Bioinformatics analysis and cloning of potent thermostable endoglucanases for biomass degradation

and

Structural characterisation and activity assay of engineered *Pyrococcus furiosus* DNA ligases

Simran Panda
MS15184

*A dissertation submitted for the partial fulfilment of
BS-MS dual degree in Science*

Indian Institute of Science Education and Research Mohali
May 2020

To,
My Grandmother

Certificate of Examination

This is to certify that the dissertation titled “**Bioinformatics analysis and cloning of potent thermostable endoglucanases for biomass degradation and Structural characterisation and activity assay of engineered *Pyrococcus furiosus* DNA ligases**” submitted by **Ms. Simran Panda** (Reg. No. **MS15184**) for the partial fulfilment of BS-MS dual degree programme of the Institute, has been examined by the thesis committee duly appointed by the Institute. The committee finds the work done by the candidate satisfactory and recommends that the report be accepted.

Dr. Rajesh Ramachandran	Dr. Shashi Bhushan Pandit	Prof. Purnananda Guptasarma
(Thesis Examiner)	(Thesis Examiner)	(Supervisor)

Dated: May 4, 2020

Declaration

The work presented in this dissertation has been carried out by me under the guidance of **Prof. Purnananda Guptasarma** at the Indian Institute of Science Education and Research Mohali. This work has not been submitted in part or in full for a degree, a diploma, or a fellowship to any other university or institute. Whenever contributions of others are involved, every effort is made to indicate this clearly, with due acknowledgement of collaborative research and discussions. This thesis is a bonafide record of original work done by me and all sources listed within have been detailed in the bibliography.

Simran Panda
(Candidate)
Dated: May 4, 2020

In my capacity as the supervisor of the candidate's project work, I certify that the above statements by the candidate are true to the best of my knowledge.

Prof. Purnananda Guptasarma
(Supervisor)

Acknowledgement

In the first place I would like to thank Prof. Purnananda Guptasarma, my supervisor, for giving me the opportunity to explore the field that I love and for all his support, not only during my thesis work but also throughout my studies and for introducing me to the world of proteins. While he gave me complete independence regarding the execution of my research work which helped to build my scientific mind to a great extent, I never found his office door shut for a quick discussion or a long pending trouble-shooting.

I want to thank my thesis committee members, Dr. Rajesh Ramachandran and Dr. Shashi Bhushan Pandit, for their time in assisting and examining this thesis.

My sincerest thanks all my professors at IISER Mohali, who trained me in every way possible which helped me in building a scientific attitude, which will remain indispensable in my future endeavours.

I would like to thank all lab members for accepting me as a part of this colourful group of people of the PG lab. Thanks for always helping me out and sharing your expertise. Thank you for all the support, care and guidance in this short period of time. I wish all of them, all the best in the future and want to convey that I had a lot of fun working with them.

I will always be indebted to all my friends throughout the five years at IISER Mohali for providing me with the right kind of advice, fun times, dance and badminton sessions amongst others. I have been very fortunate be in close friendship with genuine and self-evaluating people especially, Swastik and Sujata, who have been there for me no matter what. I sincerely thank them for all their efforts, in developing and shaping me into who I am today.

I especially would like to thank Divanshu for supporting me with all his patience and unconditional love and care, all the way from another continent.

I am obliged to my family for believing in me at every point of my life and for providing me with their never-ending support. I would like to thank my father for endless thought-provoking discussions and my mother for making every effort to keep me healthy. I would like to thank my elder brother, for casual fun discussions time and again. I am also grateful to my grandparents, uncles and aunts and my cousins for always being there for me and sharing frolic times to recharge me for upcoming challenges.

Last but not the least; I am thankful to the existence of dance, music and badminton in my life which has kept me energetic and alive in times of pressure. I would like to bow down to the elegant and marvellous nature around us which has always kept me curious and motivated to learn new things every day.

Simran

List of Figures and Photographs

Figure 1.1 Schematic representation of Biofuel Production process

Figure 1.2 Cellulosome on surface of *Clostridium thermocellum*

Figure 1.3 Schematic representation of the structure of cellulosome

Figure 1.4 Different components of a cellulosome

Figure 1.5 Structure of cellulose

Figure1.6 Schematic representation of domains of a cellulase

Figure 1.7 Action of different kinds of cellulases on cellulose

Figure 1.8 List of cellulolytic enzymes in *Clostrdium thermocellum* with NSAF values

Figure 1.9 Resazurin oxygen level indications – Blue (high), pink (medium), and yellow (low) in medium

Figure 1.10 Schematic representation of PCR cycle

Figure1.11 Concept of Gel Electrophoresis

Figure 1.12 pET23a Vector map annotated using Snapgene

Figure1.13 Representative images of medium with resazurin indicating oxygen levels

Figure 1.14 Gel Electrophoresis image of isolated gDNA

Figure 1.15 Gel electrophoresis images of different PCRs

Figure 1.16 Gel electrophoresis image of attempted reamplification PCR from PCR9

Figure 1.17 Plasmid maps of recombinant clones

Figure 1.18 Gel electrophoresis image of PCR done to amplify CelR and CelA

Figure 1.19 Gel electrophoresis image of double digested plasmids

Figure 1.20 Images showing successful preparation of XL1-Blue competent cells

Figure 1.21 Gel electrophoresis of colony PCR

Figure 2.1 The C-terminal helix mediated stabilization of closed conformation in *P.furiosus* DNA ligase

Figure 2.2 Structure comparison of two archael ligases in two different conformations

Figure 2.3 Structure alignment of different domains of T4 DNA ligase and *P.furiosus* DNA ligase

Figure 2.4 Schematic representation of hypothesis behind the PfuT4H engineering

Figure 2.5 The structure of T7 DNA Ligase

Figure 2.6 ΔPfu ligase structure

Figure 2.7 Characteristic secondary structure peak of protein in CD spectroscopy

Figure 2.8 SDS-PAGE of purified ΔPfu ligase using Ni-NTA chromatography

Figure 2.9 Anion exchange chromatogram of Ni-NTA purified ΔPfu ligase

Figure 2.10 SDS-PAGE of anion exchange fractions of ΔPfu ligase

Figure 2.11 Absorption spectrum of ΔPfu ligase

Figure 2.12 CD spectrum of ΔPfu ligase showing its secondary structure

Figure 2.13 CD spectrum showing thermal melt of ΔPfu ligase

Figure 2.14 CD spectra and MRE plots at 222nm of chemical denaturation of ΔPfu ligase

Figure 2.15 Anion exchange chromatogram of Ni-NTA purified PfuT4H ligase

Figure 2.16 SDS-PAGE of Ni-NTA purified and anion exchange fractions of PfuT4H

Figure 2.17 CD spectrum of PfuT4H showing its secondary structure

Figure 2.18 CD spectrum showing thermal melt of PfuT4H

Figure 2.19 CD spectra and MRE plots at 222nm of chemical denaturation of PfuT4H ligase

Figure 2.20 Urea-PAGE gel image of ligation assay at 25ºC

List of Tables

Table 1.1 Plant biomass composition

Table 1.2 List of hyperthermophilic and thermophilic organisms

Table 1.3 List of endoglucanases and their details from hyperthermophiles and thermophiles

Table 1.4 List of well characterised thermostable endoglucanases used for comparison

Table 1.5 Plan of Comparisons between putative and reference endoglucanases

Table 1.6 Sequence Identity matrices for different comparisons of putative and reference enzyme

Table 1.7 GH12 bacterial enzyme sequence alignment and comparison with refrence enzymes

Table 1.8 GH5 bacterial enzyme sequence alignment and comparison with refrence enzymes

Table 1.9 List of cellulases from *Clostridium thermocellum*

Table 1.10 Special properties of selected cellulases

Table 1.11 PCR program for different polymerases

Table 1.12 Reaction components for restriction digestion of inserts and plasmids

Table 1.13 Reaction components for ligation of respective vectors with inserts

Table 1.14 Colony PCR components

Table 1.15 Colony PCR program

Table 1.16 Annealing temperature and Extension time for different inserts for Colony PCR

Table 1.17 List of primers designed for different *C.thermocellum* cellulases

Table 1.18 Conditions and parameters varied in different PCRs

Table 1.19 PCR conditions for reamplification of DNA fragments

Table 1.20 Standardised parameters for PCR amplification of CelA, CelQ and CelR

Table 2.1 Different components used in ligation assay

Contents

List of Figures **i**

List of Tables **v**

Abstract **xi**

1. Bioinformatics analysis and cloning of potent thermostable cellulases for biomass degradation **1**

1.1 Introduction 3

1.2 Bioinformatics analysis for potent endoglucanases 15

1.2.1 Materials and Methods 17

1.2.2 Results and Discussion 19

1.3 Cloning of potent cellulases from *Clostridium thermocellum* 37

1.3.1 Materials and Methods 39

1.3.2 Results and Discussion 58

1.3.3 Future Outlook 76

2. Structural characterisation and activity assay of engineered *Pyrococcus furiosus* DNA ligases **77**

2.1 Introduction 79

2.2 Materials and Methods 87

2.3 Results and Discussion 95

2.4 Future Outlook 105

Bibliography **106**

Abstract

Efficient enzymatic depolymerisation is an essential step in the process of production of 2nd generation bio fuels from biomass. Cellulose being the most abundant constituent of biomass is a regenerative source of biomass energy but is robust and recalcitrant in nature. Thermostable cellulases and enzymes from biomass degrading cellular machinery, cellulosome can prove to be good candidates for use of cellulose deconstruction. Here, we present a list of potent thermostable endoglucanases shortlisted by comprehensive bioinformatics analysis using homology modelling. Further, a similar analysis was carried out to enlist potent cellulases from the thermostable anaerobe, *Clostridium thermocellum.* We also describe the attempts in cloning some of the potent thermostable cellulases from *Clostridium thermocellum* genome.

DNA ligase is an indispensably important enzyme that seals nick in DNA backbone. In molecular biology experiments, T4 DNA ligase has been used commercially. The limitation of T4 DNA ligase exists in its thermostability and in dealing with the ligation of the DNA template forming a secondary structure at low temperature. In order to overcome these shortcomings, two mesoactive thermostable DNA ligases were engineered from *Pyrococcus furiosus* WT DNA ligase. Here, we describe the expression and purification of these two engineered ligases, namely ΔPfu ligase and PfuT4H ligase. Through structural characterisation, both novel ligases were observed to be extremely thermostable as well as resistant to chemical denaturation by urea. Both of the ligases were partly denatured by high concentration of guanidium hydrochloride. On comparison of enzymatic activity with T4 DNA ligase, activity was observed to be more for both ligases at room temperature.

Chapter 1

Bioinformatics analysis and cloning of potent thermostable cellulases for biomass degradation

1.1 INTRODUCTION

With the ever-increasing demand of energy and its resources all around the world, humankind has been trying to discover and use new renewable and sustainable sources of energy. According to "BP Statistical Review of World Energy," 2019, consumption of energy around the world has also been reaching new heights, so-much-so that there was a 2.9% jump increase in primary energy consumption which is the fastest since 2010.

The limited amount, high demand and cost of fossil fuels in addition to surplus of greenhouse gases has made them impractical and out-dated as a source of energy. As security of electrical energy is essential for the ceaseless working of our lives, new renewable sources of energy have become all the more attractive (Ellabban, Abu-Rub, & Blaabjerg, 2014).

One of such renewable sources of energy is biofuel that is advised as a cost-efficient and environmentally amiable alternative as opposed to fossil fuels such as petroleum, coal and gas (Lehman, 2020). Biomass, the source of biofuel, can be replenished readily unlike fossil fuels and therefore attention needs to be given in the development of this source of energy.

1.1.1 Biomass and its composition

Any living or recently dead organism, more specifically plants and their by-products that can be used in the production of energy and other products are referred to as Biomass. Plants trap energy from the sun and this energy in biomass can be used directly or indirectly – directly to produce heat by burning it or indirectly to produce biofuel and electricity. In doing so, living biomass uses a carbon-neutral cycle, thereby resulting in no increase in the concentration of greenhouse gases in the atmosphere ("Biomass Energy,"). Furthermore, biomass is important for a lasting production of food, feed, chemicals and materials (Djajadi et al., 2018).

Woody crops, agricultural crop residues, forestry residues, algae, wood processing residues, municipal waste, and wet waste can serve as Biomass feedstocks. These can be either used directly as fuel or are converted into some form of source of energy ("Biomass Resources,"). These feedstocks are a ubiquitous and regenerative resource for the generation of second-generation biofuels (Leis et al., 2018) unlike the first generation biofuels which are made edible parts of crops – sugars and vegetable oils.

Biomass is made up of plant polysaccharides. Plant polysaccharides are of two types – plant cell wall polysaccharides and storage polysaccharides **(Table 1.1).** The second generation of biofuels aims to use the plant cell wall polysaccharides which include celluloses, hemicelluloses and pectin. Different types of monomers of cellulose and hemicelluloses are bonded together with different types of linkages along with the aromatic polymer lignin that makes plant cell wall robust and recalcitrant. Lignin serves as an anti-microbe for the plant cell wall. The exact composition of biomass is very complex and varies highly with geographical location, species of plant, type of tissue and season (Benocci, (2018)).

Plant polysaccharide	**Polymer type**	**Polymer**	**Main monomers**
Plant cell wall	Cellulose		D-glucose
	Hemicellulose	Xylan	D-xylose
		Glucuronoxylan	D-glucuronic acid, D-xylose
		Arabinoglucuronoxylan	D-xylose, L-arabinose, D-glucuronic
		Arabinoxylan	D-xylose, L-arabinose
		Galacto(gluco)mannan	D-glucose, D-mannose, D-galactose
		Mannan/galactomannan	D-mannose, D-galactose
		Xyloglucan	D-glucose, D-xylose, D-fructose, D-galactose
		β(1,3)/(1,4)-Glucan	D-glucose
	Pectin	Homogalacturonan	D-galacturonic acid
		Xylogalacturonan	D-galacturonic acid, D-xylose
		Rhamnogalacturonan I	D-galacturonic acid, L-rhamnose, D-galactose, L-arabinose, ferulic acid, D-glucuronic acid
		Rhamnogalacturonan II	D-galacturonic acid, L-rhamnose, D-galactose, L-arabinose, L-fucose, D-glucose, D-manno-octulosonic acid (KDO), D-lyxo-heptulo-

			saric acid (DhA), D-xylose, D-apiose, L-acetic acid
Storage	Inulin		D-fructose, D-glucose
	Starch	Amylose	D-glucose
		Amylopectin	D-glucose
	Various gums		D-galacturonic acid, L-rhamnose, D-galactose, L-arabinose, D-xylose, L-fucose (depending on the specific gum type)
	Lignin		monolignols: *ρ*-coumaryl alcohol, coniferyl alcohol, sinapyl alcohol

Table 1.1: Plant biomass composition (Benocci, (2018))

The conversion of the plant cell wall polysaccharides into biofuels involves breaking down of these polymers into monomeric sugar units which are then converted into biofuels such as into ethanol through fermentation by yeast (Chandel, Chandrasekhar, Silva, & Silvério da Silva, 2012) **(Figure 1.1).**

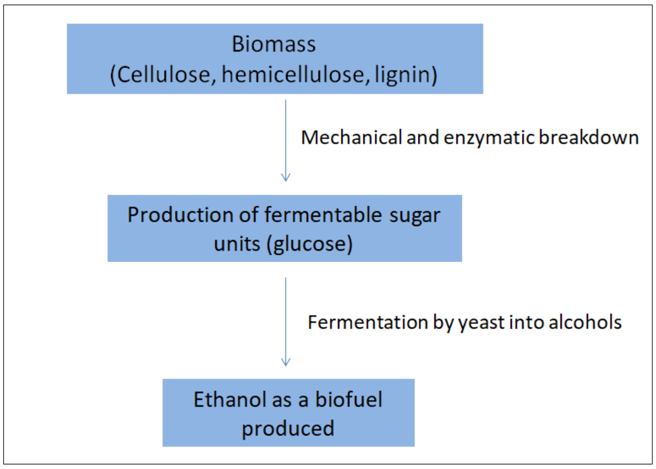

Biomass
(Cellulose, hemicellulose, lignin)
Mechanical and enzymatic breakdown
Production of fermentable sugar units (glucose)
Fermentation by yeast into alcohols
Ethanol as a biofuel produced

Figure 1.1: Schematic representation of Biofuel Production process

1.1.2 Depolymerisation of biomass into fermentable sugars

The process of deconstruction of biomass into fermentable sugar monomeric units is the most critical step towards production of biofuel. Even after years of continuous attempts, it still remains as one of the limiting factors in the bio-refinery process (Leis et al., 2018).

Attempts to depolymerise biomass is achieved by the following steps –

1. Mechanical breakdown - In order to breakdown the robust nature of biomass, physicochemical treatments such as hydrothermal pre-treatment along with others are used. This step is believed to 'loosen up' the recalcitrant biomass for efficient biological and enzymatic deconstruction (Djajadi et al., 2018).

2. Enzymatic depolymerisation – To breakdown the pre-treated biomass into monomeric units, enzymes from cellulolytic organisms or cellulolytic organisms themselves are used.

Limiting factors of depolymerisation –

Physicochemical treatments are helpful in only partially opening up the sturdy structure of tightly packed cellulose microfibrils which makes it difficult to process the biomass, even for the most potent biomass degrading enzymes because of inaccessibility. To help better diffusion of the enzymes and accessibility of substrate, amorphogenesis which is disruption of the tight cellulose packing in a non-hydrolytic manner using certain proteins, making the cellulosic fibres more accessible for enzymatic action (Arantes & Saddler, 2010).

Further, since a long-time different species of fungi belonging to ascomycetes and basidiomycetes taxa such as *Myceliophthora heterothallica* have been used for the breakdown of pre-treated biomass. However, the rate of fungal degradation is too slow for industries, that too only consuming a fraction of the polysaccharides (Benocci, (2018) ; van den Brink, van Muiswinkel, Theelen, Hinz, & de Vries, 2013) (Mäkelä, Donofrio, & de Vries, 2014). Enzymatic degradation is preferred over fungal degradation (Andlar et al., 2018).

As biomass composition is highly complex and variable, there is necessity of a technique of depolymerisation that involves thermostable (to withstand unfavourable conditions of pre-treatment), highly adaptive (to cater to different kinds of biomass) and variety of biomass degrading enzymes (to breakdown the different constituents of biomass simultaneously).

Possible solutions –

Enzymatic degradation can be brought about by a potent enzyme cocktail consisting of a variety of potent biomass degrading enzymes. It allows for selective degradation of biomass depending on the nature of the polysaccharides (Andlar et al., 2018). Consequently, cost efficient and effective enzymes are highly demanded (Leis et al., 2018).

Thermostable enzymes are desirable candidates for enzymatic cocktails. Thermostable enzymes have higher reaction rates and lower diffusional restrictions, which make the process of depolymerisation faster. Due to higher stability, these enzymes can be reused to a great extent (Markoglou & Wainer, 2003). Use of thermostable enzymes allows for saccharification of biomass at elevated temperatures guaranteeing lesser contamination, more accessibility to substrate because of reduced substrate viscosity, working simultaneously along with heat treatments (van den Brink et al., 2013). Thermostable enzymes, along with their extra-ordinary stability, are also more resistant to chemical agents that allows them to be used in industries as pre-treatment of biomass is known to release some by-product that inhibit the lesser stable mesophilic enzymes (Lasa & Berenguer, 1993).

Hyperthermophilic and thermophilic cellulolytic organisms whose optimum temperatures of growth are greater than 50°C are a good source for such thermostable enzymes. Heterologous expressions of such thermostable enzymes are a more practical option than to engineer the mesophilic enzymes for the purpose of increasing their stability.

Efficient degradation of biomass calls for synergistic activity of a combination of enzymes that are highly adaptive to different substrates. In nature, there exists a cellular machinery in some cellulolytic organisms precisely for this function, called cellulosome. Cellulosome is an extracellular multienzyme cellulolytic complex that is present on cell surface of many anaerobic cellulolytic bacteria such as *C. thermocellum, Acetivibriocellulolyticus, Bacteroides cellulosolvens, Clostridium cellobioparum, Clostridium cellulovorans, and Ruminococusalbus* (**Figure 1.2**) (Lamed, Naimark, Morgenstern, & Bayer, 1987). First described my Lamed *et al*, the cellulosome was reported to have a mass of 2.1 x 106 Daltons and to contain 14 different polypeptides. Apart from being remarkably stable (Ljungdahl et al., 1988), the cellulosome displays extraordinary cellulolytic activity i.e. it can hydrolyse both amorphous and highly ordered crystalline cellulose (Felix & Ljungdahl, 1993). The co-localisation and immobilisation of a wide variety of thermostable cellulolytic enzymes in cellulosome is crucial for the breakdown of biomass polysaccharides into sugars (Leis et al., 2018) (Hirano et al., 2016).

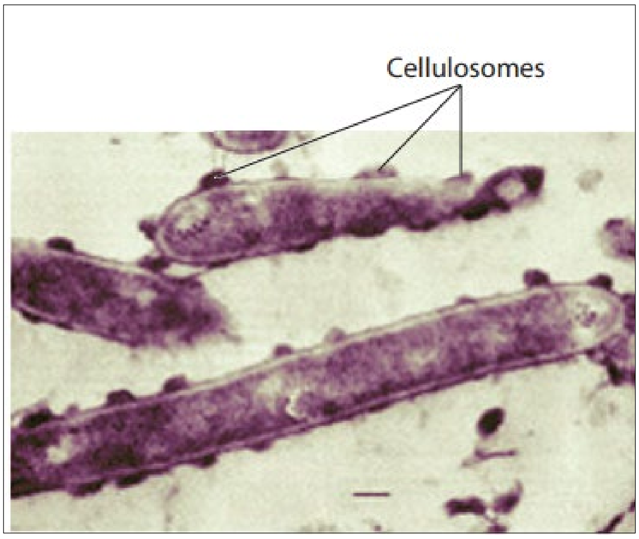

Cellulosomes

Figure 1.2: Cellulosome on surface of *Clostridium thermocellum* (Lamed et al., 1987)

1.1.2.1 Cellulosomes and their structure

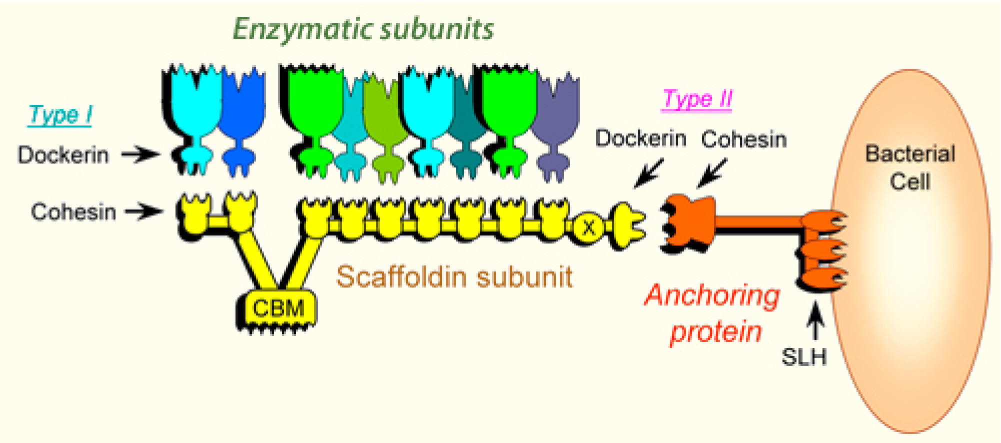

Enzymatic subunits
Type I
Dockerin
Cohesin
Type II
Dockerin
Cohesin
X
Bacterial Cell
Scaffoldin subunit
CBM
Anchoring protein
SLH

Figure 1.3: Schematic representation of the structure of cellulosome (Bayer)

The cellulosome consists of both catalytic and non-catalytic components **(Figure 1.3)**. It contains a large non-catalytic protein called scaffoldin which acts as backbone of the multi-enzyme complex, on which many cohesion modules are located. The catalytic components, i.e. the cellulolytic enzymes contain non-catalytic modules called dockerin through which they interact with cohesion modules. The tight interaction between Cohesin type I on scaffoldin and dockerin type I on enzymes helps in integration of different enzymes to make the cellulosome (**Figure 1.4**). Further a tight interaction between cohesion type II and membrane associated proteins helps in tethering the cellulosome to the cell wall of the bacterium. In addition to this, the scaffoldin as well as some of the enzymes contain a non-catalytic carbohydrate binding module (CBM) that tethers the cellulosome to biomass substrates (Fontes & Gilbert, 2010).

Cellulosomes can prove to be extremely desirable for efficient degradation of biomass on an industrial scale. Studies have been carried to overexpress cellulosomes and purify them, only to get very poor yields (Lamed et al., 1987). Development of synthetic cellulosomes seems like a plausible alternative given its advantages in degrading biomass. Studies also show that enzymes on cellulosomes can be specified in stoichiometric ratios giving high adaptability to these complexes, which is way less cumbersome than development of fusion proteins for every different kind of substrate (Leis et al., 2018). Therefore, there is a need to understand the cohesion-dockerin interaction to understand cellulosome as a whole to make efficient biomass degrading nano-machines.

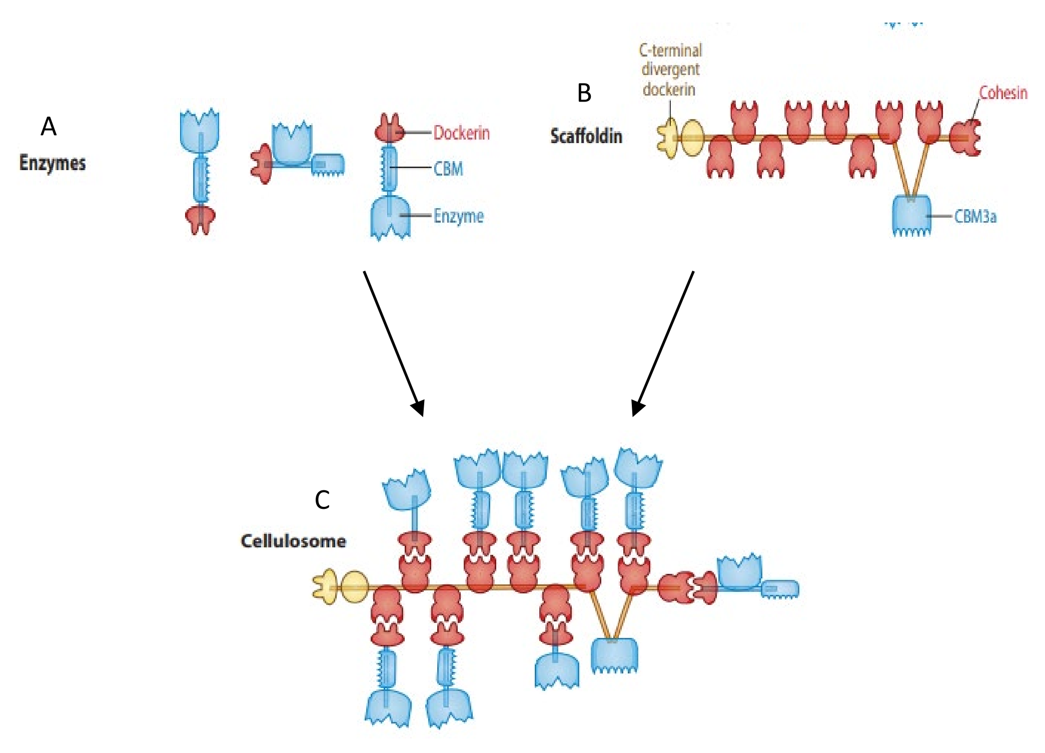

A
Enzymes
Dockerin
CBM
Enzyme
B
Scaffoldin
C-terminal divergent dockerin
Cohesin
CBM3a
C
Cellulosome

Figure 1.4: Different components of a cellulosome (Fontes & Gilbert, 2010)

1.1.2 Most abundant biomass polysaccharide – Cellulose

Cellulose is the most abundant plant cell wall polysaccharide, (Beta- glucan) made up of β-D-glucose units linked by β-(1-4)-O-glycosidic bond organized in bundles called microfibrils **(Figure 1.5**). Its molecular formulae is $(C_6H_{12}O_6)_n$. The degree of polymerization, indicated by n, is broad, ranging from several thousand to several ten thousands (Kolpak & Blackwell, 1976).

Cellulolysis refers to the breakdown of cellulose into cellodextrins/glucose units. This brought about by a class of enzymes called cellulases also called β-glucanases that perform hydrolysis (1-4)- β-D glucosidic linkage in β-D-glucans.

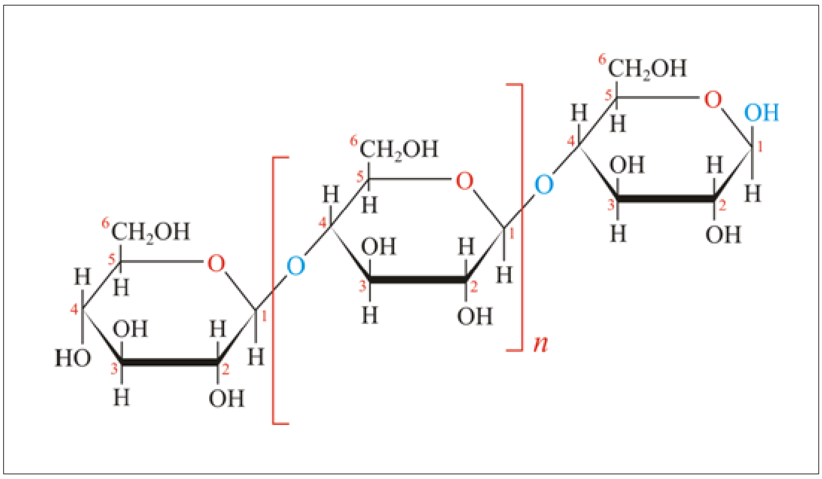


Figure1.5: Structure of cellulose

A general cellulose consists of a signal peptide (SP) followed by a catalytic domain (CD) connected to a carbohydrate-binding module (CBM) (Vuong & Wilson, 2010) by a short polypeptide sequence called a linker **(Figure1.6)** (Liu et al., 2009).

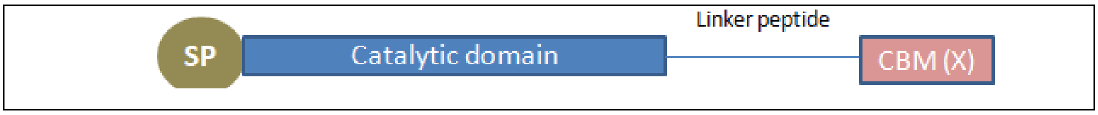


Figure1.6: Schematic representation of domains of a cellulase

Several types of cellulases are known to exist that differ in structure and mechanism of action. The process by which cellulose is depolymerised into glucose units involves a variety of cellulases- a. Endoglucanases, which catalyse the endo-hydrolysis of cellulose b. Exocellulase (also called Cellobiohydrolases), that perform hydrolysis at ends of cellulose to release successive glucose units c. β-glucosidase, which release glucose units from dimer cellobiose (Gomez del Pulgar & Saadeddin, 2014). Exocellulases are further classified – Type I act on reducing ends of cellulose and release glucose units on hydrolysis whereas Type II act on non-reducing ends releasing cellobiose on hydrolysis **(Figure 1.7**).The EC classification, known best for classification of enzymes, is given in the schematic classification (**Figure 1.8).**

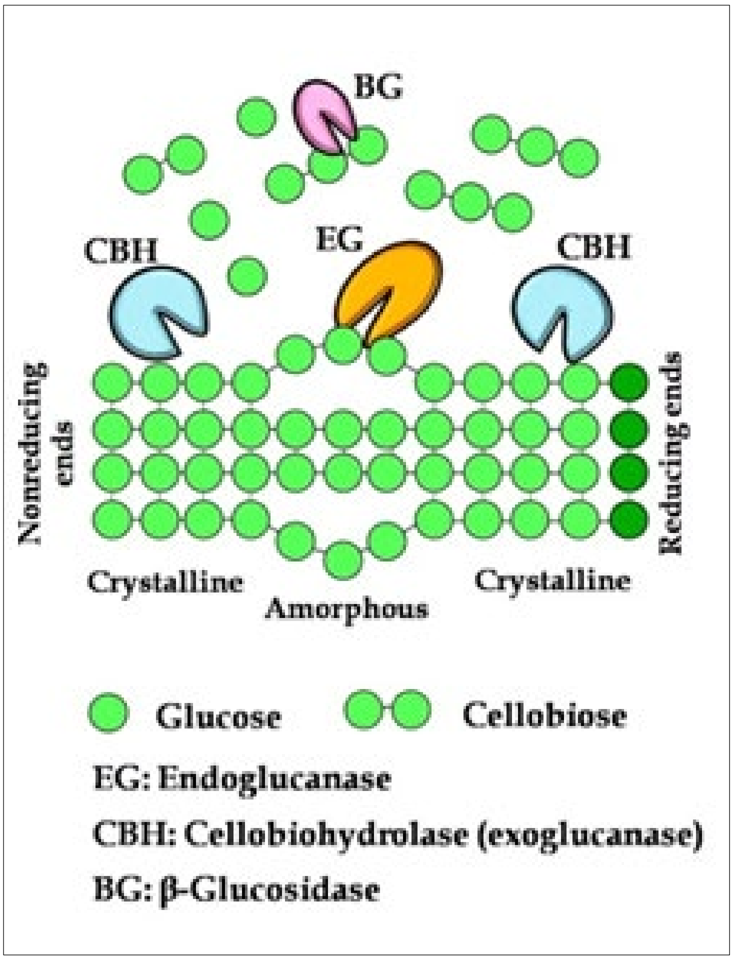

BG
CBH
EG
CBH
Nonreducing ends
Reducing ends
Crystalline
Amorphous
Crystalline
Glucose
Cellobiose
EG: Endoglucanase
CBH: Cellobiohydrolase (exoglucanase)
BG: β-Glucosidase

Figure 1.7: Action of different kinds of cellulases on cellulose ("Composition of Enzymes,")

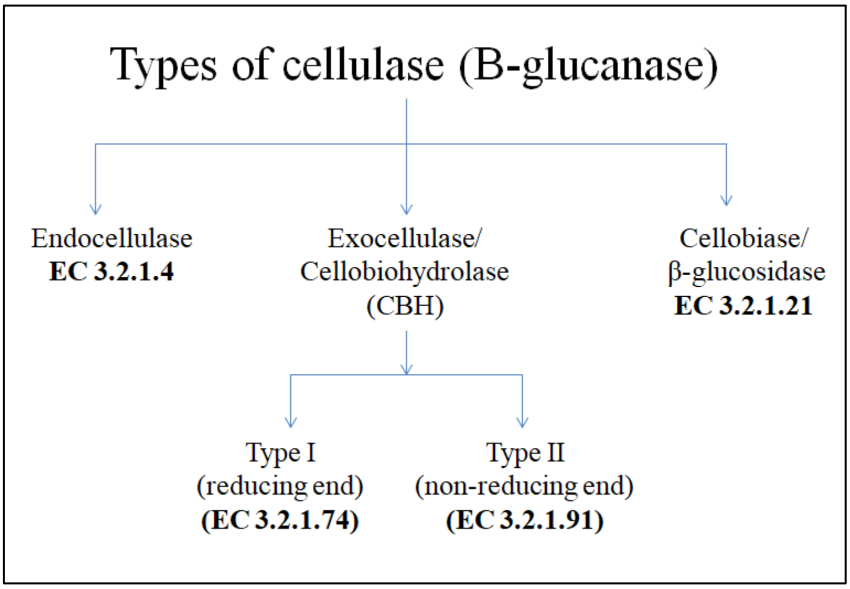

Types of cellulase (B-glucanase)
Endocellulase
EC 3.2.1.4
Exocellulase/
Cellobiohydrolase
(CBH)
Cellobiase/
β-glucosidase
EC 3.2.1.21
Type I
(reducing end)
(EC 3.2.1.74)
Type II
(non-reducing end)
(EC 3.2.1.91)

Figure 1.8: Classification of cellulases

AIM OF PROJECT

As efficient and cost effective depolymerisation of biomass still remains a challenge and cellulose is the most abundant polysaccharide in biomass, the aim of this project was to identify novel putative potential thermostable endoglucanases from hyperthemophilic and thermophilic microorganisms. This project aims on searching, bioinformatically analysing and cloning potent endoglucanase. Furthermore, it also consisted of comprehensive search, selection and cloning of potent cellulases from thermophile *Clostridium thermocellum*. This would be helpful in making enzyme cocktails for enzyme activity analysis on different biomass, as well as in studying cohesin-dockerin interaction of cellulosomes being carried out in the lab.

