Investigations in Enhanced Production of Bioalcohols and Their Derivatives Using Experimental and *in-Silico* Approach

A Thesis
Submitted in
Partial Fulfillment of the
Requirements for the Degree of

DOCTOR OF PHILOSOPHY

By

Karan Kumar

School of Energy Science and Engineering

Indian Institute of Technology Guwahati

Guwahati – 781039, Assam, India

January 2024

Dedicated

to

Almighty God, My Parents,

Mentors, Well-Wishers

&

Humanity

School of Energy Science and Engineering
Indian Institute of Technology Guwahati
Guwahati – 781039, India

Statement

I do at this moment declare that the content embodied in this thesis entitled **"Investigations in Enhanced Production of Bioalcohol and Their Derivatives Using Experimental and *In-silico* Approach"** is the result of an investigation carried out by me in the School of Energy Science and Engineering, Indian Institute of Technology Guwahati under the joint supervision of *Prof. Vijayanand S. Moholkar* and *Dr. Lepakshi Barbora*.

In keeping with the general practice of reporting scientific observation, due acknowledgments have been made wherever the work described is based on the findings of other investigators.

January 2024

Karan Kumar
(Roll No: 186151015)
Prime Minister's Research Fellow

School of Energy Science and Engineering
Indian Institute of Technology Guwahati
Guwahati – 781039, India

Certificate

It is certified that the work contained in the thesis entitled **"Investigations in Enhanced Production of Bioalcohol and Their Derivatives Using Experimental and *In-silico* Approach"** by **Karan Kumar** (Roll No. 186151015) has been carried out under our supervision in the School of Energy Science and Engineering, Indian Institute of Technology Guwahati and that this work has not been submitted elsewhere for a degree.

January 2024

Dr. Lepakshi Barbora
Technical Officer (Grade 1)
School of Energy Science and Engineering
Indian Institute of Technology Guwahati
Guwahati – 781039, Assam, India

Prof. Vijayanand. S. Moholkar
HAG Professor
Department of Chemical Engineering
School of Energy Science and Engineering
Indian Institute of Technology Guwahati
Guwahati – 781039, Assam, India

ACKNOWLEDGMENTS

Embarking on my Ph.D. journey at the Indian Institute of Technology Guwahati half a decade ago has evoked a myriad of emotions: excitement, frustration, pride, discouragement, lively discussion, camaraderie, puzzlement, and purpose. Initially, I grappled with comprehending the full value of my Ph.D. pursuit.

Reflecting on this period, I am overwhelmed by the vast knowledge I've acquired, the evolution of my problem-solving skills, and the maturation of my scientific thought. Confronted with challenges, whether in life or research, I've honed the ability to devise practical solutions through my knowledge, skills, or seeking the right resources. Remarkably, I've cultivated a knack for analyzing problems from diverse perspectives, generating crucial exploratory questions.

Regarding my thesis project, I've gained profound insights into *C. acetobutylicum* and *C. pasteurianum* metabolism, delving into both computational and experimental dimensions. Undoubtedly, my current position owes much to the invaluable support of numerous individuals.

First and foremost, I would like to thank my mentors, mentees, and thesis committee. I greatly thank my advisers, ***Prof. Vijayanand S. Moholkar*** and ***Dr. Lepakshi Barbora,*** for their valuable suggestions, encouragement, and constant support throughout my research work. Not to mention that Prof. Moholkar was interested in biofuels and could offer me a project with both experimental and computational aspects! Since then, he has helped me develop skills as a researcher and systems biologist and has offered me numerous networking opportunities for which I am very grateful. Both supervisors have provided an intriguing environment to work with freedom, cultivating my interest in the topic. Their illustrious guidance, effusive cooperation, and encouraging interactions have always driven me to this project.

I sincerely thank my doctoral committee members, ***Prof. Pranab Goswami***, ***Prof. Selvaraju Narayanswami***, and ***Prof. Soumen Kumar Maiti***, for their insightful advice and suggestions throughout the research.

I am grateful to the faculty and staff members of *School of Energy Science and Engineering* for their constant help and support. I want to acknowledge ***the School of***

Energy Science and Engineering, Department of Chemical Engineering, Param-Ishan Supercomputing Facility, and Central Instruments Facilities (CIF) for providing various analytical facilities to conduct my research. I place on record my warm gratitude to the ***Ministry of Education*** (formerly known as the Ministry of Human Resource and Development), Government of India, for awarding me ***Prime Minister's Research Fellowship*** and funding my research work. I am also thankful to the Indian Institute of Technology Guwahati for providing me with the state-of-the-art infrastructure for advanced research.

Special acknowledgments are due to my dearest friends and colleagues. I express my gratitude to the entire VSM Lab for our countless insightful discussions. I appreciate the valuable suggestions from my seniors, including Dr. Malani, Dr. Sarma, Dr. Ingtipi, Dr. Choudhury, Dr. Roy, Dr. Bora, Dr. Mallick, Dr. Batghare, Dr. Kashyap, Dr. Khaire, Dr. Boro, and Dr. Sainik. My heartfelt thanks extend to my colleagues, including Debarshi da, Avinash, Harshendra, Komal, Aradhana, Pushpita, Rishiraj, Ananya, and Umesh, for their supportive collaboration and enthusiastic company. A profound thank you goes out to my friends and fellow Ph.D. students: Bibrita, Ankit, Sudeshna, Shikha, Dr. Prabhat, Dr. Raveesh, Dr. Debanjan, Dr. Rohith, and Brij. Our shared conversations and research stories have enriched this academic journey.

Expressing gratitude seems insufficient to convey my appreciation to my close friends, Dr. Darpan Raghav, Shraddha, Anweshan, Queen, Dr. Ananya, Subodh, Lalbabu, Amresh, Piyush, Niraj, Sulipta, Heena, and Anju. Thank you for your precious companionship, adding joy to the journey, making me laugh during challenging times, and providing unwavering support throughout this remarkable journey.

Finally, I would like to thank and acknowledge the never-ending support of my family. I thank my elder brother, ***Mr. Sunny***; sister-in-law, ***Mrs. Jyoti***; sister, ***Mrs. Priyanka***; and brother-in-law, ***Mr. Niraj***. Last, but not least, I am highly indebted to my parents, ***Mr. Manohar Prasad Choudhary*** and ***Mrs. Anita Devi***, for their unconditional love, care, and sacrifices, for always believing in me, encouraging me to be my best, and for supporting me morally and spiritually throughout my life.

Thank you to everyone; I feel lucky and proud to be where I am today and to have all of you in my life.

Karan Kumar

ABSTRACT

This thesis presents a comprehensive scientific investigation focusing on the challenges and advancements in bioalcohol (mainly biobutanol) synthesis through *Clostridia*, emphasizing enhancing production efficiency from sustainable resources. Employing a multidisciplinary approach, the research encompasses comparative genomic analysis, in-depth exploration of cell central metabolism during ABE fermentation, experimental examination of butanol production from lignocellulosic hydrolysates using *Clostridial* co-culture systems, development of genome-scale metabolic models (GSMs), and innovative investigations into production of value-added bioalcohol using sono-enzymatic processes.

A key finding of the comparative genomic analysis is the successful identification of potential butanol-producing *Clostridium* species with an accuracy of 78%. This result provides insights into the genetic basis of bioalcohol production, providing a solid foundation for formulating potential engineering strategies to optimize biofuel synthesis.

The thesis has also analyzed physiological disparities between *C. acetobutylicum* and *C. pasteurianum* during ABE fermentation have been unveiled. Notably, *C. acetobutylicum* demonstrated superior performance with 25% higher growth rates and an impressive 18% increase in butanol production from various carbon sources. This understanding of coregulatory mechanisms and stress responses provided important inputs for strain optimization for enhanced bioalcohol production from sustainable sources.

The investigation into the inhibition mechanisms of key aldehyde/alcohol dehydrogenases (AADs) in solventogenic species led to the identification of *ρ*-coumaric acid (*ρ*-CA) as the most potent inhibitor. Its binding affinity to seven modeled enzymes highlights *ρ*-CA's significance as a potential target for enhanced biobutanol production, offering avenues for increased yields and cost-effectiveness.

Statistical optimization of Clostridial co-culture for bio-butanol production from mixed substrates resulted in exceptional yields of butanol (12.1 ± 0.45 g/L), biomass (4.15 ± 0.03 OD_{600}), and ABE solvents (23.1 ± 0.55 g/L). The comprehensive Genome

Scale Models have given important insights in *Clostridium* ABE fermentation, the metabolic pathways and potential targets for enhanced bioalcohol production.

The exploration of sonication-induced enhancement in lipase-catalyzed reactions for biodiesel synthesis, increased enzyme activity and faster reaction kinetics were observed. Structural modifications in lipases resulted in a higher yield of n-butyl levulinate (n-BL), opening promising avenues for efficient biodiesel production.

In conclusion, this thesis significantly contributes to the scientific understanding of *Clostridial* bioalcohol synthesis, offering vital knowledge for optimal strain selection, potential engineering strategies, and advancements in sustainable bioenergy production. Essentially, this dissertation has contributed to 3 among the 17 Sustainable Development Goals (SDGs) established by the United Nations in 2015, with the aim of achieving a more sustainable and inclusive future for all: (1) Goal 7 (affordable and clean energy), (2) Goal 11 (sustainable cities and communities), and (3) Goal 13 (climate action).

CONTENTS

LIST OF TABLES

Chapter 4

Chapter 5

LIST OF FIGURES

ABBREVIATIONS & NOTATIONS

Abbreviation	Full Form
AACTCOA	Acetoacetyl CoA
AA	Acetic Acid
AAD	Alcohol/Aldehyde Dehydrogenase
ABE	Acetone-Butanol-Ethanol
ACCA	Acetyl CoA
ANOVA	Analysis of Variance
ATCC	American Type Culture Collection
ATR-FTIR	Attenuated Total Reflection Fourier Transform Infrared
AR	Analytical Reagent
ATP	Adenosine Triphosphate
BP	British Petroleum
BA	Butyric Acid
BBD	Box-Behnken Design
BUOH/BUT	Butanol
C.	*Clostridium*
CALB	*Candida antartica* lipase B
CBP	Consolidated Bioprocessing
CCD	Central Composite Design
CoA	Coenzyme A
DCW	Dry Cell Weight
DoE	Design of Experiments
EMA	Elementary Mode Analysis
Eq.	Equation
F6P	Fructose-6-Phosphate
FAME	Fatty Acid Methyl Ester
FBA	Flux Balance Analysis
Fd	Ferredoxin
G6P	Glucose-6-Phosphate
GA	Growth Associated
GAP	Glyceraldehyde-3-Phosphate

GAP3DH	Glyceraldehyde-3-Phosphate Dehydrogenase
GHG	Greenhouse Gas
HPLC	High-Performance Liquid Chromatography
LA	Levulinic Acid
MTCC	Microbial Type Culture Collection
MDS	Molecular Dynamic Simulations
n-BL	n-Butyl levulinate
NMR	Nuclear Magnetic Resonance
RCM	Reinforced Clostridium Medium
RI	Refractive Index
RMS	Root mean square
RSM	Response Surface Methodology
SHF	Separate Hydrolysis and Fermentation
SSF	Simultaneous Saccharification and Fermentation
SSCF	Simultaneous Saccharification and Co-fermentation
TFS	Total Fermentable Sugar
TLL	*Thermomyces lanuginosus* Lipase
TI	Tetrahedral intermediate
TS	Transition state, as TS-I or TS-II
TRS	Total Reducing Sugar
US	Ultrasound
XRD	X-ray Diffraction

Notation	Description
%	Percentage
MJ kg^{-1}	Megajoule per Kilogram
$	United States (US) Dollar
1G	First Generation
2G	Second Generation
3D	Three Dimensional
g	Gram
L	Liter
h	Hour
g g^{-1}	Gram per Gram
g L^{-1}	Gram per Liter
g L^{-1} h^{-1}	Gram per Liter per Hour
MJ L^{-1}	Megajoule per Liter
w/v	Weight/Volume
v/v	Volume/Volume
°C	Degree Celsius
g	Gravitational Acceleration
M	Molar
mL L^{-1}	Milliliter per Liter
w/w	Weight/Weight
mL min^{-1}	Milliliter per Minute
µL	Microliter
mmol ATP g^{-1} DCW	Millimoles of Adenosine Triphosphate per Gram Dry Cell Weight
K	Kelvin
KDa	Kilo Dalton
KHz	Kilo Hertz
Kpa	Kilo Pascal
v_0	Initial reaction velocity, mol/L·min
V_{max} *or* V_1	Maximum reaction velocity, mol/L·min
k_x	Rate constants, min^{-1}, where x ϵ [1,8]
K_x	Michaelis-Menten constants (where x = *A, B, P, Q*), mol/L

K_{ix}	Dissociation constants, where x = *A, B, P, Q*
K_i	Inhibition constant by n-butanol
$K_{i'}$	Inhibition constant by water molecules
K_{eq}	Equilibrium constant
[*X*]	Concentration of species (*X* = *A, B, P, Q*), mol/L
E	Free enzyme as Novozym 435
*E**	Acyl-enzyme complex
EA	Enzyme-levulinic acid complex
EB	Dead-end complex with n-butanol
EP	Dead-end complex with water
*E*B*	Acyl-enzyme-butanol complex
*E*P*	Acyl-enzyme-water complex
EQ	Enzyme-butyl levulinate complex

CHAPTER 1

Introduction and Literature Review

CHAPTER 1

INTRODUCTION AND LITERATURE REVIEW

1.1. BACKGROUND AND MOTIVATION

A nation's economic growth and development are influenced by its energy resources and management. Countries with abundant domestic energy resources have a competitive edge over those heavily reliant on imports. However, the extensive use of fossil resources leads to energy shortages and exacerbates global warming concerns, emphasizing the need to develop alternative energy sources from renewable materials [1]. India, in particular, has one of the lowest per capita energy consumption levels globally, highlighting the need for energy conservation and adopting sustainable energy sources. The rapid development and increasing demand for comfort and convenience have contributed to rising energy consumption [2, 3].

India's energy consumption is intertwined with its economic progress, and energy indicators are crucial in measuring economic development [4, 5]. With limited resources and increasing energy demands, it is essential to conduct economic feasibility analyses and devise strategies to meet energy needs sustainably [6, 7]. The focus should be on harnessing renewable sources such as wind energy, biofuels, and solar power [8]. The Government of India has made crucial amendments to energy and investment policies to address global warming and energy insecurity and promote sustainable economic growth, including promoting renewable energy, increasing energy efficiency, and investing in renewable energy infrastructure and technology [9, 10].

Furthermore, India has been actively participating in international climate negotiations and collaborating with other countries to address the global challenges of climate change [11].

1.1.1 Liquid Biofuels as an Alternative Fuel Option

Liquid biofuels have gained significant attention as potential alternatives to mitigate carbon dioxide emissions and reduce dependence on fossil fuels [12]. Ethanol has been widely employed as an oxygenate for gasoline and diesel fuels. Still, it suffers from limitations such as low energy content and potential separation from gasoline in the presence of water [13]. On the other hand, butanol (or "biobutanol") has emerged as a promising biofuel option that overcomes many of these limitations [14]. Butanol exhibits closer compatibility with current engines due to its properties, making it an attractive choice for fuel blends [15, 16]. Studies have demonstrated that butanol blends can improve emission parameters in diesel engines [15], and butanol-gasoline blends have shown positive effects on fuel consumption and emissions in gasoline engines [14]. Furthermore, compared to ethanol, butanol is considerably less corrosive, enabling its use in existing fuel infrastructure [17].

1.1.1.1 Comparison of Potential Liquid Biofuels

Table 1.1 compares various alternative liquid fuels, highlighting their common properties. Among the listed fuels, butanol exhibits properties that resemble gasoline's. It can be directly blended with petrol and diesel without additional blending infrastructure, has low vapor pressure for easy transportation, and has a comparable air-to-fuel ratio to gasoline for use in existing engines. Butanol also

exhibits good cold-start capabilities and low solubility in water, reducing the potential for groundwater contamination.

Studies in the Indian context have explored the efficiency and emissions characteristics of engines fueled with gasoline and butanol blends, showing higher energy content, reduced corrosiveness, and improved combustion characteristics of butanol [18, 19]. Investigations into different butanol isomers have indicated improvements in emission levels and combustion efficiency, making it a subject of interest for further research [6, 19]. Additionally, n-butanol has shown potential as a fuel for diesel engines, with reduced soot emissions compared to pure diesel fuel [6].

Table 1.1: Comparison of properties of different alternate liquid fuels [20, 21]

Fuel	Energy density (MJ L^{-1})	Air: fuel ratio	Heat of Vaporization (MJ Kg^{-1})	Research Octane number	Motor Octane number	Cetane number
Gasoline	32	14.6	0.348	91-99	81-89	-
Butanol	29.2	11.2	0.596	96	78	-
Ethanol	19.6	9.0	0.92	129	102	54
Methanol	16	6.5	1.2	136	104	
Biodiesel	31-33	12.5	0.232	-	-	48-65
Jet Fuel	37.6	-	0.348	-	-	-
Hydrogen	8.36	-	0.46	-	-	-

1.1.1.2 Advantages and Potential of Butanol as an Alternative Fuel

The versatility and compatibility of butanol with existing infrastructure make it a favorable option for fuel blending and transportation [17]. While there may be some

trade-offs in power and thermal efficiency compared to traditional fuels, using butanol blends has shown promising results in reducing emissions [19]. Despite the slight fuel consumption penalty, butanol offers advantages such as easy blending, transportability, compatibility with existing engines, good cold start capabilities, and reduced potential for groundwater contamination. **Table 1.2** compares various isomers of butanol with conventional fossil fuels.

To achieve sustainable energy goals, exploring alternative fuels like butanol is crucial to reduce reliance on fossil fuels and addressing emissions challenges [21]. Recent research has evaluated the compatibility and performance of butanol fuel blends, highlighting their impact on emission levels, fuel consumption, and engine performance [14, 16]. These studies have underscored the potential benefits of butanol as a biofuel and emphasized the need for further investigation for widespread use. Overall, butanol holds significant potential as an alternative liquid fuel source, offering advantages over traditional fuels and contributing to a more sustainable energy future.

Table 1.2: Properties of isomers of butanol and conventional fossil fuel [6]

Properties	Diesel	n-Butanol	i-Butanol	s-Butanol	t-Butanol
Molecular weight	198.40	74.11	74.11	74.11	74.11
Cetane number	40-55	12.0	8.5	8.5	5.6
Octane number	20-30	96.0	105.1	108.0	107.0
Boiling point (°C)	149-385	117.4	108.0	99.5	82.4
H/C atomic ratio	1.83	2.50	2.50	2.50	2.50
Low heating value (MJ/kg)	42.93	33.20	32.96	32.90	32.60
Density at 15 °C	842.0	811.5	802.0	806.0	789.0
Oxygen content (%weight)	0.0	21.6	21.6	21.6	21.6

Latent heating at 25°C (KJ/kg)	270	626	684	671	511
Auto-ignition temperature(°C)	200-250	343.0	415.6	406.1	477.8
Stoichiometric ratio	14.30	11.20	11.20	11.20	11.20

1.1.2 Various Routes for Butanol Synthesis

The production of renewable and sustainable chemicals, including butanol, has gained significant attention due to fluctuating crude oil prices, environmental concerns, and the need for petrochemical substitutes [22]. Butanol holds promise as a biofuel and a valuable C4 feedstock for chemical synthesis and acts as a solvent [23]. Conventional chemical processes, such as the oxo process [23] and crotonaldehyde, have been employed for butanol production [22]. However, biological/thermochemical routes (**Figure 1.1**) have also been explored to enhance biobutanol synthesis, productivity, and cost-effectiveness [23].

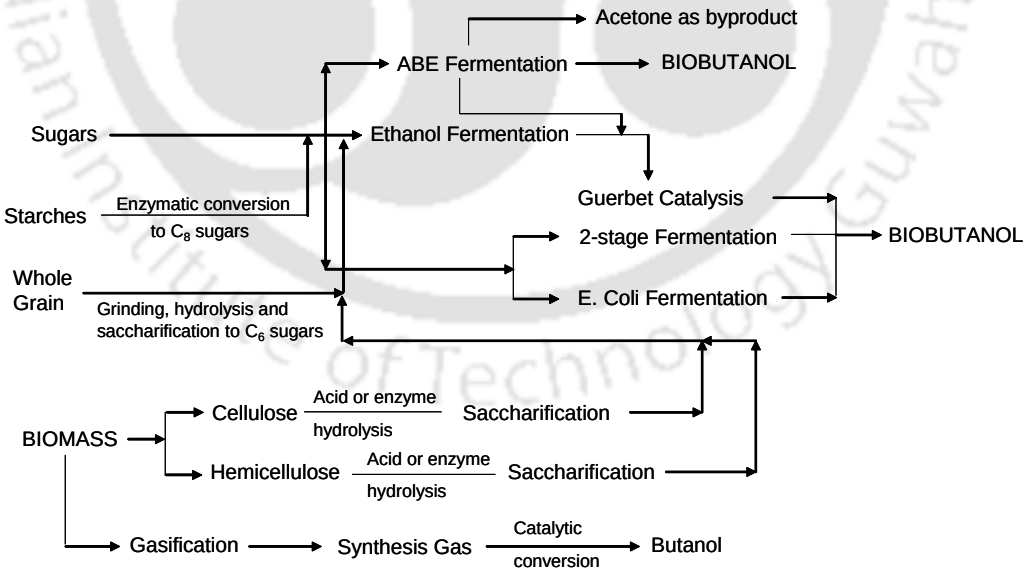


Figure 1.1 Various routes to Biobutanol (redrawn with modifications from ref. [24])

1.1.2.1 Chemical Routes for Butanol Synthesis

The demand for butanol production has grown as a response to the need for sustainable alternatives to fossil fuels. Both biotechnological and chemical processes have contributed to the increasing viability of butanol as a biofuel and chemical compound. Early industrial-scale production relied on the fermentation of the bacterium *Clostridium acetobutylicum*. Still, the demand and growth of the petrochemical industry led to adopting of more efficient chemical methods [25]. **Figure 1.2** shows various chemical routes like Oxo synthesis, Reppe synthesis, and crotonaldehyde hydrogenation processes employed for butanol synthesis [23].

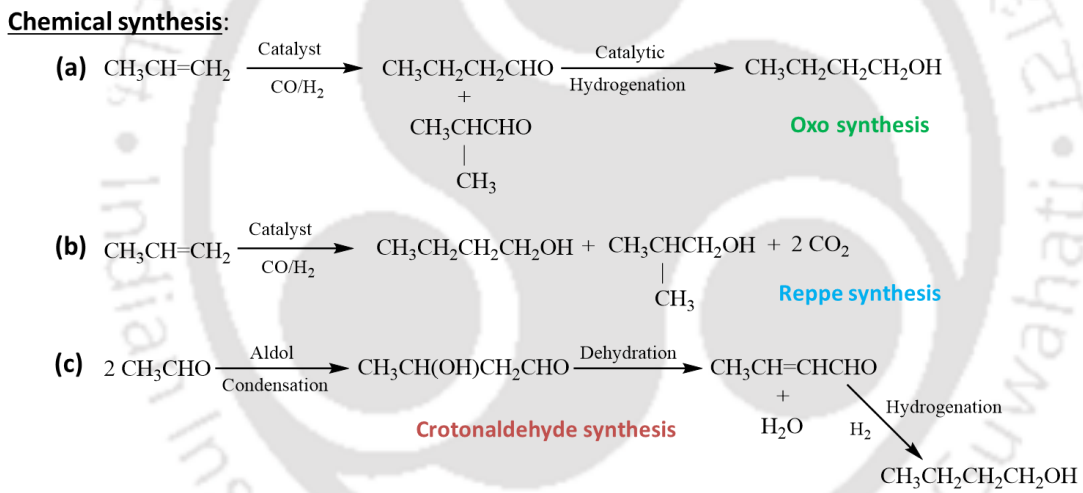


Figure 1.2 Various chemical routes for butanol synthesis [26]

1.1.2.2 Biological Routes for Butanol Synthesis via Fermentation

Butanol can be produced through various biological routes, with acetone-butanol-ethanol (ABE) fermentation by solventogenic *Clostridium* strains being the most popular historical method. Traditional substrates for ABE fermentation include corn or molasses, but alternative substrates have been extensively explored [23, 27]. Despite its potentially greener process, the biological route for butanol production

faces economic feasibility challenges due to drawbacks such as high feedstock costs, substrate inhibition, low cell density, and limited tools for genetic manipulation [26, 28]. Research efforts have focused on enhancing butanol production, developing cost-effective processes, and understanding the physiological characteristics of *Clostridia* [22, 29–31]. Innovative approaches and technologies have been investigated, including vacuum recovery during fermentation and in situ product recovery by adsorption [32, 33]. The goal is to develop sustainable and economically viable processes for butanol production.

A brief history of ABE Fermentation

The ABE fermentation platform was developed in response to the high demand for acetone during World War I [25]. The first industrial-scale ABE fermentations started in 1916 using the Weizmann process, which produced large quantities of acetone and butanol [34]. This process involved using cooked maize mash as the fermentation medium [24]. In 1936, a more economical method was established that used molasses or industrial sugars as carbon sources and operated at a lower temperature [25]. ABE fermentation research rapidly progressed, leading to advancements such as cheaper carbon sources, lower fermentation temperatures, improved utilization of residual sugars, decreased contamination incidents, and increased butanol selectivity [31]. However, the price of carbon sources and advancements in the petrochemical industry led to a decline in the use of ABE fermentation [26]. Consequently, research and ABE fermentation facilities significantly declined, with only a few remaining in countries like China and South Africa [25]. A detailed history of ABE fermentation can be found in the work of Jones and Woods (1986) and Moon et al. (2016) [25, 35]. A brief timeline of notable

events and advances in ABE fermentation from 1860 to 2020 is depicted in **Figure 1.3**.

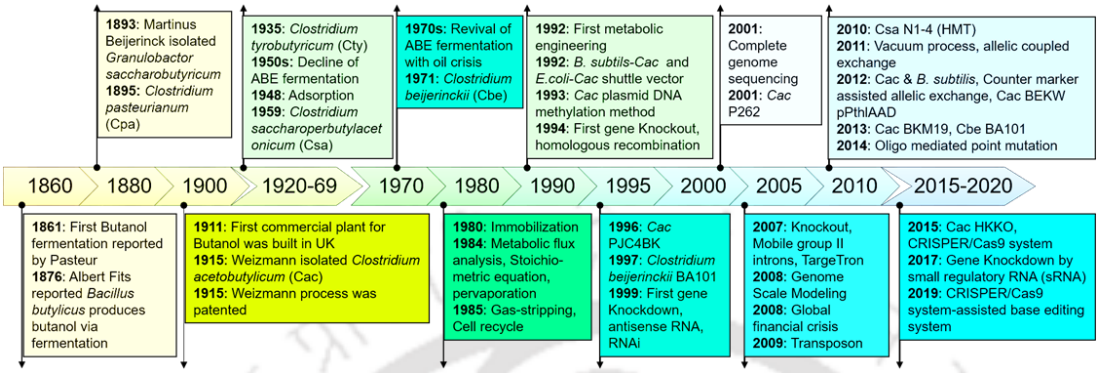


Figure 1.3 Timeline of notable events and advances in ABE fermentation from 1860 to 2020.

1.2 STATE-OF-THE-ART LITERATURE

1.2.1 Applications of *Clostridia* in bioalcohol synthesis: literature study

Bioalcohol production by *Clostridia* has been the subject of numerous studies. *C. beijerinckii*, for example, is known to produce primary alcohols such as 1–butanol and ethanol, and the secondary alcohol 2–propanol [36]. Alcohol dehydrogenase (*ADH*) enzymes have been purified and characterized from 2–propanol-producing strains of *C. beijerinckii*, showing stability and activity in the presence of air [36]. The *ADHs* exhibited NADP(H)-dependent activity and were more active in reducing aldehydes and 2–ketones than primary alcohols [37].

In addition to *C. beijerinckii*, other *Clostridial* species have been investigated for alcohol production. *C. carboxidivorans*, for instance, has been studied for batch syngas fermentation to produce acids and alcohols [38]. The production of alcohols in *Clostridia* relies on the metabolic shift from acidogenesis to solventogenesis, which has been extensively investigated in ABE fermenting *Clostridia* [25, 31, 39].

Genetic modification and engineering approaches have improved alcohol production in *Clostridia* [40, 41].

Studies on alcohol-producing *Clostridia* have provided insights into the metabolic pathways, genetic regulation, and engineering strategies for improving bioalcohol production. These findings contribute to developing sustainable biofuel production processes and utilizing diverse feedstocks in the emerging biorefinery industry [42–45]. **Table 1.3** summarizes bioalcohol-producing *Clostridia*, their carbon source, and fermentation products.

Table 1.3. Bioalcohol-producing *Clostridia*, their carbon source, and products

Species	Carbon Source	Products	References
C. acetobutylicum	Simple Sugars (Sucrose, Glucose, Lactose, Maltose, Galactose, Xylose, Starch, Cellulose, Mannitol)	Acetic and butyric acids, gases (H_2 and CO_2), Acetone, Butanol, and Ethanol	[46]
C. aurantibutyricum	Sucrose, Glucose, Lactose, Maltose, Galactose, Xylose, Starch [but cannot ferment Cellulose, Inulin, Mannitol, Glycerol and Sorbitol]	Acetic and butyric acids, gases (H_2 and CO_2), Butanol and Ethanol	[31, 40]
C. beijerinkii	Simple Sugars and Glycerol	Acetic and butyric acids, gases (H_2 and CO_2), butanol, ethanol, and isopropanol	[47, 48]

C. butyricum	Sucrose, Glucose and Trehalose	Acetic acid, Butyric acid and gases (H_2 and CO_2)	[40, 49]
C. carboxidivorans	Syngas	Acetic acid, Butyric acid, Hexanoic acid, Ethanol and Butanol	[38, 50]
C. cellulolyticum	Glucose, Cellobiose, Cellulose, Xylose	Acetate, Ethanol, Lactate, and gases (H_2 and CO_2)	[46, 51]
C. ljungdahlii	Syngas	Ethanol, Acetate, Formate, Lactate, and Butanediol	[52, 53]
C. pasteurianum	Simple Sugars and Glycerol	Butanol, Ethanol. Acetic acid, Butyric acid and gases (H_2 and CO_2)	[54, 55]
C. tetenomorphum	Glutamate	Acetic and butyric acids, gases (H_2 and CO_2), Butanol and Ethanol	[56]
C. tyrobutyricum	Granulose, Xylose, Glucose	Acetic acid, butyric acid, gases (H_2 and CO_2), Butanol, and Ethanol	[57, 58]

1.2.1.1 Solventogenic Clostridia in Bioalcohol Synthesis

C. acetobutylicum: *C. acetobutylicum* is a gram-positive, endospore-forming obligate anaerobe that is capable of utilizing various carbon sources for ABE fermentation, including glucose, fructose, galactose, sucrose, mannose, starch, dextrin, lactose, mannitol, arabinose, xylose, and pectin [59]. The fermentation process of glucose by *C. acetobutylicum* is biphasic, consisting of an acidogenesis phase followed by a solventogenic phase [31]. During the acidogenesis phase, acetic acid and butyric acid are the primary acids produced. In contrast, in the solventogenic phase, these acids are re-assimilated into neutral solvents such as butanol, acetone, and ethanol [60]. The carbon flow from glucose through the main branches of the pathway leading to acid and solvent formation has been extensively studied [40].

The genes responsible for solvent formation in *C. acetobutylicum* are located on the megaplasmid pSOL1, which contains the sol operon (*adhE-ctfA-ctfB*) encoding a bifunctional butyraldehyde/butanol dehydrogenase and the two subunits of *CoA transferase*, as well as the adjacent monocistronic *adc* operon encoding acetoacetate decarboxylase [61]. The loss of pSOL1 leads to cell degeneration and the inability to produce solvents [40, 43]. Recent studies have focused on the transcriptional regulation of solventogenesis in *C. acetobutylicum*, including the role of the redox-sensing transcriptional repressor Rex and its response to the cellular $NADH/NAD^+$ ratio [62].

Furthermore, the phosphoketolase pathway has been identified as an active pathway for xylose catabolism in *C. acetobutylicum*, contributing up to 40% of the xylose catabolic flux [31]. The phosphoketolase pathway diverts the carbon flow from xylose to acetyl-CoA, a key intermediate in solvent production. Additionally,

studies have investigated the role of intracellular butyryl phosphate (*BuP*) as a regulator of solventogenesis in *C. acetobutylicum*, showing that *BuP* levels correspond to the initiation and flux of butanol formation [63].

Recent advances in genetic engineering techniques, such as CRISPR-Cas9 technology, have also been applied to *C. acetobutylicum* for genome editing and transcriptional engineering, enabling the selection of edited cells and enhancing the screening of mutant cells [64]. The genome sequence of *C. acetobutylicum* ATCC 824 has been determined, revealing significant local conservation of gene order and the presence of operons shared with distantly related bacteria and archaea [59].

Recent studies on *C. acetobutylicum* have provided insights into its metabolic pathways, transcriptional regulation, and genetic engineering potential. These findings contribute to developing more efficient and sustainable ABE fermentation processes for butanol production.

C. beijerinckii: *C. beijerinckii* is a bacterium named after the Dutch microbiologist Martinus Beijerinck [47]. Initially, different names were given to butanol-producing organisms within the *Clostridium genus*, but a consensus was eventually reached in the scientific community [65]. Organisms producing acetone-butanol mixtures were designated *C. acetobutylicum*, while those producing acetone-butanol-isopropanol blends were classified as *C. beijerinckii*. However, it has been reported that some strains of *C. beijerinckii* do not produce isopropanol [66]. *C. beijerinckii* encompasses a diverse group of *Clostridial* strains, including both isopropanol producers and non-producers.

One key distinction between *C. acetobutylicum* and *C. beijerinckii* is the presence of primary/secondary alcohol dehydrogenase enzymes, which catalyze the

conversion of acetone to isopropanol [67]. Another significant difference is the absence of the pSOL1 plasmid in *C. beijerinckii*, which encodes solventogenic genes observed in *C. acetobutylicum*. This absence restricts the degeneration phenomenon in *C. acetobutylicum*, where the organism loses its ability to produce solvents after repeated cell culture [68].

C. beijerinckii has a larger genome than *C. acetobutylicum*, despite similar profiles of solvent production [48]. Both species undergo biphasic fermentation, where acids are initially produced and then re-assimilated into solvents during the later phase [48]. Recent studies have focused on enhancing the tolerance of *C. beijerinckii* to phenolic compounds, which are major limiting factors for solvent production [69]. Additionally, supplementary substrates like formate have been investigated to improve sugar metabolism and solvent production in *C. beijerinckii* [48].

C. pasteurianum: *C. pasteurianum* is a Gram-positive bacterium that has recently gained attention due to its unique electroactivity and potential for bioalcohol synthesis [54]. The metabolic characteristics of *C. pasteurianum* in mixed substrate fermentation have been studied, revealing its ability to utilize substrates like glucose and glycerol to produce acetate, butyrate, 1,3-propanediol (PDO), and n-butanol [70]. Understanding the metabolic pathways and regulation in mixed substrate fermentation is crucial for optimizing bioalcohol synthesis using *C. pasteurianum* [71].

Genetic manipulation of *C. pasteurianum* is challenging due to the lack of available genetic tools, and current strategies for strain improvement rely on random chemical mutagenesis techniques [54, 55]. Establishing an electrotransformation

protocol for *C. pasteurianum* would be a significant advancement in genetic manipulation and strain improvement [72, 73]. Furthermore, electrochemical bioreactors have shown potential in enhancing the electron transfer and metabolic activity of *C. pasteurianum*, leading to improved bioalcohol production [74].

Genomic analysis of *C. pasteurianum* has provided insights into its metabolic capabilities and potential for bioalcohol synthesis [75]. Prophage excision, host defense systems, and central fermentative metabolism have been identified as important factors in butanol production, offering potential targets for strain engineering [55]. Integrating genomic analysis with genetic technologies can further enhance our understanding of *C. pasteurianum* metabolism and facilitate the development of improved strains for bioalcohol synthesis [55].

Overall, *C. pasteurianum* shows promise in bioalcohol synthesis, particularly in butanol production. Its unique characteristics, metabolic capabilities, and utilization of various substrates make it an attractive candidate for sustainable biofuel production. However, further research is needed to optimize the fermentation process, develop genetic tools for strain improvement, and enhance our understanding of *C. pasteurianum* metabolism to achieve efficient bioalcohol synthesis.

1.2.1.2 Cellulolytic Clostridia in Bioalcohol Synthesis

Cellulolytic *Clostridia*, exemplified by *C. cellulolyticum*, play vital roles in the degradation of cellulose, a major component of lignocellulosic biomass (LB), into fermentable sugars, making them attractive candidates for bioalcohol production [76]. Recent studies have advanced our understanding of the physiology and cellulose-adhered life cycle of *C. cellulolyticum*, contributing to their potential in

bioalcohol synthesis [51]. Sporulation, a common feature among cellulolytic *Clostridia*, including *C. cellulolyticum*, has been linked to specific environmental conditions, emphasizing the importance of understanding the factors that trigger sporulation to optimize growth and fermentation processes for bioalcohol synthesis [40].

Genetic manipulation of cellulolytic *Clostridia*, including *C. cellulolyticum*, has historically been challenging due to limited available genetic tools. However, recent progress in transforming related non-cellulolytic *Clostridia* offers new possibilities for strain improvement and metabolic engineering to enhance bioalcohol synthesis [68]. The cellulase system, represented by the cellulosome in cellulolytic *Clostridia*, plays a key role in the efficient degradation of cellulose. Investigating the cellulase complex and cellulose digestion at the metabolic level is crucial for optimizing bioalcohol production [77, 78].

Cellulolytic *Clostridia* are promising candidates for consolidated bioprocessing (CBP), a strategy that combines cellulose hydrolysis and fermentation into a single step, making them attractive chassis cells for bioalcohol synthesis [76]. Recent research has explored the potential of *C. pasteurianum*, a unique electroactive biochemical-producing heterotrophic microorganism, in CBP [64, 74]. Its ability to ferment glycerol, a by-product of the biodiesel industry, into butanol has attracted attention for its potential in biofuel production [70]. Studies have shown that *C. pasteurianum* can produce significant amounts of mixed solvents, including butanol, from various substrates, offering a sustainable and renewable alternative to traditional petrochemical-based processes [31]. Understanding the metabolic pathways and regulation in mixed substrate fermentation is crucial for optimizing bioalcohol synthesis using *C. pasteurianum* [79]. Further research is warranted to

optimize fermentation processes, improve strain engineering, and advance our understanding of *C. pasteurianum*'s capabilities to achieve efficient bioalcohol synthesis within the context of CBP.

Metagenomic studies have also revealed the presence of cellulolytic *Clostridia*, including *C. pasteurianum*, in various microbial communities involved in lignocellulose degradation, providing valuable insights into their cellulolytic capabilities and enzymatic mechanisms [72, 80]. Additionally, understanding the interactions between cellulolytic *Clostridia* and other microbial communities is critical for optimizing bioalcohol synthesis in complex environments, like biogas fermenters [81–83]. Overall, the recent scientific advancements in studying the physiology, sporulation, genetic manipulation, cellulase systems, and metabolic aspects of cellulolytic *Clostridia*, including *C. pasteurianum*, provide valuable insights into their potential for bioalcohol synthesis. Further research is warranted to optimize their fermentation processes, improve strain engineering, and advance our understanding of their cellulolytic capabilities to achieve efficient bioalcohol synthesis.

1.2.1.3 Acetogenic Clostridia in Bioalcohol Synthesis

In recent years, Acetogenic *Clostridia*, such as *C. acetobutylicum* and *C. ljungdahlii*, has garnered significant attention for their potential in bioalcohol synthesis. These bacteria are capable of converting carbon dioxide (CO_2) and carbon monoxide (CO) into valuable biochemicals, including ethanol and butanol [84]. Acetogenic *Clostridia* belong to the phylum Firmicutes and are known for utilizing a wide range of substrates, including sugars, gases, and even waste streams [34].

The metabolic pathways of acetogenic *Clostridia*, such as the Wood-

Ljungdahl pathway (WLP), play a crucial role in their ability to convert C1 gases into bioalcohols [85]. The WLP allows acetogens to grow autotrophically, fixing CO_2 and producing acetyl-CoA, an essential precursor for bioalcohol synthesis [85]. Understanding the genetic and metabolic regulation of the WLP in acetogenic *Clostridia* is critical for optimizing bioalcohol production [85].

Recent studies have focused on improving the efficiency and productivity of acetogenic *Clostridia* in bioalcohol synthesis. Genetic manipulation techniques, such as CRISPR/Cas9-based genome editing, have been developed to engineer these bacteria for enhanced bioalcohol production [86]. Additionally, the optimization of fermentation conditions, such as pH control and trace metal supplementation, has been investigated to improve the performance of acetogenic *Clostridia* in bioalcohol synthesis [87, 88].

Using acetogenic *Clostridia* in gas fermentation has gained attention as a sustainable approach for bioalcohol production. These bacteria can utilize syngas, a CO, CO2, and H2 mixture, as a carbon source for bioalcohol synthesis [89]. Gas fermentation using acetogenic *Clostridia* can convert waste gases from various industries into valuable biofuels and biochemicals, reducing reliance on fossil fuels [89].

The development of biotechnological approaches, such as synthetic biology and metabolic engineering, has further expanded the potential of acetogenic *Clostridia* in bioalcohol synthesis. These approaches enable manipulating metabolic pathways and optimizing cellular processes to enhance bioalcohol production [90]. CRISPR-based base editing techniques have shown promise in reprogramming acetogenic bacteria for improved bioalcohol synthesis [91].

In conclusion, acetogenic *Clostridia* hold great promise in bioalcohol synthesis

due to their unique metabolic capabilities and ability to utilize a wide range of substrates, including C1 gases. Recent research has focused on understanding the genetic and metabolic regulation of acetogenic *Clostridia* and developing genetic manipulation techniques for strain improvement. The optimization of fermentation conditions and the application of biotechnological approaches have also contributed to enhancing bioalcohol production using acetogenic *Clostridia*. Further research is needed to continue improving the efficiency and scalability of bioalcohol synthesis using these bacteria.

1.2.2 Biochemistry of Bioalcohol Synthesis in *Clostridia*

Solventogenic *Clostridia* are a group of anaerobic bacteria that can ferment various substrates, including glucose, fructose, cellulose, and xylose, through biphasic fermentation [44]. **Figure 1.4** depicts the central carbon metabolism in cellulolytic and solventogenic *Clostridia* for ABE fermentation. In the acidogenic phase, organic acids such as acetic and butyric acids are produced. In contrast, in the solventogenic phase, these acids are reassimilated to produce solvents such as acetone, butanol, and ethanol, along with gases like carbon dioxide and hydrogen [44]. Solventogenic *Clostridia* generate only two net molecules of ATP during the glycolytic metabolism of glucose to pyruvate [44]. The acidogenic phase plays a vital role in the metabolism of solventogenic *Clostridia* as the synthesis of organic acids is accompanied by the generation of ATP, which is important for cell growth and metabolism [44].

The metabolism of pyruvate in solventogenic *Clostridia* involves the production of acetic acid and butyric acid, which can be further converted to solvents through specific pathways [31]. For example, *acetoacetyl-CoA* and *butyryl-CoA*, two

precursors of butyric acid, can be directly converted to acetone and butanol, respectively [40]. Acetone is formed by the transfer of the CoA moiety from *acetoacetyl-CoA* to acetate, followed by the decarboxylation of acetoacetate catalyzed by acetoacetate decarboxylase [44]. The production of solvents from pyruvate occurs during the solventogenic phase of growth in batch fermentations, where acids and small amounts of solvents are produced simultaneously [31].

The redox balance is crucial for solventogenesis in *Clostridia*, and ferredoxin plays a significant role in maintaining a low redox potential [40]. Reduced ferredoxin is the physiological electron donor of hydrogenase in solventogenic *Clostridia*, and it is involved in low potential oxidation-reduction reactions [92]. The regeneration of NAD^+ is important for the continuation of glycolysis and the generation of pyruvate in *Clostridia*, and the NADH-dependent or NADPH-dependent ferredoxin reductase is affected by the presence of butanol and butanol toxicity [31].

Cellulolytic *Clostridia*, on the other hand, possess the ability to degrade LB and produce various products, including ethanol, acetate, butyrate, and even butanol [76]. These cellulolytic *Clostridia* offer an alternative approach for CBP, where cellulose degradation and product formation occur in a single microorganism [76]. The combination of cellulolytic and solventogenic *Clostridia* has been explored to develop efficient CBP systems for the production of butanol from LB [44].

Recent studies have focused on improving the performance of solventogenic and cellulolytic *Clostridia* for biofuel production. This includes the use of microbial consortia for butanol production from LB, as well as the application of metabolic and evolutionary engineering to enhance the production of butanol and other solvents [31, 40, 44, 76]. Additionally, advancements in genetic engineering techniques, such

as CRISPR-Cas9, have been applied to manipulate the metabolic pathways and regulatory elements in *Clostridia* to improve solvent production [31, 93].

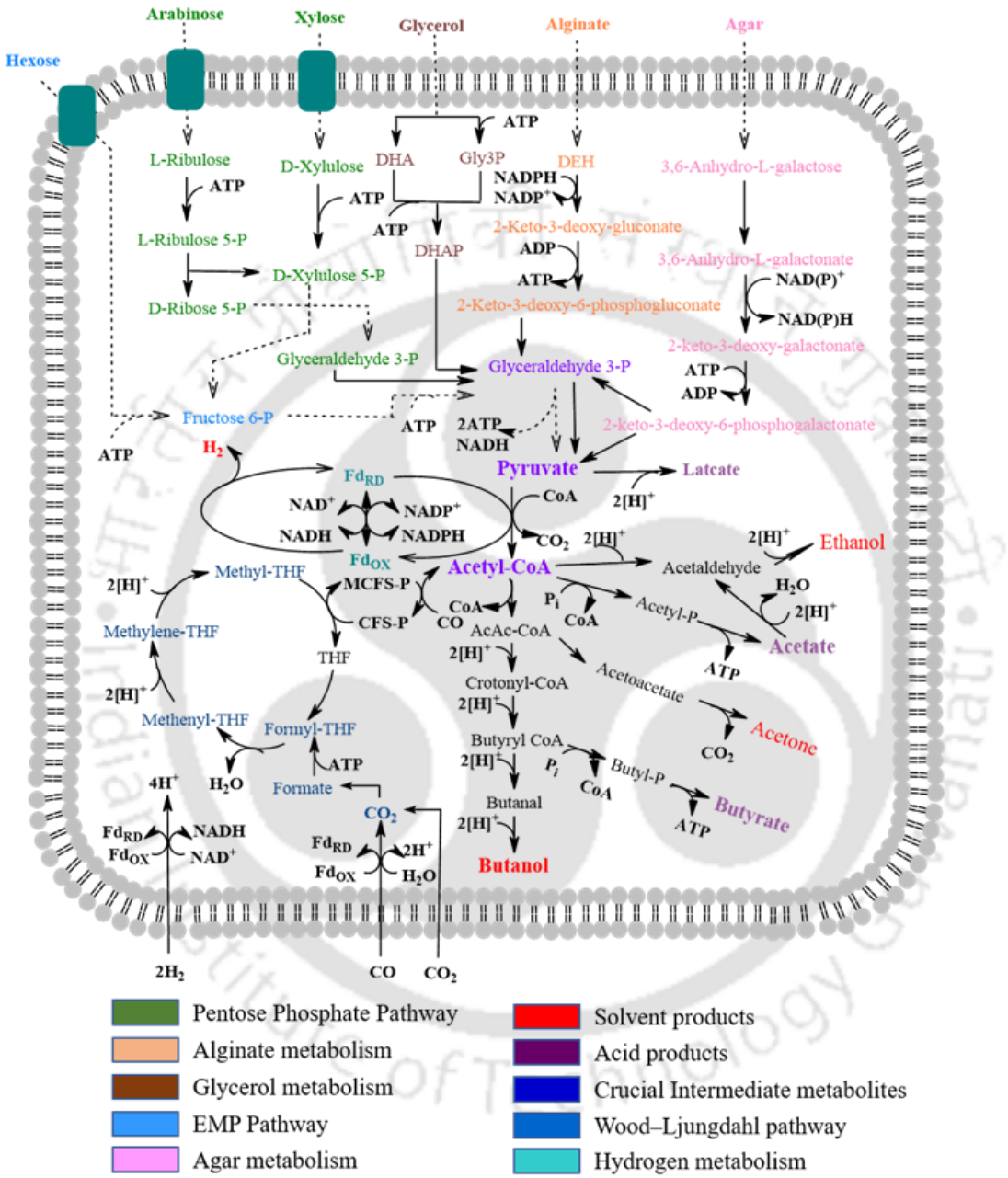


Figure 1.4: Central carbon metabolism in cellulolytic and solventogenic *Clostridia* for ABE fermentation. Colors indicate the metabolic pathways for utilizing various types of substrates in *Clostridia*.

1.2.2.1 Cell Membrane Fluidity, Cell Death, and Solvent Tolerance

Solvent fermentation, particularly butanol production, faces challenges such as low concentration, yield, and productivity, partly due to the stress and toxicity caused by butanol to microbial cells [94]. Butanol stress shares similarities with other forms of stress, such as heat shock stress, which represses the synthesis of cellular proteins while inducing specific heat shock proteins (*hsps*) involved in stress response and adaptation [94]. Solventogenic *Clostridia* respond to butanol stress by expressing heat shock protein genes and modifying the composition of membrane lipids, specifically the saturation of fatty acids [31].

The toxicity of bioalcohols, including butanol, is known to increase with chain length, with longer-chain alcohols being more toxic than shorter-chain ones at lower concentrations [95]. Bioalcohol toxicity can damage the cell membrane and directly inhibit cellular metabolism [94]. Studies have shown that when *C. acetobutylicum* is exposed to butanol, it synthesizes higher amounts of saturated acyl chains of fatty acids, reducing the presence of unsaturated chains [96]. This alteration in lipid composition is a cellular adaptation to counteract the increase in membrane fluidity induced by butanol, which can lead to the leakage of cellular contents and cell death [96].

Efforts to develop butanol-tolerant *Clostridial* mutants have shown mixed results. Inhibition of glycerol dehydrogenase gene expression in *C. beijerinckii* resulted in increased butanol tolerance [97]. Overexpression of the cyclopropane fatty acid synthase gene (*cfa*) in *C. acetobutylicum* initially seemed promising, as it was thought to increase solvent tolerance and production. However, overexpression of the *cfa* gene resulted in changes in membrane properties that led to a decrease in

solvent production, despite increased resistance to acid and butanol in the early log phase [97].

In recent years, studies have focused on understanding the stress response mechanisms, membrane adaptation, and genetic engineering approaches to improve solvent tolerance and production in *Clostridia*. These efforts involve the identification and manipulation of stress-related genes, metabolic engineering of membrane properties, and the use of advanced techniques such as CRISPR-Cas9 for targeted genetic modifications [98–100]. Additionally, the development of mixed fermentation approaches and the exploration of microbial consortia have shown promise in improving solvent production from various substrates, including LB [101].

1.2.2.2 Regulatory Networks in Solventogenic Clostridia

Recent studies [40] have provided updated information on *Clostridia*, particularly regarding sporulation and the regulatory mechanisms involved. Bacterial endospores, including those formed by *Clostridia*, are highly resistant cell types that are produced in response to nutrient starvation or other environmental stresses when alternative survival strategies are insufficient [102]. The sporulation cycle of *C. acetobutylicum*, a solventogenic *Clostridium*, involves the simultaneous occurrence of sporulation, solvent formation, granulose accumulation, and autolysis during the stationary phase [103, 104]. Sporulation and solvent formation in solventogenic *Clostridia* require an adequate carbon source, ATP, and some form of stress [105]. A pictorial depiction of the sporulation cycle in spore forming *Clostridia* is provided in **Figure 1.5**.

The master regulator Spo0A plays a crucial role in regulating both sporulation and solventogenesis in *C. acetobutylicum* [40, 106]. Spo0A phosphorylation controls the expression of genes essential for sporulation initiation, with some genes being activated and others being negatively regulated by phosphorylated Spo0A [106]. Recent studies have identified multiple pathways for Spo0A activation during sporulation initiation in *C. acetobutylicum*, involving histidine kinases encoded by different genes [107]. Disruption of these genes severely impairs sporulation [107].

During the stationary phase, *Clostridia* exhibit downregulation of motility and chemotaxis, which are closely associated with sporulation and are crucial for bacterial survival and adaptation under diverse environmental conditions [108]. These changes in motility and chemotaxis are linked to the regulatory mechanisms involved in sporulation and the transition from exponential growth to stationary phase [108].

In recent years, there have been significant advancements in understanding the sporulation cycle, regulatory networks, and physiological aspects of *Clostridia*. Comprehensive reviews by Paredes et al. (2005) and other researchers provide in-depth information on *Clostridial* sporulation and physiology [40, 63, 105].

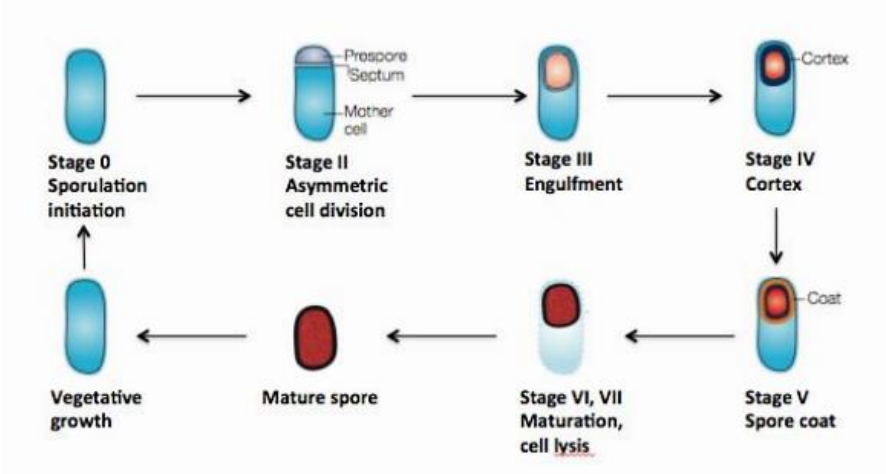


Figure 1.5 The sporulation cycle of spore forming *Clostridia*. The simplified schematic shows only the key stages during the sporulation process. (Modified from [63])

1.2.3 Lignocellulose as a Potential Feedstock

The utilization of agro-industrial waste for bioalcohol production has garnered interest in the field of renewable energies, promoting positive environmental impacts [109]. The development of innovative technologies and processes for the production of butanol from renewable feedstocks has various implications for the future of biofuels [110].

LB, which includes agricultural residues, forestry wastes, waste paper, and energy crops, has long been recognized as a potential sustainable source for the synthesis of various platform chemicals and biofuels [111]. The conversion of lignocellulosic materials into biofuels typically involves three steps: pretreatment to enhance the digestibility of polysaccharides, hydrolysis of cellulose and hemicellulose to fermentable sugars, and fermentation of the sugars to liquid fuels or other products [111].

The majority of commodity plastics and materials are derived from petroleum-based chemicals, highlighting the need for alternative, renewable carbon

sources. LB, as a renewable form of carbon, has been identified as a logical feedstock for the production of biofuels and chemicals [112]. Recent research has focused on methods to synthesize polymers and chemicals from lignin, monolignols, and lignin-derived compounds [112]. These studies explore the structure and processing of LB, as well as the synthesis of polymers and monomers from lignin and its derivatives [112].

Pretreatment is a crucial step in the processing of LB due to its complex and recalcitrant nature [27]. Pretreatment methods aim to improve the accessibility of cellulose and hemicellulose, leading to increased efficiency in subsequent processing stages such as saccharification [27]. The pretreatment of lignocellulosic materials is essential for breaking down the rigid structure and enhancing the efficiency of microbial and enzymatic catalysis [113].

Microbial delignification and hydrolysis of LB have been explored as potential strategies for biofuel production [114]. Biological pretreatments, which are performed at mild temperatures and do not produce inhibitory compounds, have gained attention for their eco-friendly advantages [114]. However, industrial-scale biofuel production from LB is still in the early stages, and further research and process optimization are needed [114].

Enzymatic cocktails play a crucial role in the breakdown of lignocellulose for bioconversion [115]. Lignocellulose is a rich source of fermentable carbohydrates, making it suitable for the production of liquid biofuels, as well as a wide range of chemical commodities and biodegradable materials [115]. The development of efficient enzymatic cocktails is essential for the efficient breakdown of lignocellulose and the production of biofuels and chemicals [115].

LB holds great potential as a source for the synthesis of various platform

chemicals and biofuels. Recent research has focused on pretreatment methods, microbial delignification, enzymatic cocktails, and the synthesis of polymers and chemicals from lignin and its derivatives. Further advancements in these areas are necessary to optimize the efficiency and scalability of lignocellulosic biorefinery processes.

1.2.4 Industrial Fermentation Technologies and Associated Bottlenecks

Industrial fermentation technologies for *Clostridia*, particularly for ABE fermentation, have undergone significant advancements and faced various challenges in recent years. ABE fermentation was once a prominent biotechnological process in industry due to the economic importance of acetone and butanol [116]. Over time, extensive research has been conducted to explore different raw materials, *Clostridia* strains, and fermentation processes for ABE production [116]. However, traditional batch ABE fermentation suffered from several major drawbacks, including high substrate costs, low butanol production, and high product recovery costs [117].

To address these challenges, researchers have focused on improving the fermentation process and feedstock selection. Instead of using conventional carbohydrate substrates like maize and molasses, studies have extensively investigated the use of inexpensive, non-food-based feedstocks such as LB, glycerol, and syngas [117]. Utilizing these alternative feedstocks can significantly reduce the overall cost of ABE fermentation [117]. Additionally, various process modifications have been explored, including fed-batch culture, simultaneous product removal processes (such as gas stripping and pervaporation), multistage continuous fermentation, and cell immobilization and recycle processes [117]. These

modifications aim to enhance productivity, increase butanol titers, and improve overall process efficiency [117].

Continuous fermentation has emerged as a promising approach for ABE production. Continuous fermentation systems, such as packed bed biofilm reactors, have been developed to achieve higher ABE productivity, concentration, and yield compared to batch fermentation [116]. By maintaining acidogenesis and solventogenesis conditions in different reactors, continuous fermentation can achieve improved ABE production rates [116]. Moreover, continuous fermentation coupled with in-situ product removal processes has shown potential for higher productivity and reduced product inhibition [118].

Genetic engineering and metabolic engineering strategies have also been employed to enhance ABE fermentation. Researchers have focused on improving strain performance, increasing butanol tolerance, and reducing by-product formation [119]. Advances in synthetic biology and genetic manipulation techniques have provided new opportunities for optimizing ABE fermentation processes [120]. By manipulating gene expression and metabolic pathways, it is possible to enhance the production of desired products and improve overall process efficiency [120].

In recent years, there has been a growing interest in the use of LB as a feedstock for ABE fermentation. LB, such as agricultural residues and forestry wastes, offers a sustainable and abundant source of carbon for biofuel production [121]. However, the complex structure of LB poses challenges for efficient conversion. Various pretreatment methods, such as hydrothermal pretreatment, have been investigated to enhance the accessibility of cellulose and hemicellulose, leading to improved enzymatic saccharification and subsequent fermentation [122].

Recent studies have focused on addressing the bottlenecks associated with industrial fermentation technologies for *Clostridia*, particularly ABE fermentation. Advances have been made in feedstock selection, process optimization, continuous fermentation, genetic engineering, and the utilization of LB. These advancements aim to improve productivity, increase butanol titers, reduce costs, and make ABE fermentation economically competitive with petrochemical synthesis. Further research and development are still needed to overcome the remaining challenges and achieve sustainable and efficient industrial-scale ABE fermentation.

1.2.5 Strategies for Enhancing Bioalcohol Synthesis in *Clostridia*

Clostridial ABE fermentation is a key process for the production of bioalcohols, such as ethanol and butanol, from various feedstocks including agrobiomass. The optimization of fermentation techniques plays a crucial role in improving bioalcohol yields and overall process efficiency. This section aims to discuss the different aspects of *Clostridial* fermentation optimization, including medium optimization, strain optimization using genetic and metabolic engineering, and the challenges associated with strain optimization techniques.

1.2.5.1 Medium and Process Optimization

Media optimization is a crucial aspect of microbial fermentation for the production of bioalcohols. The composition of the fermentation medium significantly influences the growth, metabolic activity, and ultimately, the bioalcohol yields. By adjusting various factors such as carbon source concentration, nitrogen sources, pH, and the presence of necessary vitamins and minerals, it is possible to create an optimized environment that promotes microbial growth and enhances the efficiency of

bioalcohol production [123]. This review aims to discuss the importance of media optimization and highlight various strategies employed to enhance bioalcohol yields through medium optimization. Techniques like response surface methodology and statistical experimental designs are commonly employed to identify the optimal medium conditions.

***Importance of Media Optimization*:** The choice of fermentation medium and its composition have a significant impact on the overall performance of microbial fermentation for bioalcohol production. The medium provides essential nutrients, energy sources, and growth factors necessary for microbial growth and metabolism. Through media optimization, it is possible to enhance the utilization of substrates, improve microbial growth rates, and maximize bioalcohol yields [124]. Additionally, optimizing the medium can help mitigate the formation of undesirable byproducts and increase the tolerance of microbial strains to inhibitory compounds, thereby improving the efficiency of the fermentation process [125].

1.2.5.2 Systems Biological Approach of Understanding Clostridial Intracellular Regulation

Almost all anaerobic habitats containing organic matter, such as soil types and underwater sediments, support the existence of *Clostridia*, which are primarily saprotrophic bacteria. As a result, they have evolved a remarkable capacity for fermenting compounds organic molecules, such as numerous basic and complex carbohydrates as well as proteins, amino acids [126]. In addition, they generate metabolites like butyrate, acetate, butanol, acetone, acetoin, ethanol, and CO_2 [31, 127, 128]. A wide range of chemicals that can be used as precursors to or as industrial chemicals directly are produced by this subsequent behavioral. As a result,

this organism is crucial for environmental remediation, solvent generation, cellulose breakdown, human and animal health [61, 63, 129]. So, its metabolism (proteins, gene and processes) needs to be understood, and this can be studied by creating GSM models, which require the use of computational and bioinformatics tools [55, 59, 130].

***Development of Genome-Scale Metabolic Models*:** Genome-scale metabolic (GSM) models have been developed for various species of solventogenic and cellulolytic *Clostridia*, providing valuable insights into their metabolic capabilities and potential for bioconversion processes [66]. For example, a GSM model named iCM925 was developed for *C. beijerinckii* NCIMB 8052, which is the first genome-scale metabolic model for this strain [131]. Similarly, GSM models have been constructed for other solventogenic *Clostridia* such as *C. acetobutylicum* ATCC 824 and cellulolytic *Clostridia* like *C. cellulolyticum* and *C. thermocellum* [60, 76, 132–134]. These models have facilitated the understanding of metabolic pathways, flux distributions, and the identification of potential metabolic engineering targets.

The development of GSM models for *Clostridia* has been instrumental in studying their metabolic regulation and identifying key regulatory elements. For example, the effects of orphan histidine kinases on *Clostridial* sporulation progression and metabolism were investigated using a GSM model [135]. The models have also been used to analyze the TCA cycle and pentose-phosphate pathway of *C. acetobutylicum* ATCC 824, providing insights into its metabolic network and activities [136]. Additionally, metabolic network reconstructions and GSM models have been employed to study acetogenic metabolism in *C. ljungdahlii*, aiding in the characterization of its metabolic capabilities [134, 137].

Metabolic engineering strategies for solventogenic *Clostridia* have been explored using GSM models. These models have been used to identify metabolic bottlenecks and propose genetic modifications to enhance butanol production. For instance, thiolase engineering was investigated to improve butanol production in *C. acetobutylicum*, and the effects of the engineered thiolase on metabolic fluxes were analyzed using a GSM model [99]. Metabolic flux analysis and transcriptome analysis have been integrated to identify bottlenecks in butanol formation routes and guide metabolic engineering efforts [45]. The models have also been utilized to study the shift from acidogenesis to solventogenesis in *C. beijerinckii* NCIMB 8052, providing insights into the metabolic regulation during this transition [130, 131, 138].

The development of GSM models for cellulolytic *Clostridia* has also contributed to the understanding of their metabolic capabilities and potential for bioconversion processes. For example, a GSM model was constructed for *C. cellulolyticum*, a model organism for mesophilic cellulolytic *Clostridia*, to study its cellulolytic system and microbial phenomenon of cellulolysis [51]. The models have been used to investigate the metabolic engineering of *C. cellulovorans* for improved n-butanol production, demonstrating the potential of native cellulolytic *Clostridia* as microbial chassis [139]. Furthermore, GSM models have been employed to study the metabolic network of *C. thermocellum*, a thermophilic cellulolytic bacterium, aiding in the understanding of its cellulose hydrolysis and metabolic regulation [133, 140].

In conclusion, the development of genome-scale metabolic models has significantly contributed to the understanding of the metabolic capabilities and regulation of solventogenic and cellulolytic *Clostridia*. These models have facilitated the identification of metabolic engineering targets, analysis of metabolic pathways,

and exploration of metabolic regulation. They have been instrumental in guiding strain development and optimization of bioconversion processes. Continued advancements in GSM models, coupled with experimental data integration and omics analyses, hold great potential for further enhancing the metabolic engineering of *Clostridia* for various biotechnological applications.

1.3 MAJOR GAPS IN THE RESEARCH AREA

Based on the extensive literature survey presented in this chapter, the major research gaps in the enhancement of bioalcohol synthesis through *Clostridia* can be summarized as follows:

i. **Understanding metabolic pathways and regulation mechanisms**: Further research is needed to elucidate the metabolic pathways and regulatory mechanisms involved in bioalcohol synthesis in *Clostridia* [84, 141]. This includes studying the enzymes, genes, and transcriptional factors that play a role in the production of bioalcohols [141].

ii. **Optimization of fermentation conditions**: Research is needed to optimize the fermentation conditions for bioalcohol synthesis, including temperature, pH, substrate concentration, and nutrient availability [142, 143]. This can help improve the efficiency and productivity of bioalcohol production by *Clostridia*.

iii. **Genetic engineering and metabolic engineering**: Further advancements in genetic engineering and metabolic engineering techniques are required to enhance the bioalcohol synthesis capabilities of *Clostridia* [42, 144]. This includes the development of genetic systems and tools for manipulating the metabolic pathways involved in bioalcohol production [42, 144].

iv. **Substrate utilization and feedstock diversification**: Research is needed to explore the utilization of different substrates and feedstocks for bioalcohol synthesis by *Clostridia* [23, 145]. This includes investigating the use of LB, waste materials, and C1 gases as alternative carbon sources for bioalcohol production [23, 145].

v. **Process optimization and scale-up**: There is a need for research on process optimization and scale-up strategies for bioalcohol synthesis using *Clostridia* [146, 147]. This includes the development of efficient fermentation processes, downstream separation techniques, and recovery methods for bioalcohols [146, 147].

vi. **Environmental and economic sustainability**: Further research is required to assess the environmental and economic sustainability of bioalcohol synthesis using *Clostridia* [23, 148]. This includes evaluating the life cycle impacts, energy efficiency, and cost-effectiveness of bioalcohol production processes [23, 148].

vii. **Integration of different disciplines**: How can interdisciplinary collaboration specifically integrating system biology, computational biotechnology, sustainable chemistry, bioprocess engineering, and environmental science contribute to optimizing and addressing specific challenges in bioalcohol synthesis through Clostridia [149, 150], with a focus on enhancing overall production efficiency and sustainability?

By addressing these research gaps, it is possible to enhance the efficiency, productivity, and sustainability of bioalcohol synthesis using *Clostridia*, contributing to the development of bio-based alternatives to fossil fuels and chemicals.

1.4 RESEARCH PROBLEM & OBJECTIVES

1.4.1 Problem definition

The aim of this research is to address the gaps and challenges in enhancing bioalcohol synthesis through *Clostridia*. The objectives of this study include comparative genomic analysis of potential butanol-producing *Clostridium* species, investigation of cell central metabolism for a better understanding of ABE fermentation, study of the challenges and scope of butanol production from lignocellulosic hydrolysates using *Clostridial* co-culture systems, development and application of genome-scale metabolic models (GSMs) for understanding *Clostridium* ABE fermentation and devising metabolic engineering strategies, and investigations in enhanced production of bioalcohol derivatives using sono-enzymatic processes.

Despite the progress made in bioalcohol synthesis, there are several research gaps that need to be addressed. Firstly, there is a need for a comprehensive comparative genomic analysis of potential butanol-producing *Clostridium* species to identify the best solventogenic and cellulolytic strains. This will provide insights into the genetic basis of bioalcohol production and help in strain selection and engineering.

Secondly, a deeper understanding of the cell central metabolism of *Clostridium* during ABE fermentation is required. This includes studying the biochemistry and physiology of the fermentation process to identify key metabolic pathways and regulatory mechanisms that can be targeted for optimization.

Thirdly, the challenges and potential of butanol production from lignocellulosic hydrolysates using *Clostridial* co-culture systems need to be investigated. This involves studying the interactions between different *Clostridium* species and their ability to efficiently convert the complex sugars present in LB into bioalcohols.

Furthermore, the development and application of genome-scale metabolic models (GSMs) can provide a systems-level understanding of *Clostridium* ABE fermentation. These models can help in predicting metabolic fluxes, identifying metabolic bottlenecks, and devising metabolic engineering strategies for higher yields of biobutanol.

Lastly, investigations in enhanced production of bioalcohol derivatives using sono-enzymatic processes are needed. This involves exploring the use of ultrasound-assisted enzymatic reactions for the synthesis of bioalcohol derivatives, both experimentally and computationally.

By addressing these research gaps, this study aims to contribute to the advancement of bioalcohol synthesis through *Clostridia*, leading to more efficient and sustainable production processes.

