

**Catechol-*Escherichia coli* UM146 interaction revealed through  
multi-omics**

by

Md Shiful Islam

A thesis submitted in partial fulfillment of the requirements for the degree of

Master of Science

in

Molecular Biology & Genetics

Department of Biological Sciences

University of Alberta

© Md Shiful Islam, 2024

## Abstract

Microorganisms are intricately linked with life on earth. The substantial enhancement of environment and human health is influenced largely due to their biotransformation potency. The comprehensive identification of bacterial byproducts can be accomplished by employing a systematic approach involving untargeted metabolomics techniques, with the integration of liquid chromatography high-resolution mass spectrometry (LC-HRMS), nuclear magnetic resonance (NMR), and gas chromatography-mass spectrometry (GC-MS). Untargeted metabolomics is an analytical approach to characterize the global metabolites without any prior knowledge. This study attempted to combine untargeted metabolomics and RNA sequencing (RNA-Seq) to determine unknown bacterial byproducts and genes to highlight the metabolic pathway in a particular condition.

Studying the *Escherichia coli* UMI46 strain grown on catechol-containing media in aerobic and anaerobic conditions, we identified novel genetic and metabolic changes through transcriptomic and untargeted metabolomics analysis. As highlighted in Chapter 2, many inconsistencies were noted between the NMR and MS metabolomics results, most specifically, false positive results were observed in the MS analyses. We concluded that these results were highly dependent on the available software being used and the analytical workflow employed.

Benchmarking experiments were performed to evaluate false positive and false negative rates in the identification of a set of 28 compounds in a synthetic mixture. XCMS performed better than Metaboanalyst and MZmine2 in determining minimum features with coverage of more compounds. Sensitivity and specificity were tested based on three distinct approaches including a) all features, b) putative ID, and c) putative ID & true positive (PID28). In negative mode, all three software packages provided similar sensitivity (70-75%). Variability was observed for PID28 in positive mode, XCMS (75%) has outperformed both Metaboanalyst (58.33%) and MZmine2 (66.67%). The specificity of Metaboanalyst in the criteria of all features is 44.44%, which is quite inferior compared to XCMS (96.09%) and MZmine2 (100%). As inconsistent results were observed with the synthetic mixture, pure concentrated lysine from Sigma-Aldrich (purity  $\geq 98\%$ ) was used to determine the contaminants through resin-based column chromatography and  $^1\text{H-NMR}$ . Among the three software packages in LC-HRMS analysis, entirely Metaboanalyst was able to determine the unknown contaminant (2-piperidinone) in the lysine solution.

Bacterial by-product identification using LC-HRMS analysis is challenging due to false positive results. Our analysis demonstrated that a combination of software packages is required to screen the actual features and reduce the false positive results in MS analysis.

## Preface

This thesis is an original work by Shiful Islam. No part of this thesis has been previously published. This research project was led by Dr. David Wishart from the Departments of Biological Sciences and Computing Science at University of Alberta. All experiments and research activities were performed in Dr. Wishart's Lab and The Metabolomics Innovation Centre (TMIC) at the University of Alberta. Dr. Wishart contributed to all stages of conceptualization, and coordination. I was assisted by several people in Dr. Wishart's Lab to finish the work. Dr. Rupasri Mandal, TMIC lab manager, trained me to operate the NMR analysis software (Chenomx), and assisted with GC-MS analysis. Dr. Mickel Hiebert and Dr. Lun Zhang assisted and collected the LC-MS/MS spectra used in this thesis. Dr. Brian Lee trained for the NMR analysis. Also, Dr. Lee collected all the NMR spectra used in this thesis. Dr. Dipanjan Bhattacharyya assisted with the resin-based column chromatography work for the metabolite separation and purification. *E. coli UMI46* strain was generously provided by Dr. Karen Madsen's lab from the Department of Medicine at University of Alberta. Free access of AnalyzerPro XD was obtained from SpectralWorks Ltd for GC-MS analysis.

# Table of Contents

<b>ABSTRACT</b> .....	<b>II</b>
<b>PREFACE</b> .....	<b>IV</b>
<b>TABLE OF CONTENTS</b> .....	<b>V</b>
<b>LIST OF TABLES</b> .....	<b>VIII</b>
<b>LIST OF SUPPLEMENTARY TABLES</b> .....	<b>IX</b>
<b>LIST OF FIGURES</b> .....	<b>X</b>
<b>LIST OF SUPPLEMENTARY FIGURES</b> .....	<b>XIII</b>
<b>CHAPTER 1: INTRODUCTION</b> .....	<b>1</b>
1.1. BACTERIAL METABOLITES: CHARACTERISTICS FEATURES.....	1
1.1.1. Secondary metabolites: a source of bioactive compounds .....	2
1.1.2. Microbial degradation of plant secondary metabolites.....	3
1.2. METABOLOMICS: A TOOL FOR BACTERIAL NATURAL PRODUCT IDENTIFICATION ..	4
1.2.1. Common analytical techniques in microbial metabolomics.....	5
1.2.1.1. Mass spectrometry .....	6
1.2.1.1.1. MS ionization methods .....	7
1.2.1.1.2. Mass analyzers for MS.....	8
1.2.1.1.3. Tandem Mass Spectrometry.....	9
1.2.1.1.4. Chromatography Coupled to MS .....	10
1.2.1.1.5. Analysis of mass spectrometry data .....	11
1.2.1.1.6. Challenges in untargeted mass spectrometry data analysis: false positive/false negative results .....	16
1.2.1.2. Nuclear magnetic resonance (NMR) spectroscopy.....	17
1.2.1.2.1. NMR Basics .....	17
1.2.1.2.2. Chemical shifts.....	21
1.2.1.2.3. Coupling Constants .....	21
1.2.1.2.4. One-dimensional NMR spectroscopy .....	22
1.2.1.2.5. Two-Dimensional NMR spectroscopy.....	22
1.2.1.2.6. COSY and TOCSY .....	24
1.2.1.2.7. HSQC and HMBC.....	24
1.2.1.2.8. Software and Databases for Compound Identification in NMR metabolomics .....	24
1.3. ALTERNATIVE METHODS FOR METABOLITE IDENTIFICATION .....	25
1.4. TRANSCRIPTOMICS .....	26
1.5. RATIONALE OF STUDYING CATECHOL METABOLISM IN <i>ESCHERICHIA COLI</i> .....	27
1.6. STATEMENT OF THESIS PROBLEM.....	28
1.7. THESIS OBJECTIVES .....	29
1.8. THESIS OUTLINE .....	29

<b>CHAPTER 2 .....</b>	<b>31</b>
2.1. INTRODUCTION .....	31
2.2. MATERIALS AND METHODS .....	34
2.2.1. Chemical reagents.....	34
2.2.2. Strains, media, substrate, and culture conditions.....	34
2.2.3. RNA extraction and quality check.....	35
2.2.4. Library preparation .....	36
2.2.5. RNA-seq data analysis.....	37
2.2.6. Sample preparation for metabolomics analysis .....	37
2.2.6.1. Extracellular metabolites extraction .....	37
2.2.6.2. Intracellular metabolites extraction.....	37
2.2.7. Orbitrap analytical conditions .....	38
2.2.8. LC-MS/MS data analysis.....	38
2.2.9. GC-MS analysis .....	39
2.2.9.1. GC-MS data processing .....	39
2.2.10. NMR analysis .....	40
2.2.11. In vitro enzymatic and spontaneous reactions.....	40
2.2.12. Generation of 3D protein homology model.....	41
2.2.13. Ligand preparation, docking grid generation and molecular docking.....	41
2.3. RESULTS .....	41
2.3.1. Effects of catechol exposure on growth of <i>E. coli</i> UM146 strain.....	42
2.3.2. Effect of catechol stress on genetic level.....	43
2.3.2.1. Overall differential expression of genes .....	43
2.3.2.2. Influence of catechol exposure on two-component systems.....	47
2.3.2.3. Impact of catechol exposure on stress responsive genes .....	54
2.3.2.4. Catechol causes membrane damage.....	55
2.3.2.5. Effect of Catechol on DNA damage .....	58
2.3.2.6. Impact of catechol exposure on virulence factor, biofilm formation and flagellar assembly.....	62
2.3.3. Determination of potential metabolites involved in catechol detoxification pathway .....	65
2.3.4. Polyphenol oxidase assay .....	66
2.3.5. Unveiling of genes involved in catechol detoxification pathway.....	66
2.3.5.1. ubiD gene is responsible for decarboxylase activity in <i>E. coli</i> UM146 strain.....	66
2.3.5.2. yfiH gene encodes polyphenol oxidase in <i>E. coli</i> UM146 strain.....	69
2.3.5.3. 4-hydroxybenzoate to 4-ethoxybenzoic acid ethyl ester step .....	72
2.3.6. Exploring binding sites of enzyme-metabolite interaction.....	73
2.3.7. Catechol detoxification pathway .....	73
2.3.7.1. Spontaneous reactions: an insight into chemical transformations .....	77
2.4. DISCUSSION .....	80
2.4.1. Catechol treatment: toxicity or catabolism.....	81
2.4.1.1. Does ortho or meta cleavage route exist in catechol metabolism of <i>E. coli</i> UM146? .....	82
2.4.1.2. Genetic changes due to catechol toxicity.....	83

2.4.1.3. Strategies used by E. coli UM146 to mitigate the toxic effects of catechol .....	85
2.4.2. Unknown challenging metabolites .....	85
2.4.3. Limitations of the study .....	86
2.5. CONCLUSION .....	86
<b>CHAPTER 3 .....</b>	<b>88</b>
3.1. INTRODUCTION .....	88
3.2. METHODS AND MATERIALS .....	89
3.2.1. Chemical reagents.....	89
3.2.2. Preparation of defined mixture .....	89
3.2.3. Preparation of a single concentrated pure solution.....	90
3.2.3.1. Resin based column chromatography .....	90
3.2.3.2. Liquid-liquid extraction .....	90
3.2.4. LC-HRMS analysis.....	91
3.2.4.1. Mass spectrometry data analysis.....	91
3.2.5. NMR analysis .....	92
3.2.6. Sensitivity and Specificity .....	92
3.2.7. Statistical analysis.....	92
3.3. RESULTS .....	92
3.3.1. BENCHMARKING: MIXTURE OF COMPOUNDS.....	93
3.3.1.1. Features determination .....	93
3.3.1.2. Sensitivity and specificity of Metaboanalyst, XCMS and MZmine2.....	94
3.3.2. BENCHMARKING: PURE ( $\geq 98\%$ ) LYSINE SOLUTION .....	96
3.3.2.1. Results of lysine sample on different platform.....	96
3.3.2.2. Results of liquid-liquid extraction and resin-based column chromatography .....	98
3.3.2.3. False positive results.....	101
3.4. DISCUSSION .....	103
3.5. CONCLUSION .....	106
<b>CHAPTER 4: CONCLUDING REMARKS .....</b>	<b>107</b>
4.1. SUMMARY .....	107
4.2. FUTURE DIRECTIONS .....	107
4.2.1. Chapter 2.....	107
4.2.2. Chapter 3.....	108
<b>REFERENCES.....</b>	<b>110</b>
<b>APPENDIX: SUPPLEMENTARY TABLES AND FIGURES .....</b>	<b>143</b>
APPENDIX A: CHAPTER 1 .....	143
APPENDIX A: CHAPTER 2 .....	147

## List of Tables

<i>Table 2.1. List of detoxification products determined in treated samples under aerobic and anaerobic conditions</i> .....	65
<i>Table 2.2. List of enzymes and their identity with the 3-octaprenyl-4-hydroxybenzoate carboxylase enzyme of E. coli UMI46 strain</i> .....	67
<i>Table 2.3. List of binding energies between enzymes and substrates</i> .....	73
<i>Table 3.1. List of chemical shifts and features of compounds determined in concentrated lysine (85 mM) using <sup>1</sup>H-NMR analysis</i> .....	97
<i>Table 3.2. List of features determined from 100 uM lysine using LC-MS analysis</i> .....	97
<i>Table 3.3. List of features determined from 100 uM lysine using LC-MS/MS analysis</i> .....	98
<i>Table 3.4. Comparison between MS-analyzing software results and L-lysine spectra to determine false positive results</i> .....	103

## **List of Supplementary Tables**

<i>S Table 1. List of studies that reported microbial bioactive compounds synthesis in response to secondary metabolite containing media. ....</i>	<i>143</i>
<i>S Table 2.1. List of yield and quality of each RNA samples. ....</i>	<i>147</i>
<i>S Table 2.2. List of number of reads, quality and overall mapping alignment rate. ....</i>	<i>147</i>

## List of Figures

<i>Figure 1.2. A steady increase in the most commonly utilized analytical techniques for microbial metabolite determination in PubMed Since 2000</i> .....	6
<i>Figure 1.3. Basic principles of a typical mass spectrometry</i> .....	7
<i>Figure 1.4. The use of XCMS, Metaboanalyst, MZmine2 and MS-DIAL has increased dramatically in the past 20 years</i> .....	16
<i>Figure 1.4. A typical NMR instrument, components, and inner part. A) Design of an NMR spectrometer and its components. B) Inner part of an NMR</i> .....	19
<i>Figure 1.5. A typical 700 MHz <sup>1</sup>H NMR spectrum of a natural mixture</i> .....	20
<i>Figure 1.6. A typical 600 MHz HSQC (2D NMR) spectrum of L-lysine</i> .....	23
<i>Figure 2.1. Growth curve of E. coli UM146 after catechol exposure in both aerobic (A) and anaerobic (B) conditions</i> .....	42
<i>Figure 2. 3. A volcano plot representation of the differentially expressed genes in E. coli UM146 under catechol exposure in both (A) anaerobic, and (B) aerobic conditions</i> .....	45
<i>Figure 2.4. A graphical representation of gene set enrichment of E. coli UM146 under catechol treatment in both anaerobic (A) and aerobic (B) conditions</i> .....	46
<i>Figure 2.5. A visual contrast between aerobic and anaerobic conditions with differentially expressed genes across key metabolic pathways under catechol exposure</i> .....	47
<i>Figure 2.7. Heatmap showing gene expression changes associated with two component systems induced by catechol exposure under aerobic condition</i> .....	50
<i>Figure 2.8. Effect of catechol exposure on two component systems under anaerobic condition</i> .	51
<i>Figure 2.9. Effect of catechol exposure on two component systems under aerobic condition</i> .....	52
<i>Figure 2.10. Heatmap showing gene expression changes associated with efflux pump induced by catechol exposure under both anaerobic (A) and aerobic (B) conditions</i> .....	53
<i>Figure 2.11. Heatmap showing exposure of catechol induced expression of different stress responsive genes under aerobic condition</i> .....	54
<i>Figure 2.12. Heatmap showing exposure of catechol induced expression of different stress responsive genes under anaerobic condition</i> .....	55
<i>Figure 2.13. Heatmap showing gene expression changes associated with membrane damage induced by catechol exposure under both anaerobic (A) and aerobic (B) conditions</i> .....	56

<i>Figure 2.14. Visual representation of catechol induced gene expression changes associated with solvent tolerance, reactive oxygen species, detoxification, and cellular redox balance under both anaerobic (A) and aerobic (B) conditions</i> .....	57
<i>Figure 2.15. Heatmap showing gene expression changes associated with DNA damage induced by catechol exposure under anaerobic condition</i> .....	58
<i>Figure 2.16. Heatmap showing gene expression changes associated with DNA damage induced by catechol exposure under aerobic condition</i> .....	59
<i>Figure 2.17. Effect of catechol exposure on DNA mismatch repair pathway under anaerobic (A) and aerobic (B) conditions</i> .....	60
<i>Figure 2.18. Heatmap showing gene expression changes associated with cell cycle control, division, and cell wall biosynthesis induced by catechol exposure under anaerobic (A) and aerobic (B) conditions</i> .....	61
<i>Figure 2.19. Heatmap showing gene expression changes associated with virulence factor induced by catechol exposure under both anaerobic (A) and aerobic (B) conditions</i> .....	63
<i>Figure 2.20. Effect of catechol exposure on biofilm formation under anaerobic condition</i> .....	64
<i>Figure 2.21. Effect of catechol exposure on biofilm formation under aerobic condition</i> .....	64
<i>Figure 2.22. Conserved domain comparison among phenolic acid decarboxylase, gallate decarboxylase and 3-Octaprenyl-4-hydroxybenzoate decarboxylase</i> .....	67
<i>Figure 2.23. Comparison of substrate binding pocket in ubiD protein (Escherichia coli UM 146) with (A) catechol, (B) 2-methoxyphenol, (C) phenol, and (D) 2-octaprenylphenol (positive control)</i> .....	68
<i>Figure 2.24. Pairwise alignment comparison between polyphenol oxidase and purine nucleoside phosphorylase</i> .....	69
<i>Figure 2.25. Conserved domain comparison between polyphenol oxidase and purine nucleoside phosphorylase</i> .....	70
<i>Figure 2.26. Comparison of substrate binding pocket of purine nucleoside phosphorylase and polyphenol oxidase</i> .....	71
<i>Figure 2.27. Comparison of substrate binding pocket of esterase with 4-hydroxybenzoate and 3-(acetylthio)isobutyric acid respectively</i> .....	73
<i>Figure 2.28. Detoxification pathway of catechol during anaerobic condition</i> .....	75
<i>Figure 2.29. Detoxification pathway of catechol during aerobic condition</i> .....	76

<i>Figure 2.30 Heatmap showing gene expression changes associated with detoxification pathway induced by catechol exposure under anaerobic and aerobic conditions.....</i>	<i>77</i>
<i>Figure 3.1. Venn diagram represents the number of features determined by Metaboanalyst, XCMS and MZmine2 .....</i>	<i>93</i>
<i>Figure 3.2. (A) Sensitivity and (B) specificity of Metaboanalyst, XCMS, and MZmine2 in positive mode.....</i>	<i>95</i>
<i>Figure 3.3. (A) Sensitivity and (B) specificity of Metaboanalyst, XCMS, and MZmine2 in negative mode.....</i>	<i>96</i>
<i>Figure 3.4. Presence of 2-piperidinone in a concentrated lysine solution determined using resin-based column chromatography and <sup>1</sup>H-NMR analysis.....</i>	<i>99</i>
<i>Figure 3.5. Comparison between predicted and actual chemical shifts of 2-piperidinone.....</i>	<i>100</i>
<i>Figure 3.6. Experimental spectra of L-lysine at different collision energies in negative ionization mode.....</i>	<i>101</i>
<i>Figure 3.7. Experimental spectra of L-lysine at different collision energies in positive ionization mode.....</i>	<i>102</i>

## List of Supplementary Figures

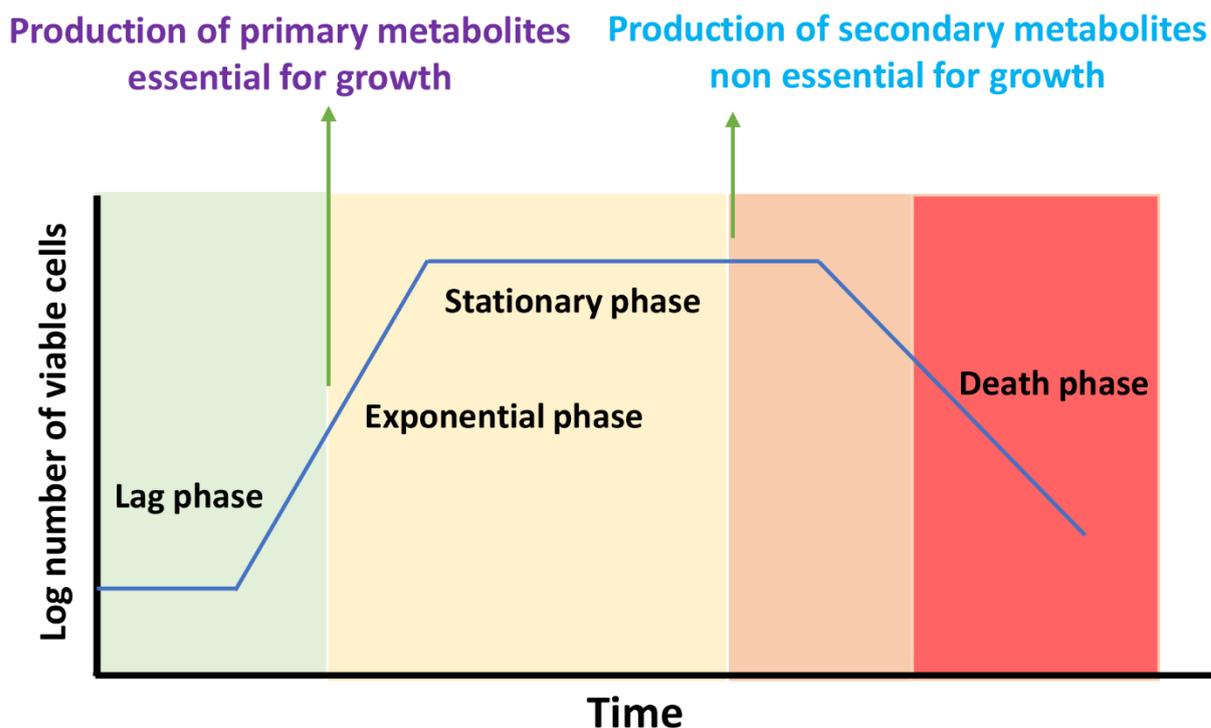
<i>S Figure 2.1. Effect of catechol exposure on glycolysis-gluconeogenesis pathway under aerobic condition</i> .....	148
<i>S Figure 2.3. Effect of catechol exposure on pentose phosphate pathway under aerobic (A) and anaerobic (B) conditions</i> .....	150
<i>S Figure 2.4. Effect of catechol exposure on biotin metabolism under aerobic condition</i> .....	151
<i>S Figure 2.5. Effect of catechol exposure on biotin metabolism under anaerobic condition</i> .....	152
<i>S Figure 2.7. Effect of catechol exposure on glutathione metabolism under anaerobic condition</i> .....	154
<i>S Figure 2.8. Effect of catechol exposure on amino sugar and nucleotide sugar metabolism under aerobic condition</i> .....	155
<i>S Figure 2.10. Effect of catechol exposure on methane metabolism under aerobic condition</i> ...	157
<i>S Figure 2.12. Effect of catechol exposure on one carbon metabolism under aerobic (A) and anaerobic (B) conditions</i> .....	159
<i>S Figure 2.13. Effect of catechol exposure on cysteine and methionine metabolism under aerobic condition</i> .....	160
<i>S Figure 2.14. Effect of catechol exposure on cysteine and methionine metabolism under anaerobic condition</i> .....	161
<i>S Figure 2.15. Comparison of fragmentation pattern of 4-hydroxybenzoate, 4-nitrocatechol, and protocatechuic acid, and 1,2,4-benzenetriol with standards</i> .....	162
<i>S Figure 2.16. Comparison of fragmentation pattern of (E) 3-methylbenzoic acid, (F) vanillic acid, (G) phenol, and (H) hydroquinone with standards</i> .....	163
<i>S Figure 2.17. Comparison of fragmentation pattern of I) 2,4-dihydroxyacetophenone, J) trans-caffeic acid, K) protocatechuic aldehyde, and L) 4-hydroxycoumarin with standards</i> .....	164
<i>S Figure 2.18. Comparison of fragmentation pattern of (M) 7-hydroxycoumarin, (N) dioxydiphenol, and (O) 7-hydroxy-2-oxo-2H-chromene-3-carboxylic acid with standards</i> .....	165
<i>S Figure 2.19. Comparison of fragmentation pattern of (Q) Catechol carbonate, and (R) 2-coumaranone with NIST2.4 database</i> .....	166
<i>S Figure 2.20. Comparison of fragmentation pattern of (S) 4-ethoxyphenol, and (T) 2-methoxy-1,3-benzodioxole with NIST2.4 database</i> .....	167

<i>S Figure 2.21. Comparison of fragmentation pattern of (U) 4-Ethoxybenzoic acid ethyl ester, and (V) Ethyl vanillin with standards of NIST2.4 database .....</i>	<i>168</i>
<i>S Figure 2.22. Comparison of fragmentation pattern of (W) 2-Phenoxyphenol, (X) 2-Methoxyphenol with standards of NIST2.4 database .....</i>	<i>169</i>
<i>S Figure 2.23. Comparison of fragmentation pattern of (Y) 2-Hydroxymethyl-1,4-benzodioxan, and (Z) O-valeryl-1,2-benzenediol with standards of NIST2.4 database .....</i>	<i>170</i>
<i>S Figure 2.24. Comparison of fragmentation pattern of (A1) Methyl 1,3-benzodioxole-2-carboxylate, and (B1) 2-propoxyphenol of NIST2.4 database .....</i>	<i>171</i>
<i>S Figure 2.25. Determination of catechol carbonate in the reaction between catechol and carbonic acid.....</i>	<i>172</i>
<i>S Figure 2.26. NMR analysis detected unknown compounds in catechol treated extracellular samples under anaerobic condition.....</i>	<i>173</i>
<i>S Figure 2.27. NMR analysis detected unknown compounds in catechol treated intracellular samples under anaerobic condition.....</i>	<i>174</i>

# Chapter 1: Introduction

## 1.1. Bacterial Metabolites: Characteristics features

Metabolites are small molecules (typically with a MW < 1500 Da) such as amino acids, organic acids, nucleic acids, sugars, lipids, and vitamins that are synthesized or broken down by all living cells (Muthubharathi et al., 2021; Wishart, 2019). They are the end products or the intermediates of biochemical reactions, catalyzed by enzymes that naturally occur within cells. Within the cell, metabolites perform a multitude of key functions, including energy generation, biosynthesis, signaling, activation, inhibition, and modification of macromolecules (Zhang et al., 2012). Metabolites can be classified into two categories: primary and secondary (Horak et al., 2019). Primary metabolites are the molecules needed for life (Seyedsayamdost, 2019). They include molecules such as lipids, amino acids, nucleic acids, short peptides, sugars, alcohols, and organic acids. Primary metabolites are commonly produced by endogenous catabolism or anabolism and have a direct role in an organism's growth, development, reproduction, and other physiological processes (Wishart, 2019). Secondary metabolites are molecules such as steroids, polyphenols, antibiotics, and pigments; that are not required for growth and development, but facilitate an organism's interaction with its environment (Brader et al., 2014). Plants, animals, and even bacteria produce both primary and secondary metabolites. Bacteria typically produce primary metabolites in the log (exponential) phase, whereas secondary metabolites predominate in a growth phase known as the late idiophase (Figure 1.1) (Horak et al., 2019).



*Figure 1.1. Graphical illustration of primary and secondary metabolites production in different growth phases in bacteria.*

Primary metabolites are conserved throughout all bacterial phyla and kingdoms while secondary metabolites are poorly conserved and can differ even among bacterial species (Karlovsky, 2008; Vaidyanathan, 2005). Because of their high turnover rates, primary metabolites are normally present at lower concentrations within cells. Indeed, the levels of primary metabolites within bacteria can be lower than the levels of secondary metabolites. The number of primary metabolites in most bacterial species typically numbers less than 3000 molecules, including lipid species (Sajed et al., 2016). It is estimated that bacteria (collectively) produce about ~50000 secondary metabolites (Thirumurugan et al., 2018; van Santen et al., 2019). On the other hand, plants are thought to (collectively) produce about 600,000 secondary metabolites (Thirumurugan et al., 2018).

### **1.1.1. Secondary metabolites: a source of bioactive compounds**

Secondary metabolites have traditionally served as the structural and functional scaffolds for many important drugs (Ntie-Kang et al., 2021). Following the discovery of penicillin (a microbially derived secondary metabolite) in 1929, a sustained period of secondary metabolite production commenced which sought to detection of bioactive secondary metabolites in plants, animals, and microbes. This work led to the development of thousands

of new drugs and drug-like compounds. Examples of plant and microbially-derived secondary metabolites used in the drug industry include antibacterial agents (Sulphonamides, Macrolides, Quinolones and Flouroquinolones), anticancer agents (azurin, arnesyltransferases inhibitors, prodiginines and epothilones), immunosuppressive agents (Mitomycins, Bleomycins, and Actinomycins ), anthelmintics agents (Milbemycin D, and Destomycin A), and antiparasitic agents (Ivermectin, Paromomycin and Hygromycin B) (Shuikan et al., 2021). By the 1990s, approximately 80% of commercial medications were natural products or their derivatives (Harvey, 2008). Because of their widespread use in drug or drug intermediate generation, as well as their roles in plastic production, herbicide and pesticide production, and in food processing, plant and microbially-derived secondary metabolites have had a profound and positive effect on the global economy (Fouillaud & Dufossé, 2022). Generating large quantities of secondary metabolites through industrial-scale fermentation or bioreactor production is an industry that is now worth more than 300 billion dollars each year ( Barbuto Ferraiuolo et al., 2021; Meyer & Schmidhalter, 2014). Industrial production of microbial and plant secondary metabolites requires a comprehensive and in-depth knowledge of the metabolite, the cell system (microbe or plant cell), and the cell's metabolism. This is because multiple mechanisms can regulate cellular metabolic processes that result in the formation of secondary metabolites (Brakhage, 2013). Many secondary metabolic pathways depend on multienzyme complexes (Vining, 2007) and the proper redirection of primary metabolic intermediates or end-products towards secondary metabolite synthesis (Bruce, 2022). Indeed, the activation of secondary metabolic pathways is generally observed only under particular cell growth conditions (Baral et al., 2018). Determining the optimal growth conditions for secondary metabolite production often requires dozens of trial-and-error efforts that examine various growth parameters and medium compositions (Bode et al., 2002; Boruta & Bizukojc, 2016). The main factors that typically affect the bioproduction efficiency of secondary metabolites include pH, level of aeration, temperature, carbon and nitrogen sources, light, and the concentration of nutrients (Frisvad, 2012). It is also known that stress response mechanisms are closely intertwined with the regulation of microbial secondary metabolism (Roze et al., 2011; Yin et al., 2013).

### **1.1.2. Microbial degradation of plant secondary metabolites**

Discovering new bioactive molecules from microbial sources can be accomplished by growing microbes on different media or substrates. Indeed, many drug discovery studies have been conducted that involve growing microbes on media containing plant-derived or microbially-derived secondary metabolites (Aura, 2008; Aura et al., 2002; Blaut et al., 2003;

Cheng et al., 1969; Fleschhut et al., 2006; Flores et al., 2015; Krishnamurty et al., 1970; Krumholz & Bryant, 1986; Meselhy et al., 1997; Řezanka et al., 2004; Schneider & Blaut, 2000; Takagaki & Nanjo, 2013, 2015). The intent of many of these studies is to transform existing, but biologically inactive secondary metabolites, into new, potentially bioactive compounds. These kinds of microbial growth studies have also been conducted to better understand the metabolic fate of plant-derived secondary metabolites in the human gut (Clavel et al., 2005; Kawabata et al., 2013; Lee et al., 2006; Puupponen-Pimiä et al., 2005; Reddy et al., 2007; Tzounis et al., 2008). An example of one such fate-determination study was described by Takagaki et al. (2015) who performed an experiment that involved exposing the gut-derived, equol-producing bacterium *Adlercreutzia equolifaciens* (MT4s-5) to a catechin/epicatechin containing media for 24 hrs under anaerobic conditions. This was done to identify the microbially-derived by-products of these plant-derived polyphenols and their potential association with anti-inflammatory activities. Catechin and epicatechin were found to have similar metabolism profiles, produced 1-(3,4-dihydroxyphenyl)-3-(2,4,6-trihydroxyphenyl)propan-2-ol from catechin-epicatechin degradation. Moreover, 1-(3,4-dihydroxyphenyl)-3-(2,4,6-trihydroxyphenyl)propan-2-ol has antioxidant and antispasmodic potential (Gleńsk et al., 2019). The characterization of these microbial biotransformation products typically requires considerable analytical work involving the use of nuclear magnetic resonance spectroscopy (NMR) and mass spectrometry coupled to liquid chromatography (LC-MS). S1 Table (supplementary section) provides a comprehensive list of many experiments conducted over the past 45 years that examined bioactive compounds produced by bacteria grown on food-derived phytochemicals and secondary metabolites. The table includes the reported experiments, the fermentation products, growth conditions, and the chemical profiling technology used to characterize the bio-transformed molecules. As seen by S1 table, many of the chemical profiling techniques employ a field of omics science called “metabolomics”.

## **1.2. Metabolomics: a tool for bacterial natural product identification**

Metabolomics is an emerging field of systems biology that uses advanced analytical chemistry techniques to identify, characterize, and quantify small non-polymeric molecules (metabolites) in a biological specimen (cell, tissue, organ, biological fluid, or organism) (Idle & Gonzalez, 2007; Krastanov, 2010; Roberts et al., 2012; Wishart, 2019). The complete set of metabolites within a cell, organ, or organism is called the metabolome (Oldiges et al., 2007). Metabolites represent the downstream products of processes involved in gene expression, protein translation, and gene-environment interactions. As such the metabolome is considered

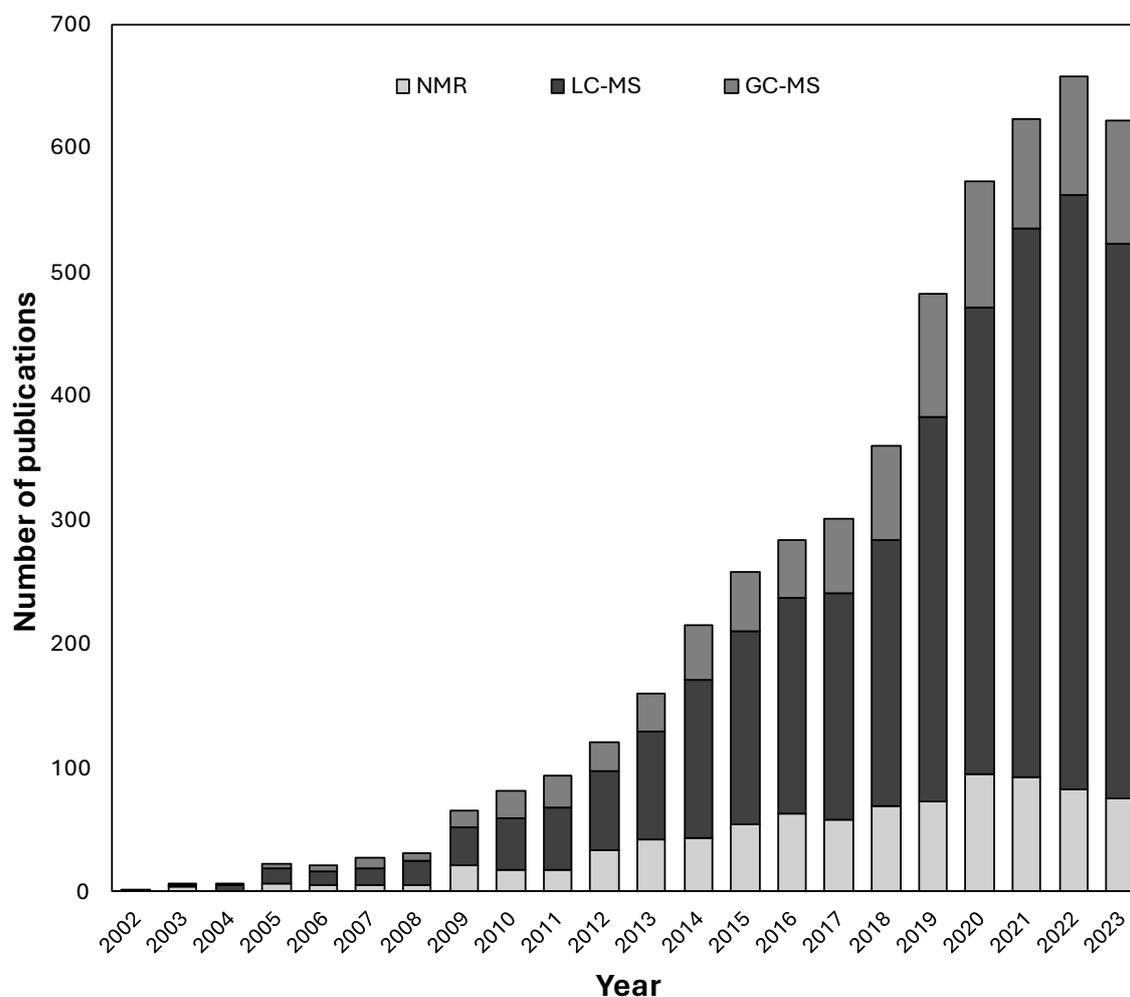
as a very useful probe of an organism's phenotype (Wishart, 2019). Generally speaking, the genome and proteome indicate what might happen in living organisms, while the metabolome indicates what is actually happening (Wishart, 2016).

Metabolomics can be performed using two distinct approaches: targeted and untargeted (Turi et al., 2018). Targeted metabolomics involves the characterization of a predefined set of known metabolites (Roberts et al., 2012; Zhao et al., 2022), while untargeted metabolomics attempts to characterize as many known and unknown metabolites as possible (Vinayavekhin & Saghatelian, 2010). In targeted metabolomics, metabolites can be accurately quantified using internal standards; while in untargeted metabolomics, metabolites are only semi-quantified. Targeted metabolomics is generally performed for hypothesis testing while untargeted metabolomics is used for hypothesis generation (Nalbantoglu, 2019). Both targeted and untargeted metabolomics can lead to the identification of >1000 metabolites (Lelli et al., 2021). The analytical techniques used for both targeted and untargeted metabolomics include nuclear magnetic resonance (NMR) spectroscopy, liquid chromatography-mass spectrometry (LC-MS), and gas chromatography-mass spectrometry (GC-MS). A more detailed discussion of the analytical methods used in metabolomics is given in the next section.

Metabolomics can be applied to a wide range of samples including cells, tissues, organs, biofluids, cell growth media, soil, and water. When metabolomics is applied to cells or cell cultures several common terms are used depending on what part of the cell culture systems is studied. Endo-metabolome profiling or metabolic fingerprinting refers to the characterization of metabolites found within cells. Exo-metabolome profiling or metabolic footprinting refers to the characterization of metabolites in the cell media or the fluids outside cells (Hoerr et al., 2012).

### **1.2.1. Common analytical techniques in microbial metabolomics**

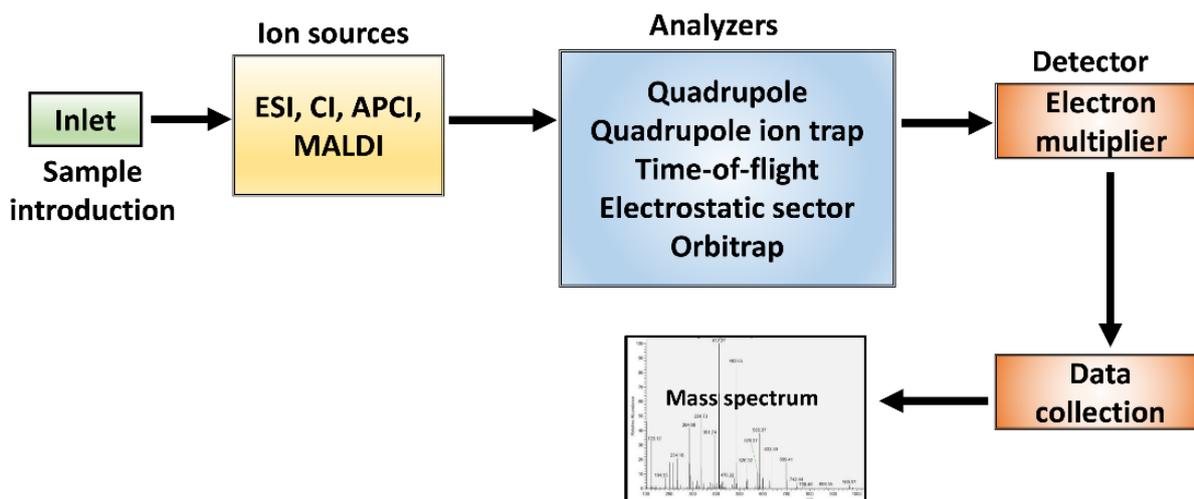
Figure 1.2 provides details about the microbial metabolomics methods and the analytical methods used, based on publications appearing in PubMed after 2000. As seen from this figure, the most popular methods for microbial metabolomics studies involve liquid chromatography-mass spectrometry (LC-MS), gas chromatography-mass spectrometry (GC-MS), and NMR spectroscopy. Each of these methods has its advantages and disadvantages. Because they are so important to the work described in this thesis, I will describe these methods in more detail.



**Figure 1.2.** A steady increase in the most commonly utilized analytical techniques for microbial metabolite determination in PubMed Since 2000. The numbers were obtained from the PubMed search using the keywords “mass spectrometry microbial / bacterial metabolomics”, “NMR microbial / bacterial metabolomics” and “GC-MS microbial / bacterial metabolomics”.

### 1.2.1.1. Mass spectrometry

Mass spectrometry (MS) is the most widely used analytical technique in metabolomics. Simply stated, MS measures the mass-to-charge ratios ( $m/z$ ) of ions and uses this information to determine the chemical composition or the chemical structure of molecules. For mass spectrometry to work, molecules or molecular fragments must be ionized (i.e., charged). Once charged, these ions can be selected, directed, and sent through a series of magnetic or electric filters that allow the ions to be detected and the  $m/z$  values of those ions to be determined (using simple physics equations).



**Figure 1.3.** Basic principles of a typical mass spectrometry. ESI, electrospray ionization; CI, chemical ionization; APCI, atmospheric-pressure chemical ionization; MALDI, matrix-assisted laser desorption/ionization.

A mass spectrometer consists of three components: an ionization source, a mass analyzer, and an ion detection system (Figure 1.3). The ionization source is both the place where ions are generated from neutral molecules and the point of entry for charged ions into the mass spectrometer. Once the ions have been generated, they are passed into the mass analyzer. The analyzer (which is typically under a high vacuum) accelerates the ions through a series of electrically charged plates or a series of magnets. These electric or magnetic fields deflect the ions or change their speed according to their charge and masses. Once the ions have passed through the analyzer, they are detected by a detector that may include electron multipliers (EM), Faraday cups (FC), photomultiplier conversion dynodes, or array detectors. The detectors allow the MS instrument to measure the number of ions and the amount of ion deflection or the period of the ion transit time. The amount of deflection or the time it takes for an ion to transit through the analyzer depends on the ratio of the mass of the ion and its charge ( $m/z$  ratio). Lighter ions with lower masses or ions with lower (single) charges will either be deflected the most or have the shortest transit times. Heavier ions with larger masses or those ions that have more than one electron will be deflected by the least or have the longest transit times. This information is used to determine the mass-to-charge ( $m/z$ ) ratio of the ions.

#### 1.2.1.1.1. MS ionization methods

Several ionization techniques are used in MS, including electrospray ionization (ESI), atmospheric-pressure chemical ionization (APCI), and matrix-assisted laser desorption/ionization (MALDI). In ESI, ions are generated by pushing a liquid with dissolved

(neutral) molecules through a tiny nozzle that is placed under a strong electric field. This generates a spray or aerosol that ionizes the neutral molecules (Ho et al., 2003). ESI is a liquid-based ionization technique that is best applied to compounds of high to moderate polarity. ESI is a soft ionization technique which results in less fragmentation in the mass spectrum. In contrast to ESI, chemical ionization is a gas-phase ionization process that uses a reagent gas to ionize sample molecules through ion–molecule reactions in the gas phase. Atmospheric pressure chemical ionization (APCI) is a chemical ionization method used in mass spectrometry which utilizes gas-phase ion-molecule reactions at atmospheric pressure. APCI is a soft ionization method that can be applied to liquid samples by coupling the gas inlet used for gas-phase chemical ionization to a liquid chromatography system (Dass, 2007; Rockwood et al., 2018). APCI performs well with compounds with low to medium polarity and with moderate molecular weight (up to ~1.5kDa). As an ionization technique, MALDI is somewhat different from ESI or APCI. This is because MALDI is best done with solid samples. In MALDI, ions are generated using pulsed laser beams that heat the sample. This heating leads to an explosive ionization (desorption) event. This simultaneous desorption and ionization process is achieved by incorporating the analyte of interest into a solid matrix of ultraviolet-sensitive crystals (Kaufmann, 1995).

#### **1.2.1.1.2. Mass analyzers for MS**

There are several types of mass analyzers used for MS. These are named based on how the ions are manipulated or the mechanisms used to select and/or accelerate ions. The most common mass analyzers include quadrupole systems, quadrupole ion trap systems, ion cyclotron resonance systems, time of flight systems, and Orbitrap mass analyzer systems. Different mass analyzers have specific advantages and disadvantages. Quadrupole mass analyzers consist of four parallel cylindrical rods with two bearing positive and two bearing negative charges. The charged ions enter the quadrupole and are accelerated by the charged DC (direct current) bias placed on the quadrupole rods. This allows the ions to pass through the center line of the analyzer (Thomas, 2019). The quadrupole ion trap (QIT) mass analyzer is a three-dimensional, dynamic ion storage device that consists of three electrodes: two end caps and a ring electrode. Ions enter the QIT analyzer through the end cap electrode, where they are stored (or trapped) and then ejected in a mass-selective manner (March, 2009). Similar to the QIT, ion cyclotron resonance (ICR) mass analyzers trap ions into a cyclic orbit inside the analyzer using a powerful magnetic field. The ions are then ejected in a mass-selective manner when an external electric field is applied (Rockwood et al., 2018). ICR (specifically

Fourier Transform ICR) mass analyzers provide the highest mass resolving power and mass accuracy of any mass analyzer, with up to parts-per-billion (ppb) mass accuracy, high dynamic range, and mass resolving power values greater than 10,000,000 (ten million) in routine analyses (Bowman et al., 2020). The time-of-flight (TOF) mass analyzer utilizes an electric field to accelerate ions through a long tube. This acceleration is done using an electrical potential, and then the time for each ion to reach the detector is measured to determine the  $m/z$  ratio of each ion (Mamyrin, 2001). TOF is the fastest MS analyzer and is well-optimized for pulsed ionization methods. TOF mass analyzers can achieve a mass accuracy of 10 ppm or better and a resolving power of up to 60,000. The Orbitrap mass analyzer, which is conceptually similar to QIT and ICR mass analyzers, uses a central rod-like electrode surrounded by a barrel-shaped outer electrode to trap and select ions (Scigelova & Makarov, 2009). Ions are electrostatically trapped in an orbital motion around the Orbitrap spindle. Orbitrap mass analyzers provide very high mass accuracy (<1-2 ppm) and high resolving power (up to 240,000 at  $m/z$  400).

#### **1.2.1.1.3. Tandem Mass Spectrometry**

Tandem mass spectrometry is a mass fragmentation technique that is also known as MS/MS or MS<sup>2</sup> (Mittal, 2015). In tandem mass spectrometry, two or more mass analyzers are coupled together. The first mass analyzer is used to select a given ion (a precursor ion), then direct that selected ion into a collision cell and measure the fragment ions arising from that collision (product ions) (Smith, 2013). MS/MS allows more information to be obtained about the structure of a molecule by characterizing the ions that arise from the molecule breaking up. Characteristic ion fragments with specific  $m/z$  values allow skilled MS chemists to determine which molecular components were part of the original (parent) molecule. MS/MS is commonly used to confirm the identification of unknown compounds in metabolomics mixtures. The most widely used analyzer in MS/MS is the triple quadrupole, triple quad, or QqQ analyzer. In a QqQ analyzer, the first (Q1) and third (Q3) quadrupoles are used as mass filters, while the second quadrupole (Q2) serves as a collision cell. Ions selected in Q1 are moved to Q2, where they produce fragments after colliding with neutral gas. Fragments produced in Q2 are then analyzed in Q3. Other kinds of MS/MS configurations exist, including the QTOF (quadrupole time of flight) system, which consists of a quadrupole MS analyzer to a TOF MS analyzer and the Orbitrap system. These kinds of MS/MS systems offer much higher resolution MS spectra than the QqQ system.

#### 1.2.1.1.4. Chromatography Coupled to MS

Mass spectrometry alone is often not enough to fully identify a compound – especially if it is being characterized from a complex mixture. To help simplify the process, chromatography techniques are often coupled to mass spectrometers to separate the mixtures and reduce the complexity of the resulting MS spectra (Lei et al., 2011). Mass spectrometers can be coupled to either liquid chromatography systems (LC-MS) or gas chromatography systems (GC-MS).

Gas chromatography (GC) is a chromatographic technique that uses an inert gas (such as Argon or Helium) as the carrier and vaporized mixtures that are carried by the gas down a long, narrow column to separate and detect the chemical components within the mixture.

The gas carrier is called the mobile phase while the column (which has an interior covered with inert organic molecules) is called the stationary phase. Gas chromatography is also known as vapor-phase chromatography (VPC), or gas–liquid partition chromatography (GLPC). The glass or metal column through which the mobile (gas) phase passes is located in an oven where the temperature of the gas can be controlled and the eluent coming off the column is monitored by a computerized mass or flame ionization detector. The chemical components within any mixture that is being analyzed by GC usually must be volatile organic molecules or gasses. For GC to be successful the chemicals in the mixture must have a molecular weight below 1250 Da and be thermally stable so they don't degrade in the heated GC system. To make non-volatile compounds more volatile, it is possible to chemically derivatize them with organo-silicon compounds, which lowers their boiling point. In GC-MS, samples are first introduced into the GC column manually or by an autosampler. Analytes are separated according to their differences in partitioning between the mobile phase and the stationary phase. After separation, the neutral molecules enter into the vacuum system of the mass spectrometer where they are ionized (usually via electron impact ionization) and sent to a mass analyzer for detection. GC (and GC-MS) is ideal for separating and identifying hydrophobic, uncharged, lower molecular weight molecules such as terpenes, alcohols, aromatic compounds, hydrocarbons, pesticides, and herbicides.

Unlike the situation with gas chromatography, the mobile phase in liquid chromatography is a liquid. Simply stated, liquid chromatography (LC) is a separation technique that separates mixtures of molecules dissolved in a liquid (the mobile phase) based on their interaction with a solid, high surface-area substrate (often coated beads or gels) placed

in a short column (called the stationary phase). Early LC methods allowed the mobile phase to flow through the stationary phase via gravity under atmospheric pressure. However, today, most liquid chromatography systems use very small packing particles and a relatively high pressure to push the mobile phase through the column. This high-pressure form of chromatography is referred to as high-performance liquid chromatography (HPLC). HPLC is both faster and provides greater chromatographic resolution than gravity/atmospheric pressure LC methods.

HPLC columns are typically packed with a stationary phase composed of irregularly or spherically shaped particles, or a porous monolithic layer. These monolithic layers or monoliths are ‘sponge-like chromatographic media’ and are made up of an sequential block of organic or inorganic parts. HPLC can be divided into three different subclasses based on the polarity of the mobile and stationary phases. Methods in which the stationary phase is more polar than the mobile phase are termed normal phase liquid chromatography (NPLC) and methods where the mobile phase is more polar than the stationary phase are termed reversed phase liquid chromatography (RPLC). The third form of HPLC is called HILIC, which is short for hydrophilic interaction chromatography. In HILIC the mobile phase and stationary phase are both relatively polar.

In LC-MS, samples are first introduced into the LC (or HPLC) column manually or by an autosampler. Analytes are then separated according to their differences in partitioning between the mobile phase and the stationary phase. After separation, the un-ionized molecules enter the vacuum system of the mass spectrometer where they are ionized (usually via electrospray ionization or ESI) and sent to a mass analyzer for detection. LC and HPLC are ideal for separating and identifying higher molecular weight molecules that are non-volatile or which are not thermally stable. In particular, HPLC-MS (often called LC-MS) is widely used to separate and characterize lipids, amino acids, fatty acids, nucleosides, nucleotides, and steroids.

#### **1.2.1.1.5. Analysis of mass spectrometry data**

MS-based metabolomics studies generate huge amounts of data and consequently, these data sets must be analyzed by computers to help with compound identification and quantification. In MS-based metabolomics, compound identification and quantification are sometimes called “peak annotation” or “metabolite annotation”.

For targeted MS-based metabolomics the software used for peak annotation is very platform- or vendor-specific. In general, these targeted MS-based metabolomics packages must be able to identify MS (or MS/MS) peaks from the raw MS data and perform accurate peak integration. After the peaks have been identified and integrated, the software will typically use lists of pre-determined, compound-specific LC retention times as well as compound-specific MRM (multiple reaction monitoring) transitions to identify specific compounds. Calibration curves and integrated peak areas determined from isotopically labeled standards are used to determine the actual concentrations of the compounds. Because targeted MS-based metabolomics always uses a predefined set of compounds, the identification process can be highly automated. In other words, targeted MS data can be analyzed very rapidly and very consistently. Targeted MS-based metabolomics usually allows users to generate lists of metabolites and concentrations for each sample in their study. These lists can then be used in multivariate statistical analysis packages to determine which metabolites have changed significantly between two (or more) groups of cohorts or which have changed significantly relative to physiologically normal values.

In contrast to targeted MS data analysis, untargeted MS data analysis is much more difficult, much less automated, and consequently much slower. In untargeted MS metabolomics, many of the thousands of features are detected but not all features are biologically interesting or even technically relevant. This is because they may represent background signals from sample processing or there may be multiple signals arising from the same analyte (adducts, isotopes, in-source fragmentation). In general, untargeted MS-based data analysis requires the collection and analysis of many more samples than targeted MS-based studies. This is because untargeted metabolomics always requires a “reference” cohort (usually healthy controls) from which to perform relative comparisons to the cohort of interest (a treated or diseased cohort). This is because untargeted metabolomics is not able to accurately quantify metabolites, which is one of its central limitations.

The standard data analysis workflow for an untargeted MS-based metabolomics experiment involves several steps. These include 1) extraction of extracted ion chromatograms (EICs); 2) noise filtering; 3) peak detection; 4) peak deconvolution (consolidating adducts and isotope peaks); 5) peak/retention time alignment; 6) feature filtering; 7) significant feature detection/selection and finally 8) significant feature annotation (i.e. compound identification and intensity measurement). Very lengthy and detailed reviews have been written about the workflows and methods required for untargeted MS-based metabolomics (Alonso et al., 2015;

Di Minno et al., 2021; Rafiei & Sleno, 2015; Schiffman et al., 2019; Vinayavekhin & Saghatelian, 2010). Rather than discussing all aspects of the untargeted MS metabolomics data analysis workflow, I will only focus on a few. Two of the most important steps are peak-picking and peak (or retention time) alignment.

Proper peak identification is critical to the success of annotating peaks and determining peak intensities. These steps are highly dependent on the parameters chosen and the applied algorithms. Several methods are commonly used for peak picking. These include centWave, matchedFilter, and MassQuant. CentWave uses a continuous wavelet transformation model to find peaks and features (Tautenhahn et al., 2008). The matchedFilter algorithm works by using a Gaussian model to extract the peaks from slices of the spectra data that have a defined mass width (for example, 0.1 m/z) (C. A. Smith et al., 2006). This method is more suitable for peak detection in low-resolution MS spectra than for high-resolution MS spectra. MassQuant is another popular algorithm. It uses the Kalman gain approach for isotopic trace detection and avoiding missing centroids. This method is particularly sensitive to low-intensity peaks (Conley et al., 2014). Most metabolomics software packages use the centwave method for peak picking. More recently, other approaches have appeared, including the ADAP method which uses continuous wavelet transform (CWT) and ridgeline detection (Du et al., 2020).

After peak picking has been completed, peak matching and retention time alignment must be performed. This peak alignment process enables the comparison of LC-MS-based metabolomic data across samples. The retention time of an ion may drift across different samples, even if those samples are analytical replicates. As a result, the drift is generally non-uniform across the retention time range and cannot be completely controlled during LC separations. For large-scale studies involving multiple samples, retention time alignment is used to correct the retention time drift and ensure that the same ion is compared across samples. One approach for peak alignment uses the peak detection results and makes efforts to find and match similar peaks. Kernel estimation is the most popular method to group peaks with similar m/z values and retention times across a given dataset (N. Kumar et al., 2021). In Kernel estimation, the algorithm assembles “well-behaved” peak groups to which very few samples have no peak assigned and these are used as landmarks for alignment. The deviations of the retention times of these landmarks from their median values within peak groups are regressed against the retention time. Those regions on a chromatogram without “well-behaved” peaks can be interpolated and aligned. The aligned peaks are grouped again to match peaks with corrected retention times. This procedure is usually carried out iteratively two or three times to

make sure the retention time drift is sufficiently corrected. Other peak alignment methods use LC-MS raw data for retention time alignment by matching the EIC (extracted ion chromatogram) chromatographic signals. Correlation-optimized warping (COW) aligns chromatographic signals by dividing the time axis into segments and performing a linear warping within each segment to maximize the overall correlation of the two chromatographic profiles (Tomasi et al., 2004). Another method called dynamic time warping (DTW) aligns chromatograms by maximizing the spectra similarity while preserving the internal ordering of the eluents (Clifford et al., 2009).

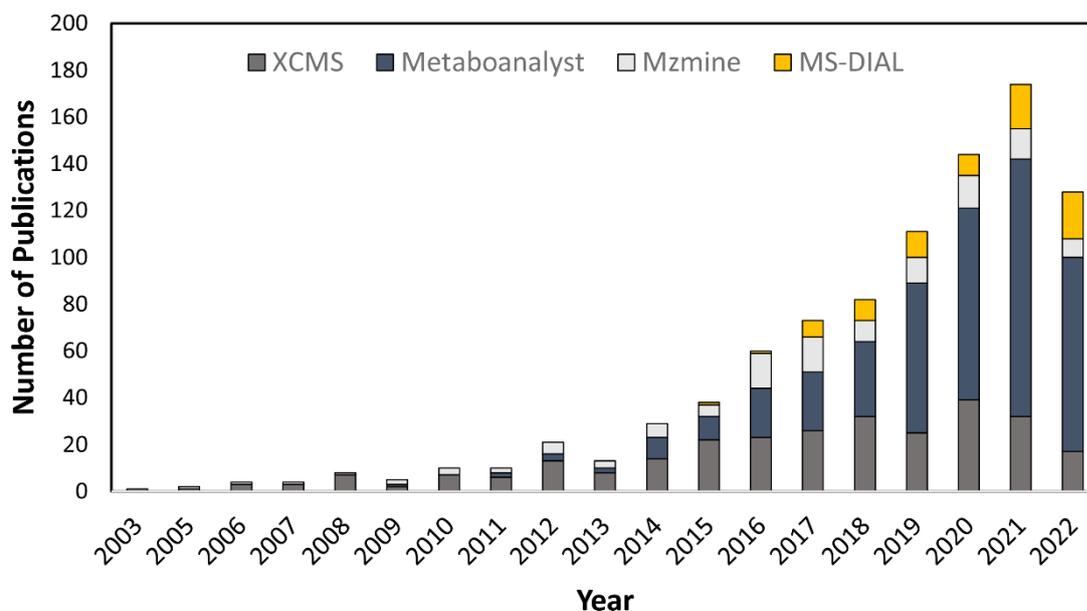
Another important step in the untargeted metabolomics data analysis workflow is peak filtering. As untargeted metabolomics datasets often have high numbers of ambiguous or erroneous features (even after peak alignment), it is often essential to employ additional filtering procedures prior to finalizing the list of peak features used for peak significance assessment. Log transformation, manual identification of high- and low-quality filters, blank subtraction, p-value, and box/density plots can all be used to filter features to help minimize or remove uninformative features. Deisotoping (removal of features corresponding to the same compound but with different isotopes) is another important feature filter that is often applied (McCardle, 2022).

A typical untargeted metabolomics experiment can easily generate 10,000 to 20,000 features. After cleaning, aligning, and filtering, the number of features may be reduced to 5,000-6,000 robust or reproducibly reliable features. However, not all of these are biologically significant, and not all of these may be changed as a result of a given intervention. Rather than attempting to annotate (identify and quantify) all 5,000 features, most researchers apply a variety of multivariate statistical techniques to reduce the number of features even further. These methods, such as principal component analysis (PCA) or partial least squares discriminant analysis (PLS-DA) can be used to identify those  $m/z$  and retention time features that have changed significantly between the two (or more) sets of cohorts. Programs such as MetaboAnalyst (Pang et al., 2022) and Workflow4Metabolomics (Giacomini et al., 2015) are examples of popular software tools that can help with these multivariate analyses. Typically, these multivariate statistical analyses will reduce the number of features from 5,000 to as few as 50-100 highly significant features. Once the most significant features have been selected, then the next step is feature annotation.

Feature annotation or metabolite identification in untargeted metabolomics is mainly achieved through mass-based search followed by manual verification. In most cases, the  $m/z$

value of a molecular ion of interest or the calculated molecular weight (MW) of the molecule is searched against database(s) such as HMDB (Wishart, Guo, et al., 2022), PubChem (Kim et al., 2016), KEGG (Ogata et al., 1999) or Metlin (Guijas et al., 2018). The molecules having molecular weights or ions having  $m/z$  values within a specified tolerance range to the query molecule are retrieved from databases as putative identifications. Unfortunately, mass-based searches rarely provide unique identifications for the ions of interest since there are often multiple compounds with the same  $m/z$  values or molecular weights. Likewise, most metabolite databases do not have complete coverage of all known molecules. Generally, less than 30% of the detected ions in a typical LC-MS-based metabolomic experiment can be uniquely identified through mass-based searching. To verify the mass-based search results, authentic compounds of those putative identifications must then be collected. By comparing the LC retention times and MS/MS spectra of the authentic compounds with the ions of interest in the sample, the identities of the metabolites can be confirmed. Using information about the species being analyzed (to eliminate the possibility of identifying plant-only compounds in bacteria or eliminating the possibility of identifying human-only drugs in rat studies) can also help minimize incorrect identifications.

Over the past decade, a wide variety of software packages have been developed to help perform untargeted metabolomics data analysis. These include SIRIUS (Dührkop et al., 2019), MetaboAnalyst 5.0 (Pang et al., 2022), MetAlign (Lommen, 2009), OpenMS (Rurik et al., 2020), Compound Discoverer (Thermo Fisher), SIEVE (Thermo Fisher), Metaboanalyst (Pang et al., 2022), XCMS online (Huan et al., 2017), MZmine (Du et al., 2020), MS-DIAL (Tsugawa et al., 2015), El-MAVEN (Agrawal et al., 2019), Progenesis QI (Waters), and MetaboScape (Bruker). Compound Discoverer, SIEVE, Progenesis QI, and MetaboScape are commercial software packages, while the others are freely available web-based or downloadable software packages. Some of these packages (such as SIRIUS) are limited to the peak annotation steps while others are more comprehensive and can perform all or nearly all the eight data processing steps needed for untargeted metabolomics. Each software program has its own approaches that offer different degrees of accuracy, and each has its own distinct advantages and disadvantages in terms of data pre-processing, analysis, visualization, and interpretation. Among all of these software packages MetaboAnalyst, XCMS online, MZmine, and MS-DIAL are perhaps the most well-known and widely used. A graph indicating the frequency of use by Metaboanalyst, XCMS online, MZmine, and MS-DIAL as indicated by PubMed citations is shown in Figure 1.4.



**Figure 1.4.** The use of XCMS, Metaboanalyst, MZmine2 and MS-DIAL has increased dramatically in the past 20 years. (A) Annual number of publications available on PubMed that contain the keywords “XCMS,” “Metaboanalyst,” “MZmine,” and “MS-DIAL”. These numbers were obtained directly from PubMed’s “Results by Year” section.

#### 1.2.1.1.6. Challenges in untargeted mass spectrometry data analysis: false positive/false negative results

Large metabolomics datasets can contain thousands of falsely identified features. These include noise signals from sample processing or multiple signals originating from the same analyte (e.g., adducts, isotopes, in-source fragmentation) or features with imperfect integration (e.g., incorrect integration regions and missing values). As a result, false positive and false negative feature identifications frequently plague untargeted metabolomics studies (Schiffman et al., 2019). The goal of many untargeted metabolomics software packages is to reduce the number of false positives and false negatives. Several studies were conducted on the performance of different software. Cobble & Fraga et. al (2014) evaluated the performance of MetAlign, XCMS, and MZmine2, and reported that significant improvements were needed in the preprocessing tools to reduce the high percentage of false peaks identified. Another study involving a comparison between PeakView, Markerview, and MetabolitePilot (three commercial software products from the MS vendor Sciex) against the freeware package XCMS Online was conducted (Rafiei & Sleno, 2015). These authors noticed significant variations in each software’s peak lists. Additionally, they reported that all four software packages failed to detect several standard metabolites. In 2017, Mayer’s et al. conducted a comparative study for chromatogram construction and chromatographic peak detection performance between XCMS

and MZmine2. They observed that both programs performed well for peak picking, however, both packages reported a significant number of false positive EIC peaks, and both failed to detect real EIC peaks as well. In addition, during the peak annotation process (conducted against the same database) selection of related isotopes and adducts for neutral masses sometimes led to false positive results. Additionally, other authors have noted that noise peaks can sometimes be incorrectly detected as a peak group while using XCMS for feature detection and integration (Schiffman et al., 2019), leading to incorrect peak area integration. As a result, untargeted metabolomics datasets may include thousands of incorrectly identified features or features with poor integration values (such as inaccurate integration regions and missing values).

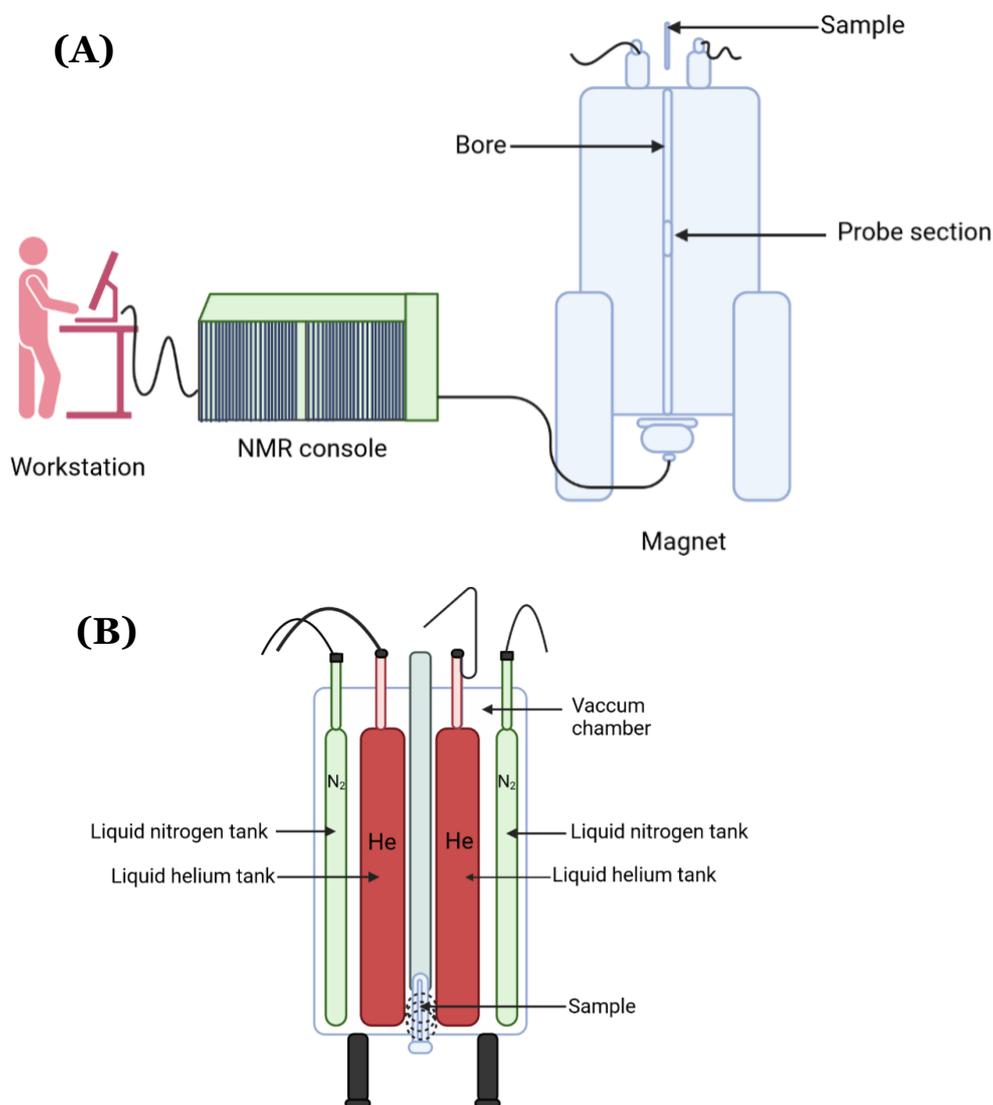
### **1.2.1.2. Nuclear magnetic resonance (NMR) spectroscopy**

NMR is a very different spectroscopic approach than mass spectrometry. Rather than measuring masses to determine molecular formulas or molecular structures, NMR measures the electromagnetic radiation absorbed by molecules in the radio frequency range to determine the structure of organic compounds (Mlynárik, 2017). Unlike MS, NMR provides much more detailed structural information about molecules, and it can be very accurate in terms of compound quantification (coefficients of variation as low as 1-2% compared to 20% for MS). Arguably, NMR was the first metabolomics platform with the vast majority of early metabolomic studies being conducted by NMR (Hoult et al., 1974; Shulman et al., 1979). However, over the past 15 years, the preferred platforms for metabolomic analysis have mostly shifted away from NMR in favor of GC-MS and LC-MS. This is because GC-MS and LC-MS are 10-100 times more sensitive than NMR (5 nM vs. 5  $\mu$ M) and as a result, it is often possible to identify 4-5 times more compounds in a metabolomics sample via GC-MS or LC-MS. Despite the disadvantage of poor sensitivity, NMR still offers some important advantages. In particular, NMR is non-destructive, it is highly reproducible, it allows identification of novel compounds, it doesn't need chemical derivatization, and it requires little to no separation (Emwas et al., 2019). NMR is also very effective in detecting certain classes of compounds, such as sugars, alcohols, amines, and low molecular weight volatile liquids which are not easily detected by GC-MS or LC-MS.

