

ANA JOSÉ DE OLIVEIRA NUNES PIRES

BREAKING THE WALL!
DECONSTRUCTION OF THE CELULLOSOME
- A JOURNEY FROM COHESIN-DOCKERIN
EXPRESSION TO NOVEL CARBOHYDRATE BINDING
MODULE STRUCTURES

Orientador: Professor Doutor Pedro Sampaio

Universidade Lusófona de Humanidades e Tecnologias
Faculdade de Engenharia

Lisboa
2015

ANA JOSÉ DE OLIVEIRA NUNES PIRES

BREAKING THE WALL!
DECONSTRUCTION OF THE CELULLOSOME
- A JOURNEY FROM COHESIN DOCKERIN
EXPRESSION TO NOVEL CARBOHYDRATE BINDING
MODULE STRUCTURES

Dissertação defendida em provas públicas na
Universidade Lusófona de Humanidades e
Tecnologias no dia 22/11/2016, perante o júri,
nomeado pelo Despacho de Nomeação n.º: 406/2016,
de 4 de Novembro, com a seguinte composição:

Presidente: Prof. Doutora Adília Charmier

Arguente: Prof. Doutora Cecília Calado (ISEL)

Orientador: Prof. Doutor Pedro Sampaio

Universidade Lusófona de Humanidades e Tecnologias
Faculdade de Engenharia

Lisboa
2015

To my Super Women: Mom, Aunt and Grandmother
And my Super Heroes: Dad and Brother

Acknowledgments

“I saw the light fade from
the sky
On the wind I heard a
sigh
As the snowflakes cover
my fallen brothers
I will say this last
goodbye

Night is now falling
So ends this day
The road is now calling
And I must away
Over hill and under tree
Through lands where
never light has shone
By silver streams that run
down to the sea

Under cloud, beneath the
stars
Over snow one winter's
morn

I turn at last to paths that
lead home
And though where the
road then takes me
I cannot tell
We came all this way
But now comes the day
To bid you farewell

Many places I have been
Many sorrows I have
seen
But I don't regret
Nor will I forget
All who took the road
with me.

Night is now falling
So ends this day
The road is now calling
And I must away
Over hill and under tree
Through lands where
never light has shone

By silver streams that run
down to the sea

To these memories I will
hold

With your blessing I will
go

To turn at last to paths
that lead home

And though where the
road then takes me
I cannot tell

We came all this way
But now comes the day
To bid you farewell

I bid you all a very fond
farewell”

The Hobbit: The
Battle of the Five Armies
– The Last Goodbye
Billy Boyd

I would like to take this opportunity to express my gratitude toward everyone that direct or indirectly support me during my Masters and in completing my thesis. The past three years were, in many aspects, an adventure. This journey has been filled with highs and lows, laughs and tears, setbacks and ultimately great accomplishments. My Masters helped me grow not only academically but also as an individual and for that I'm eternally grateful for those who accompanied me along the way.

Firstly, I would like to thank Professor Carlos Fontes and the Animal Nutrition and biotechnology department at Faculdade de Medicina Veterinária, for accepting me and

providing all the support necessary for my thesis, despite my tight schedule and my odd work hours. I loved the opportunity to develop my work here and to have such an inventive and brilliant team to help. I have learned so much and hope in my future that I will get to work with scientists as inspirational as you and to be at your level someday.

In spite of the rearranges done to my supervision, I am very grateful to have the opportunity to have both Victor Alves and Teresa Ribeiro as my supervisor and co-supervisor.

Teresa Ribeiro, I cannot thank you enough for all the support you gave me, you were always patient and supportive during my internship here. Through this journey you were, probably the person who saw it all, the good, the bad, the doubts and discoveries, and through it all you always had a lesson to teach me and a kind word to help me move forward.

To Victor Alves, I would like to thank the late opportunity to have you guiding me, as well. From the beginning you were always available to clarify all my questions, even though you were not yet my supervisor. Now I don't have word to thank enough for accepting to complete Teresa's guidance and helping me in this last challenge, writing the thesis.

This lab and all the team became a second home for me, I know that I am a quiet and shy person but I really appreciated all the time spend amongst you. To the all Nutrition team a big thank you.

To Pedro Bule, Kate Cameron and Immacolata Venditto thanks for all the support and help you gave me. For all the procedures that you taught me and all the work you have done that supports and complements my own. To Vânia Cardoso, my bench partner to everything we shared, for your friendship and for bringing Molly to work to help us relax and have fun. To Virginia Pires, your company is priceless, for all your advice and breaks thank you very much. To Prof. Shabir for being such a brilliant scientist and for all your help with Rf2 and its structure.

Our lab wouldn't be complete and working smoothly without the helpful hand of Helena Santos, thank you so much for always being there to help and support my work.

In FMV, I couldn't also have reached my goals without the help of the "Reproduction Girls". Thank you Cristina Val for being who you are, for your friendship, your company and motherly advices, is a pleasure having you by my side. To Elizabete Silva,

Mariana Baptista, Marta Baptista e Sofia Henriques for the scientific discussion and advices and all the escapes to “Alegro”.

My Masters would also be impossible without all the support from my own university, Lusófona, where I learned everything I needed to start this journey and where I had the best professors I could had. I’m sorry for not naming you all but there are a few I couldn’t miss.

To my intern supervisors Pedro Sampaio and Susana Santos, for all you taught me, for having the patience to answer all my questions.

To Professor Pedro Vidinha thank you for being the brilliant teacher you are, for teaching us not only the science itself but also helping us develop our scientific and logical thinking skills.

My thanking to Lusófona wouldn’t be complete without Susana Morgado our relentless course secretary. To you the biggest thanks you can imagine for always being there, even when everybody else abandoned us, to solve our problems, for more insignificant they might be, and ultimately save the day. Without you none of us would have even started our Masters.

Without having a place to develop my research and without all the support from my home university all of this would had been very difficult but if I hadn’t all my friends and family by my side I would never even think about starting this in the first place.

To Patricia Diniz, my best friend and lab and group partner, thank you for entering my life and for our unique friendship. This Masters wouldn’t be the same without you, without our late work and all fun nights and without your hard work and motivation. For everything I cannot be grateful enough. Thanks for being my friend.

For my borrowed family, Cristina e Marta Santos, thanks for all your love and friendship for always helping me despite not understanding half of what I’m talking about when I start explaining my Masters.

To my Family, without which I wouldn’t be here in the first place. To my “baby brother” Pedro thank you for your carefree personality, when I’m with you all my problems vanish, thanks for always believing in your big sis.

To my aunt Nena, thank you for always helping me with everything, you know how important your support and love are for me.

To my grandmother Toia, thanks for being the most persistent and strongest woman I know, it is in your story of life that I seek strength for those darkest days.

To my daddy José, we have our ups and downs but you are my super hero and role model. Thank for your love and education, for always pushing me forward despite not always in the smoothest way. I only hope someday I'll be just an inch as good as you are.

To my mommy Paula, what can say? Thanks for being able to survive living with me and dad in the same house, just kidding. I cannot thank you enough mom, I know I don't always show it but I admire your strength and kindness. Despite all the bumps in our relationship I hope you know how much I love you and how important your support is for me.

Last but the least, to “my guys”, my brothers from another mother: Marco Ramos, Ricardo Babo, Rui Gomes, Marcos Sousa, Eliana Lobão and Yuliya Marchunk thanks to all of you for your friendship. You were all always by my side, in the best and worst moments of my life, you wiped my tears and never let my glass get empty. Thanks for withstanding all my insecurities and helping me surpass them. Thanks for all the parties that help me get through all the stress. Thanks just for being the best friends a person could have.

Finally I would like to thank in advance everyone I might be forgetting but I considered myself lucky to have all these amazing people walking beside and sometimes pulling me forward along this stage of my life. My sincere gratitude to all.

Abstract

Ruminant herbivores establish highly complex symbiotic relationships with anaerobic bacteria that produce cellulosomes to efficiently degrade plant cell wall polysaccharides. Cellulosomes are composed of modular enzymes containing catalytic modules which are linked to non-catalytic modules involved in protein:carbohydrate (Carbohydrate-Binding Modules or CBMs) or protein:protein (Cohesins:Dockerins or Coh-Doc) interactions.

This project aims to aid in the quest of broadening the knowledge of the cellulosome and its components. Exploring both the Cohesin-Dockerin complex and the CBMs.

Despite already well characterized, the production and crystallization process of *Bacteroides cellulosolvens*'s Cohesin-Dockerin Complex, still needs some optimization. By forcing the complex to adopt only one of the two possible conformations, through direct mutagenesis, we intended to test and optimize protein purification.

The resulting positive mutants were tested in different growth conditions from which the best Dockerin expression were obtain for BL21 induced at 25°C in LB.

Among *Ruminococcus flavefaciens* strain FD-1 uncharacterized CBMs, we selected Rf2 and Rf4 to determine its substrate affinity to: HEC, β -Glucan, Galactomannan and Xyloglucan, using affinity gel electrophoresis.

From which we concluded that RF2 presents higher affinities for Galactomannan and Xyloglucan ranging from 56 % to 49% (w/v)⁻¹, respectively, and Rf4 only has affinity for Xyloglucan with a significant Ka of 296% (w/v)⁻¹.

Lastly, this novel CBM-Rf2 structure was solved at the As edge by single wavelength anomalous dispersion.

Keywords: Cellulosome, cohesin, dockerin, *Bacteroides cellulosolvens*, *Ruminococcus flavefaciens*, Carbohydrate binding module (CBM), CAZymes

Resumo

Os Ruminantes estabelecem relações simbióticas, altamente complexas, com bactérias anaeróbias produtoras de celulosomas de forma a degradarem eficientemente os polissacarídeos da parede celular de vegetal. Os celulosomas são compostos por enzimas modulares constituídas por módulos catalíticos ligados a módulos não-catalíticos, envolvidos em ligações proteína:carboidratos (Carbohydrate-Binding Modules ou CBMs) e interações proteína:proteína (Coesinas:Doquerinas ou Coh-Doc).

Este projeto visa ampliar o conhecimento dos celulosomas e dos seus componentes, explorando tanto o complexo Coesina-Doquerina como os CBMs.

Apesar de já se encontrar bem caracterizado, o processo de produção e cristalização do complexo Coesina-Doquerina de *Bacteroides cellulosolvens*, ainda necessita de alguma otimização. Forçando o complexo a adotar apenas uma das suas duas conformações possíveis, através de mutagénese direta, pretende-se testar e otimizar a sua purificação.

Os mutantes positivos foram testados em diferentes condições de crescimento, sendo a melhor expressão obtida em células BL21 induzidas a 25°C em meio LB.

Dos vários CBMs ainda por caraterizar de *Ruminococcus flavefaciens* estirpe FD-1 selecionámos o Rf2 e o Rf4 para, através de géis de afinidade, determinar a sua afinidade a diversos substratos: HEC, β -Glucano, Galactomanano e xiloglucano.

Concluimos que a RF2 apresenta maior afinidade para Galactomanano e xiloglucano variando entre 56 e 49 (w/v)⁻¹, respetivamente. A Rf4 apresenta apenas afinidade para o xiloglucano com um Ka significativo de 296 (w/v)⁻¹.

Por último, a estrutura deste novo CBM-Rf2 foi resolvida por difração de raios-X por dispersão anómala de comprimento de onda único.

Palavras-chave: Celulosoma, Coesina, Doquerina, *Bacteroides cellulosolvens*, *Ruminococcus flavefaciens*, módulos de ligação a hidratos de carbono (CBMs), CAZymes

List of Abbreviations and Symbols

%	Percentage
ΔH	Enthalpy
ΔS	Entropy
μg	Microgram
μl	Microliter
μm	Micrometer
μM	Micromolar
AE	Agarose electrophoresis
AGE	Affinity gel Electrophoresis
Amp	Ampicillin
BSA	Bovine serum albumin
Ca(OH)_2	Calcium hydroxide
CaCl_2	Calcium chloride
CAZymes	Carbohydrate-Active enzymes
CBD	Cellulose binding domains
CBM	Carbohydrate-binding modules
CBP	Consolidated bioprocessing
cm	Centimetre
$\text{CoCl}_2 \cdot 6\text{H}_2\text{O}$	Cobalt(ii) chloride hexahydrate
Coh	Cohesin
DNA	Deoxyribonucleic acid
Doc	Dockerin
dp	Degree of polymerization
EDTA	Ethylenediaminetetraacetic acid
FAO	Food and Agricultural Organization of the United Nations
$\text{FeCl}_3 \cdot 6\text{H}_2\text{O}$	Iron(iii) chloride hexahydrate
FPLC	Fast protein liquid chromatography

<i>g</i>	Gravitational acceleration
GH	Glycoside hydrolases
GM	Genetically modified
GMO	Genetically modified organisms
GT	Glycosyl transferases
h	Hour
HEC	Hydroxyethyl cellulose
HG	Homogalacturonan
IL	Ionic liquids
IMAC	Immobilized metal ion affinity chromatography
IPTG	Isopropyl β -D-thiogalactopyranoside
ITC	Isothermal titration calorimetry
K _a	Affinity constant
Kan	Kanamycin
KCl	Potassium chloride
K _d	Dissociation constant
KOH	Potassium Hydroxide
L	Liter
LB	Luria bertani medium
LMW	Low molecular weight
M	Molar
Mes	Mesophilic
mg	Milligram
MgCl ₂	Magnesium chloride
MgSO ₄	Magnesium sulfate
min	Minutes
mL	Milliliter
mM	Millimolar
mmol	Millimole
MnCl ₂ -4H ₂ O	Manganese(II) chloride tetrahydrate
Mt	Megaton

MW	Molecular weight
n	Reaction stoichiometry
Na ₂ SO ₄	Sodium sulfate
NaCl	Sodium chloride
NaOH	Sodium hydroxide
NaPO ₄	Sodium phosphate
ng	Nanogram
NH ₄ Cl	Ammonium chloride
NiCl ₂ ·6H ₂ O	Nickel(ii) chloride hexahydrate
nm	Nanometer
°C	Celsius degree
OD	Optical density
OH	Hydroxyl groups
PDB	Protein data bank
PE	Polyethylene
PLA	Polylactide
PL	Polysaccharide lyases
RG-I	Rhamnogalacturonan I
RG-II	Rhamnogalacturonan II
rpm	Rotations per minute
s	Seconds
	Sodium dodecyl sulfate-polyacrylamide gel
SDS PAGE	electrophoresis
SLH	S-layer homology
SOB	Super optimal broth
Ther	Thermophilic
TBE	Tris-borate-EDTA buffer
TEMED	Tetramethylethylenediamine
T	Temperature
WT	Wild type
ZnSO ₄ ·7H ₂ O	Zinc sulfate heptahydrate

Index

Introduction	1
1. Work Objectives:	6
a) Insights into the <i>Bacteroides cellulosolvers</i> 's Cohesin-Dockerin Complex	6
b) Biochemical Characterization of Rf2 and Rf4 CBM's of <i>Ruminococcus flavefaciens</i>	8
c) Insights into the Rf2 CBM structure of <i>Ruminococcus flavefaciens</i>	9
Chapter 1 - Scientific Background	10
1.1 - Plants As The Main Source Of Biomass	10
1.1.1 - Abundance	10
1.1.2 - Applicability	12
1.2 - The Plant Cell-Wall	21
1.2.1 - Structure	21
1.2.2 - Main Constituents	23
1.3 - Breaking the wall	30
1.3.1 - Conventional treatments	30
1.3.2 - Biological Treatments	32
1.4 - The cellulosome	36
1.4.1 - Key Enzymes	41
1.4.2 - Cohesin-Dockerin Interactions	44
1.4.3 - Carbohydrate Binding Modules	53
1.5 - Process Improvements	67
1.5.1 - Plant Engineering	67
1.5.2 - Engineered microbial systems	68
Chapter 2 - Materials and Methods	72
2.1 - General	72
2.1.1 - Microbial strains and plasmids	72
2.1.2 - Antibiotics	73
2.2 - Insights into the Cohesin-Dockerin Complex of <i>Bacteroides cellulosolvers</i>	74
2.2.1 - Mutants Construction	74
2.2.2 - Protein Expression Optimization	83
2.3 - Biochemical Characterization of <i>Ruminococcus flavefaciens</i> putative CBMs Rf2 and Rf4	87
2.3.1 - Macromolecule production	88
2.3.2 - Substrate Affinity Gels	89
2.4 - Insights into the Rf2 CBM structure of <i>Ruminococcus flavefaciens</i>	92
2.4.1 - Overall procedure	92
Chapter 3 - Results and Discussion	97
3.1 - Insights into the Cohesin-Dockerin Complex of <i>Bacteroides cellulosolvers</i>	97
3.1.1 - Mutants Construction	97
3.1.2 - Protein Expression	98

3.2 - Biochemical Characterization of <i>Ruminococcus flavefaciens</i> putative CBMs Rf2 and Rf4	103
3.2.1 - Substrate affinity Gels	103
3.3 - Insights into the Rf2 CBM structure of <i>Ruminococcus flavefaciens</i>	112
3.3.1 - Sequence Analysis	112
3.3.2 - Crystallization	114
Chapter 4 - Conclusions	119
Chapter 5 - Bibliographic References	123
Appendixes	I
Appendix A - Reagents and Solutions	I
1) Water	I
2) Reagents	I
3) Growth Medium	III
4) Autoinduction Growth Medium (LBE)	IV
Appendix B - Gel Support Materials	VI
1) SDS PAGE	VI
2) Native	VI
3) AG buffers	VII
4) Bleaching Solution	VII
Appendix C - Protocols Schematics	VIII
1) Mutants Construction	VIII
2) Protein Expression	XI
Appendix D - Substrate Affinity Support Data	XIII
1) Rf2	XIII
2) Rf4	XVI
Annexes	XIX
Annex I - Crystallization Solutions	XIX
1 - Crystal Screen I & II	XIX
2 - Peg/Ion I&II	XXI
3 - 80!	XXIII
4 - JCSG	XXIV

List of Figures

Figure 1 - Key issues for improving energy resources	1
Figure 2 - <i>C. thermocellum</i> type-II cohesin–dockerin interaction around the dockerin’s 22/22’ sequence positions.	7
Figure 3 - <i>B. cellulosolvens</i> Coh-Doc Constructs.....	8
Figure 4 – Rf2 organization and amino acid sequence.....	9
Figure 5 - Vegetal Biomass present in Europe from 1990 to 2010.	11
Figure 6 - Global usage of land for feed and food purposes in 2006/07.	12
Figure 7 – Representative diagram of the contrast between the conventional industry and the novel biorefineries.	13
Figure 8 - Products obtained by the fermentation of glucose.....	15
Figure 9 - EU energy potential from vegetable biomass.	18
Figure 10 – Principal options for biomass conversion to secondary energy carriers	19
Figure 11 - Plant plasma membrane and cell-wall structure.	21
Figure 12 - Main cell wall constituents.	23
Figure 13 – Organization of Cellulose.	25
Figure 14 – Xyloglucan structure	26
Figure 15 – β -Glucan Structure.	27
Figure 16 - Structures of the main pectic polysaccharides and in pectin-containing cell wall proteoglycans.....	29
Figure 17- Classification of microbial taxons bearing putative CAZymes.	33
Figure 18 - Principal cellulolytic bacteria and their major morphological features.	35
Figure 19 - Cellulosome assembly mechanism.	37
Figure 20 - <i>B. cellulosolvens</i> cellulosome system.....	39
Figure 21 - Schematic representation the cellulosome-related proteins in <i>R. flavefaciens</i> . .	40
Figure 22 – Main carbohydrate-active enzymes and their action mechanism	41
Figure 23 - Action mechanism of carbohydrate esterase.	44
Figure 24- Structure of a type-I cohesin-dockerin interaction from <i>Clostridium thermocellum</i>	46

Figure 25 The dual binding mode of the Xyn10B dockerin.....	48
Figure 26 - Structure of the type-II cohesin-Xdockerin complex	50
Figure 27- Symmetric and Asymmetric dockerin sequences	51
Figure 28 - Putative recognition residues of different dockerin domains derived from cellulosomal components of different species.....	52
Figure 29 - CBM distribution in different proteins.	55
Figure 30 - CBMs classification based on fold.	57
Figure 31 – CBM from type A. Type A CBM1 (PDB:1cbh) from <i>Hypocrea jecorina</i> L27	58
Figure 32- CBM from type B. Type B CBM4-2 (PDB:1GU3) from <i>Cellulomonas fimi</i> ATCC 484.	59
Figure 33 - CBM from type C. Type C CBM13 (PDB:1MC9) from <i>Streptomyces lividans</i>	60
Figure 34 - The three types of CBMs.....	62
Figure 35 - CBMs present in toxins, virulence factors or pathogenesis associated proteins	65
Figure 36 - Artificial multi-functional enzyme complex.	69
Figure 37 - Multi-functional enzyme complex for biomass utilization.....	70
Figure 38 - BC2-Coh molecular construction strategy.....	76
Figure 39 - BC2-Doc molecular construction strategy.....	77
Figure 40 - Overall procedure for the Doc-Coh mutant's construction.	78
Figure 41 - DNA Restriction Procedure	81
Figure 42 - Vector-Insert assembly procedure	82
Figure 43 - Overall procedure for the biochemical characterization of Rf2 and Rf4.....	88
Figure 45 – Typical AGE.	90
Figure 45 - Schematic representation of the overall Rf2 crystallization procedure.	92
Figure 46 – 1% AE of samples.....	97
Figure 47 – Growth Condition at 16°C.....	99
Figure 48 - Growth Condition at 19°C	100
Figure 49 - Growth condition at 25°C	101
Figure 50 – A Coomassie Brilliant Blue-stained 14 % SDS-PAGE gel and 10 % native gel for evaluation of protein expression.	102
Figure 51 - AGE for RF2.....	105

Figure 52 - AGE for Rf4.	106
Figure 53 – Rf2 Substrate Affinity analysis..	108
Figure 54 - Rf4 substrate affinity analysis..	109
Figure 55 - Rf2 and Rf4 substrate affinity comparison.	110
Figure 56 – Rf2 sequence organization.	112
Figure 57 – Linker sequence analysis.....	112
Figure 58 – Rf2 crystals and respective conditions.	114
Figure 59 – Rf2 mut crystals and respective conditions.....	115
Figure 60 - The three-dimensional structure of Rf2,	117
Figure 61- The three-dimensional structure of Rf2 with surface,	118
Figure 63 – Ultimate Goal for Cellulosomes Application.....	121
Figure 63- DNA Extraction Procedure	VIII
Figure 64 - AE Gel Purification Procedure 1 of 2.....	IX
Figure 65- AE Gel Purification Procedure 2 of 2.....	X
Figure 66- Protein extraction procedure.....	XI
Figure 67- Protein Expression Procedure	XI
Figure 68 - Protein Purification Procedure.....	XII

List of Tables

Table 1 – Dockerin amino acid sequences..	7
Table 2 – Categories of biomass according to its carbohydrate nature source.....	10
Table 3 – Main Emerging Biobased Plastics for non-food applications.	16
Table 4 – Applicability of PE derivatives in Western Europe	17
Table 5 – Cellulosome-producing microorganisms	36
Table 6 - GH clans of related families.....	43
Table 7 – Microbial Strains and Plasmids Characteristics.	73
Table 8 - Stock and work antibiotic solutions.	73
Table 9 – Dockerin Amino acid sequences.	74
Table 10 - Final Constructs.	75
Table 11 - Recombinant vector and mutant insert characteristics.....	79
Table 11 - Protein Expression Conditions.....	84
Table 12 – Substrates tested for the initial essay.....	87
Table 13 – Rf2 and Rf4 macromolecule production information.	89
Table 14 - Substrates concentrations.	91
Table 16 - Crystallization procedure.	94
Table 17 - Rf2 protein sequences.	94
Table 18 - Data collection and processing.....	96
Table 19 – Concentration and purity evaluation.	98
Table 20 - Substrate Affinity for several CBMs..	104
Table 21 – NCBI Blast for Rf2mut	113
Table 22 – Crystallization conditions selected.	116
Table 23 - HEC AGE results	XIII
Table 24 - Xyloglucan AGE results	XIV
Table 25 - β -Glucan AGE results	XIV
Table 26 -Galactomann AGE results.....	XV
Table 27 - HEC AGE results	XVI
Table 28 - Xyloglucan AGE results	XVII
Table 29- β -Glucan AGE results	XVII

Table 30-Galactomann AGE

XVIII

Introduction

Since the beginning of civilization mankind struggles to improve the usage of our planet resources efficiently, and yet we face ourselves with a serious question: how can we overcome a pending energy crisis?

As usual the answer is rather complex. When most of our energy system is based on fossil fuels it has been extremely difficult to find a universal and economically appealing replacement (Aransiola, Ojumu, Oyekola, Madzimbamuto, & Ikhu-Omogbe, 2014; Bharathiraja *et al.*, 2014; Khan *et al.*, 2014; Ziolkowska, 2014). Despite alternatives, like biodiesel or bioethanol, being actively studied since roughly 1970, the green fuel movement only gained momentum with the crude oil crisis and increased environmental concerns in the turn of this millennium. Even so, after more than a decade we're still quite far from replacing fossil fuels with biofuels. In order to make the later a viable possibility there are a few key issues that must be addressed for its successful implementation:

- i) Which feedstock to use?
- ii) How will it be produced?
- iii) Compared with diesel is it profitable?

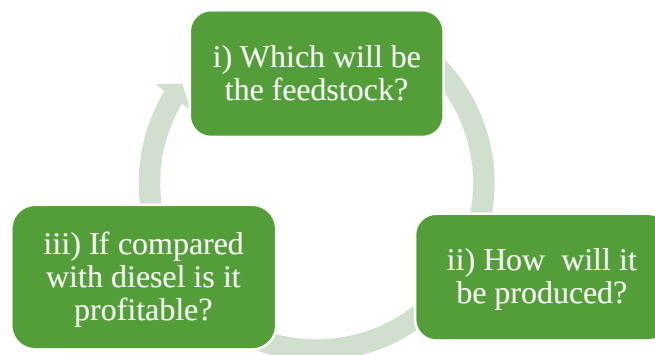


Figure 1 - Key issues for improving energy resources

To clarify these premises it's essential to understand the basics of biofuels. Biofuels can be divided in two main categories: liquid and gaseous such as bioethanol or biodiesel and methane or hydrogen respectively. Unlike the conventional fuels, besides a series of previous treatments, biofuels derived from organic materials undergo a fermentation process in order to generate the desired product.

These alternative fuels can replace petroleum-based fuels, for instance gasoline and diesel, without the need to implement significant changes on the technology already in use, providing the foremost advantage of reducing the nefarious emissions of greenhouse gases (Carere, Sparling, Cicek, & Levin, 2008).

Biodiesel has been identified as one of the notable options for at least complementing conventional fuels. Its production from renewable biological sources such as vegetable oils and fats has been reviewed extensively (Aransiola *et al.*, 2014; Rincón, Jaramillo, & Cardona, 2014).

Its advantages over petroleum diesel cannot be overemphasized: it is safe, renewable, non-toxic, and biodegradable; it contains no sulphur; and it is a better lubricant. In addition, its use engenders numerous societal benefits: rural revitalization, creation of new jobs, and reduced global warming. Its physical properties have been reviewed widely as well, some of which are dependent on the feedstock employed for its production (Aransiola *et al.*, 2014).

i) Biofuel - Which feedstock to use?

Thanks to the large range of organic matter available in your planet we are able to select the most adequate substrate from a diverse array of origins. Nevertheless, plant biomass is the most consensual source that meets our purposes (Lynd 2002).

There are several reports on biodiesel production from edible oils; thus, its competition with food consumption has been a global concern. About 6.6 Tg (34%) of edible oil was estimated for worldwide biodiesel production from 2004 to 2007, and biodiesel is projected to account for more than a third of the expected growth in edible oil use from 2005 to 2017. Consequently, employing waste and non-edible oils in biodiesel production would eliminate the competition with food consumption; it will also allow for compliance with ecological and ethical requirements for biofuel (Aransiola *et al.*, 2014; Bharathiraja *et al.*, 2014; Macías-Sánchez *et al.*, 2015; Rincón *et al.*, 2014; Ziolkowska, 2014).

Brazil, one of the largest worldwide bioethanol producers, uses sugarcane as the raw material to convert sucrose to ethanol via yeast mediated fermentation. In the USA the biofuel is produced through an enzymatic process that breaks-down grain starch to glucose, which

can then be fermented to ethanol (Maheshwari, 2008). Nevertheless, the use of this type of biomass was questioned, on the basis of promoting food and farmland use to feed our cars instead of our people. Therefore a more appealing alternative emerged, i.e. discarded cellulosic biomass (Carere *et al.*, 2008).

The plant cell wall, in which the main component is cellulose, comprises the majority of the vegetal biomass. The cellulose bio-polymer is synthesized at an approximate rate of 7.5×10^{10} tons per year, occupying the first place in the abundance scale of existing natural polymers on Earth (E. Bayer, Lamed, & Himmel, 2007).

ii) Biofuel - How to produce it?

Extracting the fermentable sugars from sugarcane or starch is a well-established method, very similar to the ones used in food industries, and has become the preferable technique for producing bioethanol (Aransiola *et al.*, 2014; Bharathiraja *et al.*, 2014).

Generally non-edible feedstocks including waste vegetable oils, animal fats and non-food crops are produced by conventional transesterification reaction. However, owing to the limitations of the conventional methods, new technologies are starting to be developed. Biodiesel could be produced by different technological processes, mainly transesterification using homogeneous catalysts as well as heterogeneous catalysts (Aransiola *et al.*, 2014; Bharathiraja *et al.*, 2014; Rincón *et al.*, 2014; Ziolkowska, 2014).

All these available methods are capable of producing biodiesel from refined oil which is the most common source of raw material for its production. However, they have their own advantages and disadvantages. The acid-catalyzed homogeneous transesterification has not been widely investigated and employed, when compared to the alkali-catalyzed process, due to its limitations such as slower reaction rates, the need for tougher conditions (higher temperatures, methanol-to-oil molar ratios and quantities of catalysts) and the formation of undesired secondary products such as dialkyl or glycerol ethers. Therefore it is less attractive to industrial purposes (Bharathiraja *et al.*, 2014).

On the other hand, the main problem associated with the heterogeneously catalyzed transesterification is the deactivation due to the presence of water, which is normally produced from the esterification reaction (Bharathiraja *et al.*, 2014).

Enzymes are believed to be a good choice to produce biodiesel; they can easily treat fatty acids as well as triglycerides to produce biodiesel from non-edible oils with higher conversions. In contrast, degrading the recalcitrant plant cell wall, pursuing the enzymatic accessibility of cellulose is extremely difficult and involves the combination of powerful chemicals, high temperatures and various enzymes before it can be fermented. Therefore their high production cost limits their employability (Bharathiraja *et al.*, 2014).

This may be overcome by using molecular technologies to enable the production of enzymes in higher quantities as well as in a virtually purified form. The most common and simple non-catalyzed biodiesel production process is by using supercritical methanol. Though the procedure is claimed to be effective, it is highly expensive. Hence, there has been more research to explore new technologies for the production of biodiesel considering their industrial economic viability (Bharathiraja *et al.*, 2014).

iii) Biofuel - Compared with diesel, is it profitable?

Ultimately, when evaluating the current alternatives a profitable and ideal alternative still does not exist (Aransiola *et al.*, 2014; Bharathiraja *et al.*, 2014; Choi, Song, Cha, & Lee, 2014; Khan *et al.*, 2014; Macías-Sánchez *et al.*, 2015; Rincón *et al.*, 2014; Ziolkowska, 2014).

On the one hand we could use an easier substrate but with substantial impacts on the food supply and fare markets. However, there are concerns regarding whether a growing population can be fed in a sustainable manner or not. The development of innovative technologies resulted in both improved genetic traits and advanced crop management. Despite these trends there is a decline of rice yields from 1985 onwards. In spite of these variations in the yield of different crops, there is still a gap between the growth of production and demand of supply. Additionally, there may be other factors but the demand of edible feedstocks for biofuel cannot be ruled out (Bharathiraja *et al.*, 2014).

As far as biofuels are concerned, it is argued that one must distinguish between biofuels driven by market forces and biofuels driven by government policy. However, it is accepted globally that biofuels produced from edible feedstocks cannot replace the petroleum fuels without impacting on food supplies (Bharathiraja *et al.*, 2014).

However, on the other hand we can re-utilize a more available and low-cost feedstock with a more complex production process, therefore increasing the overall production fee (Aransiola *et al.*, 2014).

Nonetheless most studies tend to believe that the answer is based on reducing the costs of processing Plant Biomass either through plant cell wall engineering to make it easier to break or by improving the methods to degrade it (Bharathiraja *et al.*, 2014).

Some research suggests a possible solution by genetically engineering the plant cell wall to make it easier to degrade. However such procedure would lead to, once more, the withdrawal of valuable crops soil and also entailed some risks associated with the production of Genetically Modified Organisms (GMO) (Boudet, Kajita, Grima-Pettenati, & Goffner, 2003).

In this project, it is considered that the solution to this dilemma lies in the development of a viable method for simplifying the plant cell wall breaking process. Moreover, in order to look for answers the search began with the most efficient system existent, Mother Nature, more precisely the herbivores, and their capability of degrading cellulosic matter very rapidly, in less than 24h. The process takes place in the amazing gastrointestinal compartments, namely the rumen or caecum, residence of a large consortium of symbiotic anaerobic bacteria with a unique characteristic, meaning the capability to synthesize an enzyme cocktail assembled as an extracellular mega-Dalton complex, named cellulosome. The cellulosomes are composed of modular enzymes containing catalytic modules which are linked to non-catalytic modules involved in protein:carbohydrate (Carbohydrate-Binding Modules or CBMs) or protein:protein (Dockerins or Doc) interactions. Within the cellulosome, the Carbohydrate-Active enZymes (CAZymes) (Lombard, Golaconda Ramulu, Drula, Coutinho, & Henrissat, 2014) bind to Cohesin (Coh) modules present in a molecular scaffold, the scaffoldin, via enzyme borne Doc domains (Type-I Coh-Doc interactions). Cellulosomes can also be anchored to the host microbial cell surface through type-II interactions, between scaffoldin-borne type-II Doc with type-II Coh on the cell envelope. Dockerins characteristically present two duplicated segments, each defining a Coh-binding interface displaying similar ligand specificities, thus leading to a possible dual-binding mode (Carvalho *et al.*, 2007; Fontes & Gilbert, 2010).

Inspired by this natural cellulose degrading factory, a new strategy for processing vegetal biomass arose, the so called Consolidated Bioprocessing (CBP). CBP comprises, in a one step process, the substrate hydrolysis and the fermentation by using cellulolytic bacteria (Carere *et al.*, 2008). In order for this process to be applied it's essential to fully understand the mechanism used by these microorganisms and more specifically how the cellulosome works and how can we improve it.

Assembly of CAZymes in cellulosomes potentiates protein stability and enzyme synergistic interactions. Cellulosomes comprise diverse CAZymes (glycoside hydrolases - GH, carbohydrate esterases - CE and polysaccharide lyases - PL) displaying a modular architecture in which a catalytic domain can be connected, via linker sequences, to one or more non-catalytic carbohydrate-binding modules (CBMs).

With that as the ultimate goal, this project aims to aid in the quest of broadening the knowledge of the cellulosome and its components. We can separate the work plan in two distinct paths centered on the different modules of the cellulosome: the Cohesin-Dockerin complex and the CBMs.

1. Work Objectives:

a) Insights into the *Bacteroides cellulosolvens*'s Cohesin-Dockerin Complex

Bacteroides cellulosolvens's cellulosome consists of two large scaffoldins: ScaA, an enzyme-binding primary scaffoldin with 11 Coh modules and a C-terminal Doc, and ScaB, an anchoring scaffoldin that bears 10 Coh (Xu *et al.*, 2004). In contrast to other cellulosome systems, the apparent roles of the *B. cellulosolvens* Coh are inverted, the type-II Coh are located on ScaA and mediate enzyme attachment, whereas type-I Coh are located on ScaB and are involved in cell surface attachment. Despite this cellulosome architecture being well characterized there are still unknown details on these Coh-Doc complex interactions, their production and crystallization processes, namely due to the dual binding mode capability of the type-II interactions, that hinders *in vitro* production of stable and homogeneous complexes, and as such prevents their crystallization and structure solving.

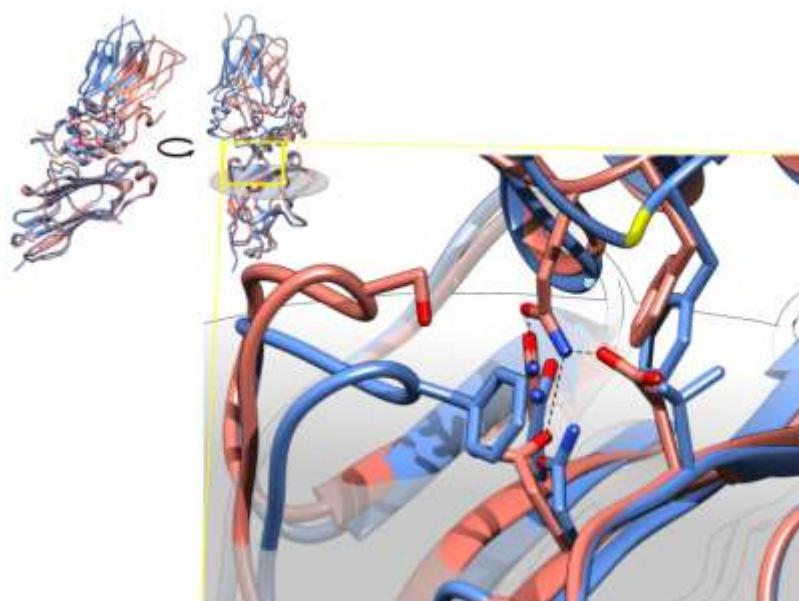


Figure 2 - *C. thermocellum* type-II cohesin–dockerin interaction around the dockerin’s 22/22’ sequence positions. The blue Doc structure seen on the top has the glycine residue highlighted in yellow and it is superposed with a salmon colored dockerin were the Gly is substituted by an Asn (shown in a stick molecular representation). Emerging from the bottom are several residues from two different cohesin molecules (superposed and also colored blue and salmon). Several sterical clashes are apparent on the salmon colored Coh-Doc complex. By mutating the glycine for an asparagine the objective is to force the dockerin into a single possible conformation on binding with its cognate cohesin, thus allowing the needed stability for successful crystallization.