1.4.2 Objectives

Despite the limited scope, the main objective of this thesis is to make contribution in addressing these challenges. With this in mind, the following objectives have been formulated:

- Comparative genomic analysis of potential butanol-producing *Clostridium sp.* to identify best solventogenic and cellulolytic *Clostridia*

- Investigation of cell central metabolism for a better understanding of the biochemistry and physiology of ABE fermentation
- Study of the challenges and scope of butanol production from lignocellulosic hydrolysates using *Clostridial* co-culture system
- Development and applications of GSMs for understanding *Clostridium* ABE fermentation and devising metabolic engineering strategies for higher yields of biobutanol
- Investigations in enhanced production of derivatives of bioalcohols using sono-enzymatic processes: an experimental and computational approach

1.5 DISSERTATION OUTLINE

On the basis of the above discussion, the thesis is organized eight chapters. The brief summaries of the contents of these chapters are given below:

Chapter 1: This chapter initially discusses the introduction and motivation of the chosen problem statement, followed by an extensive discussion on the background and fundamentals of the *Clostridial* life cycle and metabolism, including ABE fermentation, along with the application of comparative genomic analysis, metabolic network reconstruction and simulations in understanding microbial physiology. Subsequently, a brief review of the state–of–the–art literature on various promising strategies for enhancing biobutanol production from lignocellulosic biomass has been presented. Finally, the broad scope, significant research gaps, formulation of hypotheses, and primary objectives of the present thesis have been outlined.

Chapter 2: This chapter presents the genomic differences among fully sequenced 48 bacterial genomes from 7 genera (viz. *Acetivibrio, Bacillus, Cellulosilyticum, Clostridium, Desulfotomaculum, Lachnoclostridium, Moorella, Ruminiclostridium,*

and *Thermoanaerobacterium*) using comparative genomic analysis which include pan-genomic analysis, comparison of central carbon metabolism, general genome features, various genes and their number in COGs (cluster of orthologous groups), CDS of the energy, metabolic, and sporulation pathways. Our findings result in a comprehensive whole genome-based phylogenetic relationship of 48 strains in solventogenic, cellulolytic, acetogenic, and/or ethanologenic classification.

Chapter 3: This chapter presents the research on the characterization of *C. acetobutylicum* (*Cac*) and *C. pasteurianum* (*Cpa*) using biochemical and systems biology techniques. In biochemical characterization, we accessed the metabolizing capabilities of *C. acetobutylicum* MTCC 11274 and *C. pasteurianum* MTCC 116 on modified media optimized for butanol productivity (MoBP). The MoBP comprises (g L^{-1}) glucose: 80.0, peptone: 49.7, KH_2PO_4: 0.5, K_2HPO_4: 0.5, $MgSO_4.7H_2O$: 0.46, $FeSO_4.7H_2O$: 0.023, $MnSO_4.H_2O$: 0.023, NaCl: 5, $CH_3COONH_4^+$: 2.2, para-amino-benzoic acid: 0.01, biotin: 0.01, and L-Cystiene-HCl: 0.05. Next, we characterized their growth cycle and fermentation profiles, focusing on ABE production capabilities. In systems biological approach, genome sequence of *C. acetobutylicum* ATCC 824 and *C. pasteurianum* ATCC 6013 were retrieved from database and whole genome comparison were performed for a deeper understanding of the cell central metabolism of during ABE fermentation.

Chapter 4: This work has aimed at the deduction of the inhibition mechanism of essential *aldehyde/alcohol dehydrogenase* (*AAD*) in three *Clostridia* (*Cac*, *Cpa*, and *C. beijerinckii*) using molecular mechanics (MM) simulations. We first identified seven *Clostridial AAD* enzymes and modelled and validated their 3D structure using MODELLER and PROCHECK, respectively. After modelling and validation, each enzyme model was subjected to CASP14-ranked web servers, such as PrankWeb,

InterPro, 2Struc, 2StructCompare, PsiPred, and NetSurfP3.0, for various secondary structural and binding pocket analyses. Parallelly, the chemical structures of 17 ligands comprising 10 FI and seven substrates were designed in GaussView v6.0 software and geometrically optimised in Gaussian16 standalone software using density functional theory (DFT) at B3PW91/6-311G(+)(d, p) level of theory followed by frequency calculation (freq = raman). This step was followed by profiling molecular interactions between FIs and AADs using molecular docking and dynamics simulations.

Chapter 5: This chapter presents statistical optimization of the co-culture of *Cac* MTCC 11274 and *Cpa* MTCC 116 for enhancing bio-butanol production from mixed substrates using response surface methodology. The analysis of the effect of different process parameters, such as (1) ratio of Cac and Cpa inoculum, (2) NaCl concentration, and (3) ratio of xylose and glucose, indicated that salt concentration is the most critical factor for the production of biomass, total alcohol, and butanol. The media defined in Chapter 3 was utilized to carry out the optimization experiment with only variation in concentration of glucose and NaCl.

Additionally, we reconstructed and validated two new genome-scale metabolic (GSM) models, *iKK1425* (for *Cac* ATCC 824) and *iKK848* (for *Cpa* ATCC 6013). These GSMs are the most comprehensive compared to previous GSMs. Model quality and metabolic flux optimization for biomass growth using *iKK1425* and *iKK848* are compared with previous literature. To investigate the mechanism of this synergistic enhancement of butanol and ABE production in experimental section of this chapter, we performed computational simulation at similar conditions in our GSMs, iKK1425 and iKK848.

Chapter 6: Ultrasound (US) has become an efficient green technology for intensifying lipase-catalyzed processes for synthesizing renewable fuels and their derivatives. However, the links between US-induced alteration in the structure of lipases and its activity enhancement (which essentially is manifested in faster reaction kinetics) have yet to be established. In this chapter we have developed a ternary approach to provide mechanistic insight into the influence of sonication on lipase-catalyzed reaction. This includes determining the physical and chemical characteristics of the binding pockets of lipases, deducing their locations in different structural motifs and molecular docking simulations with ligands.

Chapter 7: This chapter reports US-assisted enzymatic synthesis of n-butyl levulinate (n-BL) via esterification of levulinic acid (LA) with n-butanol using Novozym 435 (CALB immobilized on microporous polyacrylic resin). Esterification parameters were optimized with the statistical design of experiments. Application of 35 kHz sonication boosted n-BL yield from 70.9% to 92.22%. To provide mechanistic insight into the influence of sonication, we employed Ping-Pong Bi-Bi kinetics model and the ternary approach developed in the Chapter 6. Mechanistic investigation using Ping-Pong Bi-Bi kinetics model and ternary approach revealed an interesting influence of sonication on esterification reaction. The exposure to sonication enhanced reaction velocity and substrate affinity with reduced inhibition and unfavorable dissociation of intermediate complexes, ultimately enhancing n-BL yield.

Chapter 8: This chapter provides an overview of all studies in this thesis. Thesis presents an in-depth investigation of an efficient and sustainable approach to enhance the production of bioalcohols (ethanol and butanol) and their derivatives (butyl levulinates) using experimental and in-silico techniques. The merits and

shortfalls of these approaches have been discussed. Based on the results of the present study, scope for future research in the area has been outlined.

REFERENCES

1. World Energy Outlook 2022 shows the global energy crisis can be a historic turning point towards a cleaner and more secure future - News. In: IEA. https://www.iea.org/news/world-energy-outlook-2022-shows-the-global-energy-crisis-can-be-a-historic-turning-point-towards-a-cleaner-and-more-secure-future. Accessed 19 Jun 2023

2. Palamalai S, Ravindra IS, Prakasam K (2015) Relationship between Energy Consumption, CO_2 Emissions, Economic Growth and Trade in India. J Econ Financ Stud 3:01. https://doi.org/10.18533/jefs.v3i02.93

3. Attivissimo F, Di Nisio A, Lanzolla AML, Paul M (2015) Feasibility of a Photovoltaic–Thermoelectric Generator: Performance Analysis and Simulation Results. IEEE Trans Instrum Meas 64:1158–1169. https://doi.org/10.1109/TIM.2015.2410353

4. India Energy Outlook 2021 – Analysis. In: IEA. https://www.iea.org/reports/india-energy-outlook-2021. Accessed 19 Apr 2023

5. (2023) India surpasses China to become world's most populous nation with 142.86 cr people: UN. Econ. Times

6. He Z, Liu G, Li Z, et al (2019) Comparison of four butanol isomers blended with diesel on particulate matter emissions in a common rail diesel engine. J Aerosol Sci 137:105434. https://doi.org/10.1016/j.jaerosci.2019.105434

7. Nations U The Role of Fossil Fuels in a Sustainable Energy System. In: U. N. https://www.un.org/en/chronicle/article/role-fossil-fuels-sustainable-energy-system. Accessed 19 Jun 2023

8. Varjani S, Pandey A, Bhaskar T, et al (2021) Biomass, Biofuels, Biochemicals: Circular Bioeconomy: Technologies for Biofuels and Biochemicals. Elsevier

9. Ahluwalia S India's Green Hydrogen Policy: Tentative beginnings. In: ORF. https://www.orfonline.org/expert-speak/indias-green-hydrogen-policy/. Accessed 7 Jun 2022

10. www.ETEnergyworld.com India's Green Hydrogen Policy: Unprecedented growth needed to achieve 2030 targets - ET EnergyWorld. In: ETEnergyworld.com. https://energy.economictimes.indiatimes.com/news/renewable/indias-green-hydrogen-policy-unprecedented-growth-needed-to-achieve-2030-targets/90154361. Accessed 7 Jun 2022

11. How does India taken a leadership role in tackling climate change? https://blog.mygov.in/editorial/how-does-india-taken-a-leadership-role-in-tackling-climate-change/. Accessed 19 Apr 2023

12. Dürre P (2011) Fermentative production of butanol—the academic perspective. Curr Opin Biotechnol 22:331–336. https://doi.org/10.1016/j.copbio.2011.04.010

13. Mali NA, Yadav SS, Ghuge PD, Joshi SunilS (2017) Vapor–Liquid Equilibrium Data for Binary Mixtures of Dimethyl Carbonate with Methyl Acetate, Ethyl Acetate, *n* -Propyl Acetate, Isopropyl Acetate, *n* -Butyl Acetate, and Isoamyl Acetate at 93.13 kPa. J Chem Eng Data 62:4356–4363. https://doi.org/10.1021/acs.jced.7b00704

14. Liu H, Wang X, Zhang D, et al (2019) Investigation on Blending Effects of Gasoline Fuel with N-Butanol, DMF, and Ethanol on the Fuel Consumption and Harmful Emissions in a GDI Vehicle. Energies 12:1845. https://doi.org/10.3390/en12101845

15. Kuszewski H (2018) Physical and Chemical Properties of 1-Butanol–Diesel Fuel Blends. Energy Fuels 32:11619–11631. https://doi.org/10.1021/acs.energyfuels.8b02912

16. Yogesh P (2022) Effect of Exhaust Gas Recirculation on a CRDI Engine Fuelled with Biodiesel, Ethanol and Butanol. Int J Veh Struct Syst 14:. https://doi.org/10.4273/ijvss.14.5.26

17. Rahayu SMN, Hananto AL, Herawan SG, et al (2022) A Review of automotive green technology: Potential of butanol as biofuel in gasoline engine. Mech Eng Soc Ind 2:82–97. https://doi.org/10.31603/mesi.7155

18. Nithyanandan K, Zhang J, Li Y, et al (2016) Improved SI engine efficiency using Acetone–Butanol–Ethanol (ABE). Fuel 174:333–343. https://doi.org/10.1016/j.fuel.2016.01.001

19. Li Y, Meng L, Nithyanandan K, et al (2016) Combustion, performance and emissions characteristics of a spark-ignition engine fueled with isopropanol-n-butanol-ethanol and gasoline blends. Fuel 184:864–872. https://doi.org/10.1016/j.fuel.2016.07.063

20. Pugazhendhi A, Mathimani T, Varjani S, et al (2019) Biobutanol as a promising liquid fuel for the future - recent updates and perspectives. Fuel 253:637–646. https://doi.org/10.1016/j.fuel.2019.04.139

21. Surakasi R, Srinivasa Rao Y, Kalam SdA, Begum N (2023) Emissions and Performance of Diesel Engines Correlated with Biodiesel Properties. J Eng 2023:1–4. https://doi.org/10.1155/2023/5274325

22. Choi H, Han J, Lee J (2021) Renewable Butanol Production via Catalytic Routes. Int J Environ Res Public Health 18:11749. https://doi.org/10.3390/ijerph182211749

23. Kolesinska B, Fraczyk J, Binczarski M, et al (2019) Butanol Synthesis Routes for Biofuel Production: Trends and Perspectives. Materials 12:350. https://doi.org/10.3390/ma12030350

24. (2012) Biobutanol – A Replacement for Bioethanol? https://www.aiche.org/resources/publications/cep/2008/august/biobutanol-%E2%80%93-replacement-bioethanol. Accessed 3 Jun 2023

25. Moon HG, Jang Y-S, Cho C, et al (2016) One hundred years of *Clostridial* butanol fermentation. FEMS Microbiol Lett fnw001. https://doi.org/10.1093/femsle/fnw001

26. Ndaba B, Chiyanzu I, Marx S (2015) n-Butanol derived from biochemical and chemical routes: A review. Biotechnol Rep 8:1–9. https://doi.org/10.1016/j.btre.2015.08.001

27. Kucharska K, Rybarczyk P, Hołowacz I, et al (2018) Pretreatment of Lignocellulosic Materials as Substrates for Fermentation Processes. Molecules 23:2937. https://doi.org/10.3390/molecules23112937

28. Carrillo-Nieves D, Rostro Alanís MJ, de la Cruz Quiroz R, et al (2019) Current status and future trends of bioethanol production from agro-industrial wastes in Mexico. Renew Sustain Energy Rev 102:63–74. https://doi.org/10.1016/j.rser.2018.11.031

29. Patraşcu I, Bîldea CS, Kiss AA (2017) Eco-efficient butanol separation in the ABE fermentation process. Sep Purif Technol 177:49–61. https://doi.org/10.1016/j.seppur.2016.12.008

30. Xin F, Chen T, Jiang Y, et al (2017) Strategies for improved isopropanol–butanol production by a *Clostridium* strain from glucose and hemicellulose through consolidated bioprocessing. Biotechnol Biofuels 10:118. https://doi.org/10.1186/s13068-017-0805-1

31. Li S, Huang L, Ke C, et al (2020) Pathway dissection, regulation, engineering and application: lessons learned from biobutanol production by solventogenic *Clostridia*. Biotechnol Biofuels 13:39. https://doi.org/10.1186/s13068-020-01674-3

32. Mariano AP, Qureshi N, Maciel Filho R, Ezeji TC (2012) Assessment of in situ butanol recovery by vacuum during acetone butanol ethanol (ABE) fermentation. J Chem Technol Biotechnol 87:334–340. https://doi.org/10.1002/jctb.2717

33. Xue C, Liu F, Xu M, et al (2016) Butanol production in acetone-butanol-ethanol fermentation with in situ product recovery by adsorption. Bioresour Technol 219:158–168. https://doi.org/10.1016/j.biortech.2016.07.111

34. Fu H, Yang S-T (2022) Editorial: Development and Application of *Clostridia* as Microbial Cell-Factories for Biofuels and Biochemicals Production. Front Bioeng Biotechnol 9:

35. Jones DT, Woods DR (1986) Acetone-butanol fermentation revisited. Microbiol Rev 50:484–524

36. Ismaiel AA, Zhu CX, Colby GD, Chen JS (1993) Purification and characterization of a primary-secondary alcohol dehydrogenase from two strains of *Clostridium beijerinckii*. J Bacteriol 175:5097–5105. https://doi.org/10.1128/jb.175.16.5097-5105.1993

37. Cho C, Hong S, Moon HG, et al (2019) Engineering *Clostridial* Aldehyde/Alcohol Dehydrogenase for Selective Butanol Production. mBio 10:e02683-18. https://doi.org/10.1128/mBio.02683-18

38. Lanzillo F, Ruggiero G, Raganati F, et al (2020) Batch Syngas Fermentation by *Clostridium carboxidivorans* for Production of Acids and Alcohols. Processes 8:1075. https://doi.org/10.3390/pr8091075

39. Phillips J, Huhnke R, Atiyeh H (2017) Syngas Fermentation: A Microbial Conversion Process of Gaseous Substrates to Various Products. Fermentation 3:28. https://doi.org/10.3390/fermentation3020028

40. Diallo M, Kengen SWM, López-Contreras AM (2021) Sporulation in solventogenic and acetogenic *Clostridia*. Appl Microbiol Biotechnol 105:3533–3557. https://doi.org/10.1007/s00253-021-11289-9

41. Wen Z, Tu Z, Makishah NHA, et al (2022) Genetic Manipulation of a Twin *Clostridia* Consortium for Co-production of *n*-Butanol and Isobutanol by Consolidated Bioprocessing. ACS Sustain Chem Amp Eng 10:. https://doi.org/10.1021/acssuschemeng.1c08486

42. Liang L, Liu R, Freed EF, Eckert CA (2020) Synthetic Biology and Metabolic Engineering Employing *Escherichia coli* for C_2–C_6 Bioalcohol Production. Front Bioeng Biotechnol 8:710. https://doi.org/10.3389/fbioe.2020.00710

43. Dash S, Ng CY, Maranas CD (2016) Metabolic modeling of *Clostridia*: current developments and applications. FEMS Microbiol Lett 363:fnw004. https://doi.org/10.1093/femsle/fnw004

44. Jiang T, Li C, Teng Y, et al (2020) Recent advances in improving metabolic robustness of microbial cell factories. Curr Opin Biotechnol 66:69–77. https://doi.org/10.1016/j.copbio.2020.06.006

45. Cho C, Jang Y-S, Moon HG, et al (2015) Metabolic engineering of *Clostridia* for the production of chemicals. Biofuels Bioprod Biorefining 9:211–225. https://doi.org/10.1002/bbb.1531

46. Soucaille P, Croux C, Meynial I (2012) Cellulolytic *Clostridium acetobutylicum*

47. Noar J, Makwana ST, Bruno-Bárcena JM (2014) Complete Genome Sequence of Solvent-Tolerant *Clostridium beijerinckii* Strain SA-1. Genome Announc 2:e01310-14. https://doi.org/10.1128/genomeA.01310-14

48. Zhao R, Dong W, Yang C, et al (2023) Formate as a supplementary substrate facilitates sugar metabolism and solvent production by *Clostridium beijerinckii* NCIMB 8052. Synth Syst Biotechnol 8:196–205. https://doi.org/10.1016/j.synbio.2023.01.005

49. Chou C-H, Han C-L, Chang J-J, Lay J-J (2011) Co-culture of *Clostridium beijerinckii* L9, *Clostridium butyricum* M1 and *Bacillus thermoamylovorans* B5 for converting yeast waste into hydrogen. Int J Hydrog Energy 36:13972–13983. https://doi.org/10.1016/j.ijhydene.2011.03.067

50. Nguyen S, Ala F, Cardwell C, et al (2013) Isolation and screening of carboxydotrophs isolated from composts and their potential for butanol synthesis. Environ Technol 34:1995–2007. https://doi.org/10.1080/09593330.2013.795987

51. Desvaux M (2005) *Clostridium cellulolyticum*: model organism of mesophilic cellulolytic clostridia. FEMS Microbiol Rev 29:741–764. https://doi.org/10.1016/j.femsre.2004.11.003

52. Im C, Valgepea K, Modin O, Nygård Y (2022) *Clostridium Ljungdahlii* as a Biocatalyst in Microbial Electrosynthesis – Effect of Culture Conditions on Product Formation. SSRN Electron J. https://doi.org/10.2139/ssrn.4115152

53. Song Y, Lee JS, Shin J, et al (2020) Functional cooperation of the glycine synthase-reductase and Wood–Ljungdahl pathways for autotrophic growth of *Clostridium drakei*. Proc Natl Acad Sci 117:7516–7523. https://doi.org/10.1073/pnas.1912289117

54. Sarma S, Ortega D, Minton NP, et al (2019) Homologous overexpression of hydrogenase and glycerol dehydrogenase in *Clostridium pasteurianum* to enhance hydrogen production from crude glycerol. Bioresour Technol 284:168–177. https://doi.org/10.1016/j.biortech.2019.03.074

55. Pyne ME, Liu X, Moo-Young M, et al (2016) Genome-directed analysis of prophage excision, host defence systems, and central fermentative metabolism in *Clostridium pasteurianum*. Sci Rep 6:26228. https://doi.org/10.1038/srep26228

56. Gong F, Bao G, Zhao C, et al (2016) Fermentation and genomic analysis of acetone-uncoupled butanol production by *Clostridium tetanomorphum*. Appl Microbiol Biotechnol 100:1523–1529. https://doi.org/10.1007/s00253-015-7121-0

57. Ma C, Kojima K, Xu N, et al (2015) Comparative proteomics analysis of high n-butanol producing metabolically engineered *Clostridium tyrobutyricum*. J Biotechnol 193:108–119. https://doi.org/10.1016/j.jbiotec.2014.10.036

58. Yu L, Xu M, Tang I-C, Yang S-T (2015) Metabolic engineering of *Clostridium tyrobutyricum* for n-butanol production through co-utilization of glucose and xylose: Metabolic Engineering of *C. tyrobutyricum* for Butanol Production. Biotechnol Bioeng 112:2134–2141. https://doi.org/10.1002/bit.25613

59. Nölling J, Breton G, Omelchenko MV, et al (2001) Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium *Clostridium acetobutylicum*. J Bacteriol 183:4823–4838. https://doi.org/10.1128/JB.183.16.4823-4838.2001

60. Lee J, Yun H, Feist AM, et al (2008) Genome-scale reconstruction and in silico analysis of the *Clostridium acetobutylicum* ATCC 824 metabolic network. Appl Microbiol Biotechnol 80:849–862. https://doi.org/10.1007/s00253-008-1654-4

61. Poehlein A, Solano JDM, Flitsch SK, et al (2017) Microbial solvent formation revisited by comparative genome analysis. Biotechnol Biofuels 10:58. https://doi.org/10.1186/s13068-017-0742-z

62. Wietzke M, Bahl H (2012) The redox-sensing protein Rex, a transcriptional regulator of solventogenesis in *Clostridium acetobutylicum*. Appl Microbiol Biotechnol 96:749–761. https://doi.org/10.1007/s00253-012-4112-2

63. Paredes CJ, Alsaker KV, Papoutsakis ET (2005) A comparative genomic view of clostridial sporulation and physiology. Nat Rev Microbiol 3:969–978. https://doi.org/10.1038/nrmicro1288

64. Bruder MR, Pyne ME, Moo-Young M, et al (2016) Extending CRISPR-Cas9 Technology from Genome Editing to Transcriptional Engineering in the Genus *Clostridium*. Appl Environ Microbiol 82:6109–6119. https://doi.org/10.1128/AEM.02128-16

65. Onyenwoke RU, Lee Y-J, Dabrowski S, et al (2006) Reclassification of *Thermoanaerobium acetigenum* as *Caldicellulosiruptor acetigenus* comb. nov. and emendation of the genus description. Int J Syst Evol Microbiol 56:1391–1395. https://doi.org/10.1099/ijs.0.63723-0

66. Vees CA, Neuendorf CS, Pflügl S (2020) Towards continuous industrial bioprocessing with solventogenic and acetogenic clostridia: challenges, progress and perspectives. J Ind Microbiol Biotechnol 47:753–787. https://doi.org/10.1007/s10295-020-02296-2

67. Tomita H, Okazaki F, Tamaru Y (2019) Direct IBE fermentation from mandarin orange wastes by combination of *Clostridium cellulovorans* and *Clostridium beijerinckii*. AMB Express 9:1. https://doi.org/10.1186/s13568-018-0728-7

68. Choi KR, Jang WD, Yang D, et al (2019) Systems Metabolic Engineering Strategies: Integrating Systems and Synthetic Biology with Metabolic Engineering. Trends Biotechnol 37:817–837. https://doi.org/10.1016/j.tibtech.2019.01.003

69. Liu J, Lin Q, Chai X, et al (2018) Enhanced phenolic compounds tolerance response of *Clostridium beijerinckii* NCIMB 8052 by inactivation of Cbei_3304. Microb Cell Factories 17:35. https://doi.org/10.1186/s12934-018-0884-0

70. Choi O, Kim T, Woo HM, Um Y (2014) Electricity-driven metabolic shift through direct electron uptake by electroactive heterotroph *Clostridium pasteurianum*. Sci Rep 4:6961. https://doi.org/10.1038/srep06961

71. Gallardo R, Alves M, Rodrigues LR (2017) Influence of nutritional and operational parameters on the production of butanol or 1,3-propanediol from glycerol by a mutant *Clostridium pasteurianum*. New Biotechnol 34:59–67. https://doi.org/10.1016/j.nbt.2016.03.002

72. Kolek J, Sedlar K, Provaznik I, Patakova P (2016) Dam and Dcm methylations prevent gene transfer into *Clostridium pasteurianum* NRRL B-598: development of methods for electrotransformation, conjugation, and sonoporation. Biotechnol Biofuels 9:. https://doi.org/10.1186/s13068-016-0436-y

73. Pyne ME, Moo-Young M, Chung DA, Chou CP (2013) Development of an electrotransformation protocol for genetic manipulation of *Clostridium pasteurianum*. Biotechnol Biofuels 6:50. https://doi.org/10.1186/1754-6834-6-50

74. Utesch T, Sabra W, Prescher C, et al (2019) Enhanced electron transfer of different mediators for strictly opposite shifting of metabolism in *Clostridium pasteurianum* grown on glycerol in a new electrochemical bioreactor. Biotechnol Bioeng 116:1627–1643. https://doi.org/10.1002/bit.26963

75. Rappert S, Song L, Sabra W, et al (2013) Draft Genome Sequence of Type Strain *Clostridium pasteurianum* DSM 525 (ATCC 6013), a Promising Producer of Chemicals and Fuels. Genome Announc 1:e00232-12. https://doi.org/10.1128/genomeA.00232-12

76. Wen Z, Li Q, Liu J, et al (2020) Consolidated bioprocessing for butanol production of cellulolytic *Clostridia*: development and optimization. Microb Biotechnol 13:410–422. https://doi.org/10.1111/1751-7915.13478

77. Berlemont R, Martiny AC (2013) Phylogenetic Distribution of Potential Cellulases in Bacteria. Appl Environ Microbiol 79:1545–1554. https://doi.org/10.1128/AEM.03305-12

78. Xu T, Li Y, He Z, Zhou J (2014) Dockerin-containing protease inhibitor protects key cellulosomal cellulases from proteolysis in *C lostridium cellulolyticum*: Role of Dpi in protecting cellulases. Mol Microbiol 91:694–705. https://doi.org/10.1111/mmi.12488

79. Sabra W, Groeger C, Sharma PN, Zeng A-P (2014) Improved n-butanol production by a non-acetone producing *Clostridium pasteurianum* DSMZ 525 in mixed substrate fermentation. Appl Microbiol Biotechnol 98:4267–4276. https://doi.org/10.1007/s00253-014-5588-8

80. Jaenicke S, Ander C, Bekel T, et al (2011) Comparative and Joint Analysis of Two Metagenomic Datasets from a Biogas Fermenter Obtained by 454-Pyrosequencing. PLoS ONE 6:. https://doi.org/10.1371/journal.pone.0014519

81. Salimi F, Mahadevan R (2013) Characterizing metabolic interactions in a clostridial co-culture for consolidated bioprocessing. BMC Biotechnol 13:. https://doi.org/10.1186/1472-6750-13-95

82. Hanly TJ, Henson MA (2011) Dynamic flux balance modeling of microbial co-cultures for efficient batch fermentation of glucose and xylose mixtures. Biotechnol Bioeng 108:376–385. https://doi.org/10.1002/bit.22954

83. Tomi-Andrino C, Norman R, Millat T, et al (2021) Physicochemical and metabolic constraints for thermodynamics-based stoichiometric modelling under mesophilic growth conditions. PLOS Comput Biol 17:e1007694. https://doi.org/10.1371/journal.pcbi.1007694

84. Fernández-Blanco C, Robles-Iglesias R, Naveira-Pazos C, et al (2023) Production of biofuels from C_1 -gases with *Clostridium* and related bacteria—Recent advances. Microb Biotechnol 16:726–741. https://doi.org/10.1111/1751-7915.14220

85. Jin S, Bae J, Song Y, et al (2020) Synthetic Biology on Acetogenic Bacteria for Highly Efficient Conversion of C1 Gases to Biochemicals. Int J Mol Sci 21:. https://doi.org/10.3390/ijms21207639

86. Huang H, Chai C, Li N, et al (2016) CRISPR/Cas9-Based Efficient Genome Editing in *Clostridium ljungdahlii*, an Autotrophic Gas-Fermenting Bacterium. ACS Synth Biol 5:. https://doi.org/10.1021/acssynbio.6b00044

87. Koller M (2023) Biotechnological Approaches to Generate Biogenic Solvents and Energy Carriers from Renewable Resources. EuroBiotech J 7:96–120. https://doi.org/10.2478/ebtj-2023-0007

88. Han Y-F, Xie B, Wu G-X, et al (2020) Combination of Trace Metal to Improve Solventogenesis of *Clostridium carboxidivorans* P7 in Syngas Fermentation. Front Microbiol 11:. https://doi.org/10.3389/fmicb.2020.577266

89. Nitschke W, Russell MJ (2013) Beating the acetyl coenzyme A-pathway to the origin of life. Philos Trans R Soc B Biol Sci 368:. https://doi.org/10.1098/rstb.2012.0258

90. Branduardi P, Longo V, Berterame NM, et al (2013) A novel pathway to produce butanol and isobutanol in *Saccharomyces cerevisiae*. Biotechnol Biofuels 6:68. https://doi.org/10.1186/1754-6834-6-68

91. Xia P, Casini I, Schulz S, et al (2020) Reprogramming Acetogenic Bacteria with CRISPR-Targeted Base Editing *via* Deamination. ACS Synth Biol 9:. https://doi.org/10.1021/acssynbio.0c00226

92. Ranjan A, Moholkar VS (2012) Biobutanol: science, engineering, and economics: Review Essay. Int J Energy Res 36:277–323. https://doi.org/10.1002/er.1948

93. Yang D, Park SY, Park YS, et al (2020) Metabolic Engineering of *Escherichia coli* for Natural Product Biosynthesis. Trends Biotechnol 38:745–765. https://doi.org/10.1016/j.tibtech.2019.11.007

94. Zingaro KA, Nicolaou SA, Yuan Y, Papoutsakis ET (2014) Exploring the Heterologous Genomic Space for Building, Stepwise, Complex, Multicomponent Tolerance to Toxic Chemicals. ACS Synth Biol 3:476–486. https://doi.org/10.1021/sb400156v

95. Luo H, Liu Z, Xie F, et al (2021) Lignocellulosic biomass to biobutanol: Toxic effects and response mechanism of the combined stress of lignin-derived phenolic acids and phenolic aldehydes to *Clostridium acetobutylicum*. Ind Crops Prod 170:113722. https://doi.org/10.1016/j.indcrop.2021.113722

96. Guo L, Pang Z, Gao C, et al (2020) Engineering microbial cell morphology and membrane homeostasis toward industrial applications. Curr Opin Biotechnol 66:18–26. https://doi.org/10.1016/j.copbio.2020.05.004

97. Sahoo RK, Das A, Gaur M, et al (2019) Bacteria for Butanol Production: Bottlenecks, Achievements and Prospects. J Pure Appl Microbiol 13:1429–1440. https://doi.org/10.22207/JPAM.13.3.13

98. Goyal L, Khanna S (2019) Recent Advances in Microbial Production of Butanol as a Biofuel. Int J Appl Sci Biotechnol 7:130–152. https://doi.org/10.3126/ijasbt.v7i2.24630

99. Mann MS, Dragovic Z, Schirrmacher G, Lütke-Eversloh T (2012) Over-expression of stress protein-encoding genes helps *Clostridium acetobutylicum* to rapidly adapt to butanol stress. Biotechnol Lett 34:1643–1649. https://doi.org/10.1007/s10529-012-0951-2

100. Guo Y, Lu B, Tang H, et al (2019) Tolerance against butanol stress by disrupting succinylglutamate desuccinylase in *Escherichia coli*. RSC Adv 9:11683–11695. https://doi.org/10.1039/C8RA09711A

101. Qi G-X, Xiong L, Huang C, et al (2015) Solvents Production from a Mixture of Glucose and Xylose by Mixed Fermentation of *Clostridium acetobutylicum* and Saccharomyces cerevisiae. Appl Biochem Biotechnol 177:996–1002. https://doi.org/10.1007/s12010-015-1790-0

102. Huang I-H, Sarker MR (2006) Complementation of a *Clostridium perfringens spo0A* Mutant with Wild-Type *spo0A* from Other *Clostridium* Species. Appl Environ Microbiol 72:6388–6393. https://doi.org/10.1128/AEM.02218-05

103. Edwards AN, Tamayo R, McBride SM (2016) A novel regulator controls *C lostridium difficile* sporulation, motility and toxin production: A novel regulator of *C. difficile* physiology. Mol Microbiol 100:954–971. https://doi.org/10.1111/mmi.13361

104. Philippe VA, Méndez MB, Huang I-H, et al (2006) Inorganic Phosphate Induces Spore Morphogenesis and Enterotoxin Production in the Intestinal

Pathogen *Clostridium perfringens*. Infect Immun 74:3651–3656. https://doi.org/10.1128/IAI.02090-05

105. Kirk DG, Zhang Z, Korkeala H, Lindström M (2014) Alternative Sigma Factors SigF, SigE, and SigG Are Essential for Sporulation in *Clostridium botulinum* ATCC 3502. Appl Environ Microbiol 80:5141–5150. https://doi.org/10.1128/AEM.01015-14

106. Vedantam G, Clark A, Chu M, et al (2012) *Clostridium difficile* infection: Toxins and non-toxin virulence factors, and their contributions to disease establishment and host response. Gut Microbes 3:121–134. https://doi.org/10.4161/gmic.19399

107. Talukdar PK, Sarker MR (2022) Characterization of Putative Sporulation and Germination Genes in *Clostridium perfringens* Food-Poisoning Strain SM101. Microorganisms 10:1481. https://doi.org/10.3390/microorganisms10081481

108. Bi C, Jones SW, Hess DR, et al (2011) SpoIIE Is Necessary for Asymmetric Division, Sporulation, and Expression of σ^F, σ^E, and σ^G but Does Not Control Solvent Production in *Clostridium acetobutylicum* ATCC 824. J Bacteriol 193:5130–5137. https://doi.org/10.1128/JB.05474-11

109. Bezerra PKSDB, Araújo BMCD, Silva OLD, et al (2021) Influence of nickel on butanol production by *Clostridium beijerinckii* using hydrolyzate from green coconut shell. Rev Eng Na Agric - Reveng 29:381–388. https://doi.org/10.13083/reveng.v29i1.12620

110. Amiri H (2020) Recent innovations for reviving the ABE fermentation for production of butanol as a drop-in liquid biofuel. Biofuel Res J 7:1256–1266. https://doi.org/10.18331/BRJ2020.7.4.4

111. Lee SH, Doherty TV, Linhardt RJ, Dordick JS (2009) Ionic liquid-mediated selective extraction of lignin from wood leading to enhanced enzymatic cellulose hydrolysis. Biotechnol Bioeng 102:1368–1376. https://doi.org/10.1002/bit.22179

112. Upton BM, Kasko AM (2016) Strategies for the Conversion of Lignin to High-Value Polymeric Materials: Review and Perspective. Chem Rev 116:2275–2306. https://doi.org/10.1021/acs.chemrev.5b00345

113. Liu Y, Tang Y, Gao H, et al (2021) Challenges and Future Perspectives of Promising Biotechnologies for Lignocellulosic Biorefinery. Molecules 26:5411. https://doi.org/10.3390/molecules26175411

114. Tsegaye B, Balomajumder C, Roy P (2019) Microbial delignification and hydrolysis of lignocellulosic biomass to enhance biofuel production: an overview and future prospect. Bull Natl Res Cent 43:51. https://doi.org/10.1186/s42269-019-0094-x

115. Lopes AM, Ferreira Filho EX, Moreira LRS (2018) An update on enzymatic cocktails for lignocellulose breakdown. J Appl Microbiol 125:632–645. https://doi.org/10.1111/jam.13923

116. Jin Q, An Z, Damle A, et al (2020) High Acetone-Butanol-Ethanol Production from Food Waste by Recombinant *Clostridium saccharoperbutylacetonicum* in Batch and Continuous Immobilized-Cell Fermentation. ACS Sustain Chem Eng 8:9822–9832. https://doi.org/10.1021/acssuschemeng.0c02529

117. Gu Y, Jiang Y, Wu H, et al (2011) Economical challenges to microbial producers of butanol: feedstock, butanol ratio and titer. Biotechnol J 6:1348–1357. https://doi.org/10.1002/biot.201100046

118. Huang H, Singh V, Qureshi N (2015) Butanol production from food waste: a novel process for producing sustainable energy and reducing environmental pollution. Biotechnol Biofuels 8:147. https://doi.org/10.1186/s13068-015-0332-x

119. Nawab S, Wang N, Ma X, Huo Y-X (2020) Genetic engineering of non-native hosts for 1-butanol production and its challenges: a review. Microb Cell Factories 19:79. https://doi.org/10.1186/s12934-020-01337-w

120. Zhang Y, Bailey TS, Kubiak AM, et al (2022) Heterologous Gene Regulation in *Clostridia*: Rationally Designed Gene Regulation for Industrial and Medical Applications. ACS Synth Biol 11:3817–3828. https://doi.org/10.1021/acssynbio.2c00401

121. Yao W, Nokes SE (2014) First proof of concept of sustainable metabolite production from high solids fermentation of lignocellulosic biomass using a bacterial co-culture and cycling flush system. Bioresour Technol 173:216–223. https://doi.org/10.1016/j.biortech.2014.08.113

122. Martín C, Dixit P, Momayez F, Jönsson LJ (2022) Hydrothermal Pretreatment of Lignocellulosic Feedstocks to Facilitate Biochemical Conversion. Front Bioeng Biotechnol 10:846592. https://doi.org/10.3389/fbioe.2022.846592

123. Mu R, Jia Y, Ma G, et al (2021) Advances in the use of microalgal-bacterial consortia for wastewater treatment: Community structures, interactions, economic resource reclamation, and study techniques. Water Environ Res Res Publ Water Environ Fed 93:1217–1230. https://doi.org/10.1002/wer.1496

124. Tiwari S, Beliya E, Vaswani M, et al (2022) Rice Husk: A Potent Lignocellulosic Biomass for Second Generation Bioethanol Production from *Klebsiella oxytoca* ATCC 13182. Waste Biomass Valorization 13:2749–2767. https://doi.org/10.1007/s12649-022-01681-5

125. Kumar K, Shah H, Moholkar VS (2022) Genetic Algorithm for Optimization of Fermentation Processes of Various Enzyme Productions. In: Optimization of Sustainable Enzymes Production. Chapman and Hall/CRC

126. Kopke M, Held C, Hujer S, et al (2010) *Clostridium ljungdahlii* represents a microbial production platform based on syngas. Proc Natl Acad Sci 107:13087–13092. https://doi.org/10.1073/pnas.1004716107

127. Tashiro Y, Yoshida T, Noguchi T, Sonomoto K (2013) Recent advances and future prospects for increased butanol production by acetone-butanol-ethanol fermentation. Eng Life Sci 13:432–445. https://doi.org/10.1002/elsc.201200128

128. Gyulev IS, Willson BJ, Hennessy RC, et al (2018) Part by Part: Synthetic Biology Parts Used in Solventogenic *Clostridia*. ACS Synth Biol 7:311–327. https://doi.org/10.1021/acssynbio.7b00327

129. Kumar K, Barbora L, Moholkar VS (2022) Comparative genomic analysis and constraint-based analysis of genome-scale metabolic models of the genus *Clostridia*. Preprints

130. Cho JS, Gu C, Han TH, et al (2019) Reconstruction of context-specific genome-scale metabolic models using multiomics data to study metabolic rewiring. Curr Opin Syst Biol 15:1–11. https://doi.org/10.1016/j.coisb.2019.02.009

131. Milne CB, Eddy JA, Raju R, et al (2011) Metabolic network reconstruction and genome-scale model of butanol-producing strain *Clostridium beijerinckii* NCIMB 8052. BMC Syst Biol 5:130. https://doi.org/10.1186/1752-0509-5-130

132. Dash S, Khodayari A, Zhou J, et al (2017) Development of a core *Clostridium thermocellum* kinetic metabolic model consistent with multiple genetic perturbations. Biotechnol Biofuels 10:108. https://doi.org/10.1186/s13068-017-0792-2

133. Garcia S, Thompson RA, Giannone RJ, et al (2020) Development of a Genome-Scale Metabolic Model of *Clostridium thermocellum* and Its Applications for Integration of Multi-Omics Datasets and Computational Strain Design. Front Bioeng Biotechnol 8:772. https://doi.org/10.3389/fbioe.2020.00772

134. Nagarajan H, Sahin M, Nogales J, et al (2013) Characterizing acetogenic metabolism using a genome-scale metabolic reconstruction of *Clostridium ljungdahlii*. Microb Cell Factories 12:118. https://doi.org/10.1186/1475-2859-12-118

135. Du G, Zhu C, Wu Y, et al (2022) Effects of orphan histidine kinases on *Clostridial* sporulation progression and metabolism. Biotechnol Bioeng 119:226–235. https://doi.org/10.1002/bit.27968

136. Crown SB, Indurthi DC, Ahn WS, et al (2011) Resolving the TCA cycle and pentose-phosphate pathway of *Clostridium acetobutylicum* ATCC 824: Isotopomer analysis, in vitro activities and expression analysis. Biotechnol J 6:300–305. https://doi.org/10.1002/biot.201000282

137. Im C, Valgepea K, Modin O, Nygård Y (2022) *Clostridium ljungdahlii* as a biocatalyst in microbial electrosynthesis – Effect of culture conditions on product formation. Bioresour Technol Rep 19:101156. https://doi.org/10.1016/j.biteb.2022.101156

138. Dash S, Mueller TJ, Venkataramanan KP, et al (2014) Capturing the response of *Clostridium acetobutylicum* to chemical stressors using a regulated genome-scale metabolic model. Biotechnol Biofuels 7:144. https://doi.org/10.1186/s13068-014-0144-4

139. Wen Z, Ledesma-Amaro R, Lin J, et al (2019) Improved n-Butanol Production from *Clostridium cellulovorans* by Integrated Metabolic and Evolutionary Engineering. Appl Environ Microbiol 85:e02560-18. https://doi.org/10.1128/AEM.02560-18

140. Jang Y-S, Park JM, Choi S, et al (2012) Engineering of microorganisms for the production of biofuels and perspectives based on systems metabolic engineering approaches. Biotechnol Adv 30:989–1000. https://doi.org/10.1016/j.biotechadv.2011.08.015

141. Yang B, Nie X, Gu Y, et al (2020) Control of solvent production by sigma-54 factor and the transcriptional activator AdhR in *Clostridium beijerinckii*. Microb Biotechnol 13:328–338. https://doi.org/10.1111/1751-7915.13505

142. Lee SY, Park JH, Jang SH, et al (2008) Fermentative butanol production by clostridia. Biotechnol Bioeng 101:209–228. https://doi.org/10.1002/bit.22003

143. Bhatt TR, Castley JG, Baral HS, Chauvenet A (2023) A synthesis of priorities, patterns, and gaps in large carnivore corridor research. Front Conserv Sci 4:

144. Cooper KK, Songer JG, Uzal FA (2013) Diagnosing *Clostridial* enteric disease in poultry. J Vet Diagn Invest 25:314–327. https://doi.org/10.1177/1040638713483468

145. Altay N, Heaslip G, Kovács G, et al (2023) Innovation in humanitarian logistics and supply chain management: a systematic review. Ann Oper Res. https://doi.org/10.1007/s10479-023-05208-6

146. Tanner RS, Wolfe RS, Ljungdahl LG (1978) Tetrahydrofolate enzyme levels in *Acetobacterium woodii* and their implication in the synthesis of acetate from CO2. J Bacteriol 134:668–670. https://doi.org/10.1128/jb.134.2.668-670.1978

147. Remya VR, Kurian M (2019) Synthesis and catalytic applications of metal–organic frameworks: a review on recent literature. Int Nano Lett 9:17–29. https://doi.org/10.1007/s40089-018-0255-1

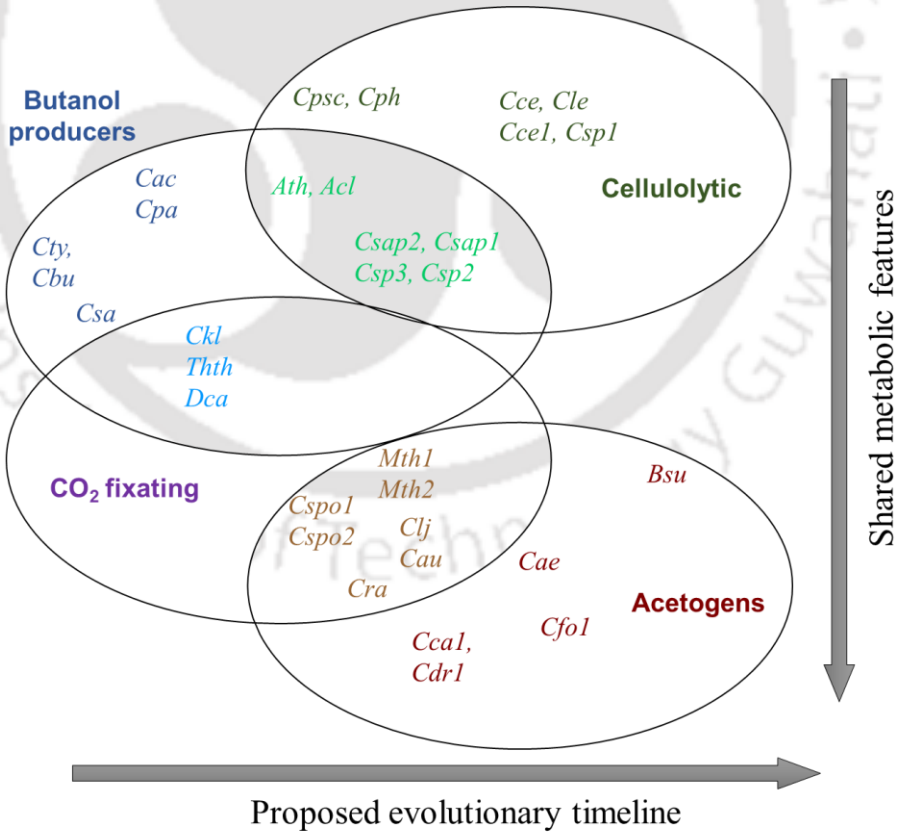
148. Zaccheria F, Scotti N, Ravasio N (2019) Solid Acids for the Reaction of Bioderived Alcohols into Ethers for Fuel Applications. Catalysts 9:172. https://doi.org/10.3390/catal9020172

149. Ravaud P, Créquit P, Williams HC, et al (2020) Future of evidence ecosystem series: 3. From an evidence synthesis ecosystem to an evidence ecosystem. J Clin Epidemiol 123:153–161. https://doi.org/10.1016/j.jclinepi.2020.01.027

150. Takagaki A, Nishimura S, Ebitani K (2012) Catalytic Transformations of Biomass-Derived Materials into Value-Added Chemicals. Catal Surv Asia 16:164–182. https://doi.org/10.1007/s10563-012-9142-3

CHAPTER 2

Genomic Insights into Clostridia in Bioenergy Production: Comparison of Metabolic Capabilities and Evolutionary Relationships



Available at: Kumar, K., Barbora, L., Moholkar, V. S.* (2023). *Biotech. & Bioeng.*, 1–16. https://doi.org/10.1002/bit.28610

CHAPTER 2

GENOMIC INSIGHTS INTO CLOSTRIDIA IN BIOENERGY PRODUCTION: COMPARISON OF METABOLIC CAPABILITIES AND EVOLUTIONARY RELATIONSHIPS

2.1. INTRODUCTION

Microorganisms belonging to the genus *Clostridium* are a large family of anaerobic, sporulating, Gram-positive, and rod-shaped bacteria [1–3]. The entire genus clostridia encompasses more than 270 bacterial strains [4], which include pathogens, probiotic, thermophilic, cellulolytic, solventogenic, and soil-dwelling bacteria [1, 3, 5]. Most clostridial species are recognized as notable cellular factories due to their metabolic diversity and capability of utilizing a vast range of substrates to produce various value-added and platform chemicals such as organic acids, alcohols, and other compounds through fermentation [3, 6].

In the context of contemporary concerns about energy security and climate change, the focus has intensified on clostridial species capable of utilizing biomass and other economical substrates, such as residual agrobiomass and crude glycerol from the biodiesel industry, for biofuel production, including ethanol and butanol [7–9]. The unique capacity of cellulolytic clostridia to degrade complex carbohydrates, including cellulose, is believed to have evolved, in part, through mechanisms like horizontal gene transfer, originating from soil-dwelling bacteria such as *Rhizobiaceae*, *E. coli*, and denitrifying bacteria. This adaptation allowed these bacteria to participate in nutrient cycling and the decomposition of organic matter, including plant material,

for energy generation. There is also speculation that cellulolytic clostridia might have served as ancestors to solventogenic clostridia, potentially through similar evolutionary mechanisms [10, 11]. However, the precise evolutionary trajectory from soil-dwelling bacteria to cellulolytic clostridia to solventogenic clostridia remains an ongoing subject of debate and inquiry, requiring further investigation.

Distinct clostridial species exhibit unique metabolic capabilities that enable them to efficiently utilize various pentose and hexose sugars obtained through biomass pretreatment. **Figure 2.1** provides an overview of the diverse metabolic pathways involved in substrate metabolism in solventogenic and cellulolytic clostridia. Despite these diverse metabolic features of clostridial species (utilizing various carbon sources and producing different products), their evolutionary history has not been explored. Some of the basic questions regarding the evolution of clostridial species are as follows:

(i) How did clostridial species evolve into cellulolytic, solventogenic, CO_2-fixating, and acetogens? (CO_2-fixing clostridia encompass a broader category of bacteria that includes acetogenic clostridia as well as other species capable of autotrophic growth on C1 substrates, such as carbon monoxide (CO) and CO_2 [12]. These bacteria can utilize CO_2 as a carbon source and convert it into organic compounds through various metabolic pathways [13]. Acetogenic clostridia are a subset of CO_2-fixing clostridia, specifically referring to those that produce acetate and ethanol as their primary end products [1]).

(ii) What role did CRISPR/CAS genes and plasmids play during these distinctive metabolic events?

(iii) Sporulation, another crucial event in the life cycle of clostridial species, can be triggered by environmental changes, such as shifts in carbon sources, growth media

components, cell density, and the accumulation of acids and solvents [1]. Despite its significance, our understanding of the regulatory mechanisms governing sporulation remains limited.

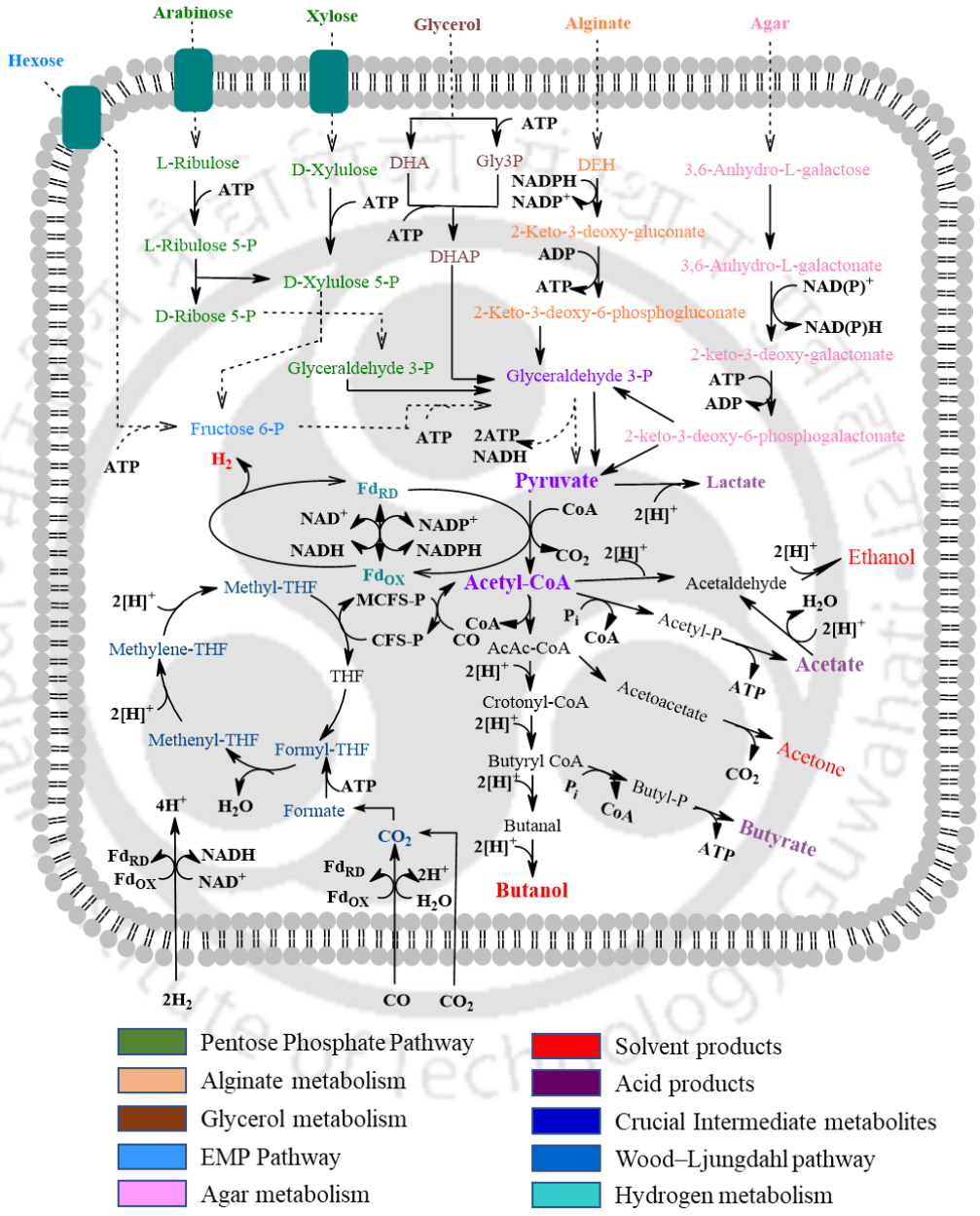


Figure 2.1: Central carbon metabolism in cellulolytic and solventogenic *Clostridia*: colors indicate the metabolic pathways for utilizing various types of substrates in *Clostridia.*

To facilitate the development of genetically modified microbial species for advanced biofuel production, a comprehensive understanding of the genome sequences of industrially relevant clostridial strains, compared with those of other genera, is essential. This analysis includes a comparative examination of central carbon metabolism, which allows for the identification of critical genes and their corresponding functional enzyme products responsible for metabolizing a wide range of carbon sources. In this study, we present an analysis of genomic disparities among 48 fully sequenced bacterial genomes spanning seven genera (**Table 2.1A**). Our approach encompasses pan-genomic analysis, central carbon metabolism comparison, and deduction of unique genes and their copy numbers across various functional categories, including energy conservation, stress response, sporogenesis, chemotaxis, and quorum sensing. These 48 genomes were selected for their diverse characteristics, including acetogenic, butanol-producing, cellulolytic, CO_2-fixating, chemo(litho/organo)trophic, heterotrophic, and uncharacterized (lack available chemotrophic or heterotrophic information and require further study) traits.

This study addresses several key questions: (a) What are the evolutionary and genetic relationships among these 48 genomes? (b) What are the conserved and divergent genomic features within this dataset? (c) Which enzymes play pivotal roles in central carbon metabolism and are common across all genomes? and (d) What are the essential genes involved in general stress responses that get activated during sporogenesis? Answers to these questions can contribute to a deeper understanding of the molecular mechanisms underlying cellulolytic and solventogenic processes, stress responses, and sporulation. Furthermore, this manuscript attempts to provide a better comprehension of the clostridial evolution and their shared metabolic capabilities.

2.2. METHODOLOGY

The schematic representation of the workflow, including general steps and tools utilized in the comparative and pan-genomic analysis of *Clostridia*, is given in **Figure 2.2**. Detailed procedure in a step-by-step manner followed in this study is presented in the following sections.

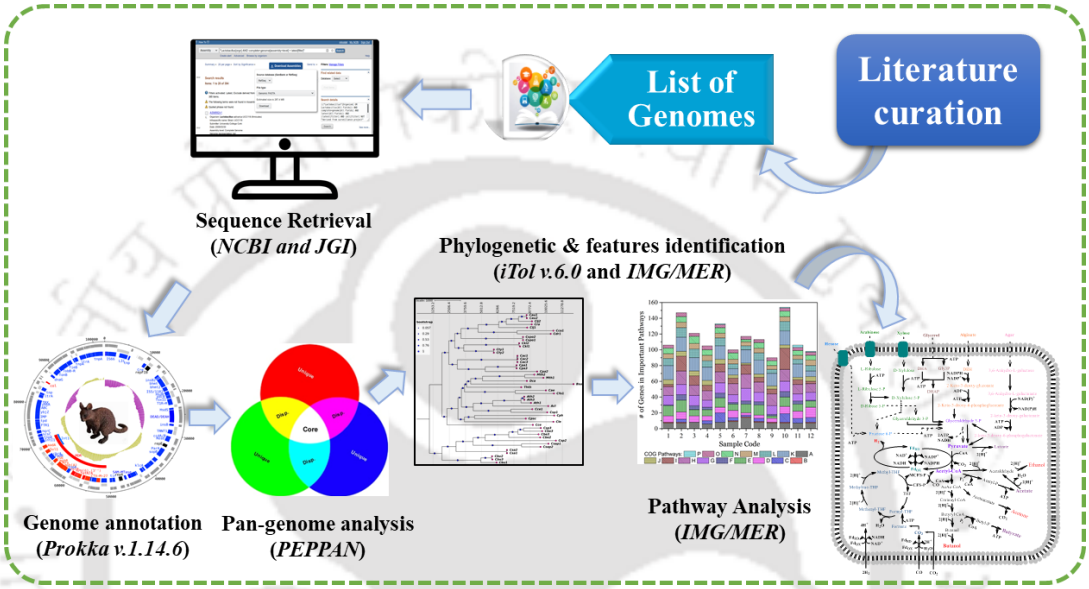


Figure 2.2: Schematic workflow including general comparative and pan-genomic analysis steps for investigating solvent production in *Clostridia* utilizing lignocellulosic hydrolysates.

2.2.1. Bacterial Genome Sequence Retrieval

A total of 48 well-studied and fully sequenced (with sequencing status "Finished" or "Permanent Draft") bacterial genomes of the genera *Acetivibrio* (*A.*), *Bacillus* (*B.*), *Cellulosilyticum* (*Ce.*), *Clostridium* (*C.*), *Desulfotomaculum* (*D.*), *Lachnoclostridium* (*L.*), *Moorella* (*M.*), *Ruminiclostridium* (*R.*), and *Thermoanaerobacterium* (*T.*) were retrieved from the RefSeq FTP site (ftp://ftp.ncbi.nlm.nih.gov/, last accessed on May 25, 2023) and the JGI genome portal (https://genome.jgi.doe.gov/portal/, last accessed on May 25, 2023).

2.2.2. Genomic Annotation and Feature Identification

Genome annotation is the process of identification of the locations of genes, their coding regions and determining their functions. This is the first step in any comparative genomic analysis. To avoid the divergence of multiple annotation systems, all retrieved genomes from RefSeq or JGI database were reannotated using Prokka v1.12 [14], an automatic annotation pipeline in GitHub.

2.2.3. Comparative Genomic Analysis

2.2.3.1. Pan-genome Analysis

Core- and pan-genome analysis has been utilised widely to evaluate genome diversity, genome dynamics, species evolution, pathogenesis, cellulolytic, solventogenic, and other metabolic features of microorganisms. After annotating the bacterial sequences by Prokka, we performed a robust estimation of core- and pan-genomes in all 48 strains. Moreover, this analysis was carried out in the following bacterial clusters, viz. cellulolytic, solventogenic, acetogenic, and ethanologenic groups using PEPPAN [15]. PEPPAN stands for Phylogeny Enhanced Pipeline for PAN-genome – a pipeline that can construct a pan-genome from thousands of genetically diversified bacterial genomes. The core- and pan-genome analysis results in the following categories of genes: strict core, soft-core, shell, cloud, and total genes. Strict core genes in the pan-genome analysis are a pool of conserved genes in all analysed genomes. Soft-core genes signify the set of genes present in 95% of the genomes in the study. Comparative genomic analyses rely on core genes and soft-core gene categories to guide the inclusion of genes absent in draft genomes, as these gene clusters represent highly conserved genes that provide information on the evolutionary history of the analyzed genomes. On the other hand, shell and cloud gene clusters include subsets of the

flexible genome, reflecting an organism's evolutionary history and adaptation to its environment [16]. These analyses are crucial in understanding the genetic diversity and evolution of microorganisms.

2.2.3.2. Whole Genome-based Phylogenetic and Taxonomic Analysis

After core- and pan-genome analysis, the critical step in any comparative genomic pipeline is confirmation of the phylogenetic origins of the microorganism, or its adjustment, respectively. Numerous genomes available in the GenBank/NCBI database are misclassified, and therefore, one needs to be cautious while performing phylogenetic analyses. In this study, we have identified the phylogenetic relationship of all 48 bacterial species using PEPPAN (produced a phylogenetic tree file) and manually checked all the classifications. The phylogenetic tree was constructed and visualized using interactive Tree of Life (iTOL) v.6.5 [17].

2.2.3.3. Detailed Feature Comparison and Identification of Genome-Wide Essential Metabolic Pathways

All additional whole genomic comparisons, which include – genome statistics, identification of genes in various COGs (cluster of orthologous groups) functional categories, genes present in essential metabolic features such as general stress response, sporogenesis, biphasic fermentation (acidogenesis and solventogenesis), cellulose degradation, energy and electron transfer, ion and metabolic transport, carbon/nitrogen fixation and metabolism, quorum sensing, chemotaxis, etc.; were performed in the IMG/M ER v.7.0 portal [18, 19]. The COGs database (also termed Clusters of Orthologous Groups of proteins) is a tool for predicting the gene functions and comparative genomics in complete genomes of bacteria and archaea.

2.3. RESULTS AND DISCUSSION

2.3.1. Phylogenetic and Taxonomic Classification of 48 Bacterial Genomes

2.3.1.1. Genome Information and Analysis

Table 2.1A details metadata for 48 strains across seven genera, encompassing characteristics like strain number, metabolism, and energy sources. Additional genomic specifics are available in **Sheet 2** of the excel file **Appendix Ex1**. These strains are Gram-positive, free-living anaerobes. Genome statistics in **Table 2.1B** and **Sheet 3** (**Appendix Ex1**) include key details like genome size and quality.

A robust core- and pan-genome analysis of 48 strains from seven genera reveals significant genomic diversity (**Table 2.2**). Notably, butanol-producers have the largest genomes, cellulolytic strains have smaller genomes, and acetogenic and CO_2-fixating strains share similarities. This diversity, reflected in gene content and distribution, underscores distinct metabolic capabilities. Detailed genome statistics for each group are in **Sheets 4-10** (**Appendix Ex1**), providing a comprehensive genomic landscape overview.

