THESE DE DOCTORAT DE

NANTES UNIVERSITE

ECOLE DOCTORALE N° 605
Biologie Santé
Spécialité : « *Biologie Moléculaire et Structurale, Biochimie* »

Par
« Lydia OGONDA »

« Clonage et ingénierie de nouvelles cellulases isolées à partir de souches bactériennes du lac Bogoria (Kenya) et applications biotechnologiques»

« Cloning and engineering of new cellulases isolated from Lake Bogoria strains and biotechnological applications»

Thèse présentée et soutenue à Nantes, le 23 novembre 2022

Rapporteurs avant soutenance :

Claire DUMON Directeur de Recherche INRAE, INSA Toulouse
Magnus WOLF-WATZ Professeur, Université de UMEA (Suède)

Composition du Jury :

Président :	Cyrille GRANDJEAN	Directeur de Recherche CNRS, Nantes Université
Examinateurs :	Cyrille GRANDJEAN	Directeur de Recherche CNRS, Nantes Université
	Sylvain TRANCHIMAND	Maître de Conférences, ENSC Rennes
Dir. de thèse :	Charles TELLIER	Professeur émérite, Nantes Université
Co-dir. de thèse :	Francis MULAA	Professeur, Université de Nairobi (Kenya)
Invité :	Edward K.MUGE	Senior Lecturer, Université de Nairobi (Kenya)

Doctoral Thesis

12/1/2023

Department of Biologie
Unité en Sciences Biologiques et Biotechnoogies (US2B)
CNRS UMR 6286.
Université de Nantes
F-44000 Nantes.
FRANCE.

Department of Biochemistry
School of Medicine,
College of Health Science
P.O BOX 30197-00100
Nairobi.
KENYA.

Printed by University of Nairobi Library Services
Printed by L'Universite De Nantes PRESS
12/1/2023

DECLARATION

I declare that this thesis is my original work. Any work cited herein have been clearly referenced

Ogonda Lydia Awuor

E14G386C H809/97387/2015

_____ _____

This research thesis has been submitted with my approval as the University supervisor:

1. Prof. Charles Tellier
   Ecole Doctoral, Biologie Sante
   Unité en Sciences Biologiques et Biotechnoogies (US2B)
   Universite de Nantes
   France

   _____ _____

2. Prof. Francis Jackim Mulaa
   Faculty of Science and Technology, (FST)
   Department of Biochemistry
   University of Nairobi
   Kenya

   _____ _____

3. Dr. Edward Kirwa Muge
   Faculty of Science and Technology, (FST)
   Department of Biochemistry
   University of Nairobi
   Kenya

   _____ _____

4. Dr. Benson Wamalwa
   Faculty of Science and Technology, (FST)
   Department of Chemistry
   University of Nairobi
   Kenya

   _____ _____

ACKNOWLEDGEMENTS

I have learnt a lot through this academic journey about A LOT of things on life, work, people and Research. Lessons that I will carry as treasures with me.

'Ubuntu' meaning a person is only a person through other persons.

I would like to express my sincerest gratitude to **God**, for it is by His grace, strength and wisdom that I was able to accomplish this work. Indeed, 'it is not by power, not by might but by His spirit'. Also for allowing me meet and interact with all the lovely people, I have met along this academic journey.

My Thesis Supervisors:

Prof. Charles Tellier, for accepting me as a PhD student at UFIP, and guidance throughout this thesis period. In addition, most especially for your assistance when I lost my dad during my research stay in France, 2017. For this, I remain forever grateful.

Prof. Francis J. Mulaa, for allowing me use your research idea and, your patience with me throughout this process. Moreover, especially for integrating me as an active member of your research team. Under your wings, I have grown immensely. It is surreal working with you and seeing how much I have grown under your mentorship.

Dr.Benson Munyali Wamalwa, for allowing me use your research idea, and sitting with me during the conceptual stages of this work. Sitting down with you for those, brainstorming /designing sessions helped me work on my logical thinking, being critical with my ideas, and myself. My scientific curiosity has wildly blossomed under your wings.

Dr.Edward Kirwa Muge, for working with me throughout the years and giving me access and room to grow under your mentorship. For your encouragement, support and patience. For teaching me the value of going beyond the bare minimum. Thank you.

My Thesis Committee Members:

I am grateful to Oliver Gonçalves and Vincent Ferrieres for being a part of my thesis committee.

UFIP Research Team:

I am grateful for past and present UFIP members that created a warm, friendly metropolitan environment. It was a real great pleasure working with you all. In addition, for your patience especially with my 'French'. Especially, I am grateful to Prof. Michel Dion for allowing me use your samples and lab. You and Annie were my first working contact in UFIP. Dr. Annie Lambert, Dr. Johann Dion, Dr. Johann Hendrickx, Dr. Christophe Dussouy, Dr. Alain Defontaine, Dr. Typhaine Violo, Dr. Maruthi Prasanna, Dr. Benoît David, Dr. Surbhi Dhingra, Dr. Iyanar Vetrivel, Dr. Estella, Dr. Ennys, Gheyouche Dr. Mahesh Velusamy, and Dr. Emilie Camberlein, I am grateful for your positive energy, kindness, and enthusiasm for science and life.

I am equally grateful to Dr. Amélie Saumonneau of D-zyme CAPACITES, for all the technical support you gave me through this work.

Finally yet importantly, I am thankful to Fabienne for all the administrative support and Carine for all your assistance with laboratory reagents and solutions. Research was definitely easier thanks to your diligence and warmth.

University of Nairobi Research Team:

In the same breathe; I remain grateful to the University of Nairobi, Department of Biochemistry staff-technicians and academic faculty for creating a warm, friendly and lovely environment for exchanging ideas and research. I am indebted to the Department of Biochemistry; for your unequivocal support to

me throughout this journey, and MOST especially, for your patience and support when I lost my dad during this journey. **THANK YOU**

Family and Friends:

My **sincerest and greatest gratitude** goes out to my late dad and my mum, for all your support throughout this process. I am thankful to my dad for encouraging me to pursue this opportunity and continue this work even amidst the challenging circumstances that were occurring then. I could not have made it this far without you both. **THANK YOU FOR BEING YOU.** I love you both so much.

I cannot close this without special mention of my spiritual support, my spiritual mum and dad; Rev Mary Achero and Rev Dr. Dodzweit Achero, Pst. Dan Achero and the entire Sound of transformation team, of the, City of Transformation-Headquarters, Syokimau, Kenya and my faith family at Serve International Ministries_Kakamega with special mention of Bishop Paul Ligono. Thank you for the love, prayers and support throughout this journey.

In addition, Pastor, Caroline Schrumpf of Eglise Protestante Unis De France, Nantes, for walking with me during the moments when I felt lost. Thank you for your love, encouragement, patience and faith in me Pastor Caroline.

I am equally grateful to Prof. Raj of KwaZulu-Natal University for sharing your doctoral journey story with me. Sitting with you over that lunch was inspiring because in your story, I saw parts of my own and your wins invigorated me with hope. I am glad I met you and Kenny during that Seminar in February 2020.

Finally yet importantly, I am thankful to Prof. Kerubo for encouraging me to sucessfully complete this journey.

DEDICATION

To God,

'It is the mind of God to hide a matter and the glory of Kings to reveal it.' Proverbs 25:2 and 'The secret things belong to God' Deuteronomy 29:29.

To my family,

I most especially lovingly dedicate this work to my Lovely Mum and late Dad. Dad, I wish you were here for the completion of this journey. HOWEVER, I am forever grateful for the part you played to see me here. FINALLY, your 6 months' premature 'daktari' is here. A manifestation of what you saw in the 1.5kg infant girl☺ .

In addition I dedicate this work to my siblings (Isaac, Faith, Dancun, Aminah, Pauline and Carey), may this work show and build faith in you for greater things. Lastly to my lovely nephews (Cian Kyeon, and Carl), May you know no bounds. As well as to my beautiful and loyal animal family☺.

Finally, I lovingly dedicate this work to my future nuclear family. May you never struggle with any giants I slayed in my journey with this work. Moreover, may this accomplishment be just but where you start. The sky is the limit!

FINANCIAL SUPPORT

I am grateful for all the financial support I have received from the following:

- The Ambassador of France to Kenya, for the award of the PhD scholarship as well as the French language training scholarship
- I am equally grateful for the University of Nairobi through the grants office for the award of the study grant 2015-2019.
- I am also indebted to the University of Nairobi, for the ticket support for the year 2017
- I am also indebted to the University of Nairobi, College of health science for the sponsorship to the Nairobi Innovation week, 2019
- In addition, I am grateful to the Universite De Nantes Ecole Doctorale, Biologie Sante (ED, BS) for the conference travel grant to make my oral poster presentation in Le Croisic ,2017
- University of Nantes within the interdisciplinary project framework program.
- Further, I am grateful to the OWSD for the Workshop sponsorship for the presentation of this work , 2017
- This is in addition to the conference grant from SFBBM, 2017
- I am equally indebted to the WFS for the study Scholarship, 2015-2016
- And last but not least, I am grateful for the short stay research grant from UFIP ,Universite De Nantes , for Apr-July 2017 & May -July 2018

TABLE OF CONTENTS

Table of Contents

LIST OF TABLES

LIST OF FIGURES

LIST OF PAPERS

This thesis is based, in addition to the following papers. Numerals indicate all mentions within the text.

I. Ogonda. L.A, Michel. D, Muge .E.K, Wamalwa .B, Mulaa. F.J, Tellier.C. Characterisation of crude cellulases from a Bacillus sp. from Lake Bogoria, Kenya (**Manuscript preprint published on RESEARCHSQUARE)** https:/ doi.org/10.21203/rs.3.rs-41635/v1 **(Chapter 4)**

II. Ogonda. L.A, Saumonneau. A, Michel. D, Muge .E.K, Wamalwa .B, Mulaa.F.J, Tellier.C. Characterization and engineering of two new GH9 and GH48 cellulases from a *Bacillus pumilus* isolated from Lake Bogoria, Biotechnol. Lett. (2021) 43:691-700 **(**https://doi.org/10.1007/s10529-020-03056-z) **(Chapter 5)**

III. Amélie Saumonneau, Nathan Lagneau, Lydia Awuor Ogonda, Catherine Dupré, Stéphanie Dutertre, Dominique Grizeau, Charles Tellier, Cyrille Grandjean, Franck Daligault. Disorganization of *Botryococcus braunii* colonies by glycoside hydrolases **(Manuscript accepted; Chapter 6)**

IV. Ogonda. L.A and Muge .E.K. harnessing the potential of underutilized aquatic bio resource for food and nutritional security in Kenya. Springer Nature publishers (Chapter 5 in the book; Food Security and Safety: African Perspectives 1st ed. 2021, Edition. ISBN-10 3030506711, ISBN-13 978-3030506711) **(Book Chapter accepted and Book Published; Appendices)**

DOI: 10.1007/978-3-030-50672-8_5

LIST OF ABBREVIATIONS

16srDNA-16s Ribosomal Deoxyribonucleic acid

3D-3Dimension

AA-auxillary activity

ALA-Aminolevulenic acid

ANOVA- Analysis of variance

AX-Arabinoxylans

BG-beta glucosidase

BOD-Biological oxygen demand

BSA-Bovine serum albumin

CAGR-Compound annual growth rate

CBD-Cellulose binding domain

CBH- Oligoxyloglucan reducing end-specific cellobiohydrolase 3.2.1.150

CBH-I Non –reducing end cellulose 1, 4 –β-cellobiosidase 3.2.1.91

CBH-II Reducing end-acting cellobiohydrolase EC 3.2.1.176

CBM-Carbohydrate binding module

CBO-Cellobiose oxidase

CBQ-Cellobiose quinone oxidoreductase

CCS-Croscarmellose sodium

CDB-Cellulose degrading bacteria

CD-catalytic domain

CDH-Cellobiose dehydrogenase

CEL- Endo-β-1, 4-glucanase 3.2.1.4

Cel-Cellulase

CMC-Carboxymethylcellulose

CMC-Carboxymethylcellulose

COD-Chemical oxygen demand

Coh-Cohesin

DH-Degree of hydrolysis

DNS-Dinitrosalicylic acid

DNS-Dinitrosalicylic acid

DP-Degree of polymerization

DS-Degree of substitution

DTT-Dithiothreitol

Dwb-Dry weight basis

ED, BS-Ecole Doctorale, Biologie Sante

EDTA- Ethylene diamine tetraacetic acid

EG- Endoglucanases

EPS-Exopolysaccharide

FAD-Flavin adenine inucleotide

FDCA-Furandicarboxylic acid

FSC-Forward scatter

FST- Faculty of Science and Technology

GC-MS/MS –Gas chromatography tandem mass spectrometry

gDNA- Genomic deoxyribonucleic acid

GH-Glycoside hydrolase

GIT-gastrointestinal tract

HCl-Hydrochloric acid

HMF-Hydroxymethylfurfural

IMAC –Immobilised metal affinity chromatography

LB-Luria bertani

LB-Luria Bertani

LCCs-Lignin carbohydrate complexes

LPMOs -Lytic polysaccharide monooxygenases

MF-mumefural

ML-Middle lamellae

MSM-Minimal salt media

MSM-Minimal salt media

Mw-Molecular weight

Na-CMC- Sodium carboxymethylcellulose

NMR-Nuclear magnetic resonance

OD- Optical Density

OD-Optical Density

OWSD-Organization for women scientists from developing world

PAGE-Polyacrylamide gel electrophoresis

PASC-Phosphoric acid swollen cellulose

PCR-Polymerase chain reaction

PET-Polyethylene terephthalate

PHA-polyhydroxyalkanoate

PLA-Polylactic acid

PMSF –Phenylmethylsulfonylflouride

PS-polysaccharides

R.E- Restriction site

RPM-Revolutions per minute

Sca-Scaffolding

SDG-sustainable development goals (SDG).

SDS - Sodium dodecyl sulphate

SDS-Sodium dodecyl sulfate

SEC-Size exclusion chromatography

SEM- Scanning electron micrograph

SEM-Scanning electron microscopy

SFBBM-

SIDP-sustainable industrial development policy

SLH- Surface layer homology

SPS-Sulfated polysaccharides

SSC-side scatter

TLC-Thin layer chromatography

TLC-Thin layer Chromatography

UFIP-Unite Fonctionalité et d'ingéniere des Proteines

UV-Ultraviolet

WFS-World Federation of scientists

WRF-white rot fungi

CHAPTER 1
INTRODUCTION

With the advent in "Green Chemistry" and the drive for sustainable industrial processes, biomass/ bio resource utilization has been on the rise. Efficient utilization of these natural feedstocks necessitates sustainability, extracting value prior to processing and value addition by creating new product streams from residual and spent biomass liquors. Several carbohydrate bio resources meet this demand. Subsequently, "Carbohydrate economy" (David Morris *et al.,* 1980) is on the rise.

One such biomass is cellulose. It is the most abundant organic bio resource (Bayer 2017, Jahangeer *et al*., 2005; Walsh 2002,). Moreover, forms an important component of the primary cell wall in plants, algae and oomycetes. In its simplest form, cellulose is a linear unbranched homopolymer of glucose residues (7000-15,000) linked in β-1, 4 glycosidic bonds. Rearrangement of cellulose results in microcrystalline, para crystalline and amorphous cellulosic structures. Cellulose is a heavily loaded glucose stockpile with numerous potential industrial applications; biofuels, paper, textile, detergent, pulp, biomaterials, food, biochemical, pharmaceuticals, bioplastics as well as biomedical applications (Jahangeer *et al*., 2005). However, in order to be industrially useful, cellulose-containing biomass has to be broken down to its monomeric hexose or pentose sugars (Akinosho *et al*., 2014) under conditions that are industrially viable and economically sustainable.

All these notwithstanding, cellulose remains largely underutilised mainly because of its rigid structure that lends it very high in-built stability and insolubility (Lederberg, 1992) making it very resistant to hydrolysis (Klemm, Heublein, Fink & Bohn, 2005). As a result, present cellulose hydrolysis methods involve the use of 'harsh' chemical methods; such as highly concentrated, strong acids like sulphuric acid (H_2SO_4) and/ highly alkali solutions such as liquid anhydrous ammonia or NaOH. These chemical processing methods are associated with a heavy environmental burden due to the high biological oxygen demand (BOD) and chemical oxygen demand (COD) in industrial effluents. In addition, chemical

processing is associated with high production costs related to maintenance of operation conditions (chemical, water and energy demands) as well as treatment of effluents.

With the advancement in GREEN CHEMISTRY, sustainable industrial development policy (SIDP), Kyoto protocol and various local (per country) climate change governance policies , the search for alternative EFFICIENT and CLEAN cellulosic biomass processing methods is on an all-time high. We have therefore identified c*ritical points to alleviate this burden on industrial cellulose processing. One such step involves substituting cellulose enzyme processing for chemical processing.*

However, evolution driven enzyme traits, greatly limit the efficacy of native enzymes outside their reaction conditions. This consequently hinders the industrial use of native enzymes for bio catalysis (Steiner and Schwab, 2012). Cellulose enzymatic hydrolysis/saccharification therefore, becomes the rate-limiting step and the main bottleneck preventing cost efficiency in cellulose utilisation.

Nevertheless, with synthetic biology, molecular biology and biotechnology, enzyme engineering is possible. This has led to the production of 'designer biocatalysts' which are tailored for the reaction as well as the reaction conditions which sometimes involve organic solvents, extreme pH ,high substrate and or product concentrations and high temperatures. In addition using these synthetic biology tools it is possible to introduce new enzymatic functionalities as well as fit enzymes to the requirements of industrial biotechnology.

In nature, microorganisms: bacteria and fungi have evolved mechanisms to combat the cellulose breakdown problem. Aerobic microorganisms for instance, secrete to the milieu a cocktail of hydrolytic enzymes (cellulases and ancillary enzymes that act synergistically to breakdown cellulose. They then take up the smaller polysaccharides (cellodextrins) or the complete hydrolysis product (glucose). In

addition to these hydrolytic enzymes, a recently discovered class of lytic polysaccharide monooxygenases (LPMOs) enzymes carry out oxidative cleavage of cellulose and non-cellulose polysaccharides. LPMOs support and improve the efficacy of the archetypal hydrolytic enzymes by acting on the exterior of the crystalline substrates. In addition, they introduce breaks that opens the substrate to hydrolysis. As a result, LPMOs alleviate the need for pre-treatment to extract the cellulose chains from the rigid crystalline structure (Horn *et al.,* 2012).

On the other hand, the more efficient cellulose utilising/degrading anaerobic organism *Clostridium thermocellum* possesses an inherent multiple enzyme complex consisting of different catalytic units known as a cellulosome (Bayer *et al.*, 2004; Bayer *et al.*, 2008; Demain, Newcomb & Wu, 2005; Doi & Kosugi, 2004; Fontes & Gilbert, 2010). In both mechanisms, the complete degradation of cellulose is a concerted effort between three different cellulase enzyme types known as β-1, 4 endoglucanase, cellobiohydrolase, and glucosidase as well as ancillary enzymes.

To date, a lot of research has been conducted on cellulases. Therefore, using *Clostridium thermocellum* as a blueprint, this study proposed to use synthetic biology tools to develop a designer cellulosome with improved saccharide production. This we did in a three stepwise enhancement process. First, we screened for novel, native processive cellulases from an extremophile (halophilic, alkaliphilic and thermophilic) environment. Secondly, using rational approaches, we bioengineered the cellulases by fusion to exogenous scaffoldin (cohesin, dockerin and CBD/CBM) modules thereby creating new enzyme architectures (and introducing new enzyme functionalities and properties) as well as complementary affinity pairs. Further, using these complementary affinity pairs, we produced an enzyme tailored to microcrystalline cellulosic biomass by clustering cellulases in precise arrangements and with specific stoichiometric ratios thus forming homogenous ‘Designer IN LINE cellulosomes’ prototypes, proposed to have improved enzyme performance.

CHAPTER 2

RESEARCH AIMS AND OBJECTIVES

Overall objective

Several efforts have been invested to reduce the cost of cellulose processing in industry by substituting chemical processing with biological treatments (cellulase enzymes). At the forefront of these efforts are initiatives to engineer individual soluble cellulase preparations and cellulase mixes/cocktails. Currently, these preparations still require synergy between a number of cellulase enzymes, high enzyme loads and extensive reaction times. In spite of the significant improvements obtained so far, cellulase engineering is still required to ensure cost reduction and efficacy, complete degradation and value added cellulolytic products. Thermostable, halophilic and alkaliphilic cellulases have several advantages for bioconversion processes. These include; increased specific activity, higher level of stability, inhibition of microbial growth in the reaction tank, increased mass transfer rate due to lower viscosity rate and greater rate of flexibility in the bioprocess (Anbar and Bayer, 2010).

This study envisioned to improve cellulase efficacy in a three stepwise enhancement process. First by screening for novel, native, free, processive cellulases from an extremophile (alkaliphilic, halophilic and thermophilic) environment in order to enhance the specific activity and stability of the cellulase bioprocess. Secondly, using rational approaches, we bioengineered the cellulases by fusion to exogenous scaffoldin modules thereby creating new enzyme architectures (and introducing new enzyme functionalities) and complementary affinity pairs. This is useful, novel and innovative because it creates a superior enzyme geometry by bringing together the various aspects of both aerobic as well as anaerobic cellulase systems. By so doing, this tool fuses the advantages of both into one catalytic tool that can be utilised for efficient cellulose degradation with minimal damage to the cellulose quality and diminished waste generation.

Moreover, using the complementary enzyme pairs produced, synthetic, prototype, Quad-functional IN-LINE cellulosomes, were also designed by clustering the selected stable cellulases (EG, CBH and BG) in precise stoichiometric ratios and specific positioning.

This synthetic cellulosome would be useful for improved cellulose processing, improved saccharide production and in general efficient microcrystalline cellulose utilization. Further, we envision the successful conversion of an aerobic non-cellulosomal, enzyme system into a synthetic cellulosome. This would provide a platform for further research studies into the more complex cellulosomes systems with more than one scaffolding, as well as provide a foundation for the anchoring of cellulases along with complementary enzymes for the efficient utilization of cellulose feedstock for chemical commodity and biofuel production.

Looking at the **BIGGER picture**, this research will provide prototypes that can be used as a blueprint to tailor make enzyme systems/tools for different lignocellulosic/other biomass feedstock for varied industries and or for varied production purposes. The value addition offered by these new geometries include faster reaction times, hence saving on production costs for value time. The high yields resulting from the processing of the lignocellulosic material in addition justify the inclusion of lignocellulose biomass (Both wastes and fresh biomaterial with non-competing interests). This would foster the development of local biotechnology companies, enable the recycling and or value addition to lignocellulosic waste, generating revenue from the new products as well as reducing environmental waste and enabling efficient lignocellulosic waste management.

The clean processing achieved by this catalytic tool also helps to avoid the need for chemicals in the processing chain and thus eliminates the waste effluent treatment costs. The carbohydrate industry via

this lignocellulose stock is used for varied industries and processes. Thus, the carbohydrate economy driven by cellulosic biomass will lend to all of the sustainable development goals (SDG).

Specific objectives:

1. To screen for extremophile (alkaliphilic and thermophilic) cellulases Cellobiohydrolase/exocellulase and β-1,4-endoglucanase) producers from Kenyan ecosystem,
2. To order and rank the cellulolytic potential of the microbial isolate library
3. To conduct genetic characterization of the identified isolate (16srRNA).
4. To conduct biochemical characterization of the native crude enzyme isolates
5. To conduct preliminary data mining and DNA sequence analysis of the selected isolate and to extract using molecular biology tools, the genes coding for the identified cellulases.
6. To conduct rational design of the **'Designer'** cellulosome
7. To extract using molecular biology tools the genes coding for scaffoldin modules from *C. thermocellum* and *C. josui*
8. Construction, expression and characterization of the novel enzyme architectures by fusion of catalytic modules to exogenous scaffoldin modules
9. To conduct molecular and biochemical characterization of the GH48 and GH9 enzymes (Primary , secondary, tertiary, and quaternary structures and Mw, sequence analysis, Optimum pH/Optimum temperature , SEC analysis, Activity on substrates, Mechanism of action, synergy respectively)
10. The fusion of complementary enzyme pairs into cellulosome complex
11. Testing the functionality and integrity of the novel enzyme architectures against the native enzyme architectures
12. Testing the functionality of the "designer cellulosomes"

13. Comparing the activity of the "designer cellulosomes" against the individual enzymes (Cellobiohydrolase/ β-1,4-endoglucanase) and enzyme mixes (Cellobiohydrolase and β-1,4-endoglucanase) on microcrystalline cellulose substrates

CHAPTER 3

LITERATURE REVIEW

3.1 CELLULOSE

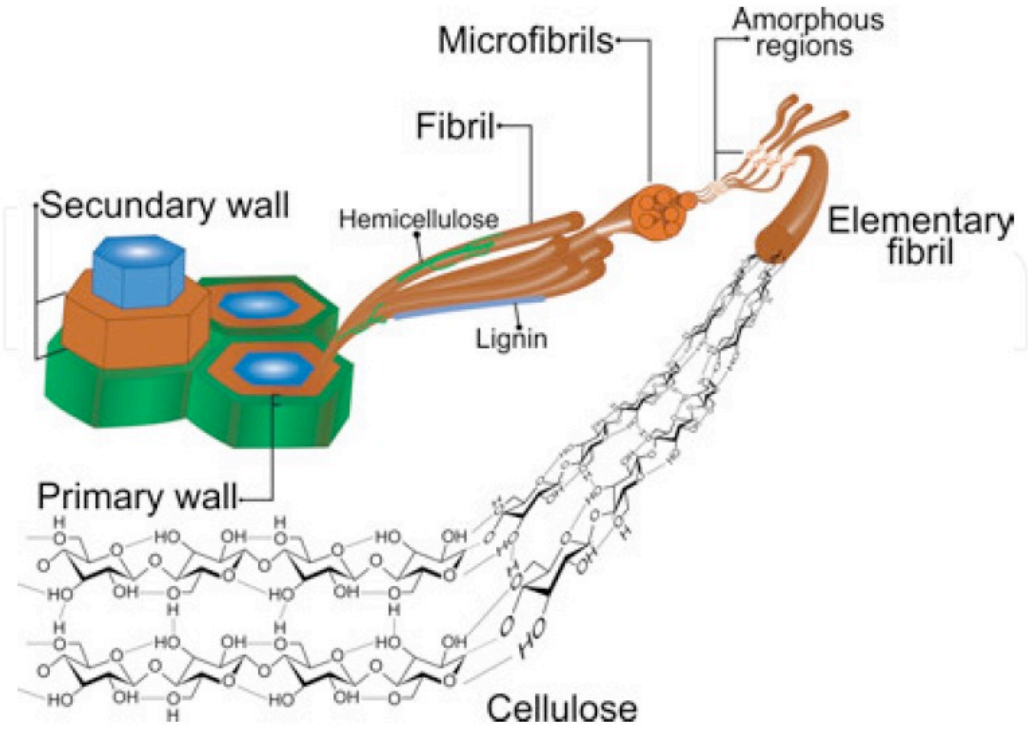


Figure 3.1: Structure of cellulose extracted from plants (Adapted from Rojas *et al.*, 2015)

Cellulose (Fig. 3.1/ Fig. 3.2) is the most abundant organic material in the globe. It is the main structural component of marine algae, oomycetes, bacteria, higher plant cell wall (Fig. 3.1) and some animals (tunicates) (Chen Hongzhang, 2014). In addition to being part of plant cell wall, a number of plant material is largely composed of cellulose. Good examples are cotton, jute, flax and wood, which comprise of 95-97%, 60-70%, 80% and 40-50% of cellulose respectively. In addition, several agricultural by-products such as cornstalk, rye, wheat, oats, rice, sugarcane bagasse and sorghum

contain cellulosic tissues. Lastly, industrial, agricultural and municipal wastes as well as animal manure also contain cellulose.

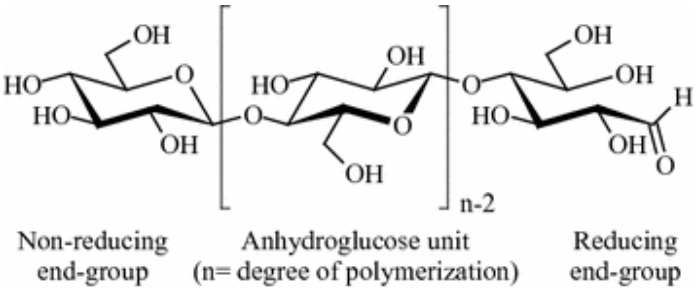


Figure 3.2: Chemical structure of microcrystalline cellulose (Adapted from Atindana *et al.*, 2017)

Since it forms part of plant material, cellulose is regularly replenished by plants following photosynthesis and plays an important role in the carbon cycle. Presently, cellulose constitutes a total volume of 324billion $m^{3.}$ This volume comprises of 40% of the annual net yield of biodegradable material from photosynthesis (Lee *et al.,* 2011). Previous reports have shown that cellulose contains 77kcal of incident radiation energy per gram of pure cellulose. Based on this energy composition, cellulose is classified as one of the potent low–cost energy sources (Lee *et al.,* 2011). All these characteristics make cellulose a promising feedstock for several industrial applications since there is regular replenishing driven by photosynthesis thereby ensuring a continuous supply of raw material to meet the markets' demands. Moreover, use of non-edible cellulosic plant material and industrial /agricultural by products eliminates any possibility of competing interest with human/animal food and feed as well as nature's balance and aesthetics. Nevertheless, cellulosic bio-resources remain greatly underutilized. This is partly because the most common industrially exploited and largely used cellulosic resources is harvestable trees. (Lee *et al.,* 2011). As a result, several other cellulosic biomass remain

unutilized. In addition, cellulose remains largely underused because of its structural complexities, which make it recalcitrant to simple degradation (Klemm *et al.,* 2005).

3.2 COMPOSITION OF CELLULOSIC MATERIAL

Cellulosic material/biomass is made up of three layers; A ***peripheral substances' layer, lignin layer and polysaccharides layer*** (including cellulose and hemicellulose)

3.2.1 Peripheral Substances' layer

The peripheral substances' layer constitutes of non-cell wall forming content. Depending on the solubility of these substances in water or organic solvents, they can be classified as either extractable or non-extractable. This layer of cellulosic materials/biomass plays a significant role in cellulosic material/biomass resistance and makes them resistant to insect attack and decay. On the other hand, however, it inhibits cellulose pulping and bleaching. The peripheral layer extractible components include; ***terpenes*** (terpene alcohols and ketones), ***resins (***non-volatile compounds such as fats, fatty acids, alcohols, phytosterols, resin acids and neutral compounds in small amounts), ***phenols*** (tannins and heartwood phenols), ***low molecular weight carbohydrates, alkaloids and soluble lignin.***

Non-extractibles in this peripheral substances' layer consist of inorganic materials present. The leading being, alkali earth carbonates, oxalates, silica, starch, pectin and protein molecules.

3.2.2 Polysaccharides layer

Cellulosic material/biomass polysaccharide layer constitutes of high molecular weight carbohydrates; cellulose (Fig.3.2) and hemicellulose (Fig.3.3). Hemicellulose is the principle non-cellulosic fraction of the polysaccharide layer and is made up of short chain polysaccharides that provide a link between cellulose and lignin. In nature, hemicellulose exist in three forms: ***xylan, mannans and galactans*** which exist as single units or collectively (Ahlgren & Ericksson 1969, Anderson & Han 1977,

Borrevik *et al.,* 1978, Casey 1960 & Lee *et al.,* 2011). **Xylan** are present as arabinoxylans, glucuronoxylans or arabinoglucuronoxylans. On the other hand, mannans are present in wood as glucomannans and galactomannans. Although relatively rare, galactans are found in the form of arabinogalactans (Borrevik *et al.*, 1978 & Cowling *et al.,* 1963).

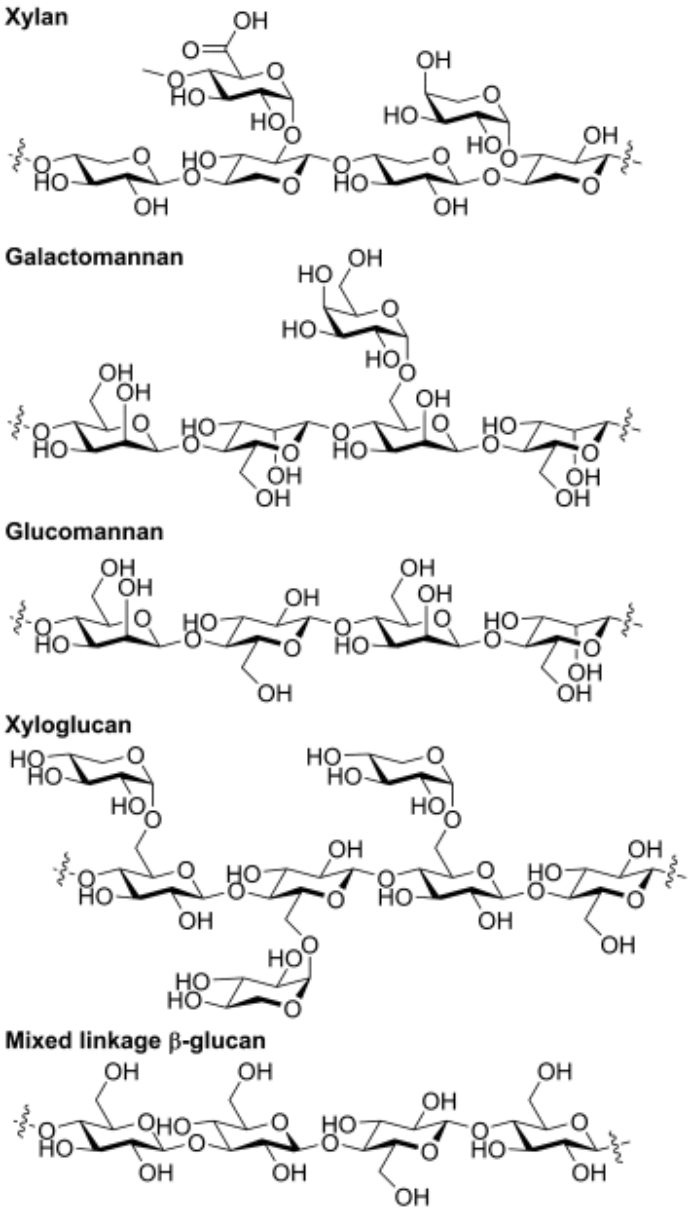


Figure 3.3a: Structure of various forms of hemicelluloses (Adapted from Shrotri *et al.*, 2017)

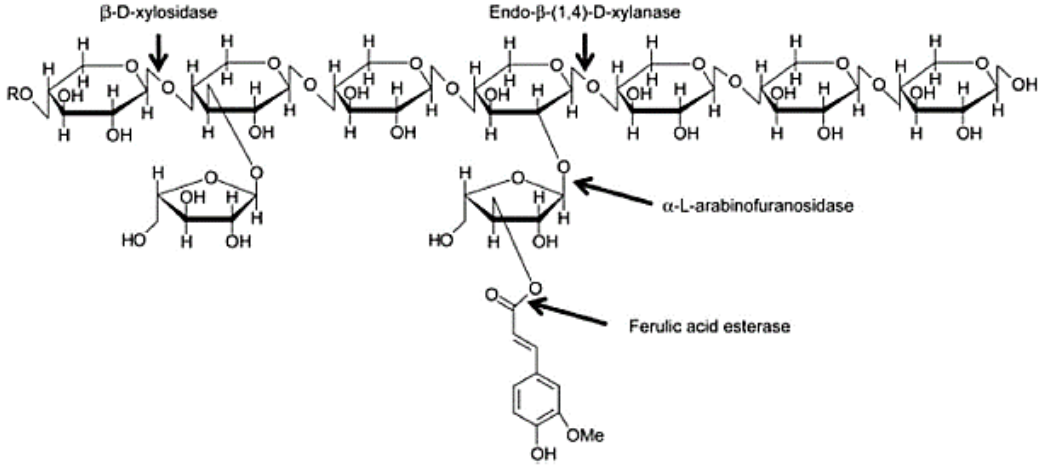


Figure 3.3b: Structure of arabinoxylans (AX) and the sites of attack by xylanolytic enzymes involved in its degradation.

The backbone of AX is composed of β-(1, 4)-linked xylose residues, that could be substituted with arabinose residues on the C (O)-2 and/or C (O)-3 position. Ferulic acid can be esterified on the C (O)-5 position of arabinose. Endo-β-(1,4)-d-xylanases (**EC 3.2.1.8**) cleave the xylan backbone internally, β-D-xylosidases (**EC 3.2.1.37**) remove xylose monomers from the non-reducing end of xylo-oligosaccharides, α-L-arabinofuranosidases (**EC 3.2.1.55**) remove arabinose substituents from the xylan backbone, and ferulic acid esterases (**EC 3.1.1.73**) remove ferulic acid groups from arabinose substituents **(Figure 3.3b)** (Dornez *et al*., 2009).

3.2.3 Lignin layer

Lignin is most concentrated in the outer layers of cellulosic materials/biomass. It is the most abundant large polymer in the cell wall other than cellulose. It is a complex, high molecular weight, amorphous and least characterized part of cellulose materials/biomass. Lignin is a high molecular 3D phenylpropane polymer held together by ether and carbon-carbon bonds. Aryl glycerol-β-aryl ether (Units 1-2, 2-6, 6-7, 7-8) is the most common interphenylpropane link found in lignin. Lignin encloses the bundle cells for example in wood fibers and sclerenchyma cells and plays as important role in structural rigidity. It does this by stiffening and holding together the polysaccharide fibers (Andren *et al*., 1976 & Lee *et al*., 2011).

Lignin content is 27-32% in woody plants and 14-25% in herbaceous plants. In addition, the degree of linkage between phenyl groups and methoxy substitution in lignin differs between softwoods, hardwoods and grass. However, the common structural features prevail in all (Berghem *et al.,* 1976 & Lee *et al.,* 2011). Degradation of lignin, yields phenylcoumaran (units 4-5 and units 15-16) (Figure 3.4) and diarylpropane (unit 3-4). Lignin also has biphenyl structure (units 11-12 and units 16-18), (Figure 3.4) which comprise 0.095-0.11 for every C_3-C_6 unit.

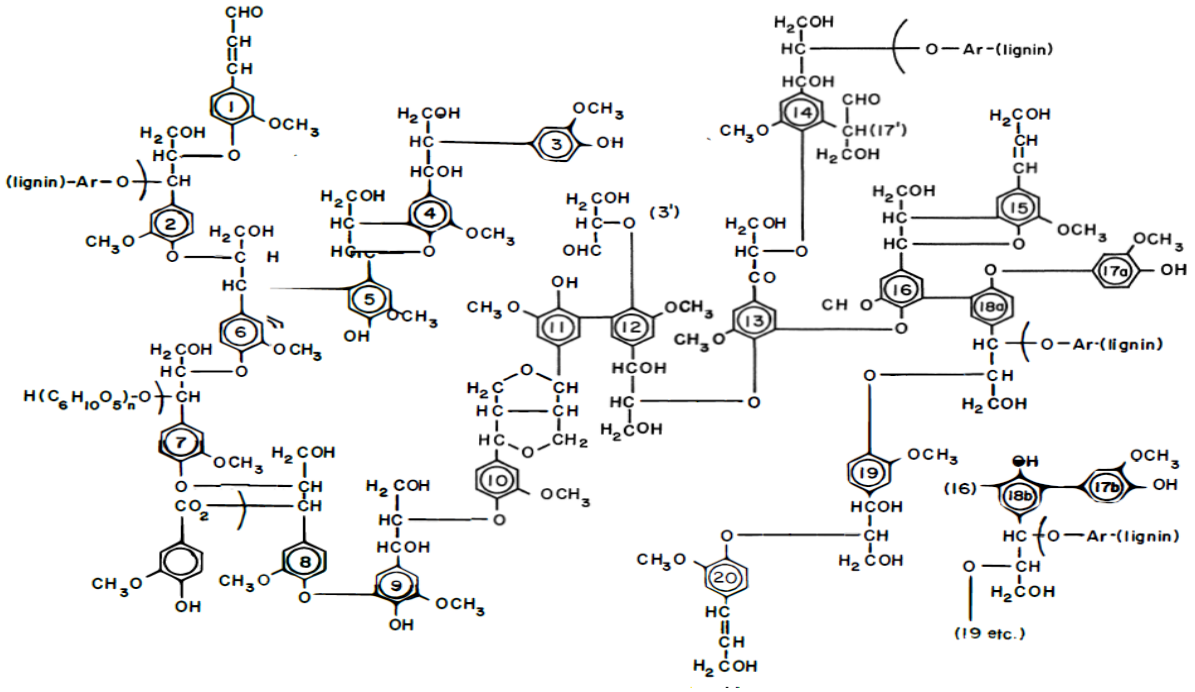


Figure 3.4: Molecular structure of spruce lignin (Adapted from Lee *et al.*, 2011)

3.3 MATERIALS COMPRISING OF CELLULOSE

As earlier mentioned in section 1.1, plant cell wall and plant material are made up of cellulose. These are composed of high molecular weight fibrils (approximately half a million) (Table 1) (Anon 1980 & Lee *et al*., 2011). Cellulose fibers are made up of a linear polymer of glucose linked in β-1,4 glycosidic bonds to form highly crystalline structures that are resistant to enzymatic hydrolysis (Arthur *et al*., 1968 & Lee *et al.,* 2011).

Table 3.1: Molecular weight and degree of polymerization of cellulosic materials

S/No	Cellulose Biomass/ Materials	Molecular weight	Degree of polymerization
1.	Native cellulose	600,000-1,500,000	3,500-10,000
2.	Chemical cotton	80000-500,000	500-3,000
3.	Wood pulp	80,000-340,000	500-2,100
4.	Rayon Filament	57,000-73,000	

Adapted from Lee *et al*., 2011

3.3.1 Plant Cell wall

Plant cell wall formation follows cell division. Plant cell wall, is classified as primary or secondary depending on the time of formation and or chemical composition. The primary cell wall is formed in the new cell plate, while the intercellular layers are formed between primary cell walls. With further cell differentiation, the secondary cell wall is then formed inside the primary cell wall and outside the protoplast. Continued differentiation leads to cell wall structure adaptation to the cell function.

Many plant cells only have the primary cell wall, while others have both the primary and secondary walls (Fig. 3.1). The main primary cell wall polysaccharides are cellulose (15-30% dry weight basis (dwb), hemicellulose, which interacts with cellulose to a form a network with microfibrils and pectin (30%). The primary cell wall is important for cells with active cell division such as mature cells that carry out photosynthesis, respiration, secretory action and cells that are important to callus reaction and cell regeneration. These cells therefore only have the primary wall.

Primary cell wall also contains proteins (such as expansins, enzymes, and glycoproteins), pigments, cutin, suberin, wax, fatty substances and ash (from ions such as calcium) (Chen, Hongzhang, 2014). Each of these components of the plant cell wall, lend it specific functions. For example, cellulose determines the main structure and framework of the plant cell wall while suberin coupled with wax;

cutin and uncoupled suberin help reduce water loss from plants. In addition, pectin serves in joining neighboring cells together in multicellular plants (Chen, Hongzhang *et al.*, 2014).

The plant primary cell wall has a number of functions:

I. To provide structural and mechanical support to the plant cell wall
II. To maintain and determine the cell morphology
III. To withstand cell swelling pressure
IV. Control the rate and direction of cell growth
V. To promote plant morphogenesis
VI. To regulate the material diffusion through the middle lamellae (ML)
VII. It acts as a carbohydrate reserve
VIII. To maintain pathogen resistance
IX. To prevent dehydration
X. Finally to activate the interaction between source signal molecules and cells

On the other hand, the secondary cell wall is important in cells specialized for mechanical reinforcement and water transportation. Unlike the primary cell wall, the secondary cell wall is made up of three main organic substances that is cellulose, lignin (instead of pectin found in the primary cell wall) and hemicellulose in the ratio of 4:3:3 although, this ratio varies between different lignocellulose materials. This highly cross-linked lignin layer that is specific to the secondary cell wall, lends its mechanical strength to support plant upward growth (Song, Shen & Li *et al.*, 2008).

In addition to being a part of the primary cell wall, the amorphous colloidal pectin molecule with high plasticity and hydrophobicity is also the main component of the intercellular layer.

An elaborate structure of wood plant cell wall is shown in Fig. 3.5 (Cowling 1975, Cowling 1976, Cysewski *et al*., 1976, Lee *et al*, 2011).

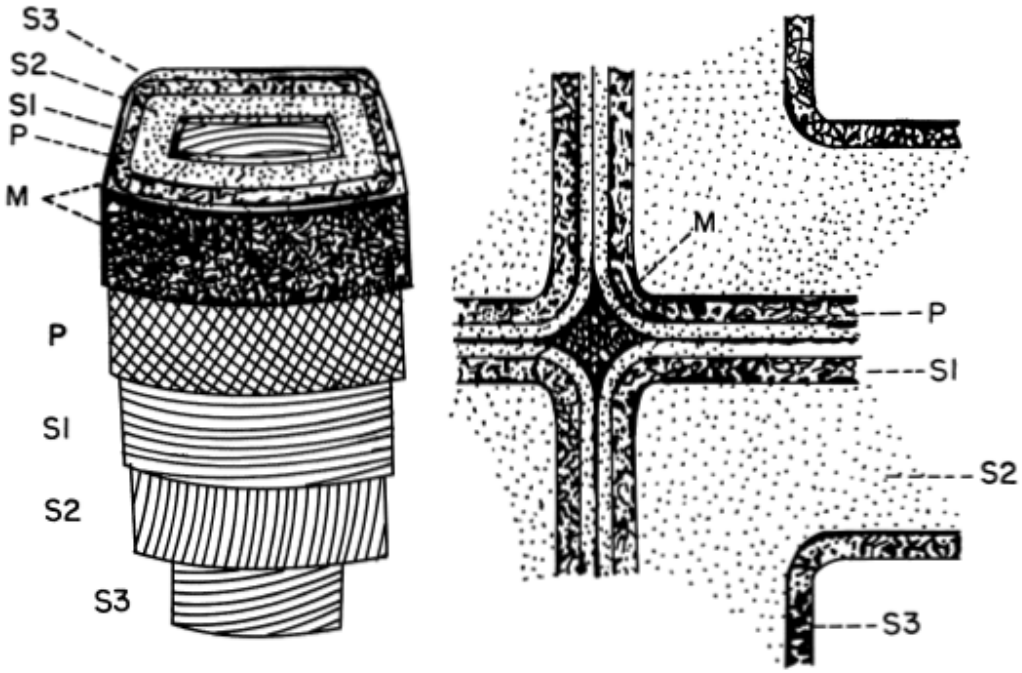


Figure 3.5: Elaborate structure of wood cell wall (Adapted from Lee *et. al.,* 2011)

It shows the primary wall (P), the thin outer layer of the secondary wall (S1), the substantial middle layer (S2), and the very thin inner layer (S3) sometimes called the tertiary wall.

In each of the layers of the secondary walls, cellulose and other cell wall constituents aggregate into long, slender bundles known as micro fibrils (Figure 3.5). It is important to note that microfibril structures are different since few if any cellulose molecules move from one micro fibril to another. The S1 layer has microfibrillar groups arranged in helices that are alternately crossed. Within the middle S2 layer however, the microfibrillar groups are arranged in bands (lamellae) that are almost parallel to the cell axis. The inner S3 layer is organized perpendicular to the S2. Further, the primary wall (P) is arranged irregularly around the cell axis. Adjacent to the P fiber layer is the heavily lignified stiff

middle lamellae (ML), which is shared by adjacent fibers (Arioli *et al*., 1998, Lee *et al*., 2011, Mueller *et al*., 1976 & Reiter 2002).

Figure 3.5 shows the distribution of chemical constituents in a typical wood plant cell wall. (Bender *et al*., 1970 & Lee *et al.,* 2011). The general scheme around the hollow lumen is such that the fibrillar elements composed of cellulose are wound in spirals hence making the fiber flexible and tenacious. In addition, the cable is waterproof and highly resistant due to the lignin and wax component (Detroy *et al.,* 1980 & Lee *et al.,* 2011). Lastly, the hemicellulose provides linkage between lignin and cellulose. The middle lamellae that is mainly composed of lignin is 1-2µ thick, amorphous and highly porous. The primary wall (P) remains thin throughout the plant growth. On the other hand, the secondary wall (S) that is majorly composed of cellulose thickens during cell growth phase.

3.3.2 Cotton

Cotton fiber is the seed hair from plants of the genus *Gossypium* and the hibiscus or mallow (*malvaceae)* family. It is the dominant natural fiber and the largest structural unit of cellulose, with fibrils of up to 500nm in diameter. In addition, it is the only natural pure cellulose consisting of up to 95-97% cellulose with a crystallinity of 70% as revealed by electron microscopy, and less than 10% raw fiber weight as wax, pectate, proteins and minerals. (Dochia *et al.,* 2012).

The strength of cotton is attributed to its highly fibrillar and crystalline structure. This strength is increased by 25% upon wetting. Its fibrillar structure consists of a primary cell wall, a secondary wall and a lumen (Losonczi *et al.,* 2004, Heikinheimo, *et al*., 2002 and Dochia *et al.,* 2012). The network formed by the entwining of microfibrils makes cotton/cellulose fibril stronger than steel wire of the same thickness. The primary wall structure, specifically the cuticle, which is the outer layer surface, has a major influence on cotton fiber properties, cotton processing and cotton uses (Figure 3.6 and Figure 3.7) below, show the archetypal components of dry mature cotton fibers. From this illustration,

it is clear that most of the non-cellulosic materials are present in the outer layer of the cotton fiber (Agrawal *et al.*, 2005; Wakelyn *et al.*, 2007 and Dochia *et al.*, 2012).

Microscopic examination of cotton fiber reveals a twisted ribbon or collapsed and twisted tube (Figure 3.6). There are about 60 twists /convolutions per centimeter of the cotton fiber. This convolutions give cotton an uneven fiber surface which increases the inter fiber friction and enables fine cotton yarns of high strength.

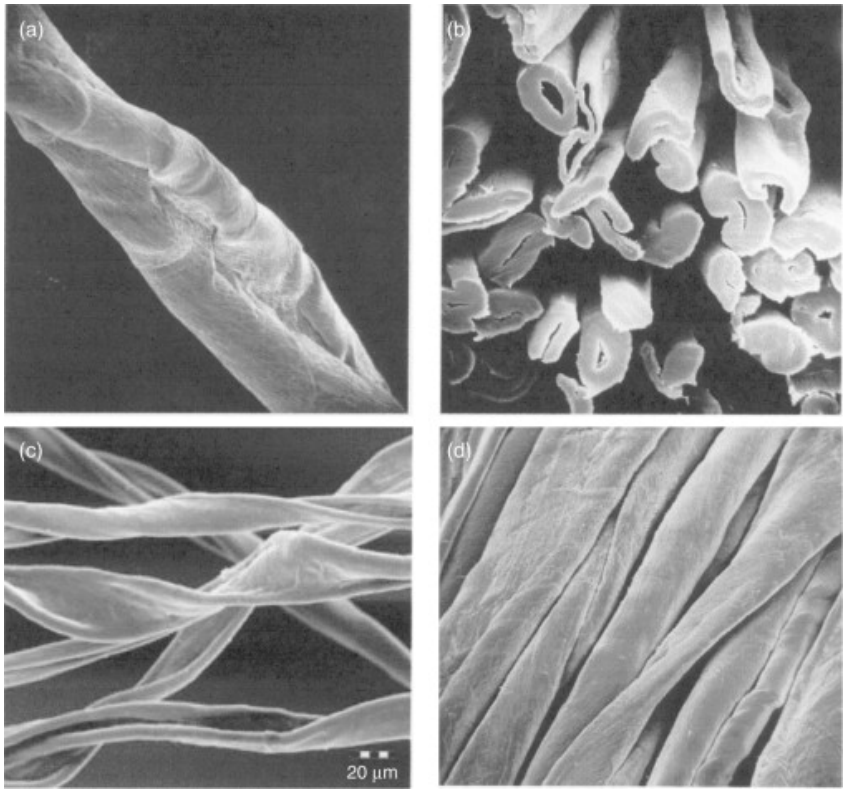


Figure 3.6: The appearance of cotton fiber under the microscope (a) Cotton fiber looking like a twisted ribbon; (b) the cotton fiber's cross section is referred to as being kidney-shaped; (c) cotton fiber looking like a collapsed and twisted tube and (d) bundle of cotton fibers (Adapted from Dochia *et al.*, 2012)

Cross sectional images of cotton fibers, show that they have a kidney shape. Scanning electron microscopy (SEM) images of cotton fiber have shown that the outer layer also known as the cuticle is a

thin film made up of fats and waxes (Figure 3.7b). This waxy layer contains smooth grooves and forms a thin sheet over the primary wall that forms grooves on the cotton surface. The primary wall consists of non-cellulosic materials and amorphous cellulose with criss-cross fibrillar pattern. The non-structured orientation of cellulose and non-cellulose components of the primary cell wall makes it open and disorderly. This however, has the advantage of making the primary wall flexible and is necessary for cell growth. An intricate interconnection is at play in the cotton primary cell wall between cellulose, hemicellulose, pectin, proteins and ions.

Unlike the primary cell wall, the secondary cell wall consists mainly of crystalline cellulose. Consequently, the secondary cell wall is highly systematic with a compact structure consisting of cellulose fibrils arranged in parallel pattern (Figure 3.7d)

Table 3.2: Chemical composition of cotton fiber

	Percent Dry Basis (%)		
Constituents	**Typical**	**Low**	**High**
Cellulose	94	88	96
Protein	1.3	1.1	1.9
Pectic substances	0.9	0.7	1.2
Ash	1.2	0.7	1.6
Wax	0.6	0.4	1.0
Malic, Citric and other organic acids	0.8		1.0

Total sugars	0.3
Pigment	Trace
Others	0.9

Adapted from Dochia *et al*., 2012 and Kanchagar *et al*., 2003

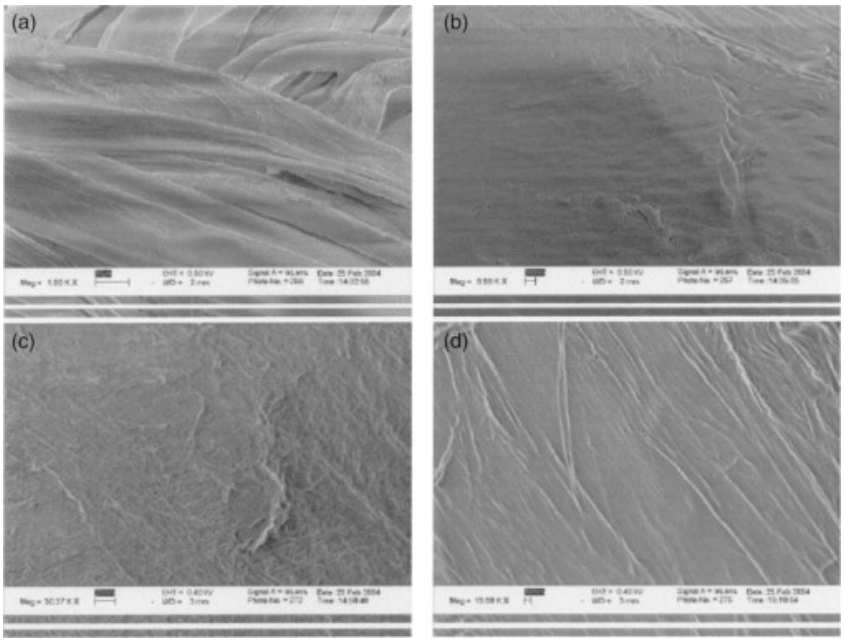


Figure 3.7: Single particle Cryo-Electron microscopy (SEM) images of the different layers. (a) Fibers from desized cotton fabric; (b) amorphous wax surface of the desized cotton fibre; (c) network of primary wall of cotton fiber and (d) secondary wall of cotton fiber

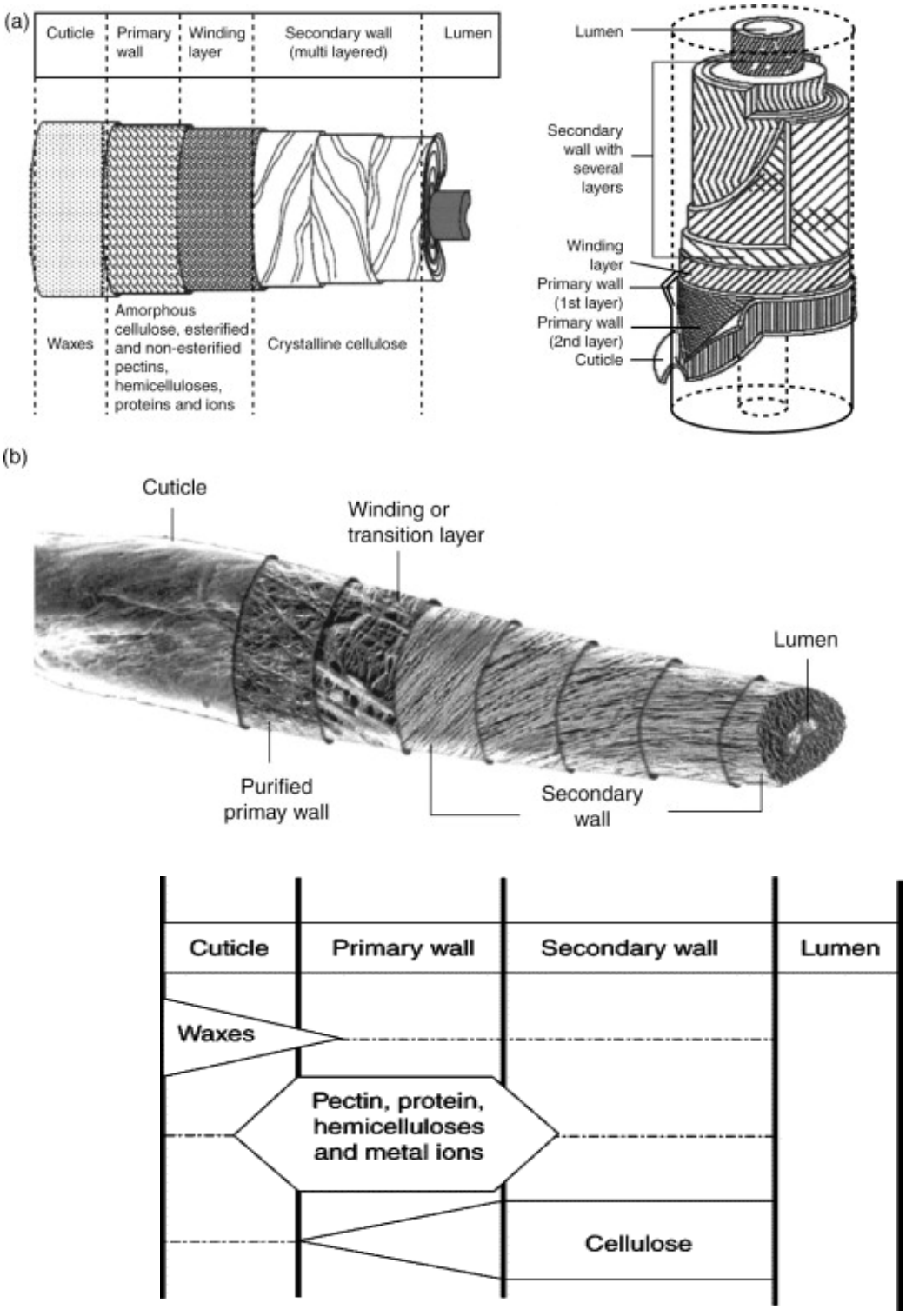


Figure 3.8: A schematic representation of mature cotton fiber showing its various layers (a) Cross section of cotton fiber. Typical components in dry, mature cotton fibers and compositions of each layer. (b) Morphological model of cotton fiber (Mangat 2009 & Agrawal 2005).

3.3.3 Wood

As mentioned earlier in section 3.2, cellulosic material is made up of three major components. Wood is no different. It is made of a peripheral substance layer, a polysaccharide layer and a lignin layer. Wood polysaccharide layer, makes up to 60-80% of the total wood and is made up of mainly cellulose and hemicellulose. Cellulose is the major component of wood fiber cell walls (Figure 3.9).

The lignin layer in wood is the most complex and least characterized layer. It makes up to 20-35 % of the total wood content and is concentrated in the outer layer fibers (Figure 3.10 and Figure 3.11).

Significant compositional differences exist between softwood and hardwood species. The following inter- wood species differences stand out:

I. The lignin content of softwoods is higher than hardwood
II. Their hemicellulose content is similar
III. Cellulosic content of hardwoods is higher than that of softwoods
IV. Hardwood hemicelluloses are rich in xylan polymers with traces of mannans, while softwood hemicellulose are rich in mannan polymers and copious amounts of xylan polymers
V. The mannan groups in the glucomannans from deciduous (hard) wood and coniferous (soft) wood are different; the former lacks galactose units.

Finally, other than the inter-wood species compositional differences, there equally exist compositional differences within single wood species.

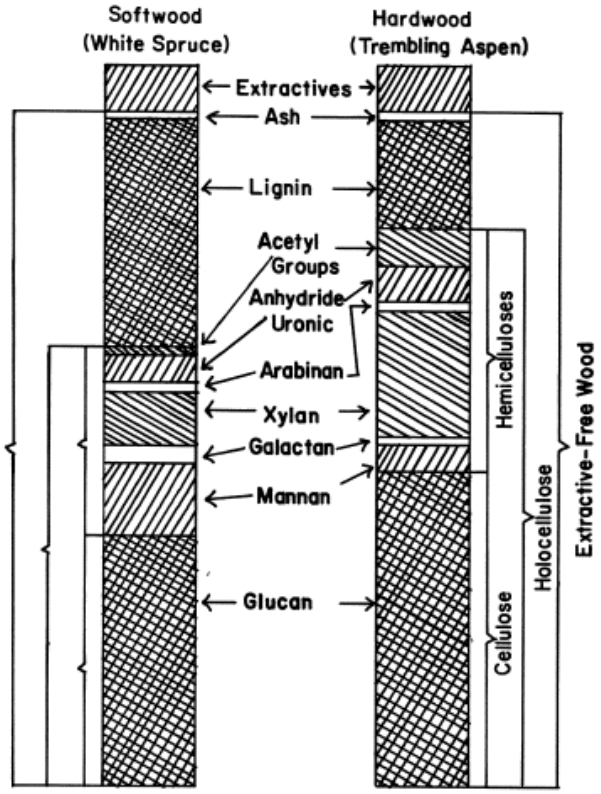


Figure 3.9: Comparison between hardwood and softwood compositions (Lee *et al.*, 2011)

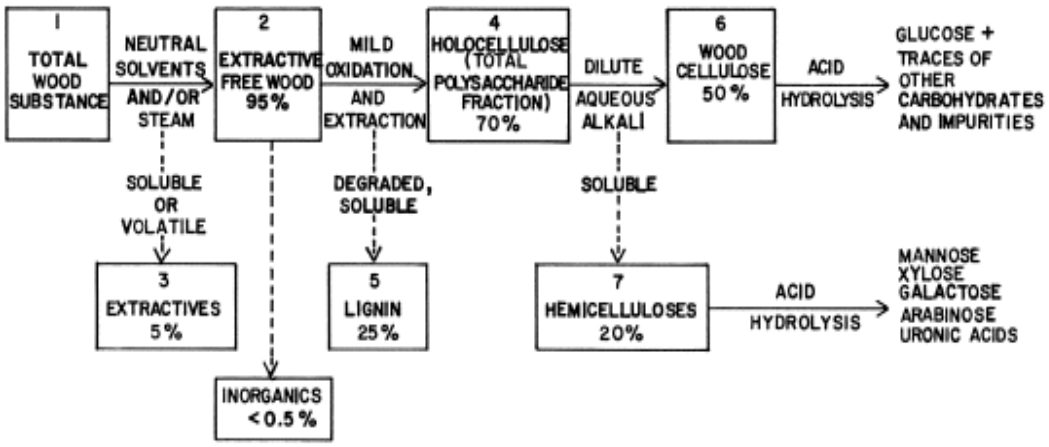


Figure 3.10: Structure showing the major components of wood (Lee *et al.*, 2011)

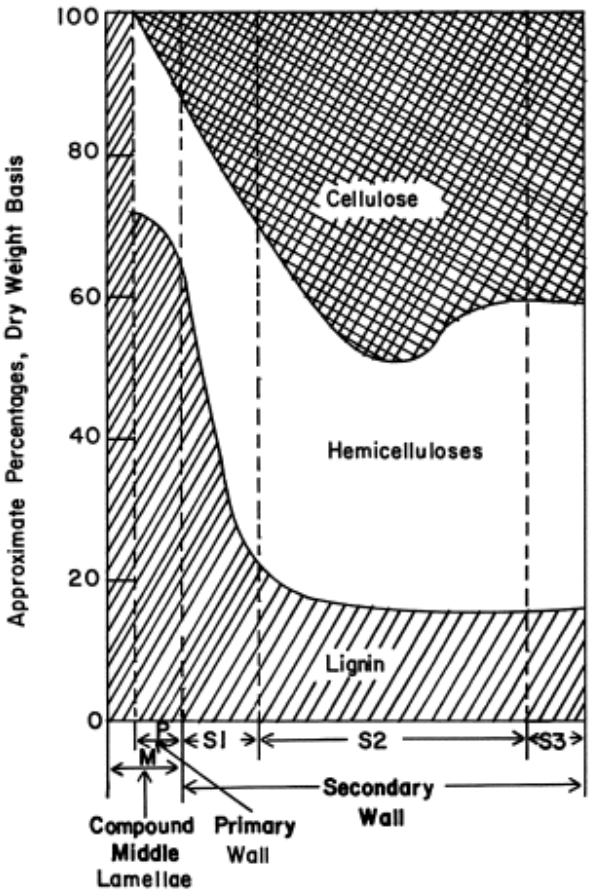


Figure 3.11: Chemical constituents and their distribution in the wood plant cell wall (Lee *et al.*, 2011)

3.3.4 Straw

Unlike wood cell wall, straw cell wall has not been well studied. From the data available (Table 3.3) however, straw composition is generally uniform and there exist compositional differences between straw and wood. Unlike wood, straw has a lower cellulose content. Nevertheless, its total carbohydrate content also known as holocellulose is equal to that of wood. This can be attributed to the high hemicellulose and ash content and low lignin content compared to wood (Lee *et al.*, 2011).

Straw is much more heterogeneous in comparison to wood. Straw fibers are principally from cells and internodes, fairly long, and slender with sharply pointed ends. In addition, straw contains short non-fibrous cells constituting of epidermal cells, platelets, serrated cells and spirals from nodes, pitch, chaff and rachises (head). In conclusion, 96% of cells in wood are fiber whilst only 35-39% of straw cells are fibers (Lee *et al*., 2011).

Table 3.3: Composition of various types of American straws

Component /Source	Rice	Barley	Wheat	Rye	Oat	Flax Shives	Soybean stalks
Moisture	8.0	8.4	6.6	7.4	7.0	8.1	8.3
Ash	16.1	6.4	6.6	4.3	7.2	3.5	2.3
Extractives							
Alcohol-benzene	4.6	4.7	3.7	3.2	4.4	4.1	3.9
Cold water	10.6	16.0	5.8	8.4	13.2	9.7	7.3
Hot water	13.3	16.1	7.4	9.4	15.3	11.4	8.8
1% NaOH	49.1	47.0	41.0	37.4	41.8	32.0	32.0
Nitrogen	0.58	1.10	0.38	0.72	0.46	0.69	0.66
Lignin	11.9	14.5	16.7	19.0	17.5	22.3	19.8
Pentosans	24.5	24.7	28.2	30.5	27.1	23.6	24.8
Cross & Bevan Cellulose							
Crude	49.8	48.2	54.4	54.9	53.6	48.4	50.1
Ash free	48.6	47.4	53.6	54.3	53.4	46.2	48.2
Pentosans	27.8	30.0	26.8	29.5	28.4	21.6	22.1
Alpha Cellulose	36.2	33.8	39.9	37.6	39.4	34.9	34.5

3.3.5 Microalgae / Microphytes

Microalgae are miniscule algae present in both freshwater and marine systems in the water columns as well as sediment. They exist individually or in chains or groups. There are four types of microalgae; red, green, brown and blue-green whose sizes vary from a few to hundreds of µM depending on species. A huge microalgae biodiversity exist and remains untapped. Presently, we have 200,000-800,000 microalgae species from different genera with only 50,000 species that have been

described. Microalgae produce diverse compounds such as carotenoids, antioxidants, and fatty acids for example omega -3 fatty acids, enzymes, polymers, peptides, toxins, sterols as well as chemical signals that contribute to defense, avoidance and prey selection. The chemical signals propagate by simple diffusion and laminar advective flow affect algal blooms.

Microalgae chemical composition is not constant and varies depending on species as well as cultivation conditions. In addition, microalgae have a high adaptability quotient that enables them to alter their chemical composition in response to environmental changes. A classical case is the substitution of phospholipids with non-phosphorus membrane lipids in phosphorus deficient environments. This is important since a number of hatcheries are in place for growing microalgae for various commercial purposes. This therefore means that environmental and /hatchery conditions such as lighting, temperature, pH, CO_2 supply, salt and nutrients can be controlled in alga culture for specific purposes.

Microalgae have a high content of carbohydrates/polysaccharides (PS) mostly sulphated (sPS) such as glycogen, starch, agar and cellulose. The chemical structure of PS produced by microalgae determine their physico-chemical and biochemical properties as well as their physical behavior and biological activities.

These polysaccharides can be converted to fermentable sugars for bioethanol production. Figure. 3.12 shows some microalgae species suitable for algal bioethanol production. Micro algal sPS (sulfated polysaccharides) are also associated with biological activities, health benefits and subsequent potential applications in regenerative medicine, therapeutics and pharmaceuticals (Raposo *et al*., 2015).

.

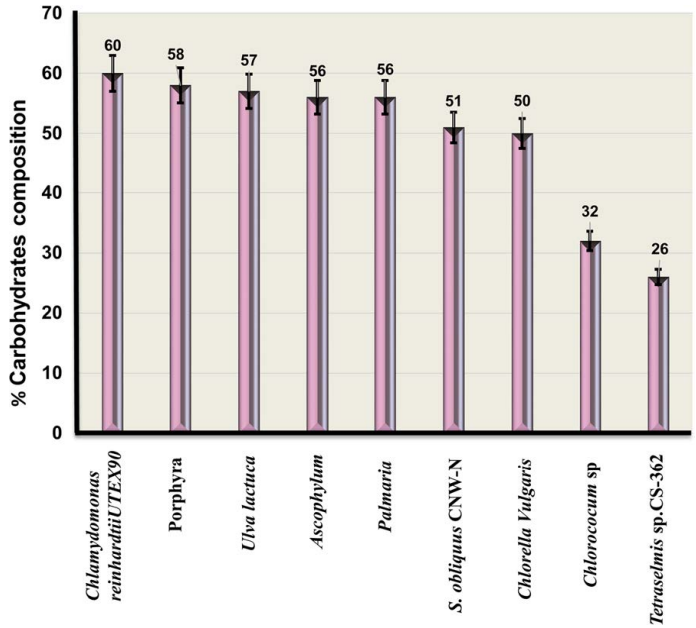


Figure 3.12: Different species of carbohydrate-rich microalgae that compose feasible feedstock for bioethanol production (Khan *et al*., 2018).

3.4 CHEMICAL, PHYSICAL AND MOLECULAR STRUCTURE OF CELLULOSE

Native cellulose is a high molecular weight (above 1.5M) linear polymer that is composed of up to 10,000 D-anhydroglucose units joined by β-1,4 glycosidic bonds to form long chain molecules .It mainly contains carbon 44.4%, hydrogen (6.17%) and oxygen (49.39%). The chemical formula of cellulose is $(C_6H_{10}O_5)_n$, n here, represents the degree of polymerization (DP), which ranges from a hundreds to tens of thousands, and represents the number of glucose groups. The length of each anhydroglucose unit is 0.515nm (5.15 Å). This makes the total length of native cellulose molecule 5µM. (Fan *et al*., 1980; Chen *et al*., 2014 & Lee *et al*., 2011).