1.2 Bioinformatics analysis for potent endoglucanases

1.2.1 MATERIALS AND METHODS

1.2.1.1 Materials

Databases used

Database	Source
NCBI	https://www.ncbi.nlm.nih.gov/
UniProt	https://www.uniprot.org/
CAZy	http://www.cazy.org/
RCSB PDB	https://www.rcsb.org/

Websites used

Website	Source
Clustal Omega	https://www.ebi.ac.uk/Tools/msa/clustalo/
ESPript 3.0	http://espript.ibcp.fr/ESPript/ESPript/

1.2.1.2 Methods

Selection of potent endoglucanases through bioinformatics analysis

i. List of novel putative potent thermostable endoglucanases

a. Literature survey was carried out to get an estimate of all hyperthermophilic and thermophilic organisms known to exist using the internet and articles from PubMed in NCBI. PubMed is a free resource that gives open access to biomedical and life-sciences articles.

b. Further, to shortlist celluloytic hyperthermophilic and thermophilic organisms, thorough search was done using CAZy database. CAZy is a frequently updated database that stands for Carbohydrate Active enZYmes that has information on enzymes that degrade, create or modify glycosidic bonds. The enzymes, using their amino acid sequence, are divided into different classes and further into families, based on the structurally related catalytic domains (Davies & Henrissat, 1995; B. Henrissat, 1991; B. Henrissat & Bairoch, 1993, 1996; Bernard Henrissat & Davies, 1997; Lombard, Golaconda Ramulu, Drula, Coutinho, & Henrissat, 2014). Further, organisms whose culturing were possible and were available to do experiments with were chosen over others.

c. To shortlist putative potent thermostable enzymes, genomes of selected hyperthermophile and thermophile were searched using particular glycoside hydrolase family (contains enzyme that hydrolyse or re-arrange glycosidic bonds) annotation in CAZy database. UniProt, a well-known protein database containing sequence and structure information, was used to retrieve the amino acid sequence of the selected enzymes if available else the GenBank accession number was used.

ii. Sequence homology modelling to search for potent thermostable endoglucanase

a. Multiple entries with identical names on the CAZy database were checked for sequence identity. If entries had the exactly same sequence, one entry was retained and others were discarded.
b. Function prediction of the shortlisted putative endoglucanases using their Protein sequence homology modelling were carried out against well-known and characterised thermostable endoglucanases using Clustal Omega. Clustal Omega is a multiple sequence alignment tool that used to check the identity and alignment of different nucleotide or protein sequences. It uses HMM profile and seeded guide trees to carry out the alignment (Madeira et al., 2019). Sequence identity matrices of different alignments were compared to get conclusive results.
c. Protein sequence and secondary structure modelling was also carried out to get a clearer result. The alignment files from Clustal Omega were visualised on ESPript with input of the PDB ID of reference sequence. The similarity of the sequences were analysed to reach conclusive results.

Search and shortlisting of *Clostridium thermocellum* cellulases

a. A list of all cellulases from the *Clostridium thermocellum* genome was made by carrying out a comprehensive literature search.
b. Shortlisting of some of these cellulases was done based on some important factors.

1.2.2 RESULTS AND DISCUSSION

Bioinformatics analysis and prediction of thermostable endoglucanases

i. Hyperthermophiles and thermophiles on earth

Table 1.2 gives a non-exhaustive list of these organisms with their domain classification, i.e. either archae or bacteria. Domain archae was found to have the most of the hyperthermophiles and domain bacteria have most of the thermophiles. As hyperthermophiles will have more thermostable enzymes (Unsworth, van der Oost, & Koutsopoulos, 2007) and heterologous expression of archael proteins in Escherichia coli is possible (R. Kim et al., 1998), we decided to incorporate archael enzymes along with bacterial enzymes into further analysis.

Type	**Domain**	**Number**
Hyperthermophile	Archae	96
	Bacteria	15
Thermophile	Archae	15
	Bacteria	109

Table 1.2: List of hyperthermophilic and thermophilic organisms (jointly prepared with Shreya D.K.)

ii. Potent endoglucanases search using CAZy and literature

Using the organism names of the hyperthermophiles and thermophiles, a scrutiny of literature articles and/or reviews submitted to NCBI and entries in the CAZy database was carried out. The CAZy database has its own classification system of the Carbohydrate active enzymes. All enzymes that hydrolyse the glycosidic bond between two moieties are classified under the glycoside hydrolase family (GH family). This is based on the specificity of the substrate and molecular mechanism, and not on the structural details of the enzymes. Under the EC classification, endoglucanases come under the family EC 3.2.1.4. In CAZy database, EC 3.2.1.4 family corresponds to many GH families – 5, 6, 7, 8, 9, 10, 12, 26, 44, 45, 48, 51, 74, 124, 148 (Lombard et al., 2014). Each of these GH families was searched for characterised as well as putative endoglucanases from hyperthermophiles and thermophiles.

Organism Name	**Domain (Archae(A)/Bacteria(B))**	**Optimum Temperature(°C)**	**Hyperthermophile(HT)/ Thermophile(T)**	**GH**	**ATCC Availability**	**GenBank ID**	**UniProt ID**	**PDB ID**	**No. of Amino acids**
Acidilobus saccharovorans 345-15	A	80-85	HT	5	No	ADL19140.1	D9Q1F2	NA	637
Caldivirga maquilingensis IC-167	A	85	HT	5	Yes	ABW01768.1	A8MBD2	NA	600
				12		ABW01309.1	NA	NA	423
						ABW01682.1	A8MD26	NA	410
						ABW01683.1	A8MD27	NA	437
						ABW02444.1	A8M9X3	NA	284
						ABW02804.1	A8MC20	NA	351
Desulfurococcaceae	A	85	HT	5	No	AEB53062.1	NA	NA	842
Dictyoglomus thermophilum H-6-12	B	78	T	12	Yes	ACI19111.1	B5YCT9	NA	288
				5		ACI18520.1	B5YAS2	NA	335
						ACI18413.1	B5YFI6	NA	545
						ACI19118.1	B5YCY1	NA	331
						ACI19154.1	B5YCY0	NA	312
						ACI19692.1	B5YCX9	NA	332
Dictyoglomus turgidum DSM 6724	B	78	T	12	No	ACK41916.1	B8DZI3	NA	288
				5		ACK41602.1	B8DYX6	NA	335
						ACK41955.1	B8DZM2	NA	312

						ACK42859.1	B8E2L5	NA	545
						ACK41954.1	B8DZM1	NA	332
Ignisphaera aggregans DSM 17230	A	92-95	HT	5	Yes	ADM27158.1	E0SQT4	NA	662
						ADM27405.1	E0SSD3	NA	393
						ADM27073.1	E0SQC3	NA	369
Pyrococcus abyssi GE5	A	94	HT	5	No	CAB49854.1	Q9V052	NA	514
Saccharolobus solfataricus P1	A	94	HT	5	No	SAI86656.1	NA	NA	597
Saccharolobus solfataricus P2	A	92	HT	12	No	AAK42142.1	Q97X08	NA	334
						AAK41590.1	Q97YG7	NA	332
Staphylothermus hellenicus DSM 12710	A	85	HT	5	No	ADI31911.1	D7D815	NA	453
Sulfolobus islandicus L.D.8.5	A	75	T	5	No	ADB88268.1	D2PGC8	NA	594
Thermococcus barossii SHCK-94	A	82.5	HT	5	Yes	ASJ05073.1	NA	NA	395
Thermofilum pendens Hrk 5	A	85-90	HT	5	Yes	ABL79067.1	A1S0T7	NA	640
Thermoproteus uzoniensis 768-20	A	95	HT	5	No	AEA11889.1	NA	NA	627
Thermotoga naphthophila RKU-10	B	80	HT	12	Yes	ADA67369.1	D2C3R5	NA	274
						ADA67370.1	D2C3R6	NA	311
Thermotoga neapolitana	B	80	HT	12	Yes	AAC95059.1	O08428	NA	257
						AAC95060.1	P96492	NA	274
Thermotoga neapolitana DSM 4359	B	80	HT	12	Yes	ACM23282.1	B9K8J9	NA	292
						ACM23283.1	B9K8K0	NA	274

Thermotoga petrophila RKU-	B	80	HT	12	Yes	ABQ47281.1	A5IM58	NA	274
						ABQ47282.1	A5IM59	NA	328
Vulcanisaeta distributa DSM 14429	A	90	HT	5	No	ADN50980.1	E1QTR3	NA	633

Table 1.3: List of endoglucanases and their details from hyperthermophiles and thermophiles

In search of novel potent endoglucanases, those enzymes were shortlisted which were not completely characterised or had very less information about them. Upon examination, it was found that mostly GH5 and GH12 families had putative endoglucanases from hyperthermophiles and thermophiles. A non-exhaustive list, containing 40 of these enzymes, along with necessary details was made **(Table 1.3).** As the end goal was heterologous expression of these enzymes, availability for their genomic DNA and/or culture was checked using the ATCC website. ATCC is a resource centre from where microorganisms, cell lines and other materials can be bought for research. The enzymes for which culture/ genomic DNA of the organism were available on the ATCC website (highlighted in grey in the **Table 1.3**) were selected for further analysis.

iii. Function prediction based on homology modelling using sequence and secondary structure

For predicting the function of the shortlisted endoglucanases, both sequence based and secondary structure based homology modelling was carried out. In order to do that, well characterised endoglucanases had to be used as reference (Cheng et al., 2012; Kataoka & Ishikawa, 2014; H. W. Kim & Ishikawa, 2011; Liberato et al., 2016; Sandgren et al., 2003). As catalytic domains of GH families differ, the comparisons had to be made within the GH families. Further, to check for the thermostability of the enzymes, different reference enzymes from mesophiles, thermophiles and hyperthermophiles were chosen, the details of which are given in the **(Table 1.4).**

S. No.	Organism Name	Domain (Archae(A)/Bacteria(B))	Optimum Temperature(°C)	Hyperthermophile(HT)/Thermophile (T)/ Mesophile(M)	GH	EC	GenBank ID	UniProt ID	PDB ID	No. of Amino Acids
1	*Pyrococcus horikoshii OT3*	A	98	HT	5	3.2.1.4	AAQ31833.1	O58925	2ZUN	458
2	*Pyrococcus furiosus DSM 3638*	A	100	HT	12	3.2.1.4	AAD54602.1	Q9V2T0	3WQ7	319
3	*Bacillus sp. BG-CS10*	B	37	M	5	3.2.1.4	ADD62401.1	D4P8C6	5XRC	569
4	*Streptomyces sp. 11AG8*	B	39	M	12	3.2.1.4	AAF91283.1	Q9KIH1	1OA4	371
5	*Thermobifida fusca YX*	B	55	T	5	3.2.1.4	AAZ54939.1	Q01786	2CKR	466
6	*Thermotoga maritima MSB8*	B	80	HT	12	3.2.1.4	AAZ54939.1	Q60032	3VHN	257

Table 1.4: List of well characterised thermostable endoglucanases used for comparison

iii(a). Sequence based Homology modelling

Homology modelling based on amino acid sequence and secondary structure of the proteins help in predicting the function of the enzyme. At levels of sequence identity ≥ 40%, precise function is conserved and all four EC numbers are conserved (Whisstock & Lesk, 2003).

Using Clustal Omega, comparison of the sequences of the putative endoglucanases was done. As comparisons had to be made within GH families, the available shortlisted endoglucanases were grouped into respective GH families, i.e. 5 and 12 and respective domain of origin i.e. Archae or bacteria. Each of these groups was compared to a thermostable archael endoglucanase, a thermostable bacterial endoglucanase and a bacterial endoglucanase of mesophilic origin (optimum temperature of growth around 37˚C) **(Table 1.5).** No comparison could be made with an archael endoglucanase of mesophilic origin as no such endoglucanase existed, to the best of my knowledge.

Enzyme group		**Reference Enzyme (Refer to Table 1.4 for enzyme details under respective S.no.)**
Archae	GH5	1, 5, 3, 2, 6, 4
	GH12	2, 6, 4
Bacteria	GH5	1, 5, 3
	GH12	2, 6, 4

Table 1.5: Plan of Comparisons between putative and reference endoglucanases

Sequence identity matrices were analysed for the different groups of enzymes. In **Table 1.6(A)**, the GH5 archael putative enzymes ADM27405.1 and ADM27073.1 show a sequence identity of more than 40%, i.e. 57.22% and 64.56% respectively when compared with a well characterised *Pyrococcus horikoshii* (archae) hyperthermophilic endoglucanase. With this information, we safely predicted that these two enzymes will probably work as thermostable endoglucanases. However, these two enzymes showed less sequence identity when compared with the bacterial thermophilic enzyme. This may be because of the thermophilic origin of reference enzyme as well as because of the different domain of origin. The same is the case when compared to a reference enzyme of mesophilic origin. This is understandable as the amino acid sequence of a hyperthermostable enzyme will be quite different from an enzyme of mesophilic origin. In other words, this observation helped us to categorise these two enzymes as thermostable. All the other putative enzymes show a less value of sequence identity. Further, when the same enzymes were compared with enzymes of a different GH family i.e. GH12, even enzymes that showed a great percentage of sequence identity drop to a very low value. These results could lead to faulty analysis and for all the upcoming analysis, comparisons were always made within GH families.

For GH12 archael putative enzymes and GH5 bacterial putative enzymes, in **Table 1.6(B)** and **(C),** we observed that all the putative enzymes show low value of sequence identity with their respective reference enzymes.

GenBank ID (refer to table)	**Sequence Identity %**					
	GH5 Archael HT enzyme (1)	**GH5 Bacterial T enzyme(5)**	**GH5 Bacterial M enzyme(3)**	**GH12 Archael HT enzyme(2)**	**GH12 Bacterial HT enzyme(6)**	**GH12 Bacterial M enzyme(4)**
ABW01768.1	18.55	16.05	17.99	18.72	16.24	13.99
ADM27158.1	18.10	14.77	17.54	15.09	17.14	14.89
ADM27405.1	57.22	18.91	17.19	18.99	13.70	16.13
ADM27073.1	64.56	23.85	14.07	12.24	14.69	15.31
ASJ05073.1	23.30	12.68	19.55	16.46	14.99	5.26
ABL79067.1	18.41	15.45	18.60	14.21	16.67	16.18

Table 1.6 (A): GH5 Archael putative endoglucanases

GenBank ID (Refer to table)	**Sequence Identity %**		
	GH12 Archael HT enzyme(2)	**GH12 Bacterial HT enzyme(6)**	**GH12 Bacterial M enzyme(4)**
ABW01309.1	24.83	20.80	20.20
ABW01682.1	31.80	26.40	20.35
ABW01683.1	24.92	20.72	20.54
ABW02444.1	30.89	26.91	19.52
ABW02804.1	24.75	24.30	23.75

Table 1.6 (B): GH12 Archael Putative endoglucanases

GenBank ID (refer to table)	Sequence Identity %		
	GH5 Archael HT enzyme (1)	GH5 Bacterial T enzyme(5)	GH5 Bacterial M enzyme(3)
ACI18520.1	19.55	23.00	24.04
ACI18413.1	17.40	18.27	30.81
ACI19118.1	19.53	17.29	22.00
ACI19154.1	16.55	23.48	24.57
ACI19692.1	21.04	22.26	23.95

Table 1.6 (C): GH5 Bacterial Putative endoglucanases

GenBank ID (refer to table)	Sequence Identity %		
	GH12 Archael HT enzyme(2)	GH12 Bacterial HT enzyme(6)	GH12 Bacterial M enzyme(4)
ACI19111.1	35.94	42.52	18.67
ADA67369.1	36.57	49.80	20.69
ADA67370.1	32.97	99.22	20.00
AAC95059.1	33.33	73.15	22.33
ACM23282.1	31.56	72.76	21.81
ACM23283.1	37.31	50.98	21.98
ABQ47281.1	36.19	49.41	20.26
ABQ47282.1	32.97	96.89	20.00
AAC95060.1	37.31	50.98	21.98

Table 1.6 (D): GH12 Bacterial Putative endoglucanases

Table 1.6: Sequence Identity matrices for different comparisons of putative and reference enzyme

For GH 12 bacterial putative enzymes, in **Table 1.6(D)**, all the putative enzymes show a sequence identity% of greater than 40% when compared with the well-studied *Thermotoga maritima* (bacterial) thermostable endoglucanase. The enzymes with entries ADA67370.1 and ABQ47282.1 show a very high sequence identity%, i.e. 99.22% and

96.89% as the organism species of these enzymes is same as that of the reference sequence, i.e. *Thermotoga.* Additionally, these enzymes also show a higher sequence identity% when aligned against a hyperthermophilic archael enzyme as compared to a mesophilic bacterial enzyme. This leads us to conclude that enzymes of hyperthermophilic origin have more amino acids in common in their sequences as compared to mesophilic enzymes, even if they belong to two different domains.

Therefore, two of the GH5 Archael endogulcanases (ADM27405.1, ADM27073.1) and all the GH 12 bacterial endoglucanases were safely assumed to function as thermostable endoglucanases.

Sequence identity might not always be the only parameter by which homology modelling can be done. Many sequences with low sequence identity show a high sequence similarity and conserved structural domains which help in proving homology between two sequences and predict function. Next, we did a secondary structure based sequence similarity analysis to get a better idea about these putative enzymes.

iii(b). Secondary structure and sequence similarity based Homology modelling

Sequence similarity refers to when the same position in the sequence have amino acids with the same property. As, the property of the side chains of amino acids play a role in the folding and functioning of proteins, even if sequences are not identical, similar sequence may very well deliver the job. Structure changes more conservatively than sequence during evolution. The conservation of structure in between two proteins can be used to safely predict function and homology of the query protein (Whisstock & Lesk, 2003).

ESPript and PDB IDs were used for secondary structure and sequence similarity based homology modelling. As per table, the alignment and analysis was carried out.

The results obtained were the same as from homology modelling using amino acid sequence. (As the files were huge, only a part of the sequence and some representative results are shown).

Putative GH12 bacterial endoglucanases showed the greatest similarity with the hyperthermophile endoglucanase from *Thermotoga maritima*. A huge amount of similarity signalled that these enzymes would most possibly function as thermostable endoglucanases. (**Table 1.7(B)**)

Additionally, they also showed a larger amount of similarity with a hyperthermophilic archael endoglucanase (**Table 1.7(A)**) as compared to a mesophilic bacterial enzyme (**Table 1.7(C)**). Therefore, these enzymes showed greater amount of similarity when they are thermostable irrespective of the domains they come from.

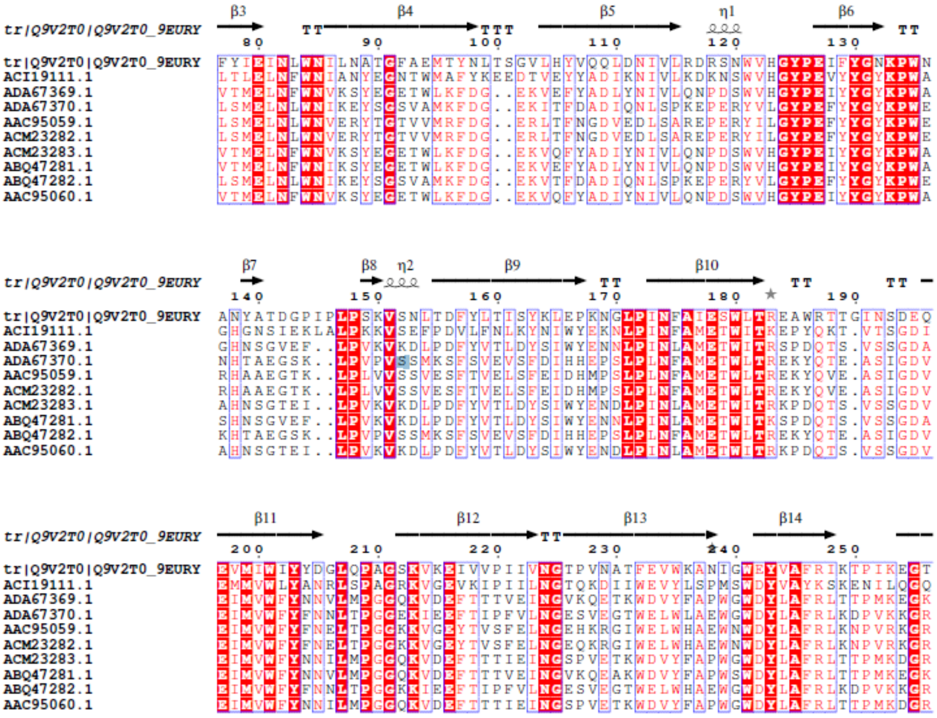

tr|Q9V2T0|Q9V2T0_9EURY
β3
β4
β5
η1
β6
β7
β8
η2
β9
β10
β11
β12
β13
β14
ACI19111.1
ADA67369.1
ADA67370.1
AAC95059.1
ACM23282.1
ACM23283.1
ABQ47281.1
ABQ47282.1
AAC95060.1

Table 1.7 (A): Alignment with *P.Furiosus* endoglucanase

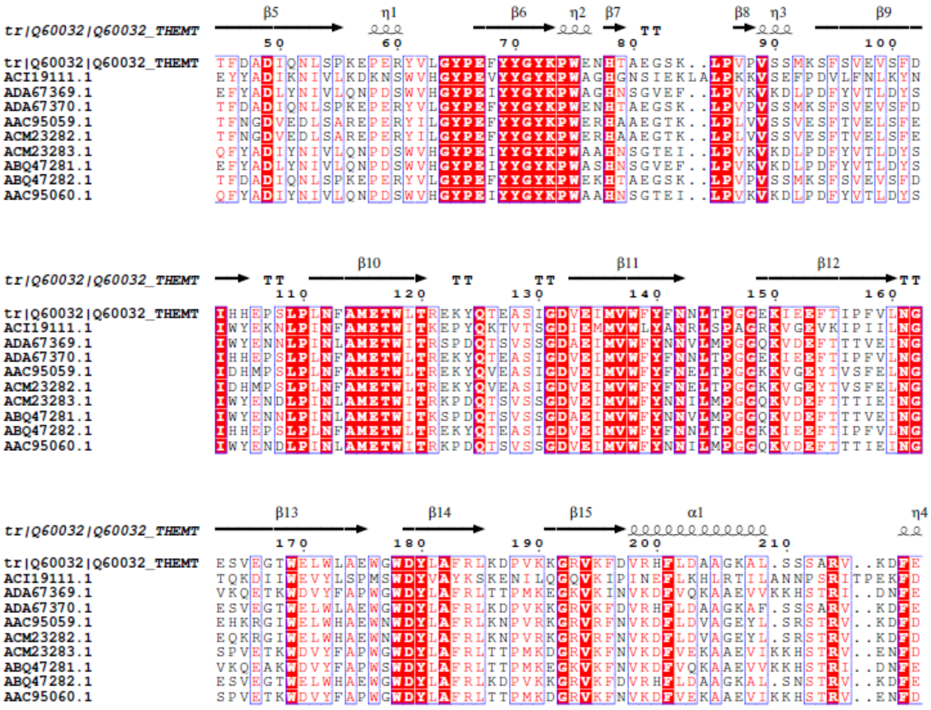

tr|Q60032|Q60032_THEMT
ACI19111.1
ADA67369.1
ADA67370.1
AAC95059.1
ACM23282.1
ACM23283.1
ABQ47281.1
ABQ47282.1
AAC95060.1

Table 1.7 (B): Alignment with *T.maritima* endoglucanase

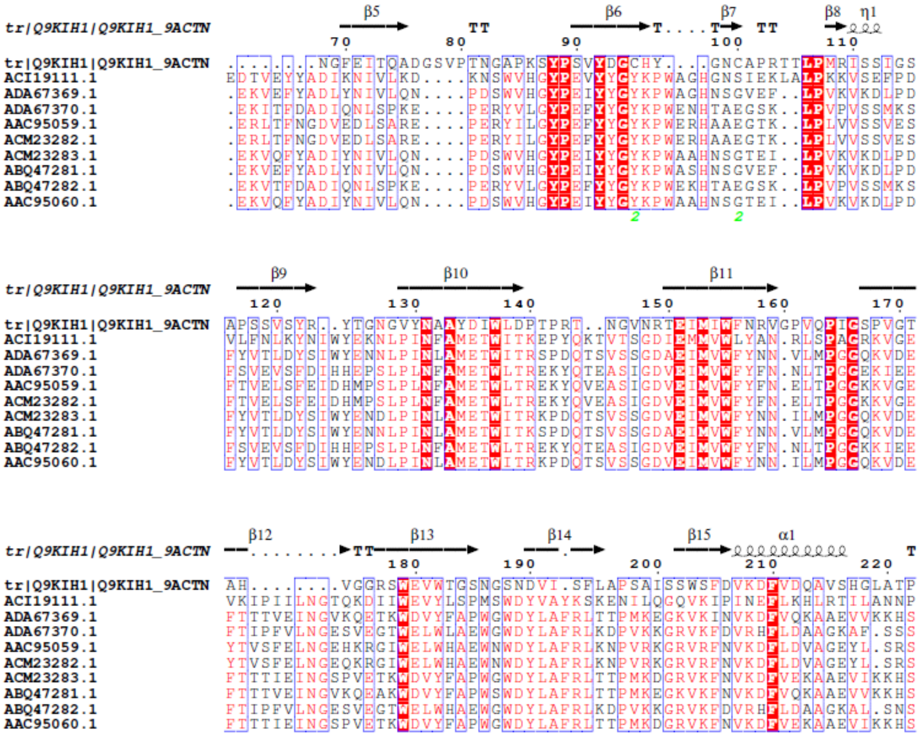
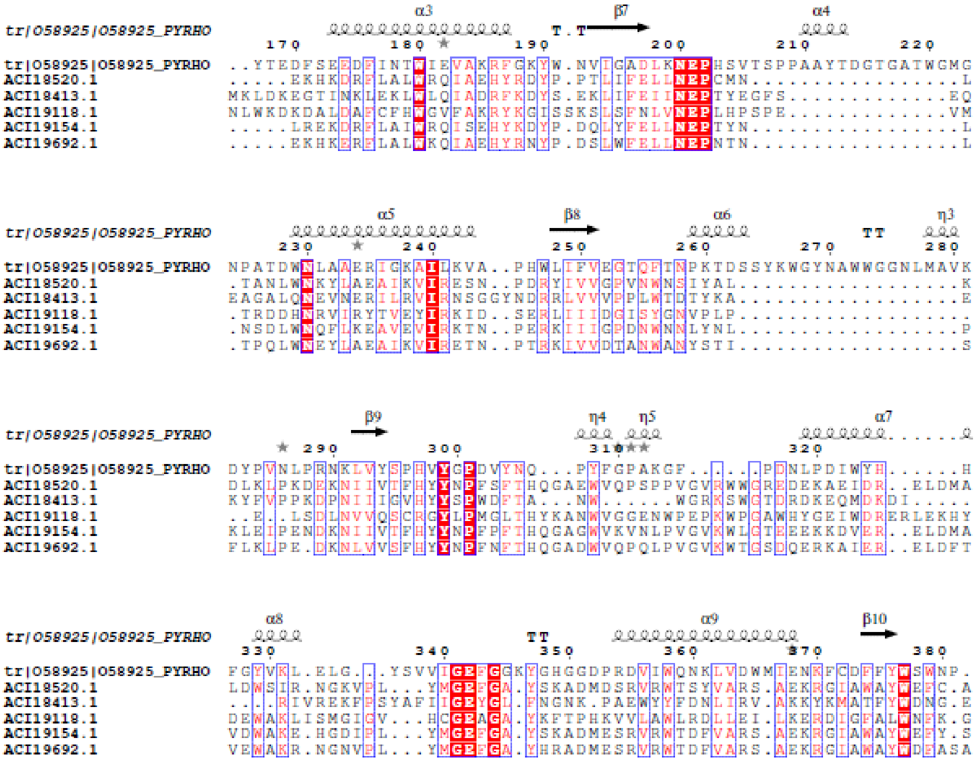

tr|Q9KIH1|Q9KIH1_9ACTN
ACI19111.1
ADA67369.1
ADA67370.1
AAC95059.1
ACM23282.1
ACM23283.1
ABQ47281.1
ABQ47282.1
AAC95060.1

Table 1.7(C): Alignmentwith *Streptomyces sp.* endoglucanase

Table 1.7: GH12 bacterial enzyme sequence alignment and comparison with refrence enzymes

For putative GH 5 archael endoglucanases, greater sequence similarity was observed when comapred with the archael hyperthermophilic endoglucanase as compared to both thermophilic as well as mesophilic bacterial endoglucanase.