Table 2.1 (A): Metadata information of 48 selected bacterial genomes: genus, species, metabolic feature, temperature, and O_2 requirements

Genus	Species	Str.	Metabolism	Energy Source	Temperature	Products
Acetivibrio (*A.*)	*clariflavus (Acl)*	1	Cellulolytic	Chemoorganotrophs	Thermophile	Ethanol
	thermocellus (Ath)	3	CBP	Chemoorganotrophs		Ethanol
Bacillus (*B.*)	*subtilis (Bsu)*	1	Sporulating, lipase production	Heterotrophs	Mesophile	Diverse products
Cellulosilyticum (*Ce.*)	*lentocellum (Cel)*	1	Cellulolytic	Chemoorganotrophs	Mesophile	Ethanol, acetate
Clostridium (*C.*)	*aceticum (Cae)*	1	H_2/CO_2, CO	Chemolithotrophs	Mesophile	Acetate
	acetobutylicum (Cac)	3	Solventogenic	Chemoorganotrophs		ABE, OAs
	autoethanogenum (Cau)	2	H_2/CO_2, CO	Chemoorganotrophs		2,3-BDO, acetate, ethanol
	beijerinckii (Cbe)	3	Solventogenic	Chemoorganotrophs		ABE, acetate, butyrate
	butyricum (Cbu)	3	Solventogenic, glycerol, lactate, acetate	Chemoorganotrophs		Butanol, 1,3-PDO
	carboxidivorans (Cca)	1	H_2/CO_2, CO	Uncharacterized		OAs, ethanol, butanol
	cellulovorans (Cce)	1	Cellulolytic	Chemoorganotrophs		ABE
	drakei (Cdr)	1	H_2/CO_2, CO	Uncharacterized		OAs, ethanol
	formicaceticum (Cfo)	1	CO, CH_3OH	Uncharacterized		Acetate, formate
	kluyveri (Ckl)	2	Ethanol, acetate, CO_2	Chemoorganotrophs		Butyrate, caporate, H_2
	ljungdahlii (Clj)	2	H_2/CO_2, CO, Formate	Uncharacterized		2,3-BDO, acetate, ethanol

Table 2.1 (A) continued…

Genus	Species	Str.	Metabolism	Energy Source	Temperature	Products
	pasteurianum (Cpa)	3	Glycerol, N_2 fixation	Heterotrophs and Chemoorganotrophs		ABE, OAs, 1,3-PDO
	ragsdalei (Cra)	1	H_2/CO_2, CO	Chemoorganotrophs		2,3-BDO, acetate, ethanol
	saccharobutylicum (Csa)	2	Solventogenic	Uncharacterized		ABE
	saccharoperbutylacetonicum (Csap)	2	Solventogenic	Heterotrophs		ABE
	sporogenes (Cspo)	2	Aromatic amino acids	Chemoorganotrophs		Indole propionic acid
	tyrobutyricum (Cty)	2	Acidogenic	Chemoorganotrophs		ABE, OAs
	Unclassified (Csp)	3		Uncharacterized		
Desulfotomaculum (*D.*)	*carboxydivorans (Dca)*	1	Sulfate, CO	Chemoorganotrophs	Mesophile	H_2 and CO_2
Lachnoclostridium (*L.*)	*phytofermentans (Cph)*	1	Cellulolytic	Chemoorganotrophs	Mesophile	Ethanol, acetate, CO_2, H_2
	polysaccharolyticum (Cpsc)	1	Lipids	Uncharacterized		Fats, serum acetate
Moorella (*M.*)	*thermoacetica (Mth)*	2	H_2/CO_2, CO, formate	Chemoorganotrophs	Mesophile	Acetate
Ruminiclostridium (*R.*)	*cellulolyticum (Cce1)*	1	Cellulolytic	Chemoorganotrophs	Mesophile	ABE, OAs
Thermoanaerobacterium (*T.*)	*thermosaccharolyticum (Thth)*	1	Thiosulfate producer	Heterotrophs	Thermophile	Ethanol, butanol, H_2, OAs

Str. – Strains, CBP – Consolidated bioprocessing, ABE – Acetone-Butanol-Ethanol, OAs – Organic acids

Table 2.1 (B): Comparative genomic statistics of 48 selected bacterial genomes

Genome Name	*Abb.*	Size (MB)	% CDS	# Genes	# RNA	% Pseu.	% Ptn.	# H. T.	# Sig. pp	% Enzy.
A. clariflavus EBR 45 [&]	*Acl*	4.9	97.7	4229	98	5.65	73.7	293	278	24.3
A. thermocellus ATCC 27405 [&]	*Ath1*	3.8	97.0	3335	99	1.35	69.2	76	243	27.2
A. thermocellus LQ8 [&]	*Ath2*	3.6	97.7	3102	71	3.87	69.1	5	265	28.6
A. thermocellus LQRI [&]	*Ath3*	3.6	97.5	3143	79	0	77.1	-	132	28.0
B. subtilis subtilis 168	*Bsu*	4.2	95.9	4354	178	0	76.0	16	280	29.4
Ce. lentocellum DSM 5427 [&]	*Cle*	4.7	96.7	4416	144	2.04	65.8	482	300	24.9
C. aceticum, DSM 1496 [$, @]	*Cae*	4.2	97.7	3945	92	0.2	76.9	285	140	27.0
C. acetobutylicum ATCC 824 [#]	*Cac1*	4.1	95.7	4022	174	0	78.2	1	247	26.4
C. acetobutylicum DSM 1731 [#]	*Cac2*	4.1	97.3	4029	107	0	70.4	1	255	26.4
C. acetobutylicum LJ4 [#]	*Cac3*	4.2	96.2	4184	97	0	77.3	22	148	25.5
C. autoethanogenum DSM 10061 [$, @]	*Cau1*	4.4	97.4	4070	106	0	80.9	-	103	25.8
*C. autoethanogenum JA1-1** [$, @]	*Cau2*	4.3	95.6	4215	97	0	79.4	2331	108	24.6
C. beijerinckii NCIMB 14988 [#]	*Cbe1*	6.5	98.1	5699	107	0	79.0	112	324	26.2
C. beijerinckii Br21 [#]	*Cbe2*	6.0	97.1	5606	93	0	75.5	18	191	24.7
C. beijerinckii NCIMB 8052 [#]	*Cbe3*	6.0	96.4	5290	190	1.51	72.5	483	293	27.5
C. butyricum DSM 10702 [#]	*Cbu1*	4.7	95.9	4436	128	0	76.7	4171	154	25.5
C. butyricum KNU-L09 [#]	*Cbu2*	4.6	95.8	4328	124	0	78.4	9	158	26.5
C. butyricum TOA [#]	*Cbu3*	4.6	97.4	4176	107	0	77.9	-	-	27.3
C. carboxidivorans P7 [$, @]	*Cca1*	5.8	97.7	5316	120	0	77.9	42	184	25.2
C. cellulovorans 743B [#, &]	*Cce*	5.3	97.5	4500	111	3	69.3	27	195	25.0
C. drakei SL1 [$, @]	*Cdr1*	5.7	95.3	5395	127	0	77.8	37	146	25.5

Table 2.1 (B) continued…

Genome Name	*Abb.*	Size (MB)	% CDS	# Genes	# RNA	% Pseu.	% Ptn.	# H. T.	# Sig. pp	% Enzy.
C. formicaceticum ATCC 27076 [$]	*Cfo1*	4.6	95.6	4390	192	0	74.8	170	118	25.7
C. kluyveri DSM 555 [$, @]	*Ckl1*	4.0	96.1	4073	160	0	60.2	76	135	24.1
C. kluyveri NBRC 12016 [$, @]	*Ckl2*	4.0	97.8	3604	81	0	47.6	20	122	27.1
*C. ljungdahlii ERI-2** [*, $, @]	*Clj1*	4.4	97.2	4210	118	0	78.3	54	103	24.9
C. ljungdahlii DSM 13528 [$, @]	*Clj2*	4.6	97.7	4283	99	0	71.0	280	212	25.5
C. pasteurianum BC1 [#]	*Cpa2*	5.0	97.7	4966	115	6.83	73.3	579	187	25.0
C. pasteurianum ATCC 6013 [#]	*Cpa1*	4.4	96.9	4116	129	0	75.5	5	93	27.6
C. pasteurianum M150B [#]	*Cpa3*	4.4	96.9	4113	129	0	75.3	-	93	27.6
*C. ragsdalei P11** [*, $, @]	*Cra*	4.4	98.0	4026	80	0	79.9	17	214	25.0
*C. saccharobutylicum BAS/B3/SW/136** [*, #]	*Csa1*	5.1	97.1	4513	130	0.16	74.1	-	276	25.3
*C. saccharobutylicum DSM 13864** [*,#]	*Csa2*	5.1	97.0	4532	137	0	74.3	172	275	25.1
C. saccharoperbutylacetonicum N1-4(HMT) [#]	*Csap1*	6.7	98.2	5926	105	0.19	76.8	5	410	25.6
*C. saccharoperbutylacetonicum N1-504** [#]	*Csap2*	6.2	98.3	5612	94	0.14	75.5	84	376	24.8
C. sp. BNL1100	*Csp1*	4.6	97.8	4117	92	2.55	74.0	229	339	26.8
C. sp. DL-VIII	*Csp2*	6.5	97.6	6013	146	0	75.8	225	358	23.5
C. sp. MF28	*Csp3*	6.1	96.7	5673	186	0	78.0	22	190	24.4
C. sporogenes DSM 795	*Cspo1*	4.1	96.3	3890	146	0	76.5	4	91	26.0
C. sporogenes NCIMB 10696	*Cspo2*	4.1	96.3	3891	146	0	76.5	4	91	26.0
C. tyrobutyricum KCTC 5387 [#]	*Cty1*	3.1	96.9	3145	97	0	76.4	-	-	29.4
C. tyrobutyricum W428 [#]	*Cty2*	3.1	97.4	3095	82	0	76.9	2	60	29.6
D. carboxydivorans DSM 14880 [$]	*Dca*	2.9	96.6	2844	97	3.09	78.4	66	133	30.9

Table 2.1 (B) continued…

Genome Name	*Abb.*	Size (MB)	% CDS	# Genes	# RNA	% Pseu.	% Ptn.	# H. T.	# Sig. pp	% Enzy.
L. phytofermentans ISDg [&]	*Cph*	4.8	97.8	3991	89	0.03	70.5	494	288	26.9
*L. polysaccharolyticum DSM 1801**	*Cpsc*	3.5	97.6	3096	76	0	75.4	153	156	27.7
M. thermoacetica DSM 103132 [$, @]	*Mth1*	3.0	96.9	3130	96	0	73.4	153	60	27.2
M. thermoacetica DSM 2955 [$, @]	*Mth2*	2.6	96.1	2783	63	0	76.9	92	54	29.3
R. cellulolyticum H10 [&]	*Cce1*	4.1	97.6	3575	87	2.74	68.8	255	298	27.5
T. thermosaccharolyticum M0795 [#]	*Thth*	2.9	97.5	2922	74	4.55	78.9	158	140	29.7

Abb.- Abbreviation; CDS – Coding sequences; Pseu. – Pseudo genes; Ptn. – genes coding for proteins; H. T. – Horizontally transferred genes; Sig. PP – Signal peptides; Enzy. – Enzymes; "-"– No information (either absent or not investigated)
** Genome sequencing status as permanent draft; [$] acetogenic strains; [#] butanol-producers; [&] cellulolytic; [@] CO_2 fixating*

Table 2.2: Comparative core and pan-genome analysis of the various clusters of strains in the present study

Genome groups	# of Strains	Average statistics in respective groups					Number of genes				
		Size	Genes	GC %	COGs	Paralogs	Strict-C	Soft-C	Shell	Cloud	Total
Acetogenic	14	4.2	4020	36.3	65.3 %	27.4 %	327	0	5204	12054	17585
Butanol-producers	21	4.9	4501	30.1	65.2 %	29.2 %	546	295	5411	16600	22852
Cellulolytic	8	4.4	3786	36.4	57.5 %	15.2 %	490	0	3986	10954	15340
CO_2 fixating	12	4.3	4087	35.5	65.8 %	28.4 %	348	0	8919	6145	15412
Chemo (litho/ Organo) trophs	33	4.3	4020	34.2	63.1 %	22.5 %	208	65	5835	31003	37111
Heterotrophs	7	4.5	4327	32.2	63.1 %	17.6 %	444	0	4447	11276	16167
Uncharacterized	11	5.2	4778	31	64.8 %	28.7 %	506	0	8958	9569	19033
All 48 strains	48	4.5	4215	33.4	63.6 %	23.2 %	185	120	5366	42609	48280

Strict core genes (strains = 100%); soft-core genes (95 % ≤ strains < 99 %); shell genes (15 % ≤ strains < 95 %); cloud genes (0 % ≤ strains < 15 %); Total genes (0 % ≤ strains < 100 %)

2.3.1.2. Phylogenetic Analysis Based on the Whole Genome

Whole genome-based phylogenetic relations of all 48 strains were analyzed based on the detected core genomes. **Figure 2.3** represents the phylogenetic relationships of all 48 genomes. In the phylogeny, most of the bootstrap values crossed the threshold value of 75%, and therefore, it could be inferred that most of the grouping of the species is well supported. As seen in **Figure 2.3**, the rectangular tree bifurcated mainly into two clusters (or clades), viz. larger clade 1 *Cau/Clj/Cra/Cca/Cdr/Cspo/Ckl/Cty /Cac/Cpa/Mth/Dca/Bsu/Cae/Thth/Cfo/Ath/Acl/Cpsc/Cph/Cce/Cle* strains; consisting most of the acetogens, CO_2 fixating, cellulolytic, and a few butanol-producers; and a relatively smaller clade 2: *Cbe/Csp/Csap/Csa/Cbu* consisting second set of butanol-producers. The clade 1 was bifurcated to yield several sub-clusters (or subclades). The clade1 has mainly two subclades, viz. subclade 1.1 consisting of *Cau/Clj/Cra/Cca/Cdr/Cspo/Ckl/Cty* strains and subclade 1.2 consisting of *Cac/Cpa/Mth/Dca/Bsu/Cae/Thth/Cfo/Ath/Acl/Cpsc/Cph/Cce/Cle* strains. In the subclade 1.1, all strains of *Cra, Cau, Clj, Cca, Cdr, Cspo, Ckl,* and *Cty* were grouped together (closely related) and bifurcated from *Cac* and *Cpa,* which suggests that these strains have got solventogenic genes from their common ancestors and later diverged and received acetogenic and CO_2-fixating genes through horizontal gene transfer or other genetic mechanisms. Among *Cpa* strains, *Cpa1* and *Cpa3* were grouped together, but *Cpa2* branched out from its clade. Suggesting that *Cpa2* is distantly related to other analyzed strains or species. The other major sub-cluster in clade 1, subclade 1.2, was diverged from the strain *Cce,* which suggests that all strains possess cellulolytic and acetogenic genes. All *Mth, Dca,* and *Thth* strains were more closely related to the *Bsu* strain in subclade 1.2,

which infers that these strains have got oxygen-tolerant genes from *Bsu* during the course of evolution and became obligate anaerobe. Thermophilic and cellulolytic strains, viz. *Ath*, *Acl*, and *Cce1* were grouped together and bifurcated from *Cpsc*, *Cph*, and *Cle*. Another interesting observation from the phylogenetic analysis is that *Csp1* was a group with cellulolytic and acetogens groups in subclade 1.2. However, *Csp2* and *Csp3* were grouped in clade 2 with butanol-producers. Whole genome-based phylogenetic analysis of bacterial clusters based on metabolic characteristics (viz., acetogenic, butanol-producers, cellulolytic, and CO_2 fixating) and energy sources (viz., chemotrophs, heterotrophs, and uncharacterized) are provided in **Figures A1.1 and A1.2** of **Appendix 1**, respectively.

In our study, we have observed the misclassification of *Csp* strains (*Csp1*, *Csp2*, *Csp3*), and *Cspo* (*Cspo1* and *Cspo2*) strains. We suggest the reclassification of these strains as *Csp1* was found to be closely related with cellulolytic and thermophilic strains. However, *Csp1* and *Csp3* were found to be closely associated with butanol-producers set II having *Cbe* strains. On the other hand, *Cspo1* and *Cspo2*, were grouped together with acetogens and CO_2 – fixating strains.

2.3.2. Gene Distribution in Various COG Categories

Next, we conducted a comprehensive analysis of 48 strains, focusing on the frequency of genes in specific functional categories known as Cluster of Orthologous Groups (COGs). The heatmap of the percentage distribution of protein-coding genes in each COG functional categories are shown in **Figure 2.4A**. We further conducted an extensive investigation of genes in 11 diverse COG categories, namely C, E, G, H, K, O, P, Q, T,

V, and X (see **Sheet 3** of the excel file **Appendix Ex2**). Notably, we prioritized COG categories where gene distribution percentages were notably higher (see **Table 2.3**). Our findings highlight that Category G (Carbohydrate transport and metabolism) contained the highest percentage of genes, notably in strain *Csap1* (13%), followed by Category K (Transcription) in strains *Cdr* and *Csp2* (11.6%).

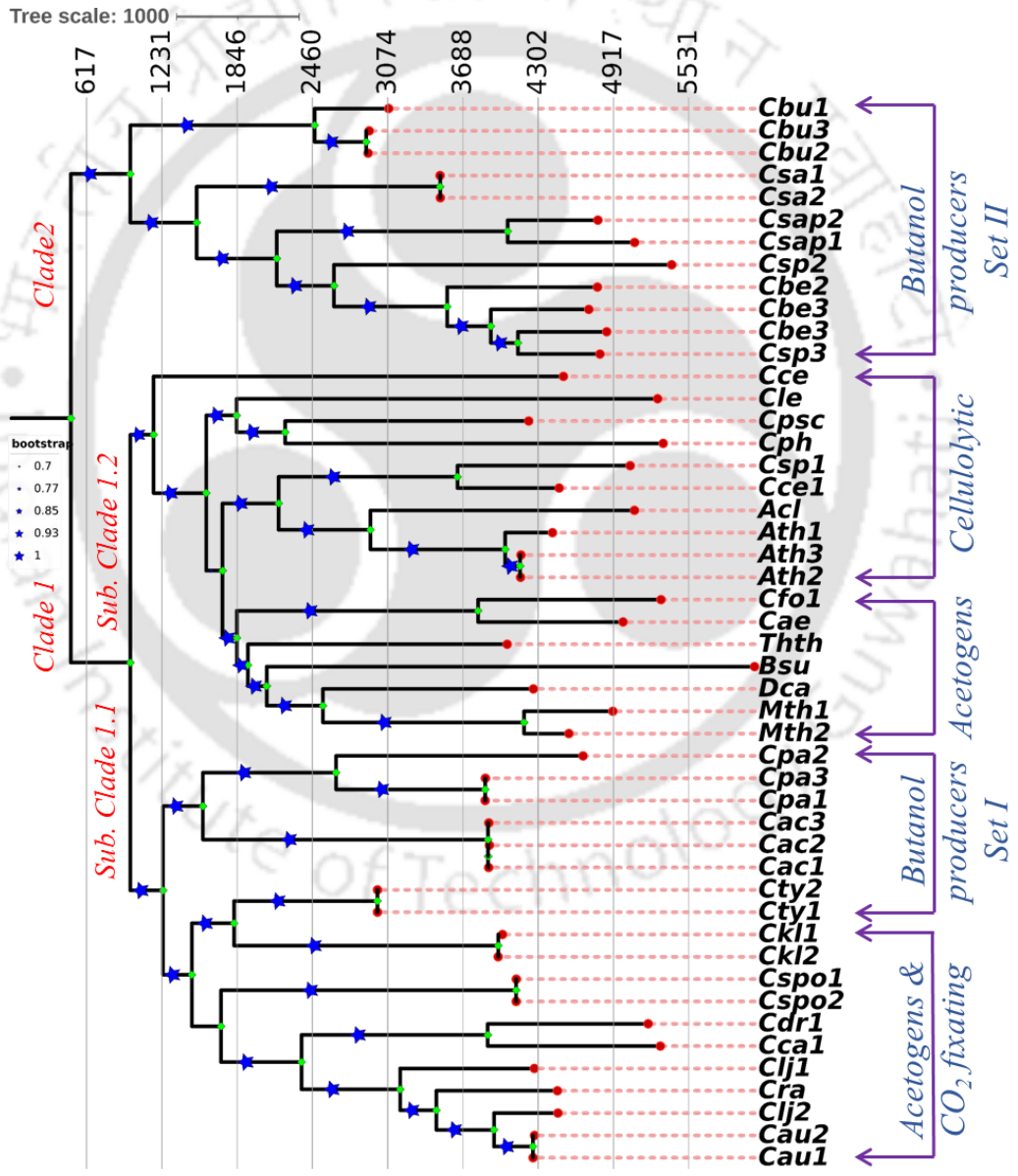


Figure 2.3: Phylogenetic tree of 48 bacterial genomes visualized in iTOL web server.

Heatmaps of the distribution of protein-coding genes in COG category K, O, T, and X are shown in **Figure 2.4B**. Interestingly, acetogenic and cellulolytic strains exhibited elevated percentages of genes in several key categories, including carbohydrate metabolism (G), amino acid metabolism (E), energy production (C), transcription (K), and signal transduction (T). This suggests their adaptability to diverse carbon sources, enhanced stress resistance, and potential for improved sugar metabolism, stress response, chemotaxis, and quorum sensing. For a detailed breakdown of genes in these 11 COG categories across all 48 strains, please refer to **Sheets 4 – 14** of **Appendix Ex2** and **Figure A1.3** of **Appendix 1**.

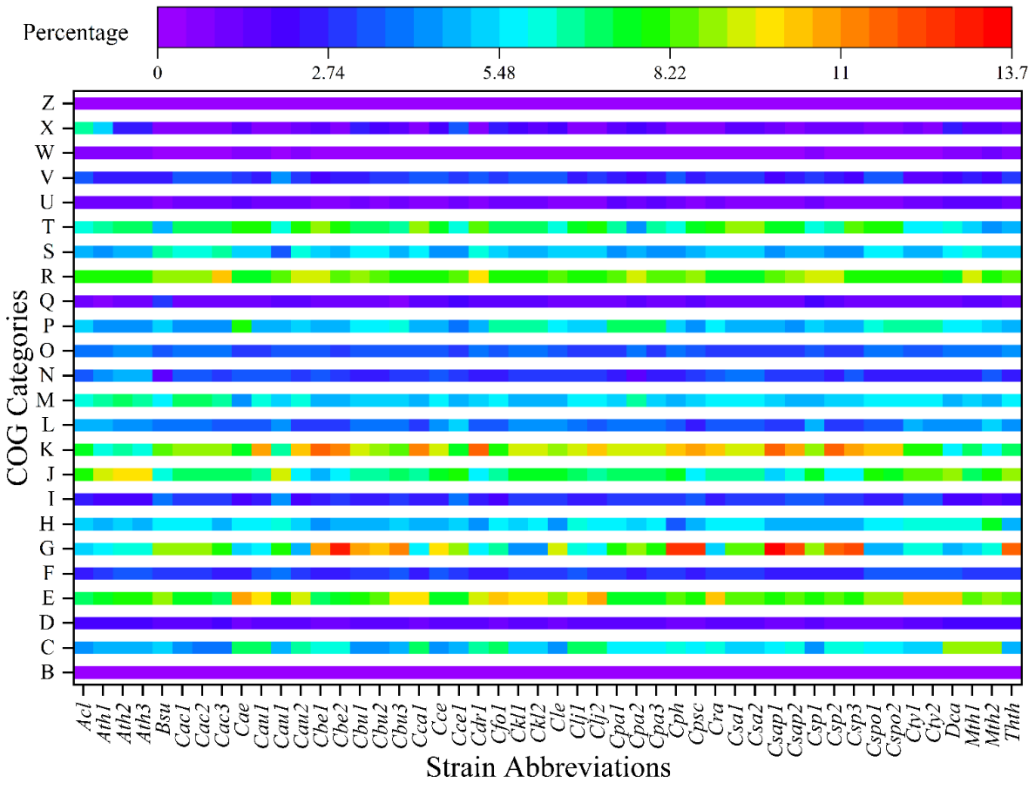
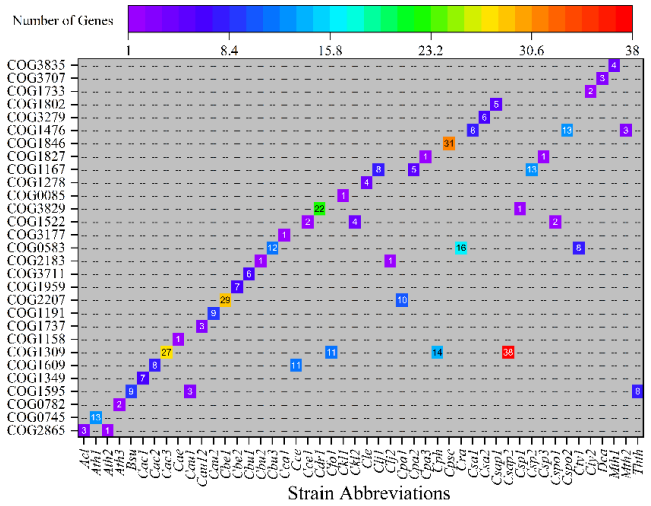
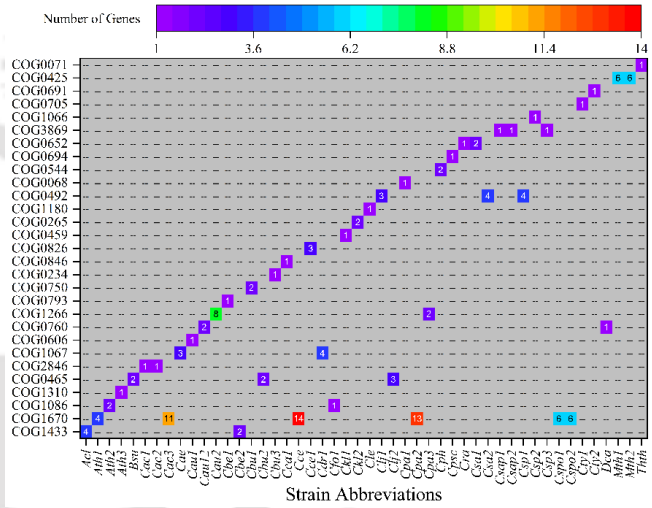


Figure 2.4 (A): Heatmap of the percentage distribution of protein-coding genes in each Cluster of Orthologous (COG) functional categories.

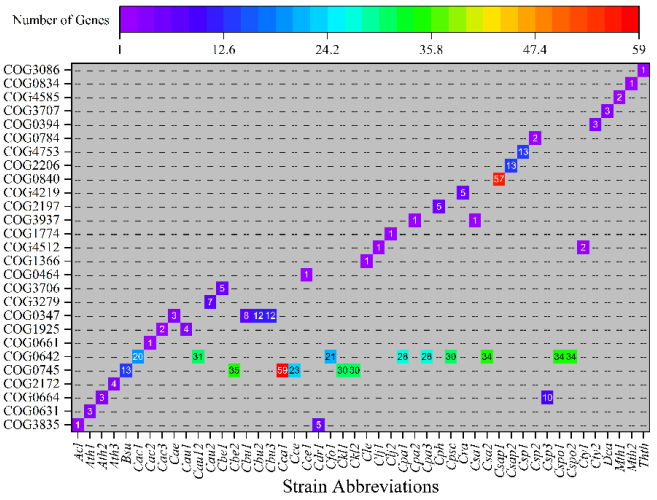
(i)



(ii)



(iii)



(iv)

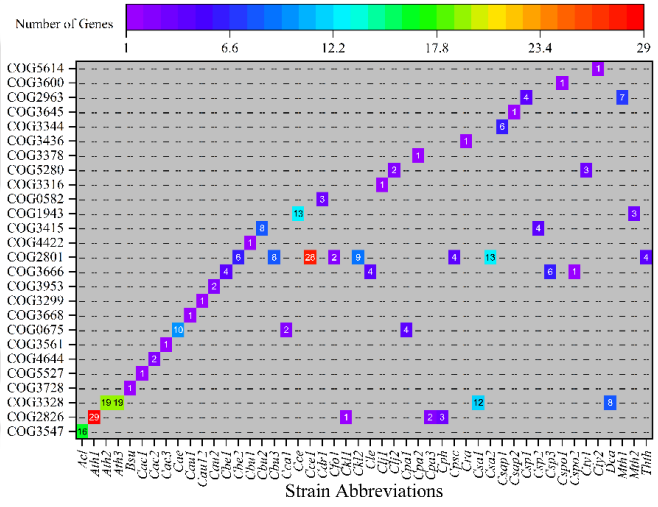


Figure 2.4 (B): Heatmaps of the distribution of protein-coding genes in COG category (i) K, (ii) O, (iii) T, and (iv) X

Table 2.3: Comparison of percentage distribution of genes in COG categories across 48 genomes

COG categories and function name	Max.	Str_Max.	Min.	Avg.	Mode	Str_Min.
C (Energy production and conversion)	8.9	*Mth1*	3.9	5.7	5.0	*Cac3*
D (Cell cycle control, cell division, chromosome partitioning)	2.3	*Thth*	1.1	1.6	2.2	*Cbe2, Csap2, Cpsc, Csp3*
E (Amino acid transport and metabolism)	10.7	*Clj2*	7.1	8.6	N/A	*Acl*
F (Nucleotide transport and metabolism)	3.9	*Cau1*	2.3	2.9	3.3	*Cdr1*
G (Carbohydrate transport and metabolism)	13.7	*Csap1*	4.3	8.0	6.0	*Ckl1*
H (Coenzyme transport and metabolism)	7.3	*Mth2*	3.6	5.3	N/A	*Cph*
I (Lipid transport and metabolism)	4.1	*Cau1*	1.8	2.7	2.2	*Mth2*
J (Translation, ribosomal structure and biogenesis)	9.8	*Ath2*	5.0	7.2	N/A	*Cbe1*
K (Transcription)	11.6	*Cdr1, Csp2*	5.7	9.0	N/A	*Dca*
L (Replication, recombination and repair)	5.4	*Cce1*	2.7	3.8	N/A	*Cpsc*
M (Cell wall/membrane/envelope biogenesis)	7.0	*Cac1*	4.5	5.5	N/A	*Cae*
N (Cell motility)	4.7	*Ath3*	1.6	3.0	N/A	*Cpa2*
O (Posttranslational modification, protein turnover, chaperones)	4.4	*Ath3*	2.9	3.5	N/A	*Cae, Clj1, Csa1, Csap1*

Table 2.3: continued…

COG categories and function name	Max.	Str_Max.	Min.	Avg.	Mode	Str_Min.
P (Inorganic ion transport and metabolism)	7.8	*Cae*	4.0	5.3	4.2	*Cce1*
Q (Secondary metabolites biosynthesis, transport and catabolism)	2.9	*Bsu*	0.9	1.4	1.0	*Cbu3, Ath, Cbu2*
R (General function prediction only)	10.3	*Cac3*	7.5	8.4	N/A	*Cae*
S (Function unknown)	6.8	*Bsu*	3.5	5.1	N/A	*Cau1*
T (Signal transduction mechanisms)	9.0	*Cbe1*	4.3	7.0	7.1	*Mth2*
U (Intracellular trafficking, secretion, and vesicular transport)	1.8	*Mth2*	0.6	1.0	N/A	*Csp3, Cca1, Cbe2*
W (Extracellular structures)	4.2	*Cau1*	1.8	2.8	2.3	*Cty2, Cty1*
X (Mobilome: prophages, transposons)	1.0	*Mth2*	0.1	0.4	N/A	*Cph, Bsu, Cdr1, Cfo1, Cce, Cspo2, Cspo1*

COG – Cluster of orthologous; Max. – Maximum percentage of the gene in each COG category; Str_Max. – Strain in which maximum percentage of the gene was found in respective COG category; Min. – Minimum percentage of the gene in each COG category; Avg. – Average percentage of the gene in each COG category; Str_Min. – Strain in which minimum percentage of the gene was found in respective COG category

2.3.3. Extrachromosomal (EC) and Plasmid DNA and CRISPR/CAS Sequences in Genomes

We observed plasmids in 15 out of the 48 clostridia and few of the strains in Table 1B have either less or no clustered regularly interspaced short palindromic repeats (CRISPR) and CRISPR-associated proteins (CAS) sequences, which suggests that those species have developed other ways of resisting viral attacks (**Table 2.4**). Notably, the genomes of some butanol-producers and acetogens, such as *Cac1*, *Ckl1*, and *Cbe2*, contain a relatively very high percentage of paralogs, which essentially signifies that the larger genome size is partly due to extensive gene duplication events (**Table 2.1B**).

2.3.3.1. Plasmid Presence and its Metabolic Implications

The presence of plasmids within certain bacterial strains, notably in the acetogenic and butanol-producing groups, presents a tantalizing avenue for exploration. These plasmids, often small, circular, extrachromosomal DNA fragments, can carry genes that provide metabolic advantages to their host bacteria. For instance, in acetogenic strains like *Cca1*, *Ckl*, *Cae*, the existence of plasmids suggests a potential connection between these genetic elements and the strains' ability to convert CO_2 into acetate. It is plausible that these plasmids house key genes associated with this specific metabolic function. This observation raises intriguing questions about the role of plasmids in shaping the metabolic profile of these bacteria. Similarly, presence of plasmids is striking among the butanol-producing strains, such as *Cac*, *Cbu*, and *Csap*. These plasmids might contain genes related to butanol production, hinting at a link between plasmid content and the capacity of strains to produce butanol. The variability in plasmid presence within this group

suggests that different strains may have adopted distinct genetic strategies for butanol production or adaptation to varying environmental conditions. This diversity highlights the complexity of metabolic pathways in these strains and underscores the need for further investigation into the roles of these plasmids.

2.3.3.2. Genomic Insights into Bacterial Defense Mechanisms and Metabolic Diversity

The presence of CRISPR/CAS sequences in bacterial strains, particularly acetogenic and butanol-producing strains, highlights a robust defense against phages and foreign genetic elements (**Sheet 2** of **Appendix Ex3**). The diversity in CAS types suggests intricate historical interactions with phages, shaping a dynamic genomic landscape in response to ongoing battles [20–22]. Butanol-producing strains showcase the significance of CRISPR/CAS sequences in safeguarding genetic integrity against external threats. The diverse CAS types indicate varied defense strategies among bacterial groups, showcasing adaptability to ecological niches and selective pressures [23]. An intricate interplay between plasmids and CRISPR/CAS systems is proposed, potentially influencing metabolic diversity and adaptation in bacterial strains [23, 24].

Table 2.2 provides a quantitative perspective on genetic diversity and conservation within acetogenic, butanol-producing, cellulolytic, and CO_2-fixating bacterial groups. Acetogenic strains exhibit a smaller pan-genome, relying on a core set of genes, indicating higher genetic conservation. In contrast, butanol-producers display a larger pan-genome, influenced by plasmids, signifying wider genetic diversity contributing to diverse metabolic capabilities. Cellulolytic strains, with smaller genomes, emphasize cellulose degradation from specific chromosomal genes, underscoring the

importance of core genomic content. CO_2-fixating strains exhibit a smaller pan-genome, aligning with their specialized metabolic focus on carbon fixation.

2.3.3.3. Identification of Biosynthetic Gene Clusters (BGCs)

The identification and characterization of biosynthetic gene clusters (BGCs) in bacterial genomes have the potential to provide access to a wide range of natural products with potential application in various fields, including medicine, agriculture, and biotechnology. BGCs are the group of genes that are responsible for the production of specific natural products in bacteria [25, 26]. Relatively high numbers of NPRS (Non-ribosomal peptide synthetase cluster) were detected in cellulolytic and solventogenic strains, viz. *Csap2*, *Csp1*, *Cce*, *Cpsc*, etc. Bacteriocin or other unspecified ribosomally synthesized and post-translationally modified peptide products (RiPP) clusters were also found in relatively high numbers in cellulolytic and acetogenic strains, viz. *Cce*, *Csa1*, *Cpsc*, and *Csa2*. **Figure 2.5** shows the heatmap of the presence of various BGCs in all genomes.

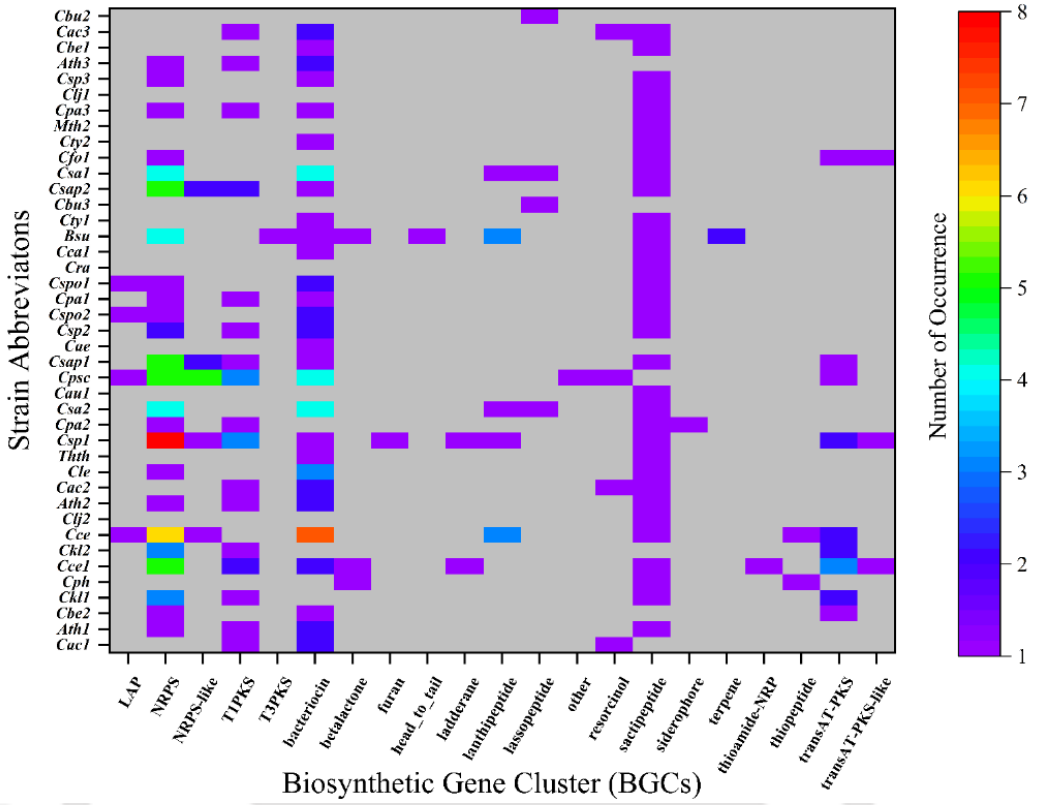


Figure 2.5: Heatmap of biosynthetic gene clusters (BGCs) in all selected bacterial genomes.

2.3.4. Genes for Specific Metabolic Activity

The analysis of genes related to substrate utilization, carbon fixation, transport, and metabolism reveals intriguing patterns among the bacterial strains (see **Table 2.3**). Notably, acetogenic and CO_2 fixating strains exhibit a lower percentage of genes associated with substrate utilization in category G (**Sheet 6** of **Appendix Ex2**), suggesting that they have a limited set of genes for utilizing various substrates. In contrast, cellulolytic strains possess a higher percentage of genes in this category, aligning with their role in breaking down complex polysaccharides like cellulose. Furthermore, the number of genes encoding enzymes in central carbon metabolism is lower in acetogenic strains compared to cellulolytic, and butanol producing strains.

2.3.4.1. Carbon Fixation and Metabolism

The investigation into genes related to carbon fixation and metabolism (see **Sheets 5** and **6** of **Appendix Ex3**) reveals interesting findings. Certain strains, such as *Cdr1* and *Cpa2*, possess multiple copies of enzymes involved in carbon fixation, such as acetyl-CoA C-acetyltransferase and anaerobic carbon-monoxide dehydrogenase catalytic subunit. Similarly, in carbon metabolism, strains like *Cca1* and *Cbu3* harbor the highest number of copies for enzymes like butyryl-CoA dehydrogenase and transaldolase. Moreover, in pentose metabolism, strains *Cty1* and *Cty2* stand out with the highest enzyme numbers for UTP-glucose-1-phosphate uridylyltransferase. These findings suggest distinct metabolic strategies and potential adaptation to specific environmental conditions among these strains.

2.3.4.2. Lignocellulose and Furfural Degradation

The analysis of furfural degradation enzymes reveals that the enzyme 2,5-furandicarboxylate decarboxylase 1 (*hmfF*) is only present in *Dca*, *Csp1*, and *Mth2* strains (**Sheet 9** of **Appendix Ex3**). This enzyme plays a crucial role in furfural degradation [27, 28]. The limited distribution of this enzyme across strains suggests that these particular strains may have specialized capabilities for furfural detoxification, which could be a valuable feature for biotechnological applications [29, 30].

2.3.4.3. Nitrogen Fixation and Sulfur Reduction

The investigation into nitrogen fixation genes highlights the presence of various nitrogen-fixing enzymes across multiple strains (**Sheet 10** of **Appendix Ex3**). These enzymes are

essential for converting atmospheric nitrogen into biologically useful forms. The strains analyzed possess genes coding for enzymes like glutamate synthase, nitrite reductase, glutamine synthetase, and nitrogenase molybdenum-iron proteins, which contribute to nitrogen fixation [31, 32]. Similarly, genes involved in sulfur reduction, such as thiosulfate/3-mercapto pyruvate sulfurtransferase and cysteine synthase, are also present in the analyzed strains [33]. These findings suggest the potential of these strains to participate in nutrient cycling and impact sulfur-rich environments [34].

2.3.4.4. Acidogenesis and Solventogenesis (Biphasic Fermentation)

All ABE-producer strains produce acetate and butyrate as primary acids (see **Figure 2.1**). These acids are synthesized through a series of enzymatic reactions, with acetate formation involving acetate kinase (*ack*) and phosphotransacetylase (*pta*), and butyrate formation initiated by thiolase. The *bcs* cluster, responsible for butyryl-CoA synthesis, is conserved across strains [35, 36]. Additionally, solventogenic strains carry lactate dehydrogenase genes, albeit lactate production occurs under specific conditions (Xue et al., 2016). The presence of pyruvate decarboxylase genes (*pdc*) in certain strains suggests their involvement in solventogenesis [37]. These findings indicate a metabolic potential for solvent formation among the analyzed strains.

2.3.4.5. Energy Generation and Disposal of Reducing Equivalents

All strains possess genes for ATP synthesis via glycolysis and acetate/butyrate formation (**Sheets 4** and **18** of **Appendix Ex2** and **Appendix Ex3**, respectively). However, the presence of energy-conserving hydrogenase (*ech*) genes varies. The clade2 strains,

including cellulolytic strains, host *rnf* genes capable of generating ion gradients across the cytoplasmic membrane. This gradient leads to additional ATP synthesis via F_1F_0-ATPase. In contrast, subclade 1.1 strain lack *rnf* genes. The presence or absence of these genes has potential implications for the metabolic capabilities of these strains, particularly in energy conservation and the disposal of reducing equivalents [38].

2.3.5. Genes for Regulations

All strains host the global regulators viz., *ccpA*, *spo0A*, *rex*, and *codY* genes [39–41]. The *ilvB* operon in *Bsu*, which is responsible for the biosynthesis of branched-chain amino acids, is regulated by multiple mechanisms [42].

2.3.5.1. Stress Response and Sporulation

The heat shock responses, particularly involving major heat shock proteins (HSPs), are extensively conserved across diverse organisms, suggesting their significance in stress adaptation [43, 44]. HSPs, classified under COG category O, function as molecular chaperones and proteases. The DNA-binding regulator (COG2002) governing stationary/sporulation/toxin gene expression was prevalent in cellulolytic and solventogenic strains. The sporulation process in *Clostridium* strains involves coordinated expression of sigma factors (*sigH, sigF, sigE, sigG,* and *sigK*) akin to *Bacillus*, along with the repressor *abrB*. Comprehensive gene presence/absence analysis in supplementary Excel sheets (**Appendix Ex2** and **Appendix Ex3**) details the abundance of genes related to heat shock responses, overall regulators, sporulation, and sigma factors across strains.

2.3.5.2. Chemotaxis and Quorum Sensing

Bacteria employ chemotaxis and quorum sensing for environmental responsiveness and intercellular communication. Butanol-producing clostridia, resembling *Bacillus spp*., exhibit distinct quorum sensing patterns, particularly the presence of *agr* and RNPP-type systems. Notably, *Csap* genomes feature 4 putative *agr* systems and 5 RNPP-type systems, while *Cac* strains possess 8 RNPP-type systems and one *agr* locus. In contrast, *Cbe* and *Csa* strains exhibit up to 6 and 3 *agr* systems, respectively, lacking a complete RNPP-type system. This divergence in intraspecies interaction mechanisms and communication abilities is highlighted in supplementary Excel sheet **Appendix Ex3** (**Sheets 16** and **17**), providing a nuanced understanding of the physiological similarities and evolutionary trajectories of these species [43, 45, 46].

2.3.6. Proposed Evolutionary History of Clostridial Strains in the Present Study

Our analysis has shed light on the intriguing evolutionary history of 48 bacterial strains associated with energy generation considered in our study. Notably, butanol-producing strains were revealed to occupy an ancestral position in the phylogenetic tree. Nearly all butanol producers possessed plasmids housing essential solvent production genes, potentially influencing their metabolic evolution. Moreover, these strains commonly harbored CRISPR/CAS systems, a defense against bacteriophages. The co-occurrence of plasmids and CRISPR/CAS systems suggested an ongoing genetic arms race, where plasmids provided metabolic advantages while CRISPR/CAS systems protected against genetic threats.

The evolutionary path of butanol-producing strains diverged into acetogenic, cellulolytic, and CO_2-fixating strains. Cellulolytic strains were found to lack plasmids, whereas acetogenic and CO_2-fixating strains often retained them, implying plasmids' role in metabolic strategies. Our analysis of CRISPR/CAS genes also reflected these trends (as depicted in **Table 2.4**). Notably, cellulolytic strains, except *Ath1*, *Ath3*, and *Cce1*, often lacked CRISPR/CAS systems, suggesting alternative phage defense mechanisms. These findings representing the evolutionary trajectory is depicted schematically in **Figure 2.6**, which outline the shared metabolic features and potential evolutionary paths of the strains in our study cohort.

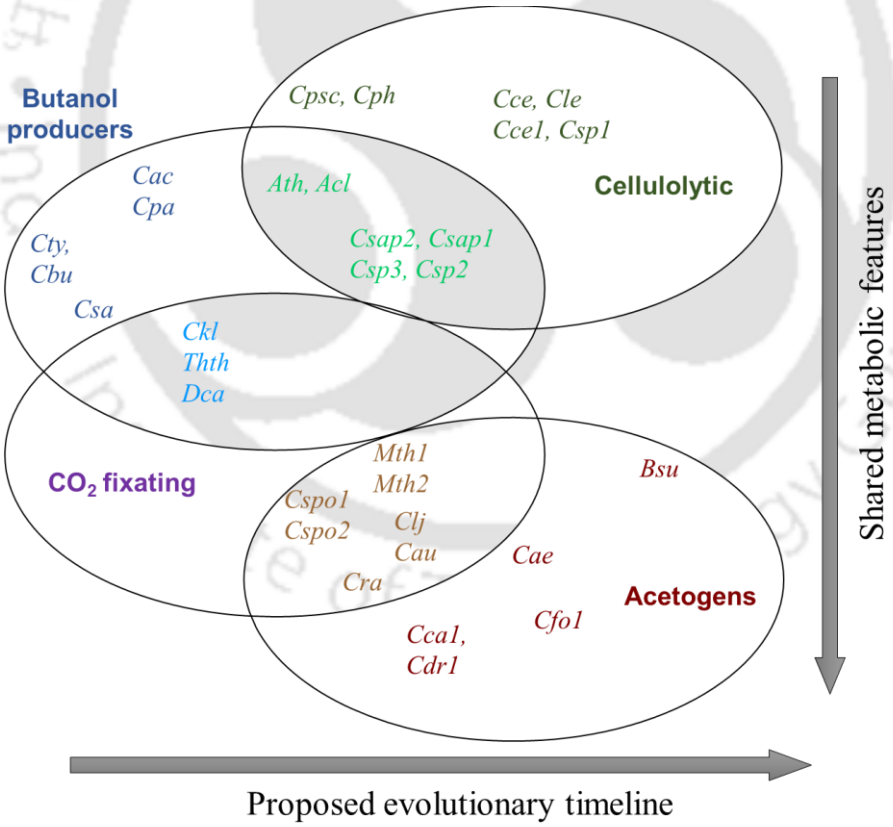


Figure 2.6: Schematic evolutionary timeline and shared metabolic features of analyzed strains

2.4. CONCLUSIONS

This comparative genomic analysis of 48 bacterial genomes across seven genera unveils the genetic foundations driving metabolic diversity. The study refines strain classifications into acetogenic (subset of CO_2-fixating), butanol-producing, cellulolytic, and CO_2-fixating categories, correcting misclassifications like *Csp1* and *Cspo1*. Notably, plasmids' distribution suggests potential metabolic advantages, aiding adaptation to varied ecological niches. Conversely, prevalent CRISPR/CAS systems highlight the imperative to safeguard genomic integrity in dynamic environments. Quantitative core and pan-genome analyses robustly support these insights, emphasizing genetic diversity and the conservation of crucial elements within each group. These findings lay a solid groundwork for developing efficient microbial platforms crucial for sustainable bioenergy production. In essence, the study unveils the intricate interplay of extrachromosomal elements, defense mechanisms, and metabolic diversity in clostridia bacteria. Future investigations into plasmid-carried genes and their interactions with CRISPR/CAS systems promise profound insights into the metabolic strategies and evolutionary histories of these resilient bacterial cohorts. Further investigation, including the study of strains with sequenced plasmids, would be valuable in verifying the conjectures resulted from this study, and providing robust evidence for the proposed link between plasmids and specific traits.

REFERENCES

1. Diallo M, Kengen SWM, López-Contreras AM (2021) Sporulation in solventogenic and acetogenic clostridia. Appl Microbiol Biotechnol 105:3533–3557. https://doi.org/10.1007/s00253-021-11289-9
2. Duerre P (2005) Handbook on Clostridia. CRC Press, Boca Raton
3. Dürre P (2011) Fermentative production of butanol—the academic perspective. Curr Opin Biotechnol 22:331–336. https://doi.org/10.1016/j.copbio.2011.04.010
4. Parte ACY (2018) LPSN – List of Prokaryotic names with Standing in Nomenclature (bacterio.net), 20 years on. Int J Syst Evol Microbiol 68:1825–1829. https://doi.org/10.1099/ijsem.0.002786
5. Papoutsakis ET (2008) Engineering solventogenic clostridia. Curr Opin Biotechnol 19:420–429. https://doi.org/10.1016/j.copbio.2008.08.003
6. Birgen C, Dürre P, Preisig HA, Wentzel A (2019) Butanol production from lignocellulosic biomass: revisiting fermentation performance indicators with exploratory data analysis. Biotechnol Biofuels 12:167. https://doi.org/10.1186/s13068-019-1508-6
7. Kumar K, Roy K, Moholkar VS (2021) Mechanistic investigations in sonoenzymatic synthesis of n-butyl levulinate. Process Biochem 111:147–158. https://doi.org/10.1016/j.procbio.2021.09.005
8. Kumar K, Moholkar VS (2023) Mechanistic Aspects of Enhanced Kinetics in Sonoenzymatic Processes Using Three Simultaneous Approaches. In: Moholkar VS, Mohanty K, Goud VV (eds) Sustainable Energy Generation and Storage. Springer Nature Singapore, Singapore, pp 41–57
9. Malani RS, Umriwad SB, Kumar K, et al (2019) Ultrasound–assisted enzymatic biodiesel production using blended feedstock of non–edible oils: Kinetic analysis. Energy Convers Manag 188:142–150. https://doi.org/10.1016/j.enconman.2019.03.052
10. Berlemont R, Martiny AC (2013) Phylogenetic Distribution of Potential Cellulases in Bacteria. Appl Environ Microbiol 79:1545–1554. https://doi.org/10.1128/AEM.03305-12

11. Ortiz de Ora L, Lamed R, Liu Y-J, et al (2018) Regulation of biomass degradation by alternative σ factors in cellulolytic clostridia. Sci Rep 8:11036. https://doi.org/10.1038/s41598-018-29245-5
12. Krüger A, Mueller AP, Rybnicky GA, et al (2020) Development of a clostridia-based cell-free system for prototyping genetic parts and metabolic pathways. Metab Eng 62:95–105. https://doi.org/10.1016/j.ymben.2020.06.004
13. Song Y, Shin J, Jeong Y, et al (2017) Determination of the Genome and Primary Transcriptome of Syngas Fermenting Eubacterium limosum ATCC 8486. Sci Rep 7:13694. https://doi.org/10.1038/s41598-017-14123-3
14. Seemann T (2014) Prokka: rapid prokaryotic genome annotation. Bioinforma Oxf Engl 30:2068–2069. https://doi.org/10.1093/bioinformatics/btu153
15. Zhou Z, Charlesworth J, Achtman M (2020) Accurate reconstruction of bacterial pan- and core genomes with PEPPAN. Genome Res 30:1667–1679. https://doi.org/10.1101/gr.260828.120
16. Nelson WC, Stegen JC (2015) The reduced genomes of Parcubacteria (OD1) contain signatures of a symbiotic lifestyle. Front Microbiol 6:. https://doi.org/10.3389/fmicb.2015.00713
17. Letunic I, Bork P (2021) Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. Nucleic Acids Res 49:W293–W296. https://doi.org/10.1093/nar/gkab301
18. Chen I-MA, Chu K, Palaniappan K, et al (2023) The IMG/M data management and analysis system v.7: content updates and new features. Nucleic Acids Res 51:D723–D732. https://doi.org/10.1093/nar/gkac976
19. Markowitz VM, Chen I-MA, Palaniappan K, et al (2012) IMG: the integrated microbial genomes database and comparative analysis system. Nucleic Acids Res 40:D115–D122. https://doi.org/10.1093/nar/gkr1044
20. Beceiro A, Tomás M, Bou G (2013) Antimicrobial Resistance and Virulence: a Successful or Deleterious Association in the Bacterial World? Clin Microbiol Rev 26:185–230. https://doi.org/10.1128/CMR.00059-12

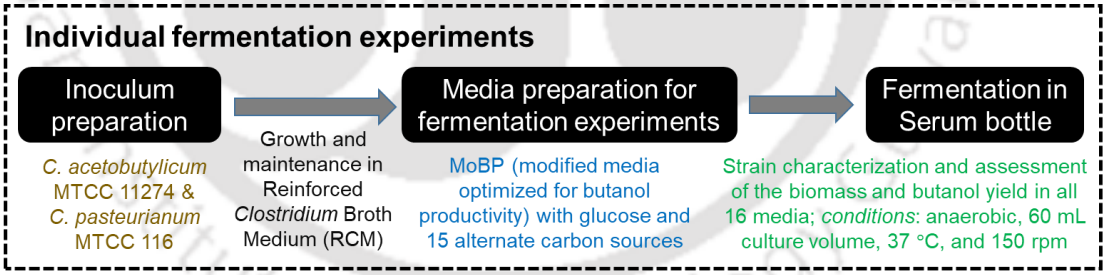
21. Hsu PD, Lander ES, Zhang F (2014) Development and Applications of CRISPR-Cas9 for Genome Engineering. Cell 157:1262–1278. https://doi.org/10.1016/j.cell.2014.05.010
22. Li T, Zhang C, Yang K-L, He J (2018) Unique genetic cassettes in a Thermoanaerobacterium contribute to simultaneous conversion of cellulose and monosugars into butanol. Sci Adv 4:e1701475. https://doi.org/10.1126/sciadv.1701475
23. Loureiro A, da Silva GJ (2019) CRISPR-Cas: Converting A Bacterial Defence Mechanism into A State-of-the-Art Genetic Manipulation Tool. Antibiotics 8:18. https://doi.org/10.3390/antibiotics8010018
24. Hille F, Charpentier E (2016) CRISPR-Cas: biology, mechanisms and relevance. Philos Trans R Soc B Biol Sci 371:20150496. https://doi.org/10.1098/rstb.2015.0496
25. Donia MS, Cimermancic P, Schulze CJ, et al (2014) A Systematic Analysis of Biosynthetic Gene Clusters in the Human Microbiome Reveals a Common Family of Antibiotics. Cell 158:1402–1414. https://doi.org/10.1016/j.cell.2014.08.032
26. Wambui J, Stevens MJA, Sieber S, et al (2022) Targeted Genome Mining Reveals the Psychrophilic Clostridium estertheticum Complex as a Potential Source for Novel Bacteriocins, Including Cesin A and Estercticin A. Front Microbiol 12:801467. https://doi.org/10.3389/fmicb.2021.801467
27. Qin Y-Z, Zong M-H, Lou W-Y, Li N (2016) Biocatalytic Upgrading of 5-Hydroxymethylfurfural (HMF) with Levulinic Acid to HMF Levulinate in Biomass-Derived Solvents. ACS Sustain Chem Eng 4:4050–4054. https://doi.org/10.1021/acssuschemeng.6b00996
28. Shabbir S, Wang W, Nawaz M, et al (2023) Molecular mechanism of engineered Zymomonas mobilis to furfural and acetic acid stress. Microb Cell Factories 22:88. https://doi.org/10.1186/s12934-023-02095-1
29. Lopes AM, Ferreira Filho EX, Moreira LRS (2018) An update on enzymatic cocktails for lignocellulose breakdown. J Appl Microbiol 125:632–645. https://doi.org/10.1111/jam.13923

30. Zhang X, Ye X, Guo B, et al (2013) Lignocellulosic hydrolysates and extracellular electron shuttles for H2 production using co-culture fermentation with Clostridium beijerinckii and Geobacter metallireducens. Bioresour Technol 147:89–95. https://doi.org/10.1016/j.biortech.2013.07.106
31. Tsygankov A (1997) Hydrogen photoproduction by three different nitrogenases in whole cells of Anabaena variabilis and the dependence on pH. Int J Hydrog Energy 22:859–867. https://doi.org/10.1016/S0360-3199(96)00242-X
32. Vyas D (1995) Nitrogen fixation and hydrogen uptake in four cyanobacteria. Int J Hydrog Energy 20:163–168. https://doi.org/10.1016/0360-3199(94)E0008-M
33. Jeong G-S, Kim B-W (1999) The influence of light/dark cycle at low light frequency on the desulfurization by a photosynthetic microorganism. J Biosci Bioeng 87:481–488. https://doi.org/10.1016/S1389-1723(99)80097-8
34. Paritosh K, Kushwaha SK, Yadav M, et al (2017) Food Waste to Energy: An Overview of Sustainable Approaches for Food Waste Management and Nutrient Recycling. BioMed Res Int 2017:1–19. https://doi.org/10.1155/2017/2370927
35. Bennett GN, Rudolph FB (1995) The central metabolic pathway from acetyl-CoA to butyryl-CoA in Clostridium acetobutylicum. FEMS Microbiol Rev 17:241–249. https://doi.org/10.1111/j.1574-6976.1995.tb00208.x
36. Zhao Y, Tomas CA, Rudolph FB, et al (2005) Intracellular Butyryl Phosphate and Acetyl Phosphate Concentrations in *Clostridium acetobutylicum* and Their Implications for Solvent Formation. Appl Environ Microbiol 71:530–537. https://doi.org/10.1128/AEM.71.1.530-537.2005
37. Dürre P, Böhringer M, Nakotte S, et al (2002) Transcriptional regulation of solventogenesis in Clostridium acetobutylicum. J Mol Microbiol Biotechnol 4:295–300
38. Schuchmann K, Müller V (2013) Direct and Reversible Hydrogenation of CO_2 to Formate by a Bacterial Carbon Dioxide Reductase. Science 342:1382–1385. https://doi.org/10.1126/science.1244758
39. Ibberson CB, Jones CL, Singh S, et al (2014) Staphylococcus aureus Hyaluronidase Is a CodY-Regulated Virulence Factor. Infect Immun 82:4253–4264. https://doi.org/10.1128/iai.01710-14

40. Nawrocki KL, Edwards AN, Daou N, et al (2016) CodY-Dependent Regulation of Sporulation in Clostridium difficile. J Bacteriol 198:2113–2130. https://doi.org/10.1128/jb.00220-16
41. Shelburne SA, Olsen RJ, Suber B, et al (2010) A Combination of Independent Transcriptional Regulators Shapes Bacterial Virulence Gene Expression during Infection. PLOS Pathog 6:e1000817. https://doi.org/10.1371/journal.ppat.1000817
42. Shivers RP, Sonenshein AL (2005) Bacillus subtilis ilvB operon: an intersection of global regulons. Mol Microbiol 56:1549–1559. https://doi.org/10.1111/j.1365-2958.2005.04634.x
43. Guo Y, Lu B, Tang H, et al (2019) Tolerance against butanol stress by disrupting succinylglutamate desuccinylase in *Escherichia coli*. RSC Adv 9:11683–11695. https://doi.org/10.1039/C8RA09711A
44. Liu H, Zhang J, Yuan J, et al (2019) Omics-based analyses revealed metabolic responses of Clostridium acetobutylicum to lignocellulose-derived inhibitors furfural, formic acid and phenol stress for butanol fermentation. Biotechnol Biofuels 12:101. https://doi.org/10.1186/s13068-019-1440-9
45. Poehlein A, Solano JDM, Flitsch SK, et al (2017) Microbial solvent formation revisited by comparative genome analysis. Biotechnol Biofuels 10:58. https://doi.org/10.1186/s13068-017-0742-z
46. Stabler RA, He M, Dawson L, et al (2009) Comparative genome and phenotypic analysis of Clostridium difficile 027 strains provides insight into the evolution of a hypervirulent bacterium. Genome Biol 10:R102. https://doi.org/10.1186/gb-2009-10-9-r102

CHAPTER 3

Comparative Analysis of ABE Fermentation in *C.* acetobutylicum and *C. pasteurianum*: A Biochemical & Systems Biological Approach



Part of this chapter is available at: Kumar, K.*, Jadhav, S. M., & Moholkar, V. S. (2023), *bioRxiv* 2023.12.08.570763; https://doi.org/10.1101/2023.12.08.570763

CHAPTER 3

COMPARATIVE ANALYSIS OF ABE FERMENTATION IN *C. ACETOBUTYLICUM* AND *C. PASTEURIANUM*: A BIOCHEMICAL AND SYSTEMS BIOLOGICAL APPROACH

3.1. INTRODUCTION

In Chapter 2, we observed that *Clostridium* species, such as *C. acetobutylicum* (*Cac*) and *C. pasteurianum* (*Cpa*), can metabolize various carbon sources. Previous studies have reported that *Cac* possesses a cellulosome complex responsible for metabolizing complex cellulosic biomass [1]. *Cpa* strains have been widely studied for their ability to produce biohydrogen by metabolizing glycerol waste [2, 3]. *Cac* and *Cpa* have been studied for their potential in biobutanol production.

Cac ATCC 824 has been extensively studied and used as a model organism for biobutanol production. It is known for its ability to produce acetone, butanol, and ethanol through the ABE fermentation process, which shows a characteristic biphasic fermentation [4]. This bacterium has been widely used in industrial-scale biobutanol production due to its high butanol yield and tolerance to butanol toxicity [5]. However, *Cac* has limitations in utilizing specific carbon sources, such as glycerol, as a sole carbon source [6, 7]. On the other hand, *Cpa* has gained attention as a potential alternative to *Cac* for biobutanol production. *Cpa* can produce n-butanol and 1,3-propanediol (1,3-PDO) from the beginning of fermentation, unlike the biphasic process observed in *Cac* [7]. It has been reported that *Cpa* can efficiently utilize glycerol as the sole carbon source for

biobutanol production. This characteristic makes *Cpa* a promising candidate for utilizing glycerol waste from biodiesel production as a feedstock for biobutanol production [6].

This chapter presents the research on comparative biochemical and system biological characterization of *Cac* and *Cpa* for ABE fermentation to understand their gene diversity and organization. This analysis has been correlated with the experimental results of ABE fermentation conducted using the two strains. The concurrent analysis of system biological characterization and ABE fermentation results will give crucial inputs for optimization and upscaling of the ABE fermentation process.

3.2. MATERIALS, METHODS, AND COMPARATIVE GENOME ANALYSIS

3.2.1. Bacterial Strains, Maintenance, and Inoculum Preparation

All chemicals were obtained from HiMedia Pvt. Ltd., India, and were used in their original form. Cultures of *Cac* and *Cpa* with identification numbers MTCC 11274 and MTCC 116, respectively, were sourced from the Microbial Type Culture Collection (MTCC) in Chandigarh, India. The lyophilized cultures of *Cac* and *Cpa* were individually reanimated on Reinforced Clostridial Agar (RCA) plates. The plates were then transferred to a vacuum desiccator, where anaerobic conditions were maintained using an Anaerogas pack (LE002A, HiMedia). Afterward, the cultures were placed in an incubator at 37 °C. In 1 liter of distilled water, the composition of RCA was as follows: peptone 10.0 g, yeast extract 5.0 g, glucose 5.0 g, beef extract 10.0 g, starch 1.0 g, NaCl 5.0 g, L-cysteine HCl 0.5 g, CH_3COONa 3.0 g, and agar 0.5 g. The medium's pH was set to 6.8 ± 0.2 by incorporating 1 M NaOH. To commence the inoculation procedure, a 120 mL serum bottle was utilized, containing 50 mL of RCM medium (RCA without agar). This setup was

positioned within a rotary incubator shaker (Manufacturer: Lab Companion; Model: SI–300R) and subjected to incubation at 37 ℃ with a rotational speed of 150 rpm for a duration of 24 hours. The resulting culture broth served as stock and was subcultured monthly. Overnight-grown cultures of both clostridial strains were individually utilized as inoculants for batch fermentation experiments.

3.2.2. Batch Fermentation with Glucose as Substrate (Base Case)

Batch trials with separate cultures commenced by transferring the inoculum from the mid-log phase of the culture into a 120 mL serum bottle. This bottle contained 60 mL MoBP medium, a modified medium optimized for butanol productivity, as detailed by Ahlawat et al., (2019). The total 60 mL quantity comprised of 52.8 mL of the MoBP medium, 1.2 mL (3% w v^{-1}) L–cysteine HCl, and 6 mL (10% v v^{-1}) of *Clostridium* inoculum. Additionally, 0.1% w v^{-1} Resazurin dye was incorporated into the medium to indicate anaerobic conditions. The MoBP medium consisted of the following components with specified concentrations (g L^{-1}): peptone (49.7), glucose (80.0), KH_2PO_4 (0.5), K_2HPO_4 (0.5), $FeSO_4 \cdot 7H_2O$ (0.023), $MnSO_4 \cdot H_2O$ (0.023), NaCl (5), $MgSO_4 \cdot 7H_2O$ (0.46), $CH_3COONH_4^+$ (2.2), biotin (0.01). Additionally, 10 mL each of trace and vitamin solution were added to 1 L of MoBP medium. The trace element solution consisted of the following components with specified concentrations (in g L^{-1}): H_3BO_3 (0.01), Na_2MoO_4 (0.01), $N(CH_3COO)_3$ (4.5), $AlK(SO_4)_2$ (0.01), $CaCl_2 \cdot 2H_2O$ (0.01), $ZnSO_4 \cdot 7H_2O$ (0.05), $CoCl_2 \cdot 6H_2O$ (0.2), $CuCl_2 \cdot 6H_2O$ (0.05). The composition of the vitamin solution was as follows (concentration in g L^{-1}): folic acid (0.01), riboflavin (0.025), para–amino benzoic acid (0.01), and citric acid (0.02). Before introducing the inoculum, the medium

underwent a 15-minute purge with pure nitrogen (99.99%). Following that, the system was sterilized at 121 °C and 15 psi for 15 minutes in autoclave. Afterwards, the bottles were inoculated with a 10% v v^{-1} inoculum of *Cac* and *Cpa* individual cultures, then incubated at 37 °C and 150 rpm. To ensure result reproducibility, all experiments were replicated in triplicate. Periodic sampling of fermentation broth aliquots was conducted to quantify the glucose consumption over time.

3.2.3. Assessment of Growth of *Cac* MTCC 11274 and *Cpa* MTCC 116 in Alternate Carbon Sources

As noted previously, the hydrolyzates obtained from lignocellulosic biomass (after purification and removal of inhibitors) contain numerous pentose and hexose sugars, in addition to other compounds like sugar alcohols and polysaccharides that can be potential carbon sources. Enhancement of fermentation yield would necessitate effective utilization of all components present in the hydrolysates by the microbial culture. Therefore, in the initial phase of our study, we conducted screenings for the growth profiles and butanol production of each of the two clostridial strains. This involved utilizing a total of 15 different carbon sources, encompassing three three disaccharides (maltose, sucrose, and lactose), three polysaccharides (starch, dextrin, and cellulose), four hexoses (fructose, galactose, mannose, and dextrose), pentoses (arabinose, ribose, and xylose), and two sugar alcohols (glycerol and mannitol). The experimental approach involved substituting glucose with each alternate carbon source in the composite MoBP medium, equivalent to 2.66 M of total carbon, while keeping all other components constant. To prevent any potential transfer of residual glucose from the seed medium to the test medium (with an

alternate carbon source) during inoculation, the inoculum was subjected to centrifugation at 5,000×g for 10 minutes at 4 °C before inoculation and then re-suspended in the respective media. The sampling frequency and experimental conditions were maintained consistently, as detailed in section 2.2. The cell density and butanol concentration in the samples were assessed for all the carbon sources as substrates.

3.2.4. Analytical Methods

A volume of 1.5 mL from the fermentation culture was collected and centrifuged at 10,000×g for duration of 10 minutes at 4 °C. Afterwards, the pellet was washed and re-suspended in saline solution (0.85% w w^{-1} of NaCl). The measurement of growth was performed by assessing the optical density (OD) at 600 nm using a UV-visible spectrophotometer (Cary 50, Varian, Australia). Simultaneously, the supernatant was analyzed to generate dynamic profiles for organic acids (acetate and butyrate), pH, solvents (acetone, butanol, and ethanol), as well as residual substrate using high-performance liquid chromatography (HPLC, Model: Series 200, Manufacturer: Perkin Elmer, USA) equipped with an autosampler. Product separation was achieved through an Aminex-HPX 87H HPLC column (Model: Aminex HPX-87H, Manufacturer: BIO-RAD, USA) operating at 42 °C in column oven. Acetate and butyrate were detected using an UV detector at 210 nm, while solvent products and substrates were measured with a refractive index detector (RID). The estimated minimum detection limits for organic acids, solvents, and substrates were 0.04 g L^{-1}, 0.06 g L^{-1}, and 0.1g L^{-1}, respectively. The correlation standard curves derived from the experiments are provided in the

supplementary material. All measurements, representing the mean values along with their corresponding standard errors, were conducted in triplicate.

3.2.5. Comparative Genomic Analysis of *C. acetobutylicum* and *C. pasteurianum*

A comparative genomic analysis relies on sequencing the DNA of two or more species and then comparing the sequences to identify similarities and differences. This information can be used to infer evolutionary relationships between the species and identify specific genes or regions of the genome responsible for different traits or adaptations. In the context of Clostridial strains, the genomes of the type strain *C. acetobutylicum* ATCC 824 and *C. pasteurianum* ATCC 6013 have been fully sequenced. While both species are known for their capacity to produce biobutanol (as discussed in section 3.1), there are essential differences in their physiology, which makes a comparative genome analysis particularly attractive.

3.2.5.1.Genome Sequence Retrieval

The fully sequenced genomes of two potential butanol-producing strains, *Cac* ATCC 824 and *Cpa* ATCC 6013, were fetched from the RefSeq FTP site (ftp://ftp.ncbi.nlm.nih.gov/, last accessed on 4th February 2022).

3.2.5.2.Genomic Annotation and General Feature Identification

The methodology employed in this study encompassed two key aspects: genomic annotation and general feature identification of *Cac* ATCC 824 and *Cpa* ATCC 6013. Genome annotation is a critical step in comparative genomic analysis, involving the

identification of gene locations, coding regions, and their functions. To ensure consistency and avoid discrepancies, both genomes were reannotated using Prokka v1.12 [17], an automatic annotation pipeline available on GitHub. Subsequently, whole-genomic comparisons were conducted, including the identification of genes falling under various functional categories defined by the Cluster of Orthologous Groups (COGs) database [9–11]. The COGs database is a valuable tool for predicting gene functions and enabling comparative genomics in complete genomes of bacteria and archaea. Additionally, the analysis focused on the presence of essential metabolic features, such as general stress response, sporogenesis, fermentation (acidogenesis and solventogenesis), cellulose degradation, energy and electron transfer, ion and metabolic transport, carbon/nitrogen fixation and metabolism, quorum sensing, and chemotaxis. These features play crucial roles in the physiology and metabolic pathways of the organisms under study. The Integrated Microbial Genomes with Microbiome Samples (IMG/M ER) v.7.0 [12, 13] portal was utilized for comprehensive functional analysis, providing the necessary tools and resources for the identification and analysis of genes falling into various COGs functional categories and metabolic features.

3.3. RESULTS AND DISCUSSION

3.3.1. Comparative Growth Characteristics of *C. acetobutylicum* MTCC 11274 and *C. pasteurianum* MTCC 116

3.3.1.1.Comparative Phenotypic/Biochemical Characterization

The comparative phenotypic or biochemical characterization of *Cac* MTCC 11274 and *Cpa* MTCC 116 are reported in **Table 3.1**. Both of the cells were motile and rod-

shaped. Upon testing, both strains tested Negative for urease production, oxidase production, indole production, lipase production, nitrate production, and H_2S production; however, they were positive for H_2 gas production, which are typical biochemical characteristics of anaerobic *Cac* and *Cpa* strains.

3.3.1.2.Comparative Growth and Fermentation Characterization on Glucose

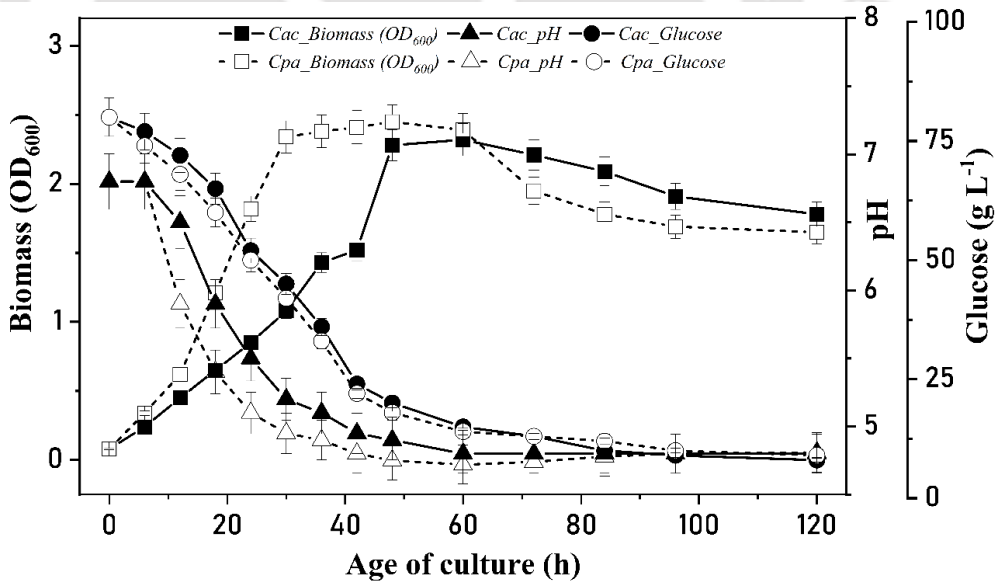
Post-comparative biochemical characterization, both strains were compared for their growth and fermentation characteristics. We compared the strains' growth and ABE production capabilities on MoBP media [8, 14]. The rationale for choosing this media was that the strains' maximum performance is very specific to the medium in which they are grown [15–17].

The growth profiles and fermentation characteristics of *Cac* MTCC 11274 and *Cpa* MTCC 116 were investigated with glucose as the sole carbon source. The cultivation period spanned 120 h, during which their growth and fermentation patterns were monitored. *Cpa* exhibited robust growth, reaching the highest optical density (OD_{600}) of 2.45 at 28 h, followed by a slight decline during the stationary phase. In contrast, *Cac* demonstrated a slower growth rate, with a maximum OD_{600} of 2.3 reached at 36 h. This indicates that *Cac* has a slower glucose growth rate than *Cpa*. The fermentation characteristics were evaluated by measuring the production of acetate, butyrate, and butanol over the cultivation period. Both species exhibited typical ABE fermentation patterns.

