Master's Thesis

Masters Programme in Chemistry and Molecular Science

Physical Chemistry

Acetaldehyde Production in *Escherichia coli*: Headspace Analysis of Responses to Ethanol, Glucose, and Diabetic Fecal Extract

Ville Salminen

13.2.2026

Supervisor(s): Docent Markus Metsälä

Examiner(s): Docent Markus Metsälä
Professor Markku Vainio

UNIVERSITY OF HELSINKI

FACULTY OF SCIENCE

PL 33 (A.I. Virtasen aukio 1)
00014 Helsingin yliopisto

Abstract

Faculty: Faculty of Science
Degree Programme: Masters Programme in Chemistry and Molecular Science
Study Track: Physical Chemistry
Author: Ville Salminen
Title: Acetaldehyde Production in *Escherichia coli*: Headspace Analysis of Responses to Ethanol, Glucose, and Diabetic Fecal Extract
Level: Master's Thesis
Month and Year: 13.2.2026
Number of Pages: 55
Keywords: acetaldehyde, ethanol, glucose, fecal extract, *Escherichia coli*, diabetes mellitus
Supervisor(s): Docent Markus Metsälä
Where Deposited: HELDA - Digital Repository of the University of Helsinki
Additional Information:
Abstract:

In my Bachelor's thesis, a correlation was found between the blood glucose levels of a diabetic test subject and acetaldehyde concentration in their exhaled breath. This observation led to the hypothesis that poor glycemic control may increase bacterial production of acetaldehyde in the intestine, leading to its local accumulation. Acetaldehyde is a known carcinogen, and its local accumulation has been associated with an increased risk of cancer. This thesis aimed to validate the hypothesis *in vitro*.

The study aimed to validate previous findings on bacterial acetaldehyde production from glucose and ethanol. Additionally, the study explored whether the stool of individuals with diabetes mellitus contains something that increases acetaldehyde production. Metabolic responses were analyzed using liquid-based headspace measurements under aerobic conditions. *Escherichia coli* was used as a model organism. Proton-transfer-reaction mass spectrometry and gas chromatography-mass spectrometry were used as analytical techniques. The results confirmed acetaldehyde production in response to glucose. The response to ethanol appeared to be a false positive, because the metabolic pathway by which *Escherichia coli* oxidizes ethanol yields only minimal amounts of acetaldehyde. In addition, the fecal extracts produced a weaker-than-expected response. These findings did not support the hypothesis. However, the topic is relevant, and further studies should be considered. Transitioning to anaerobic conditions is essential as it would simulate the intestinal environment and likely yield data more relevant to the hypothesis.

Tiivistelmä

Tiedekunta: Matemaattis-luonnontieteellinen tiedekunta
Koulutusohjelma: Kemian ja molekyylitieteiden maisteriohjelma
Opintosuunta: Fysikaalinen kemia
Tekijä: Ville Salminen
Työn nimi: Asetaldehydin muodostus *Escherichia coli* -bakteerissa: Kaasufaasianalyysi vasteista etanolille, glukoosille ja diabeetikon ulostenäytteestä valmistetulle uutteelle
Työn laji: Maisterintutkielma
Kuukausi ja vuosi: 13.2.2026
Sivumäärä: 55
Avainsanat: asetaldehydi, etanoli, glukoosi, ulosteuute, *Escherichia coli*, diabetes mellitus
Ohjaaja(t): Dosentti Markus Metsälä
Säilytyspaikka: HELDA - Helsingin yliopiston digitaalinen arkisto
Muita tietoja:
Tiivistelmä:

Kandidaatintutkielmassani havaittiin korrelaatio diabeetikon verensokeritasojen ja hänen uloshengityksensä asetaldehyditasojen välillä. Tämä havainto johti hypoteesiin, jonka mukaan huono sokeritasapaino voi lisätä bakteerien asetaldehydituotantoa suolistossa, mikä johtaa sen paikalliseen kertymiseen suolistossa. Asetaldehydi on tunnettu karsinogeeni ja sen paikallinen kertyminen on yhdistetty kohonneeseen syöpäriskiin. Tämän tutkielman tarkoituksena oli testata hypoteesia *in vitro*.

Tutkimuksessa pyrittiin vahvistamaan aiempia havaintoja bakteerien kyvystä tuottaa asetaldehydiä glukoosista ja etanolista. Lisäksi tutkittiin sisältääkö diabeettikkojen uloste jotakin, mikä edistää asetaldehydin tuotantoa. Metabolisia vasteita analysoitiin aerobisissa olosuhteissa kasvatusliemen yläilmatilasta. Malliorganismi oli *Escherichia coli*. Analyysimenetelminä käytettiin protoninsiirtoreaktioon perustuvaa massaspektrometriaa ja kaasukromatografia-massaspektrometriaa. Tulokset vahvistivat asetaldehydin muodostumisen glukoosista. Etanoliin liittyvä vaste osoittautui virheelliseksi positiiviseksi tulokseksi, koska *Escherichia coli* -bakteerin etanolin hapettamiseen käytetty metaboliareitti tuottaa vain vähäisiä määriä asetaldehydiä. Lisäksi ulosteuutteet aiheuttivat oletettua heikomman vasteen. Nämä havainnot eivät tukeneet hypoteesia. Jatkotutkimukset ovat kuitenkin perusteltuja. Siirtyminen anaerobisiin olosuhteisiin on keskeistä, sillä se simuloisi paremmin suoliston olosuhteita ja todennäköisesti tuottaisi hypoteesin kannalta merkityksellisempää tietoa.

The role of the infinitely small in nature is infinitely great.
— Louis Pasteur

Contents

*ChatGPT has been used to point out grammatical errors and misspelling.

List of Symbols and Abbreviations

k_H	Henry's constant
K_m	Michaelis constant
m/z	Mass-to-charge ratio
Acetyl-CoA	Acetyl-coenzyme A
AdhE	Alcohol dehydrogenase enzyme
ArcA	Response regulator
ArcAB	Two-component anoxic redox control system
ArcB	Membrane-associated sensor kinase
ATP	Adenosine triphosphate
cAMP	Cyclic adenosine monophosphate
cAMP-CRP	cAMP in complex with CRP
CCR	Carbon catabolite repression
CRP	Cyclic adenosine monophosphate receptor protein
CyuA	L-cysteine desulfidase enzyme
DM	Diabetes mellitus
DSMZ	The Leibniz Institute DSMZ - German Collection of Microorganisms and Cell Cultures GmbH
EI	Electron ionization
Fe-S cluster	Iron-sulfur cluster
FNR	Fumarate and nitrate reductase regulation protein
GC-MS	Gas chromatography-mass spectrometry

Glycolysis	Embden-Myerhof-Parnas pathway
H_3O^+	Hydronium ion
HPS	Heat shock protein
NA	Nutrient agar
NB	Nutrient broth
PDHC	Pyruvate dehydrogenase enzyme complex
PflB	Pyruvate formate lyase enzyme
ppb	Parts per billion
PTR-MS	Proton-transfer-reaction mass spectrometry
ROS	Reactive oxygen species
RpoH	Regulator of stress response in heat shock
RpoS	Regulator of stress response in starvation
sccm	Standard cubic centimeters per minute
TCA cycle	Tricarboxylic acid cycle
TdcG	L-serine deaminase III enzyme
VOC	Volatile organic compound

1. Introduction

1.1 Background and Rationale

This thesis is based on observations from my Bachelor's thesis, where dietary acetaldehyde exposure was compared to alcohol-derived exposure.[1] The IARC (International Agency for Research on Cancer) has classified acetaldehyde associated with alcohol consumption as a Group 1 carcinogen.[2–4]

The research question in my Bachelor's thesis was whether acetaldehyde containing food could raise salivary acetaldehyde to biologically significant levels. The test subjects were exposed to dietary acetaldehyde, and saliva and breath samples were collected. A diabetic test subject showed significantly different saliva acetaldehyde levels compared to the others, raising the question of why this happened. To study this, a glucose tolerance test was conducted for the diabetic test subject, which led to a serendipitous finding. The correlation found between the blood glucose of the diabetic test subject and acetaldehyde in their exhaled breath is presented in Figure 1.1.[1]

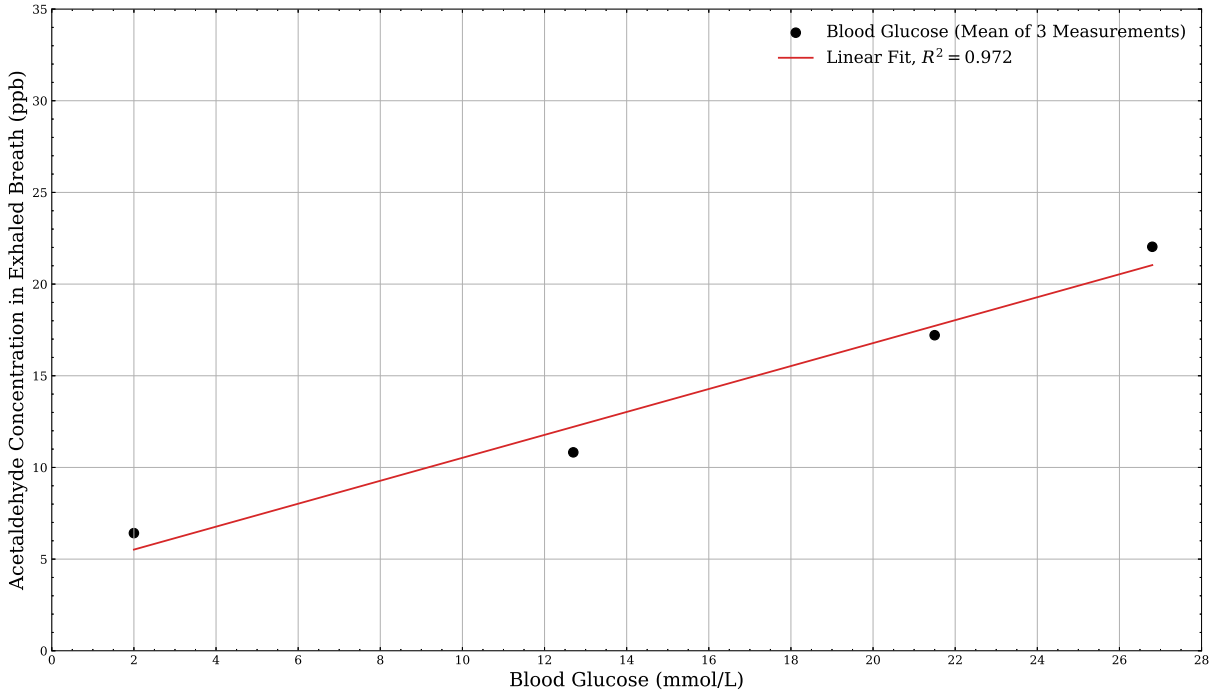


Figure 1.1: Correlation between the blood glucose of the test subject and acetaldehyde in their exhaled breath.[1]

Based on the result shown in Figure, 1.1 and the known link between high glucose levels and poor glycemic control in diabetes mellitus (DM), the hypothesis of this thesis was formed. The hypothesis is that poor glycemic control may increase the bacterial production of acetaldehyde in the intestine, leading to its local accumulation. This accumulation could then increase the risk of colorectal cancer in people with DM. Glycemic control refers to the maintenance of blood glucose levels within a desirable range,[5] with fasting levels of 4.0-6.0 mmol/l and 2-4 hours after a meal < 7.8 mmol/l.[6] Diabetes mellitus is a group of metabolic diseases characterized by long-term elevated blood sugar levels resulting from defects in insulin secretion, insulin action, or both. Several types are known, but the main types are type 1, type 2, and gestational DM.[6]

To investigate this, a brief literature review was conducted. No direct answers were found, but few studies explored the association between type 2 DM and colorectal cancer.[7–10] These studies provided indirect support for the hypothesis and highlighted the need for further experimental validation.

1.2 Significance and Aims

Diabetes mellitus is one of the fastest-growing global health challenges, with over 500 million diagnosed cases worldwide. The IDF (International Diabetes Federation) predicts that without sufficient action to address the situation, the number will rise to 800 million by 2045.[11] The highest estimates state that 800 million has already been reached.[12] In Finland approximately 500 thousand people have DM. Around 400 thousand have type 2, and 50 thousand have type 1.[13]

People living with DM have a higher risk of developing several serious complications such as cardiovascular diseases, nerve damage, kidney damage, lower-limb amputation, and eye disease.[11,13] These complications are caused by chronic high blood glucose levels resulting from insulin deficit or resistance, but with appropriate management of DM these complications can be delayed or prevented.[11]

In addition to the previously mentioned colorectal cancer, DM has also been linked to an increased risk of other cancers such as liver, pancreatic, endometrial, breast, and bladder cancers. The risk is higher for individuals with type 2 DM.[14,15] The mechanisms of this association are not yet fully understood. Studies suggest that hyperinsulinemia, insulin resistance, chronic inflammation, oxidative stress, and hyperglycemia can contribute to tumor progression.[14] The role of insulin, IGF-1 (insulin-like growth factor 1), and inflammation in cancer development has been extensively studied.[14,15] The hyperglycemia hypothesis has received less attention, although hyperglycemia is one of the most widely recognized metabolic changes in DM.[14] It is defined as an excess of circulating glucose.[6,11,14]

The hypothesis of this thesis is based on a concept similar to hyperglycemia. To study the hypothesis, a study was designed to validate previous findings on the ability of bacteria to produce acetaldehyde from glucose and ethanol via liquid-based headspace measurements. A new perspective was also introduced: whether there is something in the stool of individuals with DM that increases bacterial acetaldehyde production. The aim of the study was also to lay the groundwork for further research. Therefore, *Escherichia coli* (*E. coli*) was selected as a model organism. It was cultured on nutrient agar (NA) and in nutrient broth (NB) under aerobic conditions to maintain procedural simplicity. Glucose and ethanol solutions were prepared, and the fecal extracts were obtained from the FinnDiane Study Group. Kajsa Roslund's academic dissertation *Investigation of Volatile Compounds Produced by Pathogenic Oral Bacteria* laid the groundwork for the study design.[16]

In the theoretical part of the thesis, the volatile organic compounds (VOCs) produced in *E. coli*, the role of acetaldehyde in health and disease, the metabolism of *E. coli*, and the factors that affect it are presented. These topics are meant to provide a broad theoretical background that justifies the experimental part of this thesis. In addition, proton-transfer-reaction mass spectrometry (PTR-MS) and gas chromatography-mass spectrometry (GC-MS) are introduced as analytical techniques for bacterial VOC analysis. These techniques were used in the experimental part of the thesis.

The theoretical part is followed by the experimental part, which consists of the following sections: Materials and Methods, Results, and Conclusions. The Materials and Methods section describes cultivation methods, analytical instruments used, study design, and data processing. The Conclusions section summarizes the results, discusses the limitations of the study, and provides suggestions for future research.

2. Theoretical Background

2.1 Volatile Organic Compounds (VOCs) in Bacterial Metabolism

Bacterial metabolism refers to all the biochemical reactions that occur in a bacterial cell and are essential for cells to live, function, and replicate.[17] These reactions include energy-yielding and energy-requiring reactions.[17,18] In general, the basic metabolic mechanisms are similar in all living cells.[17]

VOCs are characterized as compounds with low molecular mass, high vapor pressure, and low boiling point.[18] In the context of bacterial metabolism, VOCs are byproducts or intermediates of various metabolic pathways.[18–20] Bacteria produce a wide range of VOCs such as terpenes, amines, ketones, aldehydes, and alcohols. The amount and composition of VOC can vary according to species, growth stage, and environmental factors.[18] In the following section, the VOC profile of *E. coli* is examined based on the results from individual studies. In addition, these studies are compared to the VOCs listed in the mVOC 4.0 database. The mVOC 4.0 database gathers over 2000 VOCs from around 1000 microbial species, offering a comprehensive and informative platform.[21] The compound names have been presented as they appear in the original source.

2.1.1 VOCs Produced in *Escherichia coli*

E. coli has been widely studied in relation to VOCs. Already in 1977, Hayward et al. proposed a test for detection of *E. coli*. In that study, *E. coli* was detected based on its production of ethanol from lactose. Additionally, they studied production of ethanol from arabinose, but the production was lower.[22]

A year later, Coloe et al. proposed an improvement for the test Hayward et al. developed. They explained that ethanol production from lactose is not specific for *E. coli* and use of arabinose would be better in the test. The relationship between ethanol production and culture aeration was also studied. In static (microaerobic) and shaken (aerobic) cultures, the detected ethanol was proportional to the cell count. From shaken cultures they were able to detect *E. coli* based on the ethanol concentration after four hours and from static culture after six hours. The

difference was due to the increase in the growth rate in the shaken cultures.[23]

These early studies laid the foundation for the VOC profiling. They also show that factors that need attention, such as aeration, have remained relevant for almost fifty years. In more recent research, the VOC profile of *E. coli* has been gradually characterized. For example, in 2006 Randal et al. identified *E. coli* VOCs that included acetaldehyde, ethanol, acetone, hydrogen sulfide, methanethiol, and dimethyl sulfide.[24] In the same year, Scotter et al. reported a profile that included acetaldehyde, ethanol, pentanol, acetone, hydrogen sulfide, methanethiol, indole, 2-aminoacetophenone, and propene.[25]

In 2008, Bunge et al. tentatively identified VOCs that included methanol, acetaldehyde, ethanol, methanethiol, acetone, acetic acid and indole. Additionally, their measurements revealed five specific signals for *E. coli* at 39, 53, 91, 93, and 132 amu (all unidentified).[20] Three years later, Thorn et al. reported a broader profile that included 1-butanol, 1-pentanol, 2-aminoacetophenone, acetoin, acetone, butanoic acid, dimethyl disulfide, ethanol, ethyl acetate, ethyl butanoate, formaldehyde, hydrogen sulfide, indole, isoprene, methyl mercaptan, phenylacetic acid, pyrrole and trimethylamine. In the context of this thesis, the results of that study are interesting. They detected a background signal of acetaldehyde over 100 ppb (parts per billion, $1/10^9$) from the control medium without bacteria. They discussed that this background signal may have masked acetaldehyde signal in the test samples, and therefore acetaldehyde was not detected.[19]

In 2013, Bos et al. concluded a systematic review on volatile metabolites of pathogens. Thirty-one articles were included in their review, of which seventeen articles included *E. coli*. Through their review, they found that *E. coli* can be identified with methanol, pentanol, ethyl acetate, and indole. They also showed that there is a strong evidence that *E. coli* VOCs include butanol, ethanol, propanol, propionic acid, acetaldehyde, formaldehyde, acetone, ethyl butanoate, dimethyl disulfide, hydrogen sulfide, methyl mercaptan, 2-aminoacetophenone, and trimethylamine.[26]

The study conducted in 2017 by Ratiu et al. produced a VOC profile that differs from the profiles reported by Thorn et al. and Bos et al. Ratiu et al. investigated the effect of growth medium on *E. coli* metabolism. In that study, *E. coli* was cultivated in three different growth mediums: trypcase soy broth, Mueller Hinton, and minimal salts enriched with glucose. They identified 52 VOCs,[27] of which only indole, dimethyl disulfide, and propanol were also reported by Thorn et al. and Bos et al.

In 2024, Zheng et al. produced a VOC profile of 36 compounds, of which 26 were positively identified. These included ethylene glycol, 1-butanol, 3-methyl-1-butanol, acetaldehyde, propanal, nonanal, benzaldehyde, 2-methyl-2-butenal, acetic acid, propionic acid, ethyl acrylate, bornyl acetate, methyl 2-methylbutanoate,

acetone, butan-2-one, cyclohexanone, 2,3-pentanedione, acetoin, 2-methyl pyrazine, 2-ethyl-5-methylpyrazine, 2,5-dimethylpyrazine, and pyrrolidine.[28]

In the mVOC 4.0 database, the following VOCs are associated with *E. coli*: benzaldehyde, dibenzofuran, 2-phenylacetaldehyde, 2-phenylacetic acid, 3-methylphenol, phenol, phenylmethanol, methanethiol, sulfane, acetic acid, 3-hydroxybutan-2-one, butanoic acid, naphthalene, 1-butanol, ethanol, octanol, and acetaldehyde. The database also provides a signature VOC profile, which consists of VOCs that are characteristic for *E.coli*. This profile includes 1-propanol, 2-heptanone, 2-nonanone, 2-pentanone, 2-undecanone, 2,3-butadione, 3-methylbutanal, butyric acid, cis-7-tetradecen-1-ol, dimethyl disulfide, ethyl acetate, ethyl butanoate, ethyl hexanoate, indole, isopentanol, methoxy-phenyl-oxime, N,N-dipropyl-1-propanamine, p-xylene, pentyl cyclopropane, and propanoic acid.[21]

The database provides the most comprehensive overview of the VOC profile. Most importantly in the context of this thesis, the database includes acetaldehyde. Acetaldehyde was not consistently detected in all of the studies discussed. This inconsistency likely stems from differences in analytical techniques, cultivation and sampling methods, or genetics between *E. coli* strains. These differences fall outside the scope of this thesis and are therefore not discussed in greater detail.

