore, it is possible that something similar occurred in *E. coli* NST74 with the tryptophanase gene, leading to the breakdown of L-tryptophan.

An alternative explanation is that pyruvate is pooled within the cells to enable flux between phosphoenolpyruvate and pyruvate in either direction, leading to many different pathways (Moxley and Eiteman, 2021). As the shikimate pathway becomes overwhelmed with the amount of L-tryptophan and L-phenylalanine, this pyruvate might be transported out of the cell in an attempt to decrease the reactions and slow metabolism. This could explain why it is found in a high concentration extracellularly.

Extracellular succinate was also increased in the 10mM added tryptophan condition compared to the control (*p*<0.0354) (figure 6.6B). Succinate is part of the TCA cycle, and therefore this indicates an increase in flux through the cycle (Thakker et al., 2012). This supports the hypothesis that there is a large build-up of pyruvate, as this could also be pushed through the TCA cycle, resulting in a large increase of succinate.

One interesting point to note is that L-tryptophan itself was not found to be significant (figure 6.7), even though it was purposefully added at different concentrations. This could indicate that the L-tryptophan is being rapidly catabolised and used within the pathway. Another possibility is that the MRM used to identify L-tryptophan was incorrect, however as an authentic standard was used this is unlikely. However, the MS settings may not have been optimal for transmission of L-tryptophan ions and therefore this could be optimised and analysed to see if there is any effect on L-tryptophan detection.

### 6.4.3 Culturing with added L-tryptophan – Untargeted MS

The samples analysed here were the same samples generated in section 6.4.2. *E. coli* NST74 was cultured with L-tryptophan at various concentrations, 0µM, 10µM, 100µM, 1mM, and 10mM. Growth curve readings can be found in figure 6.3. It was
decided to explore an untargeted mass spectrometry method, to analyse off target effects that the addition of L-tryptophan had on the metabolism of *E. coli* NST74, and so the samples were analysed via LC-IM-q-TOF mass spectrometry.

**Significantly different extracellular metabolites with untargeted MS**

For the extracellular samples, 12 metabolites were identified as significantly different (p<0.05) when statistical analysis was performed using ANOVA with fishers LSD post hoc analysis. 4 were identified in negative mode (figure 6.8), whilst 8 were identified in positive mode (figures 6.10 and 6.11). This is also illustrated in table 6.6.
Table 6.6: Extracellular metabolites that were identified as being significantly different (p<0.05) in E. coli NST74 grown in various concentrations of added L-tryptophan when analysed with ANOVA and Fishers LSD post hoc analysis. *Glufosinate-ammonium identified is a contaminant.

<table>
<thead>
<tr>
<th>Metabolite</th>
<th>Mode</th>
<th>P value</th>
<th>FDR</th>
<th>Fishers’ LSD post hoc analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Indole acetaldehyde</td>
<td>Positive</td>
<td>6.5215E-11</td>
<td>6.065E-9</td>
<td>10mM T - 100uM T; 10mM T - 10uM T; 10mM T - 1mM T; 10mM T - NST74+MM1; 1mM T - 10uM T; 1mM T - NST74+MM1</td>
</tr>
<tr>
<td>1-Methyladenosine</td>
<td>Positive</td>
<td>8.3902E-10</td>
<td>3.9014E-8</td>
<td>10mM T - 100uM T; 10mM T - 10uM T; 10mM T - 1mM T; 10mM T - NST74+MM1</td>
</tr>
<tr>
<td>L-Valine</td>
<td>Positive</td>
<td>1.7059E-8</td>
<td>5.2883E-7</td>
<td>100uM T - 10mM T; 1mM T - 100uM T; 10uM T - 10mM T; 1mM T - 10mM T; 100uM T - NST74+MM1 - 10mM T; 1mM T - 10uM T; 1mM T - NST74+MM1</td>
</tr>
<tr>
<td>Indole-3-acetamide</td>
<td>Positive</td>
<td>2.4512E-8</td>
<td>5.6991E-7</td>
<td>10mM T - 100uM T; 1mM T - 100uM T; 10mM T - 10uM T; 10mM T - NST74+MM1; 1mM T - 10uM T; 1mM T - NST74+MM1</td>
</tr>
<tr>
<td>Glufosinate-ammonium*</td>
<td>Positive</td>
<td>1.0968E-6</td>
<td>2.0401E-5</td>
<td>10mM T - 100uM T; 10mM T - 10uM T; 10mM T - 1mM T; 10mM T - NST74+MM1</td>
</tr>
<tr>
<td>D-Tryptophan</td>
<td>Positive</td>
<td>1.5171E-5</td>
<td>2.3515E-4</td>
<td>10mM T - 100uM T; 1mM T - 100uM T; 10mM T - 10uM T; 10mM T - NST74+MM1; 1mM T - 10uM T; 1mM T - NST74+MM1</td>
</tr>
<tr>
<td>Pyridoxine</td>
<td>Positive</td>
<td>7.7101E-4</td>
<td>0.010243</td>
<td>10mM T - 100uM T; 1mM T - 100uM T; 10mM T - 10uM T; 10mM T - 1mM T; 10mM T - NST74+MM1 - 1mM T; 10mM T - NST74+MM1</td>
</tr>
<tr>
<td>L-Tryptophan</td>
<td>Positive</td>
<td>0.0023341</td>
<td>0.027133</td>
<td>10mM T - 100uM T; 10mM T - 10uM T; 10mM T - 1mM T; 10mM T - NST74+MM1</td>
</tr>
<tr>
<td>D-Tryptophan</td>
<td>Negative</td>
<td>1.1897E-15</td>
<td>1.0112E-13</td>
<td>10mM T - 100uM T; 1mM T - 100uM T; 10mM T - 10uM T; 10mM T - 1mM T; 10mM T - NST74+MM1; 1mM T - 10uM T; 1mM T -NST74+MM1</td>
</tr>
<tr>
<td>Indole acetaldehyde</td>
<td>Negative</td>
<td>6.5803E-11</td>
<td>2.7966E-9</td>
<td>10mM T - 100uM T; 100uM T - 10uM T; 100uM T - NST74+MM1; 10mM T - 10uM T; 10mM T - NST74+MM1</td>
</tr>
<tr>
<td>N-Acetyl-DL-tryptophan</td>
<td>Negative</td>
<td>1.3643E-6</td>
<td>3.8655E-5</td>
<td>10mM T - 100uM T; 10mM T - 10uM T; 10mM T - 1mM T; 10mM T - NST74+MM1</td>
</tr>
<tr>
<td>2,3-Dihydroxybenzoate</td>
<td>Negative</td>
<td>7.9582E-5</td>
<td>0.0016911</td>
<td>10mM T - 100uM T; 10mM T - 10uM T; 10mM T - 1mM T; 10mM T - NST74+MM1</td>
</tr>
</tbody>
</table>
Figure 6. 8. Significantly different metabolites from extracellular negative data acquired on IM-qTOF A) D-tryptophan is increased in the 10mM added L-tryptophan condition, with some increase in 1mM added L-tryptophan. B) Indole acetaldehyde, C) N-Acetyl-DL-tryptophan and D) 2,3-Dihydroxybenzoate are also increased in the 10mM added L-tryptophan condition.
All the significantly different extracellular metabolites acquired in negative mode on the IM-q-TOF follow the same trend. This is an increase in metabolite concentration, with increased concentration of L-tryptophan. The control, with no added L-tryptophan, has none of these metabolites present. This suggests the higher the concentration of L-tryptophan, the higher the concentration of extracellular D-tryptophan, indole acetaldehyde, N-acetyl-DL-tryptophan and 2,3-dihydroxybenzoate (figure 6.8).
Figure 6. Principal component analysis (PCA) graphs for the extracellular data acquired in negative polarity on the IM-q-TOF. A) The PCA scores plot of the samples shows good separation of the 10mM added tryptophan and the other conditions. PC1 explains over 65% variation, and PC2 explains almost 25%, giving a total 90.4% variance. B) The green line on top shows the accumulated variance explained; the blue line underneath shows the variance explained by individual PC.
Figure 6. 10. Four out of eight significantly different metabolites from extracellular positive 
data acquired on IM-qTOF A) Indole acetaldehyde concentration increases as the 
concentration of added L-tryptophan rises. B) Methyl adenosine is not present, except in the 
condition with the high concentration of added L-tryptophan. C) L-valine is present in all 
conditions in consistent amounts, except the 10mM added L-tryptophan condition where it is 
absent. D) Indole-3-acetamide is increased in the 1mM added L-tryptophan condition and 
also present in the 10mM added L-tryptophan condition.
Figure 6. 11. Four out of eight significantly different metabolites from extracellular positive data acquired on IM-qTOF A) Glufosinate-ammonium is most likely a contaminant B) D-Tryptophan is increased in the 1mM and 10mM added L-tryptophan conditions. C) Pyridoxine is present in all samples, but the concentration varies. It is decreased in the 100µM added L-tryptophan. D) L-tryptophan is present in the 10mM added L-tryptophan condition.
For the extracellular data analysed in positive mode on the IM-q-TOF, 8 metabolites were found to be significantly different, as shown in table 6.6 and figures 6.10 and 6.11. The majority of these followed the same trend, increasing with more L-tryptophan added. This is true for all the identified metabolites except L-valine and indole-3-acetamide.

L-valine is found in consistent amounts in all conditions including the control but is not present in the 10mM added tryptophan sample (figure 6.10C). Indole-3-acetamide is found to be higher in the 1mM added tryptophan condition and then 10mM condition (figure 6.10D). The potential reasons for these will be discussed later.
PCA for extracellular data acquired in positive polarity

A) The PCA scores plot of the samples shows good separation of the 10mM added L-tryptophan and the other conditions. PC1 explains over 44% variation, and PC2 explains over 15%, giving a total 59% variance.

B) The green line on top shows the accumulated variance explained; the blue line underneath shows the variance explained by individual PC.

Figure 6. 12. Principal component analysis (PCA) graphs for the extracellular data acquired in positive polarity on the IM-q-TOF A) The PCA scores plot of the samples shows good separation of the 10mM added L-tryptophan and the other conditions. PC1 explains over 44% variation, and PC2 explains over 15%, giving a total 59% variance. B) The green line on top shows the accumulated variance explained; the blue line underneath shows the variance explained by individual PC.
Significantly different intracellular metabolites with untargeted MS

For the intracellular samples, 11 metabolites were identified as significantly different (p<0.05) when statistical analysis was performed with ANOVA with fishers LSD post hoc analysis (table 6.7). 6 were identified in negative mode (figures 6.13 and 6.14), whilst 5 were also identified in positive mode (figures 6.16 and 6.17). This is also shown in table 6.7.
Table 6. Five intracellular metabolites were identified to be significantly different (p<0.05 from ANOVA with Fishers LSD post hoc analysis) in positive mode from the IM-q-TOF. In negative mode this increased to 6 significantly different intracellular metabolites.