For both putative GH5 bacterial **(Table 1.8)** and GH12 archael endoglucanasesthe results remained inconclusive as the similarity between the sequences under any comparison was observed to be low.



```
tr|O58925|O58925_PYRHO  ..YTEDFSEEDFINTWIEVAKRFGKYW.NVIGADLKNEPHSVTSPPAAYTDGTGATWGMG
ACI18520.1              .....EKHKDRFLALWRQIAEHYRDYP.PTLIFELLNEPCMN.................L
ACI18413.1              MKLDKEGTINKLEKLWLQIADRFKDYS.EKLIFEIINEPTYEGFS.............EQ
ACI19118.1              NLWKDKDALDAFCFHWGVFAKRYKGISSKSLSFNLVNEPLHPSPE.............VM
ACI19154.1              .....LREKDRFLAIWRQISEHYKDYP.DQLYFELLNEPTYN.................L
ACI19692.1              .....EKHKERFLALWKQIAEHYRNYP.DSLWFELLNEPNTN.................L

tr|O58925|O58925_PYRHO  NPATDWNLAAERIGKAILKVA..PHWLIFVEGTQFTNPKTDSSYKWGYNAWWGGNLMAVK
ACI18520.1              .TANLWNKYLAEAIKVIRESN..PDRYIVVGPVNWNSIYAL..................K
ACI18413.1              EAGALQNEVNERILRVIRNSGGYNDRRLVVVPPLWTDTYKA..................E
ACI19118.1              .TRDDHNRVIRYTVEYIRKID..SERLIIIDGISYGNVPLP...................
ACI19154.1              .NSDLWNQFLKEAVEVIRKTN..PERKIIIGPDNWNNLYNL..................P
ACI19692.1              .TPQLWNEYLAEAIKVIRETN..PTRKIVVDTANWANYSTI..................S

tr|O58925|O58925_PYRHO  DYPVNLPRNKLVYSPHVYGPDVYNQ...PYFGPAKGF......PDNLPDIWYH......H
ACI18520.1              DLKLPKDEKNIIVTFHYYNPFSFTHQGAEWVQPSPPVGVRWWGREDEKAEIDR..ELDMA
ACI18413.1              KYFVPPKDPNIIIGVHYYSPWDFTA...NW......WGRKSWGTDRDKEQMDKDI.....
ACI19118.1              ..E..LSDLNVVQSCRGYLPMGLTHYKANWVGGENWPEPKWPGAWHYGEIWDRERLEKHY
ACI19154.1              KLEIPENDKNIIVTFHYYNPFPFTHQGAGWVKVNLPVGVKWLGTEEEKKDVER..ELDMA
ACI19692.1              FLKLPE.DKNLVVSFHYYNPFNFTHQGADWVQPQLPVGVKWTGSDQERKAIER..ELDFT

tr|O58925|O58925_PYRHO  FGYVKL.ELG...YSVVIGEFGGKYGHGGDPRDVIWQNKLVDWMIENKFCDFFYWSWNP.
ACI18520.1              LDWSIR.NGKVPL...YMGEFGA.YSKADMDSRVRWTSYVARS.AEKRGIAWAYWEFC.A
ACI18413.1              ....RIVREKFPSYAFIIGEYGL.FNGNK.PAEWYYFDNLIRV.AKKYKMATFYWDNG.E
ACI19118.1              DEWAKLISMGIGV...HCGEAGA.YKFTPHKVVLAWLRDLLEI.LKERDIGFALWNFK.G
ACI19154.1              VDWAKE.HGDIPL...YMGEFGA.YSKADMESRVRWTDFVARS.AEKRGIAWAYWEFY.S
ACI19692.1              VEWAKR.NGNVPL...YMGEFGA.YHRADMESRVRWTDFVARS.AEKRGIAWAYWDFASA
```

Table 1.8(A): Alignment with *P.horikoshii* endoglucanase

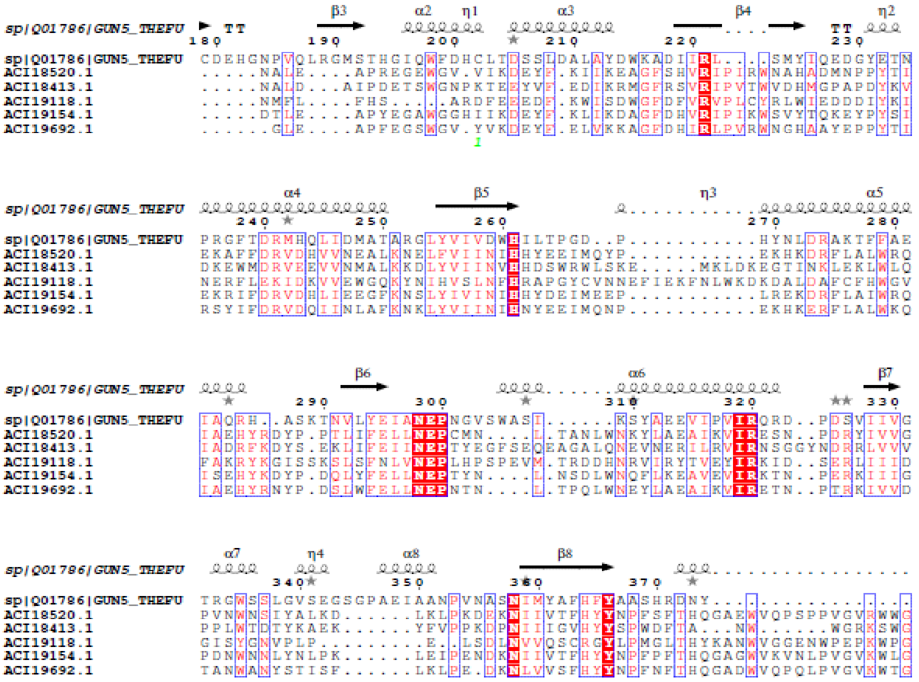


Table 1.8(B): Alignment with *T.fusca* endoglucanase

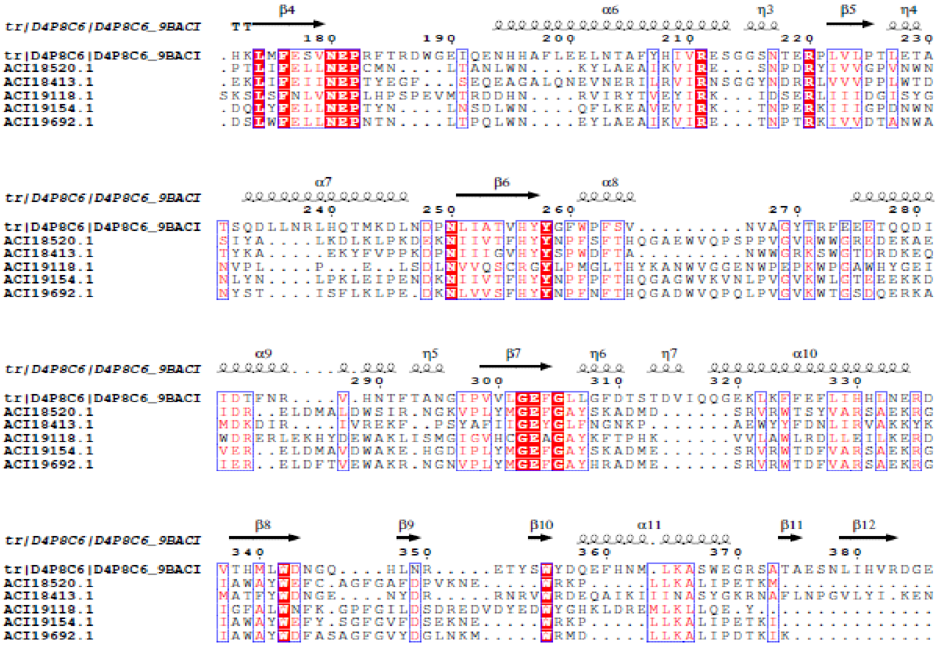


Table 1.8(C) : Alignment with *Bacillus sp.* Endoglucanase

Table 1.8 : GH5 bacterial enzyme sequence alignment and comparison with refrence enzymes

iv. Concluding remarks and Outlook

As seen in the case of two GH5 archael and GH12 bacterial endoglucanases, sequence identity and similarity had high values from which it was safely assumed that these enzymes would work as thermostable endoglucanases. To get a more lucid picture, structure of these endoglucanases can be predicted (ab-initio as well as based on homology) and can aligned to see the similarities between the different endoglucanases.

For the case of GH5 bacterial and GH12 archael endoglucanases, the identity and similarity amongst the sequences were low. As selection of the sequences to be analysed were done in small numbers and based on availability, it still remains inconclusive where these enzymes are thermostable or otherwise. Moreover, well characterised and studied endoglucanases were chosen as reference enzymes which, by a chance event, could have been possibly closer in terms of ancestry to GH5 archael enzymes and GH12 bacterial enzymes than others which gave such a result output. In order to get a more productive output, more endoglucanases should be incorporated into the list and comparisons with other well studied endoglucanases can be carried out.

Cellulases from *Clostridium thermocellum*

A comprehensive search was done using available literature and CAZy database for all the cellulases present in the genome of *Clostridium thermocellum.* It was also available at ATCC under the name *Hungateiclostridium thermocellum strain ATCC 27405.* All three different kinds of cellulases, with or without dockerin and CBM were listed **(Table 1.9)** (Hirano et al., 2016; Katayeva, Golovchenko, Chuvilskaya, & Akimenko, 1992; Lombard et al., 2014).

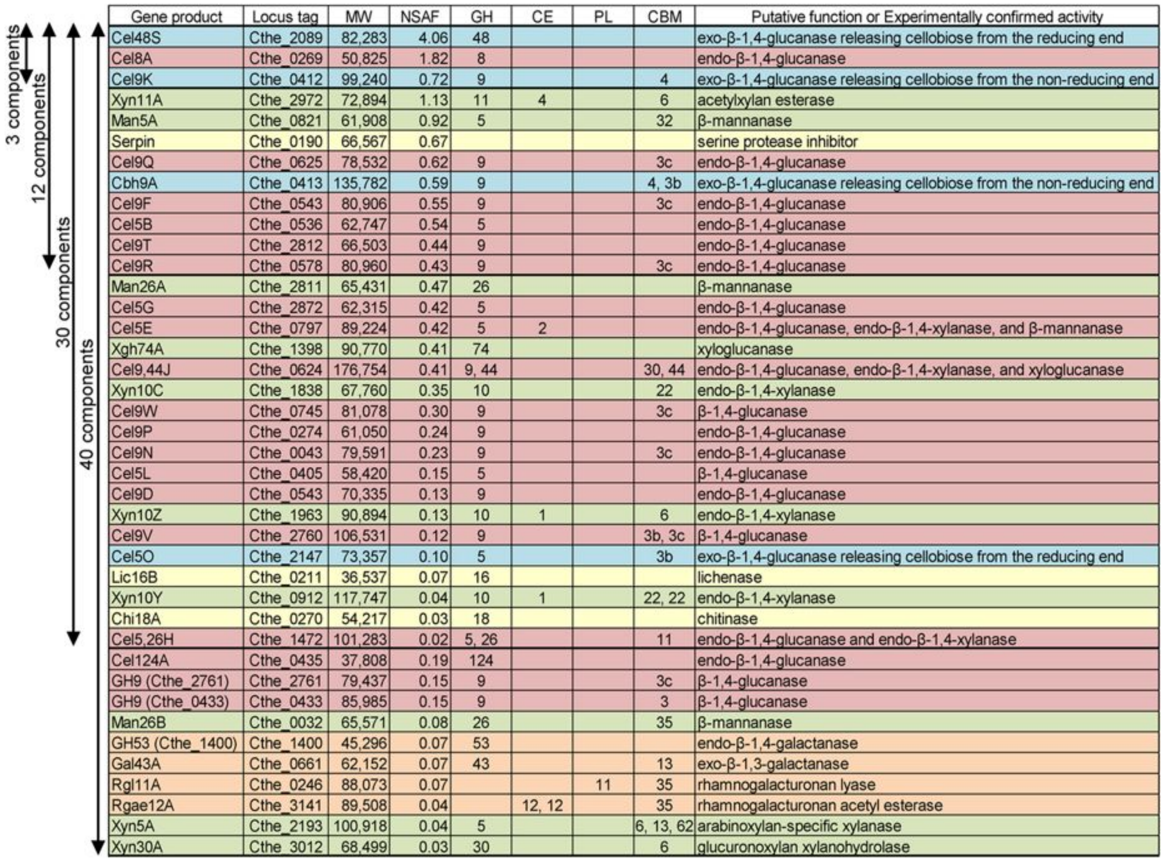
Gene Name	Type Endoglucanase(Endo)/Exocellulase (Exo)/ β-glucosidase(B)	GenBank ID	UniProt ID	PDB ID	No. of Amino acids	Dockerin position	CBM Type/Position	GH family
CelA	Endo	ABN51508.1	A3DC29	1CEM	477	411-477	-	8
CelB	Endo	ABN51772.1	P04956	NA	563	496-562	-	5
CelF	Endo	CAA43035.1	P26224	NA	739	664-737	3c/ 480-639	9
CelG	Endo	ABN54070.1	Q05332	NA	566	497-564	-	5
CelQ	Endo	ABN51860.1	A3DD31	NA	710	640-709	3c/ 468-627	9
CelR	Endo	ABN51814.1	A3DCY5	NA	736	661-734	3c/ 480-638	9
CelT	Endo	ABN54011.1	A3DJ82	2YIK	611	543-611	-	9
CelH	Endo	ABN52701.1	P16218	2BVD	900	827-900	11/ 655 – 900	26, 5
CtCel 124	Endo	ABN51673.1	A3DCJ4	4DH2	350	33-103	-	124
CelI	Endo	ABN51281.1	Q02934	2XFG	887	-	3/ 529 –684 736 –887	9
CelY	Endo	ABN51312.1	A3DBI3	NA	938	-	3/ 787 –938	48
Cel9 W	Endo	ABN51980.1	A3DDF1	NA	730	661-728	3c/ 482 – 641	9
Cel9	Endo	ABN51859.1	A3DD30	2EJ1	1601	1286 - 1352	-	9

Cel5E26	Endo	AAA23224.1	P10477	4IM4	814	409–479	-	5
Cthe_2761	Endo	ABN53960.1	A3DJ31	NA	707	635–705	3c/ 473 – 623	9
Cthe_0433	Endo	ABN51671.1	A3DCJ2	NA	789	720 –788	3/539-701	9
CelD	Endo	CAA28255.1	A3DDN1	NA	649	579 –649	-	9
Cel9P	Endo	ABN51513.1	A3DC34	NA	563	476 –553	-	9
Cel9N28	Endo	ABN51284.1	A3DBF5	NA	742	675 –742	3c/ 496 – 658	9
Cel5L	Endo	ABN51643.1	A3DCG4	NA	526	460 –526	-	5
Cel9V	Endo	ABN53959.1	A3DJ30	2WOB	961	893 –961	3b/ 560-720 3c/ 738-890	9
CbhA	Exo	ABN51651.1	A3DCH2	3PDD	1224	1157 – 1223	4/ 1001-1147	9
CelS	Exo	ABN53296.1	A3DH67	NA	741	673 –739	-	48
CelK	Exo	ABN51650.1	A3DCH1	NA	895	828 –894	4/40 – 199	9
Cel5O	Exo	ABN53350.1	A3DHC1	NA	660	584 –654	3/40 – 186	5
bglA	B	CAA42814.1	P26208	5OGZ	448	NA	NA	1
bglB	B	CAA33665.1	P14002	-	755	NA	NA	3

Table 1.9: List of cellulases from *Clostridium thermocellum*

Some of these cellulases had exceptional property such as CelH could function as both endoglucanase and xylanase and CelY was supposedly a transmembrane cellulase.

In a study on the cellulosomes of *Clostridium thermocellum*, the enzymatic composition of these complexes had been reported in terms of relative ratios of normalized spectral abundance factors (NSAF) of each cellulosomal component **(Figure 1.8)** (Hirano et al., 2016).


3 components
12 components
30 components
40 components

Gene product	Locus tag	MW	NSAF	GH	CE	PL	CBM	Putative function or Experimentally confirmed activity
Cel48S	Cthe_2089	82,283	4.06	48				exo-β-1,4-glucanase releasing cellobiose from the reducing end
Cel8A	Cthe_0269	50,825	1.82	8				endo-β-1,4-glucanase
Cel9K	Cthe_0412	99,240	0.72	9			4	exo-β-1,4-glucanase releasing cellobiose from the non-reducing end
Xyn11A	Cthe_2972	72,894	1.13	11	4		6	acetylxylan esterase
Man5A	Cthe_0821	61,908	0.92	5			32	β-mannanase
Serpin	Cthe_0190	66,567	0.67					serine protease inhibitor
Cel9Q	Cthe_0625	78,532	0.62	9			3c	endo-β-1,4-glucanase
Cbh9A	Cthe_0413	135,782	0.59	9			4, 3b	exo-β-1,4-glucanase releasing cellobiose from the non-reducing end
Cel9F	Cthe_0543	80,906	0.55	9			3c	endo-β-1,4-glucanase
Cel5B	Cthe_0536	62,747	0.54	5				endo-β-1,4-glucanase
Cel9T	Cthe_2812	66,503	0.44	9				endo-β-1,4-glucanase
Cel9R	Cthe_0578	80,960	0.43	9			3c	endo-β-1,4-glucanase
Man26A	Cthe_2811	65,431	0.47	26				β-mannanase
Cel5G	Cthe_2872	62,315	0.42	5				endo-β-1,4-glucanase
Cel5E	Cthe_0797	89,224	0.42	5	2			endo-β-1,4-glucanase, endo-β-1,4-xylanase, and β-mannanase
Xgh74A	Cthe_1398	90,770	0.41	74				xyloglucanase
Cel9,44J	Cthe_0624	176,754	0.41	9, 44			30, 44	endo-β-1,4-glucanase, endo-β-1,4-xylanase, and xyloglucanase
Xyn10C	Cthe_1838	67,760	0.35	10			22	endo-β-1,4-xylanase
Cel9W	Cthe_0745	81,078	0.30	9			3c	β-1,4-glucanase
Cel9P	Cthe_0274	61,050	0.24	9				endo-β-1,4-glucanase
Cel9N	Cthe_0043	79,591	0.23	9			3c	endo-β-1,4-glucanase
Cel5L	Cthe_0405	58,420	0.15	5				β-1,4-glucanase
Cel9D	Cthe_0543	70,335	0.13	9				endo-β-1,4-glucanase
Xyn10Z	Cthe_1963	90,894	0.13	10	1		6	endo-β-1,4-xylanase
Cel9V	Cthe_2760	106,531	0.12	9			3b, 3c	β-1,4-glucanase
Cel5O	Cthe_2147	73,357	0.10	5			3b	exo-β-1,4-glucanase releasing cellobiose from the reducing end
Lic16B	Cthe_0211	36,537	0.07	16				lichenase
Xyn10Y	Cthe_0912	117,747	0.04	10	1		22, 22	endo-β-1,4-xylanase
Chi18A	Cthe_0270	54,217	0.03	18				chitinase
Cel5,26H	Cthe_1472	101,283	0.02	5, 26			11	endo-β-1,4-glucanase and endo-β-1,4-xylanase
Cel124A	Cthe_0435	37,808	0.19	124				endo-β-1,4-glucanase
GH9 (Cthe_2761)	Cthe_2761	79,437	0.15	9			3c	β-1,4-glucanase
GH9 (Cthe_0433)	Cthe_0433	85,985	0.15	9			3	β-1,4-glucanase
Man26B	Cthe_0032	65,571	0.08	26			35	β-mannanase
GH53 (Cthe_1400)	Cthe_1400	45,296	0.07	53				endo-β-1,4-galactanase
Gal43A	Cthe_0661	62,152	0.07	43			13	exo-β-1,3-galactanase
Rgl11A	Cthe_0246	88,073	0.07			11	35	rhamnogalacturonan lyase
Rgae12A	Cthe_3141	89,508	0.04		12, 12		35	rhamnogalacturonan acetyl esterase
Xyn5A	Cthe_2193	100,918	0.04	5			6, 13, 62	arabinoxylan-specific xylanase
Xyn30A	Cthe_3012	68,499	0.03	30			6	glucuronoxylan xylanohydrolase

Figure 1.8: List of cellulolytic enzymes in *Clostrdium thermocellum* with NSAF values (Hirano et al., 2016)

The aim was to shortlist some of these cellulases in order to produce potent cellulases of all the three types as well as facilitate the cohesion-dockerin interaction studies in the lab. Factors that were taken into account while selection of the cellulases were variability (multiples of all three kinds of cellulases to be included), presence of dockerin and/or CBM, high NSAF and exceptional property. The selected cellulases are enlisted in **Table 1.10** below as well as highlighted in grey in **Table 1.9.**

Cellulase	Special Property
Cel A	Highest NSAF amongst endoglucanases, CBM absent
Cel F	Endoglucanase with both CBM and dockerin
Cel Q	Endoglucanase with both CBM and dockerin
Cel R	Endoglucanase with both CBM and dockerin

Cel H	Had both endoglucanase and xylanase activity
Cel I	Endoglucaanse with two CBM, dockerin absent
Cbh A	Exocellulase with both CBM and dockerin
Cel S	Highest NSAF amongst all cellulosomal enzymes, exocellulase without CBM
bgl A	β-glucosidase
Bgl B	β-glucosidase

Table 1.10: Special properties of selected cellulases

Concluding remarks

A total of 10 cellulases – 6 endoglucanases, 2 exocellulases and 2 β-glucosidases - were shortlisted to heterologously express in *Escherichia coli* in order to make enzyme cocktails for biomass degradation by checking their activity as well as to understand different cohesion-dockerin interactions.

1.3 Cloning of potent cellulases from *Clostridium thermocellum*

1.3.1 MATERIALS AND METHODS

1.3.1.1 Materials

(i) *Clostridium thermocellum* growth medium

RCB medium (Reinforced Clostridial Broth)

Components	**Amount**
Pancreatic Digest of Casein	10.00 g
Beef Extract	10.00 g
Yeast Extract	3.00 g
Dextrose	5.00 g
Sodium Chloride	5.00 g
Soluble Starch	1.00 g
Sodium Acetate	3.00 g
Resazurin (0.025% solution)	4.00 mL
L-Cysteine Hydrochloride (25.0% solution)	2.00 mL
DI Water	Upto 1.00 L

Manual medium (ATCC 1190 medium)

Components	**Amount**
Potassium dihydrogen Phosphate (KH_2PO_4)	1.5 g/L
Disodium Phosphate($Na_2HPO_4.12H_2O$)	4.2 g/L
Ammonium Chloride (NH_4Cl)	0.5 g/L
Magnesium Chloride ($MgCl_2.6H_2O$)	0.18 g/L
Yeast Extract	2.0 g/L
Glucose	8.0 g/L
Vitamin Solution	0.5 ml/L
Wolfe's Modified Mineral Elixir	5.0 ml/L
Resazurin (0.1%)	1.0 ml/L
Reducing Solution	40.0 ml/L
Distilled deionized water	Upto 1.0 L

Reducing Solution

Components	**Amount**
Sodium Hydroxide (NaOH)(0.2 N)	200.0 ml
Sodium sulphide ($Na_2S.9H_2O$)	2.5 g
L-Cysteine.HCl	2.5 g

Vitamin Solution

Components	**Amount**
Biotin	20.0 mg
p-Aminobenzoic acid	50.0 mg
Folic acid	20.0 mg
Pantothenic acid calcium salt	50.0 mg
Nicotinic acid	50.0 mg
Vitamin B12	1.0 mg
Thiamine. HCl	5.0 mg
Pyridoxine hydrochloride	100.0 mg
Thioctic acid	50.0 mg
Riboflavin	5.0 mg
Distilled Water	500 ml

Wolfe's Modified Mineral Elixir

Components	**Amount**
Nitrilotriacetic acid	1.5 g
Magnesium Sulphate ($MgSO_4.7H_2O$)	3.0 g
Manganese Sulphate ($MnSO_4.H_2O$)	500.0 mg
Sodium Chloride (NaCl)	1.0 g

Ferrous sulphate ($FeSO_4.7H_2O$)	100.0 mg
Cobalt nitrate ($Co(NO_3)_2.6H_2O$)	100.0 mg
Calcium Chloride ($CaCl_2$)	100.0 mg
Zinc Sulphate ($ZnSO_4.7H_2O$)	100.0 mg
Copper Sulphate ($CuSO_4.5H_2O$)	10.0 mg
Aluminium Potassium Sulphate ($AlK(SO_4)_2$)	10.0 mg
Boric acid	10.0 mg
Sodium Molybdate ($Na_2MoO_4.2H_2O$)	10.0 mg
Sodium Selenite (Na_2SeO_3)	1.0 mg
Distilled water	Upto 1.0 L

(ii) Glycerol stock

Components	**Amount**
Glycerol	60 ml (60%)

Volume was made upto 100 ml by adding distilled water, and the mix was autoclaved for 15 minutes at 121˚C under 15 psi.

(iii) Genomic DNA (gDNA) isolation

For isolation of genomic DNA from the cultured *Clostridium thermocellum* cells, Quick-DNA™ Fungal/Bacterial Miniprep Kit from Zymo Research was used.

(iv) Polymerase Chain Reaction (PCR)

Primers

Stock concentration	**Working concentration**
10µM	0.2µM

dNTPs

Stock concentration	**Working concentration**
10mM	200µM

Polymerase and their buffers

Enzyme/Buffer		**Stock Concentration**	**Working Concentration**
Taq Ayonex	Buffer	5X	1X
	Polymerase	-	0.1 µl in 20 µl reaction
VegaPol	Buffer	5X	1X
	Polymerase	-	0.2 µl in 20 µl reaction
Deep Vent	Buffer	5X	1X
	Polymerase	-	0.1 µl in 20 µl reaction
Taq Flexi	Buffer	5X	1X
	Polymerase		0.15µl in 20 µl reaction

(v) DNA Agarose Gel Electrophoresis

50X TAE Buffer

Component	**Amount**
Tris-Cl	242g
Glacial acetic acid	57.1 mL
0.5M EDTA (pH 8.0)	100 mL

0.8% Agarose Gel

Component	**Amount**
Agarose	0.8 gm
1x TAE buffer	100 ml

Ethidium Bromide Stock Solution (1% w/v)

Component	**Amount**
Ethidium bromide	0.1g
Deionized Water	10ml

6X DNA Gel Loading Buffer

Component	**Amount**
Bromophenol Blue	0.25%
Glycerol	30%

DNA Ladder

3µl of 1Kb DNA ladder(with 1x DNA loading dye) was used.

(vi) Cloning

Gel Extraction and PCR clean up

For isolation DNA fragments excised from agarose gels, QIAquick® Gel Extraction Kit QIAquick® PCR & Gel Clean-up Kit from Qiagen was used.

LB media

Component	**Amount for 1 Litre (g)**
NaCl	10g
Tryptone	10g
Yeast Extract	5g

Components were dissolved in 1 litre distilled water and then autoclaved at 15 psi and 121˚C for 15 minutes.

LB-Agar Plates

Component	**Amount for 1 Litre (g)**
NaCl	10g
Tryptone	10g
Yeast Extract	5g
Agar	20g

Components were dissolved in 1 litre distilled water and then autoclaved at 15 psi and 121˚C for 15 minutes. After it cooled down to lukewarm state, desired antibiotic amount was added and the media was poured into agar plates under sterile conditions.

Antibiotics

Antibiotic	**Concentration of stock (1000X)**
Ampicillin	100 mg/mL in water
Chloramphenicol	35mg/mL in methanol
Kanamycin	25mg/mL in water
Tetracycline	12.5mg/mL in 70% ethanol

Antibiotics were filtered using a syringe driven filter (0.22 µm) and stored as aliquots at -20°C. The working concentration of antibiotics was 1X.

Plasmid isolation

For isolation of plasmid in order to use as vector in the cloning process, QIAprep Spin Miniprep Kit from Qiagen was used.

Plasmid	**Restriction sites used**	**Gratefully sourced from**
pET23a:E1-E2-E3-E4	Nde I, Xho I	Arpita Mrigwani
pET23a:N1-N2-N3-N4	Nhe I, Xho I	

Restriction Digestion

For restriction digestion, Fast Digest restriction enzymes (NdeI, NheI, XhoI) and buffer from ThermoFischer were used.

Ligation

For ligation, T4 DNA ligase and buffer were used.

Competent cells Preparation

$CaCl_2$ Buffer

Component	Composition
$CaCl_2$ (60mM)	1.764g
PIPES (10mM)	0.605g
Glycerol (15%)	30ml

Volume was made upto 200 ml with distilled water and components were mixed well. The pH was set to 7. The solution was autoclaved for 15 minutes at 15 psi and 121˚C.

Bacterial Strain Used

Strain	Antibiotic resistance
XL1-Blue	Tetracycline

1.3.1.2 Methods

(A). Growing *Clostridium thermocellum*

i. Media preparation

RCB medium

All components were mixed accordingly in to make 200 ml of media. The mixture was transferred to a glass serum bottle and boiled for about 15-20 minutes (to decrease solubility of oxygen in liquid phase) and flushed with N_2 gas for 5 minutes (to flush out remaining oxygen). The bottle is capped with isobutyl rubber cap and sealed with aluminium cap with a crimper. Then the medium was autoclaved at 15 psi in 121˚C for 15 minutes.

ATCC Manual Medium

Preparation of components

a. Reducing Solution - For making the reducing solution was the 200 ml of sodium hydroxide solution was boiled for 20 minutes and purge with N_2 gas for 5 minutes in a serum bottle. Then the mixture was allowed to cool down and sodium sulphide and cysteine was added accordingly. The serum bottle was immediately sealed with isobutyl rubber cap and aluminium cap. The solution was then autoclaved at 15 psi in 121˚C for 15 minutes.

b. Vitamin Solution – The vitamins were added to 500 ml of distilled autoclaved water in a laminar hood and mixed well. The solution was stored in a 4˚C refrigerator away from light.

c. Wolfe's Modified Mineral Elixir – Designated amount of nitriloacetic acid was added to 500 ml of autoclaved distilled water (dissolved using titration with 2N Sodium hydroxide until pH of 6.5 was achieved). The rest of the ingredients were added after this and the volume was made up to 1 litre. The mixture was stored in the 4˚C refrigerator.

d. Preparation of the manual media

200 ml of manual media was prepared. Accordingly calculations were done and solid components were added to 100 ml of distilled water along with resazurin. The mixture was boiled for 20 minutes and purge with N_2 gas for 5 minutes in a serum bottle. Further, it was sealed with isobutyl rubber cap and aluminium cap. The mixture was autoclaved at 15 psi in 121˚C for 15 minutes.

After the solution cooled, the calculated amounts reducing solution, vitamin solution and mineral solution were added under sterile and anaerobic conditions using a syringe.

The medium was mixed well and left overnight to check for contamination.

Resazurin is oxygen indicator in the media. Before boiling the media should be blue in colour (indicating presence of oxygen), after boiling it should turn pinkish in colour (indicating lower levels of oxygen) and after autoclaving and adding reducing solution it should turn yellow (indicating almost anoxic conditions). **(Figure 1.9)**

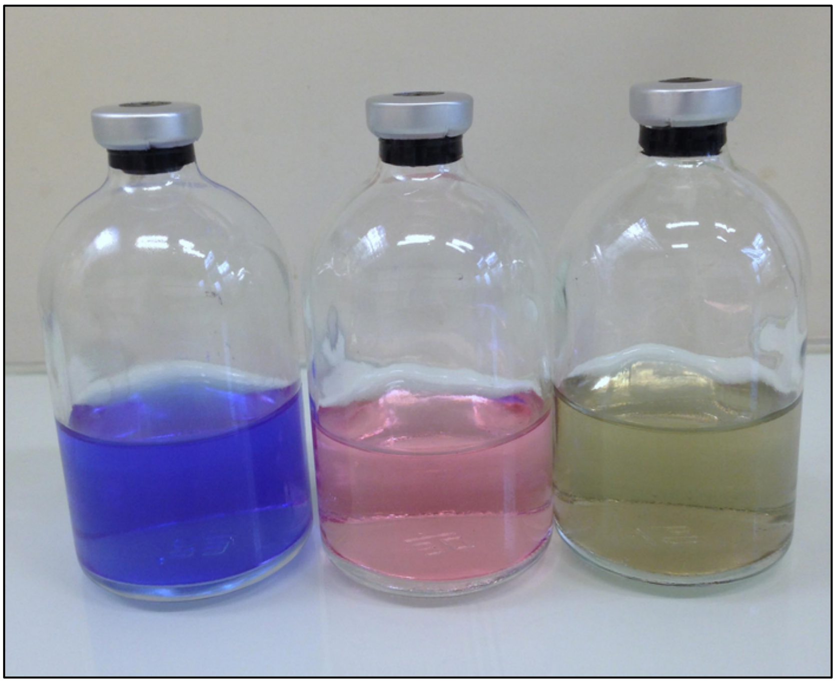


Figure 1.9: Resazurin oxygen level indications – Blue (high), pink (medium), and yellow (low) in growth medium (Wagner et al., 2019)

ii. Inoculation and culturing

An inoculation of 1% of that of the volume of the medium was added to the medium under sterile and anaerobic conditions using a syringe. The inoculated media was kept in an incubator set at 50˚C without shaking for growth of *Clostridium thermocellum.*

(B). Genomic DNA (gDNA) isolation from cultured *Clostridium thermocellum* cells

i. The cultured cells were centrifuged at 5000 x g for 10 minutes in falcon tubes

ii. The cell pellet was resuspended in 200 µl of 1X PBS. This mixture was added to a ZR Bashing Bead lysis tube. 750 µl of Bashing bead buffer was added to the tube and the mixture was mixed well using a vortex for 5-7 minutes.

iii. The Bashing Bead lysis tube was centrifuges at 10,000 x g for 60 seconds.
v. The supernatant was transferred to a Zymo-Spin III-F filter in a collection tube and was centrifuged at 10,000 x g for 60 seconds.
vi. 1200 µl of Genomic lysis buffer was added to the flowthrough.

vi. This mixture was transferred to a Zymo-Spin IICR column in a collection tube and was centrifuged at 10,000 x g for 60 seconds. After discarding the flowthrough, the same step was repeated.

vii. 200 µl of DNA Pre-Wash buffer was added to the Zymo-Spin IICR column and centrifuged at 10,000 x g for 60 seconds. Flow through was discarded.

viii. 500 µl of gDNA Wash buffer was added Zymo-Spin IICR column and centrifuged at 10,000 x g for 60 seconds.

ix. The Zymo-Spin IICR column was transferred to a new 1.5 ml microcentrifuge tube and 40 µl of DNA Elution buffer was added to centre of the column and left at room temperature for 5 mins. After that, it was centrifuged at 10,000 x g for 60 seconds to elute the gDNA.

x. Nano-drop measurements were taken to estimate the concentration of isolated gDNA.

(C). Polymerase Chain Reaction (PCR)

PCR is a molecular biology technique which is used to amplify a short stretch of DNA in large quantities from a template.

It consists of three steps:

i. Denaturation - The template DNA is denatured at a very high temperature and separate into single strands

ii. Annealing - Forward and reverse primers complementary to both strands of desired DNA stretch anneal at 3' ends at a favourable temperature.

iii. Extension – A thermostable DNA polymerase sits on the primer-DNA complex and synthesises new DNA strands in 5'→3' using the older DNA strand as template.

These three steps are repeated over 30-35 cycles to get multiple copies of the template DNA strand. **(Figure 1.10)**

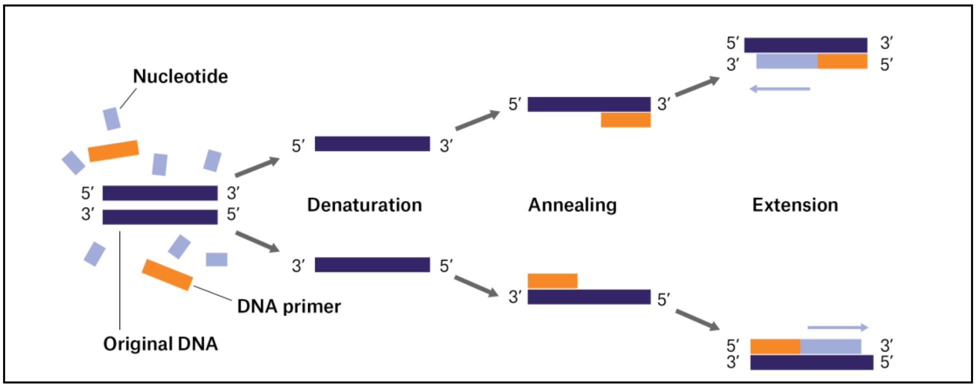


Figure 1.10: Schematic representation of PCR cycle (Alexander Goldberg, 2019)

PCR amplification was used to amplify the different cellulases genes from *Clostridium thermocellum*. As standardisation of the PCR amplification was an aim of the thesis, apart from dNTPs, primers and enzyme concentration, the concentrations of gDNA template and type of polymerase used was varied to get the optimum conditions for gene amplification.