2.1.2 The Role of Acetaldehyde in Health and Disease

In this section, the focus is on acetaldehyde and its effects on health and disease are discussed. In addition, the causes of these effects are explained. In the following Tables 2.1 and 2.2, the most important terms that may be unfamiliar to chemists are presented with explanations to aid the reader.

Table 2.1: Glossary of key terms related to biochemical properties.

Term	**Explanation**
DNA repair process	A process that identifies and corrects DNA damage.
DNA replication process	The mechanism by which a cell copies its DNA.
Electrophilic carcinogen	A carcinogen that forms adducts with cellular macromolecules.[29]
Genotoxicity	The ability to induce DNA damage, mutations, or both.[29]
Oxidative stress	A condition in which there is an imbalance between reactive oxygen species and antioxidants.[30]

Table 2.2: Glossary of key terms related to acetaldehyde-induced molecular and physiological effects.

Term	Explanation
Arrhythmia	An irregular heartbeat.[31]
Base-pair mutation	A change in a single base pair of DNA.
Cardiomyopathy	A disease of a heart muscle.[32]
Deletion	A part of a chromosome is lost or removed.
DNA adduct	An addition product of a molecule binding to DNA.[29]
DNA cross-links	Interstrand or intrastrand bonds that can cause alterations in DNA replication and repair processes.
DNA-polymerase	A key enzyme in DNA replication.
DNA single-strand break	A break in one of the two strands.
Formation of micronuclei	A genetic abnormality.[33]
Frameshift mutation	A change in a reading frame of a gene.
Increased cytokine production	The release of a signalling molecules that can cause inflammantion.[30]
Mitochondrial damage	Damage that leads to reduced cellular energy production and can cause cell death.[30]
Neurotoxicity	The ability to cause harmful effects to the nervous system.[34]
Point mutation	A change in a single nucleotide base.
Rearrangement	A change in a structure of a chromosome.
Replication fork	A structure that forms when the double helix opens.
Vasodilation	The widening of blood vessels.[31]

As mentioned in the Introduction, acetaldehyde associated with alcohol consumption is a carcinogen in humans.[3,4] Multiple sources have summarized studies that found dose-response relationships between alcohol consumption and cancer of the oral cavity, pharynx, larynx, esophagus, stomach, and colorectum.[31,35–38] This increased cancer risk can be explained by the local accumulation of acetaldehyde in these tissues due to the ethanol metabolism of the normal microbiota.[35,36,38,39] Genetic factors, poor oral hygiene, and smoking can further increase cancer risk by enhancing exposure to acetaldehyde. Genetic factors affect acetaldehyde detoxification and can therefore lead to increased local accumulation. Poor oral hygiene increases the biomass of acetaldehyde-producing bacteria, and smoking introduces additional acetaldehyde.[36,37,39]

The carcinogenicity of acetaldehyde can be explained by its biochemical properties. Acetaldehyde is electrophilic and genotoxic, alters DNA replication and repair processes, and induces oxidative stress.[39] These are properties that human carcinogens commonly show.[29]

As an electrophilic compound, acetaldehyde can form adducts with DNA and proteins. DNA adducts are a form of DNA damage.[29] Through DNA adducts, acetaldehyde induces DNA interstrand and intrastrand cross-links, base-pair mutations, deletions, rearrangements, and frameshift mutations.[33] Furthermore, acetaldehyde alters DNA replication and repair processes through these adducts.[29,33] This alteration occurs, for example, by interfering with DNA polymerase function or inducing replication fork collapse.[33] These disruptions lead to genomic instability, which promotes carcinogenesis.[29] In addition to its effects via adducts, acetaldehyde also directly alters DNA replication and induces DNA single-strand breaks, point mutations, and formation of micronuclei.[33] These effects result from the genotoxicity of acetaldehyde.[29] The discussed mechanisms are summarised in Figure 2.1.

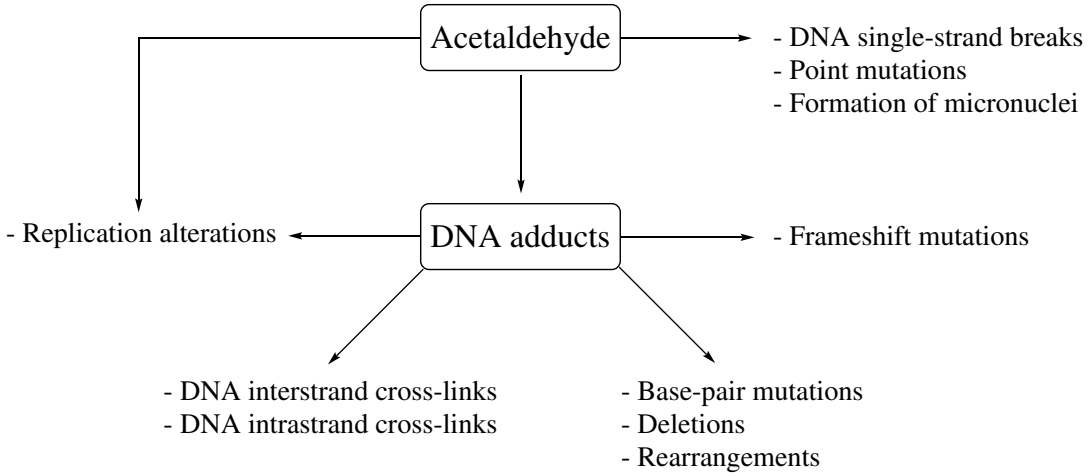


Figure 2.1: Graphical summary of the text, illustrating both direct and DNA adduct-mediated mechanisms.[33]

Acetaldehyde induces oxidative stress by promoting overproduction of reactive oxygen species (ROS) through direct and indirect mechanisms. The direct effect results from the increase of the expression and activity of the NADP oxidase-2 (nicotinamide adenine dinucleotide phosphate oxidase-2) enzyme complex, which contributes to mitochondrial ROS production.[30] An indirect effect occurs through alcohol metabolism. The first phase of alcohol metabolism is the oxidation of ethanol to acetaldehyde. One of the enzymes that oxidizes ethanol to acetaldehyde is cytochrome P450 enzyme CYP2E1.[30,31,40] The enzyme produces ROS, for example, hydrogen peroxide, during its catalytic cycle. This oxidative stress induced by acetaldehyde promotes lipid peroxidation, DNA adduct formation, DNA cross-links, oxidative DNA damage, and DNA strand breaks. In addition, oxidative stress can trigger protein inactivation, increased cytokine production, and mitochondrial damage.[30] These disruptions lead to genomic instability, which promotes carcinogenesis as mentioned above.[29]

Acetaldehyde also has several non-carcinogenic effects, in addition to its carcinogenity. It plays a role in the development of alcoholic cardiomyopathy.[30–32] In particular, acetaldehyde induces cardiac hypertrophy and dilated cardiomyopathy.[31,32] Acetaldehyde disrupts myocardial con-

traction by reducing calcium entry into the cell, altering calcium release from the sarcoplasmic reticulum, and decreasing the expression of key calcium-regulating proteins. Additionally, acetaldehyde stimulates the release of signaling molecules such as epinephrine, norepinephrine, histamine, and bradykinin. These signaling molecules contribute to cardiovascular symptoms such as vasodilation, arrhythmias, and fluctuations in blood pressure.[31]

In addition to cardiovascular effects, acetaldehyde plays a role in neurotoxicity.[31,41,42] It contributes to behavioral and physiological effects commonly associated with alcohol ingestion, including uncoordination, memory impairment, and sleepiness.[31] Several hypotheses have been proposed about the biochemical processes that explain these effects, but none of them have been supported by strong experimental evidence.[41,42]

2.2 Metabolism of *Escherichia coli*

E. coli is facultative anaerobe,[43–45] meaning it can survive and grow in the presence and absence of oxygen.[45] Therefore, its central metabolism is divided into respiration, microaerobic respiration and fermentation.[44]

Respiration refers to the biological oxidation of organic compounds to yield adenosine triphosphate (ATP). In aerobic respiration, oxygen is used as an electron acceptor.[17] *E. coli* can perform respiration under aerobic and microaerobic conditions.[44,45] In microaerobic respiration, nitrate, sulfate, carbon dioxide or fumarate can be used as an electron acceptor.[17,44] In fermentation, ATP is generated through the enzymatic breakdown of organic compounds. The organic compounds are used as electron donors, and the metabolic end products as electron acceptors.[17]

In the following sections, specific pathways leading to the formation of acetaldehyde and the key enzymes involved are introduced. First, the central metabolism of E. coli with glucose is introduced. Secondly, amino acids that *E. coli* can ferment are discussed. Glucose is used as an example of sugars, as it is one of the most common substrates used to study metabolism.

2.2.1 Pathways Leading to Acetaldehyde Formation

Under aerobic conditions, *E. coli* oxidizes glucose completely by the following Equation 2.1:

$$\mathrm{C_6H_{12}O_6} + 6\mathrm{O_2} \rightarrow 6\mathrm{CO_2} + 6\mathrm{H_2O}. \tag{2.1}$$

The reaction is exothermic, with a reaction enthalpy of approximately -2880 kJ per mol of glucose.[17] However, the maximum efficiency is between 30–40%, which leads to a yield of

800–1000 kJ per mole of glucose. The remaining energy cannot be conserved biologically and is dissipated due to incomplete coupling between glucose oxidation and ATP synthesis.[46]

The complete oxidation includes three biochemical pathways: Embden-Myerhof-Parnas (glycolysis), tricarboxylic acid cycle (TCA cycle), and oxidative phosphorylation.[17] Glycolycis converts glucose to two molelecules of pyruvate.[17,43] Pyruvate is then converted into acetyl-coenzyme A (acetyl-CoA) by the pyruvate dehydrogenase complex (PDHC). Acetyl-CoA is further processed in the TCA cycle.[17,44] In microaerobic and anaerobic conditions, this pathway is downregulated and PDHC is inactive. In microaerobic respiration, acetate is the main product. Acetate is converted from acetyl-CoA. Additionally, formate is produced. The formation of acetyl-CoA and formate from pyruvate is catalyzed by pyruvate formate lyase enzyme (PflB).[43,44]

Acetaldehyde is not formed in aerobic or microaerobic respiration, but it is an intermediate in fermentative pathways.[43] In anaerobic conditions, *E. coli* shifts its metabolism to mixed-acid fermentation to maintain redox balance.[43,47] Mixed-acid fermentation results in production of succinate, formate, acetate, lactate, and ethanol.[43,44,47] Acetaldehyde is an intermediate in the pathway, whose end product is ethanol.[43] Ethanol is formed in a two-step reaction catalyzed by alcohol dehydrogenase (AdhE). In the first step, acetyl-CoA is converted to acetaldehyde, and in the second step acetaldehyde is reduced to ethanol.[47] Figure 2.2 illustrates the central metabolism of *E. coli*.

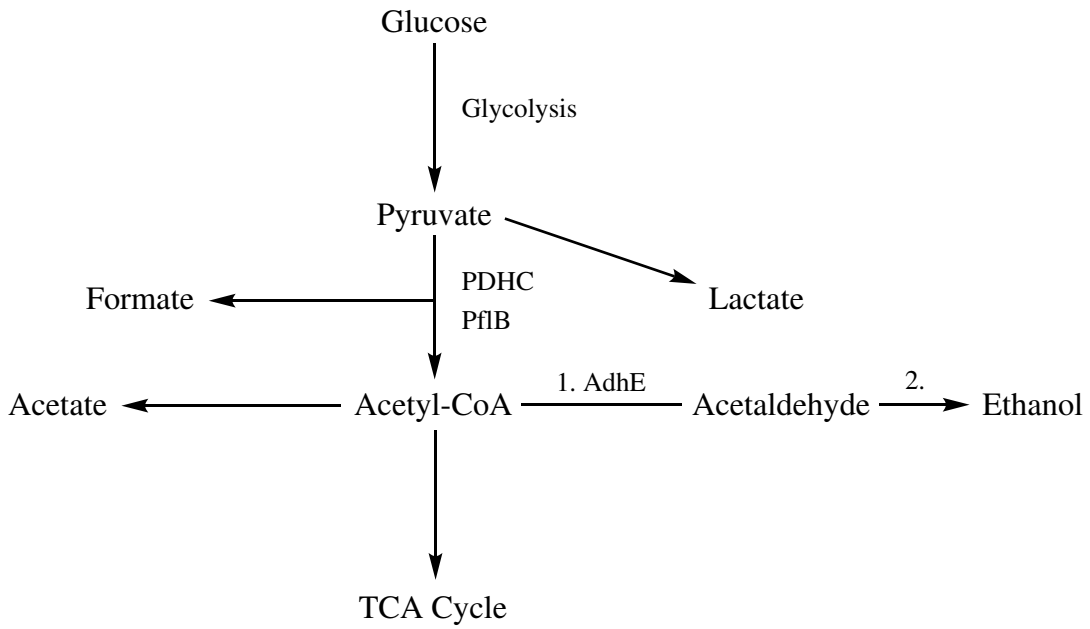


Figure 2.2: The central metabolism of *E. coli*. Only enzymes mentioned in the text are illustrated. Respiratory and fermentative pathways are not distinguished. Succinate is a TCA intermediate, and its production starts from glycolytic intermediate. It is therefore not shown.

Additionally, acetaldehyde can be formed through the fermentation of amino acids L-cysteine and L-serine. These amino acids are fermented because they can be converted directly to pyruvate in a single enzymatic step.[48] The enzyme that converts cysteine is L-cysteine desulfidase (CyuA) and the enzyme that converts L-serine is L-serine deaminase III (TdcG).[47,48] The resulting pyruvate is further metabolized through mixed-acid fermentation,[48] where acetaldehyde is an intermediate.

2.2.2 Key Enzymes

The key enzymes involved in the production of acetaldehyde are CyuA, TdcG, PflB, and AdhE.[47] CyuA operates together with L-cysteine importer protein to utilize L-cysteine. The enzyme catalyzes degradation of L-cysteine by delsulfidating it. The reaction produces pyruvate, ammonia, and hydrogen sulfide. The enzyme has an iron-sulfur cluster (Fe-S cluster), which is highly sensitive to molecular oxygen. Oxygen inactivates the cluster. D-cysteine and analogs of L-cysteine are not substrates for CyuA as it is highly specific.[48] The Michaelis constant (K_m) is 200 µM. A high K_m indicates that the enzyme has a low affinity for L-cysteine and therefore does not use up the cysteine pool in the cell (55 µM).[47]

Similarly to CyuA, TdcG operates with an importer protein.[48] The enzyme catalyzes the degradation of L-serine by deaminating it. The reaction produces pyruvate and ammonia.[47] TdcG also contains an oxygen sensitive Fe-S cluster.[49] The K_m is 4800 µM for TdcG and is substantially higher than K_m of CyuA.[47] Therefore, the kinetics of the enzymes differ despite a mechanistically similar role.

PflB is a central enzyme in the metabolism of *E. coli*.[43] The enzyme catalyzes cleavage of pyruvate into acetyl-CoA and formate. The reaction is non-oxidative and CoA-dependent.[47,50] Mechanistically, the reaction is unique because PflB utilizes a radical-based mechanism. An active site Glycine734 radical initiates the cleavage of carbon-carbon bond of pyruvate.[51] The K_m of PflB is 2370 µM.[47]

AdhE is multifunctional enzyme with two catalytic functions. The aldehyde dehydrogenase domain is located at N-terminus, and the iron-dependent alcohol dehydrogenase is at C-terminus. This unusual quaternary structure allows AdhE to catalyze the sequential reduction of acetyl-CoA to acetaldehyde and then to ethanol under anaerobic conditions.[47] Notably, AdhE is also expressed under aerobic conditions. Under aerobic conditions, the enzyme catalyzes ethanol degradation, oxidizing ethanol sequentially to acetyl-CoA. However, this pathway is not sufficient to support growth in wild type *E. coli*.[47]

2.3 Key Factors Influencing Metabolism of *Escherichia coli*

Metabolism of *Escherichia coli* is not static. Environmental factors, such as oxygen availability and nutrient availability, affect metabolic pathways. Additionally, cellular stress responses can alter gene expression and therefore impact metabolism.

The following sections discuss these three factors in detail. Firstly, the role of oxygen-responsive regulators is introduced, including fumarate and nitrate reductase regulation (FNR) protein and two-component anoxic redox control (ArcAB) system. In the second section, carbon catabolite repression (CCR) and the role of cyclic adenosine monophosphate (cAMP) in complex with cyclic adenosine monophosphate receptor protein (CRP) are introduced. Finally, starvation response and the heat shock response of *Escherichia coli* are discussed.

2.3.1 Oxygen Availability

As a facultative anaerobe, *E. coli* must regulate its gene expression and enzymatic activity to adapt to its environment's oxygen availability. This regulation is achieved with transcriptional regulators FNR and ArcAB.[43] FNR mediates the transition from aerobic to anaerobic growth. This protein mediates the transition by activating genes involved in anaerobic metabolism and repressing genes involved in aerobic metabolism. Generally, FNR changes from the inactive state to the active state in the absence of oxygen.[47,52,53] This change from the inactive state to the active state is controlled via a oxygen-sensing domain in its N-terminus.[52] The domain contains a $[4Fe4S]^{2+}$ cluster under anaerobic conditions, and the FNR is active. In the active state, it is able to bind to specific DNA sites and activate or repress the target genes. Oxidation of this cluster into $[2Fe2S]^{2+}$ leads to the inactivation of FNR.[47,52,53] In the inactive state, DNA binding is reduced, and therefore it cannot control transcription.[52]

The two-component ArcAB system complements oxygen sensing as being a sensor for oxygen consumption.[54] The primary role of ArcAB is in anaerobic repression of genes involved in aerobic metabolism.[47,54] Generally, the ArcAB system modulates the switch to anaerobic respiration or fermentation.[54] The system is composed of membrane-associated sensor kinase ArcB and the response regulator ArcA.[43,47,54] As the need for oxygen consumption decreases, the activity of the electron transport chain decreases. Decreased activation leads to ArcB activation via interaction with ubiquinone coenzyme. This quinone-induced regulation is an area of active research. Active ArcB interacts through phosphorylation with ArcA and then active ArcA suppresses aerobic metabolic pathways.[54]

These systems control over 80% of metabolic flux when oxygen is limited.[54] Of the enzymes and metabolic pathways discussed, FNR plays a role in activating the expression of PflB, AdhE, and TdcG. Although CyuA shares functional similarity with TdcG, FNR does not

regulate its expression. ArcAB only plays a role in activating the expression of PflB among the enzymes discussed.[47] In summary, these systems down-regulate the TCA cycle and promote fermentation.[54]

2.3.2 Nutrient Availability

Understanding how bacteria utilize available nutrients is essential in bacterial studies.[55] *E. coli* prefers glucose as a primary carbon source.[55,56] The availability of glucose represses the utilization of other sugars. CCR is the global regulatory mechanism that mediates this repression. When glucose is not available, CCR does not repress the expression and activity of genes required for the utilization of secondary carbon sources.[55] This leads to a sequential use of sugars.[55–57] First, the sugar yielding highest growth rate is consumed, followed by the next highest, and so on.[57]

Regulation of this system occurs through a multiprotein phosphorylation cascade that couples glucose uptake and metabolism.[57] A central protein of this cascade is $EIIA^{Glc}$.[56] Phosporylation of $EIIA^{Glc}$ activates adenylate cyclase, which leads to increase of cAMP concentration in the cell. As cAMP concentration increases, cAMP binds to its receptor protein CRP. The cAMP-CRP complex then activates the expression of genes involved in metabolism of non-glucose sugars.[56,57] Generally, these cAMP-CRP-regulated genes are expressed when glucose is not available. The differential activation of promoters by this complex explains the sugar utilization hierarchy.[57]