<table>
<thead>
<tr>
<th>Intracellular metabolites</th>
<th>Mode</th>
<th>P Value</th>
<th>FDR</th>
<th>Post-hoc tests</th>
</tr>
</thead>
<tbody>
<tr>
<td>Indole-3-acetamide</td>
<td>Positive</td>
<td>3.1267E-6</td>
<td>2.8453E-4</td>
<td>10mM T - 100uMT; 10mM T - 10uM T; 10mM T - 1mM T; 10mM T - NST74+MM1</td>
</tr>
<tr>
<td>Adenine</td>
<td>Positive</td>
<td>1.9596E-5</td>
<td>8.916E-4</td>
<td>10mM T - 100uMT; 10mM T - 10uM T; 10mM T - 1mM T; 10mM T - NST74+MM1</td>
</tr>
<tr>
<td>Cytosine</td>
<td>Positive</td>
<td>2.8063E-4</td>
<td>0.0085124</td>
<td>10mM T - 100uMT; 10mM T - 10uM T; 10mM T - 1mM T; 10mM T - NST74+MM1</td>
</tr>
<tr>
<td>5-Hydroxy-L-tryptophan</td>
<td>Positive</td>
<td>5.0331E-4</td>
<td>0.01145</td>
<td>10mM T - 100uMT; 10mM T - 10uM T; 10mM T - 1mM T; 10mM T - NST74+MM1</td>
</tr>
<tr>
<td>Adenosine</td>
<td>Positive</td>
<td>0.0013794</td>
<td>0.025106</td>
<td>10mM T - 100uMT; 10mM T - 10uM T; 10mM T - 1mM T; 10mM T - NST74+MM1</td>
</tr>
<tr>
<td>D-Tryptophan</td>
<td>Negative</td>
<td>2.7922E-5</td>
<td>0.001006</td>
<td>10mM T - 100uM T; 10mM T - 10uM T; 10mM T - 1mM T; 10mM T - NST74+MM1</td>
</tr>
<tr>
<td>2,3-Dihydroxybenzoate</td>
<td>Negative</td>
<td>3.0485E-5</td>
<td>0.001006</td>
<td>10mM T - 100uM T; 10mM T - 10uM T; 10mM T - 1mM T; 10mM T - NST74+MM1</td>
</tr>
<tr>
<td>Indole acetaldehyde</td>
<td>Negative</td>
<td>1.2629E-4</td>
<td>0.0027784</td>
<td>10mM T - 100uM T; 10mM T - 10uM T; 10mM T - 1mM T; 10mM T - NST74+MM1</td>
</tr>
<tr>
<td>Adenine</td>
<td>Negative</td>
<td>2.0077E-4</td>
<td>0.0033127</td>
<td>10mM T - 100uM T; 10mM T - 10uM T; 10mM T - 1mM T; 10mM T - NST74+MM1</td>
</tr>
<tr>
<td>4,6-Dioxoheptanoic acid</td>
<td>Negative</td>
<td>0.0011122</td>
<td>0.01468</td>
<td>10mM T - 100uM T; 10mM T - 10uM T; 10mM T - 1mM T; 10mM T - NST74+MM1</td>
</tr>
<tr>
<td>beta-Pseudouridine</td>
<td>Negative</td>
<td>0.0015167</td>
<td>0.016684</td>
<td>10mM T - 100uM T; 10mM T - 10uM T; 10mM T - 1mM T; 10mM T - NST74+MM1</td>
</tr>
</tbody>
</table>
Six intracellular metabolites were identified as being significantly different when analysed in negative polarity on the IM-q-TOF. They all follow the same trend. This is an increase in these metabolites, with the increasing concentration of added L-tryptophan. There is a similar pattern with all these metabolites, the higher concentration is found in the 10mM added L-tryptophan condition, indicating that the larger the amount of added L-tryptophan the higher the amount of intracellular D-tryptophan, 2,3-dihydroxybenzoate, indole acetaldehyde, adenine, 4,6-dioxoheptanoic acid and beta-Pseudouridine.
Four out of six significantly different metabolites from intracellular negative data acquired on IM-qTOF A) D-tryptophan B) 2,3-Dihydroxybenzoate C) Indole acetaldehyde D) Adenine. All four metabolites follow a similar pattern of being absent or present in low concentrations in the lower concentrations of added L-tryptophan and are present in high concentrations in the 10mM added L-tryptophan condition.
Figure 6.14. Two out of six significantly different metabolites from intracellular negative data acquired on IM-qTOF A) 4,6-Dioxoheptanoic acid B) beta-Pseudouridine. Both A and B follow the same pattern of increased concentration in the highest concentration (10mM) of added L-tryptophan.

For all of the intracellular negative results, there is a trend of an increase in these metabolites with the increasing concentration of added L-tryptophan. There is a similar pattern with all these metabolites, the higher concentration is found in the 10mM added L-tryptophan condition, indicating that the higher the amount of added L-tryptophan the higher the amount of intracellular D-tryptophan, 2,3-dihydroxybenzoate, indole acetaldehyde, adenine, 4,6-dioxoheptanoic acid and beta-Pseudouridine (figure 6.13 and 6.14).
PCA for intracellular data acquired in negative polarity

A) The PCA scores plot of the samples shows good separation of the 10mM added tryptophan and the other conditions. PC1 explains over 60% variation, and PC2 explains almost 17%, giving a total 78.7% variance.

B) The green line on top shows the accumulated variance explained; the blue line underneath shows the variance explained by individual PC.

Figure 6. 15. Principal component analysis (PCA) graphs for the intracellular data acquired in negative polarity on the IM-q-TOF A) The PCA scores plot of the samples shows good separation of the 10mM added tryptophan and the other conditions. PC1 explains over 60% variation, and PC2 explains almost 17%, giving a total 78.7% variance. B) The green line on top shows the accumulated variance explained; the blue line underneath shows the variance explained by individual PC.
Figure 6.16. Four out of five significantly different metabolites from intracellular positive data acquired on IM-qTOF A) Indole-3-acetamide B) Adenine C) Cytosine D) 5-Hydroxy-L-tryptophan. All follow the same trend of being present in the highest concentration in the highest amount of added L-tryptophan (10mM).
For all of the intracellular positive results there is a trend, this of an increase in these metabolites with the increasing concentration of added L-tryptophan. There is a similar pattern with all these metabolites, the control has none and the higher concentration is found in the 10mM added L-tryptophan condition, indicating that the higher the amount of added indole-3-acetamide, adenine, cytosine, 5-hydroxy-L-tryptophan and adenosine (figure 6.16 and 6.17).
PCA for intracellular data acquired in positive polarity

Figure 6. 18. Principal component analysis (PCA) graphs for the intracellular data acquired in positive polarity on the IM-q-TOF A) The PCA scores plot of the samples shows good separation of the 10mM added tryptophan and the other conditions. PC1 explains over 67% variation, and PC2 explains almost 15%, giving a total 82% variance. B) The green line on top shows the accumulated variance explained; the blue line underneath shows the variance explained by individual PC.
**L-phenylalanine**

Whilst L-phenylalanine was not identified as one of the significantly different metabolites, it remained a metabolite of interest and therefore was analysed. Despite being the same samples in Section 6.4.2, the data generated using the LC IM-q-TOF method yielded different results. L-phenylalanine was found to be present only on extracellular samples in negative polarity mode (figure 6.19).

It was found to show the same trend, increasing L-phenylalanine with increasing concentration of added L-tryptophan. However, despite this, the data indicates huge variability.
Figure 6. 19. Abundance of L-phenylalanine in extracellular samples run in negative mode on the IM-q-TOF. Error bars indicate standard deviation and indicate large variation.

Discussion of significantly different extracellular metabolites

There were several significant differences found in the extracellular samples. These will be briefly discussed below.

L-tryptophan

L-tryptophan was found in the extracellular samples. It was absent in every condition except the highest concentration of 10mM added tryptophan. It should be present in
each sample except the control, as it was purposefully added to each sample. As it is not, it raises the question of whether it is being broken down, perhaps to indole and pyruvate as discussed earlier. This suggests that at the lower concentrations of added tryptophan, the metabolism of *E. coli* NST74 is able to cope and metabolise the L-tryptophan, however at the highest concentration, the cells are at metabolic capacity and are unable to deal with the excess, therefore it remains in the supernatant, and is detected.

**D-tryptophan**

D-tryptophan is an enantiomer of L-tryptophan and has been shown to change the physiochemical properties of bacterial cell walls (Su et al., 2021). Interestingly in the data from the IM-q-TOF, D-tryptophan was found in 1mM and 10mM added tryptophan. It is most likely that this is a misidentification of L-tryptophan, as distinguishing between enantiomers is very difficult as enantiomers should not display statistically different CCS values (Dodds et al., 2017), however it has been possible to identify them through cyclic IMS (Cooper-Shepherd et al., 2022). As mentioned in the introduction, in ion mobility MS, ions are accelerated with an applied electric field whilst a buffer gas opposes the ion motion in the drift tube. Ions with different sizes and charges will have different ion mobilities, and subsequently different drift times. Therefore, ion mobility MS provides a new dimension of separation and shape information, because separation is based on the conformation of a molecule in addition to its mass (Kanu et al., 2008).

In positive mode data, both L and D-tryptophan were found to be present with CCS values of 151.3 (L-tryptophan) and 142.6 and 151.7 (D-tryptophan) therefore it is possible that this is a misidentification, at least partially, as the CCS tolerance is set to +/- 1%. In positive mode, L-tryptophan was not identified, whereas D-tryptophan was with CCS values of 147.9 and 150.6. It is likely that one of these is a misidentification and is L-tryptophan. This illustrates that in the McLean library the CCS values for D and L tryptophan are different (Picache et al., 2019). As both L- and D-tryptophan have the same m/z and similar retention time, it is only possible to
separate them based on CCS value, which is calculated in the drift tube of the IM-q-TOF.

Whilst the above explanation is most likely, another possibility is isomerization. A paper, by Okazaki in 1968 using a strain of Pseudomonas, shows that isomerisation occurred when bacteria were exposed to high levels of D-tryptophan plus another added amino acid leading to an increase in L-tryptophan. It could be possible that the L-tryptophan, in the presence of high levels of L-phenylalanine is being converted to D-tryptophan.

**N-acetyl-DL-tryptophan**

There is also an increase in N-acetyl-DL-tryptophan as the concentration of added tryptophan increases. One viable explanation for this is linked with the increase of D-tryptophan as mentioned above, as N-acetyl-DL-amino acids can be an intermediate of L-amino acid conversion to D-amino acid and vice versa (Espaillat et al., 2013).

**2,3-dihydroxybenzoate**

The concentration of 2,3-dihydroxybenzoate (DHB) also increases in the condition with the highest added concentration of tryptophan. This is interesting as DHB is involved in the biosynthesis of enterobactin, which involves six enzymes, and starts from chorismate. Chorismate reacts to yield isochorimsate, which then forms DHB, there are then subsequent reactions to yield enterobactin. This indicates that the metabolic pathways that branch from chorismate occurred in the condition with the highest concentration of added L-tryptophan (Khasheii et al., 2021). Enterobactin is a sideophore that usually forms in low iron conditions (Moynié et al., 2019), and therefore could be an indication that the high concentration of L-tryptophan was causing some amount of stress to the E. coli cells. It could also indicate that the excess chorismate, caused by the inhibition of anthranilate synthase, was directed towards this pathway.
Indole acetaldehyde and Indole-3-acetamide

There is also an increase in the concentration of indole acetaldehyde as the concentration of tryptophan increases. This is not surprising as tryptophan is broken down to indole and pyruvate, with indole-3-acetic acid and indole acetaldehyde being intermediates in this pathway (Zhang et al., 2019). However, as mentioned earlier when discussing the triple quadrupole data, *E. coli* NST74 has a modification in tna2, and therefore this breakdown should not be possible. The results, however, indicate otherwise, as both pyruvate and indole intermediates are found in higher concentrations in the conditions with higher levels of L-tryptophan. This could indicate that the tna-2 modification reduces the efficiency of the enzyme but does not stop it working completely.