PCR Program Set Up –

Different programs were used for different polymerases. The annealing temperature and the extension were varied according to primers used and gene length to be amplified, respectively **(Table 1.11).**

Enzyme	Step	Temperature(˚C)	Time
TaqAyonex / DeepVent	Initial Denaturation	95	5 minutes
	Denaturation	95	30 seconds
	Annealing x35	Variable	45 seconds
	Extension	72	Variable
	Final Extension	72	10 minutes
VegaPol	Initial Denaturation	98	2 minutes
	Denaturation	98	30 seconds
	Annealing x35	Variable	20 seconds
	Extension	72	Variable
	Final Extension	72	5 minutes

Table 1.11: PCR program for different polymerases

(D). DNA Gel Electrophoresis

It is a molecular biology technique used to separate out a mixture of DNA based on its length. A current pulls the negatively charged DNA towards the positive electrode. The DNA strands get separated on the matrix of agarose gel.

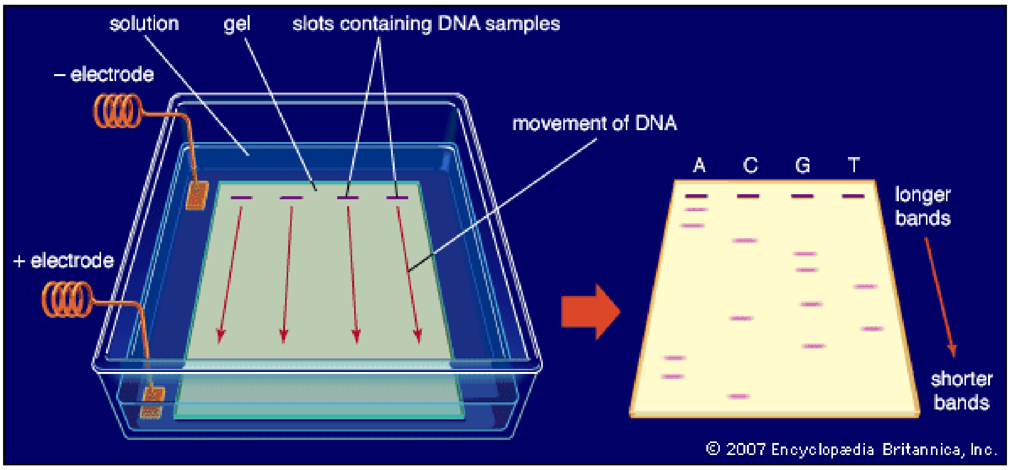

solution
gel
slots containing DNA samples
– electrode
+ electrode
movement of DNA
A
C
G
T
longer bands
shorter bands
© 2007 Encyclopædia Britannica, Inc.

Figure1.11: Concept of Gel Electrophoresis (Rogers, April 28, 2017)

The protocol for DNA agarose gel electrophoresis is

i. 1000ml 1X TAE buffer was made from 50X TAE buffer by adding 20 ml of it to distilled deionised water.

ii. 0.8g of agarose was weighed and added to 100 ml of 1X TAE buffer in a conical flask. The solution was boiled for 2 minutes in a microwave.

iii. 2µl of EtBr (with gloved hands and caution) was added to the solution after it cooled down a little.

iv. The solution was poured into an assembled gel apparatus and a comb was fixed to make wells. Formation of bubbles was prevented.

v. After the gel solidifies, the comb was removed, and the gel was placed in the electrophoresis tank filled with 1X TAE Buffer.

vi. To prepare the samples, a specified amount of 6X DNA loading dye was added to them, to make the final concentration of loading dye to 1X.

vii. An appropriate DNA ladder (3µl of 10 kb DNA ladder) was added to one of the wells for reference.

viii. The gel was run at 60V until the DNA bands separated and was visualised under UV light.

(E.) Cloning

Cloning is a molecular biology technique where in a gene of interest is recombinantly added to a vector plasmid of choice. A host organism is then used to get multiple copies of this recombinant DNA.

pET-23a was chosen as vector to clone the cellulase genes. It has an inducible promoter and 6x Histidine tag at the 3’ end of the MCS, both which would make it easier for the expression and purification of the recombinant proteins/enzymes (**Figure 1.12)**. The vectors used already had an insert, pET23a:E1-E2-E3-E4 and pET23a:N1-N2-N3-N4, in order to make verification of digestion by restriction enzymes easier (big fall out expected on digestion of the vector) Restriction enzymes and restriction sites were chosen for each gene based on the absence the restriction site in the gene itself.

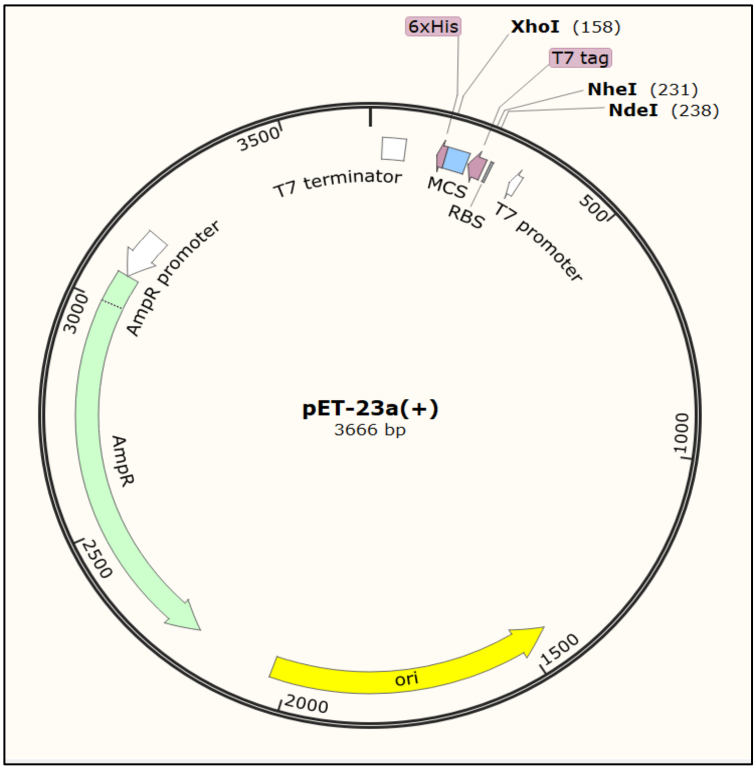

6xHis
XhoI (158)
T7 tag
NheI (231)
NdeI (238)
3500
T7 terminator
MCS
RBS
T7 promoter
500
AmpR promoter
3000
pET-23a(+)
3666 bp
1000
AmpR
2500
1500
ori
2000

Figure 1.12: pET23a Vector map annotated using Snapgene

Primer designing

Primer designing for all the cellulose genes was done manually using Snap gene. All primers were made of appropriate length and GC content. Respective restriction sites and 4-5 bases of overhang (for the restriction enzymes to get enough space to sit on the sequence and cut) were added to the correct position on the primers and some silent mutation were done in the sequence to get the optimum GC content. Primers were checked for primer-dimer formation and secondary structure formation using Gene ruler. The primers were ordered from Sigma-Aldrich and appropriate amount of autoclaved distilled water was added to make primer stock of 100µM concentration. The T_m for each of the primers was used to select different annealing temperature for each pair of primers.

Using the specific primers and standardised PCR conditions, PCR for the genes CelA, CelR and CelQ was carried out and agarose gel electrophoresis was done.

Gel Extraction of DNA fragments from agarose gel

The amplified DNA fragments of CelA, CelQ and CelR were excised from the agarose gel by the following these steps -

i. Desired DNA was excised using a scalpel and put into a 2ml microcentrifuge tube.
ii. The excised gel fragment was weighed and 3 times the volume of Buffer QG was added to 1 volume of fragment (100mg ~100 µl)
iii. Using a heat block, the gel was melted at 50˚C. Once the gel gets completely dissolved, 1 volume of isopropanol was added.
iv. A QIAquick spin column was placed in a 2ml collection tube and the solution was transferred this and centrifuged for 1 minute at 17,900 x g. The flowthrough was discarded.
v. 750 µl of Buffer PE was added to the column and centrifuged at 17,900 x g for 1 minute. The flow through was discarded and the step was repeated.
vi. To elute the DNA, the column was placed in a new microcentrifuge tube and 30-50 µl of warm water was added to the centre of the column. After 5 minutes in room temperature, it was centrifuged for 1 minute at 17,900 x g.

Plasmid isolation

Plasmids to be used as vectors i.e., pET23a:E1-E2-E3-E4 and pET23a:N1-N2-N3-N4 were isolated using the following protocol -

i. A 10 ml overnight LB culture of desired plasmid containing host E. coli cells at 37˚C was set up.
ii. The cells were harvested in a 2ml microcentrifuge tube by centrifuge the culture at 6800 x g for 3 minutes. The supernatant was discarded.
iii. The pelleted cells were resuspended in 250µl Buffer P1.
iv. 250µl of Buffer P2 was added and mixed gently by inverting the tube (care was taken not to prolong this step beyond 5 minutes).
v. 350µl of Buffer N3 was added and was mixed immediately by inverting the tube a few times. This mixture was centrifuged at 17,900 x g for 10 minutes.
vi. 800 µl of the supernatant was transferred very carefully to a QIAprep 2.0 spin column by pipetting and was centrifuged at 17,900 x g for 60 seconds. The flowthrough was discarded.

vii. To wash the column, 750 µl of Buffer PE was added to the column and centrifuged for 1 minute at 17,900 x g. The flowthrough was discarded and the step was repeated.

viii. To elute the plasmid, the column was placed in a fresh 1.5 ml microcentrifuge tube and 30-50 µl of warm autoclaved distilled water was added to the centre of the column and left at room temperature for 5 minutes. Then it was centrifuged at 17,900 x g for 1 minute.

ix. Nano-drop measurements were taken to check the concentration of the isolated plasmid.

Restriction Digestion

Restriction digestion of the DNA fragments is done using restriction enzymes to cut the DNA at specific sites in order to create sticky or blunt ends. When the same restriction enzymes are used to cut both vector and insert, generation of similar sticky ends help in ligation of the two into one recombinant plasmid. Restriction digestion was done for all the inserts and plasmids according to the following table.

Plasmid/Insert	**Volume of DNA (µl)**	**Enzymes and their volume (µl)**	**10X FD Buffer(µl)**	**Distilled Deionised Water(µl)**	**Total(µl)**
CelQ	25	Nde I – 1.5 Xho I – 1.5	5	17	50
CelA	37	Nde I – 1.5 Xho I – 1.5	5	5	50
CelR	36	Nhe I – 1.5 XhoI – 1.5	5	6	50
pET23a:E1-E2-E3-E4	18	Nde I – 2 Xho I – 2	5	23	50
pET23a:N1-N2-N3-N4	18	Nhe I – 2 XhoI – 2	5	23	50

Table 1.12: Reaction components for restriction digestion of inserts and plasmids

Restriction digest reaction was carried out at 37˚C for 4 hrs.

PCR Clean Up

Double Digested inserts (CelA, CelQ, CelR) were cleaned using the following protocol -

i. 5 volumes of Buffer PB was added to 1 volume of PCR mix.
ii. A QIAquick spin column was placed in a 2ml collection tube and the solution was transferred this and centrifuged for 1 minute at 17,900 x g. The flowthrough was discarded.
iii. 750µl of Buffer PE was added to the column and centrifuged at 17,900 x g for 1 minute. The flow through was discarded and the step was repeated.
iv. To elute the DNA, the column was placed in a new microcentrifuge tube and 30-50 µl of warm water was added to the centre of the column. After 5 minutes in room temperature, it was centrifuged for 1 minute at 17,900 x g.

Ligation of inserts and vectors

Ligation was done with compatible set of double digested insert and vector, according to the restriction enzymes used to cut them. The reaction components were calculated in a 1:3 insert to vector ratio using an online tool and calculations were made accordingly ("ligation calculator - insilico.,").

Ligation Reaction	**Components**				
	10X T4 ligase buffer(µl)	**T4 DNA ligase(µl)**	**Amount of Vector (µl)**	**Amount of insert(µl)**	**Total (µl)**
CelA+pET23a:E1-E2-E3-E4	1	0.5	1.85	7.5	10.85
CelQ+pET23a:E1-E2-E3-E4	1.6	0.8	1.85	11.75	16
CelR+pET23a:N1-N2-N3-N4	2	1	3.1	14	20.1

Table 1.13: Reaction components for ligation of respective vectors with inserts

The reaction was carried out overnight at 25˚C.

Competent cells Preparation

In order to get multiple copies of the recombinant plasmid, it needs to be put into a cloning host. Cloning host used in this case was XL1-Blue strain of Escherichia coli which has tetracycline resistance. The XL1-Blue cells were made chemically competent to take up DNA using $CaCl_2$ method.

i. A primary culture of 5 ml of the strain of bacterial cells to be made competent was setup and grown at 37˚C overnight.
ii. Next, a 200 ml of secondary culture was set up from the primary culture. The culture was made to grow upto an OD of 0.4 after which the culture was aliquotted into pre-chilled autoclaved falcon tubes and left on ice for 10 minutes.
iii. The cells were harvested at 1600 x g and 4˚C for 7 minutes.
iv. The pellet was resuspended gently in 40 ml of chilled $CaCl_2$ buffer.
v. The cells were again harvested at 1100 x g and 4˚C for 5 minutes after which the pellet was resuspended gently in 20 ml of chilled $CaCl_2$buffer.The cells were left on ice for 30 mins.
vi. Next the cells were again pelleted 1100 x g and 4˚C for 5 minutes.
vii. The pellet was resuspended gently in 1.5 ml of chilled $CaCl_2$ buffer.
viii. After thorough, the aliquots of 100µl was made in separate autoclaved 1.5 ml microcentrifuge tubes using a wide-mouth pipette tip. The competent cells were stored in -80˚C fridge.

Transformation and Plating

Transformation is the process by which competent bacterial cells take up foreign DNA. The three ligation mixes were transformed into chemically competent XL1-Blue cells using the following protocol.

i. Total amount of ligation mix was added to 100 µl aliquot of competent cells and was kept on ice for 20 minutes.
ii. Using a water bath, a heat shock for 90 seconds at 42˚C was given to the cells.
iii. The mixture was kept on ice for 10 minutes.
iv. Under sterile conditions, 1 ml of LB media was added to the cells and the mixture was incubated at 37˚C with shaking to rejuvenate cells from the shock.
v. Cells were pelleted by centrifuging the mixture at 2500 rpm for 5 minutes.

vi. 1 ml of supernatant was pipette out leaving 100 µl in the tube.
vii. The cells were resuspended in this 100 µl of supernatant.
viii. The mixture was plated on tetracycline and ampicillin plates under sterile conditions to put a selection pressure in order to allow only those XL1-Blue cells to grow which would have taken up the recombinant plasmid.
ix. The plates were incubated at 37˚C for 16 hours and then checked for the growth of any bacterial colonies.

Colony PCR

Colony PCR is done to check for successful clones after transformation. The colonies are used as template and vector specific and/or insert specific primers are used for amplification. Successful integration of insert has bigger fragment of amplified DNA.

A colony from the plates with transformed cells in picked up and streaked onto new tetracycline and ampicillin plates with a sterile toothpick under sterile conditions and remaining culture from the toothpick is mixed with appropriate amount of distilled water and other components are added then. In this case, vector (pET23a) specific primers flanking the MCS region was used for amplification (**Table 1.14, 1.15, 1.16).**

Component	**Amount(µl)**
Distilled Water	11.5
T7 Forward Primer	1
T7 Reverse Primer	1
dNTPs	0.8
$MgCl_2$ (Working conc – 0.25mM)	1.6
Taq Flexi Buffer	4
Taq Flexi enzyme	0.15
Total	20.05

Table 1.14: Colony PCR components

Step	**Temperature(˚C)**	**Time(µl)**
Initial Denaturation	95	5 minutes
Denaturation	95	30 seconds
Annealing x35	Variable	45 seconds
Extension	72	Variable
Final Extension	72	10 minutes

Table 1.15: Colony PCR program

Gene	**Annealing Temperature**	**Extension Time**
CelA	58	2 minutes
CelQ	55	3 minutes
CelR	58	3 minutes

Table 1.16: Annealing temperature and Extension time for different inserts for colony PCR

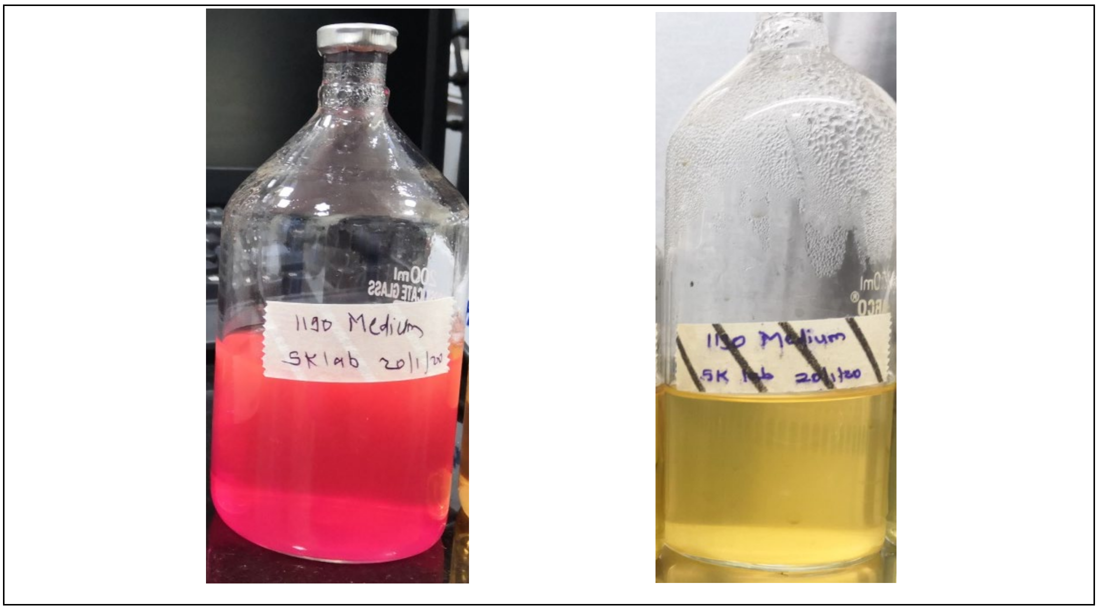
Further, the PCR mix is run on 0.8% agarose gel along with a control PCR to check for positive colonies.

1.3.2 RESULTS AND DISCUSSION

(A.) Culturing, maintenance and genomic DNA isolation of *Clostridium thermocellum*

Culturing of *Clostridium thermocellum*

200 ml each of manual (ATCC suggested) and RCB medium was made. The medium turned bluish in colour. After boiling the medium and purging the medium with N_2 gas the medium turned pinkish in colour, indicating lower levels of oxygen (**Figure1.13**). After autoclaving, the medium was light yellow coloured.



(A) After boiling **(B)** After adding reducing solution

Figure1.13: Representative images of medium with resazurin indicating oxygen levels

Protocol was followed and the culture was allowed to grow for 72 hours in a 50˚C incubator without shaking after which turbidity was observed.

Maintenance of *Clostridium thermocellum*

According to the protocol, 1 ml of the culture was added to 1 ml of anoxic 60% glycerol in a cryo-vial and was stored in -80˚C fridge.

Genomic DNA (gDNA) isolation from Clostridium thermocellum culture

Using the Zymo research bacterial Miniprep kit, genomic DNA was isolated from both the manual medium culture and the RCB medium culture. The elution was done in 40µl of elution buffer in room temperature. Nano-drop measurements of absorption at 280nm and 260 nm were recorded to check the concentration and purity of DNA isolated 1µl of the elute was run on 0.8% agarose gel (**Figure 1.14**).

Culture	**Concentration(ng/ µl)**
Manual medium	280
RCB medium	393

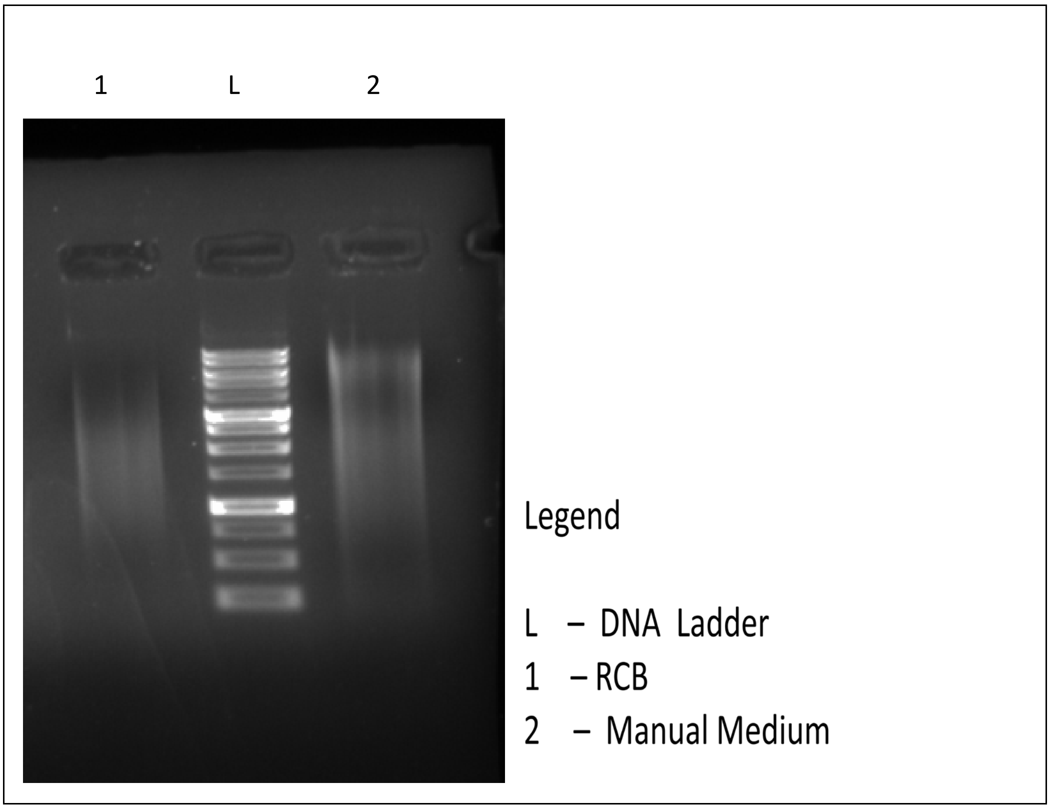


Figure 1.14: Gel Electrophoresis image of isolated gDNA

The genome of *C. thermocellum* is 3.84mbp in size. There is a great chance of the shearing of the gDNA in the process, therefore we observed a smear on the gel. Nevertheless, the genes are a very small part of the genome and until the fragments did not break the genes from within their sequences (which has a lower probability), this gDNA could be used for PCR amplification.

During my entire thesis, culturing and genomic isolation was done four times. The very first batch of isolated DNA was referred to as the old stock and all subsequent batches as new stock of DNA.

(B.) Standardisation of gene amplification from genomic DNA of *Clostridium thermocellum*

Primer designing

A total of 10 cellulases from the *Clostridium thermocellum* genome were selected to clone (**refer to section 1.2.1, Table 1.10**). Primers for the PCR amplification of each of these genes were constructed manually using Snap gene. Detail of the gene length and the designed primers for each enzyme is given in **Table 1.17.**

Gene Name	**Nucleotide length (in bp)**	**Restriction sites**	**Primers (5’→3’)**
Cel A	1434	NdeI	ATATATCATATGGTGAAGAACGTAAAAAAAAGAG
		XhoI	ATTAATCTCGAGATAAGGTAGGTGGGGTATG
Cel F	2220	NheI	ATATATGCTAGCATGTTGAAGAAAATTTTGGCG
		XhoI	ATTAATCTCGAGCTGTTCAGCCGGG
Cel Q	2133	NdeI	ATATATCATATGGTGAAAAAGACATTATGC
		XhoI	ATTAATCTCGAGTTCTACCGGAAATTTATC
Cel R	2211	NheI	ATATATGCTAGCATGGTGAAAAAACTCATTATC
		XhoI	ATTAATCTCGAGTGAATTTCCGGGTATGG
Cel H	2703	NheI	ATATATGCTAGCATGAAGAAAAGGCTTTTAG
		XhoI	ATTAATCTCGAGTATGGGTATTTCACTGATG
Cel I	2640	NdeI	ATATATCATATGAGGCTGGTTAACAGTTTGGG
		XhoI	ATTAATCTCGAGTCTTAGCCATGTTGCGTACAG
Cbh A	3675	NdeI	ATATATCATATGAAATTTAGAAGGTC
		XhoI	ATTAATCTCGAGTCGATATGGCAATTC
Cel S	2226	NdeI	ATATATCATATGGTAAAAAGCAGAAAGATTTC
		XhoI	ATTAATCTCGAGGTTCTTGTACGGCAATG
bgl A	1347	NheI	ATATATGCTAGCATGTCAAAGATAACTTTCCC
		XhoI	ATTAATCTCGAGGAAACCGTTGTTTTGATTAC
Bgl B	2265	NdeI	ATATATCATATGGCGGTAGATATCAAG
		XhoI	ATTAATCTCGAGTTCCACGTTGTTTATTTG

Table 1.17: List of primers designed for different *C. thermocellum* cellulases

Standardisation of PCR amplification of genes from *Clostridium thermocellum* genome

To amplify the listed cellulose genes from the genome of *Clostridium thermocellum*, PCR with many variable conditions was carried out using the genomic DNA (gDNA) of *Clostridium thermocellum* as template. Throughout from PCR 1- 11 many parameters were varied - gDNA stock, gDNA stock isolated from *Clostridium thermocellum* grown in which medium (referred to as gDNA stock medium), gDNA concentration, DMSO concentration, polymerase used and annealing temperatures (**Table 1.18)**. A 10kb DNA ladder was used for reference in all the PCRs.

Parameter	**Constant/Variable**								
	PCR1	**PCR2**	**PCR3**	**PCR4**	**PCR5**	**PCR6**	**PCR7**	**PCR8**	**PCR9**
Genes	CelH, CelR	CelH, CelR	CelQ, CelS, bglB	CelR	CelR	CelR	CelR	CelR	CelR, CelQ, CelA
gDNA stock	Old	Old	New	New	New	New	New	New	Old, New
gDNA stock medium (Manual Medium – MM/ RCB medium – RCB)	MM	MM	MM	MM, RCB	MM, RCB	MM, RCB	MM	MM	MM
gDNA concentration (in ng/ µl)	100	50, 100, 150	100	50	100	100	20, 100	200	100
DMSO	-	-	-	-	-	10%	2, 6%	-	-
Polymerase used (Vegapol(V), Taq(T), DeepVent(D)	V	V, T	V, T	V, T	V,T	V,T	D	D,T	V,T

Annealing temperature (in °C)	58, 60, 62	58, 60, 62	54, 56, 58, 60	58, 60, 62	56, 58, 60	56, 58	52, 55, 58	55	58(CelR, CelA), 55(CelQ)
Extension Time (in sec)	90	90(V), 180 (T)	90(V), 180(T)	90(V), 180 (T)	90(V), 180 (T)	90(V), 180 (T)	180	180 (D,T)	90(V-CelR, CelQ), 60 (V-CelA), 180 (T-CelR, CelQ), 120 (T- CelA)

Table1.18: Conditions and parameters varied in different PCRs

Below, are the gel electrophoresis images for all the PCRs.

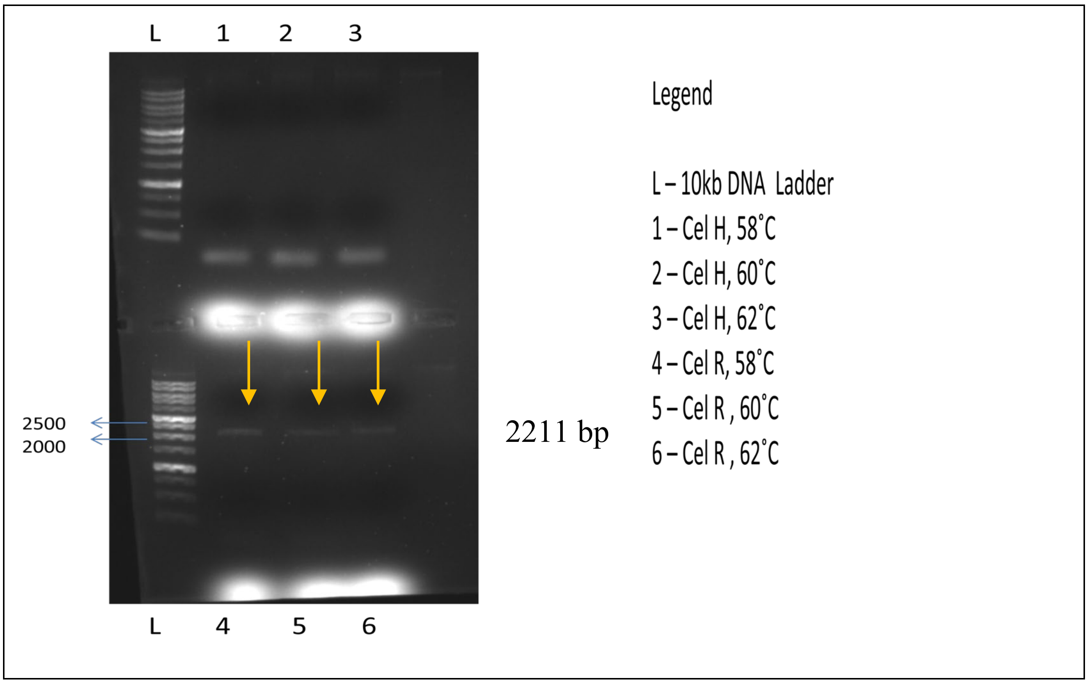

L
1
2
3
2500
2000
2211 bp
L
4
5
6
Legend
L – 10kb DNA Ladder
1 – Cel H, 58°C
2 – Cel H, 60°C
3 – Cel H, 62°C
4 – Cel R, 58°C
5 – Cel R, 60°C
6 – Cel R, 62°C

Figure 1.15 (A): Gel electrophoresis image of PCR 1

Faint DNA amplification band was reported at the correct size for CelR (2211 bp) (**Figure 1.15 (A)**).

No clear PCR amplification band was observed in PCR2 to PCR 8 **(Figure 1.15 (B) to (G))**. Reamplification of DNA was attempted in PCR 2 for CelR from the amplification in PCR 1, but results were negative. Primers dimers were observed all the bottom of the gel in case of some PCRs (PCR3, PCR7) (**Figure 1.15 (B), (F))**. In PCR 6, 7 DMSO was used to prevent the formation of primer dimers, yet no amplification was observed **(Figure 1.15 (E),(F))**. In some of the gel images of certain PCRs (lanes 7,8 PCR5, PCR7, PCR8), a smear can be observed which may be due the presence of the template genomic DNA itself (**Figure 1.15 (D), (F), (G))**. A non-specific amplification was observed in lanes 8,9 in PCR 3 (**Figure 1.15 (B))**.