CCR in *E. coli* and utilization of sugars are well characterized. Already in 1942, Monod showed the glucose-lactose diauxic shift in *E. coli*.[56] More recent studies have investigated the hierarchy of non-glucose sugars. In 2014, Aidelberg et al. studied this hierarchy. They reported the following hierarchy of six sugars: lactose > arabinose > xylose > sorbitol > rhamnose > ribose.[55] A similar hierarchy was reported by Ammar et al. in 2018. In that study, they showed that lactose represses arabinose and xylose, and arabinose represses xylose.[57]

Alternative carbon sources for *E. coli* include amino acids.[58] *E. coli* posesses transporters for all amino acids and therefore can utilize them as carbon sources.[48] Some studies suggest that hierarchichal utilization also occur with amino acids.[59] In 2007, Sezonov et al. studied physiology of *E. coli* in Luria-Bertani broth, which is a widely used rich medium. They concluded that growth in Luria-Bertani broth is based on the utilization of amino acids. Based on a previous knowledge, they proposed the following hierarchy of amino acids: L-serine > L-aspartate > L-tryptophan > L-glutamate > glycine > L-threonine > L-alanine. In addition, they suggested that L-proline, L-arginine, L-glutamine, L-asparagine, L-cysteine, and L-lysine may also be utilized and therefore be in the hierarchy.[58]

Additionally, carbon sources such as ethanol can be used. However, as discussed in Section

2.2.2, ethanol oxidation alone is not sufficient to support growth.[47] When *E. coli* grows on mixture of carbon sources, it exhibits simultaneous as well as hierarchical utilization patterns. This behavior is poorly understood, and therefore under active research.[59]

In 2019, Zampieri et al. developed a constraint-based modeling approach to study the metabolic adaptation of *E. coli* in complex medium. They used a glucose minimal medium supplemented with an undefined mixture of amino acids and oligopeptides. They discovered that during the fastest growth phase *E. coli* preferentially utilized amino acids over glucose. This preference is mediated by accumulation of amino acid degradation metabolites, pyruvate and oxaloacetate, which directly inhibit the uptake of glucose.[60]

The discussed nutrients enter the central metabolism in different points. Degradation products of sugars enter the upper or middle part of the glycolycis pathway, while degradation products of amino acids enter the lower part.[59] In context of this thesis, the exact entry point does not matter because the flux is towards metabolites past pyruvate. In summary, nutrient utilization in *E. coli* is dynamic.[59,60] The selection of nutrients depends on various factors, and regulatory mechanisms in complex media are still not fully understood.[59]

2.3.3 Stress Responses

When key nutrients become limited, bacteria experience starvation. Starvation leads to a stress response that causes a metabolic shift from generating energy for growth to conserving energy.[54] The global regulator of this stress response is a sigma factor RpoS. Generally, RpoS influences the regulation of genes that promote survival.[54,61]

In starvation, RpoS regulates central metabolic pathways,[54,61] but its activity is also influenced by these pathways. The two-component ArcAB system plays a role in this two-way relationship through regulatory RNA ArcZ. ArcZ is expressed when ArcA is inactive, which leads to an increase in RpoS translation. When ArcA is active, ArcA represses the expression of AcrZ, resulting in reduced RpoS translation.[54]

The first phase of starvation protocol is nutrient scavenging. This triggers an increase in the production of proteins that can "find" the particular limited nutrient. For the scavenging of alternative carbon sources, the cAMP-CRP complex is used. The second response of starvation protocol is entry into stationary phase. Bacterial cells enter the stationary phase if scavenging fails.[61] The ArcAB system also has an important role during stationary phase, as it represses aerobic pathways. This repression helps cells converse energy and protects them from potentially harmful accumulation of metabolic by-products.[54]

When *E. coli* encounter temperature shift, heat shock induces synthesis of more than 20 heat shock proteins (HSPs). Induction is followed by an adaptation period in which synthesis

decreases to a steady state.[62] HSPs are chaperons that prevent aggregation and refolding of proteins.[62,63]

The regulator of heat shock response is sigma factor RpoH. HSPs are synthetized as a consequence of a rapid increase in RpoH levels and stimulation of its activity.[62,63] The decrease in RpoH levels and the inhibition of its activity repress the response.[62] Inhibition is carried out by HSPs through a negative feedback loop.[63]

The effect of heat shock response on metabolism is complex. For example, glycolysis and TCA cycle are downregulated.[63] The ArcAB system plays a role in the downregulation of TCA cycle. Under fermentativie conditions, the activity of key enzymes is altered, resulting in increased formate and lactate production and decreased ethanol production. Additional effects occur, but those fall outside the scope of this thesis.[64] Figure 2.3 summarizes Sections 2.2 and 2.3 by showing the pathways leading to acetaldehyde formation, the key enzymes involved, and the role of FNR and ArcAB.

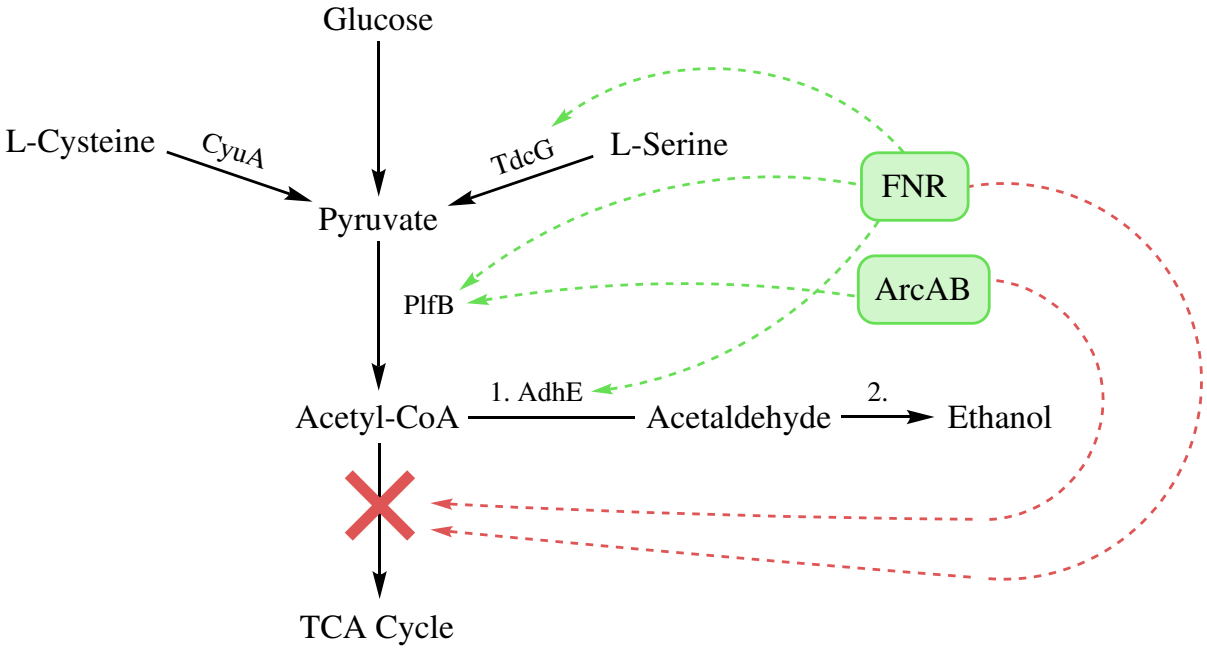


Figure 2.3: Graphical summary of the text. Activation of expression is represented by a green dashed arrow, and downregulation by a red dashed arrow. Oxygen availability and stress responses are not distinguished in downregulation of the TCA cycle.

3. Mass Spectrometry Techniques for VOC Analysis

Mass spectrometry plays a key role in VOC analysis, offering a wide mass range and high sensitivity, and ability to perform an instant quantitative analysis. The applications include breath analysis, food chemistry, metabolomics, forensics, and environmental sciences.[65] The following sections briefly introduce the principles and applications of PTR-MS and GC-MS in bacterial VOC studies. Furthermore, Henry's law is discussed, as it explains the chemistry underlying the liquid-based headspace measurements.

3.1 Proton-Transfer-Reaction Mass Spectrometry (PTR-MS)

PTR-MS is an analytical technique for real-time analysis of VOCs.[65,66] It offers simultaneous monitoring and quantification of compounds from diverse chemical groups with high sensitivity and in a non-invasive manner.[65] PTR-MS is non-invasive because it does not require pretreatment of the sample prior to ionization and detection. Therefore, the gaseous sample can be directed straight into the reaction chamber, called the drift-tube.[16,66]

In the drift-tube, compounds are ionized through a proton-transfer-reaction with a precursor ion.[66] Typically, the hydronium ion (H_3O^+) is used as the precursor ion.[65,66] The precursor ions are introduced into the drift-tube from the ion source, where they are produced.[65] Equation 3.1 describes the proton-transfer-reaction, where the hydronium ion is the precursor ion and R is the volatile compound.[16]

$$\mathrm{H_3O^+ + R \rightarrow RH^+ + H_2O} \tag{3.1}$$

For the reaction to occur as shown in Equation 3.1, the compound must have a higher proton affinity than water. The proton affinity of water is 166.5 kcal/mol.[65,66] Most organic compounds have a higher proton affinity and are therefore ionized.[16] Compounds that do not react with H_3O^+ include components of air and small hydrocarbons.

After the ionization, the resulting ions are separated based on their mass-to-charge ratio (*m/z*) into a mass analyzer,[66] and the data are displayed as a mass spectrum.[65] The ion signals in the spectrum are proportional to the number densities of product ions $[RH^+]$ relative to the precursor ions $[H_3O^+]$. This relies on the approximation that the number density of product ions $[RH^+]$ is much lower than the number of available primary ions. Therefore, the concentration of a VOC in the drift tube can be calculated by Equation 3.2:

$$[R] = \frac{[RH^+]}{[H_3O^+]} \frac{1}{kt}, \tag{3.2}$$

where $[R]$ is the number density of a VOC, k is the reaction rate constant, and t is the drift time. The reaction rate constant is approximated to be $2 \times 10^{-9}\mathrm{cm}^3/\mathrm{s}$, while the drift time is calculated from the system parameters.

Due to its real-time analysis capability, PTR-MS has become widely used in bacterial VOC studies.[65] For example, Roslund et al. profiled volatile compounds produced *in vitro* by pathogenic oral bacteria. In that study, PTR-MS was used in *in vitro* headspace measurements. VOC profiles were measured from the headspace of agar and the bacterial species used were *Porphyromonas gingivalis*, *Prevotella intermedia*, *Prevotella nigrescens*, and *Tannerella forsythia*. Spectra of pure reference substances were also measured to support tentative identification of the compounds. The study concluded that while PTR-MS is suitable for tentative identification of the produced compounds, absolute identification requires complementary technique.[67]

O'Hara et al. investigated the effect of different growth media on the types and intensities of VOCs. In that study, PTR-MS was found to be a sufficiently sensitive technique for detecting bacterial VOCs in vitro. VOCs can be detected even after a relatively short time from the start of incubation. The authors also pointed out the potential of PTR-MS for in vivo microbial diagnosis, which could lead to a faster treatment of bloodstream infections. Unlike Roslund et al., O'Hara et al. did not discuss the methodological limitations of PTR-MS.[68]

Rajendran et al. discuss the methodolocal limitations similarly to Roslund et al. Rajedran et al. used PTR-MS to determine the VOCs produced by different lactic acid bacterial strains. They concluded that while PTR-MS allowed the discovery of differences in VOCs produced between strains and fermentation temperature, complementary techniques are needed to support the identification of compounds.[69] Acetaldehyde was detected in the study. However, its detection was not discussed in detail. In particular, interference with an overlapping carbon dioxide peak was not addressed.

These studies discussed demonstrate that PTR-MS enables real-time monitoring of bacterial

VOCs and strain differentiation.[67,68] However, complementary techniques are required for compound identification.[69] This is because PTR-MS primarily provides *m/z* information without direct structural information. In addition, fragmentation of certain compounds and overlapping *m/z* signals can complicate spectral interpretation.

3.2 Gas Chromatography-Mass Spectrometry (GC-MS)

GC-MS is a commonly used analytical technique in VOC studies and serves as a gold standard for laboratory-based analysis of VOCs. It allows unambiguous compound identification through chromatographic separation and mass spectral analysis. A typical VOC GC-MS measurement has the following steps:

1. Sample collection from the headspace.
2. Trapping of compounds on an adsorptive surface, in solvent extracts, or in a chromatographic syringe.
3. Direct injection or thermal desorption of the sample in the GC inlet for separation.
4. Identification of separated compounds by MS.[65]

The separation of compounds is based on their interaction with a stationary phase. The interaction include adsorption, ionic interactions, diffusion, and solubility. Compounds with weaker interactions elute faster, and compounds with stronger interactions remain in the column longer. The amount of time taken for a compound to pass through the column is referred as "retention time".[70] This separation process results in a chromatogram, where each peak corresponds to a compound detected at a specific retention time.[16] The area under the peak corresponds to concentration.

After elution, the compounds are ionized and fragmented. Electron ionization (EI) is typically used as the ion source. In EI, compound collides with a high-energy electron.[65] If the kinetic energy of the electron is higher than the ionization energy of the compound, the collision results in a release of an electron from the compound and ionization. Collision also causes fragmentation because of dissociation of the compound. Equation 3.3 illustrates the ionization process described above, where R represents the compound.[16]

$$\mathrm{R} + \mathrm{e}^- \rightarrow \mathrm{R}^+ + 2\mathrm{e}^- \tag{3.3}$$

After ionization and fragmentation, the fragments are separated according to their *m/z* into a mass analyzer.[66] This results in a mass spectrum.[65]

In bacterial VOC studies, GC-MS is commonly used as a primary analytical technique or as a complementary tool.[27,71–73] For example, Roslund et al. identified biomarkers for four pathogenic oral bacteria. In that study, GC-MS was used to confirm compounds that were initially tentatively identified using PTR-TOF-MS in the first article of the dissertation. In addition, the VOC profiles of oral bacteria were further investigated with GC-MS. VOC profiles were measured similarly compared to the first article and from same the bacterial species.[67] Based on the results, they concluded that GC-MS provides concrete identification of VOCs produced by bacteria.[71]

Tait et al. investigated VOCs produced by bacteria. In that study, GC-MS was used to analyze VOCs produced by bacteria with the aim of determining the impact of experimental parameters on the VOC profiles generated. The type of culture from which headspace volatiles were measured was one of the varied experimental parameters. The bacterial species used were *E. coli*, *Klebsiella pneumoniae*, and *Staphylococcus aureus*. They concluded that the identification of bacteria through GC-MS appears promising, but has limitations. For example, the methodologies used to generate and extract bacterial VOCs have an effect on the detected VOCs. The type of GC column also has an effect, but its lesser. Therefore, VOC profiles are not always directly comparable between studies where GC-MS is the only method.[73]

These studies discussed demonstrate typical applications of GC-MS in bacterial VOC analysis. Although GC-MS offers concrete compound identification,[71,72] VOC profiles obtained with GC-MS are often highly specific to the experimental design and are not directly comparable between studies.[73]

3.3 Headspace sampling and Henry's law

In 1803 William Henry experimentally found that under constant temperature the solubility of gases in water increases proportionally with the pressure applied. By applying additional atmospheres, Henry demonstrated that the volume of gas absorbed is directly related to the pressure. Henry also noted that at higher temperatures gases dissolve less in water.[74] These observations became known as Henry's law. Equation 3.4 describes Henry's law in terms of the concentration of the dissolved gas:

$$c = k_H p, \tag{3.4}$$

where c is the concentration of the dissolved gas, p is the partial pressure above the solvent, and k_H is Henry's constant in units of M/Pa.[75] Henry's constant is specific to the solute-solvent pair, and it is temperature dependent.[76]

Henry's law can also be expressed in terms of the mole fraction of the solute. Equation 3.5 describes Henry's law in terms of the mole fraction of the solute:

$$p = xK_H, \tag{3.5}$$

where x is the mole fraction of the solute and K_H is Henry's constant in units of pressure.

The form of Henry's law used depends on the application,[76] which explains why Henry's constant can appear in different units across sources. In Table 3.1, examples of Henry's constants for acetaldehyde are listed as reported in the original sources.

Table 3.1: Henry's constants for acetaldehyde at a temperature of 25 °C.

Value	**Unit**
0.17[77]	mol/(m^3 Pa)
13.1[78]	M/atm
4.14[79]	atm

In the context of headspace sampling, Henry's law explains the partitioning of a compound between the liquid phase and the gas phase. A higher Henry's constant corresponds to a higher partial pressure in the headspace. In contrast, compounds with a low Henrys constant tend to stay dissolved in the liquid phase. Therefore, the headspace concentration measured in VOC analyzes is strongly dependent on the Henry's constant of the compound.

It is important to distinguish Henry's constant from the boiling point of a compound. The boiling point is a property of a pure substance and describes the temperature at which a bulk phase transition from liquid to gas occurs. In contrast, Henry's constant is a property of a mixture and describes the equilibrium distribution of the compound between the phases. It takes into account interactions between compound and the solvent. Consequently, a compound can be classified as a VOC based on its boiling point, but it can still exhibit a low headspace concentration.

As the headspace is sampled, the removal of compound from the gas phase does not irreversibly deplete the system. The system re-establishes equilibrium if the concentration of the compound in the liquid phase is sufficiently high. This behavior underlies the reliability of liquid-based headspace measurements.

4. Materials and Methods

The theoretical part of this thesis described the VOCs produced in *E. coli*, the role of acetaldehyde as a carcinogen, the metabolism of *E. coli*, factors influencing its metabolism, and techniques for VOC analysis. This background forms the foundation for the experimental work of this thesis, which is presented in the following sections.

4.1 Bacterial Strain and Cultivation Methods

The bacterial strain used in this study was *E. coli* DSM 3925,[80] obtained from the Leibniz Institute DSMZ - German Collection of Microorganisms and Cell Cultures GmbH (DSMZ). The certificate of origin is presented in Appendix A. The dried culture was activated according to the instructions provided by DSMZ. The instructions are also included in Appendix A. Table 4.1 describes the composition of the media used.

Table 4.1: Composition and manufacturer of media used in this study.

Medium	**Composition (per L)**	**Manufacturer**
NA	Peptone 5 g, Beef extract 3 g, Agar 15 g	Difco™
NB	Peptone 5 g, Beef extract 3 g	Difco™

For each measurement, the strain was initially cultured by streaking onto NA from the preserved NB. The plates were incubated under aerobic conditions at 37 °C for 24 hours. Figure 4.1 illustrates the incubated *E. coli* DSM 33277.

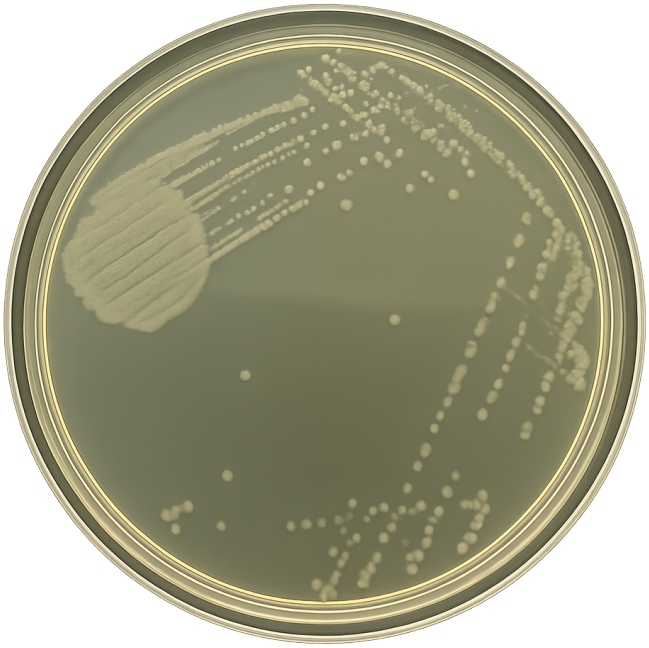


Figure 4.1: *E. coli* DSM 33277 colonies on NA.

After incubation, one colony was scraped from the plate. The colony was transferred into 10 ml of NB into 15 ml SARDSTED^{TM} tube and incubated at 37 °C for 24 h. After incubation, 100 µl of the culture was transferred into 9.9 ml of media containing 100 µl of the substrate. The substrate was ethanol, glucose, or fecal extract depending on the measurement. The medium was contained in 100 ml borosilicate glass flask (VWR), equipped with a dual-inlet cap (Duran®, GL 45) for the measurements. This method is a modification of the method presented by Tagaino et al.[81]

4.2 Instruments Used and Experimental Design

4.2.1 Proton-Transfer-Reaction Time-of-Flight Mass Spectrometer (PTR-TOF-MS)

The PTR-MS instrument used in this study was commercially available PTR-TOF-MS (PTR-TOF 1000, Ionicon), which has a resolving power of 1500. Figure 4.2 presents schematic overview of the instrument.