Consequently this project aims to force the complex to adopt only one of the conformations by alternatively mutating the critical residues at the 22 and 22’ sequence positions of the dockerin as shown in Table 1, substituting a Gly-for an Asp- The rationale for this mutation is based on predicted sterical clashes between the novel asparagine residue against cohesin residues, previously shown to occur on *Clostridium thermocellum* type-II Coh-Doc complexes by the group’s research team (Figure 2).

Table 1 – Dockerin amino acid sequences. The dockerin has a duplicated sequence which conventionally is numbered from a conserved Gly residue (position #1) whereas the corresponding residues on the duplicated segment are numbered by adding a superscript prime symbol (#1’). The residues marked in red represent the canonical pair with a decisive importance on the Coh-Doc interaction at the 11 and 11’residues. The residues in positions 22 and 22’ are marked in blue for the original amino acid and in green for the mutated one.

Dockerin Protein Sequence (N –C)	
BC2Doc-wt	GDLNGDGVINMADVMILAQSFC ²² KAIGNPGVNEKADLNNDGVINMADAI ILAQYF ^{22'} GK
BC2Doc-mut1	GDLNGDGVINMADVMILAQSFN ²² KAIGNPGVNEKADLNNDGVINMADAI ILAQYF ^{22'} GK

BC2Doc- mut2	GDLNGDGVINMADVMILAQSF ^G KAIGNPGVNEKADLNNDGVINMADAI ILAQYFNK
-----------------	---

In order to test and optimize protein purification, an HisTag was added either to the N-terminal Dockerin or to C-terminal Cohesin, thus originating 4 constructs (Figure 3).

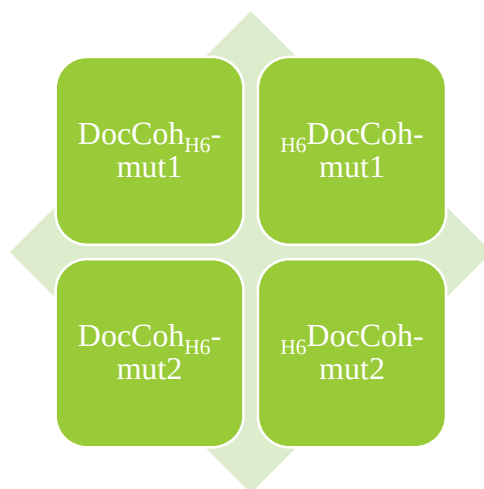


Figure 3 - *B. cellulosolvens* Coh-Doc Constructs. The constructs will test the influence of positioning the histidine tag either on the dockerin (H₆Doc) or on the cohesin (Coh_{H6}) side, on the protein expression. To force the Doc to adopt a single complex conformation, the original sequences were mutated on the residue position 22 (mut1) and 22' (mut2).

The Coh-Doc complex expression will be tested to determine the best growth conditions for the complex's large scale production.

b) Biochemical Characterization of Rf2 and Rf4 CBM's of *Ruminococcus flavefaciens*

CBM's are non-catalytic domains that usually are found associated with enzymes capable of degrading plant cell wall polysaccharides. They act as an anchor that helps the enzyme bind to the substrate. CBMs direct the appended catalytic modules to their target substrates, thus potentiating catalysis. The genome of the ruminal cellulolytic bacterium, *Ruminococcus flavefaciens* strain FD-1, contains over 200 modular proteins bearing the cellulosomal-signature dockerin module. Among these we selected the, yet to be characterized, Rf2 and Rf4 to determine its substrate affinity to:

- HEC
- β -Glucan
- Galactomannan

- Xyloglucan

c) Insights into the Rf2 CBM structure of *Ruminococcus flavefaciens*

Lastly, we intend to further characterize this unclassified Rf2 CBM using *in silico* analysis to probe other similar proteins, as well as promote its crystallization to determine its structure through x-ray bio-crystallography.

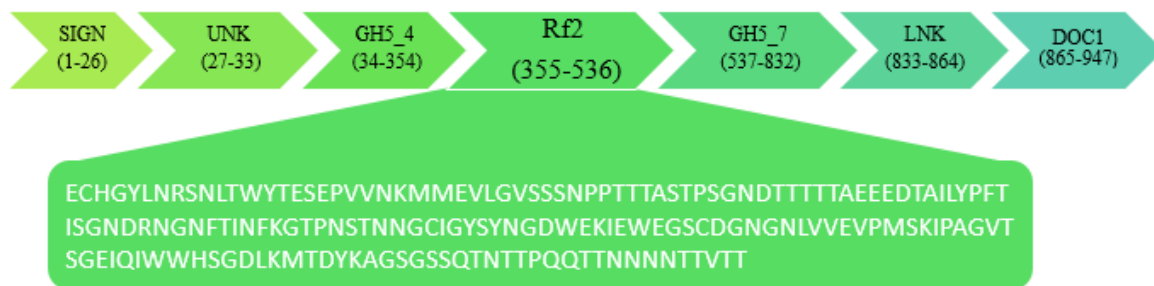


Figure 4 – Rf2 organization and amino acid sequence. The Rf2 is a protein from *Ruminococcus flavefaciens*, and this specific location of the gene flanked by two GH is consistent with other CBM's therefore supporting the conjecture that this sequence is in fact a CBM.

Chapter 1 - Scientific Background

1.1 - Plants As The Main Source Of Biomass

Appreciatively, our world is vastly populated by trees and other plants constituting the most ample form of biomass with an estimated production of 10^{11} metric tons per year and energy content of about 2.425×10^{18} KJ. The increased cost of fossil fuels heightened the interest on alternative forms of energy production such as biomass derived energy (Hyeon, Jeon, & Han, 2013; Singh *et al.*, 2014).

Although vegetal biomass is not homogeneous, with a variable composition according to its phylogeny, it shares some common main component carbohydrates, which can be sorted on three key categories as showed in Table 2 (Mateos & González, 2007).

Table 2 – Categories of biomass according to its carbohydrate nature source (Mateos & González, 2007).

Category	Carbohydrates		Exemple
Sugary	Monosaccharides	Glucose	Fruit Pulp
		Fructose	
	Disaccharides	Sucrose	Sugar cane, sweet sorghum and beet
Starch	Polysaccharides	Inulin	Potato tubercles, chicory
		Starch	Cereal grains
Lignocellulosic	Polysaccharides	Hemicellulose and Cellulose	Woods

Since the two first categories are primarily used as food supplies we are going to focus more on lignocellulosic biomass and some vegetal residues from starch production to reuse as feedstock to other processes.

1.1.1 - Abundance

According to data extracted from Food and Agricultural Organization of the United Nations (FAO), biomass available above-ground has been increasing over the years, reaching in Europe an amazing value of 2.2×10^6 million metric tons as indicated on Figure

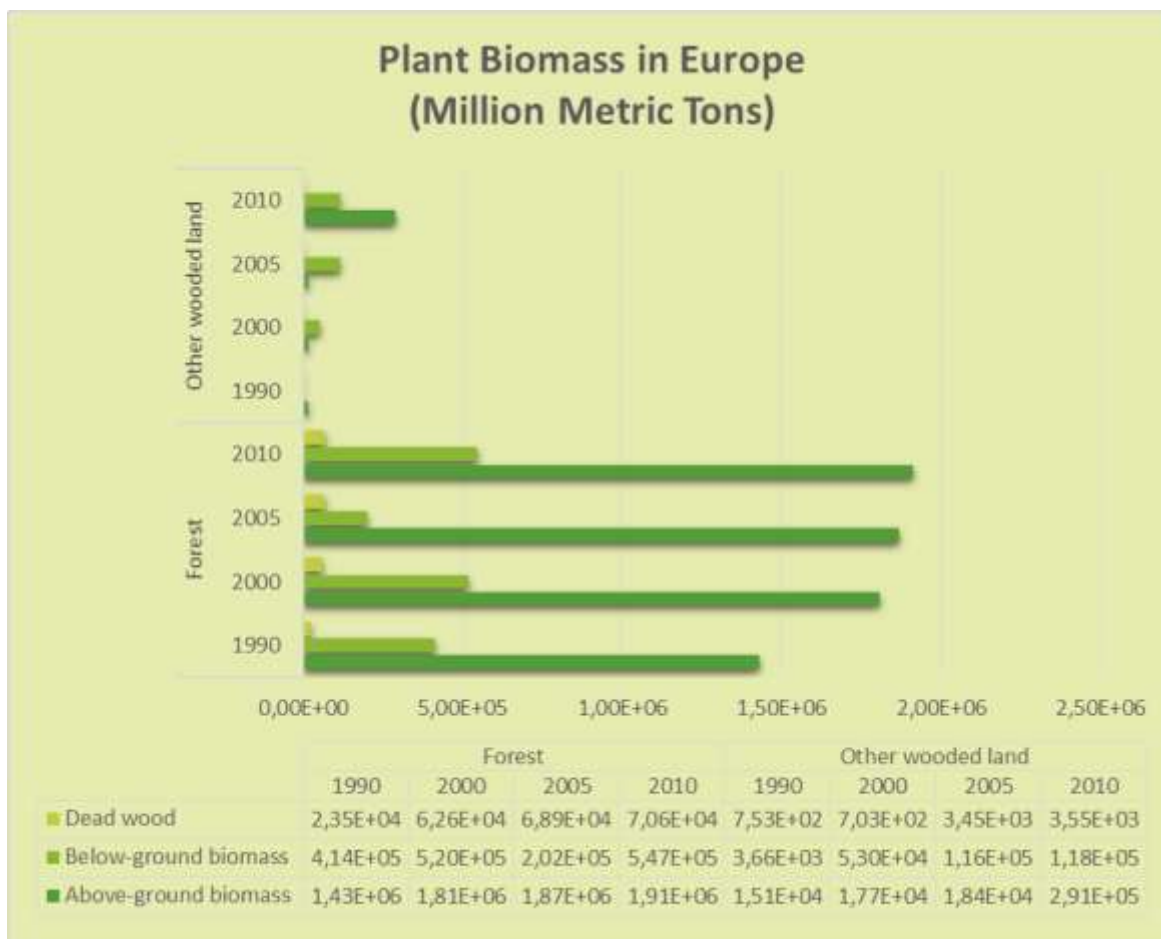


Figure 5 - Vegetal Biomass present in Europe from 1990 to 2010. Above-ground biomass: All living biomass above the soil including stem, stump, branches, bark, seeds, and foliage. Below-ground biomass: All biomass of live roots. Fine roots of less than 2 mm diameter are excluded because these often cannot be distinguished empirically from soil organic matter or litter. Dead wood: All non-living woody biomass not contained in the litter, either standing, lying on the ground, or in the soil. Dead wood includes wood lying on the surface, dead roots, and stumps larger than or equal to 10 cm in diameter or any other diameter used by the country (FAO, 2010).

According to both Figures 5 and 6 the majority of vegetal biomass available consists of living biomass and is being used worldwide for feeding purposes. Besides that, by consuming live biomass we are interfering with the carbon cycle so it's advisable to focus only on dead or waste biomass. Even with that in mind the numbers are surprisingly high with a value of 7.42×10^4 million metric tons.

Global land use for food production and renewable resources 2008

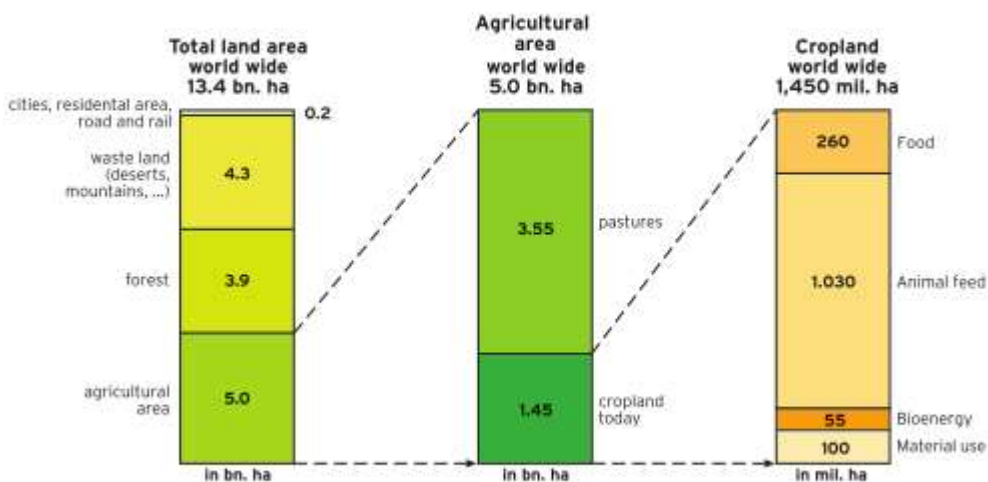


Figure 6 - Global usage of land for feed and food purposes in 2006/07. From a total of 13.4 bn. ha only 5.0 bn. ha are used for agricultural purposes from which 1.45 are used to crop purpose such as food, animal feed, material use and only 55 mil. ha to bioenergy (Raschka & Carus, n.d.).

1.1.2 - Applicability

Most of vegetable waste, like dead vegetation, scrap from paper or furniture industries or even agricultural residues, is not used to its full potential ending its life cycle at landfills in order to be burned. By leveraging this type of biomass as feedstock to other processes it is possible to reevaluate raw material that otherwise would be considered as garbage. Reusing these wastes would be cheaper for the different industries and therefore more appealing than other materials. Yet the variable composition of plant biomass would require some previous sorting in order to steer the correct substrate to the right application.



Figure 7 – Representative diagram of the contrast between the conventional industry and the novel biorefineries. Adapted from (Lucia, Argyropoulos, Adamopoulos, & Gaspar, 2006).

1.1.2.1 - Biorefineries

The new concept of biorefinery consists of reusing indigestible vegetal biomass to produce biofuels and other petroleum derived chemicals. In contrast to petroleum, with its limited resources and fixed composition, plant biomass is a self-renewable resource, recyclable and environmentally friendly that, due to its heterogeneous composition, can be broken into many usable components (Menon & Rao, 2012). The main goal in this industry is to extract from each biomass's component the largest benefit for producing both food and non-edible fractions, thus merging agro-food industries with biochemistry synthesis applications (Menon & Rao, 2012).

1.1.2.1.1 - Animal feeding

The usage of plant biomass as animal feeding is one of the options that requires some specific expertise as it deals with living beings. Therefore for this kind of application it would be advisable to use only agricultural residues or dead vegetation. As such, plant biomass could be directly given to ruminants since they are able to digest it due to their diverse and complex resident microbiome's driven enzymatic capacity (Singh *et al.*, 2014).

On the other hand, monogastric animals are not able to use these residues since they lack the endogenous enzymes capable of hydrolyzing the structural carbohydrates present in the plant cell wall.

In order to fully utilize waste biomass as supplements to animal feeding formulation, namely monogastric species, it is essential to hydrolyze the complex carbohydrate bonds. Nowadays the approaches used to break these chemical connections resort to either very harsh chemical and physical treatments while the alternative has proven difficult, due to the limitations on current biological methods (Mateos & González, 2007).

1.1.2.1.2 - Chemical Conversion

Plant derived biomass consist in various different carbohydrates linked together, such as cellulose, hemicellulose and pectins. If we break these complex sugars into basic monosaccharides it is possible to create several different compounds (Menon & Rao, 2012). There are two alternative routes to transform carbohydrates into simpler products: fermentation processes and chemical transformation (Corma, Iborra, & Velty, 2007); (Demirbas, 2001).

For instance, through glucose fermentation, obtained from breaking down cellulose into glucose, we have an array of starting components like lactic acid, succinic acid or glutamic acid that can be modified afterwards to generate additional products (Figure 8).

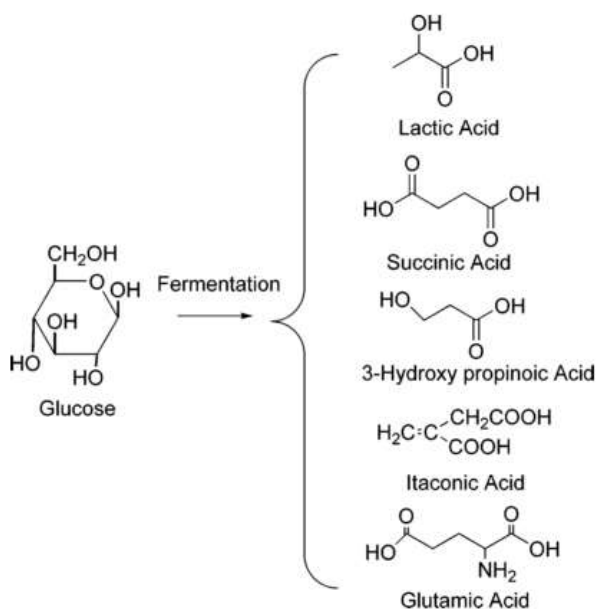


Figure 8 - Products obtained by the fermentation of glucose (Corma *et al.*, 2007).

By transforming these four products through a series of chemical reactions it is possible to obtain other, such as bioplastics that otherwise would be extremely difficult to produce and depended on petrochemical based substrates (Lucia *et al.*, 2006)

For centuries mankind has been using polymers from nature as food or to produce furniture and clothing. Despite public perception that bioplastics are a recent concept emerging after environmental global concerns arose, the first artificial plastic, celluloid, was invented in 1860. In 1940 the first ethylene based compounds were developed, but many other discoveries on these compounds were never commercialized due to the ascension of synthetic polymers made from crude in 1950 (Shen, Worrell, & Patel, 2010).

Until recently the petro-based plastics industry experienced an exponential growth, where almost everything that we use is made of or contains some type of plastic. Increasing concerns about fossil fuels scarcity, associated with climate changes and the desire to reduce human impact on earth, caused major changes in conventional industries and the rebirth of bio-based plastics. From all bioplastics presented on the Table 3, the most relevant types are Polylactide (PLA) and Polyethylene (PE) based (Shen *et al.*, 2010).

Table 3 – Main emerging biobased plastics for non-food applications (Shen *et al.*, 2010).

No	Biobased plastics (group)	Type of polymer	Types/Structure/Production Method
1.	Starch plastics	Polysaccharides	Partially fermented starch; Thermoplastic starch (TPS); Chemically modified starch (e.g., starch acetate) Starch blends; Starch composites
2.	Cellulose polymers	Polysaccharides	Organic cellulose esters; Regenerated cellulose
3.	Poly(lactide) (PLA)	Polyester	Biobased monomer (lactide) by fermentation, followed by polymerisation
4.	Poly(trimethylene terephthalate) (PTT)	Polyester	Biobased 1,3-propanediol (1,3-PDO) by fermentation plus petrochemical terephthalic acid (or DMT)
5.	Polyamides (PA)	Polyamide	Biobased monomer 11-aminoundecanoic acid from castor oil
a. PA11			Monomer sebacic acid from castor oil
b. PA610			Biobased monomer caprolactam by fermentation of sugar
c. PA6			Biobased adipic acid by fermentation
d. PA66			Biobased monomer obtained from oleic acid via azelaic (di)acid
e. PA69			
6.	Poly(hydroxyalkanoates) (PHAs)	Polyester	Direct production of PHA by fermentation
7.	Polyethylene (PE)	Polyolefin	Biobased monomer (ethylene) obtained from ethanol which is produced by fermentation of sugar.
8.	Poly(vinyl chloride) (PVC)	Polyvinyl	Monomer vinyl chloride can be obtained from biobased ethylene (from ethanol).
9.	Other thermoplastics*		
a.	Other polyesters (PBT, PBS, PBSL, PBPA, PBST, PBAT, PET, PEIT PVAc, Polyacrylates, PTN, PTL, thermoplastic elastomers)	Polyester	Various carboxylic acids, various alcohols
b.	Other ethylene-based compounds (e.g., polystyrene and EPDM rubber)	Various	Ethylene by dehydration of bio-ethanol, reacted with other compounds
c.	Methanol-based compounds (e.g., phenolic resins, urea formaldehyde resins, melamine formaldehyde resins)	Various	Syngas by gasification of biomass, and synthesis of methanol, reacted with other compounds
d.	Propylene-based compounds (e.g., PP, polyacrylates, PUR, PA)	Various	Thermochemical propylene production via bio-naphtha plus steam-cracking or via biomethanol, followed by a methanol-to-olefins process and polymerisation.
10.	Polyurethanes (PUR)	Polyurethanes	Biobased polyol from vegetable oils, plus petrochemical isocyanate.
11.	Thermosets	Cross-linked polymers	
a.	Epoxy resins	Epoxy resin	Diglycidyl ether of bisphenol A derived from bisphenol A and epichlorohydrin (ECH). ECH produced by glycerine-to-epichlorohydrin (GTE) process; glycerine is a byproduct of bio-diesel production.

The first bio-based plastic to be produced on a large scale was PLA, which consists of an aliphatic polyester originated through the fermentation of sugars, usually from sugar

cane, potato starch or tapioca starch. The lactic fermentation can produce different percentages of both lactic acid isomers, L or D, which can be adapted by using a specific *Lactobacillus* strain. Currently PLA is commercialized as a mix of 95% L-lactide and 5% D-lactide and is widely used in packaging, textiles, diapers, agricultural mulch films and cutlery (Shen *et al.*, 2010).

Bio-based polyethylene is alike the petro-based PE, where the only difference between them is the origin of the ethanol used to produce it. Nowadays bio-based PE is only produced from dehydrating ethanol derived from sugar-cane fermentation at temperatures ranging from 300°C to 600°C. This type of PE as numerous applications as presented in table 4 (Shen *et al.*, 2010).

Table 4 – Applicability of PE derivatives in Western Europe (* Excluding injection molded HDPE caps, closures and petro tanks) (Shen *et al.*, 2010).

Market Segment	HDPE	LDPE	LLDPE
Films	18%	74%	82%
Blow molding small parts	19%	1%	5%
Blow molding large parts	12%		
Pipes and extruded products	19%	4%	3%
Extruded coating		11%	1%
Caps and closures	4%		
Petro Tanks	3%		
Injection molded parts	14% *	4%	5%
Cable		4%	3%
Textiles	3%		
Other	8%	2%	1%
Total: 12.7 (Mt)	5.2 (Mt)	4.3 (Mt)	3.1 (Mt)

1.1.2.1.3 - Bioenergy

The exploitation of plant biomass as a resource to produce energy is one of the most common applications. Global efforts to lead nations into environmentally friendly lifestyles and policies due to increasingly depletion of fossil fuels caused the proliferation of alternative sources to sustain our energy needs.

In the last 15-20 years the European Community has been developing policies and studies in order to implement a global use of bioenergy. The last report from 2010 calculated the energy potential of vegetable biomass as shown on Figure 9 (Bottcher, 2010).

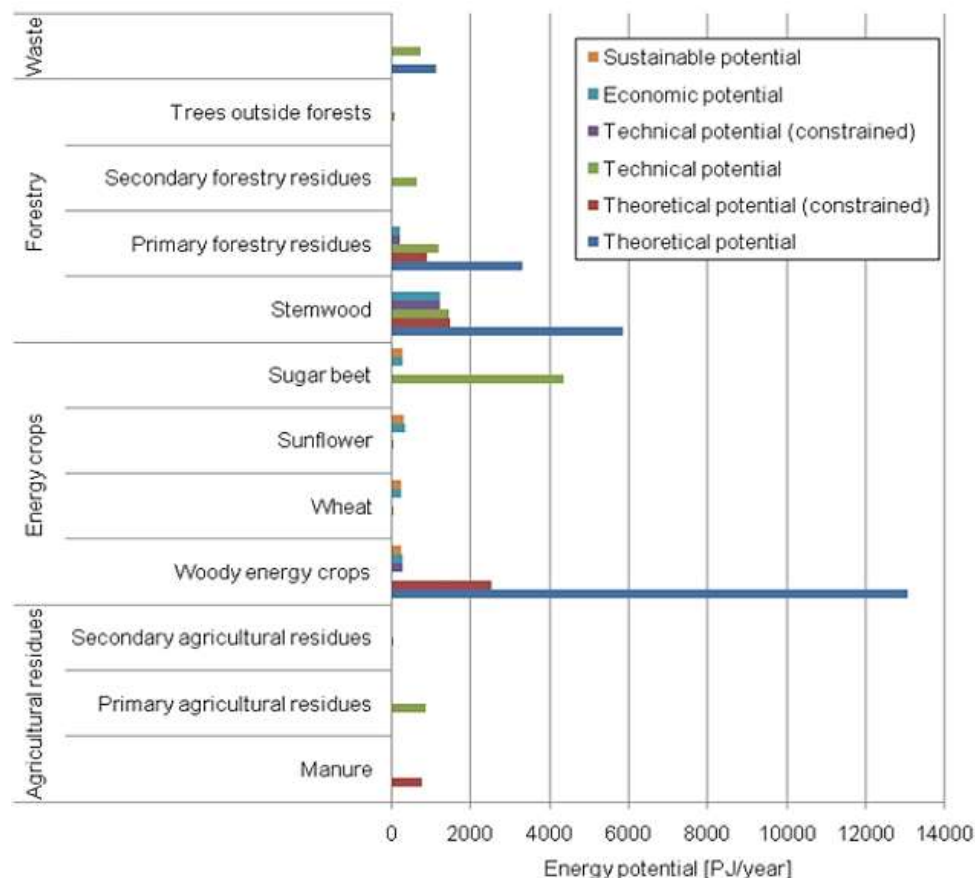


Figure 9 - EU energy potential from vegetable biomass. Predictions for all EU27 countries for the different biomass types and sources, according to potential types. Values in PJ per year. For the averages from forestry biomass sources, the conversion of tons DM wood to PJ were harmonized to a value of 15.48 (Bottcher, 2010).

Using only the biomass as feedstock, several types of energy like heat, electricity and fuels can be produced. In order to convert biomass into one of these options several technologies can be applied. These methods can be divided in three main categories: Thermochemical, Biochemical and Extraction as shown in Figure 10 (Faaij, 2006)

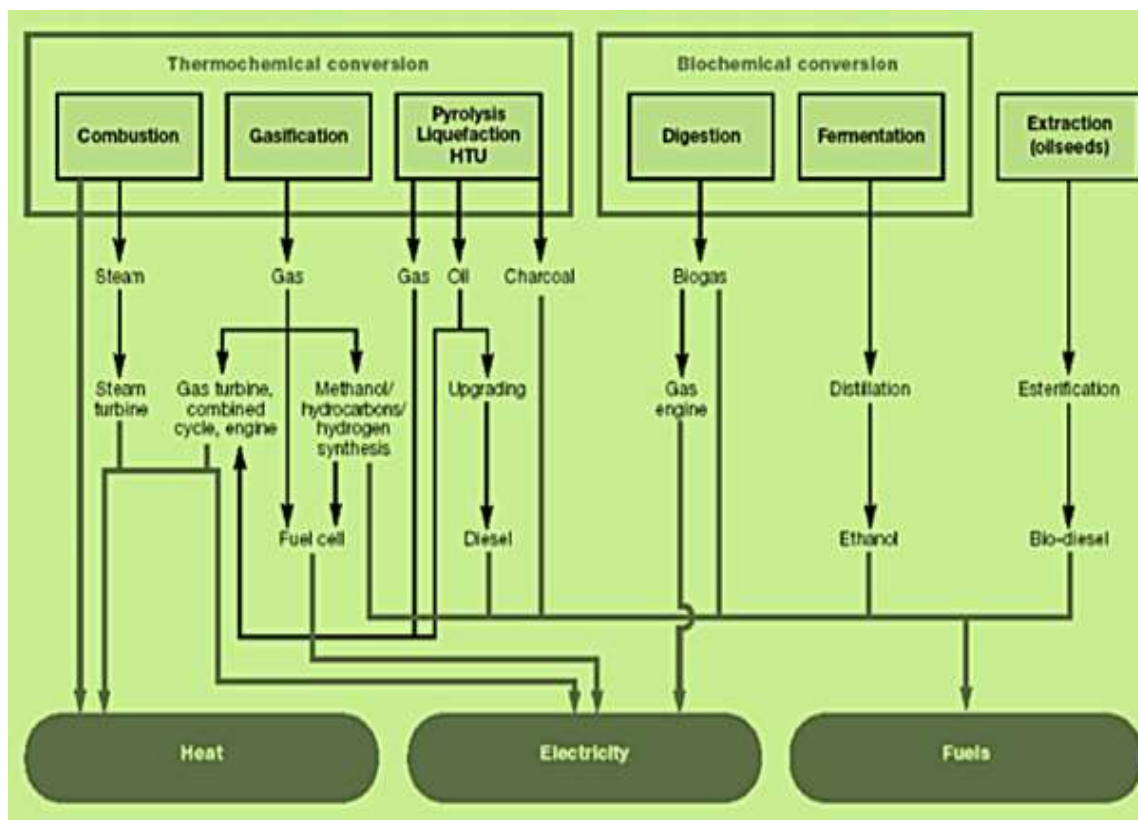


Figure 10 – Principal options for biomass conversion to secondary energy carriers (Faaij, 2006)

The thermochemical conversion uses high temperatures in order to promote the chemical reaction that generates energy (Faaij, 2006). Despite the benefits of this first method, the requirement of high temperatures increases the cost associated with this type of conversion and decreases its sustainability. Therefore its applicability would be more suitable for domestic use, like in heat production through convectional fireplaces and pellet's boilers.

Through digestion or fermentation of the biomass, new compounds can be produced and used as fuels to produce electricity, in a process called biochemical conversion (Faaij, 2006). By improving the technology of biochemical conversion of biomass, this method is envisioned as the most suitable alternative to fossil fuels.

Finally the extraction method consists of extracting oil from the biomass, which undergoes esterification to produce the so called biodiesel. This process is probably the most well established in Europe (Faaij, 2006). Nonetheless it requires crops usage, eventually implying the dislodgment of lands from the food industry, which is one of the main hurdles for the bio-diesel industries.

In order for plant biomass to become a viable feedstock for meeting future demand for liquid fuels, efficient and cost-effective processes must exist to breakdown cellulosic materials into their primary components (Elkins, Raman, & Keller, 2010). Therefore it is essential to understand specifically plant biomass composition, specifically its most recalcitrant component; the plant cell wall.

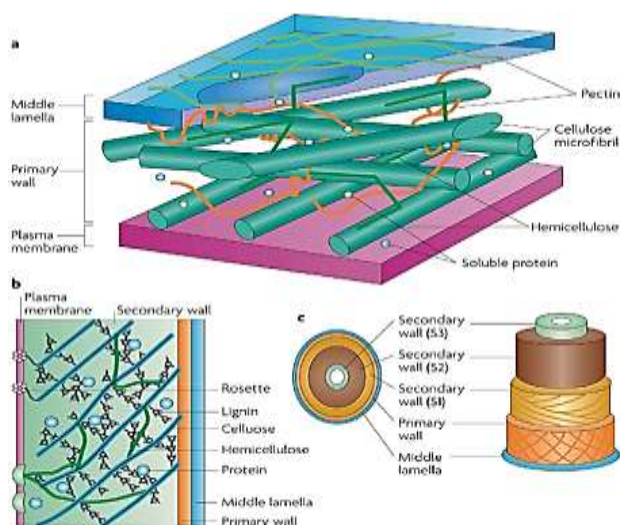
1.2 - The Plant Cell-Wall

The main difference between plants and other living organism lies, precisely, in their cell wall. The plant cell wall is a rigid structure that is responsible for the cell form, its growth and stability, osmotic balance as well as constituting a barrier against pathogenic agents such as virus, bacteria and fungus (Maza, 2012).

Plant's cell wall adapts itself according to the development state of the plant. During the initial growing and division, the cell wall becomes more elastic enabling the elongation and form definition of the cell. In contrast, fully differentiated cells have a more rigid and thicker wall providing mechanical strength to prevent biochemical degradation or even physical damage (Knudsen, 1997; Maza, 2012).

1.2.1 - Structure

Most cellular walls from differentiated vegetal cells are composed of four layers: an external coat called middle lamella and three inner layers, the primary, secondary and, sometimes, even a tertiary wall (Maza, 2012; Sticklen, 2008a).



Nature Reviews | Genetics

Figure 11 - Plant plasma membrane and cell-wall structure. Most plant's cell wall is composed by a plasma membrane, a primary and secondary wall and a middle lamella. Their main constituents vary according to its function (Sticklen, 2008b).

1.2.1.1 - Middle lamella

The middle lamella is basically a thinner cover of pectin's and lipids that act as binders between the other cells as well as protectors from various external aggressions (Maza, 2012).

1.2.1.2 - Primary wall

The primary wall consists of a two segment's system, where microfibrils of cellulose are embedded on protein and non-cellulosic polysaccharide matrix. The long cellulose chains are organized side by side creating micelles. This almost crystalline structure is maintained by hydrogen bonds between the numerous hydroxyl groups (-OH) available. The association of various micelles originates the cellulose's microfibrils, which can reach several micrometers of length and between 3 and 10µm of diameter (Maza, 2012).

Regarding to the polymeric matrix there are several theories. Initially Keegstra *et al.* (1973) suggested that the various polysaccharides and proteins, such as xylan, xyloglucan, pectins and structural proteins, formed a macromolecular net through covalent bonds amongst each other, while the cellulose microfibrils were connected by hydrogen-bonds. In 1989, Hayashi and Fry proposed the most popular model. In this model a single xyloglucan chain occupied the gaps between cellulose microfibrils, tethering them together, while the pectins and structural proteins filled the remaining space. Further models were later proposed (Talbot & Ray, 1992) or (Ha, Apperley, & Jarvis, 1997), sharing in common the coating of cellulosic fibrils with xyloglucan. In summary, the primary wall structure consists of a complex network of cellulose microfibrils superficially attached by hemicelluloses through non-covalent interactions, all embedded on a gel like phase comprising pectins (Cosgrove, 2005; Maza, 2012; Somerville *et al.*, 2004). This dynamic network is responsible for the majority of the physical properties displayed by plant cell walls.

1.2.1.3 - Secondary wall

In contrast with the primary wall, secondary walls differ depending on cell type or development stage and are normally present to increase the mechanical support. The

successive deposition and fouling of cellulose fibrils in addition to other compounds, like lignin, is responsible for the establishment of a secondary wall (Knudsen, 1997; Maza, 2012).

1.2.2 - Main Constituents

The plant cell wall is mainly composed by polysaccharides (90%), glycoproteins (2-10%), phenolic esters (< 2%) and ionic and covalently bound minerals (1-5%) (Rose, 2003). In accordance to our project goals we are going to focus only in the polysaccharides which, according to Wallace & Somerville (2014), can be divided in five distinctive groups: cellulose, hemicelluloses, pectins, proteins and lignin.

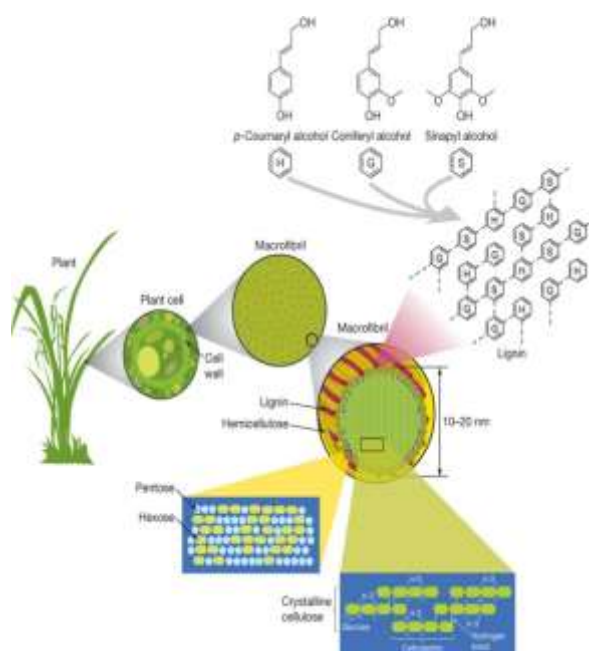


Figure 12 - Main cell wall constituents. Cellulose, hemicellulose and lignin form structures called microfibrils, which are organized into macrofibrils that mediate structural stability in the plant cell wall. Cellulose is a $\beta(1-4)$ -linked chain of glucose molecules. Hemicellulose, the second most abundant component of lignocellulose, is composed of various 5- and 6-carbon sugars such as arabinose, galactose, glucose, mannose and xylose. Lastly, lignin is composed of three major phenolic components, namely p-coumaryl alcohol (H), coniferyl alcohol (G) and sinapyl alcohol (S). Lignin is synthesized by polymerization of these components and their ratio within the polymer varies between different plants, wood tissues and cell wall layers (Rubin, 2008).

1.2.2.1 - Cellulose

Cellulose is a linear polymer with several thousand glucose residues linked together through a glycosidic bound β -1,4-D-anhydroglucopyranose ($C_{6n}H_{10n+2}O_{5n+1}$, n = glucose's polymerization's degree). This specific connection along with a 180° rotation of each glucose unit, within the chain, relatively to the previous one is responsible for a flat assembly of this polysaccharide. The structure stability is acquired due to hydrogen bonds between parallel glucans creating a crystalline microfibril that provides the mechanical strength and degradation resistance (Cosgrove, 2005; Maza, 2012; Thygesen, Oddershede, Lilholt, Thomsen, & Ståhl, 2005).