#### **1.2.1.2.1. NMR Basics**

Atomic nuclei are composed of protons and neutrons. They have an intrinsic property called spin, which is expressed by the quantum number  $I$ . Three types of spin are usually seen in nuclei, including integral spins ( $I = 1, 2, 3 \dots$ ), fractional spins ( $I = 1/2, 3/2, 5/2 \dots$ ), and no

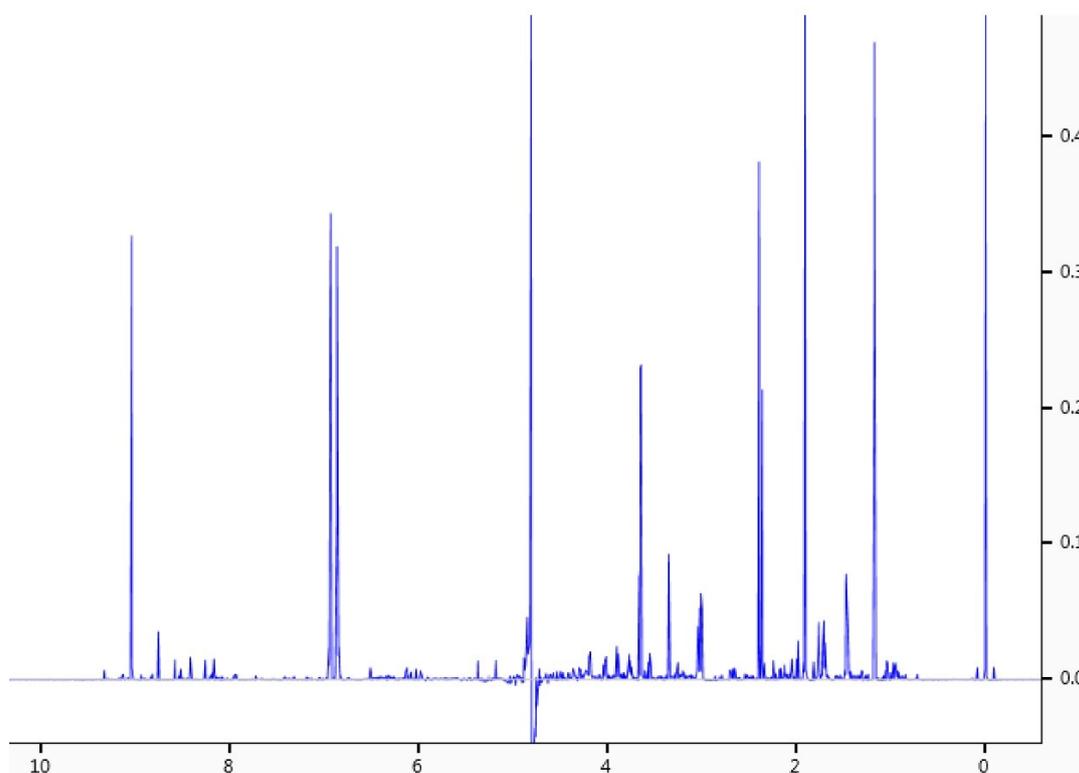
spin, where  $I = 0$ . Nuclei with spin  $I = 0$  are not detectable by NMR spectroscopy. Different isotopes can have different spins depending on the proportion of protons and neutrons. Isotopes of particular interest in NMR spectroscopy are those that have an  $I = \frac{1}{2}$  such as  $^1\text{H}$ ,  $^{13}\text{C}$ ,  $^{19}\text{F}$ , and  $^{31}\text{P}$ . NMR involves the measurement of the absorption of electromagnetic (EM) radiation by certain nuclei at specific frequencies or specific wavelengths. This is because NMR active nuclei respond to (i.e. absorb energy from) the oscillating magnetic fields in EM radiation. NMR active nuclei (those with non-zero spins) behave like tiny bar magnets. These properties allow nuclei not only to respond to EM radiation but also to align with external magnetic fields. When an external magnetic field is applied to a sample, a small fraction ( $<0.01\%$ ) of nuclei will align with the external magnetic field. This alignment “conditions” the sample so that those nuclei can absorb the EM radiation. In particular, those nuclei that are aligned with the external magnetic field, are in a low energy state that makes them susceptible for EM absorption at certain frequencies (called resonant frequencies). When broad-spectrum radio frequency EM radiation is passed through the sample, the nuclei in the low energy state will absorb the EM radiation at a characteristic frequency. The amount of energy absorbed, or the frequency absorbed depends on the type of nucleus ( $^1\text{H}$  nuclei absorb at higher frequencies than  $^{13}\text{C}$ ) and the local electric environment defined by local atomic bonds and local molecular geometry.



**Figure 1.4.** A typical NMR instrument, components, and inner part. A) Design of an NMR spectrometer and its components. B) Inner part of an NMR.

An NMR spectrometer consists of three basic components: a magnet, a probe, and a console (Figure 1.5). The magnet must be very powerful (10-15 Tesla or 100,000X more powerful than a kitchen magnet) so that it can align a sufficient number of nuclear spins in the sample. Most modern NMR spectrometers use superconducting magnets, which require very low temperatures to work. Therefore, an extensive “passive” cooling system must be used which resembles a gigantic thermos bottle. In particular, the magnet (which is shaped like a cylinder with a hole through it) is surrounded by an inner jacket filled with liquid helium (at -270 °C) which is surrounded by an additional liquid nitrogen tank (at -196 °C) to keep the magnet in its superconducting state (Figure 1.5B). A cylindrical tube, called a probe, goes into the center of the hole of the doughnut-like superconducting magnet. The NMR probe is

designed to contain the NMR sample and hold the sample in the “sweet spot” in the center of the magnet where the field is maximized. NMR probes also contain a collection of saddle-shaped coils which performs multiple functions, including radio frequency (RF) EM generation and RF reception (the coils act as RF transceivers). In other words, the coil excites the nuclear spins by irradiating the sample with RF EM radiation. The absorption of this EM radiation is also detected or received by the coils whereupon the signal is passed on to the console. The console is a collection of computers and radio frequency generators/receivers that is responsible for generating the RF radiation, recording the RF absorption/NMR signals, processing the NMR signals, and controlling the magnetic field. The console controls all the experimental conditions and communicates with the NMR probe and the NMR magnet. Most modern NMR spectrometers measure the RF absorption signals as a time-dependent change in radio frequencies. These time-dependent changes are then transformed by a mathematical transformation (called a Fourier Transformation) to produce a frequency-dependent signal that indicates which frequencies were absorbed and how strongly they were absorbed. The resulting (one-dimensional) NMR spectrum resembles an HPLC or GC chromatogram with multiple peaks at different frequencies (Figure 1.6). These NMR spectra reveal information about a molecule’s chemical shifts and their coupling constants, both of which are valuable for determining a molecule’s atomic structure.



**Figure 1.5.** A typical 700 MHz  $^1\text{H}$  NMR spectrum of a natural mixture.

#### 1.2.1.2.2. Chemical shifts

The position of any given peak seen in an NMR spectrum (Figure 1.5) is referred to as the peak's chemical shift. The chemical shift is defined as “the measured nuclear magnetic resonance frequency of a nucleus arising from its local (molecular) electronic environment”. The chemical shift is affected by the type of nucleus or atom being measured (the gyromagnetic ratio of that nucleus), the type of bonding with which the atom is involved (single, double, triple), the proximity of electropositive (Si) or electronegative (O, Cl or F) atoms or atomic groups to the atom of interest, the geometry or shape of the molecule (linear vs. aromatic) and the solvent with which the compound is dissolved. A single molecule may have as few as one (such as formic acid) or as many as dozens of peaks (such as glucose) corresponding to the number of  $^1\text{H}$  or  $^{13}\text{C}$  atoms in the molecule. Each of those peaks is a characteristic chemical shift of that molecule. The complete set of chemical shifts defines a chemical shift “signature” for that molecule. Most chemical shift signatures are unique to a molecule, which enables their unambiguous identification through comparing the observed chemical shifts with known chemical shift tables (for different molecules). In NMR spectroscopy, a molecule is considered “assigned” if all of the  $^1\text{H}$  or  $^{13}\text{C}$  shifts for each of the atoms in that molecule are known or fully determined. Formally or numerically, the chemical shift is defined as the difference between the resonant frequency of an atomic nucleus and the signal of the reference or standard molecule (Mlynárik, 2017). In this regard, chemical shifts are relative measures and so the chemical shift can vary greatly depending on the standard molecule that is chosen. The standard molecules most often chosen for both  $^1\text{H}$  and  $^{13}\text{C}$  NMR are TMS (tetramethylsilane) or DSS (sodium trimethylsilyl propanesulfonate). The chemical shift  $\delta$  is usually expressed in parts per million (ppm) by frequency because it is calculated using the following equation:  $\delta = [v(\text{sample}) - v(\text{ref})] / v(\text{ref})$ , where  $v(\text{sample})$  is the absolute resonance frequency of the sample and  $v(\text{ref})$  is the absolute resonance frequency of a standard reference compound. The chemical shifts in  $^1\text{H}$  NMR spectra typically range from 0 ppm to 14 ppm while the chemical shifts in  $^{13}\text{C}$  NMR spectra typically range from 0 ppm to 220 ppm.

#### 1.2.1.2.3. Coupling Constants

$^1\text{H}$  NMR spectra (Figure 1.6) are also characterized by clusters of peaks (often called multiplets) that have distinct intensity patterns. These clusters of peaks arise from a phenomenon known as J-coupling. J-coupling arises from an interaction between nuclei containing spin. J-couplings are also known as scalar couplings. This interaction is mediated through bonds. The coupling constant, J (usually measured in Hz) is a measure of the

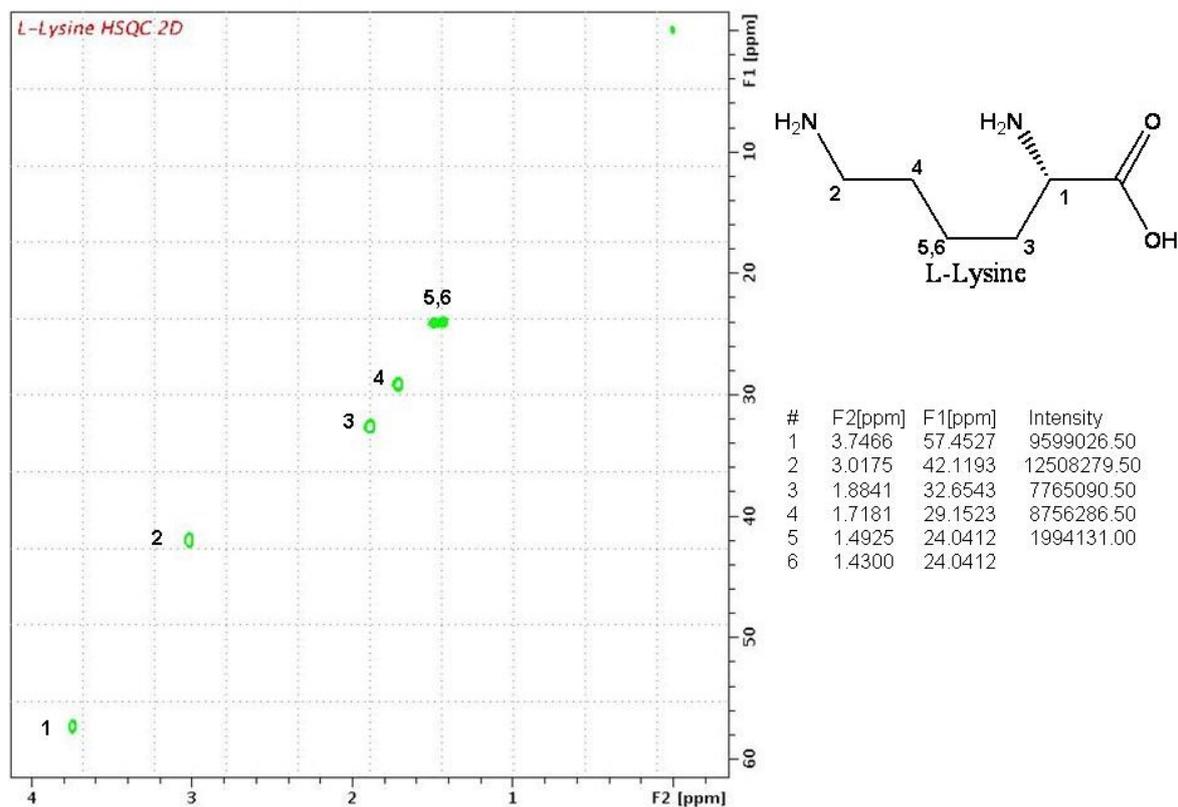
interaction between a pair of chemically bonded spins ( $^1\text{H}$ ,  $^{13}\text{C}$  or  $^{15}\text{N}$ ). Coupling is controlled by geometry and the orbitals involved between the coupling nuclei. In  $^1\text{H}$  NMR, clusters of coupled peaks come in different varieties, including pairs of peaks (doublets), groups of three peaks (triplets) and groups of four peaks (quartets). In  $^1\text{H}$  NMR, doublets have equal intensity or intensity ratios of 1:1, while triplets of intensity ratios of 1:2:1, and quartets have intensity ratios of 1:3:3:1. The relative intensities of peaks in a coupling pattern are given by a binomial expansion. J-coupling provides information about relative bond distances and angles as well as information on the connectivity of chemical bonds.

#### **1.2.1.2.4. One-dimensional NMR spectroscopy**

The simplest form of NMR spectroscopy is called one-dimensional or 1D NMR. 1D NMR usually requires a single, broad-frequency RF pulse to excite all the  $^1\text{H}$  or  $^{13}\text{C}$  nuclei in the molecule of interest. The resulting spectrum is a 2D line plot with signal intensity plotted on the Y-axis and frequency (or chemical shift) plotted on the X-axis. An example of a 1D NMR or  $^1\text{H}$  NMR spectrum is shown in Figure 1.6. The position of the peaks (chemical shifts), the multiplicity of the peaks (J-couplings), and the coupling patterns along with the peak or peak cluster intensities (indicating the number of  $^1\text{H}$  or  $^{13}\text{C}$  atoms) can all be retrieved from 1D NMR data. This information can be used to assign the chemical shifts to a molecule or to determine its structure. The detailed process of chemical shift assignment or NMR-based structure determination is complex and far beyond the scope of this introduction, but interested readers may find more about this in the following references (Cavalli et al., 2007; Nerli et al., 2018; Yao et al., 1997; Yesiltepe et al., 2018).

#### **1.2.1.2.5. Two-Dimensional NMR spectroscopy**

For small molecules, one-dimensional NMR is often sufficient to detect all individual peaks for all observable functional groups and to determine a molecule's atomic structure. However, for larger, more complicated molecules or mixtures of compounds, the interpretation of a 1D NMR spectrum may be too challenging due to the presence of several overlapping resonances (Silverstein et al., 2015). Two-dimensional NMR gives much more information than 1D NMR. This is because the data are defined by two frequencies instead of one (Silverstein et al., 2015). An example of a 2D NMR spectrum is shown in Figure 1.7. As seen in this example, a 2D NMR spectrum is somewhat like a topographic map with contour lines used to indicate peak intensities. The X and Y axes are frequencies while the Z axis (peak intensity) is depicted via contour lines.



**Figure 1.6.** A typical 600 MHz HSQC (2D NMR) spectrum of L-lysine (downloaded from HMDB database).

2D NMR spectra are collected by exciting nuclei with multiple RF pulses separated by different time periods. This multi-step RF excitation leads to magnetization transfer between nuclei and the appearance of coupled peaks or pairs of peaks at different frequencies on the X and Y axes. There are two types of 2D NMR: homonuclear correlation spectroscopy, and heteronuclear correlation spectroscopy. In homonuclear correlation spectroscopy, magnetization transfer occurs between nuclei of the same type (usually  $^1\text{H}$  to  $^1\text{H}$ ) (*Handbook of Pharmacogenomics and Stratified Medicine*, 2014). The most often utilized homonuclear correlation spectroscopic experiments are  $^1\text{H}$ - $^1\text{H}$  correlation spectroscopy (COSY), total correlation spectroscopy (TOCSY), and nuclear overhauser enhancement spectroscopy (NOESY). In heteronuclear correlation spectroscopy, magnetization transfer occurs from one nucleus to another nucleus of a different type (usually  $^1\text{H}$  to  $^{13}\text{C}$ ). Heteronuclear multiple quantum coherence (HMQC), and heteronuclear multiple-bond correlation spectroscopy (HMBC) are the two most commonly performed types of heteronuclear correlation spectroscopy.

#### **1.2.1.2.6. COSY and TOCSY**

The  $^1\text{H}$ - $^1\text{H}$  COSY experiment provides information about off-diagonally correlated protons that are directly coupled to each other (Emwas et al., 2019; Rahman, 2015). In the COSY experiment, magnetization is transferred between J-coupled spins. However, in the TOCSY experiment, magnetization transfer to all protons in a spin system occurs (protons that are directly or indirectly coupled to each other) (Emwas et al., 2019). For instance, if a spin system has components A, B, C, and D, then a COSY spectrum will show three off-diagonal peaks between A and B, B and C, and C and D. In contrast, a TOCSY spectrum will show six off-diagonal peaks between A+B, A+C, and A+D, as well as B+C, B+D, and C+D. In other words, correlations will be seen between all the spins in a TOCSY experiment. In this regard, a TOCSY experiment is often more informative than a COSY experiment.

#### **1.2.1.2.7. HSQC and HMBC**

Heteronuclear multiple quantum coherence (HMQC), and it's a slightly more efficient variant known as heteronuclear single quantum coherence (HSQC), are 2D NMR experiments that detect correlations between different nuclei (usually  $^1\text{H}$  and  $^{13}\text{C}$  or  $^1\text{H}$  and  $^{15}\text{N}$ ) that are separated by one bond (Emwas et al., 2019; Öman et al., 2014). These 2D experiments provide information about the relative relationships between protons ( $^1\text{H}$ ) that are directly bonded to carbons ( $^{13}\text{C}$ ) in a chemical structure. On the other hand, heteronuclear multiple bond coherence (HMBC) detects a correlation between  $^1\text{H}$  and  $^{13}\text{C}$  nuclei through two, three, or sometimes four bonds (Emwas et al., 2019; Öman et al., 2014; Rahman, 2015). This 2D experiment provides information about the relative relationships between protons ( $^1\text{H}$ ) that are directly and indirectly bonded to carbons ( $^{13}\text{C}$ ) in a chemical structure. HMBC spectra can show 2-4 bond coupling between protons and carbons. In this regard, the HMBC experiment can provide more structural information than an HSQC or HMQC experiment. However, both types of experiments can be used in concert to decipher which proton signals belong to which  $^1\text{H}$  atoms and which carbon signals belong to which  $^{13}\text{C}$  atoms.

#### **1.2.1.2.8. Software and Databases for Compound Identification in NMR metabolomics**

NMR can be used for the analysis of pure compounds or mixtures. In metabolomics, NMR is primarily used to analyze mixtures, although it can also be used to help identify novel (or unknown) compounds, if those compounds have been purified or partially purified. The analysis of mixtures by NMR is slightly different than the analysis of pure compounds by NMR. This is because the NMR signals in a mixture can arise from multiple chemicals, each

of which has its own characteristic NMR spectra or characteristic chemical shift signatures. As a result, the NMR spectrum of a mixture, which is made up of several different metabolites, is really the sum of the individual NMR spectra for each of the pure metabolites present in the mixture. To determine which compounds are which and how abundant each of those compounds is in the mixture, one must use a technique called spectral deconvolution (Jarvis, 1982). Spectral deconvolution is a process designed to decompose or deconstruct a spectrum into its individual components. In NMR metabolomics, spectral deconvolution is used to separate out or identify distinct pure compound spectra from the more complex mixture spectra. Spectral deconvolution of mixtures by NMR is more robust than for LC-MS or GC-MS. This is because it is rare for any two substances to have the same number of NMR peaks, the same chemical shifts, or the same peak intensities (whereas it is much more common in LC-MS or GC-MS for compounds to have the same retention times and/or the same masses). Using NMR-based spectral deconvolution, it is possible to manually or automatically identify and quantify many compounds in an NMR spectrum by accurately matching and fitting the observed peaks in the chemical mixture to a library of pure reference component peaks (which have been calibrated to an internal concentration standard) (Ebrahimi et al., 2014). There are several software packages that support NMR spectral deconvolution for manual or semi-automated compound identification and quantification. These include commercial packages such as the Chenomx NMR suite from Chenomx Inc. (<https://www.chenomx.com/>) and the Bruker TopSpin package 4.3.0 (<https://www.bruker.com/en/products-and-solutions/mr/nmr-software/topspin.html>). Several freely available, automated NMR methods for 1D NMR spectral deconvolution also exist, including BATMAN (Hao et al., 2014), MagMet (Rout et al., 2023), and Bayesil (Ravanbakhsh et al., 2015). Likewise, a 2D NMR package called COLMAR (F. Zhang et al., 2009) also exists for deconvolution of 2D homonuclear and heteronuclear spectra. Each package has its own software-specific spectral databases. Some of the open-source packages use NMR spectral databases such as the Human Metabolome Database (HMDB) (Wishart, Guo, et al., 2022), the Madison-Qingdao Metabolomics Consortium Database (MQMCD) (Cui et al., 2008), the Natural Product Magnetic Resonance Database (NP-MRD) (Wishart, Sayeeda, et al., 2022), and the Biological Magnetic Resonance Bank (BMRB) (Romero et al., 2020).

### **1.3. Alternate methods for metabolite identification**

The last 15 pages of this document have outlined some of the most common (direct) methods for determining the chemical composition of mixtures in metabolomics or for the

identification of unknown or novel compounds in biological samples. These include LC-MS, GC-MS, tandem mass spectrometry, and NMR spectroscopy. However, it is also possible to use other forms of biological information or other types of databases to help reduce the complexity of the problem. For instance, if one knows the kind of organism being studied, it is possible to restrict the search for metabolites to only those metabolites known to be produced or metabolized by that organism. For instance, if one is studying *E. coli* growing in a fermenter system, one only needs to look for metabolites that are produced by *E. coli*. Likewise, if one is studying metabolites in lab rats or lab mice, then one can immediately exclude metabolites that might come from cosmetics, or drugs, or even plants, given that lab animals are confined to cages and eat restricted diets and therefore are not exposed to these chemicals. So, genetics or knowledge of genetically encoded metabolites can be used to restrict the search space used by MS or NMR-based metabolomics. Another route that can help infer the identity of metabolites is to measure changes in gene expression (transcriptomics). This is because changes in gene expression can lead to changes in metabolite production (or consumption) or the activation of new pathways that can be used to infer or predict metabolites. This will be explained further below.

#### **1.4. Transcriptomics**

Transcriptomics is a branch of genetics that studies the global measurement of an organism's transcriptome, the complete set of RNA transcripts, including messenger RNAs (mRNAs), non-coding RNAs, and small RNAs in a biological system (Dong & Chen, 2013). Transcriptomics allows the measurement of the RNA expression of hundreds to thousands of transcripts in tissues, biofluids, and cells. Transcriptomics provides a simple route to measure gene expression changes (and by proxy, protein expression changes) that arise due to external perturbations. Prior to 2010, most transcriptomics measurements were done using microarray technologies. Microarrays measured the hybridization of sample RNA to defined or synthetic RNA segments attached to specific positions on manufactured chips. The advent of next-generation sequencing has changed the field dramatically, as now it is possible to measure the transcriptome using a more robust technique called RNA-sequencing or RNA-Seq (Wang et al., 2009). The RNA-Seq workflow involves several steps, including sample preparation, the running of DNA sequencing equipment, and data analysis. In the first step, a given sample's RNA is isolated, then converted to complementary DNA (cDNA). Next, the sample is prepared for sequencing and then sequenced using an NGS platform (Kukurba & Montgomery, 2015). RNA-Seq has several advantages over other transcriptomics technologies. Unlike

hybridization-based microarray technology, RNA-Seq is not restricted to identifying and quantifying transcripts from completely sequenced organisms. Additionally, RNA-Seq has significantly greater reproducibility than other microarray technologies (van der Kloet et al., 2020). In addition to offering a more precise and in-depth assessment of RNA abundance than microarrays, RNA-Seq also detects alternatively spliced genes and can determine allele-specific expression arising from polymorphisms in the transcript (Kukurba & Montgomery, 2015). RNA-Seq can be used to quantify several populations of RNA, including total RNA, pre-mRNA, and noncoding RNA (i.e. rRNA, microRNA, and long ncRNA), in addition to assisting with the analysis and counting of mRNA transcripts (Kukurba & Montgomery, 2015).

Transcriptomics can be combined with metabolomics to help identify changes to gene expression and the corresponding changes to metabolism that arise from changes to an organism's nutrient sources (Zhang et al., 2022), chemical exposures (Hudson et al., 2021), genetic changes (Hassan et al., 2020), and exposures to pathogens (Wu et al., 2023). Transcriptomics and metabolomics can also be combined to help identify novel metabolites, especially with regard to novel microbial metabolites (Dutta et al., 2020; Fan et al., 2023).

### **1.5. Rationale of studying catechol metabolism in *Escherichia coli***

This study focuses on catechol metabolism in *Escherichia coli* (*E. coli*), which will contribute to a comprehensive understanding of microbial metabolism and physiology. Catechol, a common aromatic molecule found in a variety of natural instances, acts as a fundamental model for elucidating general degradation mechanisms for more complex aromatic substances. *E. coli* is a gram-negative, rod-shaped facultative anaerobic member of the Enterobacteriaceae family of bacteria (Desmarchelier & Fegan, 2002). *E. coli* has been one of the most extensively studied species since Theodor Escherich discovered it in 1885 (FAQ, 2011). *E. coli*, including other facultative anaerobic bacteria, constitutes about 0.1% of the gut microbiota (Eckburg et al., 2005). The regeneration time of *E. coli* is about 20 minutes, and their growth and culture in a laboratory setting are inexpensive. Due to the availability of its complete genome (Blattner et al., 1997) as well as the comprehensive understanding of its genome and metabolism, this is the most studied microorganism in biological research (Ruiz & Silhavy, 2022). *E. coli* is used in a wide range of applications, including the production of recombinant proteins, enzymes, and bioactive molecules (Gupta & Shukla, 2016; Huang et al., 2012). *E. coli* has also been modified to function as biosensors and detecting molecules (d'Oelsnitz et al., 2022). The metabolic changes and diversity of *E. coli* allow it to adapt in a

variety of biological niches, including the human gut and multiple environmental habitats. These features prompt us to study *E. coli*. Understanding catechol metabolism in *E. coli* will provide insights into the organism's potential as a bioremediation agent to control environmental pollution issues, as well as enable the development of medicinal enhancements.

## 1.6. Statement of thesis problem

Microbes are the microscopic chemical ‘terra-forming’ factories that populate every conceivable niche of the earth. They are responsible for transforming thousands of natural and synthetic chemicals into compounds that can either be consumed, assembled or further degraded so that other organisms can thrive or survive. In particular, microbes in the human gut play a key role in the production of chemicals (often derived from food) that can turn out to be helpful or harmful to the human body (Wishart et al., 2023). However, a complete understanding of the microbe-diet-metabolome relationship is still a long way away (Wishart et al., 2023). This is because the type and variety of chemicals in our diet, and the type and variety of known microbes and microbial chemical reactions are still poorly understood.

To better understand which chemicals can be produced by human gut microbes, especially when grown on specific food-derived chemicals, I decided to apply a combination of metabolomics and transcriptomics to the problem. This is a relatively new area of microbial metabolomics, and therefore few protocols are in place or few methods from which to draw upon. Therefore, my focus for this thesis was on developing methods and technologies on a pilot scale that could eventually be applied more broadly or more generally. Specifically, I chose to determine how a plant-derived chemical called catechol could be chemically transformed by a single human gut microbe, *E. coli*. This small-scale analysis was designed to combine NMR and MS-based metabolomics with RNA-Seq transcriptomics to identify the metabolites produced (via untargeted metabolomics) and the metabolite pathways activated (via transcriptomics) when *E. coli* is grown in catechol-containing media. This study and the results that were found are summarized in the second chapter of this thesis. As far as we are aware, no other metabolic or metabolomic study has been conducted looking at *E. coli* growing in catechol.

This microbial metabolomics/transcriptomics study described in Chapter 2 also led to a second study that looked at the problems and challenges associated with using untargeted metabolomics to identify novel compounds. As highlighted in Chapter 2, many inconsistencies were noted between the NMR and MS metabolomics results. Specifically, many false positive

compounds were identified in the MS analyses. These results and the number of false positives identified were highly dependent on the software being used and the analytical workflow being employed. Therefore, to identify the sources for these errors, I explored how and why these errors could arise by analyzing a simple set of defined mixtures (using both NMR and MS). This was done to better understand some of the challenges facing untargeted metabolomics when applied to novel compound identification. This study and the results that were found are summarized in the third chapter of this thesis. Therefore, based on these ideas and the direction that arose from this research, I have defined three thesis objectives.

## 1.7. THESIS OBJECTIVES

1) Identify and characterize the known and previously unknown metabolites produced by *E. coli* after growth in a catechol-containing medium in both anaerobic and aerobic conditions.

2) Use RNA-Seq to identify upregulated and downregulated genes arising from catechol metabolism in *E. coli* and use these transcriptomics changes to help identify known (and previously unknown) metabolites.

3) Determine the source of the variability arising from untargeted metabolomics (MS and NMR) studies in terms of false positives and false negatives using different software packages applied to defined chemical mixtures.

## 1.8. THESIS OUTLINE

This thesis consists of four chapters. The first chapter (the current chapter) introduces bacterial metabolism and provides a general background on metabolomics (both targeted and untargeted), with a specific focus on the different analytical platforms and analytical techniques used in my thesis research. This chapter also discusses some of the challenges facing untargeted metabolomics in terms of false positives and false negatives for compound identification.

The second chapter describes what is known about catechol metabolism in *E. coli*. It also describes the study design I used to characterize catechol metabolism in *E. coli*. Using a combination of untargeted MS and NMR-based metabolomics and RNA-Seq analysis of transcriptomic changes. The chapter describes the detailed growth conditions, the metabolomics and transcriptomics techniques, the relevant results of this study, and the lessons learned about the limitations of untargeted metabolomics.

Third, chapter describes the untargeted MS-based metabolomics benchmarking experiment. This study was conducted to determine the variability of MS-based results in terms of false positives and false negatives with regard to compound identification during data analysis. The chapter describes and discusses the different analytical techniques used, the types of defined mixtures prepared and analyzed, and the performance of the different software packages.

The final chapter summarizes the results described in chapters 2 and 3 and discusses potential improvements to the study design of both projects.

## Chapter 2

### Exploring Catechol Impact on *E. coli* UM146: A Multi-Omics Perspective on Detoxification and Stress Response Pathways

#### 2.1. Introduction

Phenolic compounds have a benzene ring with one or more hydroxyl groups attached. In nature, this group of chemicals is abundantly present (Minatel et al., 2017). There are more than 4,000 identified phenolic plant chemicals (Tsao, 2010), many of which have applications including pigments, lignin precursors, anti-inflammatory, antiproliferative, and antioxidant properties (Albuquerque et al., 2021). Some of these substances such as catechol, catechin, caffeic acid, gallic acid, ferulic acids, etc. are found in food or are employed in medicine (Delgado et al., 2019; Nisar, 2022). Catechol is one of the essential phenolic flavonoids, present in a trace amount in human food sources, such as green vegetables, green tea, and fruits (Knezevic et al., 2021). It serves as a raw ingredient in the production of several medications, pesticides, and fragrances (Fiege et al., 2000). Catechol is widely distributed in atmospheric air, and originates from vehicle emissions, cigarette smoke, combustion pollution plumes, and biomass burning (Rana et al., 2023). It has been found in drainage water from bituminous shale, crude wood tar, wastewater from coal conversion processes, and effluent from coal-tar chemical synthesis (Lofrano et al., 2009). Its content in wastewater from coal carbonization and gasification ranges from a few mg/L to 2000 mg/L, followed by the concentration able to reach 5,300 mg/L at low temperatures of wastewater (Subramanyam & Mishra, 2007, 2008a, 2008b). Therefore, humans or animals obtain catechol directly from food or through inhalation everyday. Around 14,000 workers in the United States get exposed to catechol (On the Evaluation, 1999). According to the International Agency for Research on Cancer (IARC) (1986), catechol is present in cigarette smoke at a level of 100–360  $\mu\text{g}$  per cigarette (“Tobacco Smoking,” 1986). It is estimated that humans consume or endogenously produce 0.3 mg/kg of catechol daily (McDonald et al., 2001). Although catechol has a diverse set of applications, based on the substantial evidence of carcinogenicity in test animals, the International Agency for Research on Cancer (IARC 1999) classified catechol as “possibly carcinogenic to humans” (Group 2B). Several strains of rats have been shown to develop adenocarcinomas of the glandular stomach after receiving long-term dietary exposure to catechol (IARC, 1999); however, no relevant epidemiological study has been found for humans. The EPA (Environmental Protection Agency) of the United States has classified phenolic chemicals,

including catechol, as priority pollutants (Moussavi et al., 2010; Subramanyam & Mishra, 2007). Catechol is severely poisonous to fish at doses of 5 to 25 mg/L (Kumar et al., 2005; Rigo et al., 2010). Several studies have reported the toxic nature of catechol towards cell lines including cats, fish, humans, mice, rats, rabbits, trout, and water fleas (Garton & Williams, 1948; Hattula et al., 1981; Neilson et al., 1991; Pellack-Walker et al., 1985; Rahouti et al., 1999; Svenson & Hynning, 1997; Van Den Heuvel et al., 1999).

Bacteria have diverse metabolic capabilities of efficiently degrading and transforming a wide range of compounds such as pesticides, hydrocarbons, lignin monomers, amino acids, quinones, and flavonoids (Koppel et al., 2017; Stevens & Maier, 2016; Weng et al., 2021). Bacteria influence bioremediation in nature by degrading, transforming, accumulating, and concentrating many toxic compounds (Coelho et al., 2015; Zhong & Zhou, 2002). Also, bacteria that exist in the human gut play a significant role in the transformation, and metabolism of phenolic compounds (Corrêa et al., 2019; Mosele et al., 2015). The microbial metabolism of catechol has been extensively studied in *Azotobacter* species (*T. aromatica* type strain K172 and *Azoarcus* sp. Strain EbN1), and *Pseudomonas* species (*Pseudomonas putida* and *Pseudomonas cepacia* ATCC 29351) (Ding et al., 2008; Hamzah & Al-Baharna, 1994; Ornston & Stanier, 1966; Suvorova & Gelfand, 2019). Degradation of catechol during aerobic condition follows two routes: meta-cleavage, or ortho-cleavage (Aghapour et al., 2013; Singh et al., 2015). Within the ortho-cleavage route, catechol 1,2-dioxygenase breaks down catechol between its two ring hydroxyls to produce cis, cis-muconate. Subsequently, this cis,cis-muconate undergoes a series of three steps to yield the intermediate 3-oxoadipate (Fritsche & Hofrichter, 2005). Succinyl-CoA and acetyl-CoA are produced after 3-oxoadipate. Catechol ortho-cleavage route is widely spread across soil bacteria and fungi (Njiru et al., 2022). Besides the intradiol-ortho cleavage route, the catechol meta-cleavage pathway is also observed in bacteria from different genera, such as *Ralstonia*, *Azotobacter*, and *Pseudomonas* species (Hamzah & Al-Baharna, 1994; Hughes & Bayly, 1983; Sala-Trepat & Evans, 1971; Shi et al., 2021). Catechol 2,3-dioxygenase enzyme cleaves the aromatic ring of catechol adjacent to the two hydroxyls, and forms (2Z,4E)-2-hydroxy-6-oxohexa-2,4-dienoate. In the next step, 2-hydroxymuconic semialdehyde hydrolase catalyzes the conversion of (2Z,4E)-2-hydroxy-6-oxohexa-2,4-dienoate into (2E)-2-hydroxypenta-2,4-dienoate (Bertini et al., 1994; Sala-Trepat & Evans, 1971). 2-oxopent-4-enoate hydratase, 4-hydroxy-2-oxovalerate aldolase, and acetaldehyde dehydrogenase enzymes work in sequence to convert 2-oxopent-4-enoate to 4-hydroxy-2-oxovalerate, pyruvate, and acetaldehyde (and eventually acetyl-coA) (acylating).

However, in anoxic conditions, catechol is degraded into benzoyl-CoA through coenzyme A ligation, dehydroxylation, and carboxylation (Harwood et al., 1998). Denitrifying bacterium *Thauera aromatica*, and sulphate-reducing bacteria *Desulfobacterium* sp. strain use catechol as only electron donor and carbon source (Ding et al., 2008; Gorny & Schink, 1994). Carboxylation of catechol forms protocatechuate in *Desulfobacterium* species (Gorny & Schink, 1994). Protocatechuyl-CoA is converted into 3-hydroxybenzoyl-CoA in a second oxygen-sensitive step by the reductive removal of the p-hydroxyl group. Procatechuate-CoA ligase and protocatechuyl-CoA reductase are involved in the anaerobic protocatechuate metabolism in *Thauera aromatica* strain AR-1, resulting in 3-hydroxybenzoyl-CoA (Philipp et al., 2002). *T. aromatica* is known to include most of the genes necessary for the anaerobic metabolism of aromatic compounds such as phenol, 3-hydroxybenzoate, 4-hydroxybenzoate, and benzoate (Breinig et al., 2000).

*E. coli* has the ability to break down numerous aromatic compounds such as phenylacetic acid, 3- and 4-hydroxyphenylacetic acid, phenylpropionic acid, 3-hydroxyphenylpropionic acid, and 3-hydroxycinnamic acid, phenylethylamine, tyramine, and dopamine (Díaz et al., 2001). *E. coli* is a highly diverse species of bacterium found in many distinct areas, including the intestines of humans, animals, food, soil, and water. According to NCBI (National Center for Biotechnology Information), more than 2,000 strains of *E. coli* have been identified and characterized. Most varieties of *E. coli* are benign, leading to mild diarrhea in humans and animals. Certain pathogenic strains such as *E. coli* O157:H7 cause foodborne infections, and bloody diarrhea (Tortorello, 1999). *E. coli* UMI46 strain found in the gut intestine involved in inflammatory bowel disease (IBD), including Crohn's disease (CD) (Kittana et al., 2023). It is an adherent invasive (AIEC) strain, possessing adhesins and fimbriae to adhere intestinal epithelial cells to rapidly colonize and invade cell lines (Krause et al., 2011). Adhesins are the preliminary step of biofilm formation. Biofilm formation is a distinctive attribute of nearly 99% of the bacteria, a factor involved in 80% of all human body infections (Davies, 2003; Paraje, 2011). In the USA, around 65% of bacterial infections are related to biofilm (Srey et al., 2013). Therefore, there is a need for antibiofilm agents which can prevent biofilm formation. Several studies were conducted to measure the effect of phytochemicals on biofilm formation in *E. coli* CECT 434 (Baptista et al., 2019) and *E. coli* PBIO729 (Buchmann et al., 2023).

To date, there is no study conducted growing *E. coli* UMI46 strain in catechol containing minimal media. This study will explore the response of *E. coli* UMI46 strain

towards catechol addressing key research questions, including whether *E. coli UMI46* can degrade catechol, the characterization of major byproducts, and an in-depth analysis of gene-expression changes. In the event of catechol displaying toxicity towards the *E. coli UMI46* strain, this study will determine the strategies used by the *E. coli UMI46* strain to survive and grow. Moreover, the impact of catechol on virulence factors of *E. coli UMI46* will be determined through transcriptomics analysis. Previous studies of catechol metabolism in bacteria were conducted using purified enzyme kinetics assay systems. Determination of metabolites from enzymatic reactions works by monitoring the affinity of enzyme-metabolite interactions and the rate of substrate depletion over time. Modern metabolomics techniques can be applied to identify compounds using liquid chromatography high-resolution mass spectrometry (LC-MS/MS), gas chromatography-mass spectrometry (GC-MS), and nuclear magnetic resonance (NMR). In this experiment, a combination of untargeted metabolomics and transcriptomics techniques were conducted to systematically elucidate the metabolites and genes expressed in the *E. coli UMI46* strain after exposure to catechol. In addition, this study will provide more insights into gene regulation and genetic responses of *E. coli* towards aromatic compounds, which could be applicable in bioremediation and sustainable industries.

## **2.2. Materials and Methods**

### **2.2.1. Chemical reagents**

Salts required to make M9 minimal media were purchased from Fisher Scientific. Catechol, glucose, phenol, protocatechuic acid, pyrocatechuic acid, benzoic acid, benzaldehyde, 1,2,4-benzenetriol, and quinone were purchased from Sigma-Aldrich. All other reagents were purchased from Sigma-Aldrich unless otherwise specified. HPLC grade water, deuterated 2,2-dimethyl-2-silapentane-5 sulfonate (DSS-d6), potassium phosphate monobasic, potassium phosphate dibasic, and D<sub>2</sub>O (99.9%) were purchased from Sigma-Aldrich (Oakville, Canada). The 2-chloropyrimidine-5-carboxylic acid (98%) was purchased from ArkPharm (Libertyville, USA). The Amicon (1.5 mL) 3 kDa molecular weight cut-off (MWCO) filtration units were purchased from Millipore Sigma (St. Louis, United States). The NMR tubes (3 mm) were purchased from Bruker Ltd. (Milton, Canada). Compounds for NMR standards were purchased from Millipore, Sigma, AK Scientific Inc., or Tokyo Chemical Industry Co. Ltd.

### **2.2.2. Strains, media, substrate, and culture conditions**

*E. coli UMI46* strain was used in this study. This strain was previously isolated from the ileum of Crohn's patients, and their genome sequence is publicly available in NCBI

(Accession number: NC\_017632). The overnight culture of the *E. coli UMI46* strain was grown from a single, independent colony in LB medium (Sigma-Aldrich). Overnight culture was sub-cultured into flasks containing M9 minimal media. 1L of M9 minimal media was prepared using 6g sodium phosphate (Na<sub>2</sub>HPO<sub>4</sub>), 3g potassium phosphate (KH<sub>2</sub>PO<sub>4</sub>), 0.5g sodium chloride (NaCl), 1g ammonium chloride (NH<sub>4</sub>Cl), 1 mL 1M magnesium sulphate (MgSO<sub>4</sub>), 1ml 0.1M calcium chloride (CaCl<sub>2</sub>), and 10 ml 10% (w/v) glucose (1 mg/mL). Salts in M9 minimal media were sterilized by autoclaving for 20 minutes at 121°C. Glucose, magnesium sulphate, and calcium chloride were filtered and added to the media after autoclaving salts. The initial pH of the media was maintained to ~7.0.

Sterile catechol was added to the M9 minimal media (with and without glucose) for the treatment group, and only glucose was added to the M9 minimal media (no catechol) for the control group. The choice of catechol concentration was based on several factors such as solubility, sedimentation (higher concentration of catechol leads to the formation of sediments), detectable presence of secondary metabolites (metabolites below 5 ppm not noticeable in NMR), and shoulder peaks (typically very low in concentration) near catechol. Initially, different concentrations of catechol (0.75 mg/ml, 1 mg/ml, and 1.5 mg/ml) were used to observe the impact of catechol on the growth of *E. coli UMI46* strain. Based on the results of different concentrations, 0.75 mg/mL catechol was used for transcriptomics and metabolomics study.

The experiment was performed in both aerobic and anaerobic conditions. Anaerobic conditions were maintained by purging with nitrogen (N<sub>2</sub>) gas. All cultures were incubated at 37°C under a shaking incubator (200 rpm) for 24hrs. Growth was measured in a spectrophotometer (Model: Spectrophotometer UV/Vis Eppendorf BioPhotometer Plus 6132). Samples intended for transcriptomic analysis were collected during the exponential growth phase, contrasting with the metabolomics study, which were collected after 24 hrs. Three replicates were used for each experimental condition (encompassing the control and treatment group) to maintain statistical significance.

### **2.2.3. RNA extraction and quality check**

Total RNA was isolated using Trizol and QIAgen RNeasy mini kit. TRizol manufactured protocol was modified for RNA isolation. Collected samples at the exponential growth phase were centrifuged for 10 minutes at 4,000 rpm to separate cells from the media. Cell volume was maintained at 10% of the lysis reagent. An optimized amount (~350 µL) of

TRizol was added and homogenized using a pipette. 70  $\mu$ L of chloroform was added to the homogenate, and vigorously vortexed for 10-15 seconds. Samples were incubated on ice for 15 minutes and centrifuged at 12,000g for 15 minutes at 4°C to get phase separation. The aqueous phase was transferred to a new tube and precipitated with 200  $\mu$ L isopropanol. Samples were incubated on ice for 10 minutes and centrifuged for 15 minutes at 12,000g at 4°C. Supernatant was removed, and pellet was washed with 500  $\mu$ L of 75% ethanol by flicking. Washed pellet was centrifuged for 10 minutes at 7,500g at 4°C. Collected pellet was air-dried for 2-3 minutes and eluted with 50  $\mu$ L of RNase-free water. Eluted RNA re-purified using Qiagen RNeasy mini kit and followed the manufacturing protocol (RNeasy® Mini Kit quick start, catalog number: 74104). DNase treatment was performed at the end of the RNA isolation using Invitrogen™ TURBO DNA-free™ Kit. The quality of the isolated RNA was analyzed at optical density (OD) 260/280 ratio using Nanodrop (Thermo Scientific). Based on the nanodrop results, final quality check was performed using Bioanalyzer (Agilent 2100 Bioanalyzer system). Isolated RNA with RIN (RNA integrity number) >7 was considered for RNA sequencing.

#### **2.2.4. Library preparation**

Genome Quebec performed the library preparation and sequencing. Total RNA and its integrity were measured using the LabChip GXII (PerkinElmer) instrument. QIAseq FastSelect (-5S/16S/23S Kit 96rxns) was used to remove rRNA from 250 ng of total RNA. The NEBNext RNA First Strand Synthesis and NEBNext Ultra Directional RNA Second Strand Synthesis Modules were used to create cDNA (New England BioLabs). The NEBNext Ultra II DNA Library Prep Kit for Illumina (New England BioLabs) was used for the remaining library preparation steps. Adapters and PCR primers were purchased from New England BioLabs. Libraries were quantified using the Revised Primers-SYBR Fast Universal kit and the Kapa Illumina GA (Kapa Biosystems). The LabChip GXII (PerkinElmer) instrument was used to determine the average size of the fragment. After normalizing and pooling the libraries, samples were denatured in 0.05N NaOH and neutralized using HT1 buffer. According to the manufacturer's guidelines, the pool was loaded at 225 pM on an Illumina NovaSeq S4 lane using the Xp protocol. The run was performed for 2x100 cycles (paired-end mode). At a 1% level, a phiX control library was mixed with libraries. RTA v3.4.4 was used for base calling. The samples were then demultiplexed to produce fastq reads using the program bcl2fastq2 v2.20. More than 10 million reads were generated per sample.

### 2.2.5. RNA-seq data analysis

RNA-seq data analysis was performed using Galaxy Australia server (<https://usegalaxy.org.au/>). FASTQC in GALAXY was used to read quality reports. Adaptor sequences were removed using Cutadapt. Reads were mapped to the *E. coli UMI46* reference genome (NCBI accession: NC\_017632) using Bowtie2. It uses a quick and memory-efficient method to align sequence reads with long reference sequences. Stringtie was used to map reads to each gene. Identification of differentially expressed genes was estimated using DESeq2. Finally, functional enrichment was conducted using ECMDB 2.0 (<https://ecmdb.ca/>), BRENDA (<https://www.brenda-enzymes.org/index.php>), Rhea (<https://www.rhea-db.org/>) Uniport (<https://www.uniprot.org/>), EcoCyc (<https://ecocyc.org/>) and ShinyGO 0.80 (<http://bioinformatics.sdstate.edu/go/>). To compare the gene expression patterns of treated and untreated (control) samples, differential gene expressions (DEGs) were set to log<sub>2</sub> fold-change  $\geq 0$  and adjusted p-value  $\leq 0.05$ . Volcano plot was generated using GALAXY. Heatmap was displayed using the R programming language.

### 2.2.6. Sample preparation for metabolomics analysis

#### 2.2.6.1. Extracellular metabolites extraction

Control and treated samples were centrifuged at 14,000 rpm for 20 minutes at 4°C. 200  $\mu$ L supernatant was aliquoted to a new Eppendorf tube with 50  $\mu$ L buffer X (750 mM phosphate buffer with 5 mM DSS and 10% D<sub>2</sub>O). The mixture was vortexed for 1 minute and centrifuged at 10,000 rpm for 5 minutes at 4°C. Samples were lyophilized and fractionated using three solvents, including hexane, ethyl acetate, and methanol to remove salts. After salt removal, individual extracts were run separately in LC-MS/MS, GC-MS, and NMR.

#### 2.2.6.2. Intracellular metabolites extraction

Control and treated samples were centrifuged at 5,000 rpm for 5 minutes at 4°C. Phosphate-buffered saline (PBS) was added to the collected pellet and resuspended using vortexing. Samples were placed into -80°C freezer for an hour. Samples were kept on ice for 10-15 minutes for thawing. Thawed samples were vortexed for 1-2 minutes. This freezing-thawing-vortexing cycle was repeated 7-10 times to break the cells to release the intracellular metabolites. Samples were centrifuged at 14,000 rpm for 20 minutes at 4°C.

### 2.2.7. Orbitrap analytical conditions

LC-MS/MS analyses were conducted with an Ultimate 3,000 UHPLC system (Thermo Scientific®, MA, USA) coupled to an Orbitrap mass spectrometer which is equipped with a heated electrospray ionization (H-ESI) source. Prior to analysis, external mass calibration was done in accordance with manufacturer instructions. Chromatographic separation was conducted on a ZORBAX C18 column (2.1 × 100 mm I.D., particle size 3.5 µm). The column temperature was set at 50°C, maintained by a Dionex UltiMate 3,000 RS analytical column heater. Mobile phases were binary mixtures: water (eluent A) and methanol (eluent B), both with 0.1% formic acid. Gradient elution started at 100% A and 0% B, followed by: 0–13 min 100% B, 13–15 min 100% B, 15–15.2 min 100% B, 15.2–20 min 0% B. The total run time was 20 min. The flow rate was set at 5 µL s<sup>-1</sup> and the injection volume was 5 µL. Metabolites were analyzed under both positive and negative electrospray ionization (ESI) modes. The ESI conditions in each run were set: spray voltage 3.20 KV in positive mode and -3.2 KV in negative mode, heated capillary temperature 320°C, sheath gas flow rate 55 U, auxiliary gas flow rate 20 U, sweep gas flow rate 1 U, auxiliary heater temperature 370°C, and S-lens RF level 55. The settings for dd-MS<sup>2</sup> data acquisition were as follows: resolution - 17500 fwhm; automatic gain control (AGC) target - 1e<sup>5</sup>; maximum injection time - 50 ms; loop count - 5; and isolation window - 4 m/z. Data acquisition conditions were set over a mass range of 80–1200 m/z to enhance the compound identification.

### 2.2.8. LC-MS/MS data analysis

Raw data obtained from the Orbitrap instrument were converted to mzML format using msConvert. The spectra were processed in Metaboanalyst 5.0, MZmine2, XCMS, and MS-DIAL using the following modules: (1) Metaboanalyst: peak picking (min\_peakwidth - 6.0, max\_peakwidth - 25.0, ppm - 10, mzdiff - 0.01, snthresh - 10.0, prefilter - 6.0, noise - 1E7, value of prefilter - 1E6), peak alignment and peak annotation settings were kept default; (2) XCMS: noise filtration was set to 1E6, and rest of the settings were kept default; (3) MZmine2: mass detector (centroid), ADAP chromatogram builder (min group size in #of scans - 5, group intensity threshold - 1E6, m/z tolerance 0-5 ppm), feature detection - chromatogram deconvolution, spectral deconvolution - hierarchical clustering, isotope peak grouper - deisotoped, filtering - peak filter, alignment - RANSAC aligner, and normalization - liner; (4) MS-DIAL: ionization type - soft ionization, separation type - chromatography, MS method type - conventional, data type - centroid, ion mode - negative / positive, minimum peak height - 1E6, and rest of the settings were kept default.

Selection and determination of metabolites were performed based on several criteria: (1) comparing control, treated, and blank samples. Also, pure catechol solution was analyzed to check if there were any degraded products from catechol. Metabolites identified in control, blank, and pure catechol solution were removed from the treated group; (2) Metaboanalyst, XCMS, MZmine2, and MS-DIAL suggested compounds further confirmed with fragmentation pattern matching [either from MONA (<https://mona.fiehnlab.ucdavis.edu/>) / Massbank Europe (<https://massbank.eu/MassBank/>) or predicted using CFM-ID 4.0 (<https://cfmid.wishartlab.com/>)], dot product similarity score (>0.52) (Schollée et al., 2017), and retention time comparison with the authentic standards.

### 2.2.9. GC–MS analysis

Gas chromatography–quadrupole mass spectrometry analysis was conducted by Agilent 7890 A/5975C GC system equipped with HP-5MS fused silica capillary column (Agilent J&W Scientific, 30 m × 0.25 mm × 0.25 μm). 2 μL of each extracted sample was injected at a split ratio of 10:1. The inlet, and the MS transfer line were set at 250, and 280 respectively. The initial oven temperature was held at 50°C with 1 min hold time and then ramped at a rate of 1°C/min to 60°C temperature, after which the temperature ramped to 300°C at a rate of 20 °C/min with 5 min hold time. The flow rate of the helium carrier gas was 1 mL/min. Electron ionization (EI) energy was set at 70 eV. Mass spectra data acquisition with full-scan spectra (50–600 *m/z*) were acquired at a scan speed of 2 spectra per second after a solvent delay of 3 min. Alkane standard (C8-C20) was run before and after the sample run to calculate the retention indices.

#### 2.2.9.1. GC-MS data processing

Raw data were converted to netCDF format and processed using AnalyzerPro XD within default settings. AnalyzerPro XD automatically processes the data, including raw signal extraction, deconvolution, baseline filtering, peak identification, and integration. It searches metabolites against commercially available databases such as the National Institute of Standards and Technology (NIST) 20 version NIST 2.4. The mass spectral fragmentation pattern and retention indices of each suggested compound by AnalyzerPro XD were further confirmed with the NIST 20 spectral library. Retention indices (RI) was calculated using the following formula obtained from NIST,  $I_x = 100n + 100(t_x - t_n) / (t_{n+1} - t_n)$ ; where  $t_n$  and  $t_{n+1}$  are retention times of the reference *n*-alkane hydrocarbons eluting immediately before and after chemical compound “X”;  $t_x$  is the retention time of compound “X”.

The criteria were set to screen the potential metabolites: (1) fragmentation pattern similarity score (>50%) (Prodhan et al., 2019), and (2) retention index (threshold: library or estimated RI +/- 35) (Degnan et al., 2023). Moreover, identified metabolites in control, blank, and pure catechol solution were removed from the treated group.

#### **2.2.10. NMR analysis**

200 ul of sample was aliquoted to an eppendorf tube with 50 ul buffer X (750 mM phosphate buffer with 5 mM DSS and 10% D<sub>2</sub>O). The mixture was vortexed for 1 minute and centrifuged at 10000 rpm for 5 minutes at 4°C. 1D-1H NMR spectra were obtained on a Bruker AVANCE III 700 MHz spectrometer (Bruker Biospin, Rheinstetten, Germany) equipped with a triple resonance 5 mm CryoProbe. Samples were acquired with automation using a SampleJet sample changer. The samples were stored at 5 - 10 °C and each sample was pre-warmed to 25 °C before insertion into the spectrometer. Spectra for metabolomics analysis were acquired using a <sup>1</sup>D <sup>1</sup>H NOESY (noseypr1d), with 2s pre-saturation pulse for water suppression, a 50 ms mixing time with water saturation, and a 4s acquisition time. A sweep width of 12 ppm was used. Spectra with poor line shape or linewidths greater than 1 Hz were reacquired. For profiling with Chenomx, the NOESY spectra were processed with an exponential line broadening such that the DSS peak width was 1 Hz, and manual baseline correction was applied. For manual profiling, samples were quantified using the Chenomx NMR Suite version 8 (Chenomx, Inc. Alberta, Canada) using a combination of the software-provided 700 MHz compound library and an in-house compound library acquired at 700 MHz. 2D <sup>1</sup>H-<sup>13</sup>C HSQC (hsqcedetgpcisp2.3) and 2D <sup>1</sup>H-<sup>1</sup>H TOCSY (dipsi2esgpph) NMR spectra were also acquired for some samples. The HSQC spectra were acquired with 64 scans and a resolution of 2048 by 256 points, a relaxation delay of 1.5 sec, and an acquisition time of 113 ms. The TOCSY spectra were acquired with 16 scans and a resolution of 16384 by 512 points, with a relaxation delay of 1.5 sec, and acquisition time of 975 ms, and a mixing time of 80 ms. TOCSY and HSQC data analysis was conducted in COLMAR within default settings.

#### **2.2.11. In vitro enzymatic and spontaneous reactions**

Metabolites detected in different analytical platforms were subjected to in vitro assays and reactions to confirm whether these were produced through enzymatic or spontaneous reactions. A total of 13 reactions (one enzymatic and the rest of spontaneous reactions) were performed under identical culture conditions such as 37°C temperature, 24 hrs incubation, and 200 rpm. The concentration of each metabolite was 5 mM, and these reactions were conducted

in slightly acidic condition (pH: 6-7). Individual reactions were performed between catechol and six compounds, including glycerol, valeric acid, sodium carbonate, acetic acid, glyoxylate, and acetyl-coA. Similarly, ethanol was separately incubated with protocatechuic aldehyde, quinol, and 4-hydroxybenzoate. Another 48h hrs experiment was performed with protocatechuate, ethanol, and methanol, followed by 5 mM mixture of protocatechuate and ethanol incubated for 24 hrs, afterwards methanol was added and further incubated for 24 hrs. Besides, phenol was incubated with acetyl-coA for 24 hrs following similar conditions maintained to grow bacteria. Additionally, one enzymatic reaction was conducted with different concentrations of polyphenol oxidase (1U, 5U, 25U, 50U, 500U, 1000U, and 2000U) and 5 mM catechol. 500U of polyphenol oxidase (without the catechol) was used as a control. Samples were quantified using the Chenomx NMR Suite version 8 (Chenomx, Inc. Alberta, Canada). Compounds that are not present in the Chenomx library, their chemical shifts were predicted either from HMDB 5.0 (<https://hmdb.ca/>) or NMRdb (<http://www.nmrdb.org/>).

#### **2.2.12. Generation of 3D protein homology model**

The three-dimensional structure of each enzyme involved in the catechol metabolism of *E. coli UMI46* was predicted using AlphaFold 2 (Jumper et al., 2021). AlphaFold works based on a neural network model, which was trained using a large dataset of known protein structures and corresponding sequences to predict protein structure with high accuracy. AlphaFold generated best model was selected for docking studies.

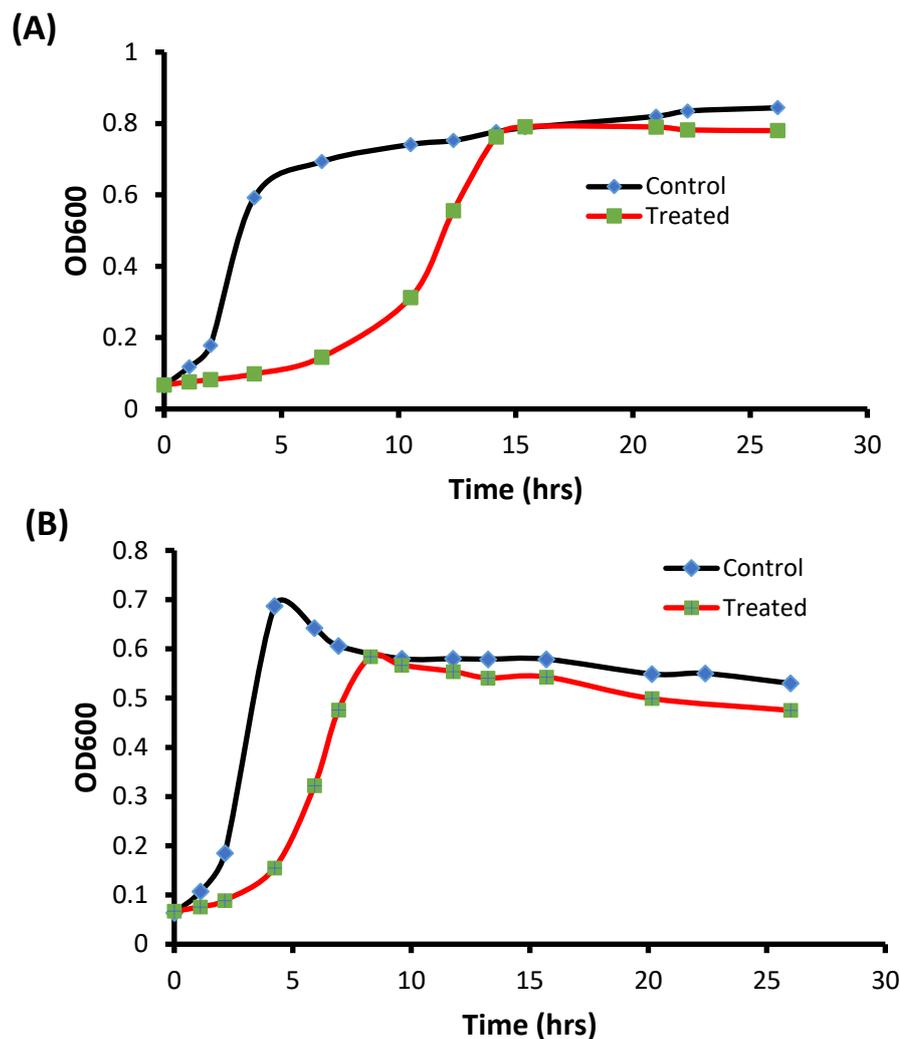
#### **2.2.13. Ligand preparation, docking grid generation and molecular docking**

PDB formatted twelve metabolite structures (adenosine, 3-(acetylthio)isobutyric acid, catechol, glycine, 4-hydroxybenzoate, hydroxyquinol, L-dopa, phenol, quinol, 2-octaprenylphenol, 2-octaprenyl-6-methoxyphenol, and riboflavin) were generated using Marvin (<https://marvinjs-demo.chemaxon.com/latest/demo.html>). MGLTools\_win32\_1.5.6 was used to create PDBQT files of each ligand and protein. Polar hydrogen addition and grid box preparation for protein was conducted in MGLTools. Autodock\_vina\_1\_1\_2\_win32.msi was used for docking to show the binding energy of ligand-receptor interaction. Discovery Studio 2021 software was used to visualize the predicted protein structure, ligands, and their interaction (Systems D., 2021).

### **2.3. Results**

### 2.3.1. Effects of catechol exposure on growth of *E. coli* UMI46 strain

*E. coli* UMI46 didn't grow in catechol containing minimal media without glucose. DNA condensation or cell aggregation was observed at levels  $\geq 1$  mg/mL of catechol (including glucose) under aerobic condition. The use of 0.75 mg/mL of catechol (including glucose) didn't result in any DNA condensation or cell aggregation in both aerobic and anaerobic conditions. Therefore, 0.75 mg/mL of catechol was used in treated condition, as well as for downstream analysis.



**Figure 2.1.** Growth curve of *E. coli* UMI46 after catechol exposure in both aerobic (A) and anaerobic (B) conditions. The optical density (OD) was measured at different time points of incubation. Red and black color depicts treated and control samples respectively. Dice and rectangular symbols were used to indicate the time points of measuring the optical density (OD).

The starting concentrations of catechol in aerobic and anaerobic conditions were 6.731 mM and 6.751 mM respectively. Following a 24 hrs incubation period, the concentration

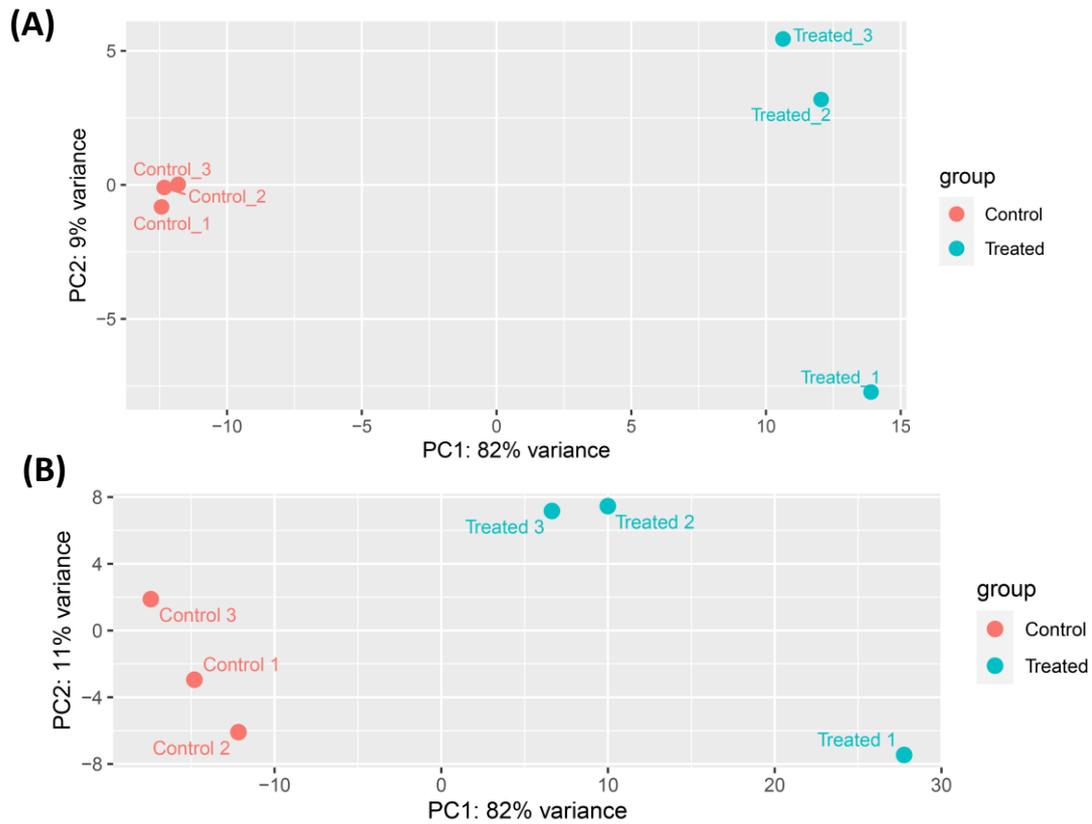
decreased to 5.797 mM (under aerobic condition) and 6.44 mM (under anaerobic condition). A total of 13.87% and 4.61% of catechol was consumed under aerobic and anaerobic conditions respectively. Figure 2.1 shows the growth curve of *E. coli UMI46* during treated and control conditions. In the presence of catechol, *E. coli UMI46* exhibited delayed and reduced growth compared to control across both aerobic and anaerobic conditions. The doubling time of *E. coli UMI46* were  $42 \pm 5$  min and  $69.5 \pm 5$  min in aerobic and anaerobic conditions respectively. In minimal media under aerobic condition, the doubling time of *E. coli B* and *E. coli MG1655* were reported 40-64 min (Plank & Harvey, 1979; Thakur et al., 2010). In response to catechol, a steady doubling time was observed, with an expansion to  $158.5 \pm 5$  min during aerobic condition and  $123 \pm 5$  min under anaerobic condition.

### **2.3.2. Effect of catechol stress on genetic level**

#### **2.3.2.1. Overall differential expression of genes**

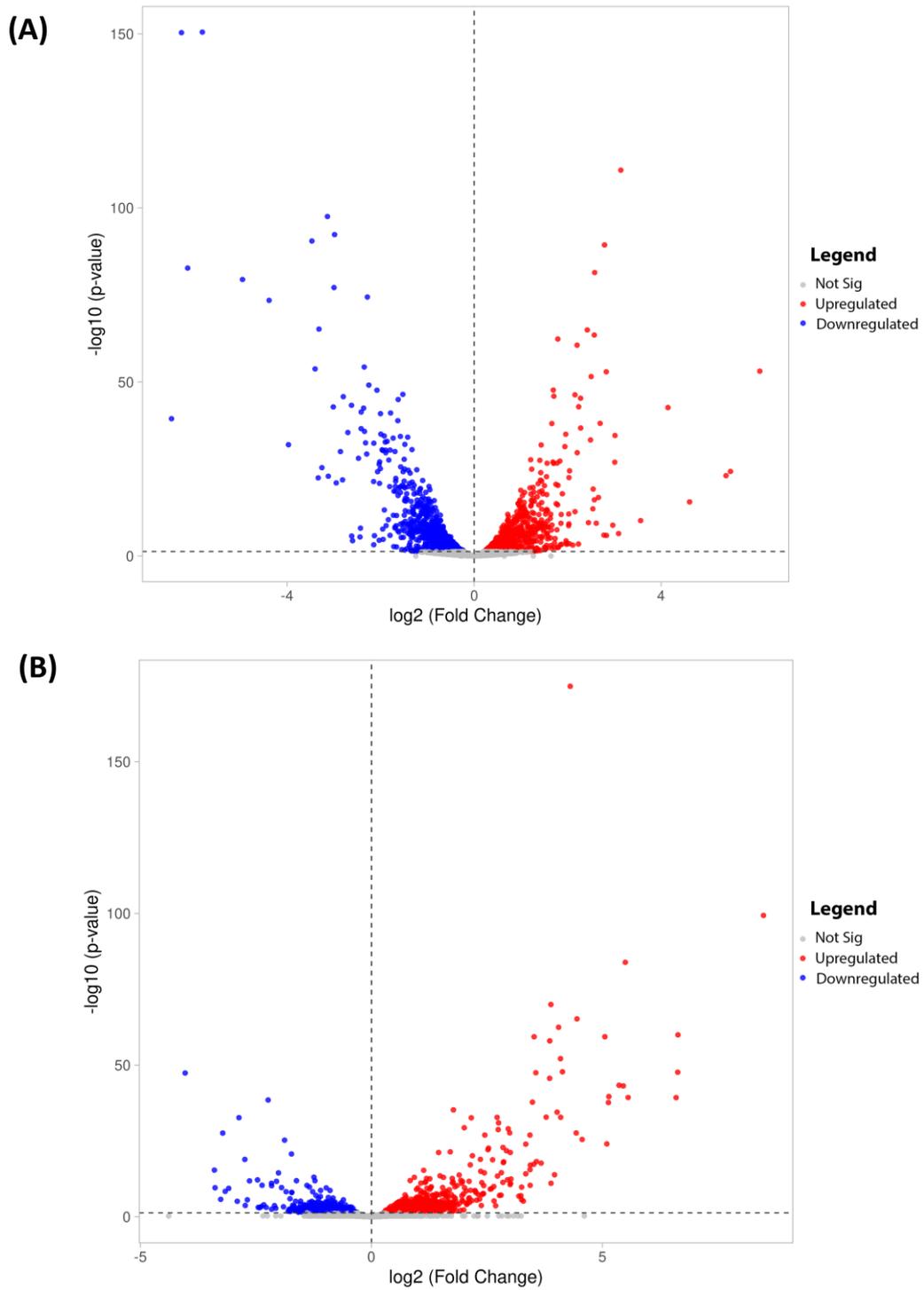
In the presence and absence of catechol, gene-expression study of the *E. coli UMI46* strain was conducted using whole-transcriptome analysis. The extracted RNAs were of high quality with RIN (RNA integrity number)  $>8.2$  (S Table 2.1). Reads of catechol-treated samples and their mapping rates were reported in S Table 2.2 (supplementary section). The quality score (Q) of all the sequences was 36. The standard base quality score ranges from 30-40, and an average quality score above 33 is considered a high-quality sequence (Sheng et al., 2017). Each sequence achieved the quality standards for downstream analysis as per the results of fastQC reports (including per tile sequence quality, per base sequence content, per sequence GC content, and per base N content). More than 30 million raw sequencing reads were generated for each sample. Overall alignment rates were above 97% (S Table 2.2). Since the sequencing data were of high quality, exclusively adaptor sequences were removed as low-quality reads.