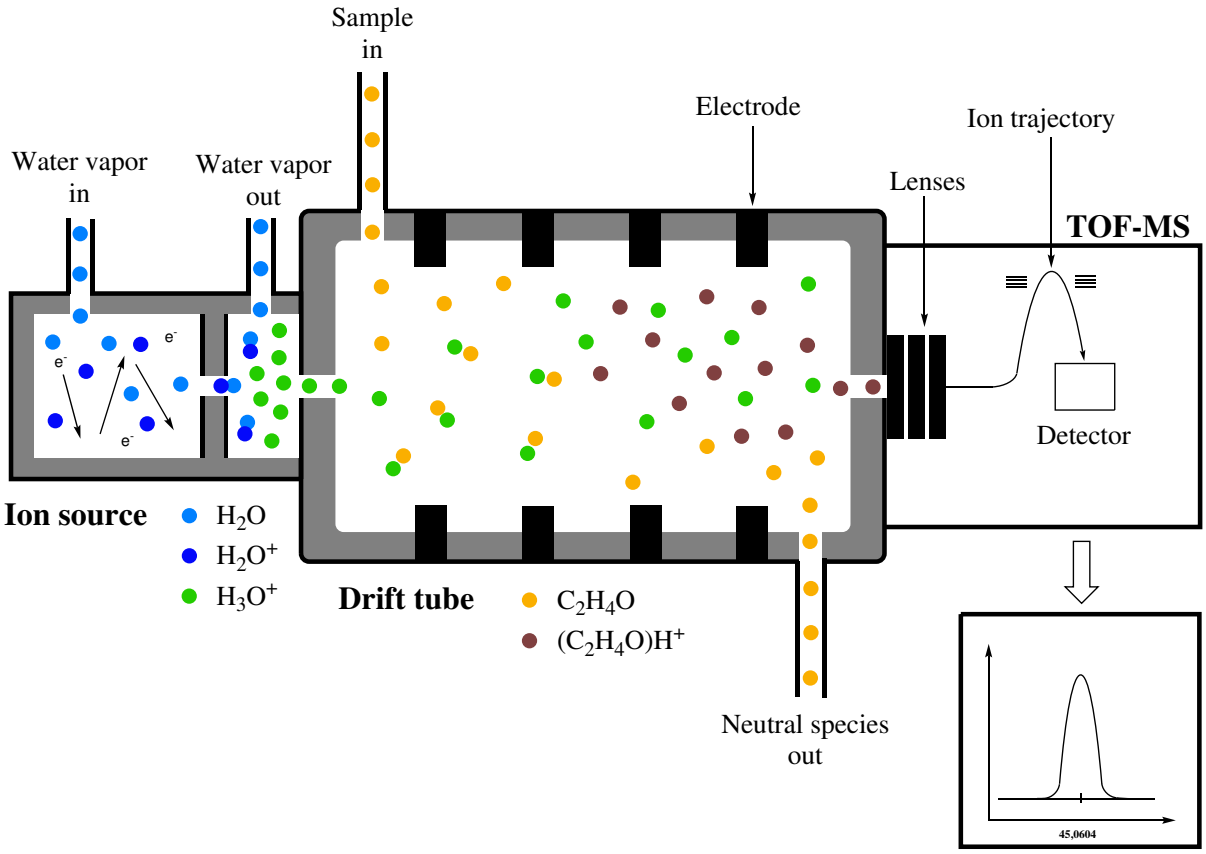


Figure 4.2: Illustration of the PTR-TOF-MS instrument used, with key components labeled. The production of the H_3O^+ ions, the ionization process of sample molecules, and their progression to a mass spectrum are shown.[72]

The measurements were performed from *m/z* 17 to 239, sampling frequency was 1 Hz, and the precursor ion was H_3O^+. An automated mass calibration was carried out using nitrosonium at *m/z* 29.9970 and an internal standard at *m/z* 203.9436. Mass calibration is discussed in greater detail in Section 4.3. The operating conditions are described in Table 4.2.

Table 4.2: Operating conditions used for measurements. Abbreviation: sccm, standard cubic centimeters per minute.

Parameter	**Value**
Field density ratio (Td)	128
Drift tube voltage (V)	550.0
Drift tube pressure (mbar)	2.20
Drift tube temperature (°C)	70.0
Ion source current (mA)	3.3
H_2O flow (sccm)	5.0
Inlet flow (sccm)	20.0
Inlet temperature (°C)	70.0

4.2.2 Gas Chromatography-Mass Spectrometer (GC-MS)

A schematic overview of a typical GC-MS instrument is presented in Figure 4.3.

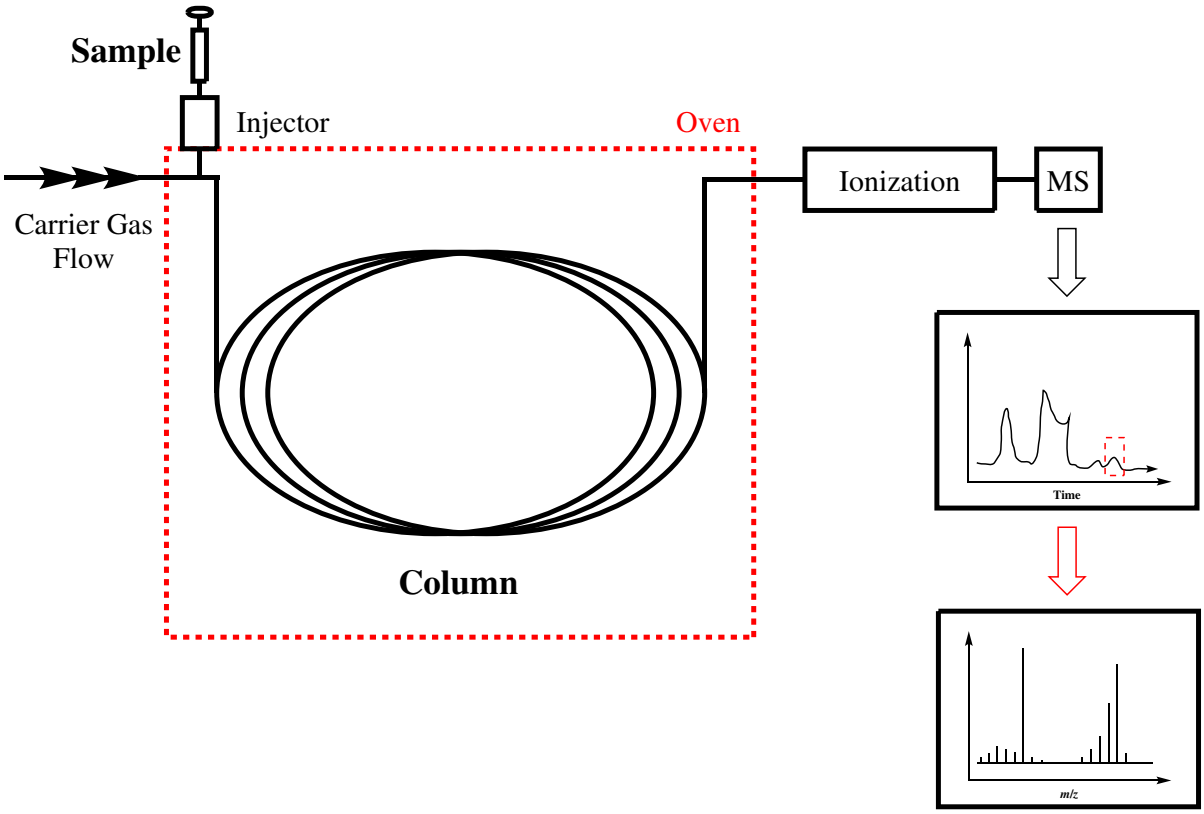


Figure 4.3: Illustration of GC-MS instrument with key components labeled. The progression from chromatographic separation and mass analysis to a chromatogram and a mass spectrum is shown.[72]

The GC-MS used (Agilent 7890A) was combined to a mass spectrometer (Agilent 5975B), which uses a quadrupole as mass analyzer. The ionization technique used was EI. A polar porous-layer open tubular column (Rt-U-BOND PLOT, Restek) was used. The column is 30 m long with an inner diameter of 0.25 mm and a film thickness of 0.8 µm. Separation in the column was achieved by the temperature program described in Table 4.3.

Table 4.3: Temperature program used for measurements.

	Rate (°C/min)	**Value** (°C)	**Hold time** (min)	**Run time** (min)
Initial		30	2	2
Ramp 1	7	100	0	12
Ramp 2	10	180	0	20

Helium was used as the carrier gas. Mass scans were made from *m/z* 29 to 300. The ion source temperature was kept at 230 °C, and the quadrupole temperature at 150 °C. The method was adapted by Julius Nokelainen from a method used in his Master's thesis.[82]

4.2.3 Gas Flow System for Headspace Measurements

A custom-built gas flow system was developed to analyze the acetaldehyde production. The work of Roslund et al. laid the groundwork for the design.[72] Figure 4.4 illustrates the system.

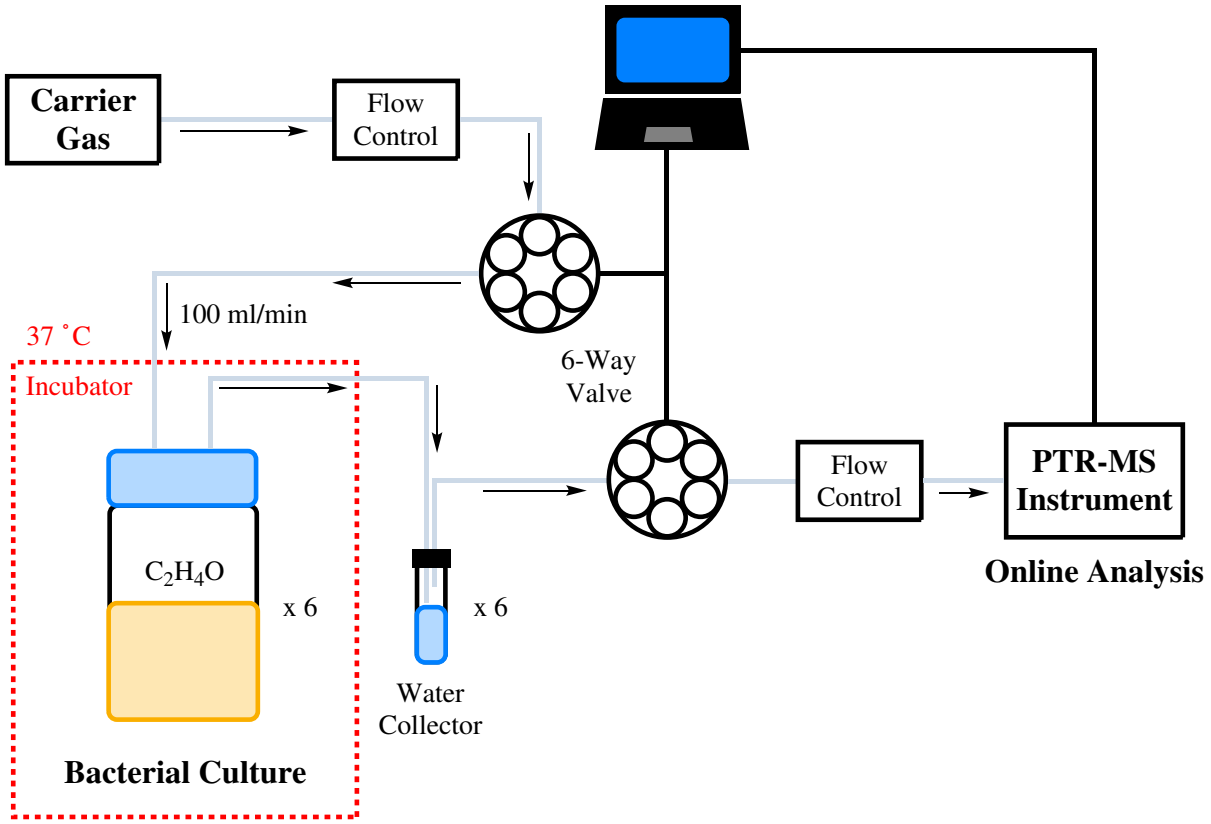


Figure 4.4: Illustration of the custom-built gas flow system with key components labeled. The direction of sample flow is shown.

The system was designed to support up to six bacterial cultures simultaneously and allow the use of two measurement methods:

1. **Parallel measurements**: In this approach, samples were taken one by one from different cultures. Each culture was sampled every x minutes. The interval x depended on the number of cultures measured and the time used per culture. It could be decreased by reducing the number of cultures or measuring for a shorter time per culture. With this method, it was possible to compare different conditions.

2. **Continuous measurements**: In this approach, a single culture was monitored continuously over time. This method was used to capture a more complete picture of metabolic fluctuations.

The system included the carrier gas source, incubator, water collectors, mass flow controllers, PTFE tubes (3 mm and 6 mm) as gas lines, six-way valves, computer, and PTR-TOF-MS. Two

six-way valves (080T6 series Flow Selection Valve, Bio-Chem Valve™) controlled the flow of gas in the system. These valves were connected to the PTR-TOF-MS measurement program (IoniTOF, Ionicon), which allowed automated switching between cultures. The switching process was controlled by an automation script that ensured that each culture was sampled at defined intervals. The automation script is presented in Appendix B.

Compressed air was used as the carrier gas and the flow to the culture was controlled using a mass flow controller (M100B Mass-Flo®, MKS Instruments) to ensure it did not exceed 100 ml/min. The second mass flow controller was used to ensure that the flow is sufficient for PTR-TOF-MS. Water collectors were used to prevent water from entering the PTR-TOF-MS.

The system was also designed to be adaptable for GC-MS measurements. For these measurements, an airtight container was placed before the PTR-TOF-MS, and a sample was taken from the container using a chromatography syringe. The setup is presented in Figure 4.5.

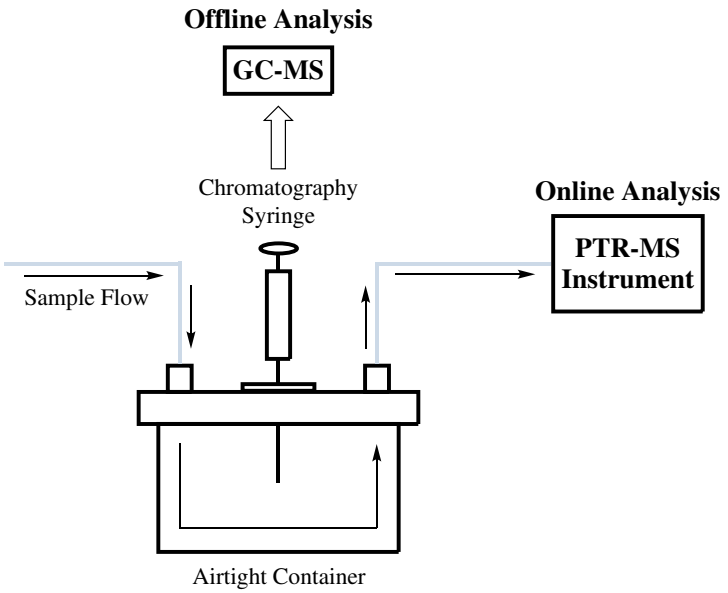


Figure 4.5: Illustration of sample collection from the setup. The direction of sample flow is shown.

4.3 Data Analysis and Processing

The *PTR-MS Viewer* software was used to process and analyze the PTR-TOF-MS measurement data, which are the consecutive mass spectra. The software provides various tools for processing and analyzing such as mass calibration, calculation of the peak data and concentrations, and compound identification. Mass calibration refers to determining the parameters of the relationship between *m/z* and time-of-flight to obtain an accurate mass scale. An accurate mass scale is essential for the calculation of the peak data and concentrations. The spectrum-per-spectrum mass calibration was carried out using two-point calibration, which

refers to using two known peaks as a reference. The two peaks used were nitrosonium at *m/z* 29.9970 and an internal standard at *m/z* 203.9436. These peaks are always present in the spectrum and of moderate intensity. Nitrosonium is formed in the ion source as molecular nitrogen and oxygen enter the ion source via diffusion or through leaks. The internal standard was provided by an Ionicon permeation unit, specifically 1,3-Diiodobenzene ($C_6H_4I_2$, CAS 626-00-6).

To extract concentration data from the measurement data, all ion counts for one peak must be integrated. The software does this automatically after user has manually selected the limits of integration. For isolated peaks, the procedure is straightforward. The situation becomes more complicated when peaks overlap. In this case the peaks need to be separated. In the data of this study, the acetaldehyde and carbon dioxide peaks overlapped. The peaks were separated using the software's *Multipeak Mode*, where peak functions were fitted to the overlapping peaks to determine the intensities. The intensity of the acetaldehyde peak was much stronger, and the carbon dioxide peak could not be resolved. Figure 4.6 presents *Peak Table Editor*, which was used for the fitting and calculation.

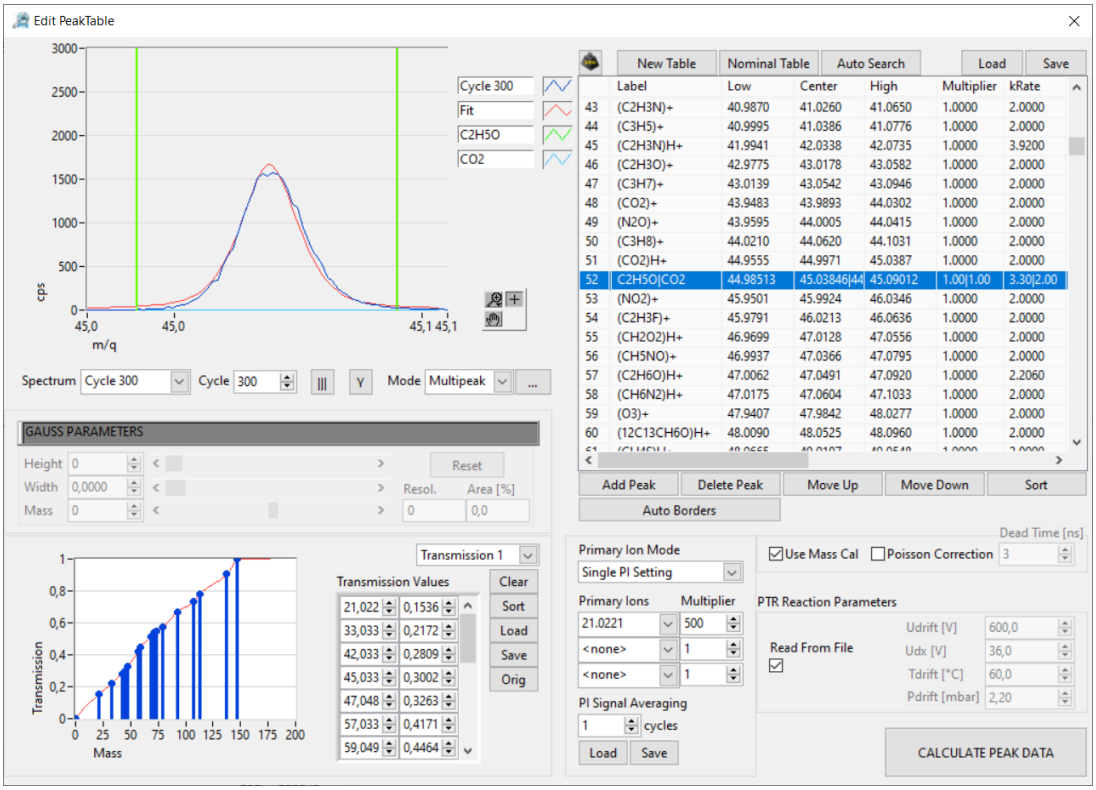


Figure 4.6: Screenshot of the *Peak Table Editor* window. Acetaldehyde and carbon dioxide peaks have been separated using the *Multipeak Mode*. The carbon dioxide peak could not be resolved due to its much lower intensity. Primary ion is H_3O^+ with a multiplier of 500.

The *Isotope Tool* was used to get information about presence of acetaldehyde in the sample by matching its theoretical isotope pattern with the measurement data. The software calculated the expected isotope masses and abundances, which were then overlaid on the spectrum for visual comparison. Figure 4.7 presents the *Isotope Tool*.