Cellulose, available commonly in global flora, normally has a polymerization's degree of roughly 15 000 glucopyranose in cellulose from native cotton and 10 000 in wood cellulose. There are several polymorphic forms of cellulose that can be interconverted into each other. Their nomenclature ranges from I to IV, in which cellulose I is the native form (Sullivan, 1997). The natural form of cellulose adopts a rather unstable crystalline structure called cellulose I, which includes the forms $I\alpha$ and $I\beta$ where both chains assemble in a parallel mode (Jamal *et al.* 2004). The concept of cellulose $I\alpha$ and $I\beta$ was reported by Simon *et al* (1988) which described that the cellulose on the surface of the crystal differed from its structure on the center. After various studies comparing both forms it was establish that they had the same backbone structure, differing only in their hydrogen bonding patterns. The meta-stability of cellulose $I\alpha$ confers it a higher reactivity rate than $I\beta$, therefore their relative proportion in each sample may affect the overall reactivity (Sullivan, 1997). The proportion of these two forms apparently alternate according to the type of plant or organism that produces it. For instance, higher plants have predominantly $I\beta$ form, whereas cellulose originated from primeval organisms has mostly $I\alpha$ (Jamal, Nurizzo, Boraston, & Davies, 2004; Thygesen *et al.*, 2005).

Besides this crystalline structure, natural occurring cellulose seems to comprehend other less organized and loosely portions, named “amorphous” cellulose. Despite the overall knowledge of the different levels of organization of the cellulose fibers present on the plant cell wall, the proportions and definite structure is still unclear (Gourlay, 2014). As shown in Figure 13, in general, cellulose is crystalline when molecules are tightly packed and is

amorphous when they are less densely packed. The crystalline areas are more insoluble and inaccessible to enzymatic attack than the amorphous areas, making the hydrolysis of these regions more complex and difficult (Warren, 1996). Most of the “amorphous phase” of cellulose corresponds to chains that are located at the microfibril surface, whereas crystalline components occupy the core (E. A. Bayer, Chanzy, Lamed, & Shoham, 1998). To sum up it's important to refer that despite its terminology amorphous cellulose still has some level of organization (Sullivan, 1997), that its ratio and order in the main structure differ according to the phylogeny and it is reported to be easier to degrade by enzymes (Gourlay, 2014).

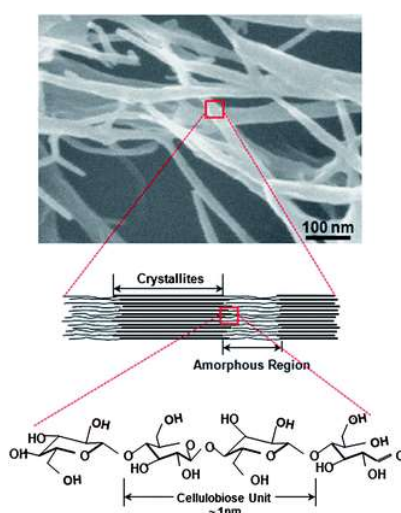


Figure 13 – Organization of Cellulose. Cellulose is crystalline when molecules are tightly packed whereas it is amorphous when they are less densely packed (Chun *et al.*, 2012).

1.2.2.2 - Hemicellulose

Following cellulose, hemicellulose is the second most abundant component in lignocellulosic materials, comprehending a diverse group of highly ramified polysaccharides. Unlike cellulose, hemicelluloses are not chemically consistent, establishing heterogeneous polymers of pentoses such as xylose and arabinose, hexoses like mannose, glucose, galactose, and also sugar acids. The type of polymers depends on the hemicellulose's origin, since normally hardwood hemicellulose contains more xylans and softwood hemicelluloses are mainly formed by glucomannans (Saha, 2003; Rose, 2003).

1.2.2.2.1 - Xyloglucan

In land plants, xyloglucan is the most predominant cross-linking hemicellulosic polymer. This polysaccharide is composed of a semi rigid β -1,4-glucan backbone commonly decked with α -1-6-xylosyl residues on the side chains. The side chain residues are often replaced at O6 by diverse alternative glycosyl residues according to the phylogeny and cell type (Fry, 1989; Wallace & Somerville, 2014); rose, 2003).

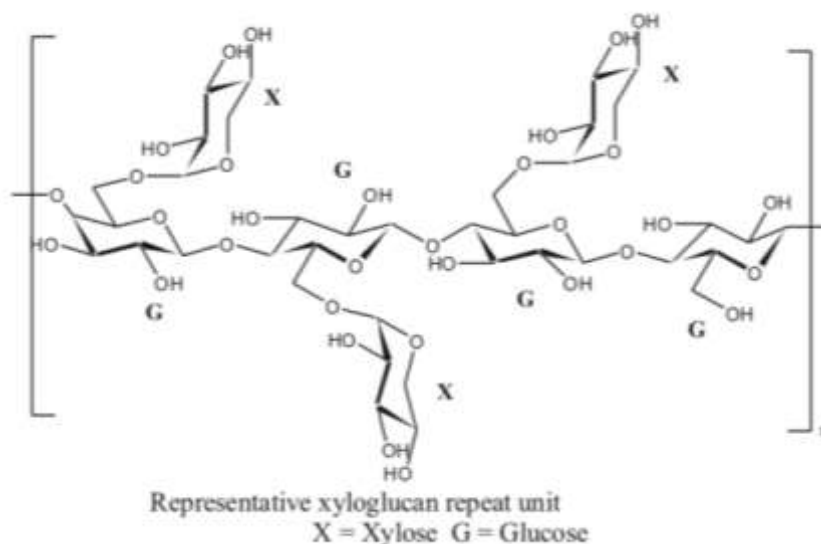


Figure 14 – Xyloglucan structure (Bhalekar, Sonawane, & Shimpi, 2013).

1.2.2.2.2 - β -Glucan

The β -glucans are included in the glucan category which consists in glucose polymers linked by either an α or β bond. In β -glucans the D-glucose is linked by a β -glycosidic bond (El Khoury, Cuda, Luhovyy, & Anderson, 2012). The uniqueness of this molecule resides on this linkage alternate conformation, causing twists in the main chain and reducing intermolecular associations (Edney 1991).

This hemicellulosic polysaccharide is mostly present in cereals such as barley, oats, rye and wheat. The overall structure of this polysaccharide varies according to the source. One of the simplest structure of β -glucans is found on the plant cell wall of some cereal grains such as barley and its composed by linear β -1,3; 1,4-D-glucans (Figure 15). In contrast,

glucans extracted from bacteria, yeast and fungi and some algae tend to present a branched structure consisting of β -(1,3)- or β -(1,4)-glucan backbone with either (1,2)- or (1,6)-linked β -glucopyranosyl side branches (El Khoury *et al.*, 2012). This polymer has raised further interest when compared with other cellulosic polysaccharides due to its nutritional and medical value (Brown & Gordon, 2001).

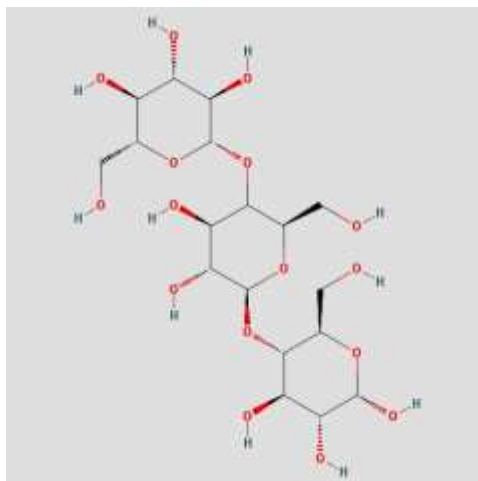


Figure 15 – β -Glucan Structure. National Center for Biotechnology Information. PubChem Compound Database; CID=46173706, <http://pubchem.ncbi.nlm.nih.gov/compound/46173706> (accessed Apr. 22, 2015).

1.2.2.2.3 - Mannose derivatives – Mannan, glucomannans and galactomannan

The polysaccharides derived from mannose are a significant constituent of plant biomass and also very useful as thickeners and stabilizers in the food and pharmaceutical industries.

Mannan comprises a backbone of β -1,4-linked mannose residues, whereas the other mannose-containing polymers have a heterogeneous linkage of mannose and other sugars. For instance glucomannans are randomly distributed mannose and glucose β -1,4 bonds, whereas galactomannans have the mannan backbone decorated with an α -1,6-linked galactosyl residue (Brett, Waldren, & Brown, 1996).

1.2.2.3 - Pectin

Pectins are highly conserved polysaccharides amongst land plants, acting as the foremost filler between cellulose and hemicelluloses. This cell wall component is composed by a backbone structure of galacturonic acid α -D-GalA linked at mutually 1 and 4 positions.

Pectic substances can be divided in three main categories, as shown in Figure 16: homogalacturonan (HG), rhamnogalacturonan I (RG-I) and rhamnogalacturonan II (RG-II). HG consists in α -D-GalA single chain that can suffer some modifications like O-acetylation at O-2 or O-3 and substitution of the O-3 with β -D-Xyl, depending on the type of plant of origin. The difference between HG and the RG-II lies on the addition of alternative polysaccharide side chains mainly with galacturonic acid and rhamnose. In contrast, RG-I has a single chain backbone with repetitions of 4- α -D-GalA-1,2- α -L-Rha-1 in which the GalA residues are extremely acetylated at O-2 or O-3 (Atmodjo, Hao, & Mohnen, 2013).

1.3 - Breaking the wall

One of the greatest hurdles of the widespread usage of plant biomass as feedstock to produce bioenergy is related to the difficulties associated with the pre-treatments required to degrade the plant cell wall. In order to be profitable, besides large scale availability and having a low purchase price, the technology used to convert biomass to energy must be effective, efficient and also low cost, all of which are not yet available.

1.3.1 - Conventional treatments

Pre-treatments consist in the preparation of the lignocellulosic biomass through solubilization or separation of its main components. The proper pre-treatment must be adjusted to the feedstock, enzymes and organism used in all stages of the process. Although efforts have been made in the past decades to reduce the cost associated with lignocellulosic pre-treatments, most methods implemented weren't fully satisfactory (Menon & Rao, 2012).

The effectiveness of these treatments reside on the balance of operational costs, the range of lignocellulosic materials that can be used, the harnessing of most feedstock and the accurate separation of each component (Agbor, Cicek, Sparling, Berlin, & Levin, 2011).

Although it is extremely difficult to evaluate and compare methods, the perfect system would produce a disrupted carbohydrate substrate that could easily be hydrolyzed, avoiding the production of other by-products that could compromise the fermentation, like inhibitors, or that might degrade sugars. Based on these ideal standards, pretreatments are rated according to the severity factor which is determined by the combined effect of three main factors: the temperature, acidity and duration (Agbor *et al.*, 2011).

Most of the pretreatments available nowadays involve either physical, chemical and biologic treatments or a combination of these techniques (Menon & Rao, 2012).

1.3.1.1 - Physical Treatments

Physical pretreatments encompass the mechanical process that literally breaks the biomass into smaller sizes. This may be obtained through several methods such as milling, irradiation or extrusion (Agbor *et al.*, 2011; Menon & Rao, 2012; Yang & Wyman, 2008).

Most of these methods require the use of more energy than the energetic potential of the biomass being treated, therefore making this option highly inadequate to its industrial implementation (Menon & Rao, 2012; Yang & Wyman, 2008). In order to make this method profitable, it must be combined with chemical treatments thus increasing the efficacy of the process by decreasing the associated costs (Agbor *et al.*, 2011).

1.3.1.2 - Chemical Treatments

The effect of several chemicals on loosening the structure of lignocellulosic biomass have been reported in diverse studies, being one of the most applied methods and according to some authors the most effective. The chemical treatments are normally divided according to the used solvents and its pH (Agbor *et al.*, 2011).

On the higher pH range we have the alkaline solvents such as NaOH, KOH and Ca(OH)_2 . These reagents increase the surface area while decreasing the degree of polymerization (dp) and crystallinity of the biomass therefore making the carbohydrates more accessible (Agbor *et al.*, 2011). This is generally achieved due to the alkali ability to break the ester and glycosidic bonds in the side chains. The conditions required for the alkaline treatment are usually less aggressive since it's possible to reduce the operation temperatures by increasing the process time. Unfortunately this method involves additional steps of neutralization to remove lignin and inhibitors (Menon & Rao, 2012).

In order to use acids, such as sulfuric, hydrochloric and phosphoric acids, to hydrolyze biomass it's necessary to dilute them due to their corrosive properties (Agbor *et al.*, 2011). The acid hydrolysis of the sugars normally present high product yields in a short amount of time but requires high temperatures which in association to its harshness decreases the overall appeal of this method (Menon & Rao, 2012). This technique also requires additional neutralization and in order to be profitable the solvent must be recovered and reused (Agbor *et al.*, 2011).

The use of green solvents has been one of the most promising alternatives to the regular solvents. The so called ionic liquids (IL) have been increasingly gaining interest in various fields and the degradation of vegetable biomass it is not an exception (Yang & Wyman, 2008). These remarkable new solvents can be adapted according to the requirements of the process in consideration, but in general they consist entirely of ionic species. These salts are composed mainly by one or more large ions and a low symmetry degree cation. The ionic liquids may be allocated to two main categories, the simple salts and the binary ionic liquids. By exposing cellulosic material to IL the samples become essentially amorphous and porous increasing the efficacy of degradation by enzymes. The advantage of this technique resides on the fact that the energy and skills requirements are less demanding as well as its friendliness to environment. However, due to being a recent technology the costs associated with the solvent itself and the all the optimization that still need to be done in order to make it profitable, its commercial application is still limited (Menon & Rao, 2012).

1.3.2 - Biological Treatments

One of the most current strategies in science is to mimic nature and the plant cell wall degradation is an excellent example of how useful environmental strategies can be used to solve our problems.

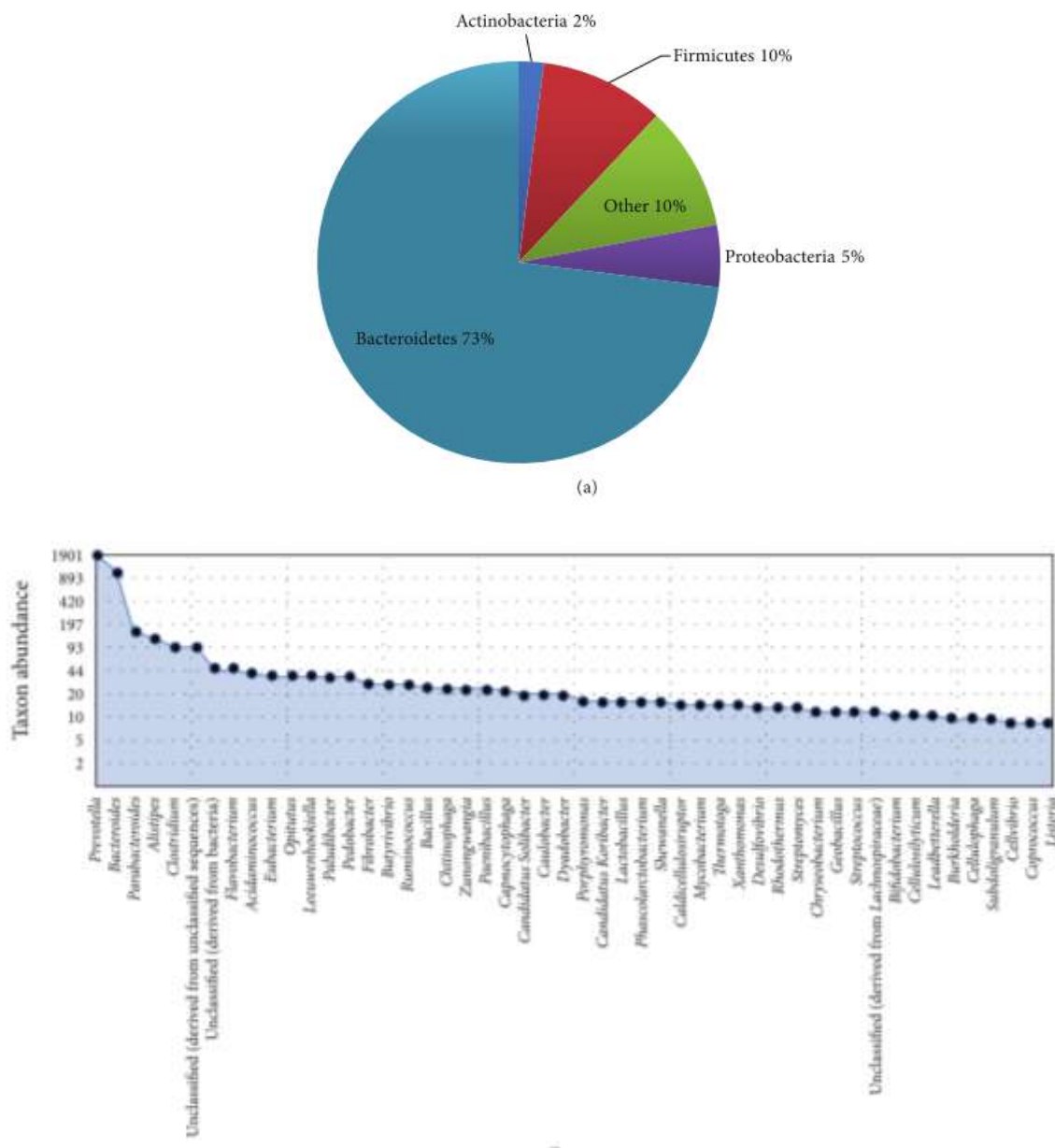


Figure 17- Classification of microbial taxa bearing putative CAZymes. (a) The pie chart shows the abundance of phylum and (b) genus abundances ordered from the most abundant to least abundant (Singh *et al.*, 2014).

Several organisms, as shown in Figure 17, have been digesting lignocellulosic materials for ages, ruminants such as sheep and cows obtain 70% of their energy from breaking the cellulosic bonds and converting them to smaller carbohydrates (Dassa, Borovok, Ruimy-Israeli, & Lamed, 2014). Another common example is the action of termites in degrading

wood. Since we are not planning to use these animals directly in the industry, it is essential to fully understand their remarkable ability.

Termites normally work in association with fungi in a symbiotic relationship where the fungi digest the cellulose while termites then consume both of them. From previous studies the usage of fungi has proven to be not as efficient as desirable, since growth conditions and the necessary area must be carefully controlled and the overall process is slow, therefore reducing its industrial appeal (Menon & Rao, 2012).

Nevertheless, low energy and mild environmental requirements associated with environmentally friendliness make the biological approach a very promising technique, prompting the research of bio-alternatives to fungi. This search brings us back to the ruminants which have a fascinating crew of bacteria in their rumen helping them to degrade lignocellulosic biomass. Despite being well known that cellulosic bacteria play a major role in degrading plant fibers on the rumen, only a few species have been identified, namely *Ruminococcus* and *Fibrobacter*, on which we are going to focus our attention (Dassa *et al.*, 2014). Besides these species, there are other microorganisms that have been classified as plant degrading and that are resumed at Figure 18. These organisms' ability to degrade the plant cell wall comes from their extracellular consortium of enzymes such as cellulases and hemicellulases. These enzymes normally display an extremely complex architecture, merging catalytic modules to non-catalytic modules which can either be interactions between two proteins like the dockerin-cohesin modules or protein-carbohydrate such as CBMs-substrate interactions (Fontes & Gilbert, 2010).

Besides the common presence of an enzymatic complex, aerobic and anaerobic microorganisms present very distinct ways of approaching and organizing their plant cell wall degrading machinery. Aerobic microorganisms present a free enzyme system where enzymes are either secreted into the extracellular milieu or are located on the outer membrane. Although these enzymes do not physically associate, they do display extensive biochemical synergy (H J Gilbert, Stalbrand, & Brumer, 2008). In contrast, anaerobes such as *Clostridia* and rumen bacteria have a particularly interesting multienzyme assembly system termed cellulosome (Fontes & Gilbert, 2010).

Oxygen relationship	Genus	Representative species ^a	Gram reaction	Morphology	Growth temp. ^b	Resting state	Motility	Features of cellulase system
Aerobic	<i>Acidothermus</i>	<i>A. cellulolyticus</i>	+	Rod	Thermo			Noncomplexed, cell free
	<i>Bacillus</i>	<i>B. pumilis</i>	+	Rod	Meso	Endospore	Flagellar	Noncomplexed, cell free
	<i>Caldibacillus</i>	<i>C. cellovorans</i>	+	Rod	Thermo	Endospore		Noncomplexed, cell free
	<i>Cellulomonas</i> ^c	<i>C. flavigena</i> , <i>C. uda</i>	+	Rod	Thermo	None	Flagellar	Noncomplexed, cell free
	<i>Cellvibrio</i>	<i>C. fulvus</i> , <i>C. gilvus</i>	-	Curved rod	Meso	None	Flagellar	Noncomplexed, cell free
	<i>Cytophaga</i>	<i>C. hutchinsonii</i>	-	Rod	Meso	None	Gliding	Noncomplexed?, cell free?
	<i>Erwinia</i>	<i>C. carotovora</i>	-	Rod	Meso	None	Flagellar	Noncomplexed, cell free
	<i>Micromonospora</i>	<i>M. chalybeata</i>	+	Filamentous rod	Meso	Spore ^d	Nonmotile	Noncomplexed, cell free
	<i>Pseudomonas</i>	<i>P. fluorescens</i> var. <i>cellulosa</i>	-	Rod	Meso	None	Flagellar	Noncomplexed, cell free
	<i>Sporocytophaga</i>	<i>S. myxococcoides</i>	-	Rod	Meso	Spore ^d	Gliding	Noncomplexed, cell free
	<i>Rhodothermus</i>	<i>R. marinus</i>	-	Rod	Thermo			
	<i>Streptomyces</i>	<i>S. reticuli</i>	+	Filamentous rod	Meso	Spore ^d	Nonmotile	Noncomplexed, cell free
	<i>Thermobifida</i>	<i>T. fusca</i>	+	Filamentous rod	Thermo	Spore ^d	Nonmotile	Noncomplexed, cell free
Anaerobic	<i>Acetivibrio</i>	<i>D. cellulolyticus</i>	-	Curved rod	Meso	None	Nonmotile	Complexed
	<i>Anaerocellum</i>	<i>D. thermophilum</i>	+	Rod	Thermo	None	Flagellar	Noncomplexed, cell free
	<i>Butyrivibrio</i>	<i>B. fibrisolvens</i>	+	Curved rod	Meso	None	Flagellar	Noncomplexed
	<i>Caldicellulosiruptor</i>	<i>C. saccharolyticum</i>	-	Rod	Thermo	None	Flagellar	Noncomplexed, cell free
	<i>Clostridium</i>	<i>C. thermocellum</i> , <i>C. cellulolyticum</i>	+	Rod	Thermo, meso	Endospore	Flagellar	Complexed, mostly cell bound ^e
	<i>Eubacterium</i>	<i>E. cellulosolvens</i>	+	Rod	Meso	None	Nonmotile	
	<i>Fervidobacterium</i>	<i>F. islandicum</i>	-	Rod	Thermo	None	Flagellar	
	<i>Fibrobacter</i>	<i>F. succinogenes</i>	-	Rod	Meso	None	Nonmotile	Complexed, cell bound
	<i>Halocella</i>	<i>H. cellulolytica</i>	-	Rod	Meso	None	Flagellar	Noncomplexed, cell free
	<i>Ruminococcus</i>	<i>R. albus</i> , <i>R. flavefaciens</i>	+	Coccus	Meso	None	Nonmotile	Complexed, cell bound
	<i>Spirochaeta</i>	<i>S. thermophila</i>	+	Spiral	Thermo	None		Noncomplexed, cell free
	<i>Thermotoga</i>	<i>T. neapolitana</i>	-	Rod	Thermo			

^a Not all strains of the indicated species are cellulolytic, and some less active or less studied cellulolytic species within these genera are not listed.
^b Meso, mesophilic; Thermo, thermophilic.
^c Most strains can also grow anaerobically.
^d Unlike true endospores, these spores have only moderate resistance to environmental stress.
^e Except for *C. stercorarium* (81, 83, 606).

Figure 18 - Principal cellulolytic bacteria and their major morphological features. Summary Table from (Lynd, Weimer, Zyl, & Pretorius, 2002).

1.4 - The Cellulosome

The cellulosome was first described in 1983, more than thirty years ago, in the *Clostridium thermocellum*, an anaerobic thermophilic bacteria known for its cellulolytic abilities (Fontes & Gilbert, 2010; Smith & Bayer, 2013).

Table 5 – Cellulosome-producing microorganisms. Mes- Mesophilic (25°-40°C) Ther- Thermophilic (> 40°C) (Doi & Kosugi, 2004).

	Microorganism	Growth Temperatures	Source
Anaerobic Bacteria	<i>Acetivibrio cellulolyticus</i>	Mes	Sewage
	<i>Bacteroides cellulosolvens</i>	Mes	Sewage
	<i>Butyivibrio fibrisolvens</i>	Mes	Rumen
	<i>Clostridium acetobutylicum</i>	Mes	Soil
	<i>Clostridium cellobioparum</i>	Mes	Rumen
	<i>Clostridium cellulolyticum</i>	Mes	Compost
	<i>Clostridium cellulovorans</i>	Mes	Wood fermenter
	<i>Clostridium josui</i>	Mes	Compost
	<i>Clostridium papyrosolvens</i>	Mes	Paper mill
	<i>Clostridium thermocellum</i>	Ther	Sewage soil
	<i>Ruminococcus albus</i>	Mes	Rumen
	<i>Ruminococcus flavefaciens</i>	Mes	Rumen
Anaerobic Fungi	<i>Neocallimastix patriciarum</i>	Mes	Rumen
	<i>Orpinomyces joyonii</i>	Mes	Rumen
	<i>Orpinomyces PC-2</i>	Mes	Rumen
	<i>Piromyces equi</i>	Mes	Rumen
	<i>Piromyces E2</i>	Mes	Faeces

Since then, several studies and reviews have been conducted in order to expand the knowledge in this field and to enhance the viability of exploiting cellulosomes to degrade plant biomass (Doi & Kosugi, 2004; Smith & Bayer, 2013). The efforts made enabled the identification of several other microorganisms able to produce cellulosomes, as seen in table 5.

Generally the cellulosome is composed of two main components; the non-enzymatic scaffolding proteins called scaffoldins and a variety of attached CAZymes. In a pivotal protein-protein cellulosome assembly interface, termed type-I Coh-Doc interaction, the cohesin domains present on the scaffoldins bind to the dockerin domains existent on the enzymes (Figure. 19) (Doi & Kosugi, 2004; Fontes & Gilbert, 2010).

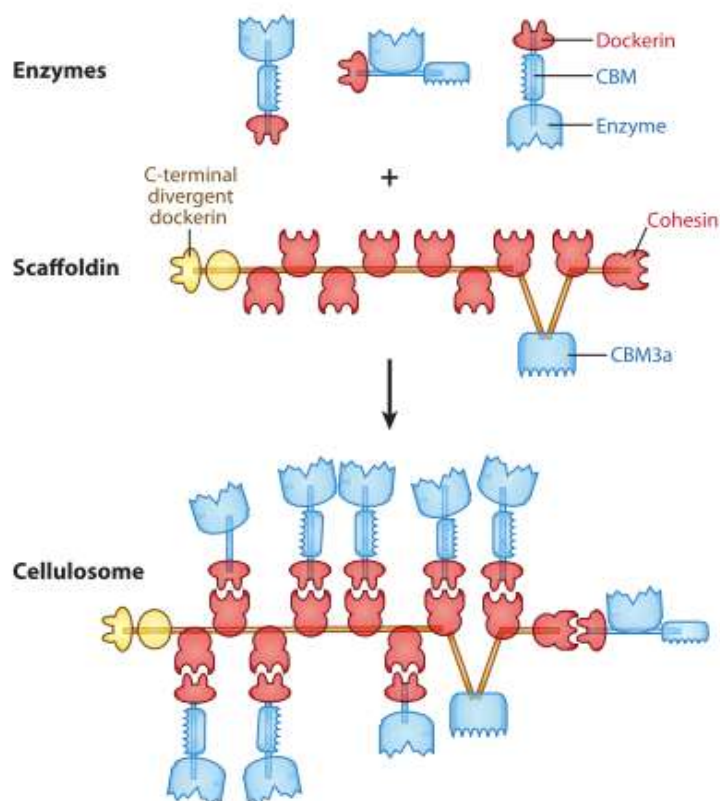


Figure 19-Cellulosome assembly mechanism. The catalytic modules (enzymes) are generally appended to a dockerin module which is coupled to a non-catalytic module composed of CBMs. The dockerins bind to the cohesins (red) of a non-catalytic scaffoldin, providing a mechanism for cellulosome assembly. Additionally the scaffoldin also contains a cellulose-specific family 3 CBM (CBM3a) and a C-terminal divergent dockerin that targets the cellulosome to the plant cell wall and the bacterial cell envelope, respectively (Fontes & Gilbert, 2010).

In more complex systems, such as those present in *C. thermocellum*, *Acetivibrio cellulolyticus* and *Bacteroides cellulosolvens*, besides the primary scaffoldin there are multiple anchoring scaffoldins, thus enabling the linkage of the cellulosome to the cell surface. This anchoring is due to C-terminal S-layer homology (SLH) modules that mediate the scaffoldin cell-wall attachment, while these anchoring scaffoldins also possess type-II cohesins involved in the so called type-II Coh-Doc interactions with the primary scaffoldin-borne C-terminal divergent type-II dockerin (Fontes & Gilbert, 2010). The simplest cellulosome system consists of a single scaffold protein with a few cohesins and a carbohydrate binding module (Doi, 2008). This assembly mechanism changes according to the organism from which it is isolated. The simplest cellulosomes are characteristic from *Clostridia* and present only type-I cohesins, lacking type-II dockerins, and therefore are not anchored to the cell surface (Doi & Kosugi, 2004; Fontes & Gilbert, 2010).

In *B. cellulosolvens* there are two scaffoldins, a primary scaffoldin termed ScaA that includes 11 type-II cohesins with a C-terminal type-I dockerin, and an anchoring scaffoldin named ScaB containing 10 type-I cohesins. This overall structure can comprise a total of 110 dockerin-enzymes and reveals an apparent Coh-Doc role reversal, as type-II interactions mediate enzyme assembly while type-I interactions mediate cell surface binding (Xu *et al.*, 2004) (Figure 20).

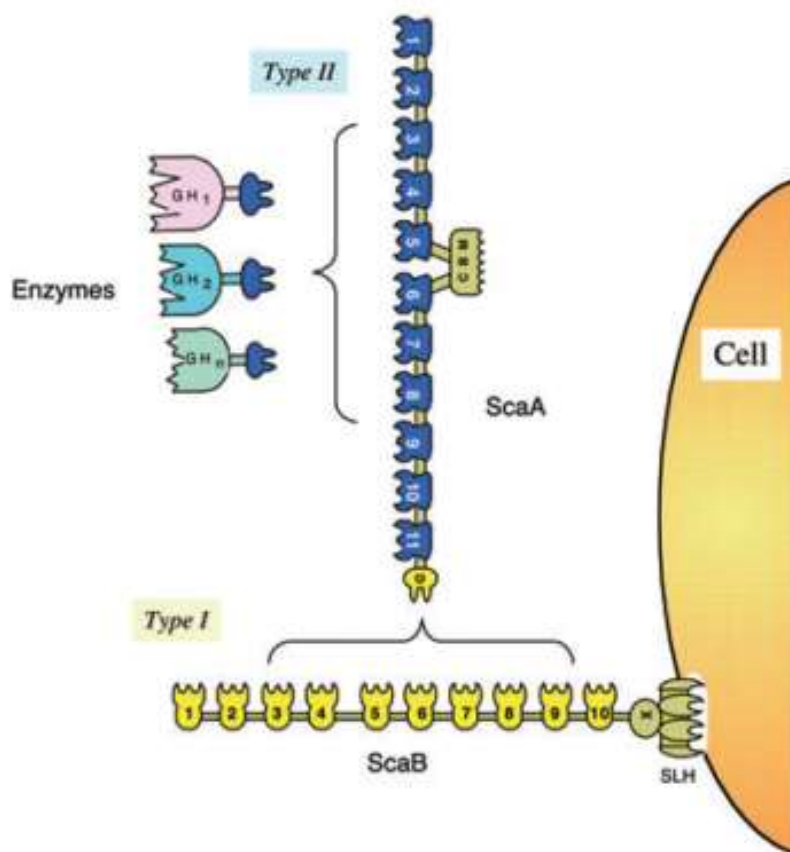


Figure 20 - *B. cellulosolvens* cellulosome system.

This cellulosome includes two known scaffoldins, ScaA (primary scaffoldin) and ScaB (anchoring scaffoldin) with 11 and 10 cohesins, respectively. The types of cohesins carried by the primary and anchoring scaffoldins are reversed relative to their canonical role. (E. a. Bayer, Lamed, White, & Flints, 2008)

One of the most complex cellulosomes is present in a significantly important rumen bacterium, *R. flavefaciens*. Its cellulosome comprises the scaffoldins ScaA, ScaB, ScaC and ScaE as well as the CttA protein which carries two CBMs (S. Y. Ding *et al.*, 2001) (Figure 21).

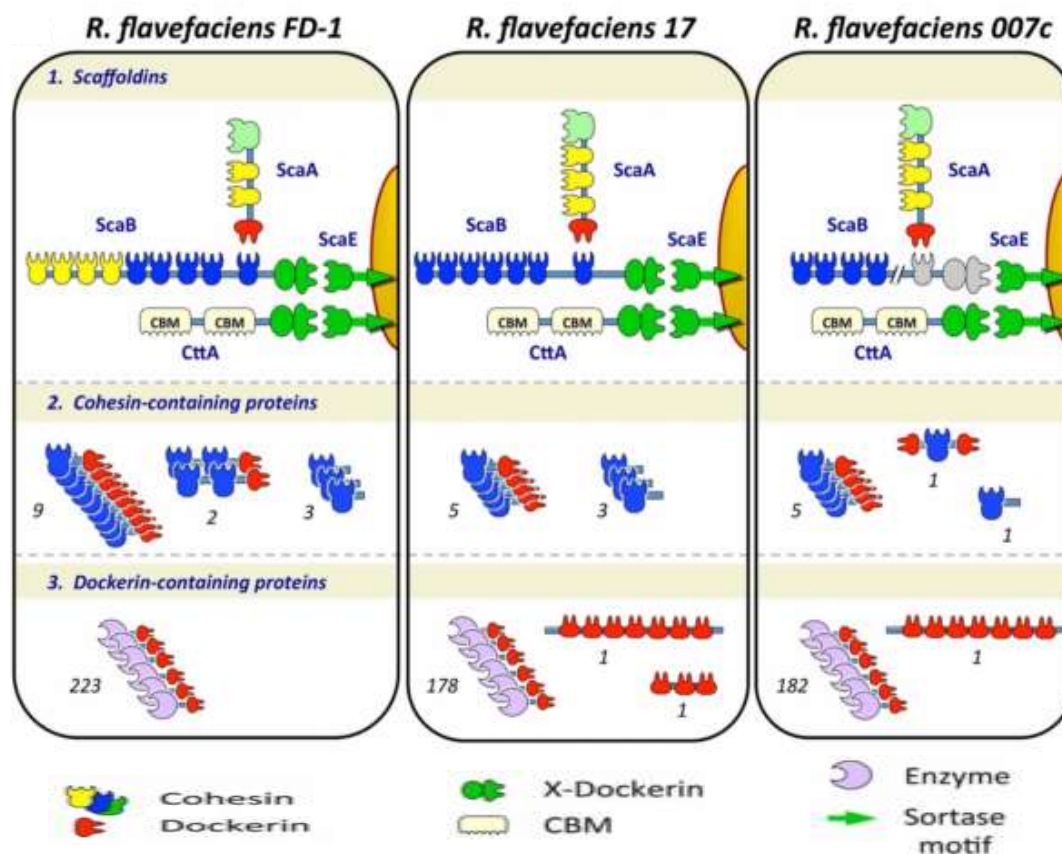


Figure 21 - Schematic representation the cellulosome-related proteins in *R. flavefaciens*. Schematic representation of scaffoldins, cohesin- and dockerin-containing proteins, which were identified in the genomes of each strain. Numbers indicate the copy number of each type of protein architecture, identified in the designated strain. Legend of pictograms is shown below (Dassa *et al.*, 2014).

These different scaffoldins have specific functions according to their type. For instance ScaA and ScaB are major subunits with several different repeats of cohesin modules. The ScaA has a specific C-terminal dockerin whereas ScaB harbors a C-terminal X-dockerin, a type-II Doc associated with an auxiliary module named, X-module. The anchoring function in these strains is attributed to ScaE thanks to its ability to bind either to ScaB or CttA. Lastly ScaC acts as an adaptor to regulate the binding of the various scaffoldin or the enzymes (Dassa *et al.*, 2014).

1.4.1 - Key Enzymes

When first described, cellulosomes were exclusively characterized as cellulose degrading biologic machines, however further studies clarified that they assemble a wide array of enzymes such as hemicellulases, pectinases and chitinases that can be divided in four main categories: glycoside hydrolases, glycosyltransferases, polysaccharide lyases and carbohydrate esterases (Doi, 2008) (Figure 22).

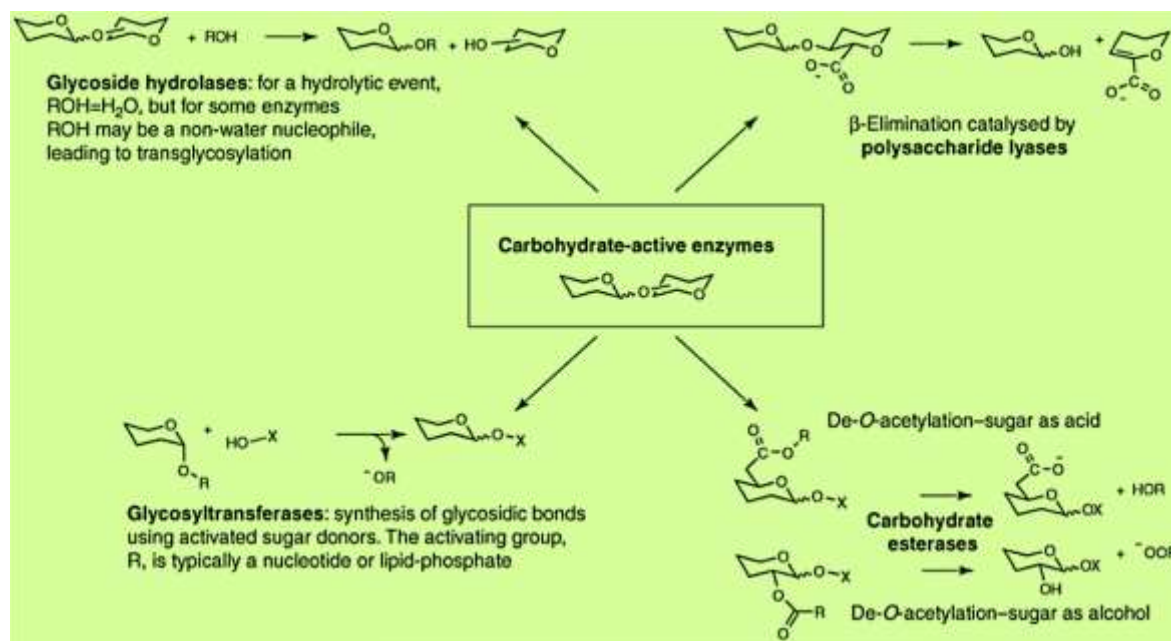


Figure 22 – Main carbohydrate-active enzymes and their action mechanism (Davies, Gloster, & Henrissat, 2005).