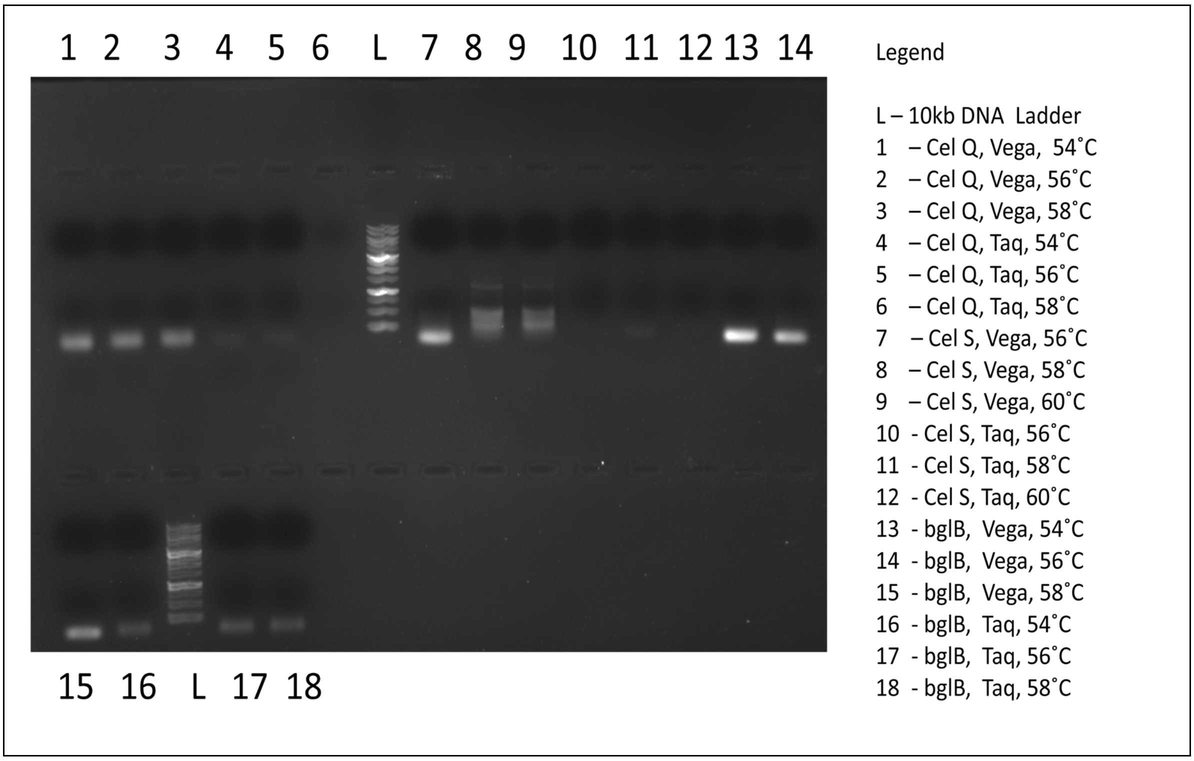

1 2 3 4 5 6 L 7 8 9 10 11 12 13 14
15 16 L 17 18
Legend
L – 10kb DNA Ladder
1 – Cel Q, Vega, 54˚C
2 – Cel Q, Vega, 56˚C
3 – Cel Q, Vega, 58˚C
4 – Cel Q, Taq, 54˚C
5 – Cel Q, Taq, 56˚C
6 – Cel Q, Taq, 58˚C
7 – Cel S, Vega, 56˚C
8 – Cel S, Vega, 58˚C
9 – Cel S, Vega, 60˚C
10 - Cel S, Taq, 56˚C
11 - Cel S, Taq, 58˚C
12 - Cel S, Taq, 60˚C
13 - bglB, Vega, 54˚C
14 - bglB, Vega, 56˚C
15 - bglB, Vega, 58˚C
16 - bglB, Taq, 54˚C
17 - bglB, Taq, 56˚C
18 - bglB, Taq, 58˚C

Figure 1.15 (B): Gel electrophoresis image of PCR 3

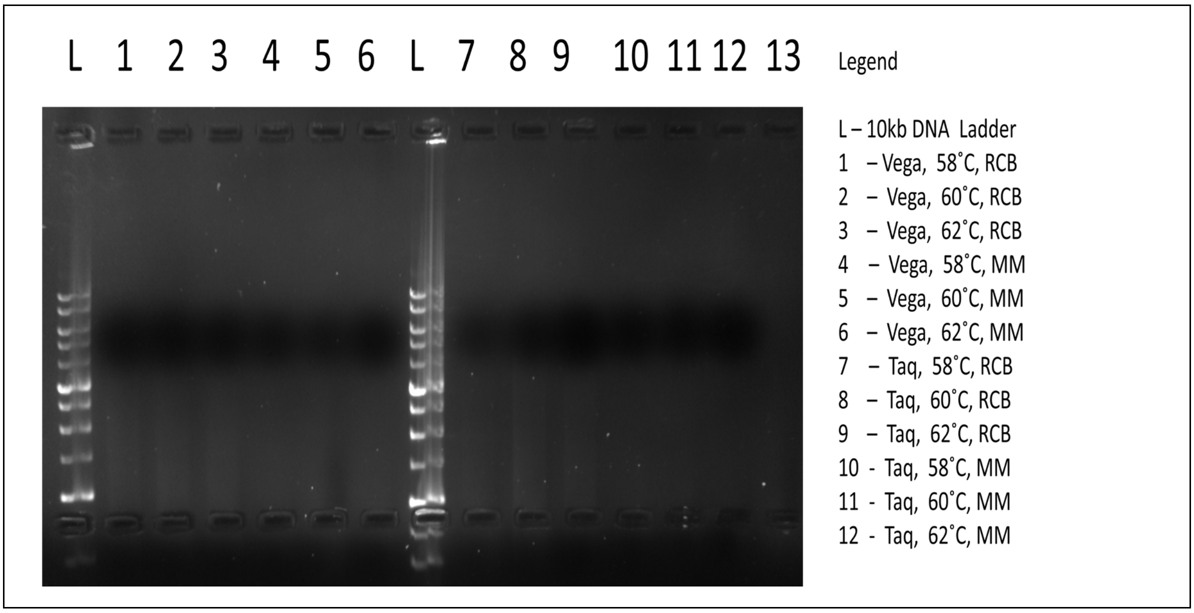

L 1 2 3 4 5 6 L 7 8 9 10 11 12 13
Legend
L – 10kb DNA Ladder
1 – Vega, 58˚C, RCB
2 – Vega, 60˚C, RCB
3 – Vega, 62˚C, RCB
4 – Vega, 58˚C, MM
5 – Vega, 60˚C, MM
6 – Vega, 62˚C, MM
7 – Taq, 58˚C, RCB
8 – Taq, 60˚C, RCB
9 – Taq, 62˚C, RCB
10 - Taq, 58˚C, MM
11 - Taq, 60˚C, MM
12 - Taq, 62˚C, MM

Figure 1.15 (C): Gel electrophoresis image of PCR 4

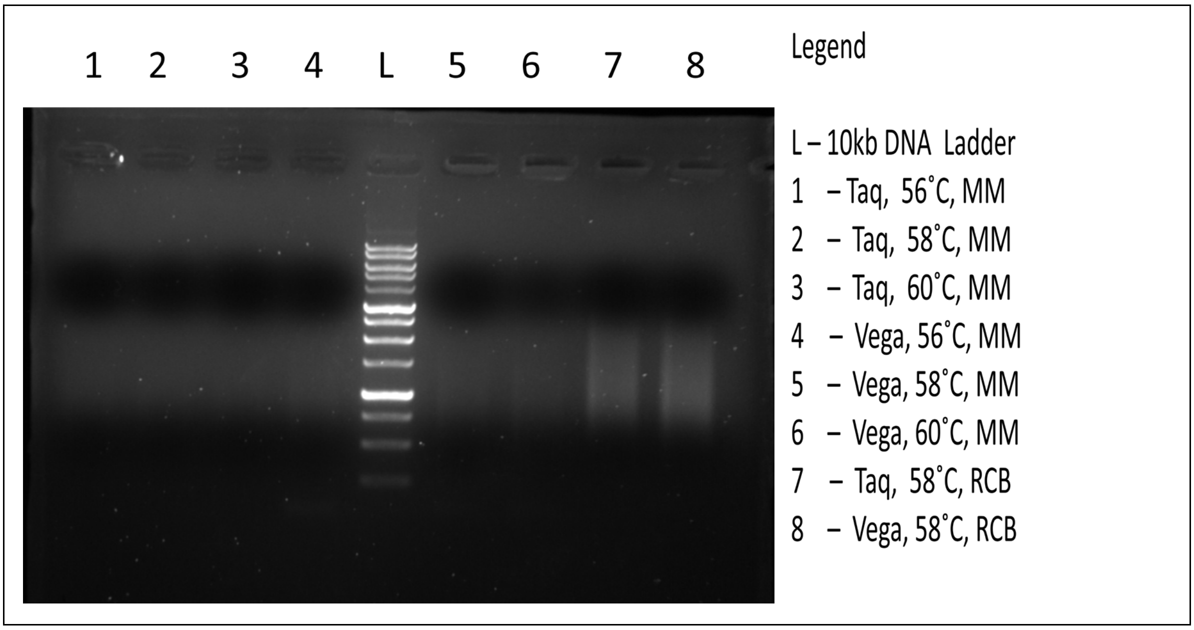

1 2 3 4 L 5 6 7 8
Legend
L – 10kb DNA Ladder
1 – Taq, 56˚C, MM
2 – Taq, 58˚C, MM
3 – Taq, 60˚C, MM
4 – Vega, 56˚C, MM
5 – Vega, 58˚C, MM
6 – Vega, 60˚C, MM
7 – Taq, 58˚C, RCB
8 – Vega, 58˚C, RCB

Figure 1.15 (D): Gel electrophoresis image of PCR 5

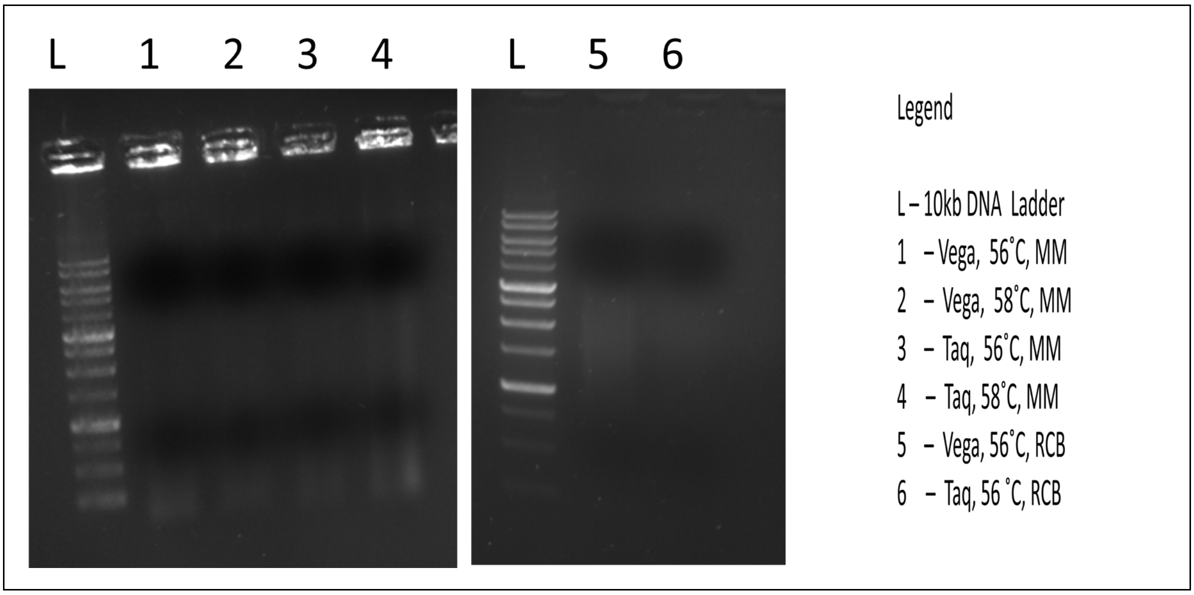

L
1
2
3
4
L
5
6
Legend
L – 10kb DNA Ladder
1 – Vega, 56˚C, MM
2 – Vega, 58˚C, MM
3 – Taq, 56˚C, MM
4 – Taq, 58˚C, MM
5 – Vega, 56˚C, RCB
6 – Taq, 56 ˚C, RCB

Figure 1.15 (E): Gel electrophoresis image of PCR 6

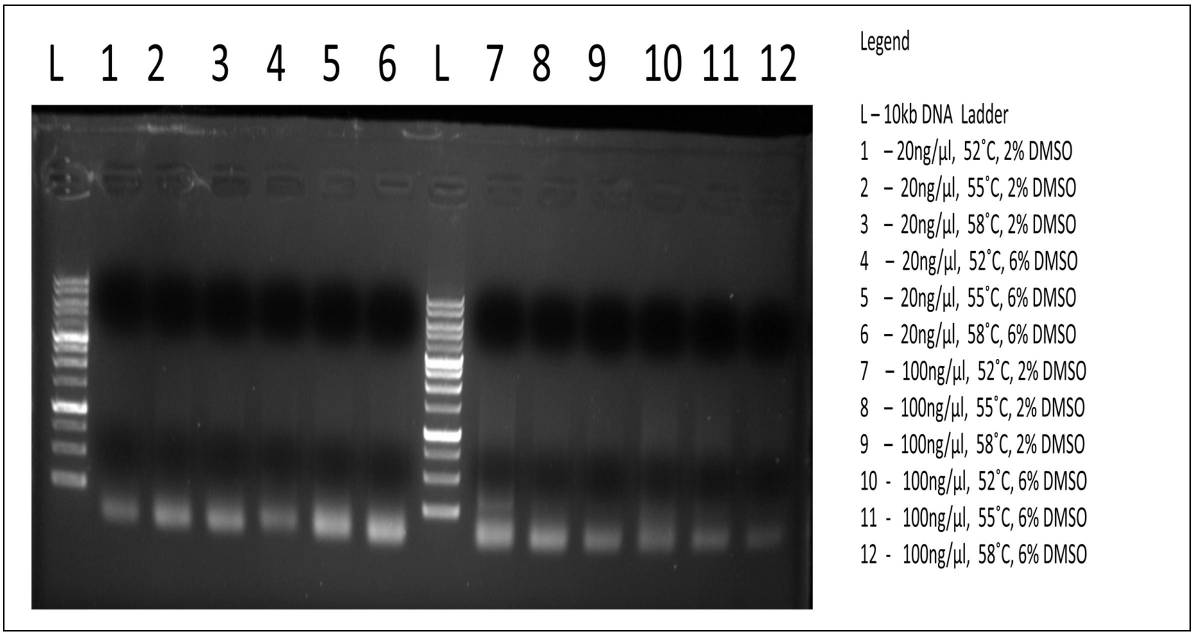

L
1
2
3
4
5
6
L
7
8
9
10
11
12
Legend
L – 10kb DNA Ladder
1 – 20ng/μl, 52˚C, 2% DMSO
2 – 20ng/μl, 55˚C, 2% DMSO
3 – 20ng/μl, 58˚C, 2% DMSO
4 – 20ng/μl, 52˚C, 6% DMSO
5 – 20ng/μl, 55˚C, 6% DMSO
6 – 20ng/μl, 58˚C, 6% DMSO
7 – 100ng/μl, 52˚C, 2% DMSO
8 – 100ng/μl, 55˚C, 2% DMSO
9 – 100ng/μl, 58˚C, 2% DMSO
10 - 100ng/μl, 52˚C, 6% DMSO
11 - 100ng/μl, 55˚C, 6% DMSO
12 - 100ng/μl, 58˚C, 6% DMSO

Figure 1.15 (F): Gel electrophoresis image of PCR 7

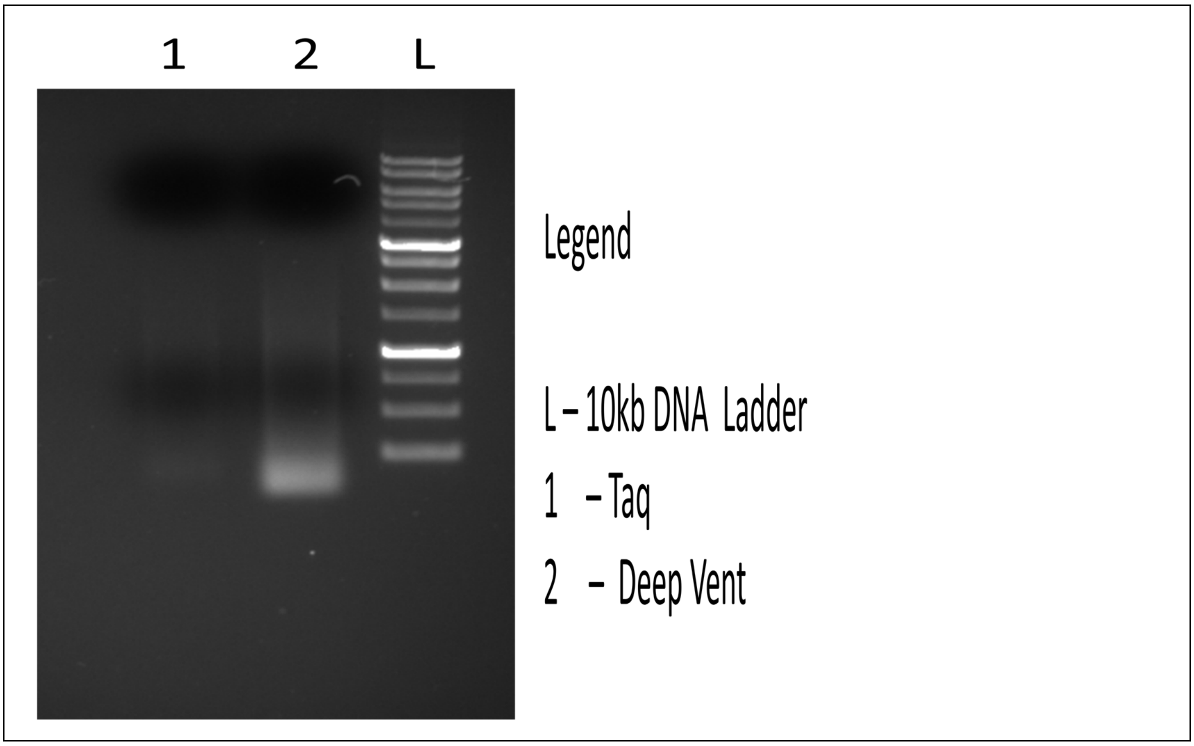

1
2
L
Legend
L – 10kb DNA Ladder
1 – Taq
2 – Deep Vent

Figure 1.15 (G): Gel electrophoresis image of PCR 8

All the PCRs with negative results were done using the new stock of DNA. To verify, if these results were indeed negative because of the different stocks of isolated gDNA, PCR 9 was done by keeping every parameter constant while using both old and new stock of DNA as template.

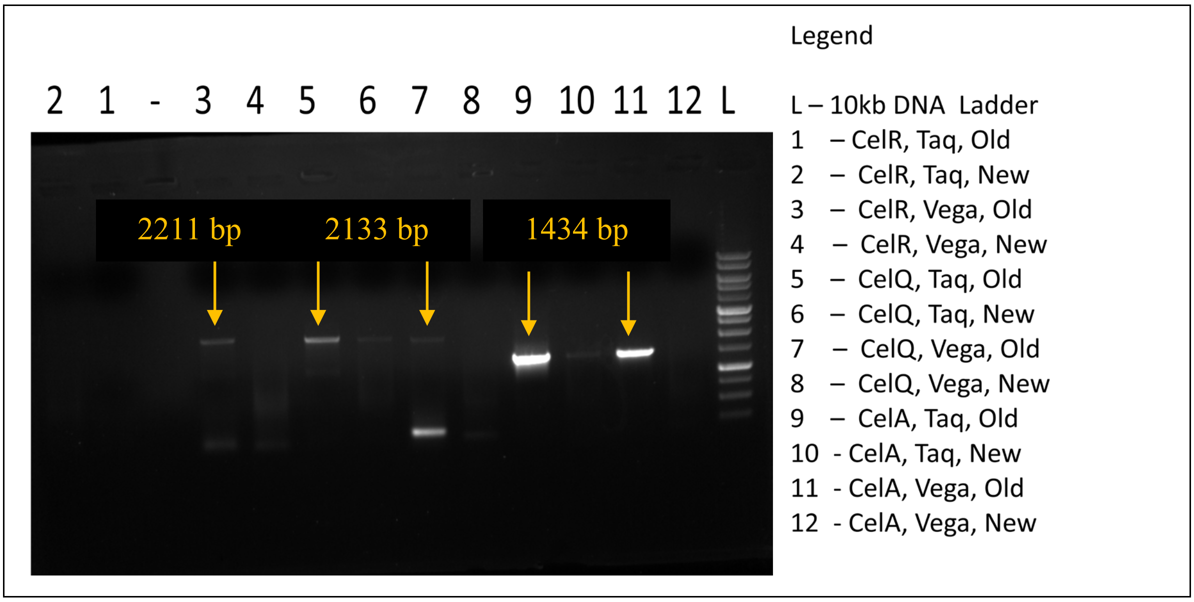

2 1 - 3 4 5 6 7 8 9 10 11 12 L
2211 bp
2133 bp
1434 bp
Legend
L – 10kb DNA Ladder
1 – CelR, Taq, Old
2 – CelR, Taq, New
3 – CelR, Vega, Old
4 – CelR, Vega, New
5 – CelQ, Taq, Old
6 – CelQ, Taq, New
7 – CelQ, Vega, Old
8 – CelQ, Vega, New
9 – CelA, Taq, Old
10 - CelA, Taq, New
11 - CelA, Vega, Old
12 - CelA, Vega, New

Figure 1.15 (H): Gel electrophoresis image of PCR 9

As evident from **Figure 1.15 (H)**, DNA amplification was observed in all those reactions where the old stock of gDNA template was used. With the same conditions, the reactions where in new stock of gDNA was used, faint or no bands were observed. It was concluded that there was some fault in the new gDNA. Further, for the following PCRs the old stock of gDNA was used.

Gel extraction of amplified DNA bands was done in order to attempt reamplification using this DNA as template. The conditions used are listed in **Table 1.19.**

Parameter	**Constant/Variable**
Genes	CelQ, CelR, CelA
Template DNA concentration	2, 10, 20ng/ µl
Polymerase used	Taq
Annealing temperature	55, 58, 60˚C
Extension Time	180 seconds (CelR, CelQ), 120 seconds(CelA)

Table 1.19: PCR conditions for reamplification of DNA fragments

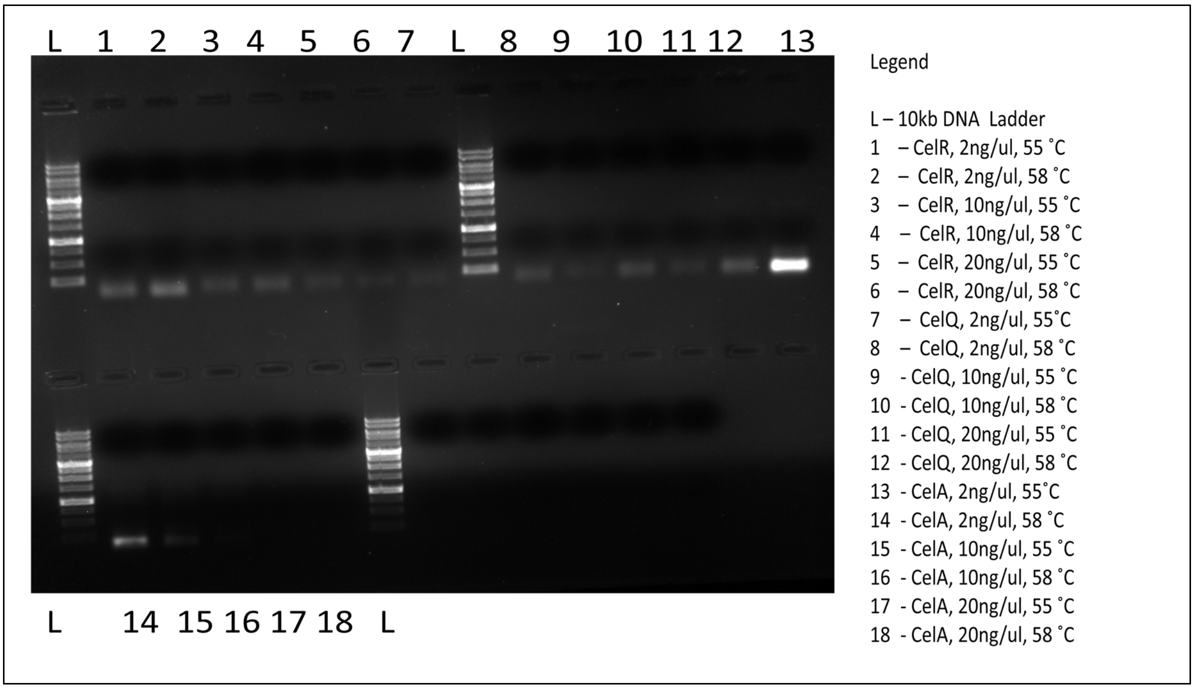
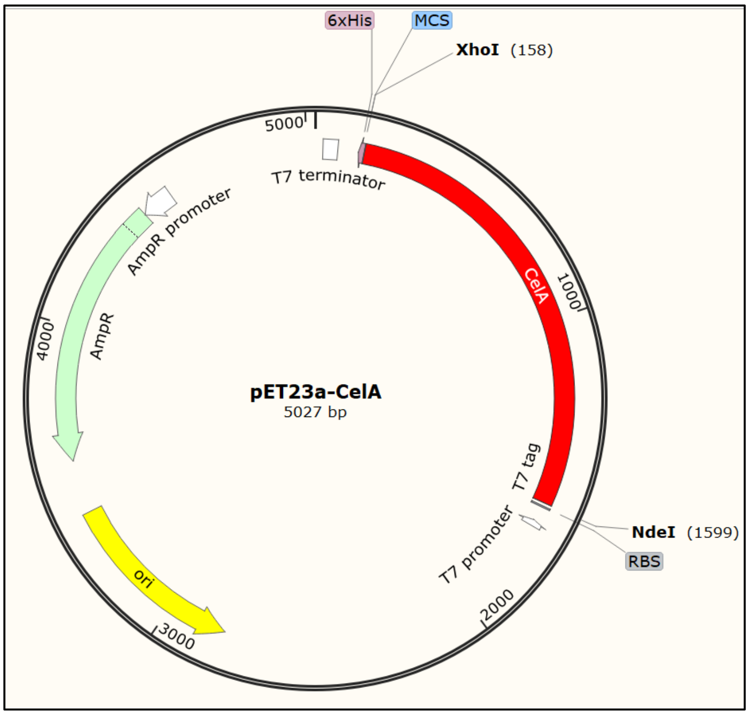

L 1 2 3 4 5 6 7 L 8 9 10 11 12 13
L 14 15 16 17 18 L
Legend
L – 10kb DNA Ladder
1 – CelR, 2ng/ul, 55 ˚C
2 – CelR, 2ng/ul, 58 ˚C
3 – CelR, 10ng/ul, 55 ˚C
4 – CelR, 10ng/ul, 58 ˚C
5 – CelR, 20ng/ul, 55 ˚C
6 – CelR, 20ng/ul, 58 ˚C
7 – CelQ, 2ng/ul, 55˚C
8 – CelQ, 2ng/ul, 58 ˚C
9 - CelQ, 10ng/ul, 55 ˚C
10 - CelQ, 10ng/ul, 58 ˚C
11 - CelQ, 20ng/ul, 55 ˚C
12 - CelQ, 20ng/ul, 58 ˚C
13 - CelA, 2ng/ul, 55˚C
14 - CelA, 2ng/ul, 58 ˚C
15 - CelA, 10ng/ul, 55 ˚C
16 - CelA, 10ng/ul, 58 ˚C
17 - CelA, 20ng/ul, 55 ˚C
18 - CelA, 20ng/ul, 58 ˚C

Figure 1.16: Gel electrophoresis image of attempted reamplification PCR from PCR9

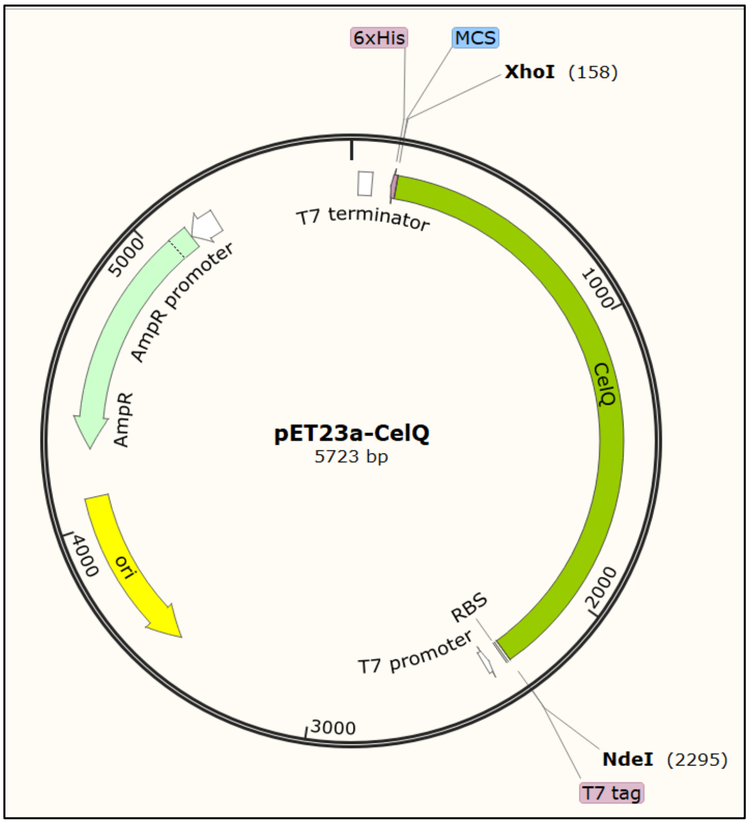
No specific bands were observed **(Figure 1.16)**. It was decided to amplify gene of interest directly from the old stock of gDNA and proceed with cloning.

(C.) Cloning of Cel R, CelQ, CelA in pET23a vector

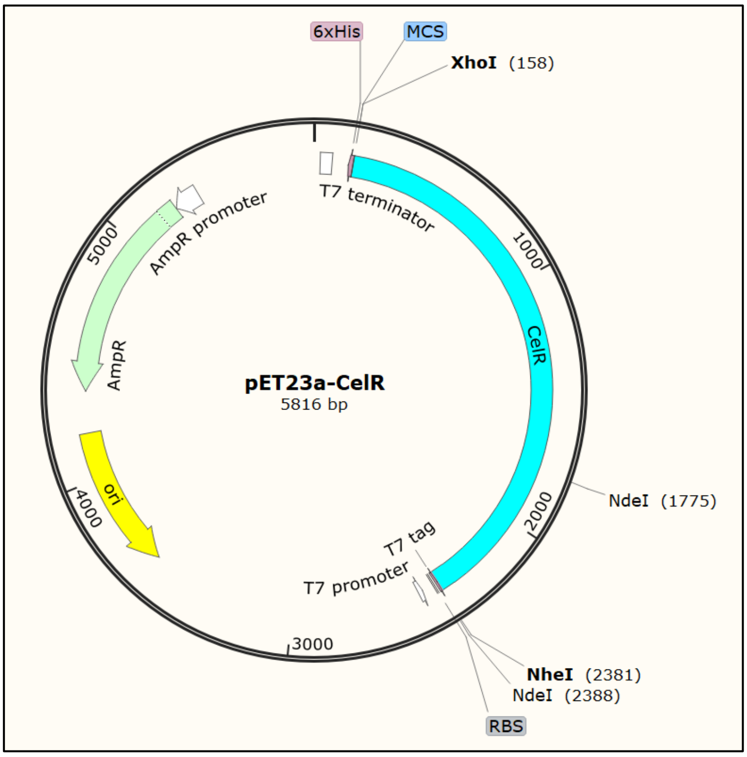
Schematic vector maps were made for all the three clones using Snapgene while checking compatibility of the two DNA fragments after using respective restriction enzyme sets **(Figure 1.17).**


6xHis
MCS
XhoI (158)
5000
T7 terminator
AmpR promoter
AmpR
4000
CelA
1000
pET23a-CelA
5027 bp
T7 tag
T7 promoter
NdeI (1599)
RBS
ori
3000
2000

pET23a:CelA


6xHis
MCS
XhoI (158)
T7 terminator
5000
AmpR promoter
AmpR
1000
CelQ
pET23a-CelQ
5723 bp
4000
ori
RBS
T7 promoter
2000
3000
NdeI (2295)
T7 tag

pET23a:CelQ


6xHis
MCS
XhoI (158)
T7 terminator
1000
CelR
pET23a-CelR
5816 bp
NdeI (1775)
2000
T7 tag
T7 promoter
3000
NheI (2381)
NdeI (2388)
RBS
4000
ori
AmpR
5000
AmpR promoter

pET23a:CelR

Figure 1.17: Plasmid maps of recombinant clones

PCR amplification and gel extraction of gene of interest

The PCR amplification of CelA, CelQ and CelR was standardised. The most favourable conditions are listed below (**Table 1.20).**

Parameter	Value
gDNA stock and medium	Old, Manual Medium
gDNA concentration	100ng/ µl
Polymerase used	VegaPol or Taq
Annealing temperature(˚C)	58(CelR, CelA), 55(CelQ)
Extension Time(seconds)	90(V-CelR, CelQ), 60(V-CelA), 180(T- CelR, CelQ), 120(T- CelA)

Table 1.20: Standardised parameters for PCR amplification of CelA, CelQ and CelR

100µl PCR reaction using VegaPol as polymerase for CelR and CelA was set up and PCR purification was done to get enough amount of insert for cloning purposes and run on 0.8% agarose gel.

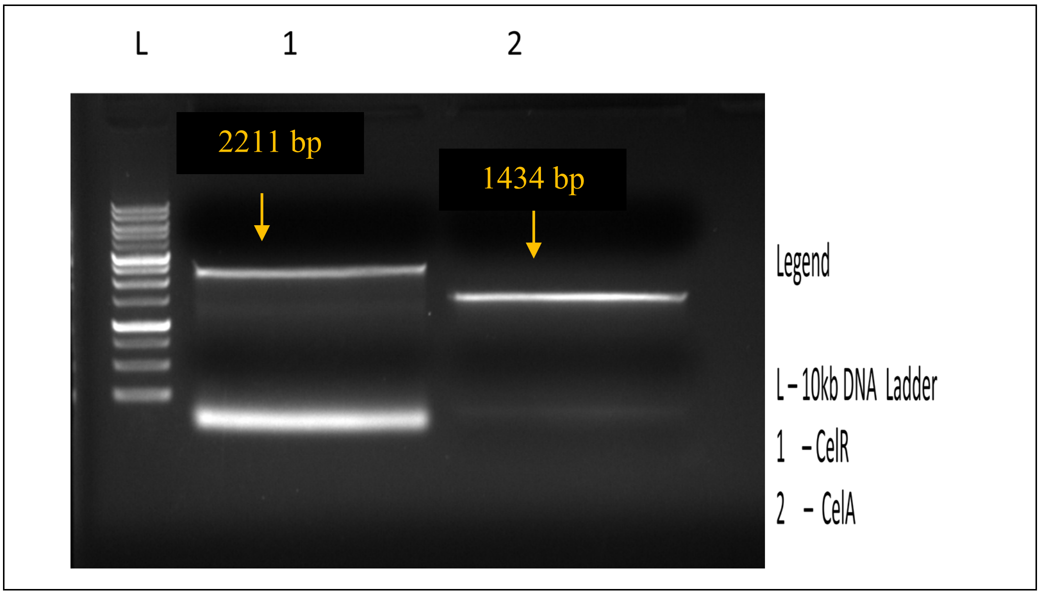


Figure 1.18 : Gel electrophoresis image of PCR done to amplify Cel R and CelA

Clear DNA bands were obtained at desired sizes **(Figure 1.18).**

Gel extraction of these two bands was done. For CelQ, the gel extracted DNA from PCR 9 was used. Nano-drop measurements were done to estimate the concentrations.

Gene	**Concentration(in ng/µl)**	**Elution Volume(µl)**
Cel A	16	30
CelQ	7	40
CelR	8	40

Plasmid isolation of pET23a:E1-E2-E3-E4 and pET23a:N1-N2-N3-N4

Two plasmids with different restriction sites (pET23a:E1-E2-E3-E4 and pET23a:N1-N2-N3-N4) to be used as vectors for the three genes of interest were isolated (elution in 50µl) and their concentrations were estimated.

Gene	Concentration(in ng/µl)
pET23a:E1-E2-E3-E4	127
pET23a:N1-N2-N3-N4	80

Restriction Digestion and Ligation of the plasmids and inserts

The PCR amplified products (CelA, CelR, CelQ) and the isolated plasmids (pET23a:E1-E2-E3-E4 and pET23a:N1-N2-N3-N4) were digested at 37˚C overnight using the respective restriction enzymes.