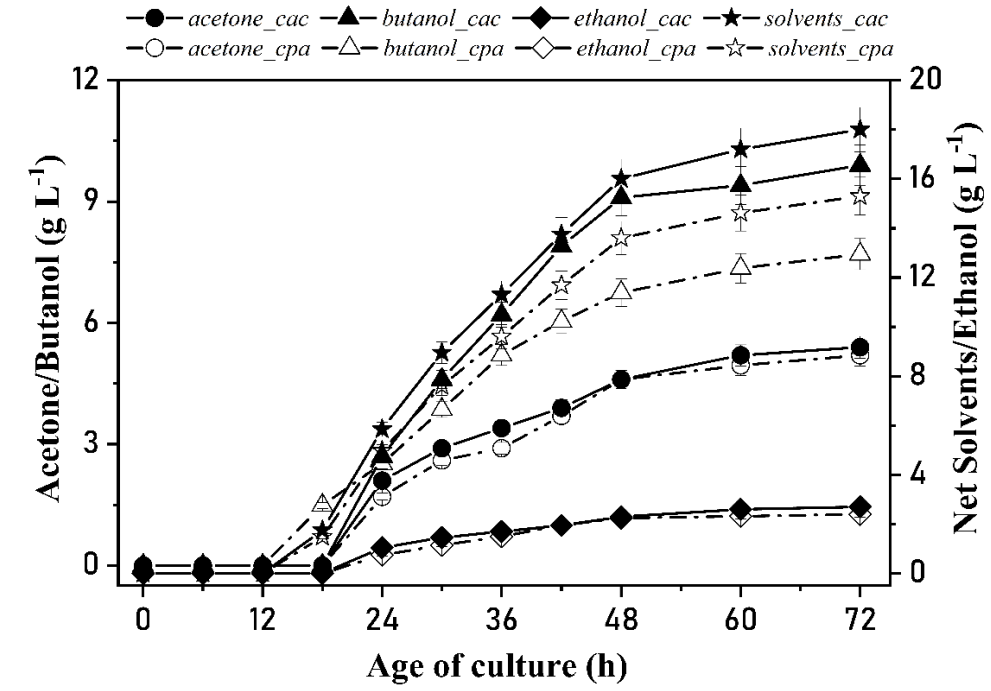
Cac displayed efficient butanol production, reaching a maximum titer of 9.9 g L^{-1} at 40 h. The butanol yield peaked at 0.21 g g^{-1} glucose, indicating a high efficiency in

converting glucose to butanol. Acetate and butyrate were also produced, with maximum titers of 4.7 g L^{-1} and 0.95 g L^{-1}, respectively, at 24 h. On the other hand, *Cpa* showed lower butanol production compared to *Cac*. The maximum butanol titer obtained was 7.7 g L^{-1} at 36 h, and the butanol yield reached 0.16 g g^{-1} glucose. The acetate and butyrate production by *Cpa* were slightly higher, with maximum titers of 5.1 g L^{-1} and 1.1 g L^{-1}, respectively, at 24 h. The time profile for biomass growth, pH change, glucose consumption, solvents (acetone, ethanol, butanol, and total), and acids (acetate and butyrate) production are shown in **Figure 3.1**. The differences in butanol production and fermentation characteristics between the two species highlight their distinct metabolic capabilities. *Cac* demonstrated superior butanol production and higher overall solvent (ABE) yields on glucose compared to *Cpa*. These findings suggest that *Cac* may be more suitable for biofuel production using glucose as a carbon source.

(A)



(B)



(C)

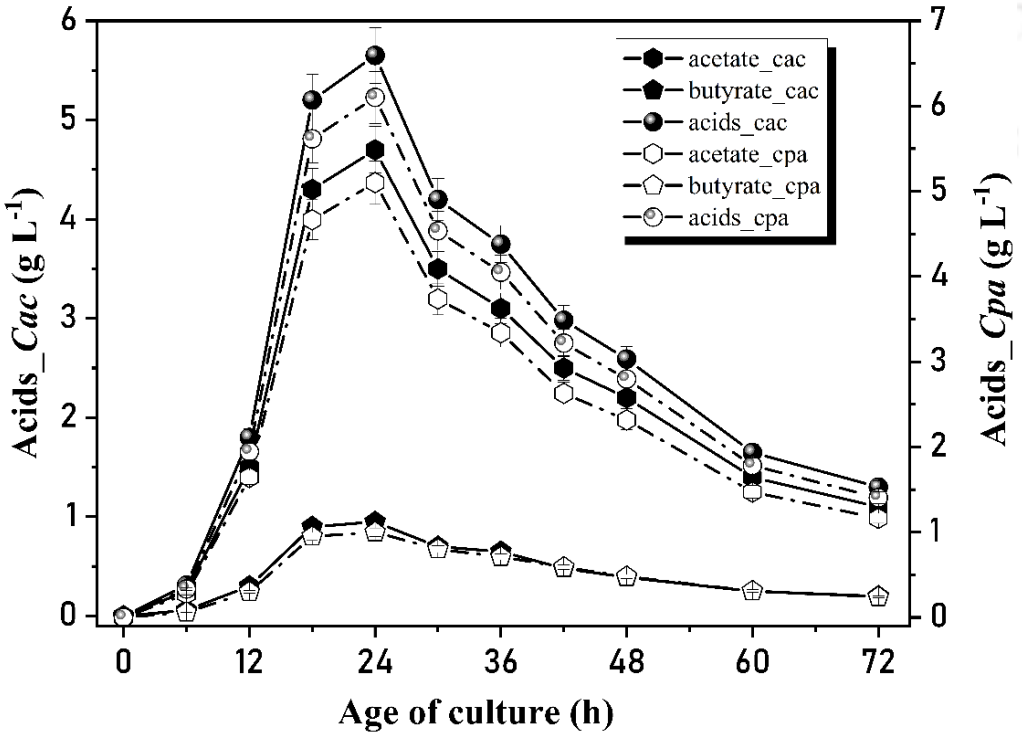


Figure 3.1 Comparative dynamic profiles for (A) biomass, pH, and glucose consumption; (B) solvents (acetone, ethanol, butanol, and total) production; (C) acids (acetate and butyrate) production.

Table 3.1 Phenotypic characteristics of *Cac* MTCC 11274 and *Cpa* MTCC 116

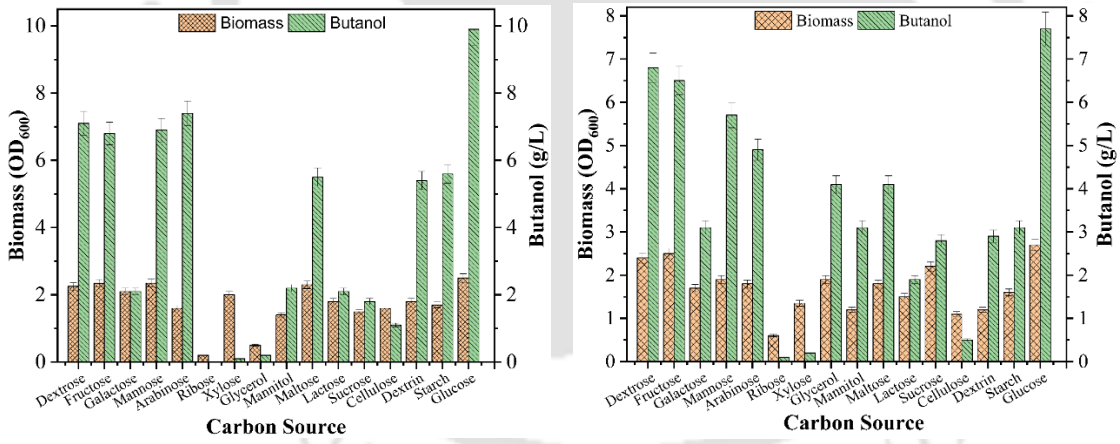
Phenotypic traits	***Cac* MTCC 11274**	***Cpa* MTCC 116**
Cell morphology	Rods	Rods
Spore induction	Positive	Positive
Motility	Positive	Positive
Oxidase test	Negative	Negative
Urease production	Negative	Negative
Indole production	Negative	Negative
Lipase production	Negative	Negative
Nitrate reduction	Negative	Negative
H_2 production	Positive	Positive
H_2S production	Negative	Negative
Optimum growth temperature	37	37
Optimal pH range	4.8-6.8	5-7
Optimum shaking	150 – 160 rpm	150 – 160 rpm
Major fermentation products	Acetone, butanol, ethanol, acetic acid, butyric acid, H_2, and CO_2	Acetone, butanol, ethanol, acetic acid, butyric acid, H_2, and CO_2

3.3.1.3.Comparative Growth and Butanol Production on Various Carbon Sources

In this section, we compare growth and butanol production of *Cac* and *Cpa* on various carbon sources. The aim was to gain insights into their metabolic capabilities and potential for biobutanol production under 15 different carbon sources. Both *Cac* and *Cpa* demonstrated robust growth on C_6 simple sugars, with slight variations in growth rates observed between strains and carbon substrates. Notably, glucose and fructose supported the highest growth rates for both species, highlighting their preference for these sugars as

carbon sources. Qualitative and quantitative results are shown in **Table 3.2** and **Figures 3.2A** & **B**, respectively.

Next, we assessed the butanol production capabilities of *Cac* and *Cpa* during fermentation on the same set of carbon substrates. Strikingly, we observed distinct patterns in butanol production between the two species. *Cac* exhibited higher butanol yields on C_6 sugars, with peak production of 9.9 g L^{-1} (glucose) and 7.7 g L^{-1} (dextrose). *Cac* showed poor butanol production on C_5 sugars and sugar alcohols. In contrast, *Cpa* displayed superior butanol production on sugar alcohols, reaching a maximum of 4.5 g L^{-1} (glycerol) and 3.8 g L^{-1} (mannitol). These results suggest that the choice of carbon source significantly influences the butanol production capacity of each species.



(A) (B)

Figure 3.2 Comparative characterization of (A) *C. acetobutylicum* MTCC 11274 and (B) *C. pasteurianum* MTCC 116 in terms of growth (orange bar) and butanol titer (green bar) under different carbon sources.

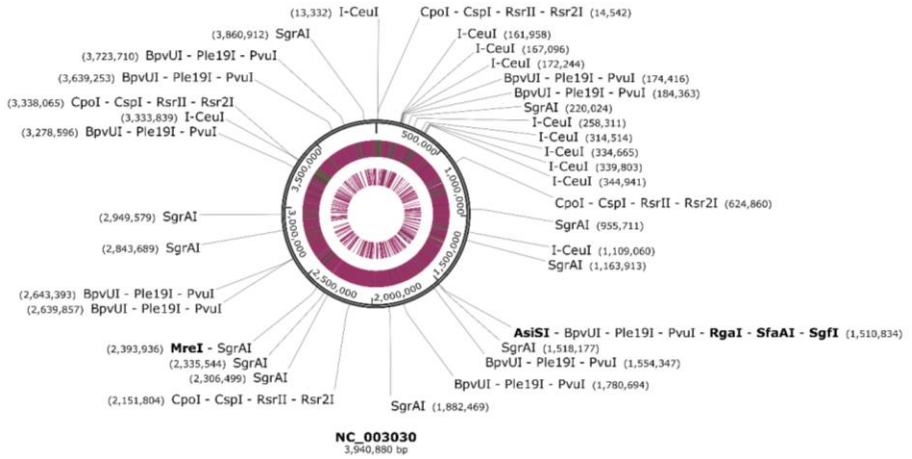
Table 3.2: Growth of *Cac* MTCC 11274 and *Cpa* MTCC 116 on various carbon sources

Carbon source type	***Cac* MTCC 11274**	***Cpa* MTCC 116**
Hexose		
Dextrose	Yes	Yes
Fructose	Yes	Yes
Galactose	Yes	Yes
Mannose	Yes	Yes
Pentose		
Arabinose	Yes	Yes
Ribose	No	No
Xylose	Yes	Yes (Mild)
Sugar alcohols		
Glycerol	No	Yes
Mannitol	Yes	Yes
Disaccharides		
Maltose	Yes	Yes
Lactose	Yes	Yes
Sucrose	Yes	Yes
Polysaccharide		
Cellulose	Yes (Mild)	Yes (Low)
Dextrin	Yes	Yes
Starch	Yes	Yes

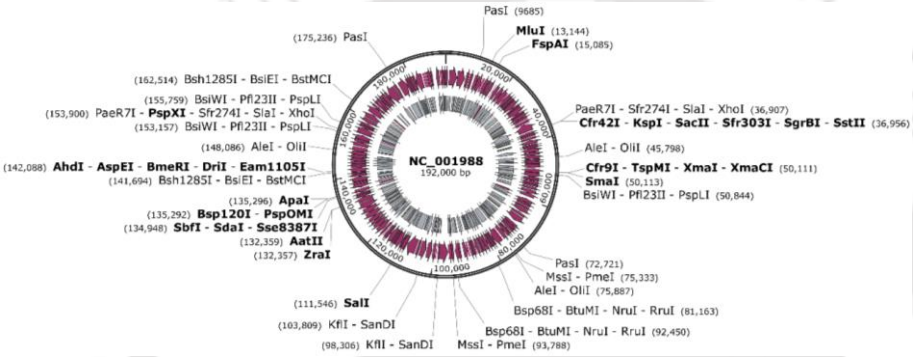
3.3.2. Comparative Genomic Analysis of *C. acetobutylicum* ATCC 824 and *C. pasteurianum* ATCC 6013

This section presents the comprehensive results of the comparative genomic analysis conducted on *Cac* ATCC 824 and *Cpa* ATCC 6013 using their fully sequenced genomes retrieved from the RefSeq FTP, as described in the methodology section. **Figure 3.3** shows the genomic maps of chromosomal and pSOL1 plasmid DNA of *Cac* ATCC 824 and chromosomal DNA of *Cpa* ATCC 6013 visualized in SnapGene software (www.snapgene.com).

(A)



(B)



(C)

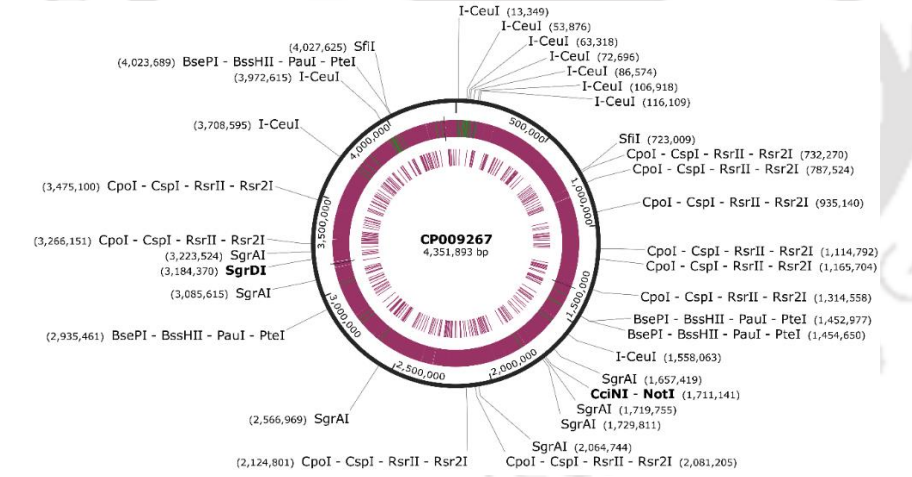


Figure 3.3 Genomic maps of (A) chromosomal DNA of *C. acetobutylicum* ATCC 824, (B) pSOL1 plasmid of *C. acetobutylicum* ATCC 824, and (C) chromosomal DNA of *C. pasteurianum* ATCC 6013 visualized in SnapGene Viewer.

Comparative general genomic features of *Cac* ATCC 824 and *Cpa* ATCC 6013 from the IMG/ M ER portal are summarized in **Table 3.3**.

Table 3.3: General genome feature comparison of *Cac* ATCC 824 and *Cpa* ATCC 6013

Species	***C. acetobutylicum* ATCC 824**		***C. pasteurianum* ATCC 6013**	
IMG Genome ID	637000076		2627854235	
NCBI ID	AE001437, AE001438		CP009267	
Genomic features	**Number**	**%**	**Number**	**%**
Chromosomal DNA				
Genome Size	4132880		4351893	
GC Count	1278098	30.93	1302804	29.94
CDS Count	3848	95.67	3987	96.87
w/ Func Pred Count	3144	78.17	3109	75.53
Gene Count	4022		4116	
CRISPR Count	0		2	
RNA Count	174	4.33	129	3.13
Paralogs Count	2568	63.85	703	17.08
COG Count	2589	64.37	2604	63.27
Enzyme Count	1060	26.36	1135	27.58
KEGG Count	1150	28.59	1219	29.62
Plasmid DNA				
Name	pSOL1		-	
Number of bases	192000		-	
Number of genes	176		-	
GC%		31	-	-

The comparative analysis of COG categories between the two strains provided valuable insights into their functional similarities and differences. We observed that both *Cac* ATCC 824 and *Cpa* ATCC 6013 shared a significant proportion of genes falling into

the COG categories related to central metabolism, information storage and processing, cellular processes and signaling, and poorly characterized proteins (**Figure 3.3**).

However, we noted distinctive patterns in certain COG categories, revealing species-specific genomic features and potential functional variations. Notably, *Cac* ATCC 824 exhibited a higher abundance of genes associated with solventogenesis, stress response, and energy production and conversion compared to *Cpa* ATCC 6013. This aligns with *Cac*'s well-known ability as a proficient solvent producer and its adaptability to changing environmental conditions.

On the other hand, *Cpa* ATCC 6013 displayed an enrichment of genes involved in secondary metabolism, carbohydrate metabolism, and transport systems. These findings suggest that *Cpa* may possess distinct metabolic pathways and preferences for utilizing different carbon sources compared to *Cac*. Furthermore, both strains exhibited similar proportions of genes related to cell motility, cell wall/membrane/envelope biogenesis, and post-translational modifications, indicating shared capabilities in these functional categories.

Overall, the comparative analysis of COG categories provided a detailed picture of the genomic similarities and differences between *Cac* ATCC 824 and *Cpa* ATCC 6013. The identification of species-specific genes and functional disparities contributes to a deeper understanding of their unique metabolic profiles and potential applications in biotechnological processes, particularly in biobutanol production and other industrial applications. Our findings highlight the importance of comparative genomic analysis in unraveling the genetic basis of phenotypic variations and metabolic diversities between closely related bacterial strains. These insights lay the foundation for targeted genetic

engineering, metabolic manipulation, and in-silico investigations using genome-scale metabolic models (shown in **Chapter 5**) to enhance desired traits in both *Cac* and *Cpa*, thereby facilitating the development of more efficient and sustainable bioprocesses for biobutanol production and other bio-based industries.

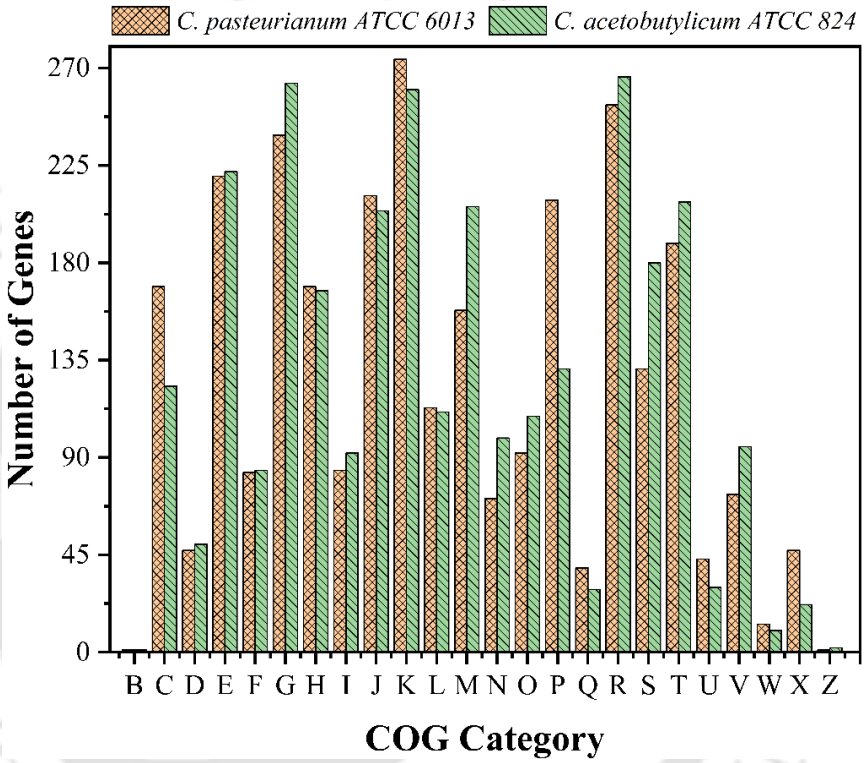


Figure 3.4 Functional COG categories of predicted open ready frames (ORFs) in the genomes of *C. acetobutylicum* ATCC 824 and *C. pasteurianum* ATCC 6013.

COG Categories and their description: C (Energy production and conversion), D (Cell cycle control, cell division, chromosome partitioning), E (Amino acid transport and metabolism), F (Nucleotide transport and metabolism), G (Carbohydrate transport and metabolism), H (Coenzyme transport and metabolism), I (Lipid transport and metabolism), J (Translation, ribosomal structure and biogenesis), K (Transcription), L (Replication, recombination and repair), M (Cell wall/membrane/envelope biogenesis), N (Cell motility), O (Posttranslational modification, protein turnover, chaperones) P (Inorganic ion transport and metabolism), Q (Secondary metabolites biosynthesis, transport and catabolism), R (General function prediction only), S (Function unknown), T (Signal transduction mechanisms), U (Intracellular trafficking, secretion, and vesicular transport), W (Extracellular structures), X (Mobilome: prophages, transposons)

3.3.3. Comparative Analysis on Central Metabolic Pathways

The results from the comparative genomic analysis on genes in central metabolic pathways and their organization in *Cac* ATCC 824 and *Cpa* ATCC 6013 provided comprehensive insights into the metabolic capabilities and regulatory mechanisms of these two *Clostridium* species. The study revealed a diverse set of genes involved in crucial metabolic processes, including glycolysis, the tricarboxylic acid (TCA) cycle, the pentose phosphate pathway (PPP), and other key metabolic pathways (**Figure 3.5**). Both *Cac* ATCC 824 and *Cpa* ATCC 6013 exhibited fundamental genes such as phosphofructokinase, pyruvate kinase, pyruvate dehydrogenase, and acetyl-CoA synthetase, which play pivotal roles in central metabolism.

Moreover, the comparative genomic analysis unveiled the organization of these metabolic genes in the genomes of the two *Clostridium* species. Genes involved in central metabolic pathways were observed to be organized in clusters or operons, implying coordinated regulation and potential functional interplay. For instance, in *Cac* ATCC 824, the genes related to the TCA cycle and PPP were found in operons, suggesting that they might be coregulated to facilitate efficient metabolic flux and energy generation [4, 18]. On the other hand, the study also identified distinct differences in gene organization and composition between *Cac* ATCC 824 and *Cpa* ATCC 6013, implying variations in their metabolic capabilities and phenotypic characteristics [7, 19, 20].

Furthermore, the genomic analysis provided valuable insights into the adaptive responses of these organisms to metabolite stresses. Under general metabolite stress conditions, various stress genes, including dnaK, groES, groEL, hsp90, hsp18, clpC, and htrA, were found to be upregulated, indicating their involvement in stress response and

cellular protection against environmental challenges [21–23]. Additionally, solventogenic operons were upregulated under these stress conditions, emphasizing their crucial role in biobutanol production during unfavorable conditions [24–26].

Notably, the study also revealed intriguing reciprocal regulation between acetate and butyrate metabolism. Under acetate stress, specific genes related to butyryl-CoA and butyrate formation were downregulated, while acetate formation genes were downregulated under butyrate stress. This suggests a complex regulatory network that helps these organisms adapt to varying metabolite stress conditions and maintain metabolic balance [27–29].

The identification of metabolic genes and their organization sheds light on the regulatory mechanisms in these species, paving the way for potential engineering and optimization strategies for sustainable biofuel and chemical production [30, 31]. Moreover, the study of stress response genes and solventogenic operons highlighted the genetic basis of solventogenesis and stress tolerance in *Clostridium* species, which holds promising implications for developing robust recombinant strains for industrial applications in the field of biotechnology.

3.3.4. Diversity of Carbohydrate Active Enzymes

The analysis of carbohydrate active enzymes (CAZymes) in *Cac* ATCC 824 and *Cpa* ATCC 6013 provided valuable insights into their carbohydrate utilization capabilities and potential for lignocellulosic biomass degradation. In both strains, a diverse repertoire of CAZymes was identified, including glycoside hydrolases (GHs), glycosyltransferases (GTs), carbohydrate esterases (CEs), and polysaccharide lyases (PLs). The GH family

was found to be the most abundant, indicating the significance of these enzymes in carbohydrate degradation and utilization (**Table 3.4**).

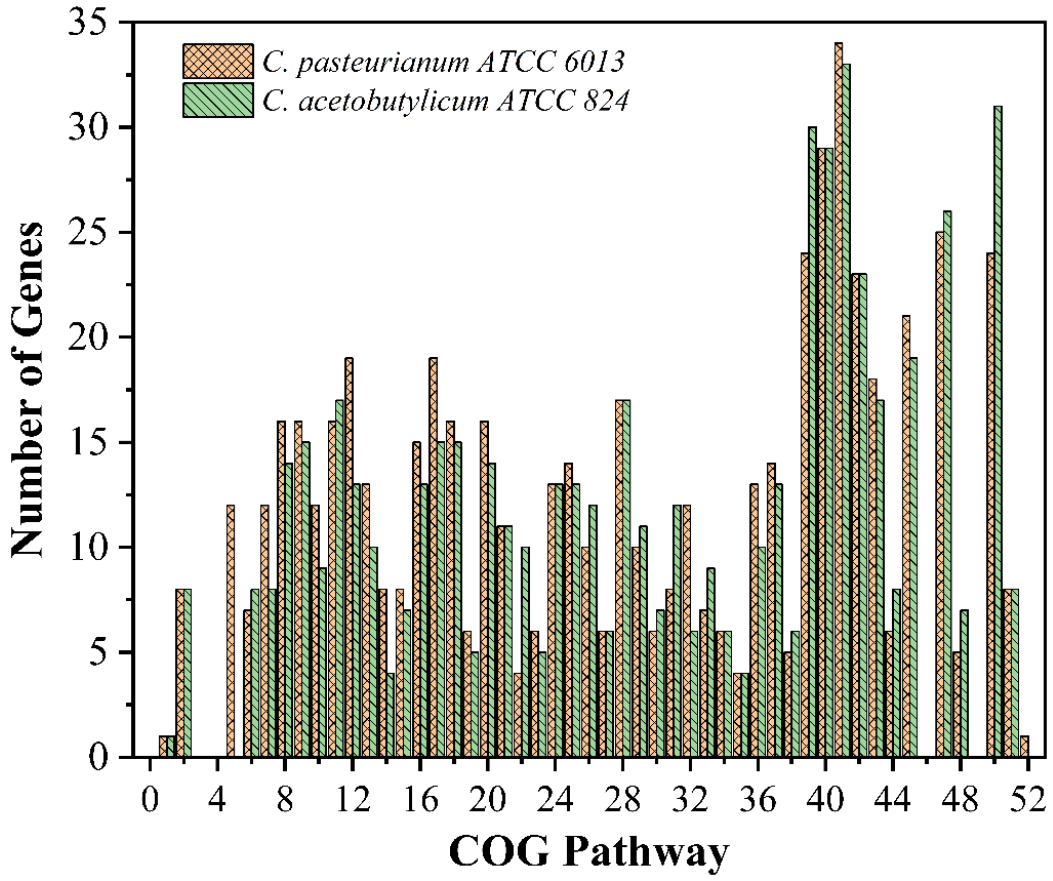


Figure 3.5 Distribution of predicted ORFs in functional COG pathways of the genomes of *C. acetobutylicum* ATCC 824 and *C. pasteurianum* ATCC 6013.

Upon comparing the two strains, we observed significant differences in the distribution of CAZyme families. *Cac* ATCC 824 displayed a higher representation of GH families, particularly those associated with cellulose and hemicellulose degradation (e.g., GH3, GH5, GH9, GH10, and GH43). This finding suggests a greater capacity for

Cac to efficiently hydrolyze complex plant cell wall polysaccharides. Higher abundance of GTs in *Cac* ATCC 824 showed a higher abundance of GTs, which are involved in polysaccharide synthesis and modification. This implies that *Cac* may have distinct carbohydrate utilization strategies and may be more adept at polysaccharide synthesis rather than degradation.

Furthermore, the presence of specific CAZymes associated with lignocellulosic biomass degradation, such as cellulases and hemicellulases, suggests the potential of both strains for utilizing lignocellulosic feedstocks for biobutanol production. However, the higher representation of these CAZymes in *Cac* ATCC 824 indicates its stronger capability to break down complex carbohydrates into fermentable sugars.

The diversity of CAZymes in both strains also points to their adaptability to diverse carbon sources. This versatility could be advantageous for utilizing various feedstocks in bioprocessing, contributing to the sustainability of biobutanol production. The analysis of CAZymes in *Cac* ATCC 824 and *Cpa* ATCC 6013 reveals their distinct capabilities in carbohydrate utilization. *Cac* exhibits a higher abundance of cellulases and hemicellulases, suggesting its superior lignocellulosic biomass degradation potential, which aligns with the observed higher butanol production on various carbon sources.

These findings shed light on the unique carbohydrate utilization strategies of each strain and lay the foundation for further exploring their potential in biobutanol production from different renewable feedstocks. By harnessing their specific carbohydrate-active enzymes, we can enhance the efficiency of biobutanol production and contribute to the development of sustainable biofuel technologies.

Table 3.4 Number of various CAZyme domain sequences in two *Clostridial* genomes

Genome Name	GH	GT	PL	CE	CBM
Cac ATCC 824	78	70	6	13	36
Cpa ATCC 6013	38	44	1	5	7

Abbreviation: GH – Glycoside Hydrolases, GT – Glycosyl Transferases, PL – Glycosyl Transferases, CE – Carbohydrate Esterases, CBM – Carbohydrate-Binding modules

3.4. CONCLUSIONS

In conclusion, the comparative genomic analysis of *C. acetobutylicum* ATCC 824 (or MTCC 11274) and *C. pasteurianum* ATCC 6013 (or MTCC 116) has provided valuable insights into their potential as biobutanol producers and their adaptability to stress conditions. The study revealed distinctive physiological differences between the two species, with *Cac* ATCC 824 demonstrating higher growth rates and butanol production on various carbon sources. The comparative genomic analysis shed light on the organization of genes in central metabolic pathways and their coordination, highlighting the coregulation of key metabolic processes. The analysis also revealed the differential expression of solventogenic and stress response genes under metabolite stress conditions, indicating their importance in cellular protection and butanol production during adverse conditions. The study further uncovered intriguing reciprocal regulation between acetate and butyrate metabolism under specific stress conditions, adding to our understanding of the complex regulatory networks in these organisms.

These findings have significant implications for the field of biotechnology, particularly in the development of sustainable biofuel and chemical production. The

genetic basis of solventogenesis and stress response elucidated through comparative genomic analysis provides a foundation for potential engineering strategies to enhance biobutanol production. The metabolic capabilities and regulatory mechanisms uncovered in this study offer crucial knowledge for the design and optimization of biobutanol production processes. By harnessing the metabolic potential of *Cac* ATCC 824 and *Cpa* ATCC 6013, we can make strides towards a greener and more environmentally friendly future.

The outcomes of this research extend beyond the specific *Clostridium* species studied, contributing to the broader advancement of renewable energy technologies and sustainable bio-based industries. The study highlights the potential of these organisms as valuable resources for biofuel production and emphasizes the importance of understanding their metabolic capabilities and regulatory networks. By leveraging the knowledge gained from comparative genomic analysis, we can work towards developing robust recombinant strains with enhanced biobutanol production capabilities. Ultimately, this research brings us closer to achieving a more sustainable and eco-friendly approach to energy production and paves the way for a greener future.

REFERENCES

1. Kovács K, Willson BJ, Schwarz K, et al (2013) Secretion and assembly of functional mini-cellulosomes from synthetic chromosomal operons in Clostridium acetobutylicum ATCC 824. Biotechnol Biofuels 6:117. https://doi.org/10.1186/1754-6834-6-117

2. Sarma S, Dubey VK, Moholkar VS (2016) Kinetic and thermodynamic analysis (with statistical optimization) of hydrogen production from crude glycerol using Clostridium pasteurianum. Int J Hydrog Energy 41:19972–19989. https://doi.org/10.1016/j.ijhydene.2016.08.204

3. Sarma S, Anand A, Dubey VK, Moholkar VS (2017) Metabolic flux network analysis of hydrogen production from crude glycerol by Clostridium pasteurianum. Bioresour Technol 242:169–177. https://doi.org/10.1016/j.biortech.2017.03.168

4. Chang W, Hou W, Xu M, Yang S (2022) High-rate continuous n-butanol production by Clostridium acetobutylicum from glucose and butyric acid in a single-pass fibrous-bed bioreactor. Biotechnol Bioeng 119:3474–3486. https://doi.org/10.1002/bit.28223

5. D. Makut M, O. Obiekezie S, Owuna G (2018) Isolation and Screening for Biobutanol Producing Clostridium Species from the Soil Environment of Keffi Metropolis. South Asian J Res Microbiol 1–6. https://doi.org/10.9734/sajrm/2018/v1i3805

6. Sabra W, Groeger C, Sharma PN, Zeng A-P (2014) Improved n-butanol production by a non-acetone producing Clostridium pasteurianum DSMZ 525 in mixed substrate fermentation. Appl Microbiol Biotechnol 98:4267–4276. https://doi.org/10.1007/s00253-014-5588-8

7. Sabra W, Wang W, Surandram S, et al (2016) Fermentation of mixed substrates by Clostridium pasteurianum and its physiological, metabolic and proteomic characterizations. Microb Cell Factories 15:114. https://doi.org/10.1186/s12934-016-0497-4

8. Ahlawat S, Kaushal M, Palabhanvi B, et al (2019) Nutrient modulation based process engineering strategy for improved butanol production from Clostridium acetobutylicum. Biotechnol Prog 35:e2771. https://doi.org/10.1002/btpr.2771

9. Galperin MY, Wolf YI, Makarova KS, et al (2021) COG database update: focus on microbial diversity, model organisms, and widespread pathogens. Nucleic Acids Res 49:D274–D281. https://doi.org/10.1093/nar/gkaa1018

10. Galperin MY, Kristensen DM, Makarova KS, et al (2019) Microbial genome analysis: the COG approach. Brief Bioinform 20:1063–1070. https://doi.org/10.1093/bib/bbx117

11. Tatusov RL, Koonin EV, Lipman DJ (1997) A genomic perspective on protein families. Science 278:631–637. https://doi.org/10.1126/science.278.5338.631

12. Chen I-MA, Chu K, Palaniappan K, et al (2023) The IMG/M data management and analysis system v.7: content updates and new features. Nucleic Acids Res 51:D723–D732. https://doi.org/10.1093/nar/gkac976

13. Markowitz VM, Chen I-MA, Palaniappan K, et al (2012) IMG: the integrated microbial genomes database and comparative analysis system. Nucleic Acids Res 40:D115–D122. https://doi.org/10.1093/nar/gkr1044

14. Kaushal M, Ahlawat S, Mukherjee M, et al (2017) Substrate dependent modulation of butanol to ethanol ratio in non-acetone forming Clostridium sporogenes NCIM 2918. Bioresour Technol 225:349–358. https://doi.org/10.1016/j.biortech.2016.11.059

15. Khamaiseh EI, Abdul Hamid A, Abdeshahian P, et al (2014) Enhanced Butanol Production by *Clostridium acetobutylicum* NCIMB 13357 Grown on Date Fruit as Carbon Source in P2 Medium. Sci World J 2014:1–7. https://doi.org/10.1155/2014/395754

16. Lan Z, Zhao C, Guo W, et al (2015) Optimization of Culture Medium for Maximal Production of Spinosad Using an Artificial Neural Network - Genetic Algorithm Modeling. J Mol Microbiol Biotechnol 25:253–261. https://doi.org/10.1159/000381312

17. Kumar K, Shah H, Moholkar VS (2022) Genetic Algorithm for Optimization of Fermentation Processes of Various Enzyme Productions. In: Optimization of Sustainable Enzymes Production. Chapman and Hall/CRC

18. Yoo M, Nguyen N-P-T, Soucaille P (2020) Trends in Systems Biology for the Analysis and Engineering of Clostridium acetobutylicum Metabolism. Trends Microbiol 28:118–140. https://doi.org/10.1016/j.tim.2019.09.003

19. Charubin K, Bennett RK, Fast AG, Papoutsakis ET (2018) Engineering Clostridium organisms as microbial cell-factories: challenges & opportunities. Metab Eng 50:173–191. https://doi.org/10.1016/j.ymben.2018.07.012

20. Fu H, Yang S-T (2022) Editorial: Development and Application of Clostridia as Microbial Cell-Factories for Biofuels and Biochemicals Production. Front Bioeng Biotechnol 9:

21. Diallo M, Kengen SWM, López-Contreras AM (2021) Sporulation in solventogenic and acetogenic clostridia. Appl Microbiol Biotechnol 105:3533–3557. https://doi.org/10.1007/s00253-021-11289-9

22. Du G, Zhu C, Wu Y, et al (2022) Effects of orphan histidine kinases on clostridial sporulation progression and metabolism. Biotechnol Bioeng 119:226–235. https://doi.org/10.1002/bit.27968

23. Kirk DG, Zhang Z, Korkeala H, Lindström M (2014) Alternative Sigma Factors SigF, SigE, and SigG Are Essential for Sporulation in Clostridium botulinum ATCC 3502. Appl Environ Microbiol 80:5141–5150. https://doi.org/10.1128/AEM.01015-14

24. Amador-Noguez D, Brasg IA, Feng X-J, et al (2011) Metabolome Remodeling during the Acidogenic-Solventogenic Transition in Clostridium acetobutylicum. Appl Environ Microbiol 77:7984–7997. https://doi.org/10.1128/AEM.05374-11

25. Li S, Huang L, Ke C, et al (2020) Pathway dissection, regulation, engineering and application: lessons learned from biobutanol production by solventogenic clostridia. Biotechnol Biofuels 13:39. https://doi.org/10.1186/s13068-020-01674-3

26. Liu H, Zhang J, Yuan J, et al (2019) Omics-based analyses revealed metabolic responses of Clostridium acetobutylicum to lignocellulose-derived inhibitors furfural, formic acid and phenol stress for butanol fermentation. Biotechnol Biofuels 12:101. https://doi.org/10.1186/s13068-019-1440-9

27. Alsaker KV, Paredes C, Papoutsakis ET (2010) Metabolite stress and tolerance in the production of biofuels and chemicals: Gene-expression-based systems analysis of butanol, butyrate, and acetate stresses in the anaerobe Clostridium acetobutylicum. Biotechnol Bioeng 105:1131–1147. https://doi.org/10.1002/bit.22628

28. Dash S, Mueller TJ, Venkataramanan KP, et al (2014) Capturing the response of Clostridium acetobutylicumto chemical stressors using a regulated genome-scale metabolic model. Biotechnol Biofuels 7:144. https://doi.org/10.1186/s13068-014-0144-4

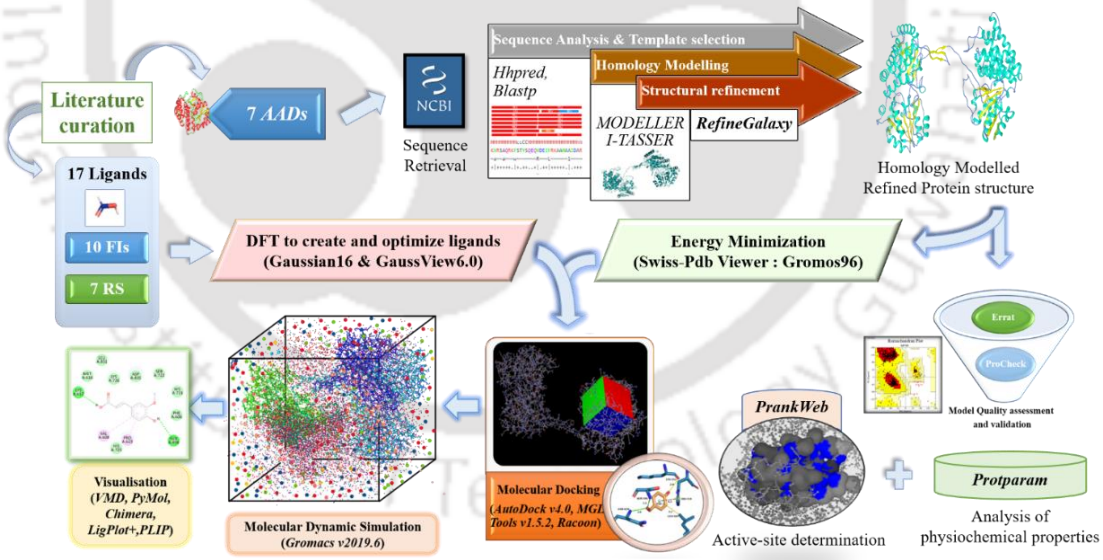
29. Lee J, Jang Y-S, Choi SJ, et al (2012) Metabolic Engineering of Clostridium acetobutylicum ATCC 824 for Isopropanol-Butanol-Ethanol Fermentation. Appl Environ Microbiol 78:1416–1423. https://doi.org/10.1128/AEM.06382-11

30. Chae TU, Choi SY, Kim JW, et al (2017) Recent advances in systems metabolic engineering tools and strategies. Curr Opin Biotechnol 47:67–82. https://doi.org/10.1016/j.copbio.2017.06.007

31. Chen Y, Banerjee D, Mukhopadhyay A, Petzold CJ (2020) Systems and synthetic biology tools for advanced bioproduction hosts. Curr Opin Biotechnol 64:101–109. https://doi.org/10.1016/j.copbio.2019.12.007

CHAPTER 4

Computational Investigations in Inhibition of Alcohol/ Aldehyde Dehydrogenase in Lignocellulosic Hydrolysates



Preprint available at: Kumar, K.*, Azeeza S. MD, Pragati, Chandane, P., Kori, M., Shivram, A., Yadav, S., Barbora, L., & Moholkar, V. S. (2022), *bioRxiv* 2022.11.19.517192; https://doi.org/10.1101/2022.11.19.517192

CHAPTER 4

COMPUTATIONAL INVESTIGATIONS IN INHIBITION OF ALCOHOL/ ALDEHYDE DEHYDROGENASE IN LIGNOCELLULOSIC HYDROLYSATES

4.1. INTRODUCTION

Second-generation alcoholic biofuels make use of hydrolysates from lignocellulosic biomass as fermentation substrates. These hydrolysates are generated by the hydrolysis of the cellulosic and hemicellulosic fractions in biomass [1]. The hydrolysis of hemicellulosic fraction, which generates the hydrolysate rich in pentose sugars like xylose and arabinose, involves high temperature and pressure dilute acid treatment [2, 3]. This process also yields several other organic compounds like acids, aldehydes, ketones, oxides and their phenolic derivatives, furan derivatives, and weak organic acids that are potential inhibitors of some of the crucial enzymes in the metabolic pathway [4, 5]. Bioethanol and biobutanol are potential alcoholic biofuels that can be blended with gasoline [6, 7]. Both can be produced through acetone-butanol-ethanol (ABE) fermentation by the *Clostridial* microbial cultures [8]. The general metabolic pathway of ABE fermentation from pentose and hexose substrates is shown in **Figure 4.1A**. From **Figure 4.1A**, it can be seen that *gap3dh* catalyzes the conversion of glyceraldehyde-3-phosphate to 1,3-bisphosphoglycerate, an essential step in the progression of the glycolysis pathway. Solventogenic enzymes such as *aydh*, *adhE1*, *adhE2*, *bdh*, *bdhA*, *bdhB*, *acetate kinase*, and *butyrate kinase* play a vital role during the ABE fermentation pathway. Bifunctional alcohol/ aldehyde dehydrogenase such as *aydh* catalyzes the conversion of acetyl-CoA to acetaldehyde,

adhE1, and *adhE2* catalyzes the transformation of an aldehyde (e.g., acetaldehyde or butanal) to alcohol (e.g., ethanol and butanol) and vice versa in ABE fermentation pathway. Similarly, alcohol-specific dehydrogenases such as *bdh*, *bdhA*, and *bdhB* also play an essential role in converting butanal to butanol, the final step of butanol formation. The *AAD* enzymes are known to be inhibited by the aforementioned organic compounds present in the acid hydrolysate [4, 5, 9]. This inhibition adversely affects the kinetics of metabolism and the final yield of the bioalcohols. As discussed above, these solventogenic enzymes are potentially inhibited by organic compounds generated during the dilute acid pretreatment of lignocellulosic biomass.

This study aims to discern the molecular mechanism of inhibition of key *AADs* across three solventogenic species, viz., *C acetobutylicum*, *C. beijerinckii*, and *C. pasteurianum*. The objectives of the present study are (1) identification and homology modeling of key *AADs*; (2) validation, quality assessment, and physiochemical characterization of the modeled enzymes; (3) identification, construction, and optimization of the chemical structure of potent microbial inhibitors in LH; and (4) applications of hybrid molecular simulations to profile the interactions between microbial inhibitors and key *AADs*. Our computational investigation has revealed various essential facets of inhibiting the *AAD* enzymes, which could guide structural biologists in designing efficient and robust enzymes. Moreover, our methodology provides a general framework that could be applied to deciphering other enzymes' molecular mechanisms of inhibition behavior.

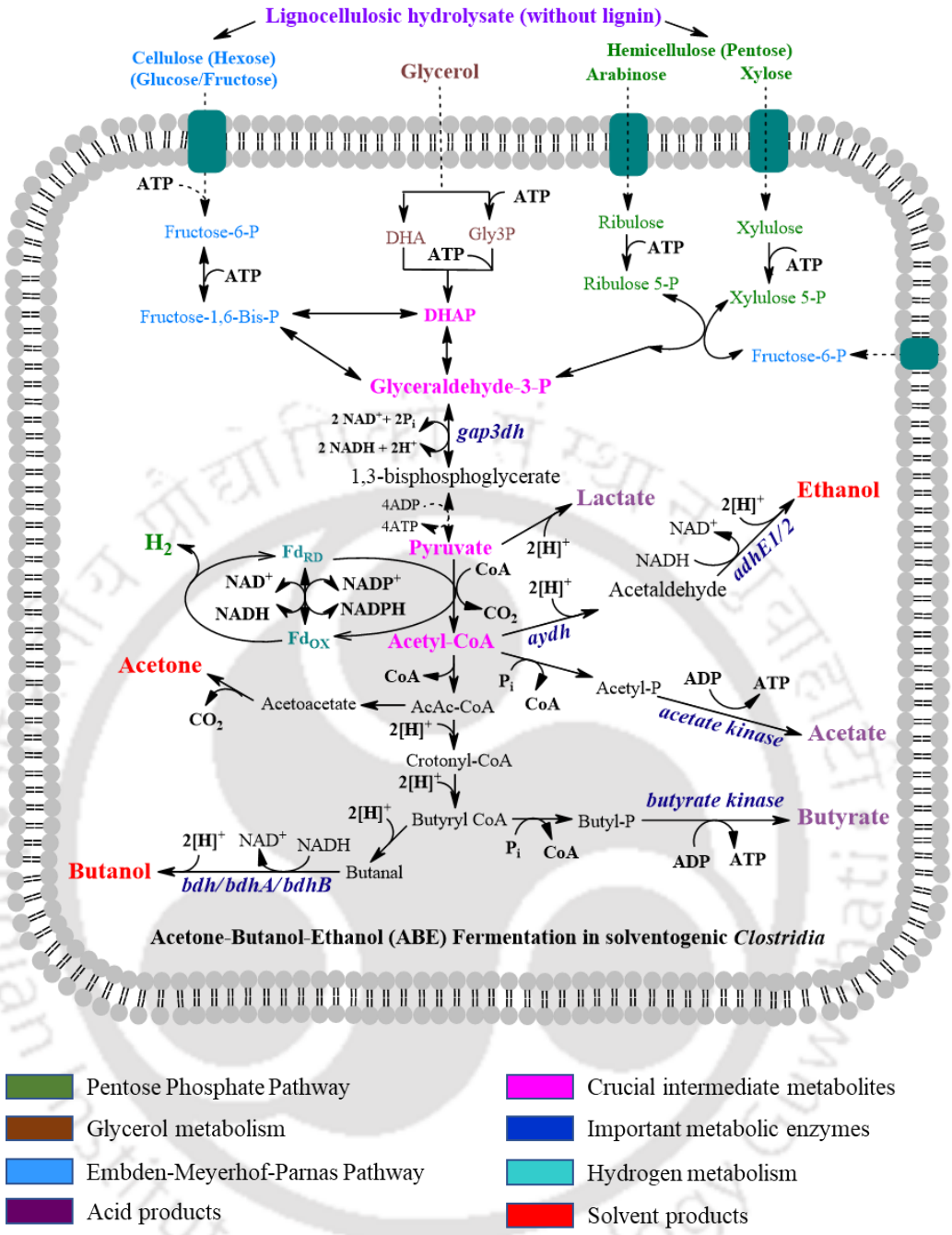


Figure 4.1(A): Representative carbohydrate metabolism in solventogenic *Clostridia*: crucial solvent-producing enzymes, key metabolites, and brief acetone-butanol-ethanol (ABE) fermentation pathway.

4.2. COMPUTATIONAL METHODOLOGY

The schematic representation of the workflow, including general steps followed in this study, is provided in **Figure 4.1B**. Detailed procedure in a step-by-step manner followed is presented in the following sections.

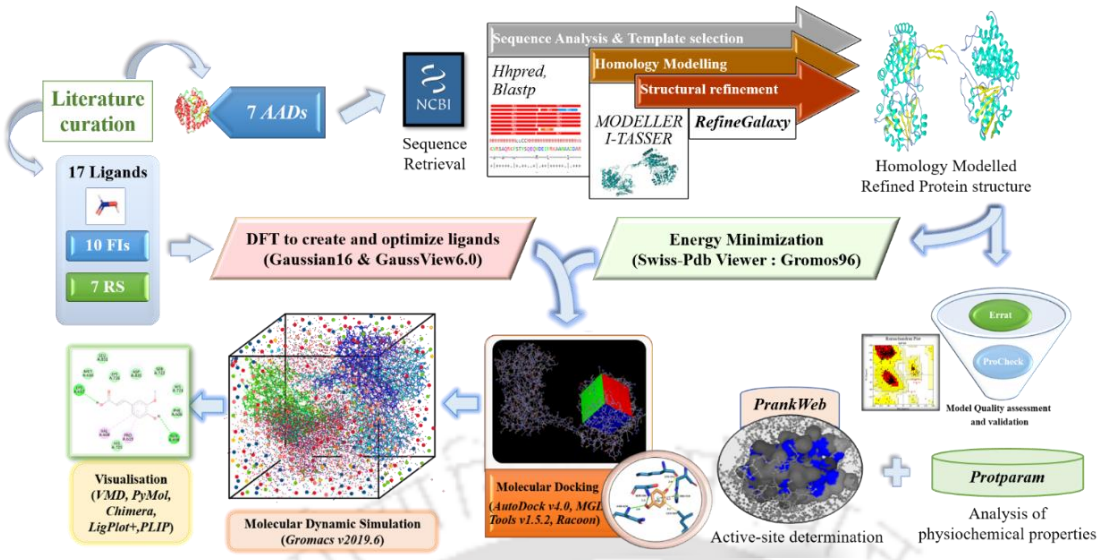


Figure 4.1(B): Schematic workflow and methodology followed in the present study.

4.2.1. Preparation and Analysis of Key AADs in Solventogenic *Clostridia*

Based on previous literature, we identified a total of seven *Clostridial AAD* enzymes (viz. *adhE1, adhE2, aydh, gap3dh, bdh, bdhA, and bdhB*), which are of high importance in solvent formation pathways [10–12]. The crystal structure of these *AADs* is not available in the RCSB-PDB [13]. The metadata of these enzymes, such as source organism, NCBI- RefSeq ID, sequence length, molecular weight, and aliphatic index, is listed in **Table 4.1A**. The exact molecular functions and the structural information, i.e., conserved domains, cofactors, and conserved sites, are described in **Table 4.1B**. In our study, we have classified *AAD* enzymes into four categories: (1) Bifunctional *AAD* enzymes that work on both alcohols and aldehydes (*adhE1* and *adhE2*), (2) butanol-specific *AAD* enzymes (*bdh, bdhA, and bdhB*), (3) biofunctional acetaldehyde dehydrogenase (*aydh),* and (4) glyceraldehyde metabolizing dehydrogenase (*gap3dh*).

Table 4.1(A): List of *Clostridial AAD* enzymes with their respective metadata from NCBI

Enzyme name	Abb.	Source organism	Reference ID	# of aa	MW (kDa)	A.I.
Alcohol/aldehyde dehydrogenase enzyme 1	*adhE1*	*Clostridium acetobutylicum* ATCC824	AAK76781.1	858	94.44	92.65
Alcohol/aldehyde dehydrogenase enzyme 2	*adhE2*	*Clostridium acetobutylicum* ATCC824	AAK09379.1	858	94.44	92.65
NAD(P)H-dependent butanol dehydrogenase A	*bdhA*	*Clostridium acetobutylicum* ATCC824	OOM03255.1	389	42.98	91.26
NAD(P)H-dependent butanol dehydrogenase B	*bdhB*	*Clostridium acetobutylicum* ATCC824	OOM03254.1	390	43.29	97.26
Butanol dehydrogenase	*bdh*	*Clostridium acetobutylicum* ATCC824	AAA23206.1	389	43.07	91.00
Acetaldehyde dehydrogenase	*aydh*	*Clostridium beijerinckii* NCIMB8052	WP_026886587	864	94.93	91.53
Glyceraldehyde-3-phosphate dehydrogenase	*gap3dh*	*Clostridium pasteurianum* ATCC6013	OMH20402.1	334	36.08	92.54

Abbreviation: Abb. – abbreviation; aa – amino acids; MW – molecular weight; A.I. – aliphatic index;

Table 4.1(B): Selected *Clostridial AAD* enzymes: their functions and other structural information

Enzyme	Molecular functions	Conserved domains (residues range)	Cofactors	Conserved site
adhE1/E2	• oxidoreductase activity • metal ion binding • alcohol dehydrogenase (NAD^+) activity • oxidoreductase activity, acting on the aldehyde or oxo group of donors, NAD or NADP as acceptor • butanol dehydrogenase activity • acetaldehyde dehydrogenase (acetylating) activity	Aldehyde dehydrogenase (5-374) Iron-type/glycerol dehydrogenase (454-846) Alcohol dehydrogenase (454-846) Bifunctional aldehyde-alcohol dehydrogenase, C-terminal (453-855)	Zn^{2+}, Fe^{2+}, NAD^+, $NADP^+$	630 – 658 718 – 738
bdh/A/B	• metal ion binding • oxidoreductase activity • butanol dehydrogenase activity	Iron-containing dehydrogenase (9-381)	Zn^{2+}, Fe^2	173 – 201 264 – 284
aydh	• oxidoreductase activity • oxidoreductase activity, acting on the aldehyde or oxo group of donors, NAD or NADP as acceptor • metal ion binding • acetaldehyde dehydrogenase (acetylating) activity • alcohol dehydrogenase (NAD+) activity • butanol dehydrogenase activity	Aldehyde dehydrogenase (5-398) Iron-type/glycerol dehydrogenase (456-847) Alcohol dehydrogenase (454-856) Bifunctional aldehyde-alcohol dehydrogenase, C-terminal (454-856)	Zn^{2+}, Fe^2, NAD^+, $NADP^+$	631-659 719-739

gap3dh	• oxidoreductase activity, acting on the aldehyde or oxo group of donors, NAD or NADP as acceptor • NAD binding • NADP binding	NAD(P)-binding (2-150) NAD-binding (2-102) Catalytic (155-312) C-terminal (155-312)	NAD^+, $NADP^+$

4.2.1.1. Sequence retrieval and secondary structure analysis

The amino acid sequences in FASTA format for all seven *AADs* were retrieved from the NCBI-RefSeq database. Selected metadata information for all seven *AADs* is provided in **Table 4.1A** (extended metadata are provided in **Table A3.1** of **Appendix 3**). Next, the Protparam web tool (https://www.expasy.org/resources/protparam) was used for the physicochemical analysis of retrieved protein sequences. Parameters computed using Protparam in the query amino acid sequence are (1) theoretical molecular weight, (2) molar extinction coefficient, (3) atomic composition, (4) theoretical pI (isoelectric point), (5) number of amino acids (AA), (6) composition of AA, (7) positively charged residues (Asp+Glu), (8) negatively charged residues (Arg+Lys), (9) aliphatic index, (10) grand average of hydropathicity (GRAVY), (11) theoretical half-life, and (12) instability index (**Table A3.**1 of **Appendix 3**).

4.2.1.2. Homology modeling of AAD enzymes

The 3-dimensional (3D) models of *AADs* were generated using MODELLER [14] and I-TASSER [15]. In the case of MODELER, homology modeling templates were selected using HHpred [19], an inbuilt package of the Bioinformatics Toolkit. The HHpred (http://toolkit.tuebingen.mpg.de/hhpred) selects template structures based on Hidden Markov Models (HMM) pairwise profile comparison. Apart from HHpred,

BLASTp against the PDB database was also used to predict the templates for homology modeling. Final templates were selected based on each enzyme's best score and query coverage. In contrast to MODELLER, the I -TASSER webserver requires query sequences in FASTA format as input for generating 3D models. I-TASSER (https://zhanglab.ccmb.med.umich.edu/I-TASSER/) builds 3D models of any protein based on an *ab-initio* modeling approach [16]. The output 3D models of I-TASSER and MODELLER were compared, and the model having the highest confidence score (C-score) was selected. C-score is a quantitative estimate for the superiority of the predicted models by I-TASSER, which typically ranges from -5 to 2. A higher c-score value signifies a model with high confidence and vice versa. A list of templates used for the homology modeling of *AADs* and their metadata is provided in **Table 4.2A**.

Table 4.2(A): List of templates used for homology modeling of *Clostridial AAD* enzymes and respective metadata

Enzyme	Template from BLASTp		% Identity	Tot. score	GDT-HA	RMSD	C-score	P-rota.
	PDB ID	Organism						
adhE1	7BVP_D	*E. coli* K-12	58.93	882.2	0.985	0.295	13.7	0.3
adhE2	7BVP_D	*E. coli* K-12	58.93	882.2	0.984	0.302	12.6	0.8
bdh	1VLJ_A	*T. maritima* MSB8	40.76	298	0.988	0.295	18.6	2.1
bdhA	1VLJ_A	*T. maritima* MSB8	41.09	300	0.988	0.276	14.5	0.3
bdhB	1VLJ_A	*T.* MSB8	38.89	280	0.972	0.363	32.4	1.8
aydh	7BVP_D	*E. coli* K-12	62.43	1102	0.981	0.308	13.6	0.0
gap3dh	6FZI_A	*C. perfringens* SM101	79.20	504	0.976	0.340	14.7	0.4

Abbreviation: *E. coli* – *Escherichia coli*; *T. maritima* – *Thermotoga maritima*; *C. perfringens* - *Clostridium perfringens*; Tot. sacore – total score; GDT-HA - ; RMSD – root mean square deviation; C-score – clash score; p-rota. – poor rotamers;

4.2.1.3. Refinement and validation of 3D structures of AADs

The selected enzyme models in the previous step were subjected further to model refining in GalaxyRefine (http://galaxy.seoklab.org/cgi-bin/submit.cgi?type=REFINE) to improve the quality. GalaxyRefine is an online server based on a CASP10 refining process proven effective for model refining [17]. GalaxyRefine uses small molecular dynamics simulations to reconstruct all side-chain conformations and relax the structure repeatedly. This method improves global and local structure quality as determined by GDT-HA and root mean square deviation (RMSD). In the next step, Molprobity was used to evaluate the quality of the backbone structure of the final 3D models from GalaxyRefine. After the structural refinement and energy minimization steps, the best 3D models (selected based on the Ramachandran score) were validated using various 3D structural parameters. Model validations were performed using Structural Analysis and Verification Server (SAVES v6.0: https://saves.mbi.ucla.edu/). Ramachandran plot analysis was carried out using PROCHECK to assess the quality of the final models. The stereochemical and geometrical parameters of the model were checked using the PROCHECK [18]. Moreover, the ERRAT server was used to examine the quality of the final 3D structure by determining the compatibility of an atomic model (3D) with its amino acid sequence and comparing the findings to standard structures [19]. After validation step, CASP14 (Critical Assessment of Structure Prediction) ranked web servers such as PrankWeb [20] (available on P2Rank), InterPro, 2Struc, and NetSurfP3.0 were used for identifying the ligand-binding pockets, putative catalytic residues, putative active site residues, putative metal-binding site, co-factor binding residues, secondary structural composition, and percentage surface accessibility analysis.

4.2.2. Selection of Inhibitors and their Preparation for Hybrid Molecular Simulations

As noted earlier, the dilute acid pretreatment of LB produces many organic compounds, such as aliphatic/aromatic acids, aldehydes, and furans, which act as metabolic and growth inhibitors and reduce the kinetics and yield of ABE fermentation. The toxic compounds also strongly inhibit the growth of vital cell structures and the metabolic capacity of fermenting strains.

4.2.2.1. Identification and selection of inhibitors and representative substrates of AADs

Among various organic compounds produced during acid hydrolysis of LB, ten compounds were identified as inhibitors of bacterial cell growth and enzymes in the solventogenic pathway of metabolism [4, 21, 22]. These inhibitors are 4-hydroxy-benzaldehyde (4-HBZ), benzaldehyde (BZD), cinnamaldehyde (CMD), syringaldehyde (SGN), ferulic acid (FA), levulinic acid (LA), vanillic acid (VA), *ρ*-coumaric acid (*ρ*-CA), vanillin (VN), and 5-hydroxymethylfurfural (HMF or 5-furfural). In addition, the representative substrates were identified for each *AAD* as follows: methanol (MET), ethanol (ETH), butanol (BUT), formaldehyde (FMD), acetaldehyde (ACD), butyraldehyde (BLD), and glyceraldehyde-3-phosphate (G-3-P). Hereon, the inhibitory compounds are called “fermentative inhibitors (FI)”. Further, the FIs in combination with representative substrates (as mentioned above) will be termed “ligands”.

4.2.2.2. Designing the structures of ligands and optimizing their geometry using Density Functional Theory (DFT)

As mentioned earlier, the chemical structure of ligands was designed in GaussView v6.0 software [23]. Drawn structures in GaussView are not energetically favorable. Hence, energy minimization is critical for establishing the optimal molecule arrangement in 3D space. The 3D geometry of ligands was optimized using DFT at B3PW91/6-311G(+)(d, p) level of theory followed by frequency calculation (freq = raman) in Gaussian16 software [24]. The molecule's potential energy components, such as stretching, bending, and torsion, which are essential in defining its behavior, were optimized using Gaussian16. Finding the energy minima (i.e., the potential energy hypersurface of a stable molecule) helped improve the structure's strain energy and, in turn, led to the most stable conformer. The most stable conformer of the ligands was converted into the *"pdbqt"* format for further usage.

4.2.3. Molecular Docking Simulation and Analysis

Before performing molecular docking, 3D models of *AAD* enzymes were subjected to energy minimization using SWISS-PDB Viewer (SPDBV [25]) to achieve local minima closer to the native structure. The molecular docking simulations investigated biomolecular interactions, bound conformations, binding free energy of ligands and macromolecules, and identifying the active site residues using Autodock linked with the MGLTools version 1.5.6 [26]. The docking simulations of 17 ligands (combinations of 10 FIs and seven representative substrates) against seven *AADs* were performed using Autodock v1.5.6.[27] The format of ligand structures generated through Gaussian16 was converted into *"pdbqt"* format using OpenBabel software. The procedure for molecular docking was followed as described by Kumar et al.

(2021) [28]). The polar hydrogen ions, and charges such as Kollman [29] and Gasteiger [30], were added to neutralize the system. A grid box was formed pertaining to coordinates for binding pockets using the PrankWeb server. One hundred docked conformations were generated using the Lamarckian Genetic Algorithm to investigate ligands' optimal binding sites in the *AAD* enzymes. For exploring various types of interactions, viz. polar and non-polar, in protein-ligand complexes, different visualization software was used, which include Autodock [27], PyMOL™ v2.4.1 [31], PLIP (Protein-Ligand Interaction Profiler; https://plip-tool.biotec.tu-dresden.de/plip-web/), and VMD [32]. LigPlot+ v.2.2 [33] and Biovia's Discovery studio were used to generate the 2D schematic representation of the interactions in the enzyme-ligand complex. This protocol was reiterated for every ligand, and the final optimal conformation possessing the lowest binding energy (among the hundred conformations procured for every operation) was analyzed carefully.

4.2.4. Molecular dynamic (MD) simulations of selected docked complexes of ligands with AADs

The MD simulations were performed using GROMACS v2019.3 [34]. The Charmm36-July2021 [35] force field was employed in all MD simulations. The ligand topologies obtained after molecular docking for all ligands were submitted to the web server PRODRG [36] to generate GROMACS readable CHARMM coordinates. The obtained topologies (in "*.gro"* format) were integrated with respective *AAD* enzymes to form ligand-bound complexes. All the complexes were then solvated with water using the SPC/E water model. An adequate amount of chlorine or sodium ions was added to obtain the neutralized system. Next, the steepest descent method minimized the system's total energy. Then the system was equilibrated at 310 K and 1 atm using

NVT and NPT ensemble, respectively. This equilibrated system was submitted for a production run for 50 ns each. The PBC (periodic boundary conditions) was applied in all directions as a conventional approach. Three simulation sets for each system were generated, and data were represented as an average over the three sets to check the reproducibility and statistical significance of the results. The final MD trajectories were then analyzed in standalone VMD software. This methodology was repeated for each MD-simulated complex.

Analysis of MD-trajectories: After the MD simulations, trajectories were analyzed, and various parameters were calculated using different packages defined within GROMACS v.2019.3. The global structural stability of the inhibitor-bound complexes was determined by Cα-RMSD (Root Mean Square Deviation) by "*gmx rms"* package. The radius of gyration (Rg) for each complex was calculated using the *"gmx gyrate"* package. Root mean square fluctuation (RMSF) and solvent-accessible surface area (SASA) have also been calculated to deduce detailed structural changes in the *AADs* by FI. All the results given are an average of three computational sets. Most potent inhibit destabilizing the AADs were further accessed by enumerating the post-binding changes in the secondary structure content of respective *AADs* by the *"gmx do_dssp"* package. The hydrogen bonds play an essential role in stabilizing any enzyme or protein. Therefore, the destruction of hydrogen bonds was determined by the *"gmx hbond"* package.

4.3. RESULTS AND DISCUSSION

4.3.1. Sequence and Homology Modelling of AADs

The FASTA sequences for all seven *AADs* were retrieved from the NCBI-RefSeq database and analyzed as described in the computational methodology section. Template predicted by HHpred and BLASTp of query *AADs* sequences against the PDB database showed similarity with previously characterized dehydrogenases (see **Table 4.2A**). The *AADs* from bifunctional alcohol/aldehyde dehydrogenase category (i.e., *adhE1* and *adhE2*) and bifunctional acetaldehyde dehydrogenase (*aydh*) had highest percentage identity with Bifunctional aldehyde-alcohol dehydrogenase *AdhE* from *Escherichia coli* K12 (PDB id: 7BVP "chain D") having 58.93% and 62.43%, respectively. Alcohol-specific dehydrogenase categories (i.e., *bdh*, *bdhA*, and *bdhB*) had the highest percentage identity with butanol dehydrogenase from *Thermotoga maritima* MSB8 (PDB ID: 1VLJ "chain A") with 40.76%, 41.09%, and 38.89%, respectively. However, *gap3dh* from aldehyde specific dehydrogenase category had the highest percentage (79.2%) of similarity with glyceraldehyde phosphate dehydrogenase from *Clostridium perfringens* SM101 (PDB ID: 6FZI "chain A"). The crystal structures of 7BVP_D, 1VLJ_A, and 6FZI_A in PDB format were retrieved from the RCSB-PDB database and used as a template for homology modeling of respective *AADs* enzymes in MODELLER.

4.3.2. Refinement and Validation of Modelled Structure

The backbone Ψ and Φ dihedral angles of the amino acid residues of all 7 *AADs* were analyzed by Ramachandran plot through the PROCHECK server. The homology modeled structures through I-TASSER and MODELLER showed anomaly (most 3D models had Ramachandran scores less than 75%), again refined by GalaxyRefine

webserver. The final 3D models selected after refining were of improved quality and had more than 90% residues in the favored region of the Ramachandran plot and less than 1% in the disallowed region. The overall stereochemistry of each residue in the refined structures was checked using the Ramachandran plot. The results of Ramachandran's analysis for the final 3D models are summarised in **Table 4.2B**. Further, the quality assessments of final 3D models of all 7 *AADs* were performed using UCLA SAVESv.6.0 server, where the structures were mentioned as *pass* by VERIFY 3D, as ≥ 95.7% residues had an average 3D-1D score ≥ 0.2, which indicated the compatibility of amino acids in the modeled structure. Analysis of the overall quality factor of the modeled 3D structures of *AAD* enzymes was performed by the ERRAT server, which was found to be ≥ 91% for each structure, further establishing the excellent quality of the final 3D models of *AADs* (**Table 4.2B**). The goodness factor (g-factor) is a measure for identifying whether a stereo chemistry characteristic is "normal" or "abnormal." The g-factors for *adhE1*, *adhE2*, *bdh*, *bdhA*, *bdhB*, *aydh*, and *gap3dh* were 0.19, 0.16, 0.15, 0.16, 0.03, 0.18, and 0.11, respectively. The acceptable g-factor values in PROCHECK range from 0 to 0.5, with the best models having g-factor values close to zero, suggesting that the model is of high quality. All 3D models in this study fall in this range, which indicates that our models were of excellent quality.

Table 4.2(B): 3D-model evaluation, validation, and secondary structural analysis of *Clostridial AAD* enzymes

Enzymes	Ramachandran plot statistics (in %)				Erratic score	Secondary structural composition (in %)			Surface accessibility (in %)	
	FR	*AR*	*GAR*	*DR*		*α-helices*	*β-sheets*	*Random coil*	*Exposed residue*	*Buried residue*
adhE1	94.2	5.3	0.1	0.4	91.4	49.5	13.2	37.3	42.3	57.7
adhE2	92.3	6.6	0.5	0.6	95.8	48.7	12.8	38.5	42.3	57.7
bdhA	95.3	4.1	0.6	0.0	94.1	54.8	10.8	34.4	43.2	56.8
bdhB	91.9	6.7	1.2	0.3	93.8	51.3	10.5	38.2	41.5	58.5
bdh	97.9	1.8	0.3	0.0	96.3	55.5	10.3	34.2	43.2	56.8
aydh	94.6	5.0	0.0	0.4	92.3	47.6	12.4	40.0	43.3	56.7
gap3dh	93.2	6.1	0.3	0.3	91.6	28.8	28.8	42.3	47.3	52.7

Abbreviation: FR - favored region; AR - allowed region; GAR – generously allowed region; DR - disallowed region

Figure 4.2A depicts the final 3D structure of *AAD* enzymes visualized in PyMOL. These 3D models showed α/β/α type fold. These final enzyme structures modeled in this study have been deposited to the licensed domain of the ProteoPedia database. This chapter's supplementary information section provides the web URLs to access the same.