Principal component analysis (PCA) was conducted to display the similarity or variability between groups (Figure 2.2). In both aerobic and anaerobic conditions, control and treated groups were distributed in the negative and positive regions of PC1, indicating a significant difference between the two groups. Variations were also observed among treated samples in both conditions. The reasons behind the variations could be sequencing depth, sequencing artifacts, viscosity, biological variation, and contamination while handling the samples.

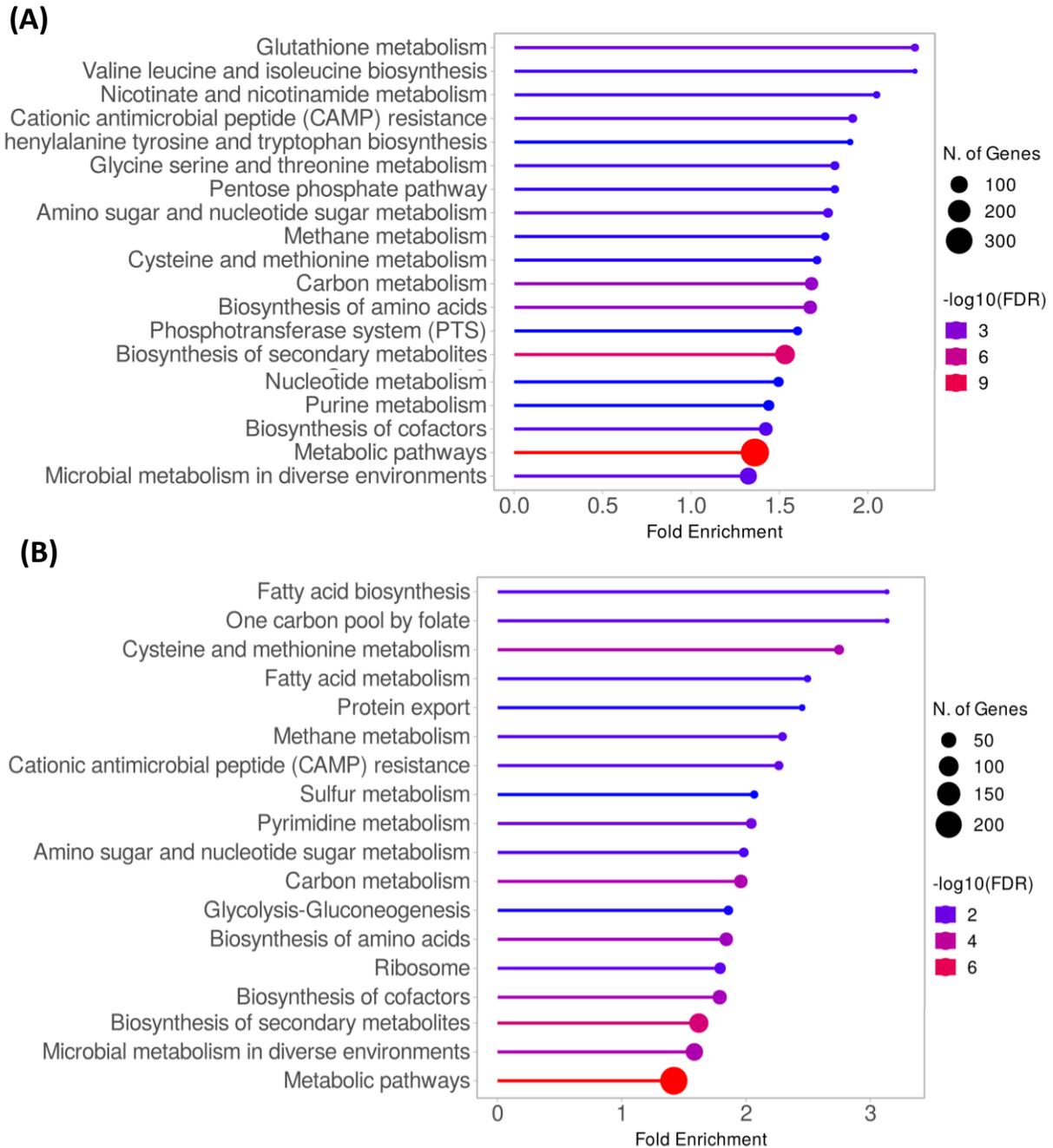


**Figure 2.2.** Principal component analysis of treated and control samples in both anaerobic (A) and aerobic (B) conditions. In the horizontal axis, PC1 represents the maximum difference between groups; and in the vertical axis, PC2 represents secondary differences between groups. 1, 2, and 3 indicate replicate number.

Variations between replicates could result in observed differences by random chances rather than true biological differences. To minimize the false positive results, multiple comparison corrections, including Benjamini–Hochberg false discovery rate was conducted under DESeq2 analysis. Besides this, z-score calculation, box plot, and Grubbs' test were employed to identify and remove the outliers. DEGs analysis represents 669 upregulated and 691 downregulated genes under anaerobic condition (Figure 2.3). A significant difference was noted in aerobic condition, with the upregulation of 746 genes and the downregulation of 392 genes (Figure 2.3).



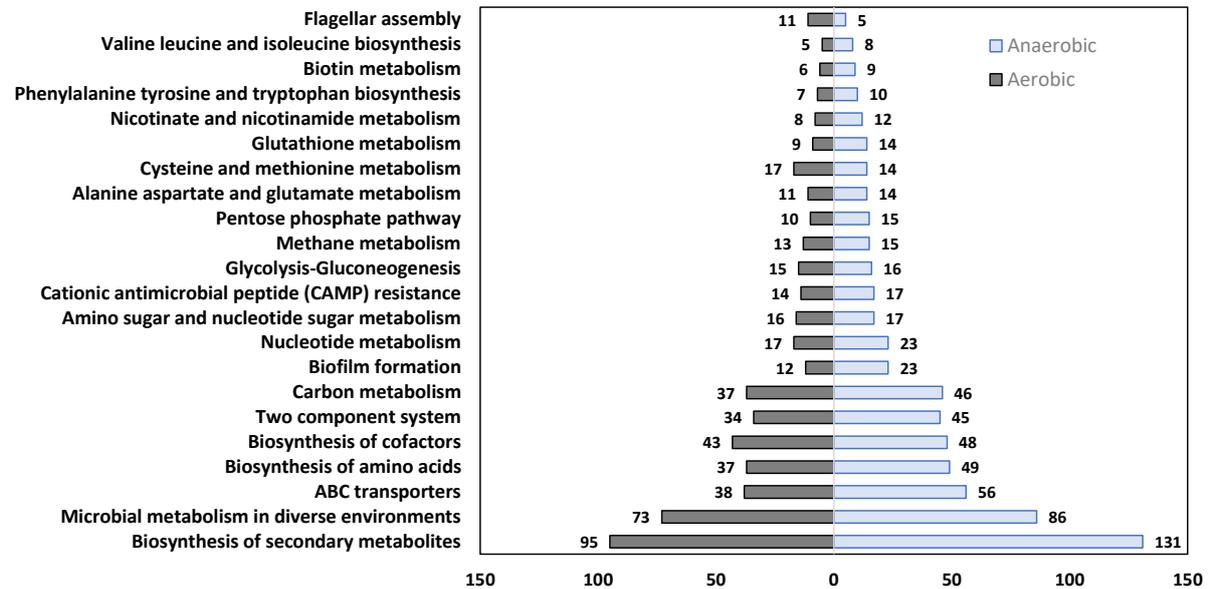
**Figure 2.3.** A volcano plot representation of the differentially expressed genes in *E. coli* UM146 under catechol exposure in both (A) anaerobic, and (B) aerobic conditions. Red, gray, and blue color represents upregulated, not significant, and downregulated genes.



**Figure 2.4.** A graphical representation of gene set enrichment of *E. coli* UMI146 under catechol treatment in both anaerobic (A) and aerobic (B) conditions. Fold enrichment is presented in X-axis. Red, purple, and blue color denote high, medium, and low FDR respectively. Enrichment FDR cut-off was set <math><0.05</math>.

Gene set enrichment analysis (GSEA) revealed a number of pathways affected due to catechol exposure in aerobic and anaerobic conditions (Figure 2.4). Notable transcriptomic changes were observed in both conditions. Figure 2.5 displays a comparison between aerobic and anaerobic conditions, detailing the number of genes linked to key pathways altered by catechol exposure. In comparison to anaerobic condition, limited gene expression was observed

across multiple pathways (except flagellar assembly, and cysteine and methionine metabolism) in aerobic condition.



**Figure 2.5.** A visual contrast between aerobic and anaerobic conditions with differentially expressed genes across key metabolic pathways under catechol exposure. Gene numbers are presented in X-axis. Light blue and grey color represent anaerobic and aerobic conditions respectively.

Figures 2.8-2.9, Figure 2.17, and Figures 2.20-2.21 illustrate two-component systems, mismatch repair system, and biofilm formation respectively, with highlighting the observed upregulated and downregulated genes. Other pathways (including glycolysis-gluconeogenesis, pentose phosphate pathway, biotin metabolism, glutathione metabolism, amino sugar and nucleotide sugar metabolism, methane metabolism, folate one-carbon metabolism, and cysteine and methionine metabolism) with indicating upregulated/downregulated genes were displayed in S Figure 2.1- S Figure 2.14.

### 2.3.2.2. Influence of catechol exposure on two-component systems

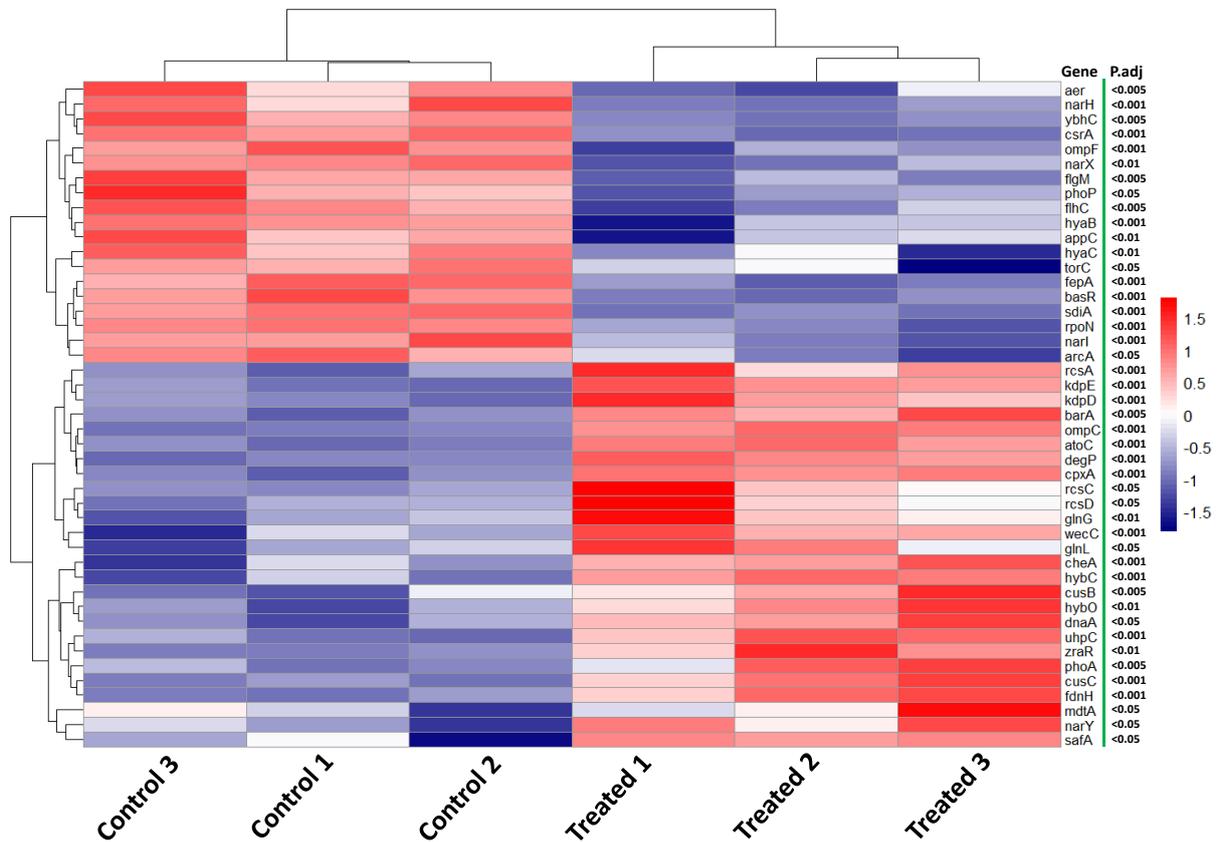
Bacteria possess a signal transduction system (referred to as two-component system) to sense, respond and adapt to changes in their environment (Skerker et al., 2005). Bacteria use this system for survival and resilience across diverse conditions. In the present study, catechol treatment led to a distinctive gene expression profiling, particularly in the two-component system across both aerobic and anaerobic conditions.

In anaerobic condition, a set of genes (including *atoC*, *barA*, *cheA*, *cpxA*, *cusB*, *cusC*, *dctA*, *dcuB*, *degP*, *dnaA*, *fdnH*, *glnG*, *glnL*, *hybC*, *hybO*, *kdpD*, *kdpE*, *mdtA*, *nary*, *ompC*, *phoA*,

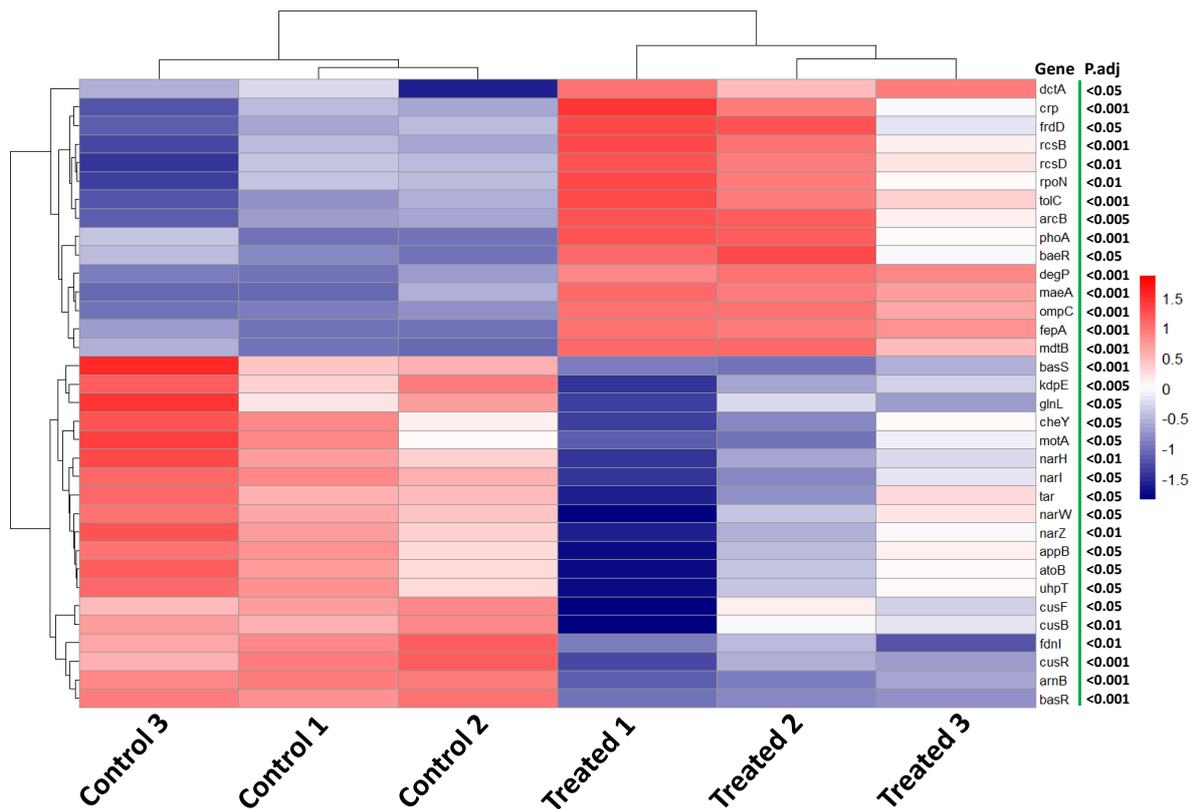
*rcsA*, *rscC*, *rscD*, *safA*, *uhpC*, *wecC*, *zraR*) exhibited upregulation, suggesting an activated response to catechol induced stress (Figure 2.6). *barA* gene is an integral component of the sensory kinase system (Chavez et al., 2010). The observed upregulation of the *barA* gene signifies its involvement in sensing extracellular signals, particularly in response to catechol. The elevated expression of the *cheA* gene (a central regulator of bacterial chemotaxis) enabled cells to navigate along or away from catechol. *ompC* and *ompF* genes encode for porins, allowing the transfer of small molecules (Mizuno et al., 1988). The differential regulation of *ompC* (upregulated) and *ompF* (downregulated) genes, indicating *E. coli* UMI 146 controlled the influx of catechol and nutrients through the adjustment of the permeability of the outer membrane. The upregulation of *rscC* and *rscD* genes suggests an activation of the Rcs system due to catechol-induced envelope stress or damage. *cpxA* gene also activated due to envelope stress. This activation leads to the upregulation of the *degP* gene, which encodes a protease involved in protein quality control and stress response. The upregulation of *glnG*, *glnL*, and *narY* genes implies catechol-induced stress causes a change in nitrogen metabolism. Moreover, an increased expression of *atoC*, *hybC*, *hybO*, *kdpD*, *kdpE*, *mdtA*, *phoA*, *uhpC*, *wecC*, and *zraR* genes denotes that *E. coli* UMI 146 facilitated cellular response to enhance stress resistance by modifying cell membrane properties and adapting metabolic pathways. The downregulation of multiple genes denoting *E. coli* UMI 146 strain made collective adjustments to prioritize essential processes such as cellular homeostasis and survival chances in diverse conditions. A more detailed overview of each of the gene's involvement in the particular two-component system is shown in Figure 2.8.

Likewise anaerobic condition, distinct changes were observed in the two-component system under aerobic condition (Figure 2.7). The upregulation of *rscB* and *rscD* genes activated the Rcs system to sense envelope damage. The increased expression of the *ompC* gene activates the outer membrane system to restrict the entry of toxic substances (such as catechol) inside the cell. The upregulation of the *baeR* gene prompted the multidrug resistance to regulate the MdtABC-efflux pump. The upregulation of *mdtB* and *tolC* genes suggested MdtABC-TolC efflux complex acted together to confer resistance to catechol. The upregulation of the *envZ* gene implies *E. coli* UMI 146 monitored and adapted to osmolarity changes due to the presence of catechol. Increased expression of the *rpoN* gene indicates alternative sigma factor 54 (RpoN) regulates stress resistance genes. A set of genes (including *arnB*, *appB*, *atoB*, *basR*, *bass*, *cheY*, *cusB*, *cusF*, *cusR*, *fdnI*, *glnL*, *kdpE*, *motA*, *narH*, *narI*, *narW*, *narZ*, *tar*, and *uhpT*) were downregulated in response to catechol in aerobic condition. The downregulation of these genes

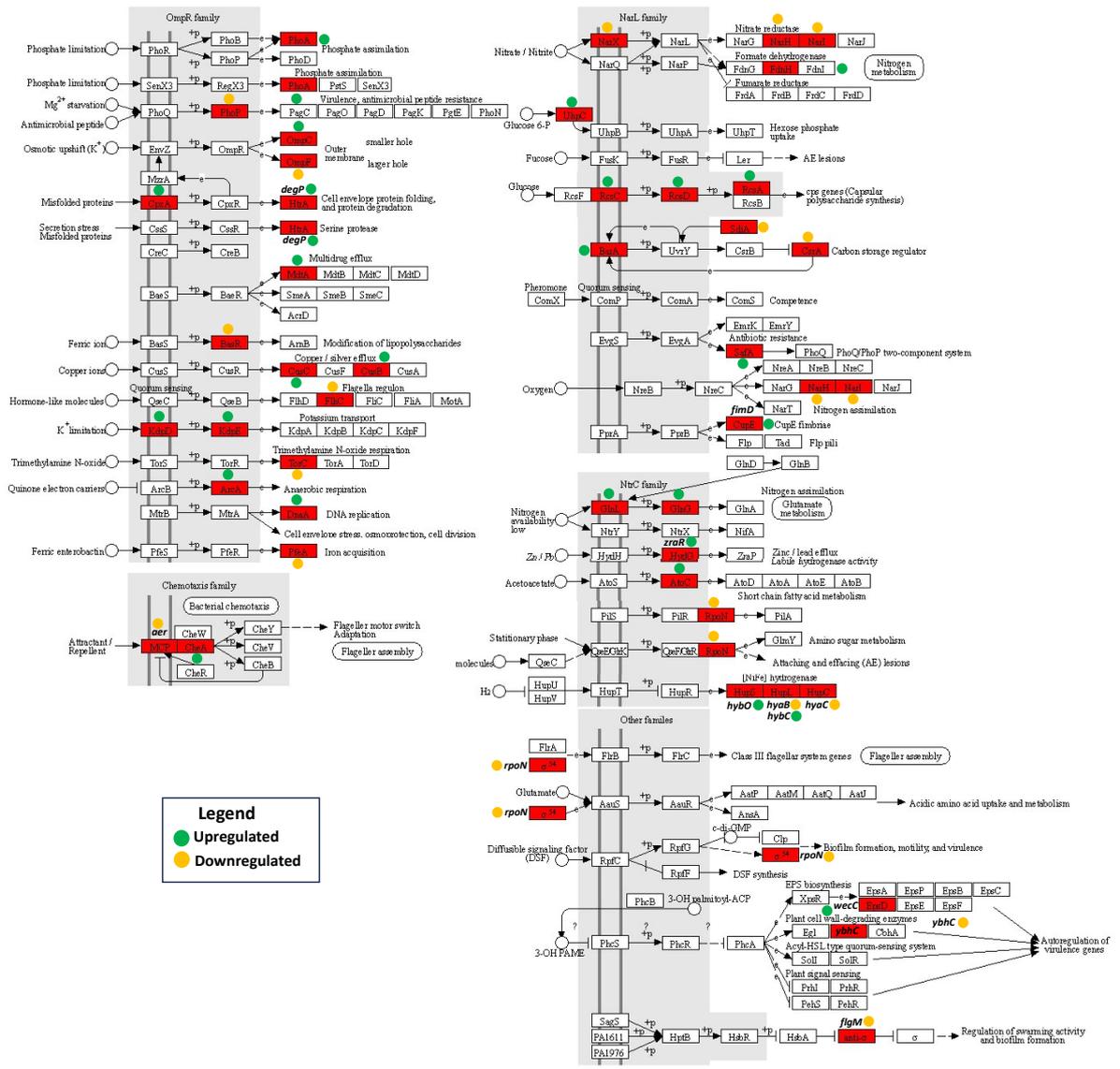
implies a strategic shift in cellular priorities (such as energy conservation, reduced metabolic load, and adaptive response) during catechol induced stress. Figure 2.9 represents more details about the individual genes involvement in the particular two-component system with highlighting their expression level (upregulation/downregulation).



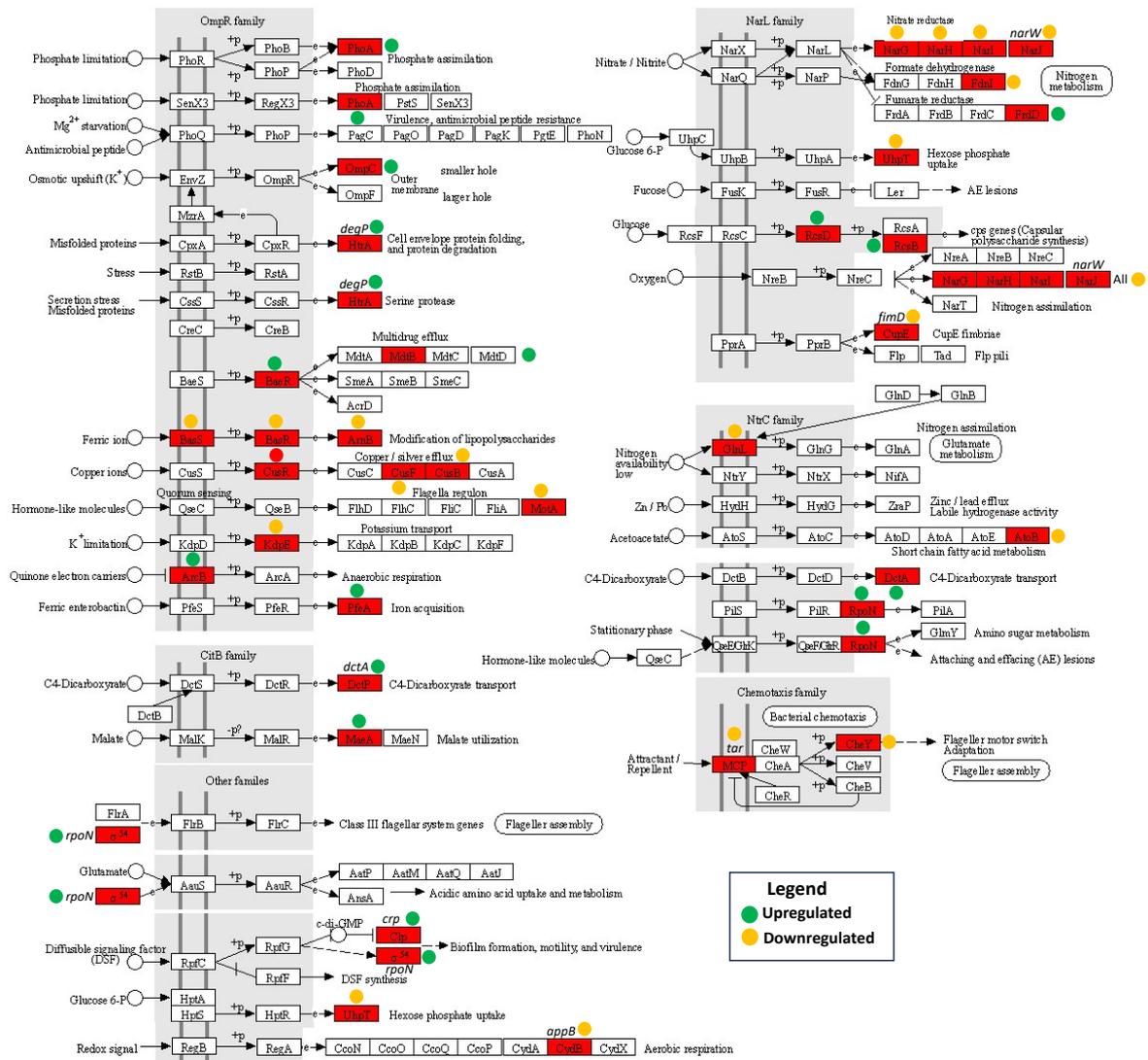
**Figure 2.6.** Heatmap showing gene expression changes associated with two component systems induced by catechol exposure under anaerobic condition. The vertical color scale at the right side of the heat map represents the intensity of data values. The highest level of expression is shown by 'Red' color, and the lowest level is represented by 'Navy' color. Adjusted p-values are displayed beside the gene name.



**Figure 2.7.** Heatmap showing gene expression changes associated with two component systems induced by catechol exposure under aerobic condition. The vertical color scale at the right side of the heat map represents the intensity of data values. The highest level of expression is shown by 'Red' color, and the lowest level is represented by 'Navy' color. Adjusted p-values are displayed beside the gene name.



**Figure 2.8.** Effect of catechol exposure on two component systems under anaerobic condition. Differentially expressed genes were determined using Deseq2 analysis comparing treated and control groups. Green, orange, and red color indicate upregulated, downregulated, and expressed genes respectively.

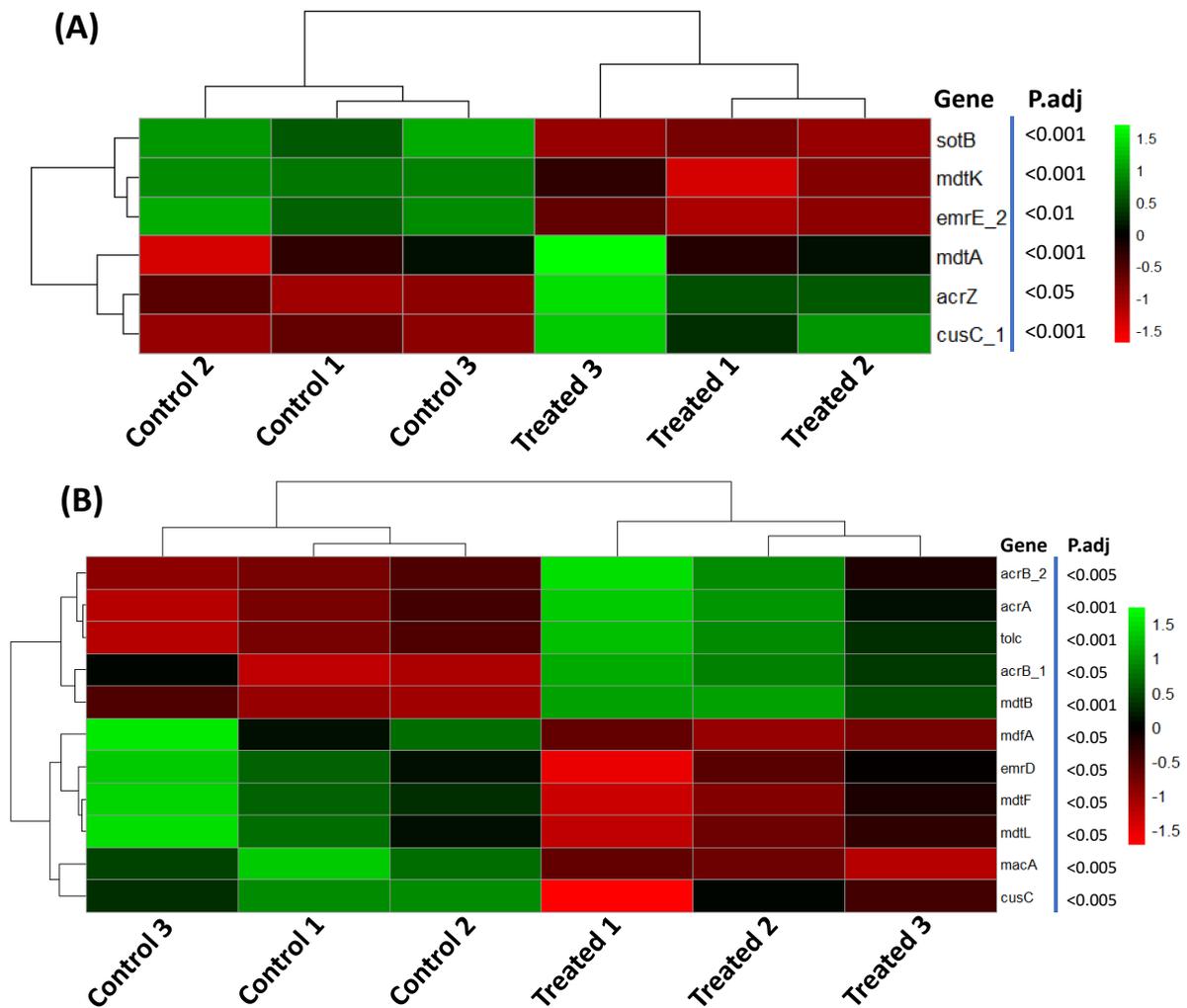


**Figure 2.9.** Effect of catechol exposure on two component systems under aerobic condition. Differentially expressed genes were determined using Deseq2 analysis comparing treated and control groups. Green, orange, and red color indicate upregulated, downregulated, and expressed genes respectively.

### 2.3.2.2. Effects of catechol on efflux pump

Efflux pumps are transmembrane protein systems that extrude substances from cellular interior to the external environment (Pages et al., 2011; Pagès & Amaral, 2009). In the present study, catechol stress significantly affected the expression levels of genes involved in efflux pumps (Figure 2.10). In anaerobic condition, catechol treatment led to the upregulation of five genes, and the downregulation of three genes associated with the efflux pump. Catechol stress induced the major multiple drug efflux system, including AcrAB-TolC, Cus, MdtABC-TolC, and MdtK. Genes such as *acrZ*, *mdtA*, and *cusC* were upregulated by 0.72, 1.09, and 1.34 log<sub>2</sub>

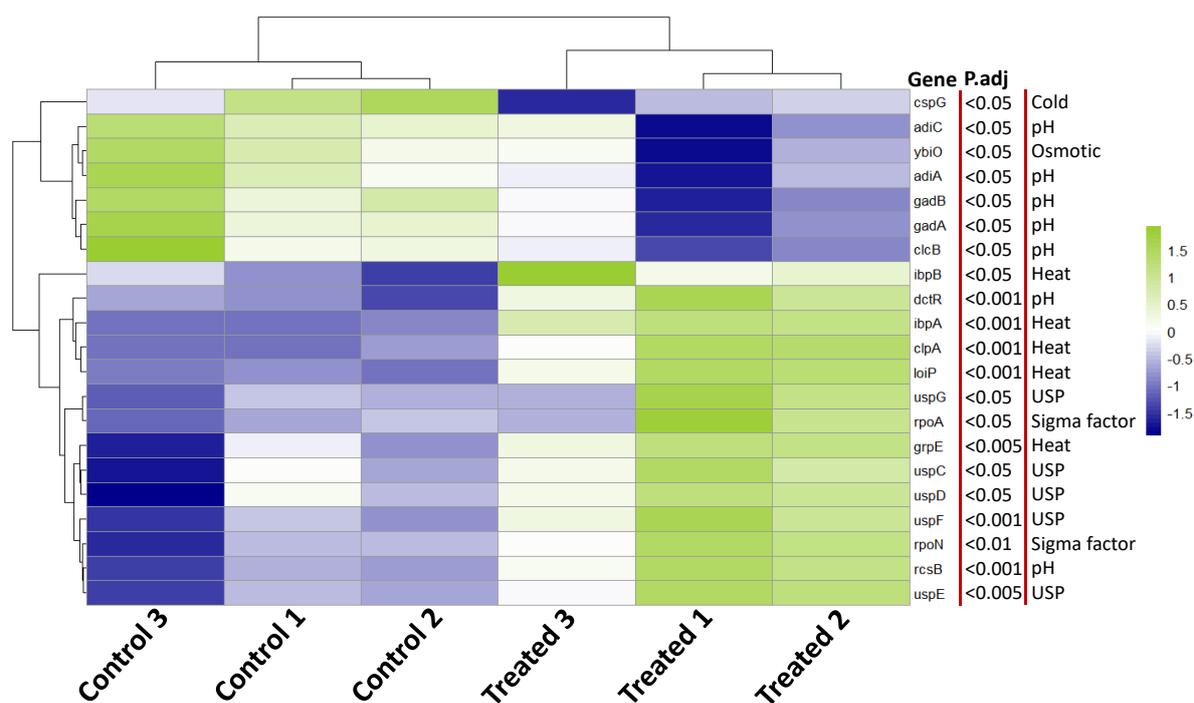
fold, respectively, and *emrE\_2*, *sotB*, and *mdtK* genes were downregulated by -0.76, -0.96, and -0.44 log<sub>2</sub> fold, respectively. Conversely, under aerobic condition, catechol caused upregulation of four genes (*acrA*, *acrB\_1*, *acrB\_2*, and *mdtB*) and downregulation of eight genes (*cusC*, *emrD*, *kefC*, *macA*, *mdfA*, *mdtF*, *mdtL*, and *tolC*) involved in efflux pump.



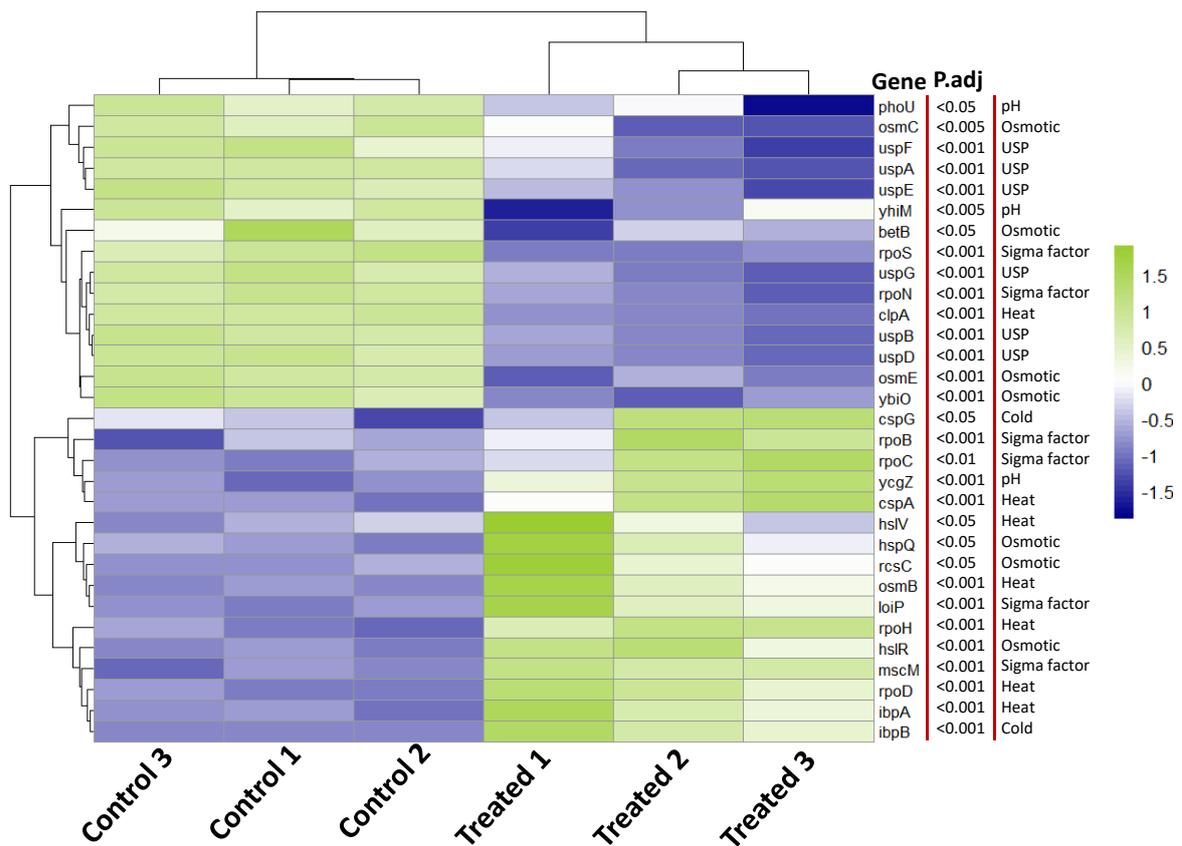
**Figure 2.10.** Heatmap showing gene expression changes associated with efflux pump induced by catechol exposure under both anaerobic (A) and aerobic (B) conditions. The vertical color scale at the right side of the heat map represents the intensity of data values. The highest level of expression is shown by 'Green' color, and the lowest level is represented by 'Red' color. Adjusted p-values are displayed beside the gene name.

### 2.3.2.3. Impact of catechol exposure on stress responsive genes

In the presence of catechol, the expression of a number of genes associated with different stress responses (including osmotic, pH, cold, and heat) was altered under both aerobic and anaerobic conditions (Figure 2.11 & Figure 2.12). Also, the USP (universal stress proteins) gene family was modulated due to catechol treatment in both conditions. *USP* genes exhibited differential expression, being upregulated in aerobic condition, and downregulated under anaerobic condition. Genes related to cold shock were upregulated in anaerobic condition and downregulated in aerobic condition. All the heat shock genes (except *clpA*) showed enhanced expression in both conditions. Under aerobic and anaerobic conditions, sigma factor-associated genes displayed elevated expression except for the *rpoN* and *rpoS* genes, which showed downregulation specifically in anaerobic condition. The genes associated with stress responses such as osmotic and pH, exhibited a varied expression pattern observed in both conditions.



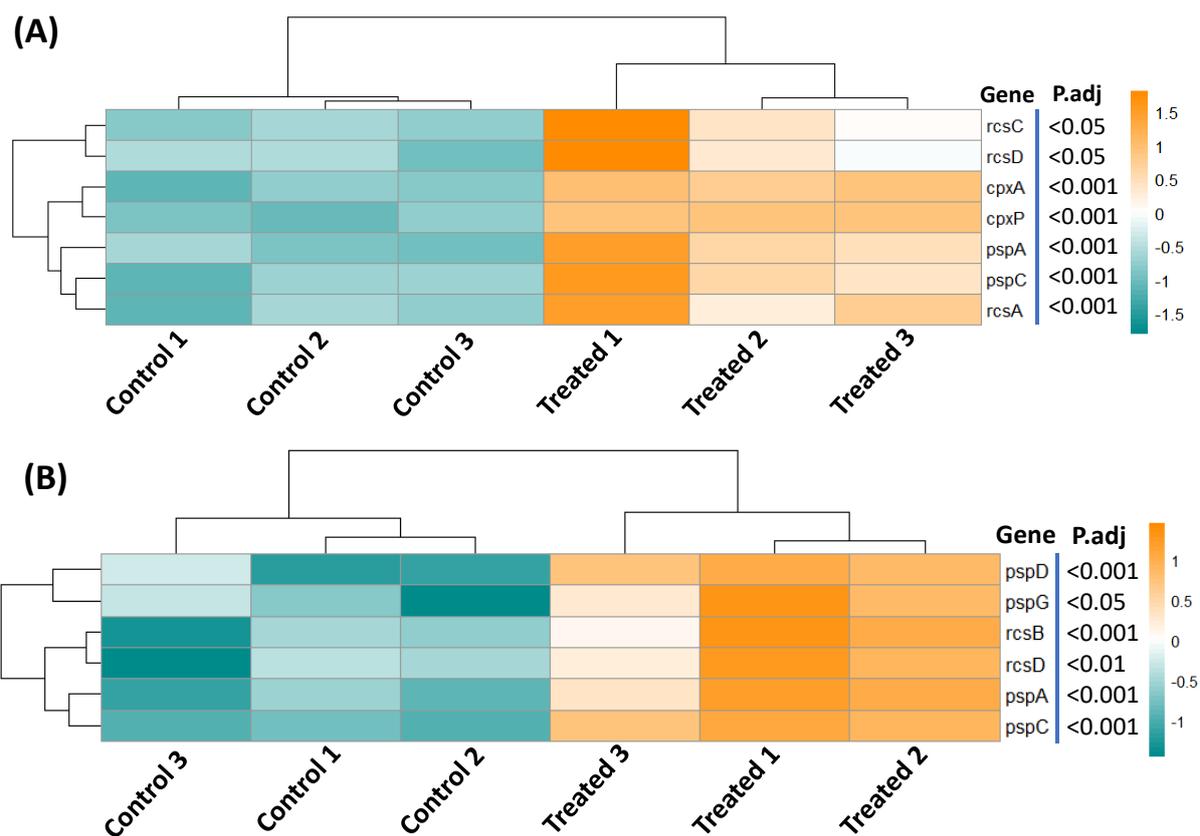
**Figure 2.11.** Heatmap showing exposure of catechol induced expression of different stress responsive genes under aerobic condition. The vertical color scale at the right side of the heat map represents the intensity of data values. The highest level of expression is shown by 'yellowgreen' color (100%), and the lowest level is represented by 'darkblue' color (100%). Adjusted p-values are displayed beside the gene name.



**Figure 2.12.** Heatmap showing exposure of catechol induced expression of different stress responsive genes under anaerobic condition. The vertical color scale at the right side of the heat map represents the intensity of data values. The highest level of expression is shown by 'yellowgreen' color (100%), and the lowest level is represented by 'darkblue' color (100%). Adjusted p-values are displayed beside the gene name.

#### 2.3.2.4. Catechol causes membrane damage

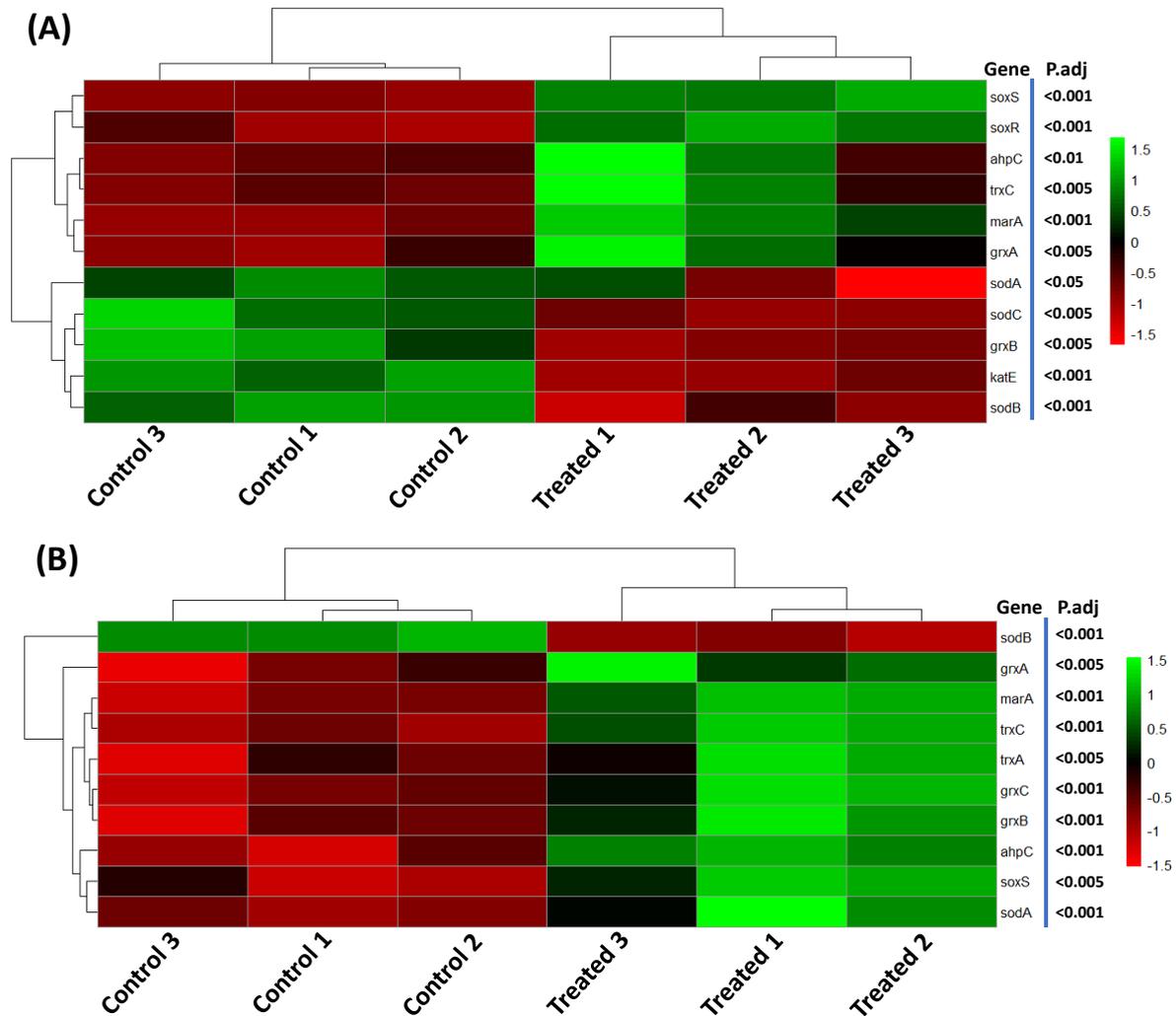
PSP (phage shock protein) system is involved in membrane stress, typically when bacterial cells are infected by bacteriophages (Joly et al., 2010). RCS system is a regulatory network involved in capsule synthesis and responds to environmental conditions such as envelope stress, and osmotic stress (J. Meng et al., 2021; Pando et al., 2017). DEGs analysis revealed that both PSP and RCS systems were altered due to catechol treatment under both aerobic and anaerobic conditions (Figure 2.13).



**Figure 2.13.** Heatmap showing gene expression changes associated with membrane damage induced by catechol exposure under both anaerobic (A) and aerobic (B) conditions. The vertical color scale at the right side of the heat map represents the intensity of data values. The highest level of expression is shown by ‘orange’ color (100%), and the lowest level is represented by ‘darkcyan’ color (100%). Adjusted p-values are displayed beside the gene name.

Moreover, catechol exposure changed the expression level of genes related to solvent tolerance, reactive oxygen species, detoxification, and cellular redox balance (Figure 2.14). Solvent tolerance-related genes such as *marA* and *soxS* showed increased expression in both conditions. Superoxide dismutase (SOD) enzyme catalyzes the conversion of superoxide radicals ( $O_2^-$ ) to hydrogen peroxide ( $H_2O_2$ ) and oxygen (Fridovich, 1997; Netto et al., 1996; Zhao et al., 2021). SOD encoding genes were downregulated during anaerobic condition, whereas only the *sodA* gene exhibited upregulation in aerobic condition. Catalase and alkyl hydroperoxide enzymes are involved in the detoxification system and protect cells from oxidative damage (Imlay & Linn, 1988). The expression of the alkyl hydroperoxide-encoding gene *ahpC* was upregulated, while the catalase enzyme-encoding gene *katE* was either downregulated (during anaerobic condition) or not significant (during aerobic condition).

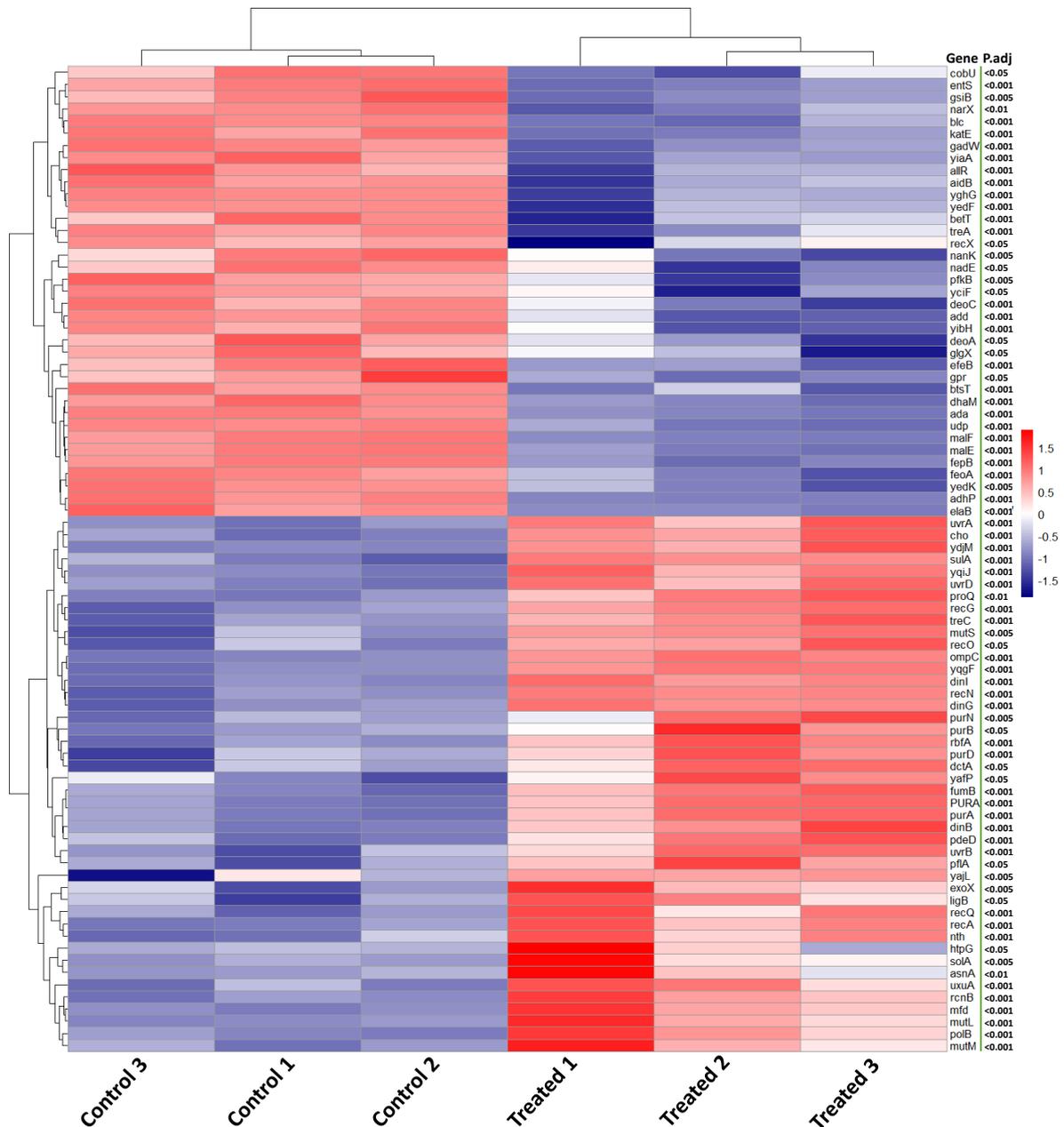
Thioredoxin (Trx) and glutaredoxin (Grx) play a key role in controlling cellular redox balance (Holmgren, 2001). Expression of all thioredoxin and glutaredoxin-encoding genes (*grxA*, *grxB*, *grxC*, and *trxC*) exhibited upregulation in aerobic condition, however under anaerobic condition, *trxC* and *grxB* genes showed upregulation and downregulation respectively.



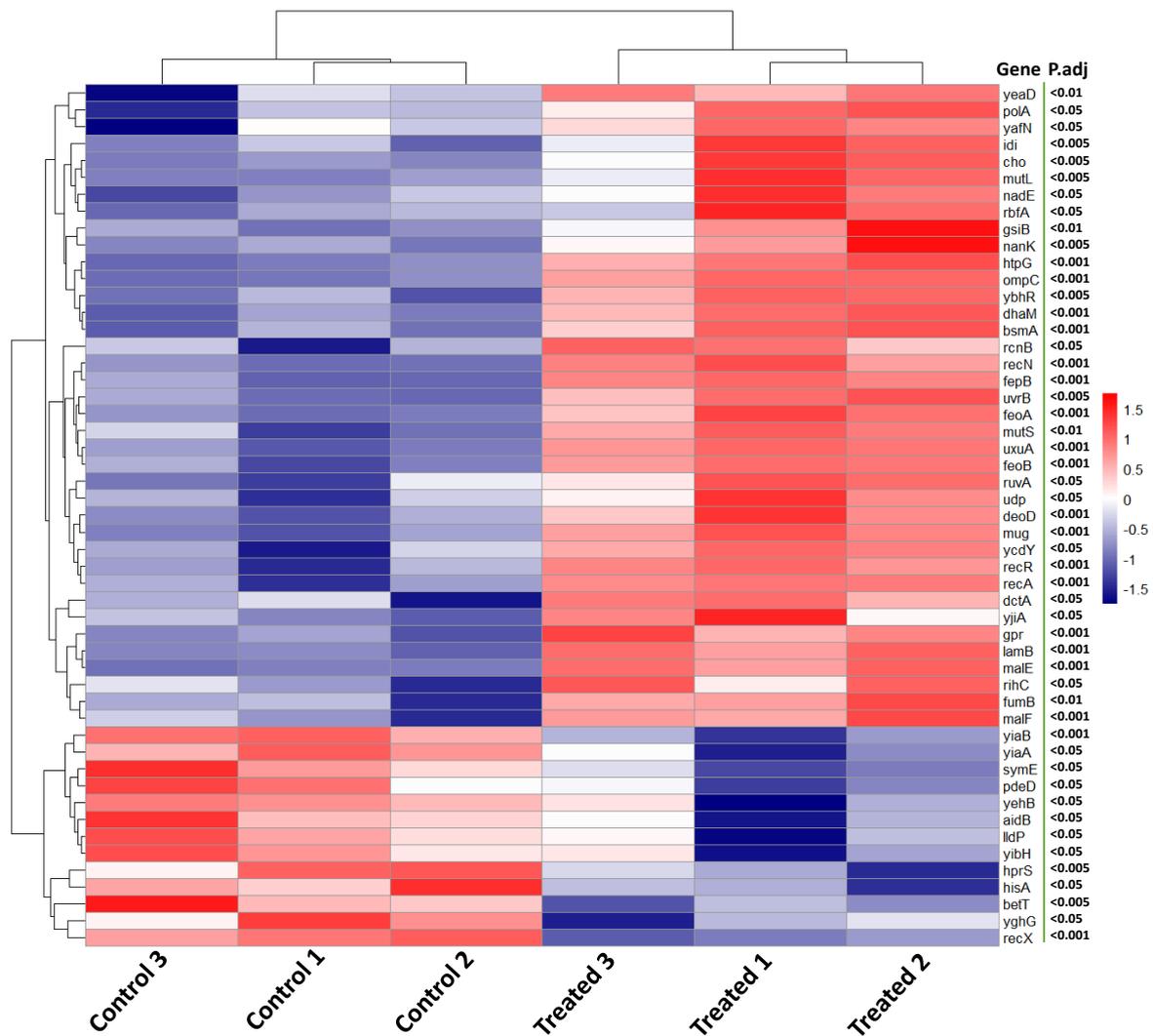
**Figure 2.14.** Visual representation of catechol induced gene expression changes associated with solvent tolerance, reactive oxygen species, detoxification, and cellular redox balance under both anaerobic (A) and aerobic (B) conditions. The vertical color scale at the right side of the heat map represents the intensity of data values. The highest level of expression is shown by 'Green' color, and the lowest level is represented by 'Red' color. Adjusted p-values are displayed beside the gene name.

### 2.3.2.5. Effect of Catechol on DNA damage

According to the DEGs analysis catechol exposure elicited DNA damage in *E. coli* UMI46, leading to alterations in 86 genes (comprising 45 genes upregulated and 41 genes downregulated) under anaerobic condition (Figure 2.15), and 55 genes (40 genes upregulated and 15 genes downregulated) during aerobic condition (Figure 2.16).



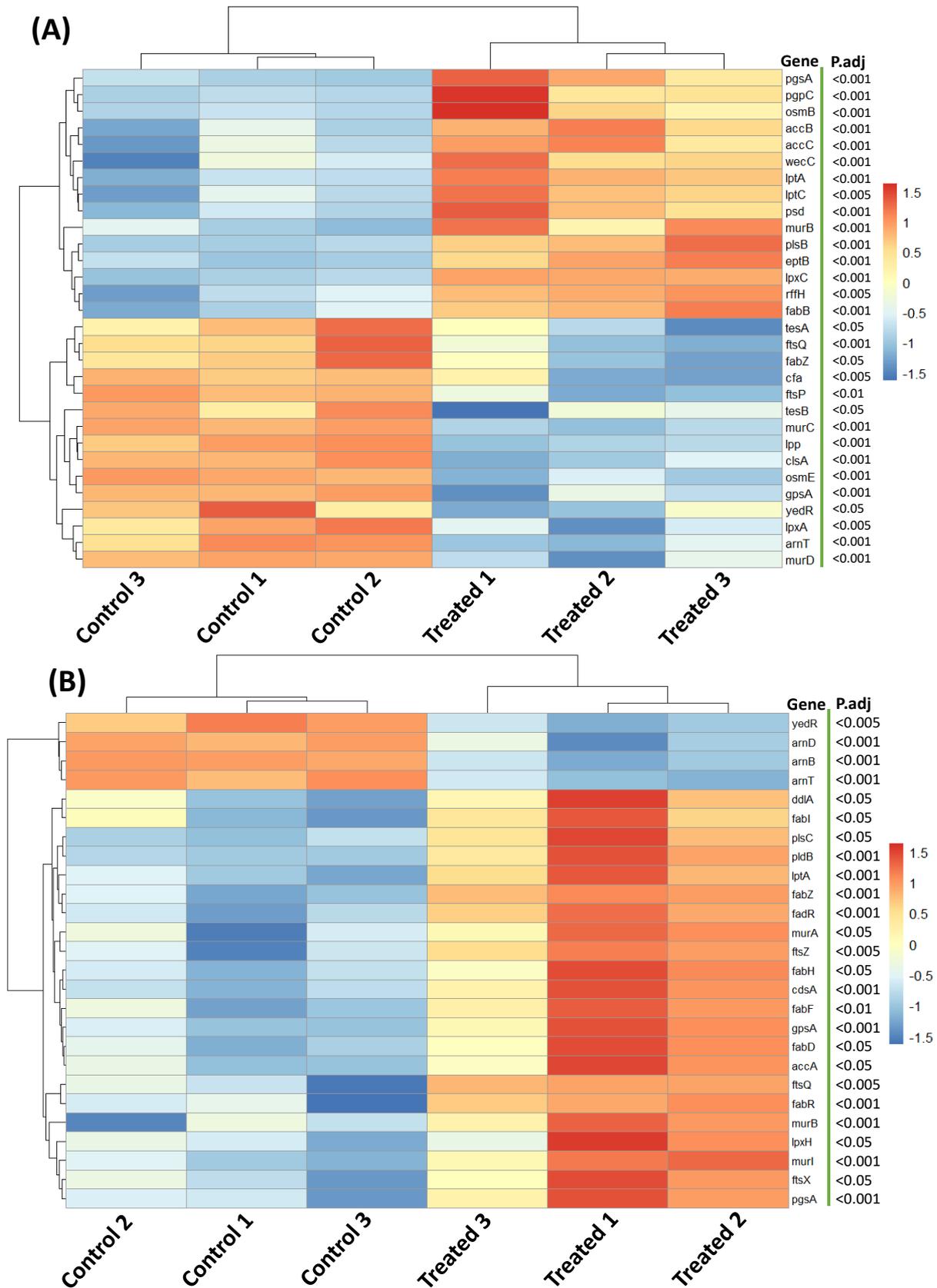
**Figure 2.15.** Heatmap showing gene expression changes associated with DNA damage induced by catechol exposure under anaerobic condition. The vertical color scale at the right side of the heat map represents the intensity of data values. The highest level of expression is shown by 'Red' color (100%), and the lowest level is represented by 'Navy Blue' color (100%). Adjusted p-values are displayed beside the gene name.



**Figure 2.16.** Heatmap showing gene expression changes associated with DNA damage induced by catechol exposure under aerobic condition. The vertical color scale at the right side of the heat map represents the intensity of data values. The highest level of expression is shown by 'Red' color (100%), and the lowest level is represented by 'Navy Blue' color (100%). Adjusted *p*-values are displayed beside the gene name.

Under catechol exposure, *E. coli* UMI46 made a strategic response to mitigate DNA damage by activating the DNA mismatch repair system. Based on DEGs analysis, all the genes (except *xseA*) were upregulated in the DNA mismatch repair pathway during anaerobic and aerobic conditions respectively (Figure 2.17).



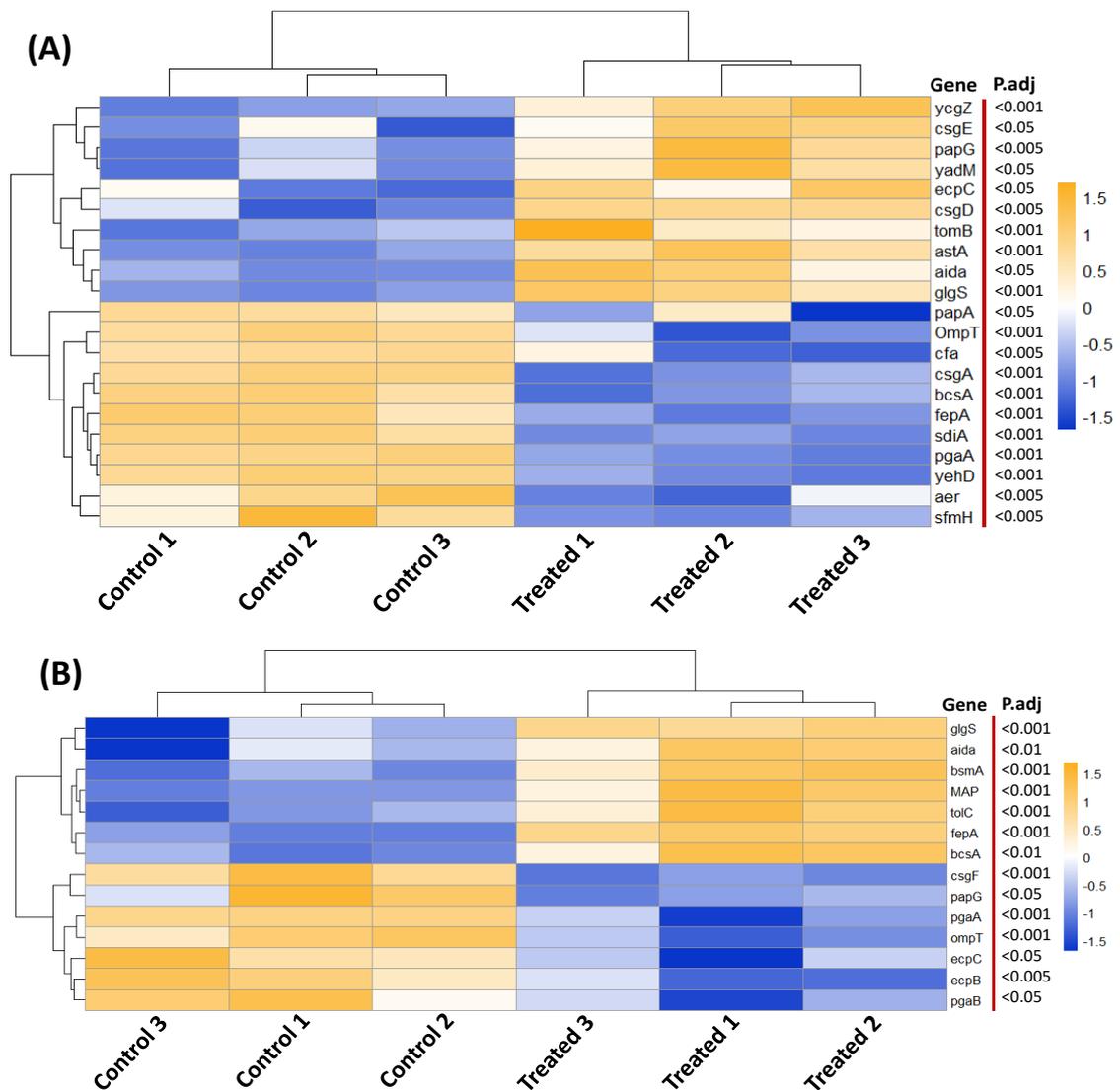


*Figure 2.18. Heatmap showing gene expression changes associated with cell cycle control, division, and cell wall biosynthesis induced by catechol exposure under anaerobic (A)*

and aerobic (B) conditions. The vertical color scale at the right side of the heat map represents the intensity of data values. The highest level of expression is shown by 'Red' color, and the lowest level is represented by 'Blue' color. Adjusted p-values are displayed beside the gene name.