Figure 4.7: Screenshot of the *Isotope Tool* window. The theoretical isotope pattern is shown with expected *m/z* and abundances. The red dots represent the theoretical isotope pattern, and the green line represents the spectrum. For the third isotopologue, the theoretical pattern does not match the spectrum due to overlap with ethanol (*m/z* 47).

After processing and analyzing the measurement data with the *PTR-MS Viewer*, the concentration data were extracted. Some additional processing was necessary for concentration data obtained from parallel measurements. Only selected data ranges were used for plotting the results. A Python script was written for this processing, which is presented in Appendix B. This script is useful for studies like this or for breath analysis. The user is asked to define the number of steps and then specify the start cycle and end cycle of each step. Once this has been repeated for all steps, the user is asked to provide a filename for the output file. The output file contains the filtered data. The Python scripts used for plotting are not presented because of their simplicity.

The data from GC-MS measurements were analyzed using *Agilent MassHunter 7* software. The identity of acetaldehyde was confirmed through automated library matching and visual inspection of the spectra. The acquired mass spectra were compared with reference spectra from the National Institute of Standards and Technology (NIST20) Mass Spectral Library. Quantitative data were not obtained because the analysis was intended only for compound identification.

5. Results and Discussion

5.1 Optimization of Glucose and Ethanol Concentrations

A comparative analysis of glucose and ethanol concentrations was conducted at the start of the study. The solution series used for the analysis were prepared within biologically relevant ranges. The four cultures were sampled sequentially, for 20 minutes each. Every fourth sampling was followed by a 20-minute flush. The results are presented in Figures 5.1 and 5.2. Each data point represents the mean of the last three minutes of the sampling period.

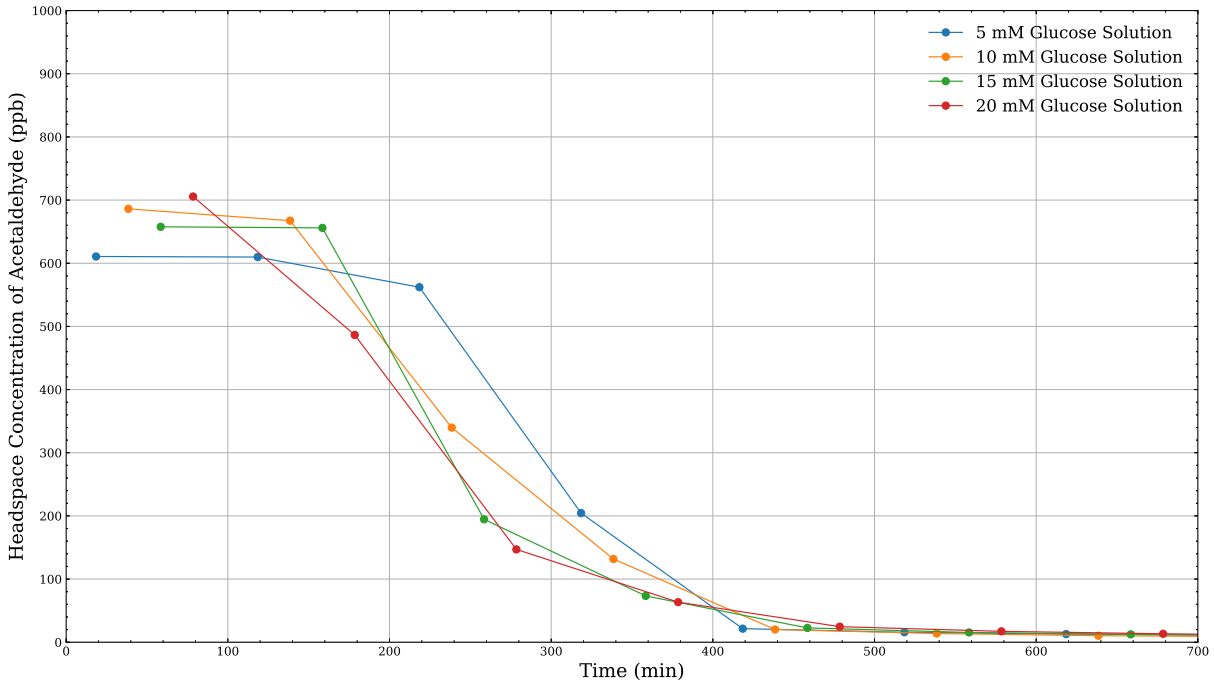


Figure 5.1: A comparative analysis of four glucose concentrations.

Figure 5.1 shows that there were initial differences in acetaldehyde concentration in the headspace with varying glucose concentrations. In all cultures, the concentration decreased to below 10 ppb by the seventh measurement cycle (12 h). This decrease is logical, as it is likely linked to growth. Under similar conditions, the growth rate is the same between cultures. It was also observed that the decrease occurred faster at higher glucose concentrations. This trend suggests that substrate utilization shifts to rely primarily on glucose without the simultaneous utilization of amino acids from NB. This observation contrasts with the findings

of Zampieri et al. in defined media, where amino acids were preferred over glucose during the fastest growth phase.[60] Without information about active genes, firm conclusions about the underlying regulation cannot be drawn.

Glucose was probably metabolized by mixed-acid fermentation, as the cultures were static and therefore microaerobic. However, it is also possible that glucose was partially utilized via respiration. The carrier gas flow may have aerated the surface layer of the medium, leaving deeper layers microaerobic. This provides a methodological explanation for why there is uncertainty regarding the active metabolic pathways.

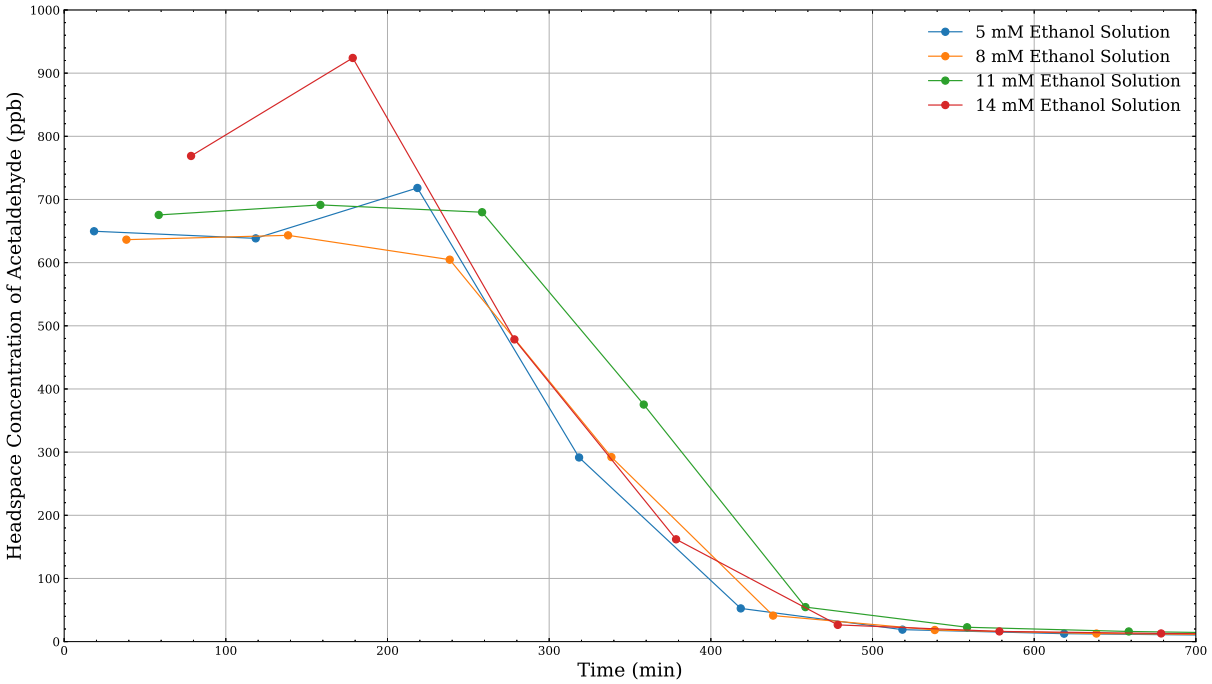


Figure 5.2: A comparative analysis of four ethanol concentrations.

A similar decrease rate was observed for ethanol in Figure 5.2. However, the theoretical background indicates that ethanol did not act as a direct carbon source for the production of acetaldehyde. As discussed in Section 2.2.2, ethanol oxidation is limited due to the low activity of AdhE under aerobic conditions. In addition, this oxidation does not occur under microaerobic or anaerobic conditions. The detected acetaldehyde was therefore most likely the result of amino acid fermentation. Consequently, the ethanol response may represent a false positive. This conclusion is further supported by the results presented in Section 5.3, where acetaldehyde production was observed even without added substrate.

5.2 Comparative Analysis of Response to Glucose and Ethanol

Based on the results presented in Section 5.1, an 11 mM ethanol solution and a 15 mM glucose solution were selected for comparative analysis. These concentrations were chosen to provide a measurable acetaldehyde response while avoiding substrate levels that led to rapid decline. In addition, the glucose concentration was selected to represent elevated blood glucose levels.

Two cultures were sampled sequentially without a flush, each for 10 minutes. The shorter interval compared to the measurements in Section 5.1 was chosen to provide improved time resolution and to capture the VOC profile more accurately. The results presented are in Figures 5.3-5.5. Data points represent the mean of the last three minutes of the sampling period.

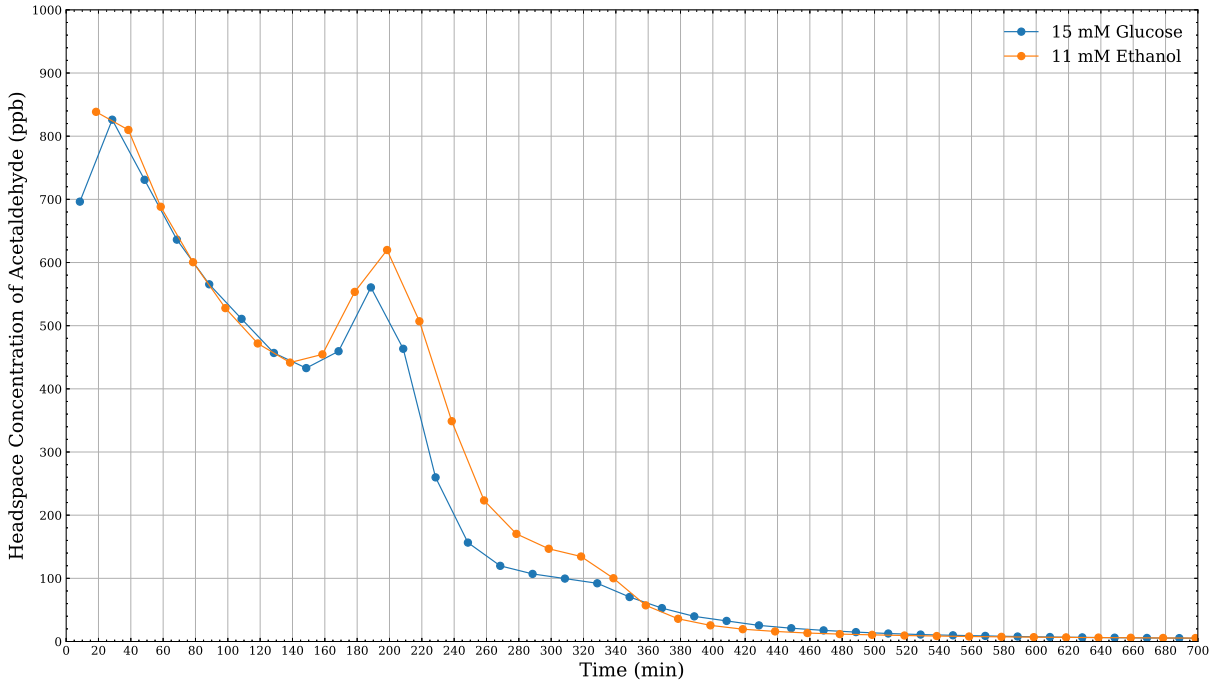


Figure 5.3: The first comparative analysis of 11 mM ethanol and 15 mM glucose concentrations.

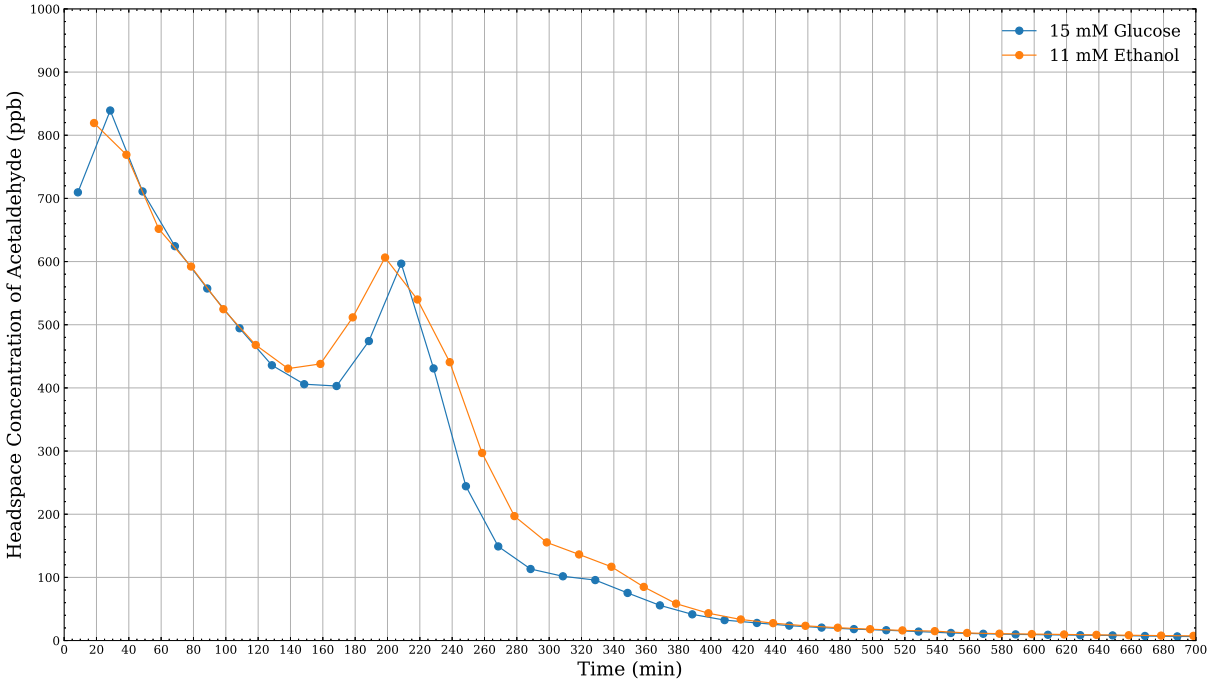


Figure 5.4: The second comparative analysis of 11 mM ethanol and 15 mM glucose concentrations.

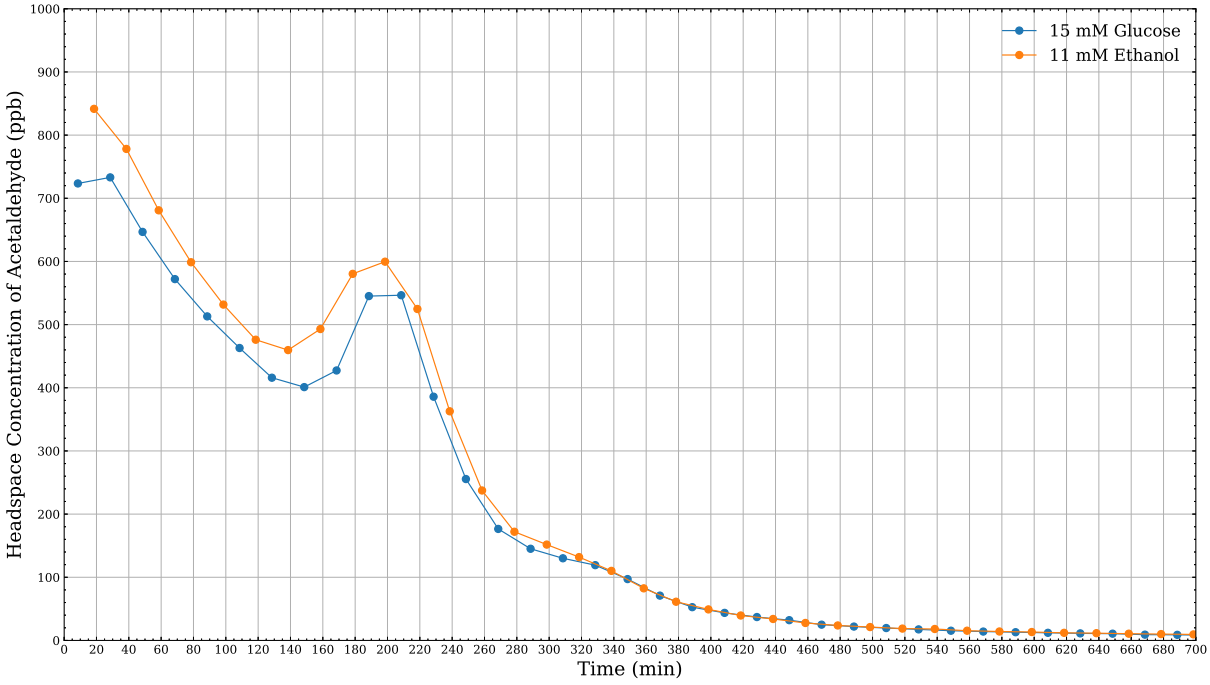


Figure 5.5: The third comparative analysis of 11 mM ethanol and 15 mM glucose concentrations.

Observations and conclusions from Figures 5.3-5.5 are consistent with those presented in Section 5.1. Additionally, the Figures show that the comparative analysis with fewer cultures and shorter intervals captured the volatile profile more accurately. The replicate measurements also closely resembled each other. Methodologically, it can be concluded that shorter intervals are preferable to obtain better data.

In each analysis, a rise in acetaldehyde concentration was observed after three hours, followed by a decrease and a stable area around five hours. The observed rise suggests a shift in metabolic flux, most likely caused by changes in oxygen availability or nutrient utilization. The stable area observed around five hours is probably linked to nutrient depletion before starvation. As discussed in Section 5.1, there is methodological explanation related to oxygen availability. In contrast, conclusions regarding nutrient utilization cannot be drawn from the volatile profile alone.

5.3 Continuous Monitoring

Continuous monitoring of a culture was conducted using 11 mM ethanol, 15 mM glucose, fecal extracts as substrates, and without added substrate. The absence of a substrate means that *E. coli* utilized amino acids from NB. The results are presented in Figures 5.6-5.9. Repeat means that the measurement has been repeated starting from the cultivation.

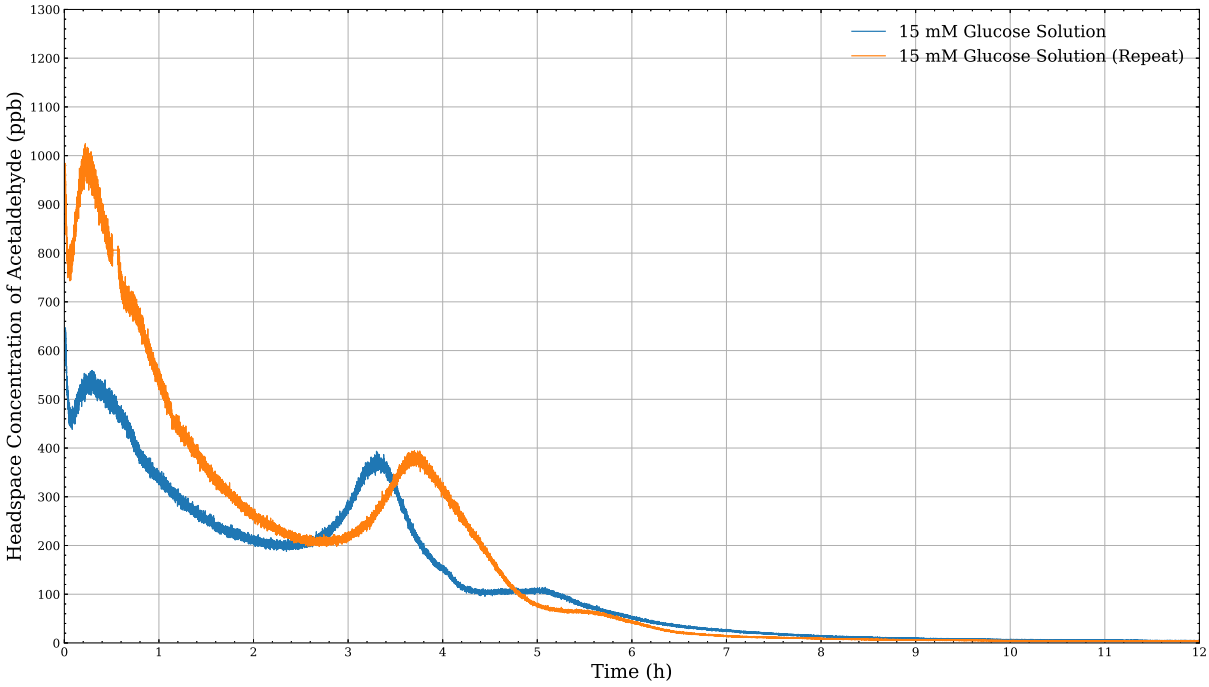


Figure 5.6: Continuous monitoring of cultures supplemented with 15 mM glucose solution.