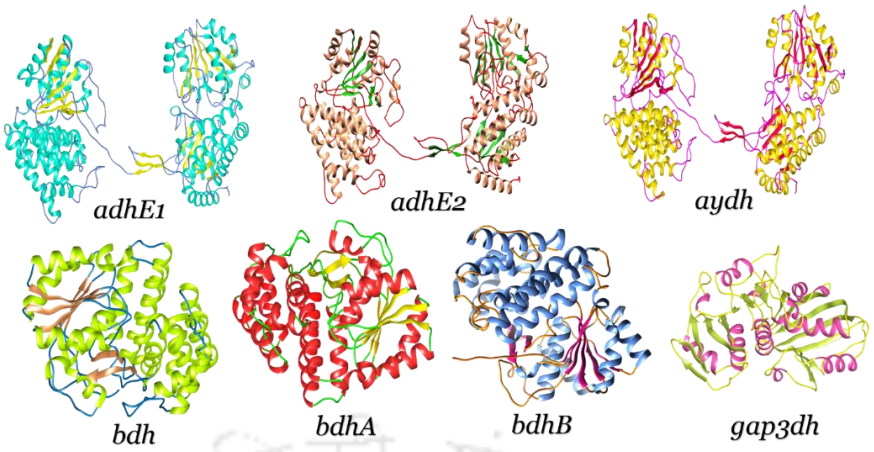


Figure 4.2 (A): Final 3D structure of 7 *Clostridial AAD* enzymes after homology modeling, refinement, and validation.

4.3.3. Secondary Structure and Binding Pocket Analysis of Modelled *AADs*

After modeling and validating the final models, each enzyme was subjected to various CASP14-ranked webservers for secondary structural and binding pocket analysis. These CASP14-ranked web servers include PrankWeb, InterPro, 2Struc, 2StructCompare, PsiPred, and NetSurfP3.0 for identifying the ligand-binding pockets, putative catalytic residues, putative active site residues, putative metal-binding site, co-factor binding residues, secondary structural composition, and percentage surface accessibility (buried and exposed residues). PrankWeb, which utilizes a machine learning algorithm to annotate the protein structure, was used for ligand-binding site prediction. If predicting ligand-binding sites, points with a high-predicted ligandability are grouped and scored using a ranking process based on the cluster's cumulative score. The secondary structural analyses performed through 2Struc and PsiPred and surface accessibilities are provided in **Table 4.2B** (detailed results in **Figure A3.1** of **Appendix 3**). All *AAD* enzymes except *gap3dh* have higher α-helical content (approximately 48 to 56%) with approximately 73 to 76% surface accessibility (percentage ratio of exposed to buried residues) of the enzymes. The *gap3dh* enzyme

has higher random coil content (42.3%) with approximately 90% surface accessibility. The catalytic residue prediction showed nucleophilic cysteine, serine, or histidine residues in each *AAD* enzyme. The residues in the active, metal-binding, and cofactor-binding sites are provided in **Table 4.2C**. Residues forming putative ligand-binding pockets are summarised in **Table A3.2** of **Appendix 3**.

Table 4.2(C): Active site and metal binding pocket analysis of *Clostridial AAD* enzymes

Enzyme	Putative active site residues	Putative metal binding site	Other residues	Catalytic residue
adhE1/E2	D485-G543-T544-P545-S548-K551-T5595-S596-G598-K617-Y618-L636-L644-D651-N655	D651-N655-H721-S735	D651 (Zinc metal, catalytic) D485 (NAD binding residue)	C244
bdh/A/B	G39-G40-S41-P70-N71-G97-G98-S99-D102-T138-L139-T142-S144-D147-I149-V151-K160-Y179-T182-V183-D194-H198-H267-H281	D194-H198-H267-H281	Dimer interface: P9-K11-V12-F14-G15-R222-K226-A242-N243-W246	S140
aydh	D486-G544-G545-S546-D549-K552-T596-S597-G599-K618-Y619-L637-L645-D652-H656-H722-H726-H736-	D652-H656-H722-H736	-	C244
gap3dh	A148-S149-C150-T151-T152-N153-C154-L155	-	NAD binding: R11-I12, D33, S119, N314 G3P binding: S149-C150-T151, T210-G211, T180, T210-G211, R233	C150, H177

4.3.4. DFT Simulations-based Generation and Optimization of Ligand Structures

As mentioned in the computational methodology section above, the chemical structures of 17 ligands comprising 10 FI and seven representative substrates were designed in GaussView v6.0 software and geometrically optimized in Gaussian16. All

geometrically DFT-optimized ligand structure was then subjected to energy minimization. The final ligand structures used for molecular docking simulation are provided in **Figure 4.2B**.

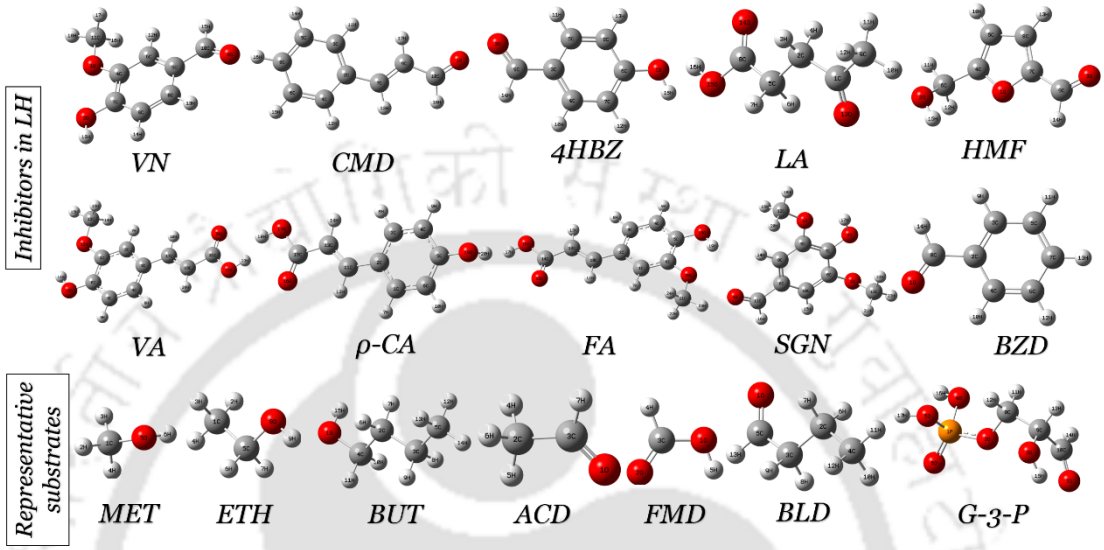


Figure 4.2(B): The final 3- dimensional structure of DFT-based geometrically optimized and energy-minimized ligands (fermentative inhibitors and representative substrates) used for molecular docking.
VN – vanillin, CMD – cinnamaldehyde, 4HBZ – 4-hydroxy-benzaldehyde, LA – levulinic acid, HMF – 5-hydroxymethylfurfural, VA – vanillic acid, ρCA – ρ-coumaric acid, FA – ferulic acid, SGN – syringaldehyde, BZD – benzaldehyde, MET – methanol, ETH – ethanol, BUT – butanol, ACD – acetaldehyde, FMD – formaldehyde, BLD – butyraldehyde, and G-3-P – glyceraldehyde-3-phosphate

4.3.5. Molecular Docking Simulations of Ligands with *Clostridial AAD*s

The molecular docking of *Clostridial AAD* enzymes (**Figure 4.2A**) with 17 ligands (**Figure 4.2B**) is explained in detail computational methodology section. The molecular docking analysis revealed the binding energies (ΔG, kcal mol^{-1}), inhibition constants (K_i), and interacting residues with corresponding ligands. The binding energies for these 105 docking studies are summarised in Table 3A, where the sign and numerical value correspond to the extent and strength of binding with enzymes, respectively (the negative sign implies favorable binding and the lower magnitude

indicates higher binding affinity). The inhibition constants for these bindings are provided in **Table A3.3** of **Appendix 3**. **Table 4.3A** shows that CMD and VA had the highest binding affinity (i.e., lowest ΔG) with bifunctional enzymes, *adhE1* and *adhE2*. In addition to CMD and VA, *ρ*-CA showed the highest binding relationship with aldehyde-specific enzymes, *aydh,* and *gap3dh*. In the case of alcohol-specific enzymes (*bdh*, *bdhA,* and *bdhB*), *ρ*-CA showed the highest binding affinity. These selections of the potent inhibitor were made in comparison with the respective representative substrate (**Figure 4.2B**). The selections of *ρ*-CA, CMD, and VA as potent inhibitors are also supported by the lowest inhibition constant (**Table A3.3** and **4.3B**) compared to other compounds in the respective category. The results of molecular docking analysis (i.e., interacting residues, type of interactions, and nature of inhibition) of the most potent inhibitors (*ρ*-CA, CMD, and VA) are summarised in **Table 4.3C**. Based on the presence of common residues during the interaction of enzyme and ligands, we judged whether the inhibition mode is competitive in the respective enzyme-inhibitor complex. As summarised in **Table 4.3C**, VA competitively inhibits *adhE1*. Similarly, *ρ*-CA inhibits *bdh*, *bdhB,* and *gap3dh* in a competitive manner. The type of inhibition in the case of *adhE2*, *bdhA*, and *aydh* by CMD, *ρ*-CA, and CMD, respectively, could not be identified, which might be non-competitive or uncompetitive. The 3D and 2D visualization of the docked complex of *ρ*-CA, CMD, and VA are shown in **Figures. A3.2** (C, D, G, H, K, L, O, and P) of **Appendix 3**.

Table 4.3(A): Binding energy (ΔG, kcal mol^{-1}) of 17 ligands with modelled *AAD* enzymes

Ligands	Binding energy (ΔG, kcal mol^{-1})						
Inhibitors	*adhE1*	*adhE2*	*bdhA*	*bdhB*	*bdh*	*aydh*	*gap3dh*
Vanillin (VN)	-4.75	-4.71	-5.44	-4.81	-5.05	-5.34	-5.04
Cinnamaldehyde (CMD)	-4.84	-5.37	-5.1	-5.47	-5.12	-5.44	-5.4
4-Hydroxy benzaldehyde (4HBZ)	-4.07	-4.86	-5.25	-4.77	-4.82	-4.66	-3.54
Levulinic acid (LA)	-3.56	-4.94	-4.42	-4.67	-3.95	-4.4	-4.01
5-Hydroxymethylfurfural (HMF)	-3.63	-4.57	-4.63	-4.59	-4.45	-4.55	-5.13
Vanillic acid (VA)	-4.89	-5.17	-5.76	-5.75	-5.18	-5.12	-5.8
ρ-Coumeric acid (*ρ*-CA)	-4.36	-5.15	-5.97	-5.88	-5.46	-5.23	-6.19
Ferulic acid (FA)	-3.92	-5.11	-5.76	-5.7	-4.98	-5.19	-5.49
Syrangaldehyde (SGN)	-4.54	-4.83	-4.79	-4.94	-5.17	-3.69	-5.13
Benzaldehyde (BZD)	-4.07	-4.61	-4.95	-4.41	-4.58	-4.55	-4.65
Representative substrates							
Methanol (MET)	-2.18	-2.36	-2.88	-2.55	-2.3	-2.6	-1.92
Ethanol (ETH)	-2.66	-3.02	-3.28	-2.76	-2.68	-2.58	-2.37
Butanol (BUT)	-2.64	-3.04	-3.25	-3.43	-3.38	NA	NA
Formaldehyde (FMD)	-3.34	-2.86	-3.13	-4.02	-3.14	-3.35	-3.05
Acetaldehyde (ACD)	NA	NA	NA	NA	NA	-2.85	NA
Butyraldehyde (BLD)	-3.22	-3.52	-3.39	-3.46	-3.17	-2.6	-2.62
Glyceraldehyde-3-phosphate (G3P)	NA	NA	NA	NA	NA	NA	-3.65

4.3.6. Comparative MD Analysis of Ligand-bound AAD Trajectories

The best docking complexes of each *AAD* enzyme (including representative substrates and inhibitors), listed in **Table 4.3B**, were tested further for their destabilization potential employing MD simulation. The definition of all such docked complexes and MD simulation systems subjected to dynamic analysis are provided in **Table 4.3B**. The MD simulation studies were conducted for 50 ns for all 24 MD systems (7 control and 17 test systems) under investigation (see **Table 4.3B**). The analysis of comparative MD simulation of ligand-bound *AADs* is shown in **Figures A3.2**(A, B, E, F, I, J, M,

N), where **Figures A3.2**(A, E, I, M) represents superimposed conformation of ligand-bound initial (i.e., 0 ns) gromacs structures. **Figures A3.2**(B, F, J, N) of **Appendix 3** represents superimposed conformation of ligand-bound final (i.e., after 50 ns) gromacs structures of *AADs* from category *adhE1/E2*, *aydh*, *gap3dh*, and *bdh/bdhA/bdhB*. The *ρ*-CA, CMD, and VA (the most potent inhibitors) were observed to migrate to different probable binding sites along the *AAD* enzymes. Still, the average ensemble image has been depicted for all the ligands under investigation in **Figures A3.2**(A, B, E, F, I, J, M, N) of **Appendix 3**. A clear demarcation of the disorganization in final *AADs* gromacs structures was observed in the presence of *ρ*-CA, CMD, and concerning representative substrates BUT, BLD, ACD, and G3P (**Table 4.3B**), which implies the strong disrupting potential of these three FIs. Average RMSD and SASA over the 50 ns simulated trajectories for all 24 MD systems are summarised in **Table 4.4**. The highest percentage change in average RMSD values was observed in the case of *ρ*-CA bound *bdhB* (50 % reduction) and VA bound *adhE1* (21 % reduction) as compared to the representative substrate, viz. BUT, FMD, and BLD (**Table 4.4**).

These reductions in the RMSD values are concurrent with the rise in SASA, which implies that *ρ*-CA and VA have higher disrupting potential by altering the backbone and imposing structural instability of *adhE1* and *bdhB*. Similar trends (rise and reduction in RMSD with representative substrates and FIs, respectively) were also observed in other ligand-bound complexes in **Table 4.4**. In conclusion, the primary analysis of investigating the global MD trajectory of all 17 ligand-bound complexes indicates *ρ*-CA as the most potent disruptor of the solventogenic *AADs*.

Table 4.3(B): System definitions for molecular docking and molecular dynamic simulations of selected ligands

SL. No.	Docking index (D.I.)	Receptor	Ligand	K_i (mM)	Residue for Salt-bridge	Complex index for MD simulation (C4MDS)
1	A0.pdb	*adhE1*	No ligand	NA	NA	C0.gro (*adhE1* + water) – Control 0
2	A1.pdb	*adhE1*	Butanol (BUT)	11.6	-	C1.gro (*adhE1* + BUT) – Test 1
3	A2.pdb	*adhE1*	Formaldehyde (FMD)	3.55	-	C2.gro (*adhE1* + FMD) – Test 2
4	A3.pdb	*adhE1*	Butyraldehyde (BLD)	4.38	-	C3.gro (*adhE1* + BLD) – Test 3
5	A4.pdb	*adhE1*	Vanillic acid (VA)	0.26	K678, K682	C4.gro (*adhE1* + VA) – Test 4
6	A5.pdb	*adhE2*	No ligand	NA	NA	C5.gro (*adhE2* + water) – Control 1
7	A6.pdb	*adhE2*	Butanol (BUT)	5.96	-	C6.gro (*adhE2* + BUT) – Test 5
8	A7.pdb	*adhE2*	Butyraldehyde (BLD)	2.64	-	C7.gro (*adhE2* + BLD) – Test 6
9	A8.pdb	*adhE2*	Cinnamaldehyde (CMD)	0.12	-	C8.gro (*adhE2* + CMD) – Test 7
10	B0.pdb	*bdh*	No ligand	NA	NA	C9.gro (*bdh* + water) – Control 2
11	B1.pdb	*bdh*	Butanol (BUT)	3.30	-	C10.gro (*bdh* + BUT) – Test 8
12	B2.pdb	*bdh*	*ρ*-Coumeric acid (*ρ*-CA)	0.1	R241	C11.gro (*bdh* + *ρ*-CA) – Test 9
13	B3.pdb	*bdhA*	No ligand	NA	NA	C12.gro (*bdhA* + water) – Control 3
14	B4.pdb	*bdhA*	Butanol (BUT)	4.14	-	C13.gro (*bdhA* + BUT) – Test 10
15	B5.pdb	*bdhA*	*ρ*-Coumeric acid (*ρ*-CA)	0.04	-	C14.gro (*bdhA* + *ρ*-CA) – Test 11
16	B6.pdb	*bdhB*	No ligand	NA	NA	C15.gro (*bdhB* + water) – Control 4
17	B7.pdb	*bdhB*	Butanol (BUT)	3.08	-	C16.gro (*bdhB* + BUT) – Test 12

18	B8.pdb	*bdhB*	*ρ*-Coumeric acid (*ρ*-CA)	0.05	-	C17.gro (*bdhB* + *ρ*-CA) – Test 13
19	D0.pdb	*aydh*	No ligand	NA	NA	C18.gro (*aydh* + water) – Control 5
20	D1.pdb	*aydh*	Acetaldehyde (ACD)	8.21	-	C19.gro (*aydh* + ACD) – Test 14
21	D2.pdb	*aydh*	Cinnamaldehyde (CMD)	0.1	-	C20.gro (*aydh* + CMD) – Test 15
22	D3.pdb	*gap3dh*	No ligand	NA	NA	C21.gro (*gap3dh* + water) – Control 6
23	D4.pdb	*gap3dh*	Glyceraldehyde-3-phosphate (G3P)	2.10	K214	C22.gro (*gap3dh* + G3P) – Test 16
24	D5.pdb	*gap3dh*	*ρ*-Coumeric acid (*ρ*-CA)	0.03	-	C23.gro (*gap3dh* + *ρ*-CA) – Test 17

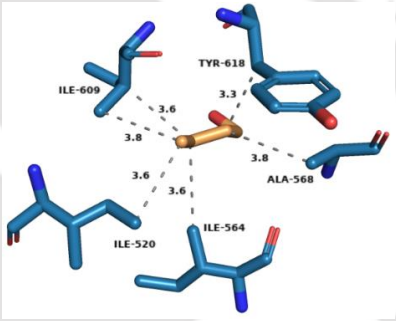
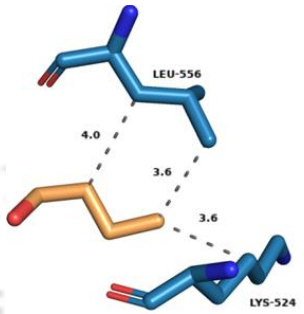
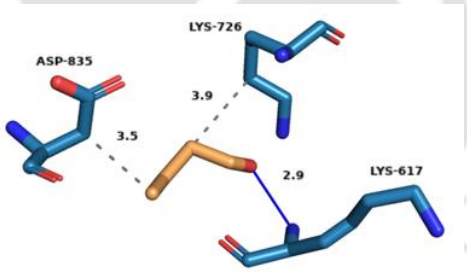
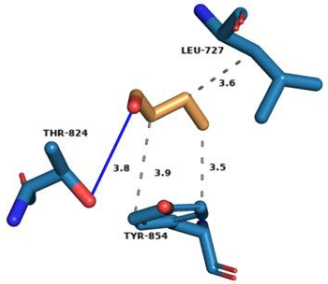
Abbreviation: *K_i – inhibition constant; MD – molecular dynamics; SASA – solvent accessible surface area; RMSD – root mean square deviation*

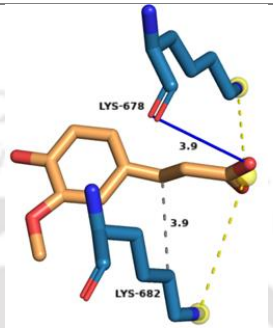
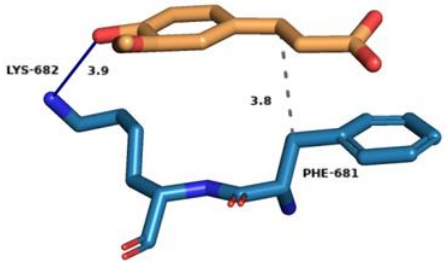
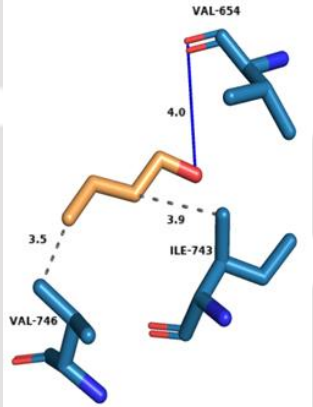
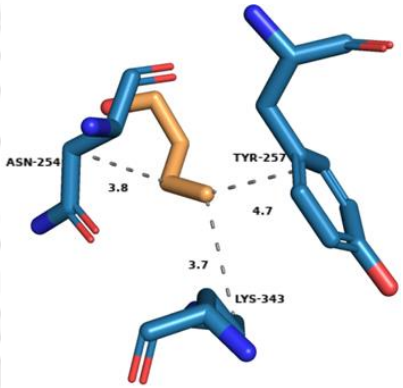
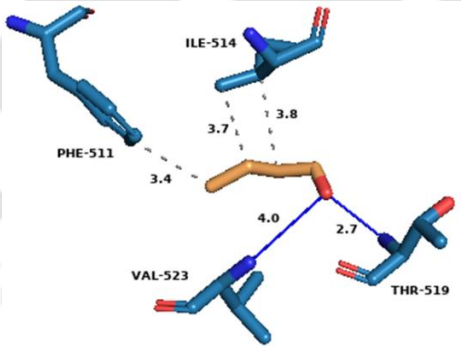
Table 4.3(C): Comparative molecular docking analysis to deduce the type of inhibition by VA, CMD, and *ρ*-CA

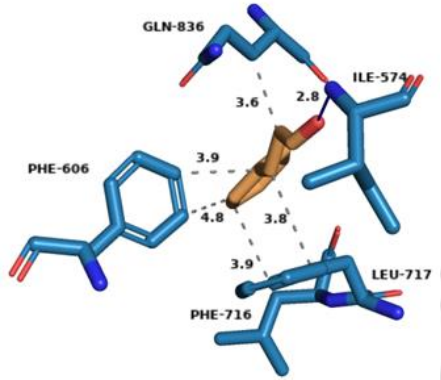
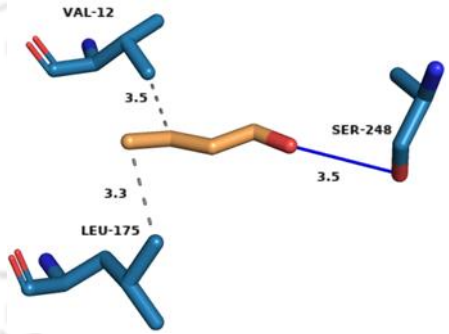
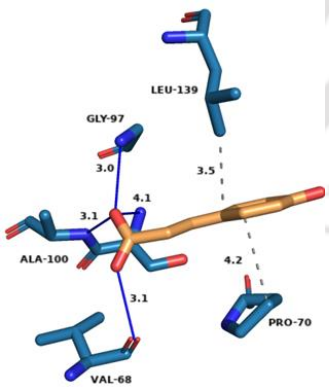
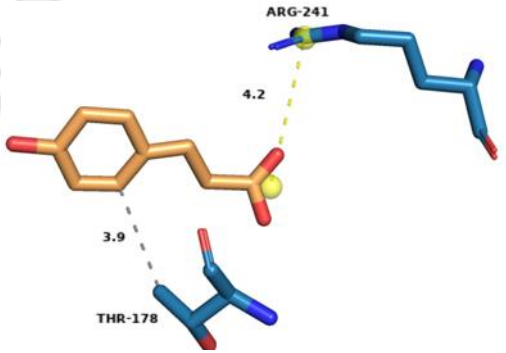
D. I.	Interacting residues	Enzyme-ligand interactions	Common residues	Is competitive?
A1.pdb	I564, I609, Y618, L620, N611	H-bonds and hydrophobic	K682 and Y683	Yes
A2.pdb	K682, Y683, R686	H-bonds		
A3.pdb	K617, K726, L831, D835	H-bonds and hydrophobic		
A4.pdb	K678, F681, K682, Y683, K780, N781, K700	H-bonds, hydrophobic, and salt bridges		

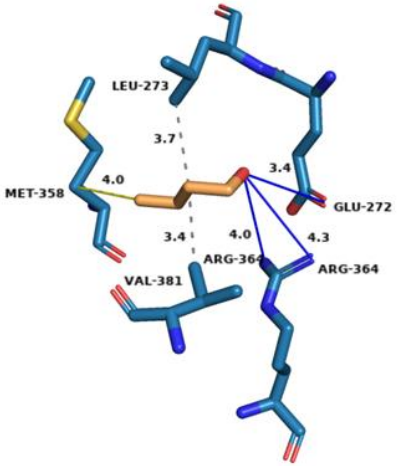
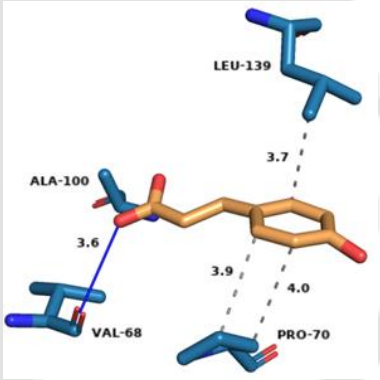
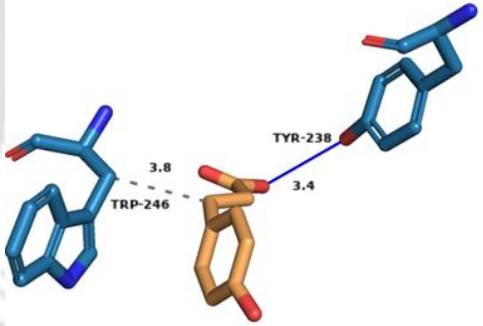
A6.pdb	E658, V661, G718, C720	H-bonds and hydrophobic		
A7.pdb	F511, I514, A549, K515, T519	H-bonds and hydrophobic	None	
A8.pdb	F606, F716, L717, T839, I574	H-bonds and hydrophobic		No
B1.pdb	I42, L139, G38, G39, S41, S99	H-bonds and hydrophobic		
B2.pdb	I42, l139, G38, G40, S41, G97, S99	H-bonds, hydrophobic, and salt bridge	I42, G38, S441, and S99	Yes
B4.pdb	H271, E272, A362, R364, S366	H-bonds and hydrophobic		
B5.pdb	I42, G39, G40, S41, S99	H-bonds and hydrophobic	None	No
B7.pdb	R129, V130, P132, K171, I133, A169	H-bonds and hydrophobic		
B8.pdb	T10, R11, K171, F172, E25, K28, R129	H-bonds, hydrophobic, and salt bridges	R129 and K171	Yes
D1.pdb	R578, L621, A622, Q873	H-bonds and hydrophobic		
D2.pdb	K68, F238, V243, I404, N115, C244	H-bonds and hydrophobic	None	No
D4.pdb	G225, L227, K214	H-bonds and salt bridges		
D5.pdb	K214, I216, A217, L227, L223, K224	H-bonds, hydrophobic, and salt bridges	L227 and K224	Yes

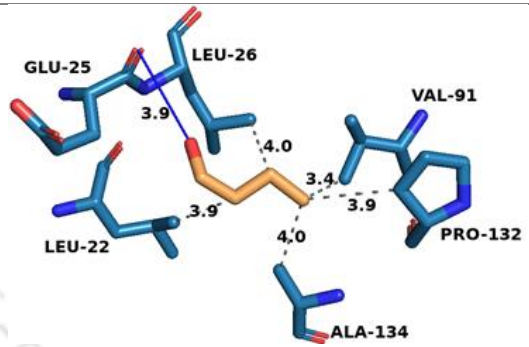
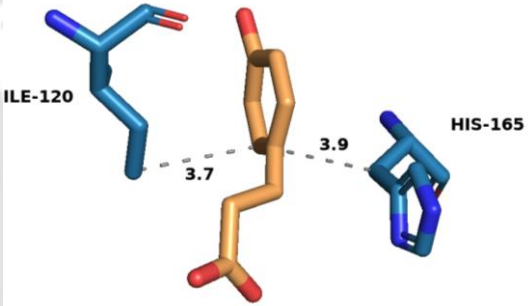
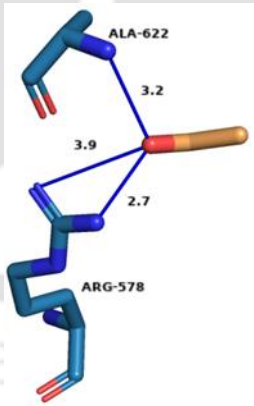
Table 4.4: Calculations of time-averaged structural integrity parameters of modeled *AADs* (with or without ligands) through 50 ns MD simulations

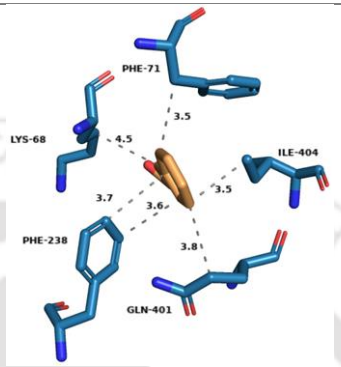
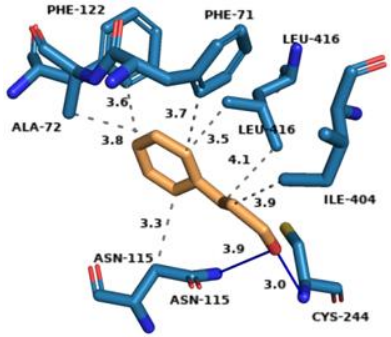
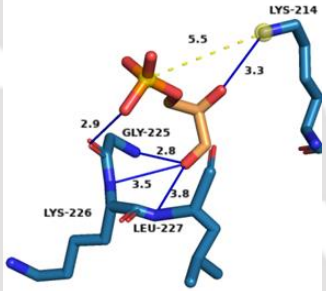
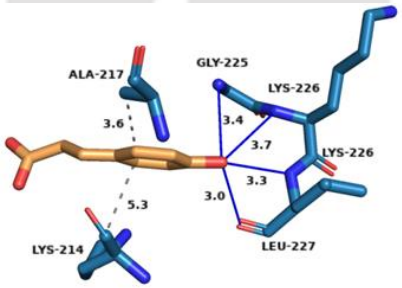
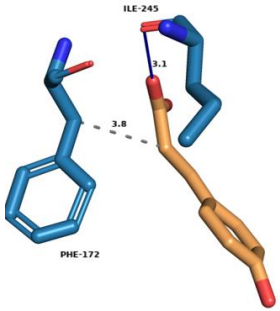
C4MDS	SASA (nm^2)	RMSD (nm)	Rg (nm)	H-Bonds with water	RMSF (nm)	Snapshot of ligand-bound dynamics	
						Initial interactions (at 0 ns)	Final interactions (at 50 ns)
C0.gro	389.9	1.03	3.64	1706	0.69	Not applicable (NA)	NA
C1.gro	395.5	1.01	3.95	1790	0.57		
C2.gro	401.3	1.12	4.20	1725	0.75	No interactions	No interactions
C3.gro	393.1	1.05	4.14	1707	0.71		

C4.gro	391.4	0.81	4.04	1707	0.33		
C5.gro	402.4	1.04	3.72	1782	0.66	NA	NA
C6.gro	400.5	1.1	4.24	1780	0.46		
C7.gro	397.1	1.33	4.05	1748	0.68		Moved out of the pocket

C8.gro	401.8	0.98	4.16	1744	0.48		Moved out of the pocket
C9.gro	183.4	0.29	2.16	1061	0.12	NA	NA
C10.gro	186.5	0.26	2.16	829	0.15	Moved out of the pocket	
C11.gro	187.2	0.28	2.19	847	0.15		

C12.gro	187.5	0.34	2.20	867	0.13	NA	NA
C13.gro	182.7	0.36	2.20	840	0.12	Moved out of the pocket	
C14.gro	187.3	0.33	2.20	837	0.15		
C15.gro	196.8	0.80	2.43	987	0.28	NA	NA

C16.gro	189.8	0.35	2.21	825	0.13	Moved out of the pocket	
C17.gro	213.3	0.40	2.20	831	0.2	Moved out of the pocket	
C18.gro	399.5	0.94	4.35	1745	0.6	NA	NA
C19.gro	401.7	1.26	4.05	1909	0.8		Moved out of the pocket

C20.gro	404.5	1.14	4.27	1863	0.74		
C21.gro	169.2	0.28	2.09	768	0.14	NA	NA
C22.gro	166.0	0.33	2.08	739	0.15		Moved out of the pocket
C23.gro	168.9	0.26	2.11	745	0.13		

Abbreviation: *SASA – solvent accessible surface area; RMSD – root mean square deviation; Rg – radius of gyration; Hbonds – hydrogen bonds; RMSF – root mean square fluctuations*

4.3.7. Detailed Comparative Structural Analysis of AADs in the Presence of ρ-CA, CMD, and VA

In addition to the calculation of time-averaged RMSD and SASA of the ligand-bound *AAD* structures, various other structural stability parameters such as Rg, RMSF, and H-bond formation were calculated for all 24 MD systems to screen the best disruptor for each *AAD* (explained in the computational methodology section). **Figures A3.3** (A, B, C) show the comparative Rg, H-bonds formation, and RMSF analysis of MD simulations trajectory for ligand-bound *AADs* and *AADs* alone. These analyses are provided for the four *AAD* categories, viz. *adhE1/E2*, *aydh*, *gap3dh*, and *bdh/bdhA/bdhB*. Calculations of time-averaged Rg, H-bonds formation, and RMSF analysis for all 24 systems are provided in **Table 4.4**. Snapshots of all MD simulation systems at the beginning (0 ns) and end (50 ns) of MD simulations were taken to check the change in interactions at these time points. The results for the same have been included in **Table 4.4** and **Table A3.4** of **Appendix 3**.

4.4. CONCLUSIONS

Our study concludes that *ρ*-CA is the most potent inhibitor since it shows the highest binding affinities with four modeled enzymes. Furthermore, VA and CMD have also been shown to inhibit two modeled enzymes each, indicating them to be the second most potent inhibitor. This analysis shows clearly that by minimizing lignocellulosic hydrolysate inhibitors, specially *ρ*-CA, VA, and CMD, the enzymatic production of butanol in terms of yield, could enhance significantly while efficiently utilizing the biomass, thus making ABE fermentation economically feasible. Moreover, this study provides a general framework for deciphering the molecular mechanism of inhibition behavior of other enzymes in the metabolic pathway of solventogenic *Clostridia*.

Supplementary Material

Listing of our ProteoPedia depositions:

adhE1: https://proteopedia.org/wiki/index.php/Image:AdhE1.pdb

adhE2: https://proteopedia.org/wiki/index.php/Image:AdhE2.pdb

aydh: https://proteopedia.org/wiki/index.php/Image:Aydh.pdb

bdh: https://proteopedia.org/wiki/index.php/Image:Bdh.pdb

bdhA: https://proteopedia.org/wiki/index.php/Image:BdhA.pdb

bdhB: https://proteopedia.org/wiki/index.php/Image:BdhB.pdb

gap3dh: https://proteopedia.org/wiki/index.php/Image:Gap3dh.pdb

REFERENCE

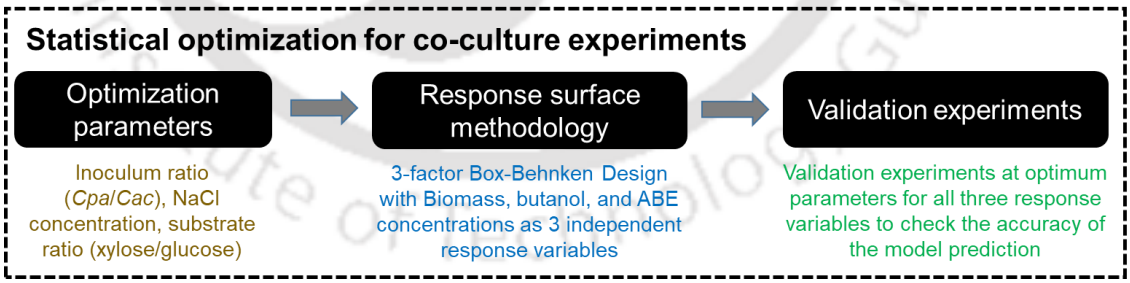
1. Kour D, Rana KL, Yadav N, et al (2019) Technologies for Biofuel Production: Current Development, Challenges, and Future Prospects. In: Rastegari AA, Yadav AN, Gupta A (eds) Prospects of Renewable Bioprocessing in Future Energy Systems. Springer International Publishing, Cham, pp 1–50
2. Jönsson LJ, Martín C (2016) Pretreatment of lignocellulose: Formation of inhibitory by-products and strategies for minimizing their effects. Bioresource Technology 199:103–112. https://doi.org/10.1016/j.biortech.2015.10.009
3. Bhatia SK, Jagtap SS, Bedekar AA, et al (2020) Recent developments in pretreatment technologies on lignocellulosic biomass: Effect of key parameters, technological improvements, and challenges. Bioresource Technology 300:122724. https://doi.org/10.1016/j.biortech.2019.122724
4. Baral NR, Shah A (2014) Microbial inhibitors: formation and effects on acetone-butanol-ethanol fermentation of lignocellulosic biomass. Appl Microbiol Biotechnol 98:9151–9172. https://doi.org/10.1007/s00253-014-6106-8
5. Kim Y, Ximenes E, Mosier NS, Ladisch MR (2011) Soluble inhibitors/deactivators of cellulase enzymes from lignocellulosic biomass. Enzyme and Microbial Technology 48:408–415. https://doi.org/10.1016/j.enzmictec.2011.01.007
6. Altun Ş, Öner C, Yaşar F, Adin H (2011) Effect of n-Butanol Blending with a Blend of Diesel and Biodiesel on Performance and Exhaust Emissions of a Diesel Engine. Ind Eng Chem Res 50:9425–9430. https://doi.org/10.1021/ie201023f
7. Liu H, Wang X, Zhang D, et al (2019) Investigation on Blending Effects of Gasoline Fuel with N-Butanol, DMF, and Ethanol on the Fuel Consumption and Harmful Emissions in a GDI Vehicle. Energies 12:1845. https://doi.org/10.3390/en12101845
8. Kolesinska B, Fraczyk J, Binczarski M, et al (2019) Butanol Synthesis Routes for Biofuel Production: Trends and Perspectives. Materials (Basel) 12:E350. https://doi.org/10.3390/ma12030350

9. Singhvi MS, Chaudhari S, Gokhale DV (2014) Lignocellulose processing: a current challenge. RSC Adv 4:8271. https://doi.org/10.1039/c3ra46112b
10. Chen J-S (1995) Alcohol dehydrogenase: multiplicity and relatedness in the solvent-producing *Clostridia*. FEMS Microbiol Rev 17:263–273. https://doi.org/10.1111/j.1574-6976.1995.tb00210.x
11. Cho C, Hong S, Moon HG, et al (2019) Engineering *Clostridial* Aldehyde/Alcohol Dehydrogenase for Selective Butanol Production. mBio. https://doi.org/10.1128/mBio.02683-18
12. Dai Z, Dong H, Zhang Y, Li Y (2016) Elucidating the contributions of multiple aldehyde/alcohol dehydrogenases to butanol and ethanol production in *Clostridium acetobutylicum*. Sci Rep 6:28189. https://doi.org/10.1038/srep28189
13. Zardecki C, Dutta S, Goodsell DS, et al (2016) RCSB Protein Data Bank: A Resource for Chemical, Biochemical, and Structural Explorations of Large and Small Biomolecules. J Chem Educ 93:569–575. https://doi.org/10.1021/acs.jchemed.5b00404
14. Gabler F, Nam S, Till S, et al (2020) Protein Sequence Analysis Using the MPI Bioinformatics Toolkit. Current Protocols in Bioinformatics 72:. https://doi.org/10.1002/cpbi.108
15. Roy A, Kucukural A, Zhang Y (2010) I-TASSER: a unified platform for automated protein structure and function prediction. Nat Protoc 5:725–738. https://doi.org/10.1038/nprot.2010.5
16. Beg MdA, Shivangi, Thakur SC, Meena LS (2018) Structural Prediction and Mutational Analysis of Rv3906c Gene of *Mycobacterium tuberculosis* H $_{37}$ Rv to Determine Its Essentiality in Survival. Advances in Bioinformatics 2018:1–12. https://doi.org/10.1155/2018/6152014
17. Heo L, Park H, Seok C (2013) GalaxyRefine: protein structure refinement driven by side-chain repacking. Nucleic Acids Research 41:W384–W388. https://doi.org/10.1093/nar/gkt458
18. Laskowski RomanA, Rullmann JAntoonC, MacArthur MalcolmW, et al (1996) AQUA and PROCHECK-NMR: Programs for checking the quality of protein structures solved by NMR. J Biomol NMR 8:. https://doi.org/10.1007/BF00228148
19. Wallner B (2006) Identification of correct regions in protein models using structural, alignment, and consensus information. Protein Science 15:900–913. https://doi.org/10.1110/ps.051799606
20. Jendele L, Krivak R, Skoda P, et al (2019) PrankWeb: a web server for ligand binding site prediction and visualization. Nucleic Acids Research 47:W345–W349. https://doi.org/10.1093/nar/gkz424
21. Kim D (2018) Physico-Chemical Conversion of Lignocellulose: Inhibitor Effects and Detoxification Strategies: A Mini Review. Molecules 23:309. https://doi.org/10.3390/molecules23020309
22. Yang Y, Hu M, Tang Y, et al (2018) Progress and perspective on lignocellulosic hydrolysate inhibitor tolerance improvement in *Zymomonas mobilis*.

Bioresources and Bioprocessing 5:6. https://doi.org/10.1186/s40643-018-0193-9
23. Dennington R, Keith TA, Millam JM (2019) GaussView Version 6
24. Frisch MJ, Trucks GW, Schlegel HB, et al (2016) Gaussian 16 Rev. C.01
25. Guex N, Peitsch MC (1997) SWISS-MODEL and the Swiss-Pdb Viewer: An environment for comparative protein modeling. Electrophoresis 18:2714–2723. https://doi.org/10.1002/elps.1150181505
26. Sanner MF (1999) Python: a programming language for software integration and development. J Mol Graph Model 17:57–61
27. Morris GM, Huey R, Lindstrom W, et al (2009) AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. J Comput Chem 30:2785–2791. https://doi.org/10.1002/jcc.21256
28. Kumar K, Roy K, Moholkar VS (2021) Mechanistic investigations in sonoenzymatic synthesis of n-butyl levulinate. Process Biochemistry 111:147–158. https://doi.org/10.1016/j.procbio.2021.09.005
29. Wang J, Kollman PA, Kuntz ID (1999) Flexible ligand docking: A multistep strategy approach. Proteins: Structure, Function, and Bioinformatics 36:1–19. https://doi.org/10.1002/(SICI)1097-0134(19990701)36:1<1::AID-PROT1>3.0.CO;2-T
30. Gasteiger J, Marsili M (1980) Iterative partial equalization of orbital electronegativity—a rapid access to atomic charges. Tetrahedron 36:3219–3228. https://doi.org/10.1016/0040-4020(80)80168-2
31. Yuan S, Chan HCS, Hu Z (2017) Using PyMOL as a platform for computational drug design. WIREs Comput Mol Sci 7:. https://doi.org/10.1002/wcms.1298
32. Humphrey W, Dalke A, Schulten K (1996) VMD: Visual molecular dynamics. Journal of Molecular Graphics 14:33–38. https://doi.org/10.1016/0263-7855(96)00018-5
33. Laskowski RA, Swindells MB (2011) LigPlot+: Multiple Ligand–Protein Interaction Diagrams for Drug Discovery. J Chem Inf Model 51:2778–2786. https://doi.org/10.1021/ci200227u
34. Abraham MJ, Murtola T, Schulz R, et al (2015) GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. SoftwareX 1–2:19–25. https://doi.org/10.1016/j.softx.2015.06.001
35. Vanommeslaeghe K, Hatcher E, Acharya C, et al (2010) CHARMM general force field: A force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields. Journal of Computational Chemistry 31:671–690. https://doi.org/10.1002/jcc.21367
36. Schüttelkopf AW, van Aalten DMF (2004) PRODRG: a tool for high-throughput crystallography of protein-ligand complexes. Acta Crystallogr D Biol Crystallogr 60:1355–1363. https://doi.org/10.1107/S0907444904011679

CHAPTER 5

Experimental and *in-Silico* Investigation of Bioaugmented Bioalcohol Production in the *Clostridial* Co-culture System



Part of this chapter is available at: Kumar, K.*, Jadhav, S. M., & Moholkar, V. S. (2023), *bioRxiv* 2023.12.08.570763; https://doi.org/10.1101/2023.12.08.570763

CHAPTER 5

EXPERIMENTAL AND *IN-SILICO* INVESTIGATION OF BIOAUGMENTED ALCOHOL PRODUCTION IN THE *CLOSTRIDIAL* CO-CULTURE FOR CONSOLIDATED BIOPROCESSING

5.1. INTRODUCTION

Consolidated bioprocessing (CBP) is a promising approach for converting cellulosic biomass into valuable products, as it integrates all four biological steps in a single bioreactor [1]. CBP eliminates the need for the external addition of saccharolytic enzymes, thereby reducing production costs associated with cellulase production. However, the development of microorganisms with rapid cellulose conversion and high-titer product formation capabilities, which are essential for CBP, remains a challenge [2]. Current strategies for achieving this include the native strategy, which aims to improve the product formation capabilities of natural cellulolytic microorganisms, and the recombinant strategy, which involves engineering non-cellulolytic organisms to express heterologous cellulase and utilize cellulose [3].

Clostridia, a metabolically diverse group of bacteria, offer an alternative method for bioaugmented alcohol production in CBP using a mesophilic *Clostridial* co-culture [4]. *C. acetobutylicum*, a solventogenic bacterium, has the ability to ferment both hemicellulose and cellulose-derived sugars to acetone, butanol, and ethanol [5]. Therefore, the co-culture of this bacterial species with a mesophilic cellulose-

degrading bacterium, such as *C. pasteurianum* [6], can be an alternative option for CBP.

C. pasteurianum is a well-known Gram-positive bacterium that has been studied for a long time by many researchers [7]. However, its ability to degrade cellulose was not discovered until recently. *C. pasteurianum* exhibits unique electrochemical activity, making it an electroactive biochemical-producing heterotrophic microorganism [7]. This discovery highlights the potential for exploring known microorganisms to uncover previously unknown electrochemical activity [8].

In order to fully exploit the potential of *C. pasteurianum* for cellulose degradation and bioaugmented alcohol production, it is crucial to understand its genetic and metabolic characteristics. Genetic and metabolic engineering of *C. pasteurianum* has been impeded by the lack of an efficient method for DNA transfer [9]. However, recent advancements have led to the development of an electrotransformation protocol for genetic manipulation of *C. pasteurianum*, opening up new possibilities for studying and engineering this bacterium [9].

Furthermore, the identification and characterization of key genes and proteins involved in cellulose degradation in *C. pasteurianum* can provide valuable insights into its cellulolytic activity. For example, the hypothetical protein CHU_3220 has been identified as essential for the degradation of the crystalline region of cellulose by *Cytophaga hutchinsonii*, a closely related bacterium [10]. Understanding the function and mechanism of similar proteins in *C. pasteurianum* can contribute to the improvement of its cellulolytic activity.

In recent years, significant progress has been made in the field of bioaugmented alcohol production and consolidated bioprocessing. Studies have focused on the development of genetically engineered microorganisms, metabolic

engineering approaches, and the optimization of co-culture systems for enhanced cellulose degradation and alcohol production [11–13]. Additionally, advancements in genome-scale metabolic modeling have provided valuable insights into the metabolic interactions and pathways involved in the co-culture fermentation [14, 15]. These experimental and in-silico investigations have contributed to a better understanding of bioaugmented alcohol production and have the potential to drive the development of efficient and sustainable processes for cellulosic biofuel production.

In this study, we aim to critically analyze the experimental and in-silico investigations in bioaugmented alcohol production for the co-culture of *C. acetobutylicum* and *C. pasteurianum* for consolidated bioprocessing. We will explore recent research findings, methodologies, and advancements in the field. By integrating experimental and computational approaches, we aim to gain insights into the metabolic interactions, optimize cellulose degradation, and enhance alcohol production in the co-culture system. The findings of this study will contribute to the development of efficient and sustainable processes for bioaugmented alcohol production from cellulosic biomass.

5.2. MATERIALS AND METHODS

5.2.1. Bacterial Strains, Maintenance, and Inoculum Preparation

Procurement of *C. acetobutylicum* (Cac) MTCC 11274 and *C. pasteurianum* (Cpa) MTCC 116, their maintenance, and inoculum preparation for further experiments were followed as described in section 3.2.1 in **Chapter 3**.

5.2.2. Optimization of Co-culture Parameters Using Statistical Experimental Design

Previous authors have shown that ABE fermentation in co-culture system grown over lignocellulosic hydrolysate is mainly influenced by the relative cell densities (or populations) of the two cultures [16, 17], and relative concentrations of hexose and pentose carbon sources [18]. In addition to these parameters, the performance of the ABE fermentation system is also reported to depend on the sodium concentration in hydrolysates [19]. A quantitative metabolomics analysis by Zhao et al. (2016) has shown that high sodium concentrations affected the acidogenesis phase of cell metabolism with accumulation of NADH and ATP, and inhibition of central carbon metabolic (viz. glycolytic and pentose-phosphate) pathways. However, the ratio of $NADP^+$/NADPH remained constant during fermentation leading to higher solvent specific productivities. The overall influence of these simultaneous phenomena was inhibition of biomass growth with greater solvent tolerance and higher solvent productivity. In view of these interesting results, we have also included the sodium concentration in the composite lignocellulosic hydrolysate medium as an optimization parameter.

Guided by these hypotheses, the optimization of ABE fermentation process parameters was conducted using response surface methodology (RSM) with a 3-factor Box-Behnken Design (BBD) of experiments. Three independent variables, namely inoculum ratio (Cac/Cpa), sodium concentration, and substrates ratio (xylose/glucose), were selected, while biomass, butanol, and ABE titers (or concentrations) were designated as the response variables. These parameters were investigated across three levels, as outlined in **Table 5.1A**, and the ranges were determined based on the outcomes of the initial batch experiments.

5.2.2.1. Box-Behnken Design (BBD)

The BBD is a design approach widely employed in response surface methodology, serving as a tool to validate a 2nd order polynomial. It is commonly utilized for studying interactive effects among process variables. The experimental set was generated using Minitab statistical software (Release 16, Trial Version) with three coded levels (-1, 0, +1), three center points, and one replicate, as specified in **Table 5.1A**. The BBD generated runs are calculated using the following mathematical expression:

$$N = 2^k + 2k + n_0 \quad \textbf{(5.1)}$$

Here, N represents the number of runs, k is the number of independent parameters, and n_0 represents the number of replicates runs at the center point. In this study, the experimental design consisted of 15 individual runs. The average response variables, including concentrations of biomass, butanol, and ABE, were recorded for each run, which was carried out in triplicates. The analysis of the individual and interactive effects of parameters on the response variable was performed using the following second-order quadratic model, as presented in **Table 5.1B**.

$$Y = \beta_0 + \sum_{i=1}^{k} \beta_i X_i + \sum_{i=1}^{k} \beta_{ii} X_i^2 + \sum_{i \neq j} \sum_{i} \beta_{ij} X_i X_j \quad \textbf{(5.2)}$$

In this mathematical expression:

Y represents the measured response variable (concentrations of biomass, butanol, and ABE), β_0 is the regression constant, k is the number of medium components or factors, β_i denotes the linear coefficient, β_{ij} signifies the interaction coefficient, β_{ii} represents the quadratic coefficient.

Experimental variables were transformed into coded values using **Equation 5.3**.

$$X = (x_i - x_o)/\Delta x \quad \textbf{(5.3)}$$

In Equation 3, the variables are defined as follows:

X is the dimensionless value of the variable, x_0 is the variable value of x at the center point, Δx represents the step change.

The relative influence of the factor components on the response variable was assessed using Analysis of Variance (ANOVA). Interactive effects between independent variables (or factor components) were evaluated through response surface plots and contour plots.

Table 5.1(A): Factors and levels of optimization variables in Box-Behnken Design

Variables	Coded values (Actual values)	
	Upper bound	Lower bound
(X_1) – Inoculum ratio (*Cac/Cpa*)	–1 (0.3)	+1 (1)
(X_2) - NaCl (g L^{-1})	–1 (0)	+1 (10)
(X_3) – Substrate ratio (Xyl/Glc)	–1 (0)	+1 (0.3)

Table 5.1(B): Experimental and model predicted values of Box-Behnken Design for various response variables in co-culture systems

Expt. No.	Inoculum ratio (X_1)	NaCl (X_2)	Substrate ratio (X_3)	Butanol (g/L) (Experiments)	Butanol (g/L) (Model)	Biomass (OD_{600}) (Experiments)	Biomass (OD_{600}) (Model)	ABE (g/L) (Experiments)	ABE (g/L) (Model)
1	0.30	5	0.00	11.90 ± 0.40	11.84	3.99 ± 0.06	3.99	22.40 ± 0.45	22.31
2	1.00	5	0.30	7.40 ± 1.15	7.46	3.86 ± 0.05	3.86	14.70 ± 0.29	14.79
3	0.30	0	0.15	5.60 ± 0.96	5.63	4.07 ± 0.08	4.05	9.90 ± 0.19	10.22
4	1.00	0	0.15	4.80 ± 0.53	4.79	4.01 ± 0.07	4.02	9.60 ± 0.19	9.61
5	1.00	5	0.00	11.20 ± 0.45	11.18	3.88 ± 0.04	3.85	21.90 ± 0.44	22.12
6	0.65	0	0.00	10.10 ± 0.29	10.13	3.79 ± 0.03	3.82	19.70 ± 0.39	19.47
7	0.30	10	0.15	5.20 ± 0.53	5.21	3.99 ± 0.06	3.97	10.00 ± 0.2	9.98
8	0.65	5	0.15	5.94 ± 1.23	5.93	3.82 ± 0.02	3.82	11.24 ± 0.22	11.75
9	1.00	10	0.15	4.98 ± 0.53	4.95	3.71 ± 0.06	3.74	9.96 ± 0.19	9.64
10	0.65	0	0.30	7.40 ± 1.03	7.35	3.81 ± 0.05	3.79	14.65 ± 0.29	14.55
11	0.30	5	0.30	7.89 ± 1.25	7.91	3.94 ± 0.07	3.98	15.78 ± 0.32	15.56
12	0.65	5	0.15	5.92 ± 1.04	5.93	3.82 ± 0.05	3.82	12.12 ± 0.24	11.75
13	0.65	5	0.15	5.94 ± 0.98	5.93	3.82 ± 0.04	3.82	11.88 ± 0.24	11.75
14	0.65	10	0.30	6.20 ± 0.49	6.17	3.66 ± 0.04	3.64	12.10 ± 0.25	12.33
15	0.65	10	0.00	10.98 ± 0.28	11.03	3.59 ± 0.03	3.61	21.38 ± 0.43	21.48

5.2.2.2. Validation experiments

Three additional fermentation test experiments were conducted at optimal parametric conditions to evaluate the accuracy of the optimal conditions predicted by the statistical analysis. The experiments were conducted in triplicate to ensure reproducibility of the results.

5.2.3. Analytical Methods for Quantifying Sample Components

Sampling intervals and their analysis using HPLC were followed as described in section 3.2.4 in **Chapter 3**.

5.2.4. *in-Silico* Predictions using Genome-Scale Metabolic (GSM) Models

5.2.4.1. Reconstruction and validation of GSM models for Cac and Cpa

Genome sequences retrieval and information sources: The genomic data for *C. acetobutylicum* ATCC 824 and *C. pasteurianum* ATCC 6013 were retrieved from the NCBI-RefSeq database (https://www.ncbi.nlm.nih.gov/refseq/; last access date 18 July 2022). The genome size of *C. acetobutylicum* ATCC 824 is approximately 3.9 Mb, with a total of 3,848 protein-coding genes, and 1,060 enzymes. On the other hand, the genome size of *C. pasteurianum* ATCC 6013 is approximately 4.4 Mb, with a total of 3,919 protein-coding genes, and 1,135 enzymes. In addition, relevant biochemical databases such as KEGG [20] and the enzyme nomenclature database (https://enzyme.expasy.org/) [21] were consulted to supplement the reconstruction process with comprehensive pathway, metabolite, and enzyme information.

Reconstruction of GSM models: The GSM models were manually reconstructed for *C. acetobutylicum* ATCC 824 and *C. pasteurianum* ATCC 6013 using the GapSeq software [22]. GapSeq facilitated the identification of gaps in the metabolic networks, which were subsequently filled to ensure a complete and functional model. The process involved leveraging the genomic annotations and drawing insights from the consulted biochemical databases. Transport reactions were included in the models based on data from KEGG and TransportDB [23], along with relevant physiological data.

Cell biomass synthesis and Growth-Associated Maintenance: The biomass synthesis equation for *C. acetobutylicum* ATCC 824 was incorporated into the network model, considering the growth-associated maintenance energy (GAM) of 40 mmol ATP per gram dry cell weight (DCW) per hour [24–26]. The reconstruction of the metabolic network for *C. pasteurianum* ATCC 6013 was guided by the available genomic annotation from the NCBI-RefSeq database. Gene-protein-reaction (GPR) associations were established based on information derived from literature and experimental data.

Gap filling and model completeness: During the reconstruction process, gap filling was carried out using GapSeq software to ensure the completeness of the GSM models. This step aimed to enable the production of ATP and the synthesis of all necessary cell biomass building blocks from the available substrates in the defined growth media.

Modeling platform and reaction equations: The network models were exported to Systems Biology Markup Language (SBML; https://sbml.org/) format for further analysis using the COBRA toolbox v3.0 [27] in MATLAB environment (MathWorks™). Reaction equations were sourced from publicly available databases such as the BiGG Database [28, 29].

Pathway completion and model validation: Pathway completion was performed manually to ensure that the models allowed for ATP production and the synthesis of all required biomass building blocks from available substrates. Missing ORFs in the original annotations were assigned putative functions using the UniProt BLAST tool (https://www.uniprot.org/blast), and GPR assignments were made based on available literature data. The models were further evaluated and validated using the COBRA

toolbox, with any remaining network gaps curated manually based on relevant literature data.

Cell biomass composition: The cell biomass macromolecular compositions for both strains were adopted from *iCac967* [30] and *iCac802* [26], and the required polymerization energy was adapted from a model developed by Lee et al. [24].

5.2.4.2. Flux Balance Analysis (FBA) and Flux Variability Analysis (FVA)

Flux balance analysis (FBA), a mathematical modeling approach, was employed to predict the metabolic flux distributions that achieve optimal growth or other defined objectives while adhering to mass balance and thermodynamic constraints [31]. The linear programming problem was formulated using the stoichiometric matrix, with the objective function set to maximize or minimize a specific cellular objective, such as growth rate or ATP production. An additional constraint is necessary in the *C. acetobutylicum* network due to the existence of cyclic acid uptake pathways [3, 32]. This constraint establishes a relationship between the rates of acetate and butyrate uptake, which is determined based on enzyme kinetic and selectivity data as well as metabolite concentrations. The concentrations of butyrate and acetate were derived from chemostat data [25, 33].

$$\frac{Rate_{butyrate\ uptake}}{Rate_{acetate\ uptake}} = \frac{0.315 \cdot [butyrate]}{[acetate]} \qquad \textbf{(5.4)}$$

Following the FBA analysis, flux variability analysis (FVA) [34, 35] was conducted to explore the range of feasible flux values for each reaction under the optimal growth conditions obtained from FBA. By fixing the objective function at the optimal value, each reaction flux was maximized and minimized while maintaining the optimal growth rate and other constraints. The resulting ranges of flux variability provided insights into the flexibility and robustness of the metabolic network,

revealing alternate optimal solutions and alternate pathways. Sensitivity analysis was performed to assess the impact of perturbations on the cellular objective and flux distributions, facilitating the identification of critical reactions and metabolic pathways. Computational tools, such as COBRA toolbox in MATLAB, were employed for these analyses, and the obtained results were validated against experimental data and compared with existing literature to ensure accuracy and reliability.

5.2.4.3. FROG and Memote analysis

To maintain the standardization and checking consistency of reconstructed GSM models Memote analysis was perfomed [36]. Memote is a GSM test suite available as an open-source software in Anaconda repository and GitHub (https://github.com/opencobra/memote). In each Memote test, the final or total score is the ratio of weighted sum of results of all individual tests and the maximum achievable score, i.e., all individual results at 100%. The final score can be calculated using the equation 1 given below:

$$Total\ score = \frac{\sum_{Sections} weight_{section} \times (\sum_{Tests} weight_{test} \times score_{test}}{Maximum\ possible\ score} \quad (5.5)$$

We performed FROG analysis [37] to increase reproducibility of the simulations, (http://runfrog.de/ available on https://www.ebi.ac.uk/biomodels/curation/fbc), in addition to the Memote analysis. FROG analysis is a community standard to foster reproducibility and curation of constraint-based models.

5.3. RESULTS AND DISCUSSION

5.3.1. Results of Statistical Optimization of Co-culture Experiments

The application of coded values from Table 1B to the quadratic regression model resulted in following expressions for the three response variables.

***Butanol* (Y_1)**: $Y_1 = -2.09X_1 + 0.38X_2 - 57.85X_3 + 0.57X_1^2 - 0.034X_2^2 + 159X_3^2 + 0.083X_1 \cdot X_2 + X_1 \cdot X_3 - 0.69X_2 \cdot X_3 + 11.25$ (**5.6**)

The optimum set of parameters for maximum (predicted) butanol titre of 11.87 g L^{-1} were as follows: X_1 = 0.30, X_2 = 5.90, X_3 = 0

***Biomass* (OD_{600}) *optimization* (Y_2)**: $Y_2 = -1.84X_1 + 0.014X_2 + 0.72X_3 + 1.37X_1^2 - 0.002X_2^2 - 2.92X_3^2 - 0.03X_1 \cdot X_2 + 0.11X_1 \cdot X_3 + 0.02X_2 \cdot X_3 + 4.43$ (**5.7**)

The optimum set of parameters for maximum biomass (predicted OD_{600}) of 4.06 were as follows: X_1 = 0.3, X_2 = 2.42, X_3 = 0.14

***ABE optimization* (Y_3)**: $Y_3 = 0.31X_1 + 0.9X_2 - 108.22X_3 - 0.58X_1^2 - 0.07X_2^2 + 312.02X_3^2 + 0.04X_1 \cdot X_2 - 2.76X_1 \cdot X_3 - 1.41X_2 \cdot X_3 + 19.52$ (**5.8**)

The optimum set of parameters for maximum (predicted) biomass growth of 22.45 g L^{-1} were as follows: X_1 = 0.46, X_2 = 6.36, X_3 = 0

The comparison between experimental and model-predicted values of the response variable is evident in **Table 5.1B**, demonstrating a close agreement and suggesting an optimal fit of the model to the experimental data. **Table 5.2A** provides the statistical analysis of the quadratic response model, presenting the coefficients of the models alongside their respective *p*- and *t*-values. Additionally, **Table 5.2B** outlines the analysis of variance (ANOVA) results for the quadratic model. ANOVA results indicate the high significance of the regression model ($p < 0.01$), with the lack of fit being non-significant ($p > 0.05$). The coefficient of determination (R^2) exceeding 0.95 in all instances suggests that approximately 95% of the variation in the response variable can be explained by the model. The R^2 value surpassing 0.95 further signifies

the optimal fit of the model to the experimental data, a conclusion supported by the close agreement between the experimental and model-predicted values of the response variable. The *t*-test, *F*-values, and *p*-values associated with the coefficients of the quadratic model serve as indicators of the relative significance of the corresponding independent variables. A significant *t*-statistic value and a *p*-value less than 0.05 indicate the significance of the coefficient and the corresponding independent variable.

The relative *F*-values of the linear, interaction, and quadratic coefficients provide insights into the significance of the individual effects of independent variables and the extent of interaction among them. According to the ANOVA results presented in **Table 5.2B**, the *F*-values for interaction coefficients are considerably smaller than the *F*-value for linear coefficients. This suggests a relatively independent or unrelated effect of individual variables on the concentration of response variables. The *p*-values for all linear and quadratic coefficients are less than 0.05, accompanied by large absolute *t*-values, indicating a significant effect of all variables on the response variables. In contrast, the *t*-values of interaction coefficients are relatively smaller than those of linear and quadratic coefficients, suggesting their comparatively lesser significance. Additionally, the *p*-values for interaction coefficients related to inoculum ratio with NaCl and inoculum ratio with the initial substrate concentration ratio exceed 0.05. This underscores the insignificance of the interaction between these variables, implying that they independently influence the response variable. The relatively small *F*-value for Lack of Fit, accompanied by its associated p-value, indicates that Lack of Fit is not significant when compared to pure error, affirming the overall significance of the model.

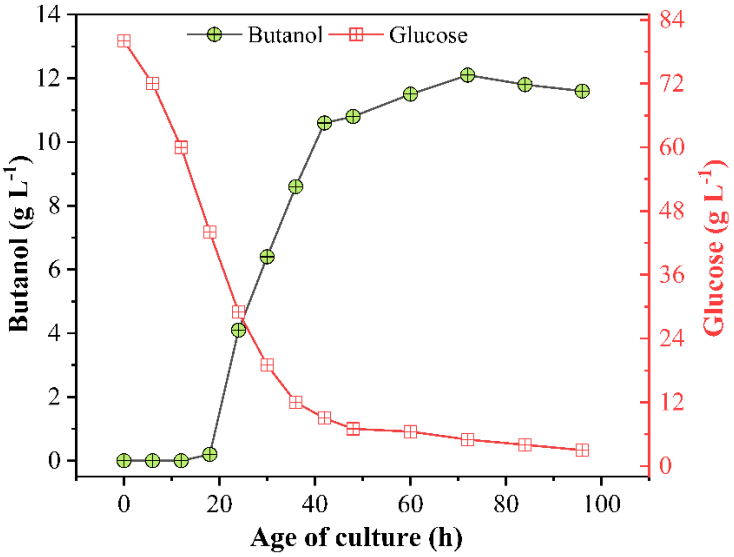
The contour plots illustrated in **Figures A4.2– A4.4(A–C)** provide a visual representation of the interactions between medium components and their effects on butanol, biomass (OD600), and ABE production. These plots delineate specific regions of the objective content (butanol, biomass, and ABE production) enclosed by contour lines. They offer insights into the interactions among different optimization parameters. e.g., (A) salt and inoculum ratio, (B) substrate ratio and inoculum ratio, (C) substrate ratio and salt. In the given range of optimization parameters, **Figures A4.2D, A4.3D, & A4.4D** depict the desirability function plots.

5.3.2. Validation of Experiments at Optimized Parameters

The optimal value of fermentation parameters for all three response variables are summarized in the section 5.3.1. The final titre of butanol, biomass, and ABE was 12.1 ± 0.45 g L^{-1}, 4.15 ± 0.03, and 23.1 ± 0.55 g L^{-1}, respectively, which slightly higher than the results predicted by BBD analysis (i.e., 11.87, 4.06, and 22.45 g L^{-1}, respectively). **Figures 5.1A** & **B** display the dynamic profiles of butanol production and glucose consumption, and the variations in total solvents, total acid production, biomass growth, and pH, respectively, during co-culture studies on optimized media to enhance butanol formation (Y_1).

Our validation experiments reveal promising findings, demonstrating approximately a 22% increase (from 9.9 to 12.1 g L^{-1}) and a substantial 61% increment (from 7.5 to 12.1 g L^{-1}) in butanol titre in co-culture experiments compared to individual fermentations experiments with *Cac* and *Cpa* grown on pure glucose, respectively.

(A)



(B)

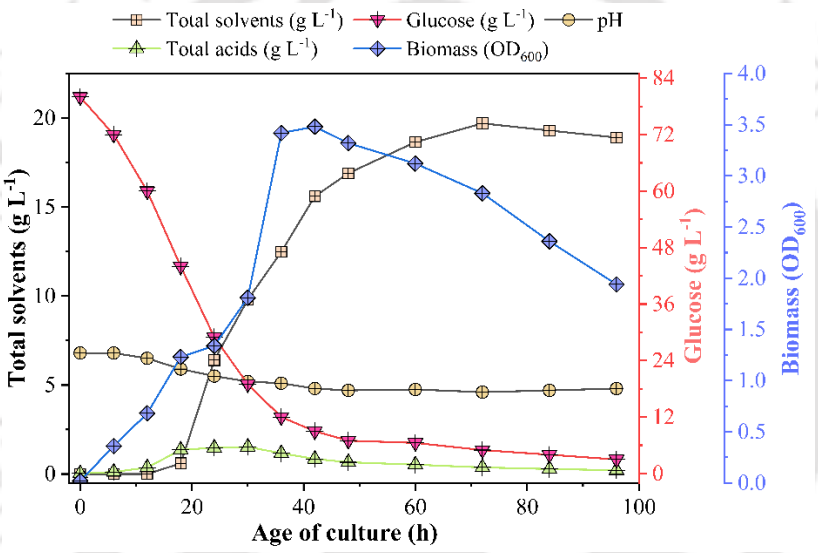


Figure 5.1 Dynamic profiles of (A) butanol production and glucose consumption, and (B) total solvents, total acid production, biomass production, and pH variation during co-culture studies on optimized media for enhanced butanol formation (Y_1).

Notably, optimizing the inoculum ratio (*Cac*/*Cpa*) suggests a strategic approach – an elevated *Cac* presence in the inoculum mixture maximizes total solvent yield, albeit with a trade-off: a decrease in butanol formation and biomass growth. This intriguing observation might have resulted from the inherent dynamics between *Cac* and *Cpa*. *Cpa* exhibits faster growth kinetics than *Cac*, a phenomenon well-documented in the

literature [38–40]. Further, the monophasic fermentation in the ABE pathway employed by *Cpa* [6] contrasts with the biphasic fermentation process utilized by *Cac* [38, 41]. Consequently, a higher volume of *Cpa* in the inoculum mixture accelerates early solvent production, contributing to elevated butanol and biomass production. Turning our attention to salt concentration optimization, our results align with previous observations by Zhao et al., (2016). We observed the highest butanol and ABE solvent yield of 12.10 ± 0.45 g L^{-1} and 12 23.1 ± 0.55 g L^{-1} for the salt addition of 5.90 and 6.36 g L^{-1}, respectively. Therefore, our study affirms that elevated sodium concentrations inhibit biomass growth while enhancing solvent tolerance and productivity. This phenomenon is attributed to the impact of high sodium concentration on the acidogenesis phase of cell metabolism, leading to the accumulation of NADH and ATP and the inhibition of central carbon metabolic pathways. These metabolic changes maintain the $NADP^+$/NADPH ratio, resulting in higher solvent specific productivities. Furthermore, our optimization results shed light on the negligible role of xylose in ABE solvent production. Maximum butanol and total ABE solvent yields are observed with zero xylose concentration, underscoring the dependency of ABE production on the presence of glucose or other hexoses in the medium.