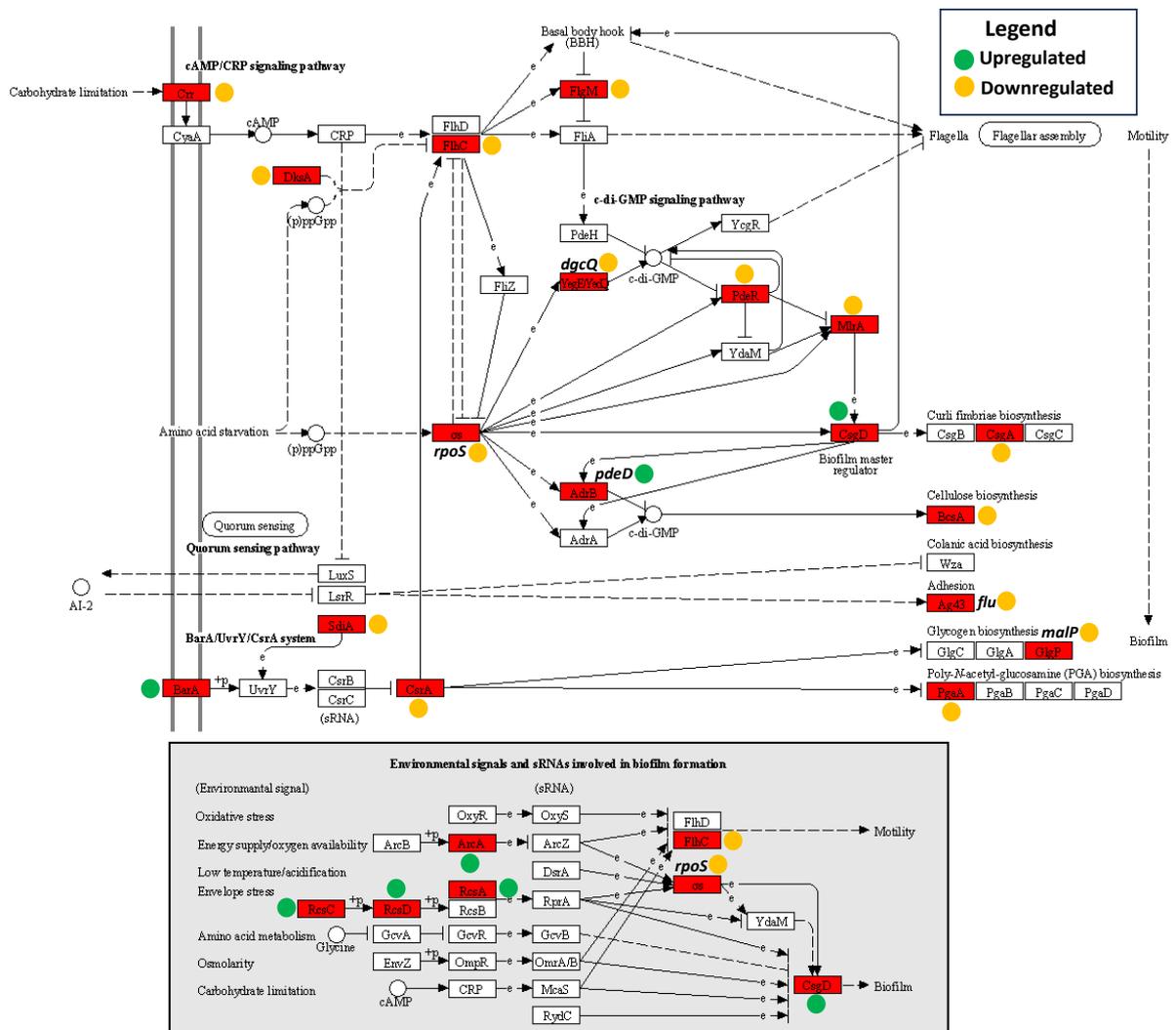
### 2.3.2.6. Impact of catechol exposure on virulence factor, biofilm formation and flagellar assembly

A set of genes related to virulence factors were modulated due to catechol treatment. The differential expression analysis revealed that catechol exposure causes a transcriptional change of 22 genes in anaerobic condition, with 10 genes upregulated and 12 genes downregulated (Figure 2.19A). Similarly, during aerobic condition, 14 genes exhibited altered expression patterns, with 6 genes upregulated and 8 genes downregulated (Figure 2.19B).

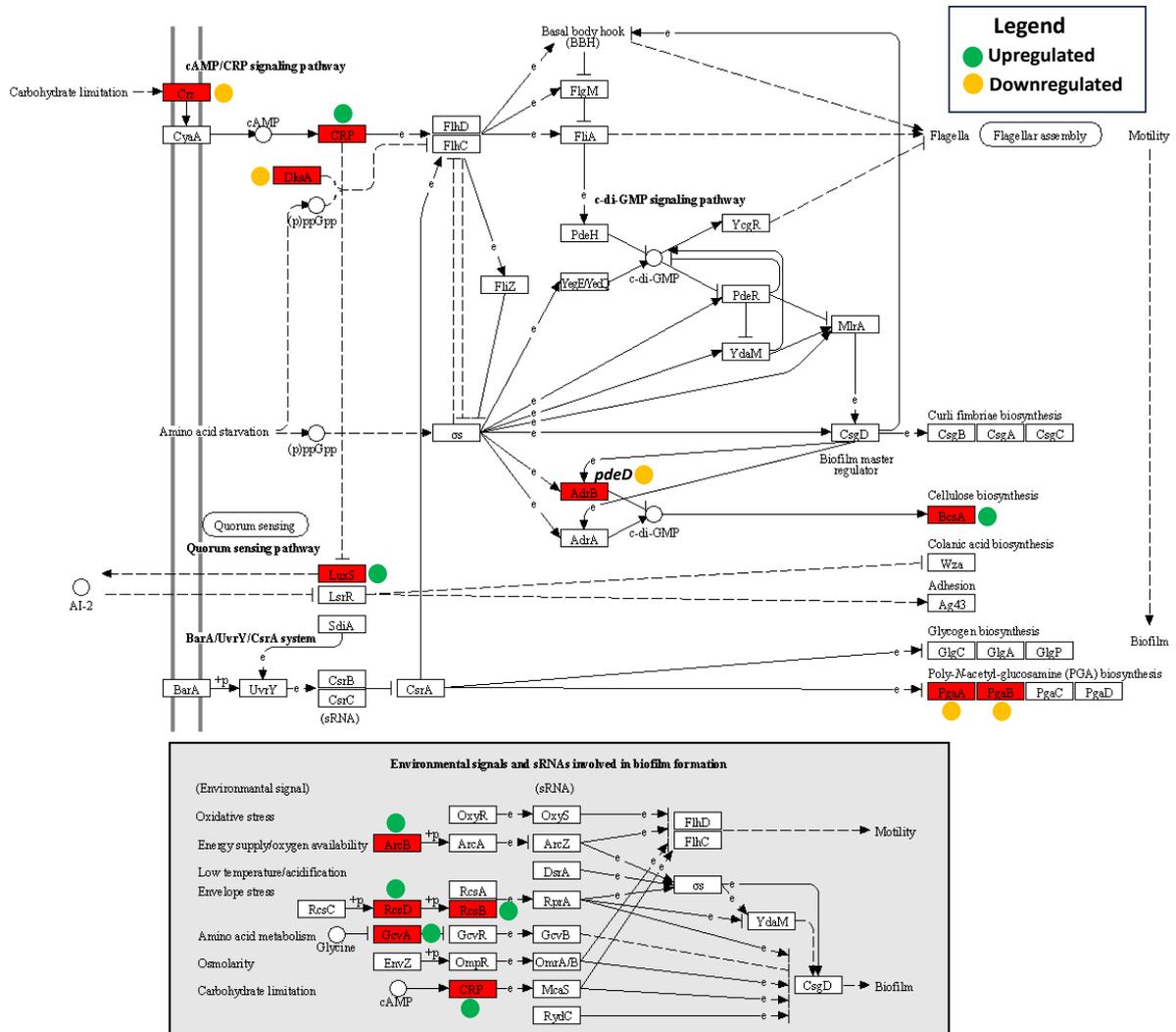


**Figure 2.19.** Heatmap showing gene expression changes associated with virulence factor induced by catechol exposure under both anaerobic (A) and aerobic (B) conditions. The vertical color scale at the right side of the heat map represents the intensity of data values. The highest level of expression is shown by 'Orange' color, and the lowest level is represented by 'Royal Blue (#0935ca)' color. Adjusted *p*-values are displayed beside the gene name.

Impact of catechol exposure on biofilm formation illustrated in Figure 2.20 (anaerobic) and Figure 2.21 (aerobic). The upregulation of *rscC*, *rscD*, *rscA*, and *csgD* genes signifies that catechol exposure causes envelope stress which leads to enhanced biofilm formation in anaerobic condition (Figure 2.20). In contrast to anaerobic condition, there was no influence of catechol exposure on the biofilm formation under aerobic condition (Figure 2.21).



**Figure 2.20.** Effect of catechol exposure on biofilm formation under anaerobic condition. Differentially expressed genes were determined using Deseq2 analysis comparing treated and control groups. Green, orange, and red color indicate upregulated, downregulated, and expressed genes respectively.



**Figure 2.21.** Effect of catechol exposure on biofilm formation under aerobic condition. Differentially expressed genes were determined using Deseq2 analysis comparing treated and control groups. Green, orange, and red color indicate upregulated, downregulated, and expressed genes respectively.

Flagella are whip-like appendages that are involved in bacterial motility (Haiko & Westerlund-Wikström, 2013). Catechol exposure results in the downregulation of genes related to the flagellar assembly under anaerobic condition (Figure 2.20), while there were no changes observed during aerobic condition (Figure 2.21).

### 2.3.3. Determination of potential metabolites involved in catechol detoxification pathway

Raw data of each control and treated samples were obtained from LC-MS/MS, and analyzed in Metaboanalyst, XCMS, MZmine2, MS-DIAL, and SIRIUS. A total of 9 metabolites were detected (via LC-MS/MS) involved in the catechol detoxification pathway in both conditions (Table 2.1). All the metabolites identified using LC-MS/MS were further confirmed with either commercially available analytical standards or fragmentation pattern comparison. GC-MS analysis determined a total of 12 metabolites in aerobic condition, and 4 metabolites in anaerobic condition (Table 2.1). However, out of 24 detoxification products, only one compound, named quinol was identified using NMR. Metabolites and their fragmentation pattern comparison were shown in S Figure 2.15-2.24.

**Table 2.1.** List of detoxification products determined in treated samples under aerobic and anaerobic conditions.

Metabolite Name	Condition	Analytical Instrument	SMILES	RT	p-value
Phenol	A/An	LC-MS/MS	<chem>Oc1ccccc1</chem>	3.58	<0.01
Quinol	A	LC-MS/MS	<chem>OC1=CC=C(O)C=C1</chem>	3.81	<0.01
Pyrogallol	A/An	LC-MS/MS	<chem>Oc1cccc(O)c1O</chem>	2.38	<0.01
4-Hydroxybenzoate	A/An	LC-MS/MS	<chem>Oc1ccc(cc1)C([O-])=O</chem>	4.4	<0.01
Protocatechuic aldehyde	A/An	LC-MS/MS	<chem>C1=CC(=C(C=C1C=O)O)O</chem>	4.42	<0.01
2,4-Dihydroxyacetophenone	A/An	LC-MS/MS	<chem>CC(=O)C1=C(C=C(C=C1)O)O</chem>	7.04	<0.01
Protocatechuate	A/An	LC-MS/MS	<chem>C1=CC(=C(C=C1C(=O)O)O)O</chem>	3.576	<0.01
4-Nitrocatechol	A/An	LC-MS/MS	<chem>C1=CC(=C(C=C1[N+](=O)[O-])O)O</chem>	5.71	<0.01
Vanillate	A	LC-MS/MS	<chem>COC1=C(C=CC(=C1)C(=O)O)[O-]</chem>	5.31	<0.01
Dioxydiphenol	A/An	LC-MS/MS	<chem>C1=CC=C(C(=C1)O)OOC2=CC=CC=C2O</chem>	6.63	<0.01
3-Methyl catechol	A/An	LC-MS/MS	<chem>CC1=C(C(=CC=C1)O)O</chem>	6.4	<0.01
4-Ethoxybenzoic acid ethyl ester	A	GC-MS	<chem>CCOC1=CC=C(C=C1)C(=O)OCC</chem>	18.34	<0.01
O-valeryl catechol	A	GC-MS	<chem>CCCCC(=O)OC1=CC=CC=C1O</chem>	18.06	<0.01
4-Ethoxyphenol	A	GC-MS	<chem>CCOC1=CC=C(C=C1)O</chem>	16.80	<0.01
2-Methoxy-1,3-benzodioxole	A	GC-MS	<chem>COC1OC2=CC=CC=C2O1</chem>	17.43	<0.01
2-Hydroxymethyl-1,4-benzodioxan	A	GC-MS	<chem>C1C(OC2=CC=CC=C2O1)CO</chem>	17.89	<0.01
2-Coumaranone	A	GC-MS	<chem>C1C2=CC=CC=C2OC1=O</chem>	16.83	<0.01
2-Phenoxyphenol	A/An	GC-MS	<chem>C1=CC=C(C=C1)OC2=CC=CC=C2O</chem>	18.48	<0.01
Methyl 1,3-benzodioxole-2-carboxylate	A	GC-MS	<chem>COC(=O)C1OC2=CC=CC=C2O1</chem>	17.05	<0.01
Ethyl Vanillin	A/An	GC-MS	<chem>CCOC1=C(C=CC(=C1)C=O)O</chem>	17.47	<0.05
2-Methoxyphenol	A	GC-MS	<chem>COC1=CC=CC=C1O</chem>	14.32	<0.01
2'-Hydroxyacetophenone	A/An	GC-MS	<chem>CC(=O)C1=CC=CC=C1O</chem>	15.29	<0.01
Catechol Carbonate	A/An	GC-MS	<chem>O=C1OC2=CC=CC=C2O1</chem>	14.76	<0.01
2-propoxyphenol	An	GC-MS	<chem>CCOC1=CC=CC=C1O</chem>	16.40	<0.01

\* p-values obtained from Welch's t-test after comparing peak area of treated and control samples. Abbreviation: A - Aerobic, An – Anaerobic; RT – Retention time; Treated – *E. coli UMI46* strain grew on glucose and catechol containing minimal media; Control – *E. coli UMI46* strain grew on glucose containing minimal media (no catechol).

#### **2.3.4. Polyphenol oxidase assay**

Enzymatic reactions of catechol substrate and polyphenol oxidase enzyme resulting in the formation of phenol, mequinol, and 2-Methyl-4-propylphenol, identified via GC-MS compared with control.

#### **2.3.5. Unveiling of genes involved in catechol detoxification pathway**

Genes that were involved in each step of the catechol detoxification pathway (Figure 2.30 & Figure 2.31), were reported either from published articles or stated in ECMD2.0, BRENDA, Rhea, and Uniprot databases. In case of the absence of queried gene product in the *E. coli UMI46* strain, BLAST (Basic Local Alignment System Tools) was conducted to find the relevant sequences, followed by evaluating the corresponding domain.

##### **2.3.5.1. *ubiD* gene is responsible for decarboxylase activity in *E. coli UMI46* strain**

Protocatechuate decarboxylase enzyme catalyzes the conversion of catechol to protocatechuate (He & Wiegel, 1996). Protocatechuate decarboxylase is annotated to 55% identical to gallate decarboxylase (*Lactobacillus plantarum*) (Marshall, Payne, et al., 2017). Protocatechuate decarboxylase along with gallate decarboxylase, phenolic acid decarboxylase, and vanillic acid decarboxylase enzymes shares the same *ubiD* domain, which is part of the 3-octaprenyl-4-hydroxybenzoate decarboxylase enzyme (Figure 2.22). NCBI protein accession numbers F9US27 and P94405 reported that decarboxylase enzymes (such as gallate decarboxylase, protocatechuate decarboxylase, phenolic acid decarboxylase, 4-hydroxybenzoate decarboxylase, and vanillic acid decarboxylase) encoded by *ubiD* gene. Therefore, *ubiD* gene encoding 3-octaprenyl-4-hydroxybenzoate decarboxylase enzyme is proposed to be catalyzing the catechol to protocatechuate step, phenol to 4-hydroxybenzoate and 2-methoxyphenol to vanillate steps respectively. Gallate decarboxylase enzyme isolated from *Lactobacillus plantarum*, reported to be an indistinguishable *ubiD* like protein which have decarboxylase activity (Jiménez et al., 2013). *UbiD* family decarboxylase enzyme from *E. coli UMI46* strain is 26%, 31% and 100% identical to protocatechuate decarboxylase, phenolic acid decarboxylase and 3-octaprenyl-4-hydroxybenzoate decarboxylase respectively (Table 2.2). However, protocatechuate decarboxylase (*Klebsiella aerogenes*) shows 99.60% sequence

identity with the UbiD family decarboxylase enzyme (*E. coli* strain 1571837), while phenolic acid decarboxylase (*E. coli* O157:H7) exhibits 65% sequence identity with UbiD family decarboxylase (*E. coli* strain 88504). Protocatechuate decarboxylase and phenolic acid decarboxylase retrieved from other species such as *Pseudomonas*, *klebsiella*, and *Bacillus*, indicates ~28-30% sequence identity to UbiD family decarboxylase enzyme of *E. coli* UM146. All these sequences were retrieved from very divergent species compared to *E. coli* UM146, therefore the similarity was low although they have the same UbiD domain.

**Table 2.2.** List of enzymes and their identity with the 3-octaprenyl-4-hydroxybenzoate carboxylase enzyme of *E. coli* UM146 strain.

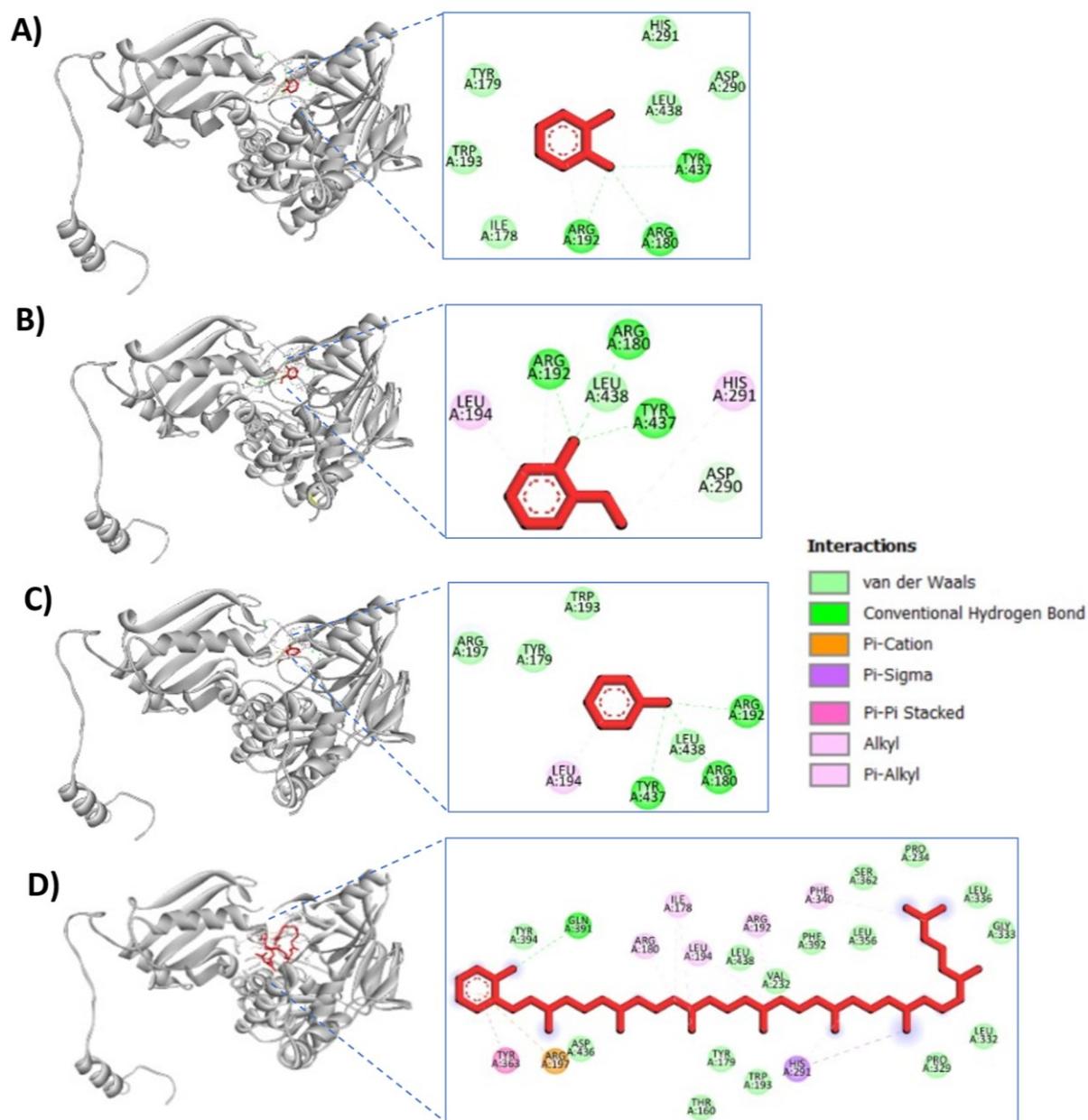
Enzyme name	Species	Accession	Identity
Gallate decarboxylase/Protocatechuate decarboxylase	<i>Lactiplantibacillus plantarum</i> WCFS1	F9US27	26%
Phenolic acid decarboxylase/ Vanillate decarboxylase/ 4-hydroxybenzoate decarboxylase	<i>Bacillus Subtilis</i>	P94405	31%
ubiD family decarboxylase/3-octaprenyl-4-hydroxybenzoate carboxylase	<i>Shigella sonnei</i> (strain Ss046)	Q3YVC4	100%

\*Similarity / identity threshold: >40%



**Figure 2.22.** Conserved domain comparison among phenolic acid decarboxylase, gallate decarboxylase and 3-Octaprenyl-4-hydroxybenzoate decarboxylase. A) Phenolic acid decarboxylase/Vanillate decarboxylase/4-hydroxybenzoate decarboxylase (Uniprot: P94405); B) Gallate decarboxylase/ Protocatechuate decarboxylase (Uniprot ID: F9US27); and C) 3-

*Octaprenyl-4-hydroxybenzoate decarboxylase (NCBI accession: WP\_000339804). Domain analysis was conducted using NCBI conserved domain.*



**Figure 2.23.** Comparison of substrate binding pocket in ubiD protein (*Escherichia coli* UM 146) with (A) catechol, (B) 2-methoxyphenol, (C) phenol, and (D) 2-octaprenylphenol (positive control). UbiD is shown in silver, substrates in red, residues in blue, and arrows in dark red color. Docking was conducted using *Autodock\_vina\_1\_1\_2\_win32.msi*. The center of the grid map was X (106), Y (70) and Z (102); and the autogrid calculation was set as  $11.594 \times -2.204 \times 0.088 \text{ \AA}$ .

Docking was performed between 3-octaprenyl-4-hydroxybenzoate decarboxylase (UbiD) enzyme and each of the metabolites (separately), including catechol, phenol and 2-methoxyphenol. UbiD enzyme binds to catechol, phenol, 2-methoxyphenol and 2-octaprenylphenol (positive control) with an affinity energy of -4.8 Kcal/mol, -4.6 Kcal/mol, -5.0 Kcal/mol and -7.2 Kcal/mol respectively (Table 2.3). All these substrates including the positive control bind to the same active site of UbiD enzyme with key residues (arginine 180, arginine 192, and aspartate 290) (Figure 2.23). Crystal structure of UbiD enzyme from *Escherichia coli* reported that this enzyme performs its activity through the same active site residues (including, arginine 180, arginine 192, and aspartate 290) (Marshall, Fisher, et al., 2017). UbiD enzyme catalyzes the para-carboxylation of catechol with residues (glutamate 289, arginine 171, arginine 188, and lysine 363) in the presence of prenylated FMN (prFMN) cofactor (Payer et al., 2017). UbiD enzyme undergoes decarboxylation of aromatic substrates (such as conversion of 4-hydroxybenzoate to phenol) by virtue of UbiD domain dynamics motion (Marshall et al., 2021).

### 2.3.5.2. *yfiH* gene encodes polyphenol oxidase in *E. coli* UMI46 strain

*yfiH* gene encodes purine nucleoside phosphorylase which is 82% identical to polyphenol oxidase (Figure 2.24). According to NCBI conserved domain analysis, purine nucleoside phosphorylase has polyphenol oxidase domain (Figure 2.25). Moreover, polyphenol oxidase possesses the domain of *yfiH* similarly to purine nucleoside phosphorylase, which is encoded by the *yfiH* gene. Since purine nucleoside phosphorylase share the same domains and is 82% identical to polyphenol oxidase, it can be inferred that purine nucleoside phosphorylase is the same polyphenol oxidase enzyme in *E. coli* UMI46 strain.

```

polyphenol_oxidase_Escherichia_coli_NYVetLIRN-212      MSKLIVPQWPLPKGVAACSSTRIGGVSLPPYDSLNLGAHCGDNPDHVEENRKRLEAAGNL 60
purine_nucleoside_phosphorylase_Escherichia_coli_UM146 MSKLIVPQWPLPKGVAACSSTRIGGVSLPPYDSLNLGAHCGDNPDHVEENRKRLEAAGNL 60
*****

polyphenol_oxidase_Escherichia_coli_NYVetLIRN-212      PSKPVWLEQVHGKDVLLKLTGEPYASKRADASYSNTPGTVCAMTADCLPVLFCNRAGTEV 120
purine_nucleoside_phosphorylase_Escherichia_coli_UM146 PSKPVWLEQVHGKDVLLKLTGEPYASKRADASYSNTPGTVCAMTADCLPVLFCNRAGTEV 120
*****

polyphenol_oxidase_Escherichia_coli_NYVetLIRN-212      SAAHAGWRGLCAGVLEETVSCFADNPENILAWLGAIGPRAFEVGAEVREAFMAADAKAS 180
purine_nucleoside_phosphorylase_Escherichia_coli_UM146 SAAHAGWRGLCAGVLEETVSCFADNPENILAWLGAIGPRAFEVGAEVREAFMAVDAEAS 180
:*****.***

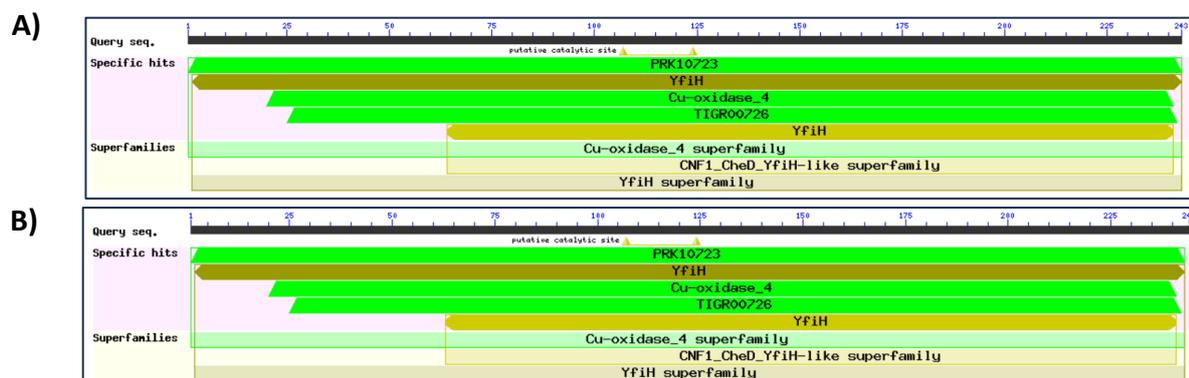
polyphenol_oxidase_Escherichia_coli_NYVetLIRN-212      TAFIQHGDKYLADIYLLARQLASVGVEQIFGGDRCTYTENETFFSYRRDKTTGRMASFI 240
purine_nucleoside_phosphorylase_Escherichia_coli_UM146 TAFIQHGDKYLADIYLLARQLASVGVEQIFGGDRCTYTENETFFSYRRDKTTGRMASFI 240
*****

polyphenol_oxidase_Escherichia_coli_NYVetLIRN-212      WLI 243
purine_nucleoside_phosphorylase_Escherichia_coli_UM146 WLI 243
***

```

**Figure 2.24.** Pairwise alignment comparison between polyphenol oxidase and purine nucleoside phosphorylase. Polyphenol oxidase (Accession: EFI4485732.1) and purine

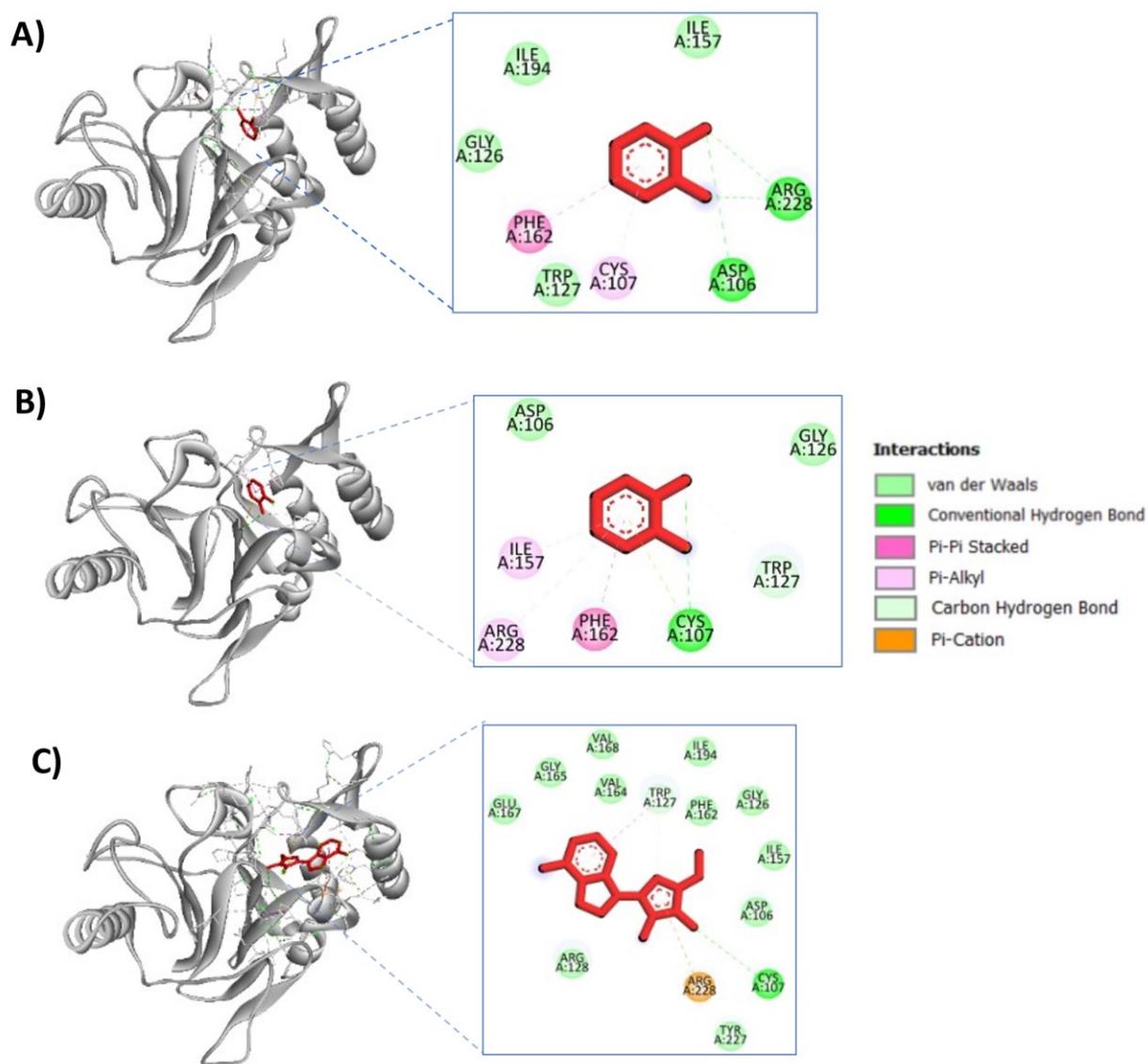
nucleoside phosphorylase from *Escherichia coli* NYVetLIRN-212 and *E. coli* UMI46 respectively. Alignment was conducted using Clustal Omega.



**Figure 2.25.** Conserved domain comparison between polyphenol oxidase and purine nucleoside phosphorylase. (A) polyphenol oxidase (Accession: EFI4485732.1) from *Escherichia coli* NYVetLIRN-212 strain and (B) purine nucleoside phosphorylase from *E. coli* UMI46 strain. Domain analysis was conducted using NCBI conserved domain.

Docking was conducted between catechol and purine nucleoside phosphorylase (Figure 2.26). Purine nucleoside phosphorylase (*E. coli* UMI46) interacts with catechol with an affinity energy of -4.5 kcal/mol (Table 2.3). Likewise, purine nucleoside phosphorylase, catechol exhibits binding to the identical active site of polyphenol oxidase (*Escherichia coli* Str. NYVetLIRN-212) with an affinity energy of -4.0 Kcal/mol (Figure 2.26). Similar to catechol, the positive control, adenosine binds to the same binding pocket of purine nucleoside phosphorylase (Figure 2.26). Phenylalanine162 is identified as the active site in both purine nucleoside phosphorylase and polyphenol oxidase, consistent with the prior findings indicating that phenylalanine261 serve as the catalytic site of catechol oxidase (synonym of polyphenol oxidase) (Siegbahn, 2004).

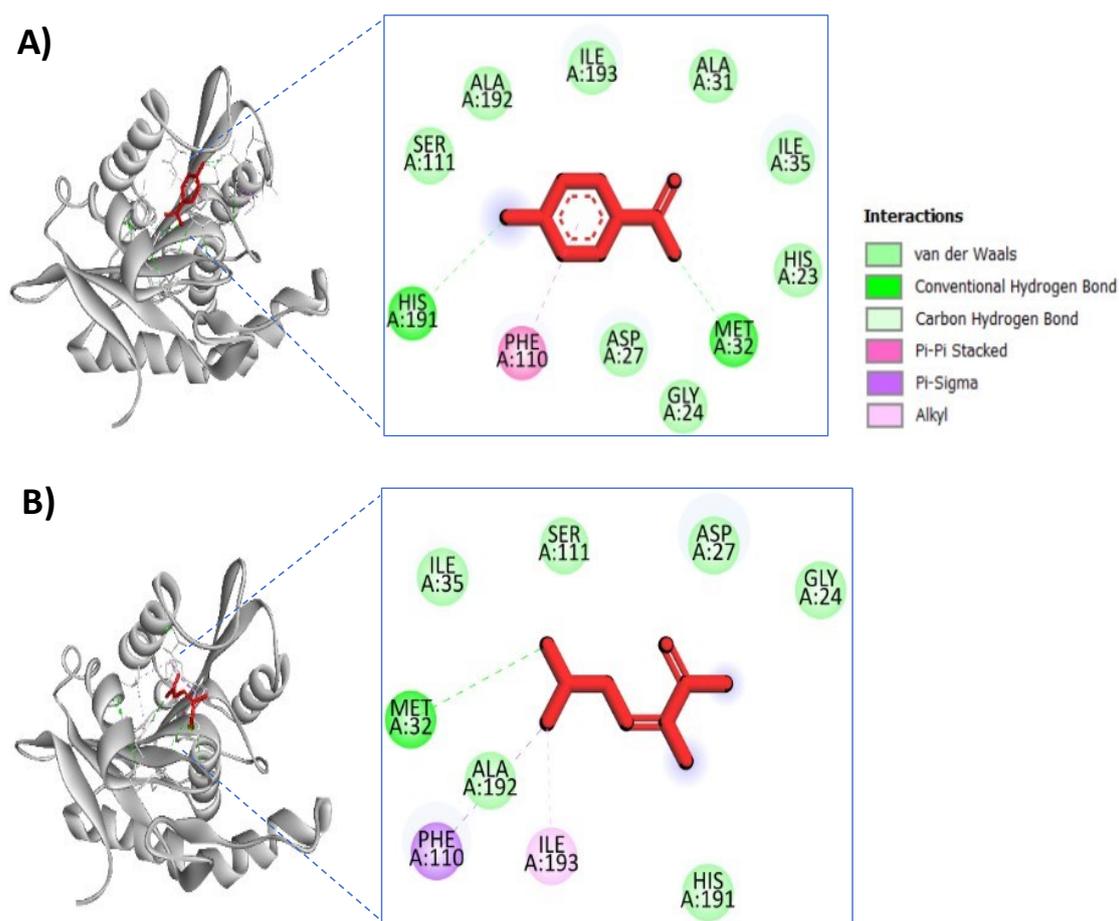
Polyphenol oxidase enzyme catalyzes the oxidation of catechol to dibenzo[1,4]dioxin-2,3-dione (Nair & Vining, 1964). There is no presence of dibenzo[1,4]dioxin-2,3-dione compound in this experiment, rather than a related dioxin compound named 2-hydroxymethyl-1,4-benzodioxan identified, which is proposed to be produced from catechol via the same enzymatic mechanism. It is also evident that polyphenol oxidase has tyrosine hydroxylase activity (performs identical function as phenol hydroxylase) (Wu et al., 2017). In the present study, the polyphenol oxidase assay with catechol resulted in phenol. Therefore, *yfiH* gene encoding polyphenol oxidase enzyme is proposed to be catalyzing the catechol to phenol step.



**Figure 2.26.** Comparison of substrate binding pocket of purine nucleoside phosphorylase and polyphenol oxidase. (A) & (C) Representation of binding interaction of purine nucleoside phosphorylase (*Escherichia coli* UM 146) with catechol and adenosine (positive control) respectively, (B) Representation of binding interaction of polyphenol oxidase (*Escherichia coli* Str. NYVetLIRN-212) with catechol. Proteins are shown in silver, substrates in red, residues in blue, and arrows in dark red color. Docking was conducted using *Autodock\_vina\_1\_1\_2\_win32.msi*. The center of the grid map was X (106), Y (104) and Z (106); and the autogrid calculation was set as  $-0.047 \times -1.566 \times 6.053 \text{ \AA}$ .

### 2.3.5.3. 4-hydroxybenzoate to 4-ethoxybenzoic acid ethyl ester step

Several studies reported that esters groups attached to the aromatic compounds can be hydrolyzed by the esterase enzyme (Bornscheuer, 2002; Saghir et al., 2022). 4-Nitrophenylacetate is converted into p-nitrophenol by esterase enzyme (Saghir et al., 2022). Esterase from *Enterobacter cloacae* catalyzes the conversion of 4-hydroxybenzoate into parabens (Valkova et al., 2003). Therefore, in the present study, esterase enzyme (encoded by the *ypfH* gene) is proposed to be catalyzing the conversion of 4-hydroxybenzoate to 4-ethoxybenzoic acid ethyl ester (Ethyl paraben). Docking was performed between esterase enzyme and substrates, including 4-hydroxybenzoate and 3-(acetylthio)isobutyric acid (positive control) (Figure 2.27). Both substrates (4-hydroxybenzoate and 3-(acetylthio)isobutyric acid) bind to the same binding pocket of esterase enzyme with an affinity energy of -4.2 Kcal/mol and -3.4 Kcal/mol respectively (Table 2.4). Aspartate27, histidine23, Serine111, and histidine191 are the active site residues of esterase enzyme which perform the hydrolytic cleavage activity. Ser-His-Asp catalytic triad is reported to be the active site of esterase enzyme (*Pseudomonas putida* IFO12996) (Elmi et al., 2005).



**Figure 2.27.** Comparison of substrate binding pocket of esterase with 4-hydroxybenzoate and 3-(acetylthio)isobutyric acid respectively. (A) & (B) Representation of binding interaction of esterase with 4-hydroxybenzoate and 3-(acetylthio)isobutyric acid (positive control) respectively. Protein is shown in silver, substrates in red, residues in blue, and arrows in dark red color. Docking was conducted using Autodock\_vina\_1\_1\_2\_win32.msi. The center of the grid map was X (42), Y (50) and Z (34); and the autogrid calculation was set as  $5.948 \times -6.948 \times -4.904 \text{ \AA}$ .

### 2.3.6. Exploring binding sites of enzyme-metabolite interaction

AlphaFold2 generated 3D structure of each of the enzymes was highly accurate with a confidence score of above 93%. The enzyme-metabolite interaction and their binding energies were reported in Table 2.3.

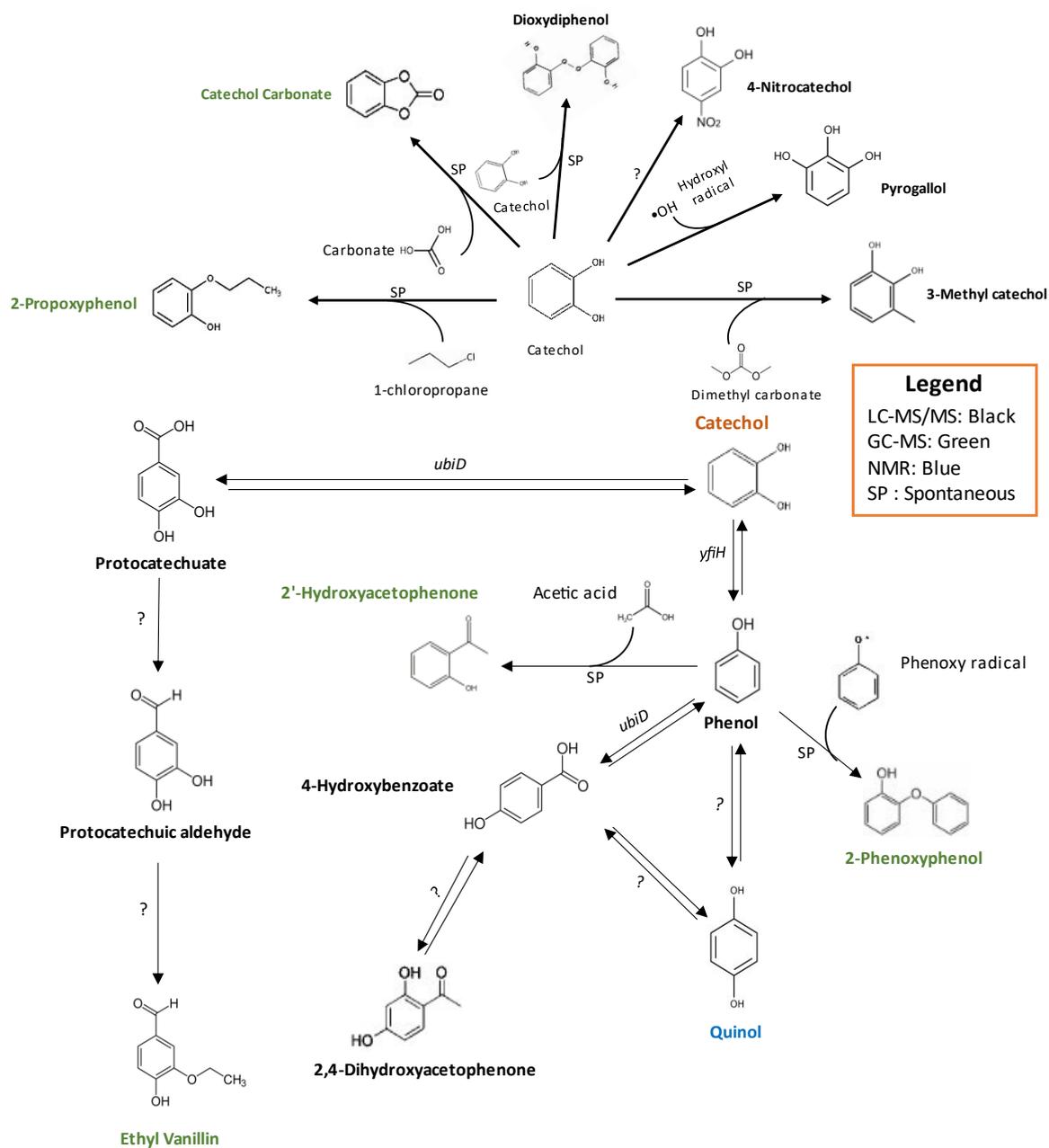
**Table 2.3.** List of binding energies between enzymes and substrates.

Enzyme Name	EC number	Species	Substrate	Binding energy, kCal/mol		
1 3-Octaprenyl-4-hydroxybenzoate decarboxylase	4.1.1.98	<i>Escherichia coli</i> UM 146	2-octaprenylphenol (Positive control)	-7.2		
			Glycine (negative control)	-2.6		
			Catechol	-4.8		
			Phenol	-4.6		
		<i>Shigella sonnei</i>	2-Methoxyphenol	-5.0		
			Catechol	-4.8		
			Phenol	-4.5		
2 Purine nucléoside phosphorylase	2.4.2.1	<i>Escherichia coli</i> UM 146	Adenosine (Positive control)	-5.1		
			Glycine (negative control)	-2.3		
			Catechol	-4.5		
		Polyphenol oxidase	1.10.3.1	<i>Escherichia coli</i> Str. NYVetLIRN-212	Adenosine (Positive control)	-5.1
					Glycine (negative control)	-2.3
3 Esterase	3.1.1.1	<i>Escherichia coli</i> UM 146	Catechol	-4.0		
			3-(acetylthio)isobutyric acid (positive control)	-3.4		
			Glycine (negative control)	-2.3		
			4-hydroxybenzoate	-4.2		

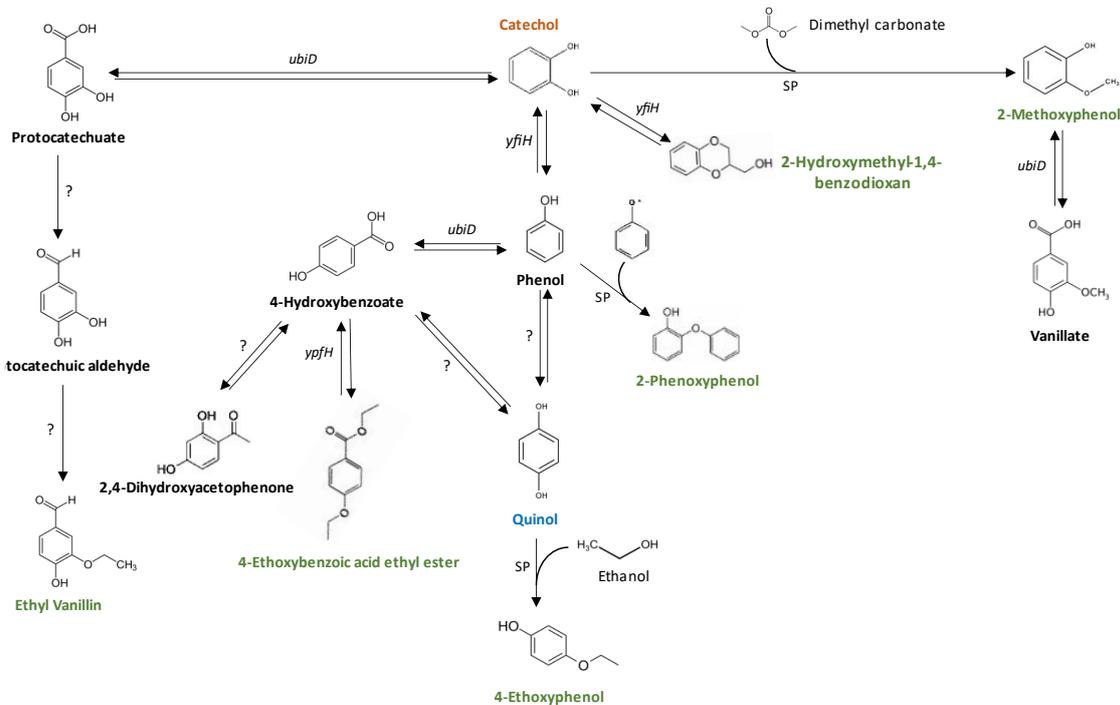
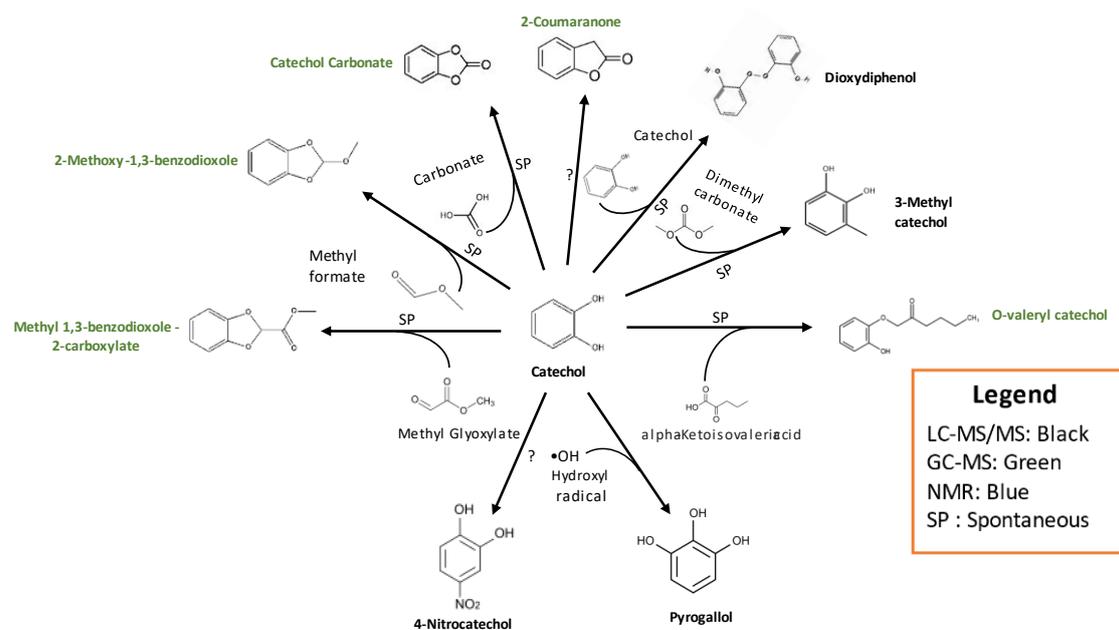
### 2.3.7. Catechol detoxification pathway

Expression profiling of candidate genes involved in catechol detoxification pathway were displayed in Figure 2.30. Catechol detoxification pathway follows 2 and 4 routes in anaerobic and aerobic conditions respectively. These include: (A) carboxylation of catechol to protocatechuate (both conditions), (B) oxidative cyclization of catechol to 2-hydroxymethyl-

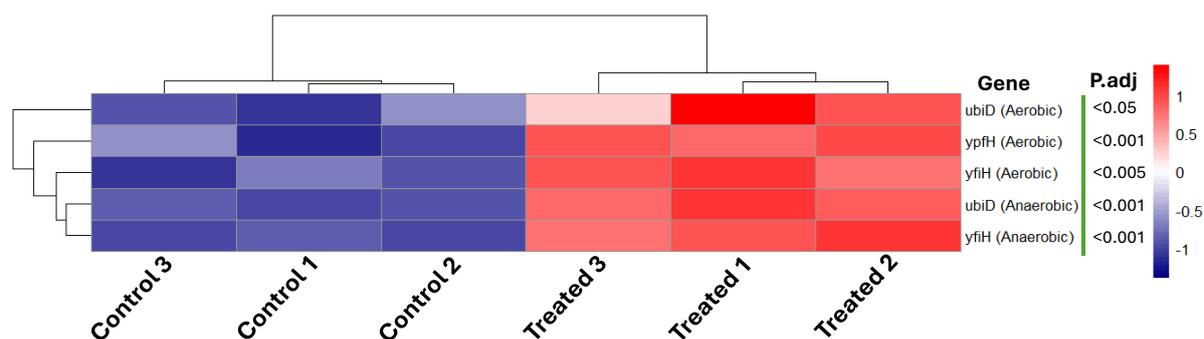
1,4-benzodioxan (aerobic condition), (C) methylation of catechol to 2-methoxyphenol (aerobic condition), and (D) dehydroxylation of catechol to phenol (both conditions) (Figure 2.28 & 2.29). (A) In the catechol to protocatechuate route, carboxylation of catechol into protocatechuate catalyzed by the *ubiD* gene encoding 3-octaprenyl-4-hydroxybenzoate decarboxylase enzyme (proposed). The underlying mechanism of the reduction of protocatechuate to protocatechuic aldehyde, and protocatechuic aldehyde to ethyl vanillin steps were unknown. (B) In catechol to 2-hydroxymethyl-1,4-benzodioxan route, oxidative cyclization of catechol into 2-hydroxymethyl-1,4-benzodioxan catalyzed by *yfiH* gene encoding purine nucleoside phosphorylase enzyme (proposed). (C) In catechol to 2-methoxyphenol route, catechol reacted with dimethyl carbonate to form 2-methoxyphenol, which afterwards converted into vanillate by *ubiD* gene encoding 3-octaprenyl-4-hydroxybenzoate decarboxylase enzyme (proposed). (D) In catechol to phenol route, the transformation of catechol into phenol is mediated by polyphenol oxidase enzyme (proposed). The conversion of phenol into other compounds follows two different routes: (i) phenol to quinol step, the underlying mechanism is unknown; and (ii) phenol to 4-hydroxybenzoate step, which is catalyzed by *ubiD* gene encoding 3-octaprenyl-4-hydroxybenzoate decarboxylase enzyme (proposed). 4-hydroxybenzoate served as a precursor to converted into several metabolites including 2,4-dihydroxyacetophenone, 4-ethoxybenzoic acid ethyl ester and quinol. The conversion of 4-hydroxybenzoate to 2,4-dihydroxyacetophenone and quinol is unknown. The *ypfH* gene encoding esterase enzyme performed the catalytic role to convert 4-hydroxybenzoate into 4-ethoxybenzoic acid ethyl ester (proposed).



**Figure 2.28.** Detoxification pathway of catechol during anaerobic condition. Arrows represent the single/reversible steps. Distinct colors indicate the metabolites detected in the respective analytical platform. Italics depict the gene name for enzymatic reactions. 'SP' indicates spontaneous reactions.



**Figure 2.29.** Detoxification pathway of catechol during aerobic condition. Arrows represent the single/reversible steps. Distinct colors indicate the metabolites detected in the respective analytical platform. Italics depict the gene name for enzymatic reactions. ‘SP’ indicates spontaneous reactions.



**Figure 2.30.** Heatmap showing gene expression changes associated with detoxification pathway induced by catechol exposure under anaerobic and aerobic conditions. The vertical color scale at the right side of the heat map represents the intensity of data values. The highest level of expression is shown by 'Red' color, and the lowest level is represented by 'Navy' color. Adjusted p-values are displayed beside the gene name.

### 2.3.7.1. Spontaneous reactions: an insight into chemical transformations

Besides the enzymatic reactions, there were several metabolites (including, methyl 1,3-benzodioxole-2-carboxylate, 2-methoxy-1,3-benzodioxole, catechol carbonate, 2-coumaranone, 3-methyl catechol, 2-propoxyphenol, o-valeryl catechol, 2-phenoxyphenol, 4-ethoxyphenol, 2-methoxyphenol, 4-nitrocatechol and dioxydiphenol) identified which are proposed to be produced from catechol via spontaneous reactions.

Catechol is a highly reactive molecule which produces 2,2-disubstituted and 2-monosubstituted 1,3-benzodioxoles from ketones and aldehydes (Jin et al., 2001). Following a similar mechanism, the formation of methyl 1,3-benzodioxole-2-carboxylate and 2-methoxy-1,3-benzodioxole is hypothesized to occur when catechol reacts with methyl glyoxylate and methyl formate respectively (Figure 2.31). Presence of methyl glyoxylate in the treated samples confirmed through comparing the fragmentation pattern generated by CFM-ID 4.0 (S Figure 2.18P). However, the exact mass of methyl formate is 60.0520. In LC-MS/MS analysis, the lowest mass range was set to 80. Therefore, the fragmentation pattern of methyl formate was not viable to examine. Formic acid was detected in treated samples using NMR. For metabolomics study, methanol solvent was used to prepare the samples. Hence, it is possible formic acid reacted with methanol to produce methyl formate.

The reaction of catechol and dimethyl carbonate has been reported to produce catechol carbonate, 1,2-dimethoxybenzene, 2-methoxyphenol, and 3-methylcatechol (Reaxys reaction ID: 10307995) (Vijayaraj & Gopinath, 2006). Dimethyl carbonate was detected using LC-MS/MS analysis. As dimethyl carbonate wasn't commercially available, a reaction was

performed between catechol and carbonic acid for 24 hrs and identified catechol carbonate using GC-MS and NMR (S Figure 2.19(Q) & S Figure 2.25). Another study reported that catechol reacts with methanol to form 2-methoxyphenol and methylcatechols in the presence of acid catalysts (Ardizzi et al., 2007). A reaction was conducted among catechol, carbonic acid and methanol without water for 24 hrs, resulting in catechol carbonate, 2-methoxyphenol and 1,2-dimethoxybenzene (determined using GC-MS). All this evidence confirmed that 2-methoxyphenol and 3-methylcatechol can be produced from catechol when it reacts with methanol and dimethyl carbonate. Notably, catechol and carbonic acid reaction form light bronze color, which suggests that the light bronze color occurring during bacterial culture (aerobic condition) originates from oxidation of catechol carbonate.

The formation of *o*-valeryl catechol is proposed to occur when catechol undergoes reaction with  $\alpha$ -ketoisovaleric acid. The presence of  $\alpha$ -ketoisovaleric acid in the treated samples was confirmed by NMR.

The mechanism by which 2-coumaranone is produced from catechol is not clearly understood. A reaction was conducted between catechol and acetic acid maintaining little acidic condition (pH 6.4~7) for 24 hrs. No catalyst was used while performing the reaction. The expected compound (2-coumaranone) wasn't detected upon NMR, LC-MS/MS and GC-MS analysis. Technically, there are numerous factors exist in bacterial culture. It is likely these factors might have a role to act as a catalyst which influenced the formation of 2-coumaranone from catechol.

2-propoxyphenol can be produced from catechol when it reacts with 1-chloropropane (Reaxys reaction ID: 20394784). The exact mass of 1-chloropropane is 78.54038 g/mol, which is lower than the mass range set to LC-MS/MS analysis. It is unable to check the presence of 1-chloropropane in the treated samples.

The mechanism behind the formation of 4-nitrocatechol from catechol is unclear. Studies reported that 4-nitrocatechol can be produced from catechol reacting with nitrate ( $\text{NO}_3^-$ ) (Rana et al., 2023). Still to date, there were no studies reported the production of nitrate by the *E. coli* UMI46 strain. Generally, nitrate reductase enzyme catalyzes the reduction of nitrate ( $\text{NO}_3^-$ ) to nitrite ( $\text{NO}_2^-$ ) (Campbell, 1999). Nitrate reductase A enzyme encoded by *narG*, *narH*, and *narI* genes. *narZ* and *narU* genes encode for nitrate reductase Z, and nitrate:nitrite transporter respectively. According to DESeq2 analysis, all these genes related to nitrate regulation are either downregulated or not significant in treated samples in both aerobic and

anaerobic conditions. More groundwork is needed to know the exact mechanism of the conversion of catechol to 4-nitrocatechol.

In the individual reactions involving catechol with glycerol, valeric acid, acetic acid, glyoxylate, and acetyl-CoA, no reaction products were detected upon analysis using NMR.

Although not directly derived from catechol, several metabolites including, 4-ethoxyphenol, phenoxyphenol, 2'-hydroxyacetophenone, protocatechuic aldehyde and ethyl vanillin are intermediates in the catechol detoxification pathway, originating from other metabolites.

Quinol and ethanol reaction was aimed to investigate the presence of 4-ethoxyphenol. The anticipated  $^1\text{H}$  NMR spectrum of quinol and ethanol reaction compared with the spectrum of 4-ethoxyphenol, indicating the presence of an unknown compound. The sample of quinol - ethanol reaction was subjected to LC-MS/MS and GC-MS for further confirmation of the presence of 4-ethoxyphenol. Results of LC-MS/MS analysis suggested a different compound, named 4-ethoxyphenyl 4-butylbenzoate. It is likely that dissociation of 4-ethoxyphenyl 4-butylbenzoate into 4-ethoxyphenol occurred during the gaseous-phase of GC-MS. Notably, no catalyst was used while performing the experiment, indicating the potential involvement of other factors or catalyst in the formation of 4-ethoxyphenol during bacterial culture. Another reaction suggested by Reaxys (reaction ID: 54356680), where quinol reacts with diethyl carbonate to produce 4-ethoxyphenol. However, there was no presence of diethyl carbonate in the treated samples. Further investigation is required to unveil the underlying mechanism of the formation of 4-ethoxyphenol from quinol.

In the presence of reactive oxygen species (ROS) such as  $\bullet\text{OH}$  and/or  $\text{SO}_4^{\bullet-}$ , phenol forms phenoxy radical, which subsequently combines with phenol to form phenoxyphenols (Huang & Zhang, 2022). Following a similar mechanism, 2-phenoxyphenol can be produced from phenol when it reacts with phenoxy radical. ROS originates from molecular oxygen (Ferooshani et al., 2017). Also, it is reported that hydrogen peroxide ( $\text{H}_2\text{O}_2$ ), superoxide ( $\text{O}_2^-$ ), singlet oxygen ( $^1\text{O}_2$ ), and hydroxyl radical ( $\bullet\text{OH}$ ) can be generated from catechol under different oxidizing conditions such as autoxidation, chemical-induced oxidation, and metal ion-mediated oxidation (Kalyanaraman et al., 1985; Meng et al., 2015; Razaviamri et al., 2021). In addition, under certain oxidation conditions, bacteria itself can produce ROS as a byproduct during aerobic respiration (H. Li et al., 2021). DEGs analysis suggested that *sodA* gene is highly

expressed during aerobic condition in catechol treated samples. The *sodA* gene encodes superoxide dismutase which produces hydrogen peroxide from superoxide (Steinman, 1988).

The formation of 2'-hydroxyacetophenone is proposed to be produced from phenol through its reaction with acetic acid (Reaxys reaction ID: 643074). The presence of acetic acid in the treated samples was detected via NMR.

Pyrogallol is produced from catechol through a hydroxylation reaction. It was reported that catechol reacts with hydroxyl radical ( $\bullet\text{OH}$ ) to form 1,2,4-trihydroxybenzene (Chen et al., 2020). Following a similar mechanism, pyrogallol can be produced from catechol.

Protocatechuic aldehyde and ethyl vanillin were determined using LC-MS/MS and GC-MS respectively. Enzymatic and spontaneous reactions were explored to elucidate the process by which protocatechuic aldehyde and ethyl vanillin produced from protocatechuate. However, the specific mechanism underlying the conversion of protocatechuate to protocatechuic aldehyde and ethyl vanillin remains unclear. To investigate further, a reaction was performed between protocatechuate and ethanol for 24hrs to assess the formation of any potential products such as protocatechuic aldehyde. No discernible changes or products were detected under these conditions. Notably, enzymes such as vanillin dehydrogenase or aromatic aldehyde dehydrogenase possess ALDH\_HBenzADH domain which catalyzes the conversion of protocatechuate to protocatechuic aldehyde (Rhea reaction ID: 72539), although aromatic aldehyde dehydrogenase enzyme is particularly absent in *E. coli UMI46*. Additionally, other relevant enzymes such as aldehyde dehydrogenase A lack the ALDH\_HBenzADH domain. Moreover, the downregulation of the *aldA* gene was observed in the catechol treated samples. Further investigation is required to elucidate the mechanism underlying the formation of protocatechuic aldehyde from protocatechuate. To investigate the mechanism behind the formation of ethyl vanillin, a reaction was conducted between protocatechuic aldehyde and ethanol without any catalyst for 24hrs. However, no additional compounds or ethyl vanillin were detected via NMR. Whether it is possible that other factors may contribute to the conversion of protocatechuic aldehyde into ethyl vanillin during culture, requiring additional research.

## 2.4. Discussion

The diverse microbial community in the human gastrointestinal system plays a major role in health, nutrition, immunity, and disease (Guinane & Cotter, 2013). This complex molecular crosstalk and the mutualistic relationship between gut bacteria and hosts maintain

intestinal homeostasis (El Aidy et al., 2013). Recent advances in characterizing the gut bacterial by-products from food digestion provide interesting findings such as gut bacterial degradation of dietary fiber to short-chain fatty acids (SCFA); bile acids to secondary bile acids; phosphatidylcholine, choline, and L-carnitine from trimethylamine-N-oxide (TMAO) (Mazhar et al., 2023; Simó & García-Cañas, 2020; Ussher et al., 2013). Uremic toxins produced by bacterial digestion of polyphenols or amino acids can cause serious harm to the heart, kidney and liver (Wishart, 2019). Some of these uremic toxins include indole derivatives such as indoxyl sulphate, which arises from the breakdown of tryptophan (Leong & Sirich, 2016). Indoxyl sulfate is widely associated with vascular and renal diseases (Leong & Sirich, 2016). Gut bacteria also produce beneficial metabolites from food digestion (Rowland et al., 2018). Isoflavonoids have a protective role in human diseases such as cardiovascular disease, osteoporosis, breast cancer, prostate cancer, and menopausal symptoms (Pejčić et al., 2023). The by-products from the metabolism of isoflavones by gut bacteria are more physiologically active than the precursors. Aromatic compounds, whether serving as precursors or produced as byproducts, are toxic in nature and hazardous to animal health (Oliphant & Allen-Vercoe, 2019). Bacteria found in natural environments or contaminated areas can break down aromatic hydrocarbons (Seo et al., 2009). Potential catalytic sources for the biodegradation of organic molecules are soil microorganisms. In aerobic and anaerobic conditions, a variety of bacteria can biodegrade hazardous organic substances, improving the clean-up of various habitats such as water, soil, and wastewater (Jothimani et al., 2003). It is necessary to decipher the complex interconnected bacterial responses to the presence of aromatic chemicals to get an overall understanding of how aromatic catabolic processes function in connection to their genome and environmental context. In the present study, the *E. coli UMI46* strain was exposed to catechol to observe whether it could consume the catechol and produce byproducts. In the case of toxicity to catechol, another objective was to determine the effect of catechol on the virulence factors in *E. coli UMI46*.

#### **2.4.1. Catechol treatment: toxicity or catabolism**

A slowed growth trend was observed in response to catechol across both aerobic and anaerobic conditions (Figure 2.1). Moreover, no diauxic growth was detected, indicating that the *E. coli UMI46* strain solely utilized glucose as carbon source. These results clearly signify that catechol inhibited the growth of *E. coli UMI46*. This observation was anticipated, given that previous studies used catechol and catechol derivatives as antimicrobial agents (Baptista et al., 2019; Kocaçalışkan et al., 2006). At a level of ~100 µg/mL, phytochemicals such as

gallic acid, eugenol, quercetin, piperine, and menthol were toxic to *E. coli* (Sharma et al., 2019). Another investigation revealed that a number of individual pure phenolic compounds, including catechin, caffeic acid, gallic acid, protocatechuic acid, quercetin, and rutin, exhibited antibacterial effects against pathogenic *E. coli* (Vaquero et al., 2007). A combined effect of gallic acid and catechin showed growth-inhibitory effects on *E. coli* (Díaz-Gómez et al., 2014). Although catechol showed toxicity towards *E. coli UMI46*, it utilized as a carbon source for a number of different bacterial species, including *Desulfobacterium sp.* Strain, *Thauera aromatica*. *T. aromatica* type strain K172, *Azoarcus sp.* Strain EbN1, *Pseudomonas putida* and *Pseudomonas cepacia* ATCC 29351, *Ralstonia*, *Azotobacter*, and *Pseudomonas* species (Ding et al., 2008, 2008; Gorny & Schink, 1994; Hamzah & Al-Baharna, 1994; Ornston & Stanier, 1966; Suvorova & Gelfand, 2019).

#### **2.4.1.1. Does ortho or meta cleavage route exist in catechol metabolism of *E. coli UMI46*?**

Most of the studies of catechol metabolism were conducted using bacteria found in nature under aerobic/anaerobic conditions, and/or by co-metabolic degradation processes (Aghapour et al., 2013, 2013; Ding et al., 2008; Gorny & Schink, 1994; Ornston & Stanier, 1966, 1966; Zheng et al., 2019). The co-metabolic degradation process refers to the transformation of a compound into a chemically changed form by the microorganism, or using enzymes by performing in-vitro reactions, resulted in a target compound. *Pseudomonas putida* degrades catechol into 3-ketoadipate, as defined by the co-metabolic process (Ornston & Stanier, 1966). In aerobic condition, catechol degradation occurs via either the intradiol ortho cleavage or extradiol meta cleavage route. The major by-product of catechol intradiol ortho cleavage is cis, cis-muconic acid, which subsequently transformed into muconolactone, 3-oxoadipate, succinate and acetyl-coA. 2-hydroxymuconic semialdehyde is produced from catechol via extradiol meta cleavage route, which afterward converted into 2-oxo-penta-4-enoate, followed by the production of 4-hydroxy-2-oxo-valerate, acetaldehyde, and pyruvate. Biological degradation of catechol in wastewater follows a meta-cleavage route (Aghapour et al., 2013). Exposure of catechol towards *E. coli UMI46*, neither ortho nor meta cleavage routes were observed. Standards of major ortho cleavage metabolites such as cis, cis-muconic acid, and 3-oxoadipate were run in LC-MS/MS and NMR. There was no observed presence of cis, cis-muconic acid, and 3-oxoadipate metabolites within the treated samples. Catechol 1,2-dioxygenase enzyme catalyzes the production of cis, cis-muconic acid. However, catechol 1,2-dioxygenase enzyme is absent in the *E. coli UMI46* strain. Several reports indicated that UbiD

enzyme have the catechol 1,2-dioxygenase activity and holds the potential to convert catechol into cis, cis-muconic acid (Jensen et al., 2021; Weber et al., 2017).