All three digested inserts were purified using PCR cleanup and were eluted in 25µl autoclaved distilled water. The digested plasmids were run on 0.8% agarose gel and the digested plasmid backbone was gel extracted.

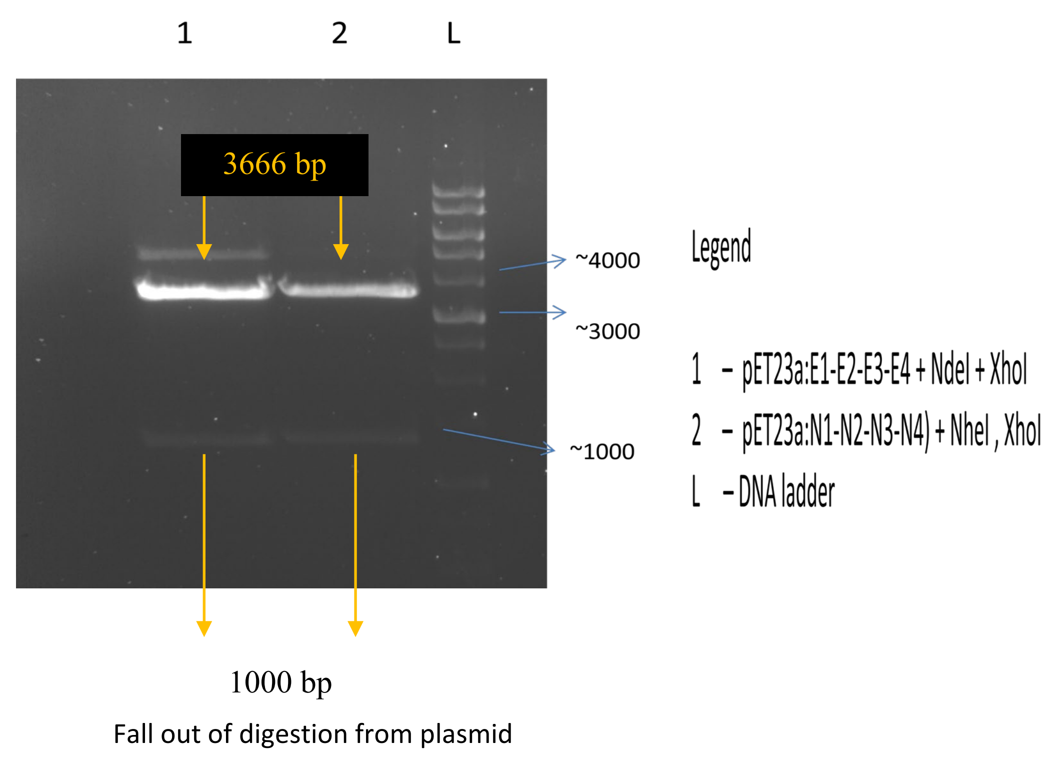
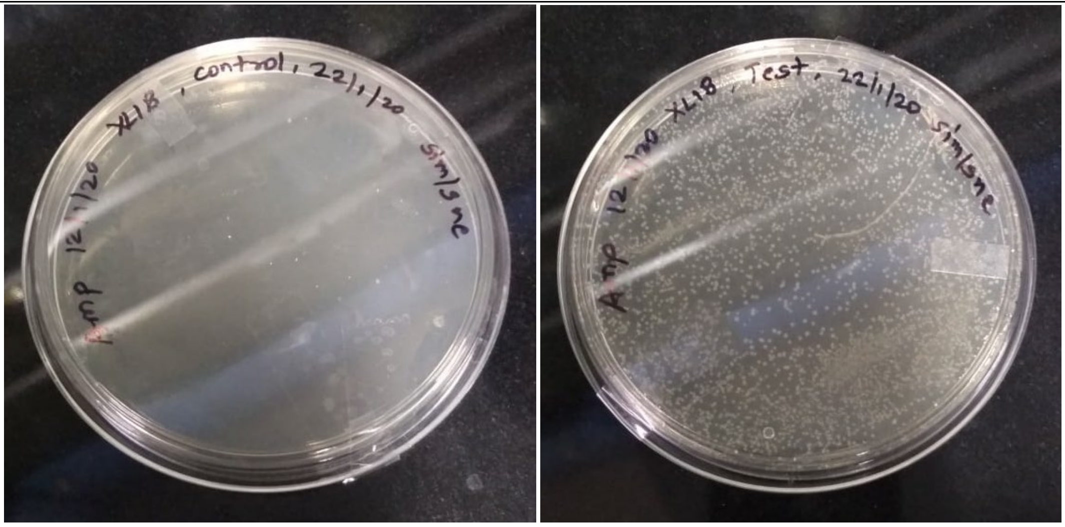

1
2
L
3666 bp
~4000
~3000
~1000
Legend
1 - pET23a:E1-E2-E3-E4 + NdeI + XhoI
2 - pET23a:N1-N2-N3-N4) + NheI , XhoI
L - DNA ladder
1000 bp
Fall out of digestion from plasmid

Figure 1.19: Gel electrophoresis image of double digested plasmids

Following a vector to insert ratio of 1:3, ligation reaction of the respective inserts and plasmids was set up at 25˚C overnight.

Competent cells preparation

XL1-Blue cells were made chemically competent using $CaCl_2$ method. To check the transformation efficiency and contamination, 100µl of untransformed cells (control) and 100µl of transformed cells (test) (with 10ng of plasmid that had ampicillin resistance) were plated on two LB-agar Amp plates.



a. Control b. Test

Figure 1.20: Images showing successful preparation of XL1-Blue competent cells (jointly prepared with Snehal Waghmere)

After 16 hours of incubation at 37˚C, the control plate had no colonies (**Figure 1.20 (a))** whereas the test plate consisted of almost a lawn like growth (**Figure 1.20(b))**. These XL1-Blue cells were used for transformation of the ligation product.

Transformation and Colony PCR

All the three ligated products were transformed into chemically competent XL1Blue cells via heat shock method. After rejuvenation, the cells were plated on LB agar plates containing ampicillin and tetracycline. After 16 hours of incubation at 37˚C, each of three plates of these plates were observed to have a single colony.

A colony PCR was set up using the cells from these colonies as template along with a control to check for successful ligation and transformation.

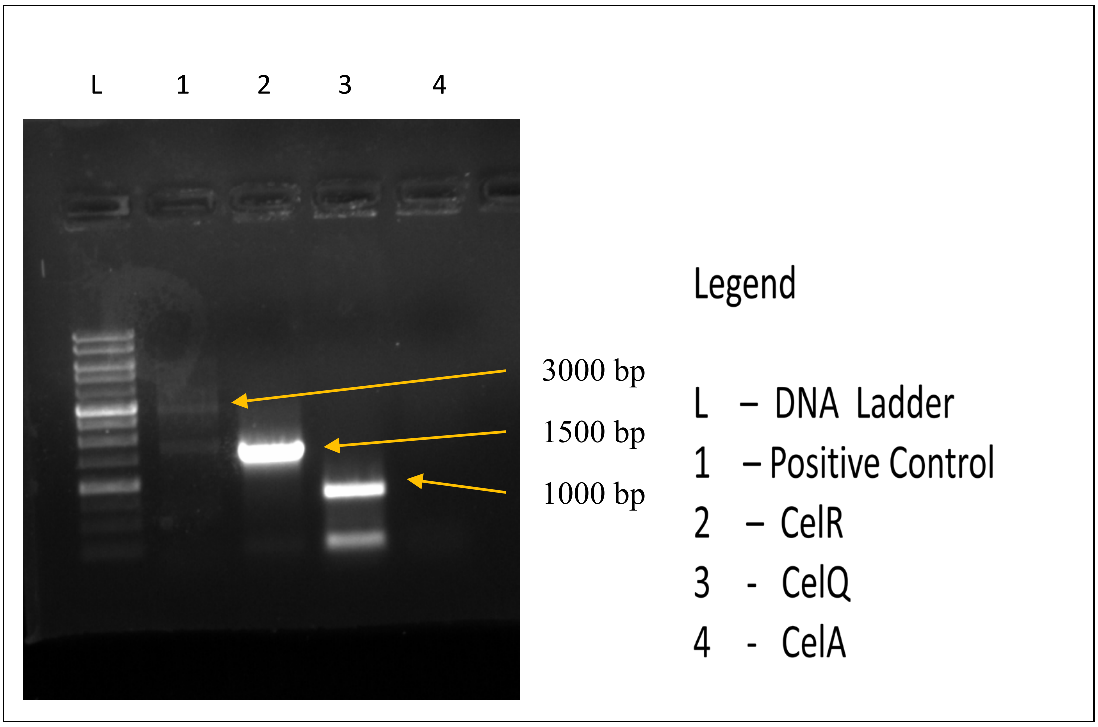


Figure 1.21: Gel electrophoresis of colony PCR

Correct band size of 3000bp was observed for positive control. Non-specific amplifications corresponding to incorrect band sizes were observed for CelR and CelQ – around 1500 and 1000 bp respectively whereas their actual sizes are 2211 and 2133 bp respectively. For CelA, there was no amplification at all **(Figure 1.21)**. It was concluded that the cloning was unsuccessful.

Concluding Remarks

In order to clone the different cellulases from *Clostridium thermocellum*, standardisation of PCR amplification of the genes CelA, CelQ and CelR was successful. However, the ligation and transformation of the desired constructs were unsuccessful.

Cloning experiments could not be repeated because of the limited amount of the old stock of gDNA.

Only the old stock of DNA was useful in PCR amplification. We speculate that the new glycerol stocks made eventually after every culture batch had become contaminated. The glycerol stocks were kept in normal aerobic cryo-vials and not in anoxic serum bottles, which may have proved to be an unfavourable condition for the *Clostridium thermocellum* cells as the *Clostridium* genus is extremely sensitive to even trace amounts of oxygen in their environment (Edwards, Suárez, & McBride, 2013). In such conditions, a contamination with a facultative anaerobe could have resulted in overtaking the growth of *Clostridium* cells by this contaminant. The nano-drop measurements of the new stocks showed a presence of DNA in them. The PCRs performed with the new stock of gDNA didn't even show a faint amplification band, which implies that the DNA isolated had either trace amounts or no genomic DNA from *Clostridium thermocellum*. Even after culturing a batch using the old glycerol stock, there was no success in getting amplification.

1.3.3 FUTURE OUTLOOK

Better handling and maintenance of new pure culture of *Clostridium thermocellum* is required to continue with this study. Successful cloning of all these cellulases would pave way for heterologous expression of these cellulases in *Escherichia coli*. These cellulases can then be characterised for their structural properties and enzymatic activity to be able to check their potency for efficient biomass degrading enzyme cocktails. Further, once expressed and purified, these enzymes can also be used to study dockerin-cohesin interactions in the context of cellulosomes of *Clostridium thermocellum*.

Chapter 2

Structural characterisation and activity assay of engineered *Pyrococcus furiosus* DNA ligases

2.1 INTRODUCTION

DNA ligase is an enzyme (EC 6.5.1.1) that catalyses the formation of phosphodiester bonds between 5’ phosphoryl and 3’ hydroxyl groups to seal nicks in DNA. It is an important enzyme – in vivo to maintain the integrity of the genome and in vitro for facilitation of molecular biology research. The ligases are divided into two classes depending on their cofactor requirements – ATP and NAD+ dependent ligases (Chambers & Patrick, 2015). Most of the ligases can seal a nick in the single strand of double stranded DNA while a few can also ligate a DNA with double stranded break.

Structure of ligase

In majority of the organisms, a DNA ligase is composed of three domains namely N-terminal DNA binding (DBD), middle adenylation domain (AdD) and a C-terminal oligomer-binding fold (OB) (Chambers & Patrick, 2015). These domains contain six conserved amino acid sequence motifs numbered as I, IIIA,IIIB, IV, V, and VI. The AdD and OB domains are together referred to as the catalytic core as they are minimally required for the activity (Tanabe, Ishino, Ishino, & Nishida, 2014). These are the only domains present in two domain ligases (bacteriophage T7 and some bacterial ligases) (Subramanya, Doherty, Ashford, & Wigley, 1996). Further, the N-terminal DBD show additive DNA binding activity along with AdD and OBD (present in human, T4 and archael ligases) (John M. Pascal, O'Brien, Tomkinson, & Ellenberger, 2004).

Molecular mechanism of Ligation

A DNA ligase works through a ping-pong mechanism consisting of three steps to seal a nick in a DNA –

1. Adenylation:

   ATP is captured by the electrostatic interaction between conserved amino acid residues RXDK residues (in motif VI) in the OB fold domain. By a conformational change in the OB domain, it donates the bound ATP to the AdD domain where the ATP covalently attached to the e-NH_2 group conserved lysine residue (KXDG) with release of PPi.

2. Nick recognition and transfer of AMP:

   The nick is recognised by scanning of DNA backbone after which the covalently bound AMP in the AdD domain is now transferred to the open 5’ phosphate group on the nicked DNA.
3. Bond formation:

   The final step in the reaction is formation of a phosphodiester bond through nucleophilic attack of the 3’ OH group on the 5’ phosphate group to seal the nick.

Ligase has been shown to form closed conformation for ATP transfer before forming another closed conformation with nicked DNA substrate (Chambers & Patrick, 2015; J. M. Pascal et al., 2006).

T4 DNA ligase has been used extensively in research for in vitro sealing of nicks in DNA. The limitation of T4 DNA ligase exists in its thermostability and in dealing with the ligation of the DNA template forming a secondary structure at low temperatures. A mesoactive thermostable ligase will be ideal for daily lab ligation reactions and for the creation of long genomic libraries.

Thermostable archael ligases

Thermostable archael ligases such *as Pyrococcus furiosus* DNA ligase and *Thermococcus sp.* DNA ligase are ATP dependant ligases consisting an additional C-terminal alpha helix rich in charged polar amino acids residues along with the three domains (Nishida, Kiyonari, Ishino, & Morikawa, 2006; Petrova et al., 2012). As the sources of these enzymes are thermophile organisms, they show optimum activity at higher temperatures. In these ligases, the ATP bound to the OB fold is transferred to AdD through formation of a closed conformation which is stabilised by the C-terminal helix forming electrostatic interaction with the oppositely charged residues in AdD and OB fold domain (Tanabe et al., 2014) (**Figure 2.1**).

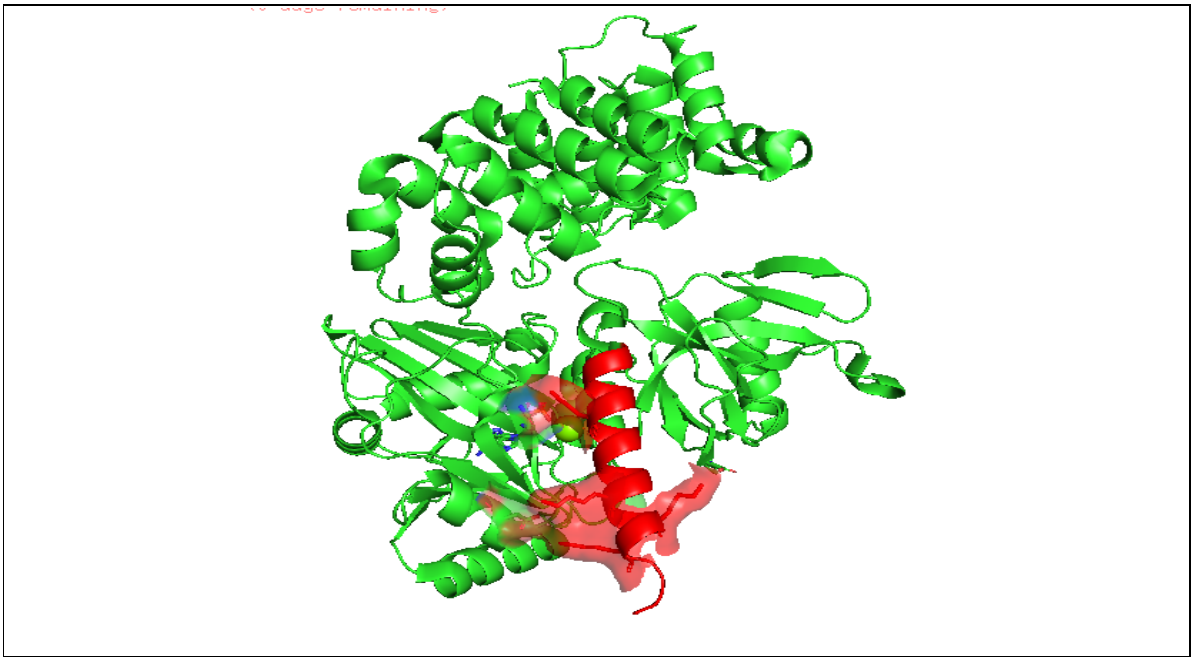
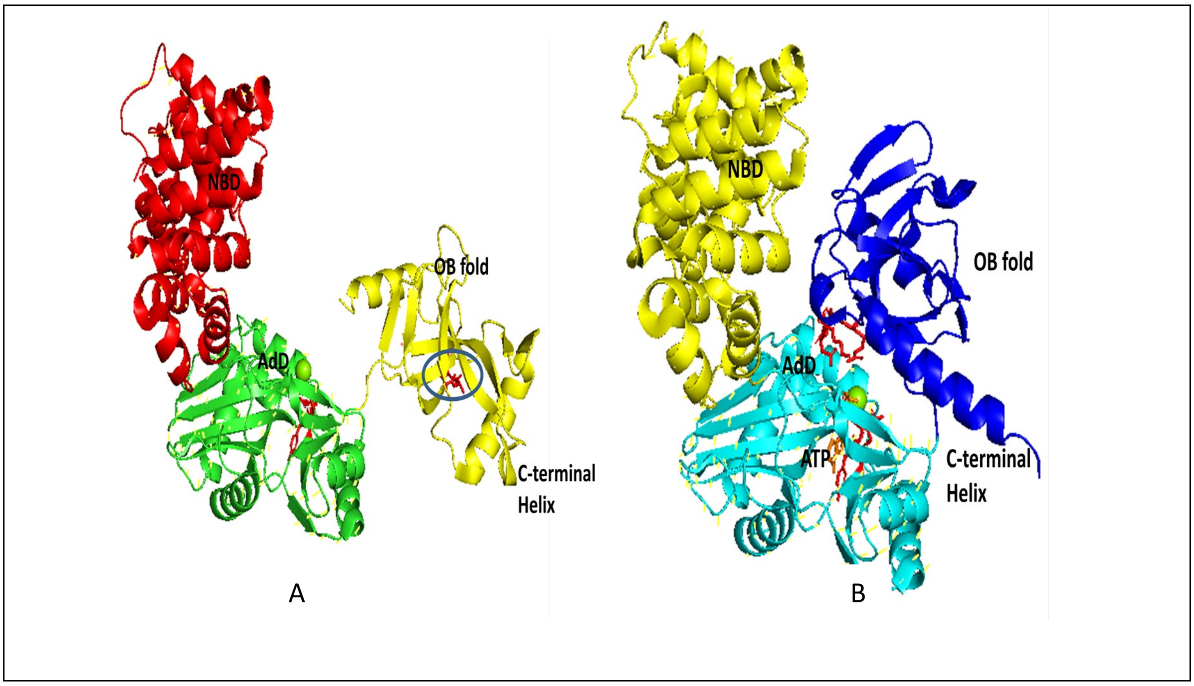


Figure 2.1: The C-terminal helix mediated stabilization of closed conformation (by interaction among K554, K558, D235, Q228 and Q547 residues) in *P.furiosus* DNA ligase

To visualise the domain rotation during adenylation in archael ligases, crystal structures of two ligases were compared. Crystal structure of a ligase without a bound ATP form *Thermococcus sp.* was compared with crystal structure of a ligase with a bound ATP from *P. furiosus* **(Figure 2.2).** The enzymes share a 95% sequence similarity and the structural comparison shows identical structure for the DBD and AdD domains which indicates that they would have the same principles of functioning. **In Figure 2.2(A), t**he crystal structure of DNA ligase from *Thermococcus sp.*is shown in an open conformation. The residues marked in blue circles are conserved residues on the OB fold which are responsible for ATP transfer from OB fold to AdD. In **Figure 2.2(B)**, the structure of DNA ligase from *P.furiosus* is shown in closed conformation with ATP bound state due to C-terminal helix mediated closure of catalytic core. In **Figure 2.2(C),** the structural alignment of both the ligases using Pymol is shown indicating the position of OB fold and C-terminal helix in both the ligases. The only difference is in the conformation of the OB domain with the C-terminal helix. It is seen that the archael ligases are in open conformation when not bound to ATP, but they form a closed figuration with ATP where it seems charged C-terminal helix is required for ATP transfer by forming closed conformation at elevated temperature.


NBD
OB fold
AdD
C-terminal
Helix
A
NBD
OB fold
AdD
ATP
C-terminal
Helix
B

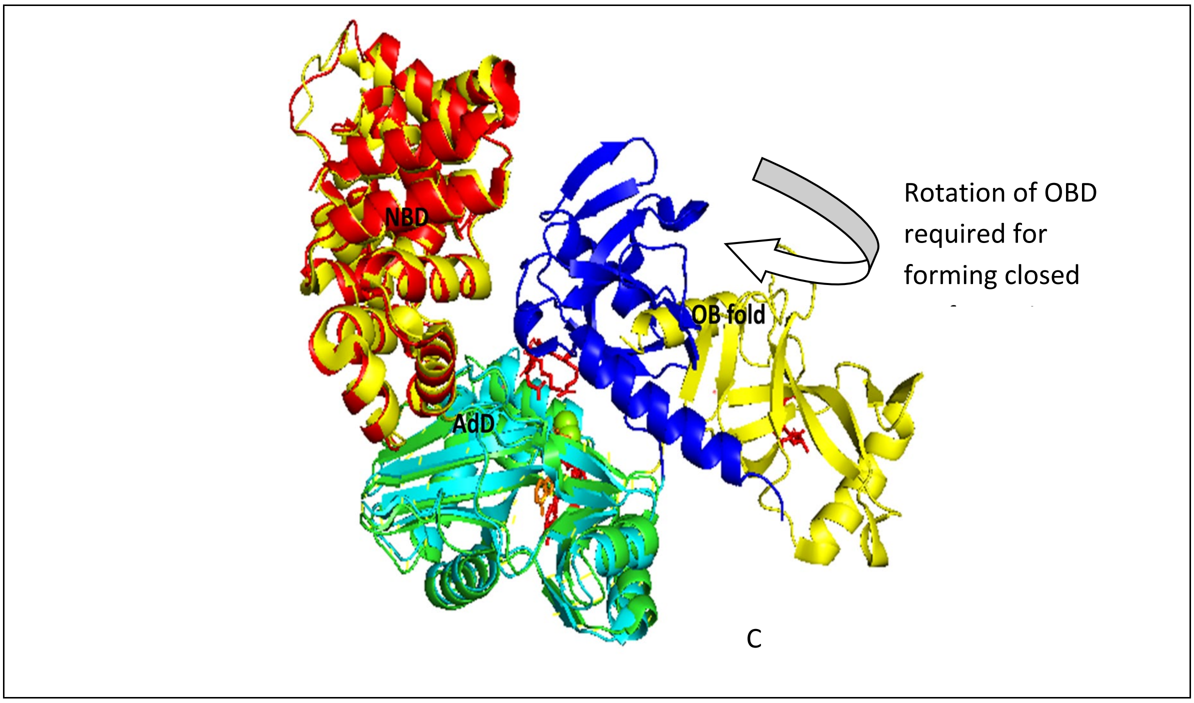

NBD
Rotation of OBD
required for
forming closed
OB fold
AdD
C

Figure 2.2: Structure comparison of two archeal ligases in two different conformations

In case of mesophilic ligases, the C-terminal helix is both shorter and neutral. The ATP transfer in this case happens without the charged C-terminal. When commercially used mesophilic T4 DNA ligase was compared to the *P. furiosus* DNA ligase, the AdD domain along with the catalytic residues and the OB fold aligned perfectly. However, a difference in the length of the N-terminal DNA binding domain and additional presence of a C-terminal helix in the *P.furiosus* ligase was observed. Further, bigger N terminal DNA binding domain in thermostable ligases could have helped in the positioning of DNA ligase on DNA strand at high-temperature conditions (**Figure 2.3**).

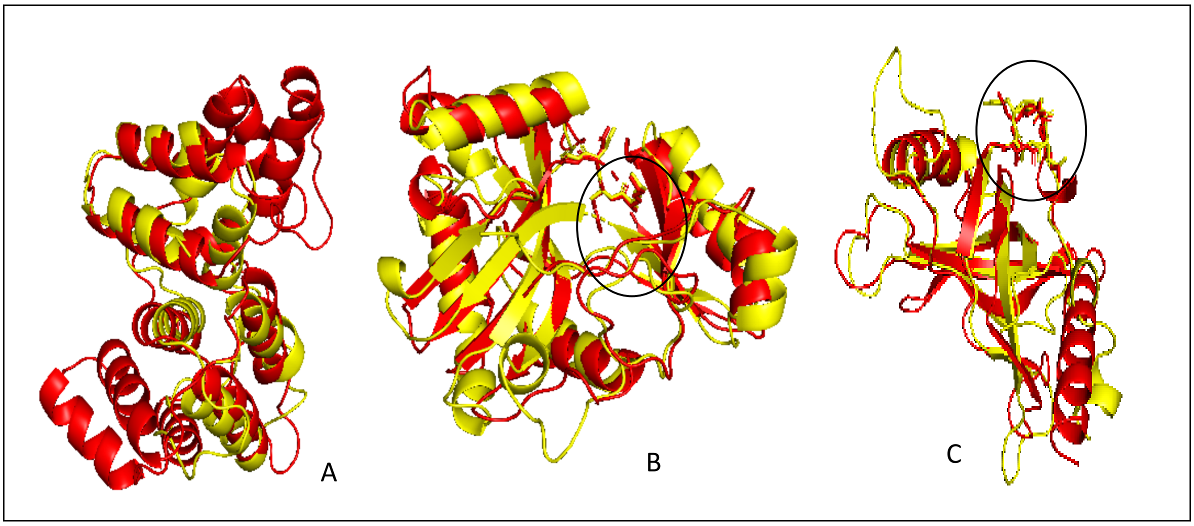


Figure 2.3: Structure alignment of (A)NBD (B)AdD (C)OBD of T4 (Yellow) and P.furiosus (Red) DNA ligase (Circles show well aligned residues responsible for activity)

Thereby, these observations point out that the temperature specificity of the activity of *P. furiosus* DNA ligase can be changed by removing or substituting the C-terminal helix of the enzyme. Using this as an inspiration, two kinds of engineering of *P. furiosus* DNA ligase were done to get mesoactive thermostable DNA ligases.

1. Chimeric PfuT4H DNA ligase

Studies have shown that performing deletion or neutralisation of few amino acids in the C-terminal helix of *P.furiosus* DNA ligase increases the catalytic activity of the enzyme by destabilising the closed conformation formation (Tanabe et al., 2014). This may have been caused due to destabilization of the closed conformation of the ligase, making it open for sealing nicks in DNA by being in an adenylated state in open conformation at a broader range of temperatures assuming that the adenylation would have happened at 37°C while being over expressed in bacterial cells.

Inspired by above observations, a chimeric *P.furiosus* ligase was designed by replacing the C-terminal helix of *P.furiosus* DNA ligase with the neutral C-terminal helix of T4 DNA ligase. As all six motifs has been shown to be conserved among all ATP using ligases, the C terminal helix after motif VI conserved region (RXDK amino acids) was replaced with corresponding amino acids of T4 DNA ligase with the hope that the neutral C-terminal helix in the chimeric enzyme would be unable to hold the closed conformation (required for ATP transfer) at high temperature and would become mesoactive. As shown in **Figure 2.4**, unlike the *P.furiosus* WT DNA ligase (whose closed conformation is stable), the chimeric PfuT4H will remain in open conformation (due to replacement of *P.furiosus* C-terminal helix with T4 DNA ligase helix) at room temperature allowing it to be ready to accept substrate at room temperature. The amino acid composition of the engineered sequence had only one positively charged amino acid in it. Thereby, the chimeric PfuT4H ligase is expected to work at lower temperatures while retaining its thermostability.

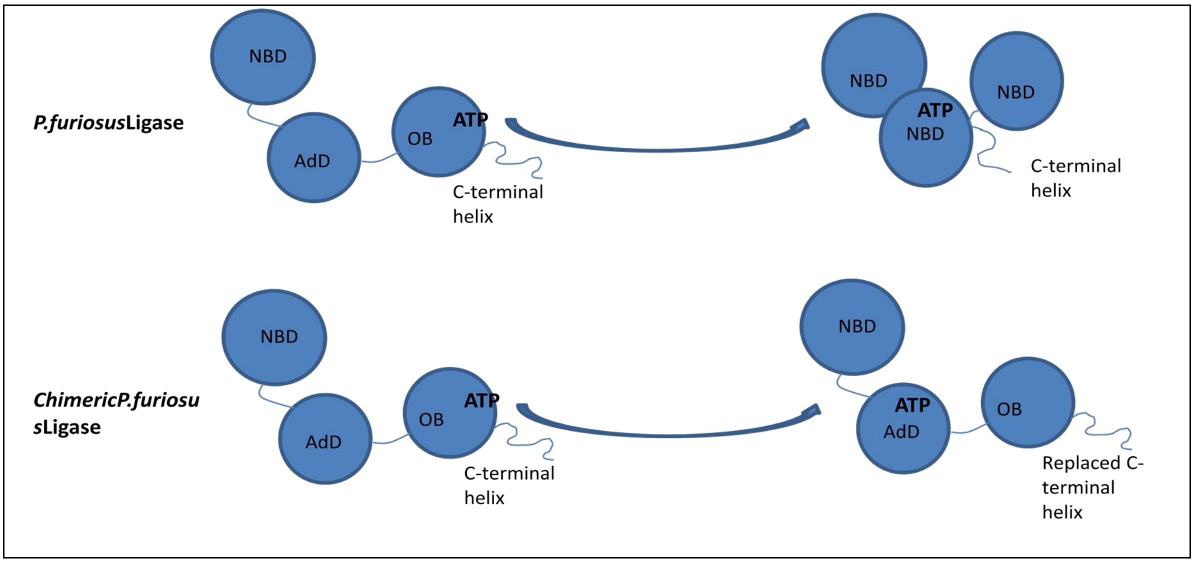

P.furiosusLigase
NBD
AdD
OB
ATP
C-terminal helix
NBD
ATP
NBD
NBD
C-terminal helix
ChimericP.furiosusLigase
NBD
AdD
OB
ATP
C-terminal helix
NBD
ATP
AdD
OB
Replaced C-terminal helix

Figure 2.4: Schematic representation of hypothesis behind the PfuT4H engineering

2. ΔPfu DNA ligase

In certain ligases, like the T7 DNA ligases only two domains are present namely the AdD and OBD lacking motif VI. The T7 DNA ligase is also an ATP dependent ligase which is as efficient as T4 DNA ligase in sticky end ligation at room temperature even though it lacks a part of the OB domain and N-terminal domain **(Figure 2.5**) (Subramanya et al., 1996).

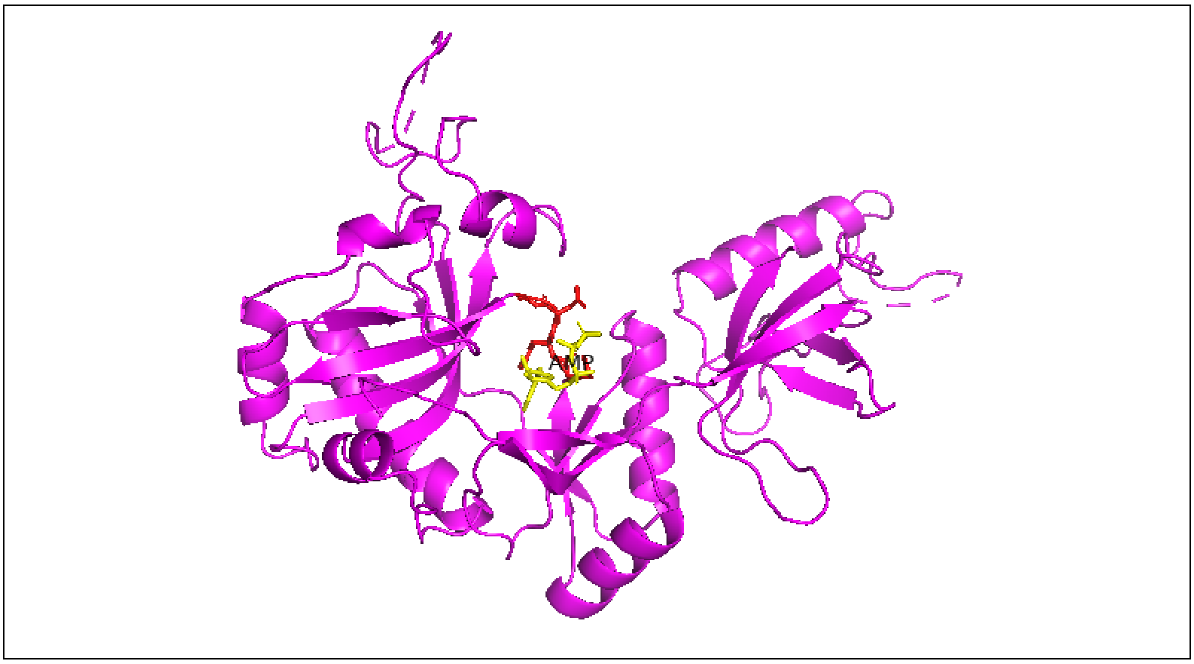


Figure 2.5: The structure of T7 DNA Ligase

The strong binding with DNA facilitated by the extra N-terminal DBD in certain ligases offers them advantage in blunt end ligation. On the other hand, the same DBD could also make them catalytically slow as the nick sensing and ligation is preceded by a process of random association dissociation cycles. In order to locate the nick in DNA backbone, the OBD scans the structure by twisting the DNA backbone to ease the supercoiling. As the OB fold locates the nick, a stable complex is made followed by phosphodiester bond formation and dissociation of the DNA and the enzyme (Doherty & Dafforn, 2000; John M. Pascal et al., 2004). Enzymes lacking the DBD may be able to scan the DNA backbone faster than the three domain ligases. This would result in faster reaction rate especially at very low temperature conditions where low temperature compensates for the stronger electrostatic interaction of DBD and C-terminal helix with DNA.