1-methyladenosine

There is an increase in 1-methyladenosine, which is a purine nucleoside. Purine nucleosides are compounds comprising a purine base attached to a ribosyl or deoxyribosyl moiety. It plays a role in regulating mRNA structure (Ogawa et al., 2020).

Pyridoxine

Vitamin B6 is an ensemble of six interconvertible vitamers: pyridoxine (PN), pyridoxamine (PM), pyridoxal (PL), and their 5′-phosphate derivatives, PNP, PMP, and PLP. Pyridoxal 5′-phosphate is a coenzyme in a variety of enzyme reactions concerning transformations of amino acid compounds (Barile et al., 2019; Tramonti et al., 2011), therefore as the L-tryptophan level increases (and so does the L-phenylalanine), the levels of PN might increase to cope with this. Another alternative
explanation is that tryptophanase, the enzyme encoded by the tnaA gene is tetrameric, with a pyridoxal phosphate (PLP) coenzyme covalently linked to the Lys270 residue at each of the four active sites (Li and Young, 2015). Therefore, if the above hypothesis is correct that tryptophanase is indeed active in the higher concentrations of added tryptophan, then this could explain the increase in pyridoxine.

**L-valine**

Interestingly L-valine is present in all samples, except the highest concentration of L-tryptophan. This was in extracellular data, and therefore the hypothesis is that in response to the excess amino acids present (L-tryptophan and L-phenylalanine), other amino acid pathways saw reduced flux and were produced at a lesser quantity (Park et al., 2007), however only L-valine was found to be significantly different.

**Discussion of intracellular metabolites**

There were several significant differences found in the intracellular samples. These will be briefly discussed below.

**D-tryptophan**

As mentioned above, it is possible that this is a misidentification of L-tryptophan, or an isomerisation reaction occurred due to the excess L-tryptophan present in the media.
5-Hydroxy-L-tryptophan (5-HTP), Indole-3-acetamide and Indole acetaldehyde

5-HTP can be produced from L-tryptophan degradation to indole and pyruvate (Hara and Kino, 2013). As all three of these metabolites are involved in the same pathway and are all only present in the 10mM added tryptophan condition, it suggests that this reaction only occurs at this high concentration.

DHB

As mentioned above, DHB is involved in the biosynthesis of enterobactin, which involves six enzymes, and starts from chorismate (Gehring et al., 1997). The increase in DHB could indicate increased stress, however this seems unlikely, as there were no issues with growth as shown in figure 6.3, which would be expected if this was a stress condition.

4,6-dioxoheptanoic acid

This is also known as succinylacetone, and is a metabolite associated with hereditary tyrosinemia (Grenier et al., 1982) therefore, it is unusual for it to be found in an *E. coli* cell. It could be a false identification. However, it is a succinate derivative, and succinate was found increased extracellularly in the data obtained on the triple quadrupole MS. Based on Grenier et al. (1982) paper, in humans succinylacetone is a chemical by-product of extreme levels of tyrosine, therefore in *E. coli* this same metabolite could be produced with extreme levels of L-phenylalanine, as the pathways are closely linked.
Beta-Pseudouridine

Pseudouridine is the most abundant posttranscriptional modification in RNAs (O’Connor et al., 2018). Therefore, this increase could be caused by the increased growth in the conditions with higher concentrations of added L-tryptophan.

Overall, the significant changes found in the IM-q-TOF data indicate that there are large amounts of L-tryptophan present, and there is evidence of L-tryptophan degradation (with the presence of indole acetaldehyde and indole-3-acetamide). There does not appear to be any negative effects on cell and viability, which indicates that the use of L-tryptophan to inhibit anthranilate could successfully be applied to *E. coli* NST74 to increase the yield of L-phenylalanine.

**Discussion of L-phenylalanine**

There are multiple reasons why L-phenylalanine may not have been found in these samples. Identification was through m/z and CCS values only. This was to ensure a high level of confidence in the identifications. If more databases had been included, such as METLIN, in the identification search, then this would have widened the criteria and led to increased identifications. This could be optimised for further analysis of this data set, and the data could be reprocessed using added metabolite databases.

Also, unlike the method on the triple quadrupole instrument, the buffers on the LC system were not optimised for this specific set of samples. Therefore, as the buffers on this LC system were made largely of acetonitrile, and the extraction solvent was 50: 50 acetonitrile: water it is possible that metabolites more suited to other solvent systems were not sufficiently extracted and transported to the MS.

One final point on this, when the samples were run on IM-q-TOF they had been thawed and frozen twice and this could have led to some metabolite degradation.
For a thorough untargeted investigation into the effect of L-tryptophan on L-phenylalanine concentration in *E. coli* NST74, this experiment should be repeated.

**Figure 6.** Metabolic pathways shown to be affected by the addition of the L-tryptophan. There is an increase in 2,3-dihydroxybenzoate which could suggest increased flux through this branch off chorismate. There is also an increase in L-tryptophan which is expected as this was added in high concentrations. The hypothesis is that this L-tryptophan is being converted to indole, as there is an increase in pathway intermediates, (not shown), such as indole-3-acetamide and indole-3-acetaldehyde. Another hypothesis is that L-tryptophan is being converted to D-tryptophan, as this also increases. There is also an increase in pathway intermediates (not shown) such as N-acetyl-DL-tryptophan and 5-hydroxy-L-tryptophan. This shows the importance of chorismate as a branching point within the shikimate pathway and formation of other products.
6.4.4 Culturing with L-tryptophan added at stationary growth phase – Targeted MS

In this experiment, *E. coli* NST74 was grown in 96 well plates and allowed to reach stationary phase, before the different concentrations of L-tryptophan were added. It was then cultured for 6 hours before metabolite extraction took place.

**Growth curves**

Growth curves were unable to be obtained due to plate reader error, and therefore normalisation of samples through metabolite extraction was unable to take place.

**L-phenylalanine**

Extracellular L-phenylalanine was increased in the samples with added L-tryptophan. In this highest concentration of added tryptophan, there was a significant difference when compared to the control with a t-test (*p* < 0.0001) (figure 6.20A).
Figure 6. 21. A) Peak area of extracellular L-phenylalanine in E. coli NST74 grown in various concentrations of added L-tryptophan. There is a significant difference (p<0.0001) between the control and 10mM added L-tryptophan condition. B) Peak area of intracellular L-phenylalanine in E. coli NST74 grown in various concentrations of added L-tryptophan. There are no significant differences. Error bars indicate standard deviation.

Whilst only one condition, 10mM added L-tryptophan showed a significant difference, there was a fold increase in extracellular L-phenylalanine in all conditions (table 6.8). In the 10mM added L-tryptophan condition this was a 5.39-fold increase compared to the control.

Table 6. 8. Mean peak area and fold change of extracellular L-phenylalanine in E. coli NST74 with L-tryptophan added at stationary phase.

<table>
<thead>
<tr>
<th>Condition</th>
<th>Mean Peak area</th>
<th>Fold change</th>
</tr>
</thead>
<tbody>
<tr>
<td>NST74 + MM1</td>
<td>178195.1</td>
<td>-</td>
</tr>
<tr>
<td>NST74 + 10µM</td>
<td>251498.4</td>
<td>1.40</td>
</tr>
<tr>
<td>NST74 + 100µM</td>
<td>255197.7</td>
<td>1.42</td>
</tr>
<tr>
<td>NST74 + 1mM</td>
<td>353315.7</td>
<td>1.97</td>
</tr>
<tr>
<td>NST74 + 10mM</td>
<td>965914.6</td>
<td>5.39</td>
</tr>
</tbody>
</table>
There are no significant differences between intracellular L-phenylalanine in the different conditions (figure 6.20B), and although not significant, there does appear to be an increasing trend of increased intracellular L-phenylalanine with increased concentrations of added L-tryptophan.

Discussion

As discussed earlier in this chapter, using L-tryptophan to increase the yield of L-phenylalanine was successful, and therefore it was decided to test whether L-tryptophan added in stationary phase would have the same effect.

In the initial experiment, the L-tryptophan was present in the media from the beginning. In the time delay experiment the L-tryptophan was added after 10 hours, which is when stationary growth phase should have been achieved, and metabolite extraction took place 6 hours after the addition of L-tryptophan. Interestingly, this data indicates that L-tryptophan is not required in the media during the growth stage to have an effect on the amount L-phenylalanine produced. In fact, extracellularly the fold change in the 10mM added L-tryptophan condition is higher when L-tryptophan is added at stationary phase (5.39-fold change increase), compared to when added at the beginning (4.61 increase).

However, the time delay data was not normalised during metabolite extraction due to an error with the plate reader, and therefore growth curves could not be obtained. This could have some impact on the data as normalisation did not occur.

Intracellularly there was no significant difference between L-phenylalanine peak area between conditions, however, there does appear to be an increasing trend of L-phenylalanine with increased added L-tryptophan. This could be due to increased production stimulated by the presence of L-tryptophan, and due to the time constraints, there may not have been enough time for L-phenylalanine to be exported extracellularly, hence the accumulation intracellularly. However, as
normalisation was unable to take place, this increase in L-phenylalanine could simply be due to an increase in growth rate.

Unfortunately, there is only one replicate for the intracellular control, this is either due to poor metabolite extraction or missed injections on the triple quadrupole mass spectrometer. Therefore, for a conclusion to be drawn, a repeat of this experiment would need to be conducted.

### 6.5 Conclusion

#### 6.5.1 General discussion

Overall L-tryptophan is a successful inhibitor of anthranilate synthase and increases the yield of L-phenylalanine. The extracellular L-phenylalanine is increased 4.61-fold when cultured with L-tryptophan from the start (section 6.4.2), and 5.39-fold when L-tryptophan is added later (section 6.4.4). The growth is positively affected, and there appear to be no major off target effects detected by both the targeted data from the triple quadrupole instrument, and the untargeted data generated through LC IM-q-TOF MS. Off target effects are unintended effects on metabolism. For example, in this research the aim is to increase the yield of L-phenylalanine through the addition of inhibitors, however if the addition of the inhibitor also results in negative effects, then its viability for future implementation should be evaluated.

One limitation of this is the expense of adding L-tryptophan to the media, but it is possible that further optimisation on the exact concentration in a fed-batch or continuous culture could result in a situation where the cost of using L-tryptophan to increase L-phenylalanine is financially viable. If it is not, then another alternative would be to conduct genetic modifications to target the anthranilate synthase enzyme which should mimic the same effect.

When comparing the two data types, targeted and untargeted, it is important to consider the strengths and limitations of both. The data from the triple quadrupole was quick and simple to analyse, it was easy to see which metabolites were affected by the added inhibitor. However, only metabolites listed in the targeted method could
be analysed, therefore, to analyse the global metabolic profile of the samples they had to be analysed on the IM-q-TOF. The untargeted data from the IM-q-TOF was much harder to analyse; data processing and analysis was more time consuming. The data obtained was immense and analysis can vary based on the databases used. As mentioned, the metabolites identified here using CCS and m/z values through the McLean CCS Compendium PCDL (Picache et al., 2019), but the use of other databases for identification is possible, and would result in increased annotations. This data can enable a much more thorough analysis, but for the purpose of this experiment, it is not necessary. As the method is high throughput, it is important that the data processing and analysis is also able to be completed in a timely manner, and therefore the triple quadrupole method is preferred.

6.5.2 Future work

As previously mentioned, the data analysed by the IM-q-TOF could undergo more extensive analysis, or the experiment could be repeated to ensure there was no metabolite degradation between thaw cycles.