Different linkages are present between hemicellulose, cellulose and lignin molecules (Table 3.4 & Table 3.5). H-Bonds join cellulose to hemicellulose or lignin. In addition, there exists chemical bonding between hemicellulose and lignin. On the other hand, hemicellulose and lignin molecules fill the fibers. This explains the lignin carbohydrates complexes (LCCs) separated during lignin extraction from natural

lignocellulose biomass (Chen *et al.*, 2014). The chemical bonding that is observed between hemicellulose and lignin is because of chemical bonding between galactose and arabinose residues on the side chains of hemicellulose, lignin and carbohydrate molecules.

3.4.1 CHEMICAL STRUCTURE OF CELLULOSE

Glucose molecules within the cellulose chain molecules assume a chair configuration that positions the hydroxyl and hydrogen groups at the equatorial and axial position respectively (Figure 3.13) below. Within this chair form of cellulose, alternate chain units are rotated 180° around the main axis. This leads to an unstrained, linear configuration with minimum steric hindrance.

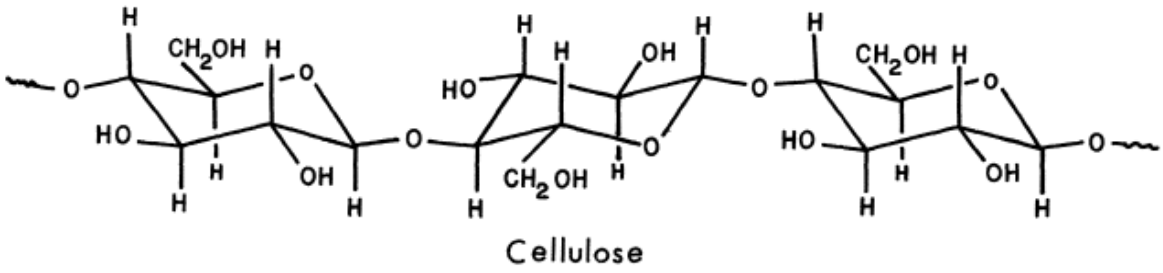


Figure 3.13: Conformational formula (Chair form) of cellulose of cellulose (poly-1, 4-D-glucosan) (Fan *et al.*, 1980 & Lee *et al.*, 2011)

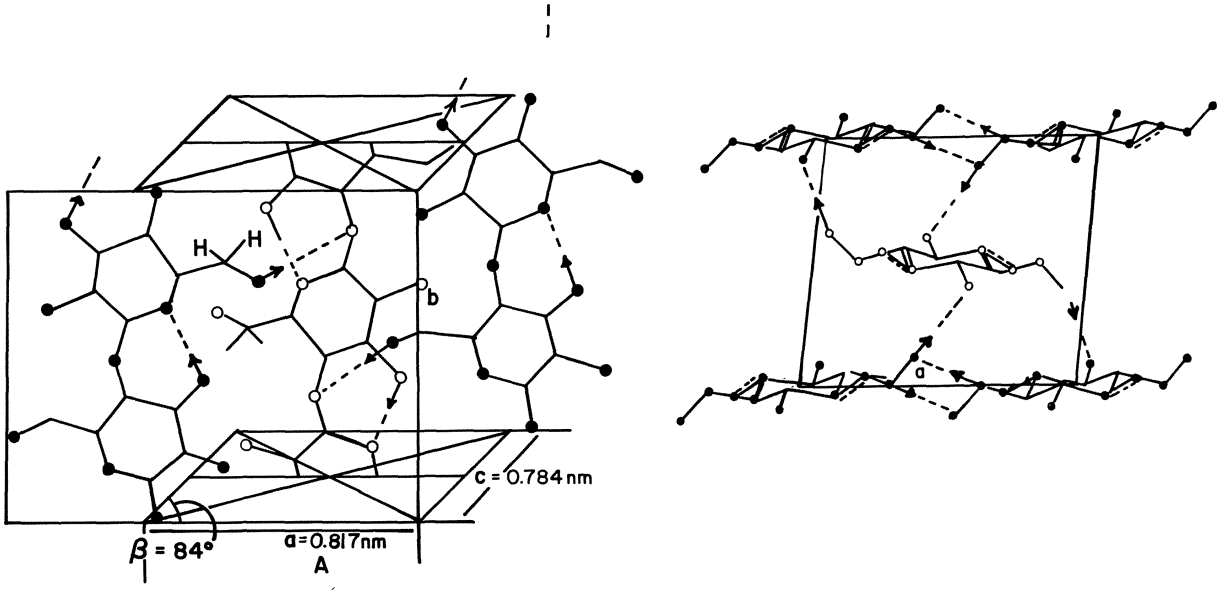


Figure 3.14*:* The arrangement of cellulose molecules and hydrogen bonds in a cellulose I –unit cell (Lianyg *et al.*, 1959 & Lee *et al.*, 2011)

Table 3.4: Stability of native cellulose

Dimensions	Length (nm)	Stability (Kcal/mole)	Nature of bond
a	0.817	15	Hydrogen
b	1.031	50	Covalent
c	0.784	8	Van der Waals

Adapted from (Sihtola and Neimo, 1975 & Lee *et al.*, 2011)

The glycosidic bonds in cellulose act as functional groups. Together with hydroxyl (-OH) groups, they influence the chemical properties of cellulose. Previous studies have shown that significant chemical reactions occur at these points.

3.4.2 PHYSICAL STRUCTURE OF CELLULOSE

This refers to the spatial arrangement of different cellulosic structural units (chain structure and aggregation structure). Chain structure also known as the primary structure is the geometric arrangement of atoms or groups in a molecular chain. It is the stereo chemical structure of one or more structural units in a single molecule polymer. This is the main structural element. Consequently, many cellulose polymer characteristics like melting point, density, solubility, viscosity and adhesion can be attributed to the chain structure. Aggregate structure or secondary structure is the inner structure of the whole polymer including the crystal structure, non-crystal structure, orientational structure and liquid crystal structure. This structure determines the service performance of macromolecular compound products (Zhan *et al.*, 2005).

3.4.2.1 Filament Structure

Small stretching units known as fibrils form the filament structure. Each fibril contains about 60-80 cellulose molecules. These fibrils aggregate and constitute the structure of some natural and synthetic fibers biomass like wood, textile fiber and fibrous protein. Fibrils also make long molecular chains that

gather into parallel running bundles. There is a wide variety of aggregated fibril sizes ranging from elementary fibril, microfilament to macrofilament.

3.4.2.2 Aggregation structure

This is also known as the supramolecular structure of cellulose. It describes the arrangement of cellulose molecules into the crystalline or amorphous structures and subsequently into elementary fibril or protofibril, about 40 Å wide, 30 Å thick, and 100 Å long, as well as microfibrils structures. X-ray diffraction studies reveal a regular arrangement of the crystalline cellulose region. This region has a high density (1.588g.cm^{-1}). Conversely, the amorphous region is irregularly, and loosely packed leaving large distances between the molecules. In addition, its density is relatively low (1.500g.cm^{-1}). Within the aggregate structure, the chain molecule lies parallel to the main spindle of cellulose.

Generally, crystallinity of cellulose is between 30-80 % and shows pleomorphism. Solid cellulose is associated with five different crystal modifications. Type I is the crystal form of natural cellulose, type II, III, IV and X are crystal forms of cellulose under artificial processing.

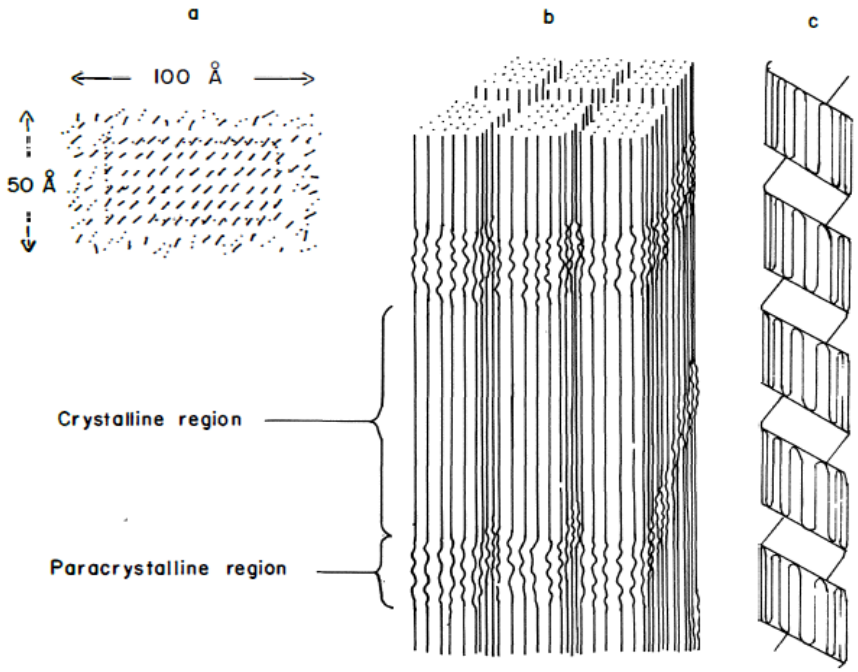


Figure 3.15: Structure of cellulose microfibrils

Cellulose transitions from one crystal form to the other through heat and /extensive chemical treatment. No clear boundary exists between the cellulose crystalline regions (known as micella, micel or microcrystal) and amorphous regions of cellulose. While the amorphous region has free hydroxyl groups (-OH), the microcrystal region -OHs in position 2, 3 and 6 are involved in hydrogen bonding.

3.4.3 PHYSICOCHEMICAL PROPERTIES OF CELLULOSE

3.4.3.1 Cellulose Chemical properties

Each glucose residue on the cellulosic molecule has three free and active -OHs (a primary –OH and 2 secondary -OHs). These are involved in a number of chemical reactions as well as hydrogen bonding. These hydrogen bonds specially the H-Bond between hydroxyl at C3 and oxygen at the adjacent molecule ring, determine the morphology and reactivity of the cellulose chains. They do this by enforcing a strictly linear rigidity and integrity of the cellulose molecule. In addition, they also ensure close range packing of the cellulose molecule to form the highly ordered crystalline cellulose.

3.4.3.2 Physical properties of cellulose

The free –OH groups of cellulose has attraction to several solutions and solvents. Water is however adsorbed only the amorphous region and not in the crystalline region.

3.5 CELLULOSE PRE-TREATMENT

Efficient cellulose biomass utilization is hindered by inaccessibility. Cellulosic biomass degradation without pre-treatment is often very low with yields of up to 20% only (L.T.Fan , M.M Gharpuray, 2007). This because processing reagents are unable to reach the reactive cellulose hydroxyl [OH^-] groups. In heterogeneous reactions, inaccessibility results due to unequal ratio of crystalline and amorphous regions. The primary [OH^-] and two secondary [OH^-] groups determine the primary cellulose reactivity. Generally, the group with least steric hindrance is the most reactive. This makes

the primary [OH^-] reactivity higher relative to the secondary [OH^-]. Consequently, any commercial processes that require cellulose hydrolysis have been greatly hindered due to low economic feasibility. In order to make cellulose hydrolysis economically viable especially in heterogeneous multi-phase reaction media, to increase the reactivity of the cellulose and to improve the quality of cellulosic products obtained therefore, a preliminary cellulose pretreatment step that can complement the cellulose hydrolysis process is essential. Such a regimen needs to be cheap, fast, and economically practical in order to complement the cellulose hydrolysis step. In addition, the pretreatment regimen has to be friendly enough to ensure the quality of the final product remains uncompromised. The lignin–carbohydrate complex, other component complexes in cellulosic biomass as well as the rigid cellulose structure itself are disrupted using various methods (Table 3.5):

Table 3.5: Different Lignocellulose Biomass Pretreatment Regimens

Physical Treatments	Chemical Treatments	Biological Treatments
Mechanical methods	**Swelling Agents**	Fungi
Ball milling		Bacteria
Two roll	**Using of Gases**	
Colloid milling	Sulfur dioxide	
Hammer milling	Nitrogen dioxide	

	Chlorine dioxide
Non-Mechanical Methods	**Alkali methods**
High pressure steaming	Ammonia
Pyrolysis	Ammonium Sulphite
Extrusion	Sodium hydroxide
High energy radiation	
Expansion	**Acid Methods**
Vibro energy milling	Phosphoric acid
Microwave treatment	Hydrochloric acid
	Sulfuric acid
	Oxidizing Agents
	Hydrogen Peroxide
	Ozone
	Cellulose solvents
	Cadoxen
	CMCs
	Solvent extraction of Lignin
	Butanol water extraction
	Benzene ethanol extraction
	Ethanol water extraction
	Ethylene glycol extraction

These are classified based on their mode of action on the cellulosic biomass. In addition, they can be done in parallel or sequentially.

3.5.1 PHYSICAL PRETREATMENTS

Physical pre-treatment regimens are classified as non-mechanical or mechanical (Table 3.5). Mechanical methods work by breaking the cellulosic biomass into finer particles. In addition, mechanical methods also reduce the crystallinity of the biomass as well as the surface area therefore making the cellulosic biomass easily accessible for hydrolysis. Non-mechanical methods on the other hand lead to cellulosic biomass decomposition due to the harsh treatment, which include pyrolysis, high pressure, irradiation, steaming and microwaving.

3.5.1.1 Ball milling

Ball milling is an efficient method of cellulose pretreatment that allows for reduction in crystallinity due to the shearing compression applied. Further, it leads to a decrease in the degree of polymerization (DP), particle size and as well as an increase in the bulk density. Moreover, ball milling is also associated with reduced capital costs due to slurry concentration and the consequent reactor volume reduction. Up to 72.9% degree of hydrolysis (DH) was achieved, for 325 magazines in 48hours .Nevertheless, the recorded ball milling efficiency varied from one cellulosic biomass to another. Solka Floc, showed decrease in crystallinity after 96 hours with no detectable changes in surface area. In addition, the DH was shown to be proportional to the ball milling time (Figure 3.16). As a result, ball milling is impractical on large scale due to time and processing costs.

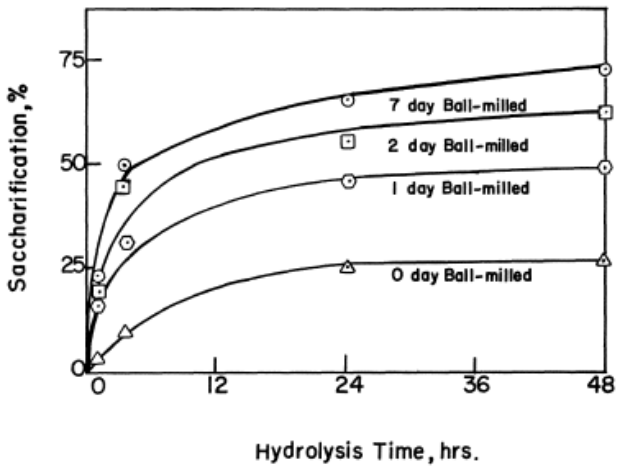


Figure 3.16: The Degree of hydrolysis of ball milled newspaper (Adapted from L.T.Fan , M.M Gharpuray, 2007)

3.5.1.2 Pyrolysis

Pyrolysis is the exposure of lignocellulosic material to high temperatures. Above 300°C, cellulose decomposes to gaseous and tarry compounds that leave char on evaporation. This decomposition is rather slow with very little product formation at intermediate temperatures. Previous studies have shown that tar yield from vacuum pyrolysis was 70% and 78% at 375°C and 425°C respectively. This tar yielded 80-85% reducing sugars with more than 50% glucose yield from cellulose following mild acid hydrolysis (in H_2S0_4 at 97°C for 2.5h). In addition, the type of atmosphere present also determines pyrolysis rate. In the presence of oxygen, dehydration and depolymerisation oxidation speeds rapidly increase.

On the contrary, inert atmosphere retards depolymerisation as well as oxidation and dehydration side reactions. Solka Floc treated by pyrolysis at 170°C in air and helium showed insignificant changes in crystallinity index. However, a significant increase in the DH of Solka Floc treated in helium was reported. This significant upward shift in the DH is attributed to the increase in DP of helium treated cellulose. Furthermore, previous studies, have shown that zinc chloride speeds up decomposition of cellulose even at lower temperatures (L.T.Fan , M.M Gharpuray, 2007).

3.5.1.3 Microwave treatment

Previous studies have showed an increase in the DH of bagasse pretreated by microwaving in sealed glass vessels by up to 3.2X compared to the untreated bagasse.

3.5.1.4 Extrusion and Expansion

These methods are similar to high pressure steaming and involve the use of moist heat expansion (extrusion) and dry heat expansion (popping). Previous research showed that extrusion at 90s is an inefficient pretreatment method for rice straw and sugarcane bagasse. On the other hand, other studies also showed that extrusion is a promising pretreatment method for newspapers that would be hydrolyzed by acid.

3.5.1.5 High Pressure Steaming

Here, the cellulosic biomass is steamed at high pressure and temperatures without the addition of chemicals. As a result, acetyl groups are cleaved off leaving an acidic medium that fosters hydrolysis.

3.5.1.6 High Energy Irradiation

High-energy irradiation, leads to the oxidative degradation of cellulose molecules, dehydrogenation, and destruction of glucose to yield carbon dioxide and cellulosic chain cleavage. There is an increase in surface area due to the decrease in the DP due to extensive depolymerisation yielding pentoses and hexoses. However, no change in crystallinity index was recorded. However, this method is species specific and not economically viable due to the costs involved. Nitrite photochemical radiation is an alternative to high-energy irradiation. It involves the irradiation of cellulose by ultraviolet light in an aqueous solution of sodium nitrite.

3.5.1.7 Hammer Milling

This technique involves the use of a rotor with an attached hammer. As the rotor turns, the hammer affects the lignocellulose biomass. Studies on newspaper showed that there was a slight increase in the susceptibility of hammer milled biomass over untreated biomass. However, extensive milling reduced the susceptibility of the newspaper to hydrolysis.

3.5.1.8 Colloid Milling

Colloid milling encompasses two disks set close to each other revolving in opposite directions as the cellulosic biomass slurry passes between the disks. Previous studies, showed slight increments in the susceptibility to hydrolysis following colloid milling. However, the associated high operational costs make it economically non-feasible on a large scale**.**

3.5.1.9 Vibro energy milling

This method is similar to ball milling. It is efficient in the reduction of size and digestibility of the biomass. Spruce and Aspen wood pretreated by vibro energy milling showed an enhanced susceptibility to enzymatic hydrolysis. However, it is also biomass specific with hardwoods being more susceptible in comparison to softwoods. Solka Floc pretreated by vibro milling for 24-48hrs showed 1.7X more reducing sugars than untreated biomass. Further, the reducing sugar quantities released were enhanced by heating at 200°C before or after pretreatment.

3.5.1.10 Two roll Milling

Two roll milled pretreated maple chips yielded 17X more reducing sugar than untreated maple chips. Further, the method has been shown to be 2.5X more efficient than ball milling. In addition, it has the advantage of reduction in the sediment volume, which allows for slurry concentration and a lower reactor volume that is associated with lower capital costs. Further, it enhances the DH of biomass by enhancing cellulosic biomass accessibility through the reduction of the degree of crystallinity and DP of cellulosic biomass. These enhancements are affected by clearance between the rolls and the processing time. Consequently, there is increased accessibility with the decrease in clearance between the mills and an increase in processing time.

3.5.2 CHEMICAL PRETREATMENTS

Chemical pretreatment is the most common pretreatment method used for reducing the crystallinity and removing the lignin complexes in cellulosic biomass, particularly for the paper industry. However, it is very expensive and harsh leading to a reduction in the quality of the final product. Therefore, there is a need for alternatives/ substitutes to chemical pretreatment methods.

Cellulose is insoluble in water, dilute acid, and dilute alkali solutions at normal temperature because the innumerable hydrogen bonds holding the chains together are not broken in these solutions. However, strong acids, strong alkalis, concentrated salt solutions, and various complexing reagents can swell, disperse and even dissolve the cellulose (Goldstein IS, 1981, Lee *et al*., 2011, H. Chen., 2014).

Acid hydrolysis, alkali hydrolysis and enzyme hydrolysis of cellulose help to break down glycosidic bonds. Alkali peeling and oxidation-reduction reactions take place at cellulose reducing ends. On the other hand, oxidative degradation of cellulose function to break the C2, C3 and C6 glucosyl rings. Carbonyls are formed at C2 when oxidized. This is then later degraded by the elimination of the O-alkoxy following alkali treatment.

3.5.2.1 Acid Pretreatment Methods

Acids are mainly used for cellulose hydrolysis rather than for pretreatment. Acids used for pretreatment are hydrochloric acid, sulfuric acid and phosphoric acid

3.5.2.1.1 Hydrochloric Acid

HCl pretreatment is done using 3% HCl at 130°C for about 4 hours

3.5.2.1.2 Phosphoric Acid

This is typically done using 85% phosphoric acid. The samples are stored at 2°C for 2 different time periods, 10 minutes and 2 hours. Digestion yield of 80% was attained in 100hrs. Phosphoric acid is also used to swell cellulose to obtain phosphoric acid swollen cellulose (PASC). D.H is directly proportional to the accessible surface area of PASC.

3.5.2.2 Sulfuric acid

Several Sulfuric acid regimens are available. Using 50% sulfuric acid at 121°C for 15min, or the use of concentrated acid for a short time interval followed by precipitation in acetone. Using the former regimen, up to 95% of cellulose was converted to glucose following enzymatic hydrolysis after 24hours.

3.5.2.2 Alkali Pretreatment Methods

3.5.2.2.1 Sodium Hydroxide

Sodium hydroxide is mainly used for enhancing digestibility of lignocellulosic biomass rather than for pretreatment. Dilute solutions of caustic soda are used for swelling which leads to an increase in surface area, decrease in DP, crystallinity, separation of lignin from cellulose and delignification. NaOH pretreatment is species specific. It affects the digestibility of straw and hardwood but with negligible effect on cotton.

3.5.2.2.2 Ammonia

Lehmann *et al*., 1950 first patented this cellulosic biomass pretreatment method. Liquid ammonia is use as a swelling agent for cellulose. It leads to the conversion of crystalline cellulose I to cellulose III. Aspen wood treated with ammonia showed an increase in percentage reducing sugars from 11% in untreated biomass to 36%. In addition, unlike NaOH, liquid ammonia demonstrated a significant increase in yield of reducing sugar. In addition, ammonia is used for increasing digestibility of lignocellulosic biomass for ruminants.

3.5.2.2.3 Ammonium Sulphite

Ammonium sulphite is mainly used, for the traditional pulping process. It is also used, for increasing digestibility of biomass.

3.5.2.3 USE PRESWELLING AGENTS

Strong electrolytic solvents are primarily the solvents used for cellulose pretreatment. There exist two different types of swelling methods; intercrystalline and intracrystalline. Water penetrates only the amorphous cellulose region. It is therefore considered an intercrystalline swelling agent. Intracrystalline swelling agents are salt or alkali swelling agents, which affect both amorphous and crystalline regions of cellulose. They aide cellulose degradation by weakening hydroxyl binding forces between cellulose molecules, therefore increasing reagents diffusion velocity. Some intracrystalline cellulosic swelling agents include; sodium hydroxide, sulfuric acid, nitrogen dioxide in dimethylsulfoxide ,zinc chloride, ruthenium red, phosphoric acid, trimethylbenzylammonium hydroxide, iron tartrate complex, methacrylate embedding, sodium zincate as well as glacial activated caustic solution .

3.5.2.4 DECREASE IN CELLULOSE CRYSTALLINITY BY CELLULOSE DERIVATIZATION

Cellulose derivatives possess a high degree of substitution and free -OH groups are substituted. Some cellulose derivatives have a low degree of substitution and a higher water absorbability. This include methyl, ethyl, and hydroxyethyl and hydroxymethyl cellulose ether. These groups lead to the swelling of the cellulose structure and decreased binding force in the macromolecules.

3.5.2.5 REDUCTION IN CELLULOSE CRYSTALLINITY BY IMMERSION IN ETHYLAMINE

Cellulose immersion in ethylamine, leads to decrease in the DP by 20%. However, ethylamine only penetrates microfilaments. As a result, it is efficient for improving the reactivity only in amorphous regions with negligible effect on the crystalline regions.

3.5.3 BIOLOGICAL PRE-TREATMENTS

This method uses microorganisms that can attack the cellulosic biomass components. Using wood cellulosic biomass, 4 categories of microorganisms: bacteria, soft rot fungi, brown rot fungi and white rot fungi, have been shown to be promising for lignocellulosic biomass pretreatment. Amongst these, the white rot fungi (WRF) stand out. WRF, remove 42% of lignin, 3% cellulose and 30% hemicellulose from Birchwood. WRF carry out co-oxidative delignification in the presence of a carbon source (cellulose or hemicellulose). They do this by attacking the phenolic residues by demethylation and ring cleavage. Cellulase free mutants of WRF have also been shown to carry out specific delignification.

This is a promising pre-treatment step but subject to modification using genetic and synthetic biology tools to accelerate the process. And or through partial physical and chemical pre-treatment.

3.5.4 ECONOMIES OF PRE-TREATMENT

The costs of chemical and physical pre-treatment regimens on wheat straw were analyzed (Table 3.6, Table 3.7). The chemical pre-treatment costs included the cost of chemicals while that of physical pre-treatments included the amount of energy consumed. Ball milling was the most promising physical pre-treatment method, with the highest DH and reducing sugar quantity. In addition, to being a clean method that is easy to operate. However, the long processing time of 8hrs make it impractical on an industrial scale. On the other hand, caustic soda and sulfite chemical pre-treatments were the ideal potential candidates for industrial scale processes based on the product yields, DH and costs.

The challenge however, remains the environmental threat posed by chemical processing due to the resultant effluent streams (chemical waste disposal), need for specialized corrosion resistant equipment, need for extensive washing and the associated water and energy costs.

Table 3.6: Cost analysis of Chemical Pre-Treatment Regimens

Type of pretreatment	Yield of sugar per kg of wheat straw, g/kg	Extent of hydrolysis after 8 h, g/l	Cost of chemical per kg of wheat straw, $/kg	Pretreatment cost based on sugar, $/kg
Caustic-AC	341.0	20.5	0.04	0.12
Sulfite-AC	252.9	16.0	0.32	1.26
Hypochlorite-AC	239.9	13.3	0.94	3.92
Peracetic acid	279.9	20.9	7.51	26.84
Butanol	72.4	4.7	2.86	39.49
Ethylene glycol-AC	241.2	18.2	11.25	46.53
Sulfuric acid	140.5	9.4	0.11	0.78
Standard	70.0	2.1	–	–

Table 3.7: Cost Analysis of Physical Pre-Treatment Regimens

Type of pretreatment	Yield of sugar per kg of wheat straw g/kg	Extent of hydrolysis after 8 h, g/l	Cost of energy consumed per kg of wheat straw $/kg	Pretreatment cost based on sugar, $/kg
Ball-milling, 8 h	255.98	9.1	1.48	5.82
Roller-milling ¼ h	160.0	6.48	2.24	14.0
Fitz-milling, fine	100.0	3.0	0.01	0.1
Extrusion, with pressure	64.29	2.5	0.01	0.16
γ-Irradiation 50 Mrad	207.02	5.6	0.1	0.48

3.6 INDUSTRIAL APPLICATIONS OF LIGNOCELLULOSIC BIOMASS/ CELLULOSE

As earlier mentioned in section 3.2, lignocellulosic biomass is made up of peripheral substances, lignin and polysaccharide (cellulose and hemicellulose) layers. Because of its rich multicomponent structure, there are diverse products that can be obtained from this biomass. Lignocellulosic biomass is useful in several industries. These include production of textiles, plastic, animal feed, paper, foils, explosives, varnishes, thickeners, glues, lumber, fuel, films and explosives (Chen, 2014; L.T.Fan , M.M Gharpuray, 2007), as well as the cosmetic industry.

The demand for biomass energy products such as pyrolysis bio-oil, ethanol, fuel and biodiesel are growing. Ethanol, which has a higher-octane value than gasoline is both an antiknock agent and an accelerant. Indeed, previous studies have shown that it could be mixed with a percentage of gasoline to avoid the need for engine overhaul and promote quicker market absorption. Raw materials for ethanol include starch (corn), sugar feedstock and lignocellulose biomass (straw). Presently, this industry is however dominated by starch and sugar feedstock due to the limitations around lignocellulose biomass that make the process economically non-feasible (Chen, 2014) . Biomass produced biodiesel industry is presently a matured technology. The biodiesel has the advantage of low sulphur content; it is environmentally friendly and biodegradable.

Lignocellulose is also used in aerospace and pharmaceutical in the form of cellulose acetate; a widely used fiber plastic. Different preparations of cellulose acetate can be made using diverse formulations and additives. It is used in aircrafts, building supplies, cosmetics, films, paints, photography, and printing and in vehicles. In addition, polyhydroxyalkanoate (PHA), a polyester used in synthesis of plastics such as polyethylene and polypropylene (PP) can be produced from lignocellulosic biomass raw materials. Plastics made from PHA are known as green plastics because they can be degraded completely to CO_2 and H_2O. These plastics are fast replacing the traditional hard plastics that are rather difficult to degrade. Other than its use in the plastic industry, PHA due to its biocompatibility is useful in the pharmaceutical industry, medical industry and in tissue engineering as an implant material. Also, as a stent in *in vivo* cell growth for organs such as heart valves, cardiovascular repair materials as well as graft material. Further, polylactic acid (PLA) and silk protein material have a role in aerospace, medicine, environmental protection and military.

Another important industry is the agricultural /food production industry. Lignocellulosic biomass can be used for feed formulation as well as fertilizer. Lignocellulose based fertilizer, is a safe and rich

alternative to chemical fertilizers, due to the rich nutrient content, trace elements and growth stimulating substances as well as organic matter that improves soil fertility and properties as well as eliminates soil pollution. This lignocellulosic biomass-based fertilizer also has the added advantage of a slow, steady and lasting fertilizer function.

Municipal wastes: paper and packaging materials, agricultural wastes, that include the associated plant material that is usually discarded when grain is harvested as well as food processing wastes such as fruit and potato peelings, coffee grounds and oats hulls and animal manure, are used for steam (fuel) generation through combustion and pyrolysis as well as for use as landfills. Cellulosic animal manure is also used as fertilizer as well as for methane gas production via anaerobic fermentation (L.T.Fan , M.M Gharpuray, 2007).

Further, a number of cellulose containing biomass such as sugarcane (Figure 3.17) are used for production of bio-based platform chemicals (AVA Biochem-1, Muttenz, Switzerland). According to a report by the US Department of Energy, platform chemicals can be defined as 12 building block chemicals that are made from sugars synthetically or by fermentation. These chemicals can then successively be converted to a number of high value bio-based chemicals and or materials. Bio-based platform chemicals offer a great potential for decarbonizing everyday products by permitting everything to be bio-based.

One such platform chemical is 5-HMF that is only made from biomass and not petro-based sources. 5-HMF is the basis of 175 downstream chemical substances. 5-HMF has numerous applications in chemical, plastics, food, biopolymers, resins, coatings, paints, varnishes, artificial fibers, special additives, pharma and automotive industries due its stability and functionality. 5-HMF has two functional reactive groups: hydroxyl and formyl group as well as a reactive furan ring structure. As a result, 5-HMF undergoes reduction, oxidation, esterification and many other reactions. It is these

possible reactions that contribute to 5-HMF's versatile derivatives and applications (Figure 3.18). Currently, the biopolymer market is the largest and most interesting for 5-HMF. This is due to increased demands in packaging, fibers and bottling which is driving the demand for bioplastics/biopolymers with an estimated market value of €1.2 billion in 2018. Following oxidation, 5-HMF yields another interesting platform chemical 2, 5-furandicarboxylic acid (FDCA). FDCA, is used as a basis for biopolymers and can substitute terephthalic acid in polyester, especially polyethylene terephthalate (PET). 2-hydroxypropane-1,2, 3-tricarboxylate (mumefural, MF) another platform chemical from 5-HMF, is also considered to be a potential anti-influenza chemical. Moreover, a valuable levulinic acid derivative: 5-aminolevulinic acid (5-ALA) is a commonly used photosensitizing drug in photodynamic therapy for skin cancer treatment. Additionally, 5-ALA is also an insecticide and a promising biodegradable herbicide. Further, 2, 5-bis (hydroxymethyl) furan, which is an important building block in the production of polymers and polyurethane, is formed following the reduction of 5-HMF formyl group.

Furthermore, the main cellulose hydrolysis product, glucose, has several uses. It is directly consumed for food in its refined form. Further, upon isomerization, it is used as a sweetener. Furthermore, glucose can be used as a substrate for alcohol fermentation, single cell protein production, biopolymer production, and enzyme production (Chen, 2014; L.T.Fan , M.M Gharpuray, 2007). Glucose, xylose and other sugars produced during hydrolysis of cellulose and hemicellulose can be used as substrates for fermentation in the production of chemicals, antibiotic drugs (such as penicillin), organic acids such as gluconic acid, L-amino acids, citric acid, pectin, acetone, butanol, ethanol, butanediol, and other liquid fuels and chemical raw materials. Phenylpropanoid from lignin can be used as raw material for organic and chemical industry and chemical products.

Indeed, the US Department of Energy proposed a shift into the more sustainable biomass based industrial structure. In this structure, 10% of chemical based products would be produced from non-food lignocellulosic biomass (agricultural stalks and urban fiber wastes) by 2020. This value is projected to rise to 50% in 2050 with 5% going to electricity production, 20% to transportation fuel and 25% to chemicals (Chen, 2014).

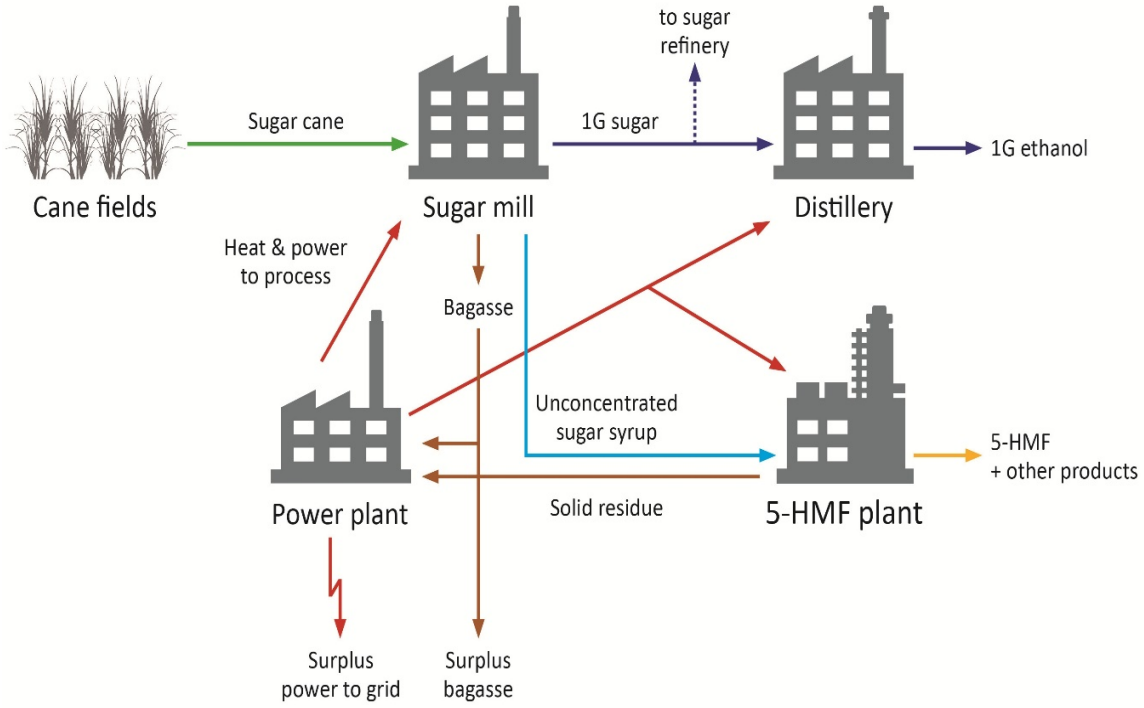


Figure 3.17 AVA Biochem's 5-HMF production via hydrothermal carbonization process. (Adapted from;https://communities.acs.org/community/science/sustainability/green-chemistry-nexus blog/blog/2014/07/17/bio-based-platform-chemicals-and-alternative-feedstocks).

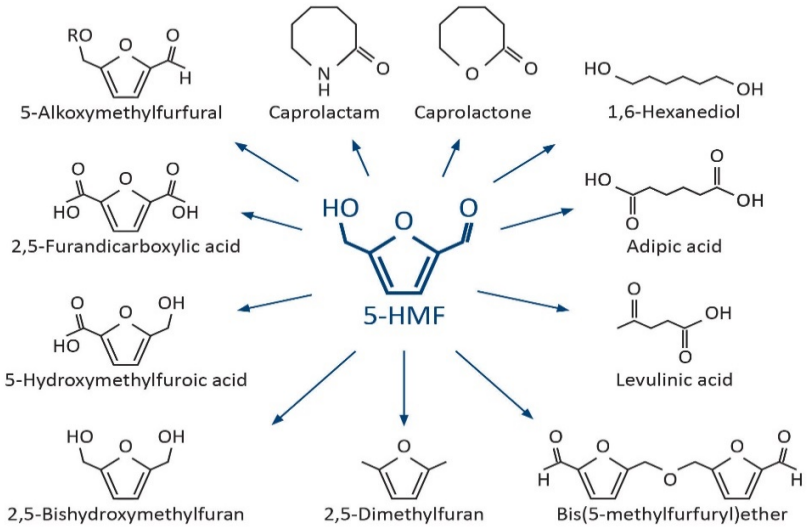


Figure 3.18: 5-HydroxymethylFurfural (5-HMF) potential as a platform chemical

3.6.1 Some Industrial Applications of Specific Cellulose Molecule

3.6.1.1 Pharmaceutical Applications of Sodium CarboxyMethylCellulose

Na-CMC, an anionic derivative is used in oral, topical and injectable formulations as well as ophthalmic pharmaceutical formulations. It is used as a binder or matrix former and yields soft granules with good compressibility and resultant tough tablets with good strength (Durig Kapish et al 2019).

Croscarmellose sodium (CCS) also serves as a superdisintegrant. CCS is mainly hydrophilic but is made fibrous and insoluble via crosslinking using salt of CMC. The final properties of CCS is determined by the source of cellulose used (Hiremath et al. 2019).

Further hydrofiber dressings are composed of soft, absorbent CMC fibres that interact with wound exudates (Aquacel®). This is used for heavy and moderately exuding wounds and wounds prone to bleeding because they are three times more absorbent than alginates and are used for abrasion, laceration following wound excision, pressure or leg ulcers and burns (Leveriza-Oh et al., 2012). Furthermore,

hydrogel in addition to wound dressing is also a useful scaffold for tissue engineering. It is used for encapsulating cells of nucleus pulposis and are a potential replacement for intervertebral disks degeneration. Blends of CMC, hydroxyapatite and chitosan are used for dental and bone regeneration (Aravamudhan et al. 2014).

Additionally, CMC is used for drug delivery in nasal and mucosal tissue. It has been successfully used in the GIT. Still, it is used for sustained nasal of apomorphine; a drug used to regulate motor response in Parkinson disease and has been shown to perform better than starch based delivery vehicle (Aravamudhan et al. 2014).In addition, Blends of CMC and sodium hyaluronate are also used to create a temporary loop stoma to facilitate stomy reversal (Beck et al 2019).

3.6.1.2 Application of Sodium carboxymethylcellulose in Food industry

CMC are thickeners that form water-soluble films compatible with several food ingredients. For example, it interacts with soy protein and renders it soluble at its PI. Additionally, due to their good water holding capacity, CMC solutions are good pseudoplastics. Moreover, they are used to make low calorie, yeast leavened wheat free baked products (Glicksman et al. 1972; Miller et al., 2008). Indeed, they control rheological properties of batter, shear thinning as well as facilitating the hydration of dry ingredients and control the moisture content of the finished product (Williams et al., 2003).

3.6.1.3 Application of cellulose in research

Na-CMC (0.1%) is used in fermentation media and is useful for minimizing mechanical damage caused by shear force that is generated by stirrer impeller (Stanbury et al. 2017).Moreover, microcrystalline cellulose (avicel) and cellulose derivatives (CMC) are used for determining cellulase activity. Cellulose polymers are extracted and purified, from plant polysaccharides, using chemical and enzymatic methods (Peria et al 2012).

3.7 CELLULOSE HYDROLYSIS

Lignocellulosic biomass utilization requires that cellulose and other component macromolecules like lignin and hemicellulose be hydrolyzed, into their constituent building blocks that could then be used for energy / fuel production and organic chemical production.Lignocellulose biomass can be transformed using chemical, physical or biological methods. Using chemical and physical /mechanical methods, cellulose is degraded into low molecular weight hydrocarbons, flammable gases or liquids via fractionation, pyrolysis, alkali and acid hydrolysis as well as redox degradation.

However, these methods are not economically viable because of their high-energy intake, accompanying low yields and enormous environmental pollution. Consequently, most of them have been relegated as auxiliary transformation methods. On the other hand, cellulose bioconversion has been adopted. This involves the biotransformation of cellulose to its constituent glucose monomers using enzymes (Chen, 2014a).

Enzyme hydrolysis has been widely studied in the last decade. Nevertheless, up to now there are no plants for commercial enzymatic hydrolysis. On the other hand, chemical (acid) hydrolysis has been carried out industrially, since 1913, with the first plant being the pine mills waste plant in South Carolina. Acid hydrolysis was also conducted in Germany using dilute sulphuric acid (H_2SO_4) and dilute hydrochloric acid (HCl) during world war II (L.T.Fan , M.M Gharpuray, 2007). Acid hydrolysis is fast with short reaction time. In addition, the use of acid eliminates the need for pre-treatments that are expensive.

Despite the extensive use of acid hydrolysis, enzymatic hydrolysis is still, the preferred cellulose hydrolysis method. This is because although slower, enzyme hydrolysis is specific therefore, it does not produce undesirable by-products. In addition, there is no need for neutralization of the product. Besides,

capital and operating costs are significantly lower since enzyme hydrolysis occurs at mild temperatures. Further, unlike the case with acids /alkalis, specialized corrosion-proof equipment is not required for enzyme hydrolysis. As an added advantage, cellulose-degrading enzymes can be engineered thanks to advances in genetic engineering. Enzymes' modification drives the realization of optimal cellulose hydrolysis enzymes and subsequent potential commercialization of cellulose enzyme hydrolysis.

A group of enzymes known as cellulases carry out the hydrolysis of cellulose. Cellulase hydrolysis is very specific and targets only precise (β, 1-4) linkages as well as glucose enzymatic monomers in specific configurations in the polysaccharide, unlike acid hydrolysis which is promiscuous and can attack both α or β 1-4, linkages in the substrate molecule. In addition, previous studies have shown that enzymes are more efficient in cellulose hydrolysis in comparison to acids. From these studies, up to 100,000X more acid was needed for the same degree of enzymatic cellulose hydrolysis at the experimental conditions.

The observed disparity increases further at the molecular level where $10^{\wedge 8}$ acid molecules are required to attain the same degree of hydrolysis (DH) achieved by a single enzyme molecule (Reese *et al.,* 1956).

All these notwithstanding, speedy cellulose hydrolysis is still a challenge since no enzyme can compete with the complete sulphuric acid (H_2SO_4) hydrolysis of native cellulose to glucose which is completed in one hour (Saeman *et al.,* 1944). This documented rapid H_2SO_4 cellulose hydrolysis is a combined action of concentrated acid at low temperatures and dilute acid at high temperatures as well as splitting and swelling. However, these levels of extensive hydrolysis of cellulose have not been demonstrated for free cellulase enzymes. On the other hand, however, highly active whole microorganisms utilize up to 50% of the cellulose (0.5%w/v) within 3 to 4 days in shake flasks. This hypothesizes that a close association between the whole microorganism hyphae and cellulose in the growing culture is responsible for the higher degree of hydrolysis (DH).

This has been explained by the accumulation of products in the free enzyme system in the absence of the organism leading to product inhibition that hampers reaction rates due to the absence of some cellulase enzymes from the reaction cocktail (Reese *et al*., 1952). Another challenge that presents with acid hydrolysis of cellulose as shown in cotton is that there is no striking increase in susceptibility to enzyme hydrolysis with an accompanied decrease in tensile strength and degree of polymerization (DP). Worth mentioning is that sulphuric acid cellulose pre-treatments leads to partial esterification of the cellulose unlike phosphoric acid. This makes phosphoric acid pre-treatment the preferred treatment method.

3.7.1 CHEMICAL HYDROLYSIS

3.7.1.1 Acid Hydrolysis

This involves the use of both concentrated as well as dilute acids in a single phase and multi-phase reactions respectively. Although acids are also used as cellulose pre-treatment agents (section 3.5.2.1 above), they are primarily used as catalysts for cellulose hydrolysis. Cellulose hydrolysis using concentrated acids takes place via the formation of cellulose acid complexes (Figure 3.19) following swelling of crystalline cellulose in acidic medium.

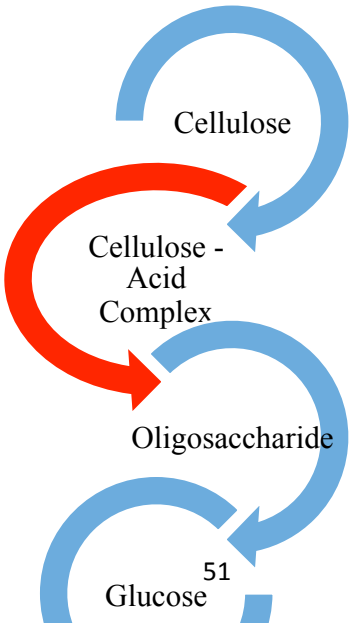


Figure 3.19: Acid hydrolysis cycle of cellulose molecule

In acid hydrolysis, different complexes are formed with different acids. It is important to note that the concentration of acid used has a direct influence on the overall course and kinetics of the reaction. In dilute hot acid hydrolysis, the following product sequence is seen. Firstly, hydrocellulose is formed followed by soluble polysaccharides and eventually simple sugars. However, 40% HCl for instance, degrades cellulose to oligosaccharides at ambient temperature. At elevated temperatures, the oligosaccharides formed are subsequently converted to glucose using a first order mechanism. Several challenges plaguing acid hydrolysis especially the inhibition of subsequent fermentation processes due to side reactions have necessitated the use of enzyme hydrolysis technologies.

3.7.1.2 Alkali Hydrolysis /Peeling/Zipping

Cellulosic β, 1-4 glycosidic bonds are mainly resistant against alkali degradation at temperatures below 170°C. As a result, most alkali hydrolysis takes place between 20-200°C with distinct degradation product profile differences. Depending on both the concentration and alkali agent used (Knil *et al.*, 2002). However, boiling in the presence of caustic soda leads to peeling of short chains from the reducing ends of the cellulose molecule via the elimination of β-alkoxycarbonyl to release D-glucoisosaccharinic acid. Subsequently, a newly protonated end that can undergo further degradation is generated. Alkali earth metals improved the yields during zipping in comparison to alkali metal ions at low alkali concentrations (0.02M). Calcium improved the yields of D-Glucoisosaccharinic acid by speeding its yield from 4-deoxy-D-glycero-2, 3 hexodiulose.

Unchecked peeling leads to complete dissolution of the cellulose molecule. In processes involving alkali cellulose hydrolysis therefore, cellulose is stabilized and the process terminated using competing reactions. There are up to 20 acid stabilizing units in peeled hydrocellulose. This stopping

reaction principle has been used to also modify cellulose end groups to produce alkali resistant molecules.

In addition, alkali scission is the hydrolysis of cellulose in alkali solutions heated at temperatures >170°C leading to decrease in weight and DP. At high temperatures, stopping reactions are more effective. In addition, Alkali hydrolysis of cellulose under aerobic conditions is used in the aging of wood pulp by molecular oxygen that is used in making viscose rayon. Several hypothesis have been put forward on the mechanism of aerobic, alkali cellulose degradation. Generally, oxidized cellulose with carbonyl groups at positions other than the end groups (C2, C3 and C6) are highly susceptible to hydrolysis, hence undergo complete hydrolysis at mild conditions.

3.7.2 ENZYMATIC HYDROLYSIS

Enzymatic cellulose hydrolysis is a heterogeneous reaction in function of the cellulase enzymes solubility and insolubility of cellulose. It is also a collective breakdown of the β-1, 4 glycosidic bonds and H-bonds. This makes cellulose degradation seem like a rather simple process. However, this is far from true given the inert and rigid structure of cellulose lent by the parallel arrangement of cellulose fibrils.

This makes enzymatic cellulose hydrolysis 100X less efficient in comparison to starch hydrolysis; a major hurdle to adequate cellulose utilization.

Improving cellulose hydrolysis efficiency is therefore a subject of continuous study. With some key areas of interest such as:

i. Understanding the molecular structure and function of cellulose degrading enzymes

ii. Understanding the catalytic mechanism of crystalline cellulose hydrolysis

iii. Understanding the synergistic effect of cellulases and auxiliary enzymes involved in cellulose hydrolysis.

Enzymatic cellulose hydrolysis is divided in three stages:

i. Accessibility: This is the contact between the cellulase enzymes and cellulose reactant molecules

ii. Cellulase adsorption and diffusion

iii. Self-assembly and synergy of exoglucanase/endoglucanase together with glucosidase in degrading the crystalline and amorphous cellulose regions.

Cellulose hydrolysis reaction speeds are governed by:

i. The cellulosic structural features; whereby crystalline regions of natural cellulose fibers are highly recalcitrant to degradation. Further, these regions have limited enzyme binding sites making cellulosic hydrolysis an extremely slow process

ii. Enzyme strain level that determines how strongly the enzyme is adsorbed to the substrate.

Cellulose degrading organisms (bacteria, algae, fungi and actinomycetes), protozoa, and animals use cellulose for growth, reproduction and bioconversion. These microorganisms are present at various stages of lignocellulose degradation. They include

1. Poor cellulose degraders that rapidly appear on plant leaves. They utilize the secretions on leaves and easily degradable starch and pectin, insect dung and small animals. These microbes grow in the soil and on leaves before they fall off. They are used due to their fast growth that allows for strong proliferation and colonization of the biomass before other microbes. However, they never appear at the end of the

cellulose degradation process following their gradual disappearance due to depletion of nutrients when other microorganisms colonize the biomass.

2. The second category of microbes, grow on cellulose and hemicellulose and are present during the early and middle stages of lignocellulose biomass degradation. Some of these include *Chaetomium sp*. and *Deuteromycota sp*.
3. A third category of microbes consists of lignin–degrading *Basidiomycota* and actinomycetes that have low population density, high stability and strong ability to utilize complex organic compounds. As a result, these microbes propagate and reproduce steadily in environments composed largely of lignin, humus and chitin.
4. A fourth category of microbes that can be detected throughout the lignocellulosic biomass degradation process consist of *Cladosporium sp*., *Trichoclerma sp*., *Penicillium sp*., *Aspergillus sp*., as well as some species of yeast, bacteria and actinomycetes

In addition to differences in microbes' profiles during the lignocellulose biomass degradation, we also have lignocellulose compositional profiles differences as follows:

i. Cellulose degrading microbes including fungi, actinomycetes, bacteria, protozoa and animals. Fungal cellulose decomposing ability is very high.
ii. Hemicellulose degrading microbes: These are more numerous than cellulose degrading microbes. Fungi dominate at the early stage after which actinomycetes later dominate.
iii. Lignin degrading microbes: These are mainly Basidiomycota and *Aphyllophorales sp*., as well as fungi such *Fomes sp*. *Polyporus sp*. and *Polysticus sp.*

Furthermore, there exist differing cellulose-hydrolyzing enzymes (cellulases) systems in nature. These cellulase systems differ based on the host organisms, enzyme architectural modalities, as well as the absence or presence of oxygen. The available cellulase systems in nature include the free cellulases,

cellulosome (which are a cluster of cellulases and their ancillary enzymes), aerobic cellulases, anaerobic cellulases as well as fungal and bacterial cellulases.

3.7.2.1 Mesophilic aerobic enzymatic hydrolysis versus anaerobic enzymatic activity

In the case of aerobic fungi and bacteria, a cocktail of cellulases; endoglucanases, exoglucanases, β-glucosidases and ancillary enzymes; lytic polymonooxygenases are secreted and act synergistically to attack cellulose. The best studied of these enzymes are the glycoside hydrolases of *Trichoderma reesei* (Doi and Kosugi, 2004; Lynd *et al.,* 2002; Warren *et al.,* 1996).

These microbes are important in the degradation of unlignified cellulosic biomass such as twigs and cotton on the surface of neutral and alkaline soils. Aerobes' cellulases are intracellular enzymes with low activity and adhere to the cell wall and cell membrane (Chen, 2014b). Anaerobic fungi and bacteria have been shown to be more efficient cellulose degraders. They possess an inherent enzyme complex known as cellulosome that is involved in lignocellulose hydrolysis. (Bayer *et al.,* 1985).

3.7.2.2 Bacterial cellulase enzyme systems versus fungal cellulase enzyme systems

Bacterial cellulases are mainly intracellular and mostly adsorb to the bacterial cell wall with only a few secreting extracellularly. *Actinomycetes* have very low cellulase activities. However, very few studies have been conducted on cellulases from *Actinomycetes.* In contrast, fungi whose cellulases have been well studied have been characterized as high cellulases producers. In addition, fungal cellulases are mainly extracellular.

3.8 CELLULOSE DEGRADING ENZYME TYPES

Efficient depolymerisation of cellulose is a concerted synergistic action of several cellulases/ glycosyl hydrolases (Table 3.8). These enzymes, can be grouped based on substrate specificity into three groups; endoglucanases; **EGs (E.C 3.2.1.4)**, exoglucanases; **EXG**/Cellobiohydrolase; **CBH**; non-reducing end acting (**E.C 3.2.1.91)** and reducing end acting **(E.C 3.2.1.74 /E.C 3.2.1.176)** and β-

Glucosidases; **BGs (E.C 3.2.1.21).** In addition, ancillary enzymes such as hemicellulases, (hemicellulose hydrolases) and ligninases, (oxidoreductases) which act on these polymers that have close interaction with cellulose hence making cellulose more accessible. In addition, we have a class of auxiliary enzymes known as Lytic polysaccharide monooxygenases (LPMOs) that catalyse the oxidative cleavage of polysaccharides (Horn *et al.,* 2012).

In nature, cellulolytic systems constitute of varied endoglucanases and exoglucanases with variable preference for the different cellulose forms (crystalline and or amorphous cellulose). This preference for varied cellulose substrates is determined by the varied enzyme domains (cellulose binding domain; CBD/CBM) structural differences.

Table 3.8: Glycoside Hydrolase Families with Documented Cellulase Activity

Glycosyl hydrolase family	Cellulase activity Type	Catalytic Mechanism	Former group name	3D structure	Catalytic Base	Catalytic Proton Donor	Clan
5	CEL/CBH I	***retaining***	Cel A	$(\beta / \alpha)_8$	Glu	Glu	GH-A
6	CEL/CBH -I	***Inverting***	Cel B	$(\beta / \alpha)_8$	Asp	Asp	
7	CEL/CBH- II	***retaining***		β-jelly roll	Glu	Glu	GH-B
8	CEL	***Inverting***	Cel D	$(\alpha / \alpha)_6$	Asp	Glu	GH-M
9	CEL/CBH- I	***Inverting***	Cel E	$(\alpha / \alpha)_6$	Asp	Glu	
12	CEL	***retaining***	Cel H	β-jelly roll	Glu	Glu	GH-C
44	CEL	***retaining***	Cel J	$(\beta / \alpha)_8$	Glu	Glu	
45	CEL	***Inverting***	Cel K	6-stranded β-barrel	Asp	Asp	

48	CEL/CBH-I/CBH-II	***Inverting***	Cel L	$(\alpha / \alpha)_6$	Not known	Glu	GH-M
51	CEL	***retaining***		$(\beta / \alpha)_8$	Glu	Glu	GH-A
74	CEL/CBH	***Inverting***		7-fold β-propeller	Asp	Asp	
124	CEL	***Inverting***		lysozyme fold			

CEL- Endo-β-1, 4-glucanase 3.2.1.4

CBH-I Non –reducing end cellulose 1, 4 –β-cellobiosidase 3.2.1.91

CBH-II Reducing end-acting cellobiohydrolase EC 3.2.1.176

CBH- Oligoxyloglucan reducing end-specific cellobiohydrolase 3.2.1.150

(Carbohydrate active enzyme database () ; Lombard *et al.*, 2013)

3.8.1 Molecular Structure and Function of Cellulase Enzymes

Previously, cellulases were isolated from *Trichoderma reesei* and partially hydrolyzed using the proteolytic enzyme papain. These studies showed that cellulases have two independent domains: a catalytic domain (CD) and a cellulose-binding domain (CBD). CD is the enzymatic domain that only hydrolyses soluble cellulose with a weak absorption capacity. However, CD does not act on the core of the larger (>56kDa) insoluble cellulose molecule. On the other hand, the cellulose binding domains (CBD) is located in either the N or the C terminal of the enzyme and adsorbs to cellulose surface as well as its smaller fragments (10kDa) (Figure 3.21 and 3.22)

The CBD and CD domains are linked by a highly glycosylated linker region with a tadpole –like structure. Bacterial linker region is rich in Pro and Thr amino acid residues whereas fungal linker region is rich in Glycine (Gly), Serine (Ser), and threonine (Thr). Bacterial and fungal CD and CBD are 135° and 180° apart respectively. Fungal cellulases have one restriction enzyme (R.E) site but bacterial cellulases have two R.E sites that cut out the linker and CBD (Chen, 2014b).

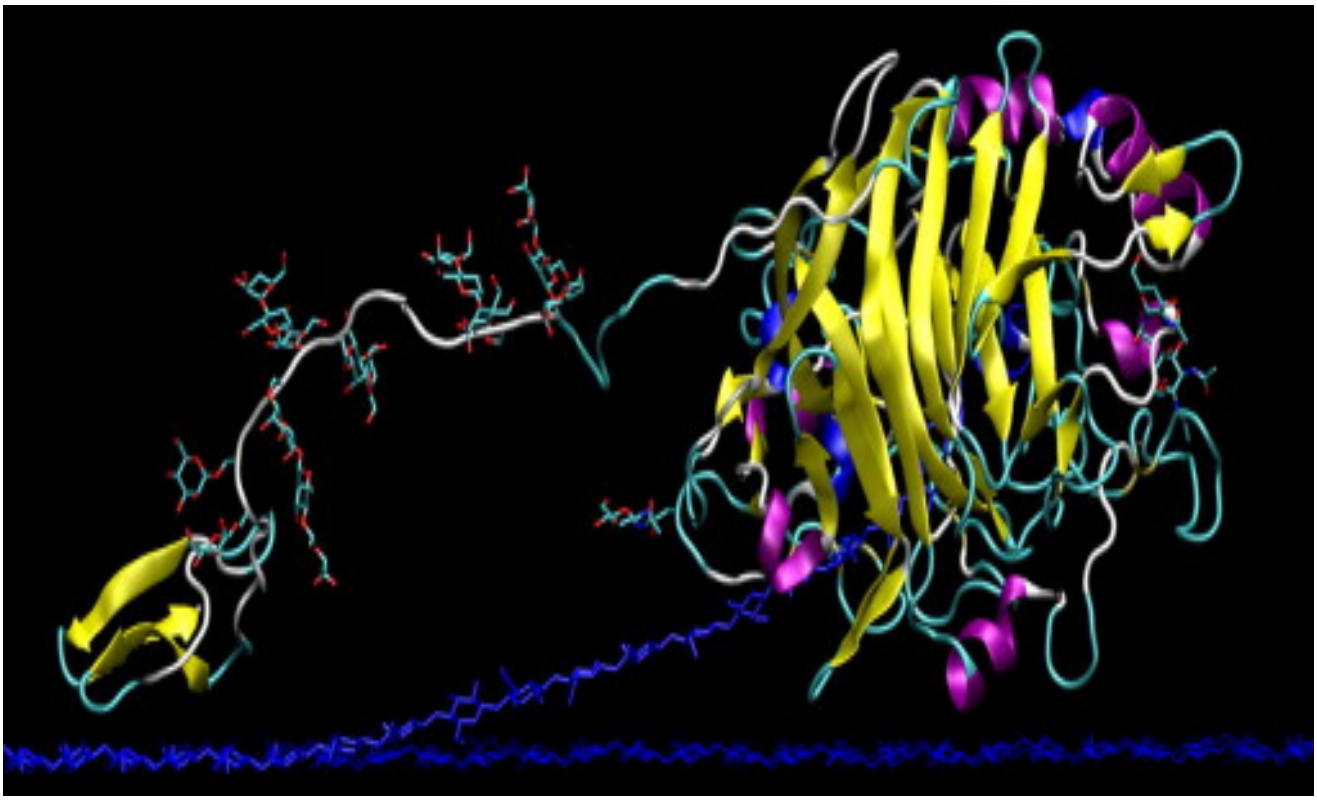


Figure 3.20: 3D Quaternary structure of endoglucanase with cellulose in its cellulose-binding domain; CBD (In –silico modelling of CBM3aBPGH9).

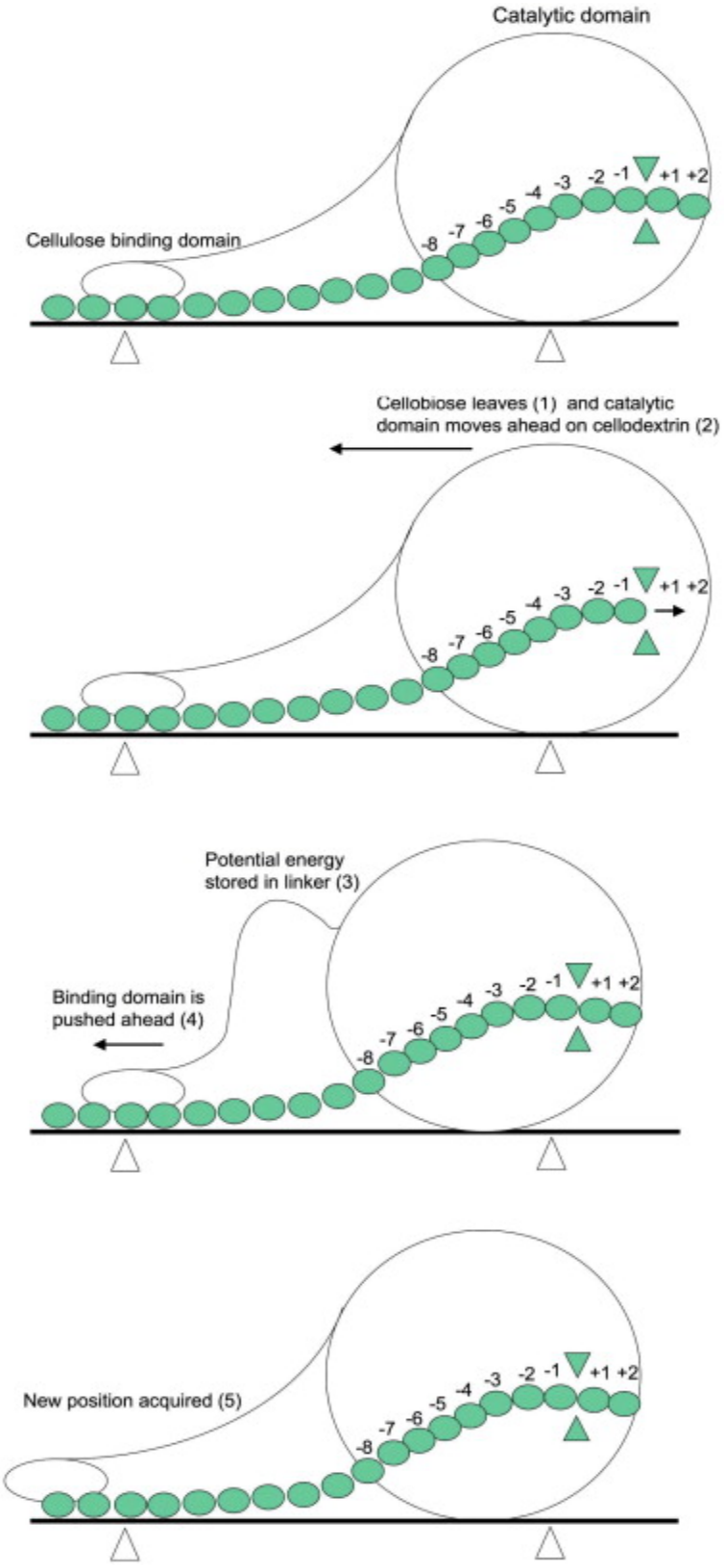


Figure 3.21: Mechanism of action of cellulose binding domain (CBD)/cellulose binding module (CBM)

3.8.2 Endoglucanase

β-1, 4 endoglucanase are a class of processive and non-processive enzymes that cleave the internal cellulosic chains to release shorter polysaccharides chains. They are also known as Cx enzymes or CMCases, because the substrate CMC is used to assay for their activity. They account for 20-30% of total cellulase preparations.

Unlike exoglucanases, endoglucanases do not have a high specificity and can hydrolyse water soluble cellulose, non-amorphous sections of phosphate pre-treated cellulose, CMC and amorphous cellulose by irregularly hydrolyzing β-1, 4 glycosidic bonds to yield glucose, cellobiose, cellotriose and other oligosaccharides with available ends for CBH (Chen, 2014b).

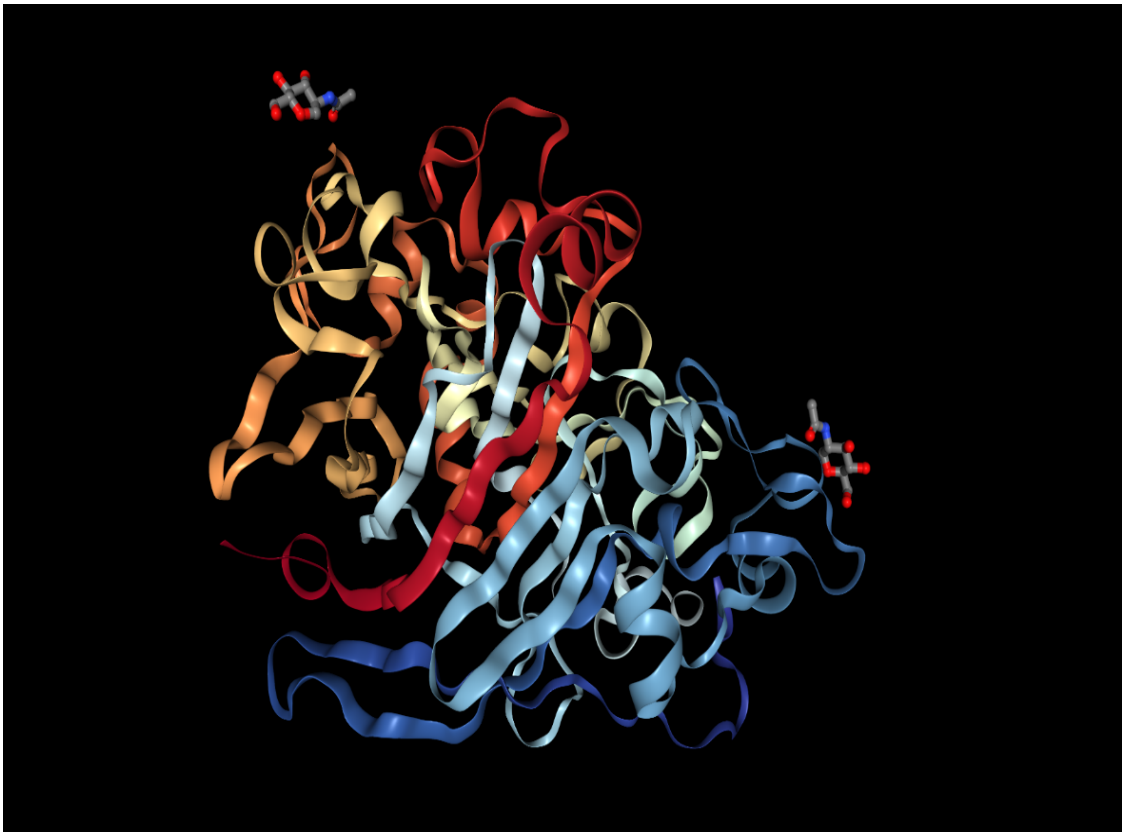


Figure 3.22: Native structure of endoglucanase I from *Fusarium oxysporum* (PDB, 3OVW). Adapted from (AS Rose *et al*., 2018)

3.8.3 Exoglucanase / Cellobiohydrolase (CBH)

Exoglucanases also known as C1/CBHI act on the reducing (E.C 3.2.1.74/E.C 3.2.1.176) or non-reducing (CBHII) (3.2.1.91) end of the short polysaccharide chains released by endoglucanases to release cellobiose and other short chain polysaccharides. Processive β-1, 4 exoglucanases (Figure 3.24) are the most abundant cellulase component in nature and form up to 80% of commercial cellulase mixtures.

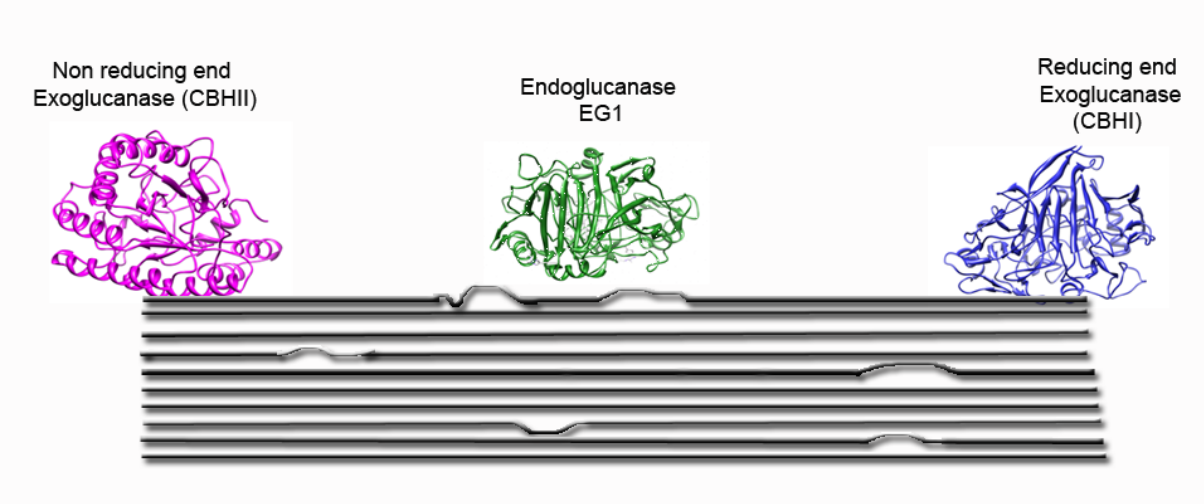


Figure 3.23: Action points of CBHI (PDB, 1CB2), CBHII (PDB, 3CBH) and Endoglucanase (PDB,1EG1) along a cellulose molecule strand (Adapted from (Quiroz & Folch, 2013)

Exoglucanases have a robust substrate specificity. Their degradation capacity is directly proportional to the oligosaccharide chain lengths. Consequently, there is a decrease in the degradation capacity with a decrease in oligosaccharide chain length. These processive cellobiohydrolases are important for the hydrolysis of crystalline cellulose and only hydrolyse natural cellulose with little to no effect on cellulosic substituents such as Na-CMC. However, processive cellobiohydrolases (Figure 3.24) are inherently slow. As a result, reducing sugar is almost undetectable when CBH breaks down cellulose on their own.