1.4.1.1 - Glycoside Hydrolases

Glycoside hydrolases (GHs) consist of a group of essential enzymes for plant cell wall hydrolysis. These enzymes break β -glycosidic and α -glycosidic bonds, and according to the catabolic mechanism can be classified as retaining or inverting enzymes. The main difference between these two denominations is that retaining enzymes maintain the configuration at the anomeric carbon, whereas inverting enzymes do not perform transglycosylations and only hydrolysis by inverting the configuration's at the anomeric center (Cantarel *et al.*, 2009).

The GH action mechanism also differs according to the substrate, thus affecting the overall final product. For instance exo-acting enzymes remove units of one or more sugars from the chain ends, whereas endo-acting enzymes cleave random bonds within the interior of the chain (Warren, 1996). In substrates like cellulose where the structure differs along the chain, the usage of both enables a better cleavage. The exo- β -1,4-glucanases release cellobiose whereas endo- β -1,4-glucanases degrade the amorphous regions of cellulose.

The GHs are classified according to both the type of catalyzed reaction and their substrate-specificity by the International Union of Biochemistry and Molecular Biology (IUB-MB; 1984). Their nomenclature consists in EC 3.2.1.x where the first three numbers correspond to enzymes hydrolyzing O-Glycoside bonds and the last two digits indicate either substrate or molecular mechanism (Henrissat, 1991). The nomenclature was revised by Henrissat in 1991, who tried to complement it by organizing the families according to sequence similarities since these bear a direct relationship to the folding of an enzyme. This classification system enables the organization in 14 clans which according to CAZymes database includes 133 families as shown in table 6 (May 2015)(Lombard et al., 2014).

Table 6 - GH clans of related families. Adapted from CAZy (Lombard et al., 2014).

Clans of Related Families	Protein Fold	GH Families
GH-A	(β/α) ₈	<u>1</u> <u>2</u> <u>5</u> <u>10</u> <u>17</u> <u>26</u> <u>30</u> <u>35</u> <u>39</u> <u>42</u> <u>50</u> <u>51</u> <u>53</u> <u>59</u> <u>72</u> <u>79</u> <u>86</u> <u>113</u> <u>128</u>
GH-B	β -jelly roll	<u>7</u> <u>16</u>
GH-C	β -jelly roll	<u>11</u> <u>12</u>
GH-D	(β/α) ₈	<u>27</u> <u>31</u> <u>36</u>
GH-E	6-fold β -propeller	<u>33</u> <u>34</u> <u>83</u> <u>93</u>
GH-F	5-fold β -propeller	<u>43</u> <u>62</u>
GH-G	(α/α) ₆	<u>37</u> <u>63</u>
GH-H	(β/α) ₈	<u>13</u> <u>70</u> <u>77</u>
GH-I	$\alpha+\beta$	<u>24</u> <u>46</u> <u>80</u>
GH-J	5-fold β -propeller	<u>32</u> <u>68</u>
GH-K	(β/α) ₈	<u>18</u> <u>20</u> <u>85</u>
GH-L	(α/α) ₆	<u>15</u> <u>65</u> <u>125</u>
GH-M	(α/α) ₆	<u>8</u> <u>48</u>
GH-N	β -helix	<u>28</u> <u>49</u>

1.4.1.2 - Glycosyl Transferases

The glycosyl transferases (GTs) are able to catalyze the transfer of sugar moieties from an activated donor to a specific acceptor forming glycosidic bonds (Sinnott, 1990). This type of mechanism can either generate an inversion or a retention of the anomeric configuration (Campbell, Davies, Bulone, & Henrissat, 1997). Until now there are 97 different identified families according to CAZy database (Lombard *et al.*, 2014).

1.4.1.3 - Polysaccharide Lyases

Polysaccharide lyases (PLs) consist of a group of enzymes responsible for the cleavage of glycosidic bonds of specific carbohydrates that contain uronic acid through β -elimination, generating an unsaturated hexenuronic acid residue and a new reducing end. They present various fold types suggesting that they may evolved from totally different scaffolds (Lombard *et al.*, 2010). Currently they are divided in 22 families in the CAZy database.

1.4.1.4 - Carbohydrate Esterases

Carbohydrate esterases are a class of enzymes that catalyze the removal of O-(ester) and N-(acetyl) through De-O or De-N-acylation of substituted saccharides respectively as shown in Figure 23 (Davies *et al.*, 2005; Lombard *et al.*, 2014). As the previous enzymes they are organized in 16 different families in CAZy (May 2015) (Lombard *et al.*, 2014).

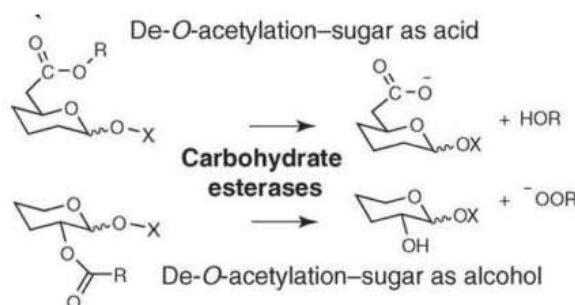


Figure 23 - Action mechanism of carbohydrate esterase. Carbohydrate esterases perform the O or N deacetylation of acetylated sugars (Correia, 2009).

1.4.2 - Cohesin-Dockerin Interactions

The interactions between cohesins and dockerins have a fundamental role on the overall assembly of the cellulosome (Doi & Kosugi, 2004). These Coh-Doc pairs exhibit one of the strongest protein-protein binding affinities known in nature, approximately 10^9M^{-1} (E. A. Bayer, Belaich, Shoham, & Lamed, 2004; Carvalho *et al.*, 2003; Fontes & Gilbert, 2010).

Dockerins (Doc) consists of 70 amino acids containing two duplicated segments with 22 residues each. These protein modules serve a noncatalytic role in several cellulosome enzymes and are usually present in a single copy module at the C terminus. In contrast, cohesins (Coh) are composed of 150 residues and these modules tend to occur in tandem repeats in scaffoldins. Both Coh and Doc modules have a high homology rate within the same species, with highly conserved residues that are directly involved in this protein:protein interaction (Fontes & Gilbert, 2010).

1.4.2.1 - Type-I Cohesin-Dockerin Interactions

The first reports of cellulosomal structures were type-I cohesins from the scaffoldins of *C. thermocellum* and *C. cellulolyticum*. These type-I cohesins consisted in 147 residues forming a nine-stranded β -sandwich in an elongated shape, with a β -barrel and jelly-roll topology. The β -strands 8, 3, 6 and 5 compose the first β -sandwich sheet whereas strands 9, 1, 2, 7 and 4 compose the second one with the β -strand 9 (C-terminus) and β -strand 1 (N terminus) running parallel while the rest are all anti-parallel. All nine β -strands are assembled around an extensive aromatic hydrophobic core (Ana Luísa Carvalho *et al.*, 2003; Shimon *et al.*, 1997). For example in *B. cellulosolvens*, the type-I cohesins are located on the ScaB cell-surface anchoring scaffoldin, showing similarities with other cohesins from *C. thermocellum* and *A. cellulolyticus* (Xu *et al.*, 2003).

In contrast the dockerin module is composed by three α -helices where the key conserved residues are located in helices 1 and 3. These three α -helices present a conformation consisting in a loop-helix motif followed by a helix-loop-helix motif, connected by a six-residue segment (Lytle, Volkman, Westler, Heckman, & Wu, 2001). The crystal structure of the type-I Coh-Doc complex revealed that Coh interacts with its Doc partner mainly along one face of its flattened β -barrel. Despite the Doc remarkable internal symmetry, when in complex with the Coh initial reports revealed a preferable binding to the cohesin through its second duplicated segment (helix 3) while only the C-terminal region of helix 1 also contributed to ligand recognition (Ana Luísa Carvalho *et al.*, 2003, 2007). Although the Coh-Doc interface mainly displays hydrophobic interactions, there is also a

network of hydrogen bonds between both proteins, were an highly conserved Ser-Thr pair in helix 3 of the Doc plays a central role in these polar interactions (Ana Luísa Carvalho *et al.*, 2003; Harry J. Gilbert, 2007) (Figure 24).

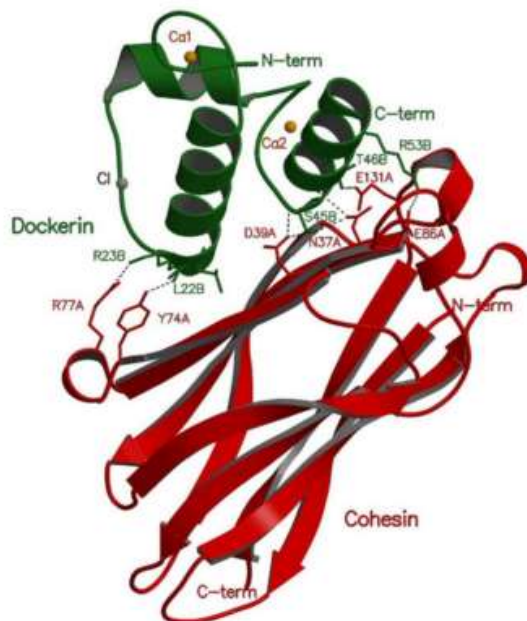


Figure 24- Structure of a type-I cohesin-dockerin interaction from *Clostridium thermocellum*. The complex is formed between a cohesin module (red) and the Ca²⁺(orange spheres) bound dockerin (green). The residues involved in domain contacts are shown as stick models (Carvalho *et al.*, 2003).

Nevertheless in 2007 Carvalho *et al.* revealed an alternative binding mechanism that allowed type-I Doc to bind to its Coh partner through two distinct surfaces, based alternatively on one of its two duplicated segments (Figure 24), thus uncovering a dual binding mode to Type-I Coh-Doc in *C. thermocellum*. This was achieved by mutating key residues in one of the Doc helices forcing the binding to occur through the other helix, not affecting the affinity constants between the two protein modules (Ana Luísa Carvalho *et al.*, 2007). Further experiments proved that a single binding mode could be obtained by substituting the Ser-Thr pair of helix 3 with two alanine amino acids, therefore inducing the mutated dockerin to rotate by 180°, with helix 1 now assuming the position of helix 3 and the Ser-Thr pair in the first duplicated segment dominating the hydrogen bond network. In summary the equivalent residues in helix 1 of mutant and helix 3 in the wild-type dockerin

interact with the cohesin module causing an almost perfect overlapping (Ana Luísa Carvalho *et al.*, 2007).

Apparently the dual binding mode is responsible for the introduction of quaternary flexibility into the multi-enzyme complex and for the enhancement of the substrate targeting and synergistic interactions between some enzymes such as exo- and endo-cellulases. Although its dual binding mode capability, the Coh-Doc binding stoichiometry is consistently 1:1 suggesting that the two binding sites are not able to recognize their ligands simultaneously. This dual binding mode may also be responsible for reducing the steric constraints that are likely to be imposed by assembling a large number of different catalytic and non-catalytic domains into a single cellulosome. The proline-threonine rich linker sequences that join cohesins within scaffoldins are also responsible for the quaternary flexibility. The inter-module linkers in free enzymes are extended and flexible, further supported by work showing the structure of adjacent cellulosomal cohesins of *A. cellulolyticus* where it is possible that cellulosomal synergy would be facilitated by the tethered motility of scaffoldins, as a result of the flexibility of the intermodular linker segment (Noach *et al.*, 2009). Additionally the linker sequences joining the cohesin domains within the *C. thermocellum* scaffoldin are quite long with more than 35 residues, therefore displaying a conformational freedom that may contribute to the synergy showed by the enzymes within the cellulosomes (Hammel *et al.*, 2005; Hammel, Fierobe, Czjzek, Finet, & Receveur-Bréchet, 2004). Furthermore, synergy optimization between specific enzymes and a better efficiency of cellulosome function may require the temporary switch of enzymatic subunits from one cellulosome position to another, but due to the tight interactions of the Coh-Doc this may be achieved by the existence of a second ligand binding surface (Fontes & Gilbert, 2010).

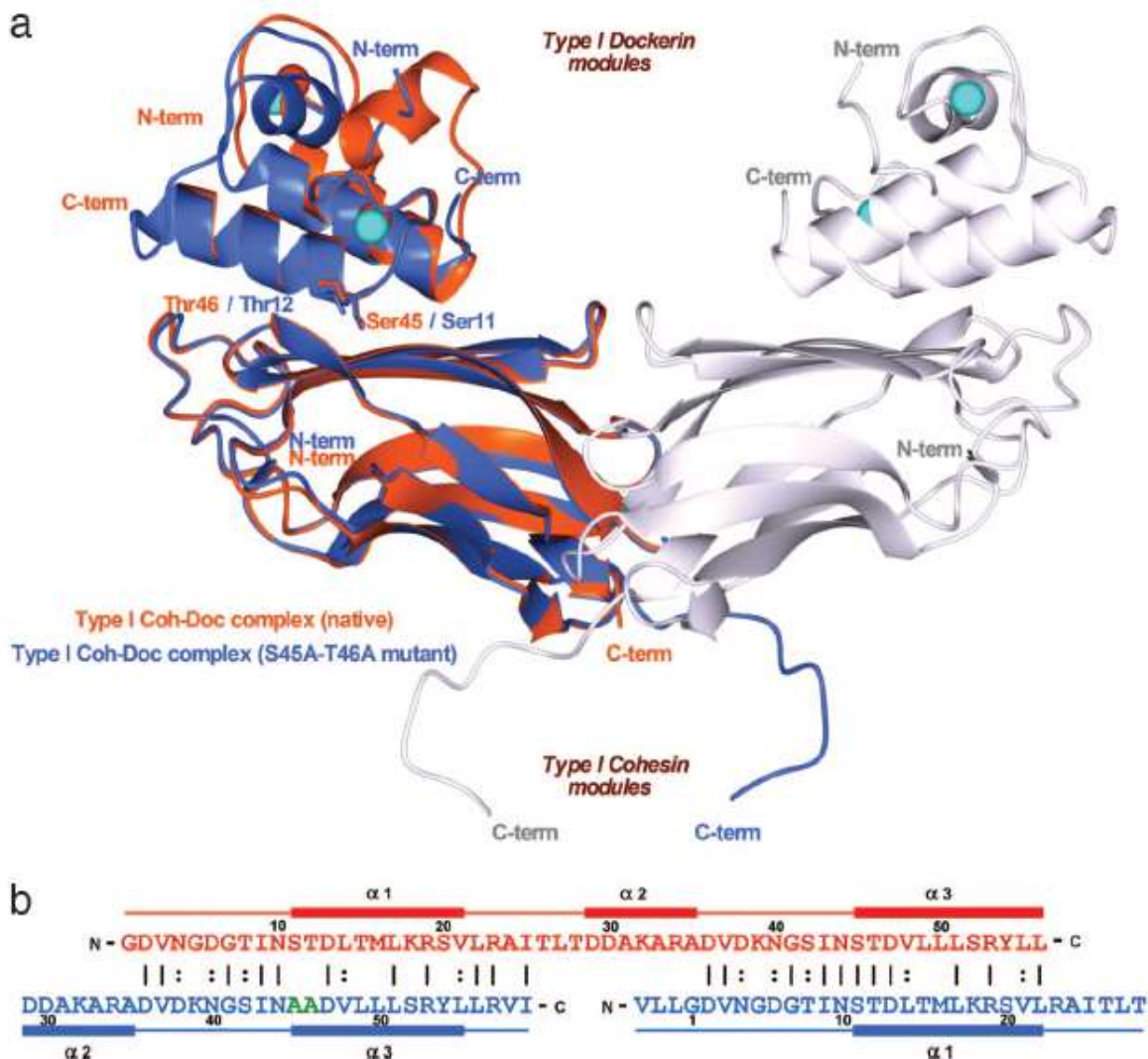


Figure 25 The dual binding mode of the Xyn10B dockerin. a) Ribbon representation of the superposition of the type-I Coh-Doc WT complex (orange) with its S45A-T46A mutant complex (blue). In the mutant complex, helix-1 (containing Ser-11 and Thr-12) dominates binding whereas, in the WT complex, helix-3 (containing Ser-45 and Thr-46) plays a key role in ligand recognition. Ser-11, Thr-12, Ser-45, and Thr-46, which interact with the cohesin module, are depicted as stick models and colored accordingly. A second molecule of the mutant complex present in the crystal asymmetric unit is represented in light-grey ribbon. The Ca^{2+} ions are depicted as spheres and colored orange, in the case of the WT complex, and light blue, in the case of the mutant. The N- and C-terminal ends are labeled and colored accordingly

b) The structure based sequence alignment of the WT (in red) and S45A-T46A mutant (in blue) type-I dockerins. Mutated residues, Ala-45 and Ala-46, are shown in green. Due to the internal 2-fold symmetry of each dockerin module, the two structures overlap almost perfectly in their $\alpha 1/\alpha 3$ regions (Carvalho *et al.* 2007).

Recent complimentary studies revealed that type-I coh modules are not exclusively present in *C. thermocellum* cellulosome scaffoldins as well as the dual binding module not being a ubiquitous feature of the type-I Doc (Smith & Bayer, 2013).

1.4.2.2 - Type-II Cohesin-Dockerin Interaction

One of the most important mechanism for the optimal uptake of nutrients and consequently for the viability of the microbe is the proper cellulosomes's attachment to the bacterial cell surface. In *C. thermocellum* the type-II Doc connects the cellulosome to the peptidoglycan layer through high-affinity interactions with type-II cohesins located in these cell-surface proteins: SdbA, OlpB, Orf2p and EXt (Béguin & Alzari, 1998; Fontes & Gilbert, 2010). The first structure obtained for type-II interactions was the type-II cohesin of *B. cellulosolvens* of ScaA, followed by the type-II cohesin from *C. thermocellum* anchoring protein SdbA. Although they were obtained from different locations they both present the same jelly-roll topology observed in type-I cohesins with the exception of the presence of an α -helix, between β -strand 6 and 7 and of two β -flaps (Carvalho *et al.*, 2005; Noach *et al.*, 2005). Despite presenting a similar fold to the type-I Doc, the type-II Doc also interacts with a neighboring module of unknown function, named the X-module, which adopts an immunoglobulin-like fold. Also both helices of type-II Doc contact with the cohesin surface over their entire length, contrasting with type-I Doc where ligand recognition is mainly associated with one of the Doc helices.

Due to the reduced charge in the interaction surface, binding in type-II interactions is predominantly hydrophobic. Still, between both dockerin helices, β -strands 8-3-6-5 of the interacting interface of the cohesin module and the X module, there is a significant hydrogen bond arrangement. Additionally, the type-II complex displays an intimate hydrophobic interface between the type-II dockerin and the Ig-like X-module fold, enabling the C-terminal region on the CipA scaffoldin to assume a rigid and elongated conformation (Adams, Pal, Jia, & Smith, 2006).

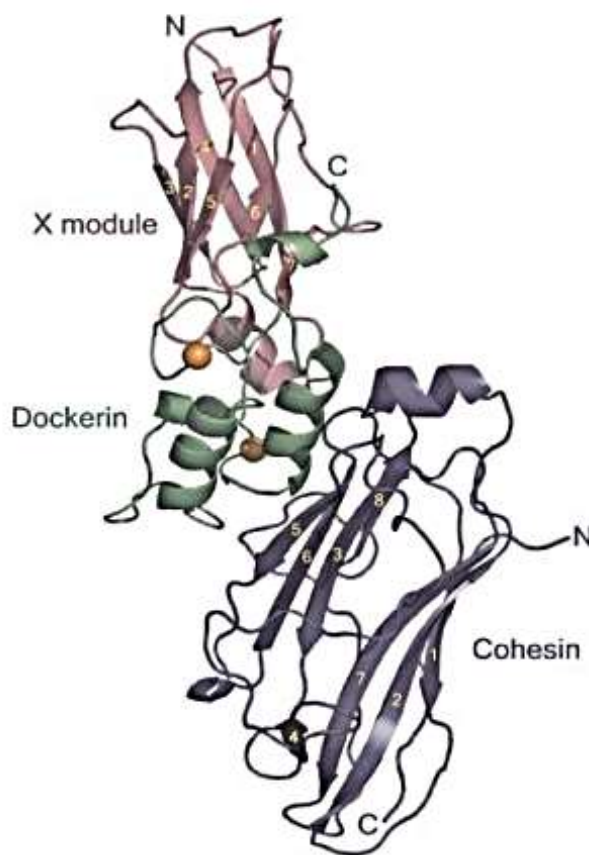


Figure 26 - Structure of the type-II cohesin-Xdockerin complex (SdbAcoh-CipAXdoc)
Ribbon representation of the type-II cohesin-dockerin complex with the cohesin module in blue, the dockerin in green and the X module in magenta. The β -strands of the X-module and the type-II cohesin are numbered in yellow. The N and C termini are labelled accordingly and the Ca^{2+} ions are depicted as orange spheres (Adams *et al.*, 2006).

The binding affinity between type-II cohesin, the X module and the dockerin was assessed by isothermal titration calorimetry (ITC) revealing a 1:1 stoichiometry and due to the fact that the high affinity of these interaction exceeds the limits of this technique the affinity constant was impossible to determine (Adams *et al.*, 2006).

The increased affinity of the type-II interaction was proposed to be associated to the X-module-mediated stabilization of the type-II dockerin combined with the hydrogen-bond contacts that exist directly between the X module and the type-II cohesin.

The discovery of the structural and biochemical data leading to the dual binding mode altered dramatically the way that cohesin-dockerin interactions were perceived. Mutational studies sustain the ability of most type-I dockerins to interact with cohesins upon

a 180° degree rotation about the interface plane, reflecting a two-fold symmetry of the dockerin structure. Through studies based on the dockerin primary sequence it was possible to identify the critical residues for cohesin interaction and enabled the deduction of a possible dual binding mode (Noach *et al.*, 2010). Therefore dockerins displaying a dual binding mode are composed of a near perfect 22-residue repeat, contrasting with single binding mode dockerins. Based on the existence of a near identical segment repeat in *A. cellulolyticus*, contrary to the situation on the type-II dockerin of *C. thermocellum*, Noach *et al.* (2010) hypothesized a dual binding mode interaction in the type-II dockerin of *A. cellulolyticus* (Figure 27).

	Ct type II dockerin (Asymmetric)	Ac type II dockerin (Symmetric)
Seq 1	DIVKDNSINLLDVAEVIRCFNA	GGTQDGAINLEDILEICKAFNT
Seq 2	DINRNGAINMQDIMIVHKHFGA	DLNRDGAISLEDVMIVAKHFNK
	** ::::** : * : * : *. *	. . :****.****:: :. * **.

Figure 27- Symmetric and Asymmetric dockerin sequences – Sequence alignment showing the 22 segment repeats of the asymmetric type-II dockerin (CipA) from *C. thermocellum* and the symmetric type-II ScaA dockerin from *A. cellulolyticus*. *-identical residues (Noach *et al.*, 2010).

Despite Noach *et al* (2010) attempted to crystallize the type-II cohesin-dockerin complex they were unable to do so, justifying their unsuccessful results with the apparent symmetry of the type-II dockerin which could lead to a dual binding mode, thus causing the formation of heterogeneous complexes and hinder the crystallization process.

To sum up, although cohesin and dockerin modules were being classified in different types, it's becoming more evident that this classification is only important to the phylogenetic similarities. The understanding of the cellulosomes and their composition may reveal unique dockerin sequences and their association with cellulosomal organization in order to potentially uncover novel CAZYme activities (Smith & Bayer, 2013).

1.4.2.3 - Cohesin-Dockerin specificity

Despite the structural similarity there is no cross-specificity between type-I and type-II cohesin-dockerin partners, thus enabling the efficient assembly and cell-surface

attachment of bacterial cellulosomes, respectively (Miras, Schaeffer, Béguin, & Alzari, 2002; Schaeffer *et al.*, 2002). Concerning type-I cohesin-dockerin interactions, it is known that the sequence duplication displayed by type-I dockerins from a variety of organisms, besides *C. thermocellum*, shows that the dual binding mode may be replicated in the majority of other microbial cellulosomes (Figure 28).

Organism	Protein(s)	Duplicated segment 1					Duplicated segment 2				
		Positions					Positions				
		11	12	18	19	23	11'	12'	18'	19'	23'
<i>C. thermocellum</i>	Enzymes ^a	S	T	K	R	K,R,G	S	T	K,H,S	R,K	K,R
	CipA	L	L	I	R	A	M	Q	H	K	A
<i>A. cellulolyticus</i>	Cel9B	S	I	R	K	G	S	L	R	Q	G
	ScaA	L	E	C	K	T	L	E	A	K	K
	ScaB	I	N	R	D	G	I	N	R	D	G
<i>B. cellulovorans</i>	Cel48A	M	A	A	Q	K	M	A	A	Q	K
	ScaA	S	D	R	Q	G	S	D	R	Q	G
<i>R. flavefaciens</i>	Enzymes (EndB-like) ^a	L	A	M	Q	N	G,N,A,E	D	Q	K,R,E,L	R,K,H,G
	CesA	S	F	R	K	N	V	A	Q	S	G
	XynE	S	L	R	R	K	V	A	K	R	G
	ScaA	V	A	N	R	D	D	K	D	P	K
	ScaC	L	A	M	Q	N	G	D	Q	K	T
<i>C. cellulolyticum</i>	Enzymes ^a	A	L,I	K	K	G	A	L,I	K	K	S,G
<i>C. josui</i>	Enzymes ^a	A	L,I	K	K	N,T	A	I	K	K,V	A,N
<i>C. cellulovorans</i>	Enzymes ^a	A,S	L,I	K	K	D,N,S,G	A	I	K	K	S,G,A
<i>C. acetobutylicum</i>	Enzymes ^a	G	R	R,T	K,Q	G	G	R	I,R	K,Q	G,N,S

Figure 28 - Putative recognition residues of different dockerin domains derived from cellulosomal components of different species. Scaffoldin-borne dockerins are highlighted in grey. Consensus residues represent the dominant amino acids that appear in the designated position from the indicated group of cellulosomal enzymes (Bayer *et al.*, 2004).

By comparing dockerin sequences a correlation is apparent between the essential Ser-Thr pair (positions 11 and 12) present in both duplicated segments, suggesting a determinant role in the interaction in *C. thermocellum*. The comparison of both *C. thermocellum* and *C. cellulolyticum* showed an invariant positioning of these pairs within the same species but a divergence between them (Pagès *et al.*, 1997). The same interspecies correlation is also maintained in the cohesin-dockerin interaction of *C. thermocellum* and *C. josui* due to its sequence similarity with *C. cellulolyticum* (Jindou *et al.*, 2004). Consequently it is evident that there is a high variability in the type-I cohesin-dockerin interactions within species (Mechaly *et al.*, 2001).

Through mutagenesis essays Pagès *et al* (1997) revealed that the type-I dockerin found in *C. cellulolyticum* cellulosomal enzymes was unable to interact with the cohesins

found in *C. thermocellum* scaffoldins and vice-versa. Generally, the residues at positions 11 and 12 of *C. cellulolyticum* dockerin were switched to the equivalent amino acids in *C. thermocellum* dockerins resulting in variants that recognized cohesins from both *Clostridia* (Pagès *et al.*, 1997). More recent studies suggest that residues at positions 18, 19 and 23 are also involved in species-specific ligand recognition. Therefore, there is a species specificity within the tight Coh-Doc interaction, contrasting with the apparent sequence similarity between different species (E. A. Bayer *et al.*, 2004).

Distinctly from the type-I interaction, the type-II Coh-Doc complex shows a relatively extensive cross-species plasticity. As such, the type-II cohesin from the *C. thermocellum* anchoring scaffoldin SdbA binds, not only to the *C. thermocellum* CipA type-II dockerin but also to both *B. cellulosolvens* and *A. cellulolyticus* type-II cohesins. Additionally, type-II Doc from *A. cellulolyticus* binds within the same species and also with *C. thermocellum* type-II cohesins (Haimovitz *et al.*, 2008). The biological importance of the flexibility of type-II cohesin-dockerin interaction is yet to be discovered.

1.4.3 - Carbohydrate Binding Modules

Essentially CBMs are non-catalytic domains most commonly found as domains of modular glycoside hydrolases or other modifying enzymes that recognize different carbohydrates (Blake *et al.*, 2006; Guillén & Sánchez, 2010; Hervé *et al.*, 2010; Shoseyov, Shani, & Levy, 2006).

The first mention of a non-hydrolytic component involved in the enzymatic degradation of crystalline cellulose was in the late 1940s and it was termed C1. Though initially CBM's were classified as cellulose binding domains (CBDs) with further studies and the increasing number of carbohydrate-active enzymes discovered that bind to carbohydrates other than cellulose, a more comprehensive terminology emerged, the CBMs (Shoseyov *et al.*, 2006).

CBMs can be found in any domain of life, usually associated with catalytic proteins that are able to recognize polysaccharides such as cellulose, chitin, β -glucans, starch, glycogen or even lipopolysaccharide and blood group A/B antigens. Occasionally these peptides are found isolated as a single protein like in the small olive pollen protein, Ole e 10,

the non-catalytic chitin-binding protein CBP21 from *Serratia marcesens* and the E7 and E8 from *Thermobifida fusca* (Guillén & Sánchez, 2010).

A CBM is thus defined as a contiguous amino acid sequence within a carbohydrate enzyme with a discrete fold having carbohydrate binding activity. This protein normally contains from 30 to 200 amino acids and exists as a single, double or triple domain. Their location within the parental protein can be both C- or N-terminal, being occasionally centrally positioned within the polypeptide chain (Guillén & Sánchez, 2010; Shoseyov *et al.*, 2006) (Figure 29).

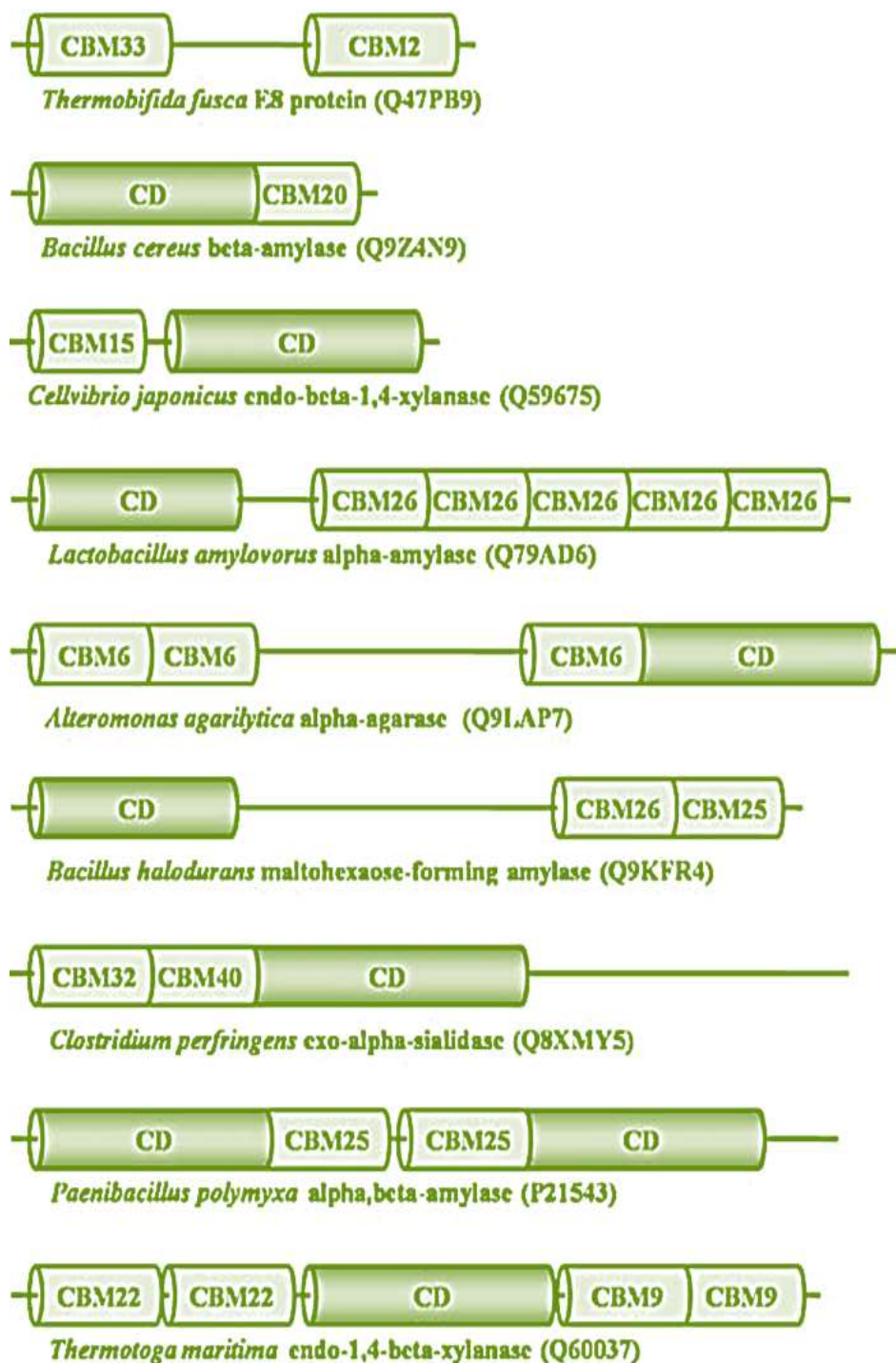


Figure 29 - CBM distribution in different proteins. They can be present in different numbers and diverse locations in the polypeptide chain. CD symbolizes the position of the catalytic domains in various glycoside hydrolases. The codes in parenthesis correspond to UniProt entries (Guillén & Sánchez, 2010).

1.4.3.1 - CBM's Classification and Nomenclature

CBMs have also been grouped into CAZy sequence-based families; currently, around half of the 59 CBM families contain members that bind to cell wall polymers. CBMs are suggested to enhance the efficiency of enzymes by mediating prolonged and intimate contact between the respective catalytic module and its target substrate.

Similar to the catalytic modules of glycoside hydrolases, CBMs are divided into families on the CAZy database (Hervé *et al.*, 2010; Lombard *et al.*, 2014; Shoseyov *et al.*, 2006) reaching a total of 71 CBM Families (January 2015). In its simplest form terminology, a CBM is named by its family, e.g. family 6 CBM from *Clostridium thermocellum* XynZ would be called CBM6, but one may also include the organism and even the enzyme from which it is derived to improve clarity. Thus this CBM6 may be defined as CtCBM6 or CtXynZCBM6. If glycoside hydrolases contain tandem CBMs belonging to the same family, a number corresponding to the position of the CBM in the enzyme relative to the N-terminus is included (Boraston, Bolam, Gilbert, & Davies, 2004). For example, *Clostridium thermocellum* enzyme Cthe_2137 contains two CBMs from family 35 and thus the first CBM is referred to as CtCBM35-1 and the second as CtCBM35-2.

In order to improve the identification process of novel CBMs, the family classification of CBMs was created. In some cases, the family classification may allow predicting of the binding specificity while aiding in identifying functional residues and revealing evolutionary relationships (Gilbert H.J., 1999). In 2004 Boraston *et al.* proposed a classification of CBM families based on their structural fold, alike what was described for the GHs superfamily classification. Therefore CBM families were categorized into seven structural family folds (β -sandwich, β -trefoil, cysteine knot, unique, OB fold, hevein fold and hevein-like fold). However, the dominant fold among CBMs is the β -sandwich fold, family fold 1 (Figure 29), and it comprises two β -sheets, each consisting of three to six antiparallel β -strands (Boraston *et al.*, 2004).

Fold family	Fold	CBM families
1	β -Sandwich	β -jelly 2,3,4,6,11,15,17,22,27,28,29,30,32,35,36,44 immunoglobulin 9,20,25,26,31,33,34
2	β -Trefoil	13, 42
3	Cysteine knot	1
4	Unique	5,12
5	OB fold	10
6	Hevein fold	18
7	Unique; contains hevein-like fold	14

Figure 30 - CBMs classification based on fold (Correia, 2009).

Besides their different fold characteristics CBMs have shown three main specificities, enabling a ligand binding property based assemblage (Guillén & Sánchez, 2010). CBMs are also grouped into three types: type A CBMs which interact with crystalline polysaccharides, primarily cellulose, type B modules which bind to internal regions of single glycan chains, and type C CBMs that recognize small saccharides in the context of a complex carbohydrate (Fontes and Gilbert, 2010).

1.4.3.1.1 - Type A CBMs – surface-binding

Type A CBMs include members of CBM families 1, 2a, 3, 5 and 10 that bind to insoluble, highly crystalline polysaccharides, such as cellulose, chitin or mannan. Therefore, showing little or no affinity for soluble carbohydrates provides a distinctive property when compared with the other CBM types. These CBMs have a flat or platform-like hydrophobic surface composed of aromatic residues that recognize the carbohydrate ligand, as shown in Figure 32. Thus, the planar conformation of the type A binding site reflects the architecture of the crystalline polysaccharides to which they bind and that also displays a flat surface. Hydrogen bonds have little effect in ligand recognition which is dominated by stacking interactions. Additionally, the interaction of type A CBMs is associated with positive entropy, demonstrating that the thermodynamic forces that drive the binding of CBMs to crystalline ligands are relatively unique among carbohydrate binding proteins (Boraston *et al.*, 2004; Guillén & Sánchez, 2010).