This inhibitor should be tested in a scaled-up environment such as a fermenter or bioreactor to check that the increase in L-phenylalanine still occurs in larger volumes, e.g. 5L rather than 200µL well plate as this is not viable commercially. Tribe (1987) stated they could obtain a yield of 4g/L L-phenylalanine in a 1L fermenter, with the addition of 10mM L-tryptophan, this could potentially increase 4.6-fold, to 18.4g/L.

As previously mentioned, it could also be a possibility to modify the L-phenylalanine transporter gene to ensure maximal transportation of L-phenylalanine out of the cell. This is possible through overexpressing the YddG protein to increase exported L-phenylalanine by three-fold (Doroshenko et al., 2007)

One final point to consider is whether other overproducing L-phenylalanine strains would be more suited to this level of optimisation, as they are capable of producing higher yields. This will be discussed in the next chapter.
Chapter 7.

Exploring a panel of inhibitors on an engineered *E. coli* strain

### 7.1 Introduction

Due to the success of the added L-tryptophan on the yield of L-phenylalanine, it was decided to look at other enzyme targets, and therefore other inhibitors, to see if the success previously achieved could be repeated. This involved using different inhibitors on different enzyme targets with the aim to alter metabolic flux and increase the yield of L-phenylalanine in *E. coli* NST74. In total, 8 inhibitors were tested on 5 different enzyme targets. Each enzyme and its inhibitor are introduced below.

*Table 7.1. Overview of the enzyme and inhibitors used in this chapter.*

<table>
<thead>
<tr>
<th>Enzyme</th>
<th>Inhibitor</th>
<th>Example organism</th>
<th>Concentration</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anthranilate synthase</td>
<td>L-tryptophan</td>
<td><em>E. coli</em></td>
<td>n/a – genetic engineering</td>
<td>Kwak et al., 1998</td>
</tr>
<tr>
<td></td>
<td>Anthranilic acid</td>
<td><em>Pseudomonas putida</em></td>
<td>n/a – genetic engineering</td>
<td>Kuepper et al., 2015</td>
</tr>
<tr>
<td></td>
<td>Methyl anthranilate</td>
<td></td>
<td>n/a</td>
<td>n/a</td>
</tr>
<tr>
<td>Isochorismate synthase</td>
<td>N-ethylmaleimide</td>
<td>Purified enzyme from <em>E. coli</em></td>
<td>Complete inactivation at 100µM</td>
<td>Daruwala et al., 1997</td>
</tr>
<tr>
<td>Chorismate pyruvate lyase</td>
<td>Vanillic acid</td>
<td>Purified enzyme from <em>E. coli</em></td>
<td>260 ± 60µM</td>
<td>Holden et al., 2002</td>
</tr>
<tr>
<td>Aminodeoxychorismate synthase</td>
<td>Nickel sulphate</td>
<td><em>Burkholderia lata</em></td>
<td>1mM</td>
<td>Li et al., 2011</td>
</tr>
<tr>
<td>Prephenate dehydrogenase</td>
<td>L-tyrosine</td>
<td><em>E. coli</em></td>
<td>n/a – genetic engineering</td>
<td>Lütke-Eversloh and Stephanopoulos, 2005</td>
</tr>
<tr>
<td></td>
<td>Diethyl malonate</td>
<td>Purified enzyme from <em>E. coli</em></td>
<td>1700± 100 µM</td>
<td>Christopherson et al., 1983</td>
</tr>
</tbody>
</table>
7.1.1 Anthranilate synthase

Anthranilate synthase has been described in detail in the previous chapter and has already demonstrated itself as an ideal enzyme target; therefore, other inhibitors of this enzyme were investigated.

Anthranilic acid was used, as the anthranilate synthase enzyme had already shown product inhibition, both in the literature (Gibson and Pittard, 1968; Knöchel et al., 1999; Spraggon et al., 2001), and in the previous chapter. Anthranilic acid had previously been found to inhibit anthranilate synthase in organisms such as *Salmonella* (Cordaro et al., 1968), *Streptomyces* (Francis et al., 1978), and *Pseudomonas putida* (Kuepper et al., 2015). Therefore, it was decided to see if this could be used on *E. coli* NST74, in order to increase the yield of L-phenylalanine by feedback inhibition.

As anthranilic acid was initially hard to obtain, methyl anthranilate was also tested. This is an ester of anthranilate and therefore has a similar structure and was thought to be a possible analogue for anthranilic acid and inhibit anthranilate synthase by feedback inhibition (K. Richards, Personal Communication).

7.1.2 Isochorismate synthase

Isochorismate synthase is an enzyme that converts chorismate into isochorismate, which then branches off into further downstream products such as salicylate (Zhang et al., 2022). There are two isochorismate synthase enzymes present in *E. coli*, encoded by *entC* and *menF*: EntC is linked to the enterobactin biosynthetic pathway, whereas MenF is specific to the menaquinone biosynthetic pathway (Kwon et al., 1996; Dahm et al., 1998).

Known inhibitors of isochorismate include Mg2+ ions at concentrations above 1mM, and Cu2+ (Daruwala et al., 1997). Another known inhibitor of this enzyme is N-ethylmaleimide (NEM) which was shown to inhibit this enzyme irreversibly at concentrations of 100µM with the purified enzyme, isochorismate synthase enzyme encoded by MenF gene from *E. coli* BL21(DE3). At 1µM there was 25% inactivation
and at 10µM there was 50% inactivation (Daruwala et al., 1997), therefore this inhibitor was tested at these ranges on *E. coli* NST74 to examine what the effect of this would be, and whether inhibiting this pathway would increase the yield of L-phenylalanine.

### 7.1.3 Chorismate pyruvate lyase

Chorismate pyruvate lyase, also known as chorismate lyase, is involved in the reaction from chorismate to 4-hydroxybenzoate and pyruvate. This pathway leads to the biosynthesis of ubiquinone. It does this by cleaving a carbon-oxygen bond to release pyruvate and aromatise the ring, producing 4-hydroxybenzoate. This exhibits strong product inhibition ($K_I=2.1\mu M$ for 4-hydroxybenzoate). In the research by Holden et al. (2002), inhibition constants were determined for 4-hydroxybenzoate, and product analogues such as vanillic acid (3-methoxy-4-hydroxybenzoic acid). Vanillic acid was found to have an inhibition constant for the purified enzyme, chorismate pyruvate-lyase at 260µM ±60µM. Therefore, this was added to see how it would affect the metabolism of the shikimate pathway in *E. coli* NST74, and to see if there would be any changes to the yield of L-phenylalanine.

### 7.1.4 Aminodeoxychorismate synthase

The biosynthesis of *p*-aminobenzoic acid from chorismate in *E. coli* occurs through two reactions. The first reaction involves chorismate and ammonia, derived from glutamine, being converted to aminodeoxychorismate and glutamate through a heterodimeric complex of pabA and pabB gene products. The second reaction involves aminodeoxychorismate being converted to *p*-aminobenzoic acid through aminodeoxychorismate lyase (Parsons et al., 2002). This is known as the PABA complex, and does not exhibit feedback inhibition (Viswanathan et al., 1995).

However, Mg$^{2+}$ is required as a cofactor and Li et al. (2011) found that in *Burkholderia lata* the addition of nickel resulted in the magnesium ions being displaced and the complex being inhibited. Therefore, nickel sulphate was added to
*E. coli* NST74 to see if the same effect could be mimicked in a different bacterial species.

### 7.1.5 Prephenate dehydrogenase

As discussed in chapter 6, chorismate mutase/prephenate dehydrogenase is a bifunctional enzyme. As well as L-phenylalanine synthesis, it is also involved in tyrosine biosynthesis. The rearrangement of chorismate to prephenate is catalysed by the N-terminal, chorismate mutase, domain of TyrA, whereas the C-terminal, prephenate dehydrogenase, domain catalyses the oxidative decarboxylation of prephenate to 4-hydroxyphenylpyruvate. Analogously, the prephenate dehydrogenase domain of the bifunctional enzyme (PheA), yields phenylpyruvate, which is subsequently transaminated to phenylalanine (Lütke-Eversloh and Stephanopoulos, 2005).

Lütke-Eversloh and Stephanopoulos (2005) performed experiments to see if the chorismate mutase/prephenate dehydrogenase complex exhibited feedback inhibition by tyrosine. It was found that chorismate mutase activity was reduced but not completely inhibited, whereas the prephenate dehydrogenase domain of TyrA showed feedback regulation by tyrosine. Therefore, tyrosine was added to the media to see how this would affect L-phenylalanine metabolism in *E. coli* NST74.

Similarly, the inhibition of the bi-functional enzyme chorismate mutase-prephenate dehydrogenase by substrate analogues, was investigated by Christopherson et al. (1983), and diethyl malonate was found to act as a linear competitive inhibitor. Diethyl malonate was found to inhibit both parts of the complex, and prevent binding of both substrates (Christopherson et al., 1983) and as such, this was also investigated.
7.2 Aims

The aims of this chapter were to probe the metabolism of *E. coli* NST74 using a panel of inhibitors for specific enzyme targets to see how metabolism would be altered, specifically in relation to L-phenylalanine production.

7.3 Method

7.3.1 Cell culture

*E. coli* NST74 was taken from glycerol stocks and revived on LB agar at 37°C for 24 hours in a stationary incubator. Three colonies were then taken from the agar plate and placed in 10mL MM1 (as described in chapter 2) suspension for 6 hours, this was one replicate. After 6 hours, the optical density of the samples was taken using 1mL of the sample in a spectrophotometer. The 96 well plate (Costar 96 Flat Bottom Transparent Polystyrene) was then set up and contained a total volume of 200µL in each well plate. This consisted of MM1 media, the sample at an OD$_{600}$ of 0.01 and the inhibitor at the desired concentration.

7.3.2 Addition of inhibitors

Inhibitors were added to the 96 well plate, in different concentrations, shown in table 7.2. The enzyme targets were chosen due to their proximity to chorismate, and the concentrations were based upon previous literature, see section 7.1 for more details.
Table 7.2. Overview of the inhibitors used and the different concentrations that were added.

<table>
<thead>
<tr>
<th>Inhibitor</th>
<th>Concentrations added</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tryptophan</td>
<td>0, 10µM, 100µM, 1mM, 10mM</td>
</tr>
<tr>
<td>Anthranilic acid</td>
<td>0, 10µM, 100µM, 1mM, 10mM</td>
</tr>
<tr>
<td>Methyl anthranilate</td>
<td>0, 100µM, 1mM, 2mM, 10mM</td>
</tr>
<tr>
<td>N-Ethylmaleimide</td>
<td>0, 1µM, 10µM, 100µM, 1mM</td>
</tr>
<tr>
<td>Vanillic acid</td>
<td>0, 10µM, 100µM, 1mM, 10mM</td>
</tr>
<tr>
<td>Nickel sulphate</td>
<td>0, 10µM, 100µM, 1mM, 10mM</td>
</tr>
<tr>
<td>Tyrosine</td>
<td>0, 100µM, 500µM, 1mM, 10mM</td>
</tr>
<tr>
<td>Diethyl malonate</td>
<td>0, 100µM, 1mM, 2mM, 10mM</td>
</tr>
<tr>
<td>Combination of Tryptophan and</td>
<td>10mM Trp, 10mM Tyr, 10mM Trp 1mM Tyr</td>
</tr>
<tr>
<td>Tyrosine</td>
<td>10mM Trp 10mM Tyr</td>
</tr>
</tbody>
</table>

7.3.3 Metabolite extraction

As in Chapter 2. Section 2.6.2.