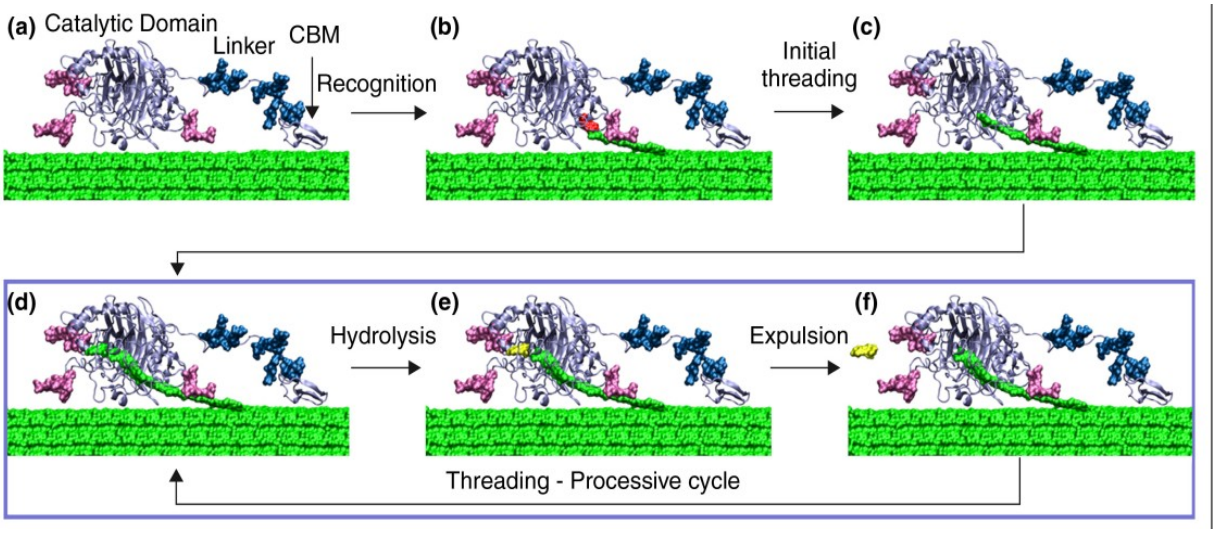


Figure 3.24: The activity of Cel7A from *T.reesei* on its substrate. Cel7A, has a small carbohydrate binding domain (CBD); 36 AAs, a long flexible linker with O-glycan (dark blue) and a large catalytic domain (CD) with N-linked glycan (pink) that can thread a single chain of cellulose into the catalytic tunnel of 50 Å.

*(*a*) Cel7A binding to cellulose, (*b*) recognition of a reducing end of a cellulose chain, (*c*) initial threading of the cellulose chain into the catalytic tunnel, (*d*) threading and formation of a catalytically active complex, (*e*) hydrolysis in a processive cycle and (*f*) product expulsion and threading of another cellobiose (shown in yellow in *e* and *f*). Adapted from (Quiroz & Folch, 2013)

Consequently, synergistic activity of CBH with endoglucanase ensures effective cellulose degradation. Furthermore, alternatives exists where pretreated cellulose that is relatively amorphous is treated with less processive enzymes (Horn *et al.,* 2012).

3.8.4 β-Glucosidase (BG)

β- Glucosidases are also known as cellobiase, or more accurately as β-glucimerase indicative of their ability to degrade all forms of β-glucose dimers. Whereas the naming does not differentiate between aryl glucosidase and dimerase, only BG that act on short chain β 1-4 oligoglucosides especially glucose dimers such as cellobiose are considered cellulolytic. This make up to 1 % of the total cellulase. These cellulolytic enzymes hydrolyse cellobiose released by exoglucanases and glucose residues from the non-reducing end of small cellulose dextrins (upto cellohexaose) to glucose. They however do not act on insoluble crystalline cellulose.

Unlike CBH that is specific for the β 1-4 glycosidic bond, BG has a low substrate specificity. They can therefore, hydrolyse 1, 1; 1, 2; 1, 4 and 1, 6 β bonds. In addition, their rate of hydrolysis increases with decrease in the size of the substrate with cellobiose having the highest hydrolysis rate. It therefore reduces the negative feedback inhibition on cellulase caused by cellobiose accumulation. Unlike CBH that works using the inversion mechanism, BG acts rapidly on dimers and trimers with the product maintaining the β-configuration. BG is also inhibited by gluconolactone (L.T.Fan , M.M Gharpuray, 2007). Moreover, with the development in synthetic biology, molecular biology and genetic engineering BG can be engineered to catalyse transglycosylation (Feng et al., 2005)

3.8.5 Ancillary Enzymes

In addition to the conventional cellulase enzymes, there exist other ancillary enzymes involved in cellulose degradation. These include lytic polysaccharide monooxygenase (LPMOs), cellobiose dehydrogenase (CDH), Cellobiose quinone oxidoreductase (CBQ), Phosphorylase as well as auxiliary enzymes that hydrolyse other components parts of lignocellulosic biomass.

3.8.5.1 Lytic Polysaccharide MonoOxygenases (LPMOs)

As earlier mentioned in section 3.2 above lignocellulosic biomass is heterogeneous constituting of three main polymer components cellulose (40-50%), hemicellulose (20-40%) and lignin (20-30%) as well as proteins , soluble sugars , lipids , pectin and minerals in relatively minute quantities (Pauly *et al.,* 2008 ; Horn *et al.,* 2012). The heterogeneity of the lignocellulosic materials impedes efficient enzymatic hydrolysis whereas the tight packing of the polysaccharide chains into crystals make the individual components inaccessible and resistant to enzyme action. Previous studies have thus hinted at additional players other than cellulases that would make the polysaccharide substrates more accessible (Reese *et al*., 1956).

Lytic polymonooxygenases (LPMOs) have been identified especially in fungi as one of the additional players in cellulose hydrolysis. These enzymes are categorized in the auxiliary activity

(AA) family in the CAZy database. They are non-hydrolytic proteins that carry out oxidative degradation of cellulose in the presence of an electron donor and divalent metal ions as cofactors (Vaaje-Kolstad *et al*., 2010). LPMOs have flat substrate binding surfaces hence potentiates the action of classical hydrolytic cellulase enzymes by acting on the surface of insoluble crystalline substrate and disrupting the polymer packing therefore making cellulose accessible for hydrolysis by cellulase (glycosyl hydrolase) enzymes.

LPMOs synergy with cellulases and chitinases, was first demonstrated in CBM33 (Reese *et al*., 1956; Horn *et al*., 2012). Using chitin, CBP21, formerly classified in the family CBM33 and currently in the family AA10, CBP21, oxidatively cleaves glycosidic bonds in chitin and produces a normal non-reducing chain end and a reducing chain end comprising a C1-oxidised sugar called aldonic acid (Vaaje-Kolstad *et al.,* 2010; Horn *et al*., 2012). Previous research, showed that the activity of CBP21 was dependent on divalent metals, boosted by adding electron donors such as ascorbic acid and inhibited by metal ions chelators like ethylene diamine tetraacetic acid (EDTA). Using isotope labelling, it was shown that oxidative cleavage by LPMOs involves molecular oxygen (Vaaje-Kolstad *et al.,* 2010; Horn *et al.,* 2012).

CelS2, an AA10 from *Streptomyces coelicolor* is a lytic cellulose monooxygenase (C1-hydroxylating) that cleaves cellulose and produces aldonic acid. CelS2 is active without metal ions probably due to high affinity to cellulose. CelS2 activity however is affected by several divalent metal ions. This was demonstrated by inhibition on addition of EDTA. This inhibition was demonstrated by restoration of activity on addition of divalent metal ions. Previous studies showed that both CBP21 and CBM33/AA10 could use several divalent metal ions. In fact, recent studies confirmed that these enzymes are copper dependent monooxygenases (Horn *et al.,* 2012).

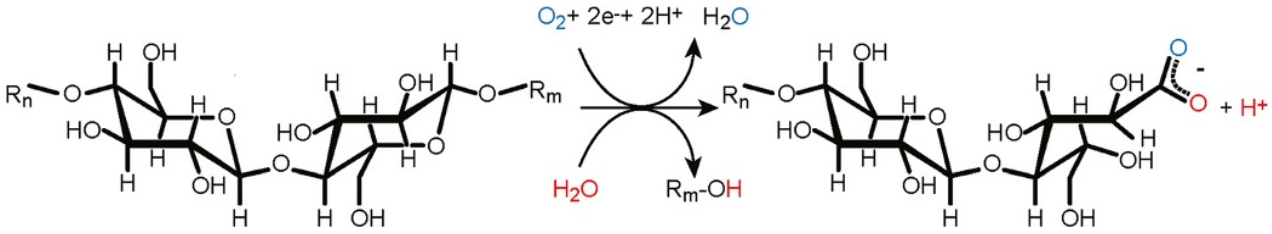


Figure 3.25: Summary of the oxidative cleavage of cellulose. In the case of cleavage by CelS2, a CBM33, and PcGH61D .The only oxidized sugars observed are aldonic acids, as indicated in this figure. Other members of the GH61 family seem to generate additional oxidized species, with oxidation at C4 or C6. Adapted from (Chen, 2014b)

Another enzyme previously classified in the GH61 family (reclassified as AA9 family) from *Thermoascus aurantiacus* (TaGH61A), *Phanerochaete chrysosporium* (PcGH61D) and *Neurospora crassa,* has been shown to be similar to CBM33, a copper dependent lytic polysaccharide monooxygenases. It catalyzes copper dependent oxidative breakdown of cellulose in the presence of electron donors such as gallic acid. These enzymes oxidize the C1, C4 and C6 carbon (Chi, 2003; Li, 2005; Martinez, 2002; Sun &Cheng, 2002; Sundaramoorthy *et al.,* 2005). CBM33 and PcGH61D show C1 oxidation whereas TaGH61A and *N. crassa* GH61s show additional products. Further, *N. Crassa* has been shown to contain C1 and C4 oxidizing GH61 (Chen, 2014a).

3.8.5.2 Cellobiose Dehydrogenase (CDH) / Cellobiose Oxidase (CBO) (E.C. 1.1.99.18)

Cellobiose dehydrogenases (CBO/CDH) are heme flavoprotein that play an important role in oxidative degradation of cellulose in filamentous fungi. They utilize flavin adenine dinucleotide (FAD) and heme as prosthetic groups. CDH has a wide substrate range that includes fiber tetra saccharide, cello oligosaccharide with five degree of polymerization, cellotriose, cellulose and lactose BUT not cellobiose and glucose. They oxidize these substrates to their corresponding lactones. CDH use cellobiose as an electron donor to reduce several compounds.

3.8.5.3 Cellobiose Quinone Oxidoreductase (CBQ)

CBQ has similar physicochemical properties as CDH but only uses flavoprotein as a prosthetic group. In addition, no oxidation has been shown for CBQ.

3.8.5.4 Phosphorylase

Some microorganisms use phosphorylases on fiber dextrin, which yield phosphorylated cellulose, and cellobiose that can then be effectively metabolized.

3.9 CELLULOSOME (Supramolecular Assembly)

Cellulosome are sophisticated nano machines fundamental in the efficient anaerobic degradation of cellulose and associated plant cell wall component polysaccharides (Bayer *et al.*, 2004; Doi and Kosugi 2004; Fontes and Gilbert, 2010). The multi-enzyme assembly of the cellulosome complex is because of close-fit binding between cohesin (Coh) modules, localized on a non-catalytic scaffoldin and dockerin (Doc) modules, attached on the C-terminus of the catalytic units (enzymes). A cellulosome can be made of upto 100 catalytic units in complex structures via secondary Coh-Doc interactions.

3.9.1 Basic Structure of a cellulosome

An archetypal cellulosome is made up of several component parts /blocks /modules (Figure 3.26)

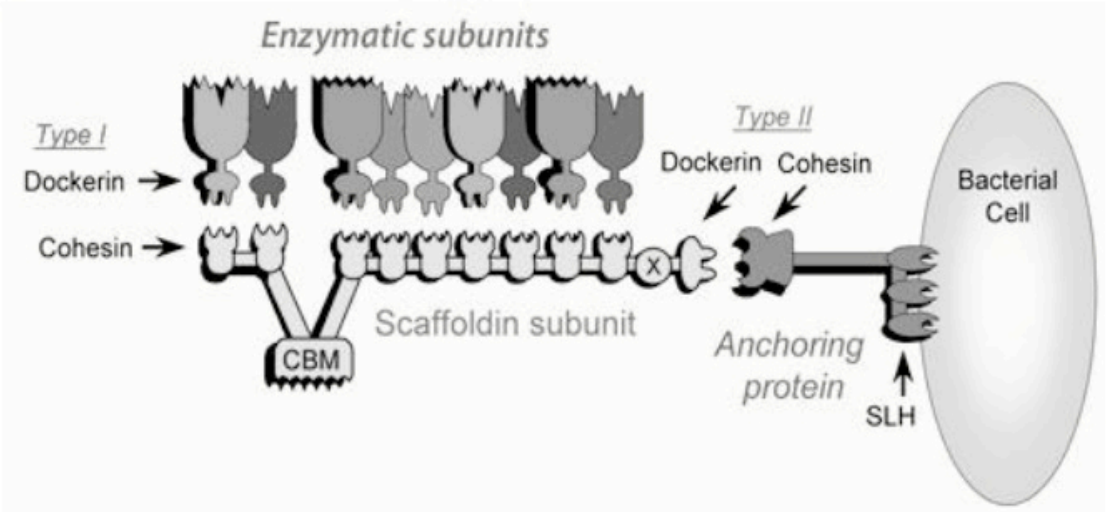


Figure 3.26: Basic structure of bacterial cellulosome from *Clostridium thermocellum* (Karmakar & Ray, 2011)

3.9.1.1 Scaffoldin

At the center of the non-catalytic core of the cellulosome is the scaffoldin that plays several functions:

i. The anchoring of the complex to the surface of the microorganism (anchoring scaffoldin)

ii. The incorporation of enzyme catalytic units onto the complex via Coh-Doc interactions (primary scaffoldin)

iii. The incorporation of additional components on to the complex via Coh-Doc interactions (adaptor scaffoldin)

iv. Targeting of the microorganism and cellulosome to the cellulose substrate via the CBM

3.9.1.2 Cohesin (Coh)

Based on sequence homology, cohesins have been classified into three groups: type I, type II and type III. Cellulosome assembly is based on the tenacious hydrophobic interaction between the cohesin and their complementary dockerin partners. This Coh-Doc interaction shows two levels of specificity. It is type and species specific, meaning type I, II and III cohesin randomly bind only to Type I, II and III Doc respectively with the same affinity. Moreover, there is no interaction between Coh and Doc from different species. However, some cohesins defy this general rule and interact with dockerins across species (Sakka *et al*., 2010).

Coh-Doc binding is a calcium (Ca^{2+}) dependent, high affinity ($Ka \geq 10^{-9}$M upto 10^{-12} M) interaction responsible for the complex cellulosomal topology (Fierobe *et al*., 1999; Fierobe *et a*l., 2001; Schaeffer *et al*., 2002). Calcium ions are critical to the interaction because of the calcium-binding motif of the dockerin domain whose sequence resembles the EF-hand motif of eukaryotic calcium-binding proteins.

(http://wws.weizmann.ac.il/Biological_Chemistry/scientist/Bayer/cohesin-dockerin)

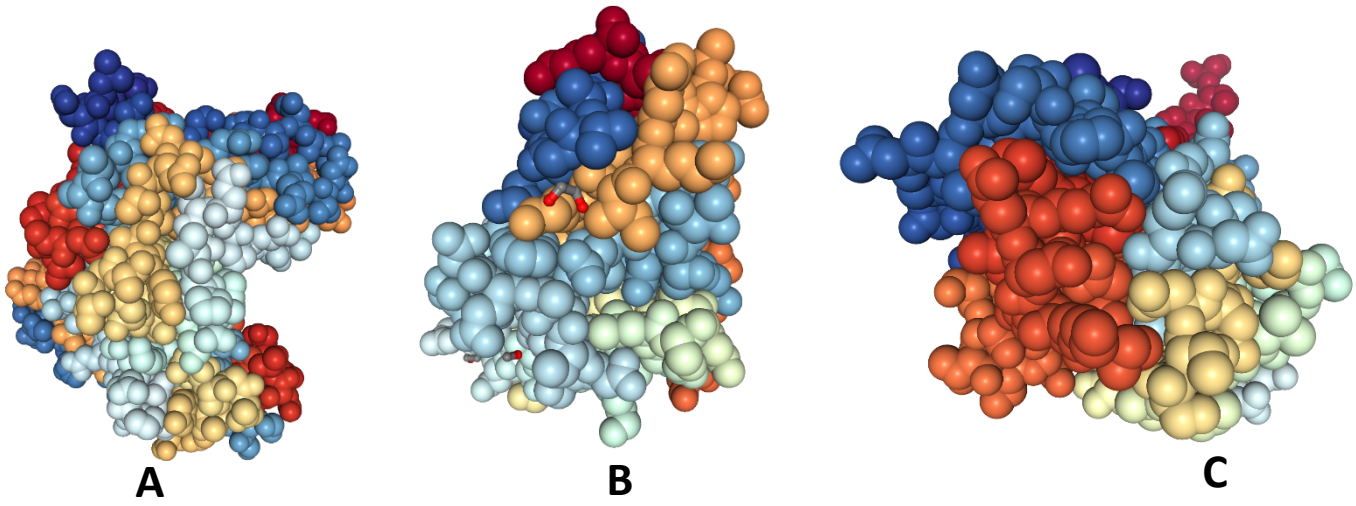


Figure 3.27: A; CipA cohesin, Type I Cohesin*;* 1AOH; from *Clostridium thermocellum.* **B;** Cohesin A11, Type II Cohesin; 1TYJ; From *Bacteriodes cellulosolvens*. **C;** Cohesin from ScaE-cell surface anchoring scaffold; type III Cohesin; 2ZF9; from *Ruminococcus flavefeciens*

3.9.1.3 Dockerin (Doc)

Just like with the cohesin, there are three types of dockerin: type I, type II and type III based on sequence homology. Doc has an internal two-fold symmetry, with tandem repeats of two calcium-binding loop-helix motifs of about 24 amino acids (distinct from EF-hand Ca^{2+}binding motifs). Progressive deletion and alanine scanning techniques on type I Doc from *Clostridium thermocellum* revealed that only one of the repeated motifs is critical for high affinity Coh-Doc interaction.

In addition, previous studies have shown that the almost perfect sequence and structural symmetry of the repeated motif had no significant role in the Coh-Doc interaction. Similar conclusions were reached using systematic mutation of highly conserved residues in the calcium-binding loop. It was established that mutations in one of the Ca^{2+} binding loops did not disrupt the cohesin recognition and subsequent Coh-Doc interactions. However, single mutations on both loops reduced dockerin affinity for its complementary cohesin significantly. These findings are compatible with type I Coh-Doc heterodimer that shows that dockerin binds to their cohesin partner using the first or second repeated motif. Using spectroscopy, it was observed that the binding of

Ca^{2+} to dockerin induced folding into its secondary structure that subsequently exposes the hydrophobic patches on the dockerin surface. (http://www.ebi.ac.uk/interpro/entry/IPR002105).

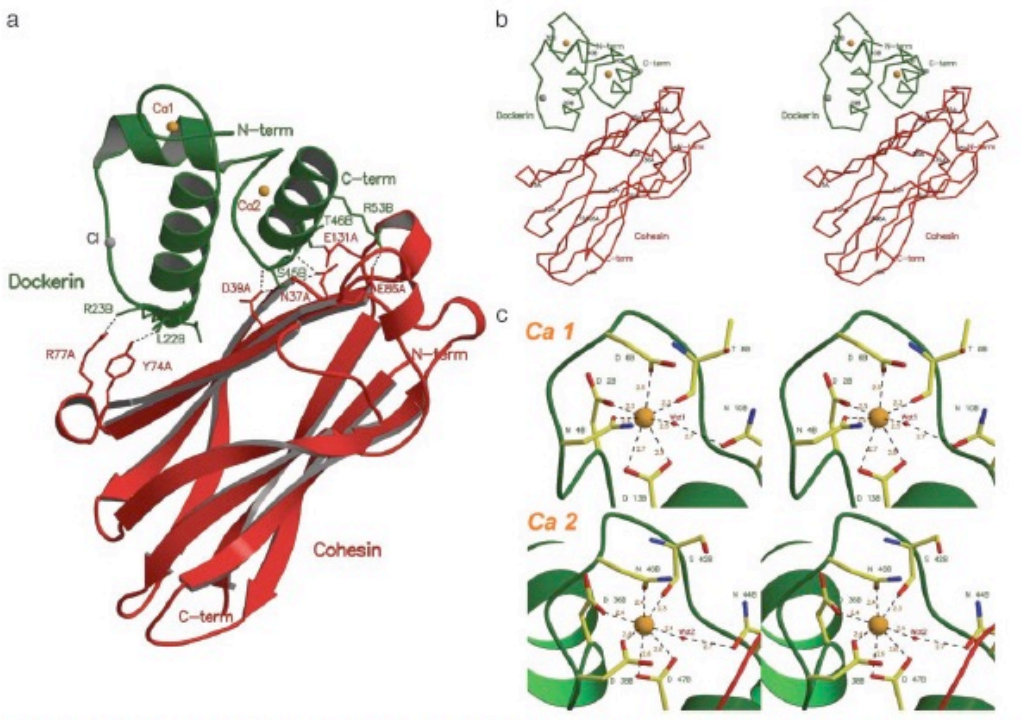


Figure 3.28*:* Structure of type 1 cohesin-Dockerin complex **(a)** The complex is formed between a cohesin 2 molecule (red) and a Ca $^{2+}$ bound dockerin (green). The residues involved in domain contacts are shown as stick models. The two Ca^{2+} binding sites of the dockerin domain are represented in orange spheres. **(b)** Representation of Coh-Doc structure with every 10^{th} residue labeled. **(c)** Ca^{2+} coordination in the dockerin domain. The Ca^{2+} bound residues are shown as stick models with green labels.

3.9.1.4 Catalytic domain /Catalytic Units

Cellulosomes contain at least 30 enzymes. The majority of these enzymes are endoglucanases (EC 3.2.1.4). However, there are also some xylanases (EC 3.2.1.8), beta-glucosidases (EC 3.2.1.21) and endo-beta-1, 3-1, 4-glucanases (EC 3.2.1.73). A majority of these enzymes contain a highly conserved type I dockerin domain of about 65 to 70 residues, which is generally (but not always) located at the C-terminus. (http://www.ebi.ac.uk/interpro/entry/IPR002105).

Depending on which enzymes are bound to the scaffoldin, there is the potential to make cellulosomes with many different compositions within a single microorganism (Doi and Kosugi, 2004).

3.9.1.5 Carbohydrate binding domain (CBM/CBD)

The cellulosomal enzymes rely on the family 3a CBM of the scaffolding subunit for collective binding to crystalline cellulose. (http://wws.weizmann.ac.il/BiologicalChemistry/scientist/Bayer/enzymes)

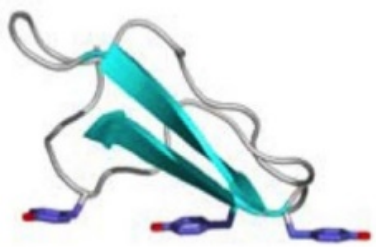


Figure 3.29: Structure of the carbohydrate binding module specific to beta -1, 4-endoglucanase (Type A) (Boraston *et al.*, 2004)

3.9.1.6 Linker

The linker region is a malleable region comprising of 5-700 amino acids (AAs) joining two adjacent domains. Bacterial linker region is highly glycosylated with a tadpole like structure. Bacterial and fungal linker regions have been shown to be rich in Pro, Thr and Gly, Ser, Thr respectively. It is important to note that the length of this region is very precise such that too long or too short linker regions affect hydrolysis efficiency of the enzyme (Diego *et al*., 2016 &Yazana *et al*., 2013).

Long linkers have specific amino acid motif repeats whereas medium linkers have repeats of Pro, Thr and Ser AAs repeats. Some linker residues such as Ser and Thr have been shown to act as binding sites for sugars (Xu *et al*., 2003). The subsequent glycosylation is an important defense

system against protease attack. In addition, it also contributes to protein-protein recognition and protein carbohydrate binding (Schwarz *et al*., 2001). Pro residues in the linker on the other hand have been shown to bind to nearing CBM modules and effect quaternary structure changes that help in positioning the cellulosome to its substrate and enhancing hydrolysis (Hammel *et al*., 2005 & Yaniv *et al*., 2012)

3.9.2 THE SIMPLE CELLULOSOMAL STRUCTURE OF *Clostridium thermocellum*

C. thermocellum is of great interest as a candidate for consolidated bioprocessing. This is due to its thermophilic properties that enable cellulose hydrolysis at elevated temperatures. As well as its ability to be co-cultured with pentose fermenting thermophiles such as *Thermoanaerobacter saccharolyticum* (Bayer, *et al.,* 2004; Schwarz, 2001). In addition, it can be genetically manipulated, and has thus been extensively studied biochemically and using proteomics (Bayer, *et al.,* 1985; Choi & Ljungdahl, 1996; Gold & Martin, 2007; Mayer, *et al.*, 1987; Raman, *et al.,* 2009; Tyurin, *et al*., 2004; Zverlov, *et al*., 2005).

The scaffolding (CipA) is a 200kDa protein containing nine type I cohesin molecules (Figure 3.26) on which the catalytic modules are anchored through type I cohesin-dockerin interactions (Kruus, *et al*., 1995). This scaffolding also contains a CBM/CBD module that binds the cellulosome complex to its cellulose substrate. CipA also contains a C-terminal type 2 dockerin domain, which binds non-covalently to type 2 cohesin domains found on cell wall anchor proteins OlpB and SdbA (Leibovitz & Beguin, 1996; Lemaire, *et al*., 1995). These anchor proteins remain associated with the cell surface by means of the S-layer homologous (SLH) domains (Adams, *et al.,* 2006).

3.9.3 COMPLEX CELLULOSOMAL STRUCTURES OF OTHER ANAEROBIC BACTERIA

Unlike *Clostridium thermocellum,* that has a relatively simple cellulosomal structure; other anaerobes such as *Acetivibrio cellulolyticus and Ruminococcus flavefaciens* contain more complex cellulosomal structures. *Ruminococcus flavefaciens* produce a cellulosome comprising three divergent cohesin-dockerin pairs mediating the association of scaffolds: scaffoldin A (ScaA),

scaffoldin B (ScaB), scaffoldinX (ScaX) in addition to multiple enzymes *(*Rincon, *et al., 2003)*. *Acetivibrio cellulolyticus* (Figure 3.30) possesses a more complex cellulosomal structure; scaffoldin C (ScaC), plays a role similar to OlpB in *C. thermocellum* and anchors the cellulosome to the cell surface by means of the surface layer homology (SLH) domains. A specific dockerin-cohesin interaction mediates the attachment of adaptor scaffold ScaB to ScaC. A second dockerin-cohesin pair mediates attachment of ScaA to ScaB. ScaA contains a CBM as well as seven cohesins; with different specificity from its ScaB and ScaC counterparts, which are capable of binding the dockerin domains located on the multiple hydrolytic enzymes. Interestingly, ScaA itself bears catalytic activity characteristic of a family 9A glycohydrolase (Xu, *et al.,* 2003; Weicozorek, 2013)

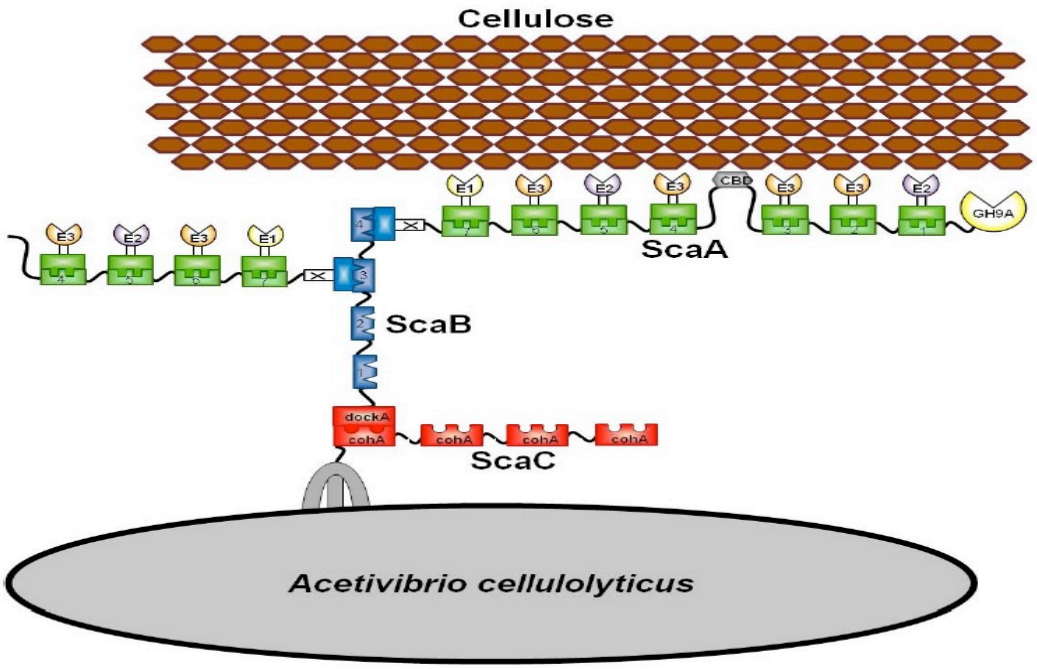


Figure 3.30: Cellulosomal structure of *Acetivibrio cellulolyticus* (Weiczorek, 2013)**. ScaA;** Primary scaffold**, ScaB;** Secondary scaffold**, ScaC;** Adapter scaffold.

3.10 STRATEGIES FOR ENGINEERING SYNTHETIC CELLULOSOMES

The great architectural variability among native cellulosomes, (See Figures 3.26 & 3.30 above), coupled with the need for organisms capable of carrying out consolidated bioprocessing; intracellularly producing, secreting or displaying to the surface enzyme complexes capable of controlling the flux of the metabolic pathway and degrading polymers, has led to extensive research

into designing and expressing of recombinant cellulosomes (Anderson, *et al*., 2011; Dueber, *et al*.,2009; Lee, *et al*., 2011; Tsai, *et al*., 2009 Fierobe, *et al*., 2002; Fierobe, *et al*., 2001; Fierobe, *et al*., 2005; Mingardon, *et al.,* 2007; Perret, *et al*., 2004; Wen, *et al*., 2010)

This clustering of enzymes in a multienzyme complex has the advantage of substrate channeling and synergy between the associated catalytic modules (Bayer, *et al*., 2004; Conrado, *et al*., 2008). In the case of cellulose hydrolysis, substrate channeling is exemplified by longer chain polysaccharides produced by non-processive cellulases becoming the substrate for processive cellulases, which then produce short chain cellodextrins and cellobiose as the primary products (Schwarz, 2001). Synergy occurs when such enzymes are localized in close proximity to one another (Bayer, *et al.,* 2004).

From a biotechnological perspective, optimizing the spatial organization of a pathway's enzymes through co-localization on protein scaffolds or fusions has the potential to greatly enhance the channeling of hydrolysis intermediates to enzymes that will use them as substrates in further reactions (Conrado, *et al.,* 2008).

Research groups, have therefore sought to design recombinant cellulosomes and investigate the effects of enzyme composition and s**patial organization on the hydrolytic activity of the resulting complexes**. Cohesins and dockerins with different specificities originating from different species have been used as building blocks to **engineer custom-designed recombinant cellulosomes or cellulosome-inspired complexes with defined architectures and enzyme compositions** (Cho, *et al,* 2004; Fierobe, *et al*., 2005; Mingardon, *et al.,* 2007; Wen, *et al*., 2010).

Synergistic activity of cellulosomic enzymes on crystalline cellulose was demonstrated by the production of a truncated version of the *Clostridium cellulovorans* scaffold protein CbpA (Mini-CbpA) as well as three enzymes, EngE, EngH, and EngS, and their subsequent assembly *in vitro* into artificial cellulosomes containing combinations of two enzymes (Murashima, Kosugi, *et al*., 2002).

This observed synergy was affected by both the type, and stoichiometric ratios of enzyme used. Optimal combinations of enzymes were determined based on increased activity on crystalline cellulose. However, the effects of enzyme positioning within the complex could not be deduced due to the non-specific binding of each enzyme to any of the two cohesins present on the scaffold. The multiple enzyme activities required to degrade crystalline cellulose and the possibility to optimize the stoichiometry and the relative positioning of the enzymes within the complex prompted the construction of recombinant cellulosomes with precise enzymatic compositions.

The construction of artificial scaffold proteins containing cohesins of different specificities has been used to precisely dictate the enzyme composition of designer cellulosomes (Fierobe, *et al.,* 2002; Fierobe, *et al.,* 2001; Fierobe, *et al.*, 2005; Mingardon, *et al*., 2007). Research describing the construction and utility of scaffold chimeras has shown that the fusion of cohesins derived from the cellulosomes of *C. thermocellum* and *C. cellulolyticum* allowed complexes with two enzymes to be generated (Fierobe, *et al*., 2001).

3.11 INDUSTRIAL APPLICATION OF CELLULASES AND OTHER LIGNOCELLULOSIC BIOMASS DEGRADING ANCILLARY ENZYMES

Market reports on world enzyme demands (2017) have shown that several key factors affect the huge enzyme demand. These include economical advances; increased per capita income in developing countries has led to increased consumer related industrial applications. Further, a study by Freedonia (January, 2018) projected a 4% annual growth in industrial enzymes demand with up to $5Bn in 2021 with accompanying gains in personal incomes. Coherent market insights showed that the textile industry was the largest market for cellulases in 2017. Another market report in 2018 showed that the textile industry, animal feed industry, food, beverage industry, and biofuels industry were some of the major areas of application. Global cellulase (CAS-9012-54-8) market research report of 2018 showed that Asia Pacific is the largest consumer of cellulase with a revenue market share of 32.84% in 2016. A further breakdown revealed that 29.7% of this cellulase demand was

allocated to animal feed, 26.3% to food and beverage and 13.8% to textile in 2016. Moreover, future forecasts show a rising cellulase demand that is growing at a compound annual rate (CAGR) of 5.5% between 2018-2025 and will reach up to $2300M by 2025. Dupont and Novozymes are the key cellulase enzymes producers for the global industrial market.

There is a huge potential for application of cellulases in several industries. These include; Health care/Medicine, Textile Industry, Pulp /Paper industry, Laundry/Detergent Industry Food (Beverages, Baking), Agriculture, Biofuel, Bio polymer Industry and Wine industry

3.11.1 MEDICINE INDUSTRY

Humans do not produce cellulase enzymes. However, cellulase enzyme blends are important for collective digestion of fibrous cellulose rich food material such as vegetables, fruits, legumes, cereals, bran, nuts, seeds, soy, diary, sprouts and herbs, carbohydrates and gluten. Some commercial consumable cellulase enzyme blends are VeganZyme®, Polyenzyme Plus®, Digestin and P-A-L Plus enzymes®

3.11.1.1 Direct Uses for cellulase in medicine

Patients with impaired digestive function and decreased gastric motility sometimes have formation of phytobezoars (gastric concretion formed by seeds, vegetable fibers, fruit skins and starch granules and fat globules trapped in the GIT). Major phytobezoars are extracted via surgery. However, minor phytobezoars are treatable using cellulases. To date, only fungal cellulases have been used for this purpose. In addition, cellulases are used for degrading pathogens' cell walls and treating microbial infections (Menendez et al., 2015).

3.11.1.2 Indirect uses for cellulase in medicine

Chitosan is important in medicine. It is used for surgical sutures, bone rebuilding, production of artificial skin, anticoagulant, antibacterial agent, hemostatic dressings as well as antidiabetic agents (together with metals) and anticancer agents, production of biopharmaceuticals,

hypocholesterolemic effectors, cosmetic elaboration and encapsulation. Cellulases/chitinases/lysozyme cocktails play a role in chitosan degradation (Menendez et al., 2015).

3.11.2 TEXTILE INDUSTRY

Cellulase (Table 3.8) is the third largest group of enzymes used in the textile industry (Paridah *et al.*, 2016). They are used for wet processing, yarn processing, fabric processing as well as garment processing. The major points of cellulase application in the textile industry are bio stoning of denim fabrics, biopolishing of textile fibers, softening of garments as well as removal of excess dye from the fabrics. *Trichoderma reesei* fungal cellulases are the most applied in textiles as well as cellulases from actinomycetes (*Thermobifida and Streptomyces*), *Pseudomonas* and *Sphingomonas* cellulases are used in textile dye degradation and decolorization.

3.11.2.1 Bio finishing and Bio polishing

Piling on cellulosic fabric surfaces gives the fabric an unattractive look. Bio polishing removes these short protruding fibers on the fabric surface. This helps improve the color, feel and look of the final fabric product by giving the product a glossy, smooth appearance in addition to color brightness hydrophilicity and moisture absorbance. Acidic cellulases are efficient for this process. Bio finishing, helps to remove the fluff from a repeatedly washed fabric to restore the feel as well as stains.

3.11.2.2 Biostoning

Denim with a worn out look is in high demand in the market. Traditionally, biostoning was achieved using pumice stones. However, there are numerous challenges presented by the use of pumice stones such as machine deterioration due to tear effects, the need to load large quantities of pumice stones as well as the subsequent offloading of the stones after the process. In addition, pumice biostoning had reduced efficiency and excessive back staining. Cellulase biostoning is an alternative to pumice stonewashing. Cellulase enzymes hydrolyse the β 1-4 glycosidic bonds and break the small cellulosic fibers from the denim to give off simple sugars. This leads to the release of the indigo dye that is trapped inside the cellulose fibers therefore giving the fabric a faded look.

Some cellulases identified as efficient biostoning agents include, acidic endoglucanase II from *Trichoderma reesei,* neutral cellulases *of Humicola insolens.*

Cellulases overcome the challenges of pumice stone washing. They increase productivity, are environmentally friendly, have less wear and tear on the machines and require short treatment times. However, like pumice stone washing, cellulases also have back staining with less back staining seen for acidic cellulases. Advanced stoning processes use enzymes cocktails of cellulases, laccases and amylases. It is important however to note that the by-products of dye breakdown by enzymes are potential carcinogens that need to be neutralized.

3.11.2.3 Bio scouring

This involves the removal of non-cellulose materials on the cotton surface. This is done in combination with other enzymes such as pectinases and amylases.

3.11.2.4 Defibrillation of lyocell

Lyocell is a biodegradable fabric made out of treated wood pulp. It is used in clothes as well as cars. Cellulases are applied on lyocell to remove fuzz and stop pill formation thereby giving the lyocell a soft and improved appearance.

3.11.2.5 Wool scouring and biocarbonisation

Bio carbonization is a promising alternative to sulphuric acid carbonization. It involves elimination of cellulosic wastes from cotton/cotton blend fabrics.

3.11.3 PAPER INDUSTRY

3.11.3.1 Pulping

Cellulases are used in the paper industry for pulping, deinking, bleaching, waste bioremediation and fiber enhancement. Traditional mechanical pulping is highly energy consuming. Cellulases, on the other hand are energy saving (up to 40% energy saving) and eco-friendly. Enzyme cocktails of endoglucanase I, II and hemicellulases are used for modification of the coarse pulp to improve fiber properties whilst reducing energy costs and environmental burden.

3.11.3.2 Deinking

Traditional chemical deinking methods used large volumes of chemicals. Consequently, the process is extremely expensive and environmentally toxic. Bio deinking eliminates the need of alkali and prevents yellowing of the paper. Cellulase/xylanase or cellulase/hemicellulase cocktail is used for deinking waste paper. Enzymatic deinking gives a bright, clean look paper with reduced environmental pollution.

3.11.3.3 Biomodification and Biocharacterisation

Hemicellulases/cellulases enzyme cocktails modify beatability, drainage and running of the paper pulp.

3.11.4 LAUNDRY AND DETERGENT INDUSTRY

Detergent industry is one of the biggest market for enzymes, making up to 30% of the total sales. The laundry detergent industry is valued at 133.3Bn in 2016. Cellulases such as Carezyme® and Celluzyme ® are used in detergent blends to make fiber modifications that can improve color brightness, softness as well as remove particulate soil matter.

Alkali cellulases are the most suitable for detergent blends due to their ability to penetrate the interfibrillar spaces and remove soil and dirt particles. In addition, they also help with pilling and improving the appearance of the fabric. In addition, detergent blend formulations containing enzyme cocktails to improve fabric care and make stain removal efficient have been produced. These include cellulases/proteases/lipases as well as amylase as is seen in the case of SaniZyme®; a bacteriostatic enzyme detergent used for the removal of blood, protein, mucous , fats ,lipids and carbohydrates from surgical equipment.

3.11.5 AGRICULTURE

Cellulases in agriculture are used in enhancing crop growth and controlling plant diseases. Enzyme cocktails of cellulase/hemicellulases and pectinases are used in agriculture. Some fungal cellulases have been shown to possess the ability to degrade plant pathogen cell walls. In addition, cellulase producers participate in rhizosphere soil decomposition, bioavailability of plant nutrient,

plant pathogen control, root colonization and nutrient and yield improvements. The exact mechanism for these however has to be elucidated.

3.11.5.1 Animal feed

Cellulases improve the digestibility of cereal foods and elevate the nutritive content of forage. Cellulases/hemicellulases cocktails are utilized as feed additives to increase milk output, animal body weight gain as well as the removal of anti-nutritional factors. Finally, the partial hydrolysis of fodder by cellulase aides with emulsification of food in animal gastrointestinal tract (GIT) leading to improved nutrient absorption.

3.11.6 FOOD INDUSTRY

Fruit contains floating polysaccharides such as hemicellulose, cellulose, lignin, pectin, starch as well as metals, proteins and tannins that lead to cloudiness in fruit and vegetable juices. This cloudiness, leads to diminished consumer acceptability. Cellulases are used in the food industry for a number of uses including juice (fruit and vegetable) clarification, to concentrate purees, to modify sensory properties (flavor, texture and aroma properties) of the fruit, to reduce viscosity of nectars to market acceptable levels, to extract carotenoids, olive oil as well as improving the quality of baked foods.

Cellulases/hemicellulases/pectinase cocktail such as Rapidase pomaliq ® from *T.reesei* and *A. niger* as well as cocktails from *Paenibacillus/Bacillus* are also used for juice clarification. Pectinase/cellulase cocktails are used for modification of sensory characteristics of juices. In addition, these enzymes also breakdown grape fruit peels releasing sugars, which are used in food industry and other industries. Moreover, cellulases are used for extracting phenolic substances from grape pomace. Cellulases also enhance the malaxing effect, phenolic content, antioxidant capacity and general quality of olive oil.

In the bakery industry, cellulase/amylase/protease/xylanase cocktails are used to increase loaf volume, produce softer breadcrumb and improve bread quality.

Similarly, cellulase/amylase/lipase/phospholipase/hemicellulases cocktails are used for dough conditioning leading to enhanced flavor, increased shelf life and volume upon baking. Natural pigments such as carotenoids are in high demand in the food industry for use as food colorants. This is because they are less toxic and available in a wide color spectrum. Cellulase are important for the extraction of these natural pigments from color rich peels from oranges, tomatoes, and carrot and sweet potatoes.

3.11.7 BIOFUEL

Cellulases convert cellulosic biomass (sugarcane bagasse, switchgrass, *Lantana camara*, *Prosopsis juliflora*, sawdust, rice straw corncob, forest waste) and are used for biofuel production.

3.11.8 BIOCHEMICAL INDUSTRY /WASTE REMEDIATION.

Cellulosic biomass (lignocellulose material and lignocellulosic waste) can be converted to numerous organic solvents and chemicals from lignocellulosic biomass (Section 3.6).

3.11.9 WINE INDUSTRY

Glucanases play an important role in alcoholic (wine and beer) beverage production. These enzymes improve the quality and yields of the fermented products. Cellulases are added during the primary fermentation step in order to hydrolyse the glucan, reduce the viscosity of the wort and improve filterability. Pectinase/glucanases/hemicellulases improve the color extraction, must clarification, filtration, quality, and stability of the wine. β-Glucosidase modifies the glycosylated precursors and improve the wine aroma (Kuhad, Gupta, & Singh, 2011). Cytolase 219®, a cocktail of pectinase, xylanase and cellulase showed up to 35% increased must extraction, 70% in must filtration , up to 70% decrease in must viscosity and 40% energy savings with 50-120 minutes pressing time reduction and enhancement in wine product stability. For beer making, endoglucanase

I and II and Exoglucanase II of *T. reesei* have been shown to reduce wort viscosity and degree of polymerization (Oksanen et *al*; Kuhad *et al*., 2011).

Table 3.9: Some Commercial Cellulases used in the textile industry

NAME	Product	Origin (Microorganism)	Time (mins)	Temperature °C	pH	Conc.	Application	REFERENCE	Company
Cellulase	Celluzyme 0.7T						Depilling	Novo Nordisk	Novo Nordisk
							Color brightening		
							Stone washing		
							Fabric softening		
							Removing particulate soil		
Cellulase	Bactosol ®			55-65°C	4.5-5.5		Stone washing,	US005,460,966A	Sandoz chemical Inc.,Basel, Switzerland
							Overdying		
Cellulase				50°C	9		Bioscouring		
Cellulase			100-120	50-60°C	4.0-5.0		Water washing jean textile	CN103835141 (A)	
Cellulase			360-	105-115			Modifying	CN103710764 (B)	

		540			and degumming jute fibres	
Cellulase		40-60	95	alkaline	Enhancing oxydol stability in oxygen bleaching	
1-20% Cellulase	Textile use enzyme preparation				Degrading fibre impurities e.g pectins, hemicellulase	CN102851989(A)
0.1-0.3 Cellulase	Textile fabric cleaning agent				Cleaning fabric	CN102851127(A)
1 Cellulase type				7	Enzymic treatment of cellulose fibres	US2012180229(A1)
Cellulase					Promoting pigment dyeing	CN102505532(A)
Cellulase					Burring readymade	CN102433774 (A)

					clothes		
Cellulase					Recycilng of cellulosic fibres wastes	CN10356474 (A)	
Cellulase					Dyeing	RU2471906 (C2)	
Cellulase					Pretreating natural fibres and textiles	CN101624583 (A)	
Cellulase	10	25		0.01g/l	Enhancing fibre wettability and absorbence	US6066494(A)	
Cellulase EG11					Treating and finishing fibre	US5874293 (A)	
Cellulase CBH11					Biofinishing cotton	US5858767(A)	
Pectinase		50	9		Scouring	US6,551,358 B2	Novozymes A/S , Bagsvaerd (DK)

Cellulase	*Trichoderma longibrachiatum RL-P37*	Stone washing with minimised strength loss , texture modifications, Fabric feel improvements and removal of dead /Immature cotton	US005989899A	Genencor International Inc.
Pectinase			US005912407A	**Novo Nordisk Biochem North America, Inc., Franklinton, NC.**

RESULTS AND DISCUSSION

CHAPTER 4

Résumé du publication 1

Les lacs (Nakuru, Bogoria, Baringo) situées dans la région volcanique de la rift valley au Kenya sont des écosystèmes intéressants de part leurs caractéristiques physicochimiques particulières liées à la présence de solfatares, sources hydrothermales (températures comprises entre 60° et 100°C), une salinité importante (2 à 6 g/l) et un pH élevé (9-10). A cela s'ajoute la présence d'une matière organique importante due à la décomposition des végétaux apportés par les cours d'eau d'alimentation et la présence d'algues (cyanobactéries). Ces caractéristiques suggèrent la présence d'un microbiote adapté à ces conditions particulières de type alcalino-thermophile.

A partir d'isolats de souches bactériennes collectées sur les rives du lac, nous avons entrepris de caractériser sur des extraits bruts de culture la présence d'activités cellulolytiques. Ces extraits sont effectivement capables de libérer des sucres réducteurs à partir de la biomasse cellulosique telle que le coton ou le papier filtre. Ces extraits bruts ont ensuite été caractérisés du point de vue cinétique en fonction de différents effecteurs : température, pH, effecteurs chimiques, sucres et alcools. Dans ces extraits, la température optimale d'activité a été déterminée à 60°C soit une légère thermophilie comparée aux cellulases mésophiles. L'évaluation de la stabilité thermique à différentes températures révèle que l'extrait brut contient plusieurs activités cellulasiques de sensibilité à la température différente. Le pH optimum de l'activité cellulasique principale est de 6, ce qui suggère que les activités mesurées proviennent d'enzymes intracellulaires libérées au cours de l'extraction, et non d'activité liée à des cellulases

secrétées, pour lesquelles le pH optimum devrait être plutôt alcalin compte tenu du milieu dans lequel se développe ces microorganismes.

De plus, ces activités cellulase présentent une bonne stabilité dans différents alcools à 40% (propanol, methanol et ethanol). En plus ces enzymes sont tolérante à d'accumulation de sucres.

Ces résultats valident la présence de cellulases dans la souche A8 provenant du lac Bogoria. Ces activités cellulase sont actives sur des substrats synthétiques mais aussi sur des substrats naturels tels que le coton ou le papier filtre. Par ailleurs, ces activités présentent un potentiel dans les applications nécessitant des températures d'utilisation élevées et la présence des solvants organiques.

Ce travail préliminaire, réalisé à l'Université de Nairobi, justifie le clonage de ces activités cellulolytiques entrepris à l'Université de Nantes et qui sera présentées dans le chapitre suivant.

Potential Cellulose Degrading Bacteria (CDB) with tolerance to high organic solvent and high sugar concentration

Lydia Awuor Ogonda [1,2,3†*] Edward Kirwa Muge [1], Benson Munyali Wamalwa[3], Francis Jackim Mulaa [1], Charles Tellier [2]

1. University of Nairobi, Department of Biochemistry, Faculty of Science and Technology, College of health sciences, P.O BOX, 30197-00100 Nairobi, Kenya
2. Universite de Nantes, UMR-CNRS 6286, US2B, 2 rue de la Houssinière ,44322 Nantes Cedex 3, France
3. University of Nairobi, Department of Chemistry, Faculty of Science and Technology, College of physical and Biological sciences, P.O BOX, 30197-00100 Nairobi, Kenya

ABSTRACT

Background/Context: Cellulosic biomass is a cheap and highly abundant bio-resource. In addition, there is a demand for a carbohydrate driven economy. However, cellulose degradation remains a major bottleneck. In efforts to overcome this challenge, a lot of research targeted towards novel cellulases as well as modifications on the already existing cellulase enzymes that could be used to efficiently degrade cellulose has been conducted. In spite of the huge potential of cellulose in industries for the production of chemicals, it largely remains underutilized due to challenges with feedback inhibition from even low concentrations of accumulating sugars, as well as low tolerance of the cellulases themselves to extreme pH, temperature and organic solvent and other reactor conditions.

Purpose: We therefore sought to characterize crude cellulases obtained from a bacteria isolated from Lake Bogoria.

Results: Our findings, established the novelty of this bacterium previously characterized as a *Bacillus pumilus*. Further, we showed the stability of the crude cellulases from this bacteria in 40% (v/v) organic solvents; propanol, methanol, ethanol. Moreover, the enzymes showed tolerance to high concentration of sugars.

Conclusion: These findings validate and show the potential of these microorganisms and/ its enzymes in cellulose hydrolysis and co-expression bio-factories with high organic solvent and high sugar yields.

Potential Implications: This work is a preliminary study to our work on the activity of the pure extracted cellulases from this microorganism. These findings substantiate the potential inclusion of either this microorganism or its isolated/extracted enzymes in the industrial processing of cellulosic biomass.

Keywords: Cellulase, *Bacillus*, Cellulose Hydrolysis, Cellulose, Cotton

4.1 BACKGROUND

Previously, chemical methods of cotton treatment were employed in the textile industry. However, these ancient methods have presented problems of environmental pollution, high temperature requirements, high alkaline pH and excessive energy costs involved in the maintenance of the high temperature and in the removal of the vast residual wastes during wastewater treatment (Verenium-cottonase ® and Novozymes).

This escalated production cost due to huge alkali chemical requirements, water requirements and energy costs as well as the taxing decontamination processes and the poor fabric quality obtained at the end of the process, have driven the need for alternative processes that can alleviate the environmental exertions, reduce processing costs and minimize chemical requirements hence making the processes more sustainable.

Biological solutions in the form of enzymes have been found to alleviate the inconveniences associated with chemical treatments as they work optimally under mild conditions with minimal residual wastes. Moreover, the fabric obtained at the end of the process, is of good quality.

In nature, multiple enzymes are employed by microorganisms to efficiently degrade plant cell wall cellulosic polysaccharide material. In fact, microorganisms employ three principal systems for enzymatic breakdown of cellulosic plant cell wall. These work either in combination or in separation and include; free enzymes, multifunctional enzymes and multi enzyme complexes such as , the cellulosome (Mcclendon *et al.,* 2012).

Applications of cellulases have increased considerably especially in the textile industry during the last two decades. These include bio processing of natural fibers, bio polishing of cotton fabrics in order to enhance the softness and feel, appearance and treatment of recycled fibres to restore fibre texture and flexibility lost during operations (Karmakar & Ray, 2011).

It is therefore vital to bioprospect for and characterize novel cellulases as well as novel cellulase producers. This study aimed at characterizing new cellulase enzymes from the thermophilic, alkalophilic and saline Lake Bogoria found in the Rift valley region of Kenya.

4.2 MATERIALS AND METHODS

4.2.1 Chemicals

G-75 sephadex (1.7 X 30cm, Pharmacia Fine chemicals, Sweden), salts used in making of microbial media were of analytical grade, TLC plates, TLC sprayer (CAMAG), sugars used here were of analytical grade, cotton, and p-anisaldehyde for TLC visualization. All experimental manipulations were conducted in triplicates.

4.2.2 Strains and media

Luria Bertani (LB) medium and Mineral salt Medium (MSM) supplemented with 0.5% (w/v) sodium carboxymethylcellulose (Na-CMC) was used for production of cellulase enzymes.

4.2.3 Enrichment medium

Microorganisms were grown in Luria bertani (LB) nutrient medium (10g/l tryptone, 5g/l yeast, 10g/l NaCl). This was done in order to enrich the culture before sub-culturing in MSM, supplemented with Na-CMC as the sole carbon source.

4.2.4 Enzyme Production

LB overnight culture (1ml) was inoculated into MSM made of tryptone 0.5% (w/v), yeast extract 0.1% (w/v), K_2HPO_4 0.08% (w/v) , KH_2PO_4 0.06% (w/v), $(NH)_2SO_4$ 0.1% (w/v), $MgSO_4.7H_2O$ 0.02%(w/v) , $CaCl_2.H_2O$ 0.005%(w/v), NaCl 0.3%(w/v), $FeCl_3$ 0.00001%(w/v) and supplemented with 0.5%(w/v) , Na-CMC to induce the expression of cellulase enzyme genes. The medium was incubated for 5 days at 37°C, 180rpm (Anish, Rahman & Rao, 2007).

4.2.5 Enzyme Purification

4.2.5.1 Concentration of the protein

After enzyme production, the supplemented MSM culture medium was centrifuged at 10,000rpm, 4°C for 20mins. The clarified supernatant containing the secreted cellulase enzyme proteins was then concentrated by precipitation in 50% ice-cold acetone overnight at -20°C. The precipitate was then pelleted by centrifugation at 15,000xg, 4°C for 15mins. The pellet was subsequently resuspended in 20mM Tris-HCl pH8.8.

4.2.5.2 Protein Purification by Gel filtration

A sephadex column (G-75, 1.7cm X 30cm, Pharmacia fine chemicals, Sweden) was used to fractionate the concentrated crude cellulase protein based on molecular weight. The sample buffer used was phosphate buffer (50mM, pH7) with 0.15M NaCl, whereas the eluent buffer

was phosphate buffer with 0.5M NaCl. Thirty fractions of 1.5ml each were collected and monitored at 280nm (UV-mini 1240, UV-Vis spectrophotometer, SHIMADZU) .The fractions were tested for protein content using biuret assay.

Fractions that peaked at 280nm were then assayed for cellulase activity. Those showing the highest activity corresponded to elution fractions 8 to 16 (LB pH7, chromatogram) and elution fractions 40 and 45 (Na-CMC, MSM pH10, chromatogram).

4.2.6 Cellulase activity assays

Cellulolytic activity was quantified using DNS assay (Miller, 1959; Dashtban et al. 2010). Due to the analytical complexities presented with pure crystalline cellulose substrates, we used Na-CMC; a cellulose derivative with a higher degree of polymerization and with better solubility for these analysis.

4.2.6.1 Dinitrosalicylic acid (DNS) assay

Gel filtration fractions, were tested for cellulase activity using the 3,5 dinitrosalicylic acid method (Miller 1959). DNS reagent (750µl), was added to a 1ml cellulase reaction test tube containing 100µl of the gel purification fractions. 40% sodium potassium tartrate (250µl) was added to the mixture and heated at 100°C for 5minutes. Optical density (O.D) was then recorded at 550nm using a spectrophotometer (UV, SHIMADZU). Using a predetermined glucose standard curve, the glucose concentrations of the samples were obtained and used to determine enzymatic activity.

4.2.6.2 Thin Layer Chromatography (TLC) assays

TLC was used to follow the cotton and filter paper cellulase hydrolysis. The spotted plates were developed in acetonitrile: water (v/v). The plates were then air dried and sprayed (TLC sprayer,

MERCK) with visualization solution (1ml *p*-anisaldehyde, 1ml 97% H_2SO_4 in 18ml ethanol) .This was followed by heating at 110°C for 30 minutes for staining.

4.2.7 Optimization of cellulase activity

Cellulase enzymatic activity parameters (temperature, pH and time), were determined.

4.2.7.1 Temperature Optimum

The temperature profile was determined by recording the cellulase activity between 20°-100°C.This was determined at pH 7 for the temperature range by incubation for 2hours. And reducing sugar released was quantified using DNS test.

4.2.7.2 pH Optimum

The optimum pH was also determined by recording the cellulase activity between pH 2-14. Different solutions (50mM) with potent buffering capacities at various points were used; Glycine-HCl buffer, pH2-3 , sodium acetate buffer pH 4-5, phosphate buffer pH 6-7, Tris-HCl buffer pH 8-9, Sodium bicarbonate –NaOH buffer pH 10 and Glycine –NaOH buffer pH 11-14. The samples were incubated at optimum temperature for 2 hours in the buffer after which reaction was stopped and reducing sugar quantified by DNS assay.

4.2.7.3 Reaction times

Optimum reaction/ incubation times was determined by recording the cellulase activity in 1-hour intervals over an 8-hour period at the optimum pH and temperature. This was preceeded, by an initial 30minutes interval reading.Thereafter, an aliquot was taken at 1hour from the sample and activity tested using DNS assay.

4.2.7.4 Enzyme Stability assays

pH and temperature stability studies were done by determining the residual enzymatic activity following pre-incubation at pH 2-14 and 20°-100°C.

4.2.7.4.1 pH stability

Enzyme was incubated in different buffers, pH 2-14 for 1-hour at room temperature. This was followed by determining of the residual activity by adding 1% Na-CMC and incubation for 2hours at 60°C. An aliquot was taken out from the reaction tube, reaction was stopped, and the amount of reducing sugar was then quantified by DNS test. DNS reagaent was added to the sample and heated at 100°C for 10mins. And O.D recorded at 550nm

4.2.7.4.2 Temperature stability

Enzyme was incubated in buffer for 1hour at different temperature 20-100°C. This was followed by determination of the residual activity by adding 1% Na-CMC and incubation for 2hours at 60°C. The amount of reducing sugar was then quantified by DNS test.

4.2.7.5 Effect of different compounds on cellulase activity

The effect of various sugars (monosaccharides, disaccharides, polysaccharides), alcohols, chemical reagents (ions, metal chelators surfactants and detergents) on cellulase activity was also determined.

4.2.7.6 Cellulase substrate specificity assays

Cellulase substrate specificity was also determined on a number of soluble and insoluble substrates (Avicel, Na-CMC, cellobiose and cotton).

4.3 RESULTS AND DISCUSSION

Thermophilic microorganisms are important sources of industrial cellulases because of their compatibility with the harsh textile processes such as softening of fabric materials, biostoning, and biopolishing (Sahoo *et al.* 2019).

4.3.1 Production and Purification of cellulase enzyme proteins

The microorganism was inoculated into LB medium and minimal salt medium supplemented with 0.5% sodium carboxymethylcellulose (Na-CMC) as the main carbon source. The role of Na-CMC was to induce the expression of the cellulase gene and cellulase production (Sang-Mok &Koo, 2001; Kubicek, 1993). The secreted cellulase enzymes were harvested by centrifugation at 4°C, 10,000rpm for 20 mins (Schallmey *et al.* 2004). This was followed by concentration as in section 4.2.5.1. The concentrated proteins were purified on a shorter column (production in a LB medium), then a final run (production in MSM) as shown in **Figure 4.1** and **Figure 4.2.**

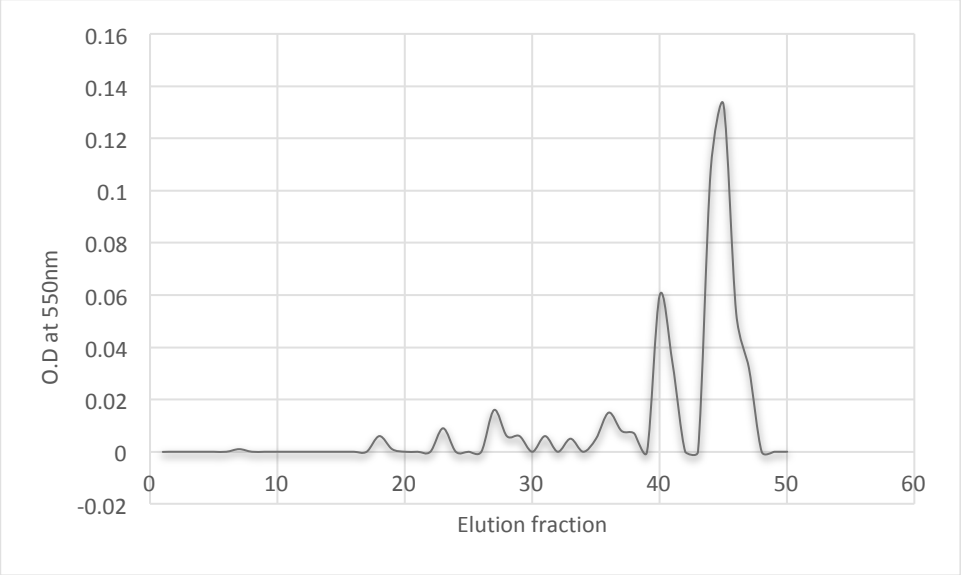


Figure 4.1: Gel filtration chromatographic purification of cellulase by Sephadex G-75 using a 5-day Na-CMC MSM growth culture at 37°C, 180rpm at pH 10.

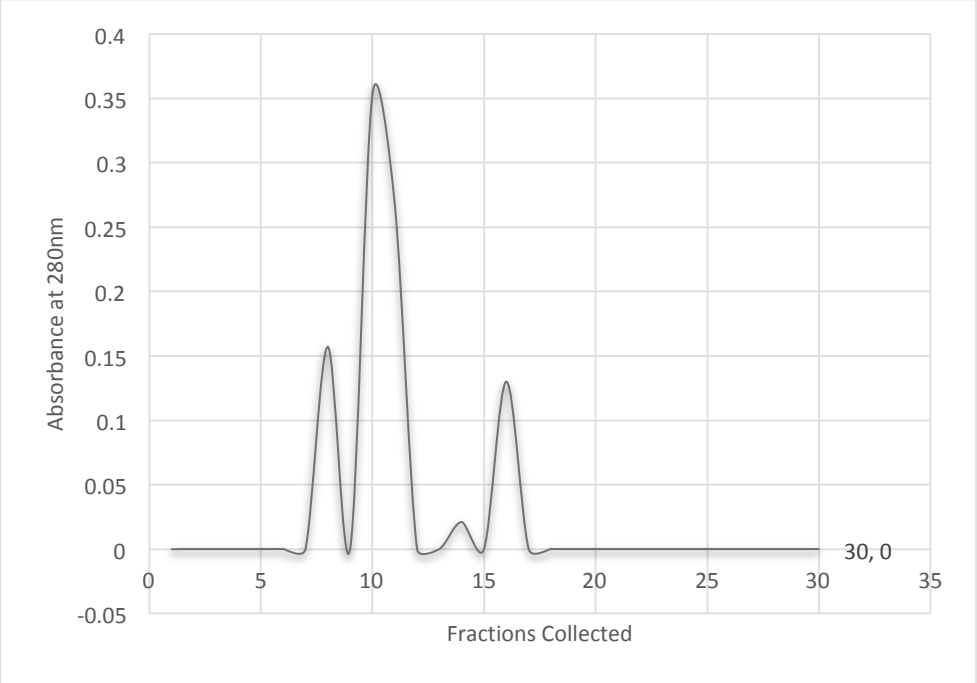


Figure 4.2: Gel filtration chromatographic purification of cellulase by Sephadex G-75 using an overnight LB growth culture at 37°C, 180rpm at pH 7.

The fractions corresponding to C10 and C16 (Figure 4.2), and fraction 41, and fraction 45 (Figure 4.1) that showed the highest reading using UV were pooled together and further analyzed for enzyme activity using *pNP*-β-cellobioside as a substrate. They recorded enzyme activity of $4.93X10^{\wedge-6}U$ (C10/C16) and $7.01X10^{\wedge-6}U$ (C41/C45) respectively.

4.3.1.2 Determining the protein concentration

Protein concentration was then determined by taking O.D readings at 280nm at the appropriate dilutions and by biuret assay readings at 550nm**. (Table 4.3; SS4)**

4.3.2 Optimization of cellulase activity

4.3.2.1 Reaction temperature

Bacterial cellulases have been shown to have optimal activity in the temperature range (35°-50°C) (Aygan *et al.* 2011). For this study, there was a significant change in enzyme activity at the various temperatures within the tested temperature range of 20°-100°C ($p<0.05$) (ANOVA). Further, the enzyme showed a working temperature range between 20°-80°C. Moreover, the highest enzymatic activity was recorded at 60°C **(Figure 4.3).** This confirms and validates the thermophilic nature of the microorganism which was isolated from *Lake Bogoria*; a hot water lake in the Rift valley region of Kenya (Ogonda *et al*. 2021). This is of interest particularly for industry as these enzymes could find application in both ambient and temperature intensive applications. Further, they could be used to substitute the temperature intensive processes by conducting the processes at the ambient conditions; subject to process optimization and cost/ benefit analysis.

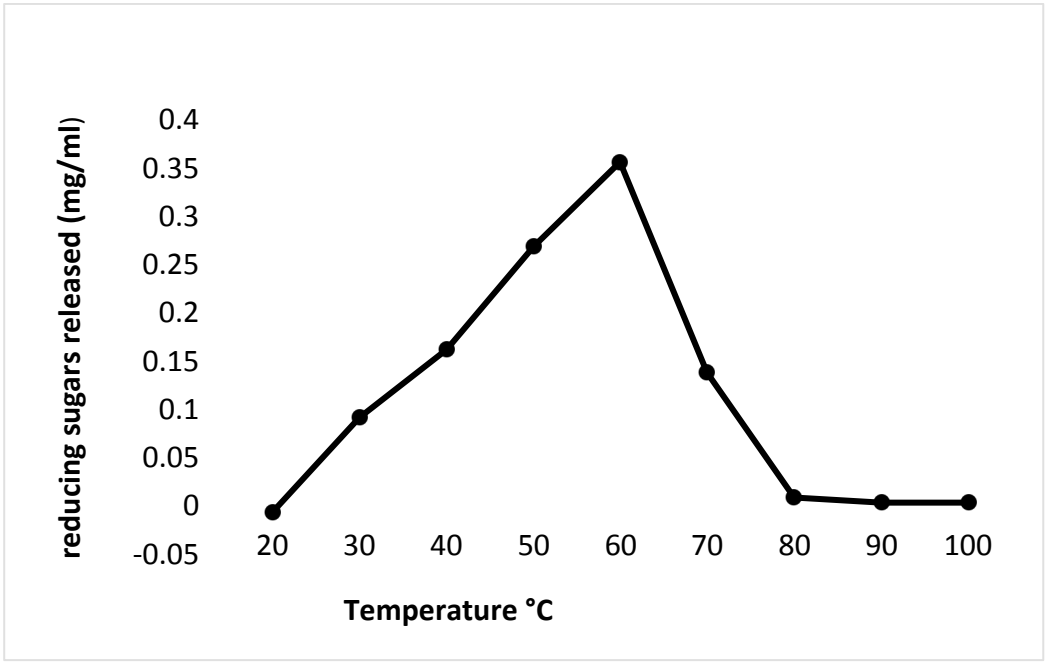


Figure 4.3: Effect of temperature on crude native cellulase enzymatic activity

Further the recorded optimum temperature is similar to those reported by Lima *et al.* 2005 but higher than those reported by Balasubramanian & Simoes, 2013.

4.3.2.2 Optimum pH

There was a significant effect on enzyme activity with change in pH ($p<0.05$) (ANOVA). The crude enzymes showed dual peaks of activity between pH3-7 and pH8-11 **(Figure 4.4).** The highest activity was observed at pH5 and pH 10. These results indicate the presence of two types of cellulolytic activity (presumably endoglucanase and exoglucanase) as previously identified from the initial gene mining. In addition, this determination is essential in showing the suitability of these two enzymes present in one gene cluster as a part of an enzyme complex for industrial application.

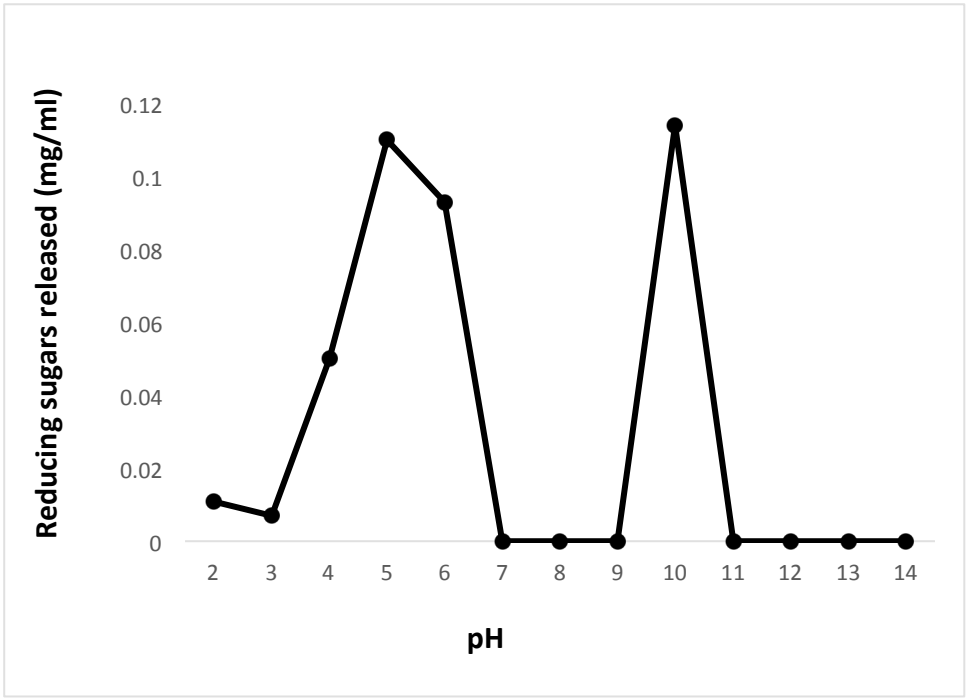


Figure 4.4: Effect of pH on crude native cellulase enzymatic activity (Glycine –HCl {pH 2-3}; Sodium acetate {pH4-5}; Phosphate buffer {pH6-7}; Tris-HCl {pH8-9}; Sodium bicarbonate-NaOH {pH10}; Glycine-HCl {pH 11-14}

4.4 Cellulase stability assays

4.4.1 Temperature stability

The crude cellulase enzymes showed triple peaks of thermal stability **(Figure 4.6)** with a sharp peak at 40°-70°C and flanking peaks at 20°-40°C and 70°-90°C. Moreover, it would be interesting to further test temperature ranges lower than 20°C as well as higher temperature ranges greater than 100°C due to the observed continued peaking at these points **(Figure 4.6)** The enzymes retained 9.50-99.99% of activity between 40°-60°C and 5.97-39.55% of activity between 70°-90°C. With the enzymes showing preference for the temperature range 40°-60°C, in which they retained the highest activity. This is an important characteristic of these enzymes, which would be of interest in synthetic biology/industry where high reaction temperatures (50°-60°C) for prolonged periods are compulsory (Lima *et al.*,2005).