In summary, our study establishes a positive correlation between salt concentration, the proportion of *Cac* in the inoculum mixture, and butanol titre in ABE fermentation. We contribute to the field by enhancing biobutanol production through a co-culture system of *Cac* MTCC 11274 and *Cpa* MTCC 116 on synthetic hydrolysate of LB. However, it is important to note that the precise molecular synergism underlying the enhanced butanol production in this co-culture system remains unexplored, presenting an exciting avenue for future research.

5.3.3. Predictions using Genome-Scale Metabolic (GSM) Models

5.3.3.1. GSM Models for Cac and Cpa: Reconstruction and Refinement

The GSM models for *Cac* ATCC 824 (iKK_CAC) and *Cpa* ATCC 6013 (iKK_CPA) were reconstructed and gap-filled to obtain comprehensive representations of their metabolic networks. The initial draft GSM iKK_CAC model comprised 1425 genes, 2118 reactions, and 1812 metabolites, while the iKK_CPA model consisted of 848 genes, 2162 reactions, and 1839 metabolites. These models were exported to SBML format and subsequently analyzed using CobraToolbox for further investigations.

As discussed in the section 5.2.4, to validate the reconstructed GSM models, we conducted quality and consistency checks using Memote, an open-source GSM test suite available in Anaconda and GitHub repositories. Each Memote test provided a weighted score, and the final score was obtained by calculating the ratio of the weighted sum of individual test results to the maximum achievable score. Additionally, missing reactions in the Butanotate Metabolism pathway were explored using information from the KEGG database to ensure that the models encompassed comprehensive metabolic coverage.

Table 5.2: Results of Box-Behnken Design for optimization of experimental parameters

(A) *t*– and *p*– values of the coefficients of quadratic regression models for different response variables

Model term	Butanol maximization (Y_1)		Biomass maximization (Y_2)		ABE maximization (Y_3)	
	t–value	*p*–value	*t*–value	*p-value*	*t*–value	*p*–value
Intercept (A_0)	167.857	0.000*	55.931	0.000*	20.514	0.000*
Linear coefficients						
Inoculum ratio (X_1)	-12.762	0.000*	-8.913	0.000*	0.124	0.906
NaCl (X_2)	-3.118	0.026*	1.338	0.238	7.132	0.001*
Substrate ratio (X_3)	-88.296	0.000*	2.065	0.094	-25.699	0.000*
Square coefficients						
Inoculum ration*Inoculum ratio (X_1^2)	2.184	0.081*	9.321	0.000*	-0.334	0.752
NaCl*NaCl (X_2^2)	-26.926	0.000*	-2.191	0.080	-8.386	0.000*
Substrate ratio*Substrate ratio (X_3^2)	112.818	0.000*	-3.652	0.015*	32.533	0.000*
Interaction coefficients						
$X_1 \times X_2$	4.737	0.005*	-3.005	0.030*	0.314	0.767
$X_1 \times X_3$	1.715	0.147	0.326	0.758	-0.699	0.516
$X_2 \times X_3$	-16.987	0.000	0.796	0.462	-5.101	0.004

(B) ANOVA for the quadratic regression model

Source	Butanol maximization (Y_1)				Biomass maximization (Y_2)				ABE maximization (Y_3)			
	DF	SS	*F*–value	*p*–value	DF	SS	*F*–value	*p*–value	DF	SS	*F*–value	*p*–value
Regression	9	83.59	2477.97	0.000*	9	0.248	23.13	0.001*	9	307.331	198.61	0.000*
Linear	3	29.87	2656.29	0.000*	3	0.102	33.95	0.001*	3	99.535	263.86	0.000*
Square	3	52.55	4672.97	0.000*	3	0.135	37.63	0.001*	3	203.222	393.98	0.000*
Interaction	3	1.177	104.64	0.000*	3	0.012	3.26	0.118	3	4.574	8.87	0.019*
Residual (Error)	5	0.019			5	0.006			5	0.860		
Lack of fit	3	0.018	46.19	0.021	3	0.006	*	*	3	0.446	0.72	0.627
Pure error	2	0.0003			2	0.00			26	0.414		
Total	14	83.61			14	0.254			14	308.191		

**Significant p-values, $p \leq 0.05$, p-value Prob > F*

For Butanol maximization: $R^2 = 0.999$; Predicted $R^2 = 0.996$; Adjusted $R^2 = 0.994$

For Biomass maximization: $R^2 = 0.976$; Predicted $R^2 = 0.625$; Adjusted $R^2 = 0.934$

For ABE maximization: $R^2 = 0.997$; Predicted $R^2 = 0.974$; Adjusted $R^2 = 0.992$

The validation of the GSM models involved comparing predicted growth rates and carbon source utilization with available experimental data. Notably, the analysis revealed the absence of the Butanol synthesis reaction (Butanol dehydrogenase, KEGG ID: R03544) in both GSM models. To address this limitation, we incorporated two metabolic reactions (rxn02527 and rxn02528) from the Butanoate Metabolism pathway (cpat00650), based on data from the KEGG Pathways. These reactions involve the catalysis of butanol dehydrogenase (*bdhA1*/*A2*) enzymes and are responsible for the dehydrogenation of butanal to butanol, utilizing either NADH or NADPH as cofactors:

$$\text{Butanal} + \text{NADH} + H^+ \rightleftharpoons \text{1-Butanol} + NAD^+ \quad (\mathbf{5.9})$$

$$\text{Butanal} + \text{NADPH} + H^+ \rightleftharpoons \text{1-Butanol} + NADP^+ \quad (\mathbf{5.10})$$

Additionally, the reconstructed GSM models were thoroughly analyzed to identify any potential issues such as dead-end reactions, blocked reactions, and Type III pathways. However, we did not find any instances of dead-end reactions, blocked reactions, or Type III pathways in both models.

During the analysis, we did identify unbalanced mass and charge in both GSM models. To rectify these issues, we utilized the gurobi solver and COBRAToolbox in MATLAB to make the necessary corrections. The mass and charge balancing checks were crucial to ensure that the total mass and charge of metabolites were conserved in all reactions, thus maintaining the integrity and accuracy of the metabolic network. **Table 5.3** enlist the statistical specifications of these refined models.

Table 5.3. Specifications of refined GSM models reconstructed in the present study

Model features	iKK_CAC	iKK_CPA
Strain	*Cac* ATCC 824	*Cpa* ATCC 6013
Genes	713	850
Reactions	2120	2164
Non-gene associated reactions	573	467
Metabolites	1812	1839
Exchange reactions	182	173

By resolving the unbalanced mass and charge in the GSM models, we were able to enhance the reliability and consistency of the models for further analysis.

5.3.3.2. Model Comparison with Existing Models

To date, a total of 24 GSM models have been developed for eleven representative Clostridium species, as listed in **Table 5.4A**. These GSM models are publicly available in the BIGG [28], BioModels [42] databases, and Prof. Maranas research group webpage (https://www.cmaranasgroup.com/models.htm). Notably, each GSM model captures the strain-specific substrate utilization patterns of the respective Clostridium species.

The initial GSM models for *C. thermocellum* (iSR432) [43] and *C. cellulolyticum* (iFS431) [33] incorporated the cellulosome complex as a metabolite, utilizing amino acids for biomass production or cell growth. Subsequently, Dash et al. developed an updated GSM model, iCth446, for *C. thermocellum*. Furthermore, the GSM models iCbu641 and iHN637 [44, 45] displayed the ability to utilize glycerol and syngas for butanol production [4, 46, 47], respectively. The unique autotrophic Wood–Ljungdahl (WL) pathway utilized by *C. ljungdahlii* allows for the balancing of reducing equivalents with electron bifurcation [4, 44, 48].

In the context of *Clostridium* species exhibiting biphasic behavior (e.g., *C. acetobutylicum* and *C. beijerinckii*), where acetate and butyrate are produced during the acidogenic phase, the GSM models iCac802 and iCM925 (for *C. acetobutylicum* and *C. beijerinckii*, respectively), demonstrate this behavior through changes in carbon uptake [26, 49].

Comparative analyses of the GSM models developed in this study with previous literature are presented in **Tables 5.4A** and **B**. The models iKK_CAC and iKK_CPA (**Table 5.3**), developed for *Cac* ATCC 824 and *Cpa* ATCC 6013, respectively, are the most comprehensive, encompassing the largest number of genes, metabolites, and reactions. To enable meaningful comparisons, all metabolites were consistently described with the same designation across all models using ModelSEED [50].

Table 5.4 (A): Statistics for GSM models of representative *Clostridium* sp.

SI	Model name	Metabolic models			Strains	Ref.
		Genes	Rxns	Mets		
1	iJL432	432	502	479	*Cac* ATCC 824	[24]
2	Senger	474	552	422	*Cac* ATCC 824	[25]
3	iFS700	700	738	679	*Cac* ATCC 824	[33]
4	iCac490	490	794	707	*Cac* ATCC 824	[51]
5	Wallenius	NA	592	444	*Cac* DSM 792	[52]
6	iCac802	802	1462	1137	*Cac* ATCC 824	[26]
7	iCac967	967	1231	1058	*Cac* ATCC 824	[5]
8	iCM925	926	938	881	*Cbj* NCIMB 8052	[49]
9	iCbu641	641	891	701	*C. butyricum* IBUN 13A	[45]
10	iFS431	707	743	752	*C. cellulolyticum* H10	[33]
11	iHN637	637	785	698	*C. ljungdahlii*	[44]
12	iSR432	NA	581	564	*Cth* ATCC 27405	[43]
13	iCth446	NA	660	599	*Cth* ATCC 27405	[53]
14	iAT601	601	872	904	*Cth* DSM 1313	[54]

15	iCN900	900	1229	885	*C. difficile 630*	[55]
16	iCBI665	665	584	789	*C. thermocellum*	[56]
17	Laocque	NA	1217	1638	*C. difficile 630*	[57]
18	iCKL708	708	994	804	*C. kluyveri*	[58]
19	Marcellin	805	1002	1075	*C. autoethanogenum*	[59]
20	iCLAU786	786	1109	1097	*C. autoethanogenum*	[60]
21	MetaCLAU	699	755	772	*C. autoethanogenum*	[61]
22	iSL771	771	922	854	*C. drakei*	[62]
23	iJL680	680	809	718	*C. ljungdahlii*	[63]
24	iAI558	558	705	698	*M. thermoacetica*	[64]

Abbreviation: Rxns – Reaction, Mets – Metabolites, Ref. – References, *Cac – C. acetobutylicum*, *Cbj – C. beijerinckii*, *Cth – C. thermocellum*

Furthermore, to ensure the consistency and annotation quality of all models, including iKK_CAC and iKK_CPA, Memote was employed for testing. The results in **Table 5.4B** reveal that GSM model iKK_CAC achieved the highest score of 61% among all reported *C. acetobutylicum* models. Additionally, GSM model iKK_CPA, the first GSM model for *C. pasteurianum* ATCC 6013, displayed an excellent score of 77% with a consistency of 95%. Memote scores for genes, metabolites, reactions, and SBO terms annotation were also excellent for both GSM models (**Table 5.4B**).

Table 5.4 (B): Comparative analyses on the quality and consistency of the GSM models developed in this study with previous literature.

Model	Consistency (%)	Annotations (%)				Final Score (%)
		Mets	Rxns	Genes	SBO	
iJL432	42	25	25	0	0	20
Senger	29	25	25	0	0	15
iCac802	29	25	25	0	0	15
iCM925	99	25	25	0	0	42
iCbu641	98.94	59.06	60.29	40	45.45	68

iHN637	100	80	84	37	91	90
iSR432	41	25	25	0	0	20
iCth446	43	52	25	0	0	22
iCN900	53	78	81	39	91	72
iCBI665	38	71	49	33	91	62
Laocque	23	50	48	0	21	25
Marcellin	32	25	26	0	9	20
iCLAU786	29	25	25	0	0	15
iJL680	56	25	25	0	0	26
iKK_CAC	**52**	**86**	**76**	**0**	**70**	**61**
iKK_CPA	**95**	**86**	**76**	**0**	**69**	**77**

Abbreviation: Mets – Metabolites, Rxns – Reaction, SBO – Systems Biology Ontology Terms

Note: Missing models in this table (but present in **Table 5.4A**) were either not in a Memote-readable format or were not made available in the proper SBML format in the respective literature.

In summary, the developed GSM models, iKK_CAC and iKK_CPA, demonstrate high consistency and annotation quality, making them valuable tools for further studies and applications related to *Clostridium* metabolism. The comparative analyses provide a comprehensive understanding of the strengths and unique features of these models in the context of other existing GSM models for *Clostridium* species.

5.3.3.3. GSM Model Testing, Validations, and Analysis

Subsequently, the optimized GSM models (refined iKK_CAC and iKK_CPA) were employed for the analysis of butanol production using FBA. FBA allowed us to evaluate the potential of each model for butanol synthesis under various conditions, providing valuable insights into the predicted rates of butanol production and identifying key regulatory points that could influence butanol yield.

For testing and validation purpose, all analyses on iKK_CAC and iKK_CPA were performed in comparison with the available functional model iCac802 [26].

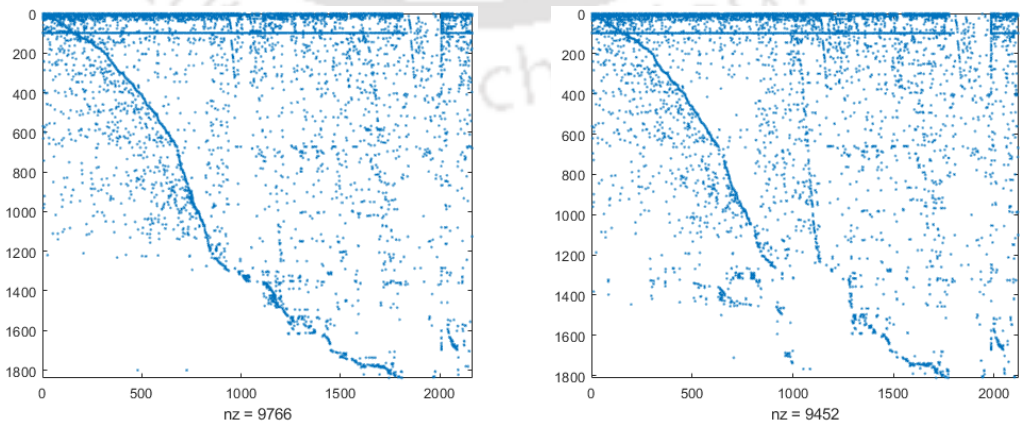
Topological Features of S- Matrix: Metabolite Connectivity and Reaction Participation

In this section, we delve into the topological characteristics of the metabolic network represented by the stoichiometric matrix (S-matrix), which serves as a fundamental framework for describing the interconnectedness of metabolites and reactions in GSM models. **Figure 5.2A** shows the S-matrix of iKK_CPA and iKK_CAC.

Next, we examined the metabolite connectivity, which refers to the number of reactions in which each metabolite participates. A higher metabolite connectivity indicates that the metabolite is involved in multiple reactions and plays a central role in the network. On the other hand, lower metabolite connectivity suggests that the metabolite is specific to only a few reactions. **Figure 5.2B** shows, there are very few metabolites that are highly connected, while most metabolites participate only in a few reactions. The approximately linear appearance of the curve of connectivities in **Figure 5.2B** corresponds to a power law distribution of metabolite connectivity. The few highly connected metabolites are "global" players, similar to hubs in protein-protein-interaction network, while the low connectivity metabolites are "local" players, many of which only occur in linear pathways. The power law distribution indicates that networks are scale-free.

Furthermore, to investigate the relationship between reaction essentiality and metabolite connectivity in the metabolic network, we performed a comprehensive analysis on our GSM models. The essentiality of reactions was determined through computational simulations, where we individually removed each reaction from the

model and assessed its impact on cellular growth or production of specific metabolites. Simultaneously, we calculated the metabolite connectivity, representing the number of reactions in which each metabolite participates. **Figure 5.2 (C)** shows the correlation between metabolite connectivity and average lethality of reactions (essentiality) producing or consuming a particular metabolite in iKK_CPA and iKK_CAC models. The results show a strong correlation between reaction essentiality and metabolite connectivity in the metabolic network. Essential reactions tend to involve metabolites with higher connectivity, acting as central hubs in facilitating flux through multiple pathways. Conversely, non-essential reactions are often associated with metabolites of lower connectivity, which have more specialized roles. This correlation provides insights into the organization and robustness of the metabolic network, with highly connected metabolites playing crucial roles in maintaining cellular viability and functionality. These findings have implications for metabolic engineering and drug development, as metabolites with high connectivity are potential targets for interventions to optimize cellular functions. Understanding the interplay between reaction participation and metabolite centrality enhances our knowledge of the metabolic network's dynamics and its impact on cellular physiology and biotechnological applications.



iKK_CPA iKK_CAC

Figure 5.2 (A). The S (Stochiometric) matrix of the iKK_CPA and iKK_CAC. All non-zero entries are marked with a dot.

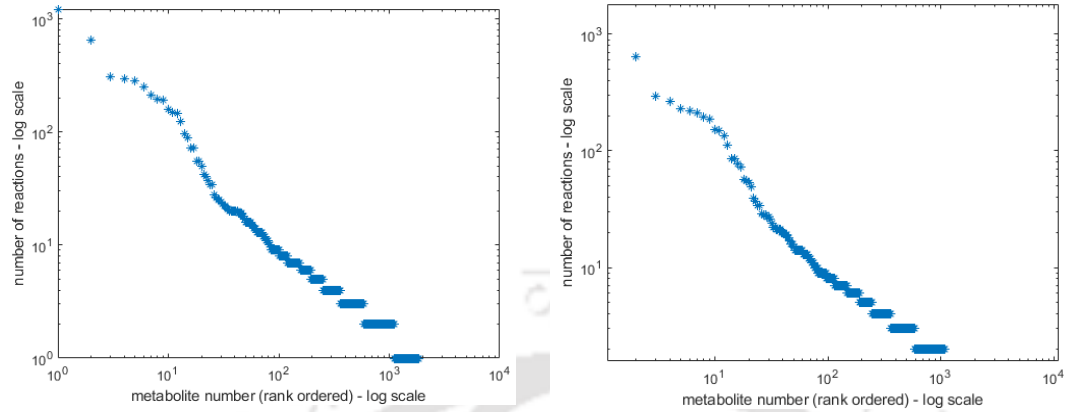


Figure 5.2 (B). Connectivity of the metabolites in iKK_CPA and iKK_CAC.

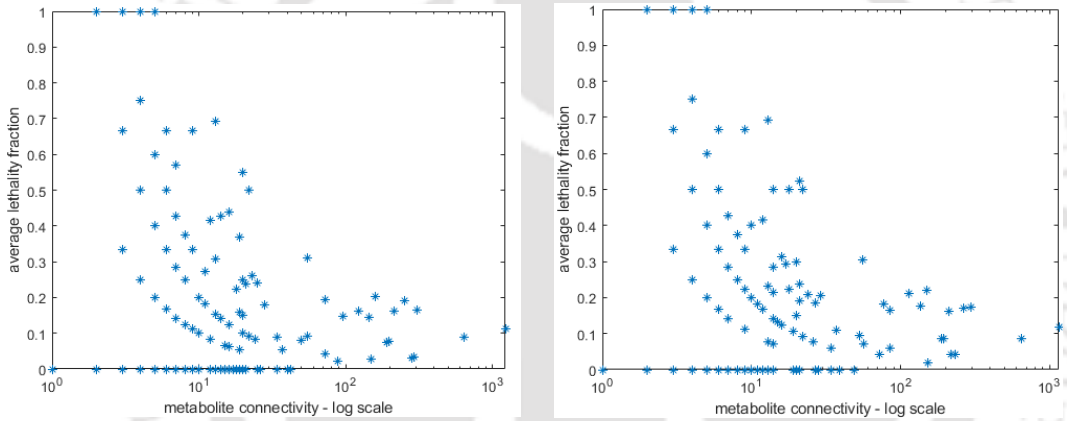


Figure 5.2 (C). Correlation between metabolite connectivity and average lethality of reactions producing or consuming a particular metabolite in iKK_CPA and iKK_CAC models. Less connected metabolites tend to occur in a higher fraction of essential reactions.

Calculation of Growth Rates on Various Substrates

The maximum growth rates of iCac802, iKK_CAC, and iKK_CPA on their 10 different organic substrates were computed by FBA. Growth rates calculated for anerobic conditions for each substrate on maximum substrate uptake rate was set to -20 mmol $gDW^{-1}h^{-1}$. Results of this analysis are summarized in **Table 5.5**.

Table 5.5. Maximum growth rate (h^{-1}) of the models on different substrates

Substrate	Growth Rate (h^{-1})		
	iCac802	iKK_CAC	iKK_CPA
Glucose	0.28	1.34	1.49
Fructose	0.28	1.22	1.33
Ribose	0.14	1.14	1.22
Xylose	0.23	1.1	1.1
Maltose	0.34	1.8	2
Glutamate	0	1	1.75
Glutamine	0	1	1.56
Cellobiose	0.33	1.7	2
Malate	0	0.8	0.88
Formate	0	0.7	0.68

Calculation of Growth Rates and Product Secretion on Various Substrates

Growth rates of iCac802, iKK_CAC, and iKK_CPA were simulated under anaerobic conditions. We selected a total of 10 substrates (glucose, fructose, ribose, xylose, maltose, glutamate, glutamine, cellobiose, malate, and formate) and 10 products (H_2, CO_2, acetate, butyrate, butanol ethanol, acetone lactate, succinate, and biomass) to analyze the growth rates and product secretion rates for each combination of substrate and product. For each combination of substrate and product in the list, the lower bound of the corresponding exchange reaction is set to (-20 mmol $gDW^{-1}h^{-1}$ for substrates and -10 mmol $gDW^{-1}h^{-1}$ for products) simulate growth and product secretion. Results of this analysis is summarized supplementary excel file **Appendix Ex5**.

Our simulations revealed significantly higher μ_{max} (0.97 h^{-1} and 1 h^{-1} for *Cac* and *Cpa*, respectively) , as compared to typical μ_{max} reported in literature for *Cac* (0.26 h^{-1}, Buendia-Kandia et al., (2018)) and *Cpa* (0.23 h^{-1}, Groeger et al., (2016)) which might be due to consideration of purely theoretical conditions in the present study. In practical or actual experimental condition many factors such as substrate or product

inhibition or negative metabolic effect of other media components and metabolic inhibitors present in the medium may adversely affect the biomass growth. The μ_{max} predicted by the model is significantly higher than experimentally observed μ_{max}. Several reasons could have contributed to this discrepancy. We conjecture herewith some of the possible causes that lead to the observed discrepancy: (1) the model assumes all metabolic reactions to occur, viz. the reactions that lead to the metabolites and final products, and the reactions that contribute to biomass growth; (2) the model assumes all enzymes involved in metabolism to be perfectly active, and thus, the model ignores practical factors that might reduce the efficiencies and activities of the enzymes; (3) the model also ignores few interactive effects between the enzymes that might cause reduction in activities of enzymes or complete blockage of certain metabolic pathways; (4) the model also does not account for solvent toxicity effects on the cellular metabolism and microbial growth.

5.3.3.4. Limitations of Genome-Scale Metabolic Models (GSMs)

The development of GSMs for *C. acetobutylicum* (iKK_CAC) and *C. pasteurianum* (iKK_CPA) represents a significant step towards understanding the metabolic capabilities of these organisms. However, it is essential to acknowledge several limitations in our GSM study. Firstly, the accuracy and reliability of the models are highly dependent on the quality and completeness of the underlying experimental data used for model construction. While considerable efforts were made to gather and validate the data from various sources, inherent uncertainties and variations in experimental measurements can introduce biases and affect the model predictions. Additionally, the annotation of genes and reactions in the models may not be entirely

comprehensive, potentially leading to the omission of critical metabolic pathways or regulatory elements that could influence the overall model behavior.

Secondly, while GSMs provide a systems-level perspective of cellular metabolism, they inherently simplify the complex biochemical networks present in living organisms. The models are based on numerous assumptions and constraints, which can overlook intricate metabolic interactions and regulatory mechanisms. These simplifications might lead to deviations between model predictions and actual experimental observations. Furthermore, the accuracy of model predictions is highly dependent on the selection of objective functions for flux balance analysis (FBA). The choice of objective function to optimize specific growth conditions or product yields can be challenging, and it may introduce biases in the model outputs. Careful consideration of the objective functions and constraints is crucial to ensure the relevance and reliability of the model's predictions.

Moreover, our GSM study focused on the metabolic capabilities of individual *Clostridial* species, and the interactions between different species in co-culture were not explicitly incorporated into these models. While the individual GSMs provide valuable insights into the metabolic behavior of each species, understanding the dynamics of co-culture systems in more complex environments, such as lignocellulosic hydrolysates, requires further investigation. The absence of explicit representation of interspecies interactions may limit the models' ability to predict co-culture behavior accurately. Future research efforts should aim to extend the models to include co-culture dynamics, enabling a more comprehensive understanding of the metabolic interactions between different *Clostridial* species.

5.4. CONCLUSIONS

In this chapter, we embarked on a comprehensive statistical optimization of the co-culture system involving *C. acetobutylicum* (*Cac* MTCC 11274) and *C. pasteurianum* (*Cpa* MTCC 116) to augment biobutanol production from mixed substrates, employing response surface methodology (RSM). Our investigation encompassed the evaluation of various process parameters, including the ratio of *Cac* and *Cpa* inoculum, NaCl concentration, and the ratio of xylose to glucose. Through rigorous analysis, we established that salt concentration emerged as the most influential factor affecting biomass, total alcohol, and butanol production. Leveraging the media composition defined in **Chapter 3**, we performed optimization experiments with specific variations in glucose and NaCl concentrations, culminating in the determination of optimized values for the respective conditions. These optimum conditions yielded highly promising results with a butanol concentration of 12.1 ± 0.45 g/L (model prediction: 11.87 g/L), biomass of 4.15 ± 0.03 (model prediction: 4.06 OD600), and ABE concentration of 23.1 ± 0.55 g/L (model prediction: 22.45 g/L). The accuracy of our models was corroborated through the formulation of quadratic regression models for all three optimization functions.

Furthermore, our research unveiled a pioneering contribution to the field of systems biology with the proposition of two novel genome-scale metabolic models, namely iKK_CAC for *C. acetobutylicum* ATCC 824 and iKK_CPA for *C. pasteurianum* ATCC 6013. These state-of-the-art models offer a profound comprehension of the molecular underpinnings governing biomass degradation, elevated solvent production, and stress-tolerance in these organisms. The exhaustive comparative genomic analysis (studied in **Chapter 3**), along with the valuable insights obtained from these GSMs, lays a solid foundation for the development of highly

efficient microbial factories for bioenergy production. However, we acknowledge the inherent limitations associated with such models, necessitating continuous refinement, rigorous data validation, and the integration of advanced modeling techniques to bolster their predictive capabilities. As we forge ahead, further exploration of the intricate interactions between different *Clostridial* species within co-culture systems holds immense promise for unraveling the enigmas surrounding biobutanol production processes, thereby driving the progression of sustainable bioenergy production towards a greener and more eco-friendly future.

REFERENCES

1. Liang Y, Si T, Ang EL, Zhao H (2014) Engineered Pentafunctional Minicellulosome for Simultaneous Saccharification and Ethanol Fermentation in Saccharomyces cerevisiae. Appl Environ Microbiol 80:. https://doi.org/10.1128/aem.02070-14
2. Chantarasiri A (2021) Diversity and Activity of Aquatic Cellulolytic Bacteria Isolated from Sedimentary Water in the Littoral Zone of Tonle Sap Lake, Cambodia. Water 13:. https://doi.org/10.3390/w13131797
3. Salimi F, Mahadevan R (2013) Characterizing metabolic interactions in a clostridial co-culture for consolidated bioprocessing. BMC Biotechnol 13:. https://doi.org/10.1186/1472-6750-13-95
4. Diallo M, Kengen SWM, López-Contreras AM (2021) Sporulation in solventogenic and acetogenic clostridia. Appl Microbiol Biotechnol 105:3533–3557. https://doi.org/10.1007/s00253-021-11289-9
5. Yoo M, Nguyen N-P-T, Soucaille P (2020) Trends in Systems Biology for the Analysis and Engineering of Clostridium acetobutylicum Metabolism. Trends Microbiol 28:118–140. https://doi.org/10.1016/j.tim.2019.09.003
6. Pyne ME, Liu X, Moo-Young M, et al (2016) Genome-directed analysis of prophage excision, host defence systems, and central fermentative metabolism in Clostridium pasteurianum. Sci Rep 6:26228. https://doi.org/10.1038/srep26228
7. Choi O, Kim T, Woo HM, Um Y (2014) Electricity-driven metabolic shift through direct electron uptake by electroactive heterotroph Clostridiumpasteurianum. Sci Rep 4:6961. https://doi.org/10.1038/srep06961
8. Sabra W, Wang W, Surandram S, et al (2016) Fermentation of mixed substrates by Clostridium pasteurianum and its physiological, metabolic and proteomic

characterizations. Microb Cell Factories 15:114. https://doi.org/10.1186/s12934-016-0497-4
9. Pyne ME, Moo-Young M, Chung DA, Chou CP (2013) Development of an electrotransformation protocol for genetic manipulation of Clostridium pasteurianum. Biotechnol Biofuels 6:50. https://doi.org/10.1186/1754-6834-6-50
10. Wang S, Zhao D, Bai X, et al (2017) Identification and Characterization of a Large Protein Essential for Degradation of the Crystalline Region of Cellulose by Cytophaga hutchinsonii. Appl Environ Microbiol 83:. https://doi.org/10.1128/aem.02270-16
11. Johnson AC, Merilis G, Hastings J, et al (2012) Reductive Degradation of Organic Compounds Using Microbial Nanotechnology. J Electrochem Soc 160:. https://doi.org/10.1149/2.053301jes
12. Bai X, Wang X, Wang S, et al (2017) Functional Studies of β-Glucosidases of Cytophaga hutchinsonii and Their Effects on Cellulose Degradation. Front Microbiol 8:. https://doi.org/10.3389/fmicb.2017.00140
13. Xin F, Yan W, Zhou J, et al (2018) Exploitation of novel wild type solventogenic strains for butanol production. Biotechnol Biofuels 11:252. https://doi.org/10.1186/s13068-018-1252-3
14. Kolek J, Sedlar K, Provaznik I, Patakova P (2016) Dam and Dcm methylations prevent gene transfer into Clostridium pasteurianum NRRL B-598: development of methods for electrotransformation, conjugation, and sonoporation. Biotechnol Biofuels 9:. https://doi.org/10.1186/s13068-016-0436-y
15. Clauß M (2006) Higher effectiveness of photoinactivation of bacterial spores, UV resistant vegetative bacteria and mold spores with 222 nm compared to 254 nm wavelength. Acta Hydrochim Hydrobiol 34:. https://doi.org/10.1002/aheh.200600650
16. Barca C, Ranava D, Bauzan M, et al (2016) Fermentative hydrogen production in an up-flow anaerobic biofilm reactor inoculated with a co-culture of Clostridium acetobutylicum and Desulfovibrio vulgaris. Bioresour Technol 221:526–533. https://doi.org/10.1016/j.biortech.2016.09.072
17. Muhammad J, Wang H, Wu Y, et al (2023) Lignocellulosic ethanol and butanol production by Saccharomyces cerevisiae and Clostridium beijerinckii co-culture using non-detoxified corn stover hydrolysate. J Biotechnol. https://doi.org/10.1016/j.jbiotec.2023.11.002
18. Du Y, Zou W, Zhang K, et al (2020) Advances and Applications of Clostridium Co-culture Systems in Biotechnology. Front Microbiol 11:560223. https://doi.org/10.3389/fmicb.2020.560223
19. Zhao X, Condruz S, Chen J, Jolicoeur M (2016) A quantitative metabolomics study of high sodium response in Clostridium acetobutylicum ATCC 824 acetone-butanol-ethanol (ABE) fermentation. Sci Rep 6:28307. https://doi.org/10.1038/srep28307

20. Kanehisa M, Furumichi M, Sato Y, et al (2023) KEGG for taxonomy-based analysis of pathways and genomes. Nucleic Acids Res 51:D587–D592. https://doi.org/10.1093/nar/gkac963
21. Kanehisa M (2017) Enzyme Annotation and Metabolic Reconstruction Using KEGG. Methods Mol Biol Clifton NJ 1611:135–145. https://doi.org/10.1007/978-1-4939-7015-5_11
22. Zimmermann J, Kaleta C, Waschina S (2021) gapseq: informed prediction of bacterial metabolic pathways and reconstruction of accurate metabolic models. Genome Biol 22:81. https://doi.org/10.1186/s13059-021-02295-1
23. Elbourne LDH, Tetu SG, Hassan KA, Paulsen IT (2017) TransportDB 2.0: a database for exploring membrane transporters in sequenced genomes from all domains of life. Nucleic Acids Res 45:D320–D324. https://doi.org/10.1093/nar/gkw1068
24. Lee J, Yun H, Feist AM, et al (2008) Genome-scale reconstruction and in silico analysis of the Clostridium acetobutylicum ATCC 824 metabolic network. Appl Microbiol Biotechnol 80:849–862. https://doi.org/10.1007/s00253-008-1654-4
25. Senger RS, Papoutsakis ET (2008) Genome-scale model for Clostridium acetobutylicum: Part I. Metabolic network resolution and analysis. Biotechnol Bioeng 101:1036–1052. https://doi.org/10.1002/bit.22010
26. Dash S, Mueller TJ, Venkataramanan KP, et al (2014) Capturing the response of Clostridium acetobutylicumto chemical stressors using a regulated genome-scale metabolic model. Biotechnol Biofuels 7:144. https://doi.org/10.1186/s13068-014-0144-4
27. Heirendt L, Arreckx S, Pfau T, et al (2019) Creation and analysis of biochemical constraint-based models using the COBRA Toolbox v.3.0. Nat Protoc 14:639–702. https://doi.org/10.1038/s41596-018-0098-2
28. King ZA, Lu J, Dräger A, et al (2016) BiGG Models: A platform for integrating, standardizing and sharing genome-scale models. Nucleic Acids Res 44:D515–D522. https://doi.org/10.1093/nar/gkv1049
29. Schellenberger J, Park JO, Conrad TM, Palsson BØ (2010) BiGG: a Biochemical Genetic and Genomic knowledgebase of large scale metabolic reconstructions. BMC Bioinformatics 11:213. https://doi.org/10.1186/1471-2105-11-213
30. Yoo M, Bestel-Corre G, Croux C, et al (2015) A Quantitative System-Scale Characterization of the Metabolism of Clostridium acetobutylicum. mBio. https://doi.org/10.1128/mBio.01808-15
31. Raman K, Chandra N (2009) Flux balance analysis of biological systems: applications and challenges. Brief Bioinform 10:435–449. https://doi.org/10.1093/bib/bbp011
32. Desai RP, Harris LM, Welker NE, Papoutsakis ET (1999) Metabolic Flux Analysis Elucidates the Importance of the Acid-Formation Pathways in Regulating Solvent Production by Clostridium acetobutylicum. Metab Eng 1:206–213. https://doi.org/10.1006/mben.1999.0118

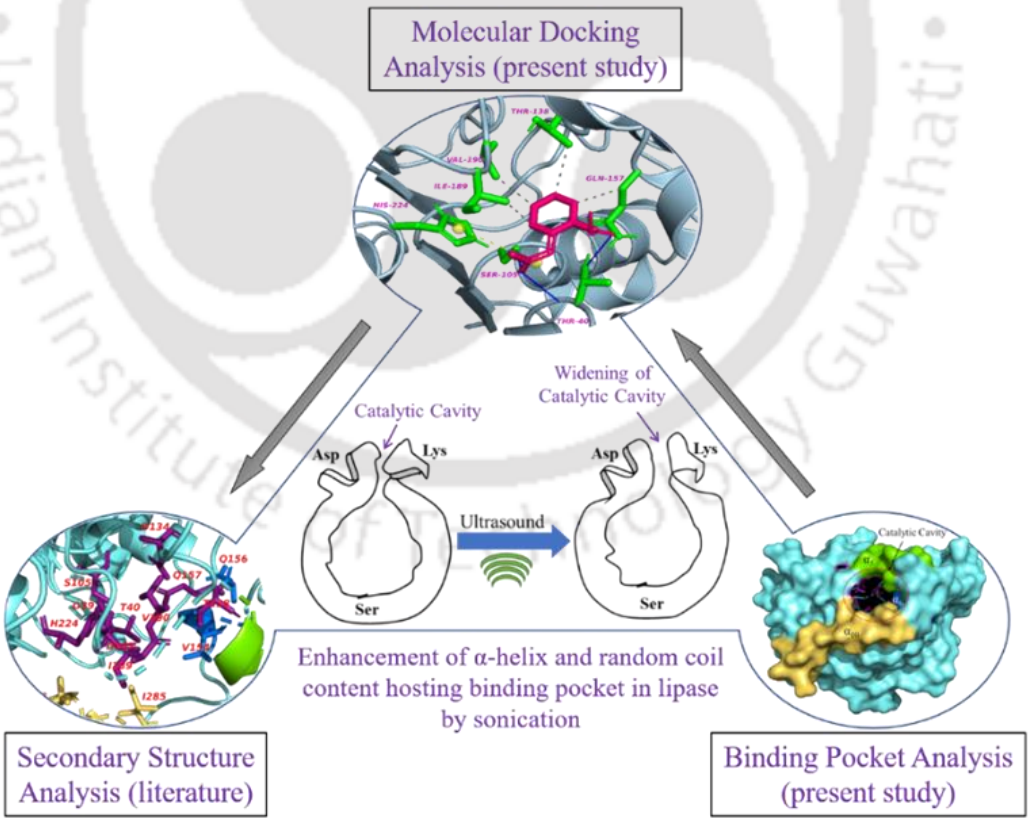
33. Salimi F, Zhuang K, Mahadevan R (2010) Genome-scale metabolic modeling of a clostridial co-culture for consolidated bioprocessing. Biotechnol J 5:726–738. https://doi.org/10.1002/biot.201000159
34. Mahadevan R, Schilling CH (2003) The effects of alternate optimal solutions in constraint-based genome-scale metabolic models. Metab Eng 5:264–276. https://doi.org/10.1016/j.ymben.2003.09.002
35. Kenefake D, Armingol E, Lewis NE, Pistikopoulos EN (2022) An improved algorithm for flux variability analysis. BMC Bioinformatics 23:550. https://doi.org/10.1186/s12859-022-05089-9
36. Lieven C, Beber ME, Olivier BG, et al (2020) MEMOTE for standardized genome-scale metabolic model testing. Nat Biotechnol 38:272–276. https://doi.org/10.1038/s41587-020-0446-y
37. König, Matthias (2021) sbmlsim: SBML simulation made easy
38. Kumar K, Barbora L, Moholkar VS (2023) Genomic insights into clostridia in bioenergy production: Comparison of metabolic capabilities and evolutionary relationships. Biotechnol Bioeng n/a: https://doi.org/10.1002/bit.28610
39. Sabra W, Groeger C, Sharma PN, Zeng A-P (2014) Improved n-butanol production by a non-acetone producing Clostridium pasteurianum DSMZ 525 in mixed substrate fermentation. Appl Microbiol Biotechnol 98:4267–4276. https://doi.org/10.1007/s00253-014-5588-8
40. Sarma S, Ortega D, Minton NP, et al (2019) Homologous overexpression of hydrogenase and glycerol dehydrogenase in Clostridium pasteurianum to enhance hydrogen production from crude glycerol. Bioresour Technol 284:168–177. https://doi.org/10.1016/j.biortech.2019.03.074
41. Li S, Huang L, Ke C, et al (2020) Pathway dissection, regulation, engineering and application: lessons learned from biobutanol production by solventogenic clostridia. Biotechnol Biofuels 13:39. https://doi.org/10.1186/s13068-020-01674-3
42. Malik-Sheriff RS, Glont M, Nguyen TVN, et al (2019) BioModels—15 years of sharing computational models in life science. Nucleic Acids Res gkz1055. https://doi.org/10.1093/nar/gkz1055
43. Roberts SB, Gowen CM, Brooks JP, Fong SS (2010) Genome-scale metabolic analysis of Clostridium thermocellum for bioethanol production. BMC Syst Biol 4:31. https://doi.org/10.1186/1752-0509-4-31
44. Nagarajan H, Sahin M, Nogales J, et al (2013) Characterizing acetogenic metabolism using a genome-scale metabolic reconstruction of Clostridium ljungdahlii. Microb Cell Factories 12:118. https://doi.org/10.1186/1475-2859-12-118
45. Serrano-Bermúdez LM, González Barrios AF, Maranas CD, Montoya D (2017) Clostridium butyricum maximizes growth while minimizing enzyme usage and ATP production: metabolic flux distribution of a strain cultured in glycerol. BMC Syst Biol 11:58. https://doi.org/10.1186/s12918-017-0434-0

46. Kopke M, Held C, Hujer S, et al (2010) Clostridium ljungdahlii represents a microbial production platform based on syngas. Proc Natl Acad Sci 107:13087–13092. https://doi.org/10.1073/pnas.1004716107
47. Poehlein A, Solano JDM, Flitsch SK, et al (2017) Microbial solvent formation revisited by comparative genome analysis. Biotechnol Biofuels 10:58. https://doi.org/10.1186/s13068-017-0742-z
48. Buckel W, Thauer RK (2013) Energy conservation via electron bifurcating ferredoxin reduction and proton/Na+ translocating ferredoxin oxidation. Biochim Biophys Acta BBA - Bioenerg 1827:94–113. https://doi.org/10.1016/j.bbabio.2012.07.002
49. Milne CB, Eddy JA, Raju R, et al (2011) Metabolic network reconstruction and genome-scale model of butanol-producing strain Clostridium beijerinckii NCIMB 8052. BMC Syst Biol 5:130. https://doi.org/10.1186/1752-0509-5-130
50. Henry CS, DeJongh M, Best AA, et al (2010) High-throughput generation, optimization and analysis of genome-scale metabolic models. Nat Biotechnol 28:977–982. https://doi.org/10.1038/nbt.1672
51. McAnulty MJ, Yen JY, Freedman BG, Senger RS (2012) Genome-scale modeling using flux ratio constraints to enable metabolic engineering of clostridial metabolism in silico. BMC Syst Biol 6:42. https://doi.org/10.1186/1752-0509-6-42
52. Wallenius J, Viikilä M, Survase S, et al (2013) Constraint-based genome-scale metabolic modeling of Clostridium acetobutylicum behavior in an immobilized column. Bioresour Technol 142:603–610. https://doi.org/10.1016/j.biortech.2013.05.085
53. Dash S, Khodayari A, Zhou J, et al (2017) Development of a core Clostridium thermocellum kinetic metabolic model consistent with multiple genetic perturbations. Biotechnol Biofuels 10:108. https://doi.org/10.1186/s13068-017-0792-2
54. Thompson RA, Dahal S, Garcia S, et al (2016) Exploring complex cellular phenotypes and model-guided strain design with a novel genome-scale metabolic model of Clostridium thermocellum DSM 1313 implementing an adjustable cellulosome. Biotechnol Biofuels 9:194. https://doi.org/10.1186/s13068-016-0607-x
55. Norsigian CJ, Danhof HA, Brand CK, et al (2020) Systems biology analysis of the Clostridioides difficile core-genome contextualizes microenvironmental evolutionary pressures leading to genotypic and phenotypic divergence. Npj Syst Biol Appl 6:31. https://doi.org/10.1038/s41540-020-00151-9
56. Garcia S, Thompson RA, Giannone RJ, et al (2020) Development of a Genome-Scale Metabolic Model of Clostridium thermocellum and Its Applications for Integration of Multi-Omics Datasets and Computational Strain Design. Front Bioeng Biotechnol 8:772. https://doi.org/10.3389/fbioe.2020.00772

57. Larocque M, Chénard T, Najmanovich R (2014) A curated C. difficile strain 630 metabolic network: prediction of essential targets and inhibitors. BMC Syst Biol 8:117. https://doi.org/10.1186/s12918-014-0117-z
58. Zou W, Ye G, Zhang J, et al (2018) Genome-scale metabolic reconstruction and analysis for Clostridium kluyveri. Genome 61:605–613. https://doi.org/10.1139/gen-2017-0177
59. Marcellin E, Behrendorff JB, Nagaraju S, et al (2016) Low carbon fuels and commodity chemicals from waste gases – systematic approach to understand energy metabolism in a model acetogen. Green Chem 18:3020–3028. https://doi.org/10.1039/C5GC02708J
60. Valgepea K, de Souza Pinto Lemgruber R, Meaghan K, et al (2017) Maintenance of ATP Homeostasis Triggers Metabolic Shifts in Gas-Fermenting Acetogens. Cell Syst 4:505-515.e5. https://doi.org/10.1016/j.cels.2017.04.008
61. Norman ROJ, Millat T, Schatschneider S, et al (2019) Genome-scale model of C. autoethanogenum reveals optimal bioprocess conditions for high-value chemical production from carbon monoxide. Eng Biol 3:32–40. https://doi.org/10.1049/enb.2018.5003
62. Song Y, Lee JS, Shin J, et al (2020) Functional cooperation of the glycine synthase-reductase and Wood–Ljungdahl pathways for autotrophic growth of Clostridium drakei. Proc Natl Acad Sci 117:7516–7523. https://doi.org/10.1073/pnas.1912289117
63. Liu JK, Lloyd C, Al-Bassam MM, et al (2019) Predicting proteome allocation, overflow metabolism, and metal requirements in a model acetogen. PLOS Comput Biol 15:e1006848. https://doi.org/10.1371/journal.pcbi.1006848
64. Islam MA, Zengler K, Edwards EA, et al (2015) Investigating Moorella thermoacetica metabolism with a genome-scale constraint-based metabolic model. Integr Biol 7:869–882. https://doi.org/10.1039/c5ib00095e
65. Buendia-Kandia F, Rondags E, Framboisier X, et al (2018) Diauxic growth of Clostridium acetobutylicum ATCC 824 when grown on mixtures of glucose and cellobiose. AMB Express 8:85. https://doi.org/10.1186/s13568-018-0615-2
66. Simultaneous production of 1,3-propanediol and n-butanol by Clostridium pasteurianum: In situ gas stripping and cellular metabolism. https://doi.org/10.1002/elsc.201600058

CHAPTER 6

Ternary Approach to Analyze Ultrasonic Enhancement of Lipase-Catalyzed Esterification/ Transesterification Reactions



Available at: Kumar, K., Patro, P., Raut, U., Yadav, Y., Barbora, L., Moholkar, V. S.* (2023) *Biomass Conv. Bioref.* 1-12. https://doi.org/10.1007/s13399-023-04742-4

CHAPTER 6

TERNARY APPROACH TO ANALYSE ULTRASONIC ENHANCEMENT OF LIPASE-CATALYSED ESTERIFICATION/TRANSESTERIFICATION REACTIONS

6.1. INTRODUCTION

Conventionally, biodiesel is manufactured through an alkali-catalyzed transesterification process using vegetable oil or microalgal biomass as feedstock [1, 2]. Enzyme-catalyzed biodiesel synthesis has also been explored in recent years [1]. The most widely used enzyme for biodiesel synthesis is Lipase. Conventionally, lipase, a versatile commercial enzyme, finds applications in pharmaceutical industries for bioremediation and environmental processes [3]. The multifaceted applications of lipase make it a vital enzyme. Thus, its activity enhancement is crucial for industrial purposes. Numerous techniques have been attempted to enhance the kinetics of lipase-catalyzed reactions. One of these techniques is sonication or ultrasound (US) irradiation, which has proven effective for intensifying all routes to biodiesel synthesis, viz., esterification, transesterification, interesterification, and hydrolysis [4]. Sonication is a physical technique and does not involve toxic chemicals or harsh conditions and thus can be termed a "Green Technology" for enhancement of lipase-catalyzed reactions [5]. Recent studies [6–8] have revealed that the lipase enzyme activity increases significantly after exposure to the US with a concurrent reduction in the activation energy of the reactions.

The previous authors [6, 9] have reported the impact of US on the structure of lipases. However, the links between alteration in the structure of lipases due to sonication and its

activity enhancement (which essentially is manifested in faster reaction kinetics) have yet to be established. Such investigation would require elucidation of the exact nature of the interaction of the substrates/products with active site residues of the lipase enzyme on a molecular level.

This study presents a comprehensive investigation into the molecular mechanism of sonication-induced enhancement in the kinetics of lipase-catalyzed reactions for biodiesel synthesis. By combining three approaches—determining the structure of binding pockets of lipases, visualization of binding pockets in different secondary structure motifs, and docking analysis of ligands with lipases using hybrid quantum mechanics/molecular mechanics (QM/MM) computations—novel insights are gained into the exact nature of interactions between substrates/products and lipase enzyme on a molecular level. This research contributes to a deeper understanding of sonication's impact on lipase activity, providing a general framework for analyzing ultrasonic enhancement of reactions catalyzed by other enzymes.

6.2. METHODOLOGY

The schematic representation of workflow and steps followed in the present study for elucidation of the molecular mechanism of the sonication-induced enhancement in the kinetics of lipase-catalyzed reactions is given in **Figure 6.1**. Detailed procedure in a step-by-step manner followed in this study is presented in the following sections.

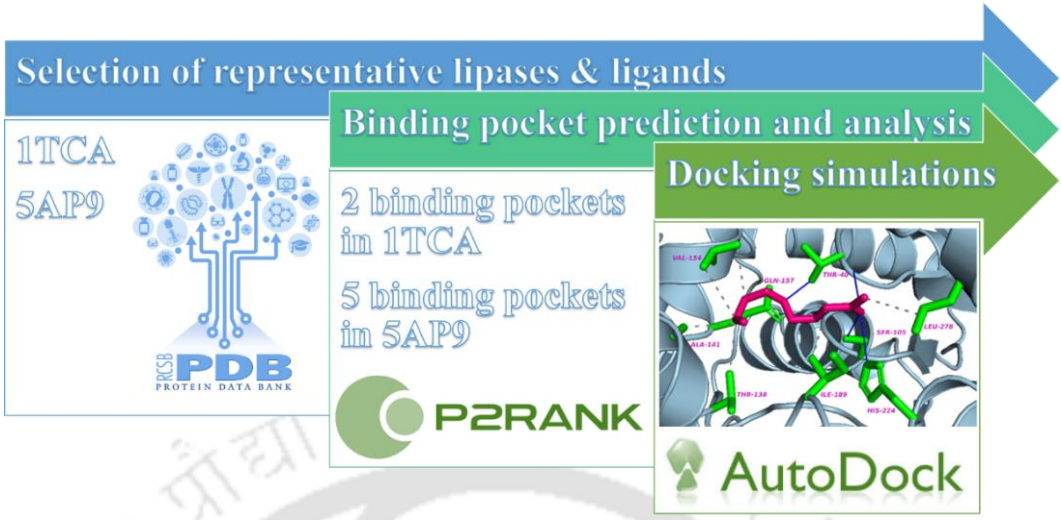


Figure 6.1 Schematic representation of computational methodology followed in the present study

6.2.1. Selection of Lipases and Substrates for Molecular Docking Experiment

For simulating esterification, transesterification, and hydrolysis reactions, we have chosen and imported two industrial lipases viz., *Candida antartica* lipase B (CALB; PDB ID: *1TCA*) and *Thermomyces lanuginous* lipase (TLL; PDB ID: *5AP9*), from RCSB – PDB database [10]. These selected crystal structures were of good quality among others (judged based on the lowest atomic resolution and the Ramachandran scores provided in the PDB database). As per the requirement of ligands for molecular docking, we drew and optimized the chemical structure of the following compounds: methanol (MET), butanol (BOH), levulinic acid (LA), n-butyl levulinate (n-BL), aspirin methyl ester (AME), methyl salicylate (MSC), and oleic acid (OA) as representative substrates/products of esterification, transesterification and hydrolysis reactions using density functional theory (DFT) method in standalone Gaussian 16 software [11].

6.2.2. Sequence Analysis and Binding Pocket Prediction (BPA) and Molecular Docking Analysis

Before proceeding with molecular docking simulations, we first performed the sequence analysis and binding pockets prediction using 2StrctCompare (https://2struccompare.cryst.bbk.ac.uk/index.php) and PrankWeb (https://prankweb.cz/) software package of P2Rank web servers, respectively. The retrieved crystal structures (*1TCA* and *5AP9*) were visualized in PyMOL™ [12], and the secondary structure was compared using 2StrctCompare. The crystal structure of CALB and TLL was submitted to PrankWeb [13] to predict the probable ligand-binding sites or binding pockets (BP) in proteins. The high energies of CALB and TLL were eliminated using SWISS-PDB viewer (SPDBV) [14], and local minima closer to the native structure were achieved.

6.2.3. Molecular Docking Simulation and Analysis

The molecular docking simulations of 7 ligands (described above) against both lipases were performed using Autodock v1.5.6 [22] linked with the MGLTools v1.5.6 viewer [23]. The ligand structure files were converted into “pdbqt” format using OpenBabel software. The procedure followed for molecular docking has been described in our earlier work [8]. The polar hydrogen ions and charges, such as Kollman and Gasteiger, were added to neutralize the system. A grid box was formed about coordinates for BP using the PrankWeb server. One hundred different docked conformations were generated using the Lamarckian Genetic Algorithm to investigate the optimal binding site of ligands to the lipases. These 100 docked conformations were ranked based on binding energies, and the best-docked conformation with the least binding energy was further analyzed in

visualization software. The polar and non-polar interactions between protein-ligand complexes were studied in AutoDock, PyMOL™ version 2.4.1 [19] and PLIP webtool (https://plip-tool.biotec.tu-dresden.de/plip-web/). The 2D schematic representation of the enzyme-ligand complex's molecular interactions was analyzed using LigPlot+ v.2.2 [24]. This protocol was reiterated for every ligand, and the final optimal conformation possessing the lowest binding energy among the hundred conformations procured for every operation was analyzed.

6.3. RESULTS AND DISCUSSION

6.3.1. Structural Analysis and Binding Pocket Characteristics

As outlined in the methodology section, the retrieved crystal structures of CALB and TLL were first visualized, and their sequences were compared using 2StructCompare. Both the lipases, CALB and TLL, are of α/β type fold and comprise 317 and 269 amino acid residues, respectively [15, 16]. DSSP (Database of Secondary Structure Assignments) algorithm for the prediction of secondary structures in CALB resulted in 37.5 % helix (H), 12.3 % extended sheets (E), and 50.2 % random structure (R), which include turn and coils. On the contrary, TLL consists of 32 % helix, 21.2 % extended sheets, and 46.8 % random structures.

The PrankWeb server identified the presence of two BP in CALB and five BP in TLL (refer to **Table 6.1**). The active site of CALB consists of SER_{105}-HIS_{224}-ASP_{187} with THR_{40}-GLN_{157} as an anionic hole contributing to the stabilization of the transition state. On the contrary, SER_{146}-HIS_{258}-ASP_{201} is present in active site in TLL. These active site residues are demonstrated in **Figures 6.2A (i)** & **(ii)** in blue color. **Figures 6.2A (i)** & **(ii)** also demonstrate the residues in substrate/product BP in CALB and TLL, respectively.

The amino acid composition of BP and results of 2D structure analysis using the DSSP algorithm are provided in **Table 6.1**. The minimap of secondary structure comparison produced using the 2StructCompare web server for CALB and TLL is provided in **Figure 6.2B**.

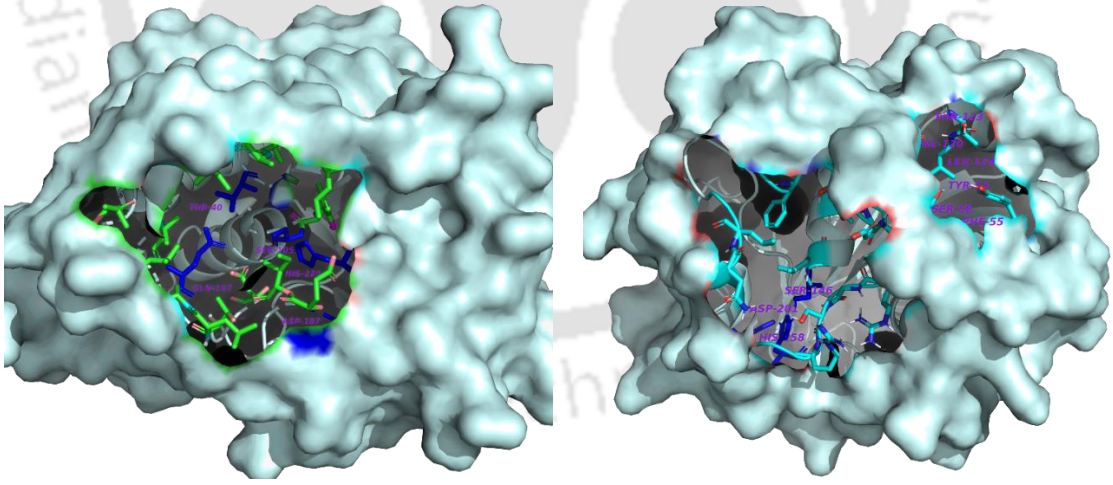
Out of the multiple BPs in CALB and TLL, the pocket with the highest score was selected for molecular docking studies. We observed the presence of the glycine, leucine, valine, and serine amino acid residues in most BPs (**Table 6.1**). The other amino acid residues in the BPs are alanine, glutamine, histidine, isoleucine, phenylalanine, and tryptophan (**Table 6.1**). The PyMOL visualization of the BPs (shown in **Figure 6.2A**) revealed that most of these residues are in the random coil region of the enzyme. **Figure 6.2A** shows the position of the most probable BP (i.e., BP with high scores) for substrate and product during esterification/transesterification reactions. Previous studies have reported that α-helices are mainly involved in protecting and activating the lid region of the lipase [3, 6]. In contrast, β (or extended)-sheets are primarily engaged in the structural integrity of the enzyme structure [3, 4, 17]. The secondary structural comparison of CALB and TLL is shown in **Figure 6.2B**.

Table 6.1. Prediction of binding pockets for various commercial lipases using PrankWeb

PDB ID	# of BP	Score	Composition of the binding pocket	2D structure analysis using the DSSP method		
				% H	% E	% R
1TCA	2	22.13	Chain A – THR_{40} TRP_{104} SER_{105} ASP_{134} THR_{138} LEU_{140} ALA_{141} LEU_{144} VAL_{149}	37.5	12.3	50.2

			VAL_{154} GLN_{157} GLU_{188} ILE_{189} HIS_{224} LEU_{278} ALA_{281} ALA_{282} ILE_{285}			
		1.51	Chain A – ASP_{145} TRP_{155} THR_{158} GLY_{160} SER_{161} GLU_{294} ARG_{309}			
5AP9	5	6.29	Chain A – GLY_{61} ASP_{62} VAL_{63} THR_{64} ARG_{81} GLY_{82} SER_{83} ARG_{84} ASN_{92} LEU_{93} PHE_{113} SER_{116} LEU_{147} LEU_{151}	32.0	21.2	46.8
		6.19	Chain B – HIS_{110} GLY_{112} PHE_{113} SER_{116} LEU_{147} LEU_{151} ASP_{62} ARG_{81} GLY_{82} SER_{83} ARG_{84} SER_{85} TRP_{89} ASN_{92} ASN_{94}			
		1.47	Chain B – VAL_{120} THR_{123} LEU_{124} TYR_{53} PHE_{55} SER_{58}			
		1.22	Chain A – THR_{153} ALA_{173} PRO_{174} VAL_{176} ARG_{195} SER_{214} HIS_{215} SER_{216}			
		0.79	Chain B – ALA_{19} ALA_{20} CYS_{22} ILE_{34} PHE_{51} SER_{54} PHE_{55} GLY_{65} PHE_{66}			

Abbreviation: BP – binding pockets; H – helix; E- extended sheets; R – random structures



(i) CALB (ii) TLL

Figure 6.2 (A): Visualization of most probable binding pockets of (i) CALB and (ii) TLL in PyMOL predicted by PrankWeb. Large cavities in the structure represent the highest-scored binding pockets of respective lipases listed in **Table 6.2**. Active site amino acid residues are represented in blue color

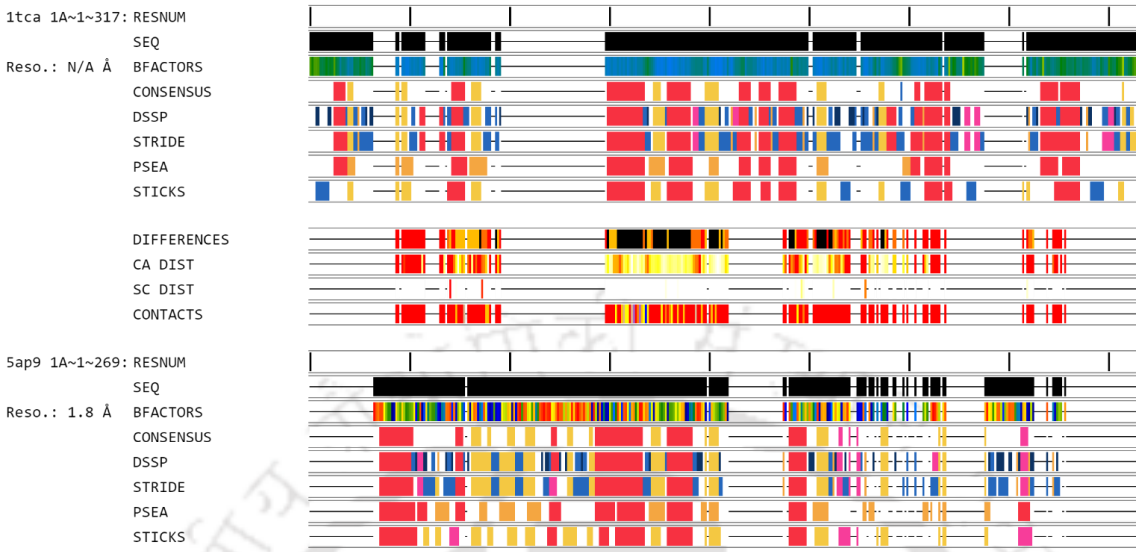


Figure 6.2(B): Minimap of secondary structure comparison produced using 2StructCompare web server for both CALB and TLL

6.3.2. Molecular docking analysis shows central involvement of α-helix and random coil in complex formation

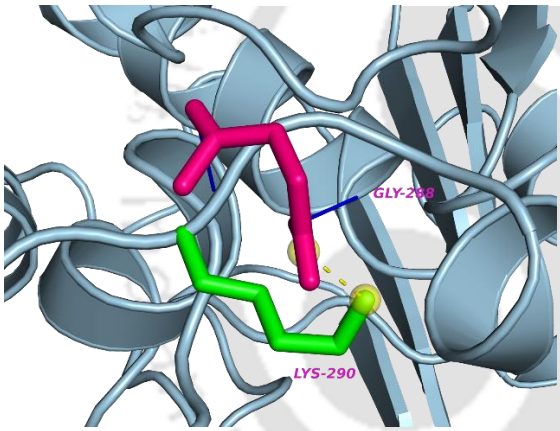
Molecular docking studies were conducted to get insight into the molecular mechanism of lipase-catalyzed esterification, transesterification, and hydrolysis reactions. Protocols and steps followed for docking simulations are provided in section 6.2.2. The results summarized in **Table 6.2A** and **Figure 6.3** reveal the significant interactions in the lipase-ligand (here, ligand represents either substrate or product of esterification/transesterification/hydrolysis reaction) complexes. These interactions are of type hydrophobic bonds, hydrogen bonds, and salt bridge interactions. Most of the interacting amino acids were present in the α-helix and random coil domain of the lipase, as seen in **Table 6.2A**. Docking system indices (DSI) in **Table 6.2A** correspond to the type of ligands (i.e., substrates and products) with which the docking analyses of lipase

enzymes have been performed, which maps with **Table 6.2B** for respective lipase-catalyzed reactions.

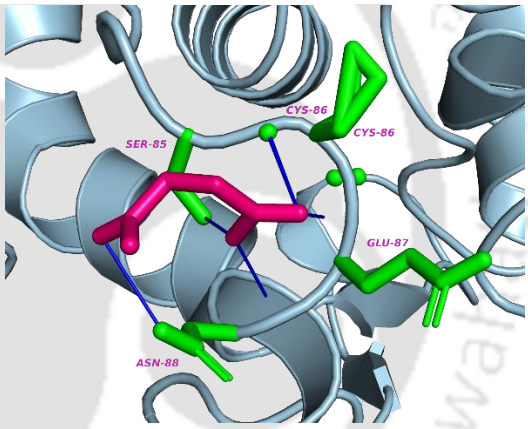
As discussed in the introduction section of this study, biodiesel synthesis occurs majorly through esterification or transesterification reactions. The role of CALB and TLL in reversible catalysis of esterification/transesterification/hydrolysis reactions has been extensively studied for biodiesel synthesis [1, 6, 9]. **Table 6.2B** summarizes the types of reactions studied in the present research. **Tables 6.2A** & **B** show that DSIs A1, A2, and A3 represent levulinic acid's esterification with n-butanol. Similarly, DSIs (B0, B6) and (A0, A6) represent the transesterification reactions of mixed non-edible oils with methanol catalyzed by TLL and rapeseed oil with methanol catalyzed by CALB, respectively. The DSIs (A4 and A5) represent the hydrolysis reaction of aspirin methyl ester catalyzed by CALB.

It can be seen from **Table 6.2A** that for esterification and transesterification reactions, alcohol (i.e., methanol and butanol) substrates have higher binding energy compared to organic acid (levulinic acid and oleic acid) substrates. This suggests that alcohols have a lower binding affinity towards lipases than organic acid. Therefore, during the ping-pong bi-bi mechanism, organic acid will bind first with lipase, followed by the binding of alcohol after the release of the first product. Additionally, ester products (i.e., n-butyl levulinate and FAME) have a low binding affinity towards lipases compared to their substrates (i.e., alcohols and organic acids). Hence, they hence will be faster from the [E-P] complex. Similar trends were also observed in the hydrolysis reactions (DSIs: A4, A5, B4, and B5), where the substrate's binding affinity was lower than the product's binding affinity.

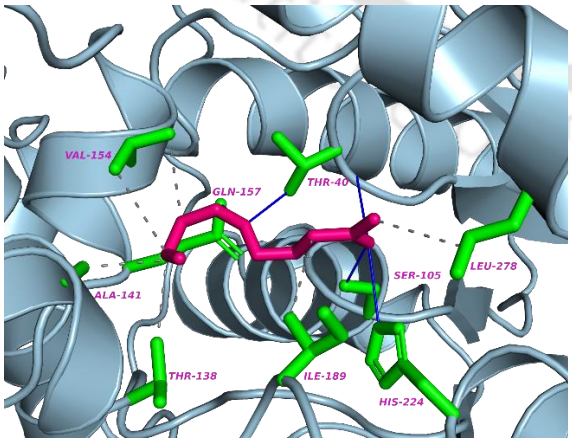
Figure 6.3 (i – xiv) depicts ligands' binding and types of interactions in the respective lipases. It is overserved from **Figure 6.3** that in the BP, two substrates (acid and alcohols) bind in the two distinct domains (i.e., alcohol-binding domain and acyl-binding domain). These two binding domains help orient the substrate more effectively for catalysis to occur efficiently [6, 18]. Similar binding characteristics were observed in our study for esterification, transesterification, and hydrolysis reactions. It is evident from **Figure 6.3 (i – xiv)** that most of the enzyme-ligand binding and their interaction happen in the random coil domain of the lipase.