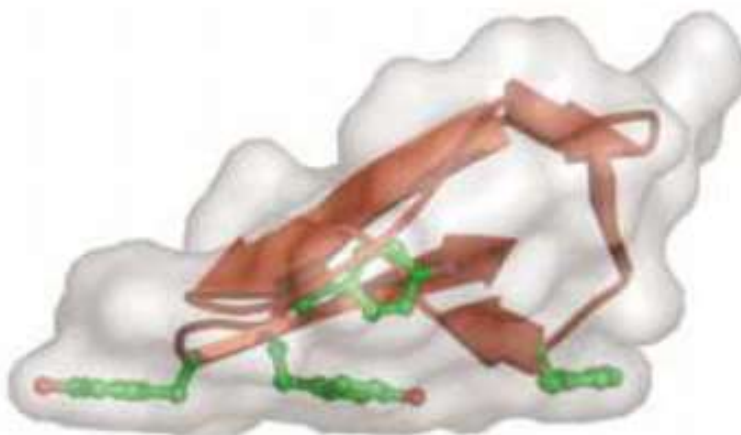


Figure 31 – CBM from type A. Type A CBM1 (PDB:1cbh) from *Hypocrea jecorina* L27 (Kraulis, 1989).

1.4.3.1.2 - Type B CBMs – glycan-chain-binding

Type B CBMs bind to amorphous cellulose or soluble complex carbohydrates such as xylan or xyloglucan, for example. These CBMs allocate the carbohydrate chain in a distinctive cleft in which aromatic residues interact with the single polysaccharide chain (Figure 32). Aromatic side chains form twisted or sandwich platforms. The orientation of these amino acids was shown to be a key determinant of ligand specificity. Biochemical studies revealed that the binding capacity of these CBMs is determined by the degree of polymerization of the carbohydrate ligand. Therefore, the affinity was shown to be higher for hexasaccharides and much lower for oligosaccharides with a degree of polymerization of three or less. Consequently, type B CBMs are usually described as “chain binders”, where the depth of these binding sites varies from very shallow to being able to accommodate the entire width of a pyranose ring. Additionally, type B CBMs comprise several sub-sites that are able to accommodate the individual sugar units of the polymeric ligand. Among others, CBMs from families 2b, 4, 6, 15, 17, 20, 22, 27, 28, 29, 34 and 36 are included in this type B group and in general, these proteins have evolved binding site topographies that are able to interact with individual glycan chains rather than crystalline surfaces. In contrast with what was observed in type A CBMs, direct hydrogen bonds play a key role in defining the affinity and ligand specificity in type B CBMs (Boraston *et al.*, 2004; Guillén & Sánchez, 2010; Hashimoto, 2006).

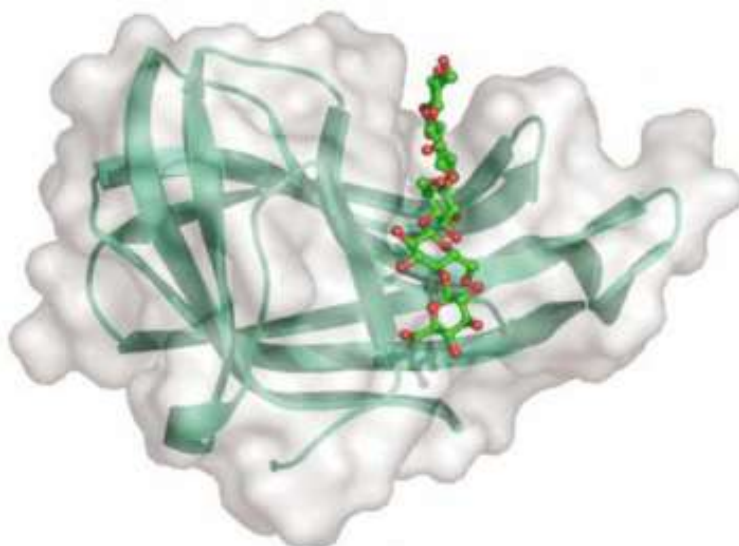


Figure 32- CBM from type B. Type B CBM4-2 (PDB:1GU3) from *Cellulomonas fimi* ATCC 484 (Boraston *et al.*, 2002).

1.4.3.1.3 - Type C CBMs – small-sugar-binding

Lastly, this unique class of CBMs have the lectin-like property of binding optimally to mono-, di- or tri-saccharides, thus lacking the extended binding-site grooves of type B CBMs (Boraston *et al.*, 2004) (Figure 33). It should be emphasized, however, that the distinction between type B and type C CBMs can be subtle. For example, the type B CBM6 module of the *Clostridium stercorarium* xylanase has a very similar fold to the type C lectin-like CBM32 family, but apparently binds longer oligosaccharide ligands (Boraston *et al.*, 2004). Nevertheless, it is apparent that the hydrogen-bonding network between protein and ligand is more extensive in type C than in type B CBMs, consistent with their lectin-like properties (Boraston *et al.*, 2004). The type C CBMs currently include examples from families 9, 13, 14, 18, 32, 40, 42, 43 and 50 (Table 30). Members of families 13 (e.g. ricin toxin B-chain), 14 (e.g. tachycitin) and 18 (e.g. WGA) were first discovered as lectins with small-sugar-binding activity and have only subsequently been included as CBMs due to their discovery in a number of glycoside hydrolases that degrade plant structural carbohydrates (Boraston *et al.*, 2004).

Identification and characterization of type C CBMs is lagging behind type A and B CBMs, probably due to their limited presence in plant cell wall active glycoside hydrolases.

Rather, type C CBMs, particularly CBMs from families 13 and 32, appear to be more prevalent in bacterial toxins or enzymes (glycoside hydrolases and glycosyl transferases) that attack eukaryotic cell surfaces or matrix glucans (Boraston *et al.*, 2004).

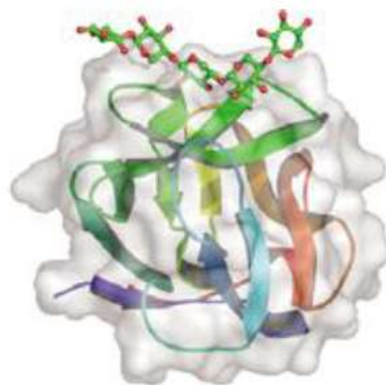


Figure 33 - CBM from type C. Type C CBM13 (PDB:1MC9) from *Streptomyces lividans* (Notenboom, 2002).

1.4.3.2 - CBM's Functions

CBMs primary function is to recognize and bind specifically to carbohydrates. The biological consequence of this event is a closer proximity between the catalytic domain and the substrate, resulting in enhanced hydrolysis of insoluble substrates, polysaccharide structure disruption and cell surface protein anchoring (Boraston *et al.*, 2004; Guillén & Sánchez, 2010). However, there are a few specific tasks worth referring.

1.4.3.2.1 - Enzyme targeting

The association of CBMs to CAZymes is responsible for the increased enzyme concentration on the polysaccharide surface as well as enhancing enzymatic activity. Therefore, the removal of the CBM causes a reduction, and in some cases abolition of binding to insoluble substrates, resulting in a partial or complete loss in catalytic activity. Nonetheless, the activity on soluble substrates is not frequently affected (Guillén & Sánchez, 2010).

According to their type, CBMs also show a tendency to escort specific enzymes. For instance, CBMs that bind to the surfaces of crystalline polysaccharides (referred to as type A

modules) can be appended to a variety of glycoside hydrolases. Whereas, CBMs that interact with single polysaccharide chains (type B or C) bind to polysaccharides that are the substrates for the cognate catalytic module of enzymes such as, cellulases, xylanases and mannanases that contain type B CBMs which bind to cellulose, xylan and mannan, respectively. Consequently, the CBM maintains proximity of the appended catalytic domain with its target substrate within complex macromolecular structures, such as the plant cell wall. It is apparent that this targeting function is even more subtle than the somewhat crude partitioning of enzymes to the different polysaccharides of plant cell walls. Although the interaction of CBMs with cellulose is occasionally irreversible, their contact with the cellulose surface is a dynamic process (Shoseyov *et al.*, 2006).

Through fluorescence recovery techniques, CBMs were labeled with fluorescent tags by Jervis *et al.* which confirmed that CBMs from *Cellulomonas fimi* are mobile on the surface of crystalline cellulose (Jervis, Haynes, & Kilburn, 1997). This dynamic binding behavior of CBMs has a functional importance as often they are part of cellulases acting processively (Lehtiö *et al.*, 2003). Although CBMs have a high affinity for carbohydrates, it is obvious that an irreversible binding would be fatal for the catalytic efficiency of the appended hydrolytic domains.

1.4.3.2.2 - Ligand binding and specificity: the role of aromatic residues.

For CBM's carbohydrate recognition it is essential the interaction of aromatic amino acid side chains with ligands, in which the location of the aromatic amino acid side chain and loop structures that shape the binding sites to mirror the conformation of the ligand are the two main key factors to determine the binding specificity (Boraston *et al.*, 2004).

Again, the CBM type effects the binding residues and how the connection is made. For instance, type A CBMs bind to crystalline cellulose through a flat platform containing aromatic residues such as tyrosines and tryptophans, separated by a distance corresponding to the length of the repeating unit, 10.3 Å in cellulose (Figure 34A). In this case, the flat aromatic ring interacts with the pyranose rings of the polysaccharides. This driven interaction

may be supplemented by a few hydrogen bonds mediated by polar residues located at the binding interface (Tomme, Warren, & Gilkes, 1995).

In contrast, type B CBM's binding sites interact with individual glycan chains rather than crystalline surfaces (Boraston *et al.*, 2004). In the binding sites of families 2b, 15, 17, 27, 29, 34 and 36, the apolar platform can be twisted due to the rotation of the planes of two to three aromatic amino acid side chains relative to one another (Figure 34 B) (Boraston *et al.*, 2004). In type B, the binding cleft can also present a sandwich form, where the aromatic amino acid side chains sandwich a sugar unit in the ligand by stacking against the face of the pyranose ring (Figure 34 C). This latest case is common to CBM families 4, 6, 9 and 22. The sandwich and twisted platforms may be used concurrently in the same CBM and can both accommodate the conformations of soluble oligosaccharide ligands (Boraston *et al.*, 2004). CBMs appear to have performed carbohydrate-recognition sites which mirror the solution conformations of their target ligands, thereby minimizing the energy required for binding (Boraston *et al.*, 2004).

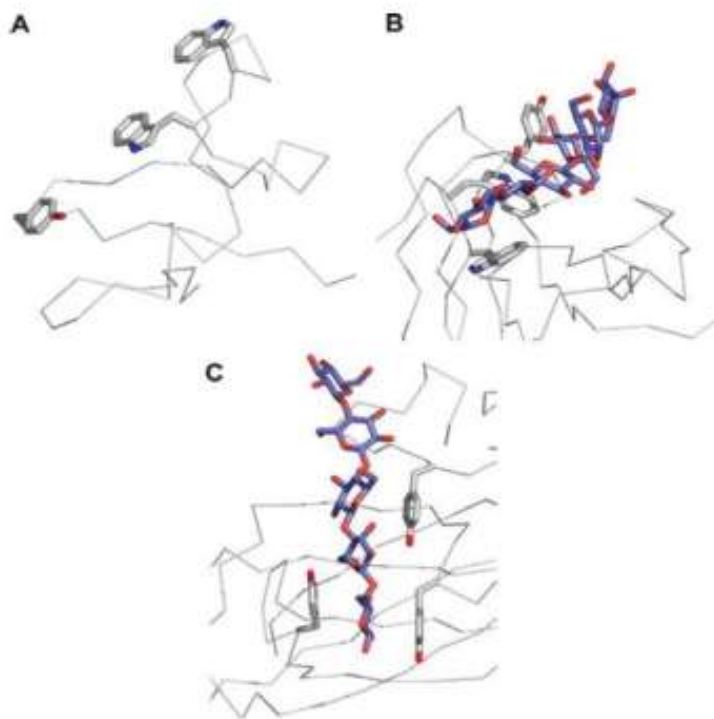


Figure 34 - The three types of CBMs - (A) - Planar Type A CBM CjCBM10 (B) - Twisted Type B CBM PeCBM29-2 (C) - Sandwich Type B CBM CfCBM4-2 (Boraston *et al.*, 2004).

When comparing both types A and B it is clear that the aromatic residues play a crucial role in ligand recognition, where a significant importance depends on the orientation of these amino acids in relation to CBMs binding specificity. For example, CBM family 2 contains type A and type B members, named CBM2a and CBM2b, which bind crystalline cellulose or xylan, respectively. The explanation to this contrasting ligand specificity results from the rotation by 90° of the side-chain of one of the surface tryptophan (Trp259) involved in the protein-carbohydrate interaction in CBM2b compared with its position in CBM2a (Simpson, Xie, Bolam, Gilbert, & Williamson, 2000). To prove this possibility, Simpson *et al.* (2000) showed that the ligand specificity of CBM2 is determined largely by a single amino acid, which controls the orientation of one of the tryptophan residues that interacts with the saccharide ligand. When the tryptophans are coplanar, the CBM recognizes the planar chains of cellulose, whereas when they are twisted into a near perpendicular arrangement, the protein recognizes the helical structure of xylan. Thus, in this CBM family, ligand specificity is determined largely by recognition of the three-dimensional shape of the polysaccharide ligand, rather than by specific hydrogen bonding patterns, as is typically seen in proteins that recognize monosaccharides (Simpson *et al.*, 2000).

Contrasting to type A CBMs, the direct hydrogen bonds in type B CBMs suggests the existence of an important influence between affinity and ligand specificity, usually described as chain binders. However there is no actual evidence indicating that water mediated hydrogen bonds are essential in CBM's ligand targeting.

Although the orientation and positioning of the aromatic residues in the binding sites of CBMs is the primary driver of specificity and affinity in these proteins, other interactions, including direct hydrogen bonds and calcium-mediated co-ordination also play a significant role in CBM ligand recognition (Boraston *et al.*, 2004).

1.4.3.2.3 - Substrate disruption

Although controversial, there has been some evidence implying the ability of CBMs for substrate disruption (Shoseyov *et al.*, 2006). The binding of CBMs to a crystalline substrate leads to polysaccharide chains disorganization therefore enhancing the substrate availability.

Further studies showed an evident substrate disruption in independent and non-catalytic chitin-binding protein CBP21, which belongs to CBM family 33 and is produced by *S. marcescens*. CBP21 promotes efficient crystalline chitin degradation by chitinases through polar interactions that modify the substrate arrangement (Guillén & Sánchez, 2010). Besides these substrates, CBMs also show non-hydrolytic disruption activity on other substrates such as starch (Southall, Simpson, Gilbert, Williamson, & Williamson, 1999) and cellulose fibers (Din *et al.*, 1991).

Additionally, enzymes located near CBMs which possess disruption functions, may facilitate the hydrolysis of recalcitrant substrates giving them an advantage over other enzymes. This idea is supported by experimental evidence that CBMs have disruption functions, although this property does not seem to be shared by many CBMs (Guillén & Sánchez, 2010).

Another interesting attribute of CBMs is their prospect role in development and modulation of plant cell walls. For example, the olive pollen protein Ole e 10, which has been recently classified as a CBM family 43 member, is an independent CBM that binds to callose (1,3- β -glucan, major component of the pollen tube wall), suggesting a role in regulation of the enzymatic activity of proteins implicated into cell wall synthesis and degradation during pollen germination (Barral *et al.*, 2005).

Finally, it has also been demonstrated that CBMs prevent the flocculation of crystalline cellulose (Shoseyov *et al.*, 2006), while Lee *et al.* (2000) provided the first physical evidence that a CBM from *Trichoderma reesei* caused peeling and smoothed surface on cotton fibers. This non-disruptive ability provides CBMs with the potential for usage in numerous biotechnological applications, such as in the textile and paper industries.

1.4.3.2.4 - CBMs in pathogenic bacteria

Besides binding to plant cell wall polysaccharides, CBMs are also able to recognize glycans in animal cells such as glycoproteins, glycolipids or other glyco-conjugates. Recently, several studies have reported the existence of CBMs in virulence factors or in proteins related to metabolism in pathogens. In association with toxins and hydrolases these

proteins could promote tissue destruction and enhance both bacterial spread and pathogenesis (Guillén & Sánchez, 2010).

For example, studies with virulence factors from *Streptococcus pneumoniae* and *Streptococcus pyogenes* provide evidence of host tissue recognition by multivalent binding of pullulanases with tandem CBM41 to alveolar type-II cells from mouse lung. The suggested biological function of these CBMs is to bind the bacteria to the reserves of intracellular glycogen in alveolar cells, allowing polysaccharide degradation by the pathogen (van Bueren, Higgins, Wang, Burke, & Boraston, 2007). A similar enhanced affinity to host lung tissue was shown with a virulence factor containing three family 47 CBMs from *S. pneumoniae* with specificity for fucosylated carbohydrates. The virulence factor was capable of binding ABH blood group antigens and more efficiently binding to LewisY antigen (Boraston *et al.*, 2006). Similarly to CBM 47, other CBMs have showed either pathogenic or virulent associated properties as seen in Figure 34.

Microorganism	Protein	Family	CBM arrangement
<i>S. pneumoniae</i>	Fucolectin-related protein	CBM47	Three tandem in C-terminus
	Pullulanase	CBM41	Two tandem in N-terminus
	Neuraminidase B	CBM40	Single N-terminus
<i>S. pyogenes</i>	Pullulanase	CBM41	Two tandem in N-terminus
<i>V. cholerae</i>	Sialidase	CBM40	Single N-terminus
<i>C. perfringens</i>	μ -toxin	CBM32	Four tandem in the middle
	Sialidase	CBM32,CBM40	N-terminal tandem with one of each family

Figure 35 - CBMs present in toxins, virulence factors or pathogenesis associated proteins (Guillén & Sánchez, 2010)

The importance of these findings on the mechanism and specificity of ligand recognition in CBMs situated in virulence factors is related to the possibility of designing new compounds that target the biological function of CBM–ligand binding as a way to control pathogenesis or dissemination of pathogenic bacteria (Guillén & Sánchez, 2010).

1.4.3.3 - CBM's Multivalence

The weak interaction between a carbohydrate-binding protein and its ligand are often compensated in nature by multivalent interactions. In these cases, multiple clustered carbohydrate-binding sites interact simultaneously with carbohydrate ligands, which present multiple recognition elements, resulting in increased association constants relative to any one of the isolated protein-carbohydrate interactions (Boraston *et al.*, 2004). These weak relations can result from a single protein having multiple binding sites or from the association of two or more univalent carbohydrate-binding proteins into multivalent quaternary structures (in random or in tandem). To date, no CBM has been found to form quaternary structures in its natural state. However, multiple CBMs, often arranged in tandem, are found frequently in glycoside hydrolases, which effectively become multivalent carbohydrate-binding proteins. Interestingly, the appearance of multiple CBMs seems to occur more often in thermo- or hyperthermophiles enzymes. This may allow overcoming the loss of binding affinity that accompanies most molecular interactions at elevated temperatures (Boraston *et al.*, 2004).

The first CBMs in tandem to be investigated were the family 2b CBMs of *Cellulomonas fimi* xylanase 11A (Bolam *et al.*, 2001). While the individual association constants for xylan were of approximately 10^4 M^{-1} , CBMs linked in tandem displayed an association constant of approximately 10^6 M^{-1} . A similar result was observed for the three family 6 CBMs of the *Clostridium stercorarium* putative xylanase (Boraston *et al.*, 2002), as well as for the tandem CBM17 and CBM28 modules from *Bacillus sp.* 1139 Cel5 (Boraston, Kwan, Chiu, Warren, & Kilburn, 2003), showing that individual CBM modules containing multiple carbohydrate-binding sites occur in a variety of CBM families.

1.5 - Process Improvements

When faced with the challenge of enhancing the applicability of plant derived biomass several opinions emerge. However most of them fit in either of these two main trends: the improvement of the source or the improvement of the degradation techniques.

1.5.1 - Plant Engineering

One of the compelling reasons for the evaluation of plants as raw materials for the bio industries is the composition and structure of the lignified walls. Therefore several efforts are being made in order to improve the composition of plant cell walls for value-added agroindustrial uses (Boudet *et al.*, 2003).

Based on the chemical flexibility of the secondary cell wall it is possible to develop new strategies to enhance its composition through genetic engineering (Boudet *et al.*, 2003). By this engineering the potential of plant biomass dramatically improves. This can be achieved through standard genetically modified (GM) procedures and through non-GM methods such as random mutagenesis and screening of natural variation (Burton & Fincher, 2014). The evolution of functional genomics enable the targeting of new plant genes that rapidly become available for this purpose and their use will open new avenues for producing tailor-made plant products with improved properties (Boudet *et al.*, 2003).

Mutants with reduced levels of cellulose or with its crystallinity altered, termed Brittle mutants, promise to require less energy for milling and to enable enzymes to better hydrolyze and access the various cell wall components. In fact cellulose crystallinity has been associated as a sign of the susceptibility of cellulose to enzymatic depolymerisation (Burton & Fincher, 2014).

However the manipulation of cellulose levels has demonstrated to be challenging. For instance the overexpression of a CesA gene in transgenic aspen, *Populus tremuloides*, resulted in silencing of the both the transgene and endogenous CesA genes, and greatly reduced cellulose levels (Burton & Fincher, 2014).

In addition to approaches involving CesA genes, the expansion of knowledge about transcription factors that regulate the expression of cellulose synthases could provide new opportunities to manipulate cellulose levels in crop plants (Burton & Fincher, 2014).

Nonetheless, current studies suggests that plants will only tolerate relatively small changes in cellulose content, and as such there might be more scope for manipulating levels of the non-cellulosic polysaccharides of the wall (Burton & Fincher, 2014).

Besides the undoubtedly importance of the advances in the cell wall biology and its implications in the commercial and industrial applicability, most of these approaches involve genetic modifications. The controversy associated with the usage of Genetic Modified Organisms is well known, not only in applications related to human food, but also for engineering forage crops, turf grass and bioenergy crops. Therefore this resistance tends to discourage new approaches that target genetic improvement of plants (Burton & Fincher, 2014).

1.5.2 - Engineered microbial systems

In order for plant biomass to become a viable feedstock for meeting future biorefineries, efficient and cost-effective processes must exist to breakdown cellulosic materials into their primary components. A one-pot conversion strategy or, consolidated bioprocessing, of biomass into other by-products such as ethanol would provide the most cost-effective route to renewable fuels. The implementation of this strategy is being actively pursued by both multi-disciplinary research centers and industrialists working at the very cutting edge of the field (Elkins *et al.*, 2010).

Even though a diverse range of bacteria and fungi possess the enzymatic machinery capable of hydrolyzing plant-derived polymers, till now none discovered are completely suitable for an industrial strength biocatalyst for the direct conversion of biomass to combustible fuels. Through combining synthetic biology with a better enlightenment of enzymatic cellulose hydrolysis at the molecular level it's possible to envision the rational engineering of microorganisms for utilizing cellulosic materials with simultaneous conversion to fuel (Elkins *et al.*, 2010; Hyeon *et al.*, 2013).

The development of a hydrolysis enzyme complex is a useful strategy for the construction of CBP-enabling microorganisms, like shown in Figures 36 and 37. This pathway involves the combined engineering of non-cellulolytic organisms with the ability to produce a valuable and high-yield product (Hyeon *et al.*, 2013). The synergistic interaction of multiple enzymes and their substrates overcomes the rate-limiting step of converting crystalline forms of cellulose to cellobiose, leading to the efficient degradation of crystalline polysaccharides in plants (Doi & Kosugi, 2004; Fontes & Gilbert, 2010).

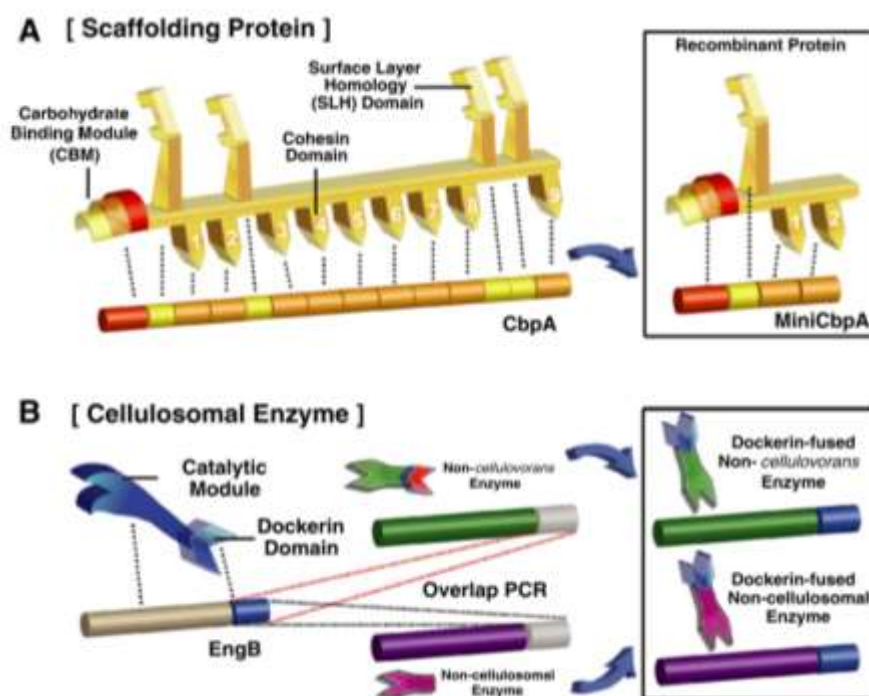


Figure 36 - Artificial multi-functional enzyme complex. This complex is composed of a recombinant scaffolding protein and dockerin-fused chimeric enzymes. (A) The recombinant scaffolding protein for complex formation includes two cohesin domains and a CBM. The cohesin domain serves as a binding moiety for the assembly of a multi-subunit enzyme complex. (B) Two types of dockerin-fused enzymes derived from a non-cellulovorans and a non-cellulosomal enzyme, respectively, are constructed using the overlapping PCR method. Each chimeric enzyme is created by the fusion of a dockerin domain to endoglucanase EngB from *C. cellulovorans* (Hyeon *et al.*, 2013).

The use of the minicellulosome as a multi-functional enzyme complex leads to the co-localization of synergistic combinations of hydrolytic enzymes (Hyeon *et al.*, 2013). This concept consists of a recombinant scaffoldin protein and cellulosomal enzymes merged as an efficient multi-functional enzyme complex for use in industrial bioprocesses. This would be achieved by the mix-and-match configuration of parts from different cellulosomes in a suitable industrial host cell system. Therefore, the construction of these engineered

complexes plays a major role in the industrial prospects. For instance, enzyme complexes containing the cellulosomal cellulase gene and the recombinant scaffoldin protein miniCbpA gene from *C. cellulovorans* were successfully co-expressed and formed in vivo in *B. subtilis*, *S. cerevisiae* and *Corynebacterium glutamicum* by harnessing the interaction between the cohesin and dockerin domains. Similarly, cellulosomes derived from other Clostridium strains such as *C. thermocellum* and *C. cellulolyticum* have been used to design mini-cellulosomes (Hyeon *et al.*, 2013) (Figures 36 and 37).

[Multi-functional Enzyme Complex Tool]

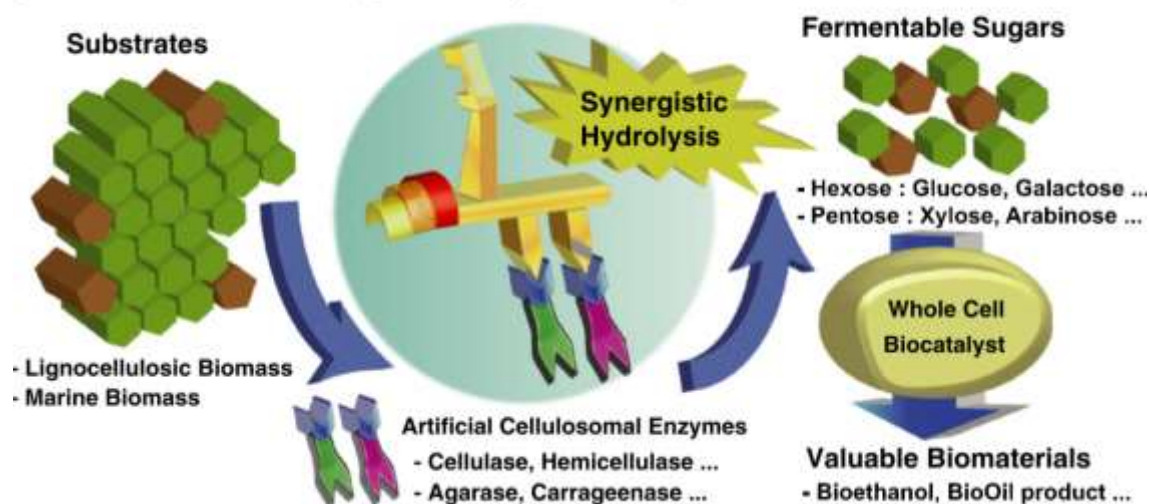


Figure 37 - Multi-functional enzyme complex for biomass utilization. The multi-functional enzyme complex assembly with a recombinant scaffolding protein and a chimeric dockerin-fused enzyme has a high level of hydrolysis activity and can convert the various biomasses to valuable biomaterials (Hyeon *et al.*, 2013).

The whole-cell biocatalysts present several advantages including the reduced carbon catabolite restriction effect, the smaller sterilization costs and the usage of a single reactor because the cells immediately utilize the sugar. Nonetheless, for a viable and cost-effective strategy to produce biomaterial using metabolically engineered strains, microorganisms with the ability to hydrolyze biomass, including lignocellulosic and marine biomass, are required (Hyeon *et al.*, 2013).

In conclusion, the applicability of a multi-functional protein complex that is suitable for expression in various industrial host cell systems has drawn considerable attention as an attractive and powerful strategy for achieving viable and cost-effective biomaterial

production and whole-cell biocatalysts with various functions (Hyeon *et al.*, 2013; Singh *et al.*, 2014).

Chapter 2 - Materials and Methods

2.1 - General

All materials, both plastics and glass as well as solutions and growth culture mediums of regular use in the lab, were sterilized under pressure and temperature by autoclave (Amaro 2000; 5510 Model) at 121°C for 20 minutes with 1 kg/cm² pressure. For the procedures that required centrifugation, Eppendorf 5415E or Eppendorf 5414D (Eppendorf™, CITY) microcentrifuges were used for volumes equal or less than 2 mL. For volumes higher than 15 mL a Beckman Coulter™ Avanti J-25I centrifuge was used. The specific times, temperatures and speeds vary according to the procedure and are listed for each method.

Absorbance readings were taken with a spectrophotometer Ultrospec 3100, Amersham Biosciences, CITY.

A Sartorius Analytic A210P scale was used for weights below 5 g, whereas for higher weights the scale used was an Acculab Atilon ATL-623 (Sartorius Group). For incubations either a Memmert Modell 500 incubator or a Gallenkamp orbital incubator with agitation were used.

2.1.1 - Microbial strains and plasmids

All the microbial strains and plasmids were obtain from different certified suppliers according to table 7 and stored at -80°C.

Table 7 – Microbial Strains and Plasmids Characteristics.

Microbial Strains and plasmids	Genotype	Reference
<i>E.coli</i>		
DH5 α	F- Φ 80lacZ Δ M15 Δ (lacZYA-argF) U169 recA1 endA1 hsdR17(rk-, mk+) phoA supE44 thi-1 gyrA96 relA1 λ -	Novagen [®]
BL21 (DE3)	F ⁻ ompT hsdS _B (r _B ⁻ m _B ⁻) gal dcm met (DE3)	Novagen [®]
Tuner [™] (DE3)	F ⁻ ompT hsdS _B (r _B ⁻ m _B ⁻) gal dcm lacY1 (DE3)	Novagen [®]
Plasmids		
pUC ₁₈	Amp ^R lacZ Bla ⁺	Gibco BRL
pET-28a	Kan ^R lacI f1oIR T7 His Tag N-terminal	Novagen [®]

2.1.2 - Antibiotics

All antibiotics solutions were dissolved in Milli-Q water and prepared and used according to the information in Table 8.

Table 8 - Stock and work antibiotic solutions.

Antibiotic	Stock Solutions		Working Concentration
	Concentration	Storage	
Ampicillin (Amp ^R)	100-200 mg/mL	4°C (maximum storage time: 2 weeks)	100 μ g/mL
Kanamycin (Kan ^R)	50 mg/mL	4°C (maximum storage time: 2 weeks)	50 μ g/mL

2.2 - Insights into the Cohesin-Dockerin Complex of *Bacteroides cellulosolvens*

2.2.1- Mutants Construction

The recombinant Coh-Doc complexes were coexpressed through an innovative strategy (Brás, 2012) already performed in *B. cellulosolvens* type-I Coh-Doc complexes (Cameron, 2014). The recombinant genes were carefully designed, such that both Coh and Doc encoding genes were under the control of separate T7 promoters and T7 terminators, allowing a restriction driven positioning of either Coh-linked or Doc-linked His6-tag used for complex purification. The constructs were cloned into the two positions (NcoI-XhoI or NheI-SalI) to generate either the Coh or Doc tagged complexes, as showed in Figure 38 and 39. A mutation of residues at positions 22 and 22' on the dockerin (respectively from the first and second repeat Doc sequences), as shown in Table 9, was performed to substitute a Glycine for an Asparagine. These Doc mutants were then synthesized in vitro (NZYTech Ltd, Portugal) and cloned in a pUC₁₈ vector.

Table 9 – Dockerin Amino acid sequences.

Dockerin	Protein Sequence (N –C)
BC2Doc-wt	GDLNGDGVINMADVMILAQSF <u>G</u> KAI GNPGVNEKADLNNDGVINMADAI ILAQYFGK
BC2Doc-m1	GDLNGDGVINMADVMILAQSF <u>N</u> KAI GNPGVNEKADLNNDGVINMADAI ILAQYFGK
BC2Doc-m2	GDLNGDGVINMADVMILAQSF <u>G</u> KAI GNPGVNEKADLNNDGVINMADAI ILAQYF <u>N</u> K

Table 10 - Final Constructs.

	DocCoh _{H6} (Coh)	H ₆ DocCoh (Doc)
Mutant 1 (Mut 1)	Coh-mut1	Doc-mut1
Mutant 2 (Mut 2)	Coh-mut2	Doc-mut2

The mutant's construction strategy consisted of using Doc and Coh recombinant vectors and two mutants inserts transformed in specialized *E. coli* DH5 α cells in order to extract the DNA to verify its quality and create a secure stock to use in the lab. The samples were digested with NheI and BamHI. The success of the restriction was confirmed by Agarose gel Electrophoresis (AE).

DNA fragments were isolated from the AE using a commercial kit GelPure (Nzytech). The target fragments were reconnected with speedy ligase enzyme in order to obtain the results showing on table 10. DH5 α cells transformed with these mutants were cultured in solid LB medium with kanamycin to confirm the success of the transformation. To ensure that the mutation was present on the recombinant vectors Doc and Coh, several colonies were selected for DNA extraction and sequencing.

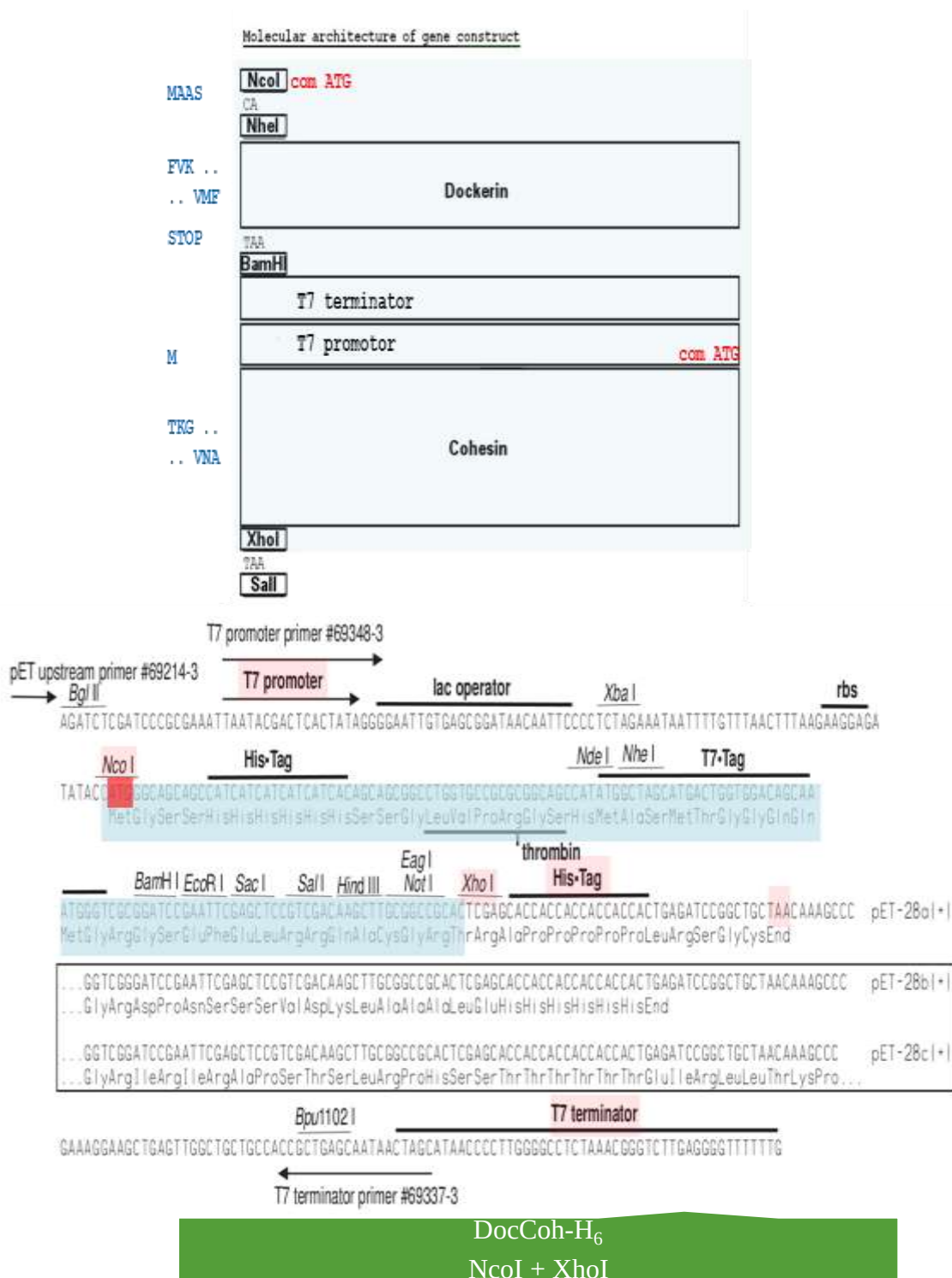


Figure 38 - BC2-Coh molecular construction strategy.