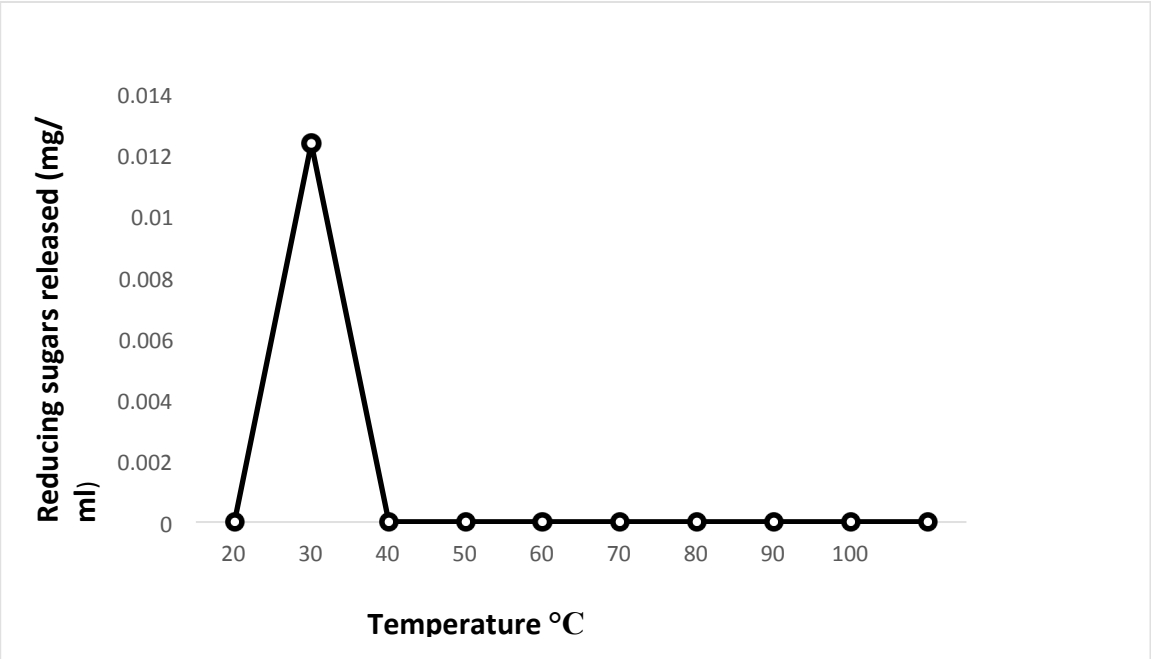


Figure 4.6: Temperature stability of crude native cellulase enzymes

4.4.2 pH stability

The crude enzymes showed stability peak between pH4-8 and was most stable at pH 6 **(Figure 4.7).** This is consistent with the results on the pure glycosyl hydrolase 9 (GH9) enzyme we extracted from this microorganism (Ogonda *et al.* 2021)

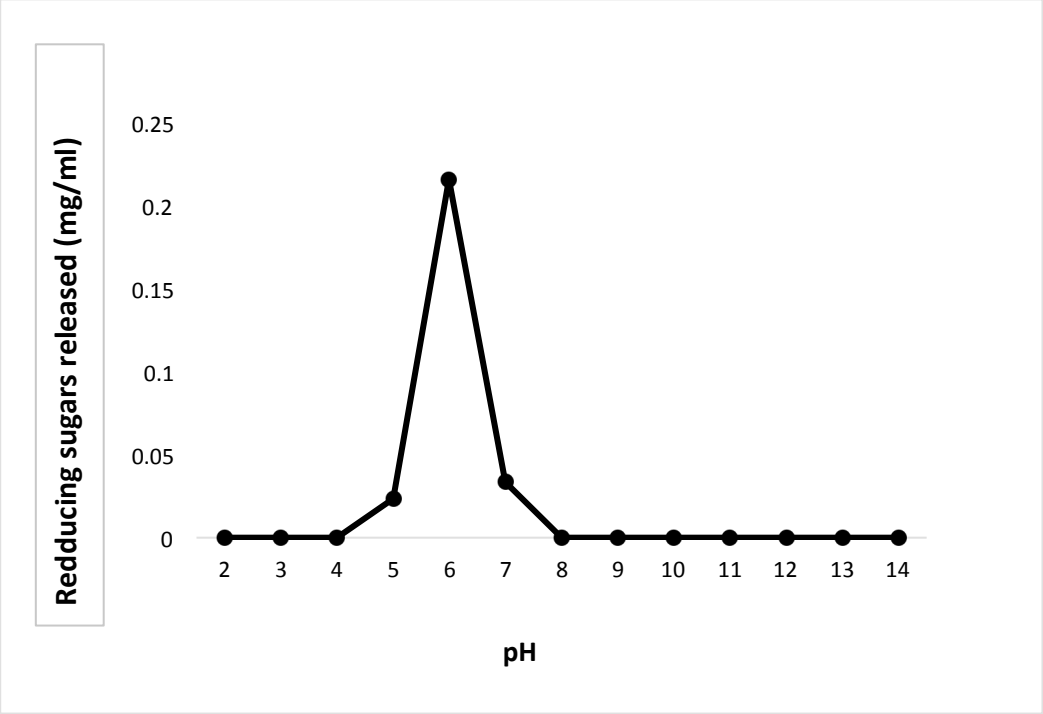


Figure 4.7*:* pH stability of crude native cellulase enzymes

4.5 Cellulase stability in various compounds

The effect of various compounds (saccharides, alcohols and chemical reagents; detergents/surfactants, ions and inhibitors). This is important especially in co-fermentation in bio-factories where cellulases could be applied. In addition, it is important to study for these because the cellulosic biomass is normally not pure and in some cases, pre-processing is needed. Moreover, to determine reagents to use in enzyme preparations. There was a significant effect on enzyme activity with the different chemical reagents ($p<0.05$) (ANOVA) **(Table 4.1) (Figure 4.8).**

4.5.1 Stability in Chemical reagents

Native crude cellulase activity was greatly inhibited by 10mM EDTA, urea, and β-mercaptoethanol and moderately by 10mM PMSF, triton 1% (v/v), SDS 1% (w/v) and 10mM

$CaCl_2$. Moreover, native crude cellulase activity was moderately enhanced by 10mM: $CuSO_4$, Imidazole and tween-20 1 %(v/v).native crude cellulase activity was greatly enhanced by 10mM: NaCl, $MgS0_4$, $MnCl_2$, $MnS0_4$, $ZnS0_4$, $CoCl_2$, KCl, DTT and $FeSO_4$. Similar effect on cellulase activity profiles have been previously reported (Balasubramanian & Simões, 2014).

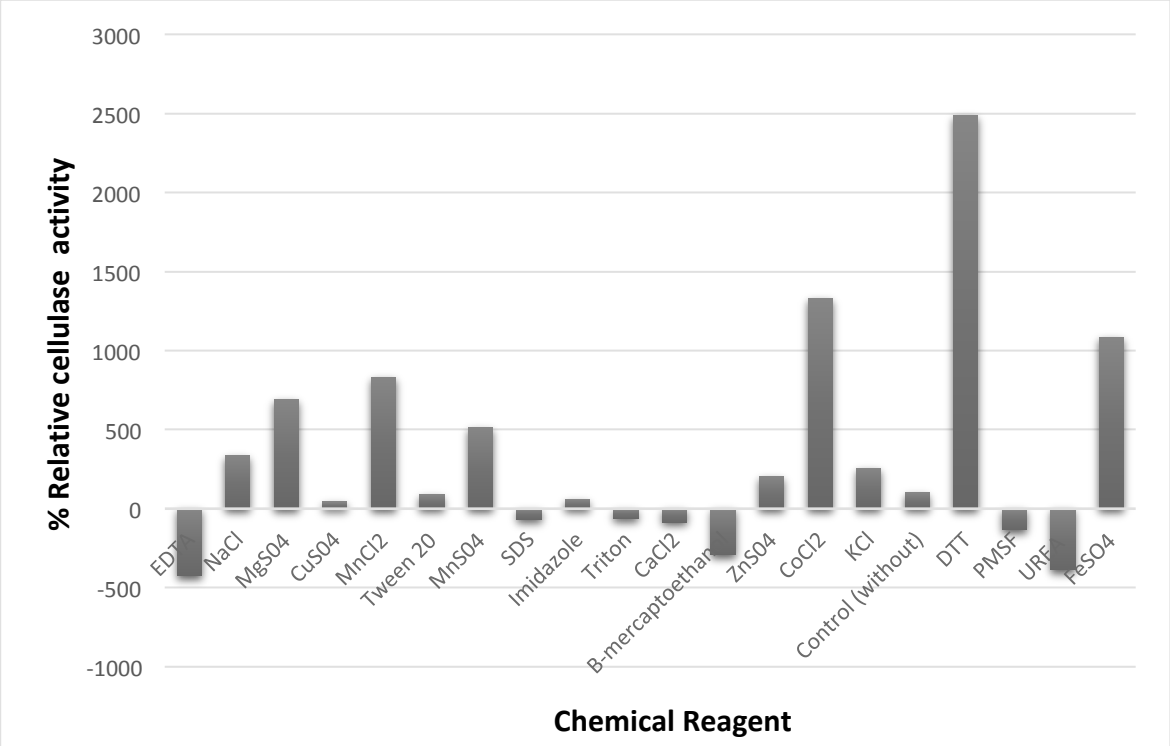


Figure 4.8: Effects of chemical reagents on cellulase enzymatic activity

Table 4.1: Effect of chemical reagents on crude cellulase enzymatic activity

Reagent	Concentration	Relative activity (%)
EDTA	10mM	0
NaCl	10mM	334
$MgSO_4$	10mM	692
$CuSO_4$	10mM	44
$MnCl_2$	10mM	831
Tween 20	1%	88

$MnSO_4$	10mM	513
SDS	1%	0
Imidazole	10mM	60
Triton	1%	0
$CaCl_2$	10mM	0
β-mercaptoethanol	10mM	0
$ZnSO_4$	10mM	206
$CoCl_2$	10mM	1329
KCl	10mM	256
Control (without chemical)		100
DTT	10mM	2485
PMSF	10mM	0
UREA	10mM	0
$FeSO_4$	10mM	1083

4.5.2 Stability in Sugars

Effect of different saccharides on native crude cellulase activity was tested **(Figure 4.9).** There is a significant effect on enzyme activity with different saccharides ($p<0.05$) (ANOVA).Cellulase activity was enhanced by mannitol, starch, inositol, sucrose, sorbitol, D-xylose, raffinose, lactose, maltodextrin, trehalose, D-glucose and cellobiose. This is contrary to previous studies that showed significant inhibitory effects on cellulase activity during cellulose hydrolysis by sugars (Xiao et al. 2004). This profile is however slightly similar to those reported by (Nigam & Prabhu, 1991) who showed that glucose, xylose and sucrose enhanced cellulase activity while cellobiose had severe inhibitory effects. The stability and activity in the presence of these sugars could be attributed to the low levels of sugars utilized (0.2µg/µl). Further, the stability and activity observed can be attributed to the presence of cellulase cocktail in the crude

enzyme extract thus enabling a high substrate consistency and minimizing the classical product (sugar) inhibition.

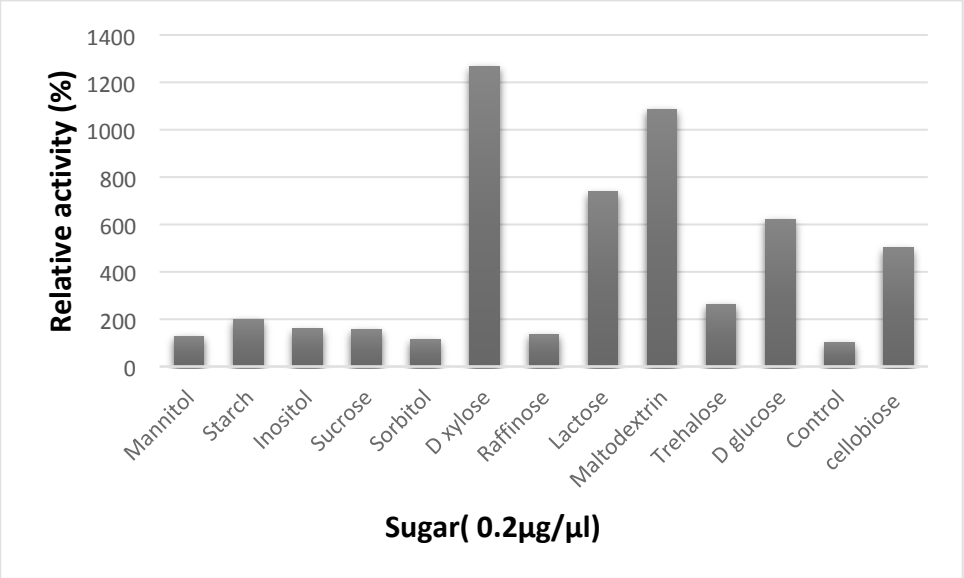


Figure 4.9: Effect of saccharides on crude native cellulase enzyme activity

4.5.3 Stability in alcohols

Effect of alcohol on native crude cellulase enzyme was tested. There is a significant effect on enzyme activity with different alcohols ($p<0.05$). Ethanol, 40% (v/v), 2-propanol and methanol enhanced native cellulase activity **(Figure. 4.10)**

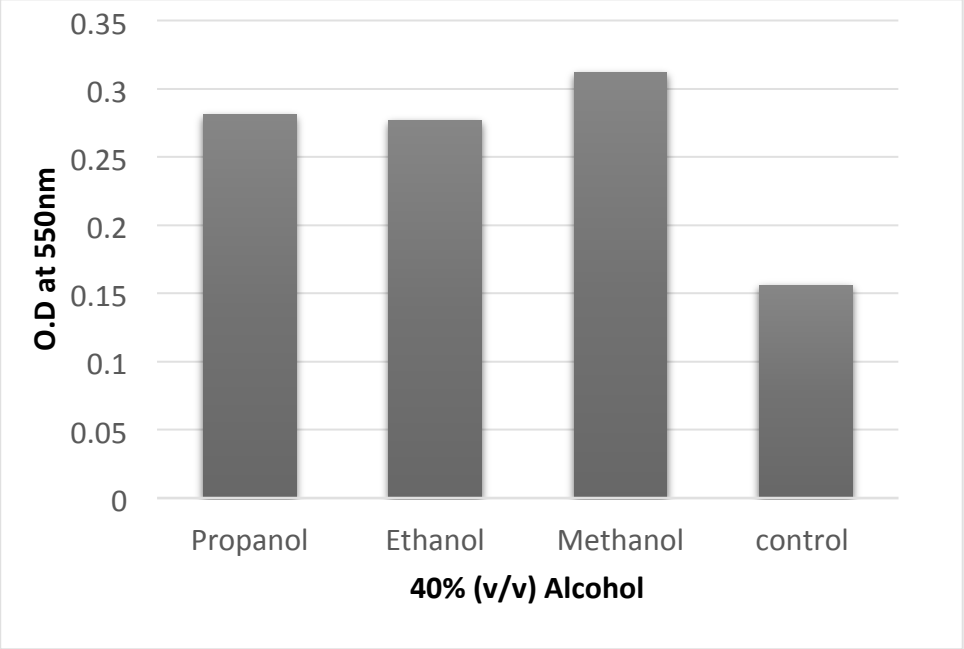


Figure 4.10: Effect of alcohol on crude native cellulase activity

4.6 Cellulase enzyme assays:

4.6.1 Cotton assay

We conducted a preliminary continuous assay using sodium acetate buffer (50mM, pH5), cotton (0.1%w/v) and crude enzyme concentrate. We weighed 0.1g, of cotton wool and filter paper and placed in 10ml sodium acetate buffer (50mM, pH 5.0).This was followed by addition of 10ml crude enzyme, and incubation at 60°C for 24 hrs. This cellulase activity assay using cotton was done over a 10-day incubation period. Aliquots of 1ml each were collected in constant time intervals, and tested for enzymatic activity by checking for reducing sugars using DNS assay **(Figure 4.11).** A similar experiment was conducted using 0.1g of cotton in 5% (v/v) H_2SO_4. The reaction products were then assayed for glucose using DNS assay. Cotton/ filter paper breakdown (cellulose) was also followed using solubilisation and weighing test.

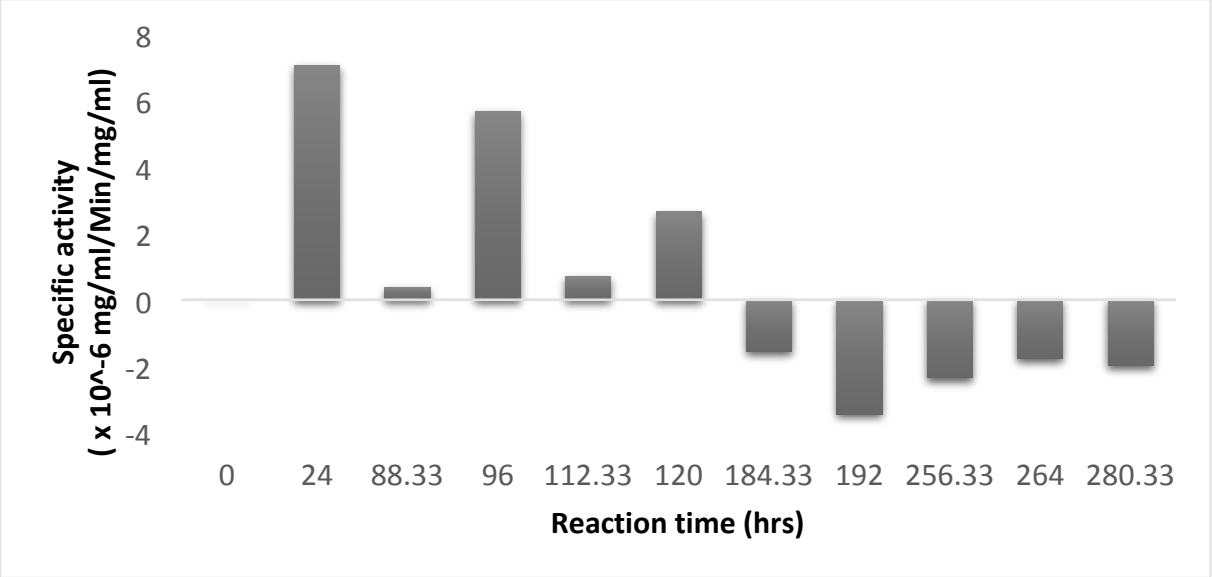


Figure 4.11: Specific activity; mg/ml glucose released per minute per mg/ml of concentrated crude cellulase protein, on cotton

Cellulose (Cotton) Hydrolysis

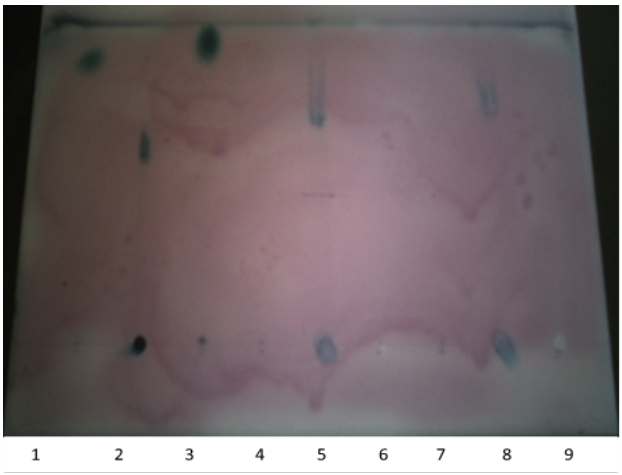


Figure 4.12: Cotton cellulase hydrolysis product profiling by TLC (**Spot 1**: Cellobiose; **Spot 2**: Na-CMC; **Spot 3**: Glucose; **Spot 4**: Control **Spot 5**: Enzyme cotton; **Spot 6**: Acid; cotton; **Spot 7**: Control filter; **Spot 8**: Enzyme filter; **Spot 9**: Acid filter)

CONCLUSION

The findings from this study show cellulase enzymes stable over a wide range of cellulose processing conditions. These cellulases have been shown to be stable in up to 40% organic solvents such as propanol, methanol and ethanol. Moreover, these cellulases have presented with a high tolerance for accumulation of sugars. In addition, the cellulase activity on cotton with the release of shorter saccharides has been demonstrated. It would therefore be important to test for the quality (functional and physical properties) of the cotton fiber achieved after this processing.

Table 4.2: Substrate specificity of crude native cellulases

Substrate	Cellulase activity	Concentration
Avicel	ND	1%
CMC	0.020284	1%
PNP-β-D-Cellobiose	0.439805	1%
Cotton	0.0069	1%

REFERENCES

Anish, R., Rahman, M. S., & Rao, M. (2007). Application of cellulases from an alkalothermophilic Thermomonospora sp. in biopolishing of denims. *Biotechnology and Bioengineering*, *96*(1), 48–56. http://doi.org/10.1002/bit.21175

Aygan A, Karcioglu L, Arikan B. Alkaline thermostable and halophilic endoglucanase from *Bacillus licheniformis* C108. Afr J Biotechnol 2011; 10(5): 78996

Balasubramanian, Natesan & Simões, Nelson. (2014). Bacillus pumilus S124A carboxymethyl cellulase; a thermo stable enzyme with a wide substrate spectrum utility. International journal of biological macromolecules. 67. 10.1016/j.ijbiomac.2014.03.014.

Cavaco-Paulo, A. (1998). Processing Textile Fibers with Enzymes: An Overview. In: Enzyme Applications in Fiber Processing, Eriksson, K.E. and A. Cavaco-Paulo (Eds.). American Chemical Society Washington, DC. ISBN-13: 9780841235472, pp: 180-189.

Dashtban, Mehdi & Schraft, Heidi & Syed, Tarannum & Qin, Wensheng (2010). Fungal biodegradation and enzymatic modification of lignin. International journal of biochemistry and molecular biology. 1. 36-50

Jatinder Kaur; Bhupinder, S. Chadha; Badhan A. Kumar; Ghatora S.Kaur ; Harvinder, S. Saini.(2007). Purification and characterization of β-glucosidase from *Melanocarpus sp. MTCC 3922.* Electronic journal of Biotechnology ISSN: 0717-3458

Karmakar, M and Ray, R.R. (2011). Current Trends in Research and Application of Microbial Cellulases. *Research Journal of Microbiology, 6: 41-53.*

Kubicek, C.P. (1992), *Adv. Biochem. Eng.,* 45, 1-27. The cellulase proteins of T. *reesei:*

structure, multiplicity, mode of action and regulation of formation

Kubicek, C.P., Messner, IL, Gruber, F., Maeh, ILL. and Kubicek-Pranz, E.M. (1993),

Enzyme Microb. Technol., 15, 90-99. The *Trichoderma reesei* cellulase regulatory puzzle--

from the interior life of a secretory fungus.

Lima AOS, Quecine MC, Fungaro MHP, Andreote FD, Maccheroni W, Araújo WL, Silva-Filho MC, Pizzirani-Kleiner AA, Azevedo JL (2005). Molecular characterization of a β-1,4-endoglucanase from an endophytic Bacillus pumilus strain. Appl Microbiol Biotechnol 68:57–65 https://doi.org/10.1007/s00253-004-1740-1

Lynd, L. R., van Zyl, W. H., McBride, J. E., & Laser, M. (2005). Consolidated bioprocessing of cellulosic biomass: an update. *Curr Opin Biotechnol, 16*(5), 577-583.

Lynd, L. R., Weimer, P. J., van Zyl, W. H., & Pretorius, I. S. (2002). Microbial cellulose utilization: fundamentals and biotechnology. *Microbiol Mol Biol Rev, 66*(3), 506-577, table of content

Miller, G. L. (1959). Use of dinitrosalicylic acid reagent for determination of reducing sugars. *Analytical Chemistry, 31*, 426-429.

Nigam, P., & Prabhu, K. A. (1991). Influence of sugars on the activity of cellulase system from two basidiomycetes cultures. *Journal of Basic Microbiology*, *31*(4), 279–283. http://doi.org/10.1002/jobm.3620310408

Verenium-cottonase ® http://www.verenium.com/prod_cottonase.html

Xiao, Z., Zhang, X., Gregg, D. J., & Saddler, J. N. (2004). Effects of sugar inhibition on cellulases and β-glucosidase during enzymatic hydrolysis of softwood substrates. *Applied Biochemistry and Biotechnology*, *115*(1-3), 1115–1126. http://doi.org/10.1385/ABAB:115:1-3:1115

SUPPLEMENTARY INFORMATION

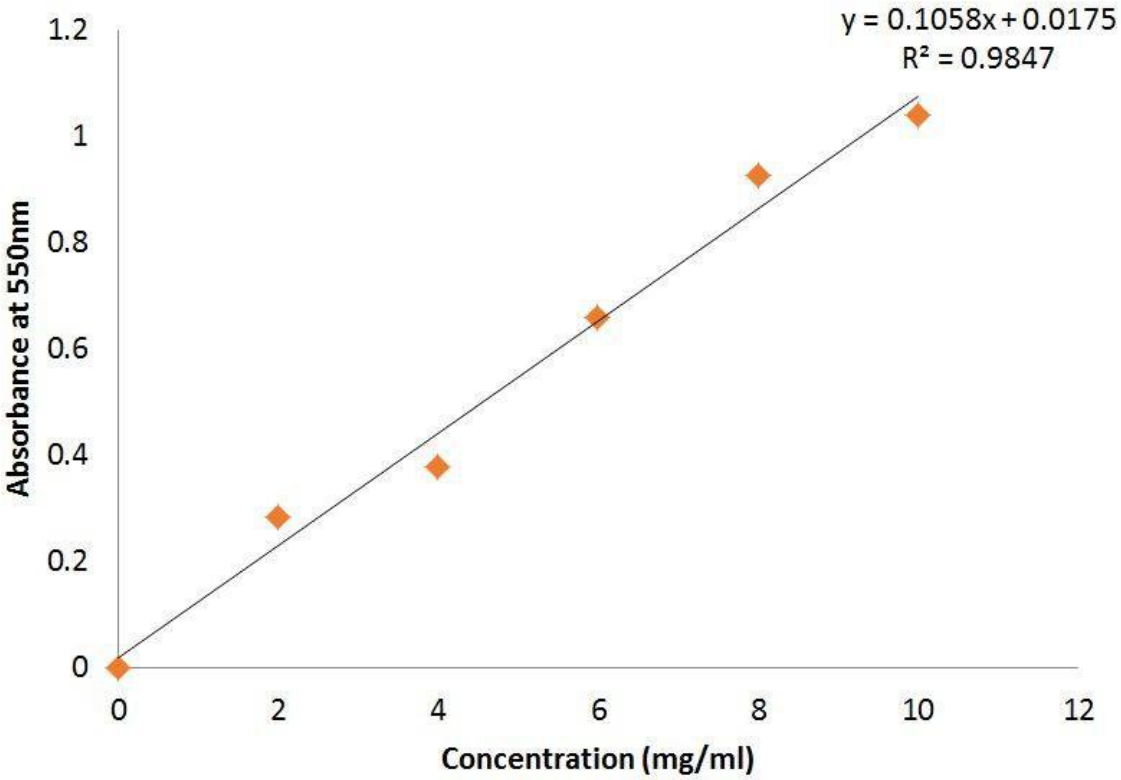


Figure 4.13: Protein (Bovine serum Albumin) standard for Biuret assay **(SS1)**

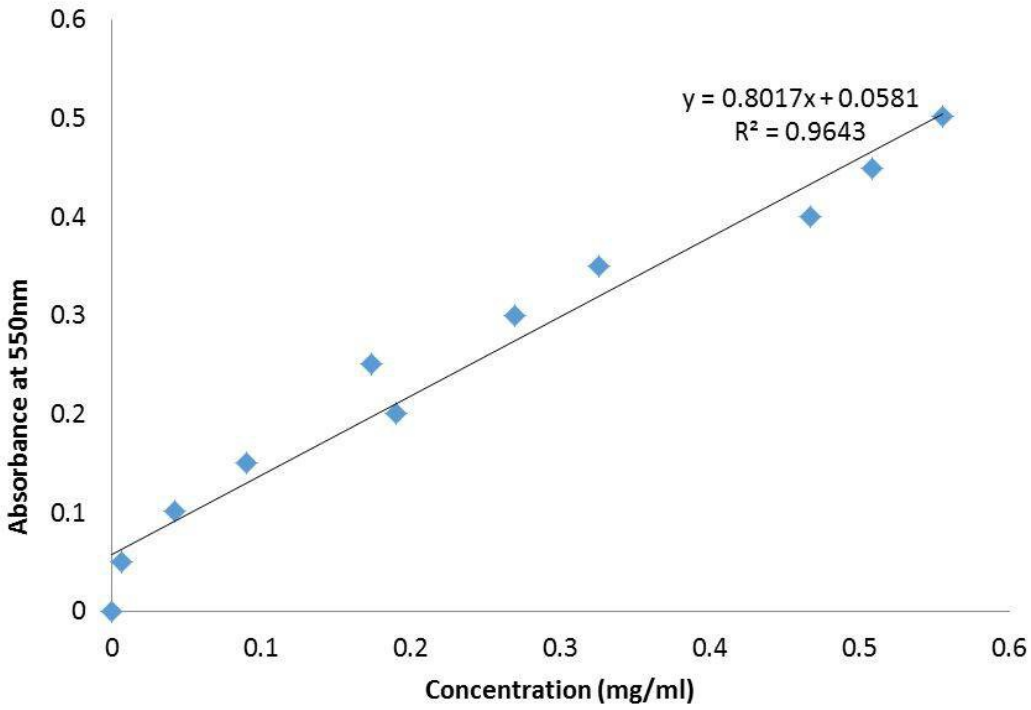


Figure 4.14: Glucose standard curve for reducing sugar (DNS) assay test **(SS2)**

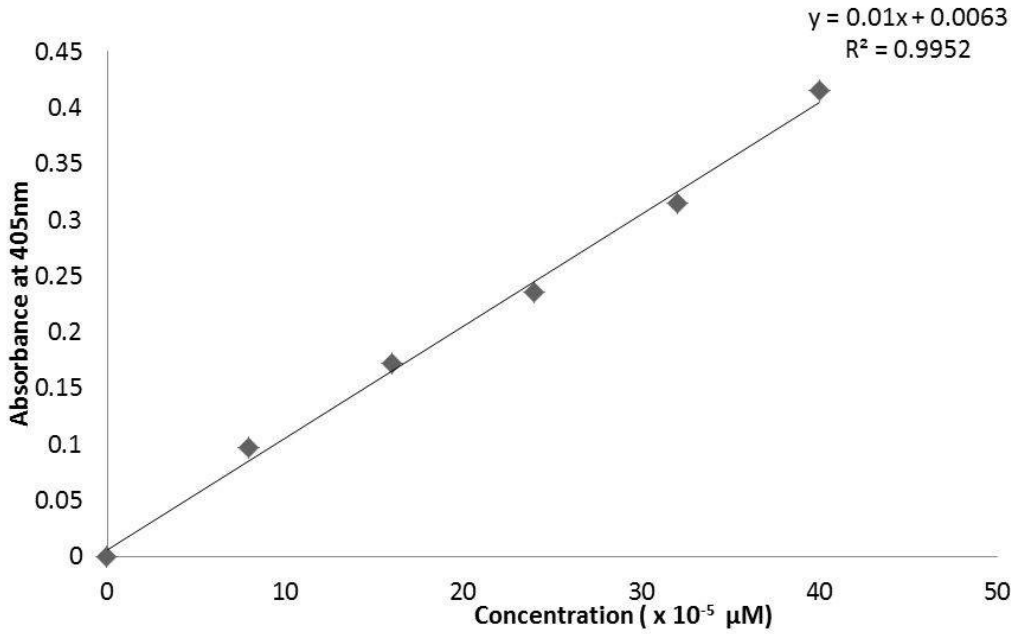


Figure 4.15: Para-nitrophenol standard assay **(SS3)**

Table 4.3: Protein concentration (SS4)

Sample	**0.D at 280nm**	**O.D at 550nm**	**Protein concentration (mg/ml) Biuret**	**Protein concentration (mg/ml) O.D 280nm**
LB pH 10	0.388	0.247	43.38	7.76
MSM pH 10	0.370	0.232	40.54	7.4
MSM pH 7	0.344	0.242	42	6.88

LB pH 10; Luria bertani media at pH 10, **MSM pH 10** and **MSM pH 7**: Minimal salt media at pH 10 and 7 respectively.

CHAPTER 5

Résumé de la publication II

Le travail précédent a validé la présence d'activités cellolytiques dans une souche de bactéries cultivées à partir de prélèvements effectués dans le lac Bogoria au Kenya. La caractérisation précise des activités cellulolytiques et de leurs propriétés est difficile à partir d'un extrait brut dans la mesure où plusieurs activités différentes peuvent être présentes. Nous avons donc entrepris un travail de clonage et de sur-expression de cellulases à partir de souches bactériennes provenant du lac Bogoria.

Dans un premier temps un criblage d'activités cellulolytiques a été effectué sur 95 souches isolées. Parmi ces souches, 42 ont présenté une activité cellulolytique à pH neutre et 23 ont conservé cette activité à pH 10. La souche alcalophile A8, identifiée comme étant un *Bacillus pumilus*, a été sélectionnée pour le clonage des activités cellulolytiques.

A partir de l'annotation du génome de souches analogues de *B. pumilus* (WP8) dans lequel il avait été repéré deux gènes de cellulases, des amorces consensus ont été construites pour amplifier, à partir de la souche A8, les gènes codant pour une endocellulase (BpGH9) et une cellobiohydrolase (BpGH48). La séquence de BpGH9 indique que cette endocellulase est modulaire et constituée d'un module catalytique associé à un domaine de fixation à la cellulose de type CBM3c. En revanche, la protéine BpGH48 est constituée d'un seul module catalytique.

Les gènes correspondant à ces deux cellulases ont été sur-exprimés dans une souche recombinante (*E.coli* BL21 (DE3)), puis les cellulases correspondantes ont été purifiées et caractérisées biochimiquement. En accord avec les tests de criblage, l'endocellulase BpGH9 présente une excellente tolérance aux pH extrêmes, même si le pH optimum se situe autour de 6. L'origine de cette adaptation alcaline est difficile à comprendre. En effet, la séquence de BpGH9

ne diffère d'une autre GH9 non alcalo-résistante que par 8 mutations, qui sont situées loin du site actif. Par contre, la cellobiohydrolase BpGH48 n'est pas alcalophile. De plus, nous avons pu montrer à l'aide de substrats synthétiques que BpGH48 est une cellobiohydrolase de type II, qui hydrolyse les chaines de cellulose à partir de l'extrémité réductrice, ce qui n'est pas très fréquent, la majorite de cellobiohydrolases dans les microorganismes cellulolytiques étant de type I (hydrolyse à partir de l'extrémité non-réductrice).

Afin de comprendre la fonction du module CBM et la synergie entre les deux modules de BpGH9, nous avons produit différentes formes de cette enzyme : une forme tronquée dans laquelle le module CBM3c a été supprimé (BpGH9t). Une construction dans laquelle le module CBM3c a été remplacé par un module de fixation à la cellulose de type CBM3a (CBM3a-BpGH9t) et une construction dans laquelle le module CBM3a a été ajouté à l'extrémité N-terminale de BpGH9 (CBM3-BpGH9). L'élimination du module CBM3c conduit à une déstabilisation de la protéine et à une perte d'activité partielle sur la carboxyméthylcellulose (Na-CMC), qui est restaurée par la fusion avec le module CBM3c. L'addition de CBM3c à BpGH9 renforce l'activité cellulolytique en particulier sur la cellulose cristalline.

Ces résultats suggèrent qu'il est possible de moduler l'activité cellulolytique des endocellulases en ajoutant ou supprimant des modules de liaison à la cellulose. Ces approches porraient être utlisées pour augmenter l'activité d'endocellulases qui sont naturellement dépourvues de CBM. Cette étude confirme également une certaine tolérance des endocellulases à la fusion, que ce soit à l'extrémité N-terminale ou C-terminale.

https://doi.org/10.1007/s10529-020-03056-z

Characterization and engineering of two new GH9 and GH48 cellulases from a Bacillus pumilus isolated from Lake Bogoria

Lydia A. Ogonda · Amélie Saumonneau · Michel Dion · Edward K. Muge · Benson M. Wamalwa · Francis J. Mulaa · Charles Tellier

Received: 30 September 2020 / Accepted: 10 December 2020

Abstract
Objectives To search for new alkaliphilic cellulases and to improve their efficiency on crystalline cellulose through molecular engineering
Results Two novel cellulases, *Bp*GH9 and *Bp*GH48, from a *Bacillus pumilus* strain were identified, cloned and biochemically characterized. *Bp*GH9 is a modular endocellulase belonging to the glycoside hydrolase 9 family (GH9), which contains a catalytic module (GH) and a carbohydrate-binding module belonging to class 3 and subclass c (CBM3c). This enzyme is extremely tolerant to high alkali pH and remains significantly active at pH 10. *Bp*GH48 is an exocellulase, belonging to the glycoside hydrolase 48 family (GH48) and acts on the reducing end of oligo-b1,4 glucanes. A truncated form of *Bp*GH9 and a chimeric fusion with an additional CBM3a module was constructed. The deletion of the CBM3c module results in a significant decline in the catalytic activity. However, fusion of CBM3a, although in a non native position, enhanced the activity of *Bp*GH9 on crystalline cellulose.
Conclusions A new alkaliphilic endocellulase *Bp*GH9, was cloned and engineered as a fusion protein

Supplementary Information The online version of this article (https://doi.org/10.1007/s10529-020-03056-z) contains supplementary material, which is available to authorized users.

L. A. Ogonda · A. Saumonneau · C. Tellier (&) Université de Nantes, CNRS, UFIP, UMR6286, 2, rue de la Houssinière, 44322 Nantes, France
e-mail: charles.tellier@univ-nantes.fr

L. A. Ogonda · E. K. Muge · F. J. Mulaa
Department of Biochemistry, School of Medicine, College of Health Sciences, University of Nairobi,
P.O BOX 30197-00100, Nairobi, Kenya

Present Address:
L. A. Ogonda
Department of Medical Biochemistry, School of Medicine, Masinde Muliro University of Science and Technology,
P.O BOX 190-50100, Kakamega, Kenya

M. Dion
Université de Nantes, IRS2, 44000, Nantes, France

B. M. Wamalwa
Department of Chemistry, School of Physical Sciences, College of Biological and Physical Sciences, University of Nairobi, P.O BOX 30197-00100, Nairobi, Kenya

Published online: 01 January 2021

(CBM3a-*Bp*GH9), which led to an improved activity on crystalline cellulose.

Keywords Cellulose · Endoglucanase · Cellobiohydrolase · Alkaliphilic · CBM · *Bacillus pumilus*

Introduction

Cellulose is the major component of the plant cell wall and thus, the most abundant renewable biomass for the production of biofuels and biochemicals. It is a linear unbranched homopolymer of glucose residues linked by b-1,4 glycosidic bonds. Despite its simple structure, efficient degradation of cellulose is prevented by the complex structure of the plant cell wall wherein cellulose is embedded in a network of hemicellulose and lignin. Consequently, although cellulose is a potential feedstock for several industrial applications in food, textile, beverage, biofuel, pharmaceutical, paper, pulp and detergent industries (Jahangeer et al. 2005), it remains largely underutilized (Klemm et al. 2005).

Various microorganisms, including bacteria and fungi, have evolved enzymes to tackle the cellulose breakdown problem (Uchiyama et al. 2020). Aerobic microorganisms secrete to the milieu, a cocktail of cellulases and ancillary enzymes, which act synergistically to breakdown cellulose. Anaerobic organisms, on the other hand, are more efficient cellulose users/degraders. They possess an inherent multiple enzyme complex consisting of different catalytic units known as a cellulosome (Bayer et al. 2004; Doi and Kosugi 2004; Fontes and Gilbert 2010). In both mechanisms, the complete degradation of cellulose is a concerted effort between three different cellulase enzyme types, namely: β-1,4 endoglucanase, cellobiohydrolase/cellobiosidase, and β-glucosidase (Zhang and Zhang 2013). Analysis of genome sequences of cellulolytic microorganisms has allowed the identification of numerous cellulases that are involved in cellulose degradation (Lombard et al. 2014). However, the biochemical characterization of these new cellulases remains a hurdle to the identification of their functional properties.

In this study, we used functional screening methods to directly identify alkaliphilic cellulases secreted by strains isolated from Lake Bogoria in Kenya. This ecosystem is interesting since its water pH is alkaline (pH 9.3–10.3) and its water temperature varies from 25 to 60 °C due to sub-aquatic hydrothermal activity (Duckworth et al. 1996). The water is also particularly rich in organic matter due to the decomposition of terrestrial plants, microalgae and cyanobacteria. This study aimed at improving cellulose utilization by screening for novel, native, extremophilic cellulases with improved saccharifying capacity. Further, we sought to bioengineer the said cellulases using protein-engineering tools in order to improve cellulose breakdown.

Materials and methods

Chemicals

Majority of the chemical materials used in this study were purchased from Merck-Sigma-Aldrich®: Sodium carboxymethylcellulose (Na-CMC), Avicel PH-101, Bovine serum albumin (BSA), 2,2′,2″,2‴- (ethane-1,2-diyldinitrilo)tetra acetic acid (EDTA), Isopropyl □-D-1-thiogalactopyranoside (IPTG), phenylmethanesulfonyl fluoride (PMSF), *para*-Nitro-Phenyl-b-D-cellobiose (*p*NP-□-D-cellobiose), aminobenzonitrile, $NaCNBH_3$. Protein and DNA reagents for molecular mass determination and chem- icals for polyacrylamide gel electrophoresis (PAGE) were ordered from Thermoscientific® and Bio-Rad®. Oligosaccharides, cellobiose, maltotriose and cel- lopentaose, were purchased from Megazyme®. All other chemicals and reagents used were of analytical grade.

Screening Bogoria strains for cellulase producers

A microbial library (95 strains) previously isolated from the hot springs of Lake Bogoria (Kenya) was used for screening (Moallic et al. 2006). Each strain previously stored at − 80 °C, was enriched on a minimal salt media (supplemented and non-supplemented) at pH 7 (Supplementary Table 1) with 1% CMC. After growth of the strains (24 h at 37 °C), the plates were flooded with 0.1% (w/v) Congo red solution and left to stand for 30 min with intermittent shaking followed by a single rinse with H_2O. The plates were then flooded with 1 M NaCl with

intermittent shaking to destain and reveal the clear zones (halos) of cellulolytic activity (Supplementary Fig. 1). Strains that were positive for cellulolytic activity at pH 7 were further screened by inoculation on solid minimal salt media (supplemented or non-supplemented) at pH 9 and 10 to identify alkaliphilic cellulase producers. Potency of cellulolytic activity was determined using the halo to colony diameter ratios. From the screening activity, the highest ranked microbial strain was selected for downstream work. The genomic DNA (gDNA) was extracted using phenol–chloroform extraction method. 16srRNA was amplified by PCR using 534R and 27F primers (Supplementary Table 2) and sequenced to identify the microbial strains.

Cloning of *BpGH9* and *BpGH48* cellulase genes

Wild type endoglucanase (*Bp*GH9) and wild type cellobiohydrolase (*Bp*GH48) genes were then amplified by PCR from gDNA of the selected isolate, A8, using primers: *Bp*GH9F, *Bp*GH9R, *Bp*GH48F *and* *Bp*GH48R (Supplementary Table 2). These primers were designed from the corresponding sequences of a closely related strain, *Bacillus pumilus* WP8, whose genome (CP010075.1) has been fully sequenced (Kang et al. 2015). *Bp*GH9 and *Bp*GH48 genes were then cloned into linearized pECa (corresponding to $pET21_{(a)}$ plasmid with insertion into *Sph*I-*Bgl*II restriction sites of Tac2 constitutive promoter, 28pb) by *BamH*I/*Xho*I restriction enzyme in order to insert the poly-histidine tagged gene at the *C*-terminal of the coding region.

Assembly of *BpGH9* gene with *CBM3a* sequence

In order to construct a recombinant endoglucanase gene (BpGH9), the wild type endoglucanase gene was fused with the sequence of type 3a Cellulose Binding Module (*CBM3a*) from *Clostridium thermocellum* optimized for *E. coli* expression (synthesized by Genscript®). These two genes are bound by a long linker sequence made of 123 nucleotides. This construct was assembled using Gibson Assembly® protocol (New England BioLabs®Inc) with recombinant primers (Supplementary Table 2). This recombinant enzyme construct (CBM3a-*Bp*GH9; Fig. 2) was designed to contain in fusion, a poly-histidine tag at the *C*-terminal.

BpGH9 gene with a truncated *CBM3c* module

In order to remove the CBM3c sequence from wild type *BpGH9* and *CBM3a-BpGH9*, 2 primer couples (Table S2) were used to amplify the DNA fragments corresponding to *Bp*GH9 without endogenous CBM3c (*Bp*GH9t) and CBM3a-*Bp*GH9 without CBM3c (CBM3a-*Bp*GH9t). The purified PCR products were introduced using Gibson Assembly® protocol (New England BioLabs®Inc.) into pECa plasmid digested by *Bam*HI and *Xho*I restriction enzymes and purified.

Production and purification of recombinant enzymes

Recombinant strains (*E.coli* BL21 (DE3)) expressing the entire assembly or truncated *Bp*GH9 and *Bp*GH48 genes were grown in 1 L of LB media with ampicillin antibiotic (100 mg/mL), and supplemented with cal- cium at 37 °C to 0.6 of $O.D_{600nm}$ (Optical Density measured at k = 600 nm). IPTG at 0.5 mM was added to induce the expression at 30 °C overnight under agitation (200 rpm). The culture was then centrifuged, the pellet frozen at – 20 °C and then resuspended in 20 mL of lysis buffer (20 mM Tris–HCl pH 8.0) containing 10 mM Imidazole, 5 mM DNaseI, 1 mg/ mL lysozyme and a cocktail of protease inhibitors (Sigma-Aldrich®). After sonication and centrifuga- tion, the recombinant proteins with His-tag in the supernatant (soluble fraction), were purified using immobilized nickel ion—affinity chromatography Ni– NTA Superflow (Qiagen). Protein expression and purity was then checked using 12% sodium dodecyl sulfate–polyacrylamide gel electrophoresis (SDS-PAGE) (Supplementary Fig. 3). Enzyme concentrations were determined by UV absorbance at 280 nm using a NanoDrop 1000 (ThermoScientific) (Supplementary Table 3). This data was confirmed using Agilent 2100 Bioanalyzer® (Supplementary Fig. 2).

Enzymatic characterization

The enzymatic activities of *Bp*GH9 (b-1,4-endoglucanase) and that of its variants (*Bp*GH9t, CBM3a-*Bp*GH9, CBM3a-*Bp*GH9t) and of *Bp*GH48 (b-1,4-cellobiosidase), were conducted using different cellulosic substrates: Na-CMC, microcrystalline cellulose substrate (Avicel PH-101), amorphous cellulosic

substrate phosphoric acid swollen cellulose (PASC) (Stern et al. 2015) and pNP-b-D-cellobiose.

The activities of the enzymes (0.5 **1**M or 1 **1**M) were tested in 50 mM citrate buffer pH 6.0 and with 1% of cellulosic substrate (Na-CMC, Avicel or PASC) or 5 mM and pNP b-D-cellobiose. This was done at 37 °C for 1 h for Na-CMC, PASC and pNP b-D-cellobiose and at 50 °C for 24 h using Avicel. For the assays with Na-CMC, Avicel or PASC, 70 µl of reaction mixture was withdrawn at regular time intervals and added to 70 µl of dinitro salicylic acid solution (DNS). The mixture was heated at 95 °C for 10 min (Miller 1959). The optical density (O.D.) was then read at 540 nm by spectrophotometry (Magellan, Tecan®), to quantify the reducing sugars released. For the assay on pNP □ -D-cellobiose, 50 µl aliquots of the reaction mixture was withdrawn at regular time intervals and added to 100 µl of 1 M Na_2CO_3. The O.D. was then read at 405 nm to quantify the pNP released.

The mode of action of the *Bp*GH48 was determined using a derivatized cellopentaose substrate with aminobenzonitrile (ABN) (Kipper et al. 2005). The cellopentaose was derivatized by reductive amination for 2 h at 70 °C in a buffer reaction prepared with 100 mg of $NaCNBH_3$, 0.6 g of 4-aminobenzonitrile (ABN) and 0.5 mL acetic acid dissolved in 10 mL of methanol. The purified *Bp*GH48 enzyme was incubated with 20 mM of derivatized and non-derivatized cellopentaose in 50 mM of citrate buffer (pH 6.0) at 37 °C for 8 h. At the start of the reaction (t0) and after at 1, 2, 4, 6 and 8 h, aliquots (1µl) of the reaction mixture were loaded onto Thin Layer Chromatography (TLC) plate with 1µl (20 mM) each of the reference molecules (glucose, cellobiose, maltotriose, cellopentaose and derivatized cellopentaose). The TLC plate was eluted in butanol/ethanol/H_2O (5/5/3) and visualized using orcinol solution (0.1% orcinol (w/v), 10% (v/v) sulfuric acid in absolute ethanol.

To determine the pH profile of □ -1,4-endoglucanase (*Bp*GH9*)* and cellobiohydrolase *(Bp*GH48*)*, enzyme assays were conducted at 37 °C within pH ranges 2–12. The pH was adjusted in a buffer reaction consisting of 50 mM NaH_2PO_4, 50 mM boric acid and 40 mM acetic acid. The enzyme was added in a pH buffer containing 1.25% (w/v) Na-CMC. Aliquots (70 **1**L) of reactions were withdrawn at t = 0, 15 min and 30 min and added to 70 **1**L of DNS solution (Miller 1959). The initial reaction velocity was measured by the reported O.D. difference per time.

Similarly, the optimum temperature for *Bp*GH9 was determined by conducting enzyme assay between 20 to 75 °C at pH 8.0 in 50 mM of citrate buffer with 1.25% (w/v) Na-CMC. In the same way, the pH stability for *Bp*GH9 was also determined by incubating the enzymes in a buffer reaction consisting of 50 mM NaH_2PO_4, 50 mM boric acid and 40 mM acetic acid at pH 10.0 and at 37 °C for t = 0 min,30 min, 60 min, 90 min and 120 min. The initial reaction velocity was then measured at pH 8.0 in 50 mM of citrate buffer with 1.25% (w/v) CMC at 37 °C.

Affinity Pull-Down assays

The functionality of the CBM3a module following its fusion to the bi-modular wild type endoglucanase was ascertained by the binding of the CBM to Avicel PH-101 (Stern et al. 2015). The enzymes (*Bp*GH9, CBM3a-*Bp*GH9, *Bp*GH9t and CBM3a-*Bp*GH9t), 3.5 µm each in 50 mM acetate buffer (pH 5.0), 12 mM $CaCl_2$, 2mM EDTA, were incubated at 4 °C for 2 h in the presence of 10% insoluble cellulose (Avicel PH-101). The mixture was then centrifuged for 2 min, the supernatant (S) recovered and the pellet washed in 10 mM Tris/HCl buffer, pH 8 containing 0.05% (v/v) Tween®20 in order to remove the non-specifically bound proteins. The mixture was again centrifuged for two minutes and the pellet (P) resuspended in 10 mM Tris/HCl buffer, pH 8. The fractions containing the pellets (P) and the supernatants (S) were denatured for 3 min at 95 °C in a loading buffer (2% (w/v) SDS, 0.08 M Tris–HCl, pH 6.8, 10% (v/v) glycerol, 0.06% (v/v) ß-mercaptoethanol) then loaded onto an SDS-PAGE gel containing 12% acrylamide under denaturing conditions.

Results

Screening Bogoria strains for alkaliphilic cellulases

Among the 95 isolates collected from Lake Bogoria (Kenya), 42 isolates were identified as neutrophilic CMCase producers (pH 7) on supplemented minimal salt media (MSM) using the Congo red plate assay.

Further, isolates were screened on supplemented, and non supplemented MSM (Supplementary Table 1) at higher pHs (pH 9 and 10) in order to identify isolates that had alkaliphilic activity. Among the 95 isolates screened, 45 and 23 isolates were identified as alkaliphilic CMCase producers at pH 9 and pH10 respectively. These isolates were then ranked based on their cellulolytic activity using the halo to colony diameter ratios in order to identify the best and highest cellulase producer (Fig. 1, Supplementary Fig. 1). The highest alkaliphilic producer A8 was then identified and used for further downstream processes. This isolate A8 was previously identified as a *Bacillus pumilus with* 99% rRNA 16S gene sequence identity (Moallic et al. 2006).

Cloning for wild type *Bp*GH9 and *Bp*GH48 protein expression

Based on the sequencing of rRNA 16S gene from A8 strain, *Bacillus pumilus* WP8 (CP010075.1) (Kang et al. 2015) was the closest strain to A8 (99.79%) for which the whole genome was sequenced. Its genome contains two cellulase genes (QR42_07950; QR42_07955) in the same gene cluster that was annotated, but not biochemically characterized, to cellulases from GH9 and GH48 families, respectively (Lombard et al. 2014). From their sequence, primers were designed for PCR amplification of *BpGH9* and *BpGH48* from strain A8 (Supplementary Table 2). After amplification and sequencing, the *BpGH9* and *BpGH48* genes obtained, coded for protein sequences with 623 and 710 aa with a molecular mass of 70.79 kDa and 79.59 kDa, respectively (Supplementary Table 3 and Supplementary Fig. 2). The translated *Bp*GH9 and *Bp*GH48 proteins showed more than 99% identity to several glycoside hydrolases from *Bacillus pumilus*. The nucleotide sequence of these two cellulases have been deposited in the NCBI Genbank database under the accession numbers KY441475.1 and KY441476.1 respectively. The closest biochem- ically characterized cellulases to *Bp*GH9 and *Bp*GH48 were the endoglucanase EglA (99% identity) (Lima et al. 2005) and a □1,4-beta-cellobiosidase from *Bacillus pumilus* (SAFR-032, 95% identity), respec- tively. A 3D structure was recently released for the GH48 cellobiosidase (PDB 5BV9), while the closest structural data to *Bp*GH9 was the 1,4□-endoglucanase CelA from *Caldicellulosiruptor bescii* (PDB 4DOE, 52% identity with the catalytic module) (Brunecky et al. 2013) and Cel9G from *C. cellulolyticum* (PDB 1G87, 49% identity) (Mandelman et al. 2003). These sequence homologies suggest that *Bp*GH9 has a modular structure containing a catalytic domain (residues 1–439) and a cellulose-binding domain (CBM3c) (Fig. 2 and Supplementary Fig. 7). On the other hand, *Bp*GH48 appears to be non-modular with only a catalytic domain.

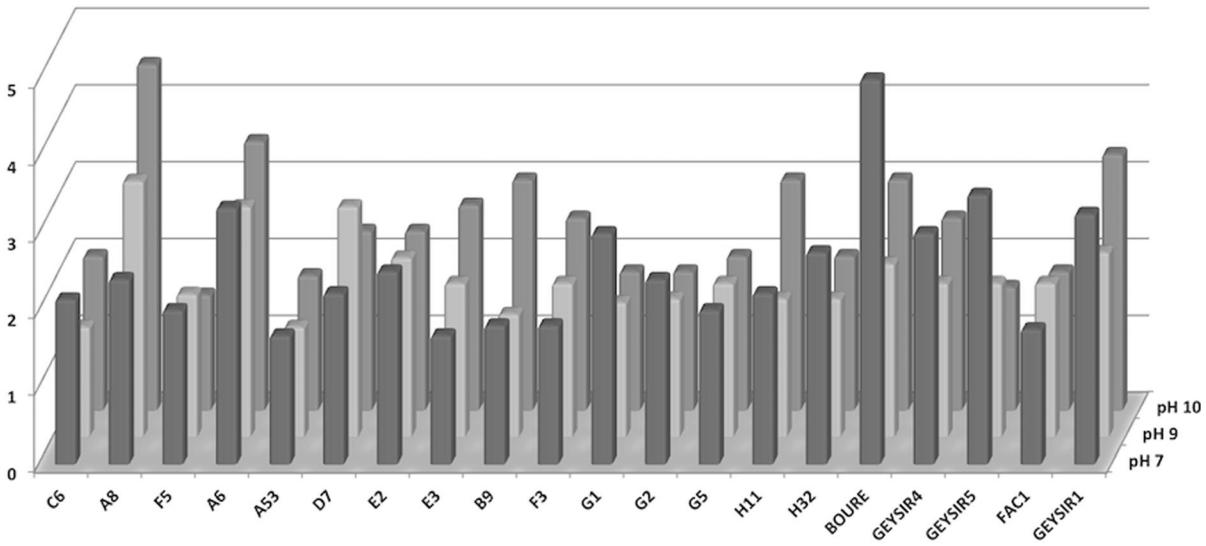


Fig. 1 Histogram of halo over colony size ratio at pH 7, 9 and 10 of cellulase producers, measured by Congo red plate screening to observe cellulase activity on minimal medium supplemented with yeast extract (0.5%) and CMC (1%)

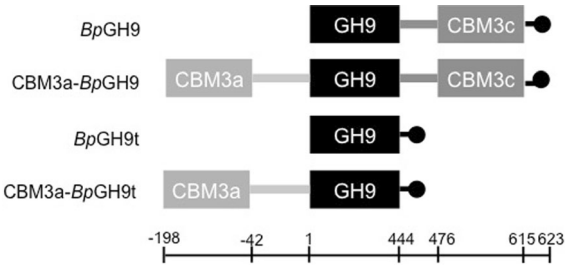


Fig. 2 Schematic diagram of the recombinants *Bp*GH9 proteins with a *C*-terminal His-tag added to all fusion proteins. Scales show the number of amino acid and the boundaries between the different domains

Enzymatic characterizations of the WT *Bp*GH9 and *Bp*GH48

The activities of both cellulases were tested on different cellulose substrates and results are summarized in Table 1. As predicted by the activity of its homologous enzymes, *Bp*GH9 is active on both crystalline cellulose substrates (Avicel) as well as on derivatized and amorphous pre-treated cellulose substrates (Na-CMC and PASC, respectively). In contrast, *Bp*GH48 has almost no activity on crystalline cellulose substrates (Avicel) as well as derivatized cellulose (Na-CMC) and pNP-cellobiose, but acts on amorphous cellulose (PASC). The major product of PASC hydrolysis by *Bp*GH48 was cellobiose, as revealed by TLC (data not shown), suggesting that it is an exo-cellobiohydrolase. Combination of *Bp*GH9 and *Bp*GH48 showed a significant synergistic effect of this pair of cellulases (Table 1).

To confirm the mode of action of *Bp*GH48, its activity was tested on cellopentaose and cellopentaose modified by reductive amination with aminobenzonitrile (Fig. 3). On free cellopentaose, the action of *Bp*GH48 led to the production of a mixture of cellobiose and cellotriose, and no further hydrolysis of cellotriose was observed since no glucose was detected (Fig. 3a). Minimal hydrolytic activity was detected on cellopentaose modified at the reductive end, which led to the liberation of cellotriose and a derivatized oligosaccharide (Fig. 3b). These results are in agreement with cellobiohydrolase activity, and further distinguishes *Bp*GH48 as a reducing end cellobiohydrolase (Cellobiohydrolase II). Consequently, there is a diminished activity when this reductive end is modified by reductive amination, and further explains why *Bp*GH48 does not hydrolyse pNP-cellobioside.

Screening of cellulolytic producers was done at high pH (7–10), this conferred an alkaliphilic environment for the extracted cellulases. We then determined the optimum pH for enzymatic activity of *Bp*GH9 and *Bp*GH48 cellulases (Fig. 4). The optimum pH of BpGH9 was determined at 6.0, where it retained 70% of its activity until pH 10 thus validating the screening process. Furthermore, its stability upon exposure to high pH is excellent since no decrease in activity was observed after 2h incubation at pH 10 **(Supplementary Fig. 5).** This pH dependence activity is quite different from that of EglA, a highly homologous enzyme (99%) to BpGH9, which showed much lower relative activity at high pH. In addition, *Bp*GH9 shows optimal activity at a temperature of 55 °C **(Supplementary Fig. 6).** On the other hand, *Bp*GH48 exhibited optimal activity at a pH of 5 and a low activity at high pH suggesting that this activity did not contribute to the screening process. This is in line with previous research that screening of

Table 1 Activities (**1**mol/min/mg) of BpGH9 and BpGH48 on different cellulosic substrates

Enzyme	CMC (1%)	PASC (1%)	Avicel (1%)	PNP-cellobiose (5 mM)
BpGH9	[a](3.0 ± 0.5)x10^5	(5.6 ± 1.0)x10^4	8.1 ± 1.0	nd
BpGH48	nd	36.8 ± 4.4	nd	nd
BpGH9 ? BpGH48	[b](5.3 ± 0.4)x10^5	(6.2 ± 0.8)x10^4	15.5 ± 0.4	nd

nd Not detected

[a]The activities were determined at pH 6 in a citrate buffer 50 mM at 37 °C, excepting with the Avicel substrate where the activities were measured at 50 °C

[b]The activities are expressed per mg of GH9. Both enzymes are in equimolar ratio

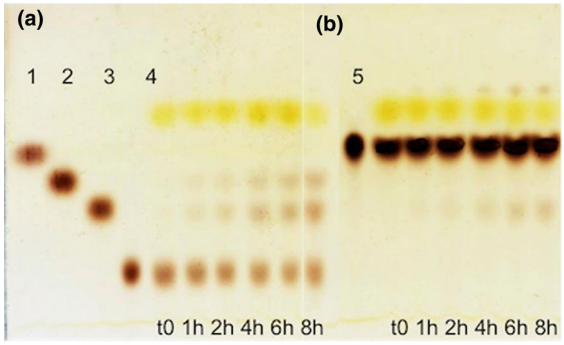


Fig. 3 Kinetics of hydrolysis of cellopentaose (4) (5 mM) hydrolysis (a) and ABN-cellopentaose (5) (b) by *Bp*GH48 at pH 8 as followed by TLC. Standards: glucose (1), cellobiose (2), cellotriose (3)

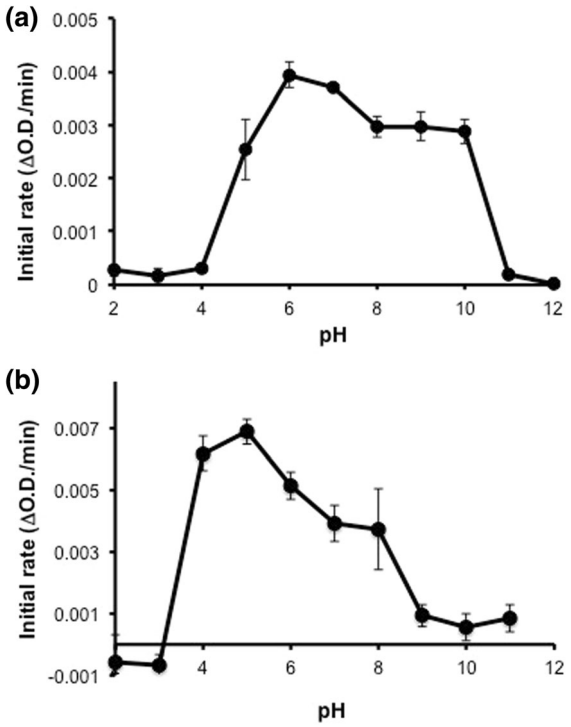


Fig. 4 pH activities profile (DO.D./min) of *Bp*GH9 (a) and *Bp*GH48 (b) on 1% CMC at 37 °C. Data points with standard deviations derive from triplicates

cellulase activity on CR plates exploits the endoglucanase activity.

Construction and characterization of a modular *BpGH9*

In order to improve the catalytic efficiency of *Bp*GH9 on crystalline cellulose, we added an exogenous cellulose binding domain to this endocellulase. The *BpGH9* gene was fused upstream with the IIIa cellulose binding module (CBM3a) gene from *C. thermocellum*, with a peptide linker sequence of 41 amino-acids (Fig. 2). This recombinant fusion protein (CBM3-*Bp*GH9) was expressed and purified by affinity chromatography using a His-tag sequence at the *C*-terminal end of *Bp*GH9. The resulting protein is 821 AAs long with a calculated molecular weight of 90.5 kDa, which corresponds to the size of the expressed protein as revealed by SDS-PAGE (Fig. 5).

The binding activity of the CBM3a in the fusion protein was examined by a pull down experiment on Avicel. Figure 5 shows that the CBM3a-*Bp*GH9 fusion was detected in the Avicel pellet, while *Bp*GH9 alone was not retained. The improved binding of the CBM3a-*Bp*GH9 to crystalline cellulose also had a positive effect on the cellulose hydrolysis since a twofold improvement in the level of degradation of Avicel was observed compared to the native *Bp*GH9 (Fig. 6a).

Two other constructs were tested to investigate the role of CBM3c and CBM3a on the catalytic activity of *Bp*GH9 (Fig. 2). A truncated form, *Bp*GH9t, composed only of the catalytic module and a modular construct with the CBM3a fused with the *Bp*GH9t, CBM3a-*Bp*GH9t, were expressed and purified as a single band in SDS-PAGE (Fig. 5).

The catalytic activity of these constructs was tested on Na-CMC substrate (Fig. 6b). Results revealed that *Bp*GH9t had reduced activity on the Na-CMC compared to the native enzyme *Bp*GH9. The fusion of CBM3a to the truncated form of *Bp*GH9t without the CBM3c however, resulted in a recovery of the activity. Furthermore, truncated constructs exhibited rapid inactivation at 37 °C (data not shown). This low thermal stability prevented us from assaying the activity on crystalline cellulose, which requires longer incubation time at 50 °C. The pull-down experiments with Avicel (Fig. 5) revealed that the truncated form of *Bp*GH9 exhibited no binding affinity to crystalline cellulose. However, adding the CBM3a to this truncated form enhanced the binding to cellulose.

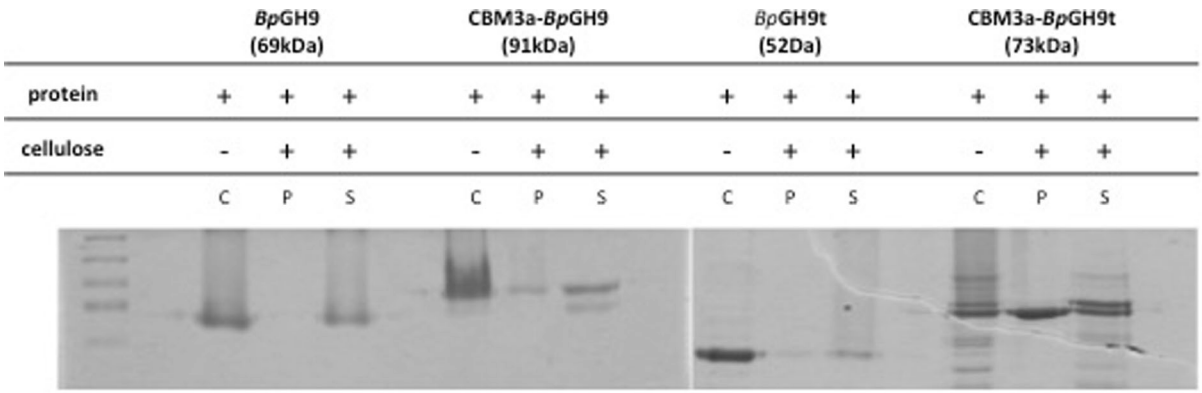


Fig. 5 Cellulose based affinity pull down assay for *Bp*GH9, CBM3a-*Bp*GH9, *Bp*GH9t and CBM3a-*Bp*GH9t over micro-crystalline cellulose (Avicel PH-101), observed by SDS-PAGE with Coomassie blue staining (*C*: control protein fraction without Avicel, *P*: pelleted fractions with Avicel, *S*: supernatant soluble fraction; MW; protein molecular weight marker).). 3.5 **1**M of each protein was incubated at 4 °C for 2 h in presence of 10% Avicel PH-101 and centrifuged before being separated by SDS-PAGE

Discussion

Microbial strains isolated from the Kenyan soda lakes, such lakes Bogoria, Magadi, Turkana and Nakuru, have attracted attention since they have the potential to produce alkaliphilic enzymes. This is due to the high alkalinity (pH 9) of their waters (Mwirichia et al. 2010). Indeed, several interesting alkaliphilic enzymes have been already characterized. Amylases active at pH 10 were extracted from a strain of *B. halodurans* (Hashim et al. 2004), lipases (Vargas et al. 2004), transpeptidases from a strain of *B. pumilus* (Moallic et al. 2006) and cellulases (Taylor et al. 2012). Some of these have found biotechnological applications in stain removal and colour preservation in textiles (Puradax HA, Genencor), or in the produc- tion of stonewashed effect on jeans (Indiage™, Genencor).

To further explore the diversity of the alkaliphilic cellulases, we identified a Bogoria strain of *Bacillus* sp., which expresses cellulolytic enzymes active at high pH (9–10). These cellulases were biochemically characterized after their cloning, which led to the identification of a new endo-cellulase from family GH9 (Cazy) and a new cellobiohydrolase from family GH48. Only the endocellulase BpGH9 was alkaliphilic and was shown to remain active up to pH 10. However, both enzymes are intracellular based on analysis with SignalP-5.0 (Almagro Armenteros et al. 2019), which suggests that the activity observed on the agar plates is probably a result of partial cell lysis.

The closest GH9 sequence to *Bp*GH9 that was biochemically characterized was EglA (99%) (Lima et al. 2005). This endo-cellulase was not alkaliphilic in contrast to *Bp*GH9. In addition, its translated sequence differs by only 8 amino acid residues (K40E, E86K, R191K, S371P, K511N, V553A, E604D, D605N). Based on the 3D model of *Bp*GH9 constructed by homology modelling (Supplementary Fig. 7), it appears that none of these mutations are located in the vicinity of the catalytic residues (D58 and E423), but they are located on the molecular surface. Therefore, it is difficult to infer the possible factors implicated in the alkaline adaptation of *Bp*GH9. Curiously, these mutations lead to a decreased charge residue content (K511N, D605N), which is in apparent contradiction with previous observations on alkaliphilic glycosidases (Zhao et al. 2011; Bai et al. 2015). However, the pH dependent activity curve of *Bp*GH9 compared to that of EglA, does not exhibit a significant shift in the optimum pH, but a better residual activity at high pH (8). Such small effects may involve a subtle change in the structure, which would require crystal structure analysis in order to be detected.

In nature, most cellulolytic enzymes are, either non-complexed cellulases such as those produced by aerobic fungi and most bacteria, or complexed cellu-lases organized in cellulosomes found in anaerobic bacteria and fungi. Most cellulases display a modular structure comprising a catalytic domain (GH) and accessory domains such as carbohydrate-binding

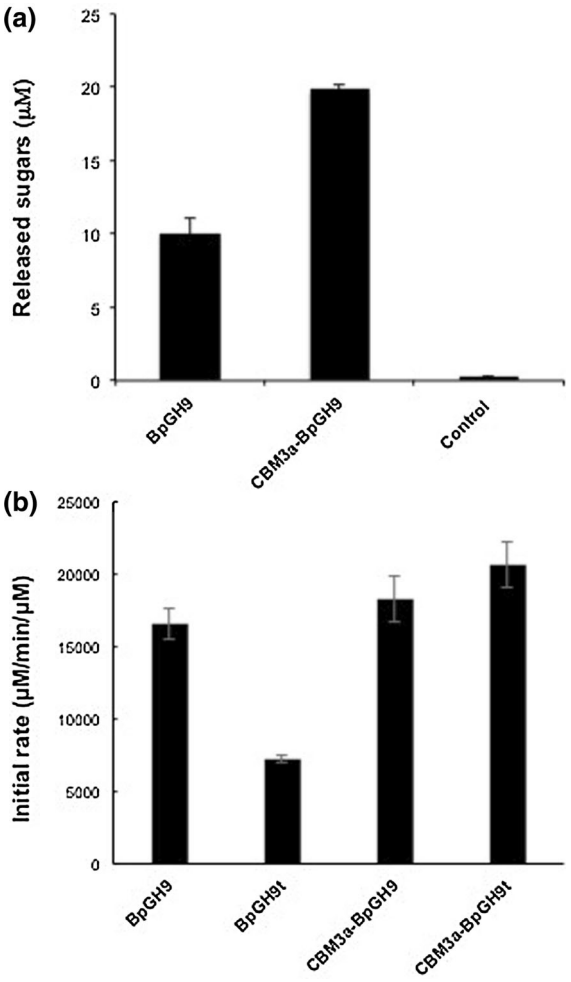


Fig. 6 Comparison of the enzymatic activities of *Bp*GH9, CBM3a-*Bp*GH9 on Avicel at 50 °C (a) and of *Bp*GH9t, CBM3a-*Bp*GH9t, *Bp*GH9, CBM3a-*Bp*GH9 on 1% Na-CMC at 20 °C (b). Enzymatic activities are defined as **1**M total reducing sugars liberated after 48 h on Avicel and **1**M total reducing sugars liberated per min and per **1**M enzyme on Na-CMC. Bar errors correspond to the standard deviations of triplicates

modules (CBMs). *Bp*GH9 from *B. pumilus* is a modular endoglucanase, that contains a catalytic domain and a *C*-terminal family 3c cellulose binding domain (CBM3c). This domain is known to be important for processivity by acting cooperatively with the catalytic domain (Chiriac et al. 2010).