7.3.4 Targeted Mass spectrometry acquisition, data processing and analysis

As in Chapter 2. Section 2.7.

7.4 Results and Discussion

7.4.1 Enzyme target: Anthranilate synthase, Inhibitor: Anthranilic acid

Anthranilic acid was added as an inhibitor of anthranilate synthase. The concentrations were added were: 0, 10µm, 100µm, 1mM, 10mM. Growth readings were obtained, and normalisation took place by adding a proportional amount of
solvent during metabolite extraction. Samples were analysed using the targeted method on the triple quadrupole instrument.

A) **E. coli NST74 with added anthranillic acid**

![Growth curves for E. coli NST74 with different concentrations of added anthranilic acid](image1)

B) **Intracellular L-phenylalanine**

![Average peak area for intracellular L-phenylalanine in E. coli NST74 with different concentrations of added anthranilic acid](image2)

C) **Extracellular L-phenylalanine**

![Average peak area for extracellular L-phenylalanine in E. coli NST74 with different concentrations of added anthranilic acid](image3)

Figure 7. 1. A) Growth curves for E. coli NST74 with different concentrations of added anthranilic acid. B) The average peak area for intracellular L-phenylalanine in E. coli NST74 with different concentrations of added anthranilic acid. There was one significant difference, between the control and 10mM added anthranilic acid concentration, however this was in the wrong direction. C) The average peak area for extracellular L-phenylalanine in E. coli NST74 with different concentrations of added anthranilic acid. Error bars indicate standard deviation.

Growth curves
At the lower concentrations, 10µM and 100µM of added anthranilic acid, there is no effect on growth as it mimics the control (figure 7.1A). At 1mM added anthranilic acid there is slightly delayed lag phase, but a similar optical density is reached after approximately 13 hours. This suggests at 1mM anthranilic acid there is some cellular stress in *E. coli* NST74, but ultimately the same optical density was reached. With the 10mM added anthranilic acid condition growth is clearly affected, and it does not grow in the same pattern as the other conditions, and stationary phase is never reached. This is not desirable for the intended purpose of increasing the yield of L-phenylalanine, as if the cells do not survive then they cannot produce L-phenylalanine.

**L-phenylalanine**

In the intracellular samples (figure 7.1B), there is a significant difference between the 10mM added anthranilic acid and the control. However, this is not in the intended direction as there is less L-phenylalanine present in the 10mM added anthranilic acid than in the control. This may have occurred as the *E. coli* struggled to grow with 10mM added anthranilic acid, and therefore the metabolic flux may have been pushed to more vital biosynthetic pathways, and therefore flux was redirected from the L-phenylalanine pathway.

For the extracellular samples, t-test was performed, and found a significant difference between the 100µM added anthranilic acid and control (*p* = 0.0008). There is also a significant difference between 1mM added anthranilic acid and the control, despite the initial growth rate being affected (*p* = 0.0003). This results in a fold change in extracellular L-phenylalanine, shown in table 7.3, with the maximum fold increase of 1.29 for the 100µM condition. This suggests that anthranilic acid inhibits anthranilate synthase by feedback inhibition, and the metabolic flux switches from this pathway towards L-phenylalanine biosynthesis leading to an increase in yield. However, ultimately there is not a huge increase in the final yield of L-phenylalanine.

There is also a significant difference in peak area of L-phenylalanine between 10mM added anthranilic acid and the control when a t-test was performed (*p*<0.0001), as with the intracellular data, this is not in the intended direction as there is more L-
phenylalanine present in the control. As with the intracellular samples, there is a very small amount of L-phenylalanine present in the 10mM condition, this is presumably because the cells struggled to grow, and therefore the metabolism was unable to reach L-phenylalanine production consistently.

Table 7.3. Average peak area and overall fold change for E. coli NST74 with different concentrations of added anthranilic acid.

<table>
<thead>
<tr>
<th>Condition</th>
<th>Average peak area of extracellular L-phenylalanine</th>
<th>Fold change</th>
</tr>
</thead>
<tbody>
<tr>
<td>NST74 + MM1</td>
<td>388607</td>
<td>-</td>
</tr>
<tr>
<td>NST74 + 10µM</td>
<td>446351</td>
<td>1.15</td>
</tr>
<tr>
<td>NST74 + 100µM</td>
<td>501081</td>
<td>1.29</td>
</tr>
<tr>
<td>NST74 + 1mM</td>
<td>490973.5</td>
<td>1.26</td>
</tr>
</tbody>
</table>

**Other metabolites**

As L-phenylalanine was found to show a statistically significantly different increase with added anthranilic acid, other metabolites were analysed. Out of both intracellular and extracellular samples, only two other metabolites were found to show significant differences.
Figure 7. 2. A) The average peak area for extracellular L-glutamate in E. coli NST74 with different concentrations of added anthranilic acid. B) The average peak area for extracellular L-tyrosine in E. coli NST74 with different concentrations of added anthranilic acid. Error bars indicate standard deviation.

One of these was extracellular L-glutamate (figure 7.2A) which showed a significant difference between the 1mM added anthranilic acid condition and the control ($p = 0.0055$).

The other was extracellular tyrosine (figure 7.2B) where both the 1mM ($p = 0.0416$) condition and the 100µM ($p = 0.0036$) condition showed significant differences when compared to the control.

This indicates that the addition of anthranilic acid increases the yield of three amino acids in E. coli NST74: L-phenylalanine, L-glutamate, and L-tyrosine.

**Discussion**

The results from this experiment indicate that anthranilic acid is a good inhibitor at concentrations of 100µM and 1mM and may act as a feedback inhibitor to
anthranilate synthase. At the above-mentioned concentrations, it does increase the yield of L-phenylalanine by 1.29, and 1.26-fold respectively, whereas at a high concentration of 10mM it has a negative effect. Whilst the fold change of L-phenylalanine does increase, it does not increase by large amounts, and therefore other inhibitors of this enzyme may be more suitable.

The increase in L-tyrosine is not surprising as this has a similar biosynthetic pathway to L-phenylalanine, branching off at prephenate to form L-tyrosine. This also can result in L-glutamate, and so the hypothesis is that the addition of anthranilic acid increases metabolic flux from chorismate to prephenate, here it branches off to increase the yield of L-phenylalanine and L-tyrosine. When producing L-tyrosine the reaction can catalyse a reaction in which L-glutamate is the end product, and this would explain the increase in this metabolite too.

Figure 7.3. A pathway diagram for L-phenylalanine, L-tyrosine, and L-glutamate.

Therefore, if the inhibition of anthranilate synthase causes increased flux from chorismate towards prephenate, this would cause an increase in L-phenylalanine, L-
tyrosine and L-glutamate. This is indicated through the results achieved with extracellular extraction samples of *E. coli* NST74 with added anthranilic acid.

7.4.2 Enzyme target: Anthranilate synthase, Inhibitor: Methyl anthranilate

Methyl anthranilate was added as an inhibitor of anthranilate synthase. The concentrations added were: 0, 100µM, 1mM, 2mM, 10mM. Growth readings were obtained, and normalisation took place by adding a proportional amount of solvent during metabolite extraction. Samples were analysed using the targeted method on the triple quadrupole instrument.
Figure 7. 4. A) Growth curves for *E. coli* NST74 with different concentrations of added methyl anthranilate. B) Average peak area for L-phenylalanine in *E. coli* NST74 with different concentrations of added methyl anthranilate, there is a significant difference between the control and 100µM added methyl anthranilate, however this is in an undesirable direction as there is more in the control. Error bars indicate standard deviation.
Growth curve

The lowest concentration of added methyl anthranilate (100µM) grew at the same rate as the control (figure 7.4A). The 1mM added methyl anthranilate condition reached stationary phase quicker, but at a lower OD$_{600}$ than the control. As the concentration of added methyl anthranilate increased to 2mM, it reached a lower optical density. When the concentration of added methyl anthranilate increased to 10mM, the *E. coli* NST74 grew very poorly, and this concentration clearly had a negative effect on the cells. This indicates that the addition of methyl anthranilate has a negative influence on growth of the *E. coli* NST74 cells.

L-phenylalanine

There is a significant difference between the control and 100µM added methyl anthranilate when a t-test was performed ($p = 0.0301$), however this is not in the desired direction, as there is less L-phenylalanine in the condition than in the control (figure 7.4B).

Interestingly, the addition of methyl anthranilate caused a decrease in the amount of extracellular L-phenylalanine, in all conditions. This is most intriguing in the 100µM added methyl anthranilate condition as this grew to the same optical density as the control, yet has an average of 50% less L-phenylalanine. Despite the poor growth rate of the 10mM added methyl anthranilate condition there was still a similar concentration of L-phenylalanine present. This suggests that the effect of methyl anthranilate on the yield of L-phenylalanine was universal, and consistent no matter the concentration.

It is important to note that intracellular data was unable to be obtained due to problems with data acquisition.
Discussion

It is not clear how methyl anthranilate is affecting the metabolism of E. coli NST74, and whether it does act as an analogue of anthranilic acid in inhibiting anthranilate synthase. It is clear, however, that it is not an ideal addition to the media for this strain of E. coli, as it decreases growth and has a negative impact on the extracellular L-phenylalanine yield.

Methyl anthranilate has been synthesised by modified E. coli strains, by converting anthranilate to N-methyl anthranilate, producing concentrations of 185μM (Lee et al., 2019). At this concentration, no negative effect on growth was observed, so it is possible that the concentrations added in this experiment were too high for the E. coli to cope. Either way, there was no positive impact on the yield of L-phenylalanine, and therefore is not suitable for this use.

7.4.3 Enzyme target: Anthranilate synthase Discussion

Overall, it is clear that anthranilate synthase is a good target to lead to an increase in the yield of L-phenylalanine. This is illustrated in chapter 6 with the addition of L-tryptophan and is also exhibited here by anthranilic acid. Cordaro et al. (1968) stated that L-tryptophan is a better inhibitor of anthranilate synthase than anthranilic acid, and this is shown with these results to be true in the case in E. coli NST74.

At concentrations of 1mM and below, anthranilic acid does not negatively affect growth, and does statistically significantly increase the yield of L-phenylalanine by 1.29-fold. Further optimisation could be performed; an example of this could include using concentrations between 1mM and 10mM anthranilic acid to observe the effect on the yield of L-phenylalanine. At 1mM there was a statistically significant increase in the yield, but in practical terms it was a low increase, and at 10mM the cell growth was negatively affected, therefore there may be a concentration in between these values that offers a further increase in yield without having the negative effect on growth rate. As with the previous L-tryptophan addition work, the anthranilic acid
work could also undergo optimisation to see if this increase in L-phenylalanine yield is apparent on a large scale, e.g. a fermenter.

As discussed, it is also clear that methyl anthranilate would not make a good addition to the media with the aim of increasing the yield of L-phenylalanine. The results suggest that methyl anthranilate does not act as a feedback inhibitor for anthranilate synthase and instead acts on the cell through another way, leading to the negative impact on L-phenylalanine production and cell growth.

This research also indicates that in *E. coli* NST74, inhibiting anthranilate synthase has a large effect on the yield of L-phenylalanine; therefore, gene knockouts of this enzyme could be produced. This would prevent the use of expensive chemicals to induce the feedback inhibition but could have the same effect. As anthranilate synthase is encoded for by the TrpE and TrpGD genes, mutations in these genes could be explored. This has already been researched with the aim of preventing bottlenecks in L-tryptophan synthesis (Gu et al., 2012).