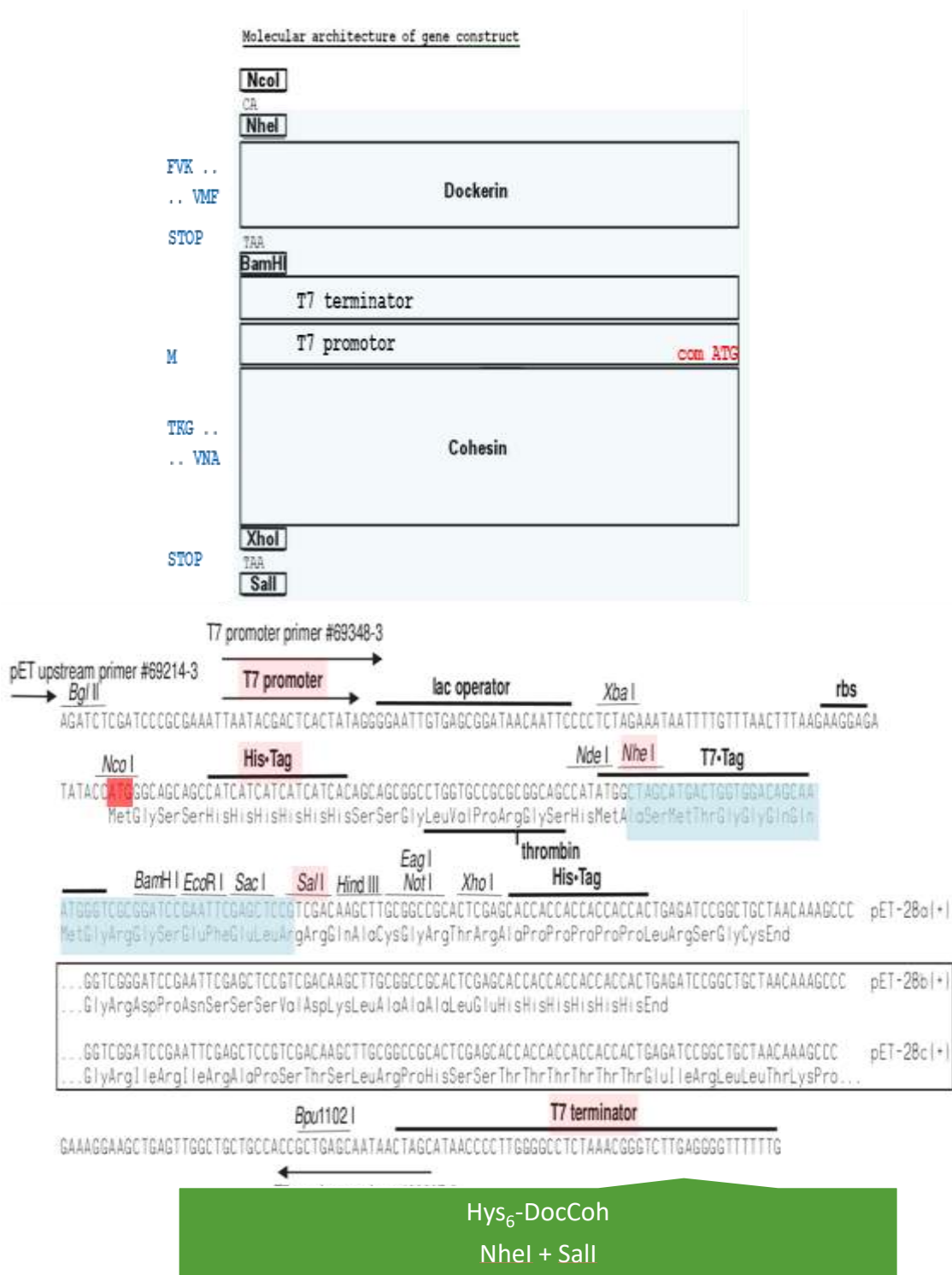


Figure 39 - BC2-Doc molecular construction strategy.

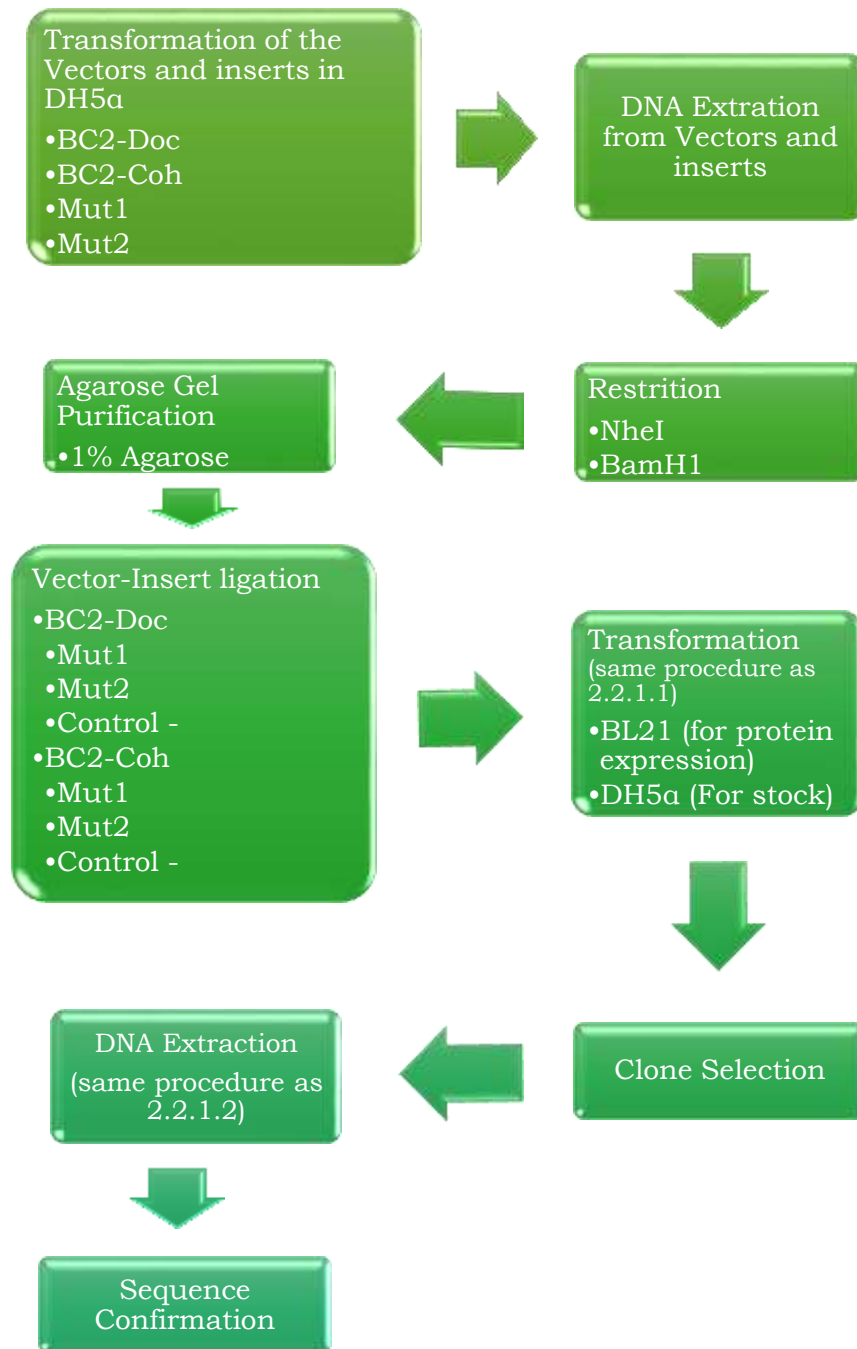


Figure 40 - Overall procedure for the Doc-Coh mutant's construction.

2.2.1.1 - Stock production

In order to have always a fresh batch of transformed recombinant vectors and mutant inserts the samples were transformed using thermal shock methodology according to the following procedure.

DH5 α competent cells were used for the first step, since these are more suitable for DNA extraction compared with BL21 and Tunner cells, which are more used for protein expression. 5 μ l of DNA sample was added for each recombinant vector and mutant insert (Coh, Doc, mut1 and mut2) and to 100 μ l of DH5 α , previously placed on ice. After a gentle mix the cells were incubated for 30 min at room temperature without agitation. In order to force the external DNA to enter the competent cells they were placed at 42 °C for 40 sec and rapidly shifted to ice for 2 min. After adding 1000 μ L of Soc, which has been pre-heated at 37 °C, the cells were incubated during 1 h at 37 °C with agitation at 200 rpm. In order to concentrate the cells and remove the culture medium, the samples were centrifuged 1 min at 500 rpm and 900 μ L of the supernatant was removed. The pellet was then resuspended and 100 μ L were cultivated in solid culture medium. The plates were incubated overnight at 37 °C. The standard procedure is outlined in Appendix C.

Table 11 - Recombinant vector and mutant insert characteristics

Plasmid			Medium
Recombinant Vectors			
Doc	pET 28		LB+Kanamicycn
Coh			
Mutant Inserts			
Mut1	pUC ₁₈		LB+Ampicilin
Mut2			

2.2.1.2 - Vector and insert DNA extraction

All the DNA extractions were performed with a commercial Miniprep kit from NZYTech Ltd, Portugal, according to the following procedure, which is based on three essential steps: bacteria lysis by alkaline hydrolysis, the plasmid adsorption to a column membrane and a final plasmid DNA wash and elution step.

The alkaline hydrolysis occurs with the addition of a solution containing NaOH and SDS, where the SDS is responsible for the solubilization of the phospholipids present in the bacteria cell membrane. The cellular contents is then resuspended by adding Tris-HCl pH 8.0, EDTA and RNase to denature chromosomal and plasmidic DNA. Through the addition of acetate and other caotropic salts, the lysis leads to the precipitation of chromosomal DNA, proteins and bacterial components. While the plasmid stays in solution, the other components precipitate by a centrifuging step.

Plasmid DNA dehydration, caused by the caotropic salts, determines an interaction with the column matrix through the phosphate residues. This matrix is composed of silica or fiber glass resin and successive washes with Tris-EDTA and ethanol with different percentages removes salts and eventual contaminant residues.

The plasmidic DNA is then recovered through its rehydration with 30 μ l of elution buffer and after centrifuging for 1 minute at $11.000\times g$.

After DNA extraction an AE run was performed to confirm the successful cloning results. For these experiments a 1 % (w/v) agarose gel was used. The agarose was dissolved in Tris-borate-EDTA buffer (TBE), and heated near ebullition. After cooling, the staining agent Green Safe Premium was added. The agarose solution was poured on an electrophoresis apparatus. After gel solidification 5 μ l of the DNA plus 2 μ l of xylanocyanol were added to the gel lanes. The gel was submersed in TBE buffer.

The run was performed at 60 V, 300 Amp during 40 min. HyperLadder III (Bioline) was used as a molecular weight marker.

The AE is based on the principle that the migration speed of DNA molecules varies according to its size, which by comparing to a commercial molecular weight marker enables the evaluation of the sample's molecular weight. This gel consists of a complex net where the DNA migrates from the negative pole to the positive by the action of electrical current.

Under a constant current the smallest molecules migrate more rapidly, whereas the largest move slowly and stay closer to the well.

2.2.1.3 - Vectors Restriction maps

After confirming in the AE that the cloning was successful, both recombinant Doc and Coh vectors and the mutated inserts were subjected to endonuclease digestion according to Figure 41 below.

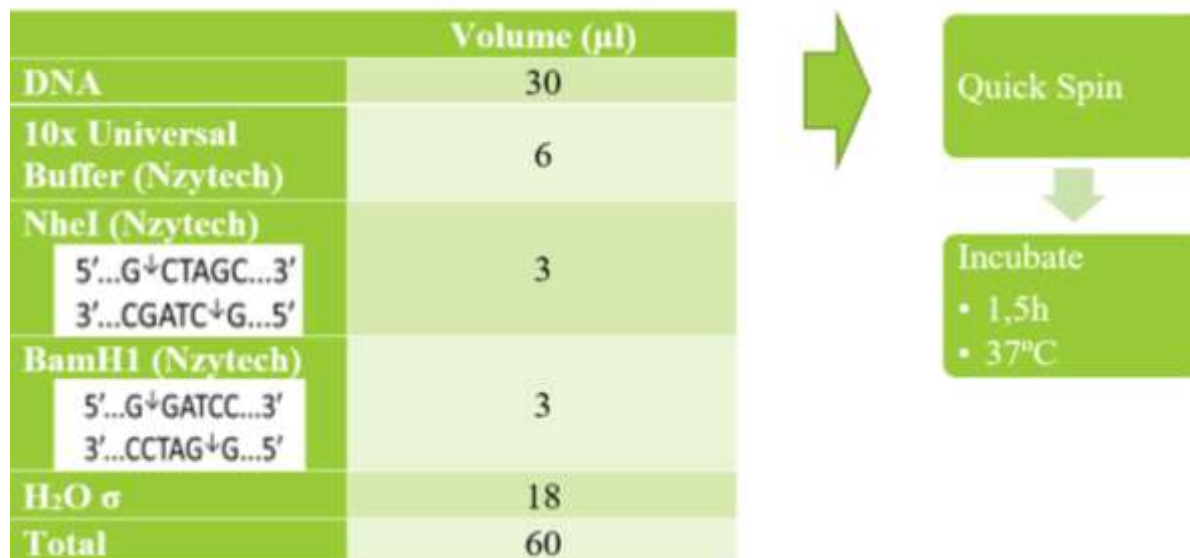


Figure 41 - DNA Restriction Procedure

2.2.1.4 - DNA Purification by AE Gel

AE purification to isolate DNA fragments was performed with the GelPure Kit from NZYtech. This procedure consists in excising the desired band from the AE. The eppendorf microtube is weighted with and without the band in order to calculate the band weight. 300µL of Binding Buffer was added for each 100 mg of AE band and incubated at 60 °C for 10 min. Since the gel matrix is dissolved through heat, it's possible to separate the different components in the mix. Isopropanol was then added with the same volume as the band weight and placed into the column. After centrifuging for 1 min at 1300 rpm, the liquid was discarded. 500 µL of wash buffer was added and the centrifuging was repeated. To complete the washing stage, an additional 600 µL of wash buffer was added followed by two centrifugations. To finish the overall extraction, the column is placed on a new tube and 30 µL of elution buffer is added. After waiting 1 min at room temperature the sample is

centrifuged for 1 min at 1300 rpm in order to release the DNA from the column. The standard procedure is outlined in Appendix C.

2.2.1.5 - Vector-Insert

Each vector was linked with each insert with speedy ligase from NZYtech according to the proportions presented in the Table below.

Volume (µl)	
Buffer	5
Vector DNA	x
Insert DNA	y
Speedy Ligase (Nzytech)	1
H ₂ O σ	Fill until the max vol
Total	20

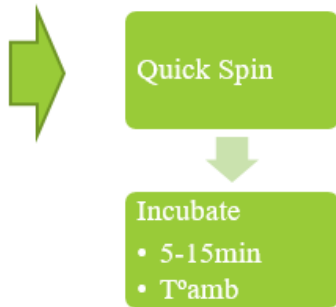


Figure 42 - Vector-Insert assembly procedure

The amount of DNA used was determined according to the equation below

Equation 1 – Vector Insert Proportion for Speedy Ligase reaction

$$\frac{Vector\ weight\ (ng) \times Insert\ Molecular\ Weight\ (kb)}{Vector\ Molecular\ Weight\ (kb)} \times \frac{Insert\ volume}{Vector\ volume} = Insert\ weight\ (ng)$$

A negative control was also performed following the previous description but without DNA that was replaced by an equal volume of water. The negative control should display much less colonies than the reaction (Mutated Clones). The DNA from the mutant clones was sent for Sanger sequencing (LIGHTrun™ from GATC Biotech, Germany).

2.2.2 - Protein Expression Optimization

The resulting positive clones were transformed into different *E. coli* strains, BL21 and Tunner, both specialized in high expression for recombinant proteins (Transformation Procedure 2.2.1.1). Two culture mediums (Luria Bertani-LB and LBE) in different conditions were tested according to Table 12.

The vectors containing the recombinant proteins grew in LB medium with kanamycin at 37 °C till an OD between 0.4 and 0.6. Then the cell culture flasks were placed on ice for 30 min. To induce protein expression, isopropyl β -D-thiogalactopyranoside (IPTG, 1 mM final concentration) was added and this mimics the allolactose, a lactose metabolite that induces transcription by activating the pBAD promoter of the T7 RNA polymerase. This mechanism is based in the Lac operon that is responsible for the lactose digestion in *E. coli*. Originally the Lac operon acts as switch to the production of β -galactosidase and if there is no lactose in the medium, there is no need to produce this enzyme therefore the gene is repressed. In this example lacY mutant hosts are used and that induction is performed with the non-hydrolysable lactose analog isopropyl- β -D-1- thiogalactopyranoside (IPTG) (Baneyx, 1999). The erlenmeyer flasks were then incubated for 16 hours at different temperatures (25 °C, 19 °C and 16 °C) at 150 rpm. The same conditions, with minor changes, were applied for growth in LBE. For growth control the absorbance was evaluated with a 1/10 v/v dilute culture medium sample, until an OD of 1 was reached. Since this is an autoinduction medium there is no need to pause the growth on ice and to add IPTG because the medium itself already contains lactose.

Table 12 - Protein Expression Conditions

Cells Growth Medium	BL21	Tunner	Temperature
LB	LB BL21 25	LB Tunner 25	25 °C
	LB BL21 19	LB Tunner 19	19 °C
	LB BL21 16	LB Tunner 16	16 °C
LBE	LBE BL21 25	LBE Tunner 25	25 °C
	LBE BL21 19	LBE Tunner 19	19 °C
	LBE BL21 16	LBE Tunner 16	16 °C

In order to evaluate the expression level of the recombinant proteins in different conditions, a sample of the recombinant protein cells was taken immediately before adding IPTG and after 16 hours of protein expression.

2.2.2.1-Protein extraction

In order to isolate the recombinant protein from the cells and separate them from the culture medium, the samples were centrifuged for 15 min at 5000 rpm. The supernatant was discard and the pellets were frozen to help the resuspension with 10 ml of 10 M imidazol. To disintegrate the cells and extract the protein, the samples were ultra-sonicated (Bandelin Sonopuls HD 2070) for 10 min at 70 % . Then to separate the soluble proteins, the samples were centrifuged for 30 min at 13.000 rpm. Both phases were stored for SDS PAGE and Native Gels analyzes.

2.2.2.1 - Protein Purification

The protein extract was then filtered with a 0.45 µm membrane (Life Sciences, city) coupled to a 10 cm³ syringe. To continue the process, the filtered extract was purified by immobilized metal ion affinity chromatography (IMAC) using Sepharose columns charged with nickel (HisTrap™, GE Healthcare, Sweden) following conventional protocols (Venditto *et al.*, 2015). This protein purification method consists in running the protein extract through a nickel containing column, since the recombinant protein has a His-tag that binds it to the nickel present in the column, thus retaining it. Through several washes with increased concentrations of imidazole, which competes with the nickel for the histidines in the resin, almost all of the impurities present on the extract are removed. The final washing buffer has a higher concentration of imidazole (300 mM) promoting the target protein elution in aliquots of 1 mL.

The collected samples were evaluated for the presence of protein through a quick qualitative Bradford test. The purity and expression were evaluated by polyacrylamide gel electrophoresis (PAGE). This procedure sorts the protein by passing it through a polyacrylamide matrix. The proteins (14 µl) were denatured by adding 6 µl SDS (Sodium Dodecylsulfate) that promotes leveling of all the electrical charges, and by boiling the protein samples. Therefore the proteins were able to migrate through a constant electric field according only to their molecular weight (MW).

For this procedure a 14% SDS PAGE (Appendix B) with Low Molecular Weight (LMW) Protein Marker (#MB0820, NZYTech, Portugal) as a molecular weight marker, were used. Some samples were also tested through 10% Native PAGE that uses the same principles as an SDS PAGE but without reducing and denaturing the samples, thus maintaining the protein's secondary structure and native charge density. For these gels Bovine Serum Albumin (BSA) was used as a marker.

After each run the gels were stained for 15 min with Coomassie Blue Staining Solution (Sigma) and then revealed with a bleaching solution containing methanol and acetic acid (Appendix B). The gel pictures were acquired with an Image Master VDS (Pharmacia-Biotech).

2.2.2.2 - Optimal growth conditions selection

All SDS PAGE were analyzed to verify the appropriate growth conditions, host cells, culture medium and temperature, in order to proceed to a larger scale production.

2.3 - Biochemical Characterization of *Ruminococcus flavefaciens* putative CBMs Rf2 and Rf4

Previous work regarding the *Ruminococcus flavefaciens* genome, allowed to scan and identify CBM sequences. These CBMs were cloned and expressed by high throughput methodologies (Venditto, Centeno, Ferreira, Fontes, & Najmudin, 2014). Two CBMs were chosen for detailed work; Rf2 and Rf4 as described in 2.3.1. These CBMs were then expressed and purified in order to evaluate the affinity of these proteins. The different substrates (Table 12) were dissolved in water to a final concentration of 0.3% (w/v) and incorporated in 10 % native gels. The affinity was qualitatively measured as “no affinity, - ” and affinity, (+)”

Table 13 – Substrates tested for the initial essay

Substrates	
Celluloses	HEC
	Lichenan
	Curdlan
Xylans	Xylan
	Wheat Arabinoxylan
	Glucurono Xylan
	Birchwood Xylan
Hemicelluloses	Xyloglucan
	Mannan
	Galactomannan (Megazyme)
	Arabinogalactan
	Galactan Lupin
	Arabinan
	Glucomannan
Pectins	Pectic Galactan
	Polygalacturonic Acid
	Potato Rhamnogalacturonan
Others	Pectin from Apples

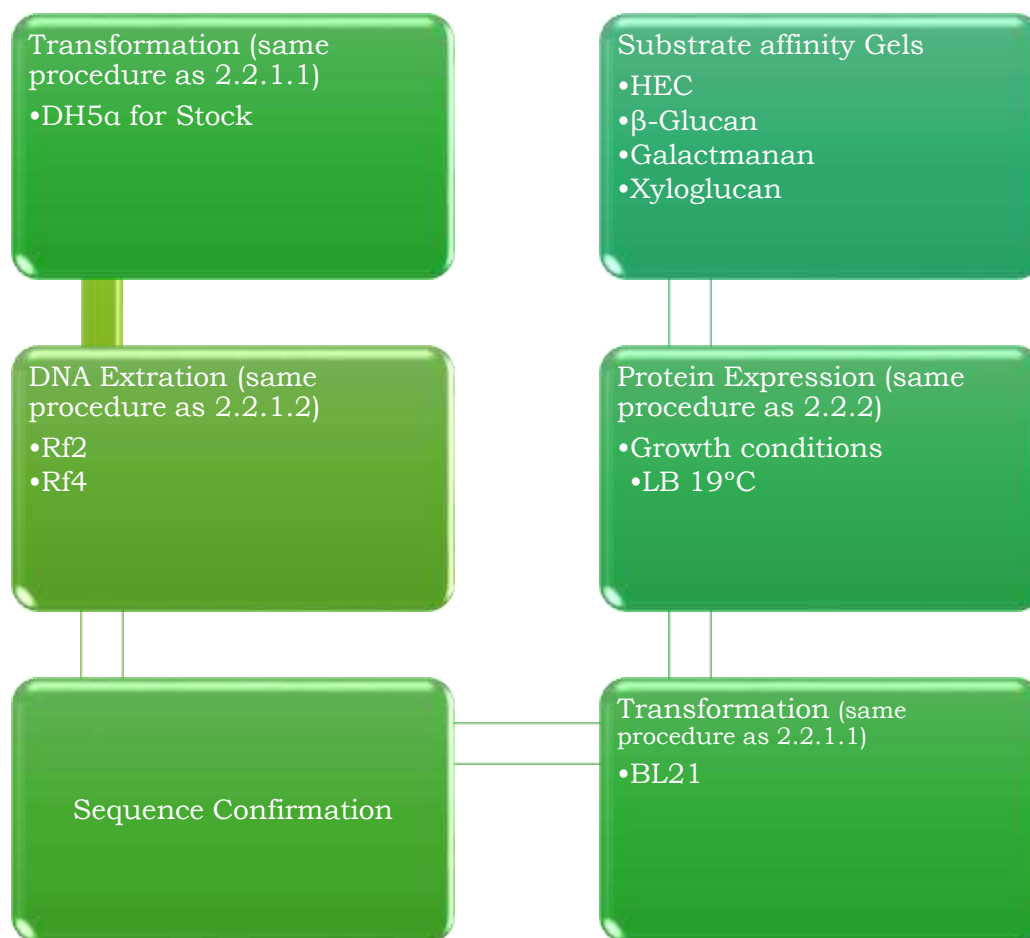


Figure 43 - Overall procedure for the biochemical characterization of Rf2 and Rf4.

2.3.1 - Macromolecule production

The gene encoding the putative CBM-Rf2 (residues 495–621 of *RfCel5*) and CBM-Rf4 were synthesized (NZYTech Ltd, Portugal) with codon usage optimized for expression in *E. coli*. The synthesized gene, containing engineered *NcoI* and *XhoI* restriction sites at the 5' and 3' ends, respectively, was subsequently subcloned into the pET-28a expression vector, generating the p*Rf2* and p*Rf4* constructs. Recombinant CBM-*Rf2* and CBM-*Rf4* contained a C-terminal His₆-tag (HHHHHH) (Table 13).

Table 14 – Rf2 and Rf4 macromolecule production information.

Source organism:	<i>Ruminococcus flavefaciens</i> FD-1
Forward primer:	CACACACCATGGGAGACGGTTACACAATCAAG
Reverse primer:	CACACACTCGAGCTTGAGAACTACAGAGTC
Cloning vector:	pET28a
Expression vector:	pET28a
Expression host:	<i>E.coli</i> BL21
Complete amino acid sequence of the construct produced CBM-RF2	ECHGYLNRSNLTWYTESEPVVNK MM EVLGVSSSNPPTTTASTPSGN DTTTTTAAEEEDTAILYPFTISGNDRNGNFTINFKGTPNSTNNGCIGYS YNGDWEKIEWEGSCDGNGNLVVEVPMSKIPAGVTSGEIQIWWHSG DLK MT DYKAGSGSSQTNTTPQQTTNNNTTVTTAKNDOQPQ HHHH <u>HH</u>
Complete amino acid sequence of the construct produced CBM-RF4	FLEGPYELDASKEKTYQNTTPGGDGEVEWSQLEGKEVAIKFTGSTPV LCFSDASYGGWTE MM KPYDIDKENGIAYYNMAKVDPDLWGDDPTTIA HMQAKTPKLTTVESVNILAAPEGEIKEPEATSKIKKINLKDAKNEDTL YVNLEGAPSTKTNGALGF MM KGDEWTQIEWSGSTDADGKLTVEIPLA DAVVGGTVEFQIWAGFKDLVDKDYISV HHHHHH
The cloning artefacts are underlined.	

2.3.2 - Substrate Affinity Gels

The first application of affinity gel electrophoresis (AGE) on CBMs was in 2000 (Tomme *et al.*, 2000). Nowadays this technique is widely used to study CBM interactions with a variety of polysaccharides such as α -glucan, β -glucan, mannan, xylan, and various pectins. This method is based on the embedding of soluble polysaccharides into a polyacrylamide matrix, where the electrophoretic mobility of a CBM in the polysaccharide-infused gel is compared with its mobility in a native polyacrylamide gel. The interaction of the CBM with the polysaccharide results in a complex of larger size and thus reduced mobility in the polysaccharide containing gel, providing a rapid and convenient readout for binding (Figure. 49). Bovine serum albumin is often used as a negative control due to its lack of affinity for carbohydrates (Abbott & Boraston, 2012).

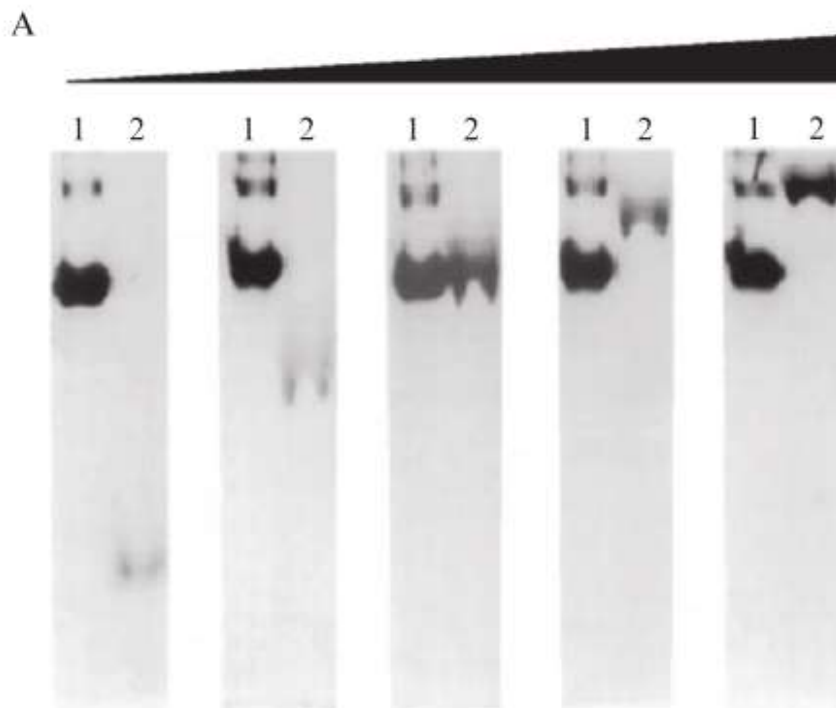


Figure 44 – Typical AGE. The polyacrylamide matrix gel is embedded with soluble polysaccharides, and the electrophoretic mobility of a CBM in the polysaccharide-infused gel is compared with its mobility in a native polyacrylamide gel. The closer the sample is to the well, the higher the affinity to the substrate is. Lane 1 – BSA as marker, Lane 2 – Substrate sample (Abbott & Boraston, 2012).

Since these CBMs had affinity mainly for cellulose hydroxyethyl (HEC) and xyloglucan substrates, an intense study for these substrates was performed in order to quantify the binding affinity.

Table 15 - Substrates concentrations.

Substrates	HEC (%)	B-Glucan (%)	Galactomannan (%)	Xyloglucan (%)
Concentrations	0.300	0.300	0.300	0.300
	0.200	0.200	0.200	0.200
	0.150	0.150	0.100	0.100
	0.100	0.100	0.075	0.050
	0.075	0.075	0.050	0.030
	0.050	0.050	0.040	0.025
	0.030	0.030	0.030	0.010
	0.025	0.025	0.025	0.005
	0.020	0.020	0.010	0.003
	0.010	0.010		0.001
	0.005	0.005		

In order to test all substrates concentrations, a stock solution of 1% (w/v) for each polysaccharide was used. The necessary substrate volume to achieve the target concentration (Table 14) was replaced in the amount of water available in the native gel recipe (Appendix B). The Native Protein Gels were then stained with Coomassie blue for band detection.

2.4 - Insights into the Rf2 CBM structure of *Ruminococcus flavefaciens*

2.4.1 - Overall procedure

In order to determine the structure of a protein it's necessary to promote its crystallization, which requires the exposure of the target protein to multiple suitable crystallization conditions. To reach that stage it is necessary to produce large quantities of good quality and highly pure proteins. The overall procedure is schematized in Figure 45.

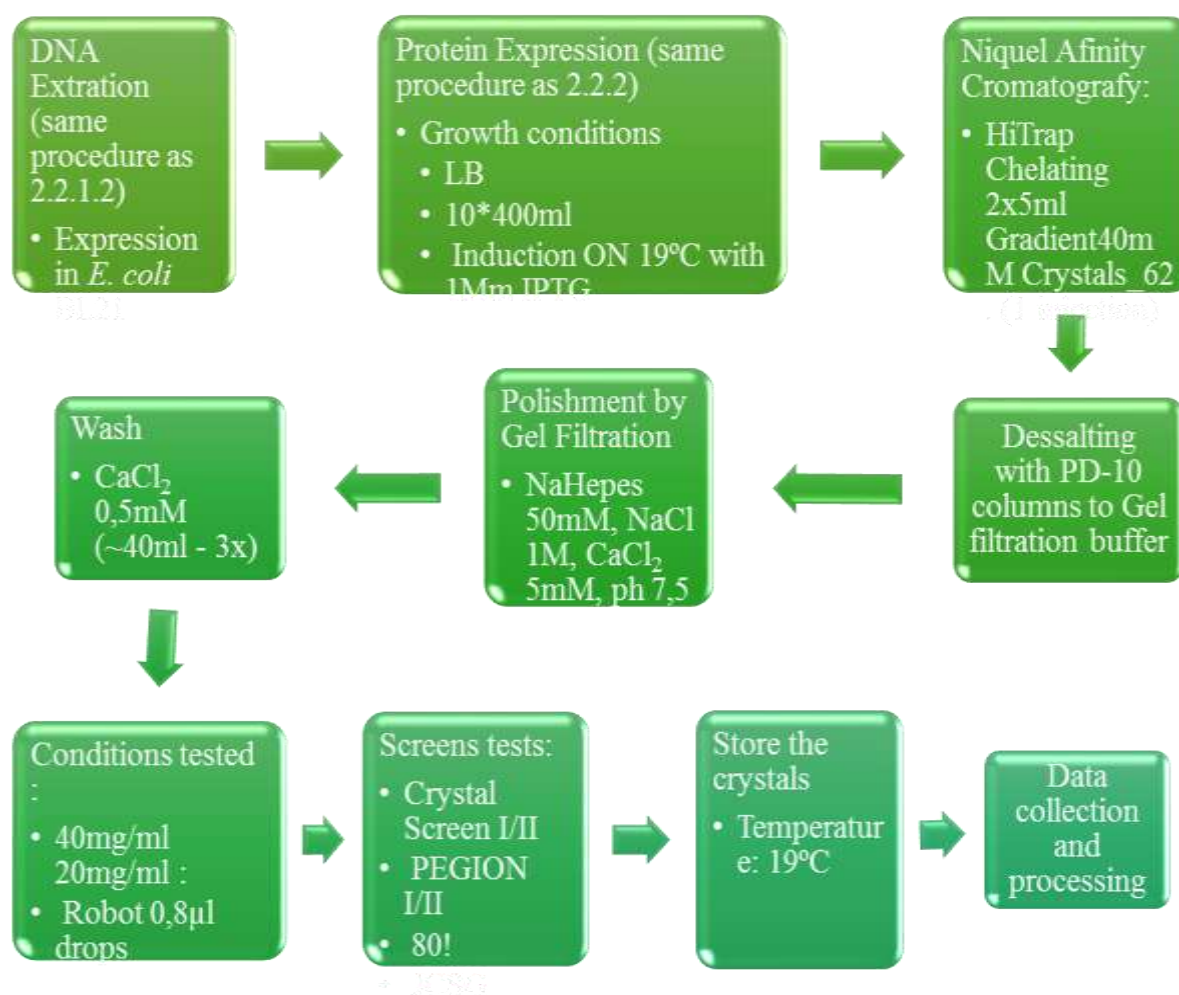


Figure 45 - Schematic representation of the overall Rf2 crystallization procedure.

2.4.1.1 - Protein production

For the crystallization studies, the recombinant CBM-Rf2 was cultured in Luria–Bertani broth (12 flasks of 400 mL of LB) at 37 °C to mid-exponential phase ($OD_{600\text{ nm}}=0.6$) and recombinant protein overproduction was induced by adding IPTG (1 mM final concentration) with incubation for a further 16 h at 19 °C. The His₆-tagged recombinant protein was purified from cell-free extracts by immobilized metal-ion affinity chromatography as described previously (Najmudin *et al.*, 2010). Purified CBM-Rf2 was buffer-exchanged into 50 mM Na-HEPES buffer, pH 7.5, containing 200 mM NaCl and 5 mM CaCl₂, and subjected to gel filtration using a Hi-Load 16/60 Superdex 75 column (GE Healthcare) at a flow rate of 1 ml/min. Purified CBM-Rf2 was concentrated using an Amicon 10 kDa molecular-mass centrifugal concentrator and washed with a 0.5 mM CaCl₂ solution. Recombinant CBM-Rf2, including the C-terminal His₆-tag (LEHHHHHH), has an approximate molecular mass of 15 kDa. The protein concentration was estimated using molar extinction coefficient (ϵ) 30.940 M⁻¹cm⁻¹ at 280 nm with a Thermo Scientific NanoDrop 2000c and verified by SDS-PAGE (Appendix B).

2.4.1.2 - Crystallization

Crystallization conditions were screened by the sitting-drop vapor-phase-diffusion method using the commercial kits Crystal Screen, Crystal Screen 2, PEG/Ion and PEG/Ion 2, and JCSG screen from Hampton Research (California, USA) and an in-house 80 factorial screen using the Oryx8 robotic nanodrop dispensing system (Douglas Instruments) (Table 16). Two drops per well containing 50 μ L of reservoir solution were prepared: one consisting of 0.7 μ L 20 mg mL⁻¹ of CBM-Rf2 and 0.7 μ L of reservoir solution, and one consisting of 1 μ L 20 mg mL⁻¹ and 1 μ L of reservoir solution. All crystals grew to maximum size within a week. The crystals were cryo-cooled in liquid nitrogen after soaking in cryo-protectant (30% (v/v) glycerol added to the crystallization buffer or just Paratone-N) for a few seconds.