MZmine2 (LC-MS/MS data analyzing software) suggested the presence of catechol meta cleavage metabolite (2-hydroxymuconate semialdehyde) in the treated samples. As 2-hydroxymuconate semialdehyde wasn't available for commercial purchase, CFMID 4.0 was used to generate the fragmentation pattern and compared with the treated samples. This comparison revealed that the *E. coli UMI46* strain didn't transform catechol into 2-hydroxymuconate semialdehyde. To further verify, expression of *xylF* and *xylG* genes, crucial for the transformation of 2-hydroxymuconate-6-semialdehyde into other forms, was examined. *xylF* and *xylG* gene encoding enzymes (such as 2-hydroxymuconate-6-semialdehyde hydrolase and 2-hydroxymuconate-6-semialdehyde dehydrogenase enzyme, respectively) involved in the conversion of 2-hydroxymuconate semialdehyde into other forms. 2-hydroxymuconate-6-semialdehyde hydrolase (encoded by *xylF* gene) catalyzes the conversion of 2-hydroxymuconate-6-semialdehyde into 2-oxopent-4-enoate; while 2-hydroxymuconate-6-semialdehyde is transformed into (2Z,4E)-2-hydroxyhexa-2,4-dienedioate, catalyzed by 2-hydroxymuconate-6-semialdehyde dehydrogenase (encoded by *xylG* gene). According to DESeq2 analysis, catechol exposure had no effect on the expression of both *xylF* and *xylG* genes. All this reasoning justifies that there was no catechol ortho/meta cleavage route existing in the *E. coli UMI46* strain.

#### **2.4.1.2. Genetic changes due to catechol toxicity**

Aromatic compounds are highly hydrophobic in nature (Díaz et al., 2001). Typically, they are toxic to microbes by accumulating in and disrupting cell membranes (Díaz et al., 2001). *E. coli* uses several strategies to facilitate biodegradation and biotransformation of aromatic compounds. Bacteria increase solvent tolerance by expressing *marA*, *robA*, and *soxS* genes to confer against hydrophobic toxic aromatic compounds (Aono et al., 1998). *E. coli* activates the efflux pumps by overexpressing *AcrA*, *AcrB*, and *TolC* genes which mediates extrusion of the aromatic toxic compounds (Aono et al., 1998; de Bont, 1998). Catechol is an aromatic compound which contains two hydroxyl groups (-OH) attached to the benzene ring. The properties of catechol are hydrophilic in nature and toxic to animals and microbes (Schweigert et al., 2001). Cyclization of catechol during redox reactions forms semiquinone radicals and ortho-benzoquinone, which leads to the accumulation of reactive oxygen species (Schweigert et al., 2001). Catechol also forms stable complexes reacting with different metal ions. Metal-catechol complex, redox cyclized products, and reactive oxygen species (ROS)

cause DNA damage in cells through different mechanisms such as directly interacting with DNA molecules, causing modifications in DNA bases, or inducing breaks in the DNA strands (Schweigert et al., 2001). Besides DNA strand breakage, ROS also can impair lipids, and proteins (Deller et al., 2008; Patridge & Ferry, 2006). Moreover, S-adducts are formed by electrophilic quinones, which impair cellular thiols and cause thiol-specific stress in cells (Rodriguez et al., 2005). Collectively, in the present study, an inhibition response model was observed, wherein catechol initially induces damage to the cell envelope and DNA. As a reaction, *E. coli UMI46* responded by activating DNA repair system, cell division, and cell wall biogenesis, and modulated other metabolic pathways (such as glycolysis-gluconeogenesis, biotin metabolism, amino sugar metabolism and nucleotide metabolism, pentose-phosphate pathway, methane metabolism, cysteine, and methionine metabolism, one carbon pool by folate). Changes in metabolic pathways due to catechol exposure were presented in S Figure 2.1-S Figure 2.14. Additionally, catechol exposure influenced several physiologically active systems including, two-component systems, efflux pump, and other systems related to stress in *E. coli UMI46*. Due to catechol treatment, two main envelope stress-responsive systems (Psp and Rcs) were identified in *E. coli UMI46*, both of which exhibited upregulation during aerobic and anaerobic conditions (Figure 2.13). Psp, Cpx, Bae, Rcs, and  $\sigma E$  become activated due to styrene exposure in *E. coli NST74* (Machas et al., 2021). Efflux systems, including AcrAB-TolC, Cus, MdtABC-TolC, and MdtK were responsive in response to catechol in *E. coli UMI46* (Figure 2.10). AcrAB-TolC RND efflux pumps were activated because of styrene exposure in *E. coli NST74* (Machas et al., 2021). Two-component systems were induced owing to catechol exposure in *E. coli UMI46*, with modulation of 34 and 45 genes in aerobic and anaerobic conditions respectively (Figure 2.8 - Figure 2.9). Gallic acid induced the expression of 33 genes related to two-component systems in *E. coli strain W3110* (Liu et al., 2022). Overexpression of *marA* and *soxS* genes suggests that *E. coli UMI46* increased solvent tolerance level to tackle the toxicity of catechol (Figure 2.14). Also, *E. coli UMI46* activated the efflux pump to mediate the expulsion of aromatic toxic compounds by increasing the expression of *AcrA* and *TolC* genes. A similar finding was observed in *E. coli NST74*, which controlled the toxicity of styrene by increasing the solvent tolerance level, and through the activation of the efflux pump (Machas et al., 2021). Besides these, catechol exposure elicited varied effects on the virulence factors of *E. coli UMI46*, causing upregulation of 10 genes and downregulation of 12 genes in anaerobic condition, while inducing upregulation of 6 genes and downregulation of 8 genes under aerobic condition (Figure 2.19). However, there was no impact of catechol exposure on the biofilm formation and flagellar assembly under aerobic

condition (Figure 2.21). Notably, under anaerobic condition, catechol treatment exhibited contrasting responses, with biofilm formation demonstrating upregulation of genes and flagellar assembly showing downregulation of genes (Figure 2.20). An example of one such mechanical stress induction to biofilm formation was observed in *E. coli CFT073* (Chu et al., 2018). Catechol has adhesion properties and can elicit siderophore activity, which overall plays an important role in biofilm formation (Saiz-Poseu et al., 2019; Tahmasebi et al., 2022). Likewise catechol, other phytochemicals such as 4-hydroxybenzoic, gallic, vanillin and epicatechin, cinnamic, sinapic, ferulic, and chlorogenic acids enhanced biofilm formation in *Pseudomonas aeruginosa*.

#### **2.4.1.3. Strategies used by *E. coli UMI46* to mitigate the toxic effects of catechol**

Bacteria use biochemical transformation strategies such as acetylation, deamination, decarboxylation, demethylation, glycosylation, hydrolysis, ring cleavage, and sulfation to convert the toxic molecules to a non-toxic/less toxic substances (P. Li et al., 2020). Following a similar way, *E. coli UMI46* strain mediated the biotransformation of toxic catechol to other compounds to mitigate the toxic effects. Catechol is converted into protocatechuate, phenol, pyrogallol, and 2-hydroxymethyl-1,4-benzodioxan via carboxylation, dehydroxylation, hydroxylation, and cyclization process respectively. Phenol produced through the dehydroxylation process, followed by another round of carboxylation and hydroxylation steps transformed into 4-hydroxybenzoate and quinol respectively. Likewise, 4-hydroxybenzoate was converted into 2,4-dihydroxyacetophenone, 4-ethoxybenzoic acid ethyl ester, and quinol.

Azoreductase, encoded by *azoR* gene, are flavoenzymes that are involved in the detoxification of azo compounds (Misal & Gawai, 2018). In this experiment, catechol treatment causes upregulation of *azoR* gene in both aerobic and anaerobic conditions. The bio-transformed or spontaneously produced compounds through the catechol detoxification pathway may react with arylhydrazines to form azo compounds, which are later detoxified by azoreductase enzyme to a simpler or non-toxic form. The formation of azo compounds through the reactions between catechol and catechol derivatives with arylhydrazines reported in (Petran et al., 2020).

#### **2.4.2. Unknown challenging metabolites**

In LC-MS/MS analysis, many annotated features were identified as false positive. It was observed that MS-data analyzing software considered daughter ions (electrically charged product of a particular parent ions) as individual parent ions, which leads to false positive results. A benchmarking experiment was conducted to discuss this detail in Chapter 3. It was

surprising that a limited numbers of metabolites were detected in positive mode. Poor separation or co-eluting peaks were observed in the positive mode, which resulted in false positive metabolites. Methanol and water along with 0.1% formic acid were used as a mobile phase in positive mode. The use of 0.1% formic acid could be the probable reason of co-eluted peaks in the chromatogram. Instead of formic acid, other solutions such as ammonium acetate or ammonium formate can use alongside water-methanol as a mobile phase in positive mode (Manier et al., 2019). This possibly improves the signals or resolution of the chromatogram, with identification of more compounds in positive mode.

Contrarily, a different issue was observed in NMR results. Many unknown compounds were detected under 700 MHz NMR (S Figure 2.26 & 2.27). The annotated compounds obtained from LC-MS/MS analysis were considered for the generation of chemical shifts. Predicted chemical shifts overlaid with the chemical shifts of unknown compounds (detected in NMR). Chemical shift comparison didn't provide any insight. The reasoning behind that is explained with a benchmarking experiment in Chapter 3. Other approaches were considered including collection of fractions using HPLC, and 2D-NMR (HSQC and TOCSY) analysis. These strategies also couldn't reveal the identity of these unknown compounds. In HSQC/TOCSY analysis, the high concentration of substrate overlapped the areas, which certainly affected the determination of other metabolites.

### **2.4.3. Limitations of the study**

This study was designed and performed using available resources and laboratory guidance. Yet, there were some limitations. First, the anaerobic condition was maintained by purging nitrogen gas instead of the anaerobic chamber. Bacteria, media, and substrates weren't kept in the anaerobic chamber before the experiment. Second, for the metabolomics study, the experiment wasn't performed using isotope-labeled substrates due to budget constraints. Third, enzyme-metabolite interactions were performed using Autodock Vina. As the binding affinities were predicted using computational algorithms, it may vary with the actual experimental results.

### **2.5. Conclusion**

Metabolomics and transcriptomics techniques were used to characterize metabolites and genes associated with the interactions of catechol and *E. coli* UMI46. Untargeted metabolomics analysis led to the determination of 16 and 23 metabolites associated with the detoxification of catechol under anaerobic and aerobic conditions respectively. Moreover,

transcriptomics analysis suggested that 2 and 3 genes are linked to the catechol detoxification pathway in anaerobic and aerobic conditions respectively. The involvement of UbiD enzyme in the catechol metabolic pathway is highly versatile and performs multiple functions. Further investigation is required to uncover the actual role of UbiD enzyme in the *E. coli UMI46* strain.

## Chapter 3

# Uncovering False Positives in Metabolomics: Insights from Benchmarking Approaches

### 3.1. Introduction

Untargeted metabolomics is a powerful analytical technology with diverse applications in a wide range of research fields, including clinical biomarker discovery, identifying bioactive compounds for disease progression, studying drug efficacy and toxicity, and environmental toxicological research (Qiu et al., 2023; Ruan et al., 2022). Liquid chromatography high-resolution mass spectrometry (LC-HRMS) has emerged as a prominent platform among other tools for untargeted metabolomics due to the high throughput, sensitivity, specificity, and broad coverage of small molecule detection (Guo & Huan, 2020; Kunzelmann et al., 2018). LC-HRMS detects tens to thousands of features, where not all the features are biologically relevant (Johnson et al., 2016; Uppal et al., 2016). Many of the identified features represent background signals due to improper sample processing (Coble & Fraga, 2014). Multiple signals from the same analyte led to false positive features (FPFs). Extracted Ion Chromatograms (EICs) and detection of chromatographic peaks from the EICs are the initial stages of the preprocessing workflow (Myers et al., 2017). These steps enable the identification and relative quantitation of analytes. Any discrepancies in these steps can affect entire data processing steps including subsequent statistical and metabolic pathway analysis (Ju et al., 2019). Screening and removing these FPFs are of foremost importance in metabolomics studies. A combination of multiple strategies is used to remove FPFs, including 1) setting signal to noise ratio (SNR) threshold level to compare the low signal intensities with background signals; 2) accurate retention time alignment of chromatographic peaks; 3) relative standard deviation filtering across samples; 4) blank and control subtraction from the samples; 5) isotope filtering; 6) incorporating quality control samples; 7) grouping of different adduct types; 8) consideration of neutral loss; 9) analyzing shape and co-elution pattern of chromatographic peaks; 10) comparing fragmentation pattern with authentic standards or databases; and 11) isomer resolution (Schrimpe-Rutledge et al., 2016). Several proprietary and open-source software packages have been developed to automate the filtering process. Arguably Metaboanalyst, XCMS, and MZmine2 are the most widely used software packages used to analyze the LC-HRMS data. These software uses different approaches to process the raw data to screen and filter the features.

Likewise LC-HRMS, nuclear magnetic resonance (NMR) is a complementary technique in metabolomics research (Letertre et al., 2021). NMR provides detailed structural information. NMR can detect metabolites in low concentrations, typically micromolar ( $\mu\text{M}$ ) to millimolar (mM) range (Goldansaz et al., 2017). A combination of both techniques (NMR and LC-HRMS) yields a more extensive and comprehensive understanding of metabolomics research.

The objective of the study was to discuss the reproducibility of untargeted metabolomics using different analytical techniques. To do this, a synthetic mixture and a single pure compound solution were analyzed to observe the variability of the results. Moreover, this study will assist in deciding the reliability of popular software packages in LC-HRMS analysis.

## **3.2. Methods and materials**

### **3.2.1. Chemical reagents**

All the chemicals required to prepare defined mixture (cytosine, alanine, kynurenine, homovanillic acid, lysine, myo-inositol, beta-hydroxybutyric acid, thymine, pyruvic acid, deoxycytidine, serine, asparagine, succinic acid, tryptophan, deoxycytidine monophosphate, citric acid, uridine, 5-methyluridine, taurine, 2-pyrocatechuic acid, benzyl alcohol, phenol, tyramine, phenylethylamine, spermidine, 2,3-butanediol, deoxyadenosine monophosphate, valeric acid) were purchased from Sigma-Aldrich. HPLC grade water, deuterated 2,2-dimethyl-2-silapentane-5 sulfonate (DSS-d6), potassium phosphate monobasic, potassium phosphate dibasic, and D2O (99.9%) were purchased from Sigma (Oakville, Canada). The 2-chloropyrimidine-5-carboxylic acid (98%) was purchased from ArkPharm (Libertyville, USA). The Amicon (1.5 mL) 3 kDa molecular weight cut off (MWCO) filtration units were purchased from Millipore Sigma (St. Louis, United States). The NMR tubes (3 mm) were purchased from Bruker Ltd. (Milton, Canada). Compounds for NMR standards were generally purchased from Millipore, Sigma, AK Scientific Inc., or Tokyo Chemical Industry Co. Ltd.

### **3.2.2. Preparation of defined mixture**

A total of 28 compounds with purity of  $\geq 98-99\%$  were used to prepare the defined mixture. 5 mM stock was prepared in water for each chemical. Required amounts were taken from each stock to prepare 75  $\mu\text{M}$  concentrated defined mixture. A 0.22-micron filter was used to clean the samples before running in LC-MS and NMR.

### 3.2.3. Preparation of a single concentrated pure solution

The compounds used to prepare the defined mixture were not entirely (100%) pure. Contaminates may arise from the compounds. Among the compounds utilized to make the defined mixture, lysine was one of them. Different concentrations (5 mM, 25 mM, 50 mM, 85 mM and 900 mM) of lysine solutions were prepared and subjected to NMR. To determine the contaminants, it was required to eliminate the lysine from the highly concentrated (900 mM) lysine solution.

#### 3.2.3.1. Resin based column chromatography

Different cation and anion exchange columns were used to remove the lysine from the concentrated lysine solution. Both cation exchange resins (DOWEX R50 WX8, H<sup>+</sup> and Amberlyte IRC-120H, H<sup>+</sup>) exhibit comparable cation exchange properties owing to their negatively charged sulphonic acid (-SO<sub>3</sub>H) functional groups. Bead size and physical appearance is different in both resins. Amberlyte IRC-120H, H<sup>+</sup> and DOWEX R50 WX8, H<sup>+</sup> comprise macropores and gel-type resins respectively, which results in slight differences in their performance. Conversely, both anion exchange columns have different functional groups. Amberlyte IRA402 chloride form and Amberlyst A-26 (OH) resins contain chloride ion and hydroxyl ions respectively. Due to the distinct functional groups, these anion exchange columns operate differently. The anionic part (carboxylate, -COO<sup>-</sup>) of lysine may interact with the positively charged functional group of cationic resins, and the cationic part (amine, -NH<sub>2</sub>) of lysine possibly bind to the negatively charged functional group of anionic resins. A concentrated lysine solution was passed through the resins and fractions were collected. After lyophilization, concentrated fractions were run in NMR and Orbitrap to determine the contaminants.

#### 3.2.3.2. Liquid-liquid extraction

Besides the column-based chromatography technique, derivatization and liquid-liquid extraction were also conducted to extract the contaminants from the lysine solution. A 5% phenyl isothiocyanate (PITC) solution was prepared by mixing 1900  $\mu$ L water, 1900  $\mu$ L of pyridine, 1900  $\mu$ L of ethanol, and 300  $\mu$ L of PITC. 10  $\mu$ L 900 mM of lysine stock was transferred to a glass vial. 2 mL of 5% PITC solution into the same glass vial was added. Derivatization was conducted at room temperature for 20 minutes. Samples were dried using a nitrogen evaporator. After that, samples were reconstituted separately using 200  $\mu$ L of water, d<sub>4</sub>-methanol, and sodium hydroxide (0.1M NaOH). 200  $\mu$ L of ethyl acetate was added to the sample reconstituted using NaOH. The tube containing sodium hydroxide and ethyl acetate

was thoroughly vortexed for 30 seconds, shaken at 1,000 rpm for 10 minutes, and centrifuged for  $18,000 \times g$  for 15 minutes. The aqueous layer was collected for NMR measurement. 50  $\mu\text{L}$  NMR buffer was added to all reconstituted samples. The tubes were thoroughly vortexed for 30 seconds and centrifuged at 1,000 rpm for 10 min. Samples were transferred to 3 mm tubes for NMR run.

#### **3.2.4. LC-HRMS analysis**

LC-HRMS analyses were conducted with an Ultimate 3,000 UHPLC system (Thermo Scientific®, MA, USA) coupled to a Orbitrap mass spectrometer which is equipped with a heated electrospray ionization (H-ESI) source. Prior to analysis, external mass calibration was done in accordance with manufacturer instructions. Chromatographic separation was conducted on a ZORBAX C18 column ( $2.1 \times 100$  mm I.D., particle size 3.5  $\mu\text{m}$ ). Column temperature was set at 50°C, maintained by a Dionex UltiMate 3000 RS analytical column heater. Mobile phases were binary mixtures: water (eluent A) and methanol (eluent B), both with 0.1% formic acid. Gradient elution started at 100% A and 0% B, followed by: 0–7.8 min 40% B, 7.8–9 min 40% B, 9–9.1 min 0% B, 9.1–10 min 0% B. The total run time was 10 min. Flow rate was set at 5  $\mu\text{L s}^{-1}$  and the injection volume was 5  $\mu\text{L}$ . Metabolites were analyzed under both positive and negative electrospray ionization (ESI) mode. The ESI conditions in each run were set to spray voltage 3 KV in positive mode and -3 KV in negative mode, heated capillary temperature 299.57°C, sheath gas flow rate 59.82 U, auxiliary gas flow rate 30 U, sweep gas flow rate 1.2 U, auxiliary heater temperature 529.80°C, and S-lens RF level 60. The settings for full scan and dd-MS<sup>2</sup> data acquisition were as follows: resolution - 50000 fwhm; automatic gain control (AGC) target -  $1e^5$ ; maximum injection time - 50 ms; loop count - 5; and isolation window - 4 m/z. Data acquisition conditions were set over a mass range of 50–750 m/z to enhance the compound identification.

##### **3.2.4.1. Mass spectrometry data analysis**

Raw data obtained from the Orbitrap instrument were converted to mzML format using msConvert. Blank was used as a control for both the defined mixture and single pure solution experiment. The spectra were processed in Metaboanalyst 5.0, XCMS, and MZmine2 using the following modules: 1. Metaboanalyst: peak picking (min\_peakwidth - 6.0, max\_peakwidth - 25.0, ppm - 10, mzdifff - 0.01, snthresh - 10.0, prefilter - 6.0, noise -  $1E7$ , value of prefilter -  $1E7$ ), peak alignment and peak annotation settings were kept default; 2. XCMS: noise filtration was set to  $1E6$ , and rest of the settings were kept default; 3. MZmine2: mass detector (centroid),

ADAP chromatogram builder (min group size in #of scans - 5, group intensity threshold - 1E6, m/z tolerance 0-5 ppm), feature detection - chromatogram deconvolution, spectral deconvolution - hierarchical clustering, isotope peak grouper - deisotoped, filtering - peak filter, alignment - RANSAC aligner, and normalization – liner. Metaboanalyst, XCMS, and MZmine2 suggested compounds further confirmed with fragmentation pattern matching, and retention time compared with the authentic standards.

### **3.2.5. NMR analysis**

200  $\mu$ L of sample was aliquoted to an eppendorf tube with 50  $\mu$ L buffer X (750 mM phosphate buffer with 5 mM DSS and 10% D<sub>2</sub>O). The mixture was vortexed for 1 minute and centrifuged at 10000 rpm for 5 minutes at 4°C. <sup>1</sup>H NMR spectra were obtained on a Bruker AVANCE III 700 MHz spectrometer (Bruker Biospin, Rheinstetten, Germany) equipped with a triple resonance 5 mm CryoProbe. Samples were acquired with automation using a SampleJet sample changer. The samples were stored at 5 - 10 °C and each sample was pre-warmed to 25 °C before insertion into the spectrometer. Spectra for metabolomics analysis were acquired using a 1D 1H NOESY (noseypr1d), with 2s pre-saturation pulse for water suppression, a 50 ms mixing time with water saturation, and a 4s acquisition time. A sweep width of 12 ppm was used. Spectra with poor line shape or linewidths greater than 1 Hz were reacquired. For profiling with Chenomx, the NOESY spectra were processed with exponential line broadening such that the DSS peak width was 1 Hz, and manual baseline correction was applied. For manual profiling, samples were quantified using the Chenomx NMR Suite version 8 (Chenomx, Inc. Alberta, Canada) using a combination of the software-provided 700 MHz compound library and an in-house compound library acquired at 700 MHz.

### **3.2.6. Sensitivity and Specificity**

Sensitivity and specificity were calculated using the following formulas:

sensitivity = True positive / (True positive + False negative), and

specificity = True negative / (False positive + True negative).

### **3.2.7. Statistical analysis**

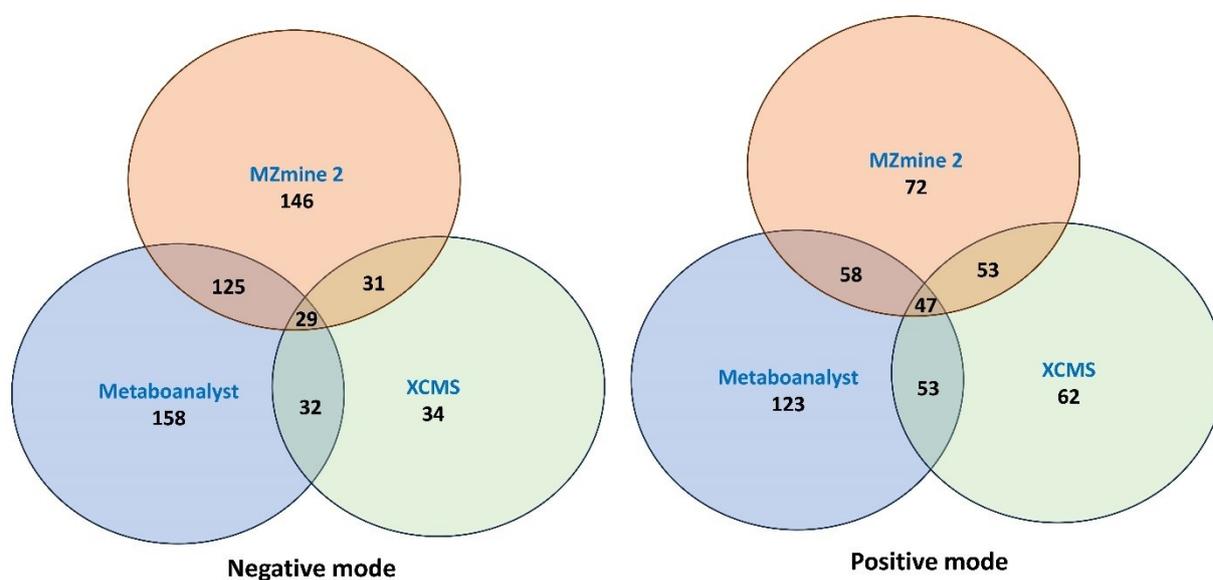
Blank was considered as control for the defined mixture and single pure lysine experiments. Metabolites were determined by comparing the experimental and control groups. Statistical test was performed using Welch's t-test comparing the peak area of each metabolite in control and treated samples.

## **3.3. Results**

### 3.3.1. Benchmarking: Mixture of compounds

#### 3.3.1.1. Features determination

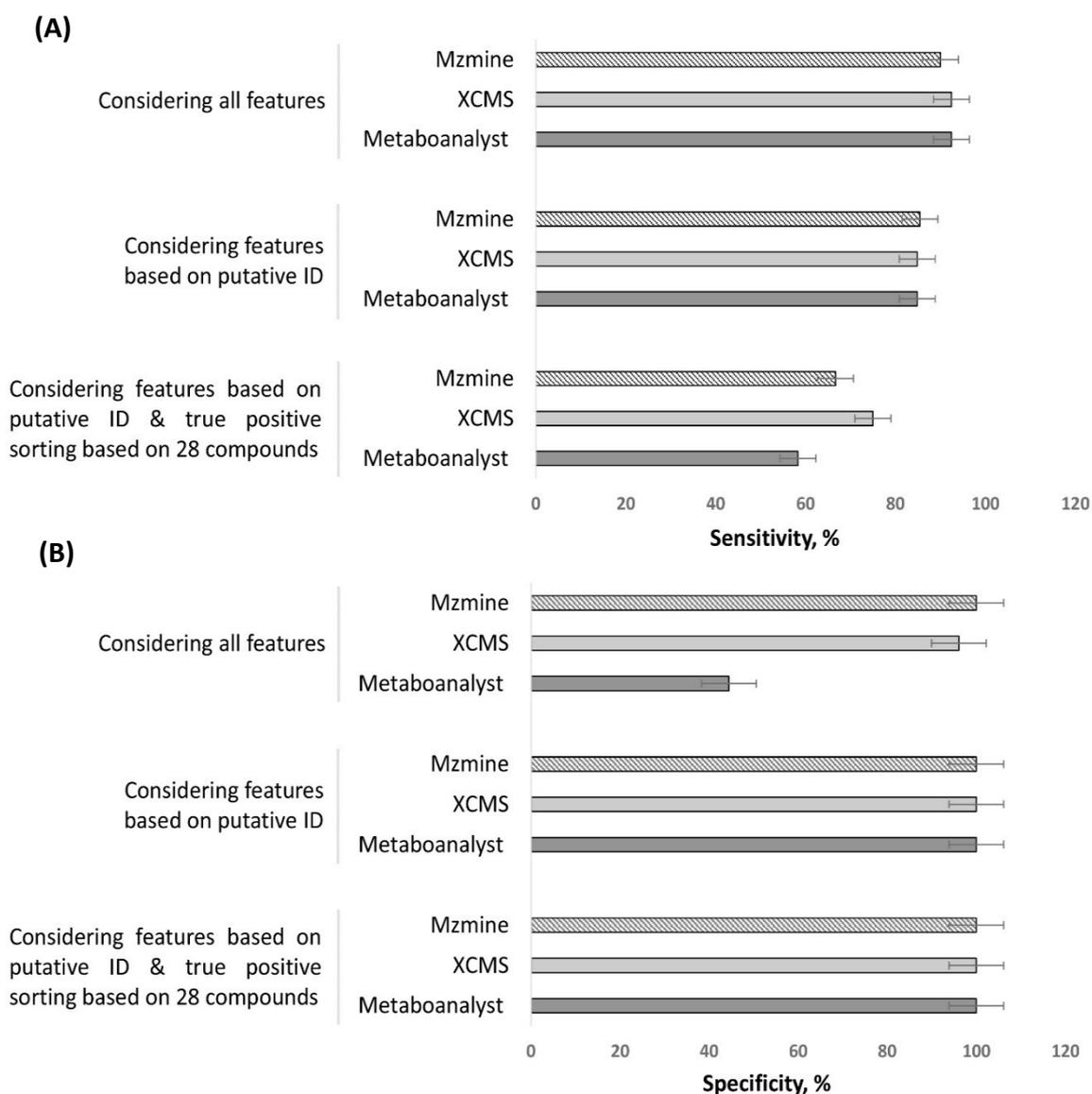
Metaboanalyst 5.0, XCMS, and MZmine2 software packages were used to detect the features in both ionization modes (negative and positive). Figure 3.1 exhibits the number of features determined by all three software. Metaboanalyst identified 158 and 123 features in negative and positive modes respectively. XCMS determined 34 features (in negative mode) and 62 features (in positive mode), which is relatively less than the number of features suggested by Metaboanalyst. In negative mode, MZmine2 exhibited 146 distinct features, with 125 overlapping with the features of Metaboanalyst (Figure 3.1). Conversely, in positive mode, MZmine2 displayed 72 features, a number in proximity to the 62 features observed by XCMS. There is an overlap of 125 (in negative mode) and 58 (in positive mode) features between Metaboanalyst 5.0 and MZmine2. Metaboanalyst 5.0 and XCMS exhibited an intersection of 32 and 53 features in negative and positive mode respectively. A total of 31 features (in negative mode) and 53 features (in positive mode) shared by MZmine2 and XCMS. Moreover, the small overlapping section signifies a consistent detection of 29 features in negative mode and 47 features in positive mode within all three software.



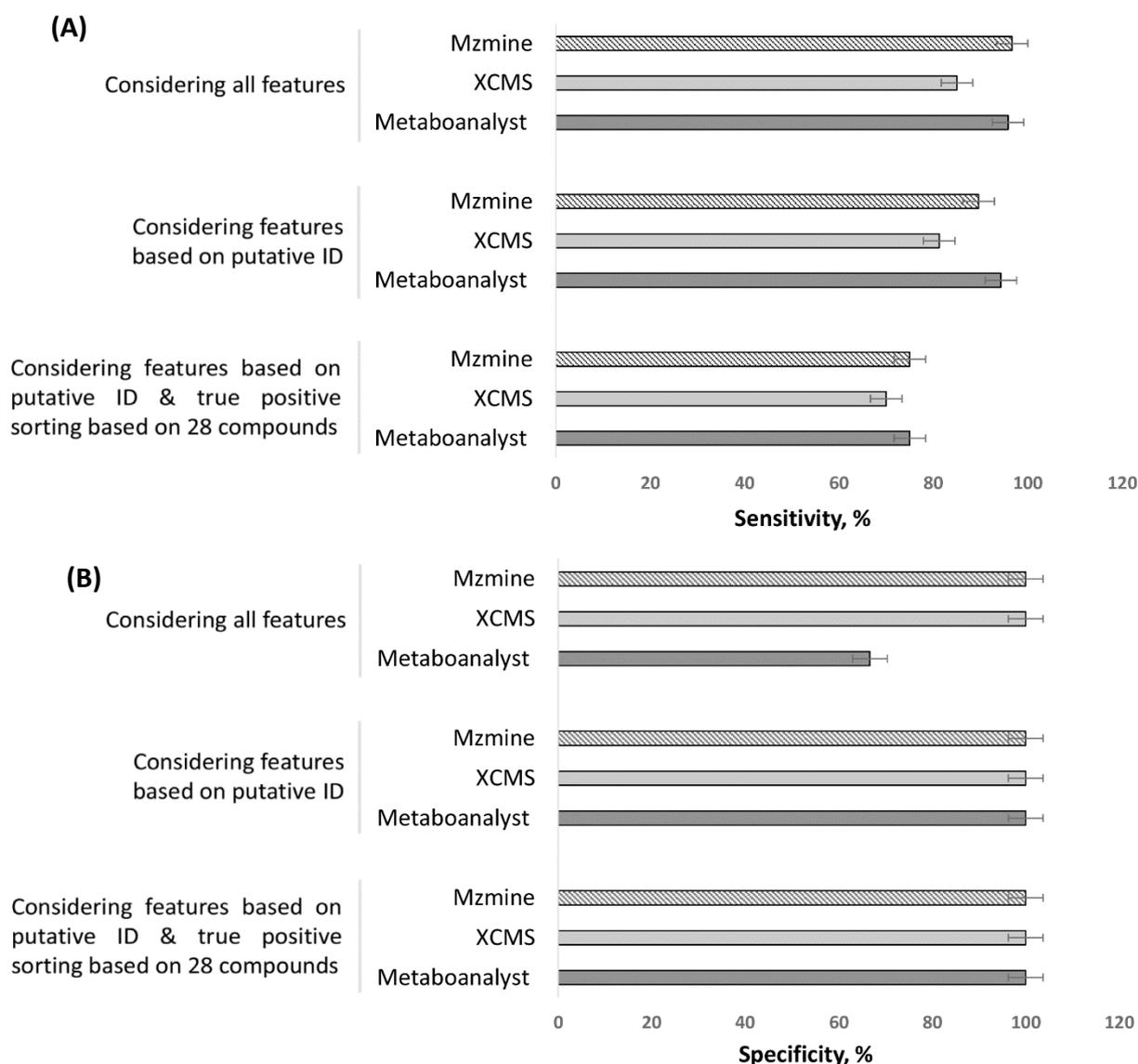
**Figure 3.1.** Venn diagram represents the number of features determined by Metaboanalyst, XCMS and MZmine2. Synthetic mixture was analyzed by LC-HRMS in both negative and positive mode.

### 3.3.1.2. Sensitivity and specificity of Metaboanalyst, XCMS and MZmine2

A set of 28 compounds was used to prepare the synthetic mixture. The exact mass and retention time were manually checked for each compound. No exact mass and retention time were detected for 8 compounds (benzyl alcohol, phenol, tyramine, phenylalanine, spermidine, 2,3-butanediol, deoxyadenosine monophosphate, and valeric acid) in negative mode and 4 compounds (2,3-butanediol, deoxyadenosine monophosphate, pyruvic acid, and valeric acid) in positive mode. All these compounds were considered as true negatives. Among the pool of 28 compounds, Metaboanalyst identified 14 compounds in negative mode and 15 compounds in positive mode, followed by XCMS detected 14 and 18 compounds in negative and positive mode respectively. Meanwhile, MZmine2 was able to determine 15 compounds in negative mode and 16 compounds in positive mode. Sensitivity and specificity were calculated in three ways, including a) all features, b) putative ID (PID), and c) putative ID & true positive (PID28). Sensitivity and specificity comparisons were displayed in Figure 3.2 and Figure 3.3. Considering all features and PID, the sensitivity of all the software was similar in positive mode. Variability was observed while calculation was conducted based on PID28, XCMS (75%) provided better sensitivity compared to Metaboanalyst (58.33%) and MZmine2 (66.67%). The specificity of Metaboanalyst in the criteria of considering all features is 44.44%, which is quite lower than XCMS (96.09%) and MZmine2 (100%). Considering PID and PID28, all these software provided 100% specificity in positive mode (Figure 3.2). In negative mode on all criteria, the sensitivity of three-software's exhibited similar results. Metaboanalyst, XCMS, and MZmine2 provided 100% specificity on PID and PID28 in negative mode (Figure 3.3). Nonetheless, Metaboanalyst demonstrated a specificity rate of 66.67%, whereas other two software yielded a higher specificity rate of 100% (Figure 3.3).



**Figure 3.2.** (A) Sensitivity and (B) specificity of Metaboanalyst, XCMS, and MZmine2 in positive mode. Features were accounted for in three ways (considering all features, considering features based on putative ID, and considering features based on putative ID & true positive sorting based on 28 compounds) to calculate the sensitivity and specificity.



**Figure 3.3.** (A) Sensitivity and (B) specificity of Metaboanalyst, XCMS, and MZmine2 in negative mode. Features were accounted for in three ways (considering all features, considering features based on putative ID, and considering features based on putative ID & true positive sorting based on 28 compounds) to calculate the sensitivity and specificity.

### 3.3.2. Benchmarking: pure ( $\geq 98\%$ ) lysine solution

#### 3.3.2.1. Results of lysine sample on different platform

According to NMR analysis, two unknown peaks were detected at 2.05 ppm and 3.95 ppm (Table 3.1). The intensity of the unknown contaminant is 0.26% compared to the lysine. Few other compounds such as methanol, ethanol, and tert-butyl alcohol were detected in the lysine sample. Besides NMR analysis, 100  $\mu\text{M}$  lysine was run in Orbitrap. There were 15 and 24 features detected in LC-MS and LC-MS/MS analysis respectively (Table 3.2 & 3.3).

Metaboanalyst, XCMS, and MZMine2 suggested one to fourteen compounds for each feature. Chemical shifts of each suggested compound were retrieved from NMRdb or HMDB. There was no potential match of the chemical shifts for those compounds (except lysine) compared with the actual lysine sample. No unknown contaminants were also identified in HSQC and TOCSY analysis.

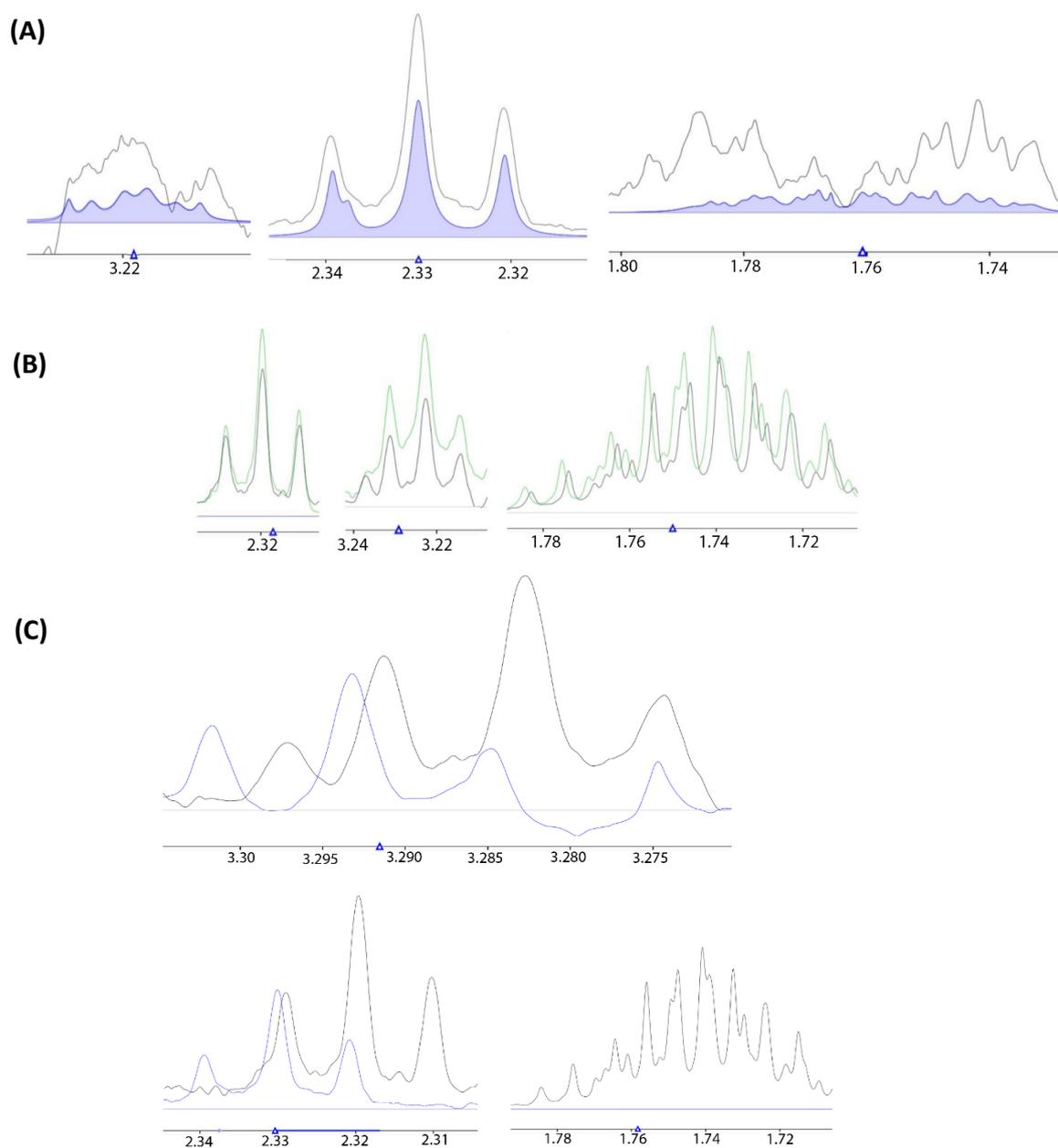
**Table 3.1.** List of chemical shifts and features of compounds determined in concentrated lysine (85 mM) using <sup>1</sup>H-NMR analysis.

ppm	Peak pattern	Compound	number of H	Area of peak	concentration, mM	Percent
0	singlet	DSS	9	100	1	1.176471
3.35	singlet	Methanol	3	0.0798	0.002394	0.002816
1.44	multiplet	Lysine	1	914.6225	82.31603	96.84238
1.5	multiplet	Lysine	1	911.0067	81.9906	96.45953
1.72	multiplet	Lysine	2	1846.519	83.09335	97.75688
1.9	multiplet	Lysine	2	1827.945	82.25751	96.77354
3.01	triplet	Lysine	2	1876.394	84.43773	99.3385
3.75	triplet	Lysine	1	943.3044	84.8974	99.87929
1.17	triplet	Buffer X, ethanol	2	0.3986	0.017937	0.021102
3.65	quartet	Buffer X, ethanol	3	0.5346	0.016038	0.018868
8.445	singlet	buffer x	1	0.0221	0.001989	0.00234
1.245	singlet	tert-butyl alcohol	9	0.0976	0.000976	0.001148
2.05	triplet	Unknown	1	2.4737	0.222633	0.261921
3.95	triplet	Unknown	1	1.6566	0.149094	0.175405

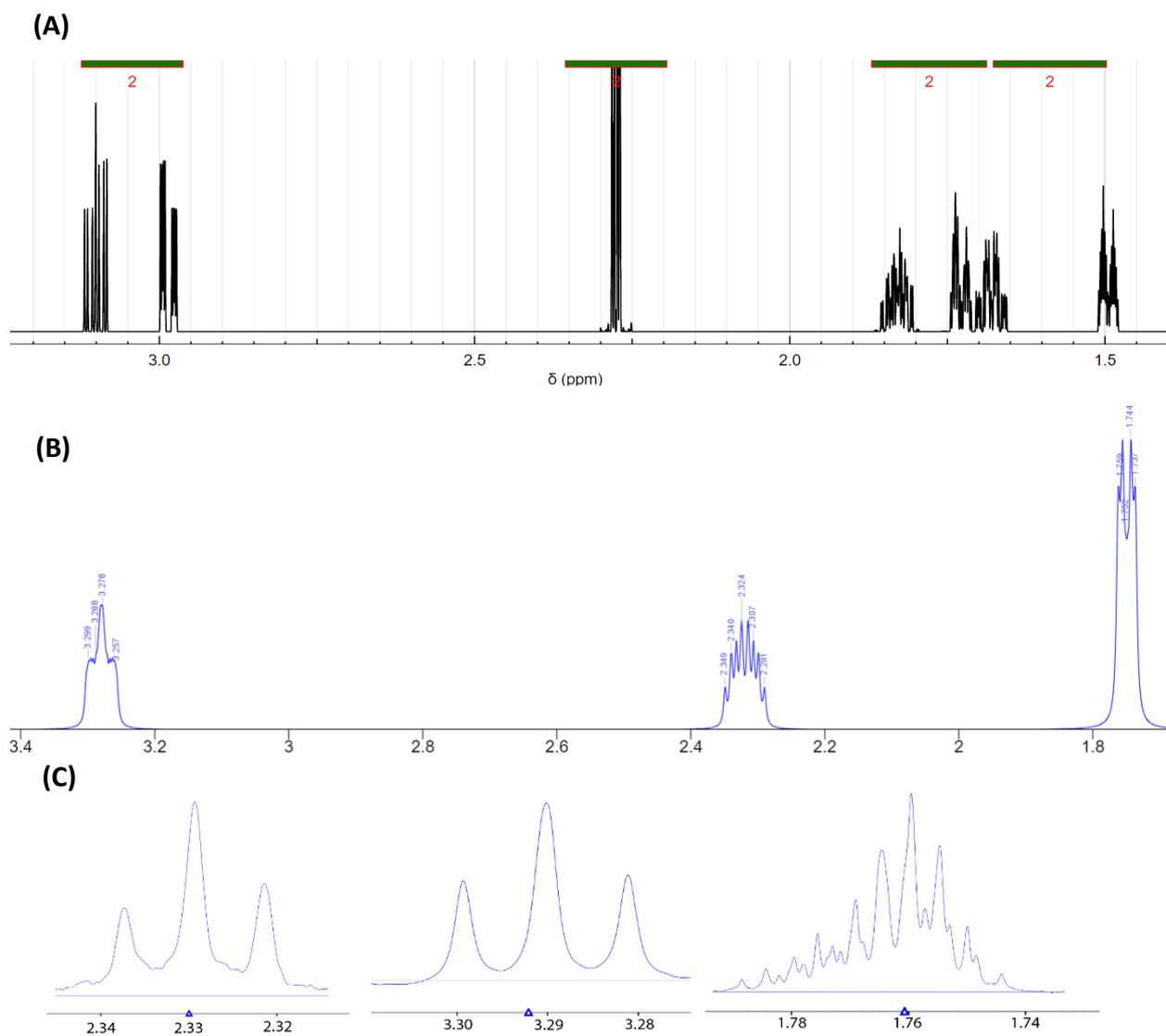
**Table 3.2.** List of features determined from 100 μM lysine using LC-MS analysis.

Software	Mode	m/z	p-value	Compound
Metaboanalyst	Neg	96.9601	5.11E-10	Sulfate
Metaboanalyst	Neg	145.0973	0	L-Lysine; D-Lysine; (3S)-3,6-Diaminohexanoate; (3S,5S)-3,5-Diaminohexanoate; 2,6-diaminohexanoic acid;
Metaboanalyst	Pos	84.0813	0	(+)-2,3-Dihydro-3-methyl-1H-pyrrole; Pentanenitrile
Metaboanalyst	Pos	240.1804	8.08E-08	5-Ethyl-4-methyl-2-octylthiazole; 2-Octyl-4-propylthiazole
Metaboanalyst	Pos	100.0761	7.36E-05	2-Piperidinone; 1-Pyrrolidinecarboxaldehyde; 2,5-Dihydro-2,4-dimethyloxazole; (2R)-2-Hydroxy-2-methylbutanenitrile;
Metaboanalyst	Pos	131.0895	0	3-Methylcyclohexanethiol; Carbachol; Pilocarpic acid; L-Pilocarpic acid; N4-Acetylamino-butanol; D-Pilocarpic acid; Vigabatrin; Pilocarpic acid; L-Pilocarpic acid; N4-Acetylamino-butanol; D-Pilocarpic acid; Vigabatrin;
Metaboanalyst	Pos	147.1124	0	acid; N4-Acetylamino-butanol; D-Pilocarpic acid; Vigabatrin;
Metaboanalyst	Pos	136.9894	2.84E-04	Methyl 2-propenyl selenide;
Metaboanalyst	Pos	169.0946	0	Pyridoxamine;
XCMS	Pos	316.9474	3.11E-14	Mitotane, TDE, 2-(m-Chlorophenyl)-2-(p-chlorophenyl)-1,1-dichloroethane, 1-Chloro-2,2-bis(4'-chlorophenyl)ethylene
MZMine	Neg	112.98		acetylenedicarboxylate
MZmine	Pos	147.1127		L-lysine (<i>S</i>)-nipecotate; (<i>R</i>)-nipecotate; D-pipecolate; L-pipecolate; vigabatrin; 4-acetamidobutanol; (<i>N</i>)-methyl-L-proline; (<i>N</i>)-(3-oxobutan-2-yl)acetamide; 4-methyl-proline; (2S,4Z)-2-aminohex-4-enoate; (1S,2S)-
MZmine	Pos	130.0861		coronamate; nipecotate; cycloleucine; pipecolate
MZmine	Pos	84.08131		cyclopentylamine
MZmine	Pos	118.9427		chromate
MZmine	Pos	169.0946		cyclo(glycyl-L-leucyl)





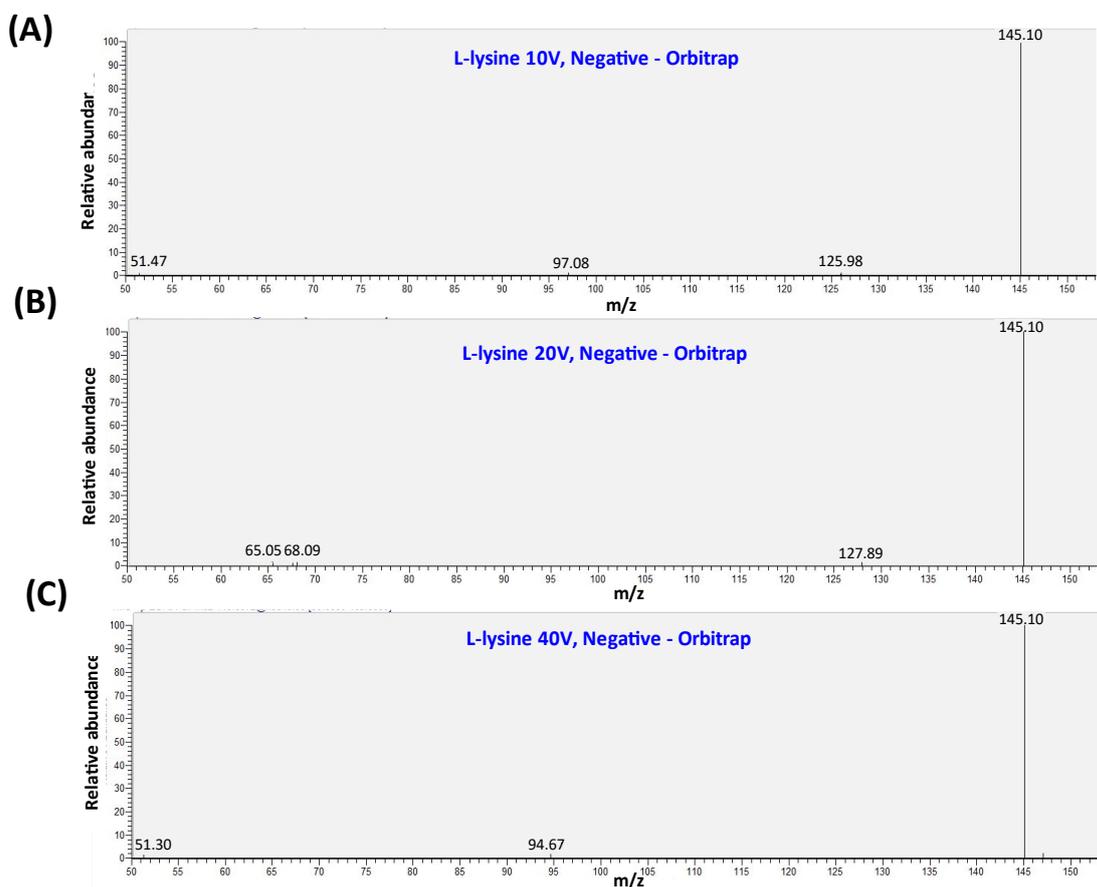
**Figure 3.4.** Presence of 2-piperidinone in a concentrated lysine solution determined using resin-based column chromatography and  $^1\text{H-NMR}$  analysis. (A) Profiling of 2-piperidinone using Chenomx. (B) Overlay of spiked 2-piperidinone. The concentration of 2-piperidinone before and after spiking was  $40.2\ \mu\text{M}$  and  $50.8\ \mu\text{M}$  respectively. Black and green color denote before and after spiking respectively. (C) Overlay of 2-piperidinone before and after separation from concentrated lysine solution. Black and blue color indicate before and after separation respectively. Chemical shift of 2-piperidinone at  $1.74\ \text{ppm}$  was submerged by concentrated lysine.



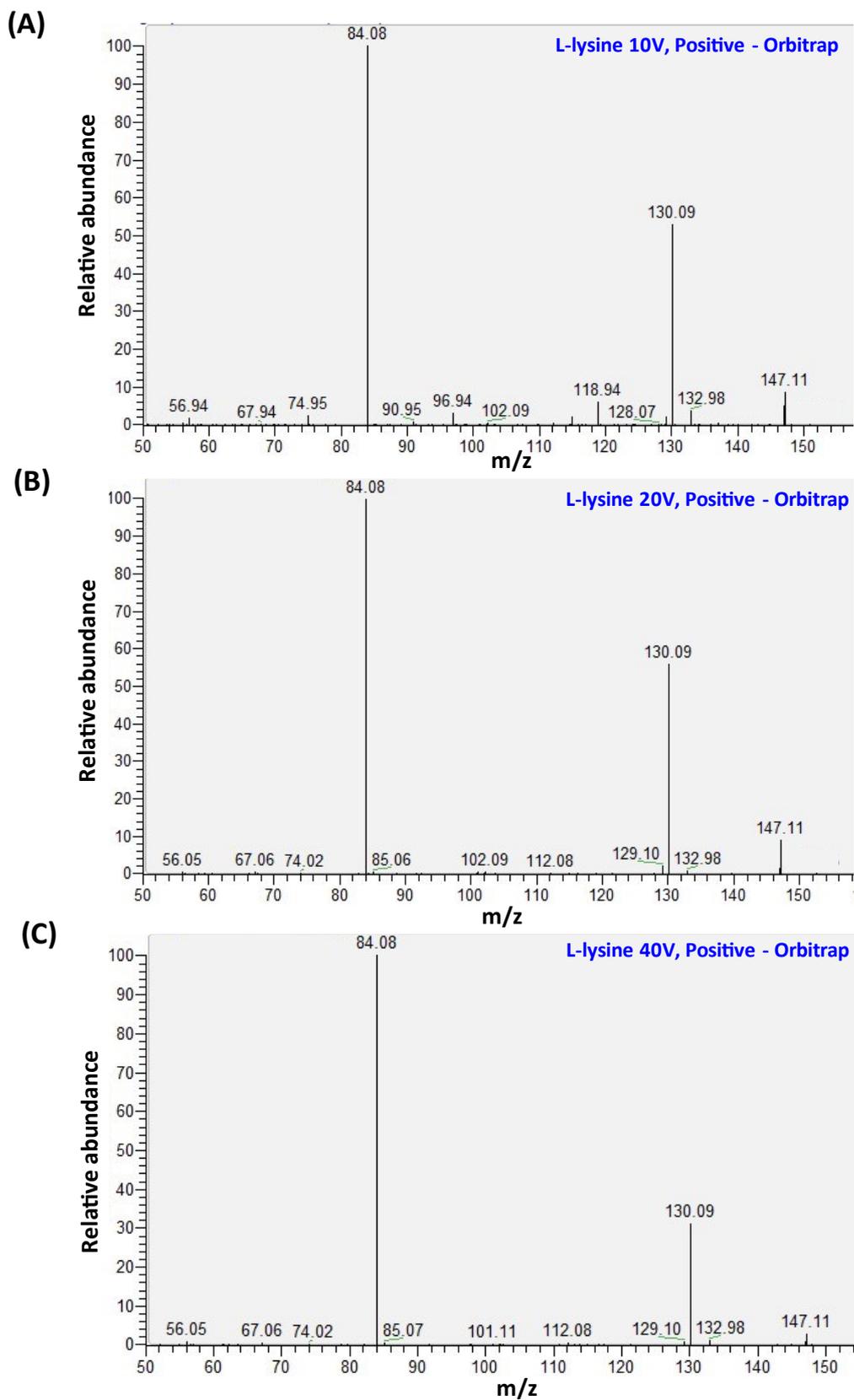
**Figure 3.5.** Comparison between predicted and actual chemical shifts of 2-piperidinone. (A) & (B) represents the predicted chemical shifts of 2-piperidinone generated using NMRdb and HMDB respectively. (C) exhibits the actual chemical shifts of 2-piperidinone.

### 3.3.2.3. False positive results

L-lysine were run in Orbitrap instrument at different collision energies, including 10V, 20V and 40V in both positive and negative ionization mode. All the spectra in negative and positive ionization mode were displayed in Figure 3.6 and Figure 3.7 respectively.



*Figure 3.6. Experimental spectra of L-lysine at different collision energies in negative ionization mode. (A) 10V, (B) 20V, and (C) 40V.*



**Figure 3.7.** Experimental spectra of L-lysine at different collision energies in positive ionization mode. (A) 10V, (B) 20V, and (C) 40V.

MS-data analysis was conducted using Metaboanalyst, XCMS, and MZmine2, and the results were presented in Table 3.2-3.3. All these software works by extracting and annotating the features. Metaboanalyst, XCMS, and MZmine2 together extracted 15-24 features for  $\geq 98\%$  pure lysine. An interesting observation was found after comparing L-lysine spectra (including the fragmentation) with the results of MS-data analyzing software (Table 3.4). Metaboanalyst, XCMS, and MZmine2 considered the daughter ions of L-lysine as individual precursor ions and afterward annotated those ions separately. Moreover, these software considered the isotopes and adducts of these daughter ions, and suggested more compounds, indeed these all are false positive results.

**Table 3.4.** Comparison between MS-analyzing software results and L-lysine spectra to determine false positive results.

m/z	Isotopes	Adducts	Comments	False positive
84.08122			Daughter Ion	Yes
129.1023			A low intensity ion beside 130.086 in Figure 3.7 (C)	Yes
130.086	$[M]^+$	$[M+H-NH_3]^+$ 146.105	Daughter Ion (146.105) + H - $NH_3$ = 130.086	Yes
131.0893	$[M+1]^+$	$[M+H]^+$	Daughter Ion (130.086) + H = 131.0893	Yes
145.0971			Precursor Ion	Yes
147.1123	$[M]^+$	$[M+H]^+$ 146.105	Precursor Ion (146.105) + H = 147.1123	No
148.1156	$[M+1]^+$		Precursor Ion (147.1123) + 1 = 148.1156	Yes
191.076		$[M+2Na-H]^+$	Precursor Ion (146.105) + 2Na - H = 191.076	Yes
169.0941		$[M+Na+NH_3]^+$ 129.079	Daughter Ion (129.079) + Na + $NH_3$ = 169.0941	Yes

### 3.4. Discussion

False positive results significantly impact downstream analysis leading to erroneous results (Pirttilä et al., 2022). In this experiment, a synthetic mixture was analyzed to test the reproducibility of Metaboanalyst, XCMS, and MZmine2. All features identified by Metaboanalyst, XCMS, and MZmine2 were subject to thorough manual checking in the chromatogram. The features determined by XCMS and MZmine2 were confirmed as positive by manual checking. Among the features suggested by Metaboanalyst regarding synthetic mixture, a subset of 5-10% features didn't display any observable peaks by manual checking in the chromatogram, considered as false positive features. Upon consideration of PID28, the rate of sensitivity of Metaboanalyst, XCMS, and MZmine2 was low (Figure 3.2 & 3.3).

Technically, the measure of software's performance lies in its ability to achieve the expected results by suggesting only required features. In the analyzed synthetic mixture, XCMS detected 34 features in negative mode and 62 features in positive mode (Figure 3.1). Conversely, Metaboanalyst identified 158 features in negative mode and 123 features in positive mode, followed by MZmine2 determined 146 and 72 features in negative and positive mode respectively (Figure 3.1). According to the scale of a number of feature detections in response to a synthetic mixture (28 compounds), XCMS exhibited limited features with coverage of more compounds compared to Metaboanalyst, and MZmine2. Although the defined mixture encompasses 28 compounds, it is unclear whether more than 30 features were determined by Metaboanalyst, XCMS, and MZmine2 (Figure 3.1). Perhaps, these features represent the formation of new compounds while compounds were mixed to prepare the synthetic mixture. Another reason could be the purity (98-99%) of the compounds. To address these, a concentrated solution of a single compound, named lysine (purity  $\geq 98\%$ ), was run in NMR and Orbitrap. Applying the  $^1\text{H-NMR}$  technique, a few compounds such as methanol, ethanol, and tert-butyl alcohol were detected in the lysine solution. These compounds were reported as common NMR contaminants (Fulmer et al., 2010; Gottlieb et al., 1997). In analyzing the lysine solution, NMR was able to detect one unknown compound at 2.33 ppm and 3.22 ppm. Besides the NMR analysis, results of LC-MS and LC-MS/MS analysis suggested 15 and 24 features respectively for the lysine solution. Annotation of each feature reported multiple compounds, comprising in total of 56 compounds. Chemical shifts of 56 compounds were overlaid with the unknown compound detected in NMR regarding the lysine solution. Chemical shift comparison was unsuccessful in exploring the identity of the unknown compound. Hence, anionic resin-based column chromatography and  $^1\text{H-NMR}$  was able to determine the identity of the unknown compound, 2-piperidinone. Among the three software packages, only Metaboanalyst was able to determine the unknown compound, 2-piperidinone. The predicted chemical shifts of 2-piperidinone didn't align properly with the unknown chemical shifts of L-lysine [Figures 3.5 (A) & (B)]. This instance exemplifies the impact of a predicted chemical shift on the results. There is a high chance that prediction may provide wrong chemical shifts for compounds. Besides the chemical shift comparison, the MS-fragmentation pattern of lysine at different collision energy compared with the results of Metaboanalyst, XCMS, and MZmine2. All these software considered the daughter ions (129.079 and 84.08123) as precursor ions, and incorporated adducts and isotopes to annotate, which eventually suggested more false positive results.

Overall, the rate of sensitivity and specificity of all three software packages close in proximity. It is challenging to conclude which software performs well and gives less false positive results. In negative mode (regarding 28 compounds mixture), XCMS worked better in determining essential features to cover the compounds used to make the synthetic mixture. Although for the lysine solution, Metaboanalyst was able to accurately identify the unknown contaminant than XCMS and MZmine2. Indeed, different software packages have distinct advantages; some work better in negative mode, and some in positive mode. For a new biomarker discovery study, ease of use of Metaboanalyst and MZmine2 could be a good option. For comparison-based (disease vs non-diseased) studies, researchers can decide to use XCMS as it detects fewer false positive features than others. Metaboanalyst and XCMS are more user-friendly platforms than MZmine2. Hence, users will get more flexibility in MZmine2 to customize the workflows. Moreover, MZmine2 has a deisotope option which is not in Metaboanalyst and XCMS. MZmine2 provides the choice to manually select the databases (PubChem, HMDB, KEGG, and MetaCyc), while Metaboanalyst and XCMS are only limited to the HMDB and METLIN databases respectively. Metaboanalyst integrates multi-omics options such as enrichment analysis, pathway analysis, and linking gene metabolite interactions. Many users prefer Metaboanalyst due to the vast options of statistics and graphical interference. Besides these software packages, people also prefer to use MS-DIAL and SIRIUS for the LC-MS/MS analysis. The advantage of using MS-DIAL as it provides the option of comparing the fragmentation pattern with the experimental/in-silico databases. SIRIUS elucidates rule-based structural information from the fragmentation patterns. It ranks features and structures based on different scores such as confidence score, SIRIUS score, and zodiac score. SIRIUS declared that the identification rate is 70% for the challenging datasets (Dührkop et al., 2019). MS-DIAL and SIRIUS were not used in this experiment as the synthetic mixture was scanned in MS1 (full scan). MS-DIAL and SIRIUS work better while the dataset is in MS<sup>2</sup> scanned. The full scan provides more features than MS<sup>2</sup>. Guo and Huan et al. discussed and compared the results of full scan, data-dependent acquisition (DDA), and data-independent acquisition (DIA). The full scan allows a comprehensive view of the entire dataset. DDA prioritizes and selects the high intense parent ions for fragmentation in subsequent scans. DIA works within the selected m/z range for each scan. The choice of selection depends upon research objectives. Besides these options in untargeted metabolomics, to improve the filtering procedures researchers developing new algorithms and strategies, including IP4M (Liang et al., 2020), asari (Li et al., 2023), UmetaFlow (Kontou et al., 2023), metaX (Wen et al., 2017), openNAU (Sun et al., 2022), DecoMetDIA (Yin et al., 2019) and DIAMetAlyzer (Alka et al.,

2022), etc. To date, there is no study conducted to compare the reliability of these packages. More groundwork requires to compare the performance of IP4M, asari, UmetaFlow, metaX, openNAU, DecoMetDIA, DIAMetAlyzer with traditional software packages, including Metaboanalyst, XCMS, MZmine2, MS-DIAL, and SIRIUS. Therefore, the selection and combination of software packages should be carefully weighed in untargeted metabolomics upon the purpose of the study.

### **3.5. Conclusion**

Reproducibility in untargeted metabolomics is very challenging. XCMS provided less false positive features compared to Metaboanalyst and MZmine2 regarding the defined mixture. In contrast, for the single pure lysine solution, only Metaboanalyst identified the unknown contaminant (2-piperidinone). Moreover, all the MS-data analyzing software counted daughter ions of L-lysine as precursor ions, which resulted in false positive metabolites. Taken together, due to different aspects, it's worth considering multiple softwares for MS-based untargeted metabolomics data analysis.

## Chapter 4: Concluding remarks

### 4.1. Summary

LC-HRMS, GC-MS, and NMR are modern cutting-edge metabolomics techniques. Now-a-days these techniques are considered to be the gold standard for characterizing small organic molecules in targeted and untargeted manner. The inclusion of untargeted metabolomics and transcriptomics is a new area in genomics to answer biological questions.

In this thesis, I provided a brief overview of bacterial metabolite identification using untargeted metabolomics techniques. Chapter 2 is a reference framework for the characterization of metabolites and genes involved in the catechol detoxification pathway in *E. coli UMI46*. Many false positive results were detected while analyzing MS data. To see the reproducibility of results, a comparison was conducted using a synthetic mixture. The results of the comparison were discussed in Chapter 3. Regarding the defined mixture which comprises 28 compounds, the performance of XCMS was better in determining minimum features with coverage of more compounds. Under the benchmarking project, another experiment was performed with  $\geq 98\%$  pure lysine to identify the contaminants in it. Contaminants were determined using resin-based column chromatography and  $^1\text{H-NMR}$ . Among the three MS data analyzing software, it was apparent that only Metaboanalyst was able to determine the unknown contaminant (2-piperidinone) in the lysine solution. Metaboanalyst, XCMS, and MZmine2 suggested other metabolites (excluding 2-piperidinone and lysine) considered false positives. The outcomes of the MS analysis of the defined mixture and single pure solution experiment underscore the importance of not relying exclusively on one software result. Across the board, none of the MS data analyzing software provides accurate results. In an effort to achieve optimum output, it is required to use different MS-data analyzing software to compare the results, and finally validate the results with authentic standards.

### 4.2. Future directions

#### 4.2.1. Chapter 2

Genes that are shown involved in the catechol detoxification pathway require further validation using quantitative real-time polymerase chain reaction (qRT-PCR) and CRISPR-Cas9 technique. The metabolomics aspect posed the main challenges in this study, involving the application of untargeted approaches to characterize both known/unknown metabolites. Each of the software utilized for metabolomics analysis indicated over 400 features for each

sample, with the potential of multiple metabolites within each feature (exact mass). Consequently, the estimated number of metabolites for each sample was approximately 600~700. It was very time-intensive to compare the fragmentation pattern and chemical shifts for each of the annotated compounds. Prior to pursuing these approaches, the collection of fractions using preparative-HPLC and then running the fractions in different analytical instruments could reduce the amount of time, and effort required to characterize these metabolites. Moreover, besides the metabolomics and transcriptomics techniques, the inclusion of proteomics followed by isolation and characterization of enzymes using the western-blot technique would give more accurate results.

Sample collection for transcriptomics and metabolomics analysis occurred at different time points. Samples for transcriptomics analysis were deliberately obtained at the exponential phase to get the total RNA, where m-RNA was predominated over ribosomal RNA (rRNA) and transfer RNA (tRNA). Due to the onset of secondary metabolite production by bacteria during the stationary phase, samples for metabolomics analysis were collected at the 24 hrs time point. To examine whether *E. coli UMI46* strain is capable of producing catechol ortho-cleavage or meta-cleavage metabolites (such as cis, cis-muconic acid, 3-oxoadipate, and 2-hydroxymuconate semialdehyde), an optimal approach would be collecting samples for metabolomics analysis at several time points (2 hrs, 4 hrs, 8 hrs, 12 hrs and 24 hrs). It is worth considering that metabolites generated through ortho or meta cleavage route, enable bacteria to consume it and transform it into another metabolite.

### **4.2.2. Chapter 3**

It is necessary to compare the features of different software packages to get a reliable result. To determine the reproducibility accurately, it is required to design and prepare the synthetic mixture using isotope-labeled compounds. Moreover, there is a need to link fragmentation pattern prediction software (such as CFM-ID 4.0) with Metaboanalyst, XCMS, and MZmine2. It would be even more remarkable if these software packages used the algorithm of CFM-ID 4.0 to generate and compare fragmentation patterns with confidence scores when presenting any features or annotated compound. This strategy can reduce the time and effort needed to accomplish such activities manually. Moreover, developers of Metaboanalyst, XCMS, and MZmine2 may also think about including the algorithm of rule-based structure prediction options like SIRIUS. One score from fragmentation pattern comparison and another score from rule-based structure prediction may allow users to screen and sort the features

easily, which would be a more robust approach in MS-based untargeted metabolomics data analysis.