One such example of a mutation that could be introduced to *E. coli* NST74 is a mutation in *trpD*9923, which would result in the synthesis of a truncated anthranilate synthase component II protein. This causes the loss of anthranilate phosphoribosyl transferase activity, meaning anthranilate can be synthesised but is not further metabolised, leading to anthranilate accumulation and tryptophan auxotrophy. (Balderas-Hernández et al., 2009). If this mutation was engineered to the *E. coli* NST74 cells, then this may lead to the same effects that the addition of L-tryptophan and anthranilic acid have on the media, thus increasing the yield of L-phenylalanine.

### 7.4.4 Enzyme target: Isochorismate synthase, Inhibitor: N-ethylmaleimide

N-Ethylmaleimide (NEM) was added as it has been shown to inhibit isochorismate synthase (Daruwula et al., 1997). The concentrations added were: 0, 1µM, 10µM, 100µM, 1mM. Growth readings were obtained, and metabolite extraction took place, with solvent added proportional to the growth rate. Samples were analysed using the targeted method on the triple quadrupole instrument.
Growth curve

A) *E. coli NST74* with added N-Ethylmaleimide

![Growth curve](image1)

B) *E. coli NST74* with added N-Ethylmaleimide

![Growth curve](image2)

*Figure 7.5. Growth curves for *E. coli* NST74 with different concentrations of added NEM. A) The plate reader experienced errors in all samples around the 10-hour mark B) The growth curves for the samples that did grow, this was the control and the condition with 1µM added NEM.*
There was an error with the plate reader at approximately 10 hours, with every reading including controls and empty wells (figure 7.5A). This made it difficult to analyse the cell growth of *E. coli* NST74 in these conditions. However, what is clear from the results is that only the control and NST74 with 1µM added NEM grew at all, and up until the 10-hour point, they appeared to have the same rate of growth (figure 7.5B). None of the other conditions showed any growth. This suggests that at concentration of 10µM and higher NEM prevents cell growth and leads to cell death.

**L-phenylalanine**

Both the control and 1µM added NEM sample had two samples where no L-phenylalanine was detected through mass spectrometry analysis, therefore only two samples in each condition were analysed for L-phenylalanine. This means the data collected and discussed here is not very robust. There is no significant difference in extracellular data between the two conditions (*p* = 0.1887) (figure 7.6A).

![Figure 7.6](image-url)
Intracellularly, there is also no significant difference between the two conditions (figure 7.6B). There was only one sample available to analyse for the control, however all 4 samples in the 1µM added NEM were able to detect L-phenylalanine. Again, this makes the data less robust.

**Discussion**

As an overall discussion of using NEM as an inhibitor, it performs very poorly. *E. coli* NST74 is unable to grow in the concentration of 10µM and above, this is not entirely surprising as NEM is able to modify cysteine residues in proteins and peptides, and therefore is toxic to cell growth (Crankshaw and Grant, 1996). However, as Daruwala et al. (1997) found NEM inhibition of isochorismate synthase occurred completely at 100µM, the concentrations of 1µM, 10µM, 100µM, 1mM were tested. The results here suggest the toxicity of NEM leads to cell death before reaching inhibition of the isochorismate synthase enzyme. This is not desired for an inhibitor and means that the inhibition of NEM on isochorismate synthase could not be fully tested as the cells died before inhibition was complete. This explains why there is no significant difference between the conditions.

This experiment does not need to be repeated due to the toxicity of NEM on *E. coli* NST74 at concentrations of 10µM above. Other inhibitors of isochorismate synthase could be tested if this target was still of interest, for example (4R, 5R)-4-hydroxy-5-(1-carboxyvinylxy)-cyclohex-1-ene carboxylate (Švarcová et al., 2015).
7.4.5 Enzyme target: Chorismate pyruvate-lyase, Inhibitor: Vanillic acid

Vanillic acid was added as an inhibitor for chorismate pyruvate lyase. The concentrations added were: 0, 10µM, 100µM, 1mM, 10mM. Growth readings were obtained, and normalisation took place by adding a proportional amount of solvent during metabolite extraction. Samples were analysed using the targeted method on the triple quadrupole instrument.

Growth curve

*E. coli* NST74 with added vanillic acid

Figure 7.7. Growth curves for *E. coli* NST74 with different concentrations of added vanillic acid.

The growth curves indicate that there is no real change in overall growth with the added vanillic acid as the same, or similar, optical density was reached in comparison to the control (figure 7.7). There is, however, a longer lag phase with the higher concentrations of added vanillic acid (1mM and 10mM), which shows that the presence of vanillic acid has some effect on the *E. coli* NST74, and the cells take longer to adapt to these higher concentrations.
L-phenylalanine

In the extracellular samples, there is no significant differences in extracellular L-phenylalanine concentration when a t-test was conducted in order to compare the added vanillic acid conditions to the control (figure 7.8A).

![Graph A) Extracellular L-phenylalanine](image1)

![Graph B) Intracellular L-phenylalanine](image2)

*Figure 7.8. A) Average peak area for extracellular L-phenylalanine in E. coli NST74 control samples and samples with different concentrations of added vanillic acid. B) Average peak area for intracellular L-phenylalanine in E. coli NST74 control samples and samples with different concentrations of added vanillic acid. Error bars indicate standard deviation.*

Similarly, no significant differences in the peak area of intracellular L-phenylalanine were detected when comparing the added vanillic acid conditions to the control (figure 7.8B).

**Discussion**

The result from this research suggests that vanillic acid is not inhibiting chorismate pyruvate lyase, or that the inhibition of chorismate pyruvate lyase does not have any impact on the yield of L-phenylalanine. As the work by Holden et al. (2022) indicates vanillic acid does inhibit chorismate pyruvate lyase, then the conclusion for this
experiment is that inhibition of chorismate pyruvate lyase in *E. coli* NST74 has no significant effect on the yield of L-phenylalanine. This means is not suitable as an inhibitor.

**7.4.6 Enzyme target: 2-amino-4-deoxychorismate synthase, Inhibitor: Nickel sulphate**

Nickel sulphate hexahydrate was added with the aim of inhibiting aminodeoxychorismate synthase. The concentrations added were: 0, 10µM, 100µM, 1mM, 10mM. Growth readings were obtained, and normalisation took place by adding a proportional amount of solvent during metabolite extraction. Samples were analysed used the targeted method on the triple quadrupole instrument.
The data obtained from the growth curve indicates that errors were made in the initial dilution as the initial OD is higher than ideal, and therefore the lag phase of the conditions is not as apparent in previous growth curve experiments (figure 7.9). It also shows that only the control and NiSO$_4$ added at the lowest concentration, 10µM, showed any growth. In comparison to the control, the growth of *E. coli* NST74 with added with 10µM NiSO$_4$ had a longer lag phase and therefore took longer to reach stationary phase. These growth curves also show some contamination in MM1 only samples, however only one well was contaminated, and that was placed next to a control sample that showed cell growth, and therefore it is assumed that the shaking in the plate reader caused this contamination.
L-phenylalanine

In each condition there was one sample that grew poorly, so these were discounted, and therefore in this experiment n=3.

**A) Extracellular L-phenylalanine**

**B) Intracellular L-phenylalanine**

*Figure 7.10. A) Average peak area for extracellular L-phenylalanine in E. coli NST74 control samples and samples with 10µM added nickel sulphate. B) Average peak area for intracellular L-phenylalanine in E. coli NST74 control samples and samples with 10µM added nickel sulphate. Error bars indicate standard deviation.*

In the extracellular L-phenylalanine there is no significant difference between the two conditions ($p = 0.0910$), but there does appear to be less L-phenylalanine in the condition with added NiSO$_4$ than the control (figure 7.10A). This suggests the addition of nickel sulphate has a negative effect on L-phenylalanine production.

The intracellular data also shows no significant difference when statistical analysis was applied in the form of a t-test between the two condition ($p = 0.4886$), however it does look like there is less L-phenylalanine in the added NiSO$_4$ condition (figure 7.10B).
Discussion

Nickel sulphate is not a good addition to the media, this is because in three of the conditions, the samples were unable to grow at all, and growth of the 10µM condition was delayed. This is most likely due to nickel stress and an intolerance to the increase level of nickel ions. Mahalakshmi et al. (2010) saw that as the concentration of nickel sulphate in the media was increased, there was substantial decrease in growth of *E. coli*, and the *E. coli* exhibited a delayed lag phase at increasing concentrations of nickel sulphate. Therefore, in *E. coli* NST74 it seems the addition of nickel at concentrations above 10µM result in cell death.

Whilst there is no significant difference between the two conditions, it still appears that there is less L-phenylalanine present both extracellularly and intracellularly in the added NiSO$_4$ condition. This could be caused by the stress of the added NiSO$_4$ redirecting the metabolic flux to ensure survival.

An alternative inhibitor for this enzyme is 6-diazo-5-oxo-L-norleucine, which could be tested in future work. This was found to inhibit a variety of glutaminases including anthranilate synthase and ADC synthase which, therefore, could have a dual effect on the yield of L-phenylalanine in *E. coli* NST74 by inhibiting two target enzymes (Viswanathan et al., 1995)

7.4.7 Enzyme target: Prephenate dehydrogenase, Inhibitor: L-tyrosine

L-tyrosine was added with the aim of inhibiting prephenate dehydrogenase. The concentrations added were: 0, 100µM, 500µM, 1mM, 10mM. Growth readings were obtained, and normalisation took place by adding a proportional amount of solvent during metabolite extraction. Samples were analysed used the targeted method on the triple quadrupole instrument.
Growth curve

The growth curve shows that at the concentration of 100µM and 500µM added L-tyrosine, growth was almost identical to the control (figure 7.11). At 1mM added L-tyrosine growth was slightly improved reaching a higher optical density at the same time. However, at the 10mM condition there is a different growth pattern and a longer lag phase. This suggests that at this concentration the *E. coli* NST74 is not responding positively to the added L-tyrosine.

![E.coli NST74 with added L-tyrosine](image)

*Figure 7.11. Growth curves for E. coli NST74 with different concentrations of added L-tyrosine.*

L-phenylalanine

The extracellular data obtained showed that some contamination was present as media only samples displayed L-phenylalanine. This was found in the MM1 with 500µM added L-tyrosine, and MM1 with 10mM added L-tyrosine conditions (figure 7.12). This leads to difficulty in quantifying the change in L-phenylalanine.
Figure 7.12. Average peak area for extracellular L-phenylalanine in E. coli NST74 control samples and samples with different concentrations of added L-tyrosine. The controls with added L-tyrosine are also included and indicate some contamination is present. Error bars indicate standard deviation.

It is hypothesised the L-phenylalanine in MM1 with 500µM L-tyrosine and MM1 with 10mM added L-tyrosine, is carry over or contamination on the mass spectrometer as the growth curves for these samples do not indicate any cell growth in these conditions (Figure 7.13). For this reason, this experiment should be repeated for conclusive results.
Figure 7.13. Growth curves for samples of media only controls for E. coli NST74 with different concentrations of added L-tyrosine, showing there was no contamination with bacterial growth.