In order to understand the role of the CBMs in the cellulose hydrolysis by *Bp*GH9, several truncated constructions of *Bp*GH9 without CBM3c and with an additional CBM3a from *C. thermocellum* were designed. As previously demonstrated in other GH9 endocellulases (Gilad et al. 2003; Petkun et al. 2015), removal of the CBM3c results in a strong reduction of the endoglucanase activity on soluble cellulose substrate (CMC). Regrettably, the effect of this truncation could not be explored on crystalline cellulose due to the lack of thermostability by the truncated enzyme (Kim et al. 2016). Moreover, the fusion of a CBM3a from *C. thermocellum* to *Bp*GH9 improves its activity on crystalline cellulose (Fig. 6a). Interestingly, in most natural endocellulases, CBMs are fused at the *C*-terminal position, but, here, the addition of a CBM3 on the *N*-terminus slightly enhances the enzyme activities on the soluble substrate CMC (Fig. 6b), in addition to increasing the activities on the insoluble substrates such as microcrystalline cellulose (Avicel). It is also interesting to note that the addition of the CBM3a module on the truncated *Bp*GH9t restores the catalytic activity to a level above to that of the native enzyme on amorphous cellulose (Na-CMC) (Fig. 6b). These data suggest that addition of CBM modules could modulate cellulolytic activity, whatever their positioning at either the *N*-terminus or the *C*-terminus of the catalytic domain. This fusion approach provides new routes to improve activity of endocellulases, which are naturally devoid of CBMs (Walker et al. 2015).

Acknowledgements LO was a recipient of a PhD scholarship (2014–2018) from the Embassy of France in Kenya, and of a stipendium (2015–2019), from the University of Nairobi. This work was supported by the Swedish International Development Cooperation Agency (Sida) sponsored Bioinnovate-Africa Program and UNEP GEF Agency Project No. 00876 on developing the microbial biotechnology industry from Kenya's Soda Lakes in line with the Nagoya Protocol.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

Ethical approval This article does not contain any studies with human participants or animals performed by any of the authors.

References

Almagro Armenteros JJ, Tsirigos KD, Sønderby CK et al (2019) SignalP 5.0 improves signal peptide predictions using deep neural networks. Nat Biotechnol 37:420–423

Bai W, Zhou C, Zhao Y, Wang Q, Ma Y, Permyakov EA (2015) Structural insight into and mutational analysis of family 11

xylanases: implications for mechanisms of higher pH cat- alytic adaptation. PLoS ONE 10:e0132834
Bayer EA, Belaich J-P, Shoham Y, Lamed R (2004) The cellulosomes: multienzyme machines for degradation of plant cell wall polysaccharides. Annu Rev Microbiol 58:521–554
Brunecky R, Alahuhta M, Xu Q et al (2013) Revealing nature's cellulase diversity: the digestion mechanism of *Caldicellulosiruptor bescii* CelA. Science 342:1513–1516
Chiriac AI, Cadena EM, Vidal T, Torres AL, Diaz P, Javier Pastor FI (2010) Engineering a family 9 processive endoglucanase from *Paenibacillus barcinonensis* display- ing a novel architecture. Appl Microbiol Biotechnol 86:1125–1134
Doi RH, Kosugi A (2004) Cellulosomes: plant-cell-wall-degrading enzyme complexes. Nat Rev Microbiol 2:541–551
Duckworth AW, Grant WD, Jones BE, Van Steenbergen R (1996) Phylogenetic diversity of soda lake alkaliphiles. FEMS Microbiol Ecol 19:181–191
Fontes CMGA, Gilbert HJ (2010) Cellulosomes: highly efficient nanomachines designed to deconstruct plant cell wall complex carbohydrates. Annu Rev Biochem 79:655–681
Gilad R, Rabinovich L, Yaron S et al (2003) CelI, a noncellulosomal family 9 enzyme from *Clostridium thermocellum*, is a processive endoglucanase that degrades crystalline cellulose. J Bacteriol 185:391–398
Hashim SO, Delgado O, Hatti-Kaul R, Mulaa FJ, Mattiasson B (2004) Starch hydrolysing *Bacillus halodurans* isolates from a Kenyan soda lake. Biotechnol Lett 26:823–828
Jahangeer S, Khan N, Jahangeer S et al (2005) Screening and characterization of fungal cellulases isolated from the native environmental source. Pak J Bot 37:739
Kang Y, Shen M, Wang H, Zhao Q (2015) Complete genome sequence of *Bacillus pumilus* strain WP8, an efficient plant growth-promoting rhizobacterium. Genome Announc 3:e01452-e1514
Kim S-J, Kim SH, Shin SK, Hyeon JE, Han SO (2016) Mutation of a conserved tryptophan residue in the CBM3c of a GH9 endoglucanase inhibits activity. Int J Biol Macromol 92:159–166
Kipper K, Väljamäe P, Johansson G (2005) Processive action of cellobiohydrolase Cel7A from *Trichoderma reesei* is revealed as ''burst''kinetics on fluorescent polymeric model substrates. Biochem J 385:527–535
Klemm D, Heublein B, Fink H-P, Bohn A (2005) Cellulose: fascinating biopolymer and sustainable raw material. Angew Chem Int Ed 44:3358–3393
Lima AOS, Quecine MC, Fungaro MHP et al (2005) Molecular characterization of a b-1,4-endoglucanase from an endo- phytic *Bacillus pumilus* strain. Appl Microbiol Biotechnol 68:57–65
Lombard V, Golaconda Ramulu H, Drula E, Coutinho PM, Henrissat B (2014) The carbohydrate-active enzymes database (CAZy) in 2013. Nucleic Acids Res 42:D490– D495
Mandelman D, Belaich A, Belaich JP, Aghajari N, Driguez H, Haser R (2003) X-Ray crystal structure of the multidomain endoglucanase Cel9G from *Clostridium cellulolyticum* complexed with natural and synthetic cello-oligosaccha- rides. J Bacteriol 185:4127–4135
Miller GL (1959) Use of dinitrosalicylic acid reagent for determination of reducing sugar. Anal Chem 31:426–428
Moallic C, Dabonne´ S, Colas B, Sine J-P (2006) Identification and characterization of a gamma-glutamyl transpeptidase from a thermo-alcalophile strain of *Bacillus pumilus*. Protein J 25:391–397
Mwirichia R, Muigai AW, Tindall B, Boga HI, Stackebrandt E (2010) Isolation and characterisation of bacteria from the haloalkaline Lake Elmenteita, Kenya. Extremophiles 14:339–348
Petkun S, Rozman Grinberg I, Lamed R et al (2015) Reassembly and co-crystallization of a family 9 processive endoglu- canase from its component parts: structural and functional significance of the intermodular linker. PeerJ 3:e1126
Stern J, Kahn A, Vazana Y et al (2015) Significance of relative position of cellulases in designer cellulosomes for opti- mized cellulolysis. PLoS ONE 10:e0127326
Taylor MP, van Zyl L, Tuffin M, Cowan D (2012) Extremophiles and biotechnology: how far have we come? In: Anitori RP (ed) Extremophiles: microbiology and biotechnology. Horizon Scientific Press, Hethersett, pp 1–24
Uchiyama T, Uchihashi T, Nakamura A, Watanabe H, Kaneko D, Samejima M, Igarashi K (2020) Convergent evolution of processivity in bacterial and fungal cellulases. Proc Natl Acad Sci 117:19896–19903
Vargas VA, Delgado OD, Hatti-Kaul R, Mattiasson B (2004) Lipase-producing microorganisms from a Kenyan alkaline soda lake. Biotechnol Lett 26:81–86
Walker JA, Takasuka TE, Deng K et al (2015) Multifunctional cellulase catalysis targeted by fusion to different carbo- hydrate-binding modules. Biotechnol Biofuels 8:220
Zhang X-Z, Zhang Y-HP (2013) Cellulases: characteristics, sources, production, and applications. Bioprocess Technol Biorefinery Sustain Prod Fuels Chem Polym 1:131–146
Zhao Y, Zhang Y, Cao Y et al (2011) Structural analysis of alkaline b-mannanase from alkaliphilic *Bacillus* sp. N16–5: implications for adaptation to alkaline conditions. PLoS ONE 6:e14608

Characterisation and engineering of two new GH9 and GH48 cellulases from a *Bacillus pumilus* isolated from Lake Bogoria

[1, 2,4] Lydia A. Ogonda. [2] Amélie Saumonneau. [3]Michel Dion. [1]Edward K. Muge. [5]Benson M. Wamalwa. [1]Francis J. Mulaa and [2]Charles Tellier *.

1. University of Nairobi Department of Biochemistry. School of Medicine. College of health sciences. P.O BOX. 30197 Nairobi. Kenya
2. Université de Nantes. CNRS. UFIP. UMR6286. F-44000 Nantes. France
3. Université de Nantes. IRS2. F-44000 Nantes. France
4. Masinde Muliro University of Science and Technology, Department of medical Biochemistry, School of medicine , P.O BOX 190-50100 Kakamega
5. University of Nairobi , Department of Chemistry , School of physical sciences , College of Biological and Physical Sciences , P.O BOX 30197-00100 Nairobi, Kenya

***Corresponding Author**

Tellier Charles. charles.tellier@univ-nantes.fr. 332 51125733. http://orcid.org/0000-0002-5788-2847

5.2 SUPPLEMENTARY MATERIALS:

Table 5.1 Composition of minimal agar media for diffusion plate assay

Table 5.2 Primers used to identified A8 strain and clone *BpGH9* and *BpGH48* cellulase genes

Table 5.3 Table of molecular weight and molar extinction coefficient of the different recombinant enzymes

Fig. 5.1 Results of the screening using Congo Red plate assay

Fig. 5.2 DNA and protein sequences of the cloned genes (*Bp*GH9, *Bp*GH48)

Fig. 5.3 SDS-PAGE gel of wild type a) *Bp*GH9 b) *Bp*GH48

Fig. 5.4 Analysis by 2100 Expert Protein 230 Bioanalyzer Agilent® of purified fractions of *Bp*GH48 and *Bp*GH9 recombinant proteins

Fig. 5.5 Residual activity of *Bp*GH9 on 1% CMC after different incubation times at pH 10.0 and 37 °C.

Fig. 5.6 Temperature effect on *Bp*GH9 activity on 1% CMC

Fig. 5.7 Model of the 3D structure of *Bp*GH9 obtained by homology modelling (Pymol align) with the structure of EglA (Lima et al. 2005). Amino acids in the sequence of *Bp*GH9 that differ from that of EglA are highlighted in red and the catalytic dyad in yellow.

Table 5.1: Composition of minimal agar media for diffusion plate assay

	Concentration % (w/v)	
Ingredients	**Minimal agar media**	**Supplemented Minimal agar media**
CMC	1	1
K_2HPO_4	0.08	0.08
KH_2PO_4	0.06	0.06
$(NH_4)_2SO_4$	0.1	0.1
$MgSO_4$	0.02	0.02
$CaCl_2.2H_2O$	0.005	0.005
NaCl	0.3	0.3
$FeCl_3$	0.0001	0.0001
Yeast Extract	-	0.1
Meat Extract	-	0.5
Agar	1.5	1.5

Table 5.2: primers used to identified A8 strain and cloned *Bp**GH9*** and *Bp**GH48*** cellulase genes

Primers for ARNr 16S gene:

534R: 5'ATTACCGCGGCTGCTGG3'

27F: 5'AGAGTTTGATCCTGGCTCAG3'

Primers for *BpGH9*** (Wp8 strain; QR42_07950) **and *Bp**GH48*** (Wp8 strain; QR42_07955) **gene cloning:**

*Bp*GH9F: 5'CGCCGGATCCATGGCATCTTACAACTATGTAGAGG3'

*Bp*GH9R: 5'ATGGCTCGAGTTTGTCTGGAAGTGTGCCAAATACTAAAGC3'

*Bp*GH48F: 5'CGCCGGATCCATGTCAATAAAAGAACGGTTTTTAACG3'

*Bp*GH48R: 5'ATGGCTCGAGGTTGATCAGACGATGATATTCAGC3'

Primers for Gibson assembling *BpGH9* and *BpGH48* cloning into pECa:

For_pECa_*Bp*GH48: 5'AGCAAATGGGTCGCGGATCCATGTCAAATAAAGAACGGTTTTTAACG3'

Rev_pECa_ *Bp*GH48: 5'TGGTGGTGGTGGTGCTCGAGGTTGATCAGACGATGATATTCAGC3'

For_pECa_*Bp*GH9: 5'AGCAAATGGGTCGCGGATCCATGGCATCTTACAACTATGTAGAGG3'

Rev_pECa_ *Bp*GH9: 5'TGGTGGTGGTGGTGCTCGAGTTTGTCTGGAAGTGTGCCAAATACTAAAGC3'

Primers for Gibson assembling of CBM3a-Linker-*Bp*GH9 into pECa:

pECa_CBMlink_For : 5'ATGGGTCGCGGATCCAATCTGAAAGTGGAGTTCTAC3'

pECa_CBMlink_Rev : 5'GTAGAACTCCACTTTCAGATTGGATCCGCGACCCAT3'

RevEndoCBM: 5'GTTGTAAG ATGCCATGGATCCCTCGAGTTTCAGATTGCCGCTAAC3'

ForCBMEndo: 5'GTTAGCGGCAATCTGAAACTCGAGGGATCCATGGCATCTTACAACTATGTAG3'

Specific Primers for Gibson deletion of CBM3a-*Bp*GH9t and *Bp*GH9t (without *CBM3c* endogenous) into pECa:

Rev_GH9t3'_pECa: 5'TCAGTGGTGGTGGTGGTGGTGCTCGAGTTCACCAAACAGCTGCACCATTTT3'

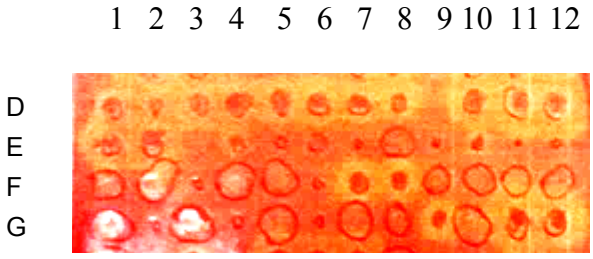
For_GH9t3'_pECa: 5'AAAATGGTGCAGCTGTTTGGTGAACTCGAGCACCACCACCACCACCACTGA3'

Table 5.3: Table of molecular weight and molar extinction coefficient of different recombinant enzymes.

	Molecular weight (kDa)	**[a]Molar extinction coefficient Ɛ ($M^{-1}.cm^{-1}$)**
*Bp*GH48	79.6	217990
*Bp*GH9	69	165255
*Bp*GH9t	52	142335
CBM3a-*Bp*GH9	90.5	200665
CBM3a-*Bp*GH9t	73	177745

[a.] **ε** were calculated calculated using http://atchimiebiologie.free.fr/proteinecalc/proteinecalc.html

Figure 5.1 Example of results of the screening using Congo plate assay at pH 7



	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**
D	D91	E1	E2	E3	B9	F3	F4	F4	F41	D71	F5	F5
E	G1	G2	G4	G5	G7	G8	G15	G24	G12	F6	H1	H2
F	H3	H4	H5	H6	H7	H8	H11	H32	C1	C2	C3	C4
G	C5	J1	J2	J3	J4	J5	J6	J7	J8	J9	BOURE	GEY4

Figure 5.2 DNA and protein sequences of the cloned genes (BpGH9, BpGH48)

***Bp*GH9 gene sequence (NCBI Genbank accession number KY441475.1)**

(Cloning sites, Stop and Start codons, His-Tag)

GGATCCATGGCATCTTACAACTATGTAGAGGTTTTACAAAAATCCATGCTGTTTTATGAAGC
GCAGCGGTCAGGCCGGCTTCCGGAAAGCAATCGTCTTAACTGGCGGGGAGATTCTGGACTA
GAGGACGGGAAGGATGTTGGGCATGACTTAACTGGCGGCTGGTATGATGCAGGAGATCAT
GTGAAATTTGGACTTCCGATGGCTTACTCAGCAGCCGTGCTTGCATGGACAGTATATGAAT
ACCGGGAAGCTTATGAAAAAGCAGAATTGCTTGATGAGATCTTAGATCAAATCAAGTGGGC
AACAGATTATTTTTTGAAAGCACATACAGGACCGAATGAATTTTGGGCGCAAGTTGGTGAT
GGAAACGCCGATCATGCTTGGTGGGGGCCAGCAGAAGTGATGCCGATGAACCGGCCGGCT
TTTAAAATTGATGAACATTGTCCTGGCACAGAAGTAGCTGCACAAACTGCGGCGGCTTTAG
CAGCAGGTTCTATCATTTTTAAAGAAACAGATGCGTCCTATGCAGCAAAGCTTTTAACACA
CGCCAAACAGCTATATGCATTTGCTGACAAGTATCGCGGCAAATATACAGATTGTGTCACC
AATGCGCAGCCATTTTACAACTCTTGGAGCGGCTATGTTGACGAACTTATTTGGGGTGGAA
TCTGGCTGTACTTGGCGACAAATGAAGAAACCTATTTAAACAAAGCATTAAAAGCGGTAGA
GGAATGGCCGAAGGATTGGGATTATACGTTTACCATGTCATGGGACAATACTTTTTTTGCTT
CACAAATCTTACTAGCAAGAATTACGAAAGAAAACAGATTTATAGAATCGACAGAGCGCA
ATCTTGATTACTGGACCACAGGTCTTGTTCAAAATGGAAAAGTAGAAAGAATCACCTATAC
GCCTGGCGGCTTGGCATGGCTGGATCAATGGGGTTCACTTCGTTATGCCGCCAATGCAGCA
TTTTTAGCCTTTGTCTATGCTGATTGGGTATCTGATCAAGAAAAGAAAAATCGATACCAATC
GTTTGCGATTAAGCAAACCCACTATATGTTAGGTGATAATCCACTGAATAGAAGCTACGTC
GTTGGGTTTGGCCAGAATCCGCCGAAGCACCCGCACCACCGTACTGCACATGGCTCATGGT
CTAACCAGCTGACAAATCCTCCATCTCACCGGCACACTCTTTATGGAGCGCTTGTAGGGGG
TCCTAATGCACAGGATCAATATGATGATGATATCTCTGATTATATATCTAACGAGGTGGCG
ACCGATTATAATGCCGCCTTTACTGGAAATATCGCCAAAATGGTGCAGCTGTTTGGTGAAG
GGCAATCAAAGCTGCCAAATTTCCCGCCTAAAGAACAAGTGGAGGATGAGTTTTTTGTAGA
GGCAGCTGTAATGCATAACGATACAACATCTACTCAAGTGAAAGCAGTGCTGTACAACAGG
TCCGGCTGGCCTGCAAGAAGCAGTCAAACACTATCCTTTAGATATTACGTCAATCTGAGCG
AGGTCTTTGCAAATGGATTCACTGAAAAGGATATTCAAGTGACAGCAGCCTACAATGAAGG
CGCTTCCTTATCGCCTTTAAAAGTATATGACGCATCAAGCCGCGTCTATTTTGCAGAAATCG
ATTTTACGGGCGTAGCCATTTCTCCTAGAGGAGAATCTGAGCATAAGAAGGAAATACAATT
TCGATTATCTGCTCCAAATGGATCGAATATATGGGATGCCTCAAATGATTATTCCTATCAAG
GATTAACATCCAATATGCAAAAAACAACAAAGATTCCTGTCTTTGACAATGGTGTTTTAGT
ATTTGGCACACTTCCAGACAAACTCGAGCACCACCACCACCACCACTGA

***Bp*GH9 protein sequence (NCBI Genbank accession number AVG70969)**

MASYNYVEVLQKSMLFYEAQRSGRLPESNRLNWRGDSGLEDGKDVGHDLTGGWYDAGDHV
KFGLPMAYSAAVLAWTVYEYREAYEKAELLDEILDQIKWATDYFLKAHTGPNEFWAQVGDG
NADHAWWGPAEVMPMNRPAFKIDEHCPGTEVAAQTAAALAAGSIIFKETDASYAAKLLTHAK
QLYAFADKYRGKYTDCVTNAQPFYNSWSGYVDELIWGGIWLYLATNEETYLNKALKAVEEW
PKDWDYTFTMSWDNTFFASQILLARITKENRFIESTERNLDYWTTGLVQNGKVERITYTPGGLA
WLDQWGSLRYAANAAFLAFVYADWVSDQEKKNRYQSFAIKQTHYMLGDNPLNRSYVVGFG
QNPPKHPHHRTAHGSWSNQLTNPPSHRHTLYGALVGGPNAQDQYDDDISDYISNEVATDYNA
AFTGNIAKMVQLFGEGQSKLPNFPPKEQVEDEFFVEAAVMHNDTTSTQVKAVLYNRSGWPAR
SSQTLSFRYYVNLSEVFANGFTEKDIQVTAAYNEGASLSPLKVYDASSRVYFAEIDFTGVAISPR
GESEHKKEIQFRLSAPNGSNIWDASNDYSYQGLTSNMQKTTKIPVFDNGVLVFGTLPDKLE

***BpGH48* gene sequence (NCBI Genbank accession number KY441476.1**

CGCGGATCCGCGACCCATGTCAAATAAAGAACGGTTTTTAACGCTTTATCATCAAATCAAA
AGCGATGCGAATGGATATTTTTCACCAGAAGGAATTCCGTATCATTCAATCGAGACACTCA
TTTGCGAAGCACCTGATTACGGGCATATGACAACTTCAGAAGCATATAGCTATTGGCTGTG
GCTTGAGGTGCTGTACGGACATTACACAGGGGACTGGAGCAAATTAGAGGCAGCTTGGGA
CAATATGGAGAAGTACATCATCCCAGTCAATGAAGATGGAAATGATGAGCAGCCGCATAT
GAACGCTTACAATCCGTCCAGCCCAGCCACATACGCAGCAGAGAAACCATTTCCGGATCAA
TACCCAAGCCAGTTGACCGGTGCTCGTCCTGCTGGCCAAGATCCAATTGATAGTGAATTGA
AATCAACCTATGGGACGAATGAAACGTATTTAATGCACTGGCTCCTTGACGTAGATAACTG
GTATGGCTATGGCAACTTGCTCAACCCATCACATACAGCCGCCTATGTGAACACATTCCAG
CGTGGTCCGCAAGAATCTGTGTGGGAGTCCATCCCTCATCCATCACAGGACGATAAATCTT
TTGGTAAGCCGAATGAAGGCTTTATGAGTTTGTTCACAAAAGAAAATCAAGCTCCAGCTGC
CCAGTGGCGCTATACGAATGCGACGGATGCTGATGCGCGCGCGATACAAGCGATGTATTGG
GCAAAGGAGCTCGGCTATAATCGGTCAGCTTATCTTGATAAAGCGAAAAAAATGGGGGATT
TCCTCCGTTACGGTATGTACGATAAATATTTTCAAACGATTGGAAGCGGGAAGCAGGGGAA
TCCGTATCCTGGAAACGGAAAGGGTGCCTGTCATTACTTAATGGCATGGTATACCTCTTGG
GGCGGAGGACTTGGGGAGTATGCAAACTGGTCATGGCGTATCGGAGCGAGCCATTGTCATC
AAGGATATCAAAATCCTGTCGCTGCTTATGCTTTATCCTCAGATAAAGGAGGCTTAAAGCC
ATCTTCAGCAACAGGGGCAAGCGATTGGGAAAAAACACTAAAGAGACAGCTTGAATTTTA
CGTATGGCTTCAATCGAAAGAAGGAGCCATAGCAGGAGGCGCAACCAATAGCTGGAATGG
AGACTATAGCGCTTATCCAGCAGGAAGAAGTACTTTTTATGATATGGCATACGAAGATGCA
CCTGTTTACCTTGACCCGCCTTCAAATAACTGGTTTGGAATGCAGGCGTGGCCAATGGAGC
GCGTAGCAGAGCTTTATTATATTTTTGTAAAAGATGGCGACAAAACATCTGAAAATGTCCA
AATGGCCAAATCTGTCATAACAAAATGGGTCAGCTATGCTTTAGATTACATTTTTATTGGCA
GCCGTCCTGTGTCAGATGAAGAGGGTTATTTCTTAGATGAGCAAGGACAGAGAATTCTAGG
AGGAACGAATGTTTCGGTTGCTACAACAAGTGCGCCTGGAGAGTTTTGGCTTCCTGGAAAT
ATTGCTTGGAGCGGACAGCCTGATACTTGGAACGGGTTCCAGTCTGCAACAGGAAATCCAA
ATTTAACAGCTGTAACTAAAGATCCGACTCAGGATACGGGCGTATTAGGAAGTCTAGCCAA
AGCTTTGATATTCTTTGCTGCAGCCACAAAGCAGGAAACAGGTGATTATACAGCGCTTGGT
GCAAGAGCGAAAGATGCAGCCGCGCAGCTCCTAGAGGTTGCATGGAATTACAATGATGGC
GTCGGCATAGTGACAGAGGAAGATAGAGAAGACTATGACCGTTTCTTTAGAAAAGAAGTG
TATTTCCCAAATGGATGGAACGGGACATTTGGTCAAGGAAATCAGATTCCTGGAGCCAGCA
CCATTCCTTCTGACCCGCAGCGAGGAGGGAATGGGGTCTATACGAGTTTTGCTGATCTTCGT
CCGAAGATTAAACAAGATCCAGCATGGTCTTCACTCGAAAGTAAATATCAGTCTTCATTTA
ATGAAGCTACAGGCAAATGGGAGAATGGCGCACCAGTTTTCACTTATCACCGCTTCTGGTC
TCAAGTGGATATGGCCACTGCCTATGCTGAATATCATCGTCTGATCAACCTCGAGCACCAC
CACCACCACCACTGA

***Bp*GH48 protein sequence (NCBI Genbank accession number AVG70970)**

MSNKERFLTLYHQIKSDANGYFSPEGIPYHSIETLICEAPDYGHMTTSEAYSYWLWLEVLYGHY
TGDWSKLEAAWDNMEKYIIPVNEDGNDEQPHMNAYNPSSPATYAAEKPFPDQYPSQLTGARP
AGQDPIDSELKSTYGTNETYLMHWLLDVDNWYGYGNLLNPSHTAAYVNTFQRGPQESVWESI
PHPSQDDKSFGKPNEGFMSLFTKENQAPAAQWRYTNATDADARAIQAMYWAKELGYNRSAY
LDKAKKMGDFLRYGMYDKYFQTIGSGKQGNPYPGNGKGACHYLMAWYTSWGGGLGEYAN
WSWRIGASHCHQGYQNPVAAYALSSDKGGLKPSSATGASDWEKTLKRQLEFYVWLQSKEGAI
AGGATNSWNGDYSAYPAGRSTFYDMAYEDAPVYLDPPSNNWFGMQAWPMERVAELYYIFVK
DGDKTSENVQMAKSVITKWVSYALDYIFIGSRPVSDEEGYFLDEQGQRILGGTNVSVATTSAPG
EFWLPGNIAWSGQPDTWNGFQSATGNPNLTAVTKDPTQDTGVLGSLAKALIFFAAATKQETGD

YTALGARAKDAAAQLLEVAWNYNDGVGIVTEEDREDYDRFFRKEVYFPNGWNGTFGQGNQIP
GASTIPSDPQRGGNGVYTSFADLRPKIKQDPAWSSLESKYQSSFNEATGKWENGAPVFTYHRF
WSQVDMATAYAEYHRLINLE

CBM3a-linker **gene sequence**

GGATCCAATCTGAAAGTGGAGTTCTACAATAGCAATCCGAGCGATACCACCAACAGCATCA
ATCCGCAGTTTAAGGTTACCAATACCGGCAGCAGCGCGATCGACCTGAGCAAGCTGACCCT
GCGTTACTATTACACCGTGGACGGTCAGAAAGATCAAACCTTCTGGTGCGATCACGCGGCG
ATCATTGGTAGCAACGGCAGCTATAACGGTATTACCAGCAACGTGAAGGGCACCTTTGTTA
AAATGAGCAGCAGCACCAACAACGCGGACACCTACCTGGAGATCAGCTTCACCGGTGGCA
CCCTGGAACCGGGTGCGCACGTTCAGATTCAAGGCCGTTTTGCGAAGAACGACTGGAGCAA
CTATACCCAGAGCAACGATTACAGCTTCAAAAGCGCGAGCCAGTTTGTGGAGTGGGATCAA
GTTACCGCGTATCTGAACGGTGTGCTGGTTTGGGGTCCGACCAACACCGCGACCCCGACCA
AGAGCGCGACCGCGACCCCGACCCGTCCGAGCGTTCCGACCAACACCCCGACCAACACCCC
GGCGAATACCCCGGTTAGCGGCAATCTGAAACTCGAG

CBM3a-linker **protein sequence**

GSNLKVEFYNSNPSDTTNSINPQFKVTNTGSSAIDLSKLTLRYYYTVDGQKDQTFWCDHAAIIGS
NGSYNGITSNVKGTFVKMSSSTNNADTYLEISFTGGTLEPGAHVQIQGRFAKNDWSNYTQSND
YSFKSASQFVEWDQVTAYLNGVLVWGPTNTATPTKSATATPTRPSVPTNTPTNTPANTPVSGN
LKLE

CBM3a-*Bp*GH9 protein: CBM3a-linker protein sequence in fusion with *Bp*GH9 protein (containing linker and CBM3c) and poly histidine tag

GSNLKVEFYNSNPSDTTNSINPQFKVTNTGSSAIDLSKLTLRYYYTVDGQKDQTFWCDHAAIIGS
NGSYNGITSNVKGTFVKMSSSTNNADTYLEISFTGGTLEPGAHVQIQGRFAKNDWSNYTQSND
YSFKSASQFVEWDQVTAYLNGVLVWGPTNTATPTKSATATPTRPSVPTNTPTNTPANTPVSGN
LKLEMASYNYVEVLQKSMLFYEAQRSGRLPESNRLNWRGDSGLEDGKDVGHDLTGGWYDAG
DHVKFGLPMAYSAAVLAWTVYEYREAYEKAELLDEILDQIKWATDYFLKAHTGPNEFWAQV
GDGNADHAWWGPAEVMPMNRPAFKIDEHCPGTEVAAQTAAALAAGSIIFKETDASYAAKLLT
HAKQLYAFADKYRGKYTDCVTNAQPFYNSWSGYVDELIWGGIWLYLATNEETYLNKALKAV
EEWPKDWDYTFTMSWDNTFFASQILLARITKENRFIESTERNLDYWTTGLVQNGKVERITYTPG
GLAWLDQWGSLRYAANAAFLAFVYADWVSDQEKKNRYQSFAIKQTHYMLGDNPLNRSYVV
GFGQNPPKHPHHRTAHGSWSNQLTNPPSHRHTLYGALVGGPNAQDQYDDDISDYISNEVATDY
NAAFTGNIAKMVQLFGEGQSKLPNFPPKEQVEDEFFVEAAVMHNDTTSTQVKAVLYNRSGWP
ARSSQTLSFRYYVNLSEVFANGFTEKDIQVTAAYNEGASLSPLKVYDASSRVYFAEIDFTGVAIS
PRGESEHKKEIQFRLSAPNGSNIWDASNDYSYQGLTSNMQKTTKIPVFDNGVLVFGTLPDKLEH
HHHHH

CBM3a-*Bp*GH9t protein: CBM3a-linker protein sequence in fusion with *Bp*GH9 protein (without linker and CBM3c) and poly histidine tag

GSNLKVEFYNSNPSDTTNSINPQFKVTNTGSSAIDLSKLTLRYYYTVDGQKDQTFWCDHAAIIGS
NGSYNGITSNVKGTFVKMSSSTNNADTYLEISFTGGTLEPGAHVQIQGRFAKNDWSNYTQSND
YSFKSASQFVEWDQVTAYLNGVLVWGPTNTATPTKSATATPTRPSVPTNTPTNTPANTPVSGN
LKLESYNYVEVLQKSMLFYEAQRSGRLPESNRLNWRGDSGLEDGKDVGHDLTGGWYDAGDH
VKFGLPMAYSAAVLAWTVYEYREAYEKAELLDEILDQIKWATDYFLKAHTGPNEFWAQVGD
GNADHAWWGPAEVMPMNRPAFKIDEHCPGTEVAAQTAAALAAGSIIFKETDASYAAKLLTHA
KQLYAFADKYRGKYTDCVTNAQPFYNSWSGYVDELIWGGIWLYLATNEETYLNKALKAVEE
WPKDWDYTFTMSWDNTFFASQILLARITKENRFIESTERNLDYWTTGLVQNGKVERITYTPGGL

AWLDQWGSLRYAANAAFLAFVYADWVSDQEKKNRYQSFAIKQTHYMLGDNPLNRSYVVGF
GQNPPKHPHHRTAHGSWSNQLTNPPSHRHTLYGALVGGPNAQDQYDDDISDYISNEVATDYN
AAFTGNIAKMVQLFGELEHHHHHH

***Bp*GH9t protein: *Bp*GH9 protein (without linker and CBM3c) and poly histidine tag**

MASYNYVEVLQKSMLFYEAQRSGRLPESNRLNWRGDSGLEDGKDVGHDLTGGWYDAGDHV
KFGLPMAYSAAVLAWTVYEYREAYEKAELLDEILDQIKWATDYFLKAHTGPNEFWAQVGDG
NADHAWWGPAEVMPMNRPAFKIDEHCPGTEVAAQTAAALAAGSIIFKETDASYAAKLLTHAK
QLYAFADKYRGKYTDCVTNAQPFYNSWSGYVDELIWGGIWLYLATNEETYLNKALKAVEEW
PKDWDYTFTMSWDNTFFASQILLARITKENRFIESTERNLDYWTTGLVQNGKVERITYTPGGLA
WLDQWGSLRYAANAAFLAFVYADWVSDQEKKNRYQSFAIKQTHYMLGDNPLNRSYVVGFG
QNPPKHPHHRTAHGSWSNQLTNPPSHRHTLYGALVGGPNAQDQYDDDISDYISNEVATDYNA
AFTGNIAKMVQLFGELEHHHHHH

Figure 5.3 SDS-PAGE gel of wild type a) *Bp*GH9 b) *Bp*GH48 before and after Ni-NTA purification. 1: Molecular weight marker; 2: Crude fraction; 3: Flow through; 4: Wash; 5 to 7: First to third elution fractions

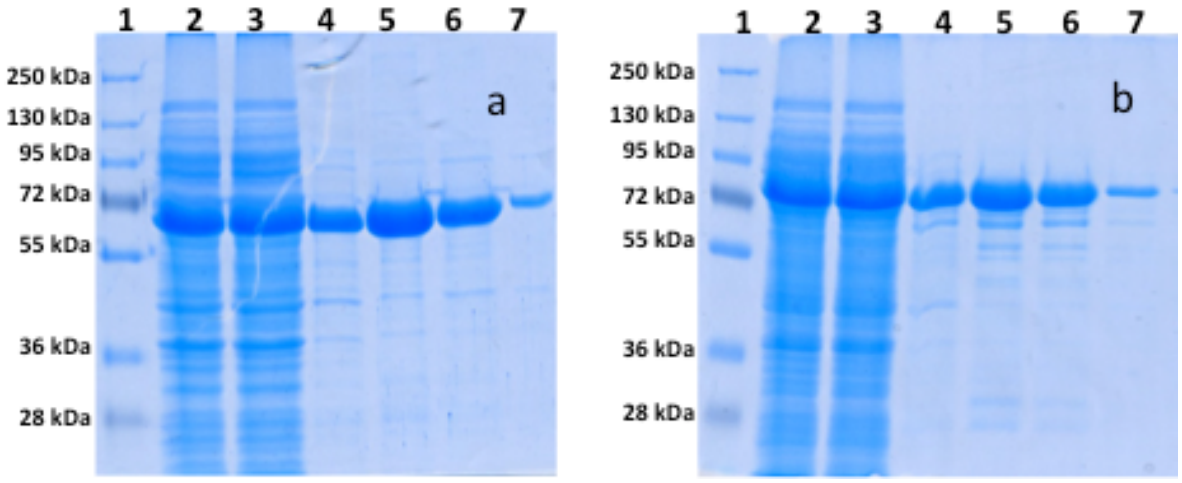


Figure 5.4 Analysis by 2100 Expert Protein 230 Bioanalyzer Agilent® of purified fractions of *Bp*GH48 and *Bp*GH9 recombinant proteins

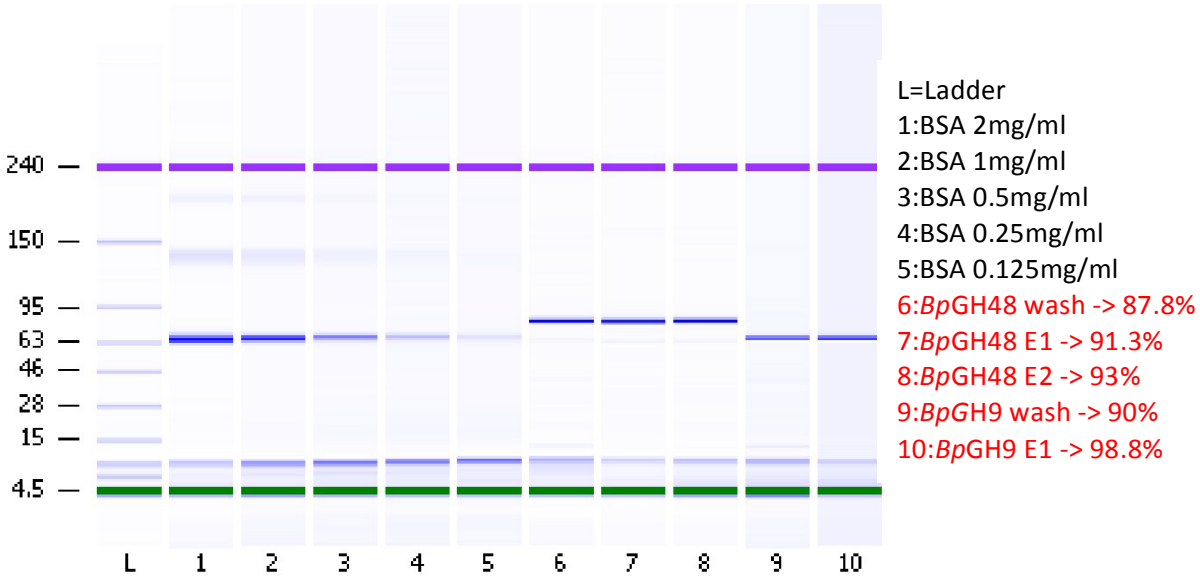


Figure 5.5 Residual activity of *Bp*GH9 on 1% CMC after different incubation times at pH 10.0 and 37 °C.

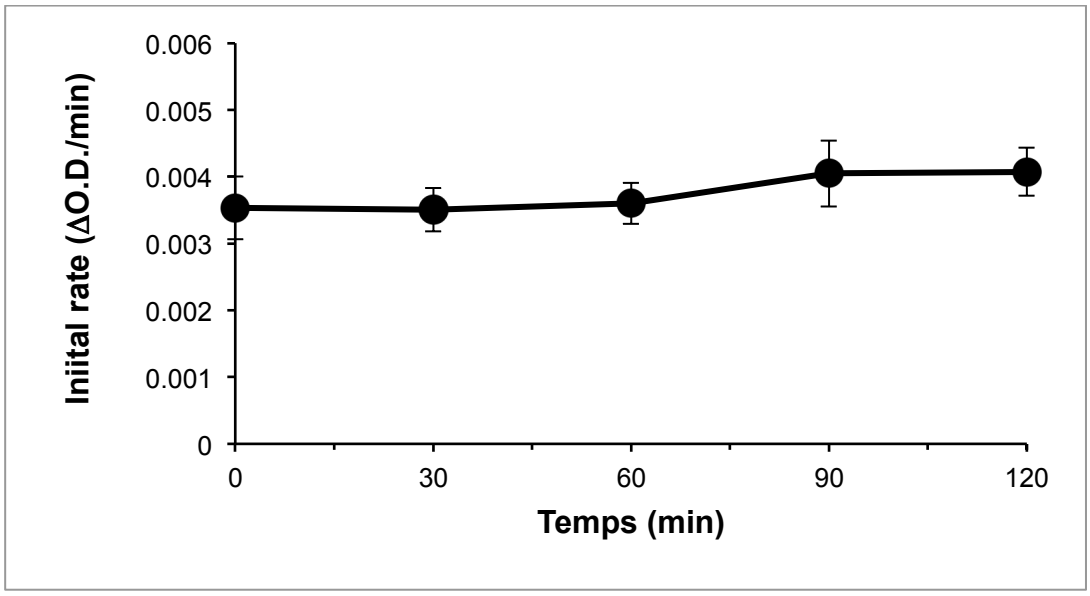


Figure 5.6 Temperature effect on *Bp*GH9 activity on 1% CMC

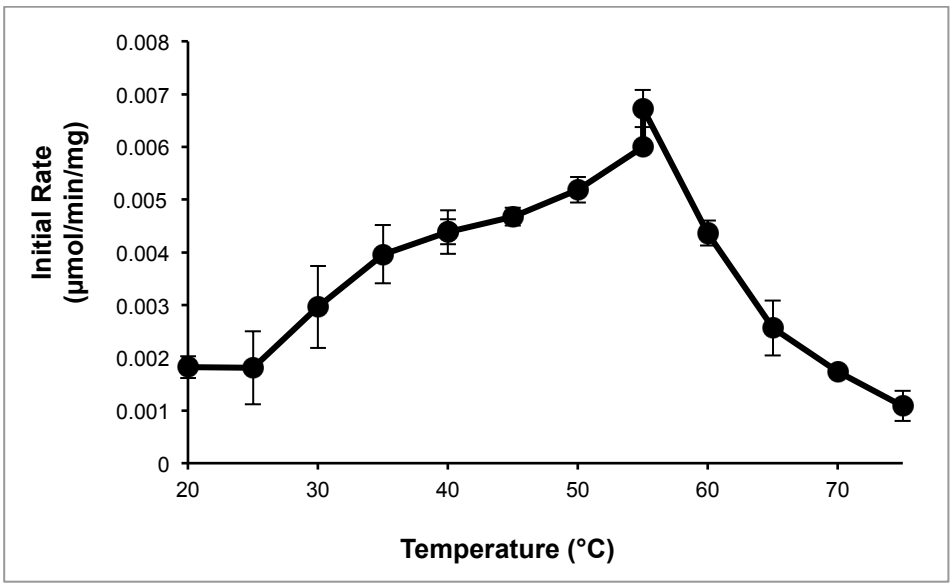


Figure 5.7 Model of the 3D structure of *Bp*GH9 obtained by homology modelling (Pymol Align) with the structure of EglA (Lima *et al.* 2005). Amino acids in the sequence of *Bp*GH9 that differ from that of EglA are highlighted in red and the catalytic dyad in yellow.

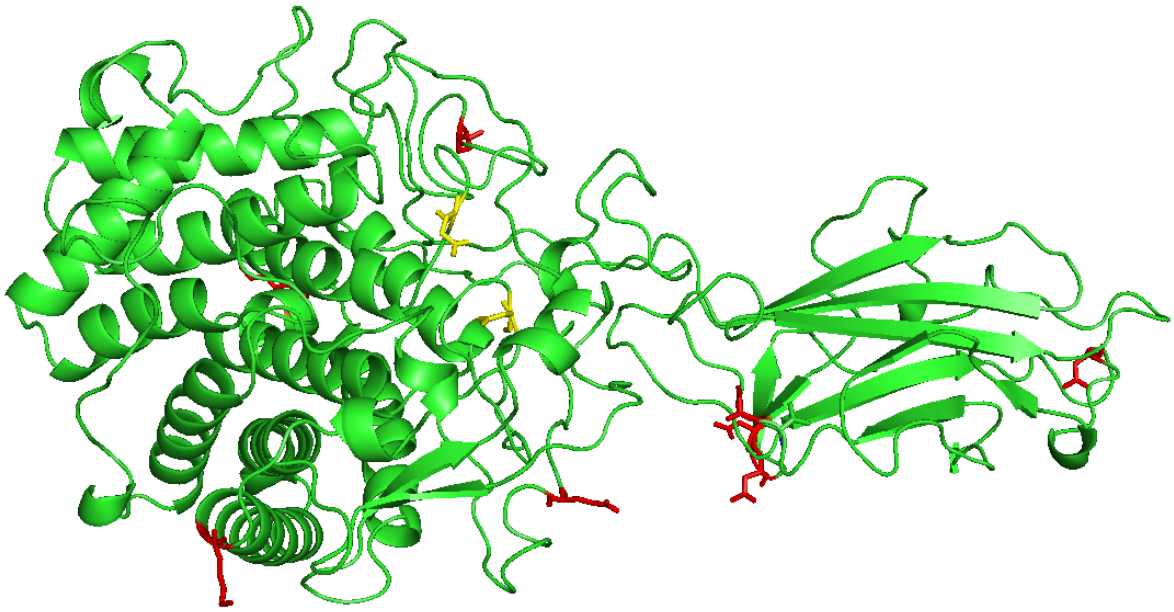


Table 5.4 showing the Growth and Cellulase producing potential of the screened microbial library

Sample	Microbe	Growth	Growth	Growth	Growth	Growth	Growth	R(H/C)	R (H/C)	R (H/C)	R (H/C)	R (H/C) -
Name	***Identity***	**LB**	**MM7Y**	**MM9Y**	**MM9Y-**	**MM10Y**	**MM10Y-**	**pH7**	**pH9Y+**	**pH9Y-**	**pH10Y+**	**pH10Y-**
B6	***B.pumilus***	+	+	+	+	+	+	++	++	++++	-	+++
C6	***ND***	+	+	+	+	+	+	++	+	++	+	+++
J4	***ND***	+	+	+	+	+	+	_	_	_	_	+++
D102	***B. halodurans***	+	+	+	-	+	+	_	_	_	_	+++
A8	***B.pumilus***	+	+	+	+	+	+	++	+++	++++	++++	+++
B9	***B.halodurans***	+	+	+	+	+	+	++	+++	++++	-	+++
F5	***ND***	+	+	+	+	+	+	+	+	+++	+	++
F5	***ND***	+	+	+	+	+	+	+	+	+++	++	++
A6	***B. licheniformis***	+	+	+	+	+	-	+++	++	++	+++	++
D71	***ND***	+	+	+	+	+	+	++	+	+++	-	+
D101	***ND***	+	+	+	-	+	+	-	-	-	-	+
C21	***B.licheniformis***	+	+	+	+	+	+	+++	+++	++++	-	+
G8	***B. flavothermus***	+	+	+	+	+	+	+	+	+	-	+
H4	***B. halodurans***	+	+	+	+	+	-	-	-	-	-	+
D71	***ND***	+	+	+	+	+	+	+	+	+++	-	+
Control	***Control***	-	-	-	-	-	-	-	-	-	-	-
A1	***B.pumilus***	+	-	-	+	+	+	-	-	-	-	-
A22	***ND***	+	-	-	-	-	+	-	-	-	-	-
A11	***ND***	+	-	+	+	+	+	-	-	-	-	-
AG	***ND***	+	-	-	+	+	+	-	-	-	-	-
A53	***B.licheniformis***	+	+	+	+	+	+	+	+	++	+	-
A72	***ND***	+	+	+	+	+	+	+	++	+++	-	-

A722	***ND***	+	-	-	+	+	+	-	-	-	-	-
A8	***B.pumilus***	+	+	+	+	+	+	+	++	+	-	-
A83	***ND***	+	+	+	+	+	-	+++	+	-	-	-
B11	***B. halodurans***	-	-	+	-	+	-	-	-	-	+++	-
B12	***ND***	-	-	-	-	-	+	-	-	-	-	-
I5	***B. licheniformis***	+	+	+	+	+	+	-	-	-	-	-
F4	***B. licheniformis***	+	-	+	+	+	+	-	-	-	-	-
C22	***B.licheniformis***	+	+	+	+	+	+	+++	++	-	-	-
C24	***ND***	+	+	+	+	+	+	++	+	-	-	-
C3	***B.pumilus***	+	+	+	+	+	+	-	+	+++	-	-
C4	***ND***	+	+	+	+	+	+	+	+	-	-	-
B9	***B. halodurans***	-	-	+	+	+	+	-	+	++++	-	-
C7	***ND***	+	+	+	+	+	+	++	++	++++	-	-
D10	***ND***	+	-	+	-	+	-	-	-	-	-	-
DG	***ND***	+	+	+	+	+	+	+	+	++	-	-
D7	***B. licheniformis***	+	+	+	+	+	+	++	++	++	++	-
D7G	***ND***	+	+	+	-	+	+	-	-	-	-	-
D91	***ND***	+	+	+	+	+	+	++	++	++	-	-
D9	***ND***	+	+	+	-	+	+	-	++++	-	-	-
D91	***ND***	+	+	+	+	+	+	++	++	-	-	-
E1	***ND***	+	+	+	+	+	-	+++	-	-	-	-
E2	***ND***	+	+	+	+	+	+	++	++	-	++	-
E3	***ND***	+	+	+	+	+	+	+	+	-	++	-
B9	***B.halodurans***	+	+	+	+	+	+	+	+	-	++	-
F3	***ND***	+	+	+	+	+	+	+	+	++	++	-
F4	***B. licheniformis***	+	+	+	+	+	+	+	+	-	-	-
F4	***B. licheniformis***	+	+	+	+	+	+	++	+	-	-	-
F41	***B. licheniformis***	+	-	+	-	+	+	-	-	-	-	-
G1	***ND***	+	+	+	+	+	+	++	+	+++	+	-
G2	***ND***	+	+	+	+	+	+	++	+	++	+	-

G4	***B. flavothermus***	+	-	-	-	+	+	-	-	-	+	-
G5	***B. flavothermus***	+	+	+	+	+	+	+	+	-	+	-
G7	***ND***	+	+	-	-	+	+	-	-	-	-	-
G15	***B. flavothermus***	+	+	+	-	+	+	-	-	-	-	-
G24	***ND***	+	+	+	+	+	+	-	-	-	-	-
G12	***ND***	+	+	+	-	+	+	-	-	-	-	-
F6	***B. licheniformis***	+	+	+	+	+	+	-	+	-	-	-
H1	***B. bogoria***	+	+	+	-	+	+	-	-	-	-	-
H2	***B. pseudofirmus***	+	+	+	-	+	+	-	-	-	-	-
H3	***ND***	+	+	+	+	+	+	-	-	-	-	-
H4	***B. halodurans***	+	+	+	+	+	+	-	-	-	-	-
H5	***ND***	+	+	+	+	+	+	-	-	-	-	-
H6	***B. crater lake***	+	+	+	+	+	+	-	-	-	-	-
H7	***ND***	+	+	+	+	+	+	-	-	-	-	-
H8	***A. crystallopoietes***	+	+	+	+	+	+	-	-	-	-	-
H11	***D. psychroalcalophilus***	+	+	+	+	+	+	++	+	-	++	-
H32	***ND***	+	+	+	+	+	+	++	+	-	+	-
C1	***ND***	+	+	+	+	+	+	-	-	-	-	-
C2	***ND***	+	+	+	+	+	+	-	-	-	-	-
C3	***ND***	+	+	+	+	+	+	-	-	-	-	-
C4	***ND***	+	+	+	+	+	+	-	-	-	-	-
C5	***ND***	+	+	+	+	+	+	-	-	+++	-	-
J1	***V. metschnikovii***	-	+	+	+	+	+	-	-	-	-	-
J2	***B. halodurans***	+	+	+	+	+	+	-	-	-	-	-
J3	***ND***	+	+	+	+	+	+	-	-	-	-	-
J5	***B. bogoria***	+	+	+	+	+	+	-	-	-	-	-
J6	***B. crater lake***	+	+	+	+	+	+	-	-	+++	-	-
J7	***B. bogoria***	+	+	+	+	+	+	-	-	-	-	-
J8	***B. halomonas***	+	+	+	+	+	+	+++	++	++++	-	-
J9	***B. bogoria***	+	+	+	+	+	+	-	+	-	-	-

BOURE	***ND***	+	+	+	+	+	+	++++	++	++++	++	-
GEYSIR4	***ND***	+	+	+	+	+	+	++	+	++	++	-
GEYSIR3	***ND***	+	+	+	+	+	+	-	-	+++	-	-
GEYSIR 2	***ND***	+	-	-	+	+	+	-	-	-	-	-
FAC2	***ND***	+	+	+	+	+	+	+	+	++	-	-
GEYSIR5	***ND***	+	+	+	+	+	+	+	-	-	-	-
GEYSIR5	***ND***	+	+	+	+	+	+	+++	+	+	+	-
FAC1	***ND***	+	+	+	+	+	+	+	+	-	+	-
GEYSIR1	***ND***	+	+	+	+	+	-	+++	++	++	+++	-
Control	***Control***	-	-	-	-	-	-	-	-	-	-	-
Control	***Control***	-	-	-	-	-	-	-	-	-	-	-
Control	***Control***	-	-	-	-	-	-	-	-	-	-	-
Control	***Control***	-	-	-	-	-	-	-	-	-	-	-
Control	***Control***	-	-	-	-	-	-	-	-	-	-	-

CHAPTER 6

Résumé de la publication III

Les microalgues sont une bio-ressource riche en molécules nutritives (protéines, vitamines, acide gras) ou métabolites ayant des applications médicales (spiruline) **(voir appendices)**. Certaines sont riches en composés hydrocarbonés qui pourraient se substituer aux combustibles fossiles, dans une approche de combustibles renouvelables et respectueux de l'environnement.

Botryococccus braunii est une microalgue verte qui a la capacité à accumuler dans la matrice entourant les colonies des hydrocarbures en C29, C31 ou C34. La structure des colonies productrices d'hydrocarbures de cette microalgue est constituée d'un ensemble de cellules associées par une matrice extracellulaire complexe (ECM) constituée (i) des parois cellulaires contenant un β-1,4/β-1,3 glucane proche de la cellulose, (ii) d'une matrice interne contenant des hydrocarbures réticulées dans laquelle se trouvent les hydrocarbures liquides d'intérêt, et (iii) d'une paroi fibrillaire entourant les colonies dans laquelle des polysaccharides de type arabino-galactanes sont majoritaires.

L'objectif de cette étude est de vérifier si un procédé utilisant des complexes enzymatiques ou multienzymatiques pourrait accroitre l'efficacité de l'extraction des hydrocarbures en attaquant la gangue protectrice des colonies. Pour ce travail, nous avons collaboré avec le laboratoire GEPEA de l'Université de Nantes, qui maitrise la culture de ces microalgues en photobioréacteurs.

Un panel de glycosidases allant de l'arabinanase, la galactanase à l'endoglucanase a été testé pour leur capacité à lyser les polysaccharides qui maintiennent l'intégrité de la colonie de *B. braunii*, afin de libérer les hydrocarbures présents dans la matrice extracellulaire, sans nuire aux développement des colonies. L'évolution de la taille des colonies après action enzymatique a été suivie par cytométrie en flux. Les résulltats indiquent que toutes ces enzymes sont plus ou moins capables de dissocier les cellules de la matrice extracellulaire tout en les maintenant en vie, suggérant la faisabilité d'une récupération in situ, semi-continue, d'hydrocarbures. L'endoglucanase BpGH9 de *Bacillus pumilus*, clonée précédemment, a été fusionnée avec CtCBM3a de *Clostridium thermocellum* et une protéine fluorescente jaune (YFP-CBM3a-BpGH9) pour sonder, par

microscopie de fluorescence, la présence de cellulose microcristalline dans la paroi cellulaire de *B. braunii* et pour augmenter l'efficacité de l'endoglucanase. Nous avons pu montrer que la partie apicale des cellules à la surface des colonies était bien marquée par la construction YFP-CBM3a-BpGH9 indiquant la présence de cellulose microcristalline dans la paroi des cellules.

Disruption of *Botryococcus braunii* colonies by glycoside hydrolases

Amélie Saumonneau[a], Nathan Lagneau[a], Lydia Awuor Ogonda[a,c], Catherine Dupré[b], Stéphanie Dutertre[d], Dominique Grizeau[b], Charles Tellier[a], Cyrille Grandjean[a], Franck Daligault[a]

Affiliations

[a]Nantes Université, CNRS US2B, UMR 6286,, F-44000 Nantes, France

[b]Nantes Université, Oniris, CNRS, GEPEA, UMR 6144, F-44600 Saint-Nazaire, France

[c]University of Nairobi Department of Biochemistry, School of Medicine, College of health sciences, P.O BOX, 30197 Nairobi, Kenya.

[d]MRic, Univ Rennes, CNRS, INSERM, BIOSIT (Biologie, Santé, Innovation Technologique de Rennes) - UMS 3480, US_S 018, F-35000 Rennes, France

Corresponding author:

Dr. Franck Daligault: franck.daligault@univ-nantes.fr

Abstract

Due to the deleterious environmental consequences of fossil-based fuels, research on renewable and environment friendly fuels has received a lot of interest these days. Microalgae are a promising alternative source of molecules previously produced from fossil resources. *Botryoccocus braunii* is a colonial green microalga having the ability to convert CO_2 by photosynthesis into long chain hydrocarbons. These are excreted and trapped in an extracellular matrix (ECM). A panel of glycosidases ranging from arabinanase, galactanase to endoglucanase were tested for their ability to lyse the polysaccharides which maintain the *B. braunii* colony integrity in order to release the hydrocarbons present in the extracellular matrix without harming the cells. The *Bp*GH9 endoglucanase from *Bacillus pumilus* was fused with *Ct*CBM3a from *Clostridium thermocellum* and yellow fluorescent protein to probe the presence of microcrystalline cellulose in the cell wall of *B. braunii* and to increase the efficacy of the endoglucanase. All the tested enzymes were more or less able to dissociate the cells

from the extracellular matrix while keeping them alive, suggesting the feasibility of a semi-continuous *in situ* recovery of hydrocarbons.

Keywords

Botryococcus braunii, microalga, glycosidase, flow cytometry, extracellular matrix, hydrocarbons.

1. Introduction

Microalgae and cyanobacteria were originally produced for food and animal feed, and have been cultivated in large scale for some decades, notably in Asia. They contain valuable compounds such as proteins, sugar, vitamins, pigments and fatty acids among other metabolites (Spolaore et al. 2006)(Borowitzka 2013)(Pulz and Gross 2004). The interest in green microalgae comes from their ability to convert carbon dioxide and sunlight to biomass and valuable products. If many microalgae produce useful proteins, carbohydrates and lipids, a few produce specific long-chain hydrocarbons such as botryococcens (Kumar et al. 2020). Some of these excreted molecules appear to have promising biological activities..(Oliva et al. 2022) (Sasaki et al. 2017).

Botryococcus braunii is a fresh water colonial oleaginous green microalga and is classified in four different races A, B, L and S, according to the chemical structure of the hydrocarbons it produces (Metzger and Largeau 2005)(Jin et al. 2016b). Although this microalga is known to deliver large amount of exopolysaccharides and carotenoids (Cheng et al. 2019), the interest in *B. braunii* rests mainly on hydrocarbons. The production of such energetically demanding hydrocarbons is supposed to be responsible to the low growth of *B. braunii* in comparison to other microalgae (Banerjee et al. 2002). Nevertheless, this microalga is considered as a great candidate for the production of liquid hydrocarbons that can represent up to 61% of dry biomass *e.g*. in race A (Jin et al. 2016b). In contrast to other oleaginous microalgae, hydrocarbons synthesized in *B. braunii* are mostly secreted in the extracellular space and not accumulated as lipid bodies in the cytoplasm (Hirose et al. 2013). However, these hydrocarbons of interest are retained in an extracellular matrix (ECM) consisting in three different components: the liquid hydrocarbon containing extracellular space, a colony retaining wall and a polysaccharide colony sheath (Weiss et al. 2012).

Because of this structural specificity and its slow growth, the exploitation of *B. braunii* is therefore not very profitable by conventional methods of hydrocarbon extraction

(Banerjee et al. 2002). The colony sheath and retaining wall have to be deconstructed in order to release the liquid hydrocarbons from the extracellular matrix of *B. braunii*. These components are mainly composed of carbohydrates (97%) where arabinose (42%) and galactose (39%) are predominant. The polysaccharide consists in a backbone made of galactose with mainly 1,2 linked side chains and ends composed of arabinofuranose. However the galactose in the backbone shows an unusual substitution pattern where linkages in position 2 and 3 are the most abundant, in comparison to the more common 1,4 or 1,6 linkages (Weiss et al. 2012). It can also be assumed that the *B. braunii* cells through their cell wall interact with the ECM, which prevent them from being efficiently released from the ECM. Thus targeting the cell wall polysaccharide consisting in β-1,4 and/or β-1,3 linked glucans in addition to the polysaccharide sheath is an appealing strategy to favor hydrocarbon release (Weiss et al. 2012).

The aim of this work was therefore to use glycosidases to hydrolyze the polysaccharides involved in the different components of a colony in a targeted and gentle manner in order to make accessible the hydrocarbons from the extracellular space. We demonstrate that specific endoglycanases (cellulases, arabinanases and galactanases) promote a significant disruption of the size of *B. Braunii* colonies. The most efficient disruptions was obtained with endocellulases that have been fused with carbohydrate binding module (CBM3a), which allowed the cellulases to target the crystalline cellulose of the cell walls.

2. Material and methods

2.1 Enzymes

Cellulase (C-0901) EC 232.734.4 from *Penicillium funicolosum*, (5 units / mg solid) and Hemicellulase H7649, EC 232.799.9 from *Aspergillus niger,* (1.5 units / mg solid) were purchased from Sigma-Aldrich® (France). endo-α1,5-arabinanase *(Bt*ABN43A) , endo-β1,4-galactanase *(Bs*Gan53A) and exo-β1,3-galactanase (*Ct*Gan43A) were acquired from Nzytech (Portugal).

2.2 Botryococcus braunii suspension cell culture

B. braunii Kützing SAG 30.81 race A was provided by the Culture Collection of Algae at Gottingen University (Sammlung von Algenkulturen, SAG). This race A strain was previously purified and studied (Jin et al. 2016a). The colonies were grown in the buffered AF6 culture medium (Provasoli and Pintner 1959) within Erlenmeyer flasks on

a rotating shaker (100 rpm). Some assays were conducted with colonies from a continuous culture within a labscale photobioreactor (0.35 L) at 20°C and under continuous illumination corresponding to a Photon Flux Density of 150 µmol m^{-2} s^{-1}. To prevent the interaction of possible exopolysaccharides with enzymatic acitvities, cell-colonies were previously rinsed and transferred to new buffered culture medium.

2.3 Gene cloning and YFP fusion

The *Bp*GH9 and CBM3a-*Bp*GH9 constructs containing an in frame C-terminal poly-histidine tag for purification by immobilized ion metal affinity chromatography (IMAC) were obtained as described previously (Ogonda et al. 2021). For *in vitro* studies and imaging, the YFP gene was inserted at the 5' extremity of the constructs (Figure S2) in the original pECa plasmid used in the previous study. These constructs were assembled using Gibson Assembly® protocol (New England BioLabs®Inc) with recombinant primers (Figure S3) to amplify by PCR (Polymerase Chain Reaction) the YFP gene and to insert it in the pECa already containing the *Bp*GH9 and CBM3a-*Bp*GH9 genes. The successful constructs were then confirmed by colony PCR and sequencing.

2.4 Production and purification of YFP-BpGH9 and YFP-CBM3a-BpGH9 recombinant enzymes

Recombinant strains of *E. coli* BL21(DE3) expressing the *Bp*GH9 variants (YFP-*Bp*GH9 or YFP-CBM3a-*Bp*GH9) were grown in 1 L of LB medium supplemented with ampicillin resistance antibiotic at 37°C to 0.6 of OD_{600nm} (Optical Density measured at λ=600nm) and IPTG at 0.5 mM was added to induce the expression at 30°C overnight under stirring (200 rpm). Then the culture was centrifuged and the pellet was frozen and resuspended in 20 ml of lysis buffer 20mM Tris-HCl pH 8.0 containing 10mM Imidazole, 5mM DNaseI and 1mg/ml lysozyme and protease inhibitors cocktail (Sigma-Aldrich®). After sonication and centrifugation, the recombinant proteins with His-tag contained in supernatant (soluble fraction) were purified IMAC: 300 µl of Ni-NTA Superflow (Qiagen), preliminary equilibrated into 20mM Tris-HCl pH 8.0, 10 mM imidazole, was added to the supernatant and stirred for 1 h at 4°C, then loaded onto a manual column. The column was washed using 25 mL of washing buffer (20mM Tris-HCl pH 8.0, 25 mM imidazole), then 5-fold 500µl of elution buffer was added (20mM Tris-HCl pH 8.0, 250 mM imidazole). The purity of the final product was checked by sodium dodecyl sulfate-polyacrylamide 12% gel (SDS-PaGE) electrophoresis. Enzyme concentrations were determined by UV absorbance at 280 nm using a NanoDrop 1000

(ThermoScientific) with the protein extinction coefficient (ε) calculated using http://atchimiebiologie.free.fr/proteinecalc/proteinecalc.html.

2.5 Enzymatic activity characterization

To assess the enzymatic activities of *Bp*GH9 constructs (YFP-*Bp*GH9 and YFP-CBM3a-*Bp*GH9), enzymatic assays were conducted using the insoluble microcrystalline cellulosic substrate Avicel PH-101 (Sigma-Aldrich®). The enzymatic activities were tested at an enzyme concentration of 3µM in 10mM Tris/HCl Buffer pH 8.0 and with 5% (w/v) of Avicel at 37°C during 24h. For the assays, 70µl of the reaction medium were withdrawn at regular time intervals and added to 70µl of DNS (Dinitro salicylic acid solution). The mixture was heated at 95°C for 10 minutes (Miller 1959). The OD was then read at 540nm by spectrophotometry (Magellan, Tecan®), reflecting the reducing end of released sugars.

2.6 Affinity pull down assay

The functionality of the CBM3a module following fusion to the bi-modular wild type endoglucanase was ascertained by binding of the CBM to the microcrystalline cellulosic substrate Avicel PH-101 according to the method by Stern et al. 2015 with modifications. The enzymes (YFP-*Bp*GH9 and YFP-CBM3a-*Bp*GH9) at 5µM were placed under orbital stirring (200 rpm /4°C) for two hours in presence of 5% (w/v) insoluble cellulose (Avicel). The mixture was then centrifuged for two minutes on a bench-top mini-centrifuge (Extragen). The supernatant was then recovered. The pellet was washed with 10 mM Tris/HCl buffer, pH8.0 containing 0.05% Tween®20 (Sigma) 0.05% (v/v) in order to remove the non-specifically bound proteins. The mixture was centrifuged again for two minutes on a bench-top mini-centrifuge. The supernatant was eliminated and the pellet was taken up with 10 mM Tris/HCl buffer, pH8.0. The fractions containing the pellets (P) and the supernatants (S) were denatured for 3 minutes at 95°C in loading buffer (SDS 2% (w/v); 0.08 M Tris-HCl; pH 6.8; 10% glycerol (v/v), ß-mercaptoethanol 0.06% (v/v)) then loaded onto an SDS-PAGE gel containing 12% acrylamide under denaturing conditions.

2.7 Fluorescence microscopy

B. braunii cell suspension was incubated with a 10µM enzyme solution of YFP-*Bp*GH9 or YFP-CBM3a-*Bp*GH9 with a ratio 2:1 (v/v). The mix loaded between blade and coverslip was observed at ×40 lens with optical microscope at optical fiber Nikon Ni-E,

in transmission light to observe the colonies. A TRITC filter (excitation: 543/22 nm; emission: 593/40 nm) was used to observe chlorophyll auto-fluorescence and a FITC filter (excitation: 482/35 nm; emission: 536/40 nm) was used to observe YFP fluorescence. The pictures were analyzed with Nis-Elements Viewer software

(https://www.nikoninstruments.com/fr_FR/Produits/Logiciels/NIS-Elements-Advanced-Research/NIS-Elements-Viewer).

2.8 Cytometry

Cytometric analyses were carried out on a Beckman Coulter CytoFLEX flow cytometer on the same cell suspension treated with *Bp*GH9 constucts and observed in fluorescent microscopy (see 2.7). For other enzymes 150µL of microalgal cell suspension were treated with 75µl of hemicellulase (20mg/ml, Merck-Sigma®), cellulase (20mg/ml, Merck-Sigma®), *Ct*Gan43A (2mg/ml, nzytech®), *Bs*Gan53A (1mg/ml, Nzytech®) and *Bt*Abn43A (1mg/ml, nzytech®). The measurements of frontal diffusion ("Forward SCatter" or FSC) and lateral diffusion ("Side SCatter" or SSC) were taken with a blue laser at 488nm. A 660nm filter was used to detect the fluorescence emission of chlorophyll in the microalgae. The flow rate was fixed at 10µL/min and the recording was done on 1000 events.

3. Results and discussion

B. braunii is a microalga which forms colonies in which cells are embedded in an ECM mainly composed of polymerized hydrocarbons in which are trapped the liquid hydrocarbons excreted by the cells (Hirose et al. 2013). By separating the cells from the ECM one can assume that extraction of the hydrocarbons should be feasible while keeping the cells alive for continuing the microalgae culture, making possible to envisage the development of a semi-continuous process. Such semi-continuous process cannot be considered with most of microalgae which must be killed prior to recover their hydrocarbon production. Indeed, lipids usually accumulate in lipid bodies inside the cell.

The cell wall, the retaining wall and the colony sheath, all are composed of polysaccharides, whose partial hydrolysis is a potential strategy for releasing the hydrocarbons from the colonies. To this aim, different glycosidases were chosen according to the carbohydrate composition of these colony components. On the one hand, arabinanases and galactanases are obvious candidates to be active on the retaining wall and colony sheath which consist of arabinose and galactose residues.

Therefore the endo-α1,5-arabinanase from *Bacteroides thetaiotaomicron* (*Bt*Abn43A) (Cartmell et al. 2011), the endo-β1,4-galactanase from *Bacillus subtilis* (*Bs*GAn53A) (Nakano et al. 1990) and the exo-β1,3-galactanase from *Clostridium thermocellum* (*Ct*Gan43A) (Ichinose et al. 2006), (Jiang et al. 2012) were selected and tested. On the another hand cellulases and hemicellulases are expected to be active on the cell wall made of β1,3 and β1,4-glucans, so commercial cellulases and hemicellulases and the cloned cellulase from *Bacillus pumilus* (*Bp*GH9)(Ogonda et al. 2021)) were included in the assay.

3.1 Binding of BpGH9 cellulase on Botyrococcus braunii

The endo-β-1,4-D-glucanase *Bp*GH9 from *B. pumilus* (NCBI Gene databank AVG70969) was chosen in order to genetically fuse this cellulase with various functional modules for detection (YFP) or binding (CBM3a). This enzyme was previously isolated and identified in our laboratory from a new strain of *B. pumilus* collected from the hot springs of Lake Bogoria in Kenya. It consists of a catalytic domain flanked by a CBM3c at its C-terminal end. It was previously shown that this enzyme was very tolerant to fusion at either the N-terminus or the C-terminus of the catalytic domain (Ogonda et al. 2021).