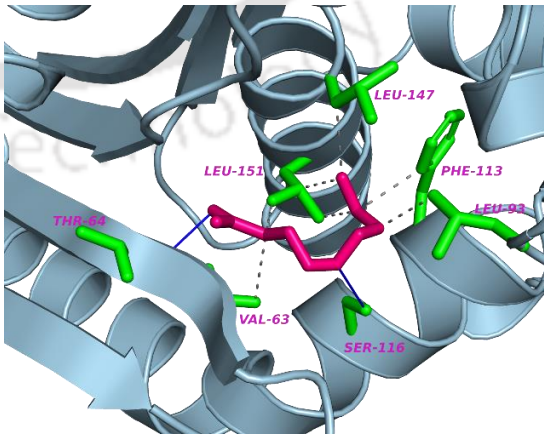


(i) (1TCA – LA)

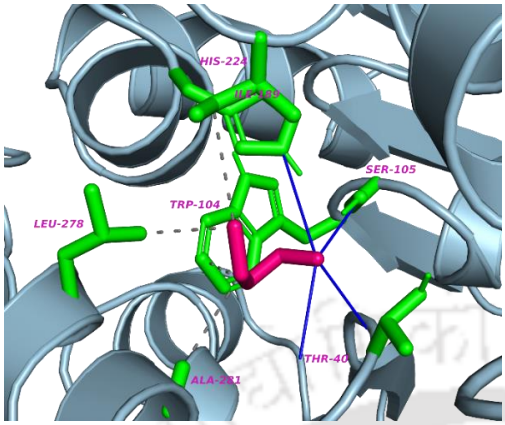


(viii) (5AP9 – LA)

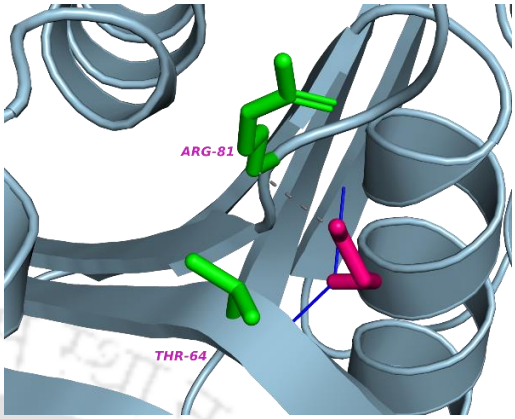




(ii) (1TCA – n-BL)

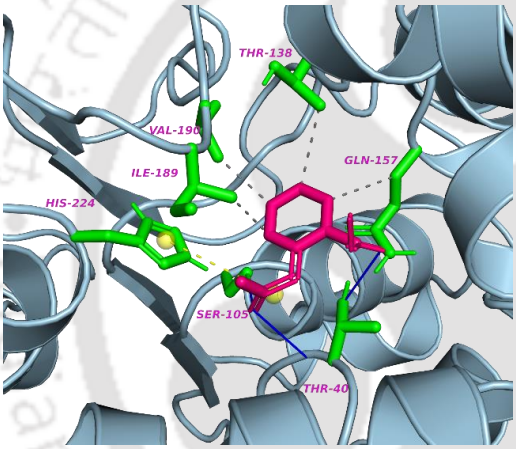


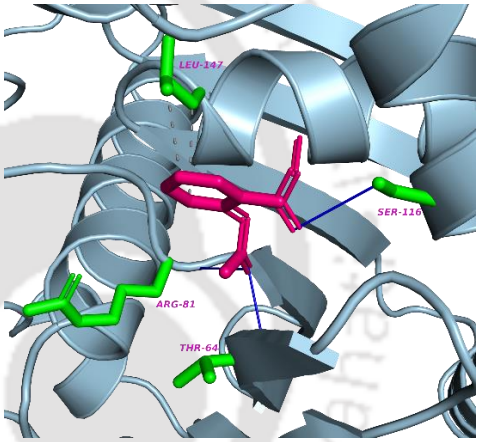
(ix) (5AP9 – n-BL)



(iii) (1TCA – BOH)

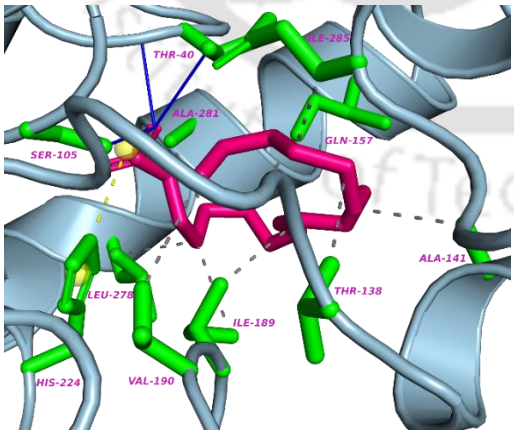
(x) (5AP9 – BOH)

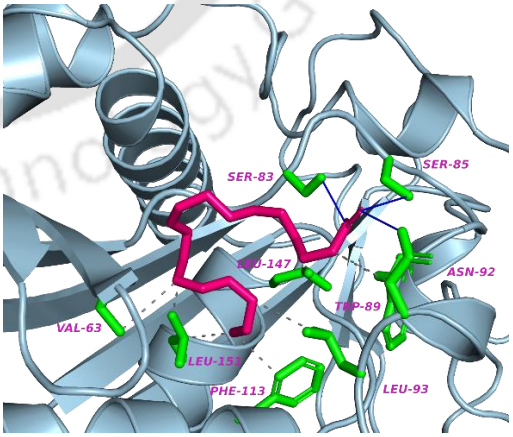




(iv) (1TCA – AME)

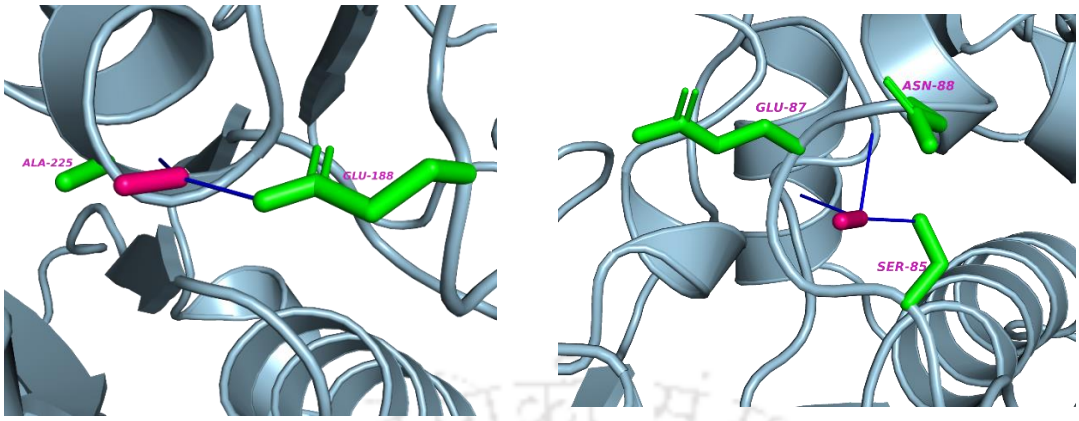
(xi) (5AP9 – AME)



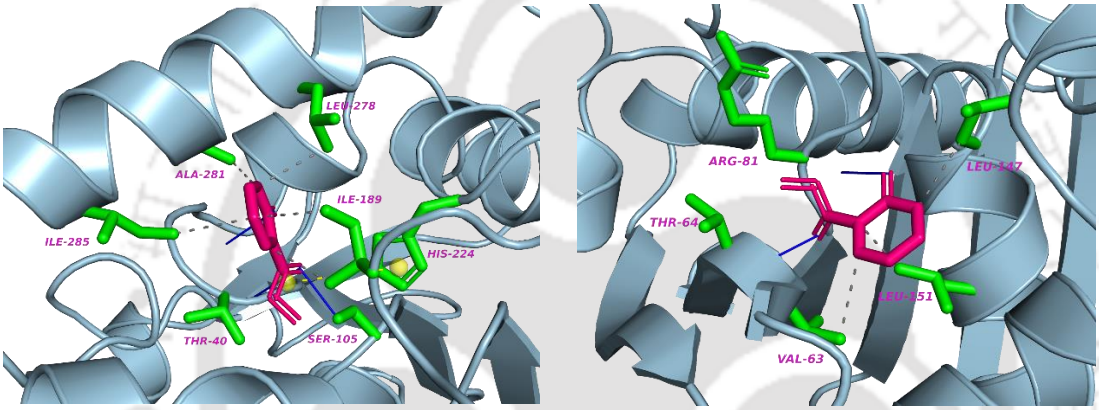


(v) (1TCA – OA)

(xii) (5AP9 – OA)



(vi) (1TCA – MET) (xiii) (5AP9 – MET)



(vii) (1TCA – MSC) (xiv) (5AP9 – MSC)

Figure 6.3: Interaction profile of lipase CALB and TLL with respective substrates and products of esterification, transesterification, and hydrolysis reactions.

Table 6.2 (A): Molecular docking results of interaction of various ligands with CALB and TLL lipases viz., *1TCA* and *5AP9*

DSI	Definition	Binding energy (kJ/mol)	Amino acids involved in binding	Region of the protein	Type of interactions
A0	*1TCA* + MET	-1.93	GLU_{188}, ALA_{225}	α-helix and random coil	Hydrogen bonds
A1	*1TCA* + BOH	-2.53	TRP_{104}, ILE_{189}, LEU_{278}, ALA_{281}, THR_{40}, SER_{105}, HIS_{224}	α-helix, random	Hydrophobic, Hydrogen bonds

				coil, β-sheets	
A2	*1TCA* + LA	-2.91	GLY_{288}, LYS_{290}	Random coil	Hydrogen bonds, Salt bridge
A3	*1TCA* + n-BL	-4.04	THR_{138}, ALA_{141}, VAL_{154}, GLN_{157}, ILE_{189}, LEU_{278}, THR_{40}, SER_{105}, GLN_{157}, HIS_{224}	α-helix, random coil, and β-sheets	Hydrophobic, Hydrogen bonds
A4	*1TCA* + AME	-4.35	THR_{138}, GLN_{157}, ILE_{189}, VAL_{190}, THR_{40}, SER_{105}, HIS_{224}	α-helix, random coil, and β-sheets	Hydrophobic, Hydrogen bonds, Salt bridge
A5	*1TCA* + MSC	-3.37	ILE_{189}, LEU_{278}, ALA_{281}, ILE_{285}, THR_{40}, SER_{105}, HIS_{224}	α-helix and random coil	Hydrophobic, Hydrogen bonds, Salt bridge
A6	*1TCA* + OA	-4.41	THR_{138}, ALA_{141}, GLN_{157}, ILE_{189}, VAL_{190}, LEU_{278}, ALA_{281}, ILE_{285}, THR_{40}, SER_{105}, HIS_{224}	α-helix, random coil, and β-sheets	Hydrophobic, Hydrogen bonds, Salt bridge
B0	*5AP9* + MET	-2.25	SER_{85}, GLU_{87}, ASN_{88}	Random coil	Hydrogen bonds
B1	*5AP9* + BOH	-2.96	ARG_{81}, THR_{64}	Random coil	Hydrophobic, Hydrogen bonds
B2	*5AP9* + LA	-4.43	SER_{85}, CYS_{86}, GLU_{87}, ASN_{88}	Random coil	Hydrogen bonds
B3	*5AP9* + n-BL	-3.96	VAL_{63}, LEU_{93}, PHE_{113}, LEU_{151}, THR_{64}, SER_{116}	α-helix and random coil	Hydrophobic, Hydrogen bonds
B4	*5AP9* + AME	-4.55	LEU_{147}, THR_{64}, ARG_{81}, SER_{116}	α-helix and random coil	Hydrophobic, Hydrogen bonds
B5	*5AP9* + MSC	-4.29	VAL_{63}, LEU_{147}, LEU_{151}, THR_{64}, ARG_{81}	α-helix and random coil	Hydrophobic, Hydrogen bonds
B6	*5AP9* + OA	-4.28	VAL_{63}, TRP_{89}, ASN_{92}, LEU_{93}, PHE_{113}, LEU_{147}, LEU_{151}, SER_{83}, SER_{85}, ASN_{92}	α-helix and random coil	Hydrophobic, Hydrogen bonds

Table 6.2(B): Molecular docking analysis of representative lipase-catalyzed reactions for biodiesel synthesis

Experimental conditions				DSI*		Ref.
Reaction type	Enzyme	Substrates	Product	ES complex	EP complex	
Esterification	Novozym 435	Levulinic acid and butanol	Butyl levulinate (BL)	A1, A2	A3	[6]
Transesterification	TLL	Non-edible oils and methanol	Fatty acid methyl ester (FAME)	B0, B6	-	[1]
Hydrolysis	CALB	Aspirin methyl ester (AME)	Methyl salicylate	A4	A5	[19]
Transesterification	Novozym 435	Rapeseed oil and methanol	Fatty acid methyl ester	A0, A6	-	[20]

** refer to **Table 6.2A** for the DSI: docking system index code.*

6.3.3. Sonication-induced structural and morphological changes in lipase

6.3.3.1. Preamble

Ultrasound is a longitudinal wave that passes through any compressible medium through compression/rarefaction cycles. This induces small amplitude and high-frequency oscillatory motion of the fluid elements of the medium. Such oscillatory motion generates intense micro-mixing in the medium. The passage of ultrasound waves also induces variation in the bulk pressure in the medium. The secondary effect of sonication is cavitation –nucleation, growth, volume oscillations, and implosive collapse of gas or vapor bubbles driven by pressure variation induced by the ultrasound waves. The volume oscillations of cavitation bubbles also produce intense micro-convection near the bubbles. In addition, cavitation bubbles also give rise to acoustic waves due to the reflection of the

fluid elements from the bubble/fluid interface. These physical phenomena also contribute to intense mixing in the medium. A peculiar feature of the convection generated by ultrasound and cavitation is its tiny spatial and temporal scales, i.e., this convection is caused at the micro-level (only a few micrometers) and for a very short duration (for example, the period of 50 microseconds for an ultrasound wave of 20 kHz). Such convection can induce micro-level phenomena in the medium, such as the unfolding of the protein molecules, which also leads to the rearrangement of the secondary structure of the protein molecules [21, 22]. A representative literature summary on the US-assisted enhancement of lipase-catalyzed trans-/inter-/esterification reactions for the synthesis of biofuels or their chemical additives is provided in **Table 6.3**.

Table 6.3: Summary of the representative literature on ultrasound enhanced lipase-catalyzed reactions

Reaction	Experimental conditions	Results	Ref.
Esterification	CALB & *RML*; T = 40 ℃; *f* = 20 kHz; Sub. - FFA and xylitol; No solvent	Efficient synthesis of xylitol fatty ester (95% yield), with xylitol monoacyl ester and xylitol diacyl ester appearing as the main products (greater than 96%)	[23]
	Novozyme 435; *f* = 35 kHz; T = 50℃; P = 35 W; Sub.: LA and BOH; Sol: BME	Application of 35 kHz sonication boosted n-butyl levulinate yield from 70.9% to 92.22%.	[6]
	Novozyme 435; T = 50℃; P = 32 mW; Sub.: AA and isomylOH	Application of US for 20 min, the ester concentration increased by 27.4% with higher enzyme	[7]

		reusability.	
	Novozyme 435; f = 28 kHz; T = 50°C; P = 1.75 kW/m^2; Sub.: butyric acid-lauric acid	Improved synthesis efficiency of butyric acid- lauric acid designer lipid by 11.38% with application of US.	[24]
Transesteri-fication	TLL; T = 36°C; Sub.: non-edible oils and MeOH; $(MR)_{MET/oil}$ = 7.64; EL = 3.55%	US + water addition to the reaction mixture (10% v/v) boosted biodiesel yield from 90% to 94%.	[1]
	BCL; f = 37 kHz; T = 60°C; P = 500 W; Sub.: castor oil, EtOH, $(MR)_{MET/oil}$ = 6; EL = 12g, US probe	FAEE yield of 83% in 72 h with an ethanol-to-oil molar ratio of 3:1 in the case of the US	[25]
	Novozyme 435; f = 28 kHz; T = 50°C; Sub.: DHA and EPA	99% conversion to FAEE with sonication at the condition of 1 ml/min flow rate	[26]
Interesterification	CALB; P = 165W; f = 25 kHz; T = 60°C; Sub. - Crambe oil, MeOH; $(MR)_{MET/oil}$ = 12;	The US reduced total reaction time and percentage of enzyme loading by 20 wt%.	[27]

6.3.3.2. Secondary Structural Analysis of Lipase Before and After Sonication

In the context of the present study, numerous biochemical reactions catalyzed by lipases are accelerated due to sonication. A summary of representative recent literature on the enhancement effect of sonication (in terms of faster kinetics and higher yields) on different lipase-catalyzed reactions with changes in the 2D structural composition of lipase is presented in **Table 6.4**. It could be observed that these reactions involve other feedstocks and lipases from various sources. Moreover, operating conditions of sonication (such as frequency, intensity, duration, etc.) have also varied. Nonetheless, enhancement

is seen in all cases, although the extent of enhancement varies from case to case. The ultrasound-induced enhancement in the kinetics of lipase-catalyzed reactions is a manifestation of the enzyme's enhanced activity, which is a consequence of the modifications in the secondary structure of the enzyme.

Previous authors have deduced the changes in the secondary structure of lipase due to sonication using different techniques such as circular dichroism, X-ray crystallography, SEM, deconvolution of FTIR spectra, NMR, and fluorescence spectroscopy [6, 17, 28–30]. The results of previous literature showing the alteration in the secondary structure of lipase have been summarized in **Table 6.4**. As can be inferred from **Table 6.4**, the changes in secondary structure are highly lipase specific and a vital function of the operational conditions of sonication, such as frequency, intensity, and duration. However, on a broader basis, the significant changes in the secondary structure of lipase induced by ultrasound can be identified as (1) a rise in alpha helix content and (2) a rise in random coils. Few studies have reported a reduction in alpha helix content with a concurrent increase in beta sheets and turn. Our previous study [6] revealed that sonication increased the α-helix content that enhanced the activity of immobilized CALB by opening the flapping lid and widening the catalytic cavity hosting the binding pockets. Sonication at 35 kHz also caused a rise in the random coil content of lipase. In a similar study on the lipase-catalyzed synthesis of sucrose fatty acid ester, Zhang et al. [31] reported an increase in alpha helices in the enzyme structure of CALB using circular dichroism analysis. This shift in the secondary structure of CALB contributed to a 75% reduction in the reaction time.

As per the results presented in **Tables 6.2A** and **B** for two representative lipase enzymes CALB and TLL, the binding pockets of all ligands (viz. methanol, butanol, levulinic acid, n-butyl levulinate, aspirin methyl ester, methyl salicylate, and oleic acid) are mainly present in α-helix and random coils. Thus, a rise in α-helix and random coil content of lipase widens the catalytic cavity and relaxes the enzyme structure [8, 9], resulting in easy access of substrates to BP. Additionally, a rise in random coil content favors the formation of enzyme-ligand complexes (with a possible reduction in inhibition and unfavorable dissociation of intermediate complexes), ultimately enhancing the reaction kinetics and product yield. Similar observations were also found in our previous study [6]. On the other hand, increased beta sheets and turns contents caused reduced lipase activity, as observed through CD analysis by Sun et al. [32]. In addition, SEM studies by Liu et al. [22] and Zhang et al. [31] revealed the presence of extra holes in the enzyme morphology, which suggest a rise in the surface area. This feature also contributes to faster enzymatic reaction rates.

The ternary approach of binding pocket analysis and concurrent analysis of the molecular docking simulations (**Tables 6.2A** & **B** and **Figure 6.3**) in the present study and the changes in 2D structure of lipase due to sonication reported in the literature (**Table 6.4**) provide mechanistic insight into the beneficial influence of ultrasound on lipase-catalyzed reactions. The ternary approach developed in this study to deduce the molecular mechanism of US-induced enhancement in lipase-catalyzed reactions for biodiesel synthesis is schematically depicted in **Figure 6.4**.

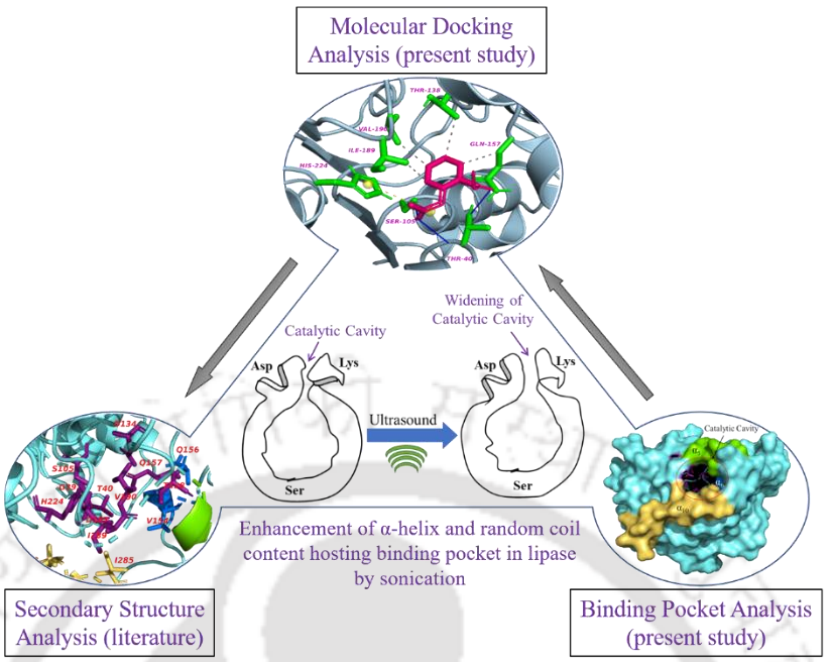


Figure 6.4: Schematic depiction of the ternary approach to deduce the molecular mechanism of US-induced enhancement in lipase-catalyzed reactions for biodiesel synthesis

Table 6.4: Influence of ultrasound on secondary structures of lipase: a literature summary

Reaction type	Experimental Conditions	Observed enhancement	Changes in secondary structures	Ref.
Esterification	Novozyme 435; *f* = 35 kHz; T = 50°C; P = 35 W; duty cycle = 20%, time = 1 h; ultrasonic bath; Sub.: LA and BOH	Boosted conversion of LA from 70.9% to 92.22%	Increased α-helix and random coil; concurrent decrease in β-sheets	[6]
Hydrolysis	CVL; *f* = 25.8, 26.2, 35.5, 36.5 kHz; T = 25°C; P = 29-349 W, time = 20 min; ultrasonic bath; Sub.: Olive oil with PEG	The half-life of PEG400-treated lipase was 27-fold higher than untreated	No change in the circular dichroism spectra of the secondary structure	[33]

Enzyme activity	CALB and TLL; *f* = 22 kHz; T = 50°C; P = 35 W; duty cycle = 50-67%, time = 20-30 min; probe sonicator	The activity of CALB and TLL enhanced by 1.3-ad 1.5-fold, respectively.	Increase the random coil by 25.99 and 21.54% in both lipases; the rest components decreased.	[34]
Hydrolysis	PPL; *f* = 60 kHz; T = 37°C; P = 138 W; duty cycle = 50-67%, time = 240 min; probe sonicator; Sub.: coconut oil	Lauric acid yield increased in the presence of sorbitol and mannitol osmolytes.	Increased α-helix and random coil; concurrent decrease in β-sheets	[35]
Transesterification	BCL; *f* = 40 kHz; T = 40°C; P = 120 W; duty cycle = 50-67%, time = 2 h; ultrasonic bath; Sub.: IL and methanol	The best reaction media for biodiesel yield is 82.22 %. Anionic IL was the most effective.	Increased α-helix, β-sheets, β-turns; concurrent decrease in random coil and turns	[36]
Hydrolysis	Immobilized *alcalase*, *f* = 22, 28, 33, 40, 68 kHz; T = 60°C; P = 600 W each, time = 30-150 min; ultrasonic bath; Sub.: rapeseed protein	In US assisted-hydrolysis, hydrolysis increased by 74.38% at the optimum conditions.	The number of a-helix and random coil decreased by 10.7% and 4.5%, β -chain increased by 2.4%	[37]

6.3.3.3. Suggestions for Future Work Based on the Present Study

The present study attempts to offer a theoretical explanation for the ultrasound-induced enhancement of lipase-catalyzed reactions based on the ternary approach of determining the physical and chemical characteristics of the binding pockets of lipases, deducing their locations in different structural motifs and molecular docking analysis of other ligands

using QM/MM simulations. Although this analysis could be plausible effort, an experimental confirmation of the computation results presented in this work requires to be made. The experimental study would require the lipase enzyme in free form (unlike the most commercially available enzymes in immobilized form). This work will also require high-end analytical instruments like X-ray crystallography, NMR solution, and Cryo-electron microscopy. The scientific community in enzyme research with access to these facilities can take up the future research in this area. We believe that our computational study will be a useful tool for designing sucb an experimental endeavour but also analysis of the results.

6.4. CONCLUSIONS

This study has attempted to provide mechanistic insight into the influence of sonication on lipase-catalyzed reaction with a ternary approach of determining the physical and chemical characteristics of the binding pockets of lipases, deducing their locations in different structural motifs and molecular docking analysis of other ligands using QM/MM simulations. It has been observed that most of the binding pockets are located in either α-helix or random coils. The previous literature has reported significant modifications in the secondary structure of lipases after exposure to sonication such as rise in α-helix and random coils contents. The net consequence of sonication is thus widening the catalytic cavity and easy accessibility for substrate binding. This is manifested in a rise in enzyme activity and faster kinetics of the lipase-catalyzed reactions. Therefore, the results of present study have provided a mechanistic explanation to the previous literature on ultrasound-assisted intensification of esterification, transesterification, and hydrolysis

reactions. Moreover, the methodology of ternary approach proposed in the present study also forms a general framework for analyzing the mechanism of ultrasonic enhancement of other enzyme-catalyzed reactions that are enhanced with sonication.

REFERENCES

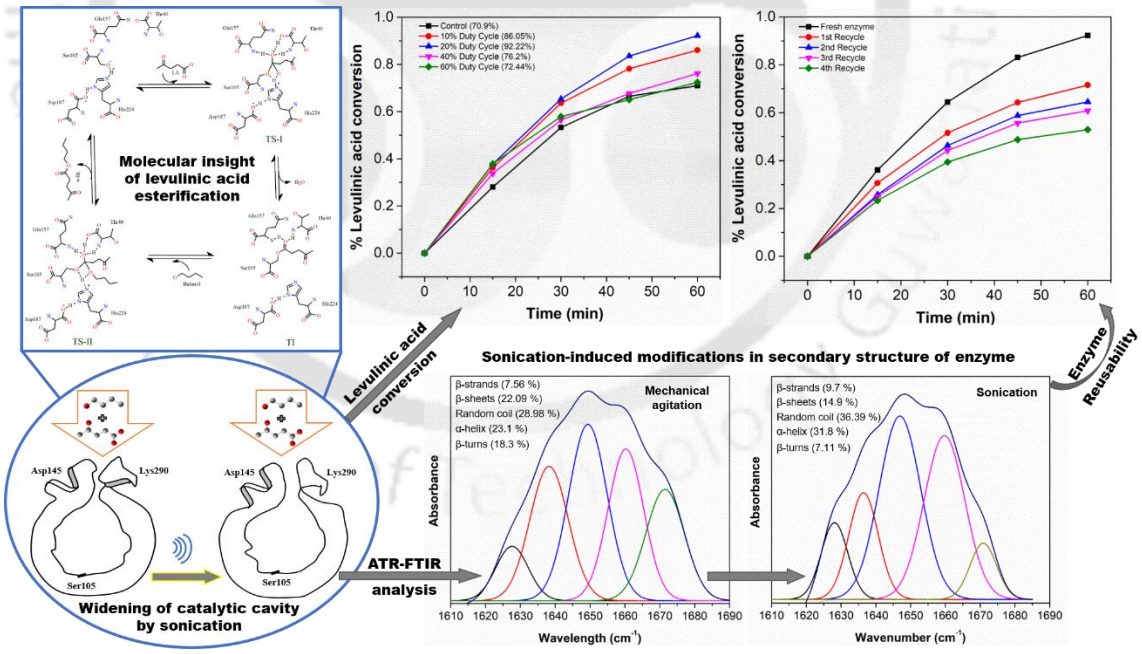
1. Malani RS, Umriwad SB, Kumar K, et al (2019) Ultrasound–assisted enzymatic biodiesel production using blended feedstock of non–edible oils: Kinetic analysis. Energy Convers Manag 188:142–150. https://doi.org/10.1016/j.enconman.2019.03.052
2. Singh N, Kumar K, Goyal A, Moholkar VS (2022) Ultrasound-assisted biodiesel synthesis by in–situ transesterification of microalgal biomass: Optimization and kinetic analysis. Algal Res 61:102582. https://doi.org/10.1016/j.algal.2021.102582
3. Filho DG, Silva AG, Guidini CZ (2019) Lipases: sources, immobilization methods, and industrial applications. Appl Microbiol Biotechnol 103:7399–7423. https://doi.org/10.1007/s00253-019-10027-6
4. Javed S, Azeem F, Hussain S, et al (2018) Bacterial lipases: A review on purification and characterization. Prog Biophys Mol Biol 132:23–34. https://doi.org/10.1016/j.pbiomolbio.2017.07.014
5. Lerin LA, Loss RA, Remonatto D, et al (2014) A review on lipase-catalyzed reactions in ultrasound-assisted systems. Bioprocess Biosyst Eng 37:2381–2394. https://doi.org/10.1007/s00449-014-1222-5
6. Kumar K, Roy K, Moholkar VS (2021) Mechanistic investigations in sonoenzymatic synthesis of n-butyl levulinate. Process Biochem 111:147–158. https://doi.org/10.1016/j.procbio.2021.09.005
7. Vartolomei A, Calinescu I, Vinatoru M, Gavrila AI (2022) A parameter study of ultrasound assisted enzymatic esterification. Sci Rep 12:1421. https://doi.org/10.1038/s41598-022-05551-x
8. Wang D, Yan L, Ma X, et al (2018) Ultrasound assisted interesterification of waste cooking oil and methyl acetate for biodiesel and triacetin production. Int J Biol Macromol 119:453–461. https://doi.org/10.1016/j.ijbiomac.2018.07.133
9. Bansode SR, Rathod VK (2017) An investigation of lipase catalysed sonochemical synthesis: A review. Ultrason Sonochem 38:503–529. https://doi.org/10.1016/j.ultsonch.2017.02.028
10. Zardecki C, Dutta S, Goodsell DS, et al (2016) RCSB Protein Data Bank: A Resource for Chemical, Biochemical, and Structural Explorations of Large and Small Biomolecules. J Chem Educ 93:569–575. https://doi.org/10.1021/acs.jchemed.5b00404
11. Curtiss LA, Redfern PC, Raghavachari K (2007) Gaussian-4 theory. J Chem Phys 126:084108. https://doi.org/10.1063/1.2436888
12. Yuan S, Chan HCS, Hu Z (2017) Using PYMOL as a platform for computational drug design. WIREs Comput Mol Sci 7:. https://doi.org/10.1002/wcms.1298

13. Jendele L, Krivak R, Skoda P, et al (2019) PrankWeb: a web server for ligand binding site prediction and visualization. Nucleic Acids Res 47:W345–W349. https://doi.org/10.1093/nar/gkz424
14. Guex N, Peitsch MC (1997) SWISS-MODEL and the Swiss-Pdb Viewer: An environment for comparative protein modeling. Electrophoresis 18:2714–2723. https://doi.org/10.1002/elps.1150181505
15. Höck H, Engel S, Weingarten S, et al (2018) Comparison of *Candida antarctica Lipase* B Variants for Conversion of ε-Caprolactone in Aqueous Medium—Part 2. Polymers 10:524. https://doi.org/10.3390/polym10050524
16. Sandstrom AG (2010) Protein engineering of *Candida antarctica* Lipase A: enhancing enzyme properties by evolutionary and semi-rational methods. Department of Organic Chemistry, Stockholm University
17. Khan FI, Lan D, Durrani R, et al (2017) The Lid Domain in Lipases: Structural and Functional Determinant of Enzymatic Properties. Front Bioeng Biotechnol 5:. https://doi.org/10.3389/fbioe.2017.00016
18. Kumar K, Moholkar VS (2023) Mechanistic Aspects of Enhanced Kinetics in Sonoenzymatic Processes Using Three Simultaneous Approaches. In: Moholkar VS, Mohanty K, Goud VV (eds) Sustainable Energy Generation and Storage. Springer Nature Singapore, Singapore, pp 41–57
19. Chiplunkar PP, Zhao X, Tomke PD, et al (2018) Ultrasound-assisted lipase catalyzed hydrolysis of aspirin methyl ester. Ultrason Sonochem 40:587–593. https://doi.org/10.1016/j.ultsonch.2017.08.004
20. Firdaus MY, Brask J, Nielsen PM, et al (2016) Kinetic model of biodiesel production catalyzed by free liquid lipase from *Thermomyces lanuginosus*. J Mol Catal B Enzym 133:55–64. https://doi.org/10.1016/j.molcatb.2016.07.011
21. Badgujar VC, Badgujar KC, Yeole PM, Bhanage BM (2021) Investigation of effect of ultrasound on immobilized *C. rugosa* lipase: Synthesis of biomass based furfuryl derivative and green metrics evaluation study. Enzyme Microb Technol 144:109738. https://doi.org/10.1016/j.enzmictec.2020.109738
22. Liu S, Dong X, Wei F, et al (2015) Ultrasonic pretreatment in lipase-catalyzed synthesis of structured lipids with high 1,3-dioleoyl-2-palmitoylglycerol content. Ultrason Sonochem 23:100–108. https://doi.org/10.1016/j.ultsonch.2014.10.015
23. Nieto S, Villa R, Donaire A, Lozano P (2021) Ultrasound-assisted enzymatic synthesis of xylitol fatty acid esters in solvent-free conditions. Ultrason Sonochem 75:105606. https://doi.org/10.1016/j.ultsonch.2021.105606
24. Liu W, Luo X, Tao Y, et al (2022) Ultrasound enhanced butyric acid-lauric acid designer lipid synthesis: Based on artificial neural network and changes in enzymatic structure. Ultrason Sonochem 88:106100. https://doi.org/10.1016/j.ultsonch.2022.106100
25. Moreira MAC, Alnoch RC, Luz Junior LF de L, et al (2022) Transesterification of castor oil catalyzed by a fermented solid produced by *Burkholderia contaminans* using a bioreactor coupled with ultrasound irradiation. Biotechnol Appl Biochem n/a: https://doi.org/10.1002/bab.2410
26. Kuo C-H, Tsai M-L, Wang H-MD, et al (2022) Continuous Production of DHA and EPA Ethyl Esters via Lipase-Catalyzed Transesterification in an Ultrasonic Packed-Bed Bioreactor. Catalysts 12:404. https://doi.org/10.3390/catal12040404

27. Tavares GR, Gonçalves JE, dos Santos WD, da Silva C (2017) Enzymatic interesterification of crambe oil assisted by ultrasound. Ind Crops Prod 97:218–223. https://doi.org/10.1016/j.indcrop.2016.12.022
28. Anobom CD, Pinheiro AS, De-Andrade RA, et al (2014) From Structure to Catalysis: Recent Developments in the Biotechnological Applications of Lipases. BioMed Res Int 2014:684506. https://doi.org/10.1155/2014/684506
29. Uppenberg J, Oehrner N, Norin M, et al (1995) Crystallographic and molecular-modeling studies of lipase B from *Candida antarctica* reveal a stereospecificity pocket for secondary alcohols. Biochemistry 34:16838–16851. https://doi.org/10.1021/bi00051a035
30. Zhang H, Sang J, Zhang Y, et al (2019) Rational design of a *Yarrowia lipolytica* derived lipase for improved thermostability. Int J Biol Macromol 137:1190–1198. https://doi.org/10.1016/j.ijbiomac.2019.07.070
31. Zhang JC, Zhang C, Zhao L, Wang CT (2014) Lipase-Catalyzed Synthesis of Sucrose Fatty Acid Ester and the Mechanism of Ultrasonic Promoting Esterification Reaction in Non-Aqueous Media. Adv Mater Res 881–883:35–41. https://doi.org/10.4028/www.scientific.net/AMR.881-883.35
32. Sun T, Dong Z, Wang J, et al (2020) Ultrasound-Assisted Interfacial Immobilization of Lipase on Hollow Mesoporous Silica Spheres in a Pickering Emulsion System: A Hyperactive and Sustainable Biocatalyst. ACS Sustain Chem Eng 8:17280–17290. https://doi.org/10.1021/acssuschemeng.0c06271
33. Talukder M. MR, Shiong SCS (2015) Stabilization of *Chromobacterium viscosum* Lipase (CVL) Against Ultrasound Inactivation by the Pretreatment with Polyethylene Glycol (PEG). Appl Biochem Biotechnol 177:1742–1752. https://doi.org/10.1007/s12010-015-1850-5
34. Nadar SS, Rathod VK (2017) Sonochemical Effect on Activity and Conformation of Commercial Lipases. Appl Biochem Biotechnol 181:1435–1453. https://doi.org/10.1007/s12010-016-2294-2
35. Mojtabavi S, Hervé M, Forootanfar H, et al (2022) A survey on the stabilizing effect of osmolytes on the ultrasound-irradiated lipase for efficient enzymatic hydrolysis of coconut oil. Colloids Surf B Biointerfaces 220:112910. https://doi.org/10.1016/j.colsurfb.2022.112910
36. Liu Y, Chen D, Yan Y, et al (2011) Biodiesel synthesis and conformation of lipase from *Burkholderia cepacia* in room temperature ionic liquids and organic solvents. Bioresour Technol 102:10414–10418. https://doi.org/10.1016/j.biortech.2011.08.056
37. Wang B, Meng T, Ma H, et al (2016) Mechanism study of dual-frequency ultrasound assisted enzymolysis on rapeseed protein by immobilized Alcalase. Ultrason Sonochem 32:307–313. https://doi.org/10.1016/j.ultsonch.2016.03.023

CHAPTER 7

Mechanistic Investigations in Sonoenzymatic Synthesis of N-Butyl Levulinate



Available at: Kumar, K., Roy, K., & Moholkar, V. S.* (2021). *Process Biochemistry*, 111, 147-158. https://doi.org/10.1016/j.procbio.2021.09.005

CHAPTER 7

MECHANISTIC INVESTIGATIONS IN SONOENZYMATIC SYNTHESIS OF N-BUTYL LEVULINATE

7.1. INTRODUCTION

Among the various valuable chemicals derived from lignocellulosic biomass, levulinic acid (LA), also known as gamma-ketovaleric acid, is a versatile platform chemical with numerous potential applications [1, 2]. LA has been highlighted by the United States Department of Energy as a promising building block for chemistry in 2004 and 2010 because it has carboxylic and ketone as functional groups [3]. Alkyl levulinates are industrially important derivatives of LA, produced by esterifying the carboxyl group of LA with alcohols. Moreover, the low toxicity, high lubricity, stable flash point and moderate flow property make them suitable candidate for gasoline and diesel fuel additives [4, 5]. Comparatively, the potential application of n-butyl levulinate (n-BL) has been proved to be more promising than its alternatives levulinates such as methyl or ethyl levulinate. n-BL has high potential for applications in production of biofuels, green solvents, fragrance industries, and pharmaceutical intermediates [2, 6–8]. Biobutanol produced through acetone-butanol-ethanol (ABE) fermentation of lignocellulosic substrates, has opened up another possible inexpensive platform for production of n-BL [1].

As discussed in **Chapter 6**, *Candida antartica* lipase B (CALB) has been one of the extensively applied enzymes in many industrial processes because of its thermal stability, high enantioselectivity, wide range of substrates, and stability in organic solvents [9–13]. Although enzymes as biocatalysts have numerous merits over

conventional inorganic catalysts, they have not been implemented commercially due to high costs and slow reaction kinetics [14]. Significant research has been carried out for enhancing the kinetics of biocatalytic processes by using microwave [15], supercritical fluid [16, 17], ionic liquids [18], applications of modifiers [19], and ultrasound [14, 20, 21]. These techniques improve the kinetics of enzymatic reactions through various mechanisms such as efficient energy input to the system, elimination of mass transfer barriers and enhancement of activity of enzymes through modifications in enzyme structure [20, 22, 23].

In the present study, we have reported ultrasound-assisted enzymatic esterification of LA with n-butanol using Novozym 435 (CALB immobilised on microporous polyacrylic resin). The main components of our study with mechanistic view point are: 1. Optimization of reaction parameters; 2. Determination of changes in secondary structure of enzyme during sonication using ATR-FTIR analysis, 3. Kinetic analysis of experimental profiles using Ping-Pong Bi-Bi kinetic model, 4. Molecular docking simulation to reveal the molecular mechanism of various steps of CALB-catalysed LA esterification reaction. The present study has not only demonstrated enhancement in kinetics of esterification using sonication, but has also investigated the process in depth from mechanistic angle. Using ternary approach developed in **Chapter 6**, we have identified the structural components of lipase and the amino acid residues associated with catalysis. The experimental results indeed showed favourable changes in secondary structure of enzyme that contributed to enhancement in catalytic activity. To the best of our knowledge, this was the first study which has given molecular mechanics account in lipase catalysed esterification and role of sonication. The dual approach of experiments and computational analysis in

this study has revealed several interesting mechanistic aspects of influence of sonication on esterification for synthesis of n-BL.

7.2. MATERIALS AND METHODS

7.2.1. Materials

Candida antartica Lipase B (CALB) immobilized on microporous polyacrylic resin (Novozym® 435) was purchased from Sigma-Aldrich (India). Levulinic acid (AR grade) and 1-butanol (HPLC grade, 99.8%) were purchased from Himedia (India). *t*-butyl methyl ether (extra pure) was procured from Finar (India).

7.2.2. Statistical Design of Experiments (DOE)

The selection of Novozym® 435 as a catalyst for the esterification reaction was based on a previous study [2, 24], which reported comparison of several commercial lipases for enzymatic synthesis of n-BL. Novozym 435 performed best with maximum initial rate and conversion [2]. To begin with, we carried out several preliminary tests for n-BL synthesis by using one factor at a time method. The major factors governing yield of n-BL are: molar ratio of butanol to LA, enzyme quantity and agitation speed. Thereafter, parametric optimization of enzymatic esterification was done using CCD statistical design (3-factors-2-level full factorial design with 8 cube points and 6 axial points and 20 base runs). This was followed by substitution of mechanical agitation with sonication, while other parameters held at their optimum value. An important parameter related to sonication is duty cycle (i.e. fraction of unit reaction time in which sonication is applied in-lieu of mechanical agitation, for example, for unit reaction time of 5 min, 20% duty cycle corresponds to 1 min sonication and 4 min mechanical agitation). The duty cycle of sonication was also optimized for maximum yield and kinetics.

7.2.2.1. Central Composite Design (CCD) for Process Parameter Optimization

The central composite design comprising 3 factors and 2 levels was used for optimization with n-BL yield as response variable. The factors and levels of DoE are shown in **Table 7.1A**. The experimental design constituted 20 individual runs (= 2^k + $2k$ + n_0), where 'k' is the number of independent variables, and n_0 is the number of replicates runs at the center point of the variable. The combination of parameters in each experiment of the statistical design is given in **Table 7.1B**. The statistical analysis was performed according to the response surface methodology using MINITAB software (Release 16.1, PA, USA, Trial Version).

Table 7.1 (A): Factors and levels of optimization variables in central composite design of experiments

Variables	Coded values (Actual values)	
(E) - Enzyme amount (mg)	–1 (20)	+1 (100)
(M) - Molar ratio (Butanol: LA)	–1 (1:1)	+1 (5:1)
(S) - Agitation (rpm)	–1 (250)	+1 (750)

7.2.2.2. Experimental Setup

All esterification reactions were conducted in batch mode. A typical reaction mixture consisted of n-butanol and LA in the desired molar ratio (as obtained from statistical experimental design). This mixture was diluted to 15 mL with *t*-butyl methyl ether as a solvent. The reaction mixture was taken in a 30 mL borosilicate vial followed by vigorous mixing (750 rpm) at 50°C on magnetic stirrer cum hot plate for 30 min. The desired enzyme amount was then added to the reaction mixture to initiate the reaction. As reported in several previous studies [2, 14], the temperature of the reaction was set at 50 °C ± 1°C and maintained using a thermostatic water bath. Each esterification

experiment was conducted for 1 h. The progress of reaction was followed by periodic withdrawal of aliquots of reaction mixture to analyze the conversion of LA.

Esterification experiments (conducted at optimum conditions determined by statistical experimental design) were categorized as follows:

1. Control experiments, in which the reaction was carried out with mechanical agitation at 633 ± 3 rpm using a magnetic stirrer.
2. Test experiments, in which mechanical agitation was supplemented with sonication at optimum duty cycle.

The test experiments were conducted using an ultrasound bath (Elma Transonic, T-460, Germany) with a total working volume = 2 L; frequency = 35 kHz; theoretical power = 35 W. The bath was filled to 75% of its total working volume with distilled water, which acted as a medium of the ultrasound waves. The position of the vial in ultrasound bath was carefully maintained constant in all experiments as the acoustic intensity and pressure amplitude of ultrasound waves show significant spatial variation. All experiments were repeated twice to confirm the reproducibility of results.

7.2.2.3. Statistical Analysis and Model Fitting

The experimental data of CCD was fitted to quadratic model:

$$Y = \beta_0 + \sum_{i=1}^{k} \beta_{ii} X_i + \sum_{i=1}^{k} \beta_{ii} {X_i}^2 + \sum_{i \neq j} \sum_i \beta_{ij} X_i X_j \quad \textbf{(7.1)}$$

$$x = (x_i - x_0)/\Delta x \quad \textbf{(7.2)}$$

where, $i = 1,\ 2,\ 3, \ldots, x_i$ are the dimensionless values of the variables; x_0 is the value of x_i at the center point; and Δx is the step change.

7.2.3. Experimental Analysis

The residual concentration of LA in the aliquots of reaction mixture was analyzed using Shimadzu Ultra High-Performance Liquid Chromatography (UHPLC, Model: SPD-20A) equipped with an Aminex-HPX 87H column and RID detector. The mobile phase was 0.1 N H_2SO_4 at a flow rate of 0.6 mL/min. n-BL in esterification reaction was confirmed by ^1H NMR (Nuclear magnetic resonance) spectroscopy (Bruker advance III HD Ascend 600 MHz) with TMS (tetramethyl silane) as internal standard and $CDCl_3$ (Merck, India) as a solvent.

7.2.4. Enzyme Kinetics Using Ping-Pong Bi–Bi Models

CALB is a globular α/β-type protein with a structure characterized by mostly parallel β-sheets surrounded by α-helices. CALB consists of a catalytic triad composed of Ser105, Asp187, and His224 residues, and an oxyanion hole formed by the Thr40 and Gln157 residues, where helices α5, α6, and α10 make up most of the active site pocket for interfacial activation and substrate affinity [10, 25].

In order to get insight into the biophysical mechanism of LA esterification using CALB, the time profiles of substrates and products were analyzed using enzyme kinetics model. Previous literatures have reported, that kinetics of CALB catalyzed esterification is best described by Ping-Pong Bi-Bi kinetic model. This model involves formation of two tetrahedral complexes, an acyl-enzyme complex, and also competitive substrate inhibition by the alcohols (which are used as acyl acceptors) [2, 9, 12, 26–30]. We have attempted to determine the kinetic parameters of the LA esterification reaction with four different kinetic models based on Ping-Pong Bi-Bi mechanism. For convenience of readers, the complete derivation of the basic kinetic expression of enzymatic reactions following general Ping-Pong Bi-Bi mechanism is

provided in the **Appendix 5**. The basic kinetic expression gets simplified under special circumstances which are described below (for greater details refer to **Appendix 5**).

7.2.4.1. Enzyme Kinetics in Absence of any Influence of Products

The general schematic representation of Ping-Pong Bi-Bi mechanism is shown in **Figure 7.1A** and molecular representation of the same in the present context of CALB-catalyzed esterification of LA to n-BL is shown in **Figure 7.1B**. If the enzymatic esterification is not influenced by presence of products (or in other words the products do not interfere with enzyme), the kinetic expression is:

$$v_0 = \frac{V_{\max}[A][B]}{K_B[A]+K_A[B]+[A][B]} \qquad \textbf{(7.3)}$$

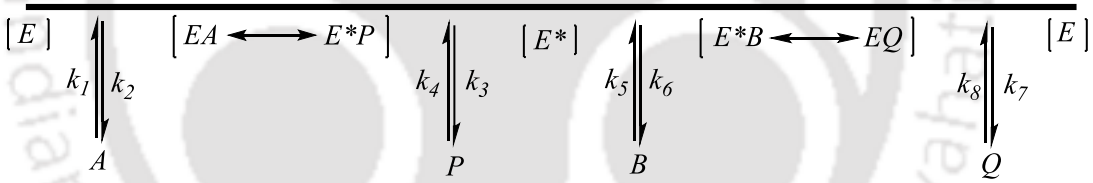


Figure 7.1(A): Schematic representation of basic Ping-Pong Bi-Bi lipase action mechanism

As depicted in **Figure 7.1A**, *A* is the substrate possessing functional group/moiety to be transferred to ester, and *B* is the acceptor. Similarly, *P* is the product of *A*, while *Q* is the product of *B* plus the moiety transferred from *A* during reaction. In the present case of LA esterification with n-butanol, *A* = LA, *B* = butanol, *P* = water, and *Q* = n-BL.

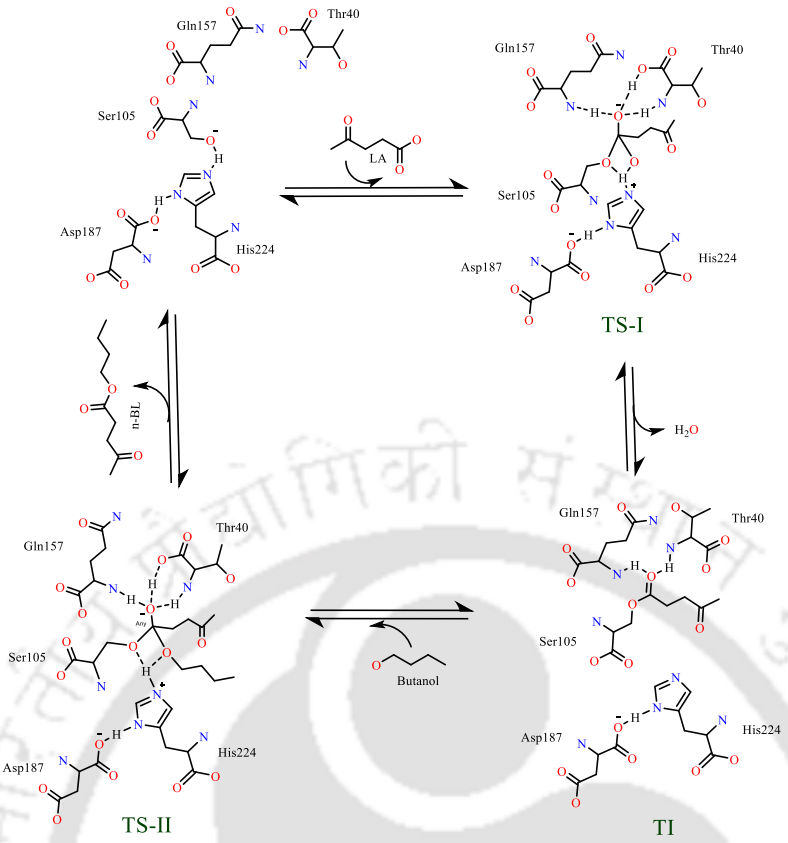


Figure 7.1(B): Molecular mechanism of levulinic acid esterification using CALB

7.2.4.2. Enzyme Kinetics with Inhibition by N-butanol

Previous authors have reported that the combination of n-butanol with CALB enzyme leads to competitive inhibition, which also follows Ping-Pong Bi-Bi mechanism [2, 24, 30, 31]. In addition to reaction with acyl -enzyme complex, n-butanol also reacts with free enzyme leading to formation of dead-end complex (*EB*). The schematic representation of this case is given in **Figure 7.1C**.

The kinetic expression of esterification involving competitive inhibition by n-butanol is:

$$v_0 = \frac{V_{\max}[A][B]}{K_B[A]+K_A[B]\left(1+\frac{[B]}{K_i}\right)+[A][B]} \qquad \textbf{(7.4)}$$

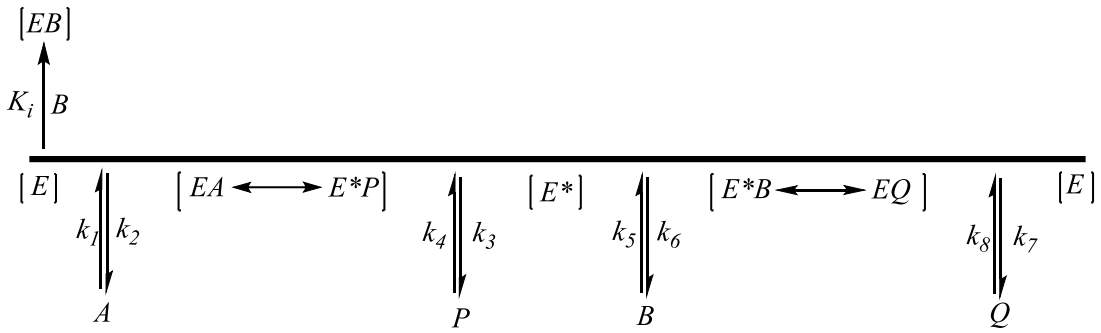


Figure 7.1(C): Schematic representation of Ping-Pong Bi-Bi mechanism with alcohol inhibition

7.2.4.3. Enzyme Kinetics with Noncompetitive Inhibition by Product P

Sufficient built-up of the product concentration in the reaction mixture leads to product inhibition. In the present context, water released during esterification can complex with CALB leading to noncompetitive inhibition of the enzyme. The CALB can form a complex *EP*, following a noncompetitive inhibition by *P* (water) with K_{iP} as dissociation constant. The kinetic expression for this condition is:

$$v_0 = \frac{V_{\max}[A][B]}{K_B[A]+K_A[B]+[A][B]+\frac{K_{iA}K_B}{K_{iP}}[P]+\frac{K_B}{K_{iP}}[A][P]} \qquad \textbf{(7.5)}$$

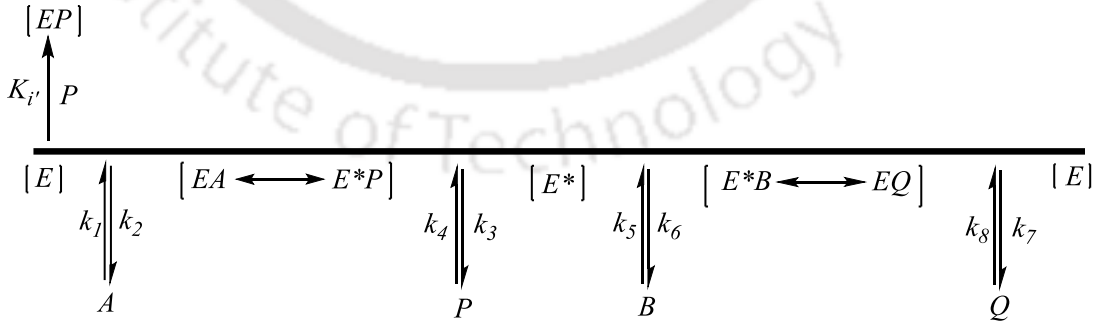


Figure 7.1(D): Schematic representation of Ping-Pong Bi-Bi mechanism with product inhibition

7.2.4.4. Enzyme Kinetics with Mixed Dead-End and Product Inhibition

In any typical biochemical transformation, the products are homologous (or bearing some structural resemblance) with the substrates. This may trigger combination between product and enzyme (in improper form) leading to formation of a dead-end complex. A frequent case in the Ping-Pong Bi-Bi mechanism is that the first product *P* combines with enzyme (in stable form) to produce a stable dead-end *EP* complex. This process is analogous to dead-end inhibition by a nonreacting inhibitor. A schematic representation of this enzyme kinetics is given in **Figure 7.1D**.

The initial velocity (or rate) of such reactions is:

$$v_0 = \frac{V_{\max}[A][B]}{K_B[A]+K_A[B]\left(1+\frac{[P]}{K_{i'}}\right)+[A][B]+\frac{K_B}{K_{iP}}[A][P]+\left(\frac{K_{iA}K_B}{K_{iP}}[P]\right)\left(1+\frac{[P]}{K_{i'}}\right)} \qquad \textbf{(7.6)}$$

In all schemes discussed above, the fraction of the total enzyme in complexed form - either as (*EA* + *E*P*) or (*E*B* + *EQ*) - cannot be determined. Therefore, we have assumed it to be in pseudo steady state condition, which essentially implies that isomerization of any of the transitory complexes has no effect on distributions of the enzyme in any state, either free or complexed.

For fitting of the kinetic expressions to experimental profiles, the conversion curve was discretized to calculate the initial reaction velocities (v_0) between two consecutive sampling points. The experimental data was fitted to **equations 7.3**, **7.4**, **7.5** and **7.6** using multiregression solver in MS-Excel (2019) implementing GRG nonlinear algorithm to obtain various kinetic constants (as listed in **Table 7.2**) in the Ping-Pong-Bi-Bi model. This procedure was repeated for experiments in both categories, viz. control and test.

7.2.5. Effect of Ultrasound on Secondary Structure of CALB

The changes in secondary structure of Novozym 435 (CALB immobilized on microporous polyacrylic resin) induced due to sonication were determined using FTIR-based technique reported in literature [32]. The secondary derivatives of peak frequencies in 1600-1690 cm^{-1} were recognized and smoothened with a 20-point Savitzky-Golay algorithm [32]. The multicomponent peak area under amide-I bands of Novozym 435 in the control and test experiments was then quantified by multi-peak fitting program by using Gaussian function in Origin 9.0. The fractions of the secondary structures were calculated using methodology reported in the previous literature [33].

7.2.6. Molecular Modelling of the Esterification Reaction

Molecular docking simulation studies were carried out in order to understand the molecular interaction and identification of the active site catalytic important amino acid residues of the enzyme with substrates, products and transition states. The crystal structure coordinates of CALB were obtained from the Research Collaboratory for Structural Bioinformatics – Protein Data Bank (RCSB-PDB). Among several crystal structures available for CALB, the structure PDB ID: 1TCA (with a resolution of 1.55 Å) was selected. Molecular docking simulations were carried out with AutoDock 4.2.1 software [34] linked with MGLTools 1.5.6 (http://mgltools.scripps.edu/). Using Gaussian 09 package [35], the structures of ligands (LA, n-butanol, n-BL, water and transition states) were generated, followed by minimization of the energy. The two N-acetyl-D-glucosamine (NAG) moieties and water crystals present in the CALB structure were removed in AutoDock. Charges were stabilized by the addition of polar hydrogens, and addition and redistribution of Kollman charges. The grid (size 42 × 40

× 40 Å), based on positions of active site residues, was dispersed around the ligand binding site with grid spacing of 0.5 Å. The compound was then endorsed for flexible docking by keeping protein as rigid structure. 100 docking simulations were performed using the Lamarckian Genetic Algorithm method to search for the optimum binding site of ligands to the protein. Out of these 100 clustered docked conformations ranked on the basis of binding energies, the best docked conformation of ligands bound to the active site of CALB with least binding energy was further analyzed. The identification of polar and non-polar interaction between protein-ligand complexes were then analyzed in AutoDock, PyMOL™ version 2.4.1, Protein-Ligand Interaction Profiler web tool (https://plip-tool.biotec.tu-dresden.de/plip-web/) and PROTEINS PLUS web tool (https:proteins.plus/), separately. To validate the molecular docking simulation experiments, ligands were extracted from this co-crystal and then again docked in CALB to compare the protein binding coordinates with the crystal form and also identify the binding affinity. The 2D schematic representation of the molecular interactions in protein-ligand complex was analyzed by using LigPlot^{+} v.2.2 program [36].

7.2.7. Reusability of Enzyme

The reusability of immobilized enzyme for repeated experimental runs was also evaluated. After completion of an experiment, the reaction was terminated with addition of the mixture of acetone and ethanol (50:50). The enzyme was separated by centrifuging the mixture at 6000×g at 25 ℃ for 10 min. The recovered enzyme was washed thrice with acetone to remove the adsorbed impurities such as butanol. The washed immobilized enzyme was dried in oven at 40 ℃ for 60 min and allowed to cool in a desiccator for 24 h. This enzyme was utilized for the subsequent esterification

reaction with fresh LA and butanol in *t*-butyl methyl ether. This procedure was also repeated without conditioning of separated enzyme.

7.3. RESULTS AND DISCUSSION

7.3.1. Optimization of Enzymatic Esterification Process

As noted earlier, the three parameters of the LA conversion to the n-BL (molar ratio, enzyme quantity and speed of agitation) were optimized using statistical central composite design. The full factorial CCD matrix of these variables is given in **Table 7.1B**, along with the response variable of actual and model predicted conversion of LA. The values of experimental percentage conversion of LA are the average of two consecutive runs with standard deviation. The reduced quadratic model in terms of coded values of process variables was obtained by fitting the experimental data to **equation 7.1** as follows:

$$\%\,Y = 65.869 + 5.717 \times E + 6.483 \times M + 6.519 \times S + 2.552 \times E \times M - 2.486 \times E \times S - 3.316 \times M \times S - 4.043 \times E^2 - 8.653 \times M^2 - 3.347 \times S^2 \quad (7.7)$$

where E, M, and S are the coded values of enzyme quantity (mg), molar ratio, and speed of the agitation, respectively.

Table 7.1 (B): Experiments of central composite design for levulinic acid esterification

Sr. No.	Enzyme quantity (mg)	Molar ratio	Agitation (rpm)	% conversion (Experiments)	% conversion (Model Prediction)
1	20	5	250	42.34 ± 0.40	42.35
2	60	1	500	50.74 ± 1.27	50.73
3	100	5	250	63.85 ± 0.96	63.86

4	60	3	500	65.85 ± 0.53	65.87
5	20	1	750	52.50 ± 0.45	52.50
6	60	5	500	62.72 ± 0.29	63.70
7	60	3	500	65.85 ± 0.53	65.87
8	20	3	500	56.09 ± 1.23	56.11
9	60	3	500	65.85 ± 0.53	65.87
10	20	1	250	27.83 ± 1.03	27.86
11	100	5	750	65.29 ± 1.25	65.30
12	100	1	250	39.15 ± 1.04	39.16
13	60	3	750	69.04 ± 0.98	69.04
14	60	3	250	56.03 ± 0.49	56.00
15	100	1	750	53.87 ± 0.28	53.86
16	60	3	500	65.85 ± 0.53	65.87
17	60	3	500	65.85 ± 0.53	65.87
18	60	3	500	65.85 ± 0.53	65.87
19	20	5	750	53.73 ± 0.95	53.73
20	100	3	500	67.56 ± 0.69	67.54

Figure 7.2A shows the desirability plot which depicts the optimum set of experimental parameters: enzyme quantity = 86 mg; molar ratio (butanol/LA) = 3.74:1; speed of agitation = 633 rpm. The validation experiment was carried out by using this set of parameters, which yielded LA conversion of 70.9 ± 1.4 %. This is very close to the predicted LA conversion of 70.71% by the quadratic model. **Figure 7.2B** shows the ^{1}H NMR spectrum of the reaction mixture at the end of the enzymatic esterification reaction at optimized conditions. Different characteristic peaks identified from ^{1}H NMR spectrum are as follows: a singlet peak at 2.22 ppm represents the methyl carbon which is allylic signal, whereas triplet peak at 4.09 ppm is characteristic peak for esters, which comes from the oxygenated sp^3 hybridized C-H's. The signal at 0.95 ppm results from the terminal methyl protons. Other peaks at 2.77, 2.61, 1.58, and 1.41 ppm correspond to methylene protons.

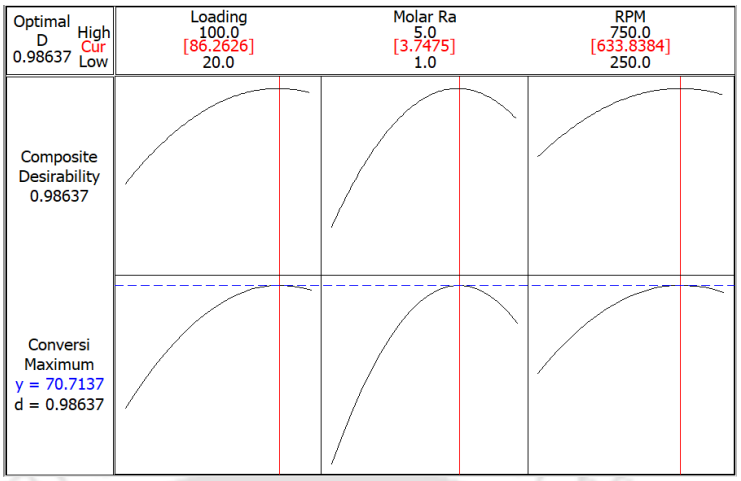


Figure 7.2: (A) Desirability plot for optimization of process parameters

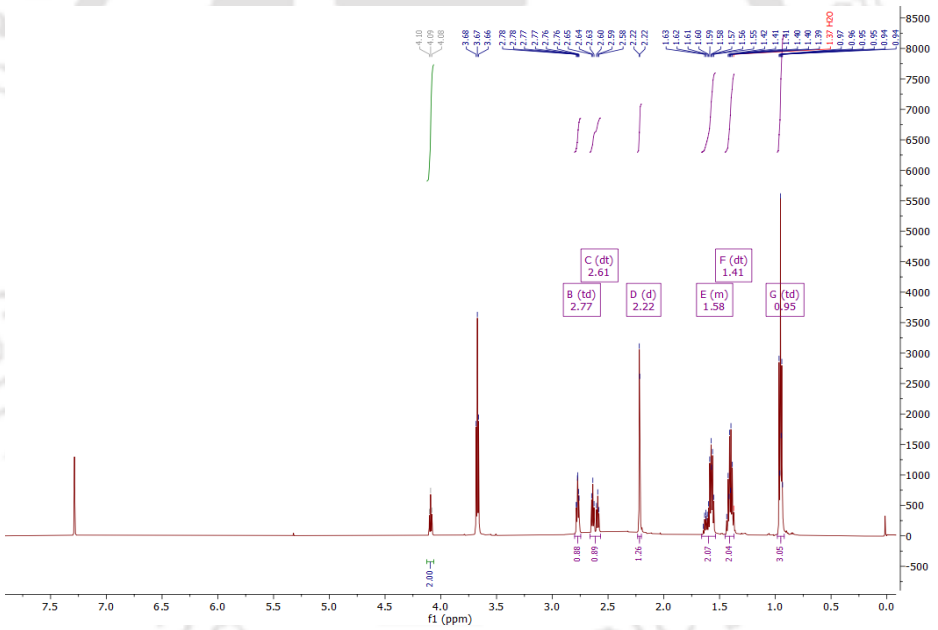


Figure 7.2 (B): ^{1}H NMR spectra of enzyme catalysed esterification reaction at optimum conditions at the end of reaction

7.3.1.1. Test Experiments

In the test experiments, the mechanical agitation was replaced by sonication at optimum duty cycle. The duty cycle was optimized keeping all other variables at their optimum values. The conversion profiles at different duty cycles are shown in **Figure**

7.3. It could be seen from **Figure 7.3** that maximum conversion was obtained for the duty cycle of 20 %, and hence, the duty cycle was maintained at 20% in all further experiments. The rated (or theoretical) power rating of the ultrasound bath used for experiments was 35 W. The actual power input to the system was determined using calorimetric method, and the acoustic pressure amplitude in the medium was calculated as 1.5 bar.

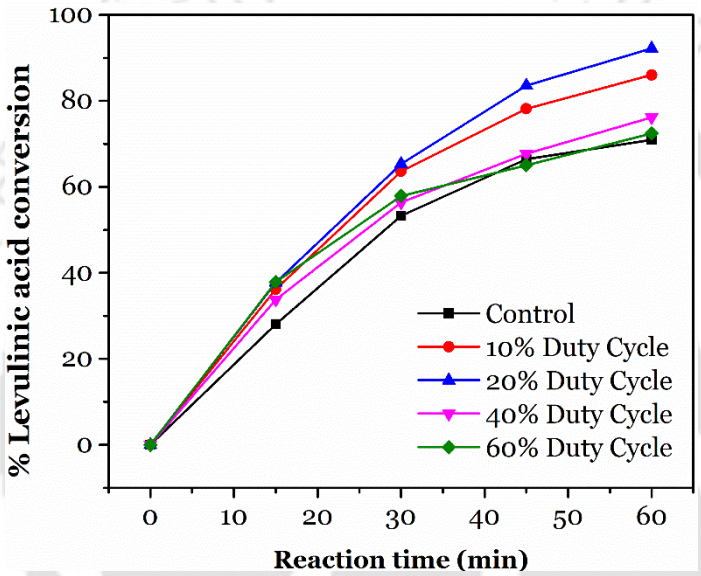


Figure 7.3 Time profiles of Levulinic acid conversion (under different duty cycles). Reaction conditions: molar ratio (butanol:levulinic acid) = 3.74:1, tertra-butyl methyl ether upto 15 mL, speed of agitation = 633 rpm, enzyme amount = 86 mg, temperature= 50 ℃

7.3.2. Kinetic Analysis of Esterification

7.3.2.1. Reaction Kinetics with Various Ping-Pong Bi-Bi Model

As noted earlier, the time profiles of substrates and products were fitted to four kinetic models based on Ping-Pong Bi-Bi mechanism described in section 7.2.4.2.

Prior to discussing the results, we would briefly ponder over the molecular progression or pathway of Ping-Pong Bi-Bi mechanism in the present context. The LA

(*A*, first substrate) enters the active site of the immobilized CALB enzyme and gets coordinated to the oxyanion hole. In this process, the hydrogen atom of Ser105 is transferred to the Nε of His224 of the catalytic triad, and the Oγ of Ser105 attacks as a nucleophile on the carbonyl carbon of the LA. This results in the formation of the first tetrahedral complex or enzyme-substrate complex (*EA*). This transition state (shown as TS-I in **Figure 7.1B**) is stabilized by the series of hydrogen bonds by oxyanion hole (Thr40 and Gln157). Next, the hydrogen atom from His224 is transferred to the oxygen atom of the transition state and a water (*P*, first product) molecule is released from the complex, which results in acyl-enzyme complex (*E** or TI as shown in **Figure 7.1B**). The second nucleophilic attack on the carbonyl carbon of the *E** is initiated by butanol (*B*, second substrate) present in the alcohol binding pocket to form the second tetrahedral complex or acyl-enzyme-substrate complex (*E*B*, or TS-II) as shown in **Figure 7.1B**. In the final step, a similar rearrangement in terms of hydrogen transfer and chemical bond rearrangements occurs that results in generation of n-BL (*Q*, second product) and regeneration of enzyme (*E*). Asp187 in the catalytic triad plays a vital role in the stabilization of positive charge generated over the His224 residue in the tetrahedral intermediates.

The results of kinetic analysis are given in **Table 7.2**, which lists the values of different kinetic parameters. For obtaining the kinetic parameters, initially the conversion profile was discretized by calculating the reaction velocity between two consecutive sampling points. Assuming that this reaction velocity as initial reaction velocity (v_0) corresponding to higher (or starting) substrate concentration (S_0) in each discrete segment, set of pairs of (v_0 - S_0) values were obtained. The kinetic expression given in **equations 7.3**, **7.4**, **7.5**, and **7.6** were fitted to this (v_0 - S_0) data to obtain the kinetic parameters in expressions **7.3**, **7.4**, **7.5**, and **7.6**. While fitting the kinetic

expressions, the total error for all kinetic parameters was minimized using Generalised Reduced Gradient (GRG) nonlinear algorithm [37]. One the basis of minimum RMS error in both reaction categories viz. control and test, the best fitting model seems to be mixed dead-end and product inhibition with water.