Nonetheless, the data obtained is still promising, and can provide a preliminary view of how L-tyrosine can affect L-phenylalanine in E. coli NST74. Figure 7.14 displays the same data as figure 7.12, with the controls excluded. This is to highlight the L-phenylalanine peak across the conditions of added L-tyrosine.
Figure 7.14. This is the same data as figure 7.12, however here the media only controls are omitted. This shows the average peak area for extracellular L-phenylalanine in E. coli NST74 control sample and samples with different concentrations of added L-tyrosine. Error bars indicate standard deviation.

As the concentration of L-tyrosine increases, the peak area of L-phenylalanine also increases. Between the control (E. coli NST74 and MM1) and 10mM added L-tyrosine, there is a significant difference ($p = 0.0082$). This can also be illustrated in table 7.4, which indicates that there is almost 7 times increase in extracellular L-phenylalanine than in the control.
Table 7.4. The average peak area and overall fold change of samples of *E. coli* NST74 with different concentrations of added L-tyrosine.

<table>
<thead>
<tr>
<th>Condition</th>
<th>Average peak area</th>
<th>Fold change</th>
</tr>
</thead>
<tbody>
<tr>
<td>NST74 + MM1</td>
<td>537329.865</td>
<td>-</td>
</tr>
<tr>
<td>NST74 + 100µM</td>
<td>628126.6725</td>
<td>1.17</td>
</tr>
<tr>
<td>NST74 + 500µM</td>
<td>910719.5167</td>
<td>1.69</td>
</tr>
<tr>
<td>NST74 + 1mM</td>
<td>1094055.718</td>
<td>2.04</td>
</tr>
<tr>
<td>NST74 + 10mM</td>
<td>3715795.173</td>
<td>6.92</td>
</tr>
</tbody>
</table>

The intracellular data also contains some contamination in the MM1 + 10mM added L-tyrosine conditions, and therefore as with the extracellular data, this experiment should be repeated to obtain clear and reliable results.
Figure 7.15. Average peak area for intracellular L-phenylalanine in E. coli NST74 control samples and samples with different concentrations of added L-tyrosine. Media only controls with added L-tyrosine are also shown and indicate contamination was present. Error bars indicate standard deviation.

Again, these results show a clear indication that the addition of L-tyrosine increases intracellular L-phenylalanine. There is a significant difference, $p = 0.0018$, between the control and 10mM condition.

Discussion

The extracellular data from this experiment indicates there is a 6.92-fold increase in L-phenylalanine between the control and 10mM added L-tyrosine condition. This suggests that the presence of L-tyrosine in E. coli NST74 increases metabolic flux
towards increased L-phenylalanine production. Intracellular L-phenylalanine is also significantly increased.

As discussed, the data obtained from this experiment does have some contamination present. The suspected cause of the contamination is the autosampler or mass spectrometer itself, as the growth curves showed no indication of contamination. This means that the samples could not be re-analysed on the MS and the experiment would need to be repeated for conclusive results. This was not feasible within the time restrictions of this work, but this preliminary data is positive.

The mechanism in which L-tyrosine is thought to inhibit prephenate dehydrogenase is feedback inhibition; this has been illustrated through much previous work such as Chen et al. (2003), and Hudson et al. (1983).

Whilst adding L-tyrosine to media to increase the yield of L-phenylalanine might not be cost effective it provides a target for future genetic modification. Therefore, if this experiment was repeated and displayed similar results, it would indicate that using genetic engineering to mutate this enzyme would be a good target. TyrA mutants have already been created which enabled TyrA to become feedback resistant to tyrosine, and therefore keep on producing (Lütke-Eversloh and Stephanopoulos, 2005). In order to increase the yield of L-phenylalanine this would not help, but instead a modification in TyrA such as a knockout, which prevented tyrosine from being created could be explored. As E. coli NST74 is a prototrophic strain, this should work as it does not require L-tyrosine to survive (Tribe, 1987).

7.4.8 Enzyme target: Prephenate dehydrogenase, Inhibitor: Diethyl malonate

Diethyl malonate was thought to target the same enzyme as L-tyrosine, prephenate dehydrogenase, by preventing binding of the active site by acting as a competitive inhibitor to chorismate (Christopherson et al., 1983). The concentrations added were: 0, 100μM, 1mM, 2mM, 10mM. Growth readings were obtained, and normalisation took place by adding a proportional amount of solvent during metabolite extraction.
Samples were analysed using the targeted method on the triple quadrupole instrument.

**Growth curve**

The results of the growth curve indicate that the addition of diethyl malonate has no significant impact on the growth of *E. coli* NST74 (figure 7.16). All conditions grow with the same pattern, with a slight decrease in optical density with the 2mM and 10mM added diethyl malonate conditions.

---

**E. coli NST74 with added diethyl malonate**

![Growth curves for E. coli NST74 with different concentrations of added diethyl malonate.](image)

*Figure 7.16. Growth curves for E. coli NST74 with different concentrations of added diethyl malonate.*

**L-phenylalanine**

The extracellular samples show a significant difference between the control and 10mM added diethyl malonate sample (*p* = 0.0172), however this is in the direction of decrease L-phenylalanine in the added diethyl malonate sample, and therefore
shows that the addition of diethyl malonate does not have a positive impact on the yield of L-phenylalanine (figure 7.17A).

The intracellular data shows no significant difference between the control and all conditions with added diethyl malonate (figure 7.17B).

**Discussion**

These results indicate that diethyl malonate does not inhibit prephenate dehydrogenase in *E. coli* NST74 despite the purified enzyme from *E. coli* being inhibited in previous work (Christopherson et al., 1983).
Another possibility is that it does inhibit the enzyme, but this does not have any effect on the yield of L-phenylalanine in *E. coli* NST74. This seems unlikely, seen as other work performed in this chapter implies that feedback inhibition of the same enzyme by L-tyrosine increases the yield of L-phenylalanine significantly.

### 7.4.9 Enzyme targets: Anthranilate synthase and prephenate dehydrogenase, Inhibitor: L-tryptophan and L-tyrosine

Due to the success of L-tryptophan as an inhibitor of anthranilate synthase (Chapter 6), and L-tyrosine as an inhibitor of prephenate dehydrogenase (Chapter 7, Section 7.4.7), it was decided to combine the two inhibitors to see how this affected the yield of L-phenylalanine in *E. coli* NST74. Due to issues with the plate reader, no growth curve readings could be obtained.

**L-phenylalanine**

The data obtained, as with the L-tyrosine data, appears to have some contamination. This involved small amounts L-phenylalanine being present within the media only controls, this is most likely due to carryover on the MS, and therefore this data would need to be repeated in order to confirm the results.
Figure 7.18. Average peak area for extracellular L-phenylalanine in E. coli NST74 control samples and samples with added L-tryptophan and L-tyrosine in different combinations. Error bars indicate standard deviation.

The results from this experiment indicate that all the conditions have significantly higher extracellular L-phenylalanine than the control (10mM added L-tyrosine, p < 0.0196; 10mM added L-tryptophan, p < 0.0001; 10mM added L-tryptophan and 1mM added L-tyrosine, p < 0.0001 and 10mM added L-tryptophan and 10mM added L-tyrosine, p < 0.0002). The fold change is illustrated in Table 7.5.
Table 7.5. Average peak area and overall fold change for extracellular E. coli NST74 samples grown with different concentrations of added L-tryptophan and L-tyrosine.

<table>
<thead>
<tr>
<th>Condition</th>
<th>Average L-phenylalanine peak</th>
<th>Fold change</th>
</tr>
</thead>
<tbody>
<tr>
<td>NST74 + MM1</td>
<td>378020.1</td>
<td>-</td>
</tr>
<tr>
<td>NST74 + 10mM Trp</td>
<td>856473.3</td>
<td>2.3</td>
</tr>
<tr>
<td>NST74 + 10mM Tyr</td>
<td>687694.5</td>
<td>1.8</td>
</tr>
<tr>
<td>NST74 + 10mM Trp, 1mM Tyr</td>
<td>1438449</td>
<td>3.8</td>
</tr>
<tr>
<td>NST74 + 10mM Trp, 10mM Tyr</td>
<td>728414.4</td>
<td>1.9</td>
</tr>
</tbody>
</table>

This data indicates that the combination of 10mM added L-tryptophan and 1mM added L-tyrosine produces the largest effect on L-phenylalanine yield, leading to a 3.8-fold increase. This fold increase is lower than when both L-tryptophan and L-tyrosine are added separately, however, repeating this experiment and obtaining growth curves could provide a reason for this.

Discussion

This data indicates that the combination of 10mM added L-tryptophan and 1mM added L-tyrosine produces the largest effect on L-phenylalanine yield, leading to a 3.8-fold increase. This fold increase is lower than when both L-tryptophan and L-tyrosine are added separately, however, repeating this experiment and obtaining growth curves could provide a reason for this. This experiment should also be repeated with a condition adding 1mM tyrosine by itself. This would ensure any changes in the yield of L-phenylalanine was directly caused by the addition of L-tryptophan and L-tyrosine at their specified concentrations.

As previously mentioned, (chapter 6), the success of L-tryptophan as an inhibitor to increase the yield of L-phenylalanine provides targets for future genetic
modifications, such as a mutation in trpD, resulting in a truncated anthranilate synthase enzyme (Balderas-Hernández et al., 2009). Similarly, TyrA mutations or knockout could be explored to mimic the effects of the added L-tyrosine (Lütke-Eversloh and Stephanopoulos, 2005). To increase the export of L-phenylalanine out of the cell, modifications to over express the YddG protein, which can increase exportation by three-fold (Doroshenko et al., 2007). Therefore, modifying all three of these targets should increase the yield of extracellular L-phenylalanine further in E. coli NST74.

7.5 Conclusion

7.5.1 Inhibitors

Out of the 8 inhibitors tested, 3 were found to increase the yield of L-phenylalanine. These were L-tryptophan, anthranilic acid and L-tyrosine. All three of these inhibit the target enzyme through feedback inhibition.
Table 7.6. Summary of inhibitors used and the fold increase this had on the yield of extracellular L-phenylalanine.

<table>
<thead>
<tr>
<th>Enzyme</th>
<th>Inhibitor</th>
<th>Concentration</th>
<th>Fold increase in extracellular L-phenylalanine yield</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anthranilate synthase</td>
<td>L-tryptophan</td>
<td>10mM</td>
<td>4.61 (5.39 with time delay)</td>
</tr>
<tr>
<td></td>
<td>Anthranilic acid</td>
<td>100µM</td>
<td>1.29</td>
</tr>
<tr>
<td>Prephenate dehydrogenase</td>
<td>L-tyrosine</td>
<td>10mM</td>
<td>6.92</td>
</tr>
<tr>
<td>Both anthranilate synthase and prephenate dehydrogenase</td>
<td>L-tryptophan and L-tyrosine</td>
<td>10mM L-tryptophan, 1mM L-tyrosine</td>
<td>3.8</td>
</tr>
</tbody>
</table>
Figure 7.19. Diagram of the synthesis of L-phenylalanine with the pathways targeted in this thesis and the effect these inhibitors had on L-phenylalanine. Green arrows symbolise an increase on the yield. Red arrows symbolise a decrease on the yield. Red X symbolises negative effect on cell viability. Yellow = symbolises no significant difference or effect on both yield and cell viability.
Product feedback inhibition was first reported in the 1950s; Pardee and Yates (1956) published a paper on pyrimidine biosynthesis in *E. coli* by a feedback mechanism and showed feedback inhibition is vital to control fluxes for optimal growth in microorganisms. In the case of amino acid feedback inhibition, the first enzyme in the pathway is an allosteric enzyme that binds to the end product. This alters its active site so that it can no longer catalyse the reaction needed to initiate the pathway. Ultimately, the pathway is prevented, as long as adequate amounts of the end product are present. Inhibition can be relieved, and the enzyme will regain its activity if the end product is no longer available (Naz et al., 2023). In the case of *E. coli NST74*, it appears that production of L-phenylalanine is improved significantly by feedback inhibition of the other amino acids.