Inspired by these observations, a T7 like ΔPfu ligase was designed by deleting N-terminal domain along with the C-terminal alpha helical extension by keeping RXDK amino acid stretch on motif VI intact for adenylation reaction **(Figure 2.6)**. Expectantly, the ligase would efficiently transfer ATP and scan DNA faster at room temperature resulting in more activity while still being thermostable. However, this truncated ligase would lose its activity at higher temperature due to lack of stable closed complex formation required for ATP transfer and the reduction in DNA binding affinity.

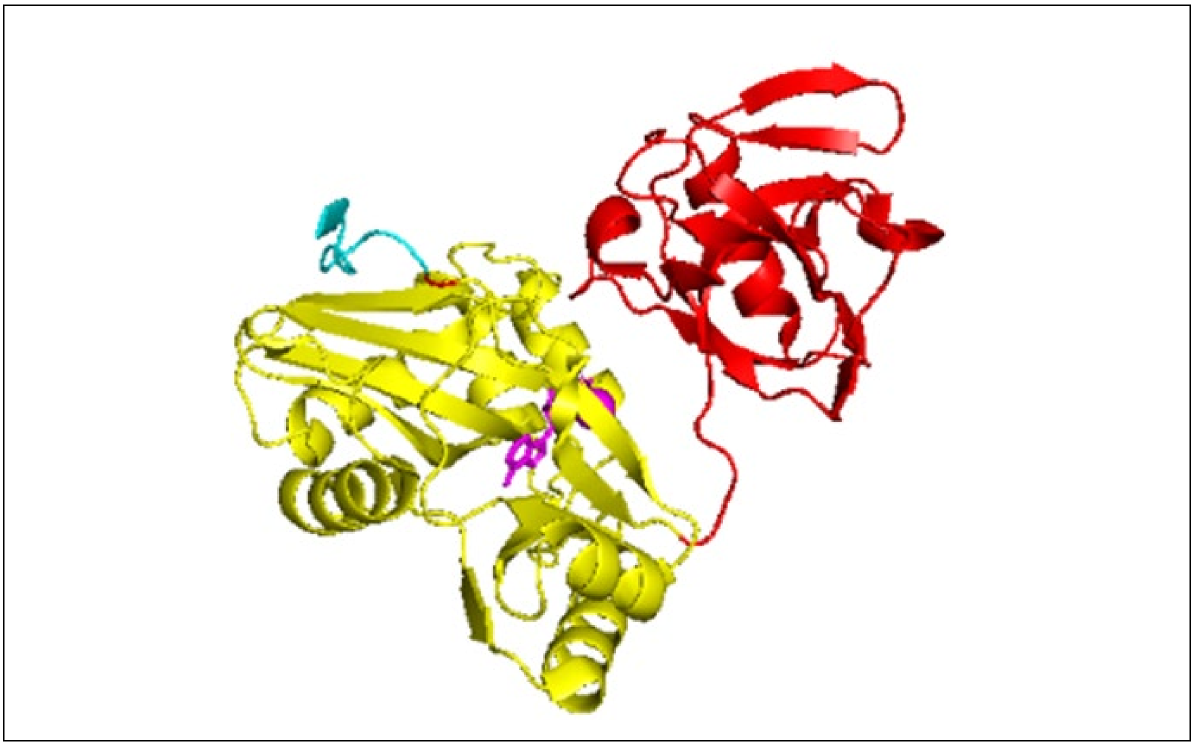


Figure 2.6:ΔPfu ligase structure (Red – Truncated OBD, Yellow- AdD)
(Without N-terminal DNA binding domain and C-terminal helix)

AIM OF PROJECT

The aim of my project was to recombinantly express and purify ΔPfu DNA ligase and PfuT4H DNA ligase, structurally characterize them and check their enzymatic activity compared to commercially used T4 DNA ligase.

2.2 MATERIALS AND METHODS

2.2.1 Materials

Clones Used

Cloned gene	Vector	Host cell	Gratefully soured from
ΔPfu ligase	pQE-30	XL1-Blue	Neeraj Dhaunta & Bhisham Thakur
PfuT4H ligase	pET28a	*BL21*(DE3)*pLysS*	

Buffers for native Ni-NTA affinity chromatography purification for 6X-His tag proteins

Qiagen non-denaturing buffers with the following composition were used.

Buffer	Component		
	Tris-Cl pH8.8 (mM)	NaCl	Imidazole (mM)
Lysis	50	1M	10
Wash	20	2M	20
Elution	20	200mM	250

Buffers and column used for Anion Exchange Chromatography

Buffer	Component	
	Tris-Cl pH 8.8 (mM)	NaCl (M)
A	20	-
B	20	1

Buffers and solutions for SDS-PAGE

Acrylamide

Acrylamide	30 g
N,N'-Methylene - bisacrylamide	0.8 g
Deionized Water	100 mL

Ammoium Persulphate (10%, APS)

APS	100 mg
Deionized Water	1 mL

Lower Tris (4X), pH 8.8

Tris	18.17 g
10% SDS	4 mL
Deionized Water	100 mL

Upper Tris (4X), pH 6.8

Tris	6.06 g
10% SDS	4 mL
Deionized Water	100 mL

5X Sample Loading Buffer

Tris-Cl (pH 6.8)	0.15 M
SDS	5%
Glycerol	25%
β-mercaptoethanol	12.5%
Bromophenol blue	0.06
Deionized water	10 mL

Gel Staining Solution

Methanol	40%
Glacial acetic acid	10%
Coomassie Brilliant Blue R-250	0.1%
Deionized water	50%

Gel Destaining Solution

Methanol	40%
Glacial acetic acid	10%
Deionized water	50%

13% SDS-PAGE composition

Percent	Stacking gel	Resolving gel
Tris (Upper and lower resp.)	2.5 ml	0.5 ml
Acrylamide	4.333 ml	0.333 ml
Water	3.067 ml	1.167 ml
APS	50 µl	12.5 µl
TEMED	10 µl	5 µl

Running Buffer (10X)

Tris base	30.0 g
Glycine	144.0 g
SDS	10.0 g
Water	1000 ml

Chemical Unfolding

Denaturant	Amount	Water
8M urea	48 gm	100 ml
8M GdmHCl	76.4 gm	100 ml

Buffers and solutions for Urea-PAGE (Summer, Grämer, & Dröge, 2009)

10X TBE (Tris/Borate/EDTA) Buffer (pH 8.3)

Component	Concentration	Amount
Tris Base	1M	12.1g
Boric Acid	1M	6.18 g
EDTA	0.02 M	0.74 g
Water	-	100 ml

15% Urea gel

Component	Amount
Urea	4.8 gm

10X TBE buffer	1 ml
30% acrylamide	5 ml
10% APS	100 µl
TEMED	5 µl

2X loading dye

Glycerol	50%
Bromophenol blue	0.1%
Xylene cyanol	0.1%

1X TBE used as running buffer

Ligation Assay

Oligonucleotides used (Bauer, Jurkiw, Evans, & Lohman, 2017)

Oligos	Sequence
DNA-OH	5’GCGCACCCTTACCACCAAGACAGGATCGTCCTTGC 3’
p-DNA-FAM	5’TGATCATGCATCGTTCCACTGTGTCCGCGACATCTACGTC 3’
Splint DNA	5’GACGTAGATGTCGCGGACACAGTGGAACGATGCATGATCAGCAAGGACGATCCTGTCTTGGTGGTAAGGGTGCGC 3’

The oligos were suspended in TE buffer (10 mM Tris-Cl, 1 mM EDTA, pH 8.0) with final concentration of 100 µM.

10X Oligonucleotide annealing buffer (OAB)

Tris-HCl (pH 7.5)	100mM
KCl	500mM
EDTA	10mM

Ligation Reaction buffer

P.furiosus DNA ligase buffer (10X) (Tanabe et al., 2014)

Tris-HCl (pH 7.5)	500mM
KCl	200mM
MgCl2	100mM
Igepal	1%
DTT	10mM
ATP	1mM

Commercially used T4 DNA ligase and its 10X buffer from NEB were used as control for ligation assay.

2.2.2 Methods

Expression of ΔPfu ligase and PfuT4H ligase

ΔPfu ligase - An 10 ml primary culture of XL1-Blue cells containing ΔPfu ligase in pQE30 plasmid was incubated overnight at 37˚C. 1% of inoculum from primary culture was added to 400 ml LB media with appropriate antibiotics and the culture was incubated for 12 hours at 37˚C.

PfuT4H ligase – 10 ml primary culture of *BL21*(DE3)*pLysS* cells containing PfuT4H ligase in pET28a plasmid was incubated overnight at 37˚C. 1% of inoculum from primary culture was added to 400 ml LB media with appropriate antibiotics. At O.D. 0.6, 1mM IPTG induction was done and culture was incubated at 37˚C for 5 hours.

Purification of ΔPfu ligase and PfuT4H ligase

i. After over-expression, the cells were harvested at 8000 rpm for 10 minutes at 25˚C.

ii. The pellet was resuspended in 20 ml of native lysis buffer.

iii. The mix was subjected to sonication followed by boiling at 85˚C for 30 minutes. Then the lysate was centrifuged at 11000 rpm for 40 minutes at 10˚C. The supernatant was separated from the pellet.

iv. To purify the protein of interest with 6X-His tag, Ni-NTA affinity chromatography was used-

a. The recharged column was washed by passing 3 column bed volume of distilled water.
b. To equilibrate the column, 1 column volume of lysis buffer was allowed to pass through the column.
c. The supernatant was allowed to pass through the column for binding of proteins with His-tag to the Ni-NTA beads.
d. To wash away non-specific proteins, 3 column volume of wash buffer was allowed to pass.

e. To elute the protein, 6ml of elution buffer was added to the column and the elute was collected in a falcon tube.

v. Elute, wash, supernatant and pellet were loaded on SDS-PAGE with 1X protein loading dye and the purification was checked. Further the protein was stored at 4˚C.

Further, the elute was subjected to Ion exchange chromatography using HiTrap SP-HP anion exhange column by applying a linear gradient of 1M NaCl.

Structural Characterisation of ΔPfu ligase and PfuT4H ligase

(A.)UV-Vis/ Absorption spectroscopy

UV-Visible spectroscopy was used to find the concentration of the protein sample. The absorbance plot can also tell us about the aggregated samples of protein. The absorbance was noted at 280 nm and concentration was calculated according the Lambert-Beer law which is defined by the following formula

$$\mathbf{A=\log 10\ (Io/I)= \epsilon * l * c}$$

where **Io** = Incident intensity; **I** = Transmitted intensity; **C** = concentration ; **ϵ** = extinction coefficient; **l** = Path length

Varian Cary 50 UV-Visible spectrophotometer was used for taking readings. The reading was taken from 250 to 600 nm and baseline correction was employed. ϵ for ΔPfu ligase was 28590 and for PfuT4H ligase was 44520 M^{-1} cm^{-1} (Information obtained using ProtParam). The path length was 0.3 cm.

(B.) CD Spectroscopy

Circular Dichroism spectroscopy is used to study secondary structure of proteins. The signal arises due to differential absorption of left and right circularly polarised light by the peptide bonds in "far-UV" spectral region (190-250 nm). The signal is sensitive to the environment of the peptide bonds giving rise to characteristic spectrum for alpha-helix, beta-sheet, and random coil **(Figure 2.7)**.

Applied photophyscis chirascan CD instrument was used to carry out CD spectroscopy to study the structure of the purified ligases and effect of temperature and chemical denaturants on the engineered ligases.

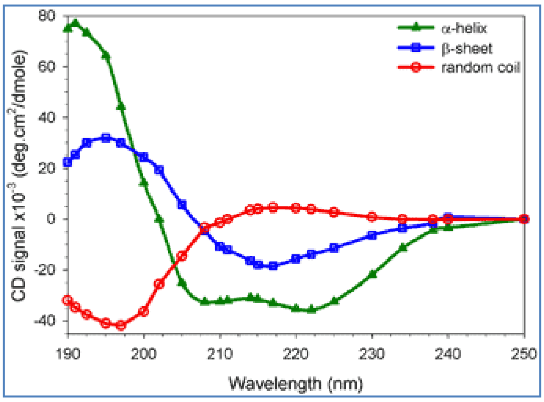


Figure 2.7: Characteristic secondary structure peak of protein in CD spectroscopy

The epsilon values (after baseline correction) obtained from the CD spectra are converted to MRE by using the following formula:

$$MRE = \frac{(\boldsymbol{\theta} \times 100 \times MRW)}{1000 \times c \times l}$$

where **MRW**= Molecular weight of proteins/No. of amino acids (for ΔPfu ligase and PfuT4H ligase MRW are 114.42 and 113.40 Da respectively); **c** = concentration of protein in mg/ml; **l** = path length of cuvette (0.1cm here)

The concentration of protein used for all the CD experiments was 0.15 mg/ml. For studying thermal stability, the proteins were subjected to heat form 25 to 90^{0}C using a ramping rate of 3 degress per minute and noting the secondary structure. For studying chemical stability, protein was incubated overnight at room temperature with varying concentrations of urea and GdmHCl from 0.25 M to 6M and the secondary structure for each of them was estimated through CD. Finally, data was plotted in MRE against wavelength using ORIGIN software. The CD spectrum was analysed using K$_2$D$_2$ website (Perez-Iratxeta & Andrade-Navarro, 2008).

Ligation assay for checking enzymatic activity of ΔPfu ligase and PfuT4H ligase

Annealing of Oligonucleotides to make double-stranded DNA with nick

Component	Using OAB(B)		Using water(W)	
	Amount	Final Concentration	Amount	Final Concentration
DNA-OH	5 µl	10 µM	3 µl	3000nM
p-DNA-FAM	5 µl	10 µM	3 µl	3000nM

Splint DNA	5 µl	10 µM	3 µl	3000nm
10X OAB	3µl	1X	-	-
Water	100µl	-	100 µl	-

The mixture was heated to 95ºC and then gradually cooled to 25ºC in a series of steps with step size of 10 degrees, each for 10 minutes. After dilution, a concentration of the annealed oligos of 60nM was achieved.

Ligation Reaction

Ligation assay was carried out to check the enzymatic activity of the engineered ligases. Annealed double stranded DNA (with 6-FAM labelling on 3' end of one of the strands) with a nick on one strand was used as substrate to be ligated. T4 DNA ligase from NEB was used as control. For all the reactions, both commercial T4 DNA ligase buffer from NEB (referred to as T) and the *P.furiosus* DNA ligase buffer (referred to as P) were used.

Component	**Amount**	**Final Concentration**
Annealed oligos	2µl	120 nM
Buffer	0.5 µl	1X
T4 DNA ligase	0.3 µl	120 units
Deionised Water	2.2 µl	-

ΔPfu ligase and PfuT4H ligase

Component	Amount	Final Concentration
Annealed oligo	2µl	120nM
Buffer	0.5 µl	1X
Ligase	0.5 µl	0.066mg/ml(for PfuT4H ligase) 0.053 mg/ml(for ΔPfu ligase)
Deionised Water	2 µl	-

Ligation reactions were set up at 25ºC and 45ºC overnight. Next day, samples were run on urea PAGE gel to denature the DNA strands and fluorescence from FAM was visualized using LAS 4000.

2.3 RESULTS AND DISCUSSION

2.3.1 Expression, Purification and characterization of ΔPfu ligase and PfuT4H ligase

2.3.1.1 ΔPfu ligase

(A.) Expression and Purification

Ni-NTA affinity chromatography of a 400ml culture of ΔPfu ligase clone grown for 12 hours at was carried out. The eluted protein showed the correct size (39.1kDA) with some additional bands of contaminants on SDS-PAGE (**Figure 2.8**).

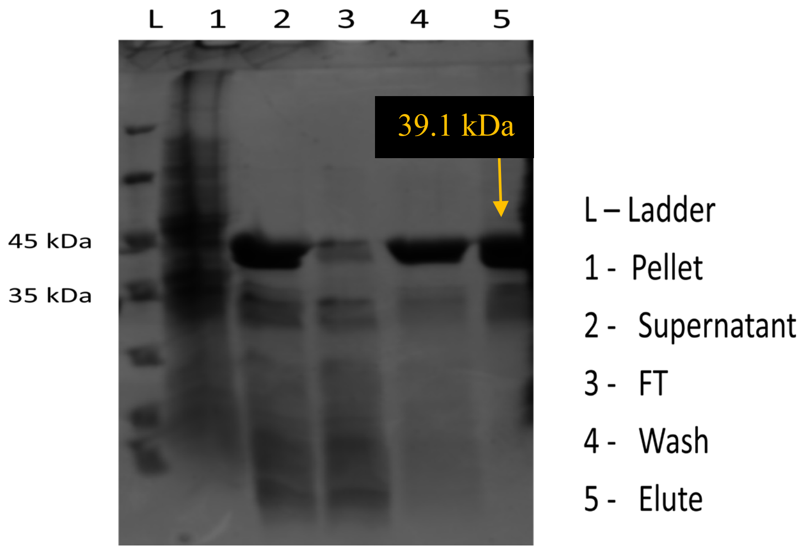

L
1
2
3
4
5
39.1 kDa
45 kDa
35 kDa
L – Ladder
1 - Pellet
2 - Supernatant
3 - FT
4 - Wash
5 - Elute

Figure 2.8: SDS-PAGE of purified ΔPfu ligase using Ni-NTA chromatography

The 6ml of eluted protein was subjected to anion-exchange chromatography by applying a linear gradient of 1M NaCl. The recombinant protein eluted at 350mM of NaCl **(Figure 2.9).**

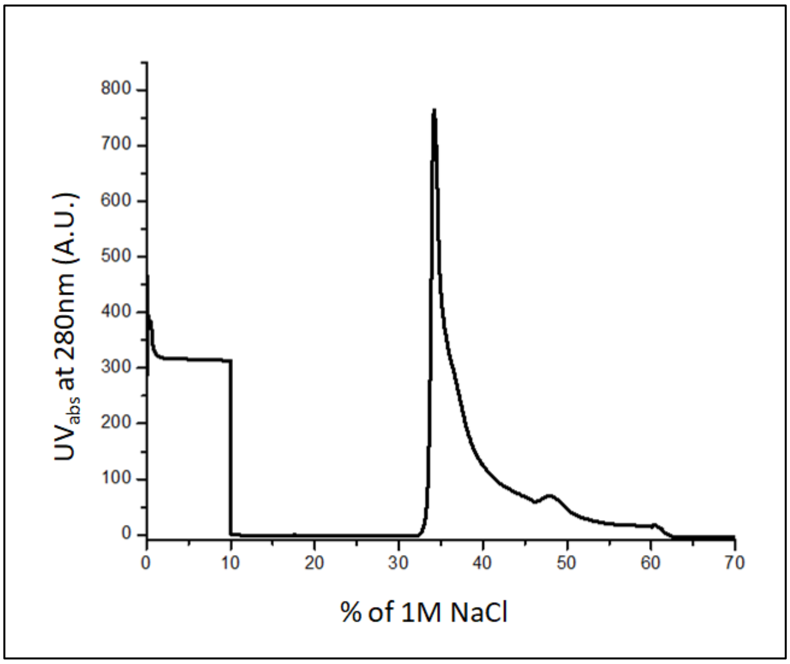


Figure 2.9: Anion exchange chromatogram of Ni-NTA purified ΔPfu ligase

When checked on SDS-PAGE for correct size and purity, contaminants appeared in the same way as they appeared on SDS-PAGE after Ni-NTA purification **(Figure 2.10).**We believe the contaminants to be proteolytic fragments of the recombinant protein which were present as they were strongly held with the protein by non-covalent interaction. These fragments got separated in the SDS-PAGE and appeared as contaminants.

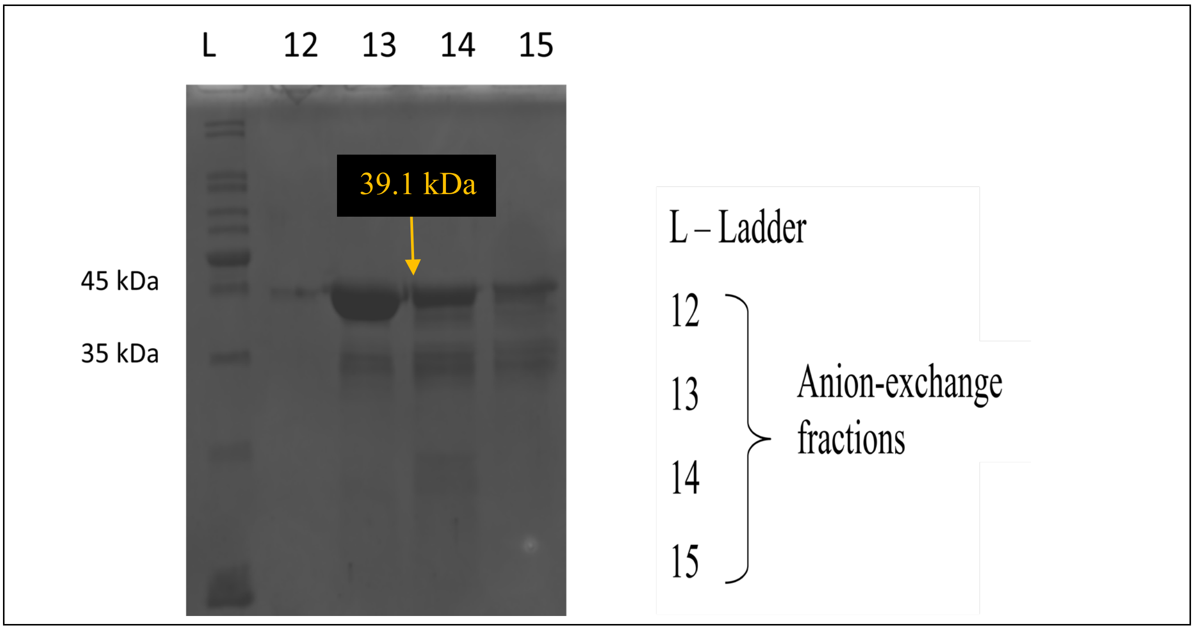

L
12
13
14
15
39.1 kDa
45 kDa
35 kDa
L – Ladder
12
13
14
15
Anion-exchange fractions

Figure 2.10: SDS-PAGE of anion exchange fractions of ΔPfu ligase

Concentration of purified protein was calculated to be 0.53 mg/ml using absorption spectroscopy. The flat line of absorption beyond 350 nm suggested that there was no aggregation in the purified protein (**Figure 2.11).**

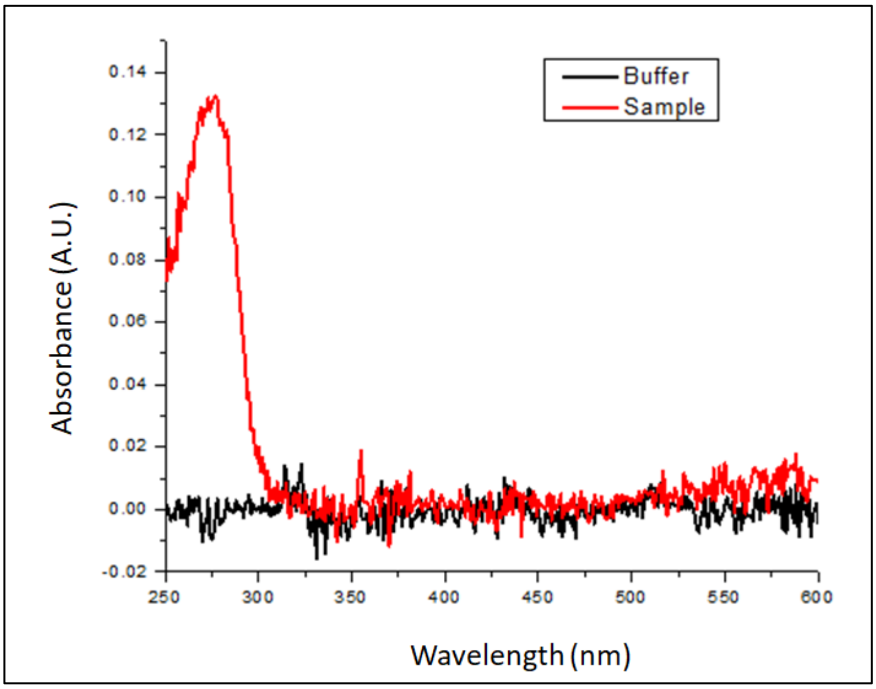


Figure 2.11: Absorption spectrum of ΔPfu ligase

(B.) Secondary structure determination of ΔPfu ligase

The CD spectrum of ΔPfu ligase shows the structure to be a mixture of alpha helices and beta sheets, predominated by beta sheets (**Figure 2.12**). Due to truncation of the N-terminal DNA binding domain, and C-terminal helix , the overall alpha helix fraction has decreased as compared to *P.furiosus* WT DNA ligase (which has 43% alpha helices and 22% beta sheets). When analysed using K_2D_2 website, the well folded structure of ΔPfu ligase was observed to be consisting of ~11% alpha helices and ~36% beta sheets with a maximum of 0.38% error.

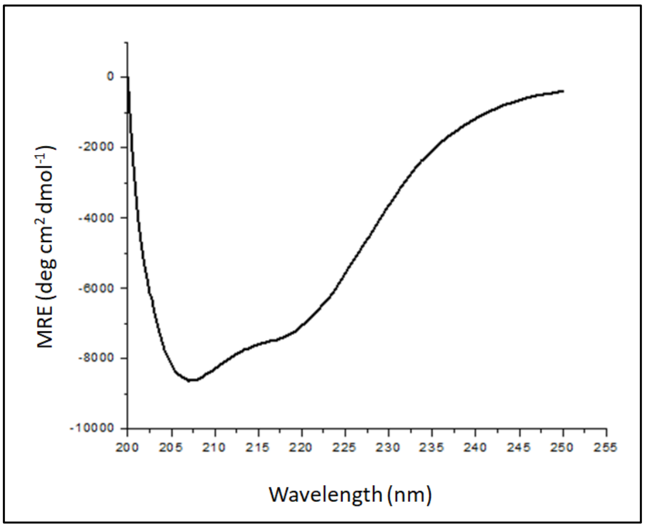


Figure 2.12: CD spectrum of ΔPfu ligase showing its secondary structure

(C.) Thermal and chemical stability estimation of ΔPfu ligase

On checking the thermodynamic stability of ΔPfu ligase using CD spectroscopy, it was observed that there were minimal changes in the MRE values across temperatures ranging from 20 to 85˚C (**Figure 2.13)**. This protein having a hyperthermophilic origin, even after getting truncation showed high thermostability.

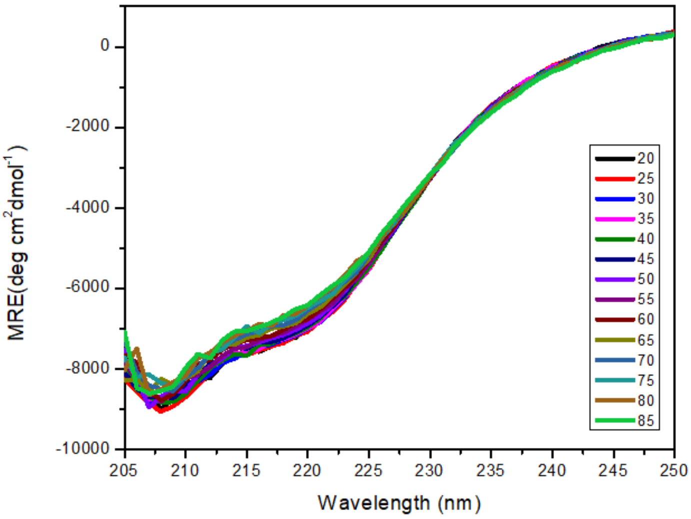
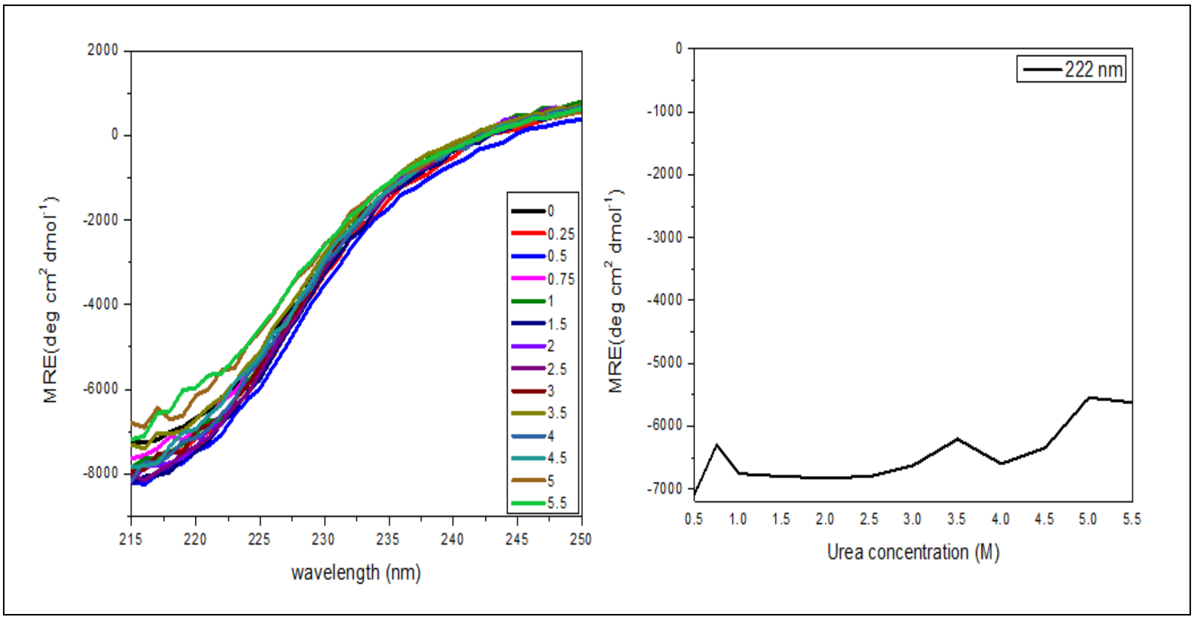
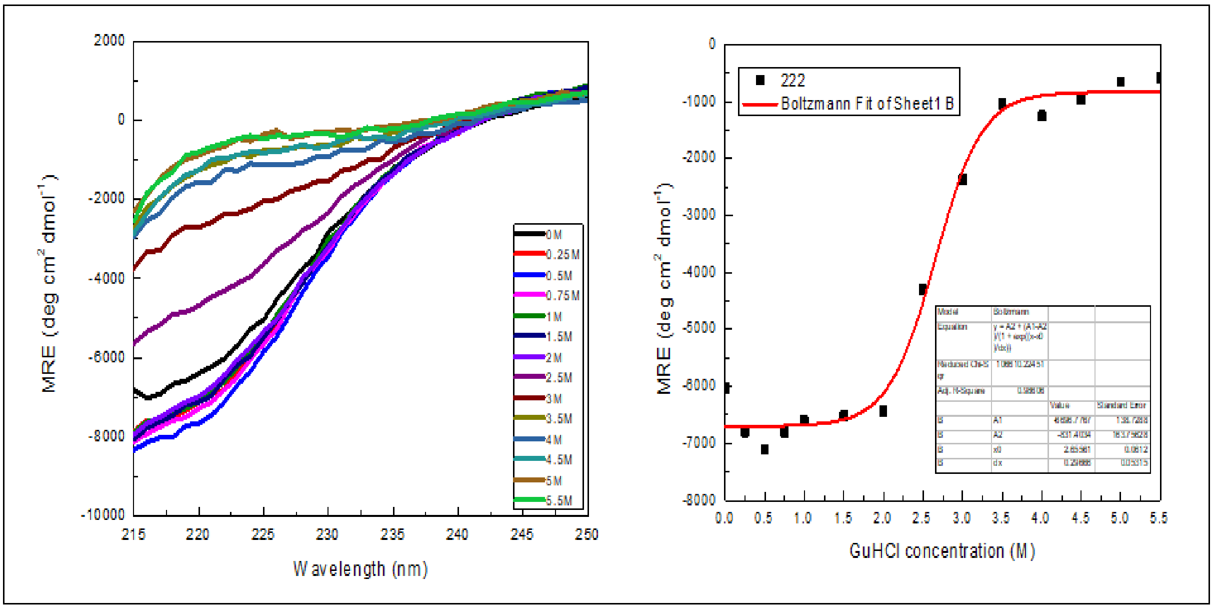


Figure 2.13: CD spectrum showing thermal melt of ΔPfu ligase

To check the chemical stability of ΔPfu ligase, the protein was subjected to different concentration of urea and GdmHCl ranging from 0.25M to 6M (**Figure 2.14).**It was observed that the protein was resistant to denaturation by urea. However, GdmHCl was able to denature the protein to a great extent. The Cm value (concentration of denaturant at which half of population of protein is unfolded) was found to be ~3M. Thermostable proteins contain extra electrostatic interactions in addition to what mesophilic proteins have, which gives them extra stability. GdmHCl in this case is a more potent denaturant than urea as it breaks the electrostatic interaction along with hydrogen bonds. A gain in structure is also observed in very low concentrations of GdmHCl (upto 0.75M). At low concentration, GdmHCl might be acting as a salt and stabilizing the structure by ionic interaction on the surface of the protein.



(a.) In presence of Urea



(b.) In presence of GdmHCl

Figure 2.14: CD spectra and MRE plots at 222nm of chemical denaturation of ΔPfu ligase

2.3.1.1 Chimeric PfuT4H ligase

(A.) Expression and Purification

IPTG induction of 400 ml culture was done to get over expression of PfuT4H ligase. As for ΔPfu ligase, Ni-NTA affinity chromatography and anion exchange chromatography was done in order to get pure protein **(Figure 2.15)**. The protein eluted with 380mM NaCl. The eluted protein showed the correct size (62.8 kDA) with some additional bands of contaminants on SDS-PAGE (**Figure 2.16).** These bands correspond to proteolytic fragments of the protein as explained earlier. The concentration of purified protein was calculated to be 0.66 mg/ml.