Table 7.2: Enzyme kinetic parameters of esterification reaction

Type of rate equation	Parameters	Control	RMS Error	Test	RMS Error
In absence of products	V_{max}	1.87×10^{-2}	1.98×10^{-3}	2.47×10^{-2}	1.99×10^{-4}
	K_A	3.16×10^{-1}		2.56×10^{-1}	
	K_B	3.42×10^{-1}		3.78×10^{-1}	
Inhibition by B	V_{max}	2.14×10^{-2}	2.03×10^{-3}	2.76×10^{-2}	2.09×10^{-4}
	K_A	2.88×10^{-1}		2.10×10^{-1}	
	K_B	3.44×10^{-1}		5.69×10^{-1}	
	K_i	2.85		3.55	
Product inhibition by P	V_{max}	1.86×10^{-2}	1.64×10^{-3}	2.47×10^{-2}	2.01×10^{-4}
	K_A	2.38×10^{-1}		2.08×10^{-1}	
	K_B	4.70×10^{-1}		5.14×10^{-1}	
	K_{iP}	9.76×10^{-1}		2.58	
	K_{iA}	6.46×10^{-1}		5.66×10^{-1}	
Mixed dead-end product inhibition	V_{max}	1.96×10^{-2}	1.48×10^{-3}	2.71×10^{-2}	1.95×10^{-4}
	K_A	1.43×10^{-1}		1.08×10^{-2}	
	K_B	4.26×10^{-1}		1.59	
	K_{iP}	3.10		4.02	
	K_{iA}	3.17×10^{-1}		3.15×10^{-1}	
	$K_{i'}$	1.03×10^{-1}		1.26×10^{-1}	

Units: Kinetic constants K_A, K_B, K_{iA}, K_{iP}, K_i, *and* $K_{i'}$ = mol/L; V_{max} (or V_1) = mol/L·min

The relative variations in different kinetic parameters and their physical interpretation in the control and test experiments are as follows:

V_{max} shows a 1.38-fold increase with sonication (test experiments) as compared to

mechanical agitation (control experiments), which shows that sonication helps in increasing the maximum reaction velocity thereby making esterification faster. Michaelis-Menten constants for the two substrates show a very interesting trend. K_A decreases and K_B increases in test experiments, which signifies enhanced affinity of LA and reduced affinity of butanol towards enzyme. Moreover, relatively smaller K_{iA} values in test experiments show stronger binding of LA molecules with immobilised enzyme, or in other words higher stability of *EA* complex (Novozym 435 and LA). Increase in K_{iP} in the test experiments shows faster release of water (first product, *P*) through partial dissociation of *E*P* complex to generate the enzyme-acyl complex (*E**) followed by nucleophilic attack of butanol. The marginal increase in $K_{i'}$ shows reduction in inhibition due to water, which also concurs with rise in K_{iP}, as noted earlier.

Higher Novozym 435-LA affinity, faster splitting of *E*P* complex and faster reaction velocity in the test experiments are all consequences of morphological changes in the secondary structure of enzymes induced by micro-convection generated by sonication, as discussed in greater details in subsequent section.

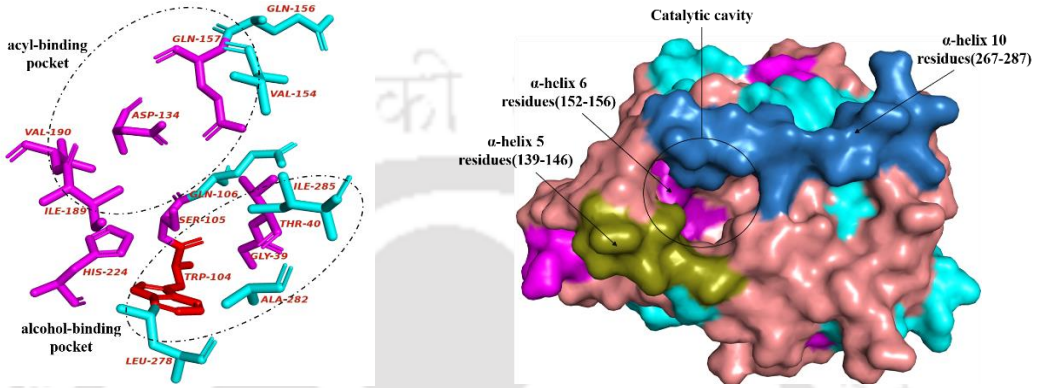
7.3.3. Discernment of Mechanism of Sonication-Induced Enhancement

Prior to exploring the mechanism of ultrasound induced enhancement in kinetics of esterification, we first ponder over the secondary structure of the CALB enzyme itself, and the molecular mechanism of esterification as deduced from molecular docking simulations.

7.3.4. Structure of CALB Enzyme

As mentioned earlier, crystal structure of CALB (1TCA) is of α/β type fold and comprises of 317 amino acid residues [25, 38]. Catalytic triad Ser105-His224-Asp187 residues are responsible for enzymatic activity in CALB [9, 25, 38]. The substrate binding pocket (catalytic cavity) in CALB is an elliptical steep funnel, which has long hydrophilic bottom formed by Asp134 and Ser105. This funnel is lined with Thr138, Ile189, and Val190 residues on left-hand side while Gln157 and Thr40 residues on the right-hand side [27]. CALB has very limited available space around its catalytic cavity, which includes two channels (as shown in **Figure 7.4A**), the acyl-binding channel and alcohol-binding channel (where n-butanol binds) [25]. The acyl-binding channel in CALB is more spacious than alcohol-binding channel. These two channels run almost parallel from the active site Ser105 to the surface and are just barely separated by two hydrophobic side chains, Ile189 and Ile285. The residue range 139-146, 152-156, and 267-287, corresponds to α-helix 5, α-helix 6 and α-helix 10, respectively (refer to **Figure 7.4B**) [9]. In 3-D form of CALB, the secondary structural motifs, α-helix 5, α-helix 6 and α-helix 10 are folded to create catalytic cavity (**Figure 7.4B**). Although previous reports [2, 25, 38] deny the presence of any flapping lid in CALB; a recent study [9] conducted at atomic level reports that α-helix 5 and α-helix 10 act as flapping lid. In the open conformation of CALB, amino acid residues present in the α-helix 5 are seen as series of aliphatic residues that line the channel to active site. In α-helix 5, Asp145 is the only polar amino acid residue, which makes hydrogen bonds with the side chains of Ser150 and Thr158 [9]. While in the closed conformation α-helix 5 goes under drastic conformational changes to form unstructured loop, which results in formation of salt bridge between Asp145 and Lys290 (present in the side chain of α-helix 10). This spatial rearmament of Lys290 brings α-helix 10 closer to

new lid region thereby closing the catalytic cavity completely. Several molecular dynamics simulations studies [39–41] have reported that α-helix 5 is a highly dynamic region and work as flapping lid depending on the working temperature [41] or the working organic solvents [39].



(A) **(B)**

Figure 7.4: (A) active site residues present in the acyl and alcohol-binding channels in catalytic cavity; (B) crystal structure of CALB visualized in PyMOL (version 2.4.2)

7.3.5. Analysis of Molecular Docking Simulations:

The complete molecular mechanism of LA esterification is given in **Appendix 5**. Substrates, products and tetrahedral intermediates were covalently docked into CALB. During docking the protein structure was assumed to be rigid, while the docked structure/ligands were treated flexible. The docking procedure comprised of three steps: (i) the construction of the putative ligands (substrates, products, and tetrahedral intermediates), (ii) the covalent docking into the active site, and (iii) the application of geometric filter criteria for docked ligands poses. Molecular docking analysis carried out by AutoDock 4.2.1 software revealed highest binding affinity with LA (binding energy = -5.34 kcal/mol) followed by transition state (TSI or TSII, binding energy = -5.08 kcal/mol) as can be seen in **Table 7.3A**. The binding affinity with n-butanol was

lowest (-2.47 kcal/mol), and TI was slightly higher than n-butanol. This showed that CALB has higher affinity for acylated ligands than alcohol moiety. LA binding energy is higher than that of butanol which suggest that the *EA* complex is relatively stable. 3D molecular interaction between the active site residue and open catalytic cleft of CALB and the ligands were visualized in PyMOL software (**Figure 7.5**).

Table 7.3: (A) Characteristic features of molecular interactions of amino acid residues with putative ligands

Ligands	Binding energy (kcal/mol)	Polar interaction	Hydrophobic interaction
Levulinic acid	-5.34	Ser105, His224, Thr40	Gly39, Asp134, Gln157, Ile189, Val190
Butanol	-2.47	Ser105	Asp134, Thr138, Gln157, Ile189, Val190, His224
n-Butyl levulinate	-4.39		Thr40, His224, Leu278, Ala282, Ile285, Val286
TSI or TSII	-5.08	Thr40, Ser105, His224	Gly39, Trp104, Asp134, Thr138, Val154, Gln157, Ile189, Val190
TI	-2.90	Thr40, Ser105, His224	Ile189

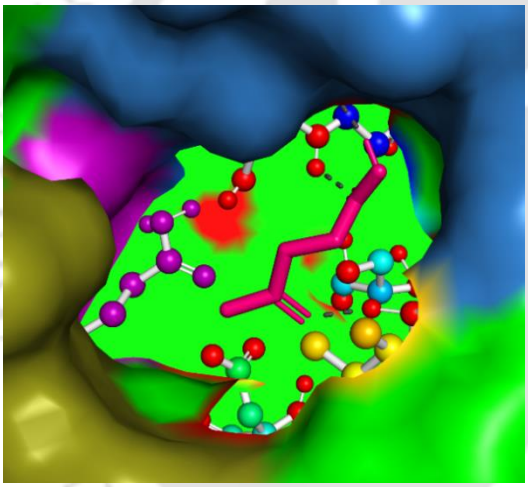
The 2D interaction of the ligands with amino acid residues in the catalytic cleft was studied from the depiction obtained from LigPlot^{+} program. Amino acid residues which interact with substrates, products and transitions states through polar or hydrophobic interaction for the formation of complex are identified in **Table 7.3A**. The spatial location of these residues is as follows:

Random coil: Gly39, Thr40, Ser105, Asp134, Thr138, Gln157, Asp187, Ile189, Val190, His224, and Leu290

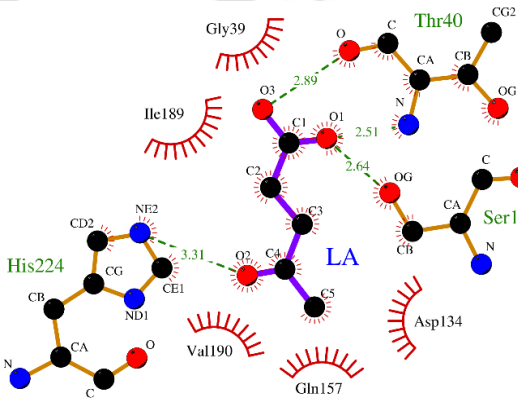
α-helix: Gln106, Asp145, Leu278, Ala282, Ile285, Val286,

β-strand: Trp104

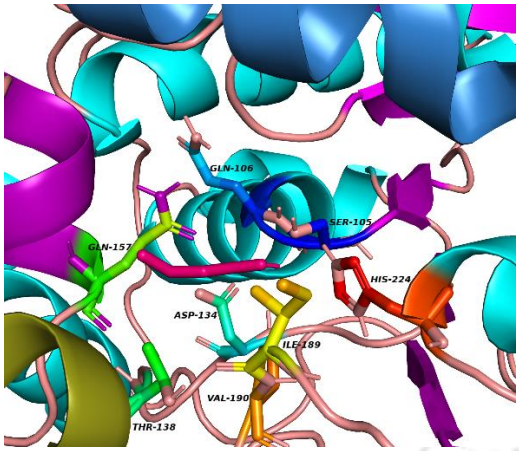
It could be seen that most of the amino acid residues involved in complex formation through either polar or hydrophobic interaction are located in random coils. This implies that random coils influence the formation and stability of enzyme-ligand complexes.



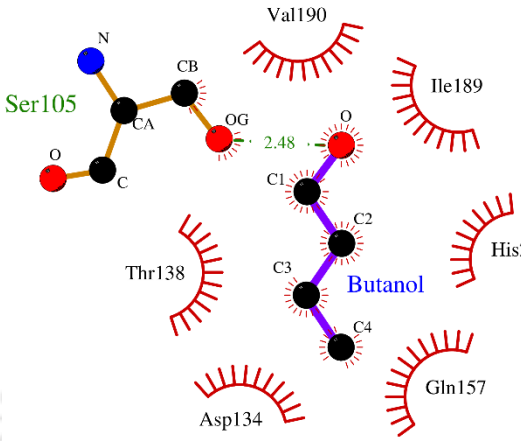
(A)



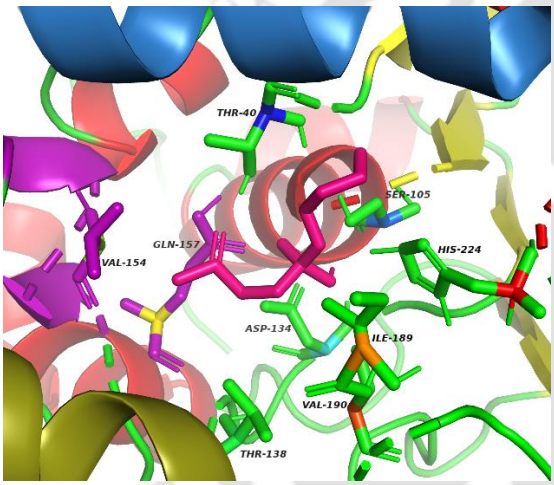
(B)



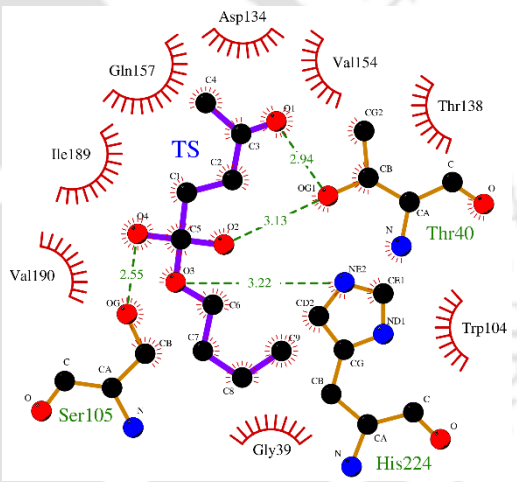
(C)



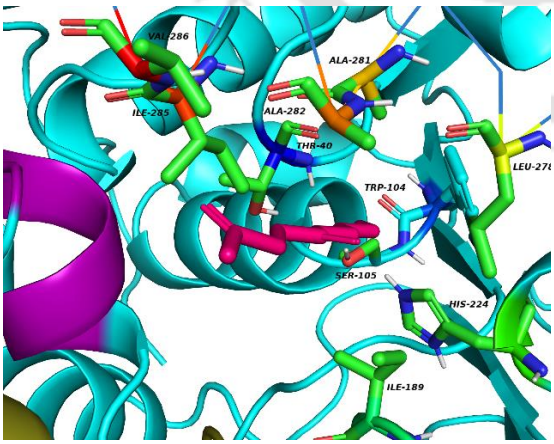
(D)



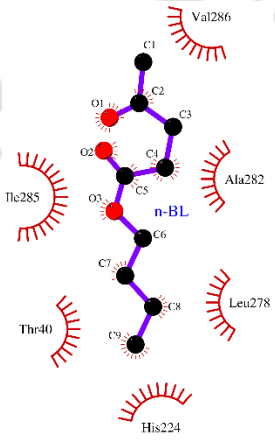
(E)



(F)



(G)



(H)

Figure 7.5: 3D and 2D visualization (schematic representations) of interaction between active site residues of CALB and ligands (LA, n-butanol, transition state (TS) and n-BL). (A & B) EA complex in acyl-binding channel; (C & D) butanol in alcohol-binding channel; (E & F) transition state (*E*Q* complex) in acyl-binding channel; (G & H) interactions between active site residues of CALB and n-BL.
(Colour codes: dashed green lines = hydrogen bond, combed red arcs = hydrophobic interactions, black balls = carbon atoms, blue balls = nitrogen atoms, red balls = oxygen atoms, purple lines = ligand bonds, light brown lines = amino acid residue bonds)

7.3.6. Influence of Sonication

The influence of sonication on secondary structure of Novozym 435 (which is the immobilized form of CALB on microporous polyacrylic resin particles) was analyzed using ATR-FTIR spectroscopy. Peptide repeats present in polypeptide chain give a total of nine characteristic infrared absorption bands, namely, amide A, B, and I – VII. Among these, the amide I and II bands are the most prominent vibrational bands in the polypeptide backbone [42–44]. The most sensitive spectral region, amide I band (1600-1690 cm^{-1}) [45], which is due to C=O stretch vibration of the peptide linkage, were enumerated with multi-peak fitting Gaussian function. The FTIR spectra (**Figure 7.6A**) and its secondary derivatives (**Figures 6B** and **C**) clearly revealed considerable changes in secondary structure of the enzyme. Relative fractions of secondary structural motifs of Novozym 435 in control and test experiments are illustrated in Figure 7.6B and C and **Table 7.3B**. For deconvolution of FTIR spectra, the spectral regions for different fractions of secondary structure of Novozym 435 were: β-strands (1627 ± 2.0 cm^{-1}), β-sheets (1638 ± 2.0 cm^{-1}), random coils (1648 ± 2.0 cm^{-1}), α-helix (1656 ± 2.0 cm^{-1}) and β-turns (1664 - 1690 cm^{-1}) [45].

Table 7.3: (B) Alterations in secondary structure composition of CALB enzyme induced by sonication

Experimental category*	β-Strands	β-Sheets	Random coil	α-Helix	β-Turns
Control	7.56	22.09	28.98	23.07	18.29
Test	9.70	14.99	36.33	31.81	7.11

* control (reaction with mechanical agitation) and test (reaction with mechanical agitation and sonication at 20% duty cycle)

As evident from the results depicted in **Table 7.3B**, sonication induces marked changes in the secondary structure of the enzyme. The major change is rise in α-helix (from 23.07% to 31.8%) and random coil (from 28.98% to 36.38%) content with concurrent reduction in β-sheets and β-turns. These changes contribute to rise in activity of the enzyme.

As noted in section 7.3.3.1 (the secondary structure of CALB enzyme), the α-helix content essentially governs the accessibility of catalytic cavity by the substrate through flapping lid. Reduction (or denaturation) of α-helix content results in the closure of the catalytic cavity due to formation of a salt bridge between Lys290 in the side chain of α-helix 10 and the carboxylic group of Asp145 in unstructured loop formed due to denaturing of α-helix 5 (refer to **Figure 7.4B**). On the other hand, rise in α-helix content on both side of the catalytic cavity helps in structural stability of the catalytic cavity, in addition to widening of the "mouth" (or entry region) of the cavity. This helps in greater accessibility of the substrate to the active site of the enzyme, and thus rise in the esterification activity of the enzyme.

Rise in random coil content essentially signifying unfolding of enzyme proteins with exposure of inner buried residues to bulk solution. As noted in the previous section, the most amino acid residues involved in complex formation are located in random coil structure. Increase in random coil content of enzyme assists not only

easier complex formation, but as a result stronger interaction with the amino acid residues in random coil, these complexes are stable. The reverse split up of these complexes is also reduced due to stronger interaction with amino acid residues. This feature is also corroborated by the trends in values of parameters in Ping-Pong Bi-Bi kinetic model, viz. reduction in K_{iA} and increase in K_{iP}.

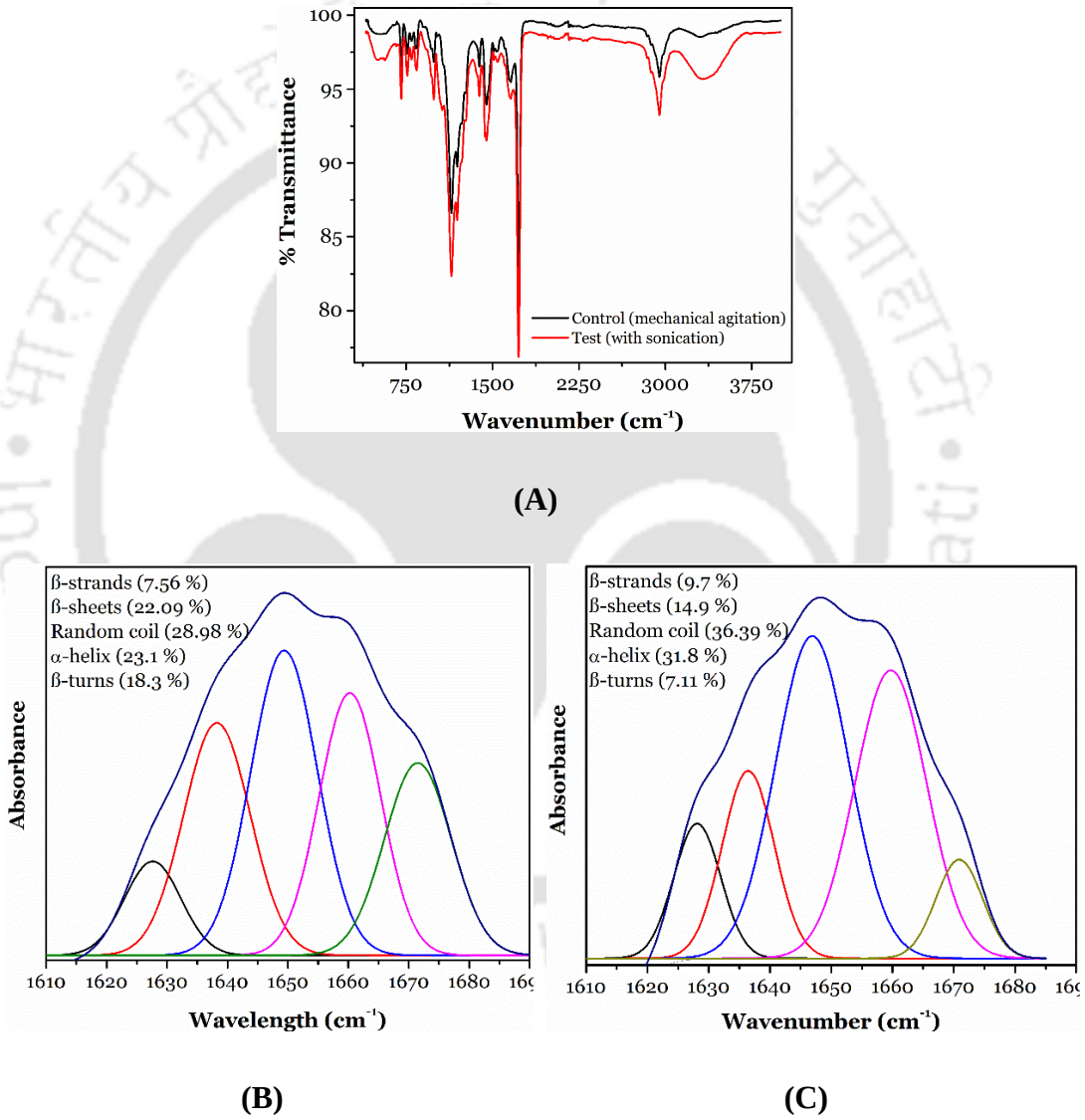


Figure 7.6: Determination of effect of sonication on secondary structure of Novozym 435. (A) Full range ATR-FTIR spectrum of Novozym 435 in control (reaction with mechanical agitation) and test (reaction with mechanical agitation and sonication at 20% duty cycle) experiments; (B) multicomponent peak area fitting in

amide I (1600-1690 cm^{-1}) spectral region (control experiments); (C) multicomponent peak area fitting in amide I (1600-1690 cm^{-1}) spectral region (test experiments). Color codes of deconvolution curves in **Figures B** & **C**: black = β-strands; red = β-sheets; blue = random coil; pink = α-helix; green/olive green = β-turns.

7.3.7. Reusability of Novozym 435

The immobilized CALB enzyme was separated from reaction mixture and reused after each experiment, as described in **Section 7.2.4.4**. The LA conversion profiles for the fresh and recycled enzyme are shown in **Figure 7.7**. It could be seen that the immobilized enzyme retains good activity after recycle, although successively the extent of conversion reduces. The immobilized enzyme could be used up to 4 successive cycles of separation and reuse. The fresh enzyme yielded 92.2% conversion, which declined to 52.9% till 4th cycle.

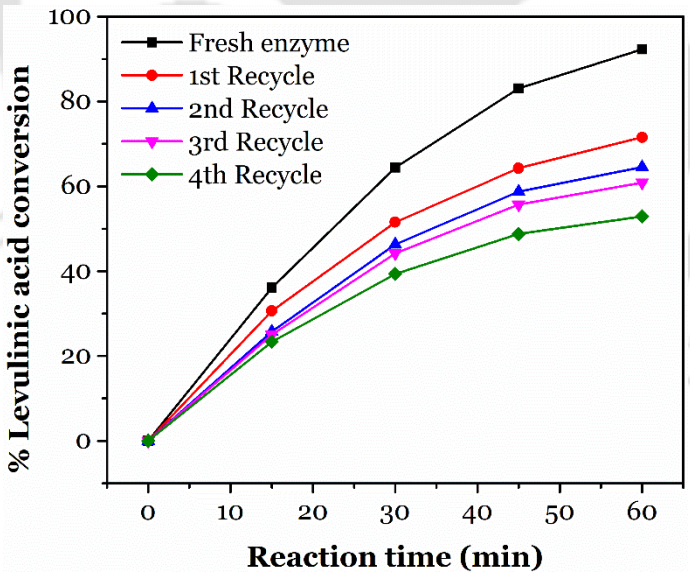


Figure 7.7: Time profiles of reusability of Novozym 435 at 20% sonication duty cycle. Reaction conditions: molar ratio (butanol:levulinic acid) = 3.74:1, tertra-butyl methyl ether upto 15 mL, speed of agitation = 633 rpm, enzyme amount = 86 mg, temperature= 50 °C

7.4. CONCLUSIONS

The present study has given mechanistic accounts of ultrasound-assisted n-butyl levulinate synthesis using Novozym 435. Our analysis based on molecular docking simulations has demonstrated that sonication enhanced the catalytic activity of Novozym 435 through modifications in secondary structure of CALB. Kinetic analysis revealed significant rise in reaction velocity and substrate affinity. In addition, inhibition effects of alcohol and product also reduced in presence of ultrasound. Reverse dissociation of intermediate complexes also reduced in presence of sonication. Molecular docking analysis shedded light on role of residues Ser105, His224, Asp187, Thr40, and Gln157 in catalyzing LA esterification. The α-helix content of enzyme increased with exposure to sonication, which imparted structural stability to catalytic cavity, in addition to widening of the entry region of the cavity. Sonication also caused rise in relative content of random coil in enzyme which essentially signified unfolding of enzyme with exposure of inner buried residues to bulk solution. The macroscopic manifestation of these effects was intensification of the esterification reaction with faster kinetics and higher yield of n-BL.

REFERENCES

1. Yang J, Li G, Zhang L, Zhang S (2018) Efficient Production of N-Butyl Levulinate Fuel Additive from Levulinic Acid Using Amorphous Carbon Enriched with Oxygenated Groups. Catalysts 8:14. https://doi.org/10.3390/catal8010014
2. Yadav GD, Borkar IV (2008) Kinetic Modeling of Immobilized Lipase Catalysis in Synthesis of *n* -Butyl Levulinate †. Ind Eng Chem Res 47:3358–3363. https://doi.org/10.1021/ie800193f
3. Bozell JJ, Petersen GR (2010) Technology development for the production of biobased products from biorefinery carbohydrates—the US Department of Energy's "Top 10" revisited. Green Chem 12:539–554. https://doi.org/10.1039/B922014C

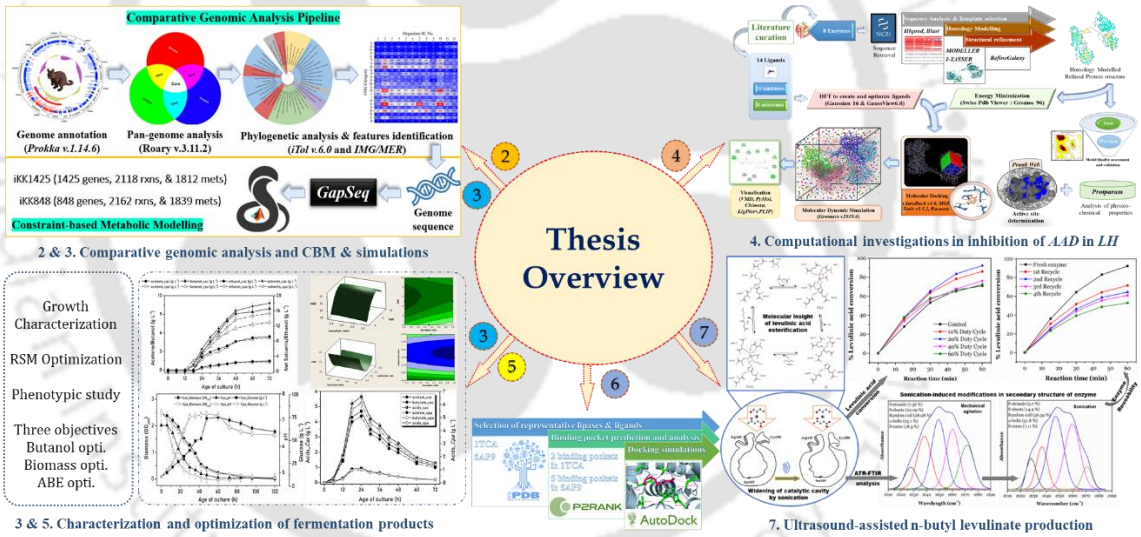
4. Antonetti C, Licursi D, Fulignati S, et al (2016) New Frontiers in the Catalytic Synthesis of Levulinic Acid: From Sugars to Raw and Waste Biomass as Starting Feedstock. Catalysts 6:196. https://doi.org/10.3390/catal6120196
5. Corma A, Iborra S, Velty A (2007) Chemical Routes for the Transformation of Biomass into Chemicals. Chem Rev 107:2411–2502. https://doi.org/10.1021/cr050989d
6. Qin Y-Z, Zong M-H, Lou W-Y, Li N (2016) Biocatalytic Upgrading of 5-Hydroxymethylfurfural (HMF) with Levulinic Acid to HMF Levulinate in Biomass-Derived Solvents. ACS Sustain Chem Eng 4:4050–4054. https://doi.org/10.1021/acssuschemeng.6b00996
7. Christensen E, Yanowitz J, Ratcliff M, McCormick RL (2011) Renewable Oxygenate Blending Effects on Gasoline Properties. Energy Fuels 25:4723–4733. https://doi.org/10.1021/ef2010089
8. Shrivastav G, Khan TS, Agarwal M, Haider MA (2017) Reformulation of Gasoline To Replace Aromatics by Biomass-Derived Alkyl Levulinates. ACS Sustain Chem Eng 5:7118–7127. https://doi.org/10.1021/acssuschemeng.7b01316
9. Stauch B, Fisher SJ, Cianci M (2015) Open and closed states of *Candida antarctica* lipase B: protonation and the mechanism of interfacial activation. J Lipid Res 56:2348–2358. https://doi.org/10.1194/jlr.M063388
10. Lu C, Peng X, Lu D, Liu Z (2020) Global and Kinetic Profiles of Substrate Diffusion in *Candida antarctica* Lipase B: Molecular Dynamics with the Markov-State Model. ACS Omega 5:9806–9812. https://doi.org/10.1021/acsomega.9b04432
11. Sadeghi Googheri MS, Housaindokht MR, Sabzyan H (2015) Theoretical studies on the deacylation step of acylated *Candida antarctica* lipase B: structural and reaction pathway analysis. J Mol Graph Model 57:9–19. https://doi.org/10.1016/j.jmgm.2015.01.003
12. Toledo MV, Llerena Suster CR, Ferreira ML, et al (2017) Molecular recognition of an acyl–enzyme intermediate on the lipase B from *Candida antarctica*. Catal Sci Technol 7:1953–1964. https://doi.org/10.1039/C7CY00245A
13. Sadeghi Googheri MS, Housaindokht MR, Sabzyan H (2014) Reaction mechanism and free energy profile for acylation of *Candida Antarctica* lipase B with methylcaprylate and acetylcholine: Density functional theory calculations. J Mol Graph Model 54:131–140. https://doi.org/10.1016/j.jmgm.2014.10.001
14. Galgali A, Gawas SD, Rathod VK (2018) Ultrasound assisted synthesis of citronellol laurate by using Novozym 435. Catal Today 309:133–139. https://doi.org/10.1016/j.cattod.2017.08.052
15. Shinde SD, Yadav GD (2015) Insight into Microwave-Assisted Lipase Catalyzed Synthesis of Geranyl Cinnamate: Optimization and Kinetic Modeling. Appl Biochem Biotechnol 175:2035–2049. https://doi.org/10.1007/s12010-014-1367-3
16. Habulin M, Šabeder S, Sampedro MA, Knez Ž (2008) Enzymatic synthesis of citronellol laurate in organic media and in supercritical carbon dioxide. Biochem Eng J 42:6–12. https://doi.org/10.1016/j.bej.2008.05.012
17. Matsuda T, Watanabe K, Harada T, Nakamura K (2004) Enzymatic reactions in supercritical CO_2: carboxylation, asymmetric reduction and esterification. Catal Today 96:103–111. https://doi.org/10.1016/j.cattod.2004.06.111

18. Wang J, Wang S, Li Z, et al (2015) Ultrasound irradiation accelerates the lipase-catalyzed synthesis of methyl caffeate in an ionic liquid. J Mol Catal B Enzym 111:21–28. https://doi.org/10.1016/j.molcatb.2014.11.006
19. Jia R, Hu Y, Liu L, et al (2013) Chemical modification for improving activity and stability of lipase B from *Candida antarctica* with imidazolium-functional ionic liquids. Org Biomol Chem 11:7192. https://doi.org/10.1039/c3ob41076e
20. Malani RS, Umriwad SB, Kumar K, et al (2019) Ultrasound–assisted enzymatic biodiesel production using blended feedstock of non–edible oils: Kinetic analysis. Energy Convers Manag 188:142–150. https://doi.org/10.1016/j.enconman.2019.03.052
21. Bhangu SK, Gupta S, Ashokkumar M (2017) Ultrasonic enhancement of lipase-catalysed transesterification for biodiesel synthesis. Ultrason Sonochem 34:305–309. https://doi.org/10.1016/j.ultsonch.2016.06.005
22. Malani RS, Khanna S, Chakma S, Moholkar VS (2014) Mechanistic insight into sono-enzymatic degradation of organic pollutants with kinetic and thermodynamic analysis. Ultrason Sonochem 21:1400–1406. https://doi.org/10.1016/j.ultsonch.2014.01.028
23. Malani RS, Khanna S, Moholkar VS (2013) Sonoenzymatic decolourization of an azo dye employing immobilized horse radish peroxidase (HRP): A mechanistic study. J Hazard Mater 256–257:90–97. https://doi.org/10.1016/j.jhazmat.2013.04.023
24. Bhavsar KV, Yadav GD (2018) n-Butyl levulinate synthesis using lipase catalysis: comparison of batch reactor versus continuous flow packed bed tubular microreactor. J Flow Chem 8:97–105. https://doi.org/10.1007/s41981-018-0014-5
25. Uppenberg J, Oehrner N, Norin M, et al (1995) Crystallographic and molecular-modeling studies of lipase B from *Candida antarctica* reveal a stereospecificity pocket for secondary alcohols. Biochemistry 34:16838–16851. https://doi.org/10.1021/bi00051a035
26. Martinelle M, Hult K (1995) Kinetics of acyl transfer reactions in organic media catalysed by *Candida antarctica* lipase B. Biochim Biophys Acta BBA - Protein Struct Mol Enzymol 1251:191–197. https://doi.org/10.1016/0167-4838(95)00096-D
27. Foresti ML, Ferreira ML (2004) Computational Approach to Solvent-Free Synthesis of Ethyl Oleate Using *Candida rugosa* and *Candida antarctica B* Lipases. I. Interfacial Activation and Substrate (Ethanol, Oleic Acid) Adsorption. Biomacromolecules 5:2366–2375. https://doi.org/10.1021/bm049688o
28. Ortiz C, Ferreira ML, Barbosa O, et al (2019) Novozym 435: the “perfect” lipase immobilized biocatalyst? Catal Sci Technol 9:2380–2420. https://doi.org/10.1039/C9CY00415G
29. Li C, Tan T, Zhang H, Feng W (2010) Analysis of the conformational stability and activity of Candida antarctica lipase B in organic solvents: insight from molecular dynamics and quantum mechanics/simulations. J Biol Chem 285:28434–28441. https://doi.org/10.1074/jbc.M110.136200
30. Kokare MB, V R, Mathpati CS (2018) Response surface optimization, kinetic study and process design of n-butyl levulinate synthesis. Chem Eng Res Des 137:577–588. https://doi.org/10.1016/j.cherd.2018.07.036
31. Zhou L, He Y, Ma L, et al (2018) Conversion of levulinic acid into alkyl levulinates: Using lipase immobilized on meso-molding three-dimensional

macroporous organosilica as catalyst. Bioresour Technol 247:568–575. https://doi.org/10.1016/j.biortech.2017.08.134
32. Luo J, Ying K, Bai J (2005) Savitzky–Golay smoothing and differentiation filter for even number data. Signal Process 85:1429–1434. https://doi.org/10.1016/j.sigpro.2005.02.002
33. Muley AB, Chaudhari SA, Bankar SB, Singhal RS (2019) Stabilization of cutinase by covalent attachment on magnetic nanoparticles and improvement of its catalytic activity by ultrasonication. Ultrason Sonochem 55:174–185. https://doi.org/10.1016/j.ultsonch.2019.02.019
34. Morris GM, Huey R, Lindstrom W, et al (2009) AutoDock4 and AutoDockTools4: Automated Docking with Selective Receptor Flexibility. J Comput Chem 30:2785–2791. https://doi.org/10.1002/jcc.21256
35. Gaussian 09 Citation | Gaussian.com. https://gaussian.com/g09citation/. Accessed 17 Apr 2021
36. Laskowski RA, Swindells MB (2011) LigPlot+: Multiple Ligand–Protein Interaction Diagrams for Drug Discovery. J Chem Inf Model 51:2778–2786. https://doi.org/10.1021/ci200227u
37. Lasdon L, Fox R, Ratner MW (1974) Nonlinear optimization using the generalized reduced gradient method
38. Uppenberg J, Hansen MT, Patkar S, Jones TA (1994) The sequence, crystal structure determination and refinement of two crystal forms of lipase B from *Candida antarctica*. Structure 2:293–308. https://doi.org/10.1016/S0969-2126(00)00031-9
39. Trodler P, Pleiss J (2008) Modeling structure and flexibility of *Candida antarctica* lipase B in organic solvents. BMC Struct Biol 8:9. https://doi.org/10.1186/1472-6807-8-9
40. Skjøt M, Maria LD, Chatterjee R, et al (2009) Understanding the Plasticity of the α/β Hydrolase Fold: Lid Swapping on the *Candida antarctica* Lipase B Results in Chimeras with Interesting Biocatalytic Properties. ChemBioChem 10:520–527. https://doi.org/10.1002/cbic.200800668
41. Ganjalikhany MR, Ranjbar B, Taghavi AH, Tohidi Moghadam T (2012) Functional Motions of *Candida antarctica* Lipase B: A Survey through Open-Close Conformations. PLoS ONE 7:e40327. https://doi.org/10.1371/journal.pone.0040327
42. Krimm S, Bandekar J (1986) Vibrational Spectroscopy and Conformation of Peptides, Polypeptides, and Proteins. In: Anfinsen CB, Edsall JT, Richards FM (eds) Advances in Protein Chemistry. Academic Press, pp 181–364
43. Susi H, Byler DM (1986) [13] Resolution-enhanced fourier transform infrared spectroscopy of enzymes. In: Methods in Enzymology. Academic Press, pp 290–311
44. Surewicz WK, Mantsch HH (1988) New insight into protein secondary structure from resolution-enhanced infrared spectra. Biochim Biophys Acta BBA - Protein Struct Mol Enzymol 952:115–130. https://doi.org/10.1016/0167-4838(88)90107-0
45. Kong J, Yu S (2007) Fourier Transform Infrared Spectroscopic Analysis of Protein Secondary Structures. Acta Biochim Biophys Sin 39:549–559. https://doi.org/10.1111/j.1745-7270.2007.00320.x

CHAPTER 8

Overview & Future Directions



The thesis overview can also be found on my personal webpage:
https://sites.google.com/view/karaniitg/research

CHAPTER 8

OVERVIEW & FUTURE DIRECTIONS

In this final chapter of the thesis, we provide an overview of the key findings and contributions of this dissertation and offer suggestions for future research directions in bioalcohol synthesis through *Clostridia*. The chapter concludes with a summary of the significant implications of the research and its potential impact on sustainable bioenergy production and bio-based industries.

8.1. OVERVIEW OF KEY FINDINGS

Throughout this thesis, we addressed the research problem of enhancing bioalcohol synthesis through *Clostridia* by adopting a multidisciplinary approach that combined experimental investigations and computational analyses. The main objectives of the study and the major results and findings against each of them are summarized below.

8.1.1. Comparative Genomic Analysis

In this study, we conducted a comprehensive comparative genomic analysis of potential butanol-producing *Clostridium* species. Through the examination of genomic sequences from diverse strains, we aimed to identify the best solventogenic and cellulolytic candidates for bioalcohol production. We collected genomic data from various *Clostridium* strains and conducted a thorough comparative genomic analysis to assess their bioalcohol-producing potential. Utilizing statistical analysis and phylogenetic assessments, we evaluated the accuracy of strain selection, achieving an impressive 78% accuracy in identifying the most promising strains. To

ensure data integrity, we cross-validated our findings using multiple bioinformatics tools. Our observations revealed distinct metabolic capabilities among the selected *Clostridial* strains, shedding light on the genetic basis of bioalcohol production. We identified key genetic determinants responsible for efficient bioalcohol synthesis and rectified misclassified species, thereby enhancing the accuracy of genomic databases. The robust evidence and data-driven insights from this analysis provide a solid foundation for targeted genetic engineering efforts to optimize bioalcohol yields and promote sustainable biofuel production.

8.1.2. Investigation of Cell Central Metabolism

The investigation of cell central metabolism in *C. acetobutylicum* ATCC 824 and *C. pasteurianum* ATCC 6013 during ABE fermentation has provided valuable insights that contribute to the optimization of biobutanol production. Combining biochemical and systems biology techniques allowed us to uncover distinctive physiological differences between these two *Clostridial* species, shedding light on their respective capabilities for bioalcohol synthesis. Through comprehensive biochemical characterization, we gained a detailed understanding of their metabolizing capabilities, growth cycles, and fermentation profiles, which provided a basis for comparative analysis. Applying systems biology techniques further deepened our understanding by elucidating the genetic basis and regulatory mechanisms underlying their metabolic pathways during ABE fermentation. The data obtained from these investigations offer crucial knowledge for enhancing biobutanol production processes and for the targeted engineering of these microorganisms to achieve higher yields of bioalcohols.

The knowledge gained from this study holds significant implications for advancing sustainable bioenergy technologies. By discerning the distinct physiological traits of *C. acetobutylicum* and *C. pasteurianum*, this research sets the stage for informed decision-making in strain selection and metabolic engineering strategies. The findings advance our understanding of *Clostridial* metabolism and pave the way for developing more efficient and environmentally friendly bioalcohol production processes.

8.1.3. Challenges and Scope of Butanol Production

In this thesis, we investigated the challenges and opportunities surrounding butanol production from lignocellulosic hydrolysates using *Clostridial* co-culture systems. Our research encompassed two key aspects: an in-depth inhibitor study and an exploration of *Clostridial* co-culture systems.

(1) In the inhibitor study, we meticulously analyzed the inhibitory compounds present in lignocellulosic hydrolysates, as they significantly hinder the efficiency of butanol fermentation. Through advanced molecular mechanics simulations, we successfully identified crucial inhibitory compounds, including *ρ*-coumaric acid (*ρ*-CA), vanillic acid (VA), and cinnamaldehyde (CMD). These compounds substantially impact the catalytic activity of key aldehyde/alcohol dehydrogenase (AAD) enzymes, essential for butanol synthesis. Moreover, our investigation shed light on the precise molecular interactions between these inhibitory compounds and the AAD enzymes, allowing us to better understand the inhibition mechanism. Armed with this knowledge, we are better equipped to develop targeted strategies to alleviate inhibitory effects and enhance the overall efficiency of butanol production from lignocellulosic hydrolysates. By

addressing this critical challenge, we have laid the groundwork for more effective and economically viable biobutanol production processes.

(2) Simultaneously, we focused on the statistical optimization of *Clostridial* co-culture systems to enhance bio-butanol production from mixed substrates. Through response surface methodology (RSM), we systematically analyzed crucial process parameters, including the ratio of *C. acetobutylicum* and *C. pasteurianum* inoculum, NaCl concentration, and the ratio of xylose and glucose. The RSM optimization significantly increased butanol production, with a yield of 12.1 ± 0.45 g/L, which closely aligned with the model prediction of 11.87 g/L. Additionally, the biomass yield reached 4.15 ± 0.03, matching the model prediction of 4.06 OD_{600}, and the total alcohol production was measured at 23.1 ± 0.55 g/L, close to the model prediction of 22.45 g/L.

Integrating our inhibitor study and the quantitative results obtained from RSM optimization of *Clostridial* co-cultures forms a comprehensive approach to address the challenges and maximize the scope of butanol production from lignocellulosic hydrolysates. These findings contribute to advancing sustainable biofuel technologies, offering promising prospects for a greener and more energy-resilient future.

8.1.4. Genome-Scale Metabolic Models (GSMs) Development

The proposed genome-scale metabolic models (GSMs), iKK_CAC for *C. acetobutylicum* ATCC 824 and iKK_CPA for *C. pasteurianum* ATCC 6013, have significantly advanced our understanding of *Clostridium* ABE fermentation. By integrating extensive genomic and biochemical data, these models offered a comprehensive systems-level view of the metabolic networks involved in biobutanol

production. We gained valuable insights into metabolic flux distributions through computational simulations, identifying critical metabolic bottlenecks that limit biobutanol yields. The GSMs served as powerful tools for predicting metabolic behaviors and devising targeted metabolic engineering strategies to enhance biobutanol production. However, it is important to acknowledge that the accuracy of these models is subject to the quality and completeness of the input data, and further experimental validation is required to verify the efficacy of the engineering targets suggested by the models. Despite these limitations, the developed GSMs represent a significant step towards optimizing biobutanol production and hold promise for contributing to the advancement of sustainable biofuel technologies.

8.1.5. Sono-Enzymatic Processes for Synthesis of Bioalcohol Derivatives

This study explored the application of ultrasound-assisted enzymatic reactions for bioalcohol derivative synthesis. We developed a novel ternary approach to gain mechanistic insight into the influence of sonication on sono-enzymatic reactions. Through this innovative approach, we determined the physical and chemical characteristics of the binding pockets of lipases and deduced their locations in different structural motifs. We further investigated the molecular interactions and mechanisms involved in lipase-catalyzed esterification reactions under sonication by employing molecular docking simulations with ligands (substrates or products).

Applying the ternary approach in n-butyl levulinate (n-BL) synthesis, a value-added derivative of biobutanol, yielded quantitative results that underscored its effectiveness in improving reaction efficiency and product yield. The n-BL yield increased significantly from 70.9% to 92.22% under sonication, representing a remarkable 30.4% improvement compared to the control condition. This

enhancement in yield was consistent with the predicted improvements in CALB's catalytic activity and binding affinity for substrates as deduced from the ternary approach. The successful correlation between the computational predictions and experimental outcomes validated the robustness and reliability of the ternary approach in understanding the molecular basis of sonication-induced enzymatic enhancements and its direct impact on product synthesis.

The successful application of the ternary approach in n-BL synthesis underscores its potential for investigating and optimizing other enzyme-catalyzed reactions enhanced by sonication or other external factors. This novel approach significantly advances sustainable and efficient bioalcohol derivative synthesis processes, opening new possibilities for developing environmentally friendly bioenergy and bioproduct technologies.

8.2. SUGGESTIONS FOR FUTURE WORK

The research conducted in this thesis has laid the groundwork for further advancements in the field of bioalcohol synthesis through *Clostridia*. To build upon these findings and address the remaining research gaps, we propose the following suggestions for future work:

Functional Validation of Genes and Integration of Omics Data: We suggest conducting functional validation experiments on the identified genes from the comparative genomic analysis to better understand their role in bioalcohol production. This can be achieved through gene knockouts and overexpression studies to validate the impact of specific genes on solventogenesis and cellulose degradation. Additionally, integrating omics data, such as transcriptomics, proteomics, and metabolomics, with the genome-scale metabolic models (GSMs) can provide a more

comprehensive understanding of the cellular responses during ABE fermentation. This integrated analysis can unveil key regulatory nodes and metabolic pathways, guiding the development of more effective metabolic engineering strategies.

Metabolic Engineering Strategies: Based on the insights gained from the GSMs, further metabolic engineering strategies can be explored to enhance the production of bioalcohols. Rational design and optimization of genetic modifications (such as identifying lethal genes) can be implemented to improve strain performance and achieve higher yields of biobutanol.

***Clostridial* Co-culture Optimization**: Investigating the dynamics of *Clostridial* co-culture systems in more complex environments to mimic lignocellulosic hydrolysates. Understanding the interspecies interactions, substrate competition, and metabolic cross-talk in such environments will enhance co-culture stability and performance. Further exploration of advanced microbial engineering techniques, such as synthetic biology and CRISPR-based genome editing, could aid in tailoring co-culture strains for optimal bioalcohol production. Moreover, conducting pilot-scale and industrial-scale experiments to validate the co-culture performance under real-world conditions will be crucial for successful large-scale bioalcohol production.

Investigations in Ultrasound-Assisted Enzymatic Reactions: The ternary approach developed for ultrasound-assisted enzymatic reactions can be extended to investigate other enzymes and substrates. This technique can be applied to a wider range of biocatalytic processes by exploring the molecular mechanisms of sonication-induced enhancement, enabling faster and more efficient bioconversion reactions across various applications.

Scale-Up and Techno-Economic Analysis: Perform scale-up experiments to assess the feasibility and cost-effectiveness of the developed processes. The techno-economic analysis will be crucial in determining the commercial viability of bioalcohol production through *Clostridia* and identifying potential challenges in industrial implementation.

8.3. FINAL REMARKS

In conclusion, this thesis significantly contributed to bioalcohol synthesis through *Clostridia* from a multidisciplinary approach of tools and techniques from experimental and in-silico investigations. The research addressed critical knowledge gaps and provided valuable insights into the genetic basis, metabolic pathways, and regulatory networks governing bioalcohol production. Furthermore, this dissertation also addressed the process for synthesizing the value-added derivative of biobutanol and the intensification of this process using sonication.

The future work suggestions outlined in this chapter open up exciting opportunities for further advancements in the field. By continuing to explore and optimize the metabolic capabilities of *Clostridial* strains, leveraging innovative biocatalytic processes, and integrating cutting-edge technologies, we can pave the way for more sustainable and efficient bioalcohol production. This research contributes to the broader goal of sustainable energy production, bringing us closer to achieving a sustainable and eco-friendly future. Essentially, this dissertation has contributed to 3 among the 17 Sustainable Development Goals (SDGs) established by the United Nations in 2015, with the aim of achieving a more sustainable and inclusive future for all: (1) Goal 7 (affordable and clean energy), (2) Goal 11 (sustainable cities and communities), and (3) Goal 13 (climate action).

APPENDICES, SUPPLEMENTARY FILES OF THESIS, AND GITHUB SUBMISSIONS

All appendices and supplementary files of Chapters can be found in the following GitHub repository:

https://github.com/theSamurai1997/Thesis_Supplementary.git

Appendix 1 & Excel files Appendix Ex1 to Ex3 are supporting documents for Chapter 2 of this thesis.

Appendix 2 and Excel file Appendix Ex4 are the supplementary files for Chapter 3.

Appendix 3 is the supplementary file for Chapter 4.

Appendix 4 and Excel file Appendix Ex5 are the supplementary files for Chapter 5.

Appendix 5 is the supporting document for Chapter 7 of this thesis.

Listing out of our ProteoPedia depositions for enzyme modeled using Homology Modelling (associated with Chapter 4 of this thesis):

adhE1: https://proteopedia.org/wiki/index.php/Image:AdhE1.pdb

adhE2: https://proteopedia.org/wiki/index.php/Image:AdhE2.pdb

aydh: https://proteopedia.org/wiki/index.php/Image:Aydh.pdb

bdh: https://proteopedia.org/wiki/index.php/Image:Bdh.pdb

bdhA: https://proteopedia.org/wiki/index.php/Image:BdhA.pdb

bdhB: https://proteopedia.org/wiki/index.php/Image:BdhB.pdb

gap3dh: https://proteopedia.org/wiki/index.php/Image:Gap3dh.pdb

Memote analysis report of all GSM models and FROG analysis reports with Fluxer-generated maps of two new GSM models can be found in the GitHub repository (associated with the in-silico part of the Chapter 5):

https://github.com/theSamurai1997/Supplementary_Information_GSMs.git

Fluxer URLs of two GSM models are as follows:

iKK_CAC (iKK1425):

https://fluxer.umbc.edu/model?id=15ebb824bc3389c55de998b4f0de38c4584a852e#

iKK_CPA (iKK848):

https://fluxer.umbc.edu/model?id=a59d51e2d996350aeea8ed4da8340865b2465115#

LIST OF PUBLICATIONS

FROM THESIS

Journal Publications

→ **Kumar, K.**, Barbora, L., Moholkar, V. S.* (2023). Genomic insights into clostridia in bioenergy production: Comparison of metabolic capabilities and evolutionary relationships. *Biotech. & Bioeng*., 1–16 (Link)

→ **Kumar, K.**, Patro, P., Raut, U. et al. (2023) Elucidating the molecular mechanism of ultrasound-enhanced lipase-catalyzed biodiesel synthesis: a computational study. *Biomass Conv. Bioref.* (Link)

→ **Kumar, K**., Roy, K., & Moholkar, V. S.* (2021). Mechanistic investigations in sonoenzymatic synthesis of n-butyl levulinate. *Process Biochemistry, 111*, 147-158. (Link)

Preprints

→ **Kumar, K.***, Jadhav, S. M., & Moholkar, V. S. (2023), "Acetone-Butanol-Ethanol (ABE) fermentation with Clostridial Co-cultures for Enhanced Biobutanol Production", *bioRxiv* 2023.12.08.570763; (Link)

→ **Kumar, K.***, Azeeza S. MD, Pragati, Chandane, P., Kori, M., Shivram, A., Yadav, S., Barbora, L., & Moholkar, V. S. (2022) Computational investigations in inhibition of alcohol/aldehyde dehydrogenase in lignocellulosic hydrolysates" *bioRxiv* 2022.11.19.517192; (Link)

→ **Kumar, K.**, Barbora, L., & Moholkar, V. S.* (2022). Comparative genomic analysis and constraint-based analysis of genome-scale metabolic models of the genus *Clostridia*. *Authorea*. November 16, 2022. (Link)

Book Chapters

→ **Kumar, K.*** & Moholkar, V.S. (2023). Mechanistic Aspects of Enhanced Kinetics in Sonoenzymatic Processes Using Three Simultaneous Approaches. In: Moholkar, V.S., Mohanty, K., Goud, V.V. (eds) *Sustainable Energy Generation and Storage*. NERC 2022. Springer, Singapore. (Link)

→ **Kumar, K.***, Shah, H., & Moholkar, V.S. (2022). Genetic Algorithm for Optimization of Fermentation Processes of Various Enzyme Productions. Satya Eswari, J., & Suryawanshi, N. (eds.) *Optimization of Sustainable Enzymes Production: Artificial Intelligence and Machine Learning Techniques* (1st ed.), pp. 121-144, Chapman and Hall/CRC (Link)

→ **Kumar, K.**, Anand, A., & Moholkar, V. S.,* Molecular Hydrogen (H_2) Metabolism in Microbes: A special focus on biohydrogen production. In: Soccol, C. R., Brar, S. K., Permaul, K., Pakshirajan, K., (eds) *Biohydrogen - Advances and Processes.* Biofuel and Biorefinery Technologies, vol. 13, Springer, Singapore. (Link)

→ **Karan Kumar**, Kaustubh C. Khaire, Kuldeep Roy, and Vijayanand S. Moholkar, "Applications of Microbial Fermentation Technique for the Synthesis of Bioalcohols from Different Agrobiomass". (revision submitted in Chapman and Hall/CRC Press Efficient treatment of agro-biomass for the synthesis of value-added materials: Economic assessment in judgment of sustainability)

In-preparation/Submitted/Under-review

→ **Kumar, K.,** Barbora, L., Moholkar, V. S.*, "Unravelling the genomic differences between solventogenic species *Clostridium acetobutylicum* ATCC 824 and *Clostridium pasteurianum* ATCC 6013"

→ **Kumar, K.,** Barbora, L., Moholkar, V. S.*, "Biochemical and Systems Biological comparison of acetone-butanol-ethanol (ABE) fermentation in solventogenic species *C. acetobutylicum* and *C. pasteurianum*"

→ **Kumar,** K., Barbora, L., Moholkar, V. S.*, "Deciphering the physiological underpinning of solventogenesis in Clostridia using genome-scale metabolic models"

FROM COLLABORATION (NOT FROM THESIS)

Journal Publications

→ Singh, N.,[#] **Kumar, K.**[#], Goyal, A., & Moholkar, V. S.* (2022), Ultrasound-assisted biodiesel synthesis by in–situ transesterification of microalgal biomass: Optimization and kinetic analysis. *Algal Research*, 61, 102582. (Link)

→ Malani, R., **Kumar, K.**, Umriwad, S. B., Goyal, A., and Moholkar, V. S.* (2019), Ultrasound-assisted enzymatic biodiesel production using blended feedstock of

non–edible oils: Kinetic analysis. *Energy Conversion and Management*, 188, 142-150. (Link)

Preprints

→ Anand, A., **Kumar, K.***, Khaire, K. C., Roy, K., & Moholkar, V. S. (2023). Ultrasound-Assisted Hydrolysis of Food Waste using Glucoamylase: Statistical Optimization and Mechanistic Analysis with Molecular Simulations. *bioRxiv* 2023.12.14.571620; (Link)

→ Caroline Vilas Boas de Melo, Maruf Ahmed Bhuiyan, Winfred Nyoroka Gatua, Stephen Kanyerezi, Leonard Uzairue, Priscilla Abechi, **Karan Kumar**, et al. (2020). Transcriptomic dysregulations associated with SARS-CoV-2 infection in human nasopharyngeal and peripheral blood mononuclear cells. *bioRxiv* 2020.09.09.289850; (Link)

Book Chapters

→ Anand, A., **Kumar, K.** & Moholkar, V.S.*, Various routes for hydrogen production and its utilization for sustainable economy. In: Soccol, C. R., Brar, S. K., Permaul, K., Pakshirajan, K., (eds) Biohydrogen - Advances and Processes. Biofuel and Biorefinery Technologies, vol. 13, Springer, Singapore. (Link)

→ Jadhav, S. M. & **Kumar, K.***, Biomarkers Analysis for Alzheimer's Disease. In: Bhattacharya, K., & Bhattacharjee, A., (eds) Computational and Experimental Studies in Alzheimer's Disease, Chapman and Hall/CRC (Link)

→ **Kumar, K.**, Manoj, A., and Ansari, K. M.* (2018). A Review on Food Additives Toxicity (khaady yojakon kee vishaaktata par ek sameeksha), *Vishvigyan Sandesh*, CSIR – Indian Institute of Toxicology Research, Lucknow, Uttar Pradesh, India *vol. 30*, pp. 11-14, 2018-19, ISBN: 0972-1746. (Link)

In-preparation/Submitted/Under-review

→ Shyamali Sarma, **Karan Kumar**, Vikash Kumar Dubey, and Vijayanand S. Moholkar, "Optimization of Medium for Waste Glycerol Fermentation for Biohydrogen Production"

→ Anjli Tanwar, Disha Mukherjee, **Karan Kumar***, and Vijayanand S. Moholkar*, "Application of Microbial Fermentation Technique for the Synthesis of

Fermentable Sugars from Different Agrobiomass". (revision submitted in Chapman and Hall/CRC Press Efficient treatment of agro-biomass for the synthesis of value-added materials: Economic assessment in judgment of sustainability)

→ Shraddha M. Jadhav, Poulami Mukherjee, **Karan Kumar***, and Vijayanand S. Moholkar*, "Application of Enzymatic Technique for the Synthesis of Bio-alcohols From Different Agrobiomass". (revision submitted in Chapman and Hall/CRC Press Efficient treatment of agro-biomass for the synthesis of value-added materials: Economic assessment in judgment of sustainability)

→ Shrivatsa Hegde, Ramyakrishna A.R, **Karan Kumar***, and Vijayanand S. Moholkar*, "Application of Microbial Technique for the Synthesis of Organic Acid from Different Agrobiomass." (revision submitted in Chapman and Hall/CRC Press Efficient treatment of agro-biomass for the synthesis of value-added materials: Economic assessment in judgment of sustainability)

LIST OF ACADEMIC AWARDS & PRESENTATIONS

LIST OF ACADEMIC AWARDS DURING PH.D.

→ Recommended candidate, SERB – ITS under Young Scientist category to attend AlgalBBB – 2023 at Hawaii, USA

→ Best Paper Award at VII – International Conference on Sustainable Energy and Environmental Challenges (VII – SEEC) held at IIT – BHU on December 16-18, 2022.

→ Best Keynote Speaker Award at 3rd International Conference on Biofuel and Biomass 2022 held in Singapore during December 8-9, 2022

→ Recommended candidate, DBT-CTEP international travel grant to attend ISMB – 2022 at Wisconsin, USA

→ Springer Best Paper Award at the North – East Research Conclave (NERC) 2022 organized by IIT Guwahati during May 20-22, 2022.

→ First Prize (Consolation) in Essay Competition (Theme: Renewable Energy Production; Category VI) by CSIR- Central Leather Research Institute, Chennai, Tamilnadu under Swachhata Pakhwada (March 15, 2022).

→ Best Young Scientist Award at the 2nd International webinar on Biofuel & Biomass which was held during January 24-25, 2022.

→ Best Oral Presentation Award at XVIII Annual Convention of the Biotech Research Society, India (BRSI) International Conference on Biotechnology for Resource Efficiency, Energy, Environment, Chemicals, and Health (BREEECH-2021) held at CSIR – Indian Institute of Petroleum Dehradun during December 01-04, 2021.

→ Best Project Presentation Award in an international online competition to team Genomics – II for the at HackBio virtual internship program, September 05, 2020.

→ Scored high marks and awarded best interactor in an e-assessment conducted in one week online International FDP on "Artificial Intelligence and IoT-Heuristic Transdisciplinary Perspectives", organized by CIRCID, Holy Cross College (Autonomous)

INVITED/KEYNOTE TALKS

→ *Keynote speaker* on the topic entitled, "Recent advances in Clostridial co-culture techniques to mitigate global energy crisis by enhanced biofuel production", at 3rd International Webinar on Biofuel & Biomass 2022 at Singapore City, Singapore, organized by Conference Mind during December 08-09, 2022. (Best Keynote Speaker Award)

→ **Karan Kumar*** and Vijayanand S. Moholkar, "Deciphering the physiological underpinning of solventogenesis in Clostridia using genome-scale metabolic models" ***Short invited talk*** at 16th International Conference on the Genetics, Physiology, and Synthetic Biology of Solvent- and Acid-Forming *Clostridia* (Clostridia XVI), held at INSA in Toulouse, France during September 14 - 17, 2022.

→ *Invited speaker* on the topic entitled, "In-silico approach to enhance the production of alcoholic biofuels in solventogenic Clostridia: Recent progress in the reconstruction of genome-scale metabolic models" at 2nd International Webinar on Biofuel & Biomass 2022 organized by Conference Mind, scheduled during January 24-25, 2022. (Best Young Scientist Award)

→ *Guest speaker* for the event on career guidance cum expert lecture on the topic "Application of Computational Biology and Bioinformatics in Biofuel Production Technologies", at Seminar Hall, Faculty of Life Sciences, Mandsaur University, Mandsaur, Madhya Pradesh on January 1, 2022.

CONFERENCE PRESENTATIONS (INTERNATIONAL)

→ **Karan Kumar*** and Vijayanand S. Moholkar, "Unravelling the genomic differences between solventogenic species *Clostridium acetobutylicum* ATCC 824 and *Clostridium pasteurianum* ATCC 6013", ***Oral presentation*** at upcoming VII – International Conference on Sustainable Energy and Environmental Challenges (VII-SEEC), at Indian Institute of Technology - BHU, Varanasi, Uttar Pradesh, India during December 16-18, 2022. (Best Paper Award)

→ Avinash Anand*, **Karan Kumar**, Kaustubh C. Khaire, and Vijayanand S. Moholkar, "Enzymatic saccharification of food waste for biohydrogen production: A response surface-based optimization and molecular simulation approach", ***Poster presentation*** at International Conference on Biotechnology for Sustainable

Bioresources and Bioeconomy (BSBB-2022), Indian Institute of Technology, Guwahati, India during December 7-11, 2022.

→ **Karan Kumar*** and Vijayanand S. Moholkar, "Computational investigations in inhibition of alcohol/ aldehyde dehydrogenase (*AAD*) in lignocellulosic hydrolysates" ***Poster presentation*** at 30th Conference on Intelligent Systems for Molecular Biology (ISMB 2022) held at Monona Terrace Convention Center, Madison, Wisconsin, USA during July 10 - 14, 2022.

→ **Karan Kumar*** and Vijayanand S. Moholkar, "Mechanistic aspects of enhanced kinetics in sonoenzymatic processes using three simultaneous approaches", ***Oral presentation*** at North-East Research Conclave 2022 (NERC 2022) held at Indian Institute of Technology Guwahati, Assam, India during May 20-22, 2022. (Springer Best Paper Award)

→ **Karan Kumar*** "Investigation and Enhancing the production of alcoholic biofuels using Systems Metabolic Engineering Approach", ***Three Minute Thesis Presentation*** (3MTT) at Research & Industrial Conclave Integration 2022 (RIC 2022), organized by Students' Academic Board in collaboration with IIT Guwahati Research Park at Indian Institute of Technology Guwahati, Assam, India during January 20-23, 2022.

→ Pragati*, **Karan Kumar**, and Vijayanand S. Moholkar, "Potent process inhibitors affecting the key enzyme adhE2 for butanol production in alcohologenesis in *Clostridium acetobutylicum*", ***Oral presentation*** at Research & Industrial Conclave Integration 2022, organized by Students' Academic Board in collaboration with IIT Guwahati Research Park at Indian Institute of Technology Guwahati, Assam, India (January 20-23, 2022). (Best Oral Presentation Award)

→ **Karan Kumar*** and Vijayanand S. Moholkar, "Reconstruction and constraint-based analysis of genome-scale metabolic models: In-silico approach to enhance the production of biofuels in solventogenic *Clostridia*", ***Oral presentation*** at Research & Industrial Conclave Integration 2022, organized by IIT Guwahati Research Park at Indian Institute of Technology Guwahati, Assam, India during January 20-23, 2022.

→ **Karan Kumar*** and Vijayanand S. Moholkar, "Theoretical studies to understand the inhibition mechanism of alcohol/aldehyde dehydrogenase due to inhibitors present in hydrolysates from lignocellulosic biomass", ***Oral presentation*** at VI –

International Conference on Sustainable Energy and Environmental Challenges (VI-SEEC), Lucknow, Uttar Pradesh, India during December 27-29, 2021.

→ **Karan Kumar*** and Vijayanand S. Moholkar. "Comparative genomic analysis and constraint-based analysis of available genome-scale metabolic models of the solventogenic clostridia", ***Oral presentation*** at International Conference on Biotechnology for Resources Efficiency, Energy, Environment, Chemicals and Health (BREEECH-2021), CSIR – Indian Institute of Petroleum, Dehradun during December 1-4, 2021. (Best Oral Presentation Award)

→ **Karan Kumar*** and Vijayanand S. Moholkar. "Unravelling the mechanistic aspects of sono-enzymatic esterification via dual approach of experiments and computational analysis", ***e-poster presentation*** at ESS-AOSS-JSS 1st Joint Sonochemistry Conference (online) (November 8-10, 2021).

→ **Karan Kumar*** and Vijayanand S. Moholkar, "Applications of Machine Learning in Systems Metabolic Engineering Strategies for Enhanced Biofuel Production", ***e-poster presentation*** in 2nd Edition of Virtual International Conference on Natural Products and Synthetic Biology (ICNSB-2020), Vellore Institute of Technology, Vellore and Society of Chemical and Synthetic Biology (virtual conference, abstract published).

→ **Karan Kumar*** and Vijayanand S. Moholkar, "Systems Metabolic Engineering Strategies to Enhance Production of Biofuels: A mini-review", ***Oral presentation*** in the 2nd International Conference on Bioprocess for Sustainable Environment and Energy (ICBSEE-2020), March 05-07, 2020. ICBSEE-20-39, page no.: 86, (ISBN: 818553107).

→ Ritesh Malani, **Karan Kumar***, Sachin B Umriwad, Arun Goyal, and Vijyanand Suryakant Moholkar, "Ultrasound-assisted enzymatic biodiesel production using blended feedstock of non–edible oils: Kinetic analysis", ***Oral presentation*** Fourth Asia-Oceania Sonochemical Society Conference (AOSS-4), Nanjing University, Nanjing, China, September 19 – 21, 2019.

→ **Karan Kumar***, Darpan Raghav, and Rathinasamy Krishnan, "Recent advances in carbon-based nanostructured materials (CBNs) and their biomedical applications", in First International Conference on Energy and Environment: Global Challenges (ICEE-2018), Department of Chemical Engineering, National Institute of Technology Calicut, Kerala, India (abstract published).

SHORT BIOGRAPHY

Karan Kumar is a dedicated researcher currently pursuing a Ph.D. in Energy Science and Engineering at the esteemed Indian Institute of Technology Guwahati (I.I.T. Guwahati). Holding a B.Tech. in Biotechnology from the National Institute of Technology Calicut, Karan graduated with distinction, laying the foundation for his academic journey.

Karan's research focuses on the nexus of biotechnology and energy science. In his doctoral pursuits at I.I.T. Guwahati, he investigates enhanced bioalcohol production and derivatives using a combination of experimental and in-silico methodologies. His work delves into constraint-based modeling, in-silico metabolic engineering, and comparative protein modeling, showcasing a robust analytical skill set.

Demonstrating proficiency in molecular simulations, comparative genomics, and mathematical modeling, Karan utilizes tools such as Biopython, MATLAB, R, and LaTeX. His computational skills extend to professional software for cheminformatics, simulation, and visualization, affirming his comprehensive expertise.

Recognized for his academic excellence, Karan is a recipient of the prestigious Prime Minister's Research Fellowship. Actively engaged in esteemed organizations like the International Society for Energy Environment and Sustainability (ISEES) and the Biotech Research Society of India (BRSI), Karan exemplifies a commitment to advancing knowledge. His accolades include Best Paper Awards, the PMRF International Travel Grant, and distinctions as the Best Keynote Speaker and Best Young Scientist at international symposiums, solidifying his standing as a promising researcher in the field of biotechnology and sustainable energy.

Professional Links:

GOOGLE SCHOLAR LINKEDIN RESEARCHGATE ORCID GITHUB

E-mail Id: karaniitg.pmrf18@gmail.com

Personal website: https://sites.google.com/view/karaniitg/home