Conversely, 5 out the 8 inhibitors tested either had no effect, or a negative effect on the yield of L-phenylalanine in *E. coli NST74*.

In the case of methyl anthranilate, it is most likely that this is not an analogue of anthranilic acid, and therefore does not bind or inhibit anthranilate synthase. NEM causes cell death through toxicity at the higher concentrations, leaving its inhibition of chorismate pyruvate lyase defunct. Similarly, NiSO₄ causes cell death and therefore its inhibition of its target enzyme in *E. coli NST74* cannot be properly evaluated. Despite the purified enzyme being inhibited by vanillic acid, chorismate pyruvate lyase either remains unaffected in *E. coli NST74*, or changes in metabolic pathways are not directed to the L-phenylalanine pathway. Similarly, diethyl malonate does not appear to inhibit prephenate dehydrogenase in *E. coli NST74*, despite showing inhibition elsewhere.

This emphasises the importance of using the correct inhibitors and looking at the correct enzyme targets. The use of inhibitors enables quick and cheap analysis to be performed. If this work had been tested using genetic modifications, then it would have been a longer process as it would have required the creation of multiple genetic constructions. It also does not allow the same flexibility that adding external inhibitors does as the concentration and level of inhibition cannot be easily altered. The addition of the inhibitors has then successfully identified targets for further genetic modification. As proof of concept has been illustrated with the inhibitors, genetic modifications could take place and have a higher likelihood of being successful. This
would prevent the enzymes becoming active again if the concentration of the inhibitor is depleted and should lead to increased yield of L-phenylalanine.

The main aim of this section was to probe the metabolism of *E. coli* NST74 with inhibitors to increase the yield of L-phenylalanine, this was successful with three inhibitors increasing the yield of L-phenylalanine by 1.29-fold (anthranilic acid), 5.39-fold (L-tryptophan) and 6.92-fold (L-tyrosine). Whilst some final work is needed to confirm (L-tyrosine) and optimise (anthranilic acid), it shows proof of concept for the inhibitor panels.

### 7.5.2 *E. coli* NST74

As mentioned, the ultimate aim was to increase the yield of L-phenylalanine in *E. coli* NST74, which has a yield of 4g/L when cultured in a 1L fermenter (Tribe, 1987), however, many strains of *E. coli* have been engineered to over produce L-phenylalanine at much higher yields.

This includes, but is not limited to, work conducted by Zhou et al. (2010), where *E. coli* WSH-Z06 with a pAP-B01 plasmid (*pheA*<sup>br</sup> and *aroF*<sup>wt</sup> genes, Kan<sup>r</sup>) resulted in yields of 35.38g/L in a 3-L fermenter. Similarly, Liu et al. (2018) used genetic engineering with *E. coli* W3110 with a XIlp21 (overexpressed *aroF<sup>WT</sup>*, *aroD*, and *pheA*<sup>br</sup>, inactivated *tyrR*, and genomic integrated *galP* and *glk*) was able to yield 72.9g/L when cultured in a 5L fermenter. Liu et al. (2019) provides a comprehensive view of other L-phenylalanine producing strains and their maximum yields.

Both these strains show that in comparison the yield of L-phenylalanine in *E. coli* NST74 is quite low, and therefore if looking for a strain for L-phenylalanine production, other strains would be preferred. However, with *E. coli* NST74 the L-phenylalanine has the ability to be converted to other high value products such as styrene (McKenna and Neilson, 2011), phenol (Thompson et al., 2021), mandelic acid (Lukito et al., 2021), 2-phenylethanol (Sekar et al., 2019), and therefore this strain is a useful strain to optimise and show proof of concept.

*E. coli* NST74 also does not contain a plasmid, unlike the L-phenylalanine strains mentioned above, which has its own advantages. Plasmid free strains are desired...
from an industrial perspective as they have lower cultivation costs as well as lower public health and environmental risk (Kroll et al., 2010). Plasmid compatibility can also be a problem, as when multiple plasmids are combined, they each require different antibiotic resistance markers, this leads to high cultivation costs and risks antibiotic resistance. Alternatively, a plasmid might become too large and unstable if introducing genes in large numbers (Koma et al., 2020). Therefore, plasmid-free L-phenylalanine producing strains are more desirable, and using inhibitors to increase the yield further could be a viable and cheaper alternative for L-phenylalanine production in industry.

One approach that has not been explored during this work, but should be, is using design of experiments to optimise the use of multiple inhibitors simultaneously in order to explore the maximum effect on the yield of L-phenylalanine. This could be expanded to different inhibitors of the specific enzymes chosen in this work but also other targets within the shikimate and L-phenylalanine pathway.
Chapter 8.

General discussions and conclusions

The hypothesis of this thesis was to reduce the bottleneck in the synthetic biology ‘design-build-test-learn’ cycle using high throughput metabolomics and mass spectrometry. This was tested using inhibitors to increase the yield of a high value product.

Metabolomics is a comprehensive analysis in which metabolites in a biological system are identified and quantified (Fiehn, 2002). The analytical technique chosen was mass spectrometry, which separates ions based on their mass \((m)\) to charge \((z)\) ratio. One advantage of MS is its wide detection range, and thus its capability of analysing small metabolites to entire proteins (Lee et al., 2010). This was then applied to a synthetic biology application. Synthetic biology is useful as it can redesign organisms for specific uses. Further to this, there are many instances where applied synthetic biology has life-changing implications; from medicinal drugs to environmental benefits (Yan et al., 2023). Therefore, if this process is more streamlined, it could result in an increase in engineered strains that produce more of a desired product, in a faster timespan.

The mass spectrometry method, described in chapter 3, is a targeted flow injection method using multiple-reaction-monitoring on a triple quadrupole instrument. For the purpose of high throughput and ease of data analysis, targeted metabolomics was the preferred option. The use of flow injection removed the need for lengthy chromatographic separation, and whilst this has its own limitations, such as ion suppression, the method was successfully optimised so the metabolites of interest could be identified. The application of the flow injection method using targeted MS on a triple quadrupole instrument provides rapid and reproducible data acquisition and simple data analysis. It took under 2.5 minutes per sample, taking 4 hours per 96 well plate. This provides fast and robust acquisition as well as providing quick and easy data analysis. Further optimisation could include optimising theautosampler speed to make it faster, as well as further optimisation of the MRM.
Another aspect explored was a low-cost automation platform to increase throughput. This involved the use of a LEGO® liquid handling robot (Chapter 4). The building and programming of the robot was a success, this robot was easy to build and modify. The programming was user friendly, and it has the ability to be compatible with multiple coding languages, thus providing it with flexibility. Optimisation is required to obtain more reproducible data, but the robot is successfully capable of performing serial dilutions and has the programming ability to produce cell culture set-ups along with metabolite extractions. Whilst this robot does have limitations, such as not being able to use single-use pipettes, these are limitations that much more expensive liquid handling robots also possess.

Overall, this work on low-cost automation shows that the advancements in open-source robotics has increased considerably, and this upward trend seems likely to continue. With this in mind, the benefits of automation should be clear to all researchers. It could mean off-loading tedious and repetitive tasks, as well as reducing the likelihood of human error. Whilst open-source automation has its difficulties in comparison to commercial robots, such as the initial build and programming, they provide thorough understanding and the results are just as credible as commercial robots, which are much more costly.

In terms of high value products, the first product evaluated was styrene. However, as styrene is a volatile product, analysis was difficult. This illustrates the importance of the availability and suitability of analytical instruments. This was one reason, amongst others, that the high value product was changed to L-phenylalanine.

L-phenylalanine is an amino acid with industrial uses as a component in L-aspartame. *E. coli* NST74 was engineered to overproduce L-phenylalanine, and in this work, inhibitors were added to alter the metabolic flux of nearby pathways. The aim here was to see if this could increase the yield of L-phenylalanine further and provide future targets for additional synthetic biology modifications. Three out of eight tested inhibitors increased the yield of extracellular L-phenylalanine, with the highest fold change close to 7 times increase.

This work provided targets for future genetic engineering which should be explored, as the inhibition of two pathways increased the yield. Therefore, mutations in these targets may show the same results as the addition of the inhibitor. The inhibitor panel
was a great success when applied to *E. coli* NST74. Future work would involve implementing this on other engineered strains.

Overall, this work illustrates how the combination of low-cost automation, high throughput flow injection mass spectrometry methods, and inhibitor panels are the start of a high throughput platform for strain optimisation and rapid testing using metabolomics. This could be implemented in many areas of synthetic biology and biotechnology.

*Figure 8. 1. DBTL cycle with the platforms developed in this work.*

In synthetic biology, the cost of DNA sequencing and synthesis have decreased dramatically (Carlson, 2014; Kosuri and Church, 2014). As a result, this is now
accessible to many labs who can design and perform genetic constructs. However, this is often a slow and iterative process involving trial and error. Automation can increase throughput, and the use of low-cost automation can ensure accessibility is not an issue. The use of the inhibitor panel can identify targets for future modification and can decrease the time taken on failed genetic constructs for inefficient targets. As stated in El Karoui et al. (2019), “We can do ‘build’. ‘Test’ is the challenge when we want to learn from the iterative design process”. The flow injection targeted method developed in this thesis provides fast data acquisition and rapid analysis, subsequently helping to decrease this bottleneck, and solves this problem for this particular organism and high value product. It could easily be optimised for other strains. As machine learning expands, it will become an emerging tool to predict the organisms behaviour and aid the learn phase (Radivojević et al., 2020). A combination of these technologies will streamline the synthetic biology DBTL cycle, and result in optimum throughput for synthetic biology and its endless possibilities.
References


Li, G. and Young, K.D. (2013). Indole production by the tryptophanase TnaA in Escherichia coli is determined by the amount of exogenous tryptophan. *Microbiology* [online] 159(Pt 2), pp.402–10. doi:https://doi.org/10.1099/mic.0.064139-0.


Mahalakshmi, T., Illamathi, M., Siva, R. and Sridharan, T.B. (2010). EFFECT OF NICKEL STRESS ON ESCHERICHIA COLI AND SACCHAROMYCES CEREFISIAE. *I Control Pollution*. [online] Available at: https://www.semanticscholar.org/paper/EFFECT-OF-NICKEL-
STRESS-ON-ESCHERICHIA-COLI&-Mahalakshmi-llamathi/16d7f30aef5f66be1a7ba5e29fa6b4ff94a2d [Accessed 19 Sep. 2023].

doi:https://doi.org/10.5702/massspectrometry.s0009.

doi:https://doi.org/10.3390/polym12122908.

doi:https://doi.org/10.1016/j.ymben.2011.06.005.

doi:https://doi.org/10.1002/biot.201300035.


doi:https://doi.org/10.3109/10408449409020137.


Phenylyalanine Future Market Insights, (2023) [online] Available at: https://www.futuremarketinsights.com/reports/phenylalanine-market#:~:text=The%20global%20phenylalanine%20market%20reached


Sekar, B. S., Lukito, B. R., and Li, Z. (2019). Production of Natural 2-Phenylethanol from Glucose or Glycerol with Coupled *Escherichia coli* Strains Expressing l-Phenylalanine