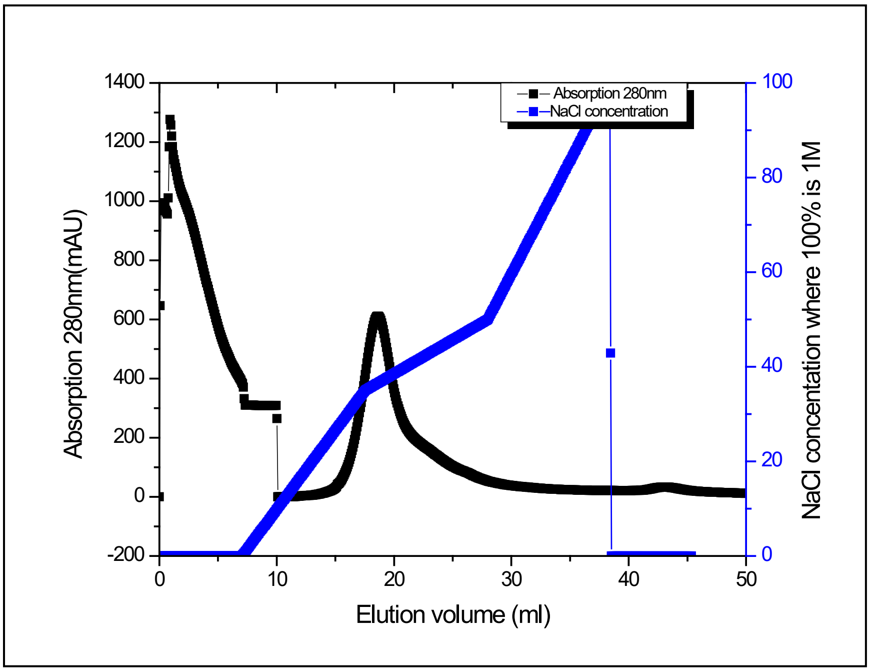


Figure 2.15: Anion exchange chromatogram of Ni-NTA purified PfuT4H ligase

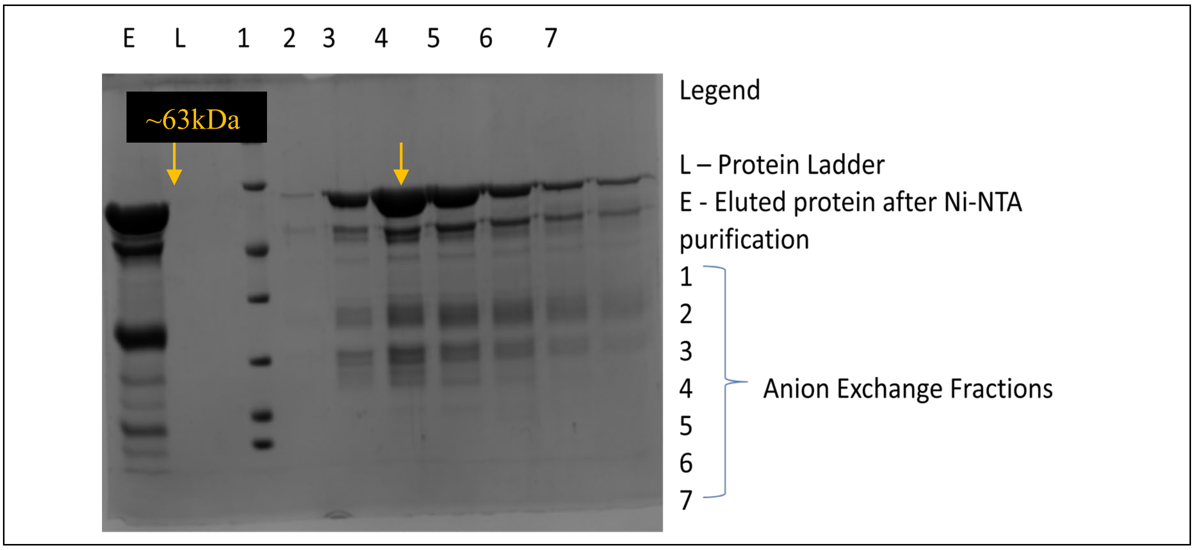

E
L
1
2
3
4
5
6
7
~63kDa
Legend
L – Protein Ladder
E - Eluted protein after Ni-NTA purification
1
2
3
4
5
6
7
Anion Exchange Fractions

Figure 2.16: SDS-PAGE of Ni-NTA purified and anion exchange fractions of PfuT4H

(B.) Secondary structure determination of PfuT4H

The peak at 222, 217 and 208 nm has a negative value of ~ 11000-12000 MRE , suggesting contributions from both alpha helices and beta sheets. As the secondary structure composition of PfuT4H is similar to the *P.furiosus* WT DNA ligase (which has 43% alpha helices and 22% beta sheets), the CD spectrum shows the well folded structure of protein with ~17 and ~28 percent contribution of beta sheets and alpha helices respectively with maximum error of 0.4% (calculated using K_2D_2) **(Figure 2.17)**.

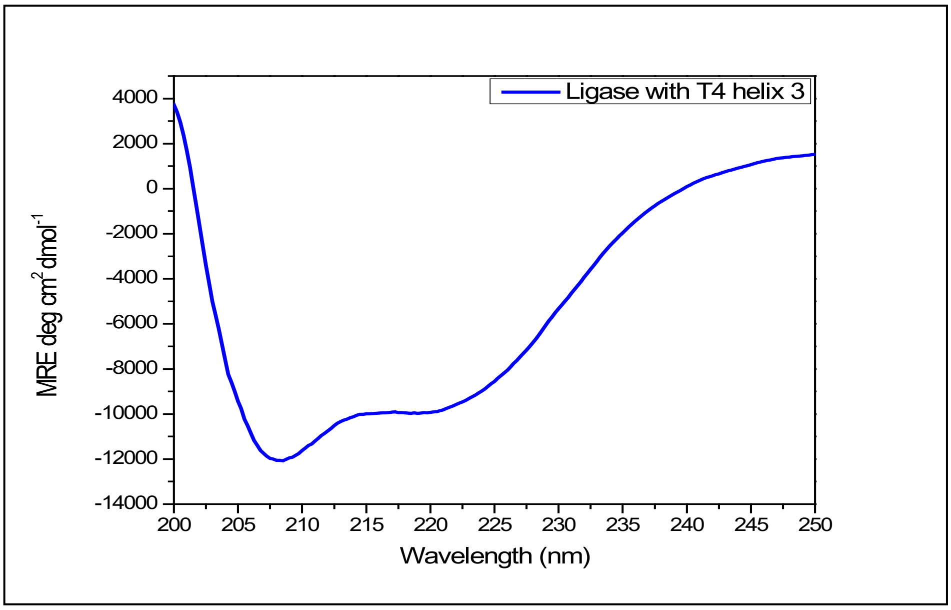


Figure 2.17: CD spectrum of PfuT4H showing its secondary structure

(C.) Thermal and chemical stability estimation of ΔPfu ligase

PfuT4H was observed to be stable across temperatures ranging from 25 to 85˚C with minimum changes in structure as evident from the millidegree value at 222nm for all temperatures (**Figure 2.18**). This is expected as it has a hyperthermophilic origin.

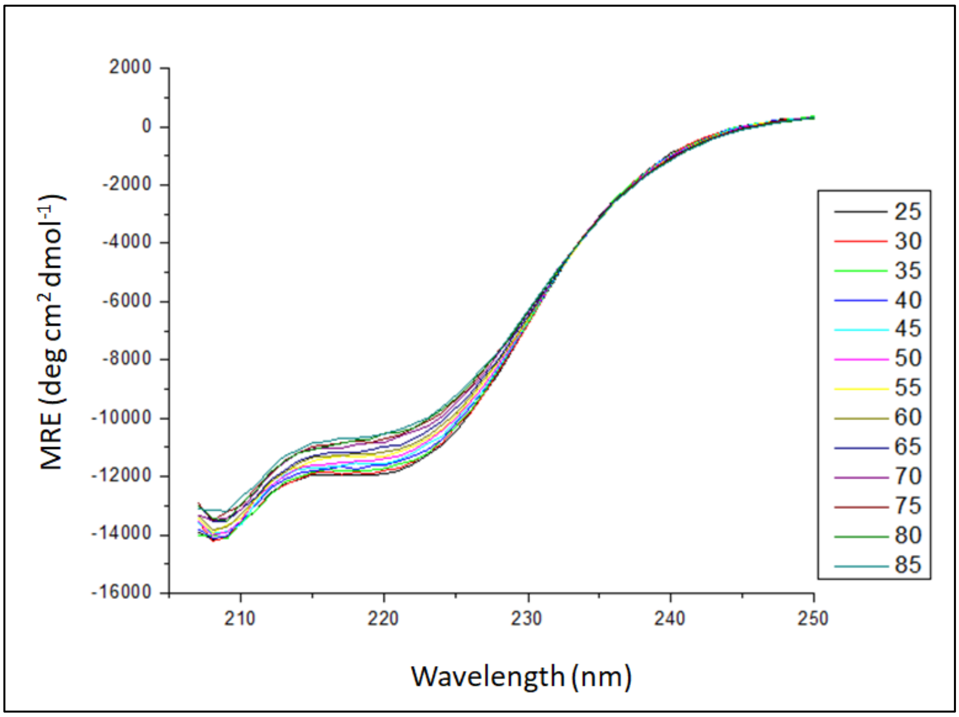
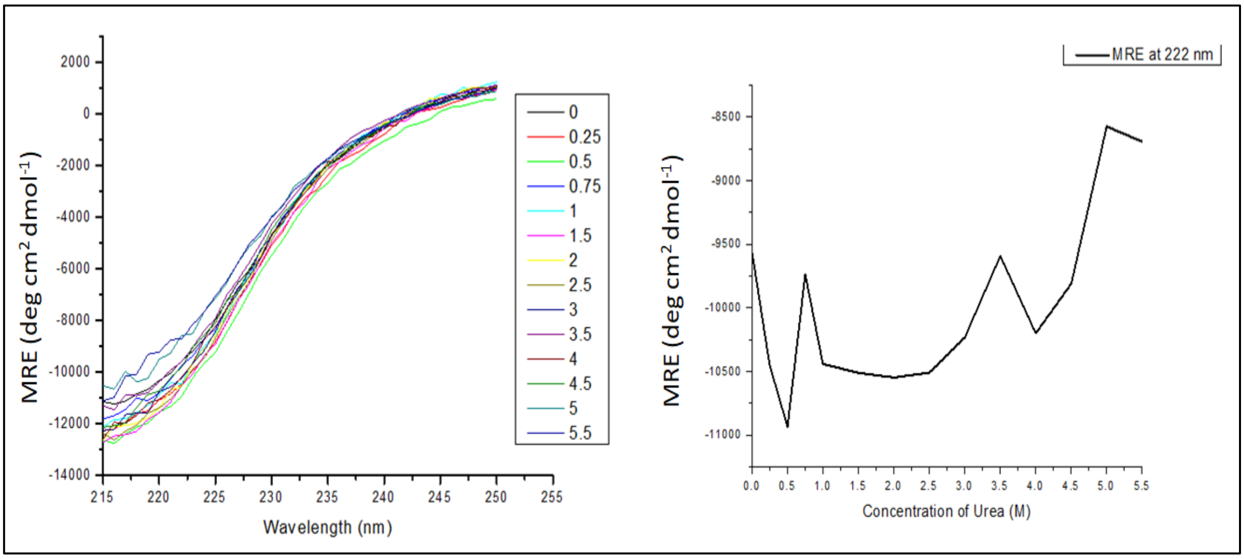
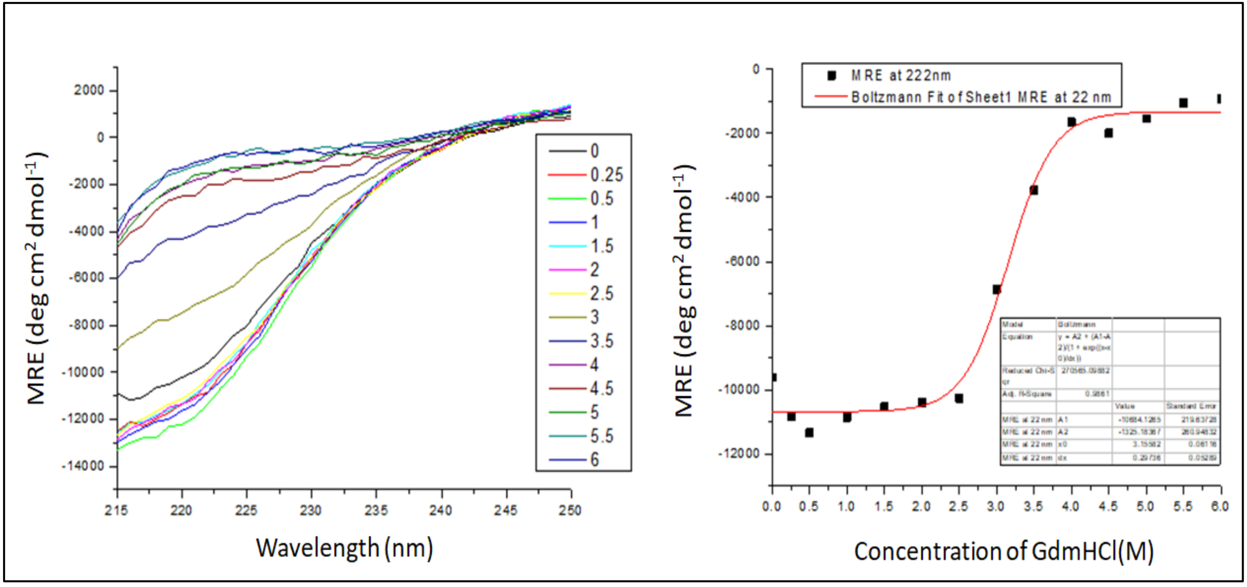


Figure 2.18: CD spectrum showing thermal melt of PfuT4H

The protein was subjected to different concentration of urea and GdmHCl ranging from 0.25M to 6M for checking chemical stability (**Figure 2.19)**. As in the case of ΔPfu ligase, GdmHCl was able to denature the protein at higher concentrations whereas urea couldn't. The Cm value was found to be ~3M for PfuT4H. As explained earlier, GdmHCl works better as a denaturant in this case as it breaks the ionic interactions that give hyperthermophilic proteins "the extra" stability. A gain in structure is also observed in very low concentrations of GdmHCl (upto 0.75M) because of the same reasons as explained in section 2.3.1.1 (C.).



(a.) In presence of Urea



(b.) In presence of GdmHCl

Figure 2.19: CD spectra and MRE plots at 222nm of chemical denaturation of PfuT4H ligase

2.3.2 Comparison of enzymatic activity of ΔPfu ligase and PfuT4H ligase with T4 DNA ligase

The activity of the ΔPfu ligase and PfuT4H ligase was determined qualitatively by ligation assay using FAM labelled oligos. To check the activity under different conditions, two parameters were varied – Oligos and ligation buffer. Both, oligos reconstituted in water (referred to as W) and in binding buffer (referred to as B) were used. Further, both 1x T4 ligation buffer (referred to as T) and 1x *P.furiosus* DNA ligase ligation buffer (referred to as P) were used **(Table 2.1)**. The activity was compared with that of commercially used T4 ligase (used as control).

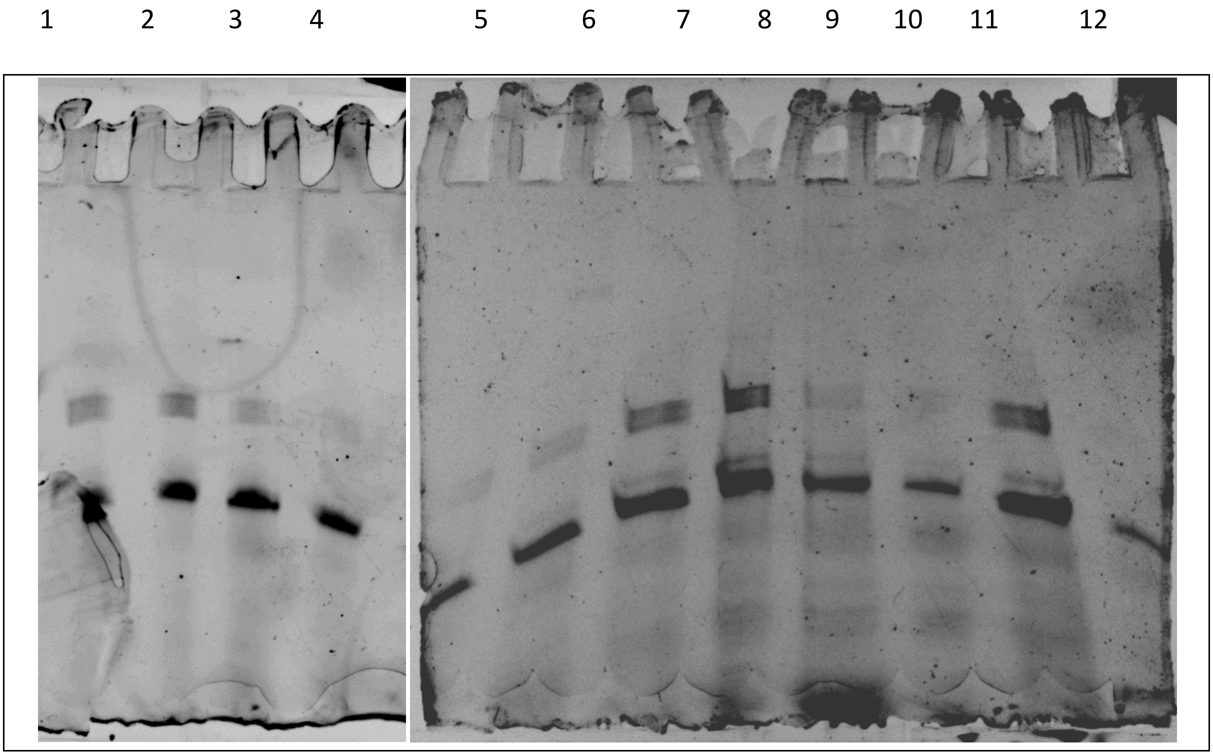


Figure2.20: Urea PAGE gel image of ligation assay at 25ºC

Ligase	Lane	Template used	Buffer used
T4 Ligase	1	W	T
	2		P
	3	B	T
	4		P
ΔPfu ligase	5	W	T
	6		P
	7	B	T
	8		P
PfuT4H ligase	9	W	T
	10		P
	11	B	T
	12		P

Table 2.1: Different components used in ligation assay

A successful ligation reaction would create longer FAM labelled oligos as compared to an unsuccessful ligation reaction, wherein the FAM labelled oligos would remain shorter in length. Since, only a fraction of ligation reactions are successful, shorter FAM labelled oligos were always expected to be present. On the urea gel after electrophoresis, the successfully ligated longer oligos would appear on top of the shorter oligos which did not get ligated.

At 25ºC, as seen in **Figure 2.20,** for commercial T4 ligase, we observed faint bands on top of the intense bands suggesting that a fraction of ligation of the oligos was successful. When compared with that of ΔPfu ligase and PfuT4H ligase, we observed more intense bands in the top half. The bands at the bottom show the same intensity. Since, concentration of reconstituted oligos was same in every reaction, we can infer qualitatively that there were more ligation products when ΔPfu ligase and PfuT4H ligase were used instead of T4 ligase. At 45ºC, for both the engineered ligases, the amount of ligated product was lesser than 25 ºC (results not shown here).

2.3.3 Concluding Remarks

When compared with T4 ligase for activity under the same conditions, both engineered ligases were observed to have better activity.

The new truncated ΔPfu ligase behaved like a mesophilic enzyme similar of T7 DNA ligase. The enzyme retained its thermostability and showed nick sealing activity at room temperature.

The new chimeric PfuT4H ligase was also found to be thermostable and had optimum activity at room temperature i.e. it was mesoactive.

2.4 FUTURE OUTLOOK

To understand whether the engineered enzymes work better than the *Pyrococcus furiosus* WT DNA ligase, same ligation experiments have to be done with the *P.furiosus* DNA ligase as another control. Further, the engineered enzymes can be subjected to ligation activity at higher temperature to check whether they are thermoactive as well.

Finally, real ligation experiments to make clones with different concentration of engineered ligases and varied buffer compositions need to be carried out to check their potential in replacing commercially used ligases.

Bibliography

Alexander Goldberg, P. D. (2019). A Brief History of PCR and Its Derivatives. Retrieved from https://blog.labtag.com/a-brief-history-of-pcr-and-its-derivatives/

Andlar, M., Rezić, T., Marđetko, N., Kracher, D., Ludwig, R., & Šantek, B. (2018). Lignocellulose degradation: An overview of fungi and fungal enzymes involved in lignocellulose degradation. *Engineering in Life Sciences, 18*(11), 768-778. doi:10.1002/elsc.201800039

Arantes, V., & Saddler, J. N. (2010). Access to cellulose limits the efficiency of enzymatic hydrolysis: the role of amorphogenesis. *Biotechnology for Biofuels, 3*(1), 4. doi:10.1186/1754-6834-3-4

Bauer, R. J., Jurkiw, T. J., Evans, T. C., & Lohman, G. J. S. (2017). Rapid Time Scale Analysis of T4 DNA Ligase–DNA Binding. *Biochemistry, 56*(8), 1117-1129. doi:10.1021/acs.biochem.6b01261

Bayer, E. Cellulosome Systems. *https://web.archive.org/web/20091217184342/http://www.weizmann.ac.il/Biological_Chemistry/scientist/Bayer/new_pages/reserch_topics/cellulosome_systems.html*. Retrieved from https://web.archive.org/web/20091217184342/http://www.weizmann.ac.il/Biological_Chemistry/scientist/Bayer/new_pages/reserch_topics/cellulosome_systems.html

Benocci, T. ((2018)). *Dissecting semi-specialist and specialist plant biomass utilization strategies using the fungi Trichoderma reesei and Podospora anserina*

Utrecht University Repository,

Bioenergy (Biofuels and Biomass). Retrieved from https://www.eesi.org/topics/bioenergy-biofuels-biomass/description

Biomass Resources. Retrieved from https://www.energy.gov/eere/bioenergy/biomass-resources

Chambers, C. R., & Patrick, W. M. (2015). Archaeal nucleic acid ligases and their potential in biotechnology. *Archaea, 2015*.

Chandel, A. K., Chandrasekhar, G., Silva, M. B., & Silvério da Silva, S. (2012). The realm of cellulases in biorefinery development. *Critical Reviews in Biotechnology, 32*(3), 187-202. doi:10.3109/07388551.2011.595385

Cheng, Y.-S., Ko, T.-P., Huang, J.-W., Wu, T.-H., Lin, C.-Y., Luo, W., . . . Wang, A. H.-J. (2012). Enhanced activity of Thermotoga maritima cellulase 12A by mutating a unique surface loop. *Applied microbiology and biotechnology, 95*(3), 661-669.

Composition of Enzymes. Retrieved from https://www.e-education.psu.edu/egee439/node/670

Davies, G., & Henrissat, B. (1995). Structures and mechanisms of glycosyl hydrolases. *Structure, 3*(9), 853-859. doi:10.1016/s0969-2126(01)00220-9

Djajadi, D. T., Jensen, M. M., Oliveira, M., Jensen, A., Thygesen, L. G., Pinelo, M., . . . Meyer, A. S. (2018). Lignin from hydrothermally pretreated grass biomass retards enzymatic cellulose degradation by acting as a physical barrier rather than by inducing nonproductive adsorption of enzymes. *Biotechnology for Biofuels, 11*(1), 85. doi:10.1186/s13068-018-1085-0

Doherty, A. J., & Dafforn, T. R. (2000). Nick recognition by DNA ligases. *J Mol Biol, 296*(1), 43-56. doi:10.1006/jmbi.1999.3423

Edwards, A. N., Suárez, J. M., & McBride, S. M. (2013). Culturing and maintaining Clostridium difficile in an anaerobic environment. *JoVE (Journal of Visualized Experiments)*(79), e50787.

Ellabban, O., Abu-Rub, H., & Blaabjerg, F. (2014). Renewable energy resources: Current status, future prospects and their enabling technology. *Renewable and Sustainable Energy Reviews, 39*, 748-764. doi:https://doi.org/10.1016/j.rser.2014.07.113

Felix, C. R., & Ljungdahl, L. G. (1993). THE CELLULOSOME: The Exocellular Organelle of Clostridium. *Annual Review of Microbiology, 47*(1), 791-819. doi:10.1146/annurev.mi.47.100193.004043

Fontes, C. M. G. A., & Gilbert, H. J. (2010). Cellulosomes: Highly Efficient Nanomachines Designed to Deconstruct Plant Cell Wall Complex Carbohydrates. *Annual Review of Biochemistry, 79*(1), 655-681. doi:10.1146/annurev-biochem-091208-085603

Gomez del Pulgar, E. M., & Saadeddin, A. (2014). The cellulolytic system of Thermobifida fusca. *Critical Reviews in Microbiology, 40*(3), 236-247. doi:10.3109/1040841X.2013.776512

Henrissat, B. (1991). A classification of glycosyl hydrolases based on amino acid sequence similarities. *Biochem J, 280 (Pt 2)*, 309-316. doi:10.1042/bj2800309

Henrissat, B., & Bairoch, A. (1993). New families in the classification of glycosyl hydrolases based on amino acid sequence similarities. *Biochem J, 293 (Pt 3)*, 781-788. doi:10.1042/bj2930781

Henrissat, B., & Bairoch, A. (1996). Updating the sequence-based classification of glycosyl hydrolases. *Biochem J, 316 (Pt 2)*, 695-696. doi:10.1042/bj3160695

Henrissat, B., & Davies, G. (1997). Structural and sequence-based classification of glycoside hydrolases. *Current opinion in structural biology, 7*(5), 637-644.

Hirano, K., Kurosaki, M., Nihei, S., Hasegawa, H., Shinoda, S., Haruki, M., & Hirano, N. (2016). Enzymatic diversity of the Clostridium thermocellum cellulosome is crucial for the degradation of crystalline cellulose and plant biomass. *Scientific Reports, 6*, 35709.

Kataoka, M., & Ishikawa, K. (2014). A new crystal form of a hyperthermophilic endocellulase. *Acta Crystallographica Section F: Structural Biology Communications, 70*(7), 878-883.

Katayeva, I., Golovchenko, N., Chuvilskaya, N., & Akimenko, V. (1992). Clostridium thermocellum β-glucosidases A and B: purification, properties, localization, and regulation of biosynthesis. *Enzyme and microbial technology, 14*(5), 407-412.

Kim, H. W., & Ishikawa, K. (2011). Functional analysis of hyperthermophilic endocellulase from Pyrococcus horikoshii by crystallographic snapshots. *Biochem J, 437*(2), 223-230. doi:10.1042/bj20110292

Kim, R., Sandler, S. J., Goldman, S., Yokota, H., Clark, A. J., & Kim, S.-H. (1998). Overexpression of archaeal proteins in Escherichia coli. *Biotechnology Letters, 20*(3), 207-210. doi:10.1023/A:1005305330517

Kolpak, F. J., & Blackwell, J. (1976). Determination of the Structure of Cellulose II. *Macromolecules, 9*(2), 273-278. doi:10.1021/ma60050a019

Lamed, R., Naimark, J., Morgenstern, E., & Bayer, E. A. (1987). Specialized cell surface structures in cellulolytic bacteria. *J Bacteriol, 169*(8), 3792-3800. doi:10.1128/jb.169.8.3792-3800.1987

Lasa, I., & Berenguer, J. (1993). Thermophilic enzymes and their biotechnological potential. *Microbiologia, 9*(2), 77-89.

Lehman, N. E. S. (2020). biofuel | Definition, Types, & Pros and Cons. Retrieved from https://www.britannica.com/technology/biofuel

Leis, B., Held, C., Andreeßen, B., Liebl, W., Graubner, S., Schulte, L.-P., . . . Zverlov, V. V. (2018). Optimizing the composition of a synthetic cellulosome complex for the hydrolysis of softwood pulp: identification of the enzymatic core functions and biochemical complex characterization. *Biotechnology for Biofuels, 11*(1), 220.

Liberato, M. V., Silveira, R. L., Prates, É. T., De Araujo, E. A., Pellegrini, V. O., Camilo, C. M., . . . Skaf, M. S. (2016). Molecular characterization of a family 5 glycoside hydrolase suggests an induced-fit enzymatic mechanism. *Scientific Reports, 6*, 23473.

ligation calculator - insilico.

Liu, Y.-S., Zeng, Y., Luo, Y., Xu, Q., Himmel, M. E., Smith, S. J., & Ding, S.-Y. (2009). Does the cellulose-binding module move on the cellulose surface? *Cellulose, 16*(4), 587-597. doi:10.1007/s10570-009-9306-0

Ljungdahl, L. G., Coughlan, M. P., Mayer, F., Mori, Y., Hon-nami, H., & Hon-nami, K. (1988). Macrocellulase complexes and yellow affinity substance from Clostridium thermocellum. In *Methods in enzymology* (Vol. 160, pp. 483-500): Elsevier.

Lombard, V., Golaconda Ramulu, H., Drula, E., Coutinho, P. M., & Henrissat, B. (2014). The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Res, 42*(Database issue), D490-495. doi:10.1093/nar/gkt1178

Madeira, F., Park, Y. M., Lee, J., Buso, N., Gur, T., Madhusoodanan, N., . . . Finn, R. D. (2019). The EMBL-EBI search and sequence analysis tools APIs in 2019. *Nucleic acids research, 47*(W1), W636-W641.

Mäkelä, M. R., Donofrio, N., & de Vries, R. P. (2014). Plant biomass degradation by fungi. *Fungal Genetics and Biology, 72*, 2-9. doi:https://doi.org/10.1016/j.fgb.2014.08.010

Markoglou, N., & Wainer, I. W. (2003). Chapter 7 - Immobilized enzyme reactors in liquid chromatography: On-line bioreactors for use in synthesis and drug discovery. In I. D. Wilson (Ed.), *Handbook of Analytical Separations* (Vol. 4, pp. 215-234): Elsevier Science B.V.

Nishida, H., Kiyonari, S., Ishino, Y., & Morikawa, K. (2006). The closed structure of an archaeal DNA ligase from Pyrococcus furiosus. *J Mol Biol, 360*(5), 956-967. doi:10.1016/j.jmb.2006.05.062

Pascal, J. M., O'Brien, P. J., Tomkinson, A. E., & Ellenberger, T. (2004). Human DNA ligase I completely encircles and partially unwinds nicked DNA. *Nature, 432*(7016), 473-478. doi:10.1038/nature03082

Pascal, J. M., Tsodikov, O. V., Hura, G. L., Song, W., Cotner, E. A., Classen, S., . . . Ellenberger, T. (2006). A flexible interface between DNA ligase and PCNA supports conformational switching and efficient ligation of DNA. *Mol Cell, 24*(2), 279-291. doi:10.1016/j.molcel.2006.08.015

Perez-Iratxeta, C., & Andrade-Navarro, M. A. (2008). K2D2: Estimation of protein secondary structure from circular dichroism spectra. *BMC Structural Biology, 8*(1), 25. doi:10.1186/1472-6807-8-25

Petrova, T., Bezsudnova, E. Y., Boyko, K. M., Mardanov, A. V., Polyakov, K. M., Volkov, V. V., . . . Popov, V. O. (2012). ATP-dependent DNA ligase from Thermococcus sp. 1519 displays a new arrangement of the OB-fold domain. *Acta crystallographica. Section F, Structural biology and crystallization communications, 68*(Pt 12), 1440-1447. doi:10.1107/S1744309112043394

Rogers, K. (April 28, 2017). Gel electrophoresis. In *Encyclopædia Britannica*

Encyclopædia Britannica, inc.

Sandgren, M., Gualfetti, P. J., Shaw, A., Gross, L. S., Saldajeno, M., Day, A. G., . . . Mitchinson, C. (2003). Comparison of family 12 glycoside hydrolases and recruited substitutions important for thermal stability. *Protein Science, 12*(4), 848-860.

Subramanya, H. S., Doherty, A. J., Ashford, S. R., & Wigley, D. B. (1996). Crystal structure of an ATP-dependent DNA ligase from bacteriophage T7. *Cell, 85*(4), 607-615. doi:10.1016/s0092-8674(00)81260-x

Summer, H., Grämer, R., & Dröge, P. (2009). Denaturing urea polyacrylamide gel electrophoresis (Urea PAGE). *JoVE (Journal of Visualized Experiments)*(32), e1485.

Tanabe, M., Ishino, S., Ishino, Y., & Nishida, H. (2014). Mutations of Asp540 and the domain-connecting residues synergistically enhance Pyrococcus furiosus DNA ligase activity. *FEBS Lett, 588*(2), 230-235. doi:10.1016/j.febslet.2013.10.037

Unsworth, L. D., van der Oost, J., & Koutsopoulos, S. (2007). Hyperthermophilic enzymes--stability, activity and implementation strategies for high temperature applications. *Febs j, 274*(16), 4044-4056. doi:10.1111/j.1742-4658.2007.05954.x

van den Brink, J., van Muiswinkel, G. C. J., Theelen, B., Hinz, S. W. A., & de Vries, R. P. (2013). Efficient Plant Biomass Degradation by Thermophilic Fungus <span class="named-

content genus-species" id="named-content-1">Myceliophthora heterothallica. *Applied and Environmental Microbiology, 79*(4), 1316-1324. doi:10.1128/aem.02865-12

Vuong, T., & Wilson, D. (2010). Engineering Thermobifida Fusca cellulases: Catalytic mechanisms and improved activity. In.

Wagner, A. O., Markt, R., Mutschlechner, M., Lackner, N., Prem, E. M., Praeg, N., & Illmer, P. (2019). Medium Preparation for the Cultivation of Microorganisms under Strictly Anaerobic/Anoxic Conditions. *JoVE (Journal of Visualized Experiments)*(150), e60155.

Whisstock, J. C., & Lesk, A. M. (2003). Prediction of protein function from protein sequence and structure. *Quarterly reviews of biophysics, 36*(3), 307-340. doi:10.1017/s0033583503003901