Table 16 - Crystallization procedure.

CBM-Rf2	
Method:	Sitting-drop vapour-diffusion
Plate type:	MRC 96-well 2-drop sitting drop crystallization plate (Molecular Dimensions, UK)
Temperature:	19°C
Protein concentration:	20 mg ml ⁻¹
Buffer composition of protein solution:	Deionized water (Sigma), 0.5 mM CaCl ₂
Volume and ratio of drop:	0.7 µL of protein and 0.7 µL reservoir solution.
Volume of reservoir:	50 µL of reservoir

Since it was impossible to crystallize the above construct, after checking the initial results, the sequence was cleaned and part of it was removed, as showed in the following table 16. The rationale for the optimization of this sequence took into account that several threonines in a row are characteristic of linkers, therefore the actual gene of the Rf2 should have started with the indicated alanine. This second construct was called Rf2-mut.

Table 17 - Rf2 protein sequences. The sequences highlighted were removed in order to improve the crystal production and the added His tags are underlined.

Rf2 wt	<u>ECHGYLNRSNLTWYTESEPVVNKMMEVLGVSSSNPPTTTASTPSGNDTTTTA</u> EEEDTAILYPFTISGNDNRNGNFTINFKGT PNSTNNGCIGYSYNGDWEKIEWEGSC DGNGNLVVEVPMSKIPAGVTSGEIQIWWHSGDLKMTDYKA <u>GSGSSQTNTTPQ</u> <u>QTTNNNNNTTVTTAKNDOQPQLEHHHHHH</u>
Rf2 mut	<u>MGAEEDTAILYPFTISGNDNRNGNFTINFKGT PNSTNNGCIGYSYNGDWEKIEW</u> EGSCDGNGNLVVEVPMSKIPAGVTSGEIQIWWHSGDLKMTDYKA <u>LEHHHHHH</u>

2.4.1.3 - Data collection and processing

X-ray diffraction data from suitable protein crystals was collected on beamline PROXIMA-1 at SOLEIL, Orsay, France, with the CBM-Rf2 crystals cooled to -173.15 °C using a Cryostream (Oxford Cryosystems) and a PILATUS 6M detector (Dectris). Precisely 180° of data were collected with a $\Delta\phi$ of 0.1° and an exposure of 0.1 sec and a further 360° using an inverse beam strategy (two equivalent complete interleaved data collections starting at $\phi = 0$ degrees and $\phi = 180$ degrees) at the As-peak edge (at an energy of 11.875 keV, $f' = -8.08$ and $f'' = 8.03$) for a single-wavelength anomalous diffraction experiment for the first crystal (x6) as described previously (Venditto *et al.*, 2015). In addition, 100° of a second higher resolution dataset (to 1.02 Å) was collected for crystal x5. All data sets were processed using XDS (Kabsch, 2010) via the command line interface xdsme (<https://code.google.com/p/xdsme/>) and AIMLESS (Evans, 2006) from the CCP4 suite (Collaborative Computational Project, Number 4, 1994 (Winn *et al.*, 2011)). Data collection statistics (as reported by AIMLESS) are given in Table 18. All the diffracting CBM-Rf2 crystals belong to the orthorhombic space group, with a single molecule in the asymmetric unit, a solvent content of ~34 % and a Matthews coefficient of ~1.85 Å³ Da⁻¹ (Matthews, 1968).

Table 18 - Data collection and processing. Values for the outer shell are given in parentheses.

Dataset	As-edge (x6)	High Res (x5)
Beamline	PROXIMA-1, SOLEIL	PROXIMA-1, SOLEIL
Space Group	P 2 ₁ 2 ₁ 2 ₁	P 2 ₁ 2 ₁ 2 ₁
Wavelength (Å)	1.04408	0.82656
Unit-cell parameters		
a, b, c (Å)	43.39, 45.21, 49.56	43.27, 45.02, 49.45
Resolution limits (Å)	33.40 – 1.29 (1.31 – 1.29)	45.02 – 1.08 (1.10 – 1.08)
No. of observations	88114 (3486)	270624 (12738)
No. of unique observations	25060 (1178)	42191 (2063)
Multiplicity	3.5 (3.0)	6.4 (6.2)
Completeness (%)	99.7 (97.3)	100 (100)
<I/σ(I)>	16.1 (6.5)	12.2 (1.3)
CC _{1/2} [†]	0.993 (0.969)	0.999 (0.759)
R _{merge} [‡]	0.053 (0.123)	0.062 (0.9122)
R _{p.i.m.} [§]	0.032 (0.083)	0.027 (0.396)

Matthews coefficient (Matthews, 1968).

[†] CC_{1/2} = the correlation between intensities from random half - dataset (Diederichs & Karplus, 2013)

[‡] $R_{merge} = \frac{\sum_{hkl} \sum_i |I_i(hkl) - \langle I(hkl) \rangle|}{\sum_{hkl} \sum_i I_i(hkl)}$, where $I_i(hkl)$ is the i th intensity measurement of reflection hkl , including symmetry-related reflections and $\langle I(hkl) \rangle$ is its average.

[§] $R_{p.i.m.} = \frac{\sum_{hkl} \{1/[N(hkl) - 1]\}^{1/2} \sum_i |I_i(hkl) - \langle I(hkl) \rangle|}{\sum_{hkl} \sum_i I_i(hkl)}$, where $\langle I(hkl) \rangle$ is the average of symmetry-related observations of a unique reflection.

Chapter 3 - Results and Discussion

3.1 - Insights into the Cohesin-Dockerin Complex of *Bacteroides cellulosolvens*

3.1.1 - Mutants Construction



Figure 46 – 1% AE of samples (lane 1 - marker, HyperLadder III; lane 2 - mut1; lane 3 - mut2; lane 4 - DocCoh_{H6} (Coh); lane 5 - _{H6}DocCoh (Doc)). The bands marked in brown represent the mutated dockerin and the bands marked in green the cohesins.

Upon AE separation, mutants purification was performed by excision of the lower molecular weight bands for the mutated doc and the high molecular weight bands for the wildtype Doc-Coh complex, marked in brown and green, respectively, on figure 46. To simplify when referring to the Doc-Coh complex with the histidine tag on the cohesin we will say Coh and when the histidine tag is on the Dockerin we will use Doc. For the purification protocol the bands were cut and weighted, resulting on the following values: Mut1-155 mg, Mut2-153 mg, Coh-80 mg and Doc-124 mg.

After purification, the samples were evaluated and quantified by spectrophotometry using NanoDropTM in order to determine its concentration and evaluate its purity. As shown in Table 19

Table 19 – Concentration and purity evaluation.

	Mut1	Mut2	Coh	Doc
[DNA]ng/ml	11.9	12.8	23.2	40.5
260/280	2.07	1.99	1.89	1.90

Regarding its purity, the samples were within the normal ranges and despite the lower DNA concentration, sequence analysis confirmed the successful construction of all four mutants.

3.1.2 - Protein Expression

After analyzing all conditions through AE and taking into account the difficulties in expressing the dockerin mutants it was concluded that for the mutants with the His-tag in the cohesin there weren't significant changes in the protein expression according to cells, temperature or growth medium.

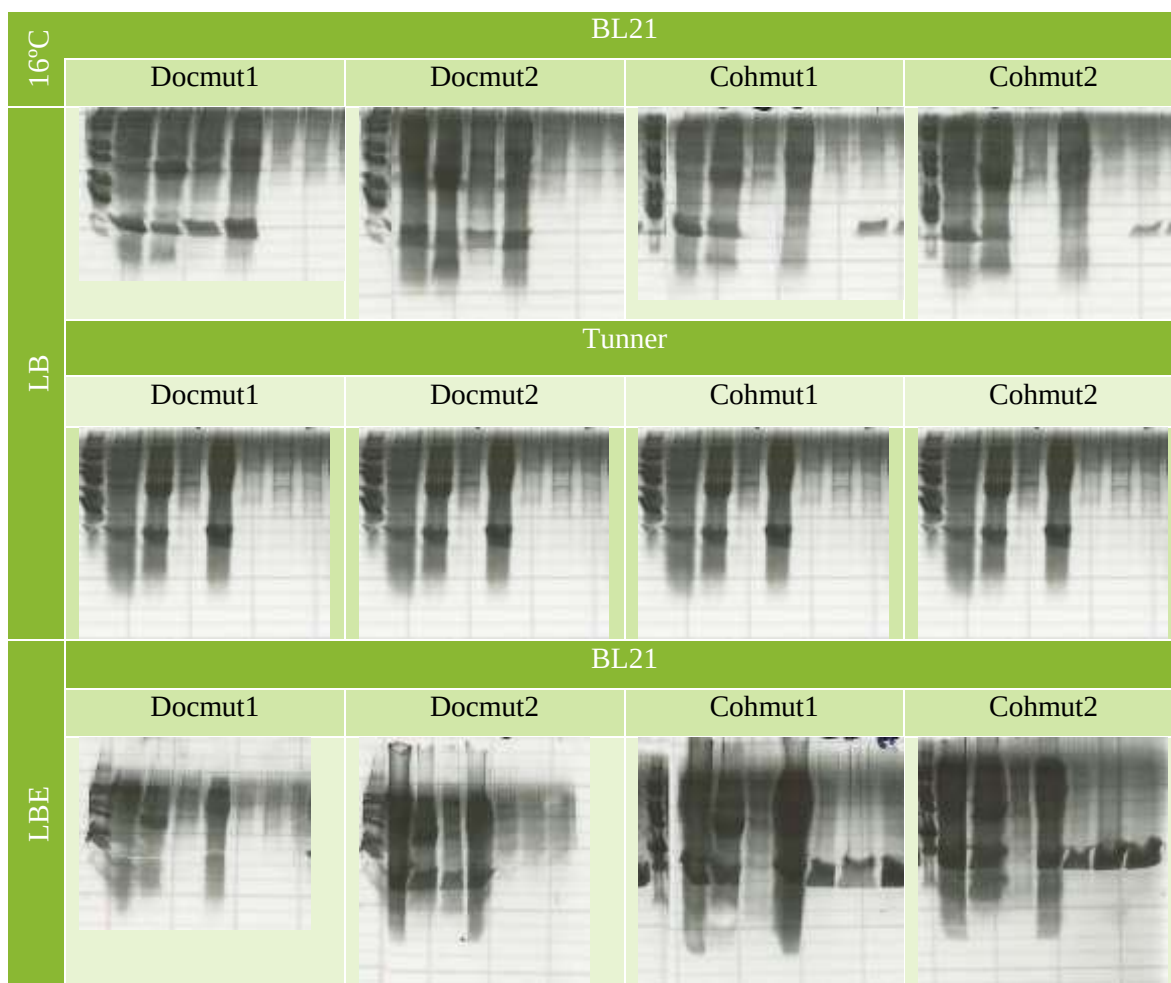


Figure 47 – Growth Condition at 16°C- Coomassie Brilliant Blue-stained 14% SDS–PAGE gel and 10%Native gel for evaluation of protein expression. Lane 1- molecular-mass markers, Lane 2- Soluble Protein; Lane 3- Insoluble Protein; Lane 4- Soluble Extract; Lane 5-Wash 1; Lane 6- Wash 2; Lane 7 and 8- Fractions 2 and 4.

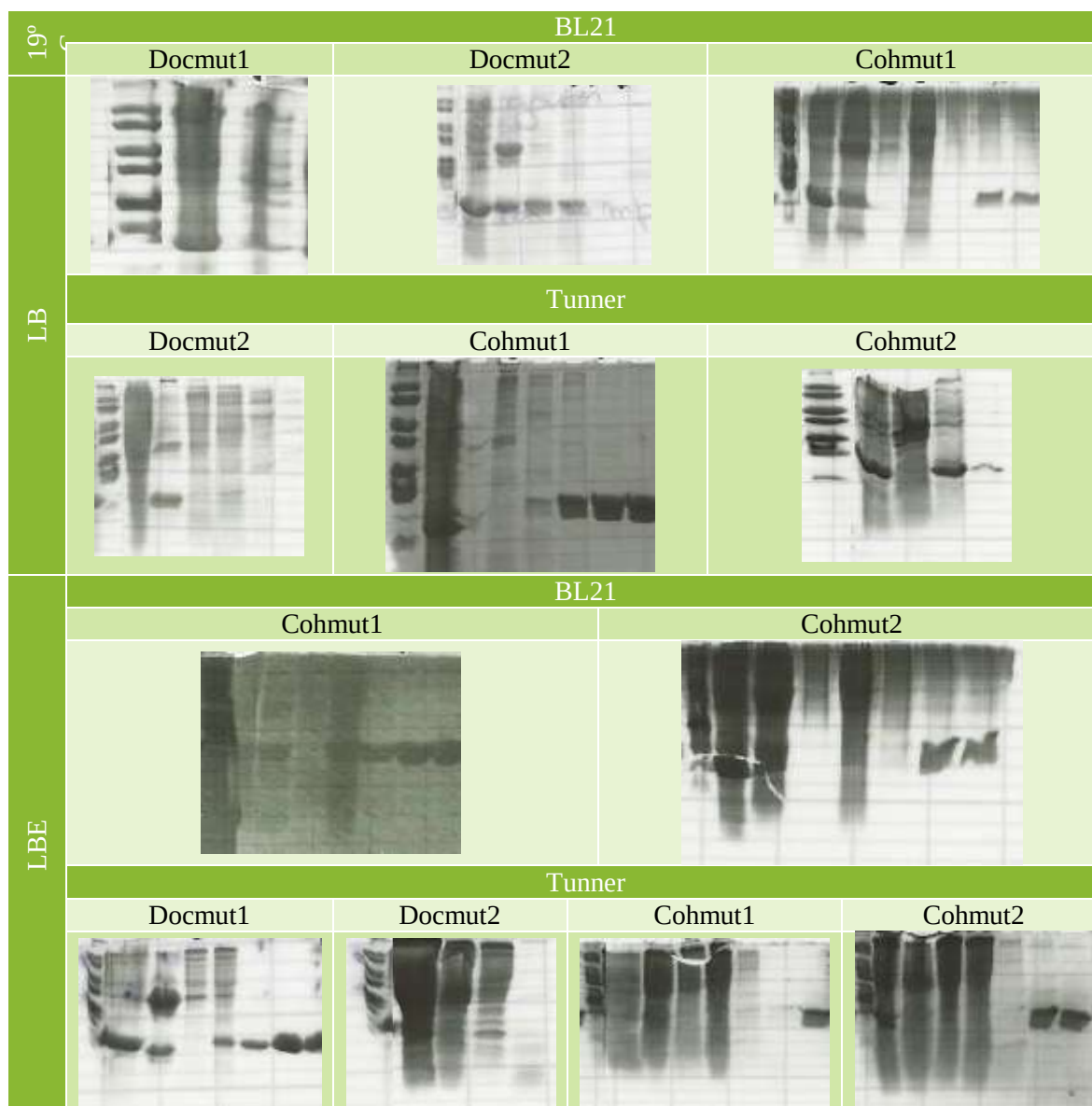


Figure 48 - Growth Condition at 19°C- Coomassie Brilliant Blue-stained 14 % SDS-PAGE gel and 10 % native gel for evaluation of protein expression. Lane 1- molecular mass marker, Lane 2 - Soluble Protein; Lane 3 - Insoluble Protein; Lane 4 - Soluble Extract; Lane 5 - Wash 1; Lane 6 - Wash 2, Lane 7 and 8 - Fractions 2 and 4.

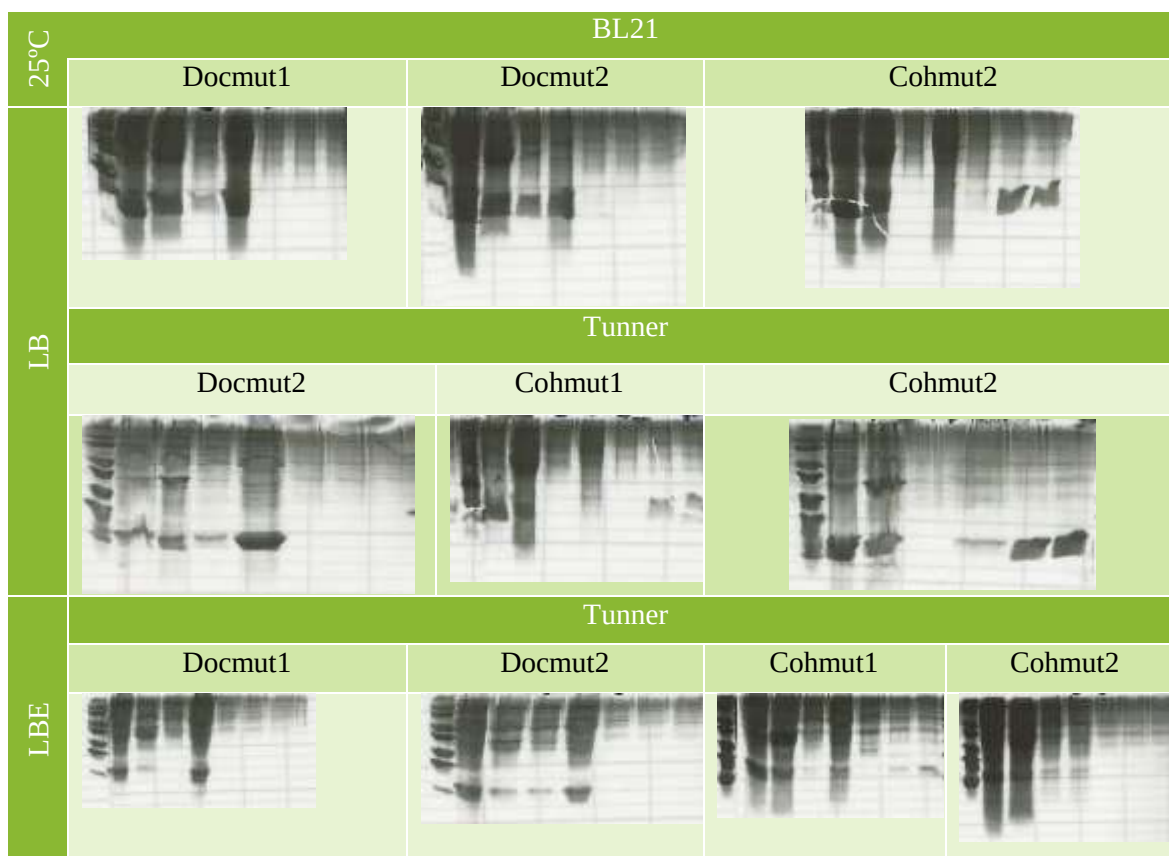


Figure 49 - Growth condition at 25°C- Coomassie Brilliant Blue-stained 14 % SDS-PAGE gel and 10 % native gel for evaluation of protein expression. Lane 1- molecular mass marker, Lane 2 - Soluble Protein; Lane 3 - Insoluble Protein; Lane 4 - Soluble Extract; Lane 5 - Wash 1; Lane 6 - Wash 2, Lane 7 and 8 - Fractions 2 and 4.

Unfortunately the dockerin was only expressed in one LB condition with induction at 25°C with BL21, and slightly on LBE 19° C with Tunner as showed in Figures 48 and 49

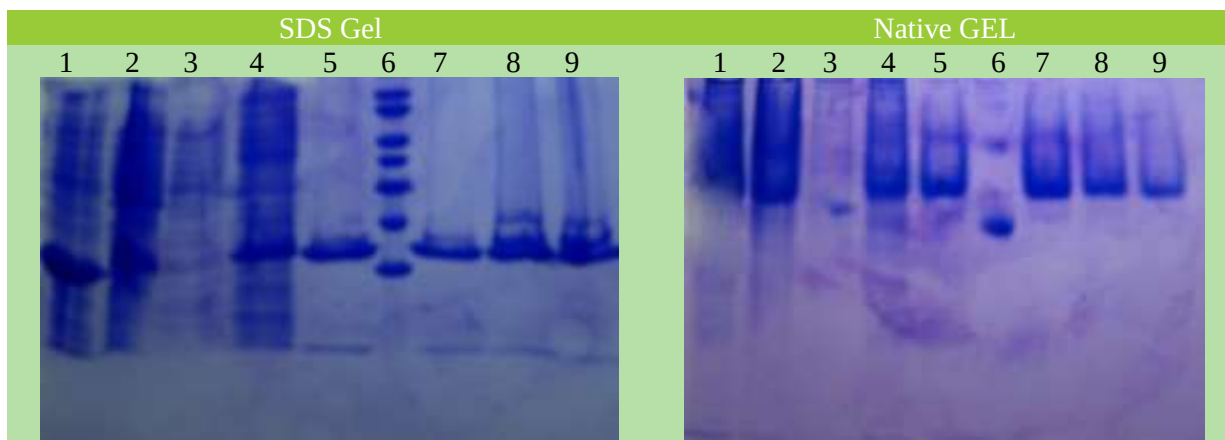


Figure 50 – A Coomassie Brilliant Blue-stained 14 % SDS-PAGE gel and 10 % native gel for evaluation of protein expression. Lane 1 - Soluble Protein; Lane 2 - Insoluble Protein; Lane 3 - Soluble Extract; Lane 4 - Wash 1; Lane 5 - Wash 2; Lane 6 - molecular-mass markers (NZYTech Ltd, Portugal) – SDS and BSA – Native ; Lane 7 - Fraction 2; Lane 8 - Fraction 4; Lane 9 - Fraction 6.

It is well known that independent dockerin expression *per se* is quite problematic as this protein is highly unstable in such conditions. As such, *in vitro* formation of the Coh-Doc pair allows the production and purification of the complex, thus stabilizing the Doc module. The dual binding mode capability of type-II Cohesin-Dockerin also causes its instability, hindering the necessary crystallization step and the pursuit of more information about these proteins function and structure.

A future line of work to continue this project aiming protein crystallization would thus entail growing our protein samples in LBE 19°C or LB 25 °C with BL21 which gave the best results for both Doc and Coh. We could also try growing them at 37 °C which is a normally good growth condition.

3.2 - Biochemical Characterization of *Ruminococcus flavefaciens* putative CBMs Rf2 and Rf4

3.2.1 - Substrate affinity Gels

Previous work regarding the whole *Ruminococcus flavefaciens* genome, allowed scanning and identifying all the CBM sequences. These CBMs were cloned and expressed by high throughput methods (Venditto *et al.*, 2014). These clones were then tested for general substrate affinity and table 20 has the summary results.

According to these results the Rf2 and Rf4 were chosen for further work due to their range of substrate affinity. By analyzing these partial results we can affirm that both Rf2 and Rf4 have affinity for cellulose and hemicellulose. Results showed affinity to HEC, lichenan xyloglucan, glucomannan and galactomannan, this last one only for Rf2. In order to determine the extent of this affinity we selected those substrates, replacing the lichenan with β -glucan due to the difficulty to dissolve the former substrate and not testing the glucomannans at all.

Table 20 - Substrate Affinity for several CBMs. All substrates were concentrated at 0,3 % w/v in native gels of 10 %. The affinity was qualitatively measured as “no affinity(-)” and affinity (+)”. For those where this distinction was not so obvious $\pm$ is used.

Substrates		Rf2 (P2)	Rf4 (P4)	Rf6 (P6)	Rf10 (P10)	Rf13 (P12)	Rf18 (P16)	Rf19 (P17)
	HEC	+	+	+	+	-	-	-
	Lichenan	+	+	+	-	-	-	-
	Curdlan	-	-	-	-	-	-	-
Xylans	Xylan	-	-	$\pm$	$\pm$	$\pm$	-	-
	Wheat Atabinoxylan	-	-	-	-	-	-	-
	Glucurono Xylan	-	-	-	-	-	-	-
	Birchwood Xylan	-	-	-	-	+	-	-
Hemicelluloses	Xyloglucan	+	+	+	-	-	+	+
	Mannan	-	-	-	-	-	-	-
	Galactomannan megazyme	+	-	-	-	-	-	-
	Arabinogalactan	-	-	-	-	-	-	-
	Galactan Lupin	-	-	-	-	-	-	-
	Arabinan	-	-	-	-	-	-	-
	Glucomannan	?	+	+	+	$\pm$	+	+
Pectins	Pectic Galactan	-	-	-	-	-	-	-
	Polygalacturonic Acid	-	-	-	-	-	-	-
	Rhamnogalacturonan Potato	-	-	-	-	-	-	-
Others	Pectin from Apples	-	-	-	-	-	-	-

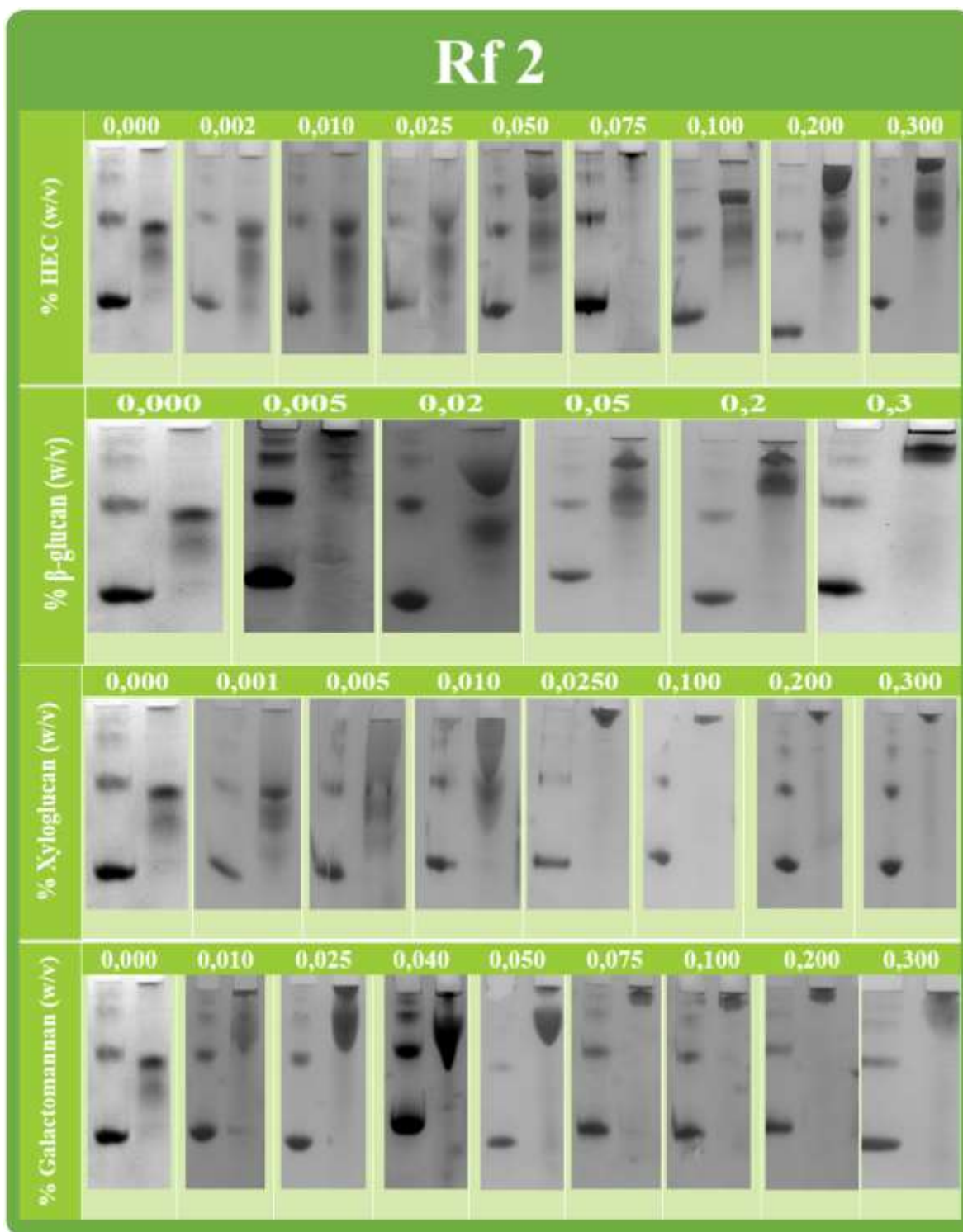


Figure 51-AGE for RF2. First band in each gel corresponds to BSA, used as control and the second band corresponds to Rf2.

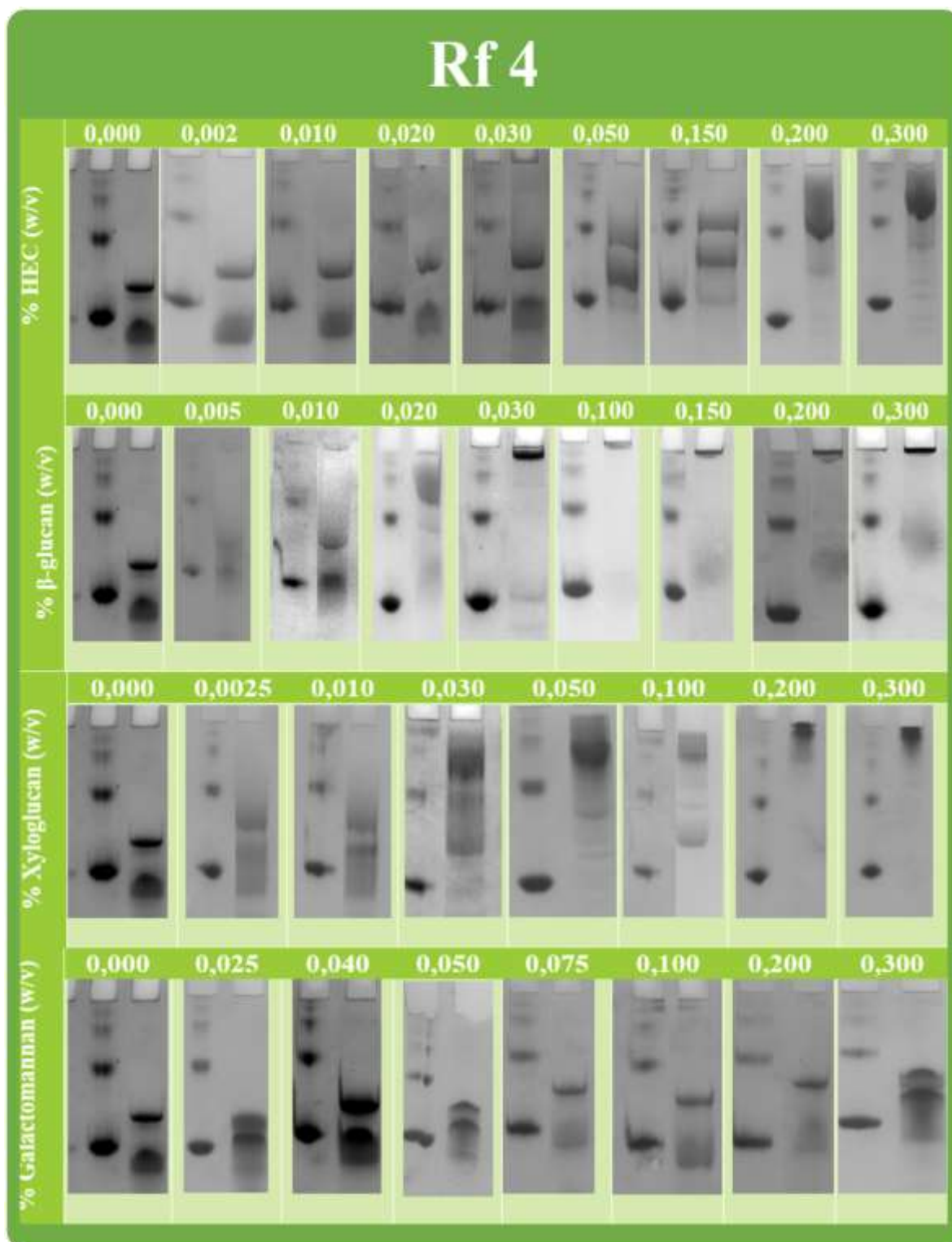


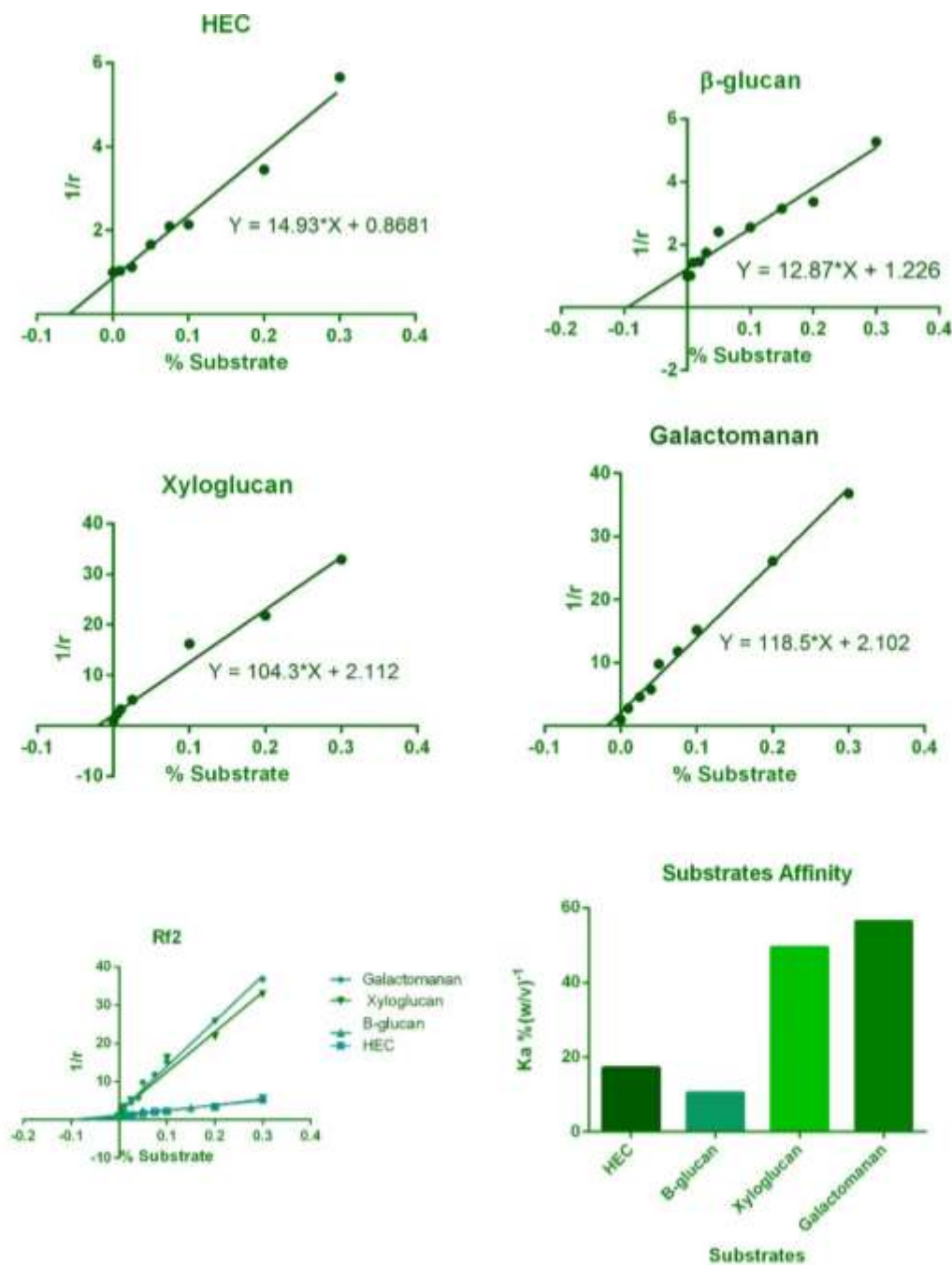
Figure 52 - AGE for Rf4. First band in each gel corresponds to BSA, used as control and the second band corresponds to Rf4.

Just by analyzing the AGE we already observe some significant differences between the different substrates for these CBMs.

For Rf2 it is clear that all substrates start to show some affinity around 0.05 % (w/v) however for xyloglucan and galactomannan that affinity is evident already at 0.005 % (w/v) and 0.01 % (w/v) respectively. Based on these results, we would conclude that substrate affinity from highest to lowest would be xyloglucan, galactomannan, HEC and β -glucan.

For Rf4 the results are not as similar between each other as for Rf2. Rf4 almost doesn't have any affinity to galactomannan and HEC binding is only evident at 0.2 % (w/v). Xyloglucan shows better results with the affinity showing from 0.05 % (w/v). However β -glucan results surpass the Rf4 affinity to any substrate, starting to show evidence of affinity at 0.02 % (w/v).

Nevertheless, in order to quantify the affinity to the substrate, the $1/R$ method was used, where R correspond to the normalized distance between the band migration of the sample compared to the control, where there is no substrate added to the gel. The dissociation constant (K_d) is the substrate percentage (w/v) to which the $1/R$ is zero, therefore the affinity constant can be obtained by the inverse of K_d . In Appendix D there are detailed tables from which we obtained the following graphs.



Ka	HEC	B-glucan	Xyloglucan	Galactomanan
$\%(w/v)^{-1}$	17.19454	10.49436	49.37611	56.36702

Figure 53 – Rf2 Substrate Affinity analysis. From these results we observe that Rf2 displays very low affinity levels with better results for galactomanan and xyloglucan.