3.1.1 Fusion of BpGH9 with YFP and CBM3a

The *Bp*GH9 was shown to be active on cellulose substrates (carboxymethylcellulose, Na-CMC and phosphoric acid swollen cellulose, PASC) but very weakly on crystalline cellulose (Avicel) despite the presence of the CBM3c module in its structure (Ogonda et al. 2021). To our knowledge the presence of crystalline cellulose in the cell wall of *B. braunii* has not been checked but has been identified in different green microalgae. If such crystalline cellulose exists in the cell wall of *B. braunii*, the fusion of *Bp*GH9 with the CBM3a from *Clostridium thermocellum* (*Ct*CBM3a), known to have high affinity for such cellulose, should improve the *Bp*GH9 activity by fixing the enzyme to its polymeric substrate.

Moreover, since the discovery of the green fluorescent protein (GFP) from the jellysfish *Aequorea victoria*, fluorescent proteins (FP) have been largely used for biotechnological applications such as cell imaging and protein interaction studies. A set of GFP variants that are fluorescing along the visible spectrum is now available like BFP (blue), CFP (cyan) and YFP (yellow). In the carbohydrate field fluorescent proteins have been fused with CBM for localizing exopolysaccharides, controlling the

accessibility of cellulose to cellulases and hemicelullases, or imaging complex carbohydrates (Kawakubo et al. 2010).

To test the presence of microcrystalline cellulose and the effect of *Ct*CBM3a on the *Bp*GH9 activity, two constructs comprising fused YFP and *Bp*GH9 with (YFP-CBM3a-*Bp*GH9) or without (YFP-*Bp*GH9) *Ct*CBM3a were expressed in *E. coli* (Figure S2).

3.1.2 Enzymatic and affinity pull down assays on Avicel

Before testing the *Bp*GH9 constructs on *B. braunii* cells, it was necessary to verify that adding YFP to CBM3a-*Bp*GH9 and *Bp*GH9 did not impair the fixation to crystalline cellulose nor the endocellulase activity.

A pull-down assay was used to show the affinity of different constructs of *Bp*GH9 for insoluble crystalline microcellulose (Avicel PH-101). When YFP-*Bp*GH9 was mixed with Avicel (Figure 1), most of the protein was found in the supernatant, showing that YFP-*Bp*GH9 possesses low affinity for Avicel. In contrast the YFP-CBM3a-*Bp*GH9 construct was located mainly in the pellet fraction, demonstrating good affinity for insoluble microcrystalline cellulose. Thus adding YFP to the *Bp*GH9 constructs did not modify the affinity of the original constructs (Ogonda et al. 2021).

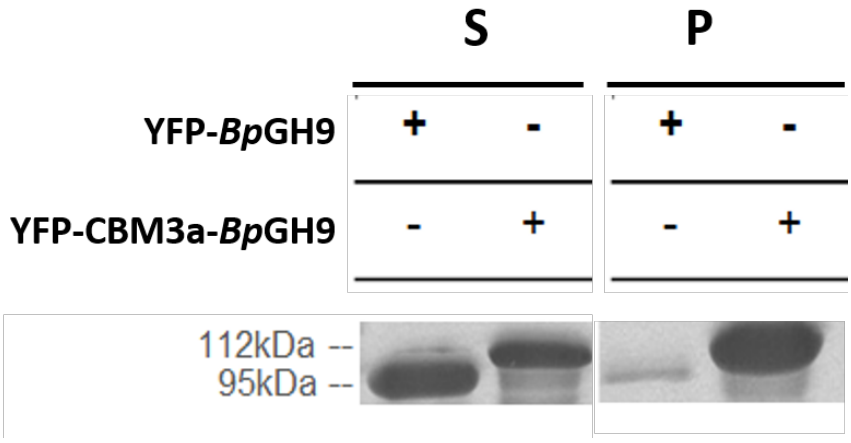


Fig. 1. Cellulose based affinity pull down assay for YFP-*Bp*GH9 (95 kDa) and YFP-CBM3a-*Bp*GH9 (112 kDa) over microcrystalline cellulose (Avicel), observed by SDS-PAGE with Coomassie Blue staining (**P**: pelleted fractions with Avicel, **S**: supernatant soluble fractions **+/-**; Presence /Absence of avicel; microcrystalline cellulose).

The activity of the constructs was determined on Avicel for checking that addition of YFP does not impair with the enzymatic activity and the effect of CBM3a. As shown on

figure 2, the YFP-*Bp*GH9 construct retains activity and the activity is increased when fused with *Ct*CBM3a (YFP-CBM3a-*Bp*GH9).

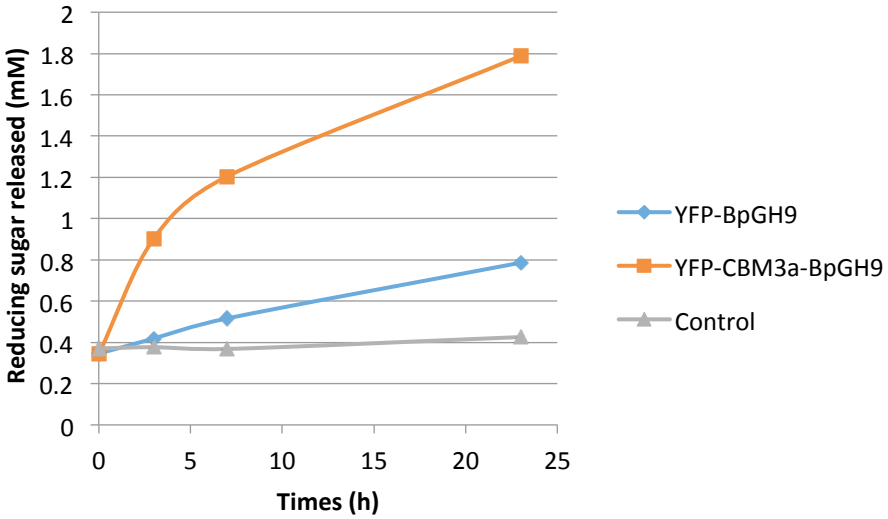


Fig. 2. Comparison of the enzymatic activities of YFP-*Bp*GH9 and YFP-CBM3a-*Bp*GH9 on Avicel at 37°C by measuring the released reducing sugar in function of time (hours), compared with a negative control (without enzyme).

3.1.3 Fluorescence microscopy

The YFP-CBM3a-*Bp*GH9 and YFP-*Bp*GH9 constructs were incubated with *Botyrococcus braunii* cell suspension and analyzed by fluorescence microscopy to verify the effect of *Ct*CBM3a on the fixation of the construct to *B. braunii* cells and thus highlight the presence of microcrystalline cellulose. By comparing pictures 3c and 3d corresponding to the YFP signal of the *Bp*GH9 constructs with and without fused *Ct*CBM3a respectively, it appears that YFP-CBM3a-*Bp*GH9 is localized at the cell surface (see arrow and square box in figure 3c), suggesting binding to crystalline cellulose present in the cell wall. In absence of *Ct*CBM3a the localization seems much more diffuse and may be due to weaker unspecific binding or binding of the native CBM3c present in *Bp*GH9 to amorphous or soluble cellulose (Figure 3d).

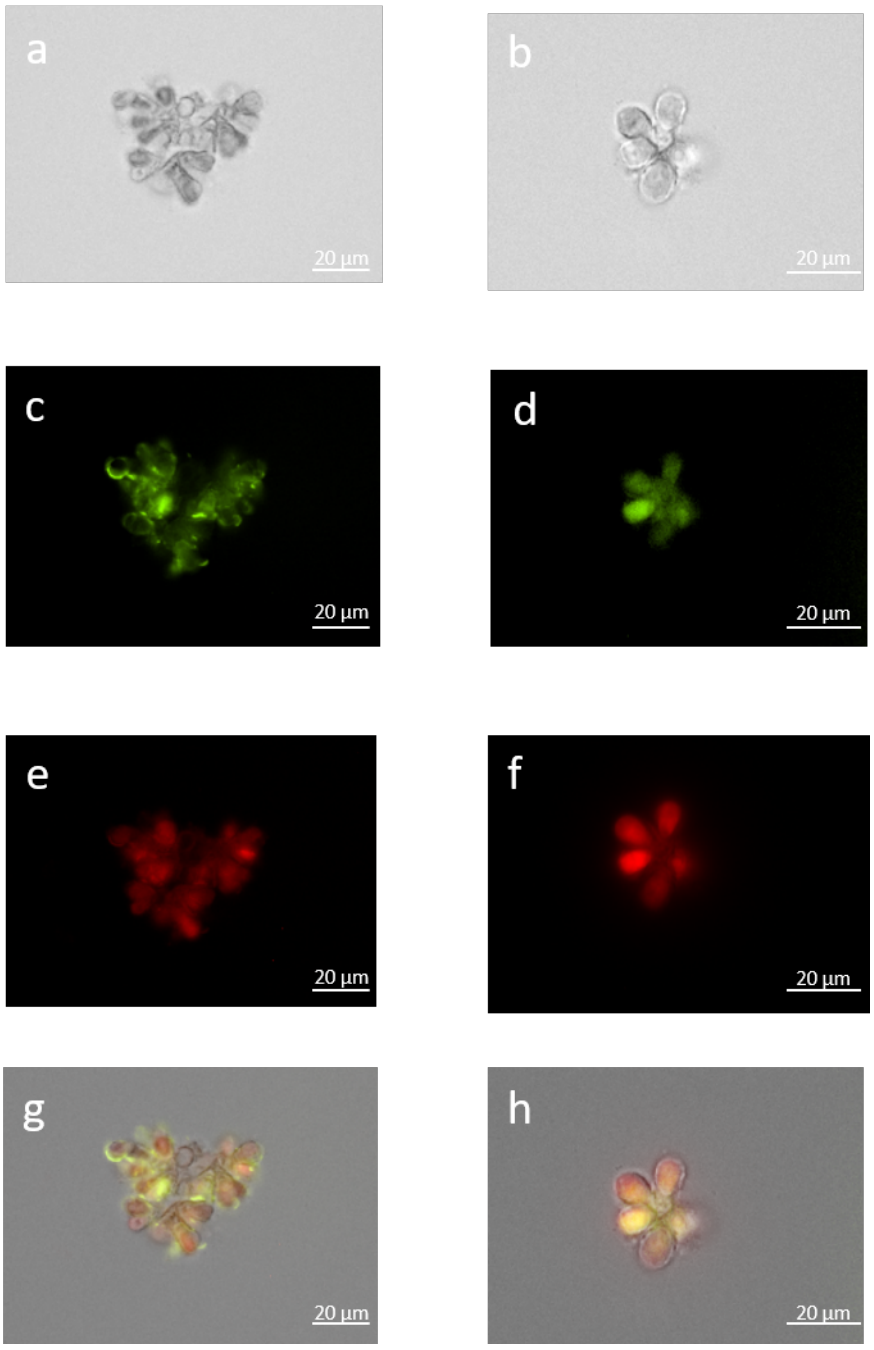


Fig. 3. Fluorescence microscopy observations of cellular localization of YFP-CBM3a-*Bp*GH9 (a, c, e, g pictures) and YFP-*Bp*GH9 (b, d, f, h pictures) in *B. braunii* cell suspension. a and b pictures correspond to the observation in transmission light, c and d correspond to the observation with FITC filter adapted to the YFP signal, e and f correspond to the observation with TRITC filter for the chlorophyll autofluorescence. g and h pictures are artificial merge of light transmission, FITC and TRITC signals. In absence of YFP the cells showed no fluorescence with FITC filter (not shown) (scale bars 20 µm).

3.2 Evaluation of enzymatic colony disruption by flow Cytometry

Cytometry is the method of choice for determining size of cells in suspension as the quantity of scattered light (Forward Scatter or FSC) increases with cell size (Peniuk et al. 2016). We therefore used this technique to evaluate the impact of glycosidases on the organization of *B. braunii* colonies. Indeed if enzymes are able to disaggregate the *B. braunii* colonies, the cell suspension should evolve from large colonies to smaller ones and eventually to single cell. By injecting a control cell suspension of *B. braunii* in the cytometer three populations P1 (Large, 18.8%), P2 (Medium, 29.7%) and P3 (small, 42.0%) were discriminated according to the scattered light (Figure 4).

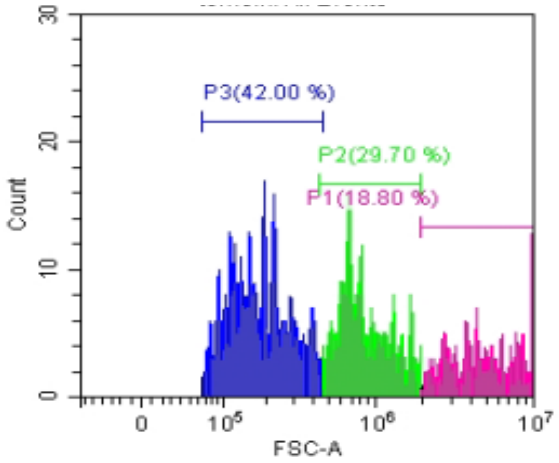


Fig. 4. Distribution of colonies size in *B. braunii* cell culture before enzyme treatment. P3; small-sized cells, **P2;** medium sized cells**, P1;** large cells

When enzymes were applied on *B. braunii* cell suspension the P3 population increased indicating that enzymes effectively reduced the size of colonies. Doubling the enzyme activity (figure S4) did not changed the percentage of P3 population significantly, which suggest that the colony disruption reach a plateau. All enzymes targeting the cell wall were active by increasing the P3 population from around 40% for the untreated cells to 60-75% (Figure 5). The three enzymes (cellulase, hemicellulase and YFP-*Bp*GH9) were almost as active but fusion of *Ct*CBM3a to *Bp*GH9 (YFP-CBM3a-*Bp*GH9) even further increased the enzyme efficiency by around 15%. Concerning the enzymes targeting the colony retaining wall and sheath, the endo-α1,5-arabinanase *(Bt*Abn43A) was almost as efficient as the cellulases. The endo-β1,4-galactanase *(Bs*Gan53A) and the exo-β1,3-galactanase *(Ct*Gan43A) were also effective but in a less extent, with an increase in P3 population from 40% (control) to 45-50%.

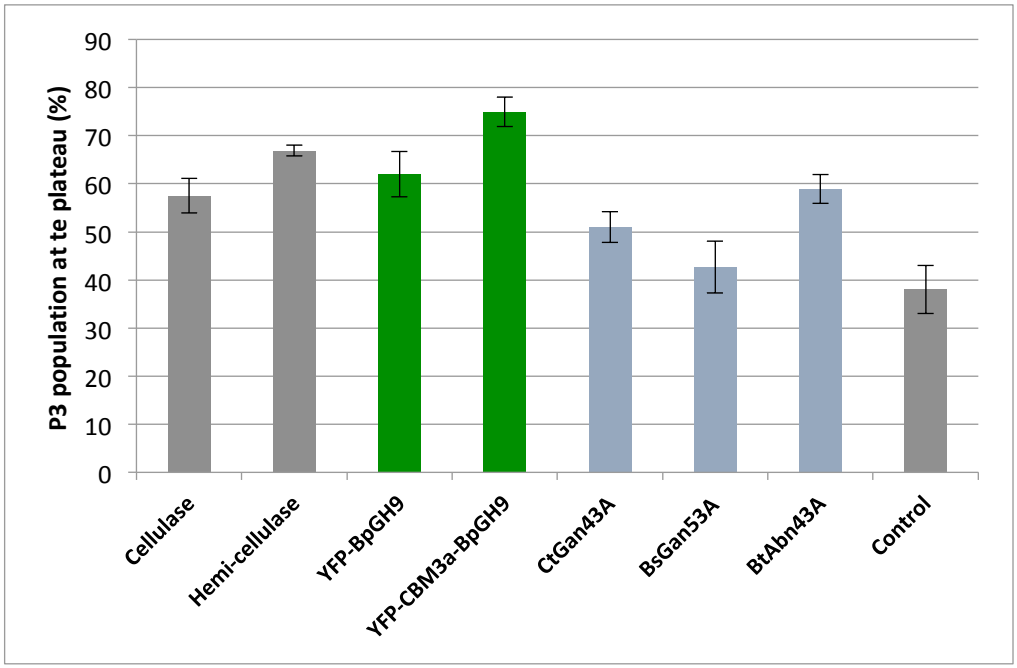


Fig. 5. Quantification of the *B. braunii* colonies disorganization by flow cytometry.

Surprinsingly the enzymes supposed to act on the cell wall (cellulase, hemicellulase and *Bp*GH9) were as much or even more effective than those supposed to act on the retaining wall and colony sheath (*Bs*Gan53A, *Bt*Abn43A, *Ct*Gan43A). This demonstrates the importance of targeting the cell wall for releasing the cells from the colony.

To confirm the effect of enzymes, the cell suspension treated with cellulase and hemicellulase were inspected by fluorescence microscopy. The presence of colonies was revealed by staining the lipids and hydrocarbons present in the ECM with Nile Red and the cell healthiness was controlled by verifying the autofluorescence from chlorophyll. The untreated cells consisted in colonies (Figure 6b) with cells surrounded by hydrocarbons as demonstrated by the presence of a green color due to Nile red (false color) in comparison to cell suspensions treated with cellulase (Figure 6d) and hemicellulase (Figure 6f) where cells appeared relatively more isolated. The autofluorescence from chlorophyll (red color) indicated that the cells were still alive after 1 h enzymatic treatment. (Wang et al. 2021)

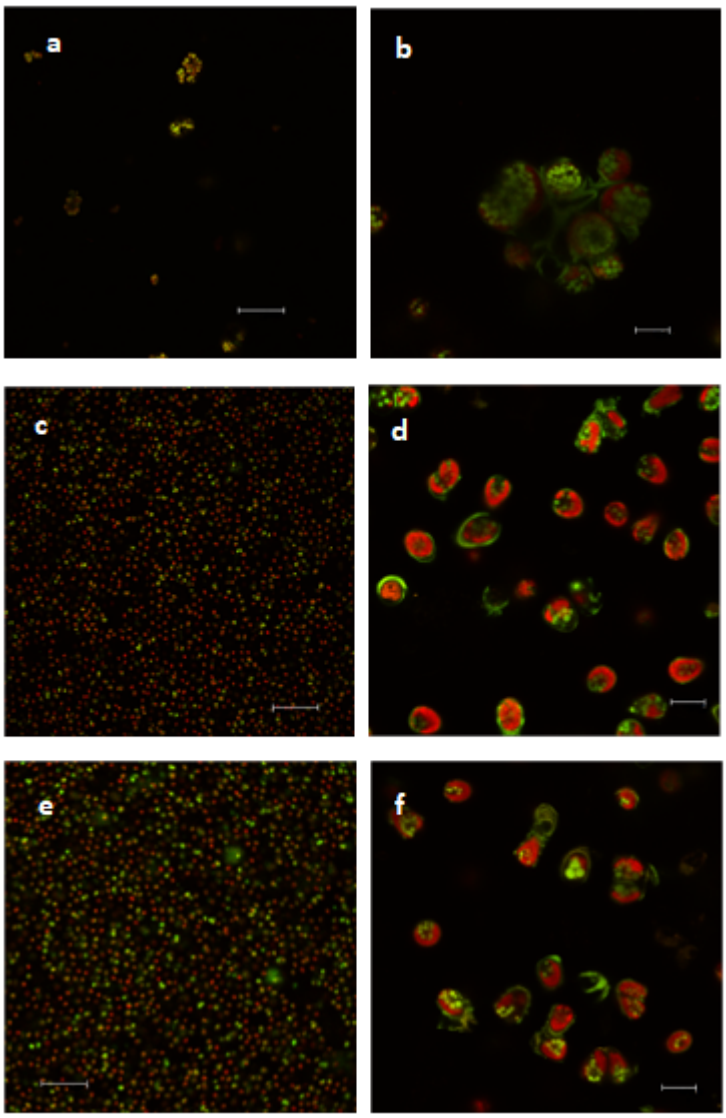


Fig. 6. Colonies *Botryococcus braunii* and cell suspensions with enzymatic treatments. Merge of red autofluorescence from chlorophyll, green fluorescence from intracellular lipids and extracellular hydrocarbons with Nile Red staining. Images **a-b** control without enzyme, enzymatic treatment **c-d** with cellulase, **e-f** with hemicellulase. (Objective x20 (a, c, e), x63 (b, d, f)); scale bars 100 µm (a, c, e), 10 µm (b, d, f) observations done after 1h incubation.

When the cells were left growing some days after the enzymatic treatment the distribution of the different populations tended to come back to the initial proportion before the treatment (data not shown).

4. Conclusions

A set of glycosidases, *i.e*. arabinanase, galactanases, cellulase, hemicellulase and endoglucanase (*Bp*GH9) were shown to be active on the disruption of *B. braunii* colonies, separating the cells from the ECM containing valuable high-energy hydrocarbons that can be used as biofuels. The fusion of *Bp*GHG9 with *Ct*CBM3a and YFP increased the endoglucanase activity and suggested the presence of

microcrystalline cellulose in the cell wall. This opens the way to the development of a semi continuous process where *B. braunii* colonies are grown, treated with enzymes to separate the hydrocarbons from the cells, the hydrocarbons extracted and the cells grown again to regenerate the ECM.

Declaration of competing interest

None.

Author contributions

Amélie Saumonneau, conceptualization and interpretation of the experiments
Nathan Lagneau, performed the pull-down experiments and the cytometry analysis
Lydia Awuor Ogonda, cloning and overexpression of cellulases and their fusion
Catherine Dupré project administration for Gepea, investigation, reviewing
Stéphanie Dutertre investigation/ confocal colocalization microscopy
Dominique Grizeau, conceptualization, methodology, investigation
Charles Tellier, conceptualization, writing and draft preparation
Cyrille Grandjean, contributes to the final version and supervises the project
Franck Daligault project administration, writing, original draft preparation

Acknowledgment

The authors are grateful to Gweninna Cueff for her technical support in fluorescence microscopy and Emmanuelle Courtois for her help in flow cytometry.

We acknowledge France-BioImaging infrastructure supported by the French National Research Agency (ANR-10-INBS-04).
This work was funded by the University of Nantes within the interdisciplinary project framework program.

References

Banerjee A, Sharma R, Chisti Y, Banerjee UC (2002) Botryococcus braunii: a renewable source of hydrocarbons and other chemicals. Crit Rev Biotechnol 22:245–279

Borowitzka MA (2013) High-value products from microalgae—their development and commercialisation. J Appl Phycol 25:743–756

Cartmell A, McKee LS, Peña MJ, et al (2011) The structure and function of an arabinan-specific α-1, 2-arabinofuranosidase identified from screening the activities of bacterial GH43 glycoside hydrolases. J Biol Chem 286:15483–15495

Cheng P, Okada S, Zhou C, et al (2019) High-value chemicals from Botryococcus braunii and their current applications–A review. Bioresour Technol 291:121911

Hirose M, Mukaida F, Okada S, Noguchi T (2013) Active hydrocarbon biosynthesis and

accumulation in a green alga, Botryococcus braunii (race A). Eukaryot Cell 12:1132–1141

Ichinose H, Kuno A, Kotake T, et al (2006) Characterization of an exo-β-1, 3-galactanase from Clostridium thermocellum. Appl Environ Microbiol 72:3515–3523

Jiang D, Fan J, Wang X, et al (2012) Crystal structure of 1, 3Gal43A, an exo-β-1, 3-galactanase from Clostridium thermocellum. J Struct Biol 180:447–457

Jin J, Dupré C, Legrand J, Grizeau D (2016a) Extracellular hydrocarbon and intracellular lipid accumulation are related to nutrient-sufficient conditions in pH-controlled chemostat cultures of the microalga Botryococcus braunii SAG 30.81. Algal Res 17:244–252

Jin J, Dupré C, Yoneda K, et al (2016b) Characteristics of extracellular hydrocarbon-rich microalga Botryococcus braunii for biofuels production: Recent advances and opportunities. Process Biochem 51:1866–1875. https://doi.org/10.1016/j.procbio.2015.11.026

Kawakubo T, Karita S, Araki Y, et al (2010) Analysis of exposed cellulose surfaces in pretreated wood biomass using carbohydrate-binding module (CBM)–cyan fluorescent protein (CFP). Biotechnol Bioeng 105:499–508

Kumar G, Shekh A, Jakhu S, et al (2020) Bioengineering of microalgae: recent advances, perspectives, and regulatory challenges for industrial application. Front Bioeng Biotechnol 8:914

Metzger P, Largeau C (2005) Botryococcus braunii: a rich source for hydrocarbons and related ether lipids. Appl Microbiol Biotechnol 66:486–496

Miller GL (1959) Use of Dinitrosalicylic Acid Reagent for Determination of reducing Sugar. Anal Chem 31:426–428

Nakano H, TAKENISHI S, KITAHATA S, et al (1990) Purification and characterization of an exo-1, 4-β-galactanase from a strain of Bacillus subtilis. Eur J Biochem 193:61–67

Ogonda LA, Saumonneau A, Dion M, et al (2021) Characterization and engineering of two new GH9 and GH48 cellulases from a Bacillus pumilus isolated from Lake Bogoria. Biotechnol Lett 43:691–700

Oliva AK, Bejaoui M, Hirano A, et al (2022) Elucidation of the Potential Hair Growth-Promoting Effect of Botryococcus terribilis, Its Novel Compound Methylated-Meijicoccene, and C32 Botryococcene on Cultured Hair Follicle Dermal Papilla Cells Using DNA Microarray Gene Expression Analysis. Biomedicines 10:1186

Peniuk GT, Schnurr PJ, Allen DG (2016) Identification and quantification of suspended algae and bacteria populations using flow cytometry: applications for algae biofuel and biochemical growth systems. J Appl Phycol 28:95–104

Provasoli L, Pintner IJ (1959) The ecology of algae. In: Spec. Publ. no. 2. Pymatuning Laboratory of Field Biology, University of Pittsburgh, pp 84–96

Pulz O, Gross W (2004) Valuable products from biotechnology of microalgae. Appl Microbiol Biotechnol 65:635–648

Sasaki K, Othman MB, Demura M, et al (2017) Modulation of neurogenesis through the

promotion of energy production activity is behind the antidepressant-like effect of colonial green alga, Botryococcus braunii. Front Physiol 8:900

Spolaore P, Joannis-Cassan C, Duran E, Isambert A (2006) Commercial applications of microalgae. J Biosci Bioeng 101:87–96

Stern J, Kahn A, Vazana Y, et al (2015) Significance of Relative Position of Cellulases in Designer Cellulosomes for Optimized Cellulolysis. PLoS ONE 10:e0127326. https://doi.org/10.1371/journal.pone.0127326

Wang Y, Wang J, Wang T, Wang C (2021) Simultaneous Detection of Viability and Concentration of Microalgae Cells Based on Chlorophyll Fluorescence and Bright Field Dual Imaging. Micromachines 12:896

Weiss TL, Roth R, Goodson C, et al (2012) Colony Organization in the Green Alga Botryococcus braunii (Race B) Is Specified by a Complex Extracellular Matrix. Eukaryot Cell 11:1424–1440. https://doi.org/10.1128/EC.00184-12

Web references

http://atchimiebiologie.free.fr/proteinecalc/proteinecalc.html.

https://www.nikoninstruments.com/fr_FR/Produits/Logiciels/NIS-Elements-Advanced-Research/NIS-Elements-Viewer

Supplementary material

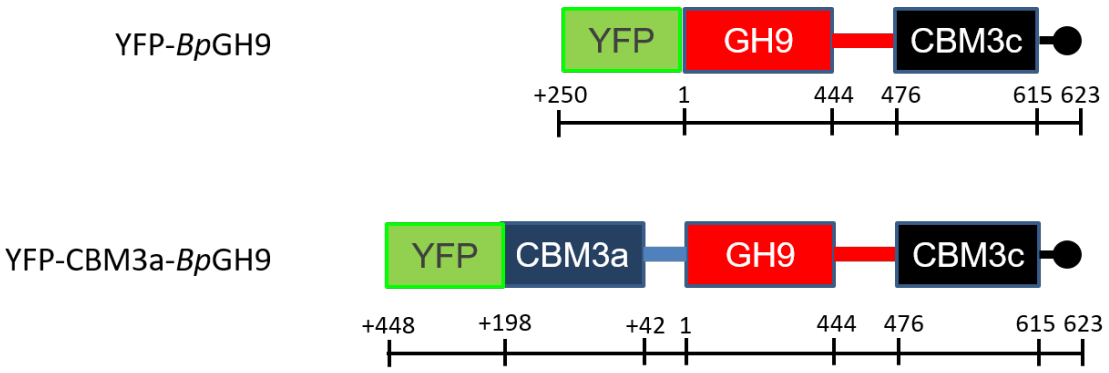


Fig. S2. Schematic diagrams of the recombinants *Bp*GH9 proteins with a C-terminal His-tag and YFP sequence added to fusion proteins. Scales shows the number of amino acid and the boundaries between the different domains.

For_PecA_YFP/CFP

GGACAGCAAATGGGTCGCGGATCCAGCAAAGGAGAAGAAC

Rev_YFP/CFP_GH9

ATAGTTGTAAGATGCCATATTAAGCTTATTACCGGTTTTG

Fig. S3. Primers used for PCR amplification

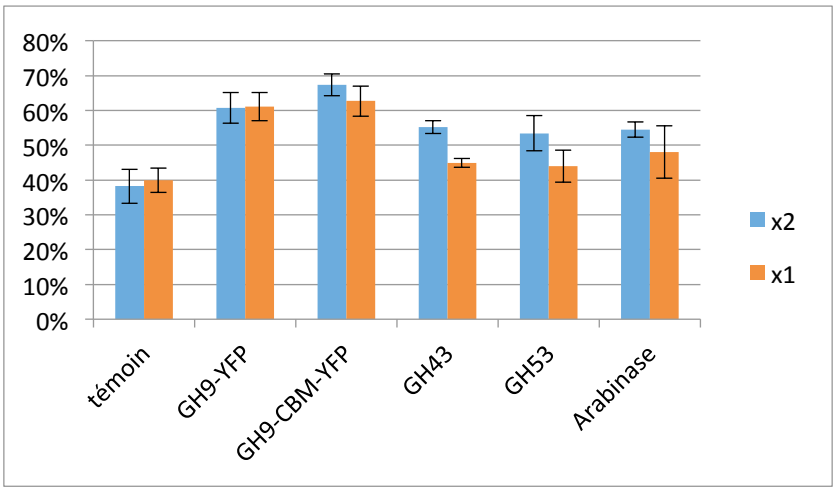


Fig. S4. Percentage of P3 population obtained by using two enzymes concentration

CHAPTER 7

Résumé du Chapitre VII

Dans ce chapitre, l'objectif est d'étendre la stratégie employée dans le chapitre 5 et 6 pour améliorer l'activité des cellulases. L'idée est de s'inspirer des complexes multienzymatiques de dégradation de la cellulose, les cellulosomes, que l'on trouve notamment dans certaines bactéries anaérobies comme *Clostridium Thermocellum*. Ces cellulosomes permettent d'agréger de nombreuses activités cellulolytiques complémentaires au sein d'un complexe supramoléculaire afin d'améliorer la digestion de la cellulose.

Les travaux de cette thèse ont permis de cloner deux nouvelles cellulases, une endocellulase et une cellobiohydrolase, qui interviennent de manière complémentaire dans la dégradation de la cellulose en cellobiose. Le laboratoire ayant préalablement cloné une β-glucosidase d'une bactérie thermophile, l'objectif était de rassembler ces activités dans un même complexe multienzymatique de manière à dégrader plus efficacement la cellulose en glucose.

Pour faciliter notamment l'expression hétérologue et le contrôle de l'organisation spatiale de ces activités dans un même complexe, nous avons décidé d'associer ces activités « en ligne » en fusionnant chaque protéine enzymatique avec les modules de reconnaisance spécifique, dockérine et cohésine, que l'on trouve dans les cellulosomes.

Ce travail a été initié lors de mon dernier séjour en France, qui, faute de temps, n'a pas pu être mené à terme. Nous présentons ci-dessous la stratégie envisagée et l'état d'avancement des travaux préliminaires nécessaires à la réalisation de ce projet.

DESIGNING AND CREATION OF ARTIFICIAL "DESIGNER"CELLULOSOME COMPLEXES WITH IMPROVED SACCHARIDE PRODUCTION

7.1 Introduction

In nature, there exists a plethora of efficient cellulase (Glycosyl hydrolase) enzyme systems **(See section 3.7.2)**. One such efficient cellulase enzyme system exists in the form of a supramolecular complex of enzymes, known as a cellulosome. The most well characterized, simple, cellulosomal system belongs to the anaerobe *Clostridium thermocellum* **(See section 3.9)**.

For this study, we sought to biomimicry the *Clostridium thermocellum* cellulosomal enzyme system in order to produce artificial cellulosomes with possibly superior saccharifying capacity. Albeit, naturally existing in anaerobes, we sought to convert a native non-cellulosomal cellulase into cellulosomal mode of action. This is because even the facultative anaerobes have proved to only have inducible cellulase enzyme production under aerobic conditions making consolidated bioprocessing an attractive process in aerobic microorganisms.

Previous research has also shown that enzymatic diversity is more essential for the degradation of crystalline cellulosic substrates (Hirano *et al*. 2016). However, native cellulosomes **(Figure 2.27 & Figure 2.32)** present with unpredictable variability and random relative positioning.

To avoid random positioning, several authors have proposed to construct minicellulosomes with different pairs of cohesion/dockerin such in the miniscafoldin **(Complex I, Figure 7.1),** (Stern *et al*., 2015). Some authors have proposed that organizing the relative position of enzymes in a mini scaffoldin could facilitate substrate channeling (You *et al.,* 2012, Stern *et al.,* 2015, Srikrishman *et al.,* 2012) and improve efficiency of the enzyme complex.

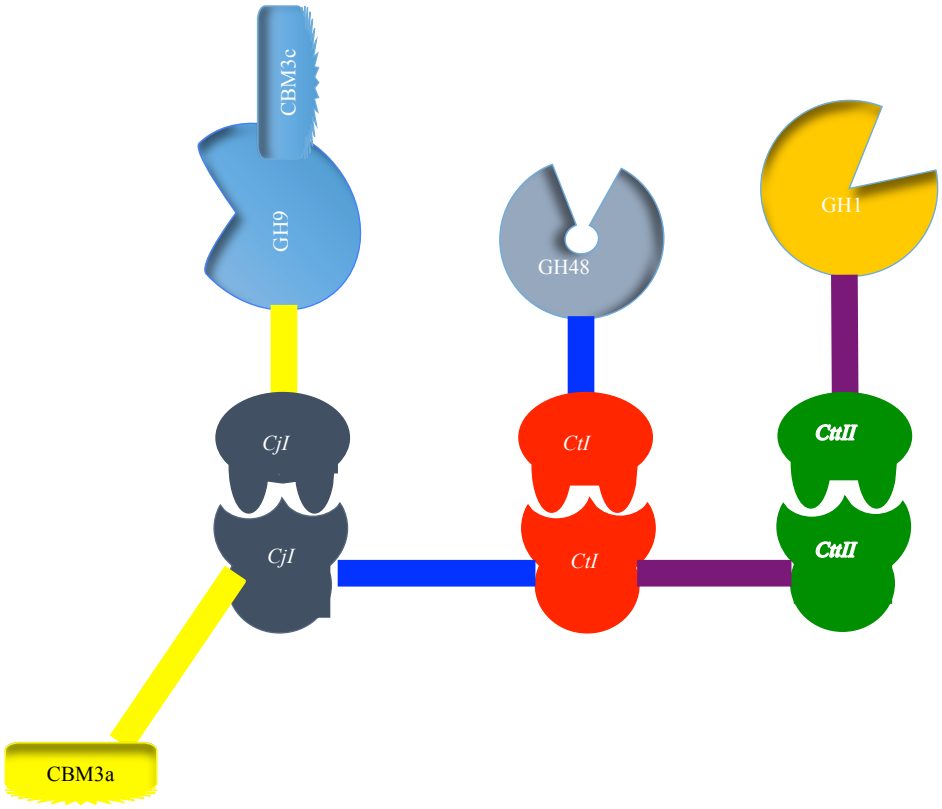


Figure 7.1 Example of Non-natural miniscaffoldin; Complex I (Adapted from Stern *et al.*, 2014)

CtI; *Clostridium thermocellum* type I cohesin/dockerin, ***CjI***; *Clostridium josui* type I cohesin/dockerin pair, ***CttII***; *Clostridium thermocellum* type II cohesin/dockerin

The work presented in the previous chapter has shown that cellulases were quite tolerant to fusion both at the C- and N-terminal end, without affecting the enzymatic activity. Such behavior suggest to us that we could envisage to fuse each enzyme at both ends with a specific cohesin/dockerin pair in order to produce IN-LINE CELLULOSOMES **(Figure 7.2)**

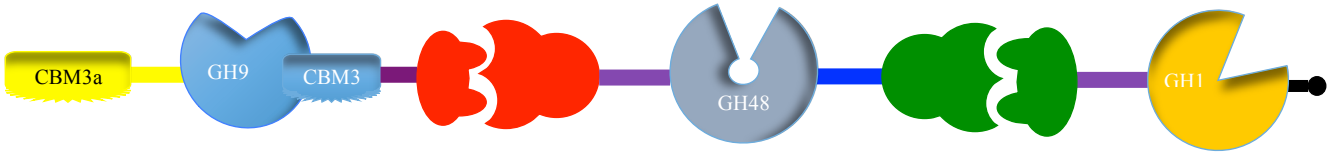


Figure 7.2 Example of an IN-LINE cellulosome construction of cellulases

7.2 Materials and methods

7.2.1 Designing of Cellulosome Complexes

In order to meet the research objectives as set out above, miniscaffoldin **(Figure 7.1)**, and novel IN-LINE cellulosomes were designed using a set of elementary building bricks including: linkers, cohesin and dockerin pairs from *Clostridium josui,* and *Clostridium thermocellum* as well as GH1, GH9 and GH48 enzymes from *Thermus thermophilus* and *Bacillus pumilus* respectively (**Figure 7.3, Figure 7.4, Figure 7.5)**

Figure 7.3 Glycosyl hydrolases i. GH9 ii. GH48 iii. GH1

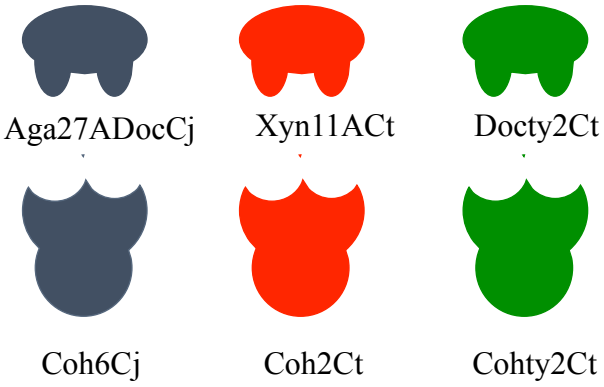


Figure 7.4 Cohesin and Dockerin Pairs

Figure 7.5 Linkers

Figure 7.6a Histidine tag & **7.6b** Cellulose Binding Module

7. 2.1.1 Towards IN-LINE cellulosomes

Using the building blocks presented above, we envisioned the construction of various in line cellulosomes in order to compare their efficiency on cellulose substrates **(Figures 7.7, 7.8, 7.9 and 7.10).**

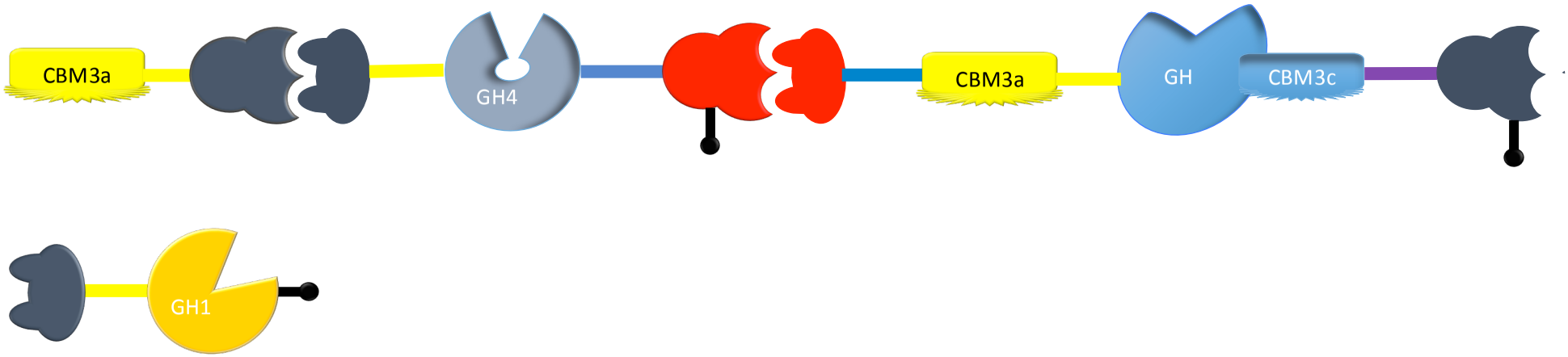


Figure 7.7 IN-LINE cellulosome, Complex II (Composed of C1; C14; C15; C13)

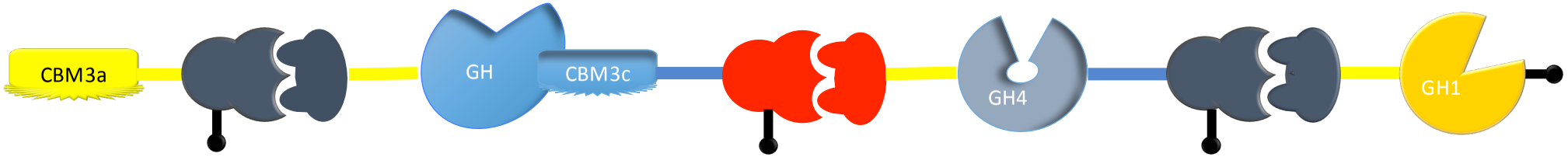


Figure 7.8 IN-LINE cellulosome, Complex III (Composed of C1; C11; C12; C13)

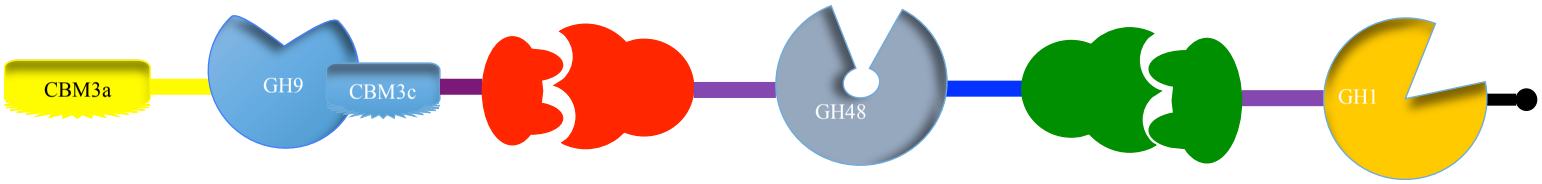


Figure 7.9 IN-LINE cellulosome, Complex IV (Composed of mini-complexes C5; C8; C7)

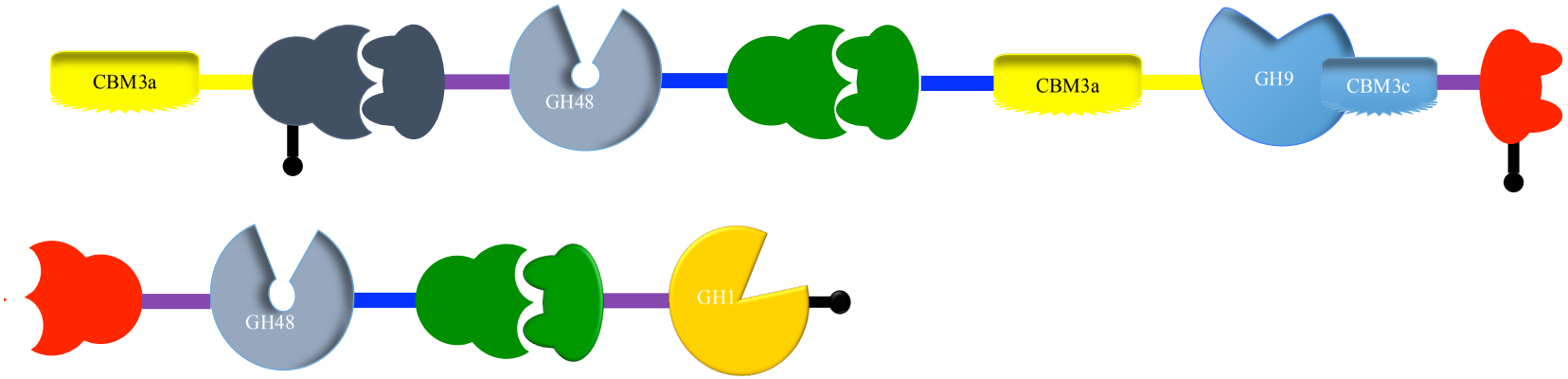


Figure 7.10 IN-LINE cellulosome, Complex V (Composed of mini-complexes C1; C9; C10; C8; C7)

7.2.1.1.1Production of Cellulosome Complexes

7.2.1.1.2Cloning the building bricks of the minicellulosomes

7.2.1.1.3 Enzymes

Three glycosyl hydrolase enzymes were used in the assembly of the complexes in this study. These included an endoglucanase, (GH9) and exoglucanase, (GH48) from an extremophile *Bacillus pumilus* obtained as previously described **(chapter 4).** In addition, to a thermophilic beta glucosidase, (GHI) enzyme obtained from a previous project conducted in the laboratory (Dion *et al*., 1999)

7.2.1.1.4 Exogenous Cellulosomal Modules

In addition to the enzyme pieces, complementary cohesin and dockerin pieces, a class 3a, cellulose binding module (CBM3a) from *Clostridium thermocellulm* and linkers of *Clostridium thermocellum*

and *Clostridium josui* were also extracted from their gDNA using polymerase chain reaction (PCR).

PCR amplification of the desired genes was done as follows:

Table 7.1 PCR set up for extraction of cellulosome building blocks

S/No.	**Reagent**	**Volume (µl)**
1	dNTPs	1
2	10x Dream Taq Buffer	5
3	10mM Forward primer	1
4	10mM Reverse primer	1
5	Template	1
6	dH_2O	40.5
7	Dream Taq Polymerase	0.5
	Total volume	**50.0**

Table 7.2: Different tailored PCR programs set to amplify different genes:

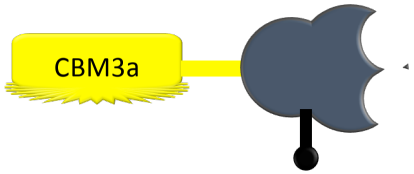
S/No.	Gene	Dream Taq polymerase	Phusion Polymerase
1.	**Aga27A Dockerin *Clostridium josui***	*95°C 5mins* *95°C 10s* *60°C 30s* *72°C 2mins* } *30 cycles* *72°C 10 mins* *4°C ∞*	*98°C 30s* *98°C 10s* *60°C 30s* *72°C 2mins* } *30 cycles* *72°C 10 mins* *4°C ∞*
2.	**Cohesin1 *Clostridium josui***	*95°C 5mins* *95°C 10s* *61°C 30s* *72°C 2mins* } *30 cycles* *72°C 10 mins* *4°C ∞*	*98°C 30s* *98°C 10s* *61°C 30s* *72°C 2mins* } *30 cycles* *72°C 10 mins* *4°C ∞*
3.	**Cohesin 2 *Clostridium josui***	*95°C 5mins* *95°C 10s* *61.8°C 30s* *72°C 2mins* } *30 cycles* *72°C 10 mins* *4°C ∞*	*98°C 30s* *98°C 10s* *61.8°C 30s* *72°C 2mins* } *30 cycles* *72°C 10 mins* *4°C ∞*
4.	**Cohesin 5 *Clostridium josui***	*95°C 5mins* *95°C 10s* *58.7°C 30s* *72°C 2mins* } *30 cycles* *72°C 10 mins* *4°C ∞*	*98°C 30s* *98°C 10s* *58.7°C 30s* *72°C 2mins* } *30 cycles* *72°C 10 mins* *4°C ∞*
5.	**Cohesin 6 *Clostridium josui***	*95°C 5mins* *95°C 10s* *60.6°C 30s* *72°C 2mins* } *30 cycles* *72°C 10 mins* *4°C ∞*	*98°C 30s* *98°C 10s* *60.6°C 30s* *72°C 2mins* } *30 cycles* *72°C 10 mins* *4°C ∞*
6.	**Cellulosome anchoring protein (CiPA) *Clostridium thermocellum***	*95°C 5mins* *95°C 10s* *60°C 30s* *68°C 6mins* } *30 cycles* *72°C 10 mins* *4°C ∞*	*98°C 30s* *98°C 10s* *60°C 30s* *68°C 6mins* } *30 cycles* *72°C 10 mins* *4°C ∞*

7.	**Dockerin Type II *Clostridium thermocellum***	*95°C 5mins* *95°C 10s* *70°C 30s* } *30 cycles* *72°C 2mins* *72°C 10 mins* *4°C ∞*	*98°C 30s* *98°C 10s* *70°C 30s* } *30 cycles* *72°C 2mins* *72°C 10 mins* *4°C ∞*
8.	**Cohesin Type II *Clostridium thermocellum***	*95°C 5mins* *95°C 10s* *67°C 30s* } *30 cycles* *72°C 2mins* *72°C 10 mins* *4°C ∞*	*98°C 30s* *98°C 10s* *67°C 30s* } *30 cycles* *72°C 2mins* *72°C 10 mins* *4°C ∞*
9.	**Cellulose binding module *(CBM3a)* *Clostridium thermocellum***	*95°C 5mins* *95°C 10s* *65.9°C 30s* } *30 cycles* *72°C 2mins* *72°C 10 mins* *4°C ∞*	*98°C 30s* *98°C 10s* *65.9°C 30s* } *30 cycles* *72°C 2mins* *72°C 10 mins* *4°C ∞*
10.	**Cohesin 2 *Clostridium thermocellum***	*95°C 5mins* *95°C 10s* *64.5°C 30s* } *30 cycles* *72°C 2mins* *72°C 10 mins* *4°C ∞*	*98°C 30s* *98°C 10s* *64.5°C 30s* } *30 cycles* *72°C 2mins* *72°C 10 mins* *4°C ∞*
11.	**Cohesin 4 *Clostridium thermocellum***	*95°C 5mins* *95°C 10s* *62°C 30s* } *30 cycles* *72°C 2mins* *72°C 10 mins* *4°C ∞*	*98°C 30s* *98°C 10s* *62°C 30s* } *30 cycles* *72°C 2mins* *72°C 10 mins* *4°C ∞*
12.	**Xylanase11A Dockerin *Clostridium thermocellum***	*95°C 5mins* *95°C 10s* *60.5°C 30s* } *30 cycles* *72°C 2mins* *72°C 10 mins* *4°C ∞*	*98°C 30s* *98°C 10s* *60.5°C 30s* } *30 cycles* *72°C 2mins* *72°C 10 mins* *4°C ∞*
13.	**Xylanase 10C Dockerin *Clostridium thermocellum***	*95°C 5mins* *95°C 10s* *64.5°C 30s* } *30 cycles* *72°C 2mins* *72°C 10 mins* *4°C ∞*	*98°C 30s* *98°C 10s* *64.5°C 30s* } *30 cycles* *72°C 2mins* *72°C 10 mins* *4°C ∞*

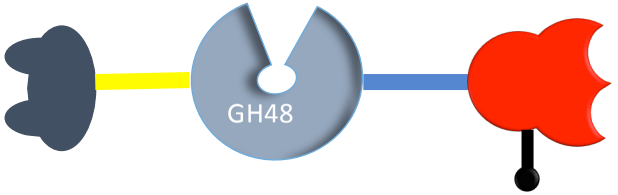
Table 7.3: Chimera constructs building bricks

Module	**Cohesins**	**Dockerins**	**CBM**	**Linkers**
1	Coh6 *Cj*	Aga27A *Cj*	CBM3a *Ct*	NLink41
2	Coh2 *Ct*	Xyn11a *Ct*		CLink41
3	Cohty2 *Ct*	Docty2 *Ct*		Xlink28
4	Coh2 *Cj*	Xyn10 *Cj*		

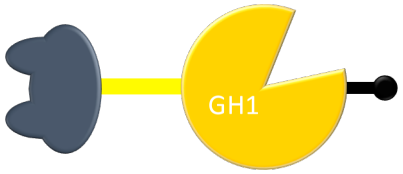
In order to form the packed complexes as designed in **section 7.2.1**, and carry out functional comparative analysis, Gibson Hi-Fi PCR assembly protocol was used to make the following modules resulting from the fusion of the building bricks



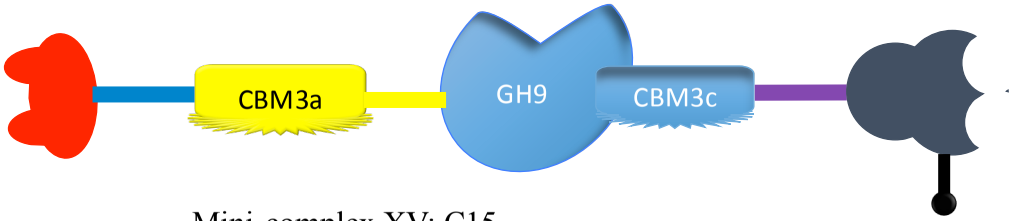
Mini-complex I; C1



Mini-complex X1V; C14

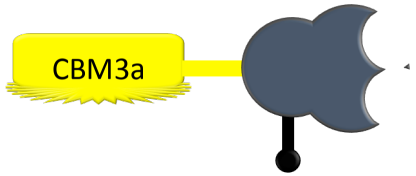


Mini-complex XIII ; C13

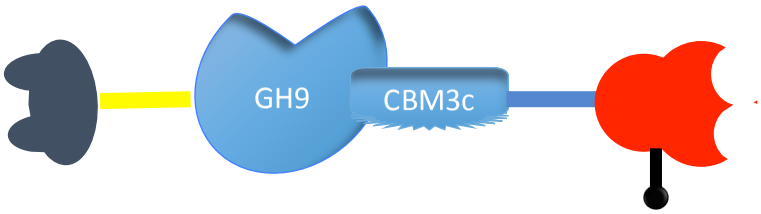


Mini-complex XV; C15

Figure 7.11 Mini-complexes for the IN-LINE cellulosome complex II

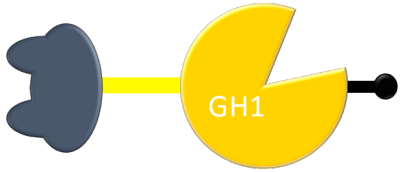


Mini-complex I; C1



Mini-complex X1; C11

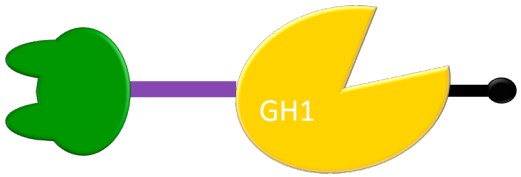
Mini-complex X11; C12



Mini-complex XIII ; C13

Figure 7.12 Mini-complexes for IN-LINE cellulosome complex III

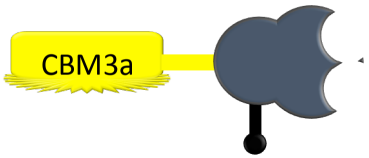
Mini-complex V ; C5



Mini-complex VII ; C7

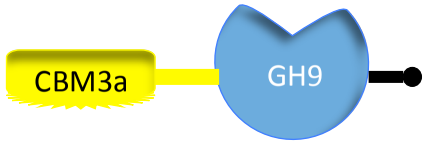
Mini-complex VIII ; C8

Figure 7.13 Mini-complexes for IN-LINE complex IV



Mini-complex I : C1

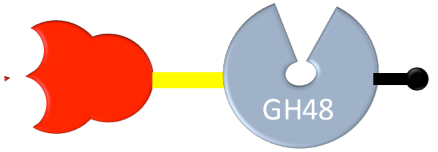
Mini-complex II : C2



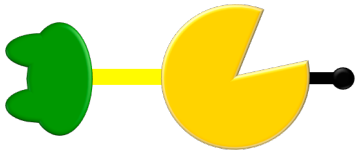
Mini-complex III :

Mini-complex IV : C4

Mini-complex V : C5



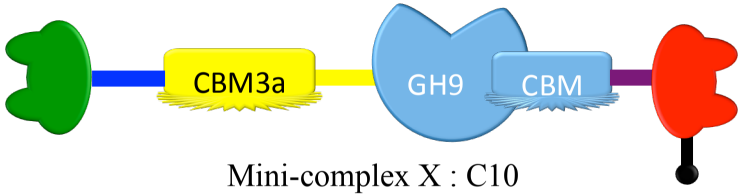
Mini-complex VI : C6



Mini-complex VII :

Mini-complex VIII : C8

Mini-complex IX : C9



Mini-complex X : C10

Figure 7.14 Mini-complexes for complex V

7.3 RESULTS and DISCUSSION

7.3.1 Design and Synthesis of the building bricks

Several complex pieces as shown in section 7.2 above were designed and synthesised from *Clostridium thermocellum* ATCC 27405 (CP000568) CiPA gene and gDNA of *Clostridium josui* (**See appendices).** The genes for Aga27A dockerin, type I cohesin 6, 5, 2, and 1 and dockerins Xyn10C and Xyn11a dockerin were extracted by PCR amplification from gDNA of *Clostridium josui.* Similarly, the genes for dockerins Xyn11a and Xyn10C and type I cohesin 2, 4 and type II cohesin and type II dockerin from *Clostridium thermocellum.* Following successful amplification **(Figure 7.15)**, the PCR fragments were sequenced **(Macrogen, Inc.)** and positive amplification confirmed. In order to make the mini-complexes and associated IN-LINE cellulosomes only select 3 cohesin/dockerin pairs; Aga27A*Cj*, Coh6*Cj*, type II Coh/Doc*Ct* and type I Coh2*Ct* / Xyn1a*Ct* **(Figure 7.3-7.6)** were cloned into PeCa plasmid and used to transform BL-21cells. Successful transformants were checked by colony PCR **(Figure 7.15a)** for all the 6 modules**.** Further, in order to analyse for the effect of the number of CBM in cellulose binding function, a chimeric CBM3a from *Clostridium thermocellum CipA* was synthesized **(Genscript, Inc)** with the native N-linker at it C–terminal **(Figure 7.3i).** For efficient translation in *E.coli*, this CBM3aCLink41 gene was codon-optimised for *Escherichia coli.* Additionally, in order to evaluate the effect of linker lengths, linkers NLink41 and XLink28 were also synthesized and codon-optimised for *Escherichia coli.*

The enzyme modules; BPGH9 **(Figure 7.3i),** cellobiohydrolase **(Figure 7.3ii)** used here were obtained as outlined above **(Chapter 5).** In addition, a GH1 enzyme **(Figure 7.3iii)** from a previous study (Dion *et al.,* 1999) was used.

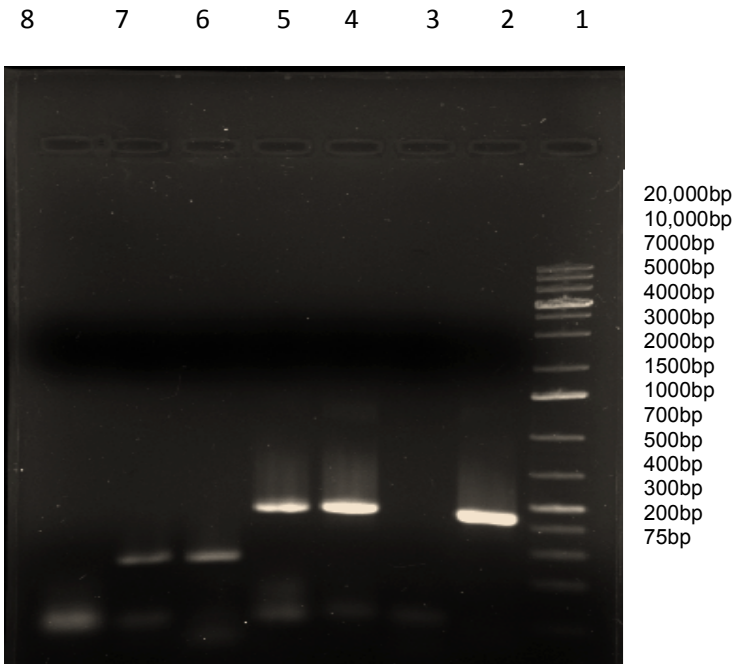


Figure 7.15a Gel of exogenous modules PCR amplicons, 1.DNA ladder, 2.Coh2*Cj*, 3.Coh5*Cj*, 4.Coh6*Cj* 5.Coh2 *Ct*, 6.Xyn11ADoc*Ct,* 7.Xyn10CDoc*Ct*, 8.CBM3a*Ct*

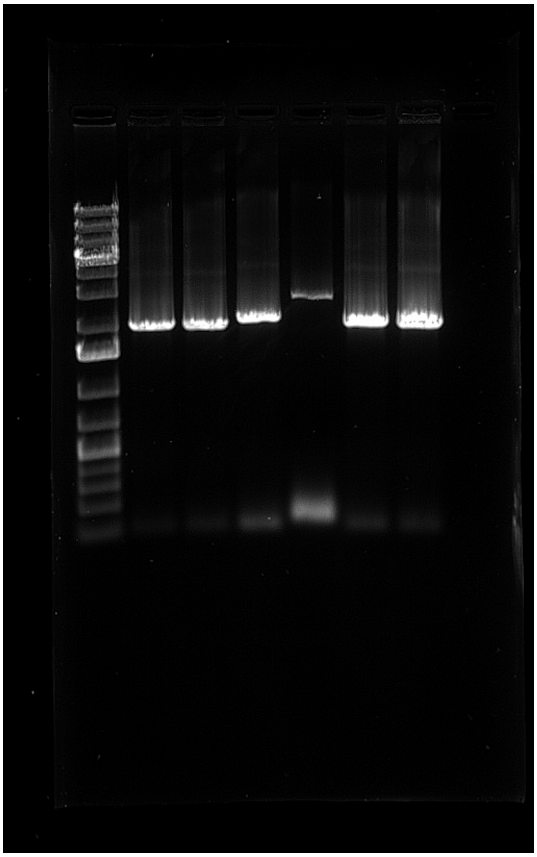


Figure 7.15b C8: Truncated BPGH9 (endoglucanase minus CBM3c)

7.3.2 Conversion of Cellulases into Cellulosomal mode

In order to convert the aerobic, non-cellulosomal extracted enzymes into cellulosomal mode, we fused the enzymes to the CBM3a module. In addition, to help with the IN-LINE cellulosome complex packing, fusion proteins of the enzymes, XLink28/NLink41/Clink41 **(Figure 7.5)** and various Cohesin/dockerin **(Figure 7.4).** The enzyme modules were constructed using restriction free Gibson Hi-Fi assembly **(Figure 7.7 to Figure 7.14).** Successful cloning was verified by colony PCR for all constructions, followed by viewing on agarose gel. This was then followed by expression and confirmation on SDS-PAGE.

Succesful conversion of the endoglucanase into cellulosomal mode was analysed by checking for the activity of this construct **(Mini-complex II; Figure 7.14)** on both crystalline cellulose and native cellulose biomass in the form of macroalgae. Results from this preliminary experiments indicated that the CBM3a module was functional as seen following pull down assay **(Chapter 5)**, additionally the function of this new construct was higher than that of the baseline construct (the wildtype BPGH9).

7.3.3 Effect of Numbers of Carbohydrate binding Module (CBM) on enzyme activity

In order to analyse for the effect of CBM numbers on cellulase function. An exogenous CBM3a from CiPA, *Clostridium thermocellum* was added to the non-native N terminus of the GH9 enzyme; CBM3aBPGH9CBM3c as is the case of Cel5A from *thermobifida fusca* (Wilson et al. 1992, 1999). In addition, a CBM3a was added to the truncated wiltype BPGH9. And finally the accessory CBM3c module was truncated from the wildtype BPGH9. The activities of these constructs were all compared to those of the wildtype **(Figure 7.16)**.

As already shown in **Chapter 5**, the truncated wiltype BPGH9 without its accessory CBM3c is unstable at high temperatures and could not be tested on microcrystalline cellulose. However, its

activity on amorphous cellulose was retained but reduced. On the other hand, addition of CBM3a on the native wildtype enzyme improved its activity (Petkun *et al.,* 2013). Interestingly however, the activity of the truncated wildtype with the CBM3a on the N-terminus was higher than that of the Native with the CBM on opposite ends **(Figure 7.16c).** This antagonistic positional synergy in the CBM highlights the need to further evaluate the effect of another construct **(Figure 7.16e)** to evaluate the effect of the agonistic positional synergy in the CBMs.

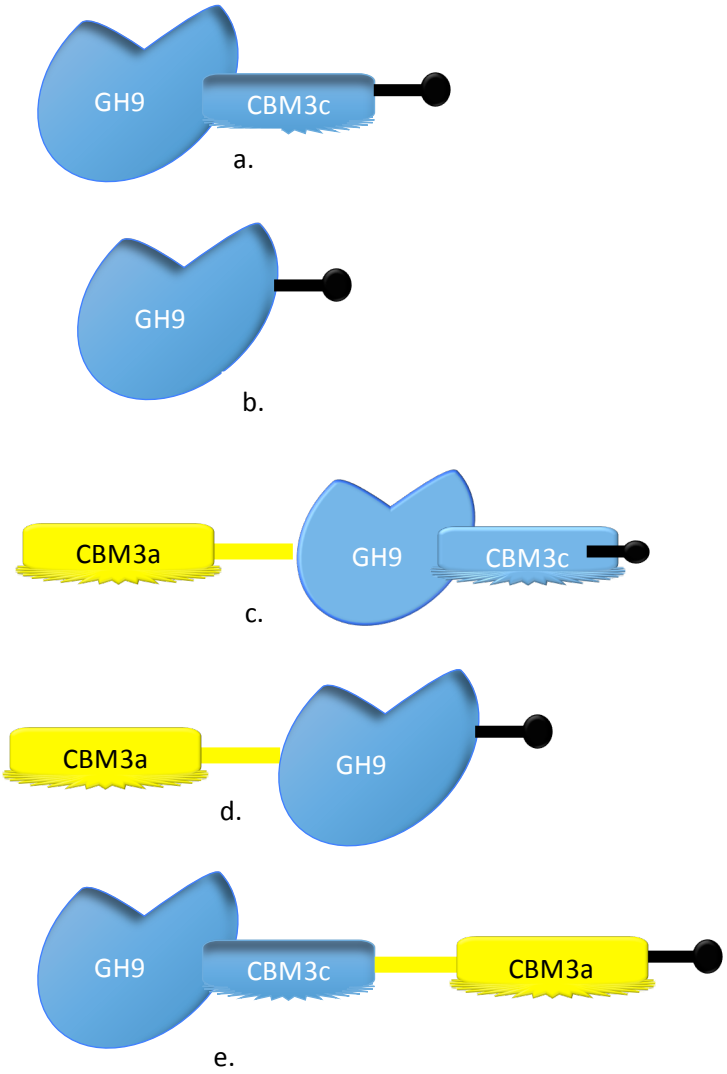


Figure 7.16: Variants of β-1, 4 endoglucanase (*BP*GH9)

7.3.4 Homology modelling and validation of the wildtype, and dockerin fused enzymes.

Preliminary homology modelling of the native enzymes BPGH9 and BPGH48 was done and validated using SWISSPROT.These structures, were viewed, and exported using PYMoL **(Chapter 5).** However, it would be imperative to conduct crystallization and structural prediction using the alphafold program using https://alphafold.ebi.ac.uk of the constructs and complexes that have been established as functionally value additive.

. This would help understand the mechanistic implications of binding using the various complementary scaffolding protein pairs during complex packing and their implications on complex function. This would additionally guide selection of alternative complementary pairs as part of the scaffoldin for designing of enzyme complexes.

7.3.5 Expression and Purification of Cellulosomal proteins

In order to purify the glycosyl hydrolase enzymes**,** and their variants BL21 (DE3) cells, and TOP10 cells containing the plasmids were grown in Luria Bertani media at 16°C supplemented with ampicillin . The proteins, were overexpressed, by adding isopropyl β-D -1, thiogalactoside to a final concentration of 0.5mM and incubated for 16hrs. This was followed by cold (4°C) centrifugation to harvest the cells. The pelleted cells were frozen, lysed by sonication and the His tagged protein purified using Ni-NTA columns. The purified fractions were dialysed against PBS buffer and analysed using SDS-PAGE **(Chapter 5)**.

7.3.6 Complex Assembly and Purification

7.3.6.1 In vitro complex assembly

Preliminary complex formation was conducted using the purified expressed modules Aga27ABPGH9CBM3c and CBM3a-Coh6 (**Figure 7.14 C1; C4)** in sample buffer (Stern *et al.*, 2015). To determine the equimolar concentrations and interactions various concentrations 91-10µg of protein each were added to PBS supplemented with 10mM $CaCl_2$ and 0.05% Tween 20 to a total volume of 30µl. These were incubated at 37°C for 2hrs and after which run on Non-denaturing PAGE. Complex formation was confirmed on non-native gel (results not shown).