## References

- Aghapour, A. A., Moussavi, G., & Yaghmaeian, K. (2013). Biological degradation of catechol in wastewater using the sequencing continuous-inflow reactor (SCR). *Journal of Environmental Health Science & Engineering*, *11*(1), 3. <https://doi.org/10.1186/2052-336X-11-3>
- Albuquerque, B. R., Heleno, S. A., Oliveira, M. B. P. P., Barros, L., & Ferreira, I. C. F. R. (2021). Phenolic compounds: Current industrial applications, limitations and future challenges. *Food & Function*, *12*(1), 14–29. <https://doi.org/10.1039/d0fo02324h>
- Alka, O., Shanthamoorthy, P., Witting, M., Kleigrewe, K., Kohlbacher, O., & Röst, H. L. (2022). DIAMetAlyzer allows automated false-discovery rate-controlled analysis for data-independent acquisition in metabolomics. *Nature Communications*, *13*(1), 1347. <https://doi.org/10.1038/s41467-022-29006-z>
- Aono, R., Tsukagoshi, N., & Yamamoto, M. (1998). Involvement of outer membrane protein TolC, a possible member of the mar-sox regulon, in maintenance and improvement of organic solvent tolerance of *Escherichia coli* K-12. *Journal of Bacteriology*, *180*(4), 938–944. <https://doi.org/10.1128/JB.180.4.938-944.1998>
- Ardizzi, M., Ballarini, N., Cavani, F., Chiappini, E., Dal Pozzo, L., Maselli, L., & Monti, T. (2007). Environmentally friendly, heterogeneous acid and base catalysis for the methylation of catechol: Chances for the control of chemo-selectivity. *Applied Catalysis B: Environmental*, *70*(1–4), 597–605. <https://doi.org/10.1016/j.apcatb.2006.02.028>
- Aura, A.-M. (2008). Microbial metabolism of dietary phenolic compounds in the colon. *Phytochemistry Reviews*, *7*(3), 407–429. <https://doi.org/10.1007/s11101-008-9095-3>
- Aura, A.-M., O’Leary, K. A., Williamson, G., Ojala, M., Bailey, M., Puupponen-Pimiä, R., Nuutila, A. M., Oksman-Caldentey, K.-M., & Poutanen, K. (2002). Quercetin Derivatives Are Deconjugated and Converted to Hydroxyphenylacetic Acids but Not Methylated by Human Fecal Flora in Vitro. *Journal of Agricultural and Food Chemistry*, *50*(6), 1725–1730. <https://doi.org/10.1021/jf0108056>

- Baptista, J., Simões, M., & Borges, A. (2019). Effect of plant-based catecholic molecules on the prevention and eradication of *Escherichia coli* biofilms: A structure activity relationship study. *International Biodeterioration & Biodegradation*, *141*, 101–113.  
<https://doi.org/10.1016/j.ibiod.2018.02.004>
- Baral, B., Akhgari, A., & Metsä-Ketelä, M. (2018). Activation of microbial secondary metabolic pathways: Avenues and challenges. *Synthetic and Systems Biotechnology*, *3*(3), 163–178.  
<https://doi.org/10.1016/j.synbio.2018.09.001>
- Barbuto Ferraiuolo, S., Cammarota, M., Schiraldi, C., & Restaino, O. F. (2021). Streptomycetes as platform for biotechnological production processes of drugs. *Applied Microbiology and Biotechnology*, *105*(2), 551–568. <https://doi.org/10.1007/s00253-020-11064-2>
- Bertini, I., Briganti, F., & Scozzafava, A. (1994). Aliphatic and aromatic inhibitors binding to the active site of catechol 2,3-dioxygenase from *Pseudomonas putida* mt-2. *FEBS Letters*, *343*(1), 56–60. [https://doi.org/10.1016/0014-5793\(94\)80606-3](https://doi.org/10.1016/0014-5793(94)80606-3)
- Blattner, F. R., Plunkett, G., Bloch, C. A., Perna, N. T., Burland, V., Riley, M., Collado-Vides, J., Glasner, J. D., Rode, C. K., Mayhew, G. F., Gregor, J., Davis, N. W., Kirkpatrick, H. A., Goeden, M. A., Rose, D. J., Mau, B., & Shao, Y. (1997). The complete genome sequence of *Escherichia coli* K-12. *Science (New York, N.Y.)*, *277*(5331), 1453–1462.  
<https://doi.org/10.1126/science.277.5331.1453>
- Blaut, M., Schoefer, L., & Braune, A. (2003). Transformation of flavonoids by intestinal microorganisms. *International Journal for Vitamin and Nutrition Research. Internationale Zeitschrift Fur Vitamin- Und Ernährungsforschung. Journal International De Vitaminologie Et De Nutrition*, *73*(2), 79–87. <https://doi.org/10.1024/0300-9831.73.2.79>
- Bode, H. B., Bethe, B., Höfs, R., & Zeeck, A. (2002). Big Effects from Small Changes: Possible Ways to Explore Nature's Chemical Diversity. *ChemBioChem*, *3*(7), 619.  
[https://doi.org/10.1002/1439-7633\(20020703\)3:7<619::AID-CBIC619>3.0.CO;2-9](https://doi.org/10.1002/1439-7633(20020703)3:7<619::AID-CBIC619>3.0.CO;2-9)
- Bornscheuer, U. T. (2002). Microbial carboxyl esterases: Classification, properties and application in biocatalysis. *FEMS Microbiology Reviews*, *26*(1), 73–81. <https://doi.org/10.1111/j.1574-6976.2002.tb00599.x>

- Boruta, T., & Bizukoje, M. (2016). Induction of secondary metabolism of *Aspergillus terreus* ATCC 20542 in the batch bioreactor cultures. *Applied Microbiology and Biotechnology*, *100*(7), 3009–3022. <https://doi.org/10.1007/s00253-015-7157-1>
- Bowman, A. P., Blakney, G. T., Hendrickson, C. L., Ellis, S. R., Heeren, R. M. A., & Smith, D. F. (2020). Ultra-High Mass Resolving Power, Mass Accuracy, and Dynamic Range MALDI Mass Spectrometry Imaging by 21-T FT-ICR MS. *Analytical Chemistry*, *92*(4), 3133–3142. <https://doi.org/10.1021/acs.analchem.9b04768>
- Brader, G., Compant, S., Mitter, B., Trognitz, F., & Sessitsch, A. (2014). Metabolic potential of endophytic bacteria. *Current Opinion in Biotechnology*, *27*, 30–37. <https://doi.org/10.1016/j.copbio.2013.09.012>
- Brakhage, A. A. (2013). Regulation of fungal secondary metabolism. *Nature Reviews Microbiology*, *11*(1), 21–32. <https://doi.org/10.1038/nrmicro2916>
- Breinig, S., Schiltz, E., & Fuchs, G. (2000). Genes involved in anaerobic metabolism of phenol in the bacterium *Thauera aromatica*. *Journal of Bacteriology*, *182*(20), 5849–5863. <https://doi.org/10.1128/JB.182.20.5849-5863.2000>
- Buchmann, D., Schwabe, M., Weiss, R., Kuss, A. W., Schaufler, K., Schlüter, R., Rödiger, S., Guenther, S., & Schultze, N. (2023). Natural phenolic compounds as biofilm inhibitors of multidrug-resistant *Escherichia coli* – the role of similar biological processes despite structural diversity. *Frontiers in Microbiology*, *14*, 1232039. <https://doi.org/10.3389/fmicb.2023.1232039>
- Campbell, W. H. (1999). NITRATE REDUCTASE STRUCTURE, FUNCTION AND REGULATION: Bridging the Gap between Biochemistry and Physiology. *Annual Review of Plant Physiology and Plant Molecular Biology*, *50*, 277–303. <https://doi.org/10.1146/annurev.arplant.50.1.277>
- Cavalli, A., Salvatella, X., Dobson, C. M., & Vendruscolo, M. (2007). Protein structure determination from NMR chemical shifts. *Proceedings of the National Academy of Sciences*, *104*(23), 9615–9620. <https://doi.org/10.1073/pnas.0610313104>

- Chavez, R. G., Alvarez, A. F., Romeo, T., & Georgellis, D. (2010). The physiological stimulus for the BarA sensor kinase. *Journal of Bacteriology*, *192*(7), 2009–2012.  
<https://doi.org/10.1128/JB.01685-09>
- Chen, X., Liu, X., Wang, H.-B., Cui, K.-P., Weerasooriya, R., He, S.-L., Li, G.-H., Pan, J., & Zhou, K. (2020). Ce<sup>3+</sup> triggers fenton-like processes in neutral solutions for effective catechol degradation. *Environmental Engineering Research*, *27*(1), 200519–0.  
<https://doi.org/10.4491/eer.2020.519>
- Cheng, K.-J., Jones, G. A., Simpson, F. J., & Bryant, M. P. (1969). Isolation and identification of rumen bacteria capable of anaerobic rutin degradation. *Canadian Journal of Microbiology*, *15*(12), 1365–1371. <https://doi.org/10.1139/m69-247>
- Chu, E. K., Kilic, O., Cho, H., Groisman, A., & Levchenko, A. (2018). Self-induced mechanical stress can trigger biofilm formation in uropathogenic *Escherichia coli*. *Nature Communications*, *9*(1), 4087. <https://doi.org/10.1038/s41467-018-06552-z>
- Clavel, T., Fallani, M., Lepage, P., Levenez, F., Mathey, J., Rochet, V., Sérézat, M., Sutren, M., Henderson, G., Bennetau-Pelissero, C., Tondu, F., Blaut, M., Doré, J., & Coxam, V. (2005). Isoflavones and functional foods alter the dominant intestinal microbiota in postmenopausal women. *The Journal of Nutrition*, *135*(12), 2786–2792.  
<https://doi.org/10.1093/jn/135.12.2786>
- Coble, J. B., & Fraga, C. G. (2014). Comparative evaluation of preprocessing freeware on chromatography/mass spectrometry data for signature discovery. *Journal of Chromatography A*, *1358*, 155–164. <https://doi.org/10.1016/j.chroma.2014.06.100>
- Coelho, L. M., Rezende, H. C., Coelho, L. M., De Sousa, P. A. R., Melo, D. F. O., & Coelho, N. M. M. (2015). Bioremediation of Polluted Waters Using Microorganisms. In N. Shiomi (Ed.), *Advances in Bioremediation of Wastewater and Polluted Soil*. InTech.  
<https://doi.org/10.5772/60770>
- Corrêa, T. A. F., Rogero, M. M., Hassimotto, N. M. A., & Lajolo, F. M. (2019). The Two-Way Polyphenols-Microbiota Interactions and Their Effects on Obesity and Related Metabolic Diseases. *Frontiers in Nutrition*, *6*, 188. <https://doi.org/10.3389/fnut.2019.00188>

- Cui, Q., Lewis, I. A., Hegeman, A. D., Anderson, M. E., Li, J., Schulte, C. F., Westler, W. M., Eghbalnia, H. R., Sussman, M. R., & Markley, J. L. (2008). Metabolite identification via the Madison Metabolomics Consortium Database. *Nature Biotechnology*, *26*(2), 162–164. <https://doi.org/10.1038/nbt0208-162>
- d'Oelsnitz, S., Kim, W., Burkholder, N. T., Javanmardi, K., Thyer, R., Zhang, Y., Alper, H. S., & Ellington, A. D. (2022). Using fungible biosensors to evolve improved alkaloid biosyntheses. *Nature Chemical Biology*, *18*(9), 981–989. <https://doi.org/10.1038/s41589-022-01072-w>
- Dass, C. (2007). *Fundamentals of Contemporary Mass Spectrometry*. John Wiley & Sons, Inc. <https://doi.org/10.1002/0470118490>
- Davies, D. (2003). Understanding biofilm resistance to antibacterial agents. *Nature Reviews Drug Discovery*, *2*(2), 114–122. <https://doi.org/10.1038/nrd1008>
- de Bont, J. (1998). Solvent-tolerant bacteria in biocatalysis. *Trends in Biotechnology*, *16*(12), 493–499. [https://doi.org/10.1016/S0167-7799\(98\)01234-7](https://doi.org/10.1016/S0167-7799(98)01234-7)
- Degnan, D. J., Bramer, L. M., Flores, J. E., Paurus, V. L., Corilo, Y. E., & Clendinen, C. S. (2023). *Evaluating Retention Index Score Assumptions to Refine GC-MS Metabolite Identification* [Preprint]. Bioinformatics. <https://doi.org/10.1101/2023.01.26.525730>
- Delgado, A. M., Issaoui, M., & Chammem, N. (2019). Analysis of Main and Healthy Phenolic Compounds in Foods. *Journal of AOAC International*, *102*(5), 1356–1364. <https://doi.org/10.5740/jaoacint.19-0128>
- Deller, S., Macheroux, P., & Sollner, S. (2008). Flavin-dependent quinone reductases. *Cellular and Molecular Life Sciences: CMLS*, *65*(1), 141–160. <https://doi.org/10.1007/s00018-007-7300-y>
- Desmarchelier, P., & Fegan, N. (2002). ESCHERICHIA COLI. In *Encyclopedia of Dairy Sciences* (pp. 948–954). Elsevier. <https://doi.org/10.1016/B0-12-227235-8/00158-9>
- Díaz, E., Ferrández, A., Prieto, M. A., & García, J. L. (2001). Biodegradation of aromatic compounds by Escherichia coli. *Microbiology and Molecular Biology Reviews: MMBR*, *65*(4), 523–569, table of contents. <https://doi.org/10.1128/MMBR.65.4.523-569.2001>

- Díaz-Gómez, R., Toledo-Araya, H., López-Solís, R., & Obreque-Slier, E. (2014). Combined effect of gallic acid and catechin against *Escherichia coli*. *LWT - Food Science and Technology*, *59*(2), 896–900. <https://doi.org/10.1016/j.lwt.2014.06.049>
- Ding, B., Schmeling, S., & Fuchs, G. (2008). Anaerobic Metabolism of Catechol by the Denitrifying Bacterium *Thauera aromatica*—A Result of Promiscuous Enzymes and Regulators? *Journal of Bacteriology*, *190*(5), 1620–1630. <https://doi.org/10.1128/JB.01221-07>
- Dong, Z., & Chen, Y. (2013). Transcriptomics: Advances and approaches. *Science China Life Sciences*, *56*(10), 960–967. <https://doi.org/10.1007/s11427-013-4557-2>
- Dührkop, K., Fleischauer, M., Ludwig, M., Aksenov, A. A., Melnik, A. V., Meusel, M., Dorrestein, P. C., Rousu, J., & Böcker, S. (2019). SIRIUS 4: A rapid tool for turning tandem mass spectra into metabolite structure information. *Nature Methods*, *16*(4), 299–302. <https://doi.org/10.1038/s41592-019-0344-8>
- Dutta, N. K., Tornheim, J. A., Fukutani, K. F., Paradkar, M., Tiburcio, R. T., Kinikar, A., Valvi, C., Kulkarni, V., Pradhan, N., Shivakumar, S. V. B. Y., Kagal, A., Gupte, A., Gupte, N., Mave, V., Gupta, A., Andrade, B. B., & Karakousis, P. C. (2020). Integration of metabolomics and transcriptomics reveals novel biomarkers in the blood for tuberculosis diagnosis in children. *Scientific Reports*, *10*(1), 19527. <https://doi.org/10.1038/s41598-020-75513-8>
- Ebrahimi, P., Nilsson, M., Morris, G. A., Jensen, H. M., & Engelsen, S. B. (2014). Cleaning up NMR spectra with reference deconvolution for improving multivariate analysis of complex mixture spectra: Improving multivariate analysis of NMR data by reference deconvolution. *Journal of Chemometrics*, *28*(8), 656–662. <https://doi.org/10.1002/cem.2607>
- Eckburg, P. B., Bik, E. M., Bernstein, C. N., Purdom, E., Dethlefsen, L., Sargent, M., Gill, S. R., Nelson, K. E., & Relman, D. A. (2005). Diversity of the human intestinal microbial flora. *Science (New York, N.Y.)*, *308*(5728), 1635–1638. <https://doi.org/10.1126/science.1110591>
- El Aidy, S., Hooiveld, G., Tremaroli, V., Bäckhed, F., & Kleerebezem, M. (2013). The gut microbiota and mucosal homeostasis: Colonized at birth or at adulthood, does it matter? *Gut Microbes*, *4*(2), 118–124. <https://doi.org/10.4161/gmic.23362>

- Elmi, F., Lee, H.-T., Huang, J.-Y., Hsieh, Y.-C., Wang, Y.-L., Chen, Y.-J., Shaw, S.-Y., & Chen, C.-J. (2005). Stereoselective Esterase from *Pseudomonas putida* IFO12996 Reveals  $\alpha/\beta$  Hydrolase Folds for D - $\beta$ -Acetylthioisobutyric Acid Synthesis. *Journal of Bacteriology*, *187*(24), 8470–8476. <https://doi.org/10.1128/JB.187.24.8470-8476.2005>
- Emwas, A.-H., Roy, R., McKay, R. T., Tenori, L., Saccenti, E., Gowda, G. A. N., Raftery, D., Alahmari, F., Jaremko, L., Jaremko, M., & Wishart, D. S. (2019). NMR Spectroscopy for Metabolomics Research. *Metabolites*, *9*(7), E123. <https://doi.org/10.3390/metabo9070123>
- Fan, T., Ren, R., Tang, S., Zhou, Y., Cai, M., Zhao, W., He, Y., & Xu, J. (2023). Transcriptomics combined with metabolomics unveiled the key genes and metabolites of mycelium growth in *Morchella importuna*. *Frontiers in Microbiology*, *14*, 1079353. <https://doi.org/10.3389/fmicb.2023.1079353>
- FAQ: E. Coli: Good, Bad, & Deadly: "What is true for E. coli is true for the elephant."* (2011). American Society for Microbiology. <http://www.ncbi.nlm.nih.gov/books/NBK562895/>
- Fiege, H., Voges, H.-W., Hamamoto, T., Umemura, S., Iwata, T., Miki, H., Fujita, Y., Buysch, H.-J., Garbe, D., & Paulus, W. (2000). Phenol Derivatives. In Wiley-VCH Verlag GmbH & Co. KGaA (Ed.), *Ullmann's Encyclopedia of Industrial Chemistry* (p. a19\_313). Wiley-VCH Verlag GmbH & Co. KGaA. [https://doi.org/10.1002/14356007.a19\\_313](https://doi.org/10.1002/14356007.a19_313)
- Fleschhut, J., Kratzer, F., Rechkemmer, G., & Kulling, S. E. (2006). Stability and biotransformation of various dietary anthocyanins in vitro. *European Journal of Nutrition*, *45*(1), 7–18. <https://doi.org/10.1007/s00394-005-0557-8>
- Flores, G., Ruiz del Castillo, M. L., Costabile, A., Klee, A., Bigetti Guergoletto, K., & Gibson, G. R. (2015). In vitro fermentation of anthocyanins encapsulated with cyclodextrins: Release, metabolism and influence on gut microbiota growth. *Journal of Functional Foods*, *16*, 50–57. <https://doi.org/10.1016/j.jff.2015.04.022>
- Forooshani, P. K., Meng, H., & Lee, B. P. (2017). Catechol Redox Reaction: Reactive Oxygen Species Generation, Regulation, and Biomedical Applications. In Y. Ito, X. Chen, & I.-K. Kang (Eds.), *ACS Symposium Series* (Vol. 1252, pp. 179–196). American Chemical Society. <https://doi.org/10.1021/bk-2017-1252.ch010>

- Fouillaud, M., & Dufossé, L. (2022). Microbial Secondary Metabolism and Biotechnology. *Microorganisms*, 10(1), 123. <https://doi.org/10.3390/microorganisms10010123>
- Fridovich, I. (1997). Superoxide Anion Radical (O<sup>•2</sup>), Superoxide Dismutases, and Related Matters. *Journal of Biological Chemistry*, 272(30), 18515–18517. <https://doi.org/10.1074/jbc.272.30.18515>
- Frisvad, J. C. (2012). Media and Growth Conditions for Induction of Secondary Metabolite Production. In N. P. Keller & G. Turner (Eds.), *Fungal Secondary Metabolism* (pp. 47–58). Humana Press. [https://doi.org/10.1007/978-1-62703-122-6\\_3](https://doi.org/10.1007/978-1-62703-122-6_3)
- Fritsche, W., & Hofrichter, M. (2005). Aerobic Degradation of Recalcitrant Organic Compounds by Microorganisms. In H.-J. Jördening & J. Winter (Eds.), *Environmental Biotechnology* (pp. 203–227). Wiley-VCH Verlag GmbH & Co. KGaA. <https://doi.org/10.1002/3527604286.ch7>
- Fulmer, G. R., Miller, A. J. M., Sherden, N. H., Gottlieb, H. E., Nudelman, A., Stoltz, B. M., Bercaw, J. E., & Goldberg, K. I. (2010). NMR Chemical Shifts of Trace Impurities: Common Laboratory Solvents, Organics, and Gases in Deuterated Solvents Relevant to the Organometallic Chemist. *Organometallics*, 29(9), 2176–2179. <https://doi.org/10.1021/om100106e>
- Garton, G. A., & Williams, R. T. (1948). Studies in detoxication. 17. The fate of catechol in the rabbit and the characterization of catechol monoglucuronide. *Biochemical Journal*, 43(2), 206–211. <https://doi.org/10.1042/bj0430206>
- Gleńsk, M., Hurst, W. J., Gliniski, V. B., Bednarski, M., & Gliński, J. A. (2019). Isolation of 1-(3',4'-Dihydroxyphenyl)-3-(2'',4'',6''-trihydroxyphenyl)-propan-2-ol from Grape Seed Extract and Evaluation of its Antioxidant and Antispasmodic Potential. *Molecules (Basel, Switzerland)*, 24(13), E2466. <https://doi.org/10.3390/molecules24132466>
- Goldansaz, S. A., Guo, A. C., Sajed, T., Steele, M. A., Plastow, G. S., & Wishart, D. S. (2017). Livestock metabolomics and the livestock metabolome: A systematic review. *PLOS ONE*, 12(5), e0177675. <https://doi.org/10.1371/journal.pone.0177675>

- Gorny, N., & Schink, B. (1994). Anaerobic degradation of catechol by *Desulfobacterium* sp. Strain Cat2 proceeds via carboxylation to protocatechuate. *Applied and Environmental Microbiology*, 60(9), 3396–3400. <https://doi.org/10.1128/aem.60.9.3396-3400.1994>
- Gottlieb, H. E., Kotlyar, V., & Nudelman, A. (1997). NMR Chemical Shifts of Common Laboratory Solvents as Trace Impurities. *The Journal of Organic Chemistry*, 62(21), 7512–7515. <https://doi.org/10.1021/jo9711176v>
- Guinane, C. M., & Cotter, P. D. (2013). Role of the gut microbiota in health and chronic gastrointestinal disease: Understanding a hidden metabolic organ. *Therapeutic Advances in Gastroenterology*, 6(4), 295–308. <https://doi.org/10.1177/1756283X13482996>
- Guo, J., & Huan, T. (2020). Comparison of Full-Scan, Data-Dependent, and Data-Independent Acquisition Modes in Liquid Chromatography-Mass Spectrometry Based Untargeted Metabolomics. *Analytical Chemistry*, 92(12), 8072–8080. <https://doi.org/10.1021/acs.analchem.9b05135>
- Gupta, S. K., & Shukla, P. (2016). Advanced technologies for improved expression of recombinant proteins in bacteria: Perspectives and applications. *Critical Reviews in Biotechnology*, 36(6), 1089–1098. <https://doi.org/10.3109/07388551.2015.1084264>
- Haiko, J., & Westerlund-Wikström, B. (2013). The role of the bacterial flagellum in adhesion and virulence. *Biology*, 2(4), 1242–1267. <https://doi.org/10.3390/biology2041242>
- Hamzah, R. Y., & Al-Baharna, B. S. (1994). Catechol ring-cleavage in *Pseudomonas cepacia*: The simultaneous induction of ortho and meta pathways. *Applied Microbiology and Biotechnology*, 41(2), 250–256. <https://doi.org/10.1007/BF00186968>
- Handbook of Pharmacogenomics and Stratified Medicine*. (2014). Elsevier. <https://doi.org/10.1016/C2010-0-67325-1>
- Hao, J., Liebeke, M., Astle, W., De Iorio, M., Bundy, J. G., & Ebbels, T. M. D. (2014). Bayesian deconvolution and quantification of metabolites in complex 1D NMR spectra using BATMAN. *Nature Protocols*, 9(6), 1416–1427. <https://doi.org/10.1038/nprot.2014.090>
- Harvey, A. (2008). Natural products in drug discovery. *Drug Discovery Today*, 13(19–20), 894–901. <https://doi.org/10.1016/j.drudis.2008.07.004>

- Harwood, C. S., Burchhardt, G., Herrmann, H., & Fuchs, G. (1998). Anaerobic metabolism of aromatic compounds via the benzoyl-CoA pathway. *FEMS Microbiology Reviews*, 22(5), 439–458. <https://doi.org/10.1111/j.1574-6976.1998.tb00380.x>
- Hassan, M. A., Al-Sakkaf, K., Shait Mohammed, M. R., Dallol, A., Al-Maghrabi, J., Aldahlawi, A., Ashoor, S., Maamra, M., Ragoussis, J., Wu, W., Khan, M. I., Al-Malki, A. L., & Choudhry, H. (2020). Integration of Transcriptome and Metabolome Provides Unique Insights to Pathways Associated With Obese Breast Cancer Patients. *Frontiers in Oncology*, 10, 804. <https://doi.org/10.3389/fonc.2020.00804>
- Hattula, M. L., Wasenius, V.-M., Reunanen, H., & Arstila, A. U. (1981). Acute toxicity of some chlorinated phenols, catechols and cresols to trout. *Bulletin of Environmental Contamination and Toxicology*, 26(1), 295–298. <https://doi.org/10.1007/BF01622093>
- He, Z., & Wiegel, J. (1996). Purification and characterization of an oxygen-sensitive, reversible 3,4-dihydroxybenzoate decarboxylase from *Clostridium hydroxybenzoicum*. *Journal of Bacteriology*, 178(12), 3539–3543. <https://doi.org/10.1128/jb.178.12.3539-3543.1996>
- Ho, C. S., Lam, C. W. K., Chan, M. H. M., Cheung, R. C. K., Law, L. K., Lit, L. C. W., Ng, K. F., Suen, M. W. M., & Tai, H. L. (2003). Electrospray ionisation mass spectrometry: Principles and clinical applications. *The Clinical Biochemist. Reviews*, 24(1), 3–12.
- Hoerr, V., Zbytnuik, L., Leger, C., Tam, P. P. C., Kubes, P., & Vogel, H. J. (2012). Gram-negative and Gram-Positive Bacterial Infections Give Rise to a Different Metabolic Response in a Mouse Model. *Journal of Proteome Research*, 11(6), 3231–3245. <https://doi.org/10.1021/pr201274r>
- Holmgren, A. (2001). Thioredoxin and Glutaredoxin: General Aspects and Involvement in Redox Regulation. In E.-M. Aro & B. Andersson (Eds.), *Regulation of Photosynthesis* (Vol. 11, pp. 321–330). Springer Netherlands. [https://doi.org/10.1007/0-306-48148-0\\_19](https://doi.org/10.1007/0-306-48148-0_19)
- Horak, I., Engelbrecht, G., van Rensburg, P. J. J., & Claassens, S. (2019). Microbial metabolomics: Essential definitions and the importance of cultivation conditions for utilizing *Bacillus* species as bionematicides. *Journal of Applied Microbiology*, 127(2), 326–343. <https://doi.org/10.1111/jam.14218>

- Hoult, D. I., Busby, S. J. W., Gadian, D. G., Radda, G. K., Richards, R. E., & Seeley, P. J. (1974). Observation of tissue metabolites using  $^{31}\text{P}$  nuclear magnetic resonance. *Nature*, *252*(5481), 285–287. <https://doi.org/10.1038/252285a0>
- Huang, C.-J., Lin, H., & Yang, X. (2012). Industrial production of recombinant therapeutics in *Escherichia coli* and its recent advancements. *Journal of Industrial Microbiology & Biotechnology*, *39*(3), 383–399. <https://doi.org/10.1007/s10295-011-1082-9>
- Huang, K., & Zhang, H. (2022). A comprehensive kinetic model for phenol oxidation in seven advanced oxidation processes and considering the effects of halides and carbonate. *Water Research X*, *14*, 100129. <https://doi.org/10.1016/j.wroa.2021.100129>
- Hudson, K. M., Shiver, E., Yu, J., Mehta, S., Jima, D. D., Kane, M. A., Patisaul, H. B., & Cowley, M. (2021). Transcriptomic, proteomic, and metabolomic analyses identify candidate pathways linking maternal cadmium exposure to altered neurodevelopment and behavior. *Scientific Reports*, *11*(1), 16302. <https://doi.org/10.1038/s41598-021-95630-2>
- Hughes, E. J., & Bayly, R. C. (1983). Control of catechol meta-cleavage pathway in *Alcaligenes eutrophus*. *Journal of Bacteriology*, *154*(3), 1363–1370. <https://doi.org/10.1128/jb.154.3.1363-1370.1983>
- Idle, J. R., & Gonzalez, F. J. (2007). Metabolomics. *Cell Metabolism*, *6*(5), 348–351. <https://doi.org/10.1016/j.cmet.2007.10.005>
- Imlay, J. A., & Linn, S. (1988). DNA Damage and Oxygen Radical Toxicity. *Science*, *240*(4857), 1302–1309. <https://doi.org/10.1126/science.3287616>
- Jarvis, D. A. (1982). Deconvolution of Absorption Spectra. *Physics Bulletin*, *33*(9), 334–334. <https://doi.org/10.1088/0031-9112/33/9/042>
- Jensen, E. D., Ambri, F., Bendtsen, M. B., Javanpour, A. A., Liu, C. C., Jensen, M. K., & Keasling, J. D. (2021). Integrating continuous hypermutation with high-throughput screening for optimization of cis,cis-muconic acid production in yeast. *Microbial Biotechnology*, *14*(6), 2617–2626. <https://doi.org/10.1111/1751-7915.13774>
- Jiménez, N., Curiel, J. A., Reverĳn, I., de las Rivas, B., & Muĳoz, R. (2013). Uncovering the *Lactobacillus plantarum* WCFS1 Gallate Decarboxylase Involved in Tannin Degradation.

*Applied and Environmental Microbiology*, 79(14), 4253–4263.

<https://doi.org/10.1128/AEM.00840-13>

- Jin, T.-S., Zhang, S.-L., Wang, X.-F., Guo, J.-J., & Li, T.-S. (2001). An Efficient and Convenient Method for Preparation of 2,2-Disubstituted and 2-Monosubstituted 1,3-Benzodioxoles from Ketones and Aldehydes with Catechol Catalysed by  $ZrO_2/SO_4^{2-}$ . *Journal of Chemical Research*, 2001(7), 289–291. <https://doi.org/10.3184/030823401103169801>
- Johnson, C. H., Ivanisevic, J., & Siuzdak, G. (2016). Metabolomics: Beyond biomarkers and towards mechanisms. *Nature Reviews Molecular Cell Biology*, 17(7), 451–459. <https://doi.org/10.1038/nrm.2016.25>
- Joly, N., Engl, C., Jovanovic, G., Huvet, M., Toni, T., Sheng, X., Stumpf, M. P. H., & Buck, M. (2010). Managing membrane stress: The phage shock protein (Psp) response, from molecular mechanisms to physiology. *FEMS Microbiology Reviews*, 34(5), 797–827. <https://doi.org/10.1111/j.1574-6976.2010.00240.x>
- Jothimani, P., Kalaichelvan, G., Bhaskaran, A., Selvaseelan, D. A., & Ramasamy, K. (2003). Anaerobic biodegradation of aromatic compounds. *Indian Journal of Experimental Biology*, 41(9), 1046–1067.
- Ju, R., Liu, X., Zheng, F., Zhao, X., Lu, X., Zeng, Z., Lin, X., & Xu, G. (2019). Removal of false positive features to generate authentic peak table for high-resolution mass spectrometry-based metabolomics study. *Analytica Chimica Acta*, 1067, 79–87. <https://doi.org/10.1016/j.aca.2019.04.011>
- Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A., Bridgland, A., Meyer, C., Kohl, S. A. A., Ballard, A. J., Cowie, A., Romera-Paredes, B., Nikolov, S., Jain, R., Adler, J., ... Hassabis, D. (2021). Highly accurate protein structure prediction with AlphaFold. *Nature*, 596(7873), 583–589. <https://doi.org/10.1038/s41586-021-03819-2>
- Kalyanaraman, B., Felix, C. C., & Sealy, R. C. (1985). Semiquinone anion radicals of catechol(amine)s, catechol estrogens, and their metal ion complexes. *Environmental Health Perspectives*, 64, 185–198. <https://doi.org/10.1289/ehp.8564185>

- Karlovsky, P. (Ed.). (2008). *Secondary metabolites in soil ecology*. Springer.
- Kaufmann, R. (1995). Matrix-assisted laser desorption ionization (MALDI) mass spectrometry: A novel analytical tool in molecular biology and biotechnology. *Journal of Biotechnology*, *41*(2–3), 155–175. [https://doi.org/10.1016/0168-1656\(95\)00009-f](https://doi.org/10.1016/0168-1656(95)00009-f)
- Kawabata, K., Sugiyama, Y., Sakano, T., & Ohigashi, H. (2013). Flavonols enhanced production of anti-inflammatory substance(s) by *Bifidobacterium adolescentis*: Prebiotic actions of galangin, quercetin, and fisetin. *BioFactors (Oxford, England)*, *39*(4), 422–429. <https://doi.org/10.1002/biof.1081>
- Kittana, H., Gomes-Neto, J. C., Heck, K., Juritsch, A. F., Sughroue, J., Xian, Y., Mantz, S., Segura Muñoz, R. R., Cody, L. A., Schmaltz, R. J., Anderson, C. L., Moxley, R. A., Hostetter, J. M., Fernando, S. C., Clarke, J., Kachman, S. D., Cressler, C. E., Benson, A. K., Walter, J., & Ramer-Tait, A. E. (2023). Evidence for a Causal Role for *Escherichia coli* Strains Identified as Adherent-Invasive (AIEC) in Intestinal Inflammation. *mSphere*, *8*(2), e0047822. <https://doi.org/10.1128/msphere.00478-22>
- Knezevic, S., Ghafoor, A., Mehri, S., Barazi, A., Dziura, M., Trant, J. F., & Dieni, C. A. (2021). Catechin and other catechol-containing secondary metabolites: Bacterial biotransformation and regulation of carbohydrate metabolism. *PharmaNutrition*, *17*, 100273. <https://doi.org/10.1016/j.phanu.2021.100273>
- Kocaçalışkan, I., Talan, I., & Terzi, I. (2006). Antimicrobial activity of catechol and pyrogallol as allelochemicals. *Zeitschrift Fur Naturforschung. C, Journal of Biosciences*, *61*(9–10), 639–642. <https://doi.org/10.1515/znc-2006-9-1004>
- Kontou, E. E., Walter, A., Alka, O., Pfeuffer, J., Sachsenberg, T., Mohite, O. S., Nuhamunada, M., Kohlbacher, O., & Weber, T. (2023). UmetaFlow: An untargeted metabolomics workflow for high-throughput data processing and analysis. *Journal of Cheminformatics*, *15*(1), 52. <https://doi.org/10.1186/s13321-023-00724-w>
- Koppel, N., Maini Rekdal, V., & Balskus, E. P. (2017). Chemical transformation of xenobiotics by the human gut microbiota. *Science*, *356*(6344), eaag2770. <https://doi.org/10.1126/science.aag2770>

- Krastanov, A. (2010). Metabolomics—The State of Art. *Biotechnology & Biotechnological Equipment*, 24(1), 1537–1543. <https://doi.org/10.2478/V10133-010-0001-Y>
- Krause, D. O., Little, A. C., Dowd, S. E., & Bernstein, C. N. (2011). Complete genome sequence of adherent invasive *Escherichia coli* UM146 isolated from Ileal Crohn's disease biopsy tissue. *Journal of Bacteriology*, 193(2), 583. <https://doi.org/10.1128/JB.01290-10>
- Krishnamurty, H. G., Cheng, K. J., Jones, G. A., Simpson, F. J., & Watkin, J. E. (1970). Identification of products produced by the anaerobic degradation of rutin and related flavonoids by *Butyrivibrio* sp. C3. *Canadian Journal of Microbiology*, 16(8), 759–767. <https://doi.org/10.1139/m70-129>
- Krumholz, L. R., & Bryant, M. P. (1986). *Eubacterium oxidoreducens* sp. Nov. Requiring H<sub>2</sub> or formate to degrade gallate, pyrogallol, phloroglucinol and quercetin. *Archives of Microbiology*, 144(1), 8–14. <https://doi.org/10.1007/BF00454948>
- Kukurba, K. R., & Montgomery, S. B. (2015). RNA Sequencing and Analysis. *Cold Spring Harbor Protocols*, 2015(11), pdb.top084970. <https://doi.org/10.1101/pdb.top084970>
- Kumar, A., Kumar, S., & Kumar, S. (2005). Biodegradation kinetics of phenol and catechol using *Pseudomonas putida* MTCC 1194. *Biochemical Engineering Journal*, 22(2), 151–159. <https://doi.org/10.1016/j.bej.2004.09.006>
- Kunzelmann, M., Winter, M., Åberg, M., Hellenäs, K.-E., & Rosén, J. (2018). Non-targeted analysis of unexpected food contaminants using LC-HRMS. *Analytical and Bioanalytical Chemistry*, 410(22), 5593–5602. <https://doi.org/10.1007/s00216-018-1028-4>
- Lee, H. C., Jenner, A. M., Low, C. S., & Lee, Y. K. (2006). Effect of tea phenolics and their aromatic fecal bacterial metabolites on intestinal microbiota. *Research in Microbiology*, 157(9), 876–884. <https://doi.org/10.1016/j.resmic.2006.07.004>
- Lelli, V., Belardo, A., & Maria Timperio, A. (2021). From Targeted Quantification to Untargeted Metabolomics. In X. Zhan (Ed.), *Metabolomics—Methodology and Applications in Medical Sciences and Life Sciences*. IntechOpen. <https://doi.org/10.5772/intechopen.96852>
- Leong, S. C., & Sirich, T. L. (2016). Indoxyl Sulfate-Review of Toxicity and Therapeutic Strategies. *Toxins*, 8(12), 358. <https://doi.org/10.3390/toxins8120358>

- Letertre, M. P. M., Giraudeau, P., & de Tullio, P. (2021). Nuclear Magnetic Resonance Spectroscopy in Clinical Metabolomics and Personalized Medicine: Current Challenges and Perspectives. *Frontiers in Molecular Biosciences*, 8, 698337. <https://doi.org/10.3389/fmolb.2021.698337>
- Li, H., Zhou, X., Huang, Y., Liao, B., Cheng, L., & Ren, B. (2021). Reactive Oxygen Species in Pathogen Clearance: The Killing Mechanisms, the Adaption Response, and the Side Effects. *Frontiers in Microbiology*, 11, 622534. <https://doi.org/10.3389/fmicb.2020.622534>
- Li, P., Su, R., Yin, R., Lai, D., Wang, M., Liu, Y., & Zhou, L. (2020). Detoxification of Mycotoxins through Biotransformation. *Toxins*, 12(2), 121. <https://doi.org/10.3390/toxins12020121>
- Li, S., Siddiqi, A., Thapa, M., Chi, Y., & Zheng, S. (2023). Trackable and scalable LC-MS metabolomics data processing using asari. *Nature Communications*, 14(1), 4113. <https://doi.org/10.1038/s41467-023-39889-1>
- Liang, D., Liu, Q., Zhou, K., Jia, W., Xie, G., & Chen, T. (2020). IP4M: An integrated platform for mass spectrometry-based metabolomics data mining. *BMC Bioinformatics*, 21(1), 444. <https://doi.org/10.1186/s12859-020-03786-x>
- Liu, L., Ma, X., Bilal, M., Wei, L., Tang, S., Luo, H., Zhao, Y., Wang, Z., & Duan, X. (2022). Toxicity and inhibition mechanism of gallic acid on physiology and fermentation performance of Escherichia coli. *Bioresources and Bioprocessing*, 9(1), 76. <https://doi.org/10.1186/s40643-022-00564-w>
- Lofrano, G., Rizzo, L., Grassi, M., & Belgiorno, V. (2009). Advanced oxidation of catechol: A comparison among photocatalysis, Fenton and photo-Fenton processes. *Desalination*, 249(2), 878–883. <https://doi.org/10.1016/j.desal.2009.02.068>
- M. Shuikan, A., N. Hozzein, W., M. Alzharani, M., N. Sandouka, M., A. Al Yousef, S., A. Alharbi, S., & Damra, E. (2021). Enhancement and Identification of Microbial Secondary Metabolites. In A. Najjari, A. Cherif, H. Sghaier, & H. Imene Ouzari (Eds.), *Extremophilic Microbes and Metabolites—Diversity, Bioprospecting and Biotechnological Applications*. IntechOpen. <https://doi.org/10.5772/intechopen.93489>
- Machas, M., Kurgan, G., Abed, O. A., Shapiro, A., Wang, X., & Nielsen, D. (2021). Characterizing Escherichia coli's transcriptional response to different styrene exposure modes reveals novel

- toxicity and tolerance insights. *Journal of Industrial Microbiology & Biotechnology*, 48(1–2), kuab019. <https://doi.org/10.1093/jimb/kuab019>
- Mamyurin, B. A. (2001). Time-of-flight mass spectrometry (concepts, achievements, and prospects). *International Journal of Mass Spectrometry*, 206(3), 251–266. [https://doi.org/10.1016/S1387-3806\(00\)00392-4](https://doi.org/10.1016/S1387-3806(00)00392-4)
- Manier, S. K., Keller, A., Schäper, J., & Meyer, M. R. (2019). Untargeted metabolomics by high resolution mass spectrometry coupled to normal and reversed phase liquid chromatography as a tool to study the in vitro biotransformation of new psychoactive substances. *Scientific Reports*, 9(1), 2741. <https://doi.org/10.1038/s41598-019-39235-w>
- March, R. E. (2009). Quadrupole ion traps. *Mass Spectrometry Reviews*, 28(6), 961–989. <https://doi.org/10.1002/mas.20250>
- Marshall, S. A., Fisher, K., Ní Cheallaigh, A., White, M. D., Payne, K. A. P., Parker, D. A., Rigby, S. E. J., & Leys, D. (2017). Oxidative Maturation and Structural Characterization of Prenylated FMN Binding by UbiD, a Decarboxylase Involved in Bacterial Ubiquinone Biosynthesis. *The Journal of Biological Chemistry*, 292(11), 4623–4637. <https://doi.org/10.1074/jbc.M116.762732>
- Marshall, S. A., Payne, K. A. P., Fisher, K., Titchiner, G. R., Levy, C., Hay, S., & Leys, D. (2021). UbiD domain dynamics underpins aromatic decarboxylation. *Nature Communications*, 12(1), 5065. <https://doi.org/10.1038/s41467-021-25278-z>
- Marshall, S. A., Payne, K. A. P., & Leys, D. (2017). The UbiX-UbiD system: The biosynthesis and use of prenylated flavin (prFMN). *Archives of Biochemistry and Biophysics*, 632, 209–221. <https://doi.org/10.1016/j.abb.2017.07.014>
- Mazhar, M., Zhu, Y., & Qin, L. (2023). The Interplay of Dietary Fibers and Intestinal Microbiota Affects Type 2 Diabetes by Generating Short-Chain Fatty Acids. *Foods*, 12(5), 1023. <https://doi.org/10.3390/foods12051023>
- McDonald, T., Holland, N., Skibola, C., Duramad, P., & Smith, M. (2001). Hypothesis: Phenol and hydroquinone derived mainly from diet and gastrointestinal flora activity are causal factors in leukemia. *Leukemia*, 15(1), 10–20. <https://doi.org/10.1038/sj.leu.2401981>

- Meng, H., Li, Y., Faust, M., Konst, S., & Lee, B. P. (2015). Hydrogen peroxide generation and biocompatibility of hydrogel-bound mussel adhesive moiety. *Acta Biomaterialia*, *17*, 160–169. <https://doi.org/10.1016/j.actbio.2015.02.002>
- Meng, J., Young, G., & Chen, J. (2021). The Rcs System in Enterobacteriaceae: Envelope Stress Responses and Virulence Regulation. *Frontiers in Microbiology*, *12*, 627104. <https://doi.org/10.3389/fmicb.2021.627104>
- Meselhy, M. R., Nakamura, N., & Hattori, M. (1997). Biotransformation of (-)-epicatechin 3-O-gallate by human intestinal bacteria. *Chemical & Pharmaceutical Bulletin*, *45*(5), 888–893. <https://doi.org/10.1248/cpb.45.888>
- Meyer, H.-P., & Schmidhalter, D. R. (2014). The History and Economic Relevance of Industrial Scale Suspension Culture of Living Cells. In H.-P. Meyer & D. R. Schmidhalter (Eds.), *Industrial Scale Suspension Culture of Living Cells* (pp. 1–38). Wiley-VCH Verlag GmbH & Co. KGaA. <https://doi.org/10.1002/9783527683321.ch00>
- Minatel, I. O., Borges, C. V., Ferreira, M. I., Gomez, H. A. G., Chen, C.-Y. O., & Lima, G. P. P. (2017). Phenolic Compounds: Functional Properties, Impact of Processing and Bioavailability. In M. Soto-Hernandez, M. Palma-Tenango, & M. del R. Garcia-Mateos (Eds.), *Phenolic Compounds—Biological Activity*. InTech. <https://doi.org/10.5772/66368>
- Misal, S. A., & Gawai, K. R. (2018). Azoreductase: A key player of xenobiotic metabolism. *Bioresources and Bioprocessing*, *5*(1), 17. <https://doi.org/10.1186/s40643-018-0206-8>
- Mittal, R. D. (2015). Tandem Mass Spectroscopy in Diagnosis and Clinical Research. *Indian Journal of Clinical Biochemistry*, *30*(2), 121–123. <https://doi.org/10.1007/s12291-015-0498-9>
- Mizuno, T., Kato, M., Jo, Y. L., & Mizushima, S. (1988). Interaction of OmpR, a positive regulator, with the osmoregulated ompC and ompF genes of Escherichia coli. Studies with wild-type and mutant OmpR proteins. *The Journal of Biological Chemistry*, *263*(2), 1008–1012.
- Mlynárik, V. (2017). Introduction to nuclear magnetic resonance. *Analytical Biochemistry*, *529*, 4–9. <https://doi.org/10.1016/j.ab.2016.05.006>

- Mosele, J., Macià, A., & Motilva, M.-J. (2015). Metabolic and Microbial Modulation of the Large Intestine Ecosystem by Non-Absorbed Diet Phenolic Compounds: A Review. *Molecules*, 20(9), 17429–17468. <https://doi.org/10.3390/molecules200917429>
- Moussavi, G., Barikbin, B., & Mahmoudi, M. (2010). The removal of high concentrations of phenol from saline wastewater using aerobic granular SBR. *Chemical Engineering Journal*, 158(3), 498–504. <https://doi.org/10.1016/j.cej.2010.01.038>
- Muthubharathi, B. C., Gowripriya, T., & Balamurugan, K. (2021). Metabolomics: Small molecules that matter more. *Molecular Omics*, 17(2), 210–229. <https://doi.org/10.1039/d0mo00176g>
- Myers, O. D., Sumner, S. J., Li, S., Barnes, S., & Du, X. (2017). Detailed Investigation and Comparison of the XCMS and MZmine 2 Chromatogram Construction and Chromatographic Peak Detection Methods for Preprocessing Mass Spectrometry Metabolomics Data. *Analytical Chemistry*, 89(17), 8689–8695. <https://doi.org/10.1021/acs.analchem.7b01069>
- Nair, P. M., & Vining, L. C. (1964). ENZYMIC OXIDATION OF CATECHOL TO DIPHENYLENEDIOXIDE-2,3-QUINONE. *Archives of Biochemistry and Biophysics*, 106, 422–427. [https://doi.org/10.1016/0003-9861\(64\)90210-3](https://doi.org/10.1016/0003-9861(64)90210-3)
- Nalbantoglu, S. (2019). Metabolomics: Basic Principles and Strategies. In S. Nalbantoglu & H. Amri (Eds.), *Molecular Medicine*. IntechOpen. <https://doi.org/10.5772/intechopen.88563>
- Neilson, A. H., Allard, A., Hynning, P., & Remberger, M. (1991). Distribution, fate and persistence of organochlorine compounds formed during production of bleached pulp. *Toxicological & Environmental Chemistry*, 30(1–2), 3–41. <https://doi.org/10.1080/02772249109357638>
- Nerli, S., McShan, A. C., & Sgourakis, N. G. (2018). Chemical shift-based methods in NMR structure determination. *Progress in Nuclear Magnetic Resonance Spectroscopy*, 106–107, 1–25. <https://doi.org/10.1016/j.pnmrs.2018.03.002>
- Netto, L. E. S., Chae, H. Z., Kang, S.-W., Rhee, S. G., & Stadtman, E. R. (1996). Removal of Hydrogen Peroxide by Thiol-specific Antioxidant Enzyme (TSA) Is Involved with Its Antioxidant Properties. *Journal of Biological Chemistry*, 271(26), 15315–15321. <https://doi.org/10.1074/jbc.271.26.15315>

- Nisar, A. (2022). Medicinal Plants and Phenolic Compounds. In F. A. Badria (Ed.), *Biochemistry* (Vol. 26). IntechOpen. <https://doi.org/10.5772/intechopen.99799>
- Njiru, C., Xue, W., De Rouck, S., Alba, J. M., Kant, M. R., Chruszcz, M., Vanholme, B., Dermauw, W., Wybouw, N., & Van Leeuwen, T. (2022). Intradiol ring cleavage dioxygenases from herbivorous spider mites as a new detoxification enzyme family in animals. *BMC Biology*, 20(1), 131. <https://doi.org/10.1186/s12915-022-01323-1>
- Ntie-Kang, F., Telukunta, K. K., Fobofou, S. A. T., Chukwudi Osamor, V., Egieyeh, S. A., Valli, M., Djoumbou-Feunang, Y., Sorokina, M., Stork, C., Mathai, N., Zierep, P., Chávez-Hernández, A. L., Duran-Frigola, M., Babiaka, S. B., Tematio Fouedjou, R., Eni, D. B., Akame, S., Arreyetta-Bawak, A. B., Ebob, O. T., ... Ludwig-Müller, J. (2021). Computational Applications in Secondary Metabolite Discovery (CAiSMD): An online workshop. *Journal of Cheminformatics*, 13(1), 64. <https://doi.org/10.1186/s13321-021-00546-8>
- Oldiges, M., Lütz, S., Pflug, S., Schroer, K., Stein, N., & Wiendahl, C. (2007). Metabolomics: Current state and evolving methodologies and tools. *Applied Microbiology and Biotechnology*, 76(3), 495–511. <https://doi.org/10.1007/s00253-007-1029-2>
- Oliphant, K., & Allen-Vercoe, E. (2019). Macronutrient metabolism by the human gut microbiome: Major fermentation by-products and their impact on host health. *Microbiome*, 7(1), 91. <https://doi.org/10.1186/s40168-019-0704-8>
- Öman, T., Tessem, M.-B., Bathen, T. F., Bertilsson, H., Angelsen, A., Hedenström, M., & Andreassen, T. (2014). Identification of metabolites from 2D 1H-13C HSQC NMR using peak correlation plots. *BMC Bioinformatics*, 15(1), 413. <https://doi.org/10.1186/s12859-014-0413-z>
- Omokhefe Bruce, S. (2022). Secondary Metabolites from Natural Products. In R. Vijayakumar & S. Selvapuram Sudalaimuthu Raja (Eds.), *Secondary Metabolites—Trends and Reviews*. IntechOpen. <https://doi.org/10.5772/intechopen.102222>
- On the Evaluation, I. W. G. (1999). *Other Data Relevant to an Evaluation of Carcinogenicity and its Mechanisms. In Re-evaluation of Some Organic Chemicals, Hydrazine and Hydrogen Peroxide. International Agency for Research on Cancer.*

- Ornston, L. N., & Stanier, R. Y. (1966). The conversion of catechol and protocatechuate to beta-ketoadipate by *Pseudomonas putida*. *The Journal of Biological Chemistry*, 241(16), 3776–3786.
- Pagès, J.-M., & Amaral, L. (2009). Mechanisms of drug efflux and strategies to combat them: Challenging the efflux pump of Gram-negative bacteria. *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics*, 1794(5), 826–833.  
<https://doi.org/10.1016/j.bbapap.2008.12.011>
- Pages, J.-M., Amaral, L., & Fanning, S. (2011). An Original Deal for New Molecule: Reversal of Efflux Pump Activity, A Rational Strategy to Combat Gram-Negative Resistant Bacteria. *Current Medicinal Chemistry*, 18(19), 2969–2980.  
<https://doi.org/10.2174/092986711796150469>
- Pando, J. M., Karlinsey, J. E., Lara, J. C., Libby, S. J., & Fang, F. C. (2017). The Rcs-Regulated Colanic Acid Capsule Maintains Membrane Potential in *Salmonella enterica* serovar Typhimurium. *mBio*, 8(3), e00808-17. <https://doi.org/10.1128/mBio.00808-17>
- Patridge, E. V., & Ferry, J. G. (2006). WrbA from *Escherichia coli* and *Archaeoglobus fulgidus* is an NAD(P)H:quinone oxidoreductase. *Journal of Bacteriology*, 188(10), 3498–3506.  
<https://doi.org/10.1128/JB.188.10.3498-3506.2006>
- Payer, S. E., Marshall, S. A., Bärland, N., Sheng, X., Reiter, T., Dordic, A., Steinkellner, G., Wuensch, C., Kaltwasser, S., Fisher, K., Rigby, S. E. J., Macheroux, P., Vonck, J., Gruber, K., Faber, K., Himo, F., Leys, D., Pavkov-Keller, T., & Glueck, S. M. (2017). Regioselective *para*-Carboxylation of Catechols with a Prenylated Flavin Dependent Decarboxylase. *Angewandte Chemie International Edition*, 56(44), 13893–13897.  
<https://doi.org/10.1002/anie.201708091>
- Pejčić, T., Zeković, M., Bumbaširević, U., Kalaba, M., Vovk, I., Bensa, M., Popović, L., & Tešić, Ž. (2023). The Role of Isoflavones in the Prevention of Breast Cancer and Prostate Cancer. *Antioxidants*, 12(2), 368. <https://doi.org/10.3390/antiox12020368>

- Pellack-Walker, P., Walker, J. K., Evans, H. H., & Blumer, J. L. (1985). Relationship between the oxidation potential of benzene metabolites and their inhibitory effect on DNA synthesis in L5178YS cells. *Molecular Pharmacology*, *28*(6), 560–566.
- Petran, A., Popa, A., Hädade, N. D., & Liebscher, J. (2020). New Insights into Catechol Oxidation – Application of Ammonium Peroxydisulfate in the Presence of Arylhydrazines. *ChemistrySelect*, *5*(30), 9523–9530. <https://doi.org/10.1002/slct.202002370>
- Philipp, B., Kemmler, D., Hellstern, J., Gorny, N., Caballero, A., & Schink, B. (2002). Anaerobic degradation of protocatechuate (3,4-dihydroxybenzoate) by *Thauera aromatica* strain AR-1. *FEMS Microbiology Letters*, *212*(1), 139–143. <https://doi.org/10.1111/j.1574-6968.2002.tb11257.x>
- Pirttilä, K., Balgoma, D., Rainer, J., Pettersson, C., Hedeland, M., & Brunius, C. (2022). Comprehensive Peak Characterization (CPC) in Untargeted LC–MS Analysis. *Metabolites*, *12*(2), 137. <https://doi.org/10.3390/metabo12020137>
- Plank, L. D., & Harvey, J. D. (1979). Generation Time Statistics of *Escherichia coli* B Measured by Synchronous Culture Techniques. *Journal of General Microbiology*, *115*(1), 69–77. <https://doi.org/10.1099/00221287-115-1-69>
- Proadhan, M. A. I., Shi, B., Song, M., He, L., Yuan, F., Yin, X., Bohman, P., McClain, C. J., & Zhang, X. (2019). Integrating comprehensive two-dimensional gas chromatography mass spectrometry and parallel two-dimensional liquid chromatography mass spectrometry for untargeted metabolomics. *The Analyst*, *144*(14), 4331–4341. <https://doi.org/10.1039/C9AN00560A>
- Puupponen-Pimiä, R., Nohynek, L., Hartmann-Schmidlin, S., Kähkönen, M., Heinonen, M., Määttä-Riihinen, K., & Oksman-Caldentey, K.-M. (2005). Berry phenolics selectively inhibit the growth of intestinal pathogens. *Journal of Applied Microbiology*, *98*(4), 991–1000. <https://doi.org/10.1111/j.1365-2672.2005.02547.x>
- Qiu, S., Cai, Y., Yao, H., Lin, C., Xie, Y., Tang, S., & Zhang, A. (2023). Small molecule metabolites: Discovery of biomarkers and therapeutic targets. *Signal Transduction and Targeted Therapy*, *8*(1), 132. <https://doi.org/10.1038/s41392-023-01399-3>

- Rafiei, A., & Sleno, L. (2015). Comparison of peak-picking workflows for untargeted liquid chromatography/high-resolution mass spectrometry metabolomics data analysis: Comparing peak picking of LC/HRMS data. *Rapid Communications in Mass Spectrometry*, 29(1), 119–127. <https://doi.org/10.1002/rcm.7094>
- Rahman, M. (2015). Antimicrobial Secondary Metabolites—Extraction, Isolation, Identification, and Bioassay. In *Evidence-Based Validation of Herbal Medicine* (pp. 495–513). Elsevier. <https://doi.org/10.1016/B978-0-12-800874-4.00023-4>
- Rahouti, M., Steiman, R., Seigle-Murandi, F., & Christov, L. P. (1999). Growth of 1044 strains and species of fungi on 7 phenolic lignin model compounds. *Chemosphere*, 38(11), 2549–2559. [https://doi.org/10.1016/S0045-6535\(98\)00462-7](https://doi.org/10.1016/S0045-6535(98)00462-7)
- Rana, M. S., Bradley, S. T., & Guzman, M. I. (2023). Conversion of Catechol to 4-Nitrocatechol in Aqueous Microdroplets Exposed to O<sub>3</sub> and NO<sub>2</sub>. *ACS ES&T Air*, acsestair.3c00001. <https://doi.org/10.1021/acsestair.3c00001>
- Ravanbakhsh, S., Liu, P., Bjordahl, T. C., Mandal, R., Grant, J. R., Wilson, M., Eisner, R., Sineelnikov, I., Hu, X., Luchinat, C., Greiner, R., & Wishart, D. S. (2015). Accurate, Fully-Automated NMR Spectral Profiling for Metabolomics. *PLOS ONE*, 10(5), e0124219. <https://doi.org/10.1371/journal.pone.0124219>
- Razaviamri, S., Wang, K., Liu, B., & Lee, B. P. (2021). Catechol-Based Antimicrobial Polymers. *Molecules*, 26(3), 559. <https://doi.org/10.3390/molecules26030559>
- Reddy, M. K., Gupta, S. K., Jacob, M. R., Khan, S. I., & Ferreira, D. (2007). Antioxidant, antimalarial and antimicrobial activities of tannin-rich fractions, ellagitannins and phenolic acids from *Punica granatum* L. *Planta Medica*, 73(5), 461–467. <https://doi.org/10.1055/s-2007-967167>
- Řezanka, T., Spížek, J., Přikrylová, V., & Dembitsky, V. M. (2004). Four New Derivatives of Trihomomonactinic Acids from *Streptomyces globisporus*. *European Journal of Organic Chemistry*, 2004(20), 4239–4244. <https://doi.org/10.1002/ejoc.200400276>
- Rigo, M., Alegre, R. M., Bezerra, J. R. M. V., Coelho, N., & Bastos, R. G. (2010). Catechol biodegradation kinetics using *Candida parapsilopsis*. *Brazilian Archives of Biology and Technology*, 53(2), 481–486. <https://doi.org/10.1590/S1516-89132010000200029>

- Roberts, L. D., Souza, A. L., Gerszten, R. E., & Clish, C. B. (2012). Targeted Metabolomics. *Current Protocols in Molecular Biology*, 98(1). <https://doi.org/10.1002/0471142727.mb3002s98>
- Rockwood, A. L., Kushnir, M. M., & Clarke, N. J. (2018). Mass Spectrometry. In *Principles and Applications of Clinical Mass Spectrometry* (pp. 33–65). Elsevier.  
<https://doi.org/10.1016/B978-0-12-816063-3.00002-5>
- Rodriguez, C. E., Fukuto, J. M., Taguchi, K., Froines, J., & Cho, A. K. (2005). The interactions of 9,10-phenanthrenequinone with glyceraldehyde-3-phosphate dehydrogenase (GAPDH), a potential site for toxic actions. *Chemico-Biological Interactions*, 155(1–2), 97–110.  
<https://doi.org/10.1016/j.cbi.2005.05.002>
- Romero, P. R., Kobayashi, N., Wedell, J. R., Baskaran, K., Iwata, T., Yokochi, M., Maziuk, D., Yao, H., Fujiwara, T., Kurusu, G., Ulrich, E. L., Hoch, J. C., & Markley, J. L. (2020). BioMagResBank (BMRB) as a Resource for Structural Biology. In Z. Gáspári (Ed.), *Structural Bioinformatics* (Vol. 2112, pp. 187–218). Springer US.  
[https://doi.org/10.1007/978-1-0716-0270-6\\_14](https://doi.org/10.1007/978-1-0716-0270-6_14)
- Rout, M., Lipfert, M., Lee, B. L., Berjanskii, M., Assempour, N., Fresno, R. V., Cayuela, A. S., Dong, Y., Johnson, M., Shahin, H., Gautam, V., Sajed, T., Oler, E., Peters, H., Mandal, R., & Wishart, D. S. (2023). MagMet: A fully automated web server for targeted nuclear magnetic resonance metabolomics of plasma and serum. *Magnetic Resonance in Chemistry: MRC*.  
<https://doi.org/10.1002/mrc.5371>
- Rowland, I., Gibson, G., Heinken, A., Scott, K., Swann, J., Thiele, I., & Tuohy, K. (2018). Gut microbiota functions: Metabolism of nutrients and other food components. *European Journal of Nutrition*, 57(1), 1–24. <https://doi.org/10.1007/s00394-017-1445-8>
- Roze, L. V., Chanda, A., Wee, J., Awad, D., & Linz, J. E. (2011). Stress-related transcription factor AtfB integrates secondary metabolism with oxidative stress response in aspergilli. *The Journal of Biological Chemistry*, 286(40), 35137–35148.  
<https://doi.org/10.1074/jbc.M111.253468>

- Ruan, X., Wang, Y., Zhou, L., Zheng, Q., Hao, H., & He, D. (2022). Evaluation of Untargeted Metabolomic Strategy for the Discovery of Biomarker of Breast Cancer. *Frontiers in Pharmacology*, *13*, 894099. <https://doi.org/10.3389/fphar.2022.894099>
- Ruiz, N., & Silhavy, T. J. (2022). How Escherichia coli Became the Flagship Bacterium of Molecular Biology. *Journal of Bacteriology*, *204*(9), e0023022. <https://doi.org/10.1128/jb.00230-22>
- Saghir, S. A., Ansari, R. A., & Munir, S. T. (2022). Fate of chemicals following exposure III: Metabolism (biotransformation). In *Reference Module in Biomedical Sciences* (p. B9780128243152000506). Elsevier. <https://doi.org/10.1016/B978-0-12-824315-2.00050-6>
- Saiz-Poseu, J., Mancebo-Aracil, J., Nador, F., Busqué, F., & Ruiz-Molina, D. (2019). The Chemistry behind Catechol-Based Adhesion. *Angewandte Chemie International Edition*, *58*(3), 696–714. <https://doi.org/10.1002/anie.201801063>
- Sajed, T., Marcu, A., Ramirez, M., Pon, A., Guo, A. C., Knox, C., Wilson, M., Grant, J. R., Djoumbou, Y., & Wishart, D. S. (2016). ECMDDB 2.0: A richer resource for understanding the biochemistry of *E. coli*. *Nucleic Acids Research*, *44*(D1), D495–D501. <https://doi.org/10.1093/nar/gkv1060>
- Sala-Trepat, J. M., & Evans, W. C. (1971). The meta cleavage of catechol by Azotobacter species. 4-Oxalocrotonate pathway. *European Journal of Biochemistry*, *20*(3), 400–413. <https://doi.org/10.1111/j.1432-1033.1971.tb01406.x>
- Schiffman, C., Petrick, L., Perttula, K., Yano, Y., Carlsson, H., Whitehead, T., Metayer, C., Hayes, J., Rappaport, S., & Dudoit, S. (2019). Filtering procedures for untargeted LC-MS metabolomics data. *BMC Bioinformatics*, *20*(1), 334. <https://doi.org/10.1186/s12859-019-2871-9>
- Schneider, H., & Blaut, M. (2000). Anaerobic degradation of flavonoids by Eubacterium ramulus. *Archives of Microbiology*, *173*(1), 71–75. <https://doi.org/10.1007/s002030050010>
- Schollée, J. E., Schymanski, E. L., Stravs, M. A., Gulde, R., Thomaidis, N. S., & Hollender, J. (2017). Similarity of High-Resolution Tandem Mass Spectrometry Spectra of Structurally Related Micropollutants and Transformation Products. *Journal of the American Society for Mass Spectrometry*, *28*(12), 2692–2704. <https://doi.org/10.1007/s13361-017-1797-6>

- Schrimpe-Rutledge, A. C., Codreanu, S. G., Sherrod, S. D., & McLean, J. A. (2016). Untargeted Metabolomics Strategies-Challenges and Emerging Directions. *Journal of the American Society for Mass Spectrometry*, 27(12), 1897–1905. <https://doi.org/10.1007/s13361-016-1469-y>
- Schweigert, N., Zehnder, A. J., & Eggen, R. I. (2001). Chemical properties of catechols and their molecular modes of toxic action in cells, from microorganisms to mammals. *Environmental Microbiology*, 3(2), 81–91. <https://doi.org/10.1046/j.1462-2920.2001.00176.x>
- Scigelova, M., & Makarov, A. (2009). Advances in bioanalytical LC–MS using the Orbitrap™ mass analyzer. *Bioanalysis*, 1(4), 741–754. <https://doi.org/10.4155/bio.09.65>
- Seo, J.-S., Keum, Y.-S., & Li, Q. X. (2009). Bacterial degradation of aromatic compounds. *International Journal of Environmental Research and Public Health*, 6(1), 278–309. <https://doi.org/10.3390/ijerph6010278>
- Seyedsayamdost, M. R. (2019). Toward a global picture of bacterial secondary metabolism. *Journal of Industrial Microbiology & Biotechnology*, 46(3–4), 301–311. <https://doi.org/10.1007/s10295-019-02136-y>
- Sharma, A., Shivaprasad, D. P., Chauhan, K., & Taneja, N. K. (2019). Control of E. coli growth and survival in Indian soft cheese (paneer) using multiple hurdles: Phytochemicals, temperature and vacuum. *LWT*, 114, 108350. <https://doi.org/10.1016/j.lwt.2019.108350>
- Sheng, Q., Vickers, K., Zhao, S., Wang, J., Samuels, D. C., Koues, O., Shyr, Y., & Guo, Y. (2017). Multi-perspective quality control of Illumina RNA sequencing data analysis. *Briefings in Functional Genomics*, 16(4), 194–204. <https://doi.org/10.1093/bfpg/elw035>
- Shi, S., Yang, L., Yang, C., Li, S., Zhao, H., Ren, L., Wang, X., Lu, F., Li, Y., & Zhao, H. (2021). Function and Molecular Ecology Significance of Two Catechol-Degrading Gene Clusters in *Pseudomonas putida* ND6. *Journal of Microbiology and Biotechnology*, 31(2), 259–271. <https://doi.org/10.4014/jmb.2009.09026>
- Shulman, R. G., Brown, T. R., Ugurbil, K., Ogawa, S., Cohen, S. M., & Den Hollander, J. A. (1979). Cellular Applications of <sup>31</sup>P and <sup>13</sup>C Nuclear Magnetic Resonance. *Science*, 205(4402), 160–166. <https://doi.org/10.1126/science.36664>

- Siegbahn, P. E. M. (2004). The catalytic cycle of catechol oxidase. *JBIC Journal of Biological Inorganic Chemistry*, 9(5), 577–590. <https://doi.org/10.1007/s00775-004-0551-2>
- Silverstein, R. M., Webster, F. X., Kiemle, D. J., & Bryce, D. L. (2015). *Spectrometric identification of organic compounds* (Eighth edition). Wiley.
- Simó, C., & García-Cañas, V. (2020). Dietary bioactive ingredients to modulate the gut microbiota-derived metabolite TMAO. New opportunities for functional food development. *Food & Function*, 11(8), 6745–6776. <https://doi.org/10.1039/D0FO01237H>
- Singh, D., Mishra, K., & Ramanathan, G. (2015). Bioremediation of Nitroaromatic Compounds. In M. Samer (Ed.), *Wastewater Treatment Engineering*. InTech. <https://doi.org/10.5772/61253>
- Skerker, J. M., Prasol, M. S., Perchuk, B. S., Biondi, E. G., & Laub, M. T. (2005). Two-component signal transduction pathways regulating growth and cell cycle progression in a bacterium: A system-level analysis. *PLoS Biology*, 3(10), e334. <https://doi.org/10.1371/journal.pbio.0030334>
- Smith, R. W. (2013). Mass Spectrometry. In *Encyclopedia of Forensic Sciences* (pp. 603–608). Elsevier. <https://doi.org/10.1016/B978-0-12-382165-2.00250-6>
- Srey, S., Jahid, I. K., & Ha, S.-D. (2013). Biofilm formation in food industries: A food safety concern. *Food Control*, 31(2), 572–585. <https://doi.org/10.1016/j.foodcont.2012.12.001>
- Steinman, H. M. (1988). Bacterial Superoxide Dismutases. In M. G. Simic, K. A. Taylor, J. F. Ward, & C. Von Sonntag (Eds.), *Oxygen Radicals in Biology and Medicine* (pp. 641–646). Springer US. [https://doi.org/10.1007/978-1-4684-5568-7\\_101](https://doi.org/10.1007/978-1-4684-5568-7_101)
- Stevens, J. F., & Maier, C. S. (2016). The Chemistry of Gut Microbial Metabolism of Polyphenols. *Phytochemistry Reviews: Proceedings of the Phytochemical Society of Europe*, 15(3), 425–444. <https://doi.org/10.1007/s11101-016-9459-z>
- Subramanyam, R., & Mishra, I. M. (2007). Biodegradation of catechol (2-hydroxy phenol) bearing wastewater in an UASB reactor. *Chemosphere*, 69(5), 816–824. <https://doi.org/10.1016/j.chemosphere.2007.04.064>