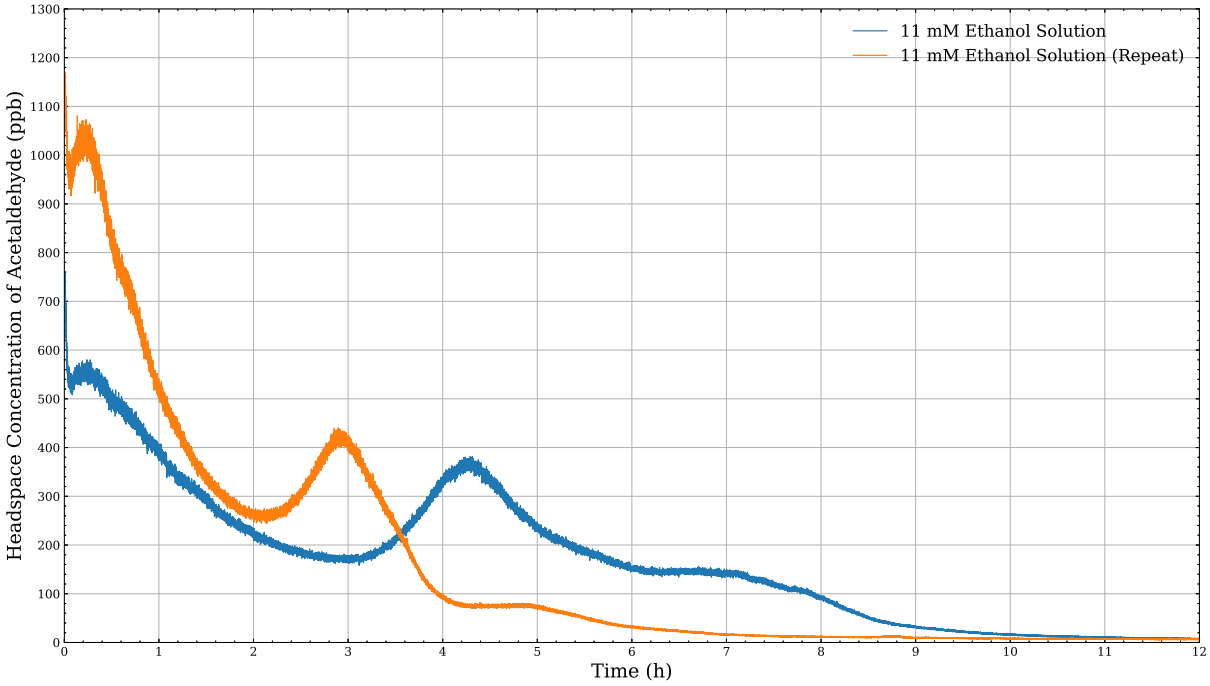


Figure 5.7: Continuous monitoring of cultures supplemented with 11 mM ethanol solution.

Figures 5.6 and 5.7 support the observations and conclusions presented in Sections 5.1 and 5.2. The rate of decrease is the same and a similar volatile profile is observed.

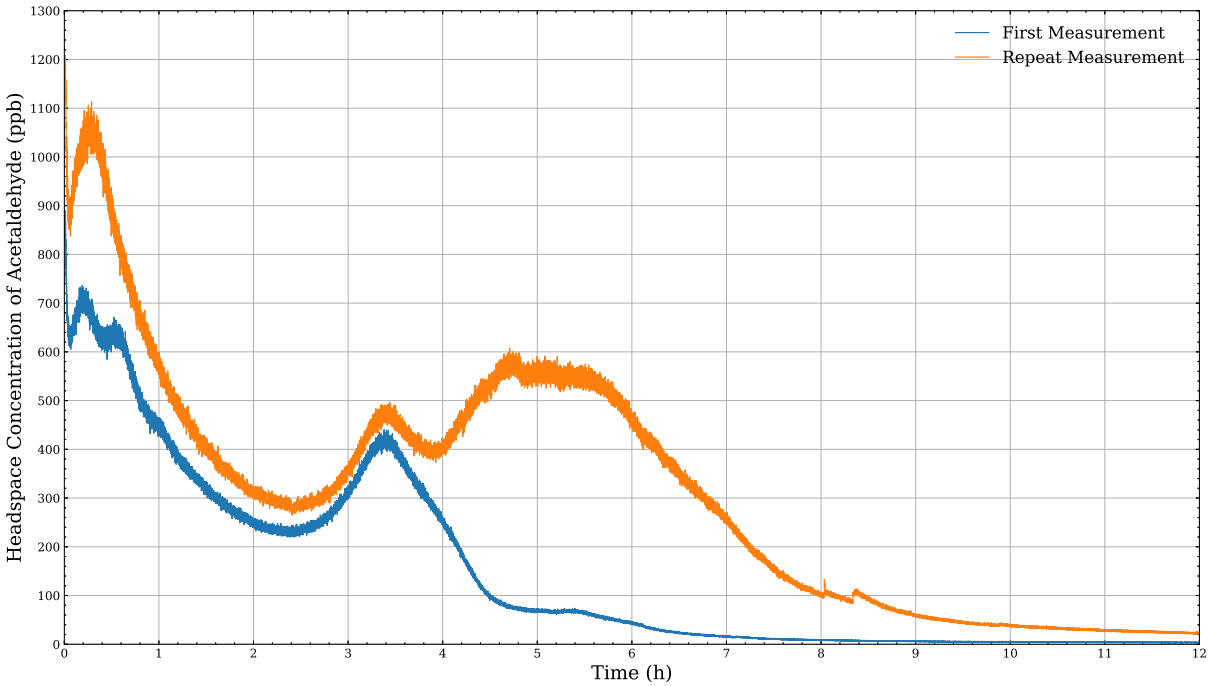


Figure 5.8: Continuous monitoring of cultures with no added substrate.

Figure 5.8 strongly supports the conclusion that ethanol represents a false positive, as it shows that acetaldehyde is produced alone from amino acids in NB. The result aligns with the theo-

retical framework regarding amino acid fermentation. Although no studies directly support this finding, some indirect evidence exists. As discussed in Section 2.1.1, Zheng et al. produced a VOC-profile for *E. coli* in which acetaldehyde was detected. Their study used tryptic soy broth, which is a widely used rich medium with nutrient composition similar to NB. Therefore, their results provide indirect support for this finding.

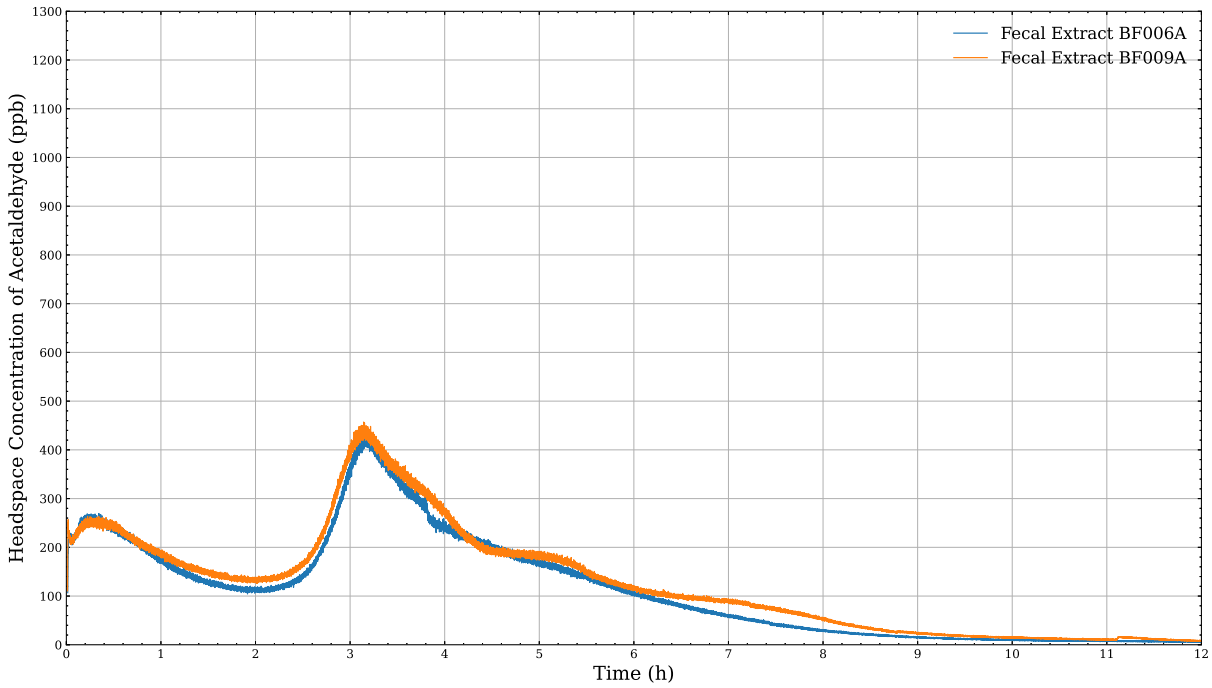


Figure 5.9: Continuous monitoring of cultures suplemented with fecal extract (samples BF006A and BF009A).

Figure 5.9 shows that the fecal extracts produced a weaker response than expected. The initial acetaldehyde concentration was even lower than that observed in Figure 5.8. However, FinnDiane Research Group has analyzed diabetic fecal samples from patients with type 1 DM and reported reduced levels of intestinal immunoglobulin A. This antibody recognizes oxidation products of LDL (low-density lipoprotein), such as the malondialdehyde-acetaldehyde adduct. A reduction in intestinal immunoglobulin A reflects weak mucosal immune defense.[83] Taking their findings into account, it is reasonable to assume that the fecal extract contains a certain amount of antibodies capable of binding acetaldehyde. This could consequently reduce the concentration of volatile acetaldehyde and explain the weaker response.

In general, continuous measurement captured the volatile profile more accurately than parallel measurements with short intervals. However, the replicates were inconsistent. The reason for this remains unclear.

5.4 GC-MS Measurements

GC-MS measurements were conducted to identify acetaldehyde unambiguously from the headspace samples. The samples were collected from continuous monitoring of cultures supplemented with glucose, ethanol, or without added substrate. The samples were collected from three different time points in each case. In glucose-supplemented culture and culture without added substrate, these sampling points correspond approximately to 0.5 h, 2.5 h, and 3.5 h. In ethanol-supplemented culture, the corresponding sampling points were approximately 0.5 h, 2 h, and 3 h. Sampling points were selected based on the VOC profiles shown in Figures 5.6-5.8. The results are presented in Figures 5.10-5.12.

The top spectrum corresponds to a sample collected near the first peak, the middle spectrum corresponds to a sample collected near the minimum, and the bottom spectrum to a sample collected near the second peak. Each measured spectrum (red) is shown above the NIST reference spectrum (blue). Retention time is reported as the mean of three samples.

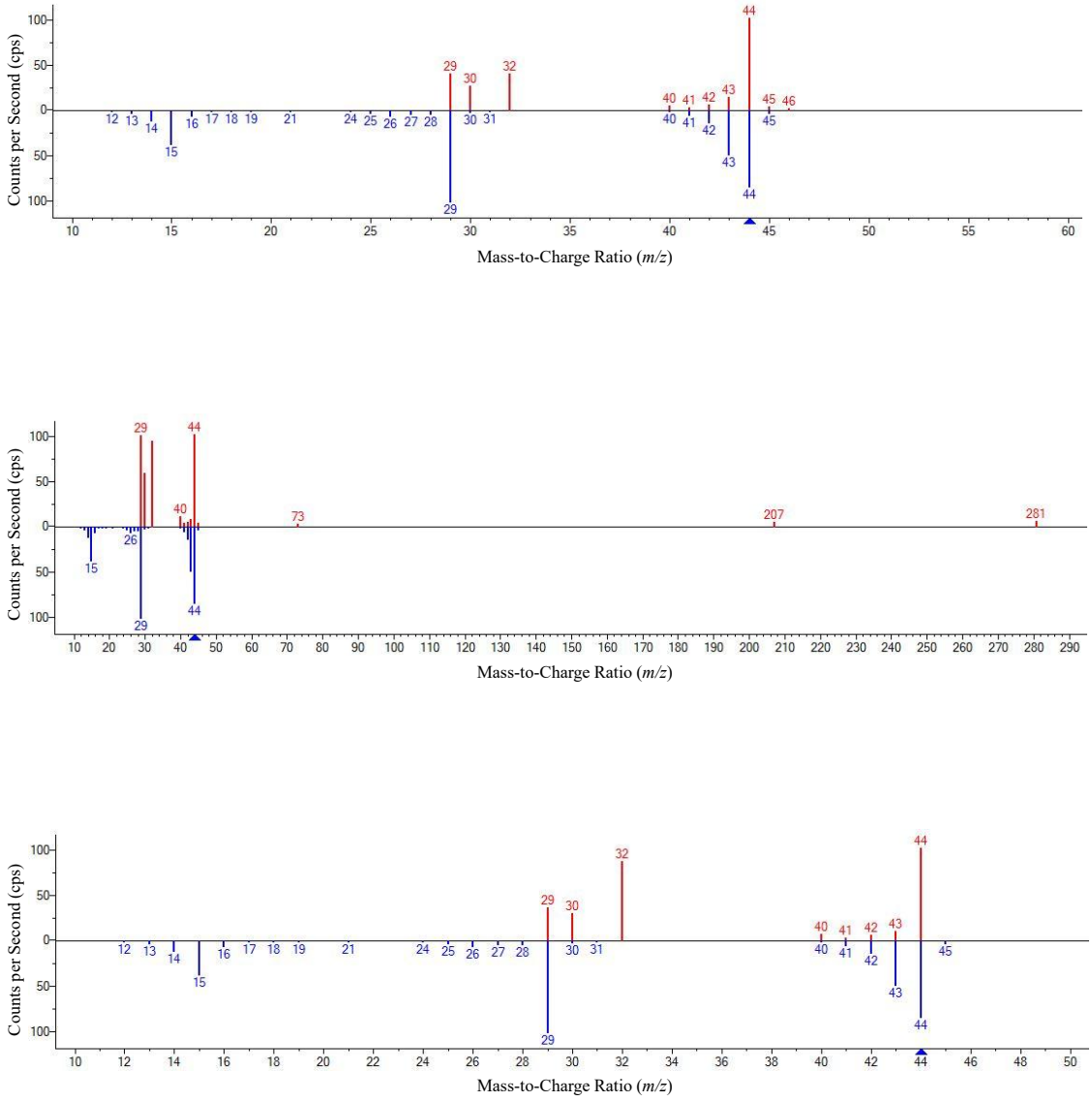


Figure 5.10: Mass spectra of acetaldehyde from three samples collected during continuous monitoring of glucose suplemented culture. Retention time: 11.48 min

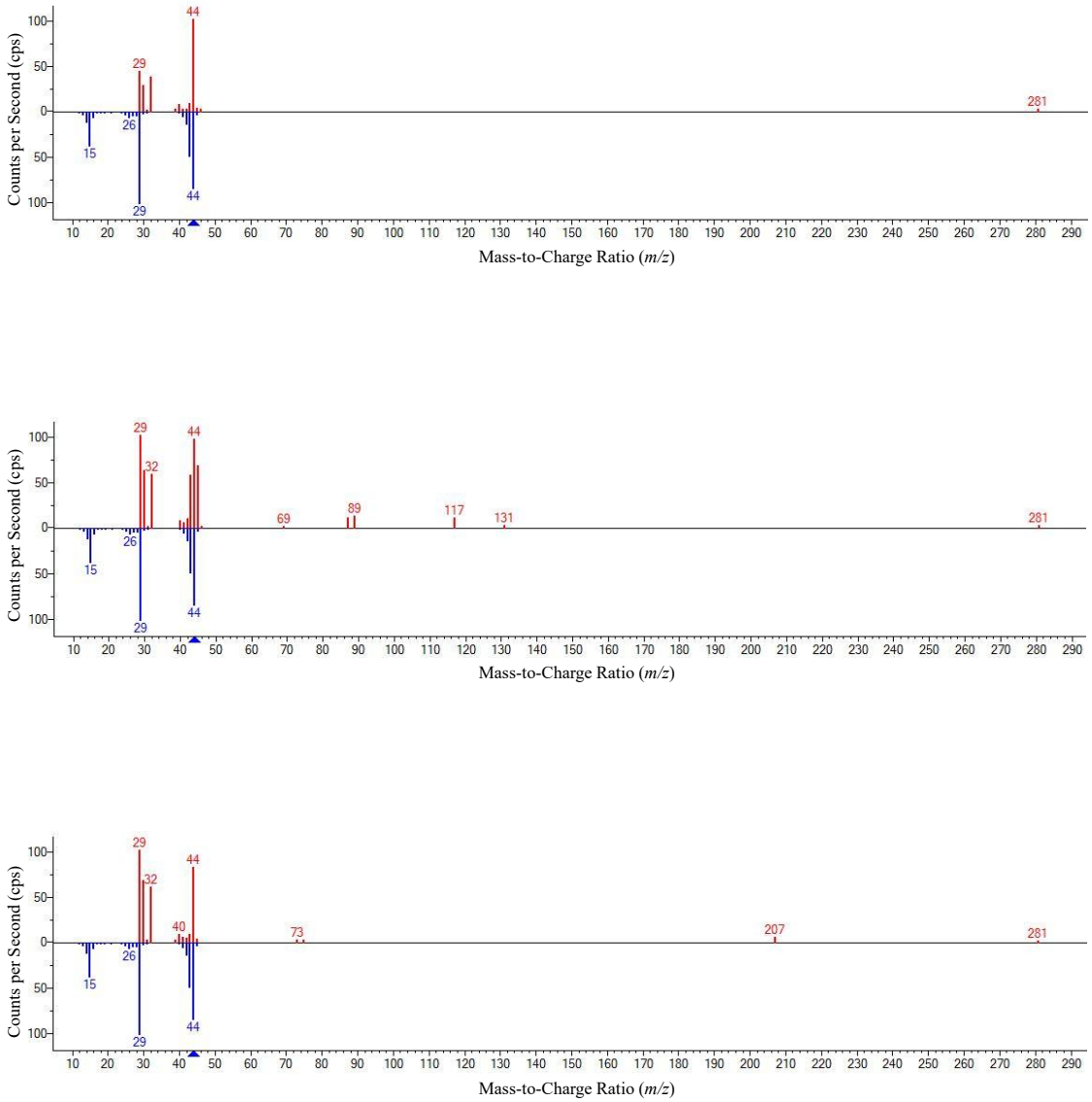


Figure 5.11: Mass spectra of acetaldehyde from three samples collected during continuous monitoring of ethanol suplemented culture. Retention time: 11.45 min