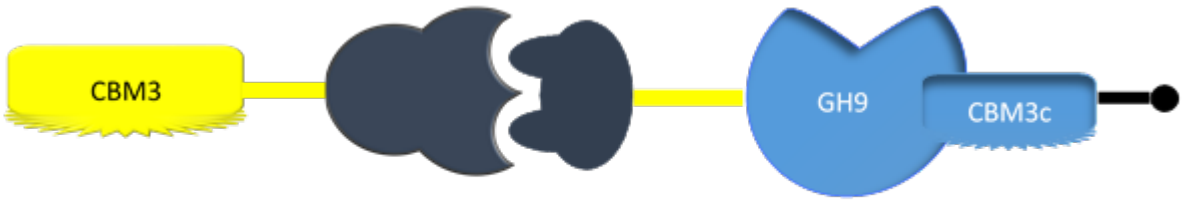


Figure 7.17a Minicellulosome I; CBM3a-Coh6; 36kDa and Aga27ABPGH9CBM3c; 84kDa

7.3.6.2. In vitro Complex Assembly on Size Exclusion Chromatography

In addition to packing the complexes *in vitro*, complexes **(Figure 7.14)** were also packed on the SEC columns. SEC standards (Blue dextran, Apoferritin, Aldolase, Ovalalbumin and PABA. At a flow rate of 0.8µl/Min. For minicomplex I (CBM3a-Coh6; 36kDa), 100µl of the sample was mixed with sample buffer (NaAcetate 50mM, 10mM, 200mM NaCl) 500µl. The samples were filtered, and injected into column. Samples that were assembled in vitro then run on the column showed successful complexation albeit eluting earlier than expected at 159.978kDa versus the expected 120kDa. This could be explained by the 3D spatial folding of the complex. In addition the single complex parts eluted out at the expected peak points, 36kDa for C1 and 84kDa for C4 **(Figure 7.17a-e)**

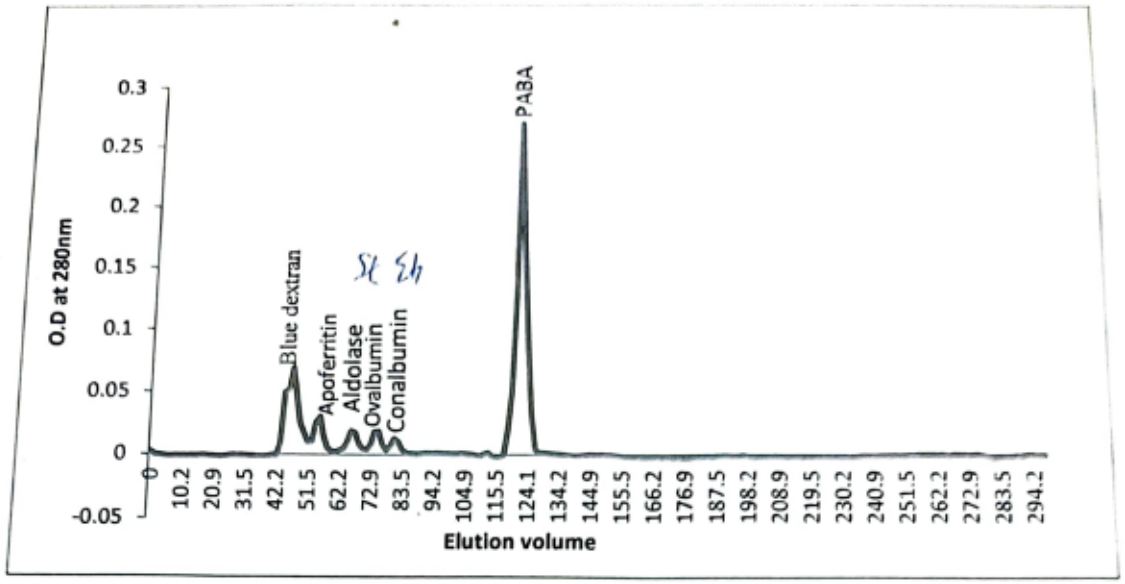


Figure 7.17b Size exclusion chromatogram standards using 50mM NaH_2PO_4, NaCl 200mM buffer at pH 8

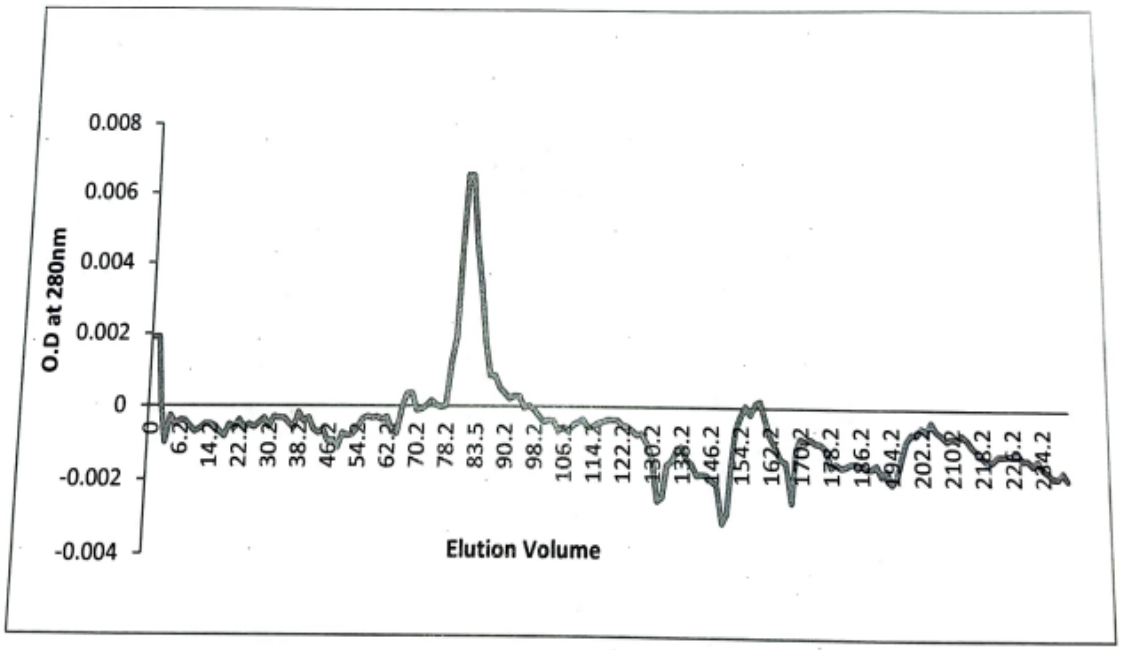


Figure 7.17c Size exclusion chromatography of C1; CBM3a-Coh6, using 50mM Na-Acetate buffer , 200mM NaCl (PH5) ; Kav=0.4850, Log molecular weight =4.3153, Determined molecular weight=40kDa , known Molecular weight=36KDa

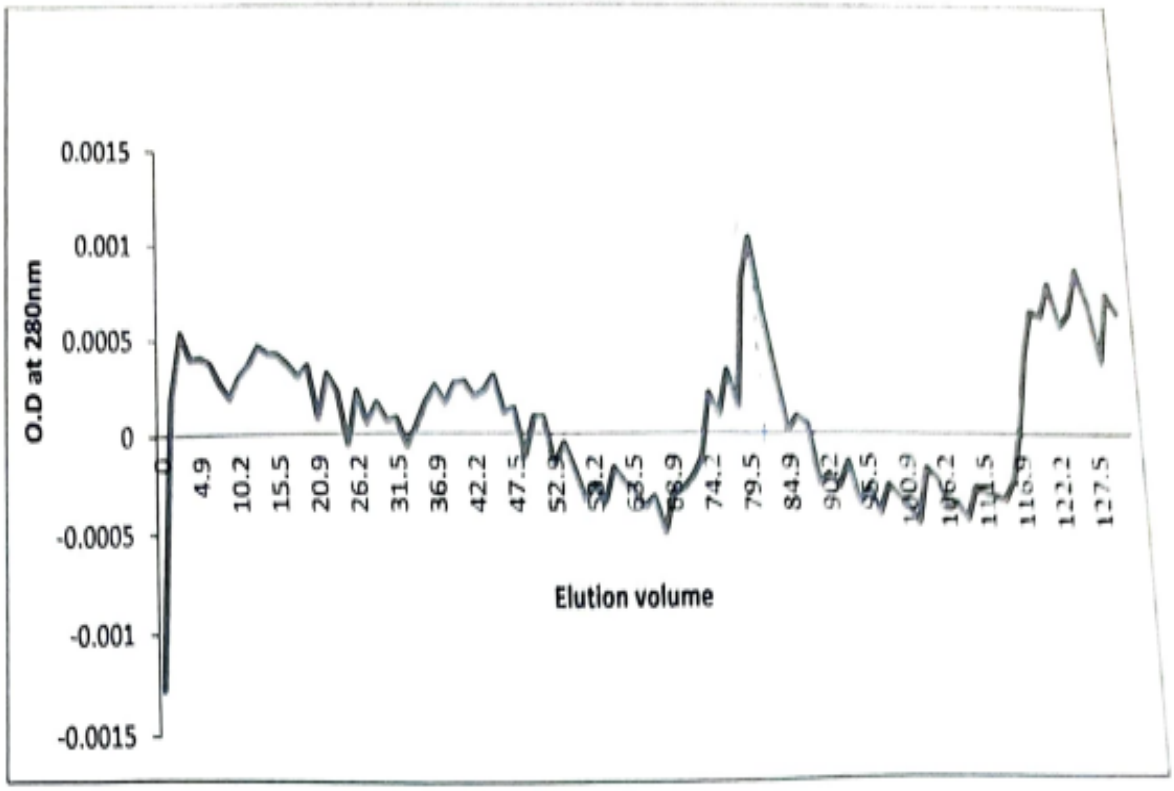


Figure 7.17d Size exclusion chromatogram of C4 (AgaBPGH9) using 50mM Na-Acetate buffer, 200mM NaCl (pH5); Kav=0.4701, Log molecular weight =4.3760, Determined molecular weight =67kDa, Known molecular weight =84kDa

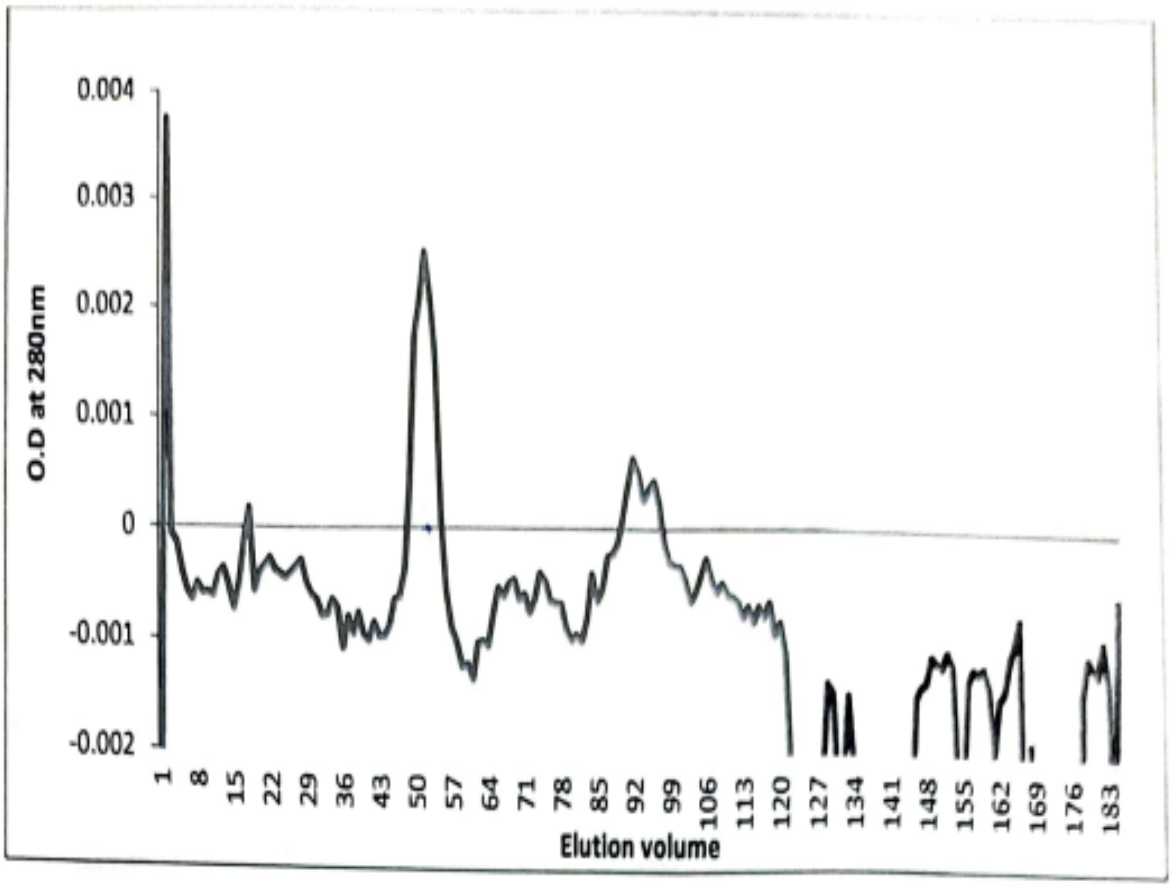


Figure 7.17e Size exclusion chromatogram of C1:C4 complex using 50mM Na-Acetate buffer , 200mM NaCl (pH5) ; Kav=0.2534, Log molecular weight =5.2638, determined molecular weight =159.978kDa, Known molecular weight =120kDa

7.3.7 Immunoflourescence Microscopy and FACS

7.3.7.1 Fluorescent tagging of enzymes

Indeed this is the first time that CBM3a anchoring to *Botryococcus braunii* EPS has been demonstrated. Our fluorescent endoglucanase fusion protein (YFP*BpGH9)* binds to its covalent fusion protein partner (CFP-CBM3a) **(Chapter 6).** This preliminary binding was carried out in order to ascertain the retention of structural integrity of proteins upon addition of the flourescent tag and hence functionality of the said modules. The YFP on *YFPBpGH9* helps to localize the enzyme into the macroalgae because of the associated fluorescence of YFP.

7.3.7.2 Fluorescent microscopy of *Botryococcus braunii* colonies

Botryococcus braunii and Flourescent tagged (YFP) Chimera construct C2 and chimera construct C3 and wild type BPGH9 **(Chapter 6),** were mixed in a ratio of 2:1. The specimen was then observed under x40 (Fibre Optic Nikon, Ni-E) to observe the algal colony structures .TRITC filters (excitation wavelength: 543/22nm; emission wavelength 593/40nm), were used to check for fluorescence of chlorophyll. In addition, FITC filters (excitation wavelength: 482/35nm; emission 536/40nm), were used to observe for fluorescence of YFP. The images were treated using Nis-Elements viewer (https://www.nikoninstruments.com/fr_FR/Produits/Logiciels/NIS-Elements-Advanced-Research/NIS-Elements-Viewer)

7.3.7.3 Analysis of degradation of EPS of Botryococcus braunii

We used flow cytometry (CytoFLEX, Beckman Coulter) to monitor the degradation of EPS of *B.braunii.* Forward scatter (FSC) and Side scatter (SSC) were detected using a 488nm filter. Similarly, a 660nm filter was also used to detect emission of fluorescence by chlorophyll that was indicative of the vitality of the alga populations. The deposit was fixed at 10µl/min and saved at 1000 events **(Chapter 6)**.

7.3.7.4 Enzyme Assays / Pull down assays

Enzyme assays on microcrystalline cellulose and its amorphous substituents PASC and Na-CMC was done according to **(Chapter 5)** and **(Chapter 6)**.

7.3.7.5 Protein –Protein docking for complex packing

In silico protein –protein docking of the matching pairs during the in vitro complex assembly is an essential next step because with these positional constraints of the different enzymes would be minimized and help in predicting of the most stable arrangements of the enzymes for maximum function and stability (Stern *et al.,* 2015). In addition, this would be helpful in finding alternative protein pairs for scaffoldin in designing and production of enzyme complexes.

7.4 CONCLUSION AND PERSPECTIVES

Several enzyme architectures, have been designed, and produced using Hi-Fi Gibson assembly. Preliminary analysis of the suitability of the select enzymes' compatibility with the cellulosomal mode are positive. Moreover, we have been able to demonstrate value addition of the new constructs particularly those of the endoglucanase **(Ogonda *et al*., 2021; Chapter 5, Chapter 6).** These results are still preliminary, and further work is needed to have the packing of all the complexes and evaluate the stepwise value addition by enzymatic assays. At this stage therefore, it is not possible to accurately deduce the advantages of an in-line cellulosome compared to a natural one. However, in the In-line organization of the enzymes, we can anticipate several functional improvements:

- Enzyme activities will be located closer to the crystalline substrate
- Spatial organization of the enzymes would be easier to control, by contrast to the natural cellulosome where the cellulosome assembly operates randomly on the scaffoldin.
- In the In-line constructions, each building block contain an enzymatic activity, which facilitates the purification process and the stoichiometric assembly. By contrast, in natural

cellulosome the scaffoldin is devoid of enzymatic activity, which makes the purification process more difficult.

- In the In-line constructions, each building block has a low molecular mass that facilitates its heterologuous expression.

In silico protein - protein docking of the matching pairs during the *in vitro* complex assembly is an essential next step because with these positional constraints of the different enzymes would be minimized and help in predicting of the most stable arrangements of the enzymes for maximum function and stability (Stern et al. 2015). In addition, this would be helpful in finding alternative protein pairs for scaffoldin in designing and production of enzyme complexes. Recent advances in protein structure predictions using artificial intelligence (Alphafold from DEEPMIND, https://alphafold.ebi.ac.uk) should help in predicting the structure of these artificial complexes and the choice of best buildind blocks. Finally, the crystallization of the mini-complexes and the packed in-line complexes would provide a real structure of these complexes in an effort to understand the mechanism of the protein-to-protein interaction. This will be helpful in building or developing new protein affinity pairs for alternative scaffoldins.

7.5 SUPPLEMENTARY INFORMATION

7.5.1 Extracting cellulosome modules from *Clostridium thermocellum (Ct)* and *Clostridium josui (Cj)* gDNA

7.5.1.1 Designed Genes sequences for Cohesin and Dockerin modules to be extracted from *Ct* and *Cj*

1. Cohesin 2 *Clostridium josui* (422bp)

CGCGCTAGCTCCTACAAATGCTCTTAAAGTAACAGTAGGAACAGCAGAAGGTAATGTTGGAGA
AACAGTAACAGTTCCTGTTACATTTGCTGATGTAGCAAAAGTAAACAACGTAGGAACATGTAA
CTTCTATCTTGCATATGATGCAAGTCTTTTGGATGTAGTATCAGTAGATGCAGGTCCAATAGTTA
AGAATGCAGCAGTAAACTTCTCAAGCAGTGCAAGCAACGGAACAATCAGCTTCCTGTTCTTGG
ATAACACAATCACTGACGAATTGATTACTTCAGATGGTGTGTTTGCAAATATCACATTCAAATT

AAAGAATGTATCAACTAAAACAACAACACCAATAACCTTCAAAGACGGAGGAGCATTTGGTGA
CGGTACTATGTCAAAGATAACTACAGTTATCAAGACAAACGGTAGTGGTC GTCGAC TGG

2. **Cohesin 6 *Clostridium josui* (426bp)**

CGC GCTAGC ACAGGACTTGGAGTAAAGATTGCTTCAGTAACAGGAAAAACTGGTGATACTAT
AACAGTACCTGTAACTCTCAGCAATGTTGCTACAGTAGGTAATGTAGGAACATGTAATTTCT
ATATTACATATGACCCAACCTTGTTGCAGGCTGTATCAGCAACAGCTGGTGATATAGTAATA
AATGCACCTGTTAACTTCTCAAGCAGTATCAATGCAACAAATGGTACTATCAGTATTCTTTTC
CTTGATAACACTATAACTGATCAACCTATCGCAAGTGATGGAGTAATCACTAACCTTACTTTC
AAAGTATTAGGTTCTTCAAGCACAACTACTCCTATCGCTTTCAAAGCAGGTGGAGCATTCGG
TAATGGTAACATGGCAAAAATCAGTGATATCACATTCACAAATGGAAGTGCAAAACTTAATG
TCGAC TGG

3. **Cohesin 2 *Clostridium thermocellum* (419bp)**

GCG CATATG CGTCAGACGGTGTGGTAGTAGAAATTGGCAAAGTTACGGGATCTGTTGGAACT
ACAGTTGAAATACCTGTATATTTCAGAGGAGTTCCATCCAAAGGAATAGCAAACTGCGACTT
TGTGTTCAGATATGATCCGAATGTATTGGAAATTATAGGGATAGATCCCGGAGACATAATAG
TTGACCCGAATCCTACCAAGAGCTTTGATACTGCAATATATCCTGACAGAAAGATAATAGTA
TTCCTGTTTGCGGAAGACAGCGGAACAGGAGCGTATGCAATAACTAAAGACGGAGTATTTGC
AAAAATAAGAGCAACTGTAAAATCAAGTGCTCCGGGCTATATTACTTTCGACGAAGTAGGTG
GATTTGCAGATAATGACCTGGTAGAACAGAAGGTATCATTTATAGACGGTGGTGTTCTCGAG
CAC

4. **Xylanase 11A Dockerin *Clostridium thermocellum* (228bp)**

TT GAATTC TTCGACTCGAATGGTGTAAATCCTACACCCACCTCTCAGCCTCAACAAGGCCAG
GTTTTGGGTGACTTGAACGGAGACAAACAAGTAAATTCAACAGACTACACAGCACTGAAGA
GACATTTGCTCAATATAACCAGACTTTCAGGAACTGCTCTTGCCAACACCGATTTAAACGGT
GACAGCAAAGTTGATTCCACAGACCTTATGATTCTTCACAGATTTCTTCTC GTCGAC GG

5. **Xylanase 10CDockerin *Clostridium thermocellum* (231bp)**

TT GAATTC AGCGTTCCGCCGCTTCCGACAGAACCGCCGGTTCAGGTTATACCCGGTGATGTA
AACGGTGACGGTCGTGTAAATTCATCCGACTTGACTCTTATGAAAAGATACCTTTTAAAATC
CATAAGCGACTTCCCGACACCGGAAGGAAAAATTGCGGCGGATTTAAACGAAGACGGCAAG
GTAAACTCGACAGACTTGTTATCGCTGAAAAAACTCGTTCTGAGAGAACTTTGA **CTCGAG** **C**
GCC

6. **Aga27A Dockerin *Clostridium josui* (243bp)**

CGC GGATCC ACTGATACAAATATAGAGTTTGGTGATGTTGACGGAAATGGCATGATTGACGC
ATTAGATTATTCATTAGTAAAACGGTATTTGCTGGGCCAGATTTCTGATTGTCCTGATTCAAA
AGGCAAGCTTGCTGCTGATGTTGATGGAGACCAGCAAATAACAGCACTGGATTTTTCATTAA
TTAAGCAATACTTACTTGGGACTATTAACAAATTTCCTGCTCAAACAGCAAGTAAAATCAAG
CCA CTCGAG ATCC

Table 7.5.1: Primers used for extracting Cohesin and Dockerin modules from *Clostridum thermocellum* and *Clostridium josui*

Primer
1. CipActFor;TTCCAGGGGCCCCTGGGAAGATCTGCCACAATGACAGTCGAG
2. CipActRev;TGGGACGTCGACTTACTTTTCGAACTGCGGGTGGCTCCAGCTTGCCTGTGCGTC GTAATCACTTGATG
3. CIPAForB; cgccAG ATC TAT GAG AAA AGT CAT CAG TAT GCT CTT AG

4. CipARevB; atccCTCGAG CTGTGCGTCGTAATCACTTGATG
5. CipActDocTy2For;GGGGATCCAGCACCAGCCAGGCGCCG
6. CipActDocTy2Rev;GGAAGCTTCTGTGCGTCGTAATCACTTG
7. CipActDocTy2RevB;GGAAGCTTCTGCGCATCATAATCGCTGCTGGTCGCGCC
8. CipActDoctY2RevC; GGCTCGAGCTGCGCATCATAATCGCTGCTGGTCGCGCC
9. SdbActCohTy2For;CTGCGGATCCGTGGTGCCCAACGAGAGCCC
10. SdbActCohTy2Rev;ATGGCTGCAGGTTCACCTTCACGCCCTTGC
11. CBM3aCtFor;CTGCGGATCCCCCACCAACACCGCCACCCCCA
12. CBM3aCtRev;ATGGCTCGAGGCTGGGGGGGATGGTGGTGGC
13. Coh2ctTy1For;GCGCATATGCGTCAGACGGTGTGGTAGTAG
14. Coh2ctTy1Rev;CACCTCGAGAACACCACCGTCTATAAATGATACC
15. Coh4ctTy1For;GCGCATATGGATTCGAATGCAGTAAGGATTAAG
16. Coh4ctTy1Rev;GCGGTCGACTACTCCACCGTCAAAGAACTGT
17. CtXyn11ADocFor;TTGAATTCTTCGACTCGAATGGCGTAAATCC
18. CtXyn11ADocRev;GGGTCGACGAGAAGATATCTGTGTAGAATC
19. CtXyn10CDocFor;TTGAATTCAGCGTTCCGCCGCTTCCGAC
20. CtXyn10CDocRev;cgccCTCGAGTCAAAGTTCTCTCAGAACGAG
21. CtXyn10CDocRevB;GGCTCGAGTCAAAGTTCTCTCAGAACGAGTTT
22. Coh1cjFor;CGCGCTAGCATCAACGTGGGCG
23. Coh1cjRev;TGGGTCGACACTACACTACCACTTGTC
24. Coh2 cjFor;CGCGCTAGCTCCTACAAATGCTCTTAAAGT
25. Coh2 cjRev;TGGGTCGACCACTACCGTTTGTCTTGATAAC
26. Coh5cjFor;CGCGCTAGCGTGACCATCATCCCCGGC
27. Coh5cjRev;GGcGTCGACacactaccatttgtcttaacaa
28. Coh5cjFor;CGCGCTAGCGTGACCATCATC
29. Coh5cjRev;TGGGTCGACacactaccatttgtcttaacaa
30. Coh6cjFor;CGCGCTAGCacaggacttggagtaaagattg
31. Coh6cjRev;TGGGTCGACattaagttttgcacttccatttg
32. AgaDocFor;cgccGGATCCACTGATACAA ATATAG
33. AgaDocRev; ATCCCTCGAG TGG CTTGATTTTACTTGCTG

7.5.2. Translated Protein sequences for Cohesin and Dockerin modules:

1. **Cohesin 2 *Clostridium josui* protein sequence**
   PTNALKVTVGTAEGNVGETVTVPVTFADVAKVNNVGTCNFYLAYDASLLDVVSVDAGPI
   VKNAAVNFSSSASNGTISFLFLDNTITDELITSDGVFANITFKLKNVSTKTTTPITFKDGGAF
   GDGTMSKITTVIKTNGSG
2. **Cohesin 6 *Clostridium josui* protein sequence**
   TGLGVKIASVTGKTGDTITVPVTLSNVATVGNVGTCNFYITYDPTLLQAVSATAGDIVIN
   APVNFSSSINATNGTISILFLDNTITDQPIASDGVITNLTFKVLGSSSTTTPIAFKAGGA
   FGNGNMAKISDITFTNGSAKLN
3. **Cohesin 2 *Clostridium thermocellum* Protein sequence**
   SDGVVVEIGKVTGSVGTTVEIPVYFRGVPSKGIANCDFVFRYDPNVLEIIGIDPGDIIVD
   PNPTKSFDTAIYPDRKIIVFLFAEDSGTGAYAITKDGVFAKIRATVKSSAPGYITFDEVG
   GFADNDLVEQKVSFIDGGV
4. **Xylanase 11A Dockerin Protein sequence**
   FDSNGVNPTPTSQPQQGQVLGDLNGDKQVNSTDYTALKRHLLNITRLSGTALANTDLNGD
   SKVDSTDLMILHRFLL
5. **Xylanase 10C Dockerin Clostridium *thermocellum* Protein sequence**
   SVPPLPTEPPVQVIPGDVNGDGRVNSSDLTLMKRYLLKSISDFPTPEGKIAADLNEDGKV

NSTDLLSLKKLVLREL

6. **Aga27A Dockerin Clostridium *josui***

TDTNIEFGDVDGNGMIDALDYSLVKRYLLGQISDCPDSKGKLAADVDGDQQITALDFSLI
KQYLLGTINKFPAQTASKIKP

7.5.3 Reference Sequences from *Clostridium thermocellum (Ct)* and *Clostridium josui (Cj)*

1. **CiP A (5475bp) *Clostridium thermocellum***

```
      GCCACA     ATGACAGTCG AGATCGGCAA AGTTACAGCA

 121 GCCGTTGGAT CAAAAGTAGA AATACCTATA ACCCTGAAAG GAGTGCCATC CAAAGGAATG
 181 GCCAATTGCG ACTTCGTATT GGGTTATGAT CCAAATGTGC TGGAAGTAAC AGAAGTAAAA
 241 CCAGGAAGCA TAATAAAAGA TCCGGATCCT AGCAAGAGCT TTGATAGCGC AATATATCCG
 301 GATCGAAAGA TGATTGTATT TCTGTTTGCA GAAGACAGTG GAAGAGGAAC GTATGCAATA
 361 ACTCAGGATG GAGTATTTGC AACAATTGTA GCCACTGTCA AATCAGCTGC AGCGGCACCG
 421 ATTACTTTGC TTGAAGTAGG TGCATTTGCG GACAACGATT TAGTAGAAAT AAGCACAACT
 481 TTTGTCGCGG GCGGAGTAAA TCTTGGTAGT TCCGTACCGA CAACACAGCC AAATGTTCCG
 541 TCAGACGGTG TGGTAGTAGA AATTGGCAAA GTTACGGGAT CTGTTGGAAC TACAGTTGAA
 601 ATACCTGTAT ATTTCAGAGG AGTTCCATCC AAAGGAATAG CAAACTGCGA CTTTGTGTTC
 661 AGATATGATC CGAATGTATT GGAAATTATA GGGATAGATC CCGGAGACAT AATAGTTGAC
 721 CCGAATCCTA CCAAGAGCTT TGATACTGCA ATATATCCTG ACAGAAAGAT AATAGTATTC
 781 CTGTTTGCGG AAGACAGCGG AACAGGAGCG TATGCAATAA CTAAAGACGG AGTATTTGCA
 841 AAAATAAGAG CAACTGTAAA ATCAAGTGCT CCGGGCTATA TTACTTTCGA CGAAGTAGGT
 901 GGATTTGCAG ATAATGACCT GGTAGAACAG AAGGTATCAT TTATAGACGG TGGTGTTAAC
 961 GTTGGCAATG CAACACCGAC CAAGGGAGCA ACACCAACAA ATACAGCTAC GCCGACAAAA
1021 TCAGCTACGG CTACGCCCAC CAGGCCATCG GTACCGACAA ACACACCGAC AAACACACCG
1081 GCAAATACAC CGGTATCAGG CAATTTGAAG GTTGAATTCT ACAACAGCAA TCCTTCAGAT
1141 ACTACTAACT CAATCAATCC TCAGTTCAAG GTTACTAATA CCGGAAGCAG TGCAATTGAT
1201 TTGTCCAAAC TCACATTGAG ATATTATTAT ACAGTAGACG GACAGAAAGA TCAGACCTTC
1261 TGGTGTGACC ATGCTGCAAT AATCGGCAGT AACGGCAGCT ACAACGGAAT TACTTCAAAT
1321 GTAAAAGGAA CATTTGTAAA AATGAGTTCC TCAACAAATA ACGCAGACAC CTACCTTGAA
1381 ATAAGCTTTA CAGGCGGAAC TCTTGAACCG GGTGCACATG TTCAGATACA AGGTAGATTT
1441 GCAAAGAATG ACTGGAGTAA CTATACACAG TCAAATGACT ACTCATTCAA GTCTGCTTCA
```

1501 CAGTTTGTTG AATGGGATCA GGTAACAGCA TACTTGAACG GTGTTCTTGT ATGGGGTAAA

1561 GAACCCGGTG GCAGTGTAGT ACCATCAACA CAGCCTGTAA CAACACCACC TGCAACAACA

1621 AAACCACCTG CAACAACAAA ACCACCTGCA ACAACAATAC CGCCGTCAGA TGATCCGAAT

1681 GCAATAAAGA TTAAGGTGGA CACAGTAAAT GCAAAACCGG GAGACACAGT AAATATACCT

1741 GTAAGATTCA GTGGTATACC ATCCAAGGGA ATAGCAAACT GTGACTTTGT ATACAGCTAT

1801 GACCCGAATG TACTTGAGAT AATAGAGATA AAACCGGGAG AATTGATAGT TGACCCGAAT

1861 CCTGACAAGA GCTTTGATAC TGCAGTATAT CCTGACAGAA AGATAATAGT ATTCCTGTTT

1921 GCAGAAGACA GCGGAACAGG AGCGTATGCA ATAACTAAAG ACGGAGTATT TGCTACGATA

1981 GTAGCGAAAG TAAAATCCGG AGCACCTAAC GGACTCAGTG TAATCAAATT TGTAGAAGTA

2041 GGCGGATTTG CGAACAATGA CCTTGTAGAA CAGAGGACAC AGTTCTTTGA CGGTGGAGTA

2101 AATGTTGGAG ATACAACAGT ACCTACAACA CCTACAACAC CTGTAACAAC ACCGACAGAT

2161 GATTCGAATG CAGTAAGGAT TAAGGTGGAC ACAGTAAATG CAAAACCGGG AGACACAGTA

2221 AGAATACCTG TAAGATTCAG CGGTATACCA TCCAAGGGAA TAGCAAACTG TGACTTTGTA

2281 TACAGCTATG ACCCGAATGT ACTTGAGATA ATAGAGATAG AACCGGGAGA CATAATAGTT

2341 GACCCGAATC CTGACAAGAG CTTTGATACT GCAGTATATC CTGACAGAAA GATAATAGTA

2401 TTCCTGTTTG CGGAAGACAG CGGAACAGGA GCGTATGCAA TAACTAAAGA CGGAGTATTT

2461 GCTACGATAG TAGCGAAAGT AAAATCCGGA GCACCTAACG GACTCAGTGT AATCAAATTT

2521 GTAGAAGTAG GCGGATTTGC GAACAATGAC CTTGTAGAAC AGAAGACACA GTTCTTTGAC

2581 GGTGGAGTAA ATGTTGGAGA TACAACAGAA CCTGCAACAC CTACAACACC TGTAACAACA

2641 CCGACAACAA CAGATGATCT GGATGCAGTA AGGATTAAAG TGGACACAGT AAATGCAAAA

2701 CCGGGAGACA CAGTAAGAAT ACCTGTAAGA TTCAGCGGTA TACCATCCAA GGGAATAGCA

2761 AACTGTGACT TTGTATACAG CTATGACCCG AATGTACTTG AGATAATAGA GATAGAACCG

2821 GGAGACATAA TAGTTGACCC GAATCCTGAC AAGAGCTTTG ATACTGCAGT ATATCCTGAC

2881 AGAAAGATAA TAGTATTCCT GTTTGCGGAA GACAGCGGAA CAGGAGCGTA TGCAATAACT

2941 AAAGACGGAG TATTTGCTAC GATAGTAGCG AAAGTAAAAT CCGGAGCACC TAACGGACTC

3001 AGTGTAATCA AATTTGTAGA AGTAGGCGGA TTTGCGAACA ATGACCTTGT AGAACAGAAG

3061 ACACAGTTCT TTGACGGTGG AGTAAATGTT GGAGATACAA CAGAACCTGC AACACCTACA

3121 ACACCTGTAA CAACACCGAC AACAACAGAT GATCTGGATG CAGTAAGGAT TAAAGTGGAC

3181 ACAGTAAATG CAAAACCGGG AGACACAGTA AGAATACCTG TAAGATTCAG CGGTATACCA

3241 TCCAAGGGAA TAGCAAACTG TGACTTTGTA TACAGCTATG ACCCGAATGT ACTTGAGATA

3301 ATAGAGATAG AACCGGGAGA CATAATAGTT GACCCGAATC CTGACAAGAG CTTTGATACT

3361 GCAGTATATC CTGACAGAAA GATAATAGTA TTCCTGTTTG CAGAAGACAG CGGAACAGGA

3421 GCGTATGCAA TAACTAAAGA CGGAGTATTT GCTACGATAG TAGCGAAAGT AAAAGAAGGA

3481 GCACCTAACG GACTCAGTGT AATCAAATTT GTAGAAGTAG GCGGATTTGC GAACAATGAC

3541 CTTGTAGAAC AGAAGACACA GTTCTTTGAC GGTGGAGTAA ATGTTGGAGA TACAACAGAA

3601 CCTGCAACAC CTACAACACC TGTAACAACA CCGACAACAA CAGATGATCT GGATGCAGTA

3661 AGGATTAAAG TGGACACAGT AAATGCAAAA CCGGGAGACA CAGTAAGAAT ACCTGTAAGA

3721 TTCAGCGGTA TACCATCCAA GGGAATAGCA AACTGTGACT TTGTATACAG CTATGACCCG

3781 AATGTACTTG AGATAATAGA GATAGAACCG GGAGAATTGA TAGTTGACCC GAATCCTACC

3841 AAGAGCTTTG ATACTGCAGT ATATCCTGAC AGAAAGATGA TAGTATTCCT GTTTGCGGAA

3901 GACAGCGGAA CAGGAGCGTA TGCAATAACT GAAGATGGAG TATTTGCTAC GATAGTAGCG

3961 AAAGTAAAAT CCGGAGCACC TAACGGACTC AGTGTAATCA AATTTGTAGA AGTAGGCGGA

4021 TTTGCGAACA ATGACCTTGT AGAACAGAAG ACACAGTTCT TTGACGGTGG AGTAAATGTT

4081 GGAGATACAA CAGAACCTGC AACACCTACA ACACCTGTAA CAACACCGAC AACAACAGAT

4141 GATCTGGATG CAGTAAGGAT TAAAGTGGAC ACAGTAAATG CAAAACCGGG AGACACAGTA

4201 AGAATACCTG TAAGATTCAG CGGTATACCA TCCAAGGGAA TAGCAAACTG TGACTTTGTA

4261 TACAGCTATG ACCCGAATGT ACTTGAGATA ATAGAGATAG AACCGGGAGA CATAATAGTT

4321 GACCCGAATC CTGACAAGAG CTTTGATACT GCAGTATATC CTGACAGAAA GATAATAGTA

4381 TTCCTGTTTG CAGAAGACAG CGGAACGGGA GCGTATGCAA TAACTAAAGA CGGAGTATTT

4441 GCTACGATAG TAGCGAAAGT AAAAGAAGGA GCACCTAACG GACTCAGTGT AATCAAATTT

4501 GTAGAAGTAG GCGGATTTGC GAACAATGAC CTTGTAGAAC AGAAGACACA GTTCTTTGAC

4561 GGTGGAGTAA ATGTTGGAGA TACAACAGTA CCTACAACAT CGCCGACAAC AACACCGCCA

4621 GAGCCGACGA TAACTCCGAA CAAGTTGACA CTTAAGATAG GCAGAGCAGA AGGAAGACCT

4681 GGAGACACGG TGGAAATACC GGTTAACTTG TATGGAGTAC CTCAAAAAGG AATAGCAAGC

4741 GGTGACTTCG TAGTAAGCTA TGACCCGAAT GTACTTGAGA TAATAGAGAT AGAACCGGGA

4801 GAATTGATAG TTGACCCGAA TCCTACCAAG AGCTTTGATA CTGCAGTATA TCCTGACAGA

4861 AAGATGATAG TATTCCTGTT TGCGGAAGAC AGCGGAACAG GAGCGTATGC AATAACTGAA

4921 GATGGAGTAT TTGCTACGAT AGTAGCGAAA GTAAAAGAAG GAGCACCTGA AGGATTCAGT

4981 GCAATAGAAA TTTCTGAGTT TGGTGCATTT GCAGATAATG ATCTGGTAGA AGTGGAAACT

5041 GACCTTATCA ATGGTGGAGT ACTTGTAACT AATAAACCTG TAATAGAAGG ATATAAAGTA

5101 TCCGGATACA TTTTGCCAGA CTTCTCCTTC GACGCTACTG TTGCACCACT TGTAAAGGCC

5161 GGATTCAAAG TTGAAATAGT AGGAACAGAA TTGTATGCAG TAACAGATGC AAACGGATAC

5221 TTTGAAATAA CCGGAGTACC TGCAAATGCA AGCGGATATA CATTGAAGAT TTCAAGAGCA

5281 ACTTACTTGG ACAGAGTAAT TGCAAATGTT GTAGTAACGG GAGATACTTC AGTTTCAACT

5341 TCACAGGCTC CAATAATGAT GTGGGTAGGA GACATAGTGA AAGACAATTC TATCAACCTG

5401 TTGGACGTTG CAGAAGTTAT CCGTTGCTTC AACGCTACTA AAGGAAGCGC AAACTACGTA

5461 GAAGAACTTG ACATTAATAG AAACGGCGCA ATTAACATGC AAGACATAAT GATTGTTCAT

5521 CACTTTG GAGCTACATC AAGTGATTAC GACGCACAGT AA

2. **CiP A Dockerin Type II (225bp)**

ACTTCACA GGCTCCAATA ATGATGTGGG TA GGAGACAT AGTGAAAGAC AATTCTATCA
ACCTGTTGGA CGTTGCAGAA GTTATCCGTTGCTTCAACGC TACTAAAGGA AGCGCAAACT
ACGTAGAAGA ACTTGACATT AATAGAAACGGCGCAATTAA CATGCAAGAC ATAATGATTG
TTCATAAGCA CTTTGGA GCT ACATCAAGTG ATTACGACGC ACAGTAA

3. **Cohesin type II *Clostridium thermocellum* (585bp)**

GTGGTGCCCAACGAGAGCCCCGAGAACAACAACATGGACAACAGCAACCAGACCAGCAAC
ACCGCCAACAACGACGGCACCAGCAGCAGCGCCGAGGCCAACATCCAGATCGTGCTGGAC
AAGAACACCGCCAAGAAGGACGAGATCATCACCGCCAAGATCATCCTGAACAACATCCCC
AAGATCGCCGGCTACCAGGTGAACATCAAGTACGACCCCAACATCCTGCAGGCCGTGGAC
CTGGACACCGGCAAGCCCCTGGAGGACAAGCAGATCCCCGGCGGCGGCGACGTGCTGAGC
AACCCCGACTACAACGTGCTGCCCCTGGCCGCCAGCGACGTGAAGAACGGCGTGATCAAC
TTCGCCAAGGCCTACGTGAACGTGGACGAGTACAAGGAGAGCAACAACCCCGAGAGCAGC
GGCGTGCTGGCCCTGATCGGCTTCAAGGTGCTGAAGGAGGAGAGCACCGTGATCAGCTTC
GCCGACACCCCCAGCATGCCCAACGCCGTGAGCGGCACCTACGTGTACGACTGGGACTTC
AACGTGCTGACCAACTACAGCGTGGGCAAGGGCGTGAAGGTGAAC

4. **CBM 3a (510bp)**

CCCACCAACACCGCCACCCCCACCAAGAGCGCCACCGCCACCCCCACCAGGCCCAGCGTG
CCCACCAACACCCCCACCAACACCCCCGCCAACACCCCCGTGAGCGGCAACCTGAAGGTG
GAGTTCTACAACAGCAACCCCAGCGACACCACCAACAGCATCAACCCCCAGTTCAAGGTG
ACCAACACCGGCAGCAGCGCCATCGACCTGAGCAAGCTGACCCTGAGGTACTACTACACC
GTGGACGGCCAGAAGGACCAGACCTTCTGGTGCGACCACGCCGCCATCATCGGCAGCAAC
GGCAGCTACAACGGCATCACCAGCAACGTGAAGGGCACCTTCGTGAAGATGAGCAGCAGC
ACCAACAACGCCGACACCTACCTGGAGATCAGCTTCACCGGCGGCACCCTGGAGCCCGTG
GTGCCCAGCACCCAGCCCGTGACCACCCCCCCCGCCACCACCAAGCCCCCCGCCACCACC
AAGCCCCCCGCCACCACCATCCCCCCCAGC

5. **Coh 2 Typ1 ct (419bp)**

AATCTTGGTA GTTCCGTACC GACAACACAG CCAAATGTTC
CGTCAGACGG TGTGGTAGTA GAAATTGGCA AAGTTACGGG
ATCTGTT GGAACTACAG TTGAAATACC TGTATATTTC
AGAGGAGTTC CATCCAAAGG AATAGCAAAC TGCGACTTTG
TGTTCAGATA TGATCCGAAT GTATTGGAAA TTATAGGGAT
AGATCCCGGA GACATAATAG TTGACCCGAA TCCTACCAAG
AGCTTTGATA CTGCAATATA TCCTGACAGA AAGATAATAG
TATTCCTGTT TGCGGAAGAC AGCGGAACAG GAGCGTATGC
AATAACTAAA GACGGAGTAT TTGCAAAAAT AAGAGCAACT
GTAAAATCAA GTGCTCCGGG CTATATTACT TTCGACGAAG
TAGGTGGATT TGCAGATAAT GACCTGGTAG AACAGAAGGT

ATCATTTATA GACGGTGGTG TT

6. **Coh 4 Typ1 *Clostridium thermocellum* (429bp)**

AATGT TGGAGATACA ACAGTACCTA CAACACCTAC AACACCTGTA ACAACACCGA
CAGATGATTC GAATGCA GTA AGGATTAAGG TGGACACAGT AAATGCAAAA CCGGGAGACA
CAGTAAGAAT ACCTGTAAGATTCAGCGGTA TACCATCCAA GGGAATAGCA AACTGTGACT
TTGTATACAG CTATGACCCGAATGTACTTG AGATAATAGA GATAGAACCG GGAGACATAA
TAGTTGACCC GAATCCTGAC AAGAGCTTTG ATACTGCAGT ATATCCTGAC AGAAAGATAA
TAGTATTCCT GTTTGCGGAAGACAGCGGAA CAGGAGCGTA TGCAATAACT AAAGACGGAG
TATTTGCTAC GATAGTAGCGAAAGTAAAAT CCGGAGCACC TAACGGACTC AGTGTAATCA
AATTTGTAGA AGTAGGCGGATTTGCGAACA ATGACCTTGT AGAACAGAAG ACACAGTTCT
TTGACGGTGG AGTA

7. **Xylanase 11A Dockerin *Clostridium thermocellum***

TTCGAC TCGAATGGCG TAAATCCTAC ACCCACCTCT CAGCCTCAAC AAGGCCAGGT
TTTGGGTGAC TTGAACGGAG ACAAACAAGT AAATTCAACA GACTACACAG CACTGAAGAG
ACATTTGCTC AATATAACCA GACTTTCAGG AACTGCTCTT GCCAACGCCG ATTTAAACGG
TGACGGCAAA GTTGATTCCA CTGACCTTAT GATTCTACAC AGATATCTTC TC

8. **Xylanase 10C Dockerin *Clostridium thermocellum***

AGCGTTCCGC CGCTTCCGAC AGAACCGCCG GTTCAGGTTA TACCCGGTGA TGTAAACGGT
GACGGTCGTG TAAATTCATC CGACTTGACT CTTATGAAAA GATACCTTTT AAAATCCATA
AGCGACTTCC CGACACCGGA AGGAAAAATT GCGGCGGATT TAAACGAAGA CGGCAAGGTA
AACTCGACAG ATTTGTTAGC GCTGAAAAAA CTCGTTCTGA GAGAACTTTG A

9. **Cohesin 2 *Clostridium josui***

GTTA TAAGTCCAGA TCCTACAAAT GCTCTTAAAG TAACAGTAGG AACAGCAGAA
1321 GGTAATGTTG GAGAAACAGT AACAGTTCCT GTTACATTTG CTGATGTAGC AAAAGTAAAC
1381 AACGTAGGAA CATGTAACTT CTATCTTGCA TATGATGCAA GTCTTTTGGA TGTAGTATCA
1441 GTAGATGCAG GTCCAATAGT TAAGAATGCA GCAGTAAACT TCTCAAGCAG TGCAAGCAAC
1501 GGAACAATCA GCTTCCTGTT CTTGGATAAC ACAATCACTG ACGAATTGAT TACTTCAGAT
1561 GGTGTGTTTG CAAATATCAC ATTCAAATTA AAGAATGTAT CAACTAAAAC AACAACACCA
1621 ATAACCTTCA AAGACGGAGG AGCATTTGGT GACGGTACTA TGTCAAAGAT AACTACAGTT
1681 ATCAAGACAA ACGGTAGTGT AA

10. **Cohesin6 *Clostridium josui***

ACTATCG ATGTTGGTGA GCCTGTTGTT ACAGGACTTG GAGTAAAGAT TGCTTCAGTA
ACAGGAAAAA CTGGTGATACTATAACAGTA CCTGTAACTC TCAGCAATGT TGCTACAGTA
GGTAATGTAG GAACATGTAA TTTCTATATT ACATATGACC CAACCTTGTT GCAGGCTGTA
TCAGCAACAG CTGGTGATAT AGTAATAAAT GCACCTGTTA ACTTCTCAAG CAGTATCAAT
GCAACAAATG GTACTATCAG TATTCTTTTC CTTGATAACA CTATAACTGA TCAACCTATC
GCAAGTGATG GAGTAATCAC TAACCTTACT TTCAAAGTAT TAGGTTCTTC AAGCACAACT
ACTCCTATCG CTTTCAAAGC AGGTGGAGCA TTCGGTAATG GTAACATGGC AAAAATCAGT
GATATCACAT TCACAAATGG AAGTGCAAAA CTTAAT

11. **AGA27A Dockerin *Clostridium josui***

1 ACTGATACAA ATATAGAGTT TGGTGATGTT GACGGAAATG GCATGATTGA CGCATTAGAT
61 TATTCATTAG TAAAACGGTA TTTGCTGGGC CAGATTTCTG ATTGTCCTGA TTCAAAAGGC
121 AAGCTTGCTG CTGATGTTGA TGGAGACCAG CAAATAACAG CACTGGATTT TTCATTAATT
181 AAGCAATACT TACTTGGGAC TATTAACAAA TTTCCTGCTC AAACAGCAAG TAAAATCAAG
241 CCATAA

CHAPTER 8

GENERAL CONCLUSIONS

From this work, we have been able to identifiy two new glycosyl hydrolases, BpGH9 and BpGH48, which were successfully cloned from the thermophile cellulolytic producer *Bacillus pumilus*. These recombinant cellulases (β-1, 4 endoglucanase and Cellobiohydrolase) were over-expressed in *E. coli* in order to characterize their kinetic properties. Both enzyme have a pH optimum around 6, which is far below the alkaline pH of the natural medium, from which the *B. pumilus* was isolated. This suggests that both enzymes are intracellular in *B. pumilus*, which is confirmed by the absence of an exportation sequence in their sequence. However, BpGH9 has high activity and stability at high pH, demonstrating that this enzyme could have a potential interest in several industrial environnments.

Moreover, it was demonstrated that these cellulases from *Bacillus pumilus* were compatible with the cellulosomal mode of action; very tolerant to fusion with other protein modules such CBM, YFP or scalfoldin, while retaining their activity. Therefore, we were able to successfully convert an aerobic, non-cellulosomal cellulase into cellulosomal mode of action.

Enzyme architechture plays an important role in the functionality and stability of glycosyl hydrolase enzymes, consequently by attaching the additional modules (Cohesin, Dockerin and/ CBM3a), we observed change in thermostability and mechanism of the endoglucanase; BpGH9 was converted into a processive enzyme. The CBM3a module here although added in its non-native position (at the N-terminal) showed additive synergy with the native catalytic, C-terminal, CBM3c hence lending to microcrystalline substrate localization and increased activity. In addition, enzyme processivity and regioselectivity can be controlled by rational design of either the individual enzymes or enzymes' complexes.

An interesting application of these molecular engineering was developed to target the cellulose in the *B.braunii* colonies. These microalgaes have a great potential to produce natural hydrocarbons, but up to now, their extraction is rather difficult and cannot be realized in a continuous mode. We have demonstrated that the engineering of cellulases could facilate microalgae colonies' disorganization without killing the cells.

In a final approach, we envisioned to design new cellulosomes, that can be controlled in terms of stoichiometry and relative positioning of enzymes on the scaffold. In so doing the composition of enzymes/ complexes produced can be predictable unlike the native set–up wherein the cellulosomes produced are of varied composition both in terms of positioning and composition. This predictability is particularly essential in industry where product uniformity is required.

This work therefore provides a basis for the rational design of GHs into complexes with defined relative positioning and stoichiometry for specific processing needs.

CHAPTER 9

GENERAL PERSPECTIVES

By now, it is evident that the construction and evaluation of such value added, functional and 'designer' cellulosome complexes is a continuous process. Evaluation of the emerging constructs against the baseline constructs is therefore essential in analyzing the value added.

In addition, it is essential to optimise conditions for the production and expression of the native cellulase enzymes as well as the new enzyme constructs. Similarly, it is important to have the optimisation of the conditions for the production and expression of the scaffoldin proteins

Further, unlike the *Clostridium thermocellum* cellulosome that is well studied and characterised, it would be necessary to conduct crystallisation and 3D structure determination of *Clostridium josui* modules herein utilised. This is in addition to the chimera mini-complexes, and non-native fusion proteins such as the CBM3a used in this study.

Moreover, we recommend the determination of the mechanism of action of the glycosyl hydrolases from this study in order to ascertain for the control for products of interest particularly during industrial application.

In addition to enzyme mechanism assays, it would be important to have structural studies of the native enzymes as well as the mini-complex pieces and complexes in order to allow for the rational development of inhibitors. This is especially essential in the industrial medical application of cellulases/ cellulase complexes and future studies that use this research as inspiration. In addition, substrate/cellulose material analogues can also be developed to make production less costly.

Finally, using these prototypes as a guide, we could have more controlled assemblies of scaffoldin and their associated enzymes into defined cellulosome complexes tailored for various

lignocellulosic substrates and or various industrial functions as well as for other enzymes using alternative scaffoldin pairs.

CHAPTER 10

REFERENCES

Adams, J. J., Pal, G., Jia, Z., & Smith, S. P. (2006). Mechanism of bacterial cell-surface attachment revealed by the structure of cellulosomal type II cohesin-dockerin complex. Proc Natl Acad Sci U S A, 103(2), 305-310.

Anderson, T. D., Robson, S. A., Jiang, X. W., Malmirchegini, G. R., Fierobe, H. P., Lazazzera, B. A (2011). Assembly of minicellulosomes on the surface of Bacillus subtilis. Appl Environ Microbiol, 77(14), 4849-4858.

Bayer, E. A., Kenig, R., & Lamed, R. (1983). Adherence of Clostridium thermocellum to cellulose. J Bacteriol, 156(2), 818-827.

Bayer, E. A., Setter, E., & Lamed, R. (1985). Organization and distribution of the cellulosome in Clostridium thermocellum. J Bacteriol, 163(2), 552-559.

Bayer, E. A., Belaich, J. P., Shoham, Y., & Lamed, R. (2004). The cellulosomes: multienzyme machines for degradation of plant cell wall polysaccharides. Annu Rev Microbiol, 58, 521-554.

Boder, E. T., and K. D. Wittrup. (1997). Yeast surface display for screening combinatorial polypeptide libraries. Nat. Biotechnol. 15: 553–557.

Cavaco-Paulo, A. (1998). Processing Textile Fibers with Enzymes: An Overview. In: Enzyme Applications in Fiber Processing, Eriksson, K.E. and A. Cavaco-Paulo (Eds.). American Chemical Society Washington, DC. ISBN-13: 9780841235472, pp: 180-189.

Cho, H. Y., Yukawa, H., Inui, M., Doi, R. H., & Wong, S. L. (2004). Production of minicellulosomes from Clostridium cellulovorans in Bacillus subtilis WB800. Appl Environ Microbiol, 70(9), 5704-5707.

Conrado, R. J., Varner, J. D., & DeLisa, M. P. (2008). Engineering the spatial organization of metabolic enzymes: mimicking nature's synergy. Curr Opin Biotechnol, 19(5), 492-499.

Csiszar, E., G. Szakacs and I, Rusznak. (1998). Bioscouring of Cotton Fabrics with Cellulase Enzyme. In: Enzyme Applications in Fiber Processing, Eriksson, K.E. and A. Cavaco-Paulo (Eds.). American Chemical Society, Washington, DC.Chapter

Delebecque, C. J., Lindner, A. B., Silver, P. A., & Aldaye, F. A. (2011). Organization of intracellular reactions with rationally designed RNA assemblies. Science, 333(6041), 470-474.

Dion, M., Fourage, L., Hallet, JN., Bernard Colas. (1999) Cloning and expression of a β-glycosidase gene from Thermus thermophilus. Sequence and biochemical characterization of the encoded enzyme. Glycoconj J 16, 27–37. https://doi.org/10.1023/A:1006997602727

Dueber, J. E., Wu, G. C., Malmirchegini, G. R., Moon, T. S., Petzold, C. J., Ullal, A. V. (2009). Synthetic protein scaffolds provide modular control over metabolic flux. Nat Biotechnol, 27(8), 753-759.

Dubois, M., K. A. Gilles, J. K. Hamilton, P. A. Rebers, and F. Smith. (1956).Colorimetric method for determination of sugars and related substances. Anal. Chem. 28:350–356.

Esther Menendez , Paula Garcia-Fraile , and Raul Rivas. (2015) Biotechnological applications of bacterial cellulases. Aims Bioengeering.Volume 2, Issue 3, 163-182.DOI: 10.3934/bioeng.2015.3.163

Fierobe, H. P., Mechaly, A., Tardif, C., Belaich, A., Lamed, R., Shoham, Y. (2001). Design and production of active cellulosome chimeras. Selective incorporation of dockerin-containing enzymes into defined functional complexes. J Biol Chem, 276(24), 21257-21261.

Fierobe, H. P., Bayer, E. A., Tardif, C., Czjzek, M., Mechaly, A., Belaich, A. (2002). Degradation of cellulose substrates by cellulosome chimeras. Substrate targeting versus proximity of enzyme components. J Biol Chem, 277(51), 49621-49630.

Fierobe, H.-P., F. Mingardon, A. Mechaly, A. Belaich, M. T. Rincon, S.Pages, R. Lamed, C. Tardif, J. P. Belaich, and E. A. Bayer. (2005). Action of designer cellulosomes on homogeneous versus complex substrates—controlled incorporation of three distinct enzymes into a defined trifunctional scaffoldin. J. Biol. Chem. 280:16325–16334.

Gold, N. D., & Martin, V. J. (2007). Global view of the Clostridium thermocellum cellulosome revealed by quantitative proteomic analysis. J Bacteriol, 189(19), 6787-6795.

Gómez-González , Soledad , José Ruiz-Jiménez , Feliciano Priego-Capote, and María Dolores Luque de Castro. (2010). Qualitative and Quantitative Sugar Profiling in Olive Fruits, Leaves, and Stems by Gas Chromatography−Tandem Mass Spectrometry (GC-MS/MS) after Ultrasound-Assisted Leaching. J. Agric. Food Chem., 58 (23), pp 12292–12299

Hui yong Feng, Jullien Drone, Lionel Hoffmann, Vinh Tran, Charles Tellier, Claude Rabiller, Michel Dion. Converting a beta-glycosidase into a beta-transglycosidase by directed evolution. Journal of Biological Chemistry, American Society for Biochemistry and Molecular Biology, 2005, 280 (44), pp.37088-37097.10.1074/jbc.M502873200 .hal-00385576

Ito, J., Kosugi, A., Tanaka, T., Kuroda, K., Shibasaki, S., Ogino, C. (2009). Regulation of the display ratio of enzymes on the Saccharomyces cerevisiae cell surface by the immunoglobulin G and cellulosomal enzyme binding domains. Appl Environ Microbiol, 75(12), 4149-4154.

Jamroz Marta K, Katarzyna Paradowska, Katarzyna Zawada, Katerina Makarova, Sławomir Ka´zmierskia, and Iwona Wawera (2014). 1H and 13C NMR-based sugar profiling with chemometric analysis and antioxidant activity of herbhoneys and honeys. J Sci Food Agric.; 94: 246–255

Karmakar, M and Ray, R.R. (2011). Current Trends in Research and Application of Microbial Cellulases. Research Journal of Microbiology, 6: 41-53.

Kosugi, A., Amano, Y., Murashima, K., & Doi, R. H. (2004). Hydrophilic domains of scaffolding protein CbpA promote glycosyl hydrolase activity and localization of cellulosomes to the cell surface of Clostridium cellulovorans. J Bacteriol, 186(19), 6351-6359.

Krauss, Jan., Vladimir V. Zverlov and Wolfgang H. Schwarz (2012). In Vitro Reconstitution of the Complete Clostridium thermocellum Cellulosome and Synergistic Activity on crystalline Cellulose. Applied and Environmental Microbiology, 78 (12) p. 4301–4307

Kruus, K., Lua, A. C., Demain, A. L., & Wu, J. H. (1995). The anchorage function of CipA (CelL), a scaffolding protein of the Clostridium thermocellum cellulosome. Proc Natl Acad Sci U S A, 92(20), 9254-9258.

Lee, H., Deloache, W. C., & Dueber, J. E. (2011). Spatial organization of enzymes for metabolic engineering. Metab Eng, 14(3), 242-251.

Leibovitz, E., & Beguin, P. (1996). A new type of cohesin domain that specifically binds the dockerin domain of the Clostridium thermocellum cellulosome-integrating protein CipA. J Bacteriol, 178(11), 3077-3084.

Lemaire, M., Ohayon, H., Gounon, P., Fujino, T., & Beguin, P. (1995). OlpB, a new outer layer protein of Clostridium thermocellum, and binding of its S-layer-like domains to components of the cell envelope. J Bacteriol, 177(9), 2451-2459.

Lilly, M., Fierobe, H. P., van Zyl, W. H., & Volschenk, H. (2009). Heterologous expression of a Clostridium minicellulosome in Saccharomyces cerevisiae. FEMS Yeast Res, 9(8), 1236-1249.

Lynd, L. (1996). Overview and evaluation of fuel ethanol from cellulosic biomass: technology, economics, the environment, and policy. Annu. Rev. Energy. Environ., 21, 403-465.

Lynd, L. R., van Zyl, W. H., McBride, J. E., & Laser, M. (2005). Consolidated bioprocessing of cellulosic biomass: an update. Curr Opin Biotechnol, 16(5), 577-583.

Lynd, L. R., Weimer, P. J., van Zyl, W. H., & Pretorius, I. S. (2002). Microbial cellulose utilization: fundamentals and biotechnology. Microbiol Mol Biol Rev, 66(3), 506-577, table of contents.

Map enzymes ® (www.mapsenzymes.com)

Mayer, F., Coughlan, M. P., Mori, Y., & Ljungdahl, L. G. (1987). Macromolecular Organization of the Cellulolytic Enzyme Complex of Clostridium thermocellum as Revealed by Electron Microscopy. Appl Environ Microbiol, 53(12), 2785-2792.

Miller, G. L. (1959). Use of dinitrosalicylic acid reagent for determination of reducing sugars. Analytical Chemistry, 31, 426-429.

Mingardon, F., Chanal, A., Tardif, C., Bayer, E. A., & Fierobe, H. P. (2007). Exploration of new geometries in cellulosome-like chimeras. Appl Environ Microbiol, 73(22), 7138-7149.

Muhammad Imran Khan, Jin Hyuk Shin and Jong Deog Kim. Microbial Cell Factoriesz (2018)

https://doi.org/10.1186/s12934-018-0879-x

Murashima, K., Chen, C. L., Kosugi, A., Tamaru, Y., Doi, R. H., & Wong, S. L. (2002). Heterologous production of Clostridium cellulovorans engB, using protease-deficient Bacillus subtilis, and preparation of active recombinant cellulosomes. J Bacteriol, 184(1), 76-81.

Murashima, K., Kosugi, A., & Doi, R. H. (2002). Synergistic effects on crystalline cellulose degradation between cellulosomal cellulases from Clostridium cellulovorans. J Bacteriol, 184(18), 5088-5095.

Miyoshi, A., Poquet, I., Azevedo, V., Commissaire, J., Bermudez-Humaran, L., Domakova, E. (2002). Controlled production of stable heterologous proteins in Lactococcus lactis. Appl Environ Microbiol, 68(6), 3141-3146.

Novozymes. (www.novozymes.com)

Pere, J., M. Siika-Aho., L. Viikari., S. Liukkonen., Gullichsen, J. (1996). Use of purified enzymes in mechanical pulping. TAPPI Pulping Con., 2: 693-696.

Perret, S., Casalot, L., Fierobe, H. P., Tardif, C., Sabathe, F., Belaich, J. P. (2004). Production of heterologous and chimeric scaffoldins by Clostridium acetobutylicum ATCC 824. J Bacteriol, 186(1), 253-257.

Raman, B., Pan, C., Hurst, G. B., Rodriguez, M., Jr., McKeown, C. K., Lankford, P. K. (2009). Impact of pretreated Switchgrass and biomass carbohydrates on Clostridium thermocellum ATCC 27405 cellulosome composition: a quantitative proteomic analysis. PLoS One, 4(4), e5271.

Rincon, M. T., Cepeljnik, T., Martin, J. C., Lamed, R., Barak, Y., Bayer, E. A. (2005). Unconventional mode of attachment of the Ruminococcus flavefaciens cellulosome to the cell surface. J Bacteriol, 187(22), 7569-7578.

Rincon, M. T., Ding, S. Y., McCrae, S. I., Martin, J. C., Aurilia, V., Lamed, R. (2003). Novel organization and divergent dockerin specificities in the cellulosome system of Ruminococcus flavefaciens. J Bacteriol, 185(3), 703-713.

Sabathe, F., Belaich, A., & Soucaille, P. (2002). Characterization of the cellulolytic complex (cellulosome) of Clostridium acetobutylicum. FEMS Microbiol Lett, 217(1), 15-22.

Sabathe, F., & Soucaille, P. (2003). Characterization of the CipA scaffolding protein and in vivo production of a minicellulosome in Clostridium acetobutylicum. J Bacteriol, 185(3), 1092-1096.

Sampson, E. M., & Bobik, T. A. (2008). Microcompartments for B12-dependent 1,2-propanediol degradation provide protection from DNA and cellular damage by a reactive metabolic intermediate. J Bacteriol, 190(8), 2966-2971.

Schwarz, W. H. (2001). The cellulosome and cellulose degradation by anaerobic bacteria. Appl Microbiol Biotechnol, 56(5-6), 634-649.

Tolan, J.S. and Foody, B. (1999). Cellulase from Submerged Fermentation. In: Advances in Biochemical Engineering/Biotechnology, Scheper, Th. (Ed.). Vol. 65. Springer, Berlin / Heidelberg, ISBN: 978-3-540-65577-0, pp: 41-67.

Tsai, Shen-Long Jeongseok Oh, Shailendra Singh, Ruizhen Chen and Wilfred Chen. (2009).Functional assembly of minicellulosome on Saccharomyces cerevisiae cell surface for cellulose hydrolysis and ethanol production. Appl. Environ. Microbiol. 2009, 75(19):6087. DOI: 10.1128/AEM.01538-09.

Tyurin, M. V., Desai, S. G., & Lynd, L. R. (2004). Electrotransformation of Clostridium thermocellum. Appl Environ Microbiol, 70(2), 883-890.

Verenium-cottonase ® http://www.verenium.com/prod_cottonase.html

Watanabe, H., & Tokuda, G. (2010). Cellulolytic systems in insects. Annu Rev Entomol, 55, 609-632.

Weiczorek, A., (2013). Engineering Lactococcus lactis for the scaffold protein-mediated surface display of recombinant enzymes. PhD . Concordia University.

Wieczorek, A. S., & Martin, V. J. (2010). Engineering the cell surface display of cohesins for assembly of cellulosome-inspired enzyme complexes on Lactococcus lactis. Microb Cell Fact, 9, 69.

Wen, F., Sun, J., & Zhao, H. (2010). Yeast surface display of trifunctional minicellulosomes for simultaneous saccharification and fermentation of cellulose to ethanol. Appl Environ Microbiol, 76(4), 1251-1260.

Wenzel, M., Schonig, I., Berchtold, M., Kampfer, P., & Konig, H. (2002). Aerobic and facultatively anaerobic cellulolytic bacteria from the gut of the termite Zootermopsis angusticollis. Journal of Apllied Microbiology, 32-40.

Wilson, D.B (1992) Biochemistry and genetics of actinomycetes cellulases. Critical review Biotechnol. 12, 45-63

Wilson, D.B., and Irwin , D.C.(1999) Genetics and properties of cellulases. Adv. Biochem. Eng 65, 1-21

Wong, S. L. (1995). Advances in the use of Bacillus subtilis for the expression and secretion of heterologous proteins. Curr Opin Biotechnol, 6(5), 517-522.

Wu, C. H., Mulchandani, A., & Chen, W. (2008). Versatile microbial surface-display for environmental remediation and biofuels production. Trends Microbiol, 16(4), 181-188.

Xu, Q., Gao, W., Ding, S. Y., Kenig, R., Shoham, Y., Bayer, E. A. (2003). The cellulosome system of Acetivibrio cellulolyticus includes a novel type of adaptor protein and a cell surface anchoring protein. J Bacteriol, 185(15), 4548-4557.

Zverlov, V. V., Kellermann, J., & Schwarz, W. H. (2005). Functional subgenomics of *Clostridium thermocellum* cellulosomal genes: identification of the major catalytic components in the extracellular complex and detection of three new enzymes. Proteomics, 5(14), 3646-3653.

CHAPTER 11

APPENDICES

11.1 Résumé d'une revue rédigée au cours de la these et parue dans: Food Security and Safety; African Perspectives (Springer Nature publisher)

Cette revue est complimentaire du chapitre 6 de cette thèse. La sécurité alimentaire est essentielle à la survie et aux moyens de subsistance des communautés. Pour qu'une communauté bénéficie de la sécurité alimentaire et nutritionnelle, elle doit avoir un accès ininterrompu à une alimentation suffisante, de qualité, sûre et nutritive. Une communauté peut être confrontée à l'insécurité alimentaire en raison d'une variété réduite d'aliments diététiques et d'une alimentation déséquilibrée. Ceci est généralement dû au coût et au manque d'informations nutritionnelles sur les aliments. Par conséquent, la malnutrition résulte de la réduction de la taille des repas ou des sauts de repas. Il est donc impératif de mener une évaluation nutritionnelle approfondie des communautés pour aider à maximiser les sources de nourriture disponibles ainsi que la recherche de modèles de sources de nourriture alternatifs en fonction de leurs besoins et des ressources alimentaires disponibles. Ce chapitre se concentre sur Dagaa, une espèce de poisson inférieure connue localement sous le nom de « Omena» (*Rastrineobola argentea*). Nous présentons des informations de composition de Dagaa qui sont importantes pour créer une prise de conscience nutritionnelle au sein des communautés et éclairer leurs décisions concernant la planification des repas. En outre, ce chapitre fournit des informations sur les utilisations alternatives de l'excédent de la ressource biologique de poisson Dagaa après le tri, basé sur la qualité, effectué lors des débarquements de poisson. En outre, nous mentionnons divers canaux potentiels pour l'inclusion et les utilisations des microalgues dans l'amélioration de la sécurité alimentaire et nutritionnelle.

Author Contributions

Lydia Awuor Ogonda is the first author of this work. Having received an invite to contribute to the book through a network of women scientists in the developing world. I designed and thought out the framework of this chapter in-Line with my current research topic on Lignocellulose efficient utilization. In addition, I wrote out the first draft of the work which was edited and panel beaten by my mentor, thesis supervisor and co-author in this work; Dr. Edward Kirwa Muge. My senior co-author, given his years of experience and role in my thesis is the corresponding author in this work.

HARNESSING THE POTENTIAL OF UNDERUTILIZED AQUATIC BIORESOURCE FOR FOOD AND NUTRITIONAL SECURITY IN KENYA

Ogonda Lydia Awuor[1, 2,] Muge Edward Kirwa[1]

1. The University of Nairobi, School of medicine, College of health sciences, P.O Box 30197-00100, University of Nairobi. Kenya
2. L'Universite de Nantes, Unité Fonctionnalité et Ingénierie De Protéines (UFIP), UMR CNRS 6286, UFR SCIENCES ET TECHNIQUES, 2, rue de la Houssinière, 44322 Nantes Cedex 03, France

*Lydia.Ogonda@gmail.com/*Mugeek@uonbi.ac.ke

Abstract:

Food security is critical to the survival and livelihood of communities. For a community to enjoy food and nutritional security, it must have uninterrupted access to sufficient, quality, safe and nutritionally dense food. A community can experience food insecurity because of reduced variety of diet foods and imbalanced diets. This is usually because of cost and the lack of food nutritional information. Consequently, malnutrition results due to reduced meal sizes or skipped meals. It is therefore imperative to conduct in-depth nutritional assessment of communities to help maximize the available food sources as well as tailor for alternative food source models based on its needs and available food resources.

This chapter will focus on *Dagaa*, a lower end fish species locally known as 'Omena' (*Rastrineobola argentea*). We will display its proximate composition information that is important for creating nutritional awareness within the communities and informing their decisions around meal planning. In addition, this chapter will inform on alternative uses of the surplus *Dagaa* fish bio resource once quality-based grading is done on the fish landings.

Keywords: *Dagaa* fish, Lake Victoria, Proximate composition, Protein hydrolysate, Food security

1.1 INTRODUCTION

FAO has termed food and nutritional insecurity as the most evident sign of poverty and extreme hardship (Béné, 2005). Although prolonged hunger is on a decline in Africa, both absolute and relative malnutrition remains rampant (Béné, 2005). In 2003, FAO classified more than 34% of the African population as malnourished following a steep rise in cases (9M) from 1996. Malnutrition has histrionic irreparable effects on socio-economic as well as physical development of the affected populations. It is especially important to mention the dire and fatal effects on African pregnant and lactating women as well as infants. The statistics show that about 20,000 women (approximately 55 each day) die from malnutrition-related (Iron deficiency anaemia) cases. In addition, over 500,000 African infants die each year in Vitamin A deficiency-related cases (Béné, 2005; UNICEF, 2004, 2009). This malnutrition associated infant mortality cases reinforce the vital role of food in child and maternal health. As a result, food is an important component of essential medicine particularly for paediatric and maternal care (UNICEF, 2017).

If improved, the fish industry is one of the sectors that has a great potential in contributing towards the alleviation of food insecurity directly as a food source in addition to indirect contribution through income generation and employment. More so, value addition activities in the fish industry such as the production of fish protein hydrolysate (FPH) further raise the capacity of this industry through the resultant bioactive products obtained.