- Subramanyam, R., & Mishra, I. M. (2008a). Co-degradation of resorcinol and catechol in an UASB reactor. *Bioresource Technology*, *99*(10), 4147–4157.  
<https://doi.org/10.1016/j.biortech.2007.08.060>
- Subramanyam, R., & Mishra, I. M. (2008b). Treatment of catechol bearing wastewater in an upflow anaerobic sludge blanket (UASB) reactor: Sludge characteristics. *Bioresource Technology*, *99*(18), 8917–8925. <https://doi.org/10.1016/j.biortech.2008.04.067>
- Sun, Q., Xu, Q., Wang, M., Wang, Y., Zhang, D., & Lai, M. (2022). *openNAU: An open-source platform for normalizing, analyzing, and visualizing untargeted metabolomics data* [Preprint]. *Bioinformatics*. <https://doi.org/10.1101/2022.08.31.506116>
- Suvorova, I. A., & Gelfand, M. S. (2019). Comparative Genomic Analysis of the Regulation of Aromatic Metabolism in Betaproteobacteria. *Frontiers in Microbiology*, *10*, 642.  
<https://doi.org/10.3389/fmicb.2019.00642>
- Svenson, A., & Hynning, P.-Å. (1997). Increased aquatic toxicity following photolytic conversion of an organochlorine pollutant. *Chemosphere*, *34*(8), 1685–1692. [https://doi.org/10.1016/S0045-6535\(97\)00025-8](https://doi.org/10.1016/S0045-6535(97)00025-8)
- Tabanelli, T., Cailotto, S., Strachan, J., Masters, A. F., Maschmeyer, T., Perosa, A., & Cavani, F. (2018). Process systems for the carbonate interchange reactions of DMC and alcohols: Efficient synthesis of catechol carbonate. *Catalysis Science & Technology*, *8*(7), 1971–1980.  
<https://doi.org/10.1039/C8CY00119G>
- Tahmasebi, H., Dehbashi, S., Nasaj, M., & Arabestani, M. R. (2022). Molecular epidemiology and collaboration of siderophore-based iron acquisition with surface adhesion in hypervirulent *Pseudomonas aeruginosa* isolates from wound infections. *Scientific Reports*, *12*(1), 7791.  
<https://doi.org/10.1038/s41598-022-11984-1>
- Takagaki, A., & Nanjo, F. (2013). Catabolism of (+)-Catechin and (–)-Epicatechin by Rat Intestinal Microbiota. *Journal of Agricultural and Food Chemistry*, *61*(20), 4927–4935.  
<https://doi.org/10.1021/jf304431v>

- Takagaki, A., & Nanjo, F. (2015). Bioconversion of (–)-Epicatechin, (+)-Epicatechin, (–)-Catechin, and (+)-Catechin by (–)-Epigallocatechin-Metabolizing Bacteria. *Biological & Pharmaceutical Bulletin*, 38(5), 789–794. <https://doi.org/10.1248/bpb.b14-00813>
- Thakur, C. S., Brown, M. E., Sama, J. N., Jackson, M. E., & Dayie, T. K. (2010). Growth of wildtype and mutant E. coli strains in minimal media for optimal production of nucleic acids for preparing labeled nucleotides. *Applied Microbiology and Biotechnology*, 88(3), 771–779. <https://doi.org/10.1007/s00253-010-2813-y>
- Thirumurugan, D., Cholarajan, A., Raja, S. S. S., & Vijayakumar, R. (2018). An Introductory Chapter: Secondary Metabolites. In R. Vijayakumar & S. S. S. Raja (Eds.), *Secondary Metabolites—Sources and Applications*. InTech. <https://doi.org/10.5772/intechopen.79766>
- Thomas, S. N. (2019). Mass spectrometry. In *Contemporary Practice in Clinical Chemistry* (pp. 171–185). Elsevier. <https://doi.org/10.1016/B978-0-12-815499-1.00010-7>
- Tobacco smoking. (1986). *IARC Monographs on the Evaluation of the Carcinogenic Risk of Chemicals to Humans*, 38, 35–394.
- Tortorello, M. L. (1999). ESCHERICHIA COLI O157 | Escherichia Coli O157:H7. In *Encyclopedia of Food Microbiology* (pp. 646–652). Elsevier. <https://doi.org/10.1006/rwfm.1999.0540>
- Tsao, R. (2010). Chemistry and Biochemistry of Dietary Polyphenols. *Nutrients*, 2(12), 1231–1246. <https://doi.org/10.3390/nu2121231>
- Turi, K. N., Romick-Rosendale, L., Ryckman, K. K., & Hartert, T. V. (2018). A review of metabolomics approaches and their application in identifying causal pathways of childhood asthma. *Journal of Allergy and Clinical Immunology*, 141(4), 1191–1201. <https://doi.org/10.1016/j.jaci.2017.04.021>
- Tzounis, X., Vulevic, J., Kuhnle, G. G. C., George, T., Leonczak, J., Gibson, G. R., Kwik-Urbe, C., & Spencer, J. P. E. (2008). Flavanol monomer-induced changes to the human faecal microflora. *The British Journal of Nutrition*, 99(4), 782–792. <https://doi.org/10.1017/S0007114507853384>

- Uppal, K., Walker, D. I., Liu, K., Li, S., Go, Y.-M., & Jones, D. P. (2016). Computational Metabolomics: A Framework for the Million Metabolome. *Chemical Research in Toxicology*, 29(12), 1956–1975. <https://doi.org/10.1021/acs.chemrestox.6b00179>
- Ussher, J. R., Lopaschuk, G. D., & Arduini, A. (2013). Gut microbiota metabolism of l-carnitine and cardiovascular risk. *Atherosclerosis*, 231(2), 456–461. <https://doi.org/10.1016/j.atherosclerosis.2013.10.013>
- Vaidyanathan, S. (2005). Profiling microbial metabolomes: What do we stand to gain? *Metabolomics*, 1(1), 17–28. <https://doi.org/10.1007/s11306-005-1104-6>
- Valkova, N., Lépine, F., Labrie, L., Dupont, M., & Beaudet, R. (2003). Purification and characterization of PrbA, a new esterase from *Enterobacter cloacae* hydrolyzing the esters of 4-hydroxybenzoic acid (parabens). *The Journal of Biological Chemistry*, 278(15), 12779–12785. <https://doi.org/10.1074/jbc.M213281200>
- Van Den Heuvel, R. L., Leppens, H., & Schoeters, G. E. (1999). Lead and catechol hematotoxicity in vitro using human and murine hematopoietic progenitor cells. *Cell Biology and Toxicology*, 15(2), 101–110. <https://doi.org/10.1023/a:1007573414306>
- van der Kloet, F. M., Buurmans, J., Jonker, M. J., Smilde, A. K., & Westerhuis, J. A. (2020). Increased comparability between RNA-Seq and microarray data by utilization of gene sets. *PLOS Computational Biology*, 16(9), e1008295. <https://doi.org/10.1371/journal.pcbi.1008295>
- van Santen, J. A., Jacob, G., Singh, A. L., Aniebok, V., Balunas, M. J., Bunsko, D., Neto, F. C., Castaño-Espriu, L., Chang, C., Clark, T. N., Cleary Little, J. L., Delgadillo, D. A., Dorrestein, P. C., Duncan, K. R., Egan, J. M., Galey, M. M., Haeckl, F. P. J., Hua, A., Hughes, A. H., ... Linington, R. G. (2019). The Natural Products Atlas: An Open Access Knowledge Base for Microbial Natural Products Discovery. *ACS Central Science*, 5(11), 1824–1833. <https://doi.org/10.1021/acscentsci.9b00806>
- Vaquero, M. J. R., Alberto, M. R., & De Nadra, M. C. M. (2007). Antibacterial effect of phenolic compounds from different wines. *Food Control*, 18(2), 93–101. <https://doi.org/10.1016/j.foodcont.2005.08.010>

- Vijayaraj, M., & Gopinath, C. (2006). Selective production of methoxyphenols from dihydroxybenzenes on alkali metal ion-loaded MgO. *Journal of Catalysis*, 243(2), 376–388.  
<https://doi.org/10.1016/j.jcat.2006.08.009>
- Vinayavekhin, N., & Saghatelian, A. (2010). Untargeted Metabolomics. In F. M. Ausubel, R. Brent, R. E. Kingston, D. D. Moore, J. G. Seidman, J. A. Smith, & K. Struhl (Eds.), *Current Protocols in Molecular Biology* (p. mb3001s90). John Wiley & Sons, Inc.  
<https://doi.org/10.1002/0471142727.mb3001s90>
- Vining, L. C. (2007). Roles of Secondary Metabolites from Microbes. In D. J. Chadwick & J. Whelan (Eds.), *Novartis Foundation Symposia* (pp. 184–198). John Wiley & Sons, Ltd.  
<https://doi.org/10.1002/9780470514344.ch11>
- Wang, Z., Gerstein, M., & Snyder, M. (2009). RNA-Seq: A revolutionary tool for transcriptomics. *Nature Reviews Genetics*, 10(1), 57–63. <https://doi.org/10.1038/nrg2484>
- Weber, H. E., Gottardi, M., Brückner, C., Oreb, M., Boles, E., & Tripp, J. (2017). Requirement of a Functional Flavin Mononucleotide Prenyltransferase for the Activity of a Bacterial Decarboxylase in a Heterologous Muconic Acid Pathway in *Saccharomyces cerevisiae*. *Applied and Environmental Microbiology*, 83(10), e03472-16.  
<https://doi.org/10.1128/AEM.03472-16>
- Wen, B., Mei, Z., Zeng, C., & Liu, S. (2017). metaX: A flexible and comprehensive software for processing metabolomics data. *BMC Bioinformatics*, 18(1), 183.  
<https://doi.org/10.1186/s12859-017-1579-y>
- Weng, C., Peng, X., & Han, Y. (2021). Depolymerization and conversion of lignin to value-added bioproducts by microbial and enzymatic catalysis. *Biotechnology for Biofuels*, 14(1), 84.  
<https://doi.org/10.1186/s13068-021-01934-w>
- Wishart, D. S. (2016). Emerging applications of metabolomics in drug discovery and precision medicine. *Nature Reviews Drug Discovery*, 15(7), 473–484.  
<https://doi.org/10.1038/nrd.2016.32>

- Wishart, D. S. (2019). Metabolomics for Investigating Physiological and Pathophysiological Processes. *Physiological Reviews*, 99(4), 1819–1875.  
<https://doi.org/10.1152/physrev.00035.2018>
- Wishart, D. S., Guo, A., Oler, E., Wang, F., Anjum, A., Peters, H., Dizon, R., Sayeeda, Z., Tian, S., Lee, B. L., Berjanskii, M., Mah, R., Yamamoto, M., Jovel, J., Torres-Calzada, C., Hiebert-Giesbrecht, M., Lui, V. W., Varshavi, D., Varshavi, D., ... Gautam, V. (2022). HMDB 5.0: The Human Metabolome Database for 2022. *Nucleic Acids Research*, 50(D1), D622–D631.  
<https://doi.org/10.1093/nar/gkab1062>
- Wishart, D. S., Oler, E., Peters, H., Guo, A., Girod, S., Han, S., Saha, S., Lui, V. W., LeVatte, M., Gautam, V., Kaddurah-Daouk, R., & Karu, N. (2023). MiMeDB: The Human Microbial Metabolome Database. *Nucleic Acids Research*, 51(D1), D611–D620.  
<https://doi.org/10.1093/nar/gkac868>
- Wishart, D. S., Sayeeda, Z., Budinski, Z., Guo, A., Lee, B. L., Berjanskii, M., Rout, M., Peters, H., Dizon, R., Mah, R., Torres-Calzada, C., Hiebert-Giesbrecht, M., Varshavi, D., Varshavi, D., Oler, E., Allen, D., Cao, X., Gautam, V., Maras, A., ... Cort, J. R. (2022). NP-MRD: The Natural Products Magnetic Resonance Database. *Nucleic Acids Research*, 50(D1), D665–D677. <https://doi.org/10.1093/nar/gkab1052>
- Wu, Q., Yan, D., Chen, Y., Wang, T., Xiong, F., Wei, W., Lu, Y., Sun, W.-Y., Li, J. J., & Zhao, J. (2017). A redox-neutral catechol synthesis. *Nature Communications*, 8, 14227.  
<https://doi.org/10.1038/ncomms14227>
- Wu, X., Xu, J., Yang, X., Wang, D., & Xu, X. (2023). Integrating Transcriptomics and Metabolomics to Explore the Novel Pathway of *Fusobacterium nucleatum* Invading Colon Cancer Cells. *Pathogens*, 12(2), 201. <https://doi.org/10.3390/pathogens12020201>
- Yao, J., Dyson, H. J., & Wright, P. E. (1997). Chemical shift dispersion and secondary structure prediction in unfolded and partly folded proteins. *FEBS Letters*, 419(2–3), 285–289.  
[https://doi.org/10.1016/S0014-5793\(97\)01474-9](https://doi.org/10.1016/S0014-5793(97)01474-9)
- Yesiltepe, Y., Nuñez, J. R., Colby, S. M., Thomas, D. G., Borkum, M. I., Reardon, P. N., Washton, N. M., Metz, T. O., Teegarden, J. G., Govind, N., & Renslow, R. S. (2018). An automated

- framework for NMR chemical shift calculations of small organic molecules. *Journal of Cheminformatics*, 10(1), 52. <https://doi.org/10.1186/s13321-018-0305-8>
- Yin, W.-B., Reinke, A. W., Szilágyi, M., Emri, T., Chiang, Y.-M., Keating, A. E., Pócsi, I., Wang, C. C. C., & Keller, N. P. (2013). bZIP transcription factors affecting secondary metabolism, sexual development and stress responses in *Aspergillus nidulans*. *Microbiology (Reading, England)*, 159(Pt 1), 77–88. <https://doi.org/10.1099/mic.0.063370-0>
- Yin, Y., Wang, R., Cai, Y., Wang, Z., & Zhu, Z.-J. (2019). DecoMetDIA: Deconvolution of Multiplexed MS/MS Spectra for Metabolite Identification in SWATH-MS-Based Untargeted Metabolomics. *Analytical Chemistry*, 91(18), 11897–11904. <https://doi.org/10.1021/acs.analchem.9b02655>
- Zhang, A., Sun, H., Wang, P., Han, Y., & Wang, X. (2012). Modern analytical techniques in metabolomics analysis. *The Analyst*, 137(2), 293–300. <https://doi.org/10.1039/c1an15605e>
- Zhang, F., Robinette, S. L., Bruschiweiler-Li, L., & Brüschiweiler, R. (2009). Web server suite for complex mixture analysis by covariance NMR: Web server suite for complex mixture analysis by covariance NMR. *Magnetic Resonance in Chemistry*, 47(S1), S118–S122. <https://doi.org/10.1002/mrc.2486>
- Zhang, Z., Zhang, F., Deng, Y., Sun, L., Mao, M., Chen, R., Qiang, Q., Zhou, J., Long, T., Zhao, X., Liu, X., Wang, S., Yang, J., & Luo, J. (2022). Integrated Metabolomics and Transcriptomics Analyses Reveal the Metabolic Differences and Molecular Basis of Nutritional Quality in Landraces and Cultivated Rice. *Metabolites*, 12(5), 384. <https://doi.org/10.3390/metabo12050384>
- Zhao, H., Zhang, R., Yan, X., & Fan, K. (2021). Superoxide dismutase nanozymes: An emerging star for anti-oxidation. *Journal of Materials Chemistry B*, 9(35), 6939–6957. <https://doi.org/10.1039/D1TB00720C>
- Zhao, Q., Shen, H., Liu, J., Chiu, C.-Y., Su, K.-J., Tian, Q., Kakhniashvili, D., Qiu, C., Zhao, L.-J., Luo, Z., & Deng, H.-W. (2022). Pathway-based metabolomics study of sarcopenia-related traits in two US cohorts. *Aging*, 14(5), 2101–2112. <https://doi.org/10.18632/aging.203926>

Zheng, X., Zhou, C., Liu, Z., Long, M., Luo, Y.-H., Chen, T., Ontiveros-Valencia, A., & Rittmann, B.

E. (2019). Anaerobic biodegradation of catechol by sediment microorganisms: Interactive roles of N reduction and S cycling. *Journal of Cleaner Production*, 230, 80–89.

<https://doi.org/10.1016/j.jclepro.2019.05.058>

Zhong, M., & Zhou, Q. (2002). [Molecular-ecological technology of microorganisms and its application to research on environmental pollution]. *Ying Yong Sheng Tai Xue Bao = The Journal of Applied Ecology*, 13(2), 247–251.

## APPENDIX: Supplementary tables and figures

### Appendix A: Chapter 1

*S Table 1. List of studies that reported microbial bioactive compounds synthesis in response to secondary metabolite containing media.*

Substrate	Product	condition	Bacteria	Profiling technology	References
(-)-catechin (C)	1-(3,4-dihydroxyphenyl)-3-(2,4,6-trihydroxyphenyl)propan-2-ol	a	Adlercreutzia equolifaciens MT4s-5	1, 2, 3	Takagaki et al. 2015
(-)-catechin (C)	4-hydroxy-5-(3,4-dihydroxyphenyl) valeric acid (3R) and (4R)-5-(3,4-dihydroxyphenyl)- $\gamma$ -valerolactone	a	Flavonifractor plautii MT42	1, 2, 3	Takagaki et al. 2015
(-)-epicatechin (EC)	4-hydroxy-5-(3-hydroxyphenyl)valeric acid; 4-oxo-5-(3,4-dihydroxyphenyl)valeric acid; 4-oxo-5-(3-hydroxyphenyl)valeric acid; 1-(4-hydroxyphenyl)-3-(2,4,6-trihydroxyphenyl)propan-2-ol	a	Microbial Consortium	4, 5, 6, 7	Takagaki et al. 2013
(-)-epicatechin (EC)	(2S)-1-(3,4-dihydroxyphenyl)-3-(2,4,6-trihydroxyphenyl)propan-2-ol (1S)	a	Adlercreutzia equolifaciens MT4s-5	1, 2, 3	Takagaki et al. 2015
(-)-epicatechin (EC)	(2S)-1-(3-hydroxyphenyl)-3-(2,4,6-trihydroxyphenyl)propan-2-ol (2S)	a	Eggerthella lenta JCM 9979	1, 2, 3	Takagaki et al. 2015
(-)-epicatechin (EC)	(4R)-4-hydroxy-5-(3,4-dihydroxyphenyl) valeric acid (3R) and (4R)-5-(3,4-dihydroxyphenyl)- $\gamma$ -valerolactone (4R)	a	Flavonifractor plautii MT42	1, 2, 3	Takagaki et al. 2015
(-)-epicatechin gallate	pyrogallol; 5-(3,4-dihydroxyphenyl)- $\gamma$ -valerolactone; 4-hydroxy-5-(3,4-dihydroxyphenyl)valeric acid; 3-(3-hydroxyphenyl)propionic acid; and m-coumaric acid	a		2, 4, 5, 6, 7	Kohri et al. 2003
(+)-catechin (+C)	4-hydroxy-5-(3-hydroxyphenyl)valeric acid	a	Microbial Consortium	4, 5, 6, 7	Takagaki et al. 2013
(+)-catechin (C)	(2R)-1-(3,4-dihydroxyphenyl)-3-(2,4,6-trihydroxyphenyl)propan-2-ol (1R)	a	Adlercreutzia equolifaciens MT4s-5	1, 2, 3	Takagaki et al. 2015
(-)-epicatechin	3-hydroxyphenylpropionic acid (3-OH-PPr) and 3-phenylpropionic acid (3-PPr); 3,4-dihydroxyphenylvaleric	a	Microbial Consortium	8, 9	Aura et al. 2008

	acid; 3-hydroxyphenylvaleric acid					
(+)-catechin (C)	3-hydroxyphenylpropionic acid (3-OH-PPr); 3-hydroxyhippuric acid	a	human faecal microbiota	10, 18, 4	Griffiths et al. 1964, Das Y et al. 1971, Das P et al. 1971, Meselhy et al. 1997, Meng et al. 2002, Mulder et al. 2005	
(+)-catechin (C)	3-hydroxyphenylpropionic acid (3-OH-PPr) ; 3-phenylpropionic acid (3-PPr); 3,4-dihydroxyphenylvaleric acid	a	Microbial Consortium	8, 9	Aura et al. 2008	
(+)-catechin (C)	(2R)-1-(3-hydroxyphenyl)-3-(2,4,6-trihydroxyphenyl)propan-2-ol (2R)	a	Eggerthella lenta JCM 9979	1, 2, 3	Takagaki et al. 2015	
(+)-catechin (C)	1-(3-hydroxyphenyl)-3-(2,4,6-trihydroxyphenyl)propan-2-ol; 1-(3,4-dihydroxyphenyl)-3-(2,4,6-trihydroxyphenyl)propan-2	a	Microbial Consortium	11, 12, 4	Groenewoud et al. 1984	
(+)-catechin (C)	(4S)-4-hydroxy-5-(3,4-dihydroxyphenyl) valeric acid; (3S) and (4S)-5-(3,4-dihydroxyphenyl)- $\gamma$ -valerolactone (4S)	a	Flavonifractor plautii MT42	1, 2, 3	Takagaki et al. 2015	
(+)-epicatechin (EC)	(4S)-4-hydroxy-5-(3,4-dihydroxyphenyl) valeric acid; (3S) and (4S)-5-(3,4-dihydroxyphenyl)- $\gamma$ -valerolactone (4S)	a	Flavonifractor plautii MT42	1, 2, 3	Takagaki et al. 2015	
anthocyanins		a	Microbial Consortium	2, 4	Flores et al. 2015	
Crotonate	Acetate and butyrate	a	Eubacterium oxidoreducens	13, 14, 1	Krumholz et al. 1986	
Eriodictyol	3-(3,4-dihydroxyphenyl)propionic acid	a	Eubacterium ramulus	2, 3	Schneider et al. 2000	
kaempferol	4-hydroxyphenylacetic acid	a	Eubacterium ramulus	2, 3	Schneider et al. 2000	
Luteolin	3-(3,4-dihydroxyphenyl)propionic acid	a	Eubacterium ramulus	2, 3	Schneider et al. 2000	
luteolin-7-glucoside	3-(3,4-dihydroxyphenyl)propionic acid	a	Eubacterium ramulus	2, 3	Schneider et al. 2000	
Naringenin	3-(4-hydroxyphenyl)propionic acid	a	Eubacterium ramulus	2, 3	Schneider et al. 2000	
Phloretin	3-(4-hydroxyphenyl)propionic acid	a	Eubacterium ramulus	2, 3	Schneider et al. 2000	
Quercetin	3,4-dihydroxyphenylacetic acid, butyrate and acetate	a	Eubacterium oxidoreducens	13, 14, 1	Krumholz et al. 1986	
Quercetin derivatives	3-hydroxyphenylacetic acid	a	Microbial Consortium	4	Aura et al. 2002	
Quercetin	3,4-Dihydroxyphenylacetic acid	a	Eubacterium ramulus	2, 3	Schneider et al. 2000, Schneider et al. 1999,	

						Aura et al. 2002
Quercetin	Alphitonin	a	Eubacterium ramulus	4, 5, 6		Blaut et al. 2003
Quercetin-3-glucoside	3,4-Dihydroxyphenylacetic acid	a	Eubacterium ramulus	2, 3		Schneider et al. 2000
Rutin	3,4-dihydroxyphenylacetic acid, 3-hydroxyphenylacetic acid	a	Microbial Consortium	4		Aura et al. 2002
Taxifolin (Dihydroquercetin)	3,4-dihydroxyphenylacetic acid	a	Eubacterium ramulus	2, 3		Schneider et al. 2000
3,4,5-Trihydroxyphenylacetic acid	3,5-dihydroxyphenylacetic acid, 3-hydroxyphenylacetic acid	a		10, 13		Griffiths et al. 1972
3,5-Dihydroxyphenylpropionic acid	3-hydroxyphenylpropionic acid	a	Microbial Consortium	10, 13		Griffiths et al. 1972
5,7 - Dihydroxy - 3',4',5'-trimethoxyflavone	3,5-dihydroxyphenylpropionic acid	a		10, 13		Griffiths et al. 1972
Tricin (5,7,4'-trihydroxy-3',5'-dimethoxyflavone)	3,5-dihydroxyphenylpropionic acid	a	Microbial Consortium	10, 13		Griffiths et al. 1972
Tricetin (5,7,3',4',5'-pentahydroxyflavone).	3,5-dihydroxyphenylpropionic acid and 3-hydroxyphenylpropionic acid	a	Microbial Consortium	10, 13		Griffiths et al. 1972
Delphinidin (3,5,7,3',4',5'-hexahydroxyflavylium chloride)	Unknown metabolite Da and Db	a		10, 13		Griffiths et al. 1972
Glycoside myricitrin	3,5-dihydroxyphenylacetic acid (1.8mg), 3-hydroxyphenylacetic acid (trace), the aglycone myricetin (0.66mg) and 3,4,5-trihydroxyphenylacetic acid (0.56mg)	a	Microbial Consortium	10, 13		Griffiths et al. 1972
Myricitrin (myricetin 3-rhamnoside).	3,5-dihydroxyphenylacetic acid (1.8mg), 3-hydroxyphenylacetic acid (trace), the aglycone myricetin (0.66mg) and 3,4,5-trihydroxyphenylacetic acid (0.56mg)	a	Microbial Consortium	10, 13		Griffiths et al. 1972
Myricetin	3,5-dihydroxyphenylacetic acid (1.58mg), 3-hydroxyphenylacetic acid (trace), 3,4,5-trihydroxyphenylacetic acid (0.62mg)	a	Microbial Consortium	10, 13		Griffiths et al. 1972
(-)-epicatechin 3-o-gallate	Gallic acid; pyrogallol; 1-(3',4'-dihydroxy-phenyl)-3-(2'',4'',6''-trihydroxyphenyl)propan-2-ol; 1-(3'-hydroxyphenyl)-3-(2'',4'',6''-trihydroxyphenyl)propn-2-ol; 5-(3',4'-dihydroxyphenyl)γ-	a	Microbial Consortium	5, 6		Meselhy et al. 1997

	valerolactone; 5-(3'-hydroxyphenyl)- valerolactone; 5-(3',4'- dihydroxyphenyl)valeric acid; 5-(3'- dihydroxyphenyl)valeric acid; 3',4'- dihydroxyphenyl)propionic acid; 3'-hydroxyphenyl propionic acid; 5-(3'- methoxyphenyl)valeric acid; 2'',3''-dihydroxyphenoxy (3',4'- dihydroxyphenyl)propionat e					
chlorogenic acid	3-(3-hydroxyphenyl)- propionic acid	a	Microbial Consortium	2, 8, 3	Rechner et al. 2004	
Naringin	3-(4-hydroxyphenyl)- propionic acid; 3- phenylpropionic acid	a	Microbial Consortium	2, 8, 3	Rechner et al. 2004	
Rutin	3- hydroxyphenylacetic acid and 3-(3-hydroxyphenyl)- propionic acid	a	Microbial Consortium	2, 8, 3	Rechner et al. 2004	
cyanidin-3-glucoside	Cy-1, cyanidin aglycone; Cy-2 protocatechuic acid; Cy-3, a minor unidentified metabolite; Cy-4, the other unidentified conjugate	a	Microbial Consortium	15, 4	Aura et al. 2005	
malvidin-3-glucoside	syringic and vanillic acid	a	Microbial Consortium	16, 17, 8	Fleschhut et al. 2006	
peonidin-3-glucoside	syringic and vanillic acid	a	Microbial Consortium	16, 17, 8	Fleschhut et al. 2006	
cyanidin-3-glucoside	protocatechuic acid	a	Microbial Consortium	16, 17, 8	Fleschhut et al. 2006	
malvin	malvidin-3-glucoside; syringic acid	a	Microbial Consortium	16, 17, 8	Fleschhut et al. 2006	
pelargonidin- 3-sophorosid-5-glucoside (Pg-glu/soph)	4-hydroxybenzoic acid	a	Microbial Consortium	16, 17, 8	Fleschhut et al. 2006	
Chlorogenic acid	Ferulic acid; m-Coumaric acid; 3- Hydroxyphenylpropionic acid	a	N/A	3	Gonthier et al. 2003	
Caffeic acid	Ferulic acid; Isoferulic acid; 3- Hydroxyphenylpropionic acid	a	N/A	3	Gonthier et al. 2003	

Abbreviations: 1, UV; 2, HPLC; 3, LC-MS/MS; 4, LC-MS; 5, <sup>1</sup>H NMR; 6, <sup>13</sup>C NMR; 7, Optical Rotation; 8, GC-MS; 9, GC×GC-TOF; 10, Paper Chromatography; 11, Column Chromatography; 12, Preparative TLC; 13, TLC; 14, Gas Chromatography; 15, HPLC-DAS; 16, HPLC-DAD; 17, HPLC/MS; 18, NMR; a, anaerobic.

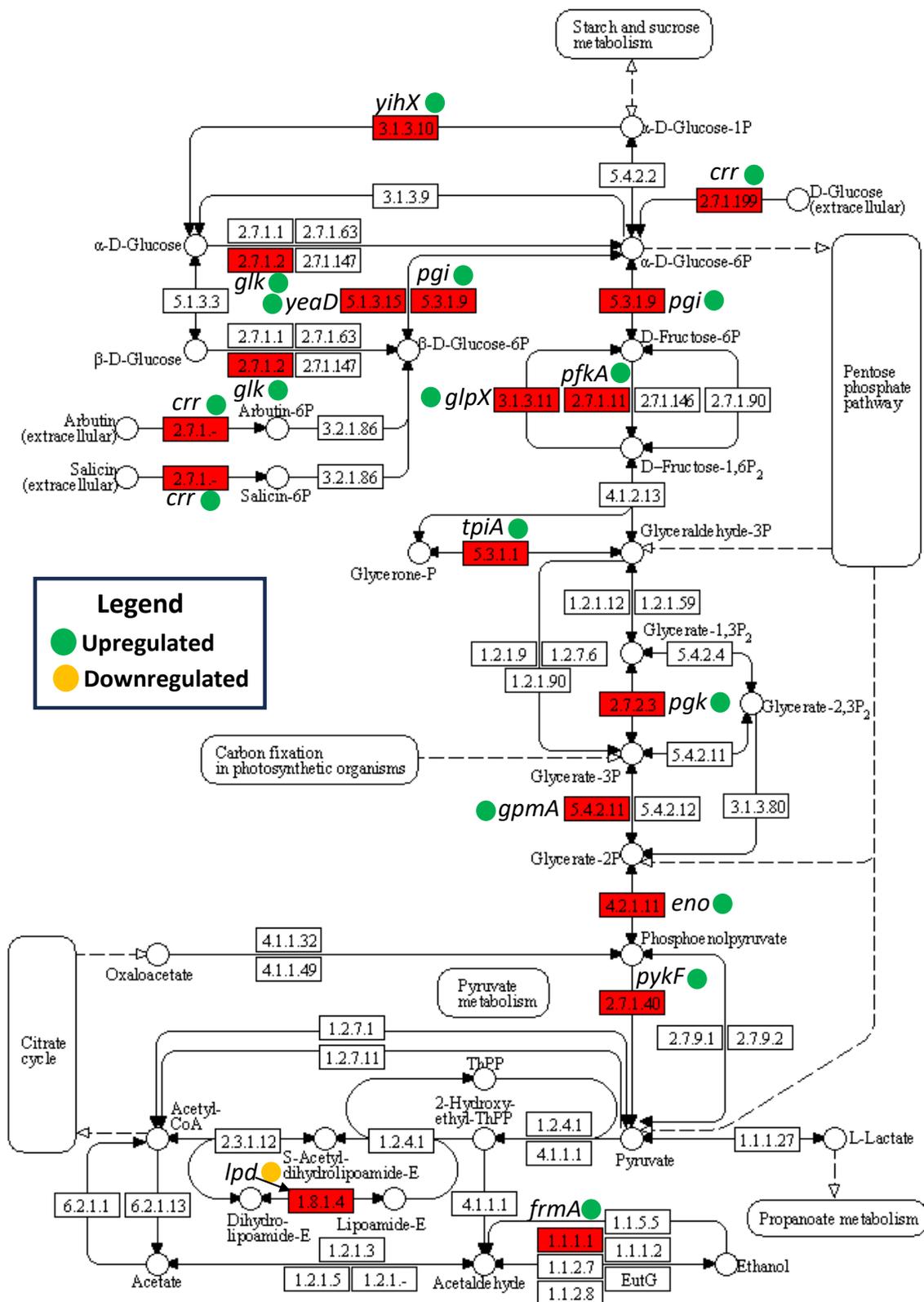
## Appendix B: Chapter 2

*S Table 2.1. List of yield and quality of each RNA samples.*

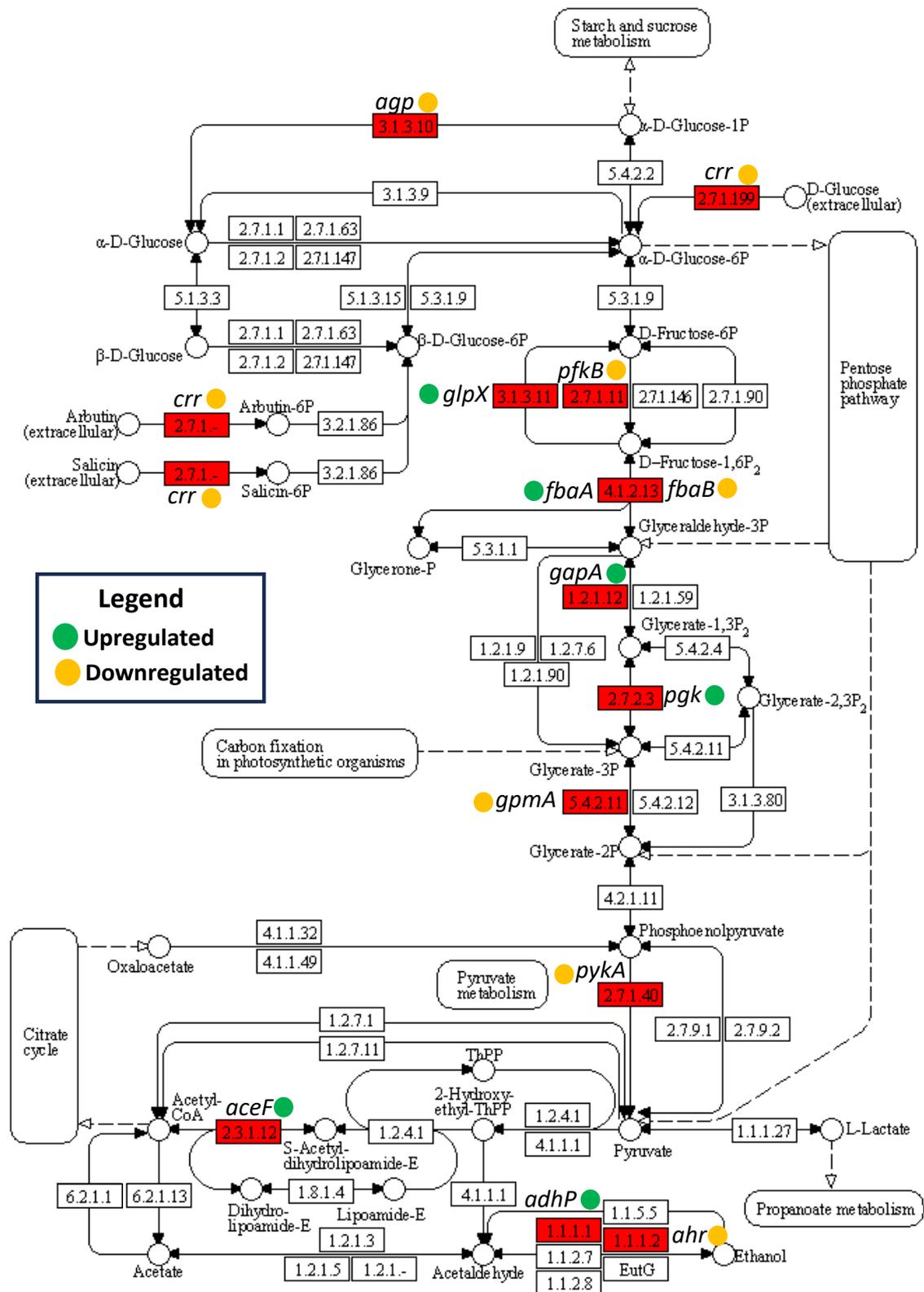
Sample type	Condition	Replicate Number	Concentration	Volume (uL)	260/280 ratio	Total RNA (ng)	RIN
Control	Anaerobic	1	80 ng/ul	31	1.97	2480	9.5
Control	Anaerobic	2	57 ng/ul	31	1.94	1767	9.7
Control	Anaerobic	3	72 ng/ul	31	1.99	2232	9.7
Treated	Anaerobic	1	32 ng/ul	31	1.98	992	9.5
Treated	Anaerobic	2	29 ng/ul	31	2	899	9.5
Treated	Anaerobic	3	52 ng/ul	31	1.99	1612	9.7
Control	Aerobic	1	180 ng/ul	31	1.98	5580	8.3
Control	Aerobic	2	159 ng/ul	31	1.99	4929	8.7
Control	Aerobic	3	255 ng/ul	31	1.98	7905	9.2
Treated	Aerobic	1	382 ng/ul	31	2.01	11842	9.1
Treated	Aerobic	2	307 ng/ul	31	2	9517	9.2
Treated	Aerobic	3	264 ng/ul	31	2	8184	9.1

*S Table 2.2. List of number of reads, quality and overall mapping alignment rate.*

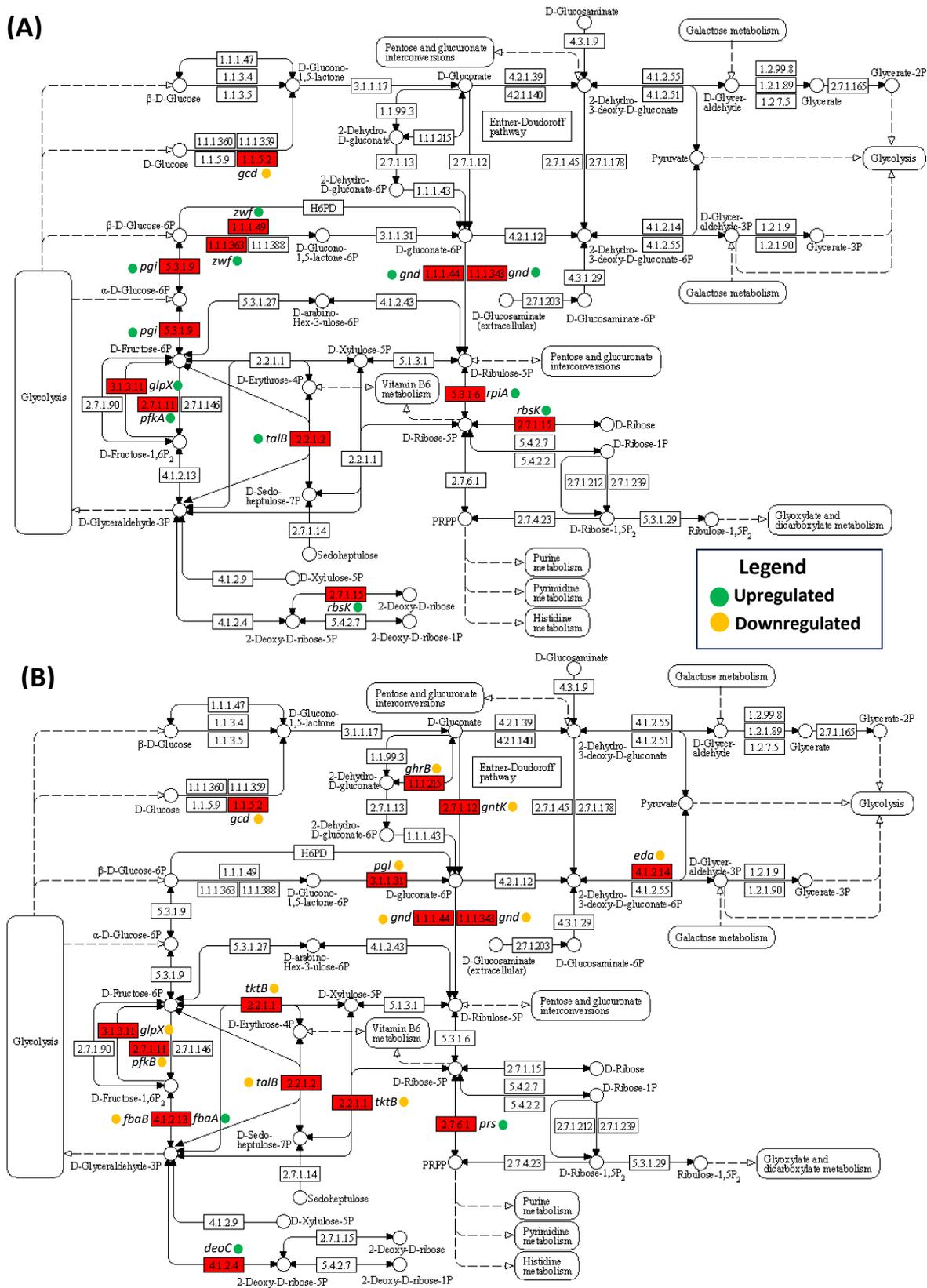
Sample type	Condition	Replicate number	Quality score, Q	Number of reads	Overall alignment rate, %
Control	Anaerobic	1	36	37403496	98
Control	Anaerobic	2	36	48287642	97.78
Control	Anaerobic	3	36	36623237	97.83
Treated	Anaerobic	1	36	35771972	98.83
Treated	Anaerobic	2	36	38005872	98.11
Treated	Anaerobic	3	36	46373360	98.1
Control	Aerobic	1	36	37283463	97.9
Control	Aerobic	2	36	36548187	98
Control	Aerobic	3	36	37472353	97.14
Treated	Aerobic	1	36	37512049	98.55
Treated	Aerobic	2	36	42123446	98.09
Treated	Aerobic	3	36	46686029	98.08



**S Figure 2.1.** Effect of catechol exposure on glycolysis-gluconeogenesis pathway under aerobic condition. Differentially expressed genes were determined using Deseq2 analysis comparing control and treated groups. Green, orange, and red color indicates upregulated, downregulated, and expressed genes respectively.

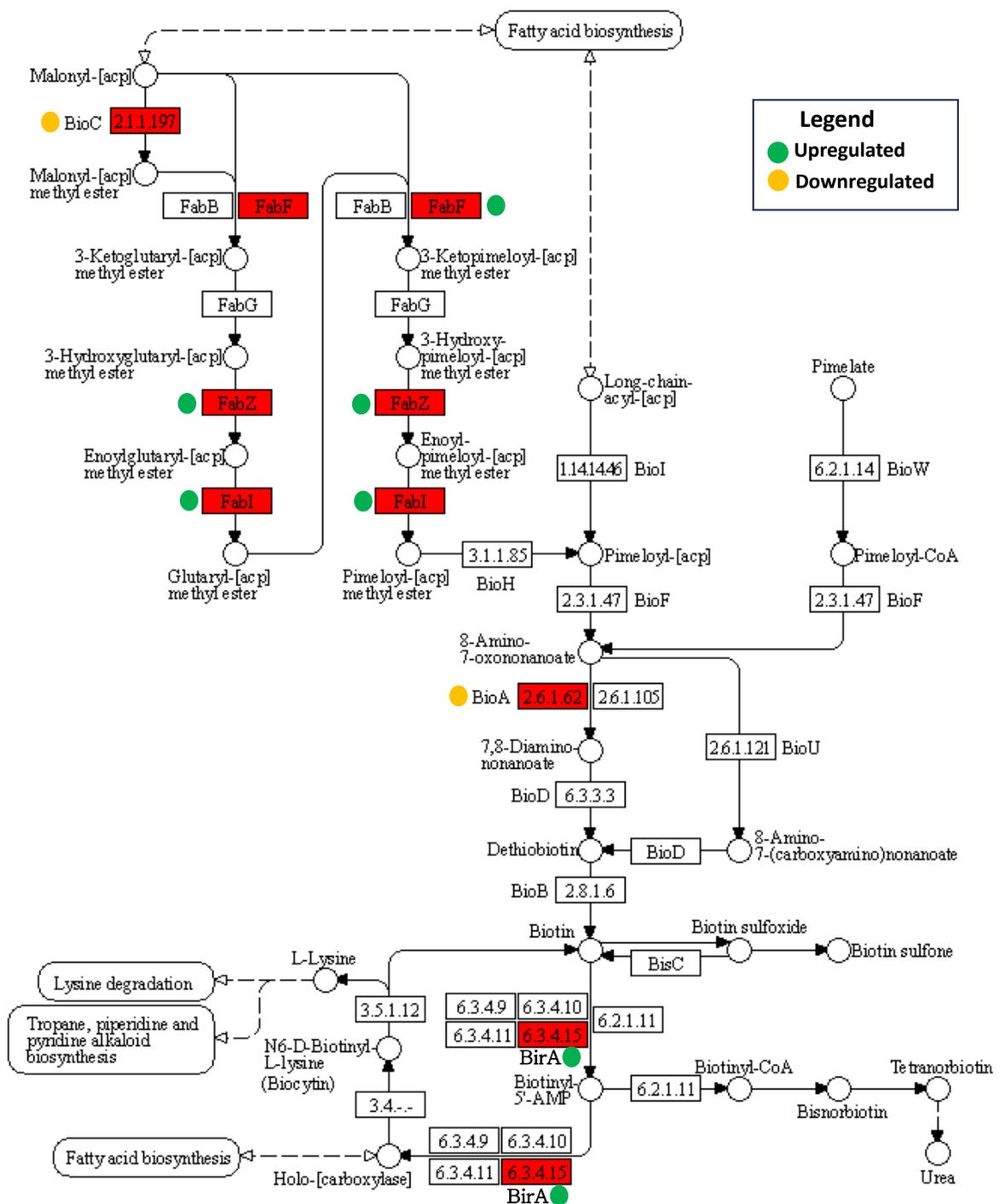


*S* Figure 2.2. Effect of catechol exposure on glycolysis-gluconeogenesis pathway under anaerobic condition. Differentially expressed genes were determined using Deseq2 analysis comparing control and treated groups. Green, orange, and red color indicates upregulated, downregulated, and expressed genes respectively.

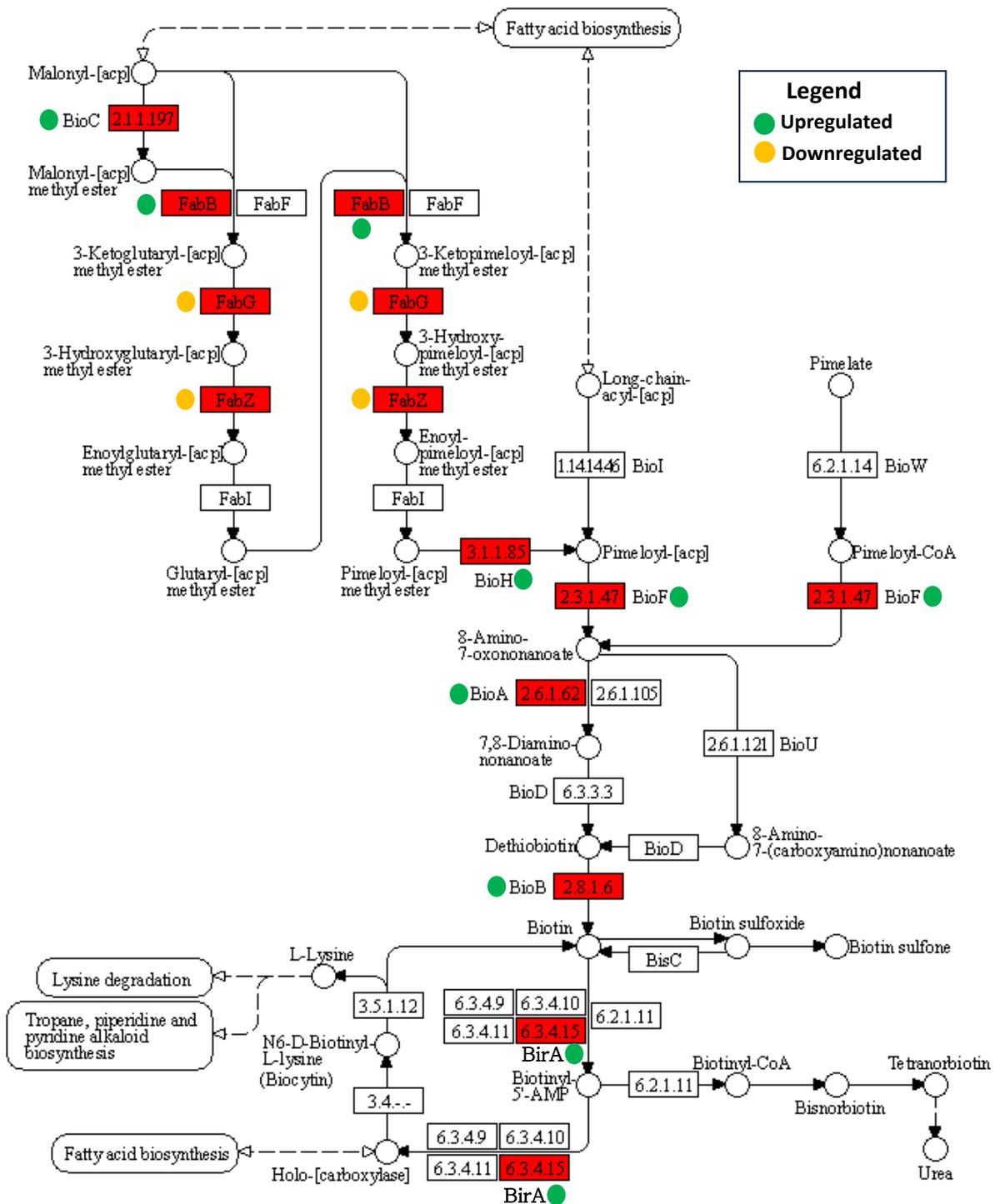


**S Figure 2.3.** Effect of catechol exposure on pentose phosphate pathway under aerobic (A) and anaerobic (B) conditions. Differentially expressed genes were determined using

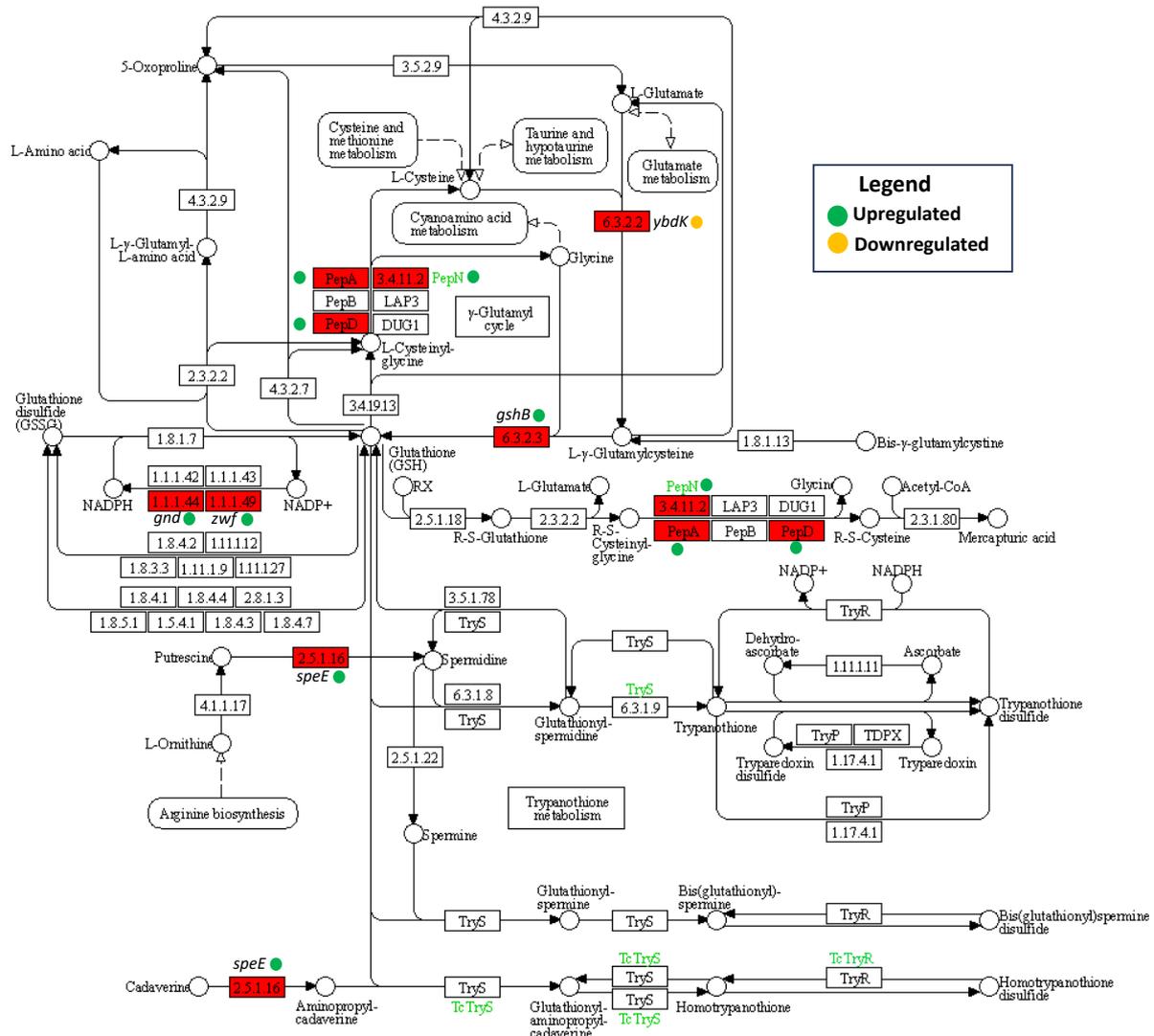
*Deseq2 analysis comparing control and treated groups. Green, orange, and red color indicates upregulated, downregulated, and expressed genes respectively.*



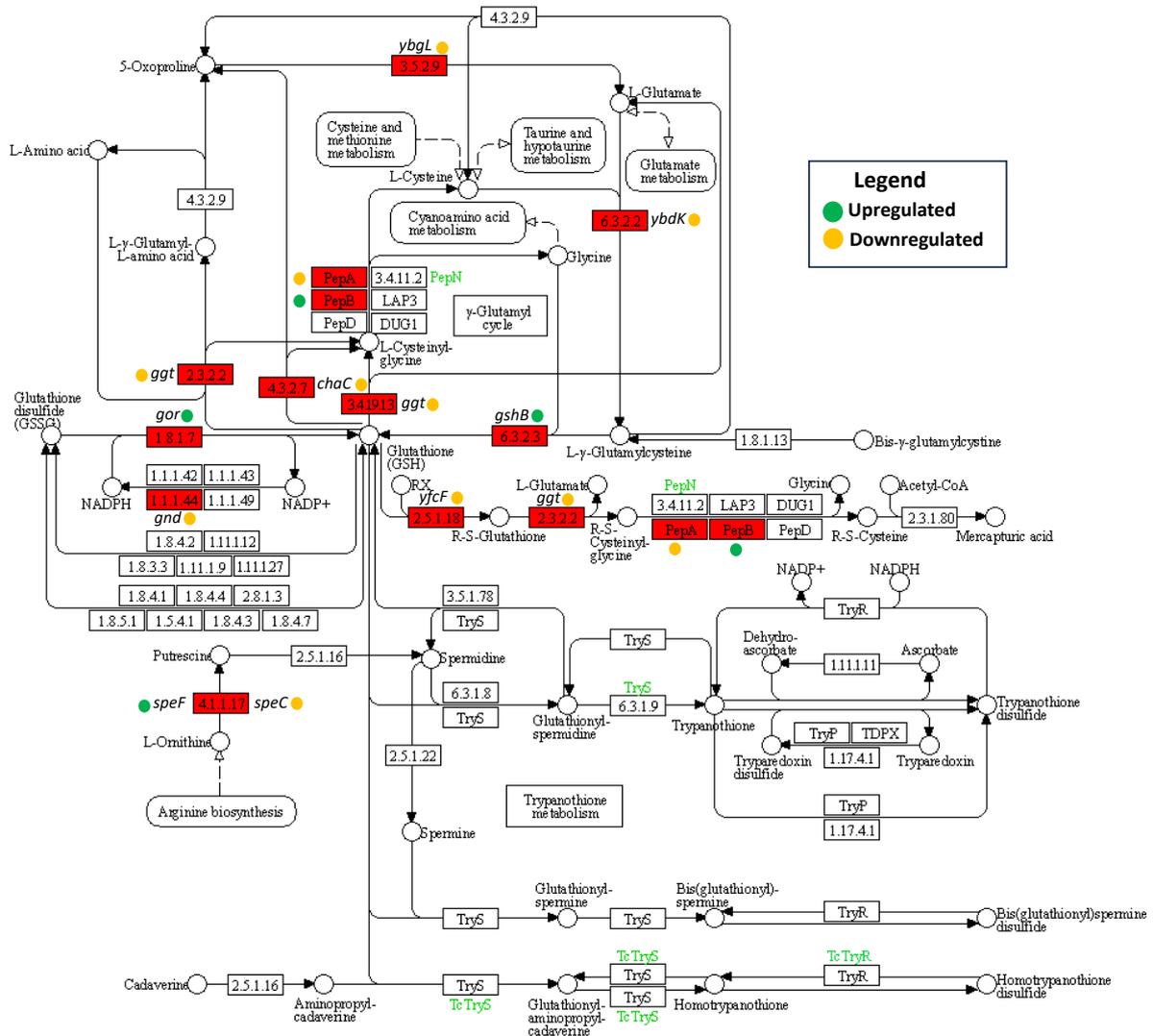
**S Figure 2.4.** Effect of catechol exposure on biotin metabolism under aerobic condition. Differentially expressed genes were determined using Deseq2 analysis comparing control and treated groups. Green, orange, and red color indicates upregulated, downregulated, and expressed genes respectively.



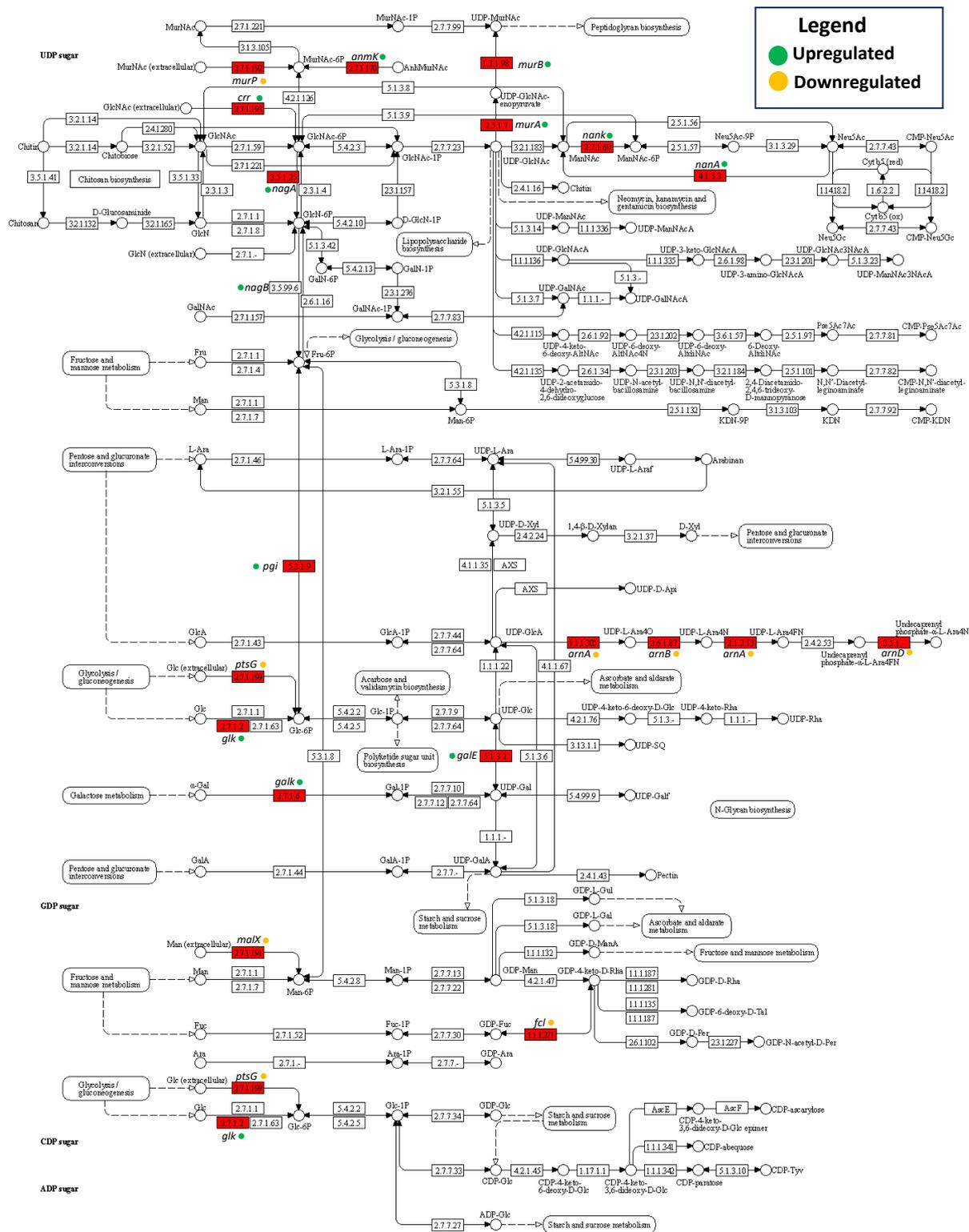
**S Figure 2.5.** Effect of catechol exposure on biotin metabolism under anaerobic condition. Differentially expressed genes were determined using Deseq2 analysis comparing control and treated groups. Green, orange, and red color indicates upregulated, downregulated, and expressed genes respectively.



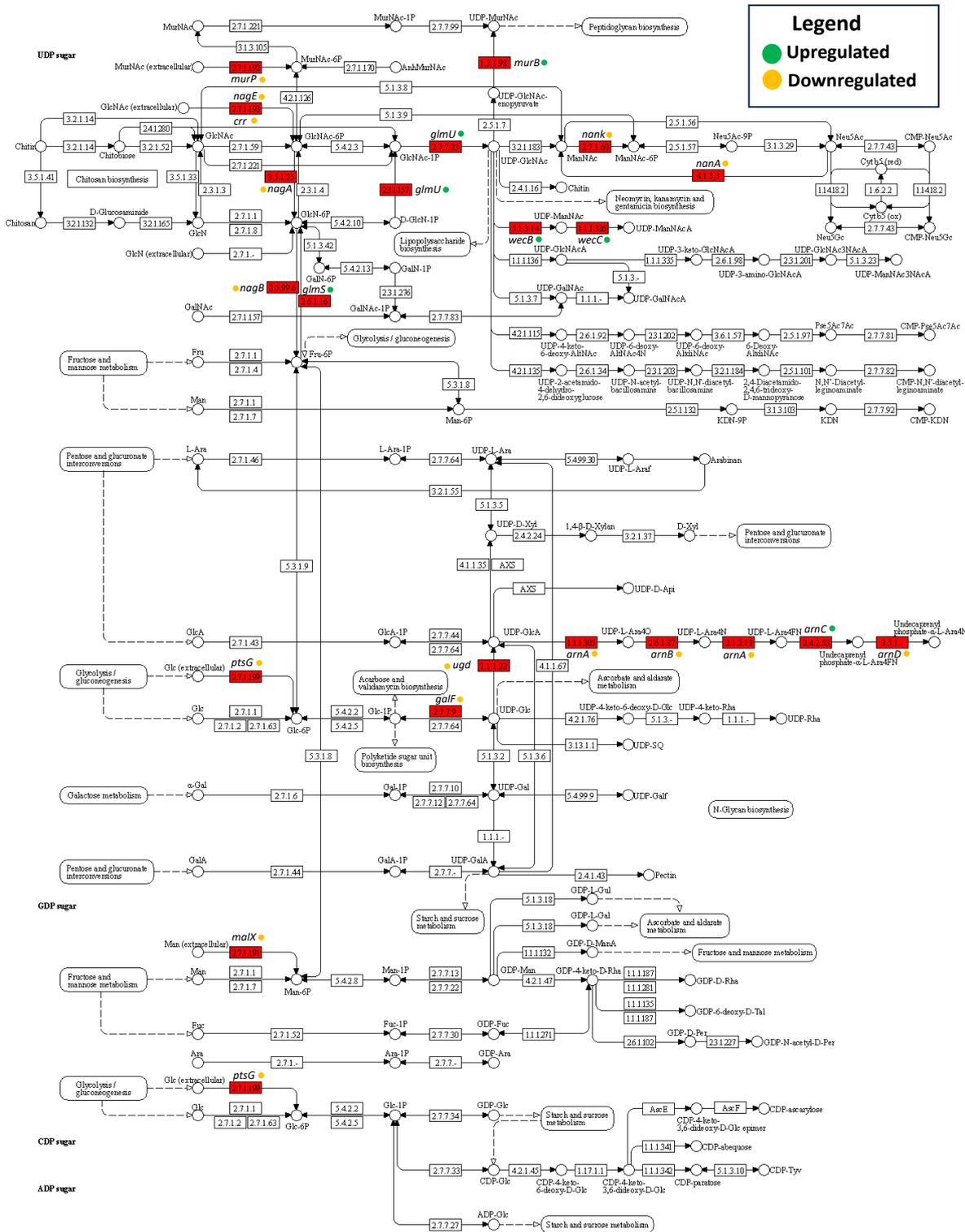
**S Figure 2.6.** Effect of catechol exposure on glutathione metabolism under aerobic condition. Differentially expressed genes were determined using Deseq2 analysis comparing control and treated groups. Green, orange, and red color indicates upregulated, downregulated, and expressed genes respectively.



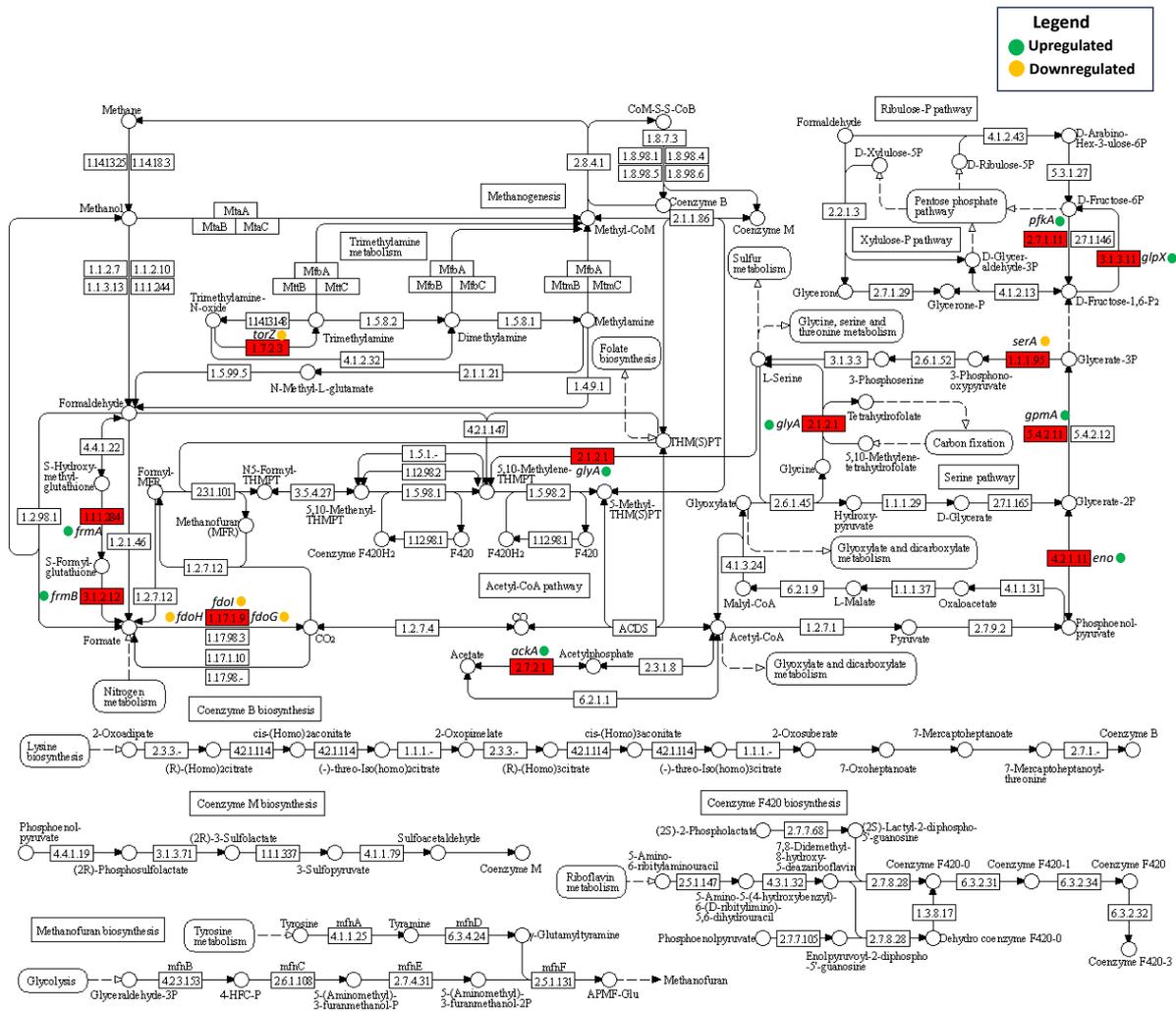
**S Figure 2.7.** Effect of catechol exposure on glutathione metabolism under anaerobic condition. Differentially expressed genes were determined using Deseq2 analysis comparing control and treated groups. Green, orange, and red color indicates upregulated, downregulated, and expressed genes respectively.