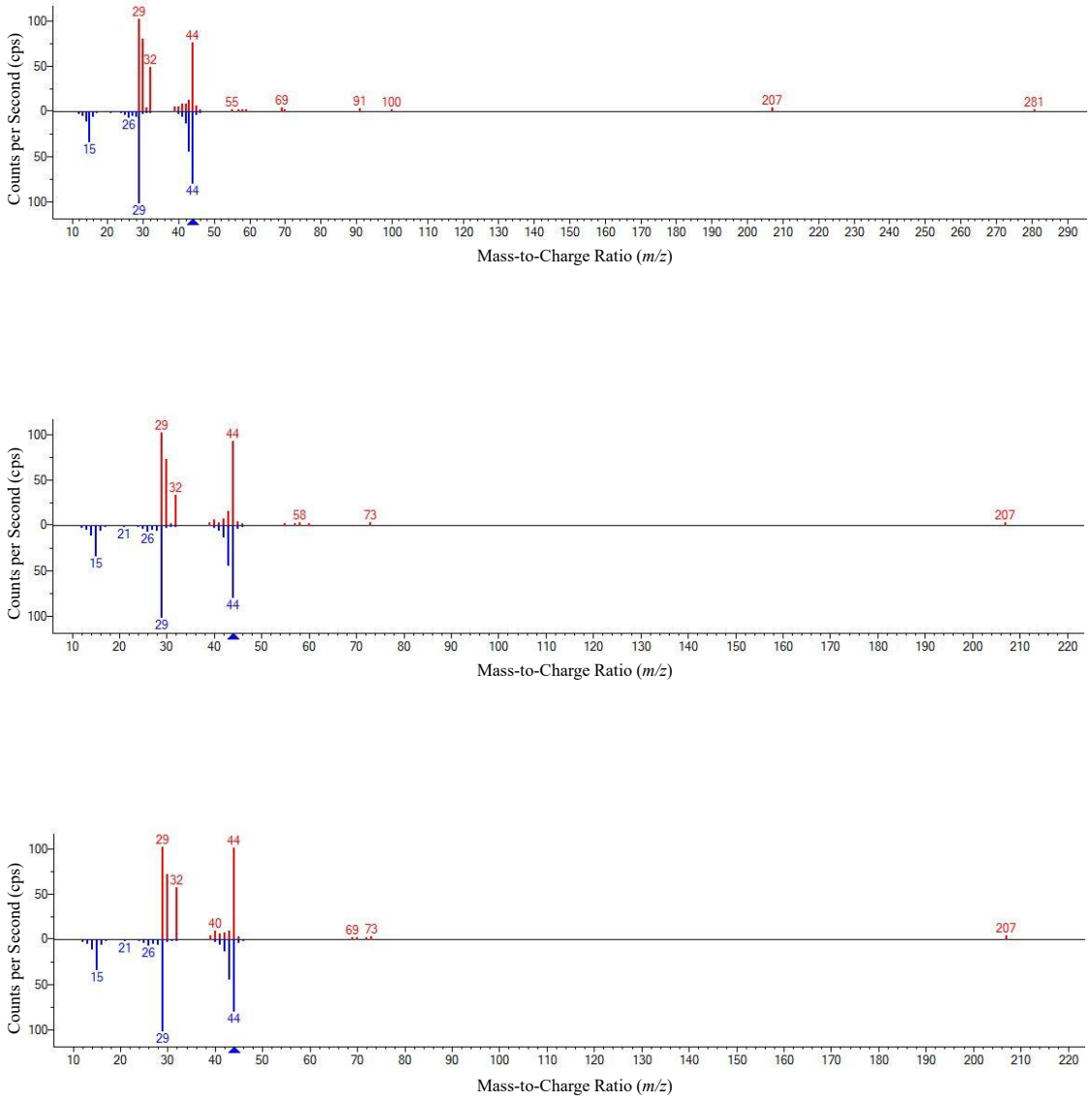


Figure 5.12: Mass spectra of acetaldehyde from three samples collected during continuous monitoring of culture without added substrate. Retention time: 11.46 min

The results confirm the presence of acetaldehyde in all samples. In Figure 5.10, the fragmentation pattern of acetaldehyde is seen more clearly. Figures 5.11 and 5.12 show that fragments of some unknown compounds interfered with the spectrum.

6. Conclusions

6.1 Summary of Findings

This thesis investigated the production of acetaldehyde by *E. coli* under different substrate conditions. The study aimed to confirm the bacterial metabolic responses to glucose and ethanol using liquid-based headspace measurements. In addition, the effect of fecal extract on acetaldehyde production was examined. All continuous measurements are plotted in Figure 6.1 to summarize the findings visually.

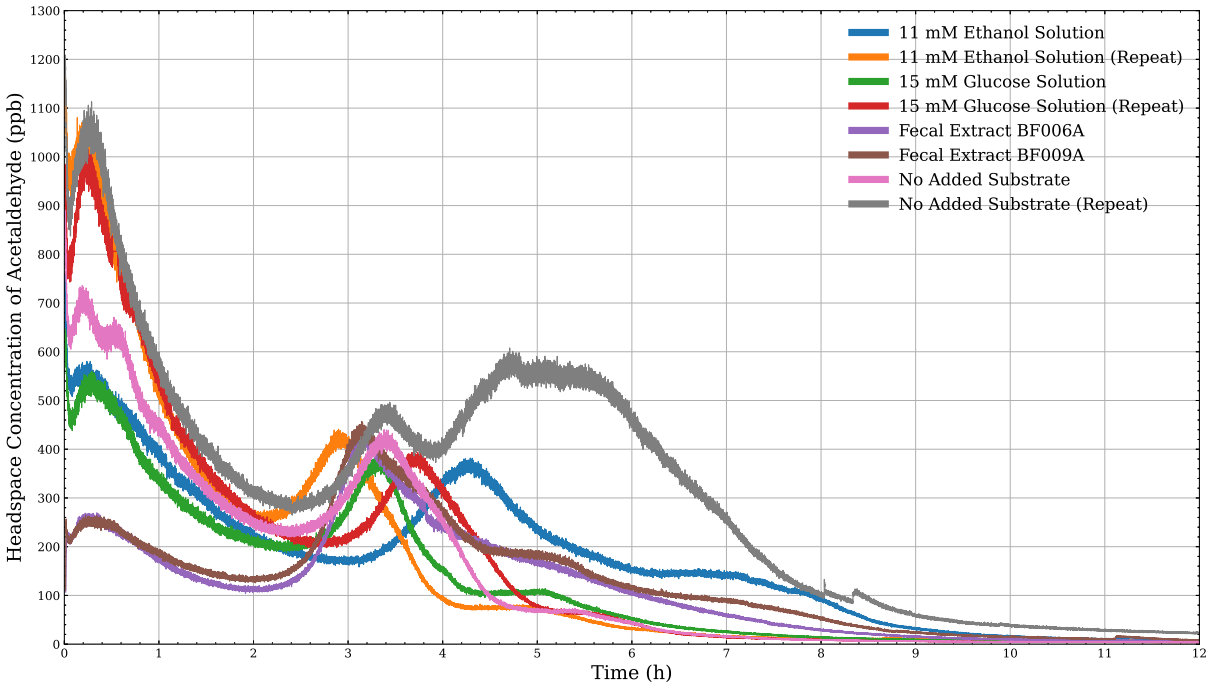


Figure 6.1: Summary of continuous measurements.

The findings partially support the initial hypothesis and align with the theoretical framework. However, a key observation was that acetaldehyde was also produced in cultures without any added substrate. This result shows that amino acid fermentation leads to acetaldehyde formation and is in line with the theory discussed in Section 2.2. This amino acid fermentation was overlooked in the original hypothesis. As a result, the simultaneous utilization of amino acids probably masked the glucose-specific response. This complicated the interpretation of the measured VOC profiles. Because no data on gene expression were available, the exact metabolic

pathways could not be confirmed. Ethanol was included as a reference rather than as a substrate related to the hypothesis. Its effect was most likely a false positive, as discussed in Section 5.3.

The response observed in cultures supplemented with fecal extract was weaker than expected. The initial acetaldehyde concentration was even lower than that observed from cultures without added substrate. However, previous findings from the FinnDiane Study Group may offer a mechanistic explanation for the observation. As discussed in Section 5.3, the FinnDiane Study Group has reported reduced levels of intestinal immunoglobulin A in fecal samples from individuals with DM.[83] Taking this finding into account, it is possible that components present in fecal extracts may have reduced the amount of volatile acetaldehyde. However, there is no comparative data from fecal extract of healthy individuals in this study. Therefore, this conclusion is speculative, and the observed response neither directly supports nor contradicts the initial hypothesis.

Although there was uncertainty about the exact conditions and which pathways were active, the developed gas-flow system proved functional. Parallel measurements with shorter intervals provided the most reliable data, and the GC-MS analyzes verified that acetaldehyde was consistently present in the headspace gas.

6.2 Limitations of the Study

This study provided insights into acetaldehyde production in *E. coli* while also highlighting several limitations and challenges. A major challenge was the detection of acetaldehyde in continuous measurements without added substrate. Although *E. coli* is recognized in the literature to ferment amino acids, acetaldehyde production from amino acid fermentation has not been discussed in previous bacterial VOC studies. Consequently, this aspect of acetaldehyde formation was overlooked in the literature review conducted before the study. As a result, this aspect was not considered in the original study design. Therefore, the conclusions on the findings presented in Sections 5.1-5.3 are partly speculative.

Another limitation was the lack of information on gene expression. Without such data, it remains uncertain which metabolic pathways were active during acetaldehyde production. The lack of information on active genes also raises the possibility that not all of the produced acetaldehyde volatilized into the headspace but was instead partially neutralized. This neutralization could occur enzymatically through aldehyde dehydrogenase B or via the accumulation of the tripeptide glutathione to physiologically active levels.[47,84] Such reactions represent a natural detoxification mechanism in *E. coli*, allowing the bacterium to neutralize harmful intermediates. This mechanism may have contributed to the unexpected changes observed in the VOC-profile.

The study also faced certain limitations in the experimental setup. One major issue was the in-

ability to show the effect of bacterial metabolic activity on acetaldehyde production by bacterial real-time growth curve. In addition, the use of carrier gas did not serve its purpose. The carrier gas probably aerated the top layer of the medium and altered oxygen availability in the culture.

Methodological limitations were also present. These limitations made the study prone to contamination and caused uncertainty about the presence of acetaldehyde. During the cultivation step, the addition of the substrate increased the risk of contamination. While the GC-MS results supported the presence of acetaldehyde in the headspace gas, fragments below *m/z* 29 were not detected. The fragmentation pattern was therefore incomplete and absolute certainty could not be achieved with the method used.

6.3 Future Research Directions

This study represents a step forward in understanding the matter. However, several improvements are needed to address its current limitations. The following directions are proposed for future research based on the findings and limitations.

The use of genetically modified strains would help identify substrate-specific responses more accurately. In such strains, genes encoding key enzymes could be knocked out or overexpressed. This would allow linking observed acetaldehyde production to specific pathways and thus help differentiate between them. Since metabolic engineering approaches for *E. coli* are well documented in the literature, suitable methods for mixed-acid fermentation pathways are readily available.[43] In general, simultaneous utilization of multiple nutrients should be prevented.

Incorporating an online growth monitoring system will be essential to improve the reliability and comparability of the data. This could be achieved by making the studied bacteria fluoresce and measuring the emitted light. This type of method has been developed by Chen et al. and has been used on the microplate scale.[85] Another key improvement for future research would involve the transition to anaerobic conditions, which would better simulate the intestinal environment. Anaerobic conditions would likely provide more realistic insight into substrate-specific effects and yield data more relevant to the hypothesis. Therefore, the use of anaerobic intestinal bacteria should be considered. If the use of aerobic conditions is considered in further studies, the use of carrier gas should be properly controlled. Aeration should be achieved by shaking the cultures to minimize oxygen gradients caused by the carrier gas. Methodological limitations should also be addressed accordingly.

Despite these challenges, the study design and the results were "good enough" within the scope of a Masters thesis. The theoretical section and the observed acetaldehyde production in response to glucose support the hypothesis. The findings in fecal extract measurements neither support nor contradict the hypothesis. The topic remains relevant and with the proposed direc-

tions further studies should be considered.

References

1. V. Salminen, BSc thesis, Faculty of Science, University of Helsinki, Helsinki, 2024.
2. IARC, *List of Classifications*, https://monographs.iarc.who.int/list-of-classifications.
3. International Agency for Research on Cancer, *Personal Habits and Indoor Combustions: IARC Monographs on the Evaluation of Carcinogenic Risks to Humans, Volume 100E*, International Agency for Research on Cancer, Lyon, France, 2012, vol. 100E, pp. 35–167.
4. U.S. Department of Health and Human Services, *15th Report on Carcinogens*, National Institute of Environmental Health Science, Research Triangle Park (NC), 2021, https://ntp.niehs.nih.gov/go/roc15.
5. L. C. Perlmuter, B. P. Flanagan, P. H. Shah and S. P. Singh, *Diabetes Care*, 2008, **31**, 2072–2076.
6. *Diabetes Care*, 2009, **32**, 62–67.
7. T. Lawler, Z. L. Walts, M. Steinwandel, L. Lipworth, H. J. Murff, W. Zheng and S. Warren Andersen, *JAMA Network Open*, 2023, **6**, DOI: 10.1001/jamanetworkopen.2023.43333.
8. J. He, D. O. Stram, L. N. Kolonel, B. E. Henderson, L. Le Marchand and C. A. Haiman, *British Journal of Cancer*, 2010, **103**, 120–126.
9. H. U. Krämer, B. Schöttker, E. Raum and H. Brenner, *European Journal of Cancer*, 2012, **48**, 1269–1282.
10. Y. Ma, W. Yang, M. Song, S. A. Smith-Warner, J. Yang, Y. Li, W. Ma, Y. Hu, S. Ogino, F. B. Hu, D. Wen, A. T. Chan, E. L. Giovannucci and X. Zhang, *British Journal of Cancer*, 2018, **119**, 1436–1442.
11. International Diabetes Federation, *IDF Diabetes Atlas 10th edition*, English, International Diabetes Federation, Brussels, 2021, pp. 1–5, www.diabetesatlas.org.
12. B. Zhou, A. W. Rayner, E. W. Gregg and et al., *The Lancet*, 2024, **404**, 2077–2093.
13. M. Arffman, P. Ilanne-Parikka, I. Keskimäki, O. Kurkela, J. Lindström, R. Sund and K. Winell, *FinDM database on diabetes in Finland*, Finnish Institute for Health and Welfare (THL), Helsinki, 2020, www.thl.fi.
14. T. Y. Ryu, J. Park and P. E. Scherer, *Diabetes & Metabolism Journal*, 2014, **38**, 330.

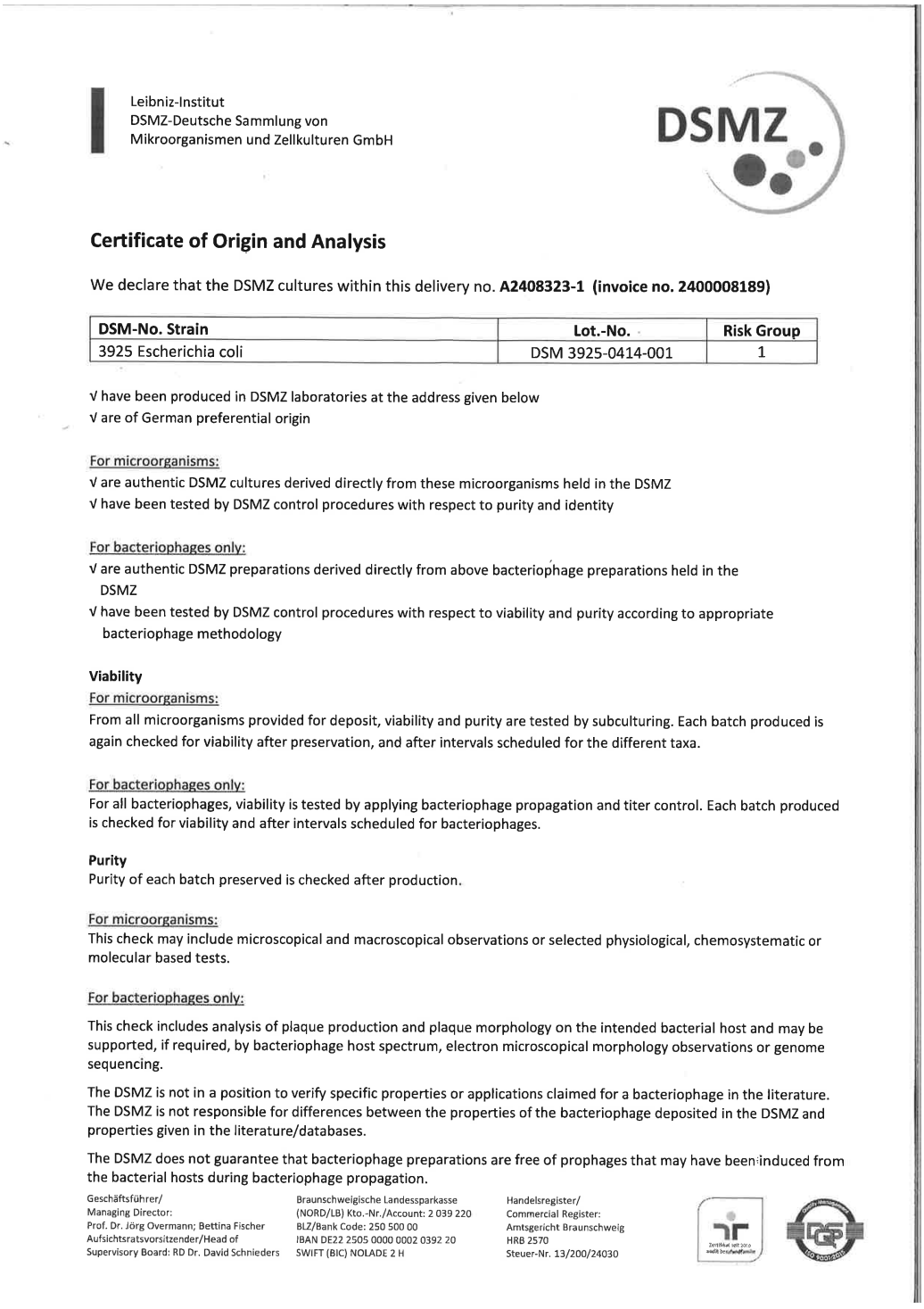
15. J. A. Johnson, B. Carstensen, D. Witte, S. L. Bowker, L. Lipscombe and A. G. Renehan, *Diabetologia*, 2012, **55**, 1607–1618.
16. K. Roslund, Ph.D. Thesis, Faculty of Science, University of Helsinki, Helsinki, 2023, http://ethesis.helsinki.fi.
17. P. J. Jurtshuk, in *Medical Microbiology*, ed. S. Baron, University of Texas Medical Branch at Galveston, Galveston, TX, 4th edition, 1996, ch. 4.
18. B. E. Herrera-Cabrera, A. Delgado-Alvarado, R. Salgado-Garciglia, L. G. López-Valdez, L. M. Sánchez-Herrera, J. Montiel-Montoya, M. Soto-Hernández, L. M. BasurtoGonzález and H. J. Barrales Cureño, in *Bacterial Secondary Metabolites*, Elsevier, 2024, pp. 177–196.
19. R. M. S. Thorn, D. M. Reynolds and J. Greenman, *Journal of Microbiological Methods*, 2011, **84**, 258–264.
20. M. Bunge, N. Araghipour, T. Mikoviny, J. Dunkl, R. Schnitzhofer, A. Hansel, F. Schinner, A. Wisthaler, R. Margesin and T. D. Märk, *Applied and Environmental Microbiology*, 2008, **74**, 2179–2186.
21. E. Kemmler, M. C. Lemfack, A. Goede, K. Gallo, S. M. T. Toguem, W. Ahmed, I. Millberg, S. Preissner, B. Piechulla and R. Preissner, *Nucleic Acids Research*, 2025, **53**, D1692–D1696.
22. N. J. Hayward, T. H. Jeavons, A. J. Nicholson and A. G. Thornton, *Journal of Clinical Microbiology*, 1977, **6**, 195–201.
23. P. J. Coloe, *Journal of Clinical Pathology*, 1978, **31**, 361–364.
24. R. A. Allardyce, V. S. Langford, A. L. Hill and D. R. Murdoch, *Journal of Microbiological Methods*, 2006, **65**, 361–365.
25. J. Scotter, R. Allardyce, V. Langford, A. Hill and D. Murdoch, *Journal of Microbiological Methods*, 2006, **65**, 628–631.
26. L. D. J. Bos, P. J. Sterk and M. J. Schultz, *PLoS Pathogens*, 2013, **9**, DOI: 10.1371/journal.ppat.1003311.
27. I.-A. Ratiu, T. Ligor, V. Bocos-Bintintan, H. Al-Suod, T. Kowalkowski, K. Rafiska and B. Buszewski, *Journal of Breath Research*, 2017, **11**, DOI: 10.1088/1752-7163/aa7ba2.
28. Y. Zheng, F. Li, C. Zhao, J. Zhu, Y. Fang, Y. Hang and L. Hu, *RSC Advances*, 2024, **14**, 25316–25328.
29. M. T. Smith, K. Z. Guyton, C. F. Gibbons, J. M. Fritz, C. J. Portier, I. Rusyn, D. M. DeMarini, J. C. Caldwell, R. J. Kavlock, P. F. Lambert, S. S. Hecht, J. R. Bucher, B. W. Stewart, R. A. Baan, V. J. Cogliano and K. Straif, *Environmental Health Perspectives*, 2016, **124**, 713–721.
30. E. E. Tsermpini, A. Plemenita Ilje and V. Dolan, *Antioxidants*, 2022, DOI: 10.3390/antiox11071374.