FAO classifies fish as a high-quality protein. Consequently, fish and its associated aquatic source foods could be a source of “rich” food for the poor African population in rural and

peri-urban areas. Particularly, the lower end underutilized fish species that do not have competing uses and are relatively cheap and easily available due to the large fish landings (Béné, 2005; FAO, WHO, & UNU, 1985; Ogonda *et al*., 2014). Other than directly through food, fish also contribute indirectly to food security through trade and export. Data has shown that fish import receipts have supported up to 50% food import bills of low-income food deficient countries. (Béné, 2005)

Lacustrine/lentic as well as lotic water ecosystems play an important role in the food chain. These water ecosystems have different characteristics and therefore support different life forms (Manyala *et al.*,2011). These different aquatic life forms vary in composition based on the occupied zone of habitation (Ogonda *et al*.,2014; Owaga *et al*., 2010; Suseno *et al.,* 2010). These intrinsic differences necessitate the individual study of the diverse ecosystems available. Further, it is important to characterize the supported life forms of different aquatic ecosystems. This dual characterization of both the ecosystems and fish populations available will aid in the nutritional information required in nutritional and meal-planning decisions that could develop and maintain this fish food preference, as has been the case in countries such as Ghana, Senegal and Gambia (Béné, 2005). Further, this information will inform on the sustainability and or efficient utilization of the fish/Aquatic bioresource following harvesting.

2.1 AQUATIC BIORESOURCE CONSUMPTION AND HEALTH WORLDWIDE

Aquatic bioresources have several important health and nutritional benefits. In addition, fish is a major animal protein source; making up to 20% of the total animal protein intake worldwide. It accounts for 50% animal protein in Indonesia, Solomon Islands and Bangladesh, 22% in Sub Saharan countries and up to 50% in West African countries (R. H. Thurstan & Roberts, 2014).

Besides fish, aquatic bioresources also comprise of phytoplankton that includes dinoflagellates, diatoms, and macrophytes such as red, green and brown algae (E. A. Shalaby, 2011). Being photosynthetic, algae are important in ocean productivity and form the basis of the aquatic food chain. Macroalgae in Egypt are mostly wild-growing along the craggy shores of the red and Mediterranean seas and are not yet used for food. However, In China, Japan, and Hawaii, macroalgae are used in medicine and for food (E. Shalaby, 2011).

3.1 HEALTH BENEFITS OF AQUATIC BIORESOURCE

3.1.1 Consumed aquatic organisms (fish plus seaweed /algae)

Over the years, the nutritional and medical benefits of marine foods have appreciated (Table 1). Unlike red meat, fish is low in saturated fats. In addition, oily fish species are high in essential fatty acids. Fish protein rich diets lower the risk of cardiovascular diseases; moreover, omega -3 fatty acids are important for neurological development and general health. Fish is also rich in minerals such as calcium, zinc, and selenium (R. H. Thurstan & Roberts, 2014). Several studies have thus been conducted to determine the optimal fish dietary intake. The recommended daily intakes (97 - 550g per capita per week), vary between different countries (Food standards agency, 2010; R.H. Thurstan & Roberts 2014).

3.1.2 Underutilized aquatic bioresources

In addition, global fish wastage is estimated to be 7.3mt (Kelleher, 2005). This accounts for a large proportion of fish landings that could be used for human consumption directly through food or indirectly through fishmeal in aquaculture. The world is moving towards efficient marine bioresource utilization that is discouraging fish wastage/discarding. Norway has successfully implemented a policy whereby unsold fish are processed into oils and fishmeal (Diamond and Beukers-Stewart, 2011). As the drive for sustainability increases, the demand

for ‘forage’ fish /small pelagic fish like *Dagaa, herring, anchoveta* for human consumption is also on the rise. The forage species support 30% of global fish capture. They are being redirected from industrial uses such as feed production in aquaculture and agriculture to direct human consumption. This is bound to increase the utilization of these fish species that are resilient to high fishing pressure and thus sustain high levels. Aquaculture production could, therefore, be greatly affected because of their diminished supplies to feed industries.

It is also important to note that pelagic fish support the complex biological systems and thus their overexploitation leads to a knock on systems of the ecosystems and other non-target species. This fishing down in the food web leads to the marine ecosystem and fish production alterations, which eventually lead to an increase in some fish species because of declined predation. The eventual outcome is a risk in disease and environmental disturbances that are unsustainable (R. H. Thurstan & Roberts, 2014).

Reformed fishery management practices coupled with efficient fish landing utilization and aquaculture is therefore important in ensuring sustainability that can outpace the human population growth. Efforts to enhance sustainability in marine practices have seen the shift in wild fish feeds with genetically modified organisms (GMO) plant feed supplements in aquaculture. In addition, fish feeds created from recycling processes have substituted wild fish feeds. Further, fishmeal is being prepared from fish processing wastes. Indeed, aquaculture operations can be environmentally costly; however, there are sustainable methods for cultivating such fish species that are less environmentally harmful. Other than shifting to sustainable aquaculture practices, sustainability in the marine practice also necessitates a shift in consumer preferences for low trophic level fish species accompanied by offshore pelagic cage culture methods that will ensure coastal space is preserved and pollution is diminished.

3.1.3 IN USE AQUATIC HEALTH SOURCES

3.1.3.1 Fish

In Africa, particularly in Kenya, the majority of the population consumes, Ugali (Maize), rice, bananas, potatoes and wheat as the staple food. Most people make their food decisions based on costs in addition to urbanization, global trends, as well as improved health awareness (Achuka, 2017). According to the same document, several indigenous Kenyan vegetables are gaining momentum whereas *Dagaa*, which for a long time has been considered a 'poor man' diet is slowly catching up.

Fish as a food source has a variety of benefits. Fish fat content ranges from 2-30% based on species, diet, gender, season, and environment. Some polyunsaturated fatty acids predominantly found in fish, promote human health. They include docosahexaenoic acid (DHA), eicosapentaenoic acid (EPA) as well as essential long-chain omega -3 fatty acids. The latter is well known for its role in safeguarding against arrhythmias, preventing atherosclerosis as well as reducing blood pressure (Pérez-Gálvez *et al.,* 2009; Shahidi *et al.,* 2000).

3.1.3.2 Seaweed / Algae (Macro and Microalgae)

Algae, like other plants, produce secondary metabolites in their growth phase induced by metabolic changes due to environmental pressures. They have been used as potential sources of high-value chemicals such as carotenoids, phenolic compounds, phycobiliprotein pigments, polysaccharides, unsaturated fatty acids, and therapeutics. In addition, macroalgae have recently been used as a novel food with nutraceutical and pharmaceutical benefits (Table 1). They are a rich source of natural bioactive compounds with antiviral, antifungal, antibacterial, antialgal, antioxidant, anti-inflammatory properties as well as activity against

hypercholesteremia and hypolipidemia. They also have antineoplastic properties in addition to being an organic fertilizer (animal feed and aquaculture) and having potential bioremediation abilities and efficient solar energy harvesters (E. Shalaby, 2011).

Cyanobacteria such as *Anabaena*, *Nostoc*, *Spirulina,* and *Oscillatoria* have been shown to be the most promising source of novel natural bioactive compounds (secondary metabolites). They comprise of 40% lipopeptides, 4.2% fatty acids, 9% amides, 4.2% macrolides and 5.6% amino acids. Cyanobacteria lipoprotein properties include antitumor, antiviral, antifungal, cytotoxic, antimalarial, antifeedant, herbicides, multi-drug resistance reversers and immunosuppressant. In addition, Blue-green algae boost metabolism and lower cholesterol (VLDL and LDL). Most pharmaceutically important compounds are peptides. Peptide synthetases are involved in the production of cyanobacterial toxins and other peptides. Another group of enzymes known as polyketide synthetases is involved in the synthesis of microcystins. Moreover, cyanobacteria are good fuel producers whereas dried *Spirulina* is also a good food source. Proximate composition of the edible macroalgae species such as *Corallina sp.* and *Porphyra sp.* have been carried out in the West. Algae have largely been used as feedstock and for extraction of agar, alginates (from Brown algae) and carragenates (from red algae) (E.A Shalaby, 2011).

Table 2 Macroalgae and their bioactive ingredients with potential health benefits (Adapted from Shalaby, 2011).

Algae	Functional ingredients	Possible health effect
Sargassum vulgore	Alginic acid, xylofucans	Antiviral activity
Himanthalia elongate	PUFAs a-Tocoferol Sterols Soluble fiber	Lower risk of certain heart diseases Antioxidant activity Lower total and LDL cholesterol Lower total and LDL cholesterol
Undaria pinnatifida	PUFAs Sterols. Soluble fiber Folates Sulfated polysaccharides Fucoxanthin	Lower risk of certain heart diseases Lower total and LDL cholesterol Lower total and LDL cholesterol. Lower risk of certain types of cancer Antiviral activity Preventive effect on cerebrovascular diseases
Porphyra spp.	*PUFAs* *Sterols* *Soluble fiber*	Lower risk of certain heart diseases Lower total and LDL cholesterol Lower total and LDL cholesterol
Chondrus crispus	PUFAs (n-3) fatty acids Sterols Soluble fiber Sulphated polysaccharides (porphyrans)	Lower risk of certain heart diseases Lower total and LDL cholesterol Lower total and LDL cholesterol Apoptotic activities
Cystoseiran spp.	Terpenes Sterols	Valuable curative properties. Lower total and LDL cholesterol

3.1.3.2.1 Antioxidative activity of Macroalgae

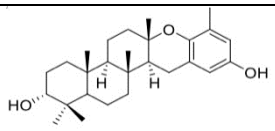
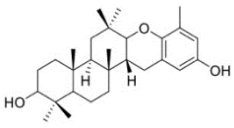
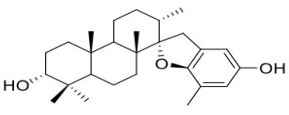
Algae is a rich source of both fat-soluble and water-soluble antioxidants against reactive oxygen species from internal metabolic reactions and from external sources such as UV irradiation, pollution, and stress. α-tocopherol/vitamin E and carotenoids are fat-soluble antioxidants whereas phycobiliproteins, polyphenols, and vitamin C are water-soluble antioxidants found in algae. Dried and fresh extracts of *Sargassum polycystum and*

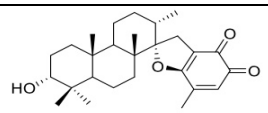
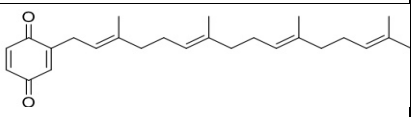
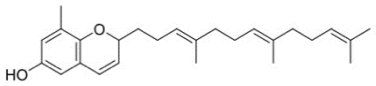
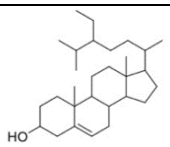
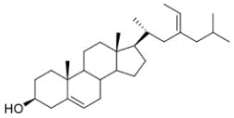
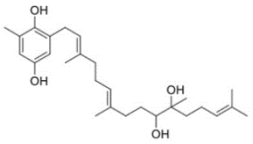
Laminaria obtuse showed that fresh algal extracts had higher antioxidative activity compared to extracts from dried algae. This could be attributed to higher triterpenes and carotenoids predominantly present in fresh algal biomass. This was supported by experiments on a brown algae *Fucus* that showed a 98% decrease in antioxidant activity with drying at 50°C for 48h. DPPH scavenging profile and lipoxygenase activity of extracts from *Sargassum fusiform* a Pacific Ocean algae showed high antioxidant activity attributed to β-sitosterol, fucose containing sulfated polysaccharide (Weng *et al*., 2000). Studies on *Bryothamnion triquetrum* (red algae) extracts showed that polyphenols (Trans-cinnamic, p-coumaric, ferulic acids) in the lyophilized extracts prevented the formation of thiobarbituric acid (TBA), a reactive substance during spontaneous lipoperoxidation in rat brain homogenates with an IC_{50} of 23.3 µg. Algae of the *Cytosera* genus contains tetraprenyltoluquinols which are tocopherol like compounds. *Laminariales sp*. showed high ferric ion reducing activity contributed by phloroglucinol. Fucoxantinein and hylopheophytin were identified as some of the active antioxidants from brown algae *Eisenia bicyclis* and *Hijikia fusiformis* respectively**.** In addition, the marine algal extracts of *Saragassum micracanthum* were shown to inhibit lipid peroxidation in rat liver homogenates with higher efficiency than vitamins C and E. The evaluated methanol, chloroform-methanol, and ethyl acetate extracts showed IC_{50} values of 0.70, 0.70 and 0.37µg/ml respectively. Fucosterol from *Silvetia siliquosa* inhibited sGPT and sGOT activities in rat blood serum by 63.16% and 25.57% respectively. This is in addition to hepatoprotective capacity by enhancing the activity of anti-oxidative enzymes glutathione peroxide, catalase and hepatic cytosolic superoxide dismutase. Brazil green seaweed species; *Ulva fascita*, *Chaetomorpha anteninna*, *Enteromorpha intestinalis,* and *Codium decorticatum* have also been shown to inhibit linoleic acid peroxidation activity. *C. anteninna* and *E. intestinalis* were the most potent with >70% inhibition of lipid peroxidation. Enzymatic

extracts of brown algae species also showed potent scavenging ability against reactive oxygen species (ROS) while its hydrogen peroxide scavenging ability was the highest at 90%. All these were higher than those recorded for commercial, synthetic antioxidants. In addition, about 16 meroterpenods with scavenging abilities (DPPH and butylcholine esterase) have also been identified from brown algae *Sargassum siliquatrum chromene.* Thirteen brown algae from the Aegean Sea have been shown to be potent radical scavengers (DPPH and Chemiluminescence). *Taonia atomaria* showed potent antioxidant activities that approached antioxidative standards that were attributed to several metabolites; stypodiol, stypoldione, sargaquinone, taondiol, isoepitaondiol and sargaol. Further, red algae; *Gracilaria sp* was also shown to slow degenerative diseases associated with aging.

Table 3 Some algal extracted compounds with antioxidative activity.

(Adapted from E. A. Shalaby, 2011 with modifications

Algal Species	**Cpd. No.**	**Name of the reported compound**	**Structure**
Taonia atomaria (Brown algae)	1	Taondiol	
	2	Isoepitaondiol	
	3	Stypodiol	

	4	Stypoldione	
	5	Sargaquinone	
	6	Sargaol	
Salvia plebeia	7	ß-Sitosterol	
P. siliquosa (Brown algae)	8	Fucosterol	
Cystoseira sp.	9	Tetraprenyltoluquinols	

3.1.3.2.2 Anticancer Activity of macroalgae

The WHO has termed cancer the number 3 killer globally and in Kenya, cancer cases have been on the rise with increasing morbidities. In 2018, cancer morbidity stood at 9.6M (WHO, 2019). The search for anticancer therapeutic agents is therefore on the rise (Table 3). Previous studies have shown 36 edible algae species with potent anti-cancer activity (Table 3).

About 501 algal extracts have been screened to date using several anticancer screening tests such as -protein tyrosine kinase (PTK), protein kinase C (PKC) and inosine monophosphate dehydrogenase (IMDH) assays. Twenty-three and 9 of the screened extracts were active against IMDH and PTK respectively. A novel chlorosulfolipid compound has been identified from *Poteriochromonas malhamensis.* In addition, three brominated chamigrane sesquiterpenes from *Laurencia majuscule* have been shown to be cytotoxic to colon cancer cell lines displaying mean cytotoxic concentrations of 10 to 100-fold lower than those of other cancer sub-panels and low cytotoxicity to normal cell lines.

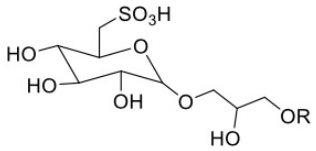
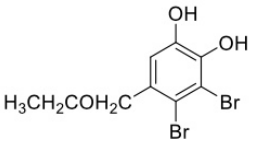
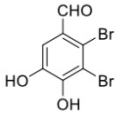
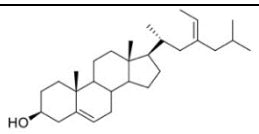
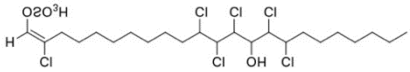
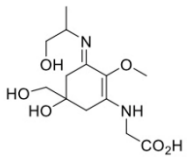
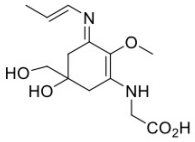
Palmitic acid extracts of red algae *Amphiroa zonata*, have been suggested to be a lead anticancer drug. These extracts have been shown to be selective for human leukemia by targeting DNA topoisomerase I but with no activity against DNA topoisomerase II. Additionally, palmitic extracts of *A. zonata* showed no cytotoxicity to normal human dermal fibroblasts (HDF).

Further, *Cladophoropsis vaucheriaeformis*; green algae, extracts showed high cytotoxicity against leukaemia. *Turbinaria ornate*; a brown alga has a sulfated fucan-like polysaccharide with an amino sugar that inhibits the growth of asynchronous cells of human non-small cell bronchopulmonary carcinoma line (NSCL-N6). The sea urchin 3’ sulfonoquinovosyl -1’-

monoacylglycerol has been shown to be effective in suppressing growth of compact tumors with an IC_{50} value of 10^{-5} M. The dominant fatty acids of the sea urchin are the saturated C16 fatty acid while fatty acids; 18:1, 14:1, 16:1, 18:1 and 14:0 are present in low concentrations.

In China, 39 marine algae were isolated of which 8 species were found to have potent cytotoxic activity. From these, 30 compounds were isolated; 14 bromophenols, 1 carotene, and 1 steroid. 3 kelp and red algae extract from *Laminaria setchellii*, *Macrocystis integrifolia*, *Nereocystis leutkeana,* and *Palmaria pamata* respectively have also been shown to have antiproliferative activity against human cervical adenocarcinoma cell line (HeLa). These studies showed that the antiproliferative activity was directly linked to the polyphenols in red algae and phlorotannins (Mycosporine like amino acids) in kelp. The extracts of *Porphyra sp*. of the red algae have also demonstrated anticancer activity through apoptosis.

Table 4 Anticancer bioactive compounds from Macroalgae

Species	Cpd No.	Name of the reported compound	Structure
Sea urchin intestine	10	3-sulfonoquinovosyl-1-monoac-ylglycerol	
Scytosiphon sp.	11	3,4-dibromo-5-(ethoxymethyl) 1,2-benzenediol	
	12	2,3-dibromo-4,5dihydroxy-benz-aldehyde	
	13	Fucosterol	
Poteriochromomas sp.	14	Malhamensilipin A (IC50 = 35 µM)	
Palmaria palmata	15	Palythinol	
	16	Palythene	

3.1.3.2.3 Antiviral activity of Macroalgae

Microalgal compounds have also been shown to have potent antiviral properties (Table 4). A polysaccharide from algae known as carrageenan has been shown to inhibit viral proliferation by blocking *in vivo* viral protein synthesis. Similarly, HIV reverse transcriptase and

consequent *in vitro* replication is hindered by sulfated polysaccharides from algae. Bulgarian Black sea *Ulva Lactuca* extracts also decreased the infectious viral titer of the H1N1 virus in chick embryos. There is inhibition of replication of *s*everal mammalian viruses by sulfated polyanions e.g heparin. HSV-1 and HSV-2 viral adsorption to cells are stopped by sulfated galactans from red algae. (Fig. 1 below)

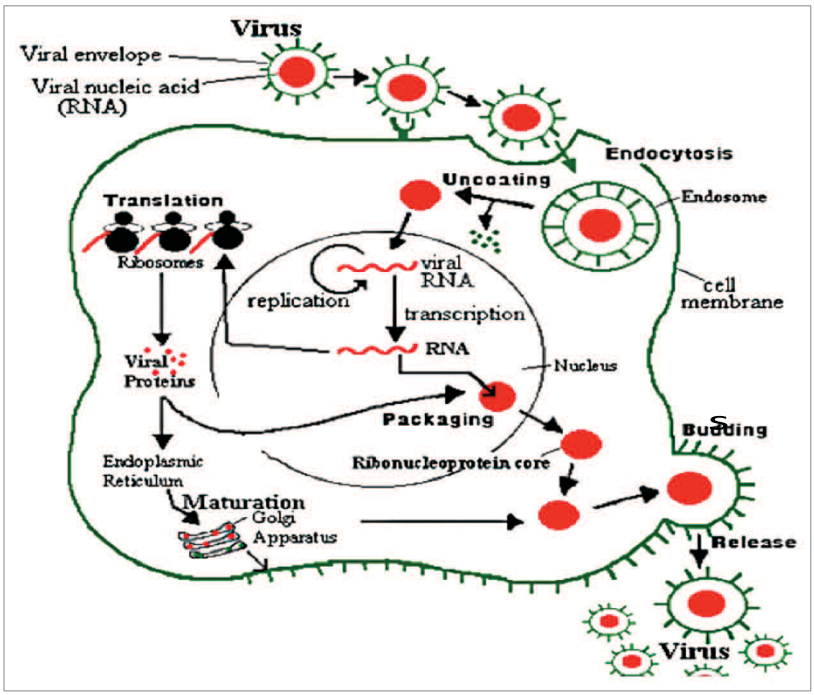
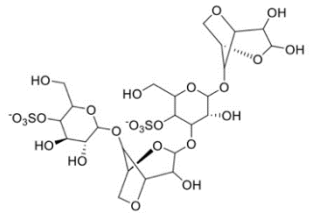
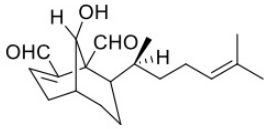
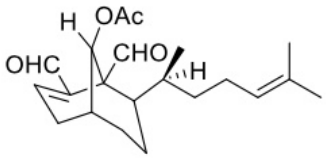


Figure 10. The reproduction of the Influenza virus (Adapted from Shalaby, 2011)

Table 5 Some algae extracted biochemical compounds with antiviral activity.(Adapted from Shalaby, 2011 with modification)

Species	Cpd. No.	Name of the reported compound	Structure
Red alga	17	κ-Carrageenan	
Dictyota menstrualis	18	(6R)-6-hydroxydichotoma-3,14-diene-1,17-dial, named Da-1	
	19	6R)-6-acetoxidichotoma-3, 14-diene-1,17-dial, named AcDa-1	

3.1.3.2.4 Antimicrobial activity of Macroalgae

A group of macroalgae from Sri Lanka were screened for antimicrobial activities (antifungal and anti-bacterial). It was established using *Staphylococcus aureus*, *E. coli*, *Cladosporium cladosporioides* and *Candida albicans* that 26 out of the 35 tested algae had antimicrobial activity. The most potent activity was evident for *Chrondrococcus hornemanni* attributed to dihalogenated monoterpenes, as well as *Ulva Lactuca* and *Gracilaria corticata* whose activities were attributed to acrylic acid. It is important to note that freshwater algae extracts showed no antimicrobial activity with extracts of up to 50mg/ml in concentration.

3.1.3.2.5 Antimicroalgal activity of Macroalgae

Macroalgae extracts are being screened to substitute synthetic algaecides to control harmful aquatic algae. Lysine, ferulic acid, fischerellin, bacillamide, B12–epi-hapalindole and rutacridone analogs are some of the active compounds with antialgal activity. This is in

addition to decaying barley that is used to control blue-green algae. Low concentrations of L-2 –azetidinecarboxylic acid (AZC) has been shown to inhibit algae growth. Some algae species have been shown to have allelopathic effects on others for example, *U. Lactuca* was shown to have negative allelopathic effects on harmful bloom-forming microalgae.

3.1.3.2.6 Plant growth stimulation by Macroalgae

Seaweed concentrates have also been shown to increase the growth and dry weight of plants by stimulating root growth, fruit set, flower production, crop quality and enhancing disease and stress resistance.

3.1.3.3 Seaweed/Algae in Kenya

Previous studies conducted on Kenya's algae population have largely been taxonomic (Coppejans, Leliaert, & Clerck, 2000; Yarish & Wamukoya, 1990).On the contrary, data on algae of economic importance is conspicuously lacking. A survey conducted on 15 sites along the Kenyan coast, showed that although Kenya had a huge potential for algae harvesting and farming, no commercial farming was currently underway unlike the neighboring countries such as Tanzania. (Yarish & Wamukoya, 1990). Moreover, the abundant Macroalgae resource along the Kenyan coast remains largely unutilized since the coastal population hardly consumes it as food. Paradoxically, Kenya continues to be a net importer of Macroalgae products when it has the potential to self-produce and even export (Yarish & Wamukoya, 1990).

4.1 FISH LANDINGS IN LAKE VICTORIA

4.1.1. Total landings, commercial value, and utilization

Globally, fish landings are divided into finfish and invertebrates (R. H. Thurstan & Roberts, 2014). Landing values are corrected for processing losses using the formula from HRMC. These processing losses are diverse globally depending on the processing methods used; fish species harvested as well as culture and market. The wild fish landings (Finfish and

invertebrates) have stabilized due to expanded aquaculture production. This has resulted in a 10% increase in fish supply to outpace the growing human population. However, no clear statistics are available for the precise proportions that go into fish processing (for oils and fishmeals) and human food. In Kenya, the high value fish landings, for example, NilePerch are largely utilized for export as a fish fillet. .However, *Dagaa* is largely used locally for food and feed. Its utilization relative to the landings however remains generally low due to poor consumer acceptability as it is considered a 'poor man's' diet.

5.1. DAGAA /OMENA

5.1. Classification and General description

Dagaa is a ray-finned fish; *Actinopterygii*, belonging to the *Cypriniformes (Carps or minnows) and Danioninae; Rastrineobola argentea* family. In the East African Region, *Dagaa* is most prevalent in the Lake Victoria Basin and the Victorian Nile. This Silverfish as it is called in English is commonly known as *Dagaa* in Tanzania, *Mukene* in Uganda and *Omena* in Kenya. The Silverfish is the most thriving local fish species with an estimated 1.3M tonne biomass in lakes Victoria in Kenya and Kyoga in Tanzania second to Nile perch (*Lates Niloticus*). This has been attributed to favorable ecological conditions in Lake Victoria despite the continual predation by Nile perch and increased fishing pressure by the fishermen population (Bureau & Resources, 2014). Other than, in the three East African countries mentioned above, *Dagaa* is also found in Lake Bulera in Rwanda where it was artificially introduced in the late '80s.

Generally, *Dagaa* is a minute and slender fish with a maximum total length (TL) of 80mm (or 8cm). It has a distinct lateral line in its body running along the caudal peduncle. Thin suborbital bones cover its cheeks and its caudal fin is yellowish in color. Unlike the live *Dagaa* fish, the dead *Dagaa* has a discrete mid-lateral stripe.

Dagaa is both an inshore and an offshore fish, however, the young fish is mainly found offshore. This is because they spend their larval stage in shallow waters and in the mature stage move to deeper waters during the daytime and then swim afloat during the nighttime. *Dagaa's* diet largely constitutes of zooplanktons and water surface insects.

5.1.1. Differences based on the catch site, size, and gender

Aquatic ecosystems differ based on the following systems; Lentic/lacustrine or lotic systems, breadth, depth; shallow or deep, mixture level; well mixed or not mixed, temperature range, oxygen concentration, wind, precipitation and evaporation levels. These ecosystems also differ seasonally with overturns during spring, fall, and stratification during winter and summer. These gradations of aquatic ecosystems demand for physiological adaptations that affect the distribution of life in the various habitation zones.

Three different zones of habitation exist in aquatic ecosystems namely:

1. Littoral zone –which is in the shallow waters that have good light penetration and support rooted plants and animals that live at the bottom area
2. Limnetic zone – which gets good light penetration and thus fosters animal plankton and plant life
3. Profundal zone, which occurs at the bottom deep-water region. This is a non-lit area without light penetration that supports organisms adapted to the dark (Manyala *et al*., 2011)

5.1.2. Basic general morphological characteristics of *Dagaa*

Morphometric analyses of *Dagaa* show that size measurements are taken in three dimensions of standard length (SL), total length (TL) and fork length (FL). Previous sturgeon methods have transformed from the traditional measurements to the fork length measurements (Manyala *et al.,* 2011). A previous study showed that samples of *Dagaa* from Usari, Nduru,

Dunga, Paga and Rota landing beaches of Lake Victoria – Kenya, had an average total length (TL) of 46mm $\pm$ 0.34 (Ogonda *et al*., 2014).

5.1.3. Proximate composition of *Dagaa*

Dagaa proximate analysis classifies the fish compounds based on their chemical composition. This involves the quantification of both macronutrients and micronutrients present in the fish as well as the moisture content. Proximate analysis is especially essential when considering new fish species as potential food proteins. Proximate composition information acts as a guide to product development in the food industry. In addition, the information is useful for quality control and regulatory purposes as well as for awareness to the general community (Thangaraj, 2016).

Given the importance of such studies, it is essential to ensure a representative and homogenous sample collection in order to have accurate results. To date, *Dagaa* proximate composition analysis has been conducted for several landing sites (Table 5) (Ogonda *et al*., 2014).

Local artisanal fishermen and fishmongers have the capacity to accurately distinguish *Dagaa* from varied landing sites based on their taste. *Dagaa* samples from Paga were identified as the most bitter while those from Usari were described as the sweetest. Bille *et al*. (2006) showed that inherent sensory (taste) characteristic differences between the same fish species from different locations are a function of proximate compositional profile differences. This was confirmed by studies by Ogonda *et al.* (2014). These detectable taste differences could be directly attributed to the effects of spontaneous lipid autolysis that begin as soon as the *Dagaa* is harvested. There was a high lipid content (7.78%) in Dagaa from Paga, this high lipid content generally provides a substrate for spontaneous auto-lipolysis facilitated by

endogenous enzymes and microorganisms before retailing. The relatively less bitter fish catch from Usari, on the other hand, was found to have relatively low lipid content.

The *Dagaa* protein composition on a wet weight basis was established to be between 19-22%. The lipid content was found to be between 1.8 -5% by the soxhlet method and 3.9-9% by Dyer and Bligh method. In addition, moisture content was found to be between 74-76%. Dry weight and ash weight content was also determined as 23-27% and 1.8-4.4% (wwb) and 10-15% (dwb).

Table 6: Proximate Composition of Dagaa on a wet weight basis (wwb): (Adapted from Ogonda *et al*., 2014).

Sample ID	Moisture %	Dry weight %	Ash content %(wwb)	Ash content % (dwb)	Lipid content %Dyer/Bligh	Lipid content %soxhlet	Protein content % kjedahl	Protein content Biuret (mg/ml)
Usari	76.35±0.05**as**	23.65±0.05**b**	1.88±0.94**ah**	11.49±0.61**a**	4.35±0.57**a**	1.77±0.00**a**	19.11±1.80**ar**	4.19±0.03**ad**
Nduru	74.05±0.95**as**	25.95±0.95**b**	2.08±0.69**ah**	10.70±0.00**ax**	7.32±0.59**a**	1.82±0.00**a**	21.07±0.19**a**	1.93± 0.15**ad**
Dunga	72.83±1.30**a**	27.17±1.30**b**	4.38±0.38**a**	10.00±0.00**ax**	3.87±0.19**a**	3.40±0.00**aw**	19.53±1.03**a**	5.80± 0.04**ad**
Paga	74.37±0.40**as**	25.63±0.40**b**	2.36±0.14**a**	10.55±0.78**ax**	7.78±0.44**a**	2.18±0.32**aw**	21.78±1.79**ar**	5.22± 0.12**ad**
Rota	76.90±0.30**as**	23.1±0.30**b**	3.08±0.81**ah**	14.58±2.94**a**	5.47±0.10**ay**	2.77±0.00**aw**	20.31±0.88**a**	4.39±0.06**ad**
Homogenous	78.40±0.89**as**	21.6±0.89**b**	4.26±0.79**ah**	10.83±3.54**ax**	3.11±0.45**ay**	4.90±0.10**aw**	18.77±0.03**a**	4.76± 0.06**ad**

From the study, there was a significant difference in the mineral (ash content), protein content, lipid content and moisture content between the different beaches. Values are shown as mean ± standard deviation for triplicate analysis of a pooled sample. **a-** values with significant differences (*$p<0.01$) using ANOVA. Values with the same double letters have a significant difference ($p<0.05$), (TUKEY).

Given this rich proximate content, *Dagaa* is well suited as an alternative fish food protein source to meet the weekly recommended fish intake of 97-550g and the daily recommended dietary intake of high-quality protein (eggs, fish and meat) which stands at 0.63g/kg body weight for western males habitually on mixed diets (FAO *et al.,* 1985). This value, however, varies with populations due to adaptation, making average estimates impossible. As a result, 'safe levels' are determined at 2 standard deviations (SD) higher than 0.75g/kg body weight. The lack of such reliable data is not a problem in the West; however, in developing countries that face malnutrition and food security and nutritional deficiencies, there is an urgent need for this information.

Information on proximate content is important in classifying/grading *Dagaa* based on quality, susceptibility to quality losses, ensuring adequate post-harvest treatment and human consumption. This is especially because recommended dietary intake corrections are essential for proteins based on quality (amino acid compositions) as well as diet digestibility especially for people living in developing countries (FAO, 2019; LIVESEY, 1987). The quality of protein is determined by the amino acid composition and most natural protein diets have been shown to provide only enough for pre-schoolchildren but insufficient amino acid content for the elderly.

Table 7: Correlation matrix of proximate parameters of *Dagaa*. (Adapted from Ogonda *et al.,* 2014)

Component	**moisture**	**Dry weight**	**Ash (wwb)**	**Ash (dwb)**	**Lipid (dyer/bligh)**	**Lipid (soxhlet)**	**Protein (Kjedahl)**
Moisture	1	-1**	a	a	b	a	b
Dry weight	-1**	1	b	b	a	b	a
Ash (wwb)	a	b	1	b	b	a*	b

Ash (dwb)	a	b	b	1	a	b	a
Lipid (dyer/Bligh)	b	a	b	a	1	b	a**
Lipid (soxhlet)	a	b	a*	b	b	1	b
Protein (kjedahl)	b	a	b	a	a**	b	1

were done using SPSS v. 16.0. **a** shows a positive correlation between the components. **b** shows a negative correlation between parameters. **1** shows a perfect correlation. * shows the correlation is significant at the 0.05 level (2-tailed). Whereas, **shows correlation is significant at the 0.01 level (2-tailed). There was a perfect negative correlation ($p<0.01$), between moisture content and dry weight, and a true positive correlation ($p<0.05$), between lipid content (soxhlet) and ash content (wwb). Another true positive correlation p (<0.01) is also seen for protein Kjeldahl values and lipid content determined by Dyer/Bligh.

The correlational analysis matrix (Table 6) indicates the relationship between the proximal constituents of *Dagaa.* Protein and Lipid content of *Dagaa* has been shown to have a direct correlation. Previous studies have also shown that Pelagic fishes were high and low in protein and fat respectively (Suseno *et al*.,2010). A study on *Dagaa* from Lake Victoria in Kenya reported protein and fat content values of 19.1-21.8% and (4.4 -7.8%) respectively. These values were higher than recorded protein and fat content values for deep-sea fishes of 11.9-20.6% and 0.01-4.84% respectively (Suseno *et al*., 2010, Ogonda *et al*., 2014). This is noteworthy given that the average size for the sampled *Dagaa* was 46mm; much lower than the average values reported for other landing sites around Lake Victoria (80mm) and even lower than most deep-sea fishes whose average length (TL) ranges from 140mm-490mm.

Using Dyer and Bligh method, fish is classified into four groups based on their fat content; lean fish (<2%), low fat (2 - 4%), medium fat (4 - 8%) and high fat (>8%) (Huss, 1988 and Ackmann, 1989). It could also, be classified based on the 5% (dwb) fat composition criteria for discriminating lean from fatty fish species according to the soxhlet method (Owaga *et al*., 2010). Using the former method, *Dagaa* was classified as fatty fish (fat content >2%) (Ogonda *et al*., 2014). *Dagaa* moisture content has also, been shown to be inversely proportional to its non-polar lipid content and directly proportional to its polar lipid content. The differential *Dagaa* lipid profile is attributed to differences in temperature, sex, age, food availability, salinity as well as geographical and seasonal variations (El Tay *et al*., 1998 and Stansby, 1981).

In comparison to deep-sea fishes, *Dagaa's* total ash content is indicative of its relatively higher mineral composition (Suseno *et al.*, 2010). This was attributed to the lower edible portions for deep-sea fishes (5%-63%). This proportion, however, was still greater (>3X) than for the *Dagaa* TL. Other studies concluded that this high total ash content in comparison to freshwater species was due to the direct inclusion of skeletal muscles in the edible *Dagaa* portion. *Dagaa* skeletal muscles have been shown to have 3600mg/100g and 10.2mg/100g respectively (Ghelichpour *et al.,* 2011 and Owaga *et al*., 2010).

5.2.1. Value addition to *Dagaa* by enzymatic hydrolysis

Lake Sardine commonly referred to as *Dagaa* and locally as *Omena,* is the second-highest fish in terms of Landings in Lake Victoria. This accounts for 63% of the total fish catch in Lake Victoria; a freshwater lake. The large *Dagaa* landings are largely wasted by postharvest losses recorded at 20-30% and rise up to 50% during the rainy season due to inefficient drying. This is unlike the Nile perch (*Lates niloticus*) that forms the highest portion of the fish landings and is mostly exported. The inefficacies in the preservation of the *Dagaa* by the traditional methods of sun drying, spicing, smoking and salting coupled with rapid quality loss resulting from auto-lipolysis and auto-proteolysis have led to a steep decline in the nutritional and economic value of this otherwise large catch obtained. This has led to poor final product quality that fetches very low returns for the artisanal fishermen due to low consumer

acceptability attributed to the soapy and bitter taste. As a result, *Dagaa* has been alternatively channeled into animal feed production instead of human consumption.

Therefore, in order to upgrade this large bioresource and as a way of value addition, pre-processing through exogenous enzymes or controlled use of endogenous enzymes is an alternative way for re-including this otherwise wasted food bioresource into the human food chain and a way of lengthening the shelf life and preserving the quality of the final product.

Previous studies have demonstrated the production of fish protein hydrolysate using commercial enzymes from Nile Perch by-products for recovery of nutritional lipids, fish protein hydrolysates and Nile perch insoluble by-products.

5.2.2 Value addition through assisted hydrolysis of *Dagaa* by an endogenous and exogenous enzyme

Mbatia *et al.,* (2014); Ogonda *et al.,*(2017), pioneered the processing of *D*agaa muscle proteins using both endogenous and exogenous enzymes in order to produce *Dagaa* fish protein hydrolysate with potential nutraceutical and pharmaceutical properties.

5.2.3 Endogenous enzyme assisted hydrolysis of *Dagaa* proteins

Value addition through hydrolysis with endogenous enzymes is conducted using intrinsic *Dagaa* microbes. However, this process is difficult to control due to variations in enzyme profiles based on age, gender, season as well as the fish species. As a result, value addition using endogenous enzymes is largely spontaneous and the resultant FPH properties differ in both quality and properties (Krisstinsson and Rasco *et al*., 2000; Guérard, 2007).

Moreover, this uncontrolled process adversely affects the organoleptic properties of the FPH and may in some cases produce toxic by-products. This notwithstanding, a number of FPH are still produced from several fish species using endogenous enzymes. Furthermore, studies have been conducted on the production of bioactive peptides from *Dagaa* using endogenous enzymes (Mbatia *et al*., 2014, Ogonda *et al*., 2014, 2017).

5.2.4 Exogenous Enzyme assisted hydrolysis of *Dagaa* proteins

The use of exogenous enzymes in the hydrolysis of fish proteins is an important method for protein recovery and adding value to fish by-products and unutilized fish that would otherwise be wasted. Moreover, it is essential in overcoming the challenges associated with endogenous enzyme hydrolysis in addition to improving the organoleptic, functional and physicochemical properties of the initial *Dagaa* substrate. The enzymes are specific and the reactions are carried out under ambient conditions. This method of hydrolysis, unlike with endogenous enzymes, assisted hydrolysis reduces the processing time and enables control over the reaction by controlling the enzyme-substrate (ES) ratio. As a result, the degree of hydrolysis (DH) can be manipulated to ensure that the product obtained is of consistent quality (molecular weight and peptide composition, functional and bioactive properties). Several enzymes are available for exogenous hydrolysis; however, some such as Alcalase, Protamex and Neutrase are preferred due to their ability to conduct rapid, extensive hydrolysis with a resultant less bitter tasting FPH product with superior bioactive properties (Wang *et al.*, 2010).

6.1 MULTIFUNCTIONAL PROPERTIES OF DAGAA BIOACTIVE PEPTIDES

6.1.1 Food functional properties of Dagaa

Dagaa fish protein hydrolysate functional properties are modified and improved by the use of specific and distinct enzyme hydrolysis conditions of time, temperature, pH and ionic concentration for attaining a predetermined percentage degree of hydrolysis (DH) (Hall and Ahmad., 1992; Petersen., 1981). Several functional properties previously exhibited by other fish protein hydrolysates include:

6.1.1.1 Solubility

There is an increase in solubility of FPH relative to the starting fish raw material. This is due to the hydrolysis of fish myofibrillar proteins that are responsible for protein insolubility in addition to increments in repulsion between the peptide units as a result of exposure to ionic residues during hydrolysis (Krisstinson et al., 2000). FPH solubility affects several other properties such as whippability, emulsification as well as oil and water holding capacity. *Dagaa* protein hydrolysate showed varied solubility under different processing regiments and pH as well as the final degree of hydrolysis. The data obtained were consistent with those from other fish species that showed a

proportional increase in solubility with an increase in the degree of hydrolysis (Mbatia *et al*., 2014; Petersen *et al*., 1981; Souissi *et al*., 2007; Santos *et al.*, 2011).

6.1.1.1.2 Emulsification

The emulsifying properties of fish protein hydrolysate can be improved following controlled hydrolysis. It is important to note that there are drastic losses in emulsifying properties with extensive hydrolysis (Krisstinson *et al*., 2000).

Unlike raw *Dagaa*, its FPH showed emulsifying properties. However, given the extensive percentage degree of hydrolysis (%DH), the stability of the emulsions was very low. This data is consistent with previous findings indicating that hydrophobicity and emulsification properties are directly proportional. Consequently, with increased hydrolysis, there is a reduction in hydrophobicity (Souissi *et al.,* 2005; Mbatia *et al*., 2014). Further, surface activity is a function of peptide length and there is diminished emulsifying ability with a decrease in surface activity following extensive hydrolysis (Josti *et al.,* 1977; Lee *et al.,* 1987; Quaglia and Orban 1990).

6.1.1.1.3 Water-holding capacity

The fish protein matrix ability to retain water molecules inside their structure is termed as water holding capacity. Most fish protein hydrolysates are important in food formulations because of their excellent water holding capacity (Krisstinson *et al*., 2000). This functional property is yet to be determined for *Dagaa* (Mbatia *et al*., 2014).

6.1.1.1.4 Fat absorption

Dagaa FPH produced by Mbatia *et al.,* (2014) showed good fat absorption properties and could be useful in the meat industry, for sausage elaboration and salad dressing industry for taste retention (Rodríguez-Ambriz *et al.,* 2005). Raw *Dagaa* fat absorption capacity is improved using enzymatic hydrolysis. The results obtained were consistent with previous findings that showed decreased emulsifying property with increased %DH (Souissi *et al*.,2007; Santos *et al*., 2011).

6.1.1.1.5 Whippability

As with other functional properties, *Dagaa* whippability was also diminished with extensive hydrolysis. This is consistent with reports by Bombara *et al*., (1994) that showed that good film cohesiveness is only obtained with high molecular mass peptides following partial hydrolysis (Mbatia *et al.,* 2014; Souissi *et al*., 2007).

The functional properties obtained by Mbatia *et al.,* (2014) for *Dagaa* were consistent with the process, which was optimized specifically for the production of antioxidative peptides. In order to have *Dagaa* FPH with functional properties for the food industry, process optimization and engineering had to be tailored to that end.

6.1.2 *Dagaa* FPH bioactive properties

Bioactive peptides speak of specific protein fragments with varying sizes from 2 to 20 amino acids that infer positive multifunctional effects in body functions and influence health beyond their basic role as a nutrient source (Hartmann & Meisel *et al*., 2007; Meisel and FitGerald, 2003). This ability of peptides to enhance human health is attributed to several peptides with known activities such as neuroactive, mineral and hormonal regulating properties immunomodulatory activities, antithrombotic, antihypertensive, antimicrobial, antioxidative and antihypertensive effects (Bernet *et al.,* 2000; Duarte *et al.,* 2006; Gormley, 2006; Je *et al.,* 2005a & b; Je *et al.,* 2007; Je *et al.,* 2009; Jun *et al.,* 2004; Jung *et al.,* 2006; Kitts and Weiler, 2003; Korhonen & Pihlanto, 2003; Liu *et al.,* 2008; Meisel and FitzGerald, 2003; Murray and FitzGerald, 2007; Nagai *et al.,* 2008; Shimizu, 2004). The capacity of these peptides to elicit biological responses is dependent on their ability to cross the epithelium, enter the blood circulation and subsequently bind to cell receptors (Pihlanto & Korhonen, 2003).

Bioactive properties were first identified in bioactive peptides from casein; casein phosphopeptides (CPPs) which were shown to enhance Vitamin D independent bone calcification in rachitic infants (Mellander, 1950). Bioactive peptides are intrinsic to the food and can be obtained from plant and animal sources through digestion. They could also be obtained following enzymatic hydrolysis, food processing and fermentation to protein hydrolysates. Milk is a bioactive peptide source. Other sources

include fish, eggs, soybean as well as maize and wheat (Yamamoto *et al*., 2003; Hartmann and Meisel, 2007).

Fish, particularly fish wastes and underutilized fish species is an essential raw material for the recovery of bioactive peptides (Samaranayaka, 2010). Some of the reported bioactive properties of the underutilized *Dagaa* fish protein hydrolysate (Mbatia *et al.*, 2014) in addition to other interesting future prospects include:

6.1.2.1.1 Antioxidative properties

There is a widespread search for antioxidants from natural sources to counter oxidative food deterioration and oxidative cell damage. This is following epidemiological and experimental data that has shown the role of free radicals in cardiovascular diseases and cancer (E. A. Shalaby, 2011). Moreover, there has been a decline in the use of synthetic antioxidants which have been demonstrated to promote carcinogenesis in addition to consumer acceptability issues around synthetic food additives (Dai *et al*., 2010). There are several mechanisms for antioxidative action which are analyzed using various assays (Table7).

Lipid oxidation is important especially in the food industry because it leads to apparition of bitterness, off-flavors, dark colors and potentially toxic reaction products that could cause serious health damages. Antioxidants act by retarding the discoloration of food as well as oxidation assisted food deterioration (Barkia *et al*., 2010; Lin and Liang 2002).

Mbatia *et al*.,(2014) produced FPH with antioxidative properties from *Dagaa*. Preliminary data showed diminished ignition of glycerol and potassium permanganate in the presence of *Dagaa* fish protein hydrolysate (FPH). Further quantitative antioxidative assays using DPPH, ferric ion reducing assay and oleic acid peroxidation assay showed that *Dagaa* FPH was superior to the synthetic antioxidant; butylated hydroxytoluene (BHT). Further, *Dagaa* FPH from endogenous enzymes equally showed higher antioxidative power compared to BHT. The antioxidative capacity was concentration-dependent and showed superior antioxidative capacity with upto 90% radical scavenging ability; RSA (DPPH assay). It is worth noting that the peptide profile differences between the endogenous and exogenous

enzyme hydrolysate functioned differently with different assays. The extensively hydrolyzed *Dagaa* FPH which has a higher hydrophilic peptide residue composition worked better in polar systems (DPPH and Ferric ion reducing assay). On the other hand, the partial hydrolysate prepared with the *Dagaa* endogenous enzymes was more potent in lipid peroxidation (Mbatia *et al.*, 2014).

Table 8: In vitro antioxidative capacity assays

Assays involving hydrogen atom transfer reactions[1]	
	ORAC (oxygen radical absorbance capacity)
	TRAP (total radical trapping antioxidant parameter)
ROO· + AH ROOH + A·	Crocin bleaching assay
ROO· + LH ROOH + L·	IOU (inhibited oxygen uptake)
	Inhibition of linoleic acid oxidation
	Inhibition of LDL oxidation
Assays by electron-transfer reaction1	
	TEAC (Trolox equivalent antioxidant capacity)
M (n) + e (from AH) → AH·+ + M (n-1)	FRAP (ferric ion reducing antioxidant parameter)
	DPPH (diphenyl-1-picrylhydrazyl)
	Copper (II) reduction capacity
	Total phenols assay by Folin-Ciocalteu reagent (FCR)
Other assays	
TOSC (total oxidant scavenging capacity) [2]	
Inhibition of Briggs-Rauscher oscillation reaction[3]	
Chemiluminescence[4]	
Electrochemiluminescence[5]Adapted and modified (Samaranayaka 2010)	

AH = Antioxidant, LH = Substrate
1Huang and others (2005)
2Winston and others (1998)
3Cervellati and others (2002)
4Ashida and others (1991)
5Chmura and Slawinski (1994)

6.1.2.2 FUTURE PERSPECTIVES

6.1.2.2.1 Antihypertensive properties of *Dagaa* protein hydrolysates

Non-infectious diseases especially high blood pressure is a major killer globally. About 1.13Bn, people worldwide live with hypertension (WHO, 2019). In order to modulate antihypertensive effects, FPH must pass the intestine wall, enter the bloodstream and reach the target organs and/control nutrient absorption and gastrointestinal (GI) function through direct effect on the gut lumen (casein phosphopeptides), binding surface receptors on epithelial cells (opioid peptides) (Miguel et al., 2008; Vermeirssen *et al.,* 2002; Shimizu *et al.,* 1997; Satake *et al.,* 2002; Vermeirssen *et al.,* 2002; Geerlings *et al.,* 2006; Miguel *et al.,* 2008; Quirós *et al.,* 2008; Ziv & Bendayan, 2000).

Previous studies on FPH has shown that they possess superior antihypertensive activity. This is attributed to the inhibition of angiotensin I converting enzyme (ACE). This has been shown to be stronger than natural peptides (Perez-Galvez, 2009). Bioactive peptides are further classified into 3 groups; pre-drug, pro-drug as well as substrate type inhibitors based on the action of ACE and other enzymes *in vivo* (Fujita *et al.,* 2000). Pre-drug ACE inhibitors require activation by gastrointestinal tract (GIT) proteases and or by ACE, while pro-drug type activity is increased by GIT proteases and ACE action (Hasan *et al.,* 2006). Given the superior antioxidative effects of *Dagaa* FPH, it will be interesting to also test for its antihypertensive effects.

6.1.2.2.2 Antitumor

Cancer according to WHO is the number three killer globally. In Kenya, the year 2019 alone, there evidently is a huge cancer burden following a rise in cancer-associated deaths. A control measure in the form of FPH with anticancer activities would be essential in tackling this cancer challenge. Fish protein hydrolysates have been identified as good inhibitors of *in vitro* proliferation of human breast cancer cell lines (Picot *et al*., 2006). Given the superior properties demonstrated by the *Dagaa* FPH, it will be interesting to also test for their anticancer properties.

7.1. CONCLUSIONS

Having the ability to live in the competitive freshwater and marine aquatic ecosystems, seaweed/algae, and *Dagaa*, as well as other underutilized fish species, are endowed with intrinsic defense strategies with structurally diverse chemical compounds. These compounds can serve as chemical prototypes for varied pharmaceutical uses and human nutrition. As a result, they contribute indirectly to food and nutritional security through the economic empowerment of the population. In addition, strengthening of the fish sector will foster food security by checking off food imports bills using fish export receipts.

Finally, fish and especially underutilized fish species; *Dagaa* also has a direct role in food security by being a food source. The rich quality macronutrient (protein), calorie and micronutrient (mineral) provided through fish diets will be essential in tackling a huge burden of malnutrition-related cases. In efforts targeted towards food and nutritional security. It is therefore essential, for the Kenyan population and/country to exploit these underutilized aquatic bioresources which comprise *Dagaa* and other fishes as well as seaweed/algae first as alternative protein food sources ; in the raw unprocessed form or in the form of FPH, in addition to mining for lead pharmaceutical compounds through processing of the excess aquatic bioresource.

8.0 ACKNOWLEDGEMENTS:

The authors wish to acknowledge Brenda Okoko and Hilder Nyaboke; from the Chemistry Department, the University of Nairobi for their contributions towards the chemical structures entered in this publication.

9.1. REFERENCES

Achuka, V. (2017). Chapati edges ugali out of the table in Kenya as the rich salivate over poor diet – Eve woman. *Eve Digital*. Retrieved from https://www.standardmedia.co.ke/evewoman/article/2001231754/chapati-edges-ugali-out-of-table-in-kenya-as-the-rich-salivate-over-poor-man-s-diet

Barkia,A.,Ali Bougatef,Hayet Ben Khaled & Moncef Nasri. (2010). Antioxidant activities of sardinelle heads and/or viscera protein hydrolysates prepared by enzymatic treatment. *Journal of Food Biochemistry* 34; 303–320

Béné, C. S. H. (2005). Fish and food security in Africa. *WorldFish Center Quarterly*,Vol.28No.3&4.RetrievedfromhttP://aquaticcommons.org/9087/1/na_2351.pdf

Bernet. F., Montel, V., Noel, B., Dupouy, J.P. (2000). Diazepam-like effects of a fish protein hydrolysate (Gabolysat PC60) on stress responsiveness of the rat pituitary-adrenal system and sympathoadrenal activity. *Psychopharmacology* 149: 34-40.

Bille, P.G and R.H, Shemkai. (2006). Process development, nutrition and sensory characteristics of spiced-smoked and sun-dried *Dagaa* (*Rastrineobola argentea*) from L. Victoria, Tanzania. *African Journal of Food,Agriculture,Nutrition and Development*; 6(2): 1-12.

Bligh, E. G., & Dyer, W. J. (1959). A rapid method of total lipid extraction and purification. *Canadian Journal of Biochemistry and Physiology*, 37, 911–917.

Bureau, I., & Resources, F. O. R. A. (2014). *The Silver Fish, Rastrineobola argetea*. 1–4.

Coppejans, E., Leliaert, F., & Clerck, O. D. E. (2000). ANNOTATED LIST OF NEW RECORDS OF MARINE MACROALGAE FOR KENYA AND TANZANIA , Since 1985 the phycology research group of (he Ghent University has. *Biologisch Jaarboek (Dodonaea)*, *67*(1), 31–93.

Dai Zhi-Yuan,Yan-Ping Zhang, Hong Zhang and Yan-Bin Lu. (2010). Preparation and characterization of mussel (*mytilus edulis*) protein hydrolysates with angiotensin-1-converting enzyme (ace) inhibitory activity by enzymatic hydrolysis.*Journal of Food Biochemistry* .ISSN 0145-8884; 1-9

Duarte. J., Vinderola, G., Ritz, B., Perdigón, G., & Matar, C. (2006). Immunomodulating capacity of commercial fish protein hydrolysate for diet supplementation. *Immunobiology* 211:341-350. 100

El-Tay, NOA., Abdeltif ,E.M & M.E, Ali. (1998). Chemical composition and quality grading of three commercial fishes from the Nile.In**:** *Proceedings of FAO Expert Consultation on Fish Technology in Africa* (Kisumu, Kenya). Report no. 574 FAO (Rome): 215 – 218.

FAO. Energy and protein requirements. Retrieved August 3, 2019, from 1987 website: http://www.fao.org/3/AA040E/AA040E09.htm

FAO, WHO, & UNU. (1985). Energy and protein requirements. In *G,LIVESEY,AFRC Institute of Food Research , Norwich Laboratory, Colney Lane ,Norwich NR4 7UA*.

Fitzgerald, A.J., Rai, P.S., Marchbank, T., Taylor, G.W., Ghosh, S., Ritz, B.W. & Playford, R.J. (2005).Reparative properties of a commercial fish protein hydrolysate preparation. *Gut* 54:775-781.

Fujita , H., Yokoyama, K. & Yoshikawa, M. (2000). Classification and antihypertensive activity of angiotensin-I-converting enzyme inhibitory peptides derived from food proteins. *J Food Sci* 65: 564-569.

Geerlings, A., Villar, I.-C., Hidalgo, Zarco., F, Sánchez M., Vera R, Zafra Gomez A, Boza J, Duarte J. (2006). Identification and characterization of novel angiotensin converting enzyme inhibitors obtained from goat milk. *J Dairy Sci* 89: 3326 –3335.

Ghelichpour, M. & Shabanpour, B (2011).The investigation of proximate composition and protein solubility in processed mullet fillets. *International Food Research Journal 18(4): 1343-1347*

Gormley, R. (2006). Fish as a functional food. *Food Science and Technology* 20, pp. 25-28.

Guérard, F.(2007). Enzymatic methods for marine by-products recovery. In: Shahidi F, editor. *Maximizing the Value of Marine By-products.* Cambridge, England: Woodward Publishing Limited. p. 107-143.

Hall, G.M. & Ahmad, N.H. (1992). Functional properties of fish protein hydrolysates.Ch. 11 In *Fish Processing Technology.* Hall, G.M. (Ed.), p. 249-265. *Blackie Academicand Professional*, N.Y., U.S.A

Hartmann , R & Meisel, H. (2007). Food-derived bioactive peptides with biological activity: from research to food applications. *Curr Opin Biotechnol* 18: 163-169.

Hasan, F., Kitagawa, M., Kumada, Y., Hashimoto, N., Shiiba,M., Katoh, S. & Terashima, M. (2006). Production kinetics of angiotensin-I converting enzyme inhibitory peptides from bonito meat in artificial gastric juice. Process Biochem 41: 505-511.

Je ,J-Y., Kim, S-Y. & Kim, S-K. (2005a). Preparation and antioxidative activity of hoki frame protein hydrolysate using ultrafiltration membranes. *Eur Food Res Technol* 221: 157-162.

Je, J-Y., Park, J-Y., Jung, W-K., , P-J. & Kim, S-K. (2005b). Isolation of Angiotensin I converting enzyme (ACE) inhibitor from fermented oyster sauce, *Crassostrea gigas.Food Chem* 90: 809-814. 104

Je ,J-Y, Park, P-J. & Kim S-K. (2005c). Antioxidant activity of a peptide isolated from Alaska pollack (*Theragra chalcogramma*) frame protein hydrolysate. *Food Res Int* 38: 45-50.

Je, J.Y., Lee, K.H., Lee, M.H. & Ahn, C.B. (2009). Antioxidant and antihypertensive protein hydrolysates produced from tuna liver by enzymatic hydrolysis. *Food Research International*, 42, 1266–1272.

Jia, J.Q., Ma, H.L., Zhao,W.R.,Wang, Z.B., Tian,W.M.,Luo, L. & He, R.H. (2010). The use of ultrasound for enzymatic preparation of ACE-inhibitory peptides from wheat germ protein. *Food Chem*. *119*, 336–342.

Jun, S.Y., Park, P.J., Jung, W.K. & Kim, S.K. (2004). Purification and characterization of an antioxidative peptide from enzymatic hydrolysate of yellowfin sole (*Limanda aspera*)

frame protein. *Eur Food Res Technol* 219: 20-26.

Jung, W-K., Karawita, R., Heo, S.J., Lee, B.J., Kim, S-K. & Jeon, Y.J. (2006). Recovery of a novel Cabinding peptide from Alaska pollack (*Theragra Chalcogramma*) backbone by pepsinolytic hydrolysis. *Process Biochem* 41: 2097-2100.

Jung, W-K,, Qian, Z.J., Lee, S.H., Choi, S.Y., Sung, N.J., Byun, H-G. & Kim, S-K. (2007). Free radical scavenging activity of a novel antioxidative peptide isolated from *in vitro* gastrointestinal digests of *Mytilus coruscus*. *J Med Food* 10: 197-202.

Kitts, D.D. & Weiler, K. (2003). Bioactive proteins and peptides from food sources. Applications of bioprocesses used in isolation and recovery. *Curr Pharm Des* 9: 1309-1323.

Korhonen, H. & Pihlanto, A. (2003). Food-derived bioactive peptides – opportunities for designing future foods. *Curr Pharm Des* 9: 1297-1308.

Kristinsson, H.G. & Rasco, B.A. (2000a). Fish protein hydrolysates: Production, biochemical and functional properties. *Crit Rev Food Sci Nutr* 40: 43–81.

Kristinsson, H.G. & Rasco, B. A. (2000b). Biochemical and functional properties of Atlantic salmon (*Salmo salar*) muscle proteins hydrolyzed with various alkaline proteases. *J Agric Food Chem* 48: 657–

666.

Lin, C.C & Liang, J.H. (2002). Effect of antioxidants on the oxidative stability of chicken breast meat in a dispersion system. *J. Food Chem. Toxiol.* 67(2), 530-533.

Liu, Z., Dong, S., Xu, J., Zeng, M., Song, H. & Zhao, Y. (2008). Production of cysteine-rich antimicrobial peptide by digestion of oyster (*Crossostrea gigas*) with alcalase and bromelin. *Food Control* 19: 231-235.

LIVESEY, G. (1987). Energy and protein requirements the 1985 report of the 1981 Joint FAO/WHO/UNU Expert Consultation. *Nutrition Bulletin*, *12*(3), 138–149. https://doi.org/10.1111/j.1467-3010.1987.tb00040.x

Manyala, J., Berghe, E., & Dadzie, S. (2011). Morphometrics, length-weight relationship and condition of Rastrineobola argentea (Pellegrin 1904) in the Winam Gulf of Lake Victoria (Kenya). *African Journal of Tropical Hydrobiology and Fisheries*, *6*(1). https://doi.org/10.4314/ajthf.v6i1.45932

Mbatia , **Ogonda, L.A, Muge, E.K , Mulaa, F.J** (2014). Antioxidative and functional properties of *Rastrineobola argentea* (*Dagaa*) fish protein hydrolysate. Discourse Journal of Agriculture and Food Sciences 2 (6), 180-189

Meisel, H. & FitzGerald, R. J. (2003). Biofunctional peptides from milk proteins:Mineral binding and cytomodulatory effects. *Pharmaceutical Design*, 9, 1289–1295

Mellander , O. (1950) The physiological importance of the casein phosphopeptide calcium salts. II. Peroral calcium dosage of infants. *Acta Soc. Med. Ups.*, *55*, 247–255.

Miguel, M., Dávalos, A., Manso, M.A., Peña, G., de la, Lasuncion, M.A. & Lopez-Fandino, R. (2008). Transepithelial transport across Caco-2 cell monolayers of antihypertensive eggderived peptides. PepT1-mediated flux of Tyr-Pro-Ile. *Mol Nutr Food Res* 52: 1507-1513.

Murray, B.A. & FitzGerald, R.J. (2007). Angiotensin-converting enzyme inhibitory peptides derived from food proteins: biochemistry, bioactivity and production. *Curr Pharm Des* 13: 773-791.

Nagai, T. & Nagashima, T. (2006). Antioxidative activities and angiotensin-I-converting enzyme inhibition of extracts prepared from chum salmon (*Oncorhynchus keta*) cartilage and skin. *Int J of Food Properties* 9: 813-822.

Ogonda, L. A., Muge, E. K., Mulaa, F. J., & Mbatia, B. N. (2014). Proximate composition of *Rastrineobola argentea* (*Dagaa*) of Lake Victoria-Kenya. *African Journal of Biochemistry Research*, *8*(1), 1–6. https://doi.org/10.5897/ajbr2013.0720

Ogonda, L. A., Muge, E. K., Mulaa, F. J., & Mbatia, B. N. (2017) Optimization of Alcalase Hydrolysis Conditions for Production of Dagaa (Rastrineobola argentea) Hydrolysate with Antioxidative Properties. Ind Chem 3:122. doi:10.4172/2469-9764.1000122

Owaga, E.E., Onyango & C.A., C.K. Njoroge. (2010) .Influence of selected washing treatments and drying Temperatures on proximate composition of *Dagaa* (*Rastrineobola argentea*), a small pelagic fish species. *African journal of Food, Agriculture ,Nutrition and development*;10(7):1-14

Petersen, R.B. (1981). The impact of enzymic hydrolysis process on recovery and use of proteins. In *Enzymes and Food Processing.* Birch, G.G., Blakebrough, N. and Parker,K.J. (Eds.), p. 296-299. Applied Science Pub. Ltd., London, Engl.

Picot,L., S. Bordenave., S. Didelot ., I. Fruitier-Arnaudin., F. Sannier ., G. Thorkelsson , J.P. Berge´., F. Gue´rard ., A. Chabeaud .& J.M. Piot .(2006). Antiproliferative activity of fish protein hydrolysates on human breast cancer cell lines. Process Biochemistry 41: 1217–1222

Quaglia, G. B. & Orban, E. (1987). Enzymatic solubilisation of proteins of sardine (Sardina

pilchardus) by commercial proteases. *Journal of the Science of Food and Agriculture*, 38, 263–269.

Quirós ,A.D., Valos, A., Lasuncin, M-A., Ramos, M. & Recio I. (2008). Bioavailability of the antihypertensive peptide LHLPLP: Transepithelial flux of HLPLP. *Int Dairy J* 18: 279– 286.

Satake, M., Enjoh, M., Nakamura, Y., Takano, T., Kawamura, Y., Arai, S. & Shimizu, M.(2002).Transepithelial transport of the bioactive tripeptide, Val-Pro-Pro, in human intestinal Caco-2 cell monolayers. *Bioscience Biotechnol and Biochem* 6: 378-384.

Samaranayaka, A.G.P, (2010). Pacific Hake (*Merluccius productus*) fish protein hydrolysates with antioxidative properties P.H.D University Of British Columbia

Shahidi, F. (2000). Antioxidants in food and food antioxidants. Nahrung .44; 158 – 163

Shalaby, E. A. (2011). Algae as promising organisms for environment and health. *Plant Signaling and Behavior*, *6*(9), 1338–1350. https://doi.org/10.4161/psb.6.9.16779

Shimizu ,M., Tsunogai, M. & Arai S. (1997). Transepithelial transport of oligopeptides in the human intestinal cell, Caco-2. Peptides 18: 681–687. *Erratum in: Peptides* 1998, 19:791.

Shimizu, M.(2004). Food-derived peptides and intestinal functions.BioFactors,21, 43–47

Suseno, S.H., Tajul, A.Y., Nadiah, W.A., Hamidah, Asti & Ali, S. (2010). Proximate, fatty acid and mineral composition of selected deep sea fish species from Southern Java Ocean and Western Sumatra Ocean,Indonesia. International Food Research Journal 17: 905-914.

Souissi, N., A. Bougatef., Y.T. Ellouz & M. Nasri . (2007). Biochemical and functional properties of sardinella (Sardinella aurita)by-product hydrolysates, Food Technol. Biotechnol. 45 :187–194.

Suseno, S. H., Yang, T. A., Abdullah, W. N., Febrianto, N. A., Asti, W. N., Bahtiar, B., … Materials, A. (2010). *Inventory and Characterization of Selected Deep Sea Fish Species as an Alternative Food Source from Southern Java Ocean and Western Sumatra Ocean , Indonesia*. *4*(8), 1557–1560.

Thangaraj, P. (2016). *Proximate Composition Analysis BT - Pharmacological Assays of Plant-Based Natural Products* (T. Parimelazhagan, Ed.). https://doi.org/10.1007/978-3-319-26811-8_5

Thurstan, R. H., & Roberts, C. M. (2014). The past and future of fish consumption: Can supplies meet healthy eating recommendations? *Marine PollutionBulletin*,*89*(1–2),5–11. https://doi.org/10.1016/j.marpolbul.2014.09.016

UNICEF. (2004). *The State of the World's Children 2004*. https://doi.org/92-806-3784-5

UNICEF. (2009). *The states of the world's children December 2009; Maternal and New born health.*

UNICEF.(2017).*https://www.unicef.org/publications/files/SOWC_2017_ENG_WEB.pdf*. Retrieved from www.soapbox.co.uk

Vermeirssen, V., Deplancke, B., Tappenden, K.A., Van Camp, J., Gaskins,H.R. & Verstraete ,W. (2002). Intestinal transport of the lactokinin Ala-Leu-Pro-Met-His-Ile-Arg through a Caco-2 Bbe monolayer. J Pept Sci 8: 95-100.

Wang, Lei ., Xueying, Mao., Xue, Cheng., Xingxing, Xiong & Fazheng, Ren. (2010). Effect of enzyme type and hydrolysis conditions on the in vitro angiotensin I-converting enzyme inhibitory activity and ash content of hydrolysed whey protein isolate. International Journal of Food Science and Technology 45:807–812

Yamamoto, N., Masahiro, E. & Mizuno, S. (2003). Biogenic peptides and their potential use. Curr Pharm Des 9: 1345-1355.

Yarish, C., & Wamukoya, G. (1990). Seaweeds of potential economic importance in Kenya: field

survey and future prospects. *Hydrobiologia*, *204–205*(1), 339–346. https://doi.org/10.1007/BF00040254

Ziv, E. & Bendayan, M. (2000). Intestinal absorption of peptides through the enterocytes. *Microsc Res* Tech 49: 346–352.

DOCTORAT BIOLOGIE BRETAGNE SANTE LOIRE

Nantes Université

Titre : Clonage et ingénierie de nouvelles cellulases isolées à partir de souches bactériennes du lac Bogoria (Kenya) et applications biotechnologiques

Mots clés : Cellulose, Endoglucanase, Cellobiohydrolase, *Bacillus pumilus*, Alkalophile, Cellulosome

Résumé : Le lac Bogoria située dans la région volcanique de la « rift valley » au Kenya est un écosystème intéressant qui héberge des bactéries de type alcalino-thermophile. A partir d'isolats de souches bactériennes collectées sur les rives du lac, nous avons vérifié la présence d'activités cellulolytiques et entrepris de les cloner. Les gènes correspondant à deux nouvelles cellulases, une endocellulase (BpGH9) et une cellobiohydrolase (BpGH48) ont été sur-exprimés dans une souche recombinante, puis les cellulases correspondantes ont été purifiées et caractérisées biochimiquement. L'endo-cellulase BpGH9 présente une excellente tolérance aux pH extrêmes. L'endocellulase BpGH9 a une structure modulaire dans laquelle le module catalytique est fusionné avec un module de reconnaissance de la cellulose (CBM3c). Des expériences d'ingénierie moléculaire impliquant la troncation et la fusion de CBMs ont permis de préciser le rôle de ces modules dans l'hydrolyse de la cellulose.
Ces nouvelles constructions moléculaires dérivées de BpGH9 ont ensuite été utilisées pour faciliter l'extraction de composés hydrocarbonés à partir de la microalgue, *Botryococcus braunii.* Les résultats indiquent qu'il est possible de désorganiser efficacement les colonies de *B. Braunii* par une hydrolyse avec des endocellulases grâce à la présence de glucanes dans la matrice extracellulaire entourant les cellules. Dans une dernière étape, nous avons exploité la tolérance à la fusion des activités cellulasiques pour construire des complexes multienzymatiques s'inspirant des édifices supramoléculaires observés dans les cellulosomes. Une approche originale dans laquelle les activités enzymatiques pourront être organisées « en ligne » est proposée dans le but d'améliorer la digestion de la cellulose.

Title : Cloning and engineering of new cellulases isolated from Lake Bogoria strains and biotechnological applications

Keywords : Cellulose, Endoglucanase, Cellobiohydrolase, *Bacillus pumilus* , Alkaliphilic ,Cellulosome

Abstract : Lake Bogoria, located in the volcanic region of the rift valley in Kenya, is an interesting ecosystem that hosts alkalino-thermophilic bacteria. From isolates of bacterial strains collected from the shores of the lake, we verified the presence of cellulolytic activities and undertook to clone them. The genes corresponding to two new cellulases, an endocellulaes (BpGH9) and a cellobiohydrolase (BpGH48) were overexpressed in a recombinant strain, then the corresponding cellulases were purified and biochemically characterized. The endocellulase has a modular structure in which the catalytic module is fused with a cellulose binding module (CBM3c). Molecular engineering experiments involving the trucation and fusion of These new molecular constructs derived from BpGH9 were then used to facilitate the extraction of hydrocarbon compounds from the microalgae, *Botriococcus braunii*. The results indicate the B. Braunii colonies can be effectively disrupted by treament with endo cellulases through the hydrolysis of glucanes in the extracellular matrix surrounding the cells.
In a final step, we exploited the fusion tolerance of cellulase activities to construct multienzyme complexes inspired by the supramolecular assemblies observed in cellulosomes. An original approach in which enzymatic activities can be organized "in line" is proposed to improve celulose digestion.