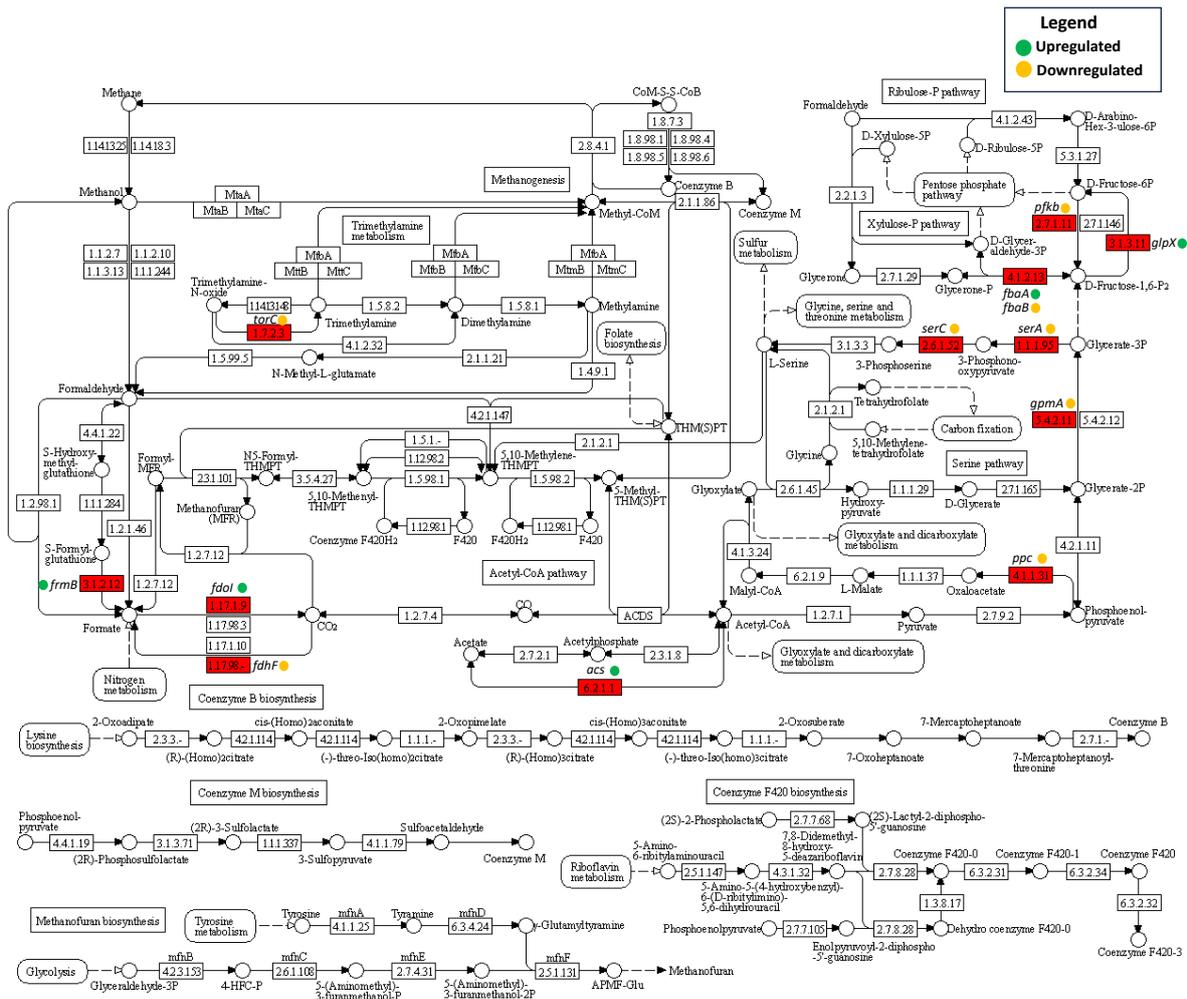
**S Figure 2.8.** Effect of catechol exposure on amino sugar and nucleotide sugar metabolism under aerobic condition. Differentially expressed genes were determined using *DeSeq2* analysis comparing control and treated groups. Green, orange, and red color indicates upregulated, downregulated, and expressed genes respectively.



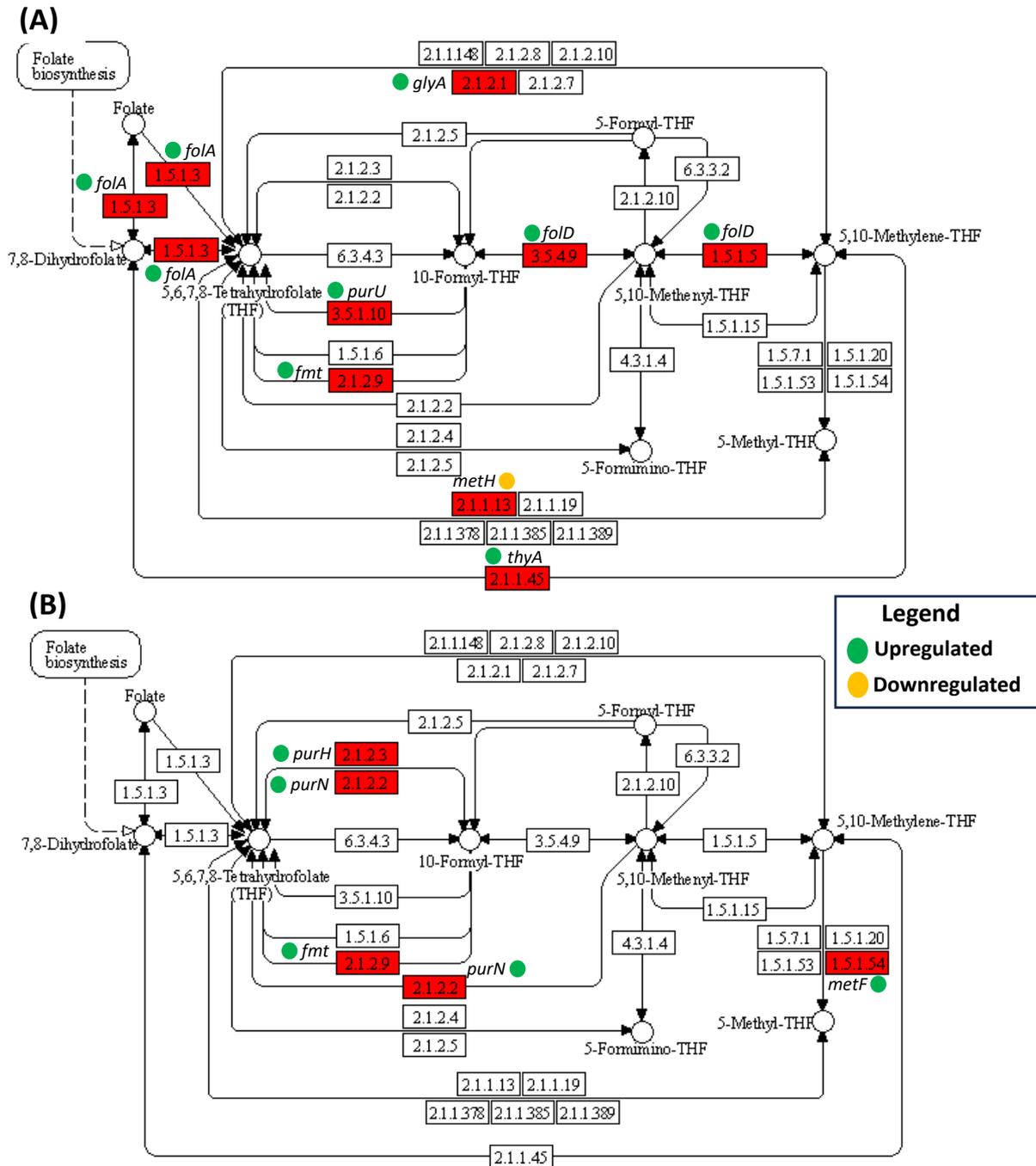
**S Figure 2.9.** Effect of catechol exposure on amino sugar and nucleotide sugar metabolism under anaerobic condition. Differentially expressed genes were determined using Deseq2 analysis comparing control and treated groups. Green and red color indicates upregulated and downregulated genes respectively.



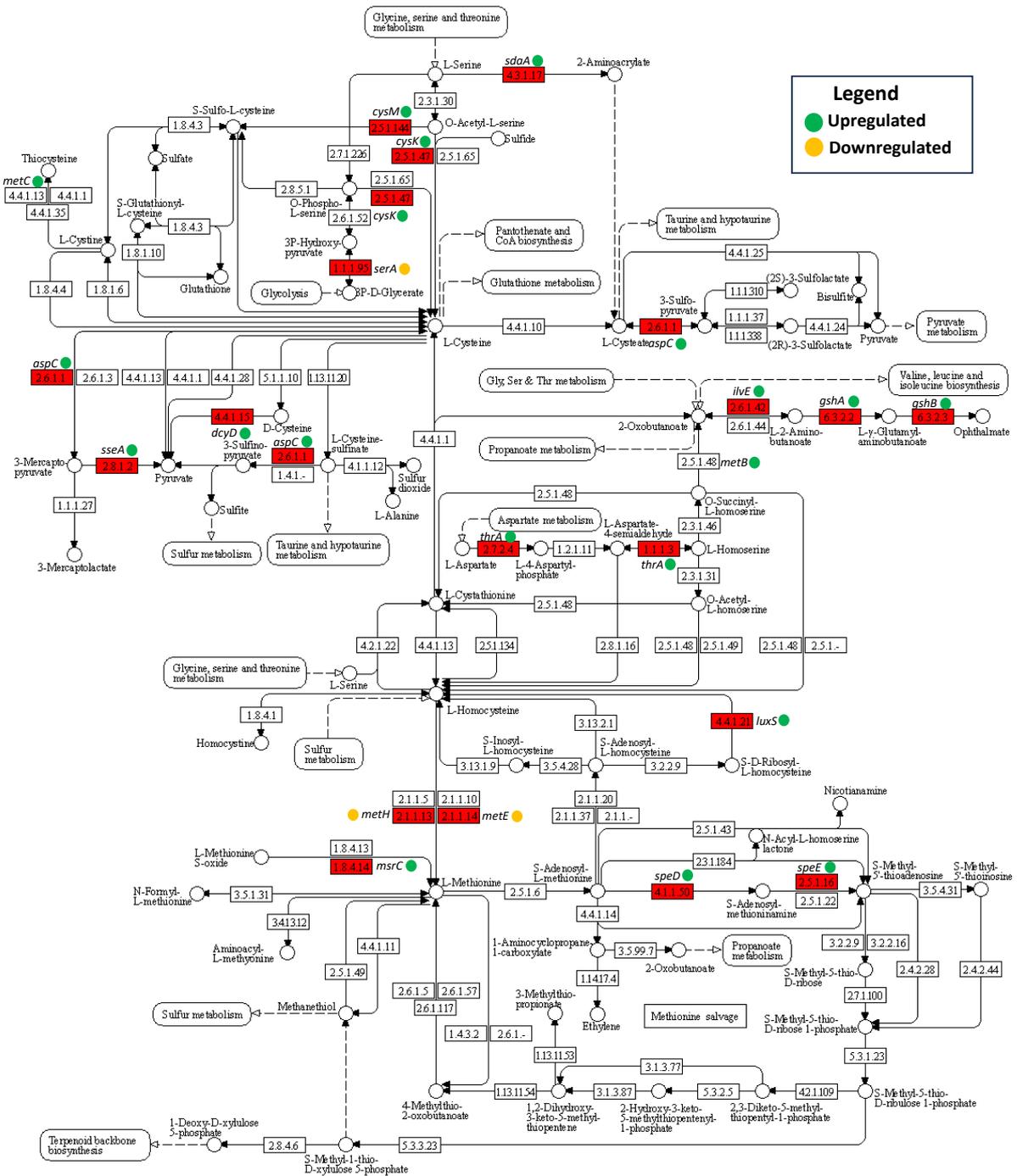
**S Figure 2.10.** Effect of catechol exposure on methane metabolism under aerobic condition. Differentially expressed genes were determined using Deseq2 analysis comparing control and treated groups. Green and red color indicates upregulated and downregulated genes respectively.



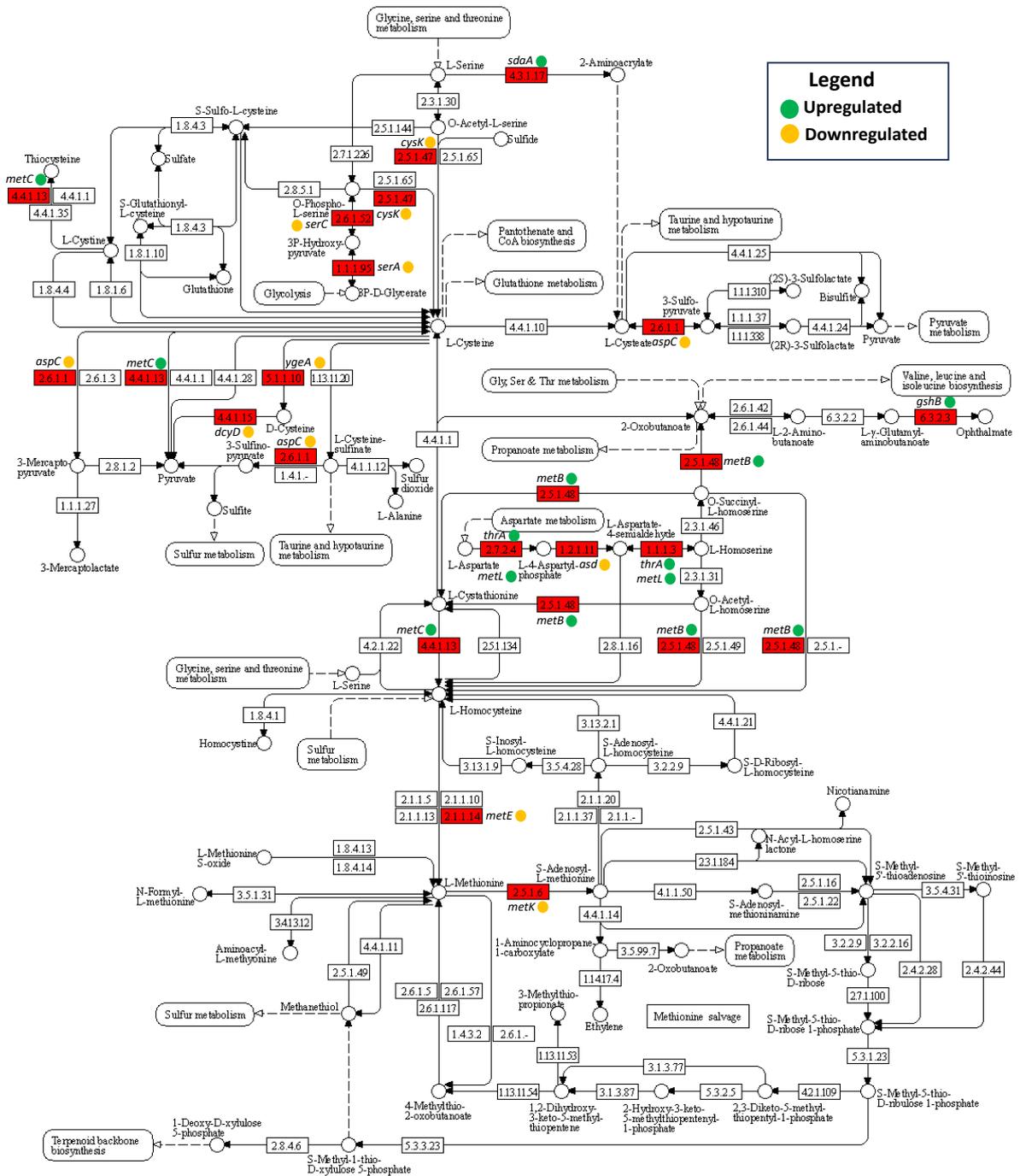
**S Figure 2.11.** Effect of catechol exposure on methane metabolism under anaerobic condition. Differentially expressed genes were determined using Deseq2 analysis comparing control and treated groups. Green, orange, and red color indicates upregulated, downregulated, and expressed genes respectively.



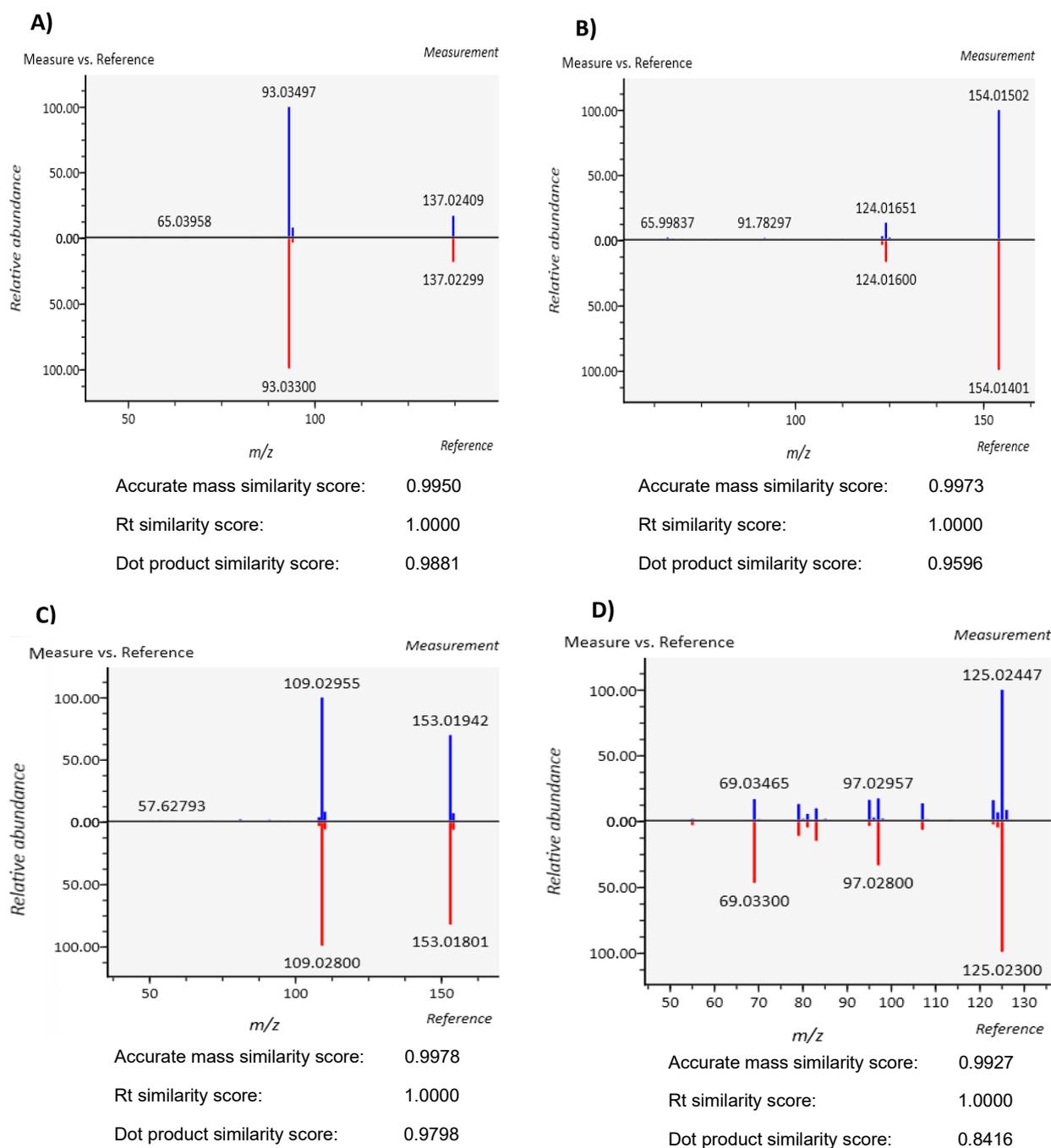
**S Figure 2.12.** Effect of catechol exposure on one carbon metabolism under aerobic (A) and anaerobic (B) conditions. Differentially expressed genes were determined using Deseq2 analysis comparing control and treated groups. Green, orange, and red color indicates upregulated, downregulated, and expressed genes respectively.



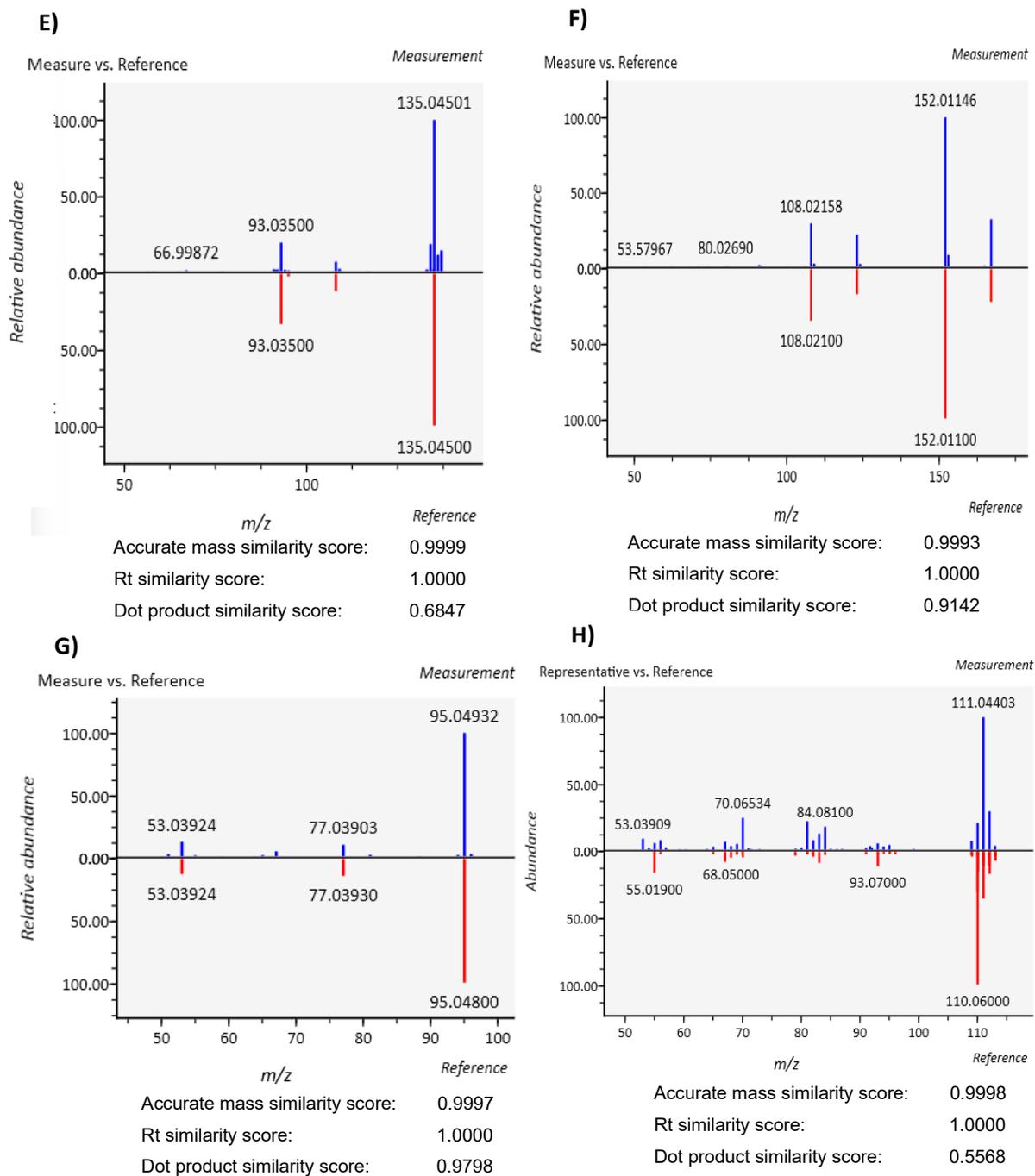
**S Figure 2.13.** Effect of catechol exposure on cysteine and methionine metabolism under aerobic condition. Differentially expressed genes were determined using Deseq2 analysis comparing control and treated groups. Green, orange, and red color indicates upregulated, downregulated, and expressed genes respectively.



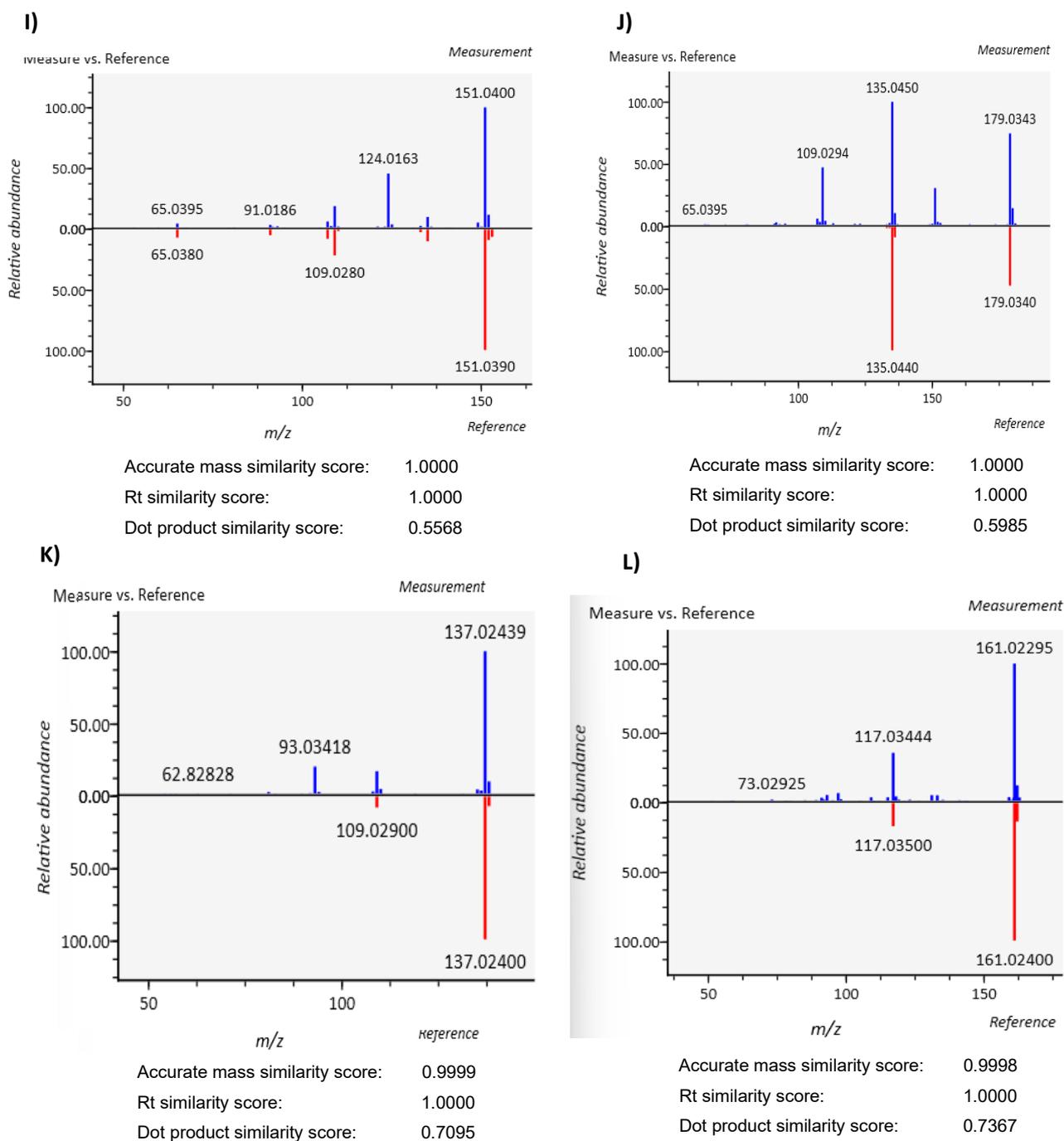
**S Figure 2.14.** Effect of catechol exposure on cysteine and methionine metabolism under anaerobic condition. Differentially expressed genes were determined using Deseq2 analysis comparing control and treated groups. Green, orange, and red color indicates upregulated, downregulated, and expressed genes respectively.



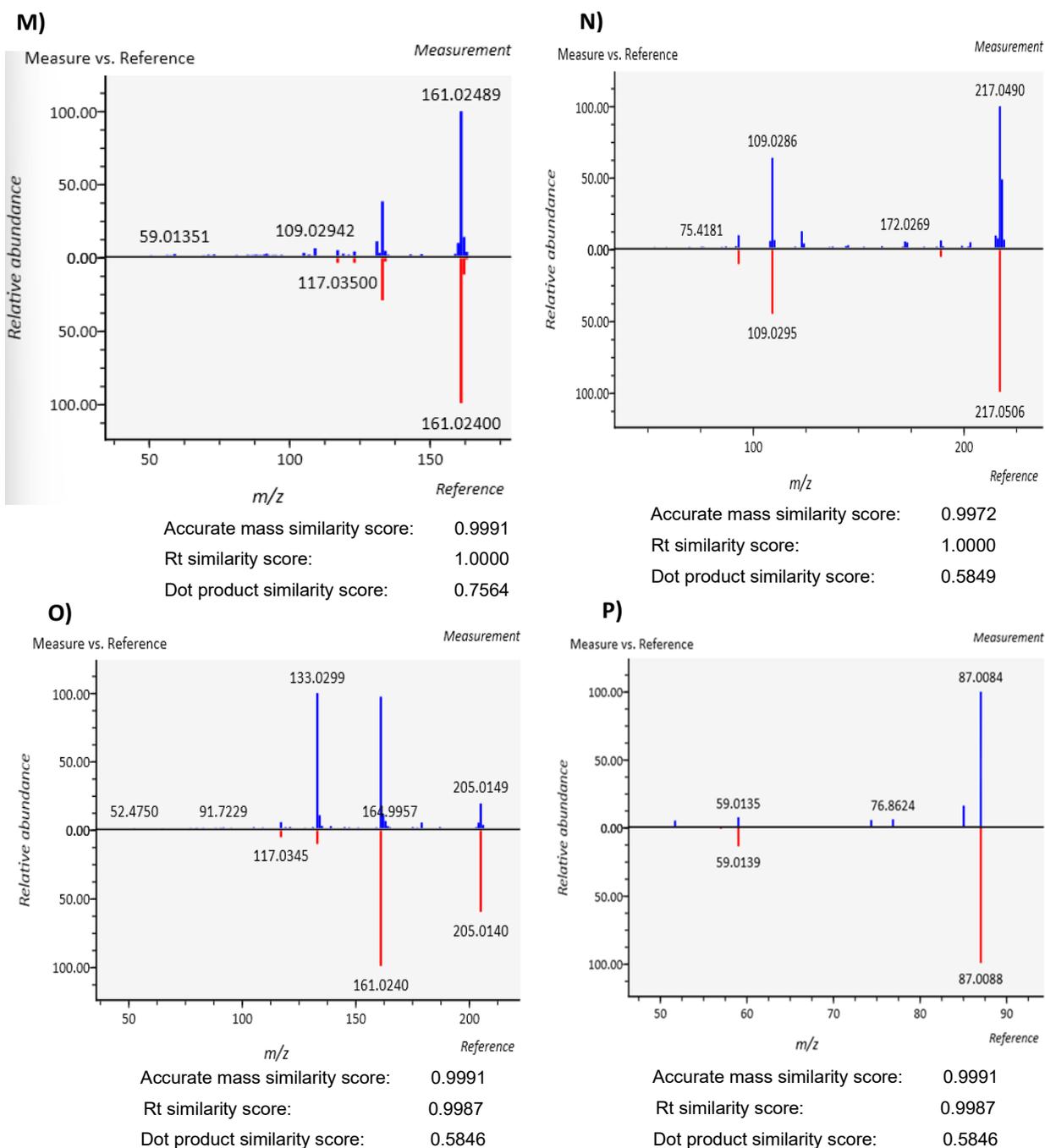
**S Figure 2.15.** Comparison of fragmentation pattern of 4-hydroxybenzoate, 4-nitrocatechol, and protocatechuic acid, and 1,2,4-benzenetriol with standards. (A), (B), (C) and (D) represents fragmentation pattern of 4-hydroxybenzoate, 4-nitrocatechol, protocatechuic acid, and 1,2,4-benzenetriol. All these metabolites were identified in LC-MS/MS in negative mode.



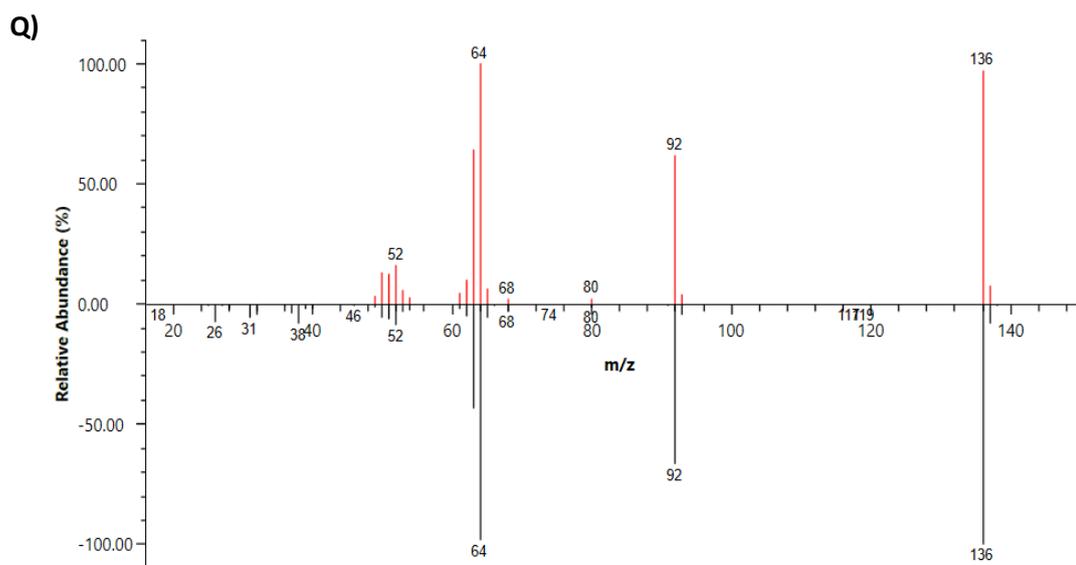
**S Figure 2.16.** Comparison of fragmentation pattern of (E) 3-methylbenzoic acid, (F) vanillic acid, (G) phenol, and (H) hydroquinone with standards. 3-Methylbenzoic acid and vanillic acid were identified in negative mode, and phenol and hydroquinone were identified using LC-MS/MS in positive mode.



**S Figure 2.17.** Comparison of fragmentation pattern of I) 2,4-dihydroxyacetophenone, J) *trans*-caffeic acid, K) protocatechuic aldehyde, and L) 4-hydroxycoumarin with standards. All metabolites were identified using LC-MS/MS in negative mode.

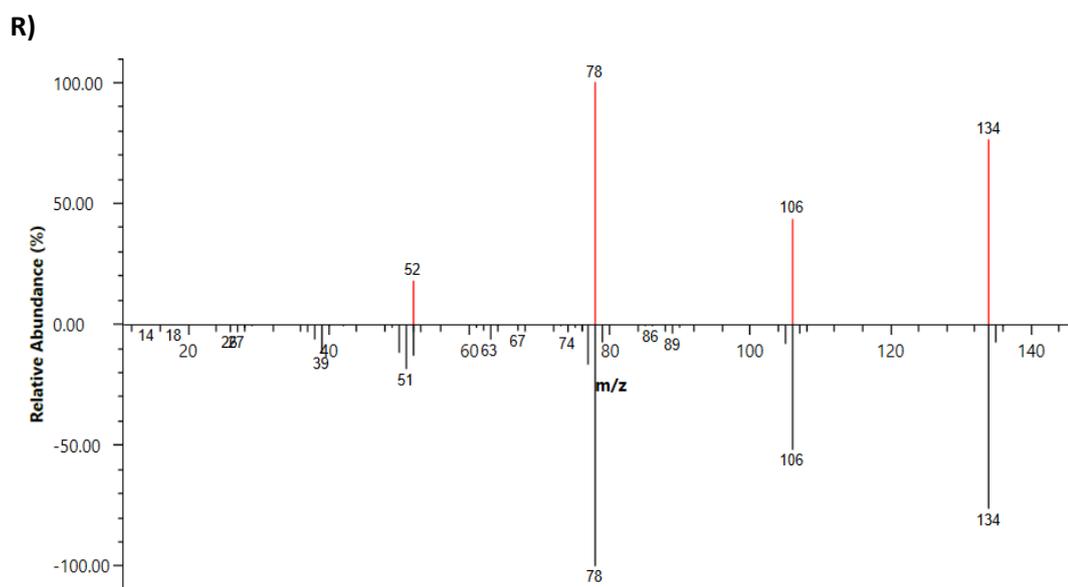


**S Figure 2.18.** Comparison of fragmentation pattern of (M) 7-hydroxycoumarin, (N) dioxydiphenol, and (O) 7-hydroxy-2-oxo-2H-chromene-3-carboxylic acid with standards. (P) Fragmentation pattern of methyl glyoxylate was predicted using CFM-ID 4.0 and compared with the experimental sample. These metabolites were identified using LC-MS/MS in negative mode.



Calculated Retention Index: 1125.65

AI predicted non-polar retention index (n-alkane scale): 1158

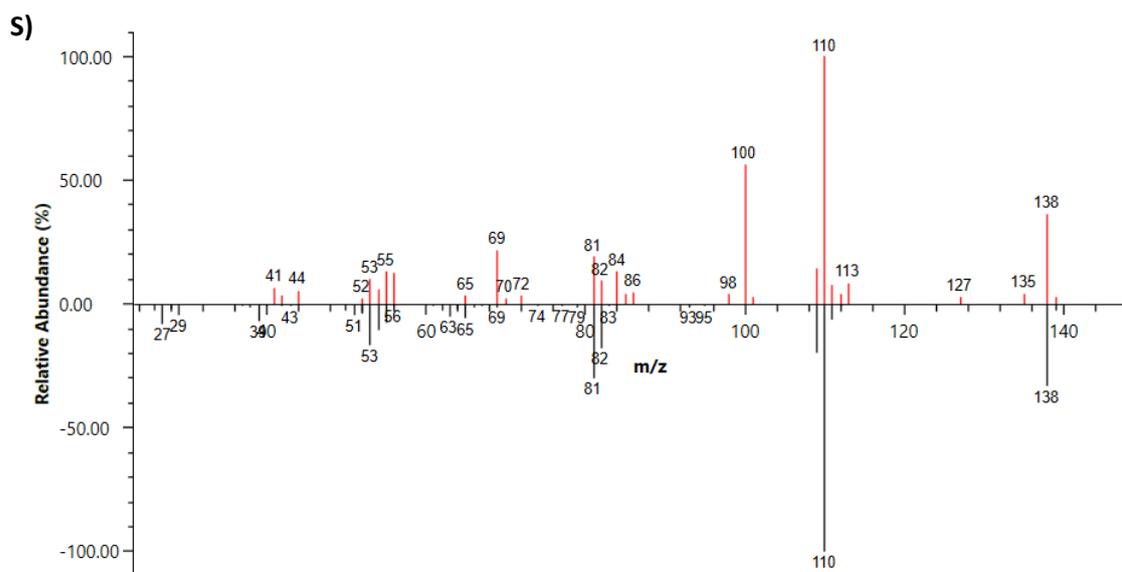


Calculated Retention Index: 1282

Semi-standard non-polar retention index (n-alkane scale): 1249

Estimated non-polar retention index (n-alkane scale): 1272

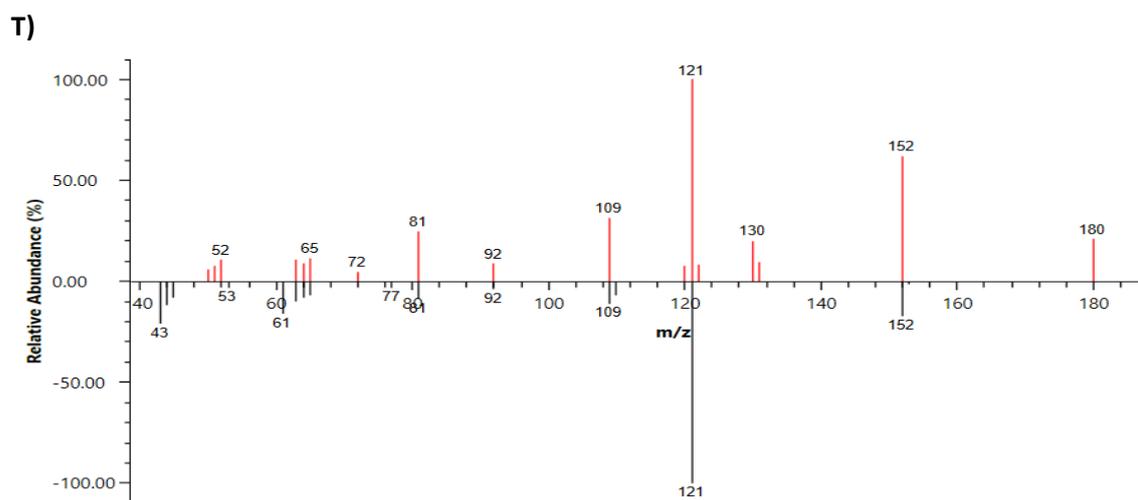
**S Figure 2.19.** Comparison of fragmentation pattern of (Q) Catechol carbonate, and (R) 2-coumaranone with NIST2.4 database. These metabolites were identified using GC-MS.



Calculated Retention Index: 1333

Standard non-polar retention index (n-alkane scale): 1350

Kovat's AI: 1300.165

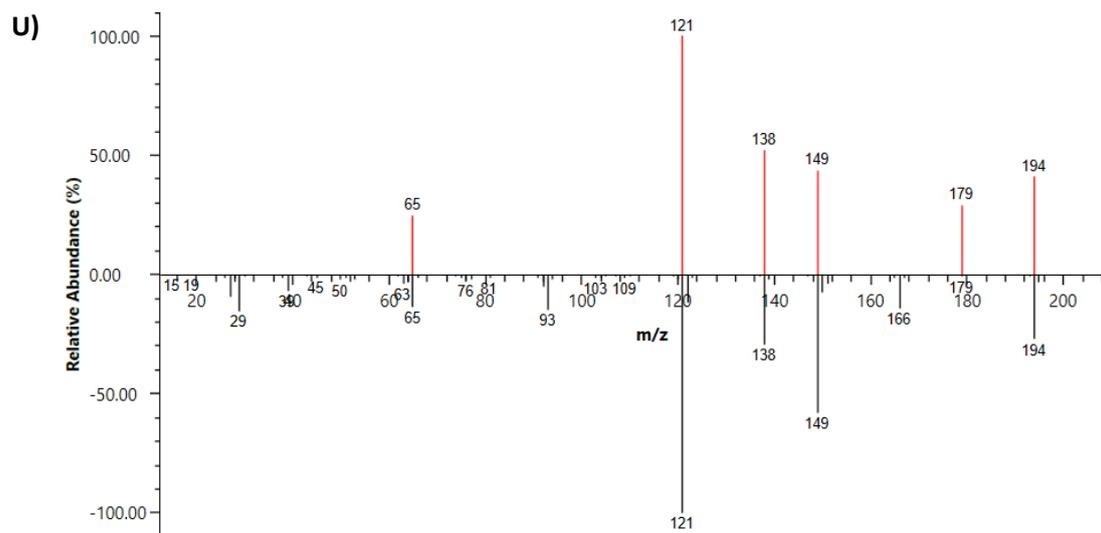


Calculated Retention Index: 1120

Estimated non-polar retention index (n-alkane scale): 1162

AI predicted non-polar retention index (n-alkane scale): 1153

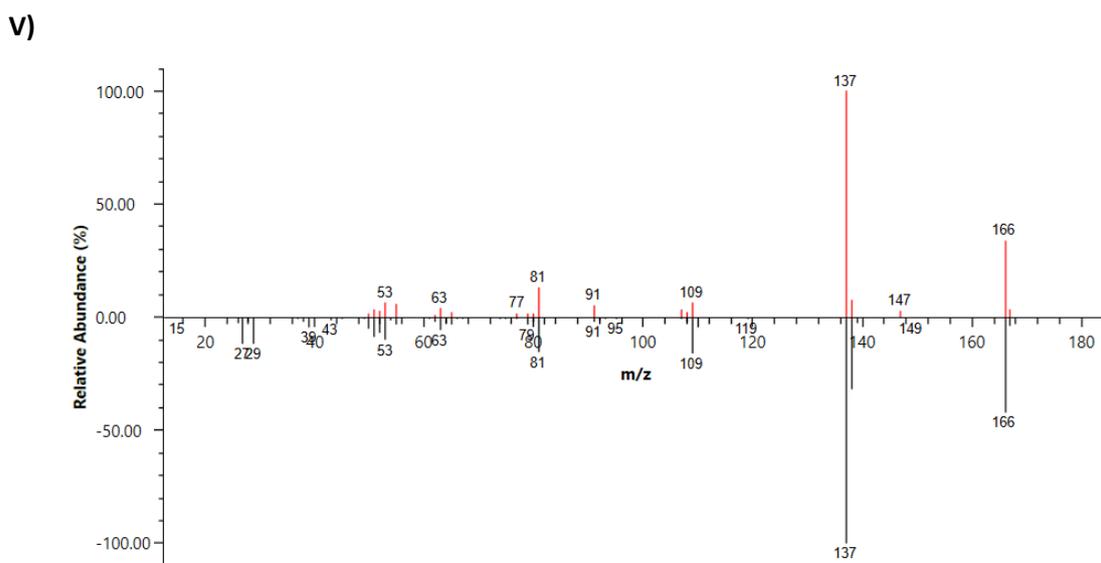
**S Figure 2.20.** Comparison of fragmentation pattern of (S) 4-ethoxyphenol, and (T) 2-methoxy-1,3-benzodioxole with NIST2.4 database. These metabolites were identified using GC-MS.



Calculated Retention Index: 1556.77

Semi-standard non-polar retention index (n-alkane scale): 1522

AI predicted non-polar retention index (n-alkane scale): 1567

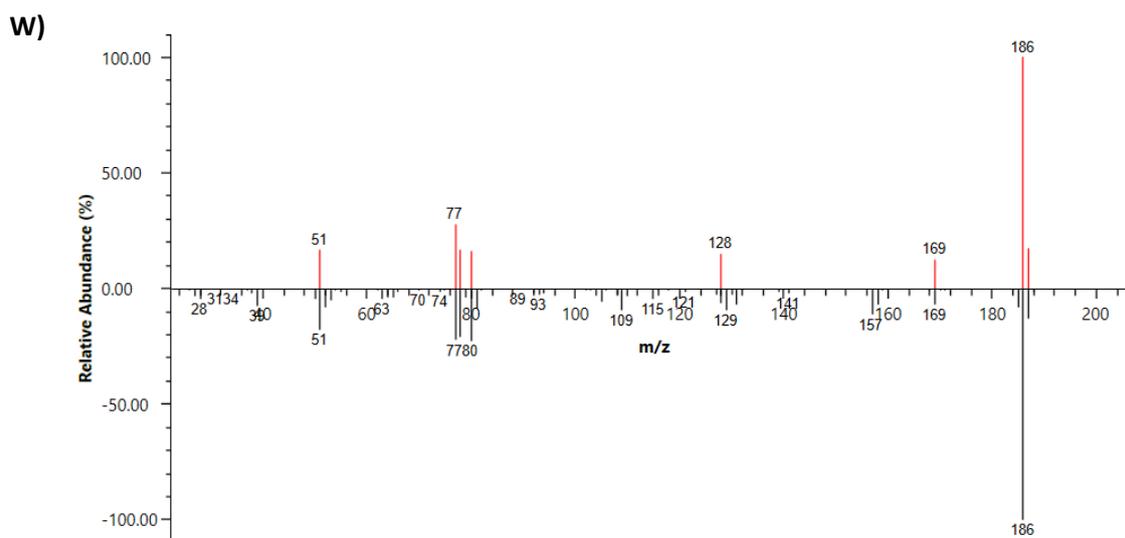


Calculated Retention Index: 1424.94

Semi-standard non-polar retention index (n-alkane scale): 1460

Standard non-polar retention index (n-alkane scale): 1433 ± 15

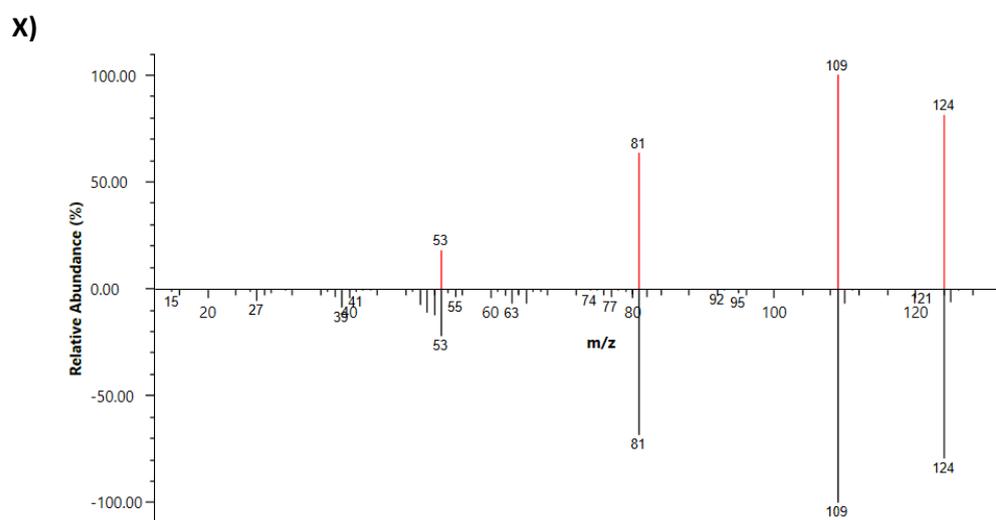
**S Figure 2.21.** Comparison of fragmentation pattern of (U) 4-Ethoxybenzoic acid ethyl ester, and (V) Ethyl vanillin with standards of NIST2.4 database. These metabolites were identified using GC-MS.



Calculated Retention Index: 1573

Semi-standard non-polar retention index (n-alkane scale): 1547

AI predicted non-polar retention index (n-alkane scale): 1567

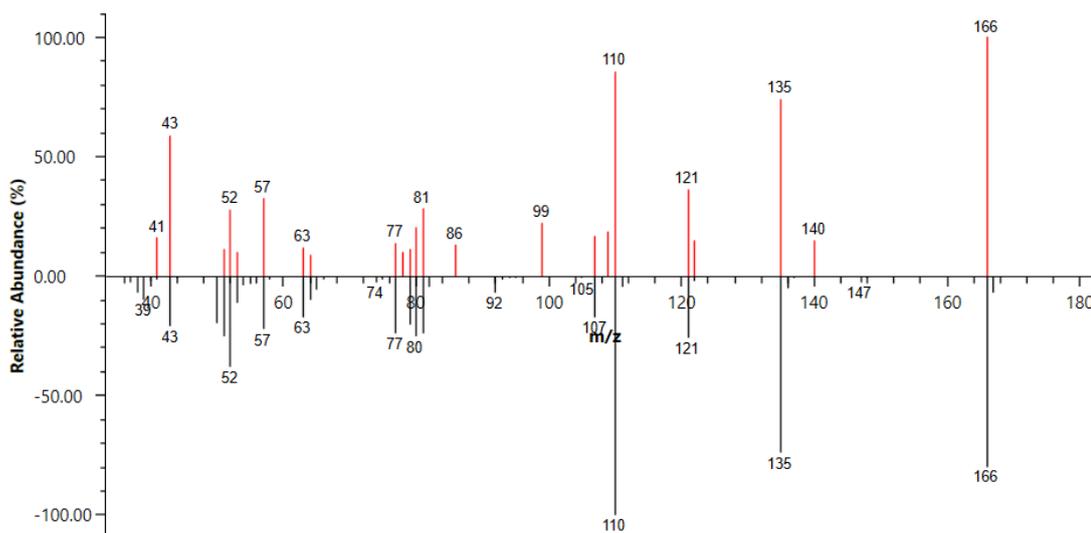


Calculated Retention Index: 1093

Semi-standard non-polar retention index (n-alkane scale): 1090 ± 3

**S Figure 2.22.** Comparison of fragmentation pattern of (W) 2-Phenoxyphenol, (X) 2-Methoxyphenol with standards of NIST2.4 database. These metabolites were identified using GC-MS.

Y)

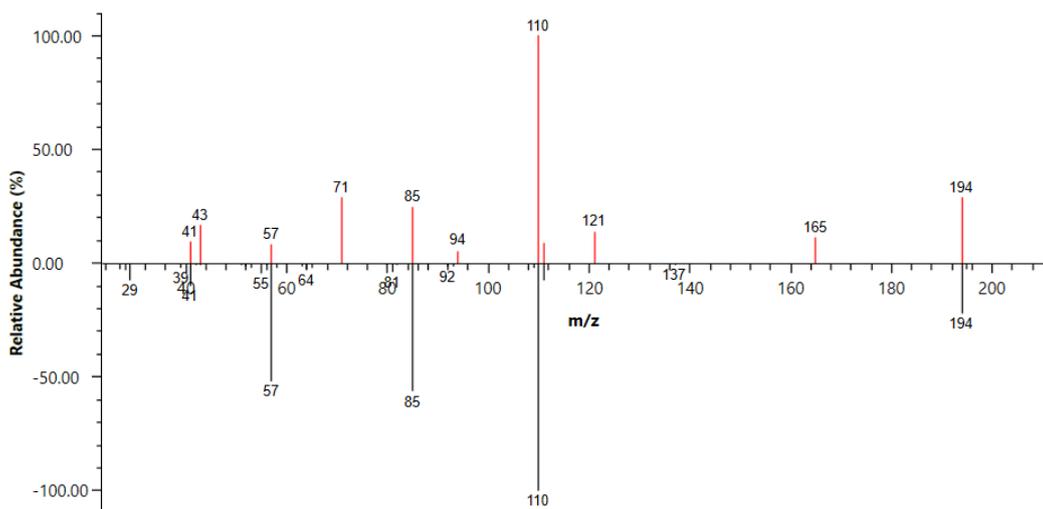


Calculated Retention Index: 1487.54

Estimated non-polar retention index (n-alkane scale): 1448

AI predicted non-polar retention index (n-alkane scale): 1452

Z)

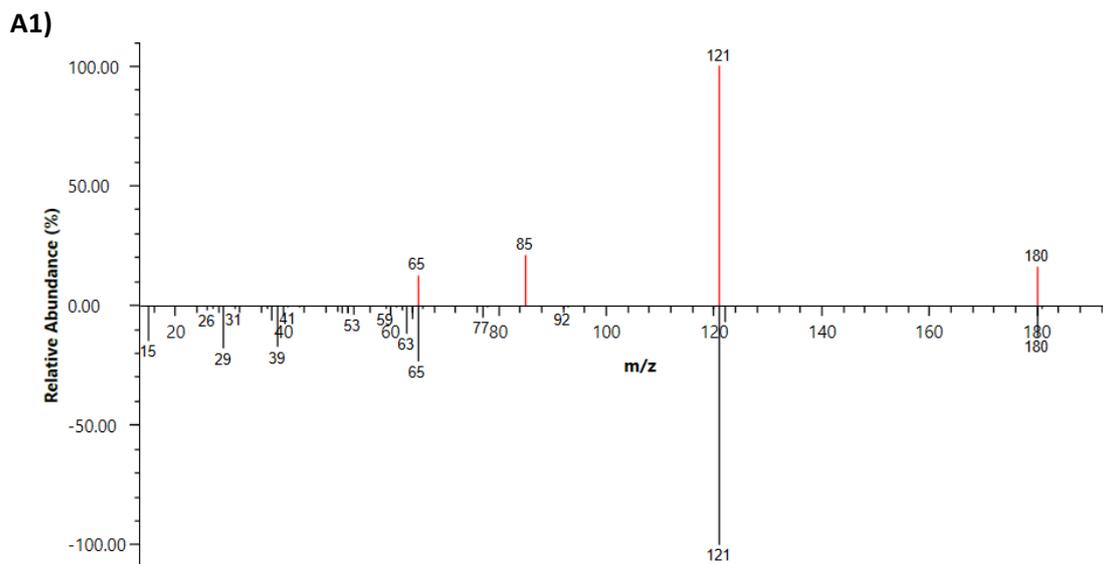


Calculated Retention Index: 1512.21

Estimated non-polar retention index (n-alkane scale): 1579

AI predicted non-polar retention index (n-alkane scale): 1534

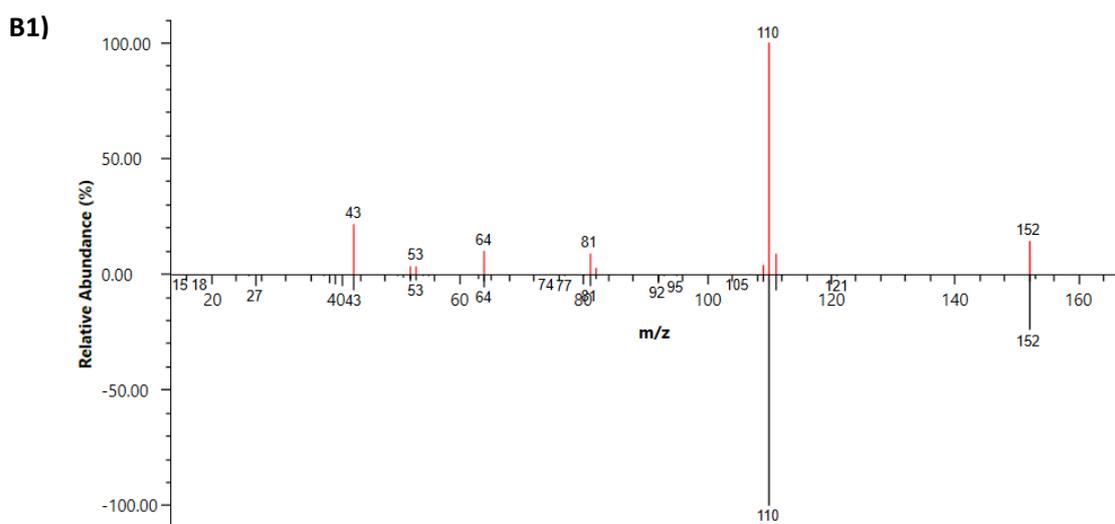
**S Figure 2.23.** Comparison of fragmentation pattern of (Y) 2-Hydroxymethyl-1,4-benzodioxan, and (Z) O-valeryl-1,2-benzenediol with standards of NIST2.4 database. These metabolites were identified using GC-MS.



Calculated Retention Index: 1367.35

Estimated non-polar retention index (n-alkane scale): 1352

AI predicted non-polar retention index (n-alkane scale):1350

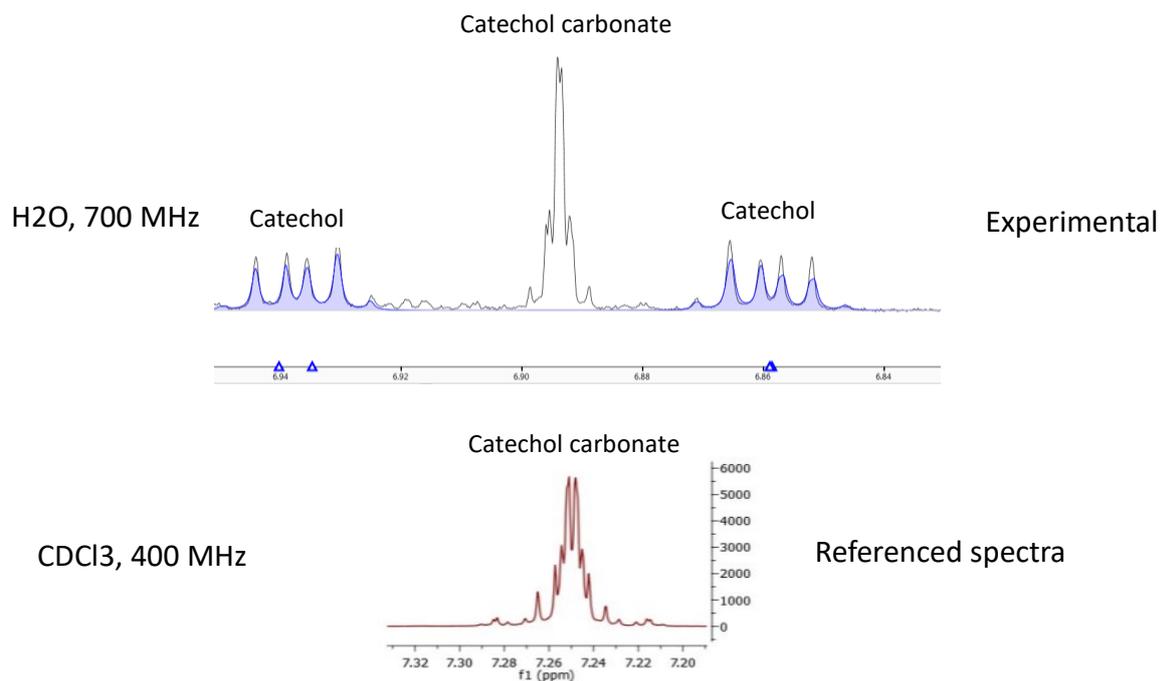


Calculated Retention Index: 1283.91

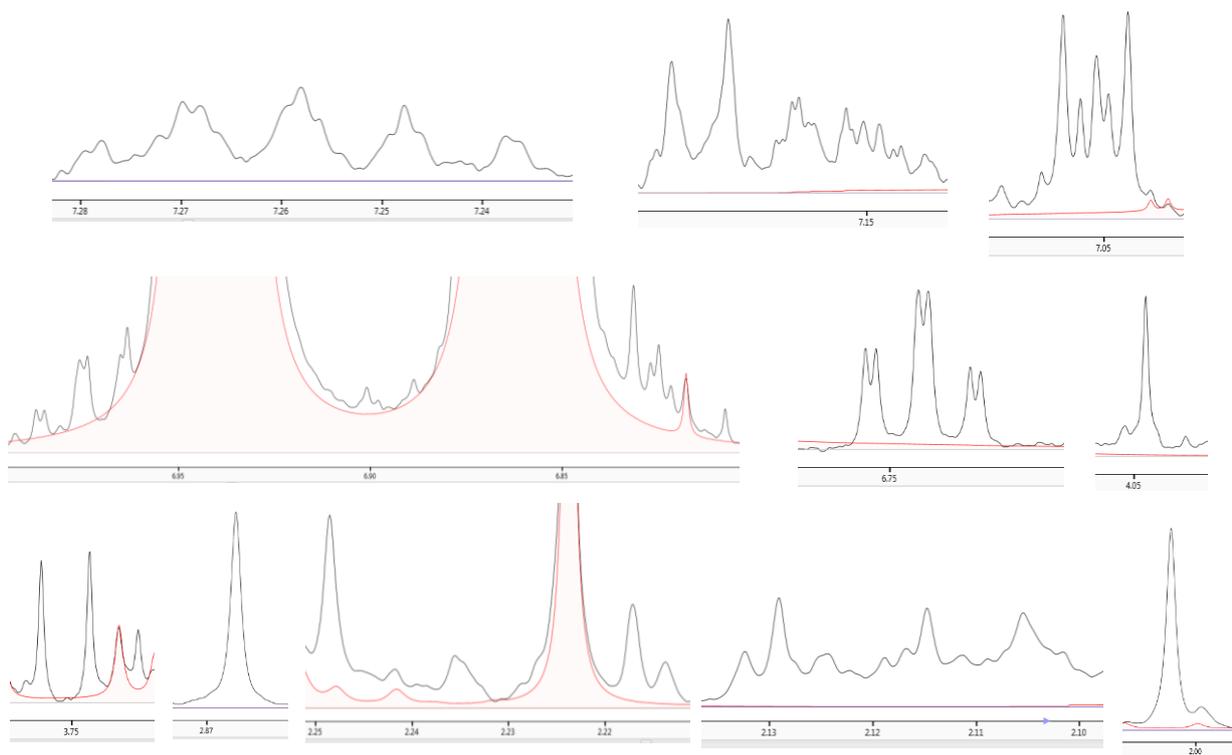
Estimated non-polar retention index (n-alkane scale): 1289

AI predicted non-polar retention index (n-alkane scale): 1251

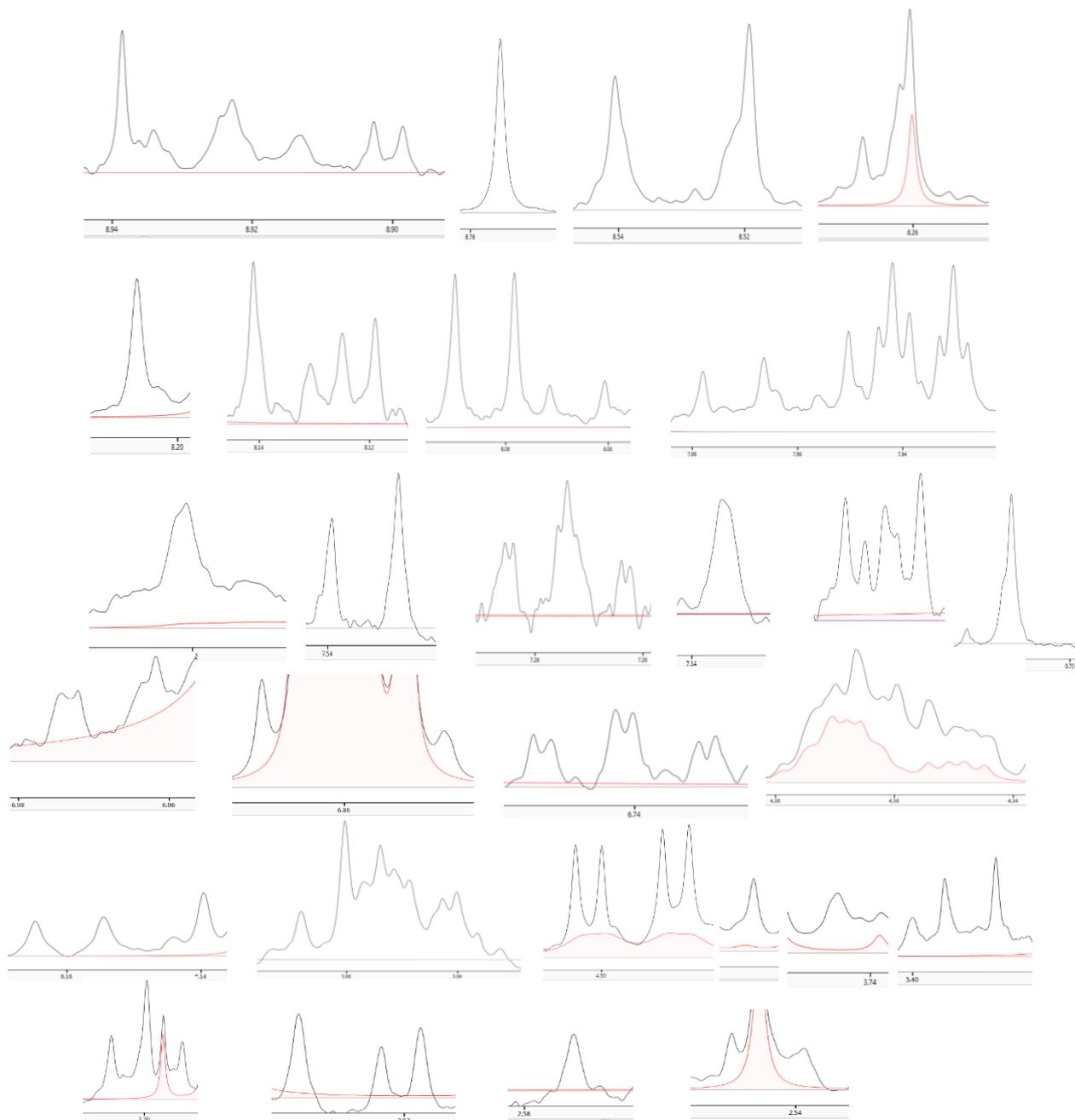
**S Figure 2.24.** Comparison of fragmentation pattern of (A1) Methyl 1,3-benzodioxole-2-carboxylate, and (B1) 2-propoxyphenol of NIST2.4 database. These metabolites were identified using GC-MS.



**S Figure 2.25.** Determination of catechol carbonate in the reaction between catechol and carbonic acid. Catechol carbonate was detected in 700 MHz NMR. Water was used to prepare the experimental sample. Referenced spectra of catechol carbonate was prepared using deuterated chloroform, reported in the article (Tabanelli et al., 2018).



**S Figure 2.26.** NMR analysis detected unknown compounds in catechol treated samples (extracellular) under anaerobic condition. Overlapping peaks (such as chemical shifts at 6.95, 6.85 and 2.21 ppm) designated by light red color.



**S Figure 2.27.** NMR analysis detected unknown compounds in catechol treated samples (intracellular) under anaerobic condition. Overlapping peaks (such as chemical shifts at 8.26, 6.96, 6.86, 4.37, 4.35 and 2.21 ppm) designated by light red color.