31. R. Guo and J. Ren, *International Journal of Environmental Research and Public Health*, 2010, **7**, 1285–1301.
32. Q. Liang, E. C. Carlson, A. J. Borgerding and P. N. Epstein, *The Journal of Pharmacology and Experimental Therapeutics*, 1999, **291**, 766–772.
33. A. Mizumoto, S. Ohashi, K. Hirohashi, Y. Amanuma, T. Matsuda and M. Muto, *International Journal of Molecular Sciences*, 2017, **18**, 1943.
34. E. Quertemont and V. Didone, *Alcohol research & health : the journal of the National Institute on Alcohol Abuse and Alcoholism*, 2006, **29**, 258–65.
35. Eeva Ollila and Clarissa Bingham, *Duodecim*, 2017.
36. Satu Väkeväinen and Mikko Salaspuro, *Duodecim*, 2003, **119**, 1072–1079.
37. M. Salaspuro, *Journal of Digestive Diseases*, 2011, **12**, 51–59.
38. A. Tsuruya, A. Kuwahara, Y. Saito, H. Yamaguchi, N. Tenma, M. Inai, S. Takahashi, E. Tsutsumi, Y. Suwa, Y. Totsuka, W. Suda, K. Oshima, M. Hattori, T. Mizukami, A. Yokoyama, T. Shimoyama and T. Nakayama, *Alcohol and Alcoholism*, 2016, **51**, 395–401.
39. M. T. Nieminen and M. Salaspuro, *Cancers*, 2018, **10**, DOI: 10 . 3390 / cancers10010011.
40. D. F. Wilson and F. M. Matschinsky, *Medical Hypotheses*, 2020, **140**, DOI: 10.1016/j.mehy.2020.109638.
41. M. Correa, E. Acquas and J. D. Salamone, *Frontiers in Behavioral Neuroscience*, 2014, **8**, DOI: 10.3389/fnbeh.2014.00249.
42. E. Quertemont and V. Didone, *Alcohol research & health : the journal of the National Institute on Alcohol Abuse and Alcoholism*, 2006, **29**, 258–265.
43. A. H. Förster and J. Gescher, *Frontiers in Bioengineering and Biotechnology*, 2014, **2**, DOI: 10.3389/fbioe.2014.00016.
44. E. W. Trotter, M. D. Rolfe, A. M. Hounslow, C. J. Craven, M. P. Williamson, G. Sanguinetti, R. K. Poole and J. Green, *PLoS ONE*, 2011, **6**, e25501.
45. A. C. André, L. Debande and B. S. Marteyn, *Cellular Microbiology*, 2021, **23**, DOI: 10.1111/cmi.13338.
46. S. Nath, *Biophysical Chemistry*, 2016, **219**, 69–74.
47. I. M. Keseler, J. Collado-Vides, A. Santos-Zavaleta, M. Peralta-Gil, S. Gama-Castro, L. Muñiz-Rascado, C. Bonavides-Martinez, S. Paley, M. Krummenacker, T. Altman, P. Kaipa, A. Spaulding, J. Pacheco, M. Latendresse, C. Fulcher, M. Sarker, A. G. Shearer, A. Mackie, I. Paulsen, R. P. Gunsalus and P. D. Karp, *Nucleic acids research*, 2011, **39**, 583–90.
48. Y. Zhou and J. A. Imlay, *mBio*, 2022, **13**, DOI: 10.1128/mbio.02965-21.
49. X. Zhang, Z. W. El-Hajj and E. Newman, *Journal of Bacteriology*, 2010, **192**, 5515–5525.

50. D.-F. YANG, Y.-T. WEI and R.-B. HUANG, *Bioscience, Biotechnology, and Biochemistry*, 2007, **71**, 746–753.
51. A. F. Wagner, M. Frey, F. A. Neugebauer, W. Schäfer and J. Knappe, *Proceedings of the National Academy of Sciences of the United States of America*, 1992, **89**, 996–1000.
52. L. J. Moore and P. J. Kiley, *Journal of Biological Chemistry*, 2001, **276**, 45744–45750.
53. G. Unden and J. Schirawski, *Molecular Microbiology*, 1997, **25**, 205–210.
54. A. N. Brown, M. T. Anderson, M. A. Bachman and H. L. T. Mobley, *Microbiology and molecular biology reviews : MMBR*, 2022, **86**, DOI: 10.1128/mmbr.00110-21.
55. G. Aidelberg, B. D. Towbin, D. Rothschild, E. Dekel, A. Bren and U. Alon, *BMC Systems Biology*, 2014, **8**, 133.
56. B. Görke and J. Stülke, *Nature Reviews Microbiology*, 2008, **6**, 613–624.
57. E. M. Ammar, X. Wang and C. V. Rao, *Scientific Reports*, 2018, **8**, 609.
58. G. Sezonov, D. Joseleau-Petit and R. D'Ari, *Journal of Bacteriology*, 2007, **189**, 8746–8749.
59. H. Okano, R. Hermsen and T. Hwa, *Current Opinion in Microbiology*, 2021, **63**, 172–178.
60. M. Zampieri, M. Hörl, F. Hotz, N. F. Müller and U. Sauer, *Nature Communications*, 2019, **10**, 3354.
61. C. N. Peterson, M. J. Mandel and T. J. Silhavy, *Journal of Bacteriology*, 2005, **187**, 7549–7553.
62. F. Arsène, T. Tomoyasu and B. Bukau, *International Journal of Food Microbiology*, 2000, **55**, 3–9.
63. X.-T. Wang, S.-D. Xiao and B.-G. Ma, in *Stress: Genetics, Epigenetics and Genomics*, Elsevier, 2021, pp. 289–296.
64. Y. Ye, L. Zhang, F. Hao, J. Zhang, Y. Wang and H. Tang, *Journal of Proteome Research*, 2012, **11**, 2559–2566.
65. Y. Li, *Natural Product Reports*, 2023, **40**, 922–956.
66. M. Mazzucotelli, B. Farneti, I. Khomenko, K. Gonzalez-Estanol, M. Pedrotti, M. Fragasso, V. Capozzi and F. Biasioli, *Green Analytical Chemistry*, 2022, **3**, 100041.
67. K. Roslund, M. Lehto, P. Pussinen, P.-H. Groop, L. Halonen and M. Metsälä, *Journal of Breath Research*, 2020, **14**, DOI: 10.1088/1752-7163/ab5559.
68. M. O'Hara and C. A. Mayhew, *Journal of Breath Research*, 2009, **3**, 027001.
69. S. Rajendran, I. Khomenko, P. Silcock, E. Betta, F. Biasioli and P. Bremer, *Applied Microbiology*, 2025, **5**, 33.
70. M. K. Gupta and P. K. Biswas, in *Basic Biotechniques for Bioprocess and Bioentrepreneurship*, Elsevier, 2023, pp. 173–182.
71. K. Roslund, M. Lehto, P. Pussinen, K. Hartonen, P.-H. Groop, L. Halonen and M. Metsälä, *Scientific Reports*, 2021, **11**, DOI: 10.1038/s41598-021-96287-7.

72. K. Roslund, M. Uosukainen, K. Järvik, K. Hartonen, M. Lehto, P. Pussinen, P.-H. Groop and M. Metsälä, *Scientific Reports*, 2022, **12**, DOI: 10.1038/s41598-022-26497-0.
73. E. Tait, J. D. Perry, S. P. Stanforth and J. R. Dean, *Journal of Chromatographic Science*, 2014, **52**, 363–373.
74. W. Henry, *Philosophical Transactions of the Royal Society of London*, 1803, **93**, 29–274.
75. E. A. Betterton and M. R. Hoffmann, *Environmental Science & Technology*, 1988, **22**, 1415–1418.
76. P. W. Atkins, J. De Paula and J. Keeler, *Atkins' physical chemistry*, Oxford University Press, Oxford, 12th, 2023, pp. 151–152.
77. F. Wieland, A. Neff, A. N. Gloess, L. Poisson, S. Atlan, D. Larrain, D. Prêtre, I. Blank and C. Yeretzian, *International Journal of Mass Spectrometry*, 2015, **387**, 69–77.
78. X. Zhou and K. Mopper, *Environmental Science & Technology*, 1990, **24**, 1864–1869.
79. H. .-. Benkelberg, S. Hamm and P. Warneck, *Journal of Atmospheric Chemistry*, 1995, **20**, 17–34.
80. L. C. Reimer, J. S. Carbasse, I. Schober, J. Koblitz, A. Podstawka and J. Overmann, *Escherichia coli (Migula 1895) Castellani and Chalmers 1919 (Version 9.3)*, dataset, BacDive, 2025, https://doi.org/10.13145/bacdive4469.20250331.9.3.
81. R. Tagaino, J. Washio, Y. Abiko, N. Tanda, K. Sasaki and N. Takahashi, *Scientific Reports*, 2019, **9**, 10446.
82. J. Nokelainen, MA thesis, Faculty of Science, University of Helsinki, Helsinki, 2025.
83. M. I. Lassenius, C. L. Fogarty, M. Blaut, K. Haimila, L. Riittinen, A. Paju, J. Kirveskari, J. Järvelä, A. J. Ahola, D. Gordin, M.-A. Härma, A. Kumar, S. R. Hamarneh, R. A. Hodin, T. Sorsa, T. Tervahartiala, S. Hörkkö, P. J. Pussinen, C. Forsblom, M. Jauhiainen, M.-R. Taskinen, P.-H. Groop and M. Lehto, *Journal of Internal Medicine*, 2017, **281**, 586–600.
84. G. P. Ferguson and I. R. Booth, *Journal of Bacteriology*, 1998, **180**, 4314–4318.
85. Y. Chen, Y. Chen, S. Tang, B. Tang and S. He, *Food Chemistry: X*, 2024, **24**, DOI: 10.1016/j.fochx.2024.101886.

Appendix A. Strain Information



Leibniz-Institut
DSMZ-Deutsche Sammlung von
Mikroorganismen und Zellkulturen GmbH

DSMZ

Certificate of Origin and Analysis

We declare that the DSMZ cultures within this delivery no. **A2408323-1 (invoice no. 2400008189)**

DSM-No. Strain	Lot.-No.	Risk Group
3925 Escherichia coli	DSM 3925-0414-001	1

√ have been produced in DSMZ laboratories at the address given below
√ are of German preferential origin

For microorganisms:
√ are authentic DSMZ cultures derived directly from these microorganisms held in the DSMZ
√ have been tested by DSMZ control procedures with respect to purity and identity

For bacteriophages only:
√ are authentic DSMZ preparations derived directly from above bacteriophage preparations held in the DSMZ
√ have been tested by DSMZ control procedures with respect to viability and purity according to appropriate bacteriophage methodology

Viability
For microorganisms:
From all microorganisms provided for deposit, viability and purity are tested by subculturing. Each batch produced is again checked for viability after preservation, and after intervals scheduled for the different taxa.

For bacteriophages only:
For all bacteriophages, viability is tested by applying bacteriophage propagation and titer control. Each batch produced is checked for viability and after intervals scheduled for bacteriophages.

Purity
Purity of each batch preserved is checked after production.

For microorganisms:
This check may include microscopical and macroscopical observations or selected physiological, chemosystematic or molecular based tests.

For bacteriophages only:

This check includes analysis of plaque production and plaque morphology on the intended bacterial host and may be supported, if required, by bacteriophage host spectrum, electron microscopical morphology observations or genome sequencing.

The DSMZ is not in a position to verify specific properties or applications claimed for a bacteriophage in the literature. The DSMZ is not responsible for differences between the properties of the bacteriophage deposited in the DSMZ and properties given in the literature/databases.

The DSMZ does not guarantee that bacteriophage preparations are free of prophages that may have been induced from the bacterial hosts during bacteriophage propagation.

Geschäftsführer/
Managing Director:
Prof. Dr. Jörg Overmann; Bettina Fischer
Aufsichtsratsvorsitzender/Head of
Supervisory Board: RD Dr. David Schnieders

Braunschweigische Landessparkasse
(NORD/LB) Kto.-Nr./Account: 2 039 220
BLZ/Bank Code: 250 500 00
IBAN DE22 2505 0000 0002 0392 20
SWIFT (BIC) NOLADE 2 H

Handelsregister/
Commercial Register:
Amtsgericht Braunschweig
HRB 2570
Steuer-Nr. 13/200/24030

Figure A.1: Certificate of origin for *E. coli* DSM 3925.

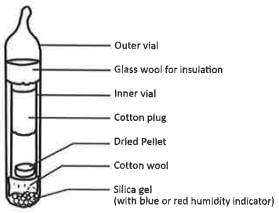
Cultures which were delivered actively growing on agar media or in liquid media

Transfer to fresh media given for that specific strain immediately after receipt. Incubate under the recommended conditions.
The volume of fresh liquid media should not be higher than 10 times the volume of the inoculum.

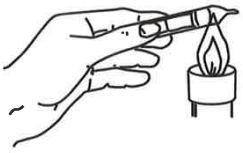
Opening of ampoules and rehydration of dried cultures

Videos may be regarded at www.dsmz.de/collection/catalogue/microorganisms/culture-technology.

1. Remove the glass ampoule carefully from the secondary packaging. Double vial preparation, sealed under vacuum:



2. Wear protective glasses when opening ampoules! Heat the tip of the ampoule in a flame.

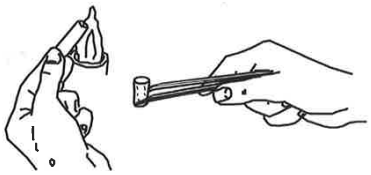


3. Place maximal three drops of water onto the hot tip to crack the glass.

4. Carefully strike off the glass tip with an appropriate tool (e.g. forceps).

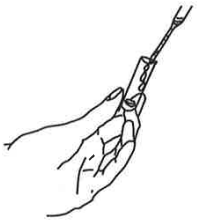
5. Remove the insulation material with forceps and take out the inner vial.

6. Lift the cotton plug using a forceps, remove it, keep it under sterile conditions and flame the top of the inner vial.



7. Add 0.5 ml of medium specified for the strain in the individual strain entry (see above). Replace the plug and allow the pellet to rehydrate for up to 30 minutes.

Subsequent handling of **anaerobic** strains is described on our website as well as in the specific strain entries in the online catalogue. For all other strains proceed as follows.



8. Mix the content gently with an inoculation loop or with a Pasteur pipette. Transfer about half of the whole amount to a test tube with 5 ml of the recommended liquid medium, streak the other half onto a respective agar plate.
(For variations see catalogue strain information.)

9. Incubate liquid and agar cultures under conditions specified for the strain.

10. Before discarding sterilize all the remains of the original ampoule.

Figure A.2: Instructions on how the dried culture obtained from DSMZ was activated.

Appendix B. Scripts

Python script used for the data filtering.

```
import pandas as pd
import numpy as np

# File path
file_path = input("CSV file: ")

# Reading the data
data = pd.read_csv(file_path, sep=';', decimal=',', usecols=['Cycle',
    'm45.03992 (C2H5O) (Conc)'])

# Splitting the data
n_steps = int(input("Number of steps: "))
steps = np.array_split(data, n_steps)

# Dataframe selection function
def select_area(df, start_cycle, end_cycle):
    return df[(df['Cycle'] >= start_cycle) & (df['Cycle'] <=
        end_cycle)]

selected_steps = []
i = 0

while i < len(steps):
    step = steps[i]
    while True:
        try:
            print(f"Step {i+1}/{len(steps)}")
            start_cycle_input = input(f"Start cycle number for step
                {i+1}: ")
            if start_cycle_input.lower() == 'p':
                i -= 1
                selected_steps.pop()
                break
            elif start_cycle_input.lower() == 'q':
                i = len(steps)
                break

            start_cycle = int(start_cycle_input)
```

```
                end_cycle = int(input(f"End cycle number for step {i+1}:
                    "))
                selected_steps.append(select_area(step, start_cycle,
                    end_cycle))
                i += 1
                break

            except ValueError as e:
                print(f"Error: {e}.")

    # Merging the steps
    result = pd.concat(selected_steps)

    # Saving the data
    output_file = input("Output file name: ")
    result.to_csv(output_file, index=False)
```

The General section and Start Step of the automation script.

```
[General]
Timing = 0
FileBase = ""
SaveFolder = "/C/Ionicon/Automation/"
SampleSuffix = FALSE
Repetitions = 0
DoStartStep = TRUE
DoEndStep = TRUE

[StartStep]
Name = "Start Step"
IsSubstep = FALSE
WritePreset = FALSE
Preset = ""
Action = 0
File = ""
WriteTraces = FALSE
TracesCalcFile = ""

[StartStep_Timing]
Duration = 600.000000
StartMode = 0
StartType = 0
StartValue = 1.000000
StopMode = 0
StopType = 0
StopValue = 1.000000
```

```
28      StartIndex = 1
29      StartEdge = 0
30      StopIndex = 1
31      StopEdge = 0
32      StartAS = 0
33      StartGC = 0
34      StopAS = 0
35      StopGC = 0
36
37      [StartStep_Measure]
38      Mode = 3
39      Setting = ""
40      DataSave = 3
41      FileName = "3"
42
43      [StartStep_DO]
44      Mode = TRUE
45
46      [StartStep_DO_1]
47      Index = 3
48      Value = TRUE
49
50      [StartStep_AO]
51      Mode = FALSE
52
53      [StartStep_FC]
54      Mode = TRUE
55
56      [StartStep_FC_1]
57      Name = "FC_FC inlet"
58      Value = 10.000000
59
60      [StartStep_Drift]
61      PressControlled = TRUE
62      PressValue = 177.199997
63      Udrift = 550.000000
64      Udx = 0.000000
65      Unc = 0.000000
66      Press_Enable = TRUE
67      Udrift_Enable = TRUE
68      Udx_Enable = FALSE
69      Unc_Enable = FALSE
70
71      [StartStep_Source]
72      Mode = FALSE
73      On = TRUE
```

```
74      Ihc = 3.300000
75      SV = 49.000000
76      PrimIonIndex = 255.000000
77      Us = 140.000000
78      Uso = 62.000000
79      On_Enable = FALSE
80      Ihc_Enable = TRUE
81      SV_Enable = TRUE
82      PrimIonIndex_Enable = TRUE
83      Us_Enable = TRUE
84      Uso_Enable = TRUE
85      FC_H2O_Enable = TRUE
86      FC_H2O = 5.000000
87      FC_O2_Enable = FALSE
88      FC_O2 = 0.000000
89      FC_NO_Enable = FALSE
90      FC_NO = 0.000000
91      FC_Krypton_Enable = FALSE
92      FC_Krypton = 0.000000
93      FC_Xenon_Enable = FALSE
94      FC_Xenon = 0.000000
95
96      [StartStep_MV]
97      Mode = FALSE
98
99      [StartStep_MV_1]
100     Label = "MPV_1"
101     Direction = 0
102     Position = 0
103
104     [StartStep_MV_2]
105     Label = "MPV_2"
106     Direction = 0
107     Position = 0
108
109     [StartStep_MV_3]
110     Label = "MPV_3"
111     Direction = 0
112     Position = 0
113
114     [StartStep_Temps]
115     T-Drift = 70.000000
116     T-Drift_Enable = TRUE
117     T-Inlet = 70.000000
118     T-Inlet_Enable = TRUE
119     T-Opt1 = 30.000000
```

```
120     T-Opt1_Enable = TRUE
121     T-Opt2 = 0.000000
122     T-Opt2_Enable = FALSE
123
124     [StartStep_AddOns]
125     Num = 0
126
127     [StartStep_GC]
128     Command = 0
129     GC_Mode = 0
130     Temp = 0.000000
131     MeasTime = 0.000000
132     InjectTime = 0.000000
133     FC = 0.000000
134     RampFile = ""
135     Command_Enable = FALSE
136     GC_Mode_Enable = FALSE
137     Temp_Enable = FALSE
138     MeasTime_Enable = FALSE
139     InjectTime_Enable = FALSE
140     FC_Enable = FALSE
141     RampFile_Enable = FALSE
142
143     [StartStep_AS]
144     Command_Enable = FALSE
145     Command = 0
146     ...
```