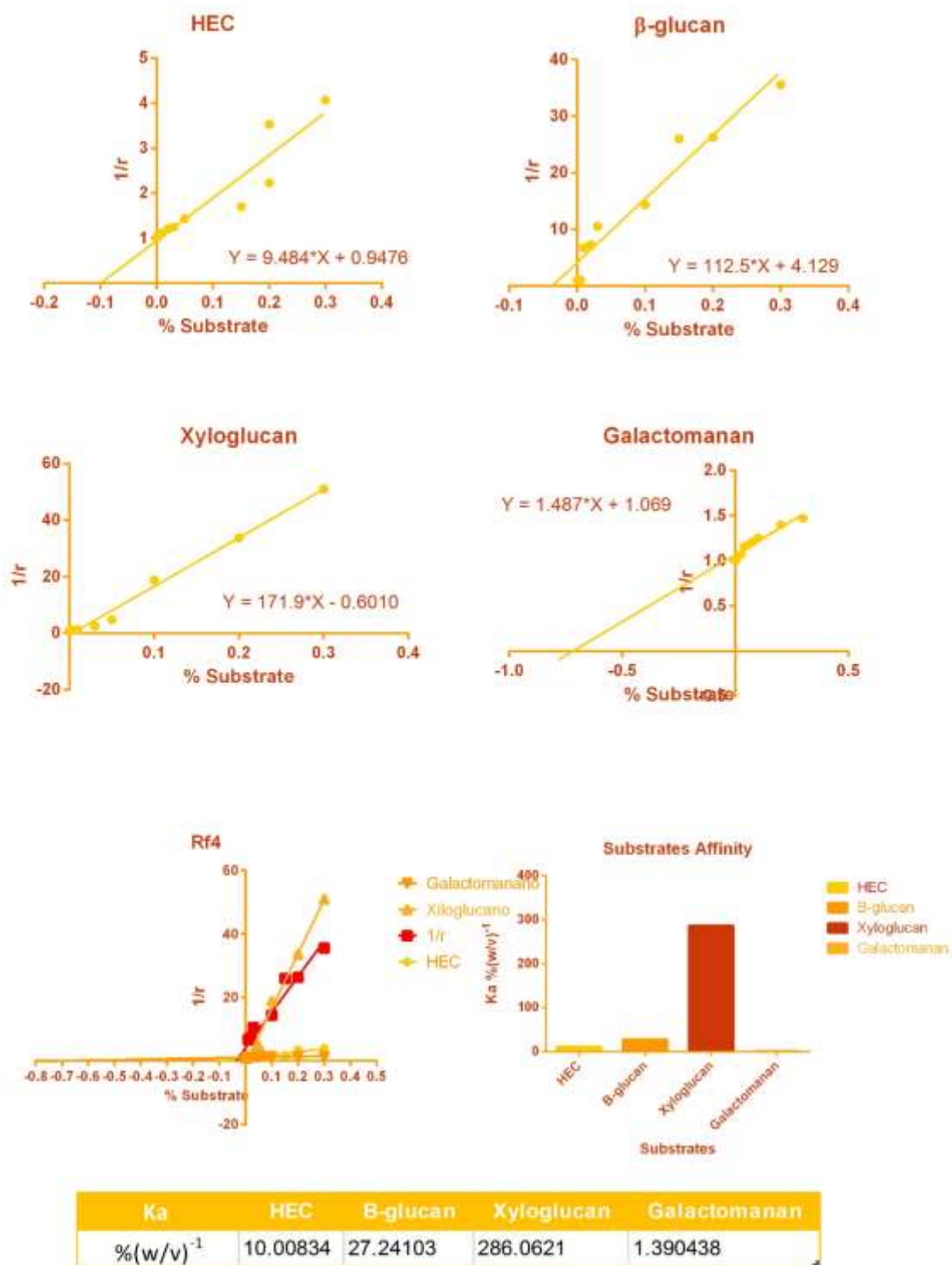


Figure 54 - Rf4 substrate affinity analysis. From these results we observe that Rf4 presented very low affinity levels with significantly higher results for xyloglucan.

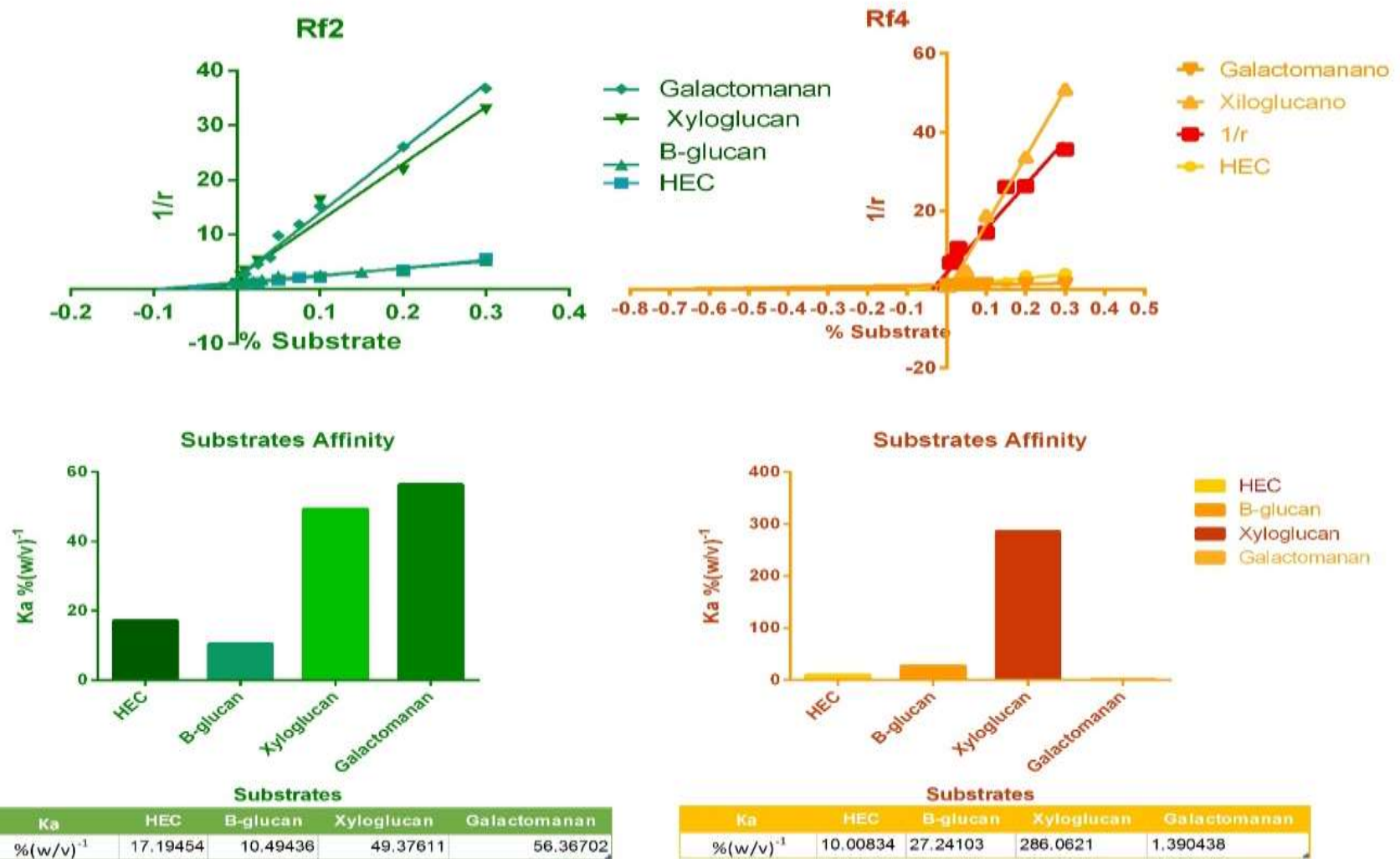


Figure 55 - Rf2 and Rf4 substrate affinity comparison. From these results we observe that Rf4 presented very low affinity levels with significantly higher results for xyloglucan, whereas Rf2 shows a low affinity for HEC and a higher affinity for galactomannan.

By analyzing both CBM results we have very similar affinities in the lower ranges for HEC and β -glucan for both Rf2 and Rf4. However for the remaining substrates we have clearly two distinct affinity results, whereas Rf2 shows a low affinity for HEC and a higher affinity for galactomannan, Rf4 presents a really low, almost null, affinity for galactomannan. Xyloglucan has the second best affinity for Rf2 and displays an extremely high affinity for Rf4.

From this we can determine that Rf2 presents higher affinities for galactomannan and xyloglucan ranging from 56 % to 49 % (w/v)⁻¹, respectively, and Rf4 only has affinity for xyloglucan with a significant K_a of 296 % (w/v)⁻¹.

Comparing the preliminary results we had from AGE and the actual quantification, we can conclude that there are no significant differences between both methods. However it is important to refer that without the calculus for the affinity value we would wrongly attribute the highest affinity of Rf2 to xyloglucan instead of galactomannan and we would never expect such a significant affinity for xyloglucan from Rf4.

3.3 - Insights into the Rf2 CBM structure of *Ruminococcus flavefaciens*

3.3.1 - Sequence Analysis

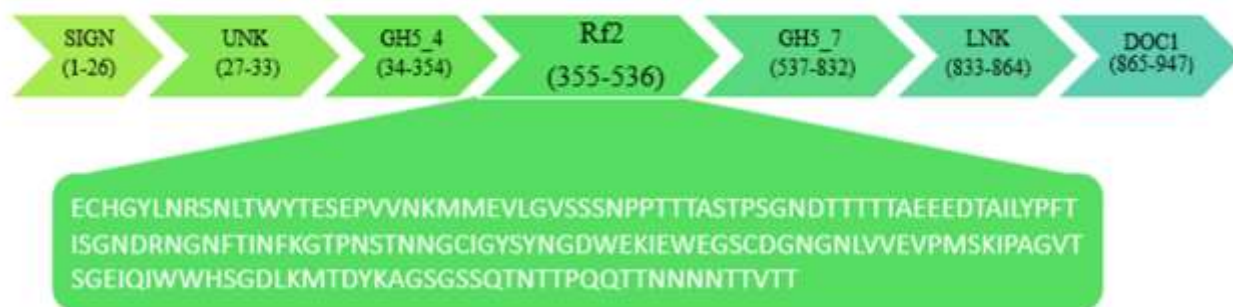


Figure 56 – Rf2 sequence organization.

By analyzing this sequence, and since tandem threonine repeats are characteristic of linkers, therefore the construct should be corrected and the Rf2 gene should start with the alanine residue following the Thr repeat (Figure 56).

ECHGYLNRSNLTWYTESEPVVNKMMMEVLGVSSSNPPTTTASTPSGNDTTTTT

Conserved domain for a metal binding protein TroA_b. These proteins are predicted to function as initial receptors in ABC transport of metal ions. They belong to the TroA superfamily of helical backbone metal receptor proteins that share a distinct fold and ligand binding mechanism.

Figure 57 – Linker sequence analysis. From a Blast analysis of these linker sequences, previously considered as part of the Rf2 sequence we obtained a conserved domain characteristic of metal binding proteins which may explain the difficulties associated with the crystallization process while using the original published Rf2 sequence.

With the correct amino acid sequence, a comparison with other known proteins could provide clues to predict its putative function and classification.

Table 21 – NCBI Blast for Rf2mut (Altschul *et al.*, 1997).

Description		Max score	Total score	Query cover	E value	Identity	Accession GenBank
<i>Ruminococcus flavefaciens</i> 17	Endoglucanase A	85.1	85.1	74%	2.00E-16	44%	CAB0588 1.1
<i>Ruminococcus flavefaciens</i>	Endoglucanase	85.1	85.1	64%	2.00E-16	51%	WP_0285 14028.1
	Endoglucanase	76.6	76.6	60%	2.00E-13	51%	WP_0248 61360.1
	Endoglucanase	76.3	76.3	67%	3.00E-13	43%	WP_0285 17111.1
	Endoglucanase	75.5	75.5	74%	4.00E-13	37%	WP_0372 99526.1
	Endoglucanase Family Protein	65.1	65.1	82%	1.00E-09	34%	WP_0099 86298.1
	Endoglucanase A	58.5	58.5	78%	2.00E-07	29%	WP_0285 19772.1
	Glycoside Hydrolase	57.8	57.8	50%	4.00E-07	39%	WP_0248 61738.1
	Glycoside Hydrolase	54.7	54.7	50%	4.00E-06	36%	WP_0373 01064.1
	Glycoside Hydrolase	53.5	53.5	50%	9.00E-06	38%	WP_0285 16165.1
	Glycoside Hydrolase	53.1	53.1	51%	1.00E-05	37%	WP_0285 17065.1
	Glycoside Hydrolase	48.1	48.1	32%	5.00E-04	44%	WP_0285 20794.1
	Glycoside Hydrolase	47.0	47.0	48%	0.001	36%	WP_0315 59475.1

According to this Blast analysis (Altschul *et al.*, 1990) *Rf2mut* shares an amino acid sequence identity of more than 25% with at least 15 other proteins like endoglucanases and glycoside hydrolases. However the highest score shows more similarities with endoglucanase A from *R. flavefaciens* 17.

3.3.2 - Crystallization

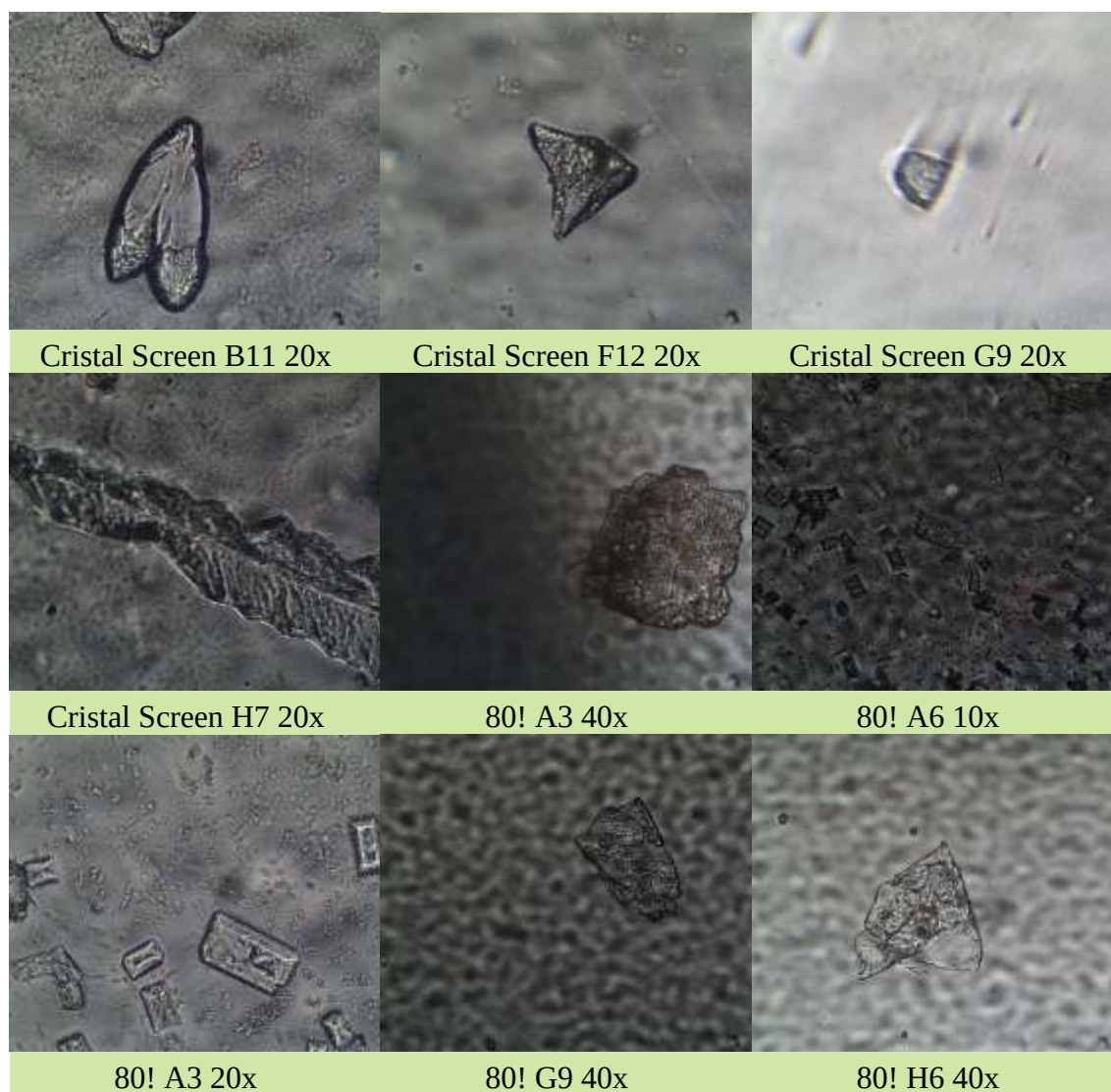


Figure 58 – Rf2 crystals and respective conditions.

From all crystals obtained (Figure 58), we selected four crystals in total, two from Crystal Screen B11 (0.2 M magnesium chloride hexahydrate, 0.1 M HEPES sodium pH 7.5 and 30 % (v/v) Polyethylene glycol 400) and another two from 80! A6 (0.2 M MgCl_2 , 0.1 M Act 4.5 and 30 % PEG 4K). The crystals were cryo-cooled in liquid nitrogen after soaking in cryoprotectant 30 % (v/v) glycerol added to the crystallization buffer.

Data were collected on beamline PROXIMA-1 at SOLEIL, Orsay, France, with the CBM-Rf2 crystals (Figure 2, Table 2) cooled to 100 K using a Cryostream (Oxford

Cryosystems) using a PILATUS 6M detector (Dectris). Diffraction experimental results, however showed that the 4 samples were salt.

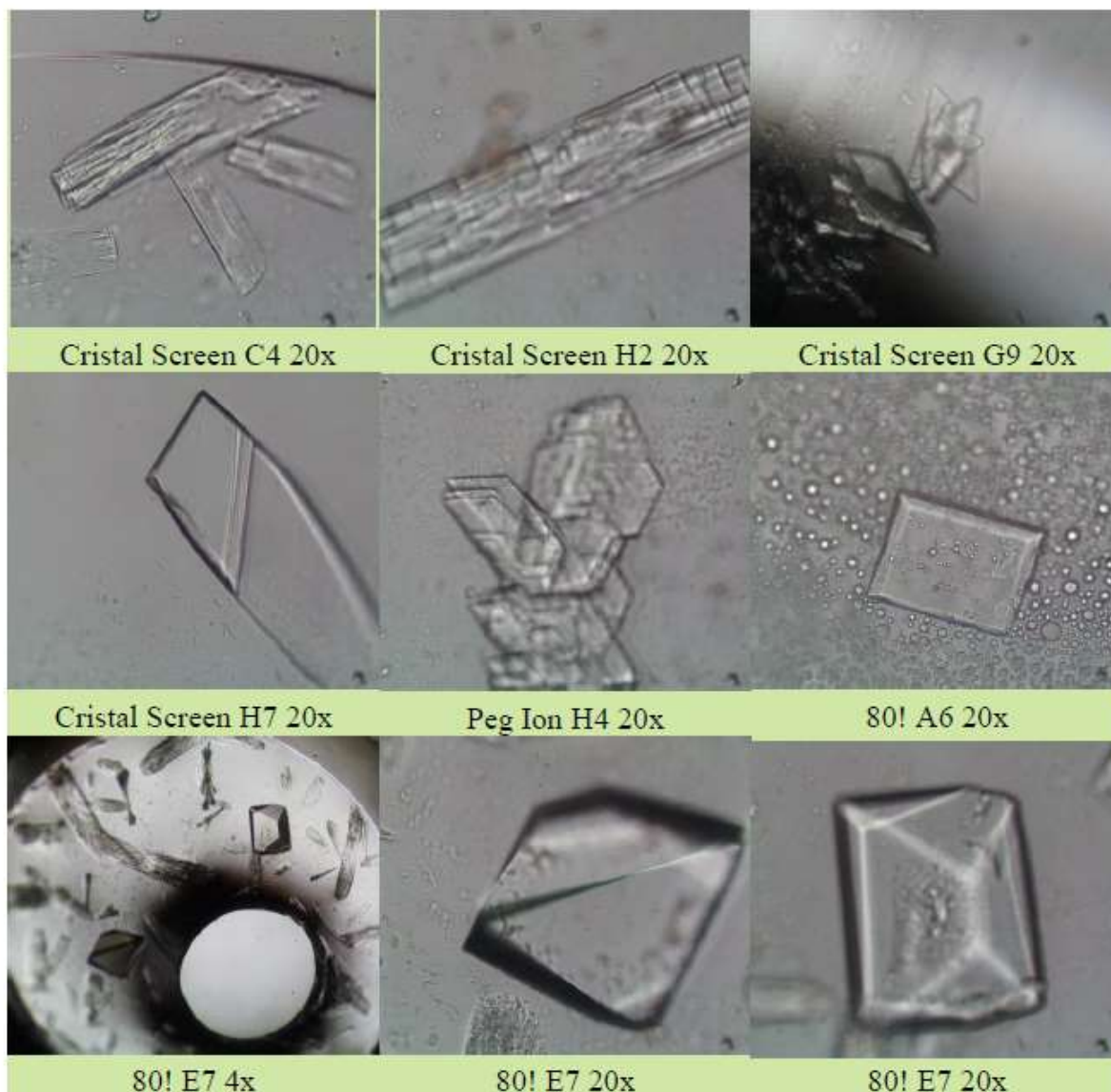


Figure 59 – *Rf2 mut* crystals and respective conditions.

Using the corrected Rf2 sequence, numerous hits were obtained as shown in Figure 59. 12 crystals were then selected and cryo-cooled in liquid nitrogen after soaking in cryoprotectant [30% (v/v) glycerol added to the crystallization buffer or just Paratone-N] for a few seconds, as showed in Table 22.

Table 22 – Crystallization conditions selected. G - Glycerol, P – Paratone.

Crystallization Screen/Condition	Condition	Cryo Protector
80! E7	0.2M S.A. 0.1M Cac 6.5 30 % PEG 8K	G
80! E7	0.2M S.A. 0.1M Cac 6.5 30 % PEG 8K	G
80! E7	0.2M S.A. 0.1M Cac 6.5 30 % PEG 8K	G
80! A6	0.2 M MgCl ₂ 0.1 M Act 4.5 30 % PEG 4K	G
Peg Ion H4	0.03 M Citric acid 0.07 M BIS-TRIS propane / pH 7.6 20% w/v Polyethylene glycol 3,350	P
Peg Ion H4	0.03 M Citric acid 0.07 M BIS-TRIS propane / pH 7.6 20% w/v Polyethylene glycol 3,350	P
Peg Ion H4	0.03 M Citric acid 0.07 M BIS-TRIS propane / pH 7.6 20% w/v Polyethylene glycol 3,350	P
JCSG C9	0.1 M Na/K phosphate pH 6.2 25 % v/v 1,2-propanediol	G
JCSG C5	0.1 M Na HEPES pH 7.5 0.8 M sodium dihydrogen phosphate	G
JCSG C5	0.1 M Na HEPES pH 7.5 0.8 M sodium dihydrogen phosphate	G
Cristal Screen C4	0.2 M Sodium acetate trihydrate 0.1 M Sodium cacodylate trihydrate pH 6.5 30% w/v Polyethylene glycol 8,000	P
Cristal Screen C4	0.2 M Sodium acetate trihydrate 0.1 M Sodium cacodylate trihydrate pH 6.5 30% w/v Polyethylene glycol 8,000	P

The Rf2mut structure was determined for the 80! E7 condition using a cacodylated derivative by single wavelength anomalous dispersion experiment using the SHELX suite via the HKL2MAP graphical interface (Pape & Schneider, 2004). Inverse beam data at the peak wavelength corresponding to the As absorption edge was used to determine heavy atom

sites using SHELXD. A single strong As site and five other minor sites were located. These sites were then used to calculate initial phases using PHASER in SAD mode (McCoy, 2004, 2007) in the CCP4 suite (Winn *et al.*, 2011) followed by density improvement using PARROT (Zhang, Cowtan, & Main, 1997). The quality of the electron density maps was excellent. Automatic model building using ARP/wARP (Langer, Cohen, Lamzin, & Perrakis, 2008) interspersed with REFMAC5 cycles (Murshudov *et al.*, 2011) placed 96 residues out of 107 with a R value of 20.5%. A second dataset was collected to a higher resolution (1.02 Å) and data was processed to 1.08 Å.

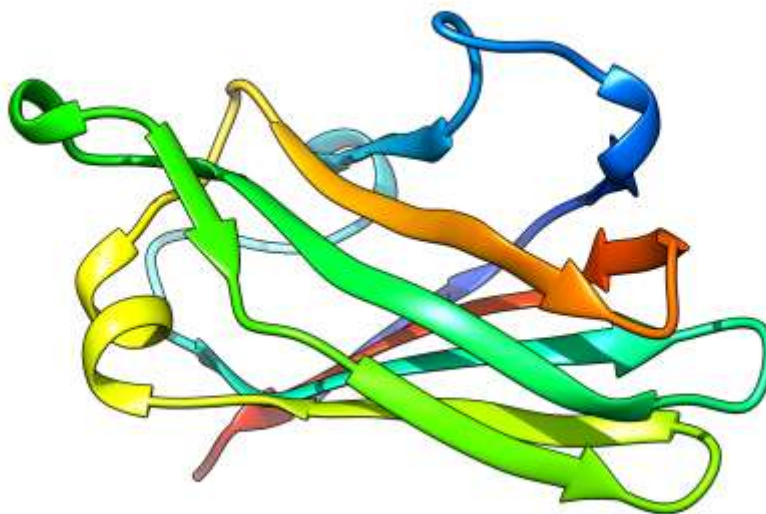


Figure 60 - The three-dimensional structure of Rf2, color ramped from N (blue) to C (red) terminus – Rf2 has a β -sandwich fold comprised of two β -sheets, each consisting of antiparallel β -strands (Image edit on UCSF Chimera).

The structure of Rf2 CBM belongs to the dominant fold among CBMs, which is the β -sandwich fold as previously described (Boraston *et al.*, 2004). It is comprised of two β -sheets, each consisting of four antiparallel β -strands (Figure 60). Unlike other CBMs included in this fold family, Rf2 structure does not present a bound metal atom, which usually serves a structural role. The ligand-recognition site in the CBM fold family members is commonly located on one of the β -sandwich surfaces.

Based upon the visual examination of the molecular surface of Rf2, this CBM might be included in the "glycan-chain-binding" or Type B CBMs, described earlier in the

Introduction and characterized by an extended groove or cleft where aromatic residues mediate ligand binding and specificity.

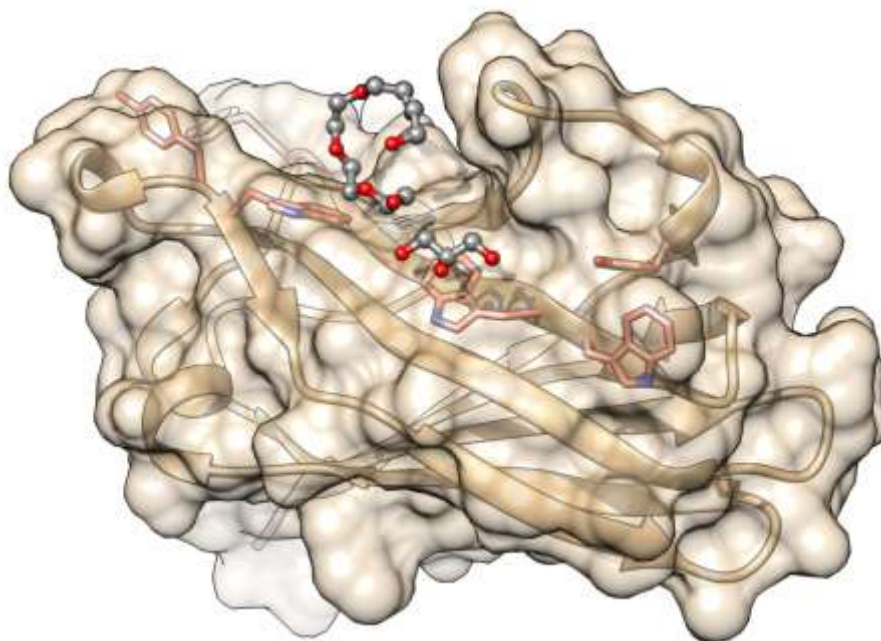


Figure 61- Molecular surface of Rf2 and the putative ligand-binding cleft. The transparent molecular surface emphasizes the putative ligand-binding site, where crystallographic bound polyethylene glycol and glycerol molecules (ball-and-stick representation) were found. The side chains of the aromatic residues referred in the text are salmon colored and shown in stick representation (Image prepared on UCSF Chimera).

In Rf2, the location of several surface aromatic residues in one of the β -sandwiches, together with the crystallographic presence of several polyethylene glycol and glycerol molecules (cryoprotectant) on all three molecules of the crystal asymmetric unit, all bound on that same β -sandwich plane, suggests a putative ligand-binding site. The mentioned aromatic residues of Rf2 include Tyr507, Tyr563, Trp564, Tyr597, Trp606 and Trp607 (based on the PDB file). The presence of protruding loops on both sides of the groove and flanking the β -sheet, confers to this putative Type B binding site a "twisted" cleft-like conformation, akin to that of family 29 CBMs (Figure 61). Speculation on this possible elongated binding cleft might account for the fitting of multiple substrates, with variable subdued interactions. Although further characterization of Rf2 is necessary, preliminary analysis suggests Rf2 will be the founder of a novel CBM family.

Chapter 4 - Conclusions

Since the first description of cellulosomes in the early 1980's by Raffi Lamed and Ed Bayer the knowledge about their structure and functioning have greatly evolved. The discrete multi-enzymatic enzyme complex is now defined as 'multi-enzyme complex produced by anaerobic bacteria for the efficient deconstruction of plant cell wall polysaccharides' (Smith & Bayer, 2013). The compilation of biochemical characterization of cellulosomal components with the aid of genomic and metagenomic information has confirmed the sophistication of cellulosomes, supporting a diversity of catalytic activities such as cellulase, hemicellulase, pectate lyase and carbohydrate esterase (E. A. Bayer *et al.*, 1998; Shi You Ding *et al.*, 2008; Smith & Bayer, 2013). The collaborative data derived from molecular biology, bioinformatics, biochemistry and structural biology has provided a deeper understanding of the molecular basis for cellulosome assembly and function (E. A. Bayer *et al.*, 1998).

However, several questions concerning cellulosome structure and function are still to be answered in order to practically apply its amazing abilities to our advantage, assisting in the development of a new biotechnological approach to face the challenges of a new era of Bio derived products. Despite the modest advances brought by our work, when compared to the amount of knowledge yet to be explored, we were able to enlighten some more aspects about this promising field. The possibility to stabilize the Coh-Doc complex enables an easier production of these two main components of the cellulosome. Although with this project we did not achieve that, we were able to optimize the mutant's construction to minimize the necessary cloning procedures and site-directed mutagenesis methods.

From our protein expression essays we were able to determine a viable path to further explore the large scale production of the studied *Ruminococcus flavefaciens* Coh-Doc complex, which includes the use of BL21 as host cells in LBE medium, grown at 25°C.

These growth conditions are very promising due to its characteristics as LBE is an autoinduction medium which does not require the addition of IPTG, lowering the costs associated with the process, and the determined optimal temperature is very close to room temperature, thus reducing the necessary energy expenditure.

CBMs potentiate the enzymatic catalysis by directing the appended catalytic modules to their target substrates. The genome of the ruminal cellulolytic bacterium, *Ruminococcus flavefaciens* strain FD-1, contains over 200 modular proteins containing the cellulosomal-signature dockerin module. Therefore by determining the substrate affinity of two of these CBM's we will expand the possible applications for these peculiar molecules.

Rf4 showed a significant higher affinity to xyloglucan of $298.65 \text{ } (\text{w/v})^{-1}$ which could indicate a good potential for testing enzymatic activity in association with the enzyme. However such high affinity suggested that more tests are needed, like discovering the 3D structure of the Rf4. By uncovering the structure we could determine the type of connections and ensure that such high affinity doesn't affect the dissociation and degradation of the substrate.

In contrast Rf2 shows less specificity and lower affinity values than Rf4, with higher affinity to galactomannan with $56\% (\text{w/v})^{-1}$ and xyloglucan with $49\% (\text{w/v})^{-1}$. Despite lower, this result may be used to our advantage to initially sort the cellulosic biomass using CBMs like Rf2 which don't have much affinity to a specific substrate but that can be applicable to two substrates.

With these two CBMs we can have contrasting approaches, by choosing from specificity and higher affinity to versatility and higher range of action. Either way, both are very promising tools in helping the task of converting cellulosic biomass into fuel.

However we should take into account the limitations of this method. The AGE is a very good low cost method to determine and scan the affinity to several substrates. The approach with this method initially is only to assess qualitatively the existence or absence of affinity to a certain substrate. Resorting to mathematical approximations we can assign a quantitative value to this affinity.

Nonetheless in order to fully determine the affinity we should consider further experiments like Isothermal Titration Calorimetry that can simultaneously determine all binding parameters in a single experiment. This technique works by directly measuring the heat that is either released or absorbed during a biomolecular binding event. By measuring the heat transfer during binding it can determine binding constants (K_D), reaction stoichiometry (n), enthalpy (ΔH) and entropy (ΔS), providing a complete thermodynamic profile of the molecular interaction.

One of this project's greatest achievements was the ability to purify, crystallize and determine the structure of the putative CBM Rf2. The tridimensional structure of Rf2 shows a putative binding cleft to which the substrates can interact, with further studies needed to relate its relatively large extension with the affinity to more than one compound and explain its low values.

The importance of these findings isn't only about justifying what we previous assessed but also as a stepping stone to expand our knowledge about CBMs. This tridimensional structure will enable further testing and molecular modeling without resorting to wet-lab procedures which saves money and time. We will be able to predict the interactions to other substrates and enzymes bioinformatically and direct our research accordingly.

In conclusion these humble contributions brings us closer to successfully developing a synthetic cellulosome (E. Bayer *et al.*, 2007; Smith & Bayer, 2013; Vazana *et al.*, 2013) that combines the stability of the type-II Coh-Doc attachment to a reusable matrix with the flexibility of the type-I interaction, enabling the attachment of several CBMs and enzymes, consequently improving the efficiency and range of substrates decomposition. The ultimate goal will be to create stable and easy to produce nanomachine complexes that can be applied to a wide range of process and originate a variety of products.

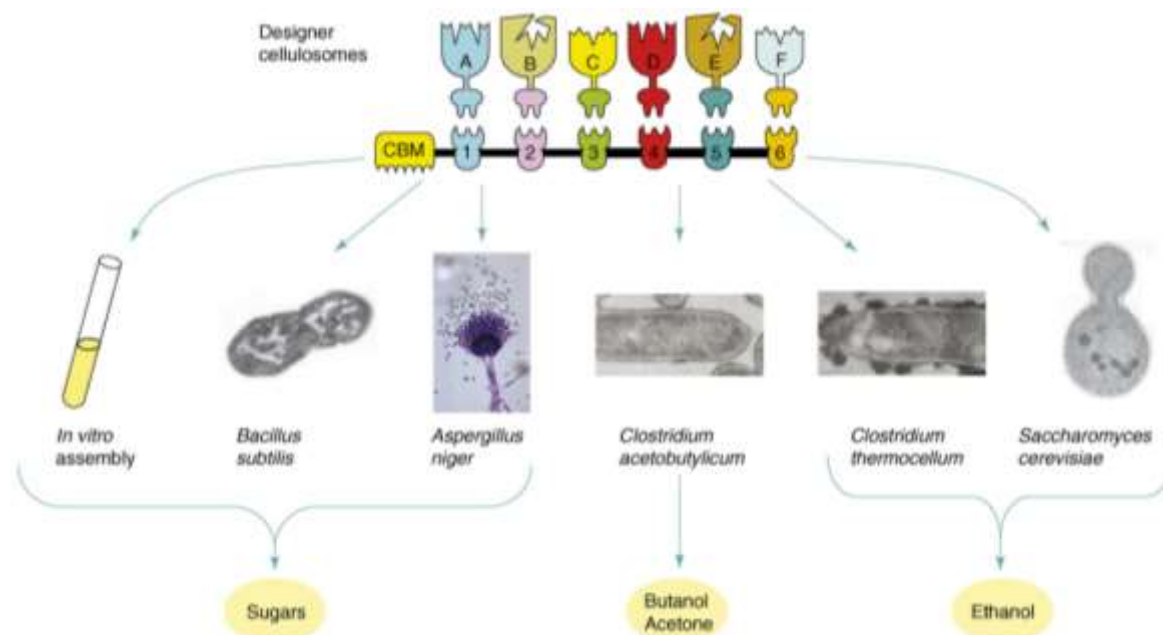


Figure 62 – Ultimate Goal for Cellulosomes Application - Engineering potent cellulolytic microbes for the production of desired end products. Genes encoding for cellulases and/or designer cellulosome components (i.e. chimeric scaffoldin and desired dockerin-containing hybrid enzymes) can be cloned

into a desired bacterial or fungal host cell, and the secreted proteins can be overexpressed for the degradation of cellulosic biomass in an industrial reactor (in vitro assembly). Alternatively, the genes can be cloned into a suitable bacterial, fungal or yeast host, and the transformed cell with either de novo or improved cellulose-degrading capacity can be grown directly on cellulosic biomass to produce a desired end product (E. Bayer *et al.*, 2007).

Chapter 5 - Bibliographic References

- Abbott, D., & Boraston, A. (2012). Quantitative approaches to the analysis of carbohydrate-binding module function. *Methods Enzymol*, 510, 211–231. doi:10.1016/B978-0-12-415931-0.00011-2
- 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. *Proceedings of the National Academy of Sciences of the United States of America*, 103(2), 305–310. doi:10.1073/pnas.0507109103
- Agbor, V. B., Cicek, N., Sparling, R., Berlin, A., & Levin, D. B. (2011). Biomass pretreatment: fundamentals toward application. *Biotechnology Advances*, 29(6), 675–85. doi:10.1016/j.biotechadv.2011.05.005
- Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W., & Lipman, D. J. (1997). Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Research*. doi:10.1093/nar/25.17.3389
- Aransiola, E. F., Ojumu, T. V., Oyekola, O. O., Madzimbamuto, T. F., & Ikhu-Omoregbe, D. I. O. (2014). A review of current technology for biodiesel production: State of the art. *Biomass and Bioenergy*, 61, 276–297. doi:10.1016/j.biombioe.2013.11.014
- Atmodjo, M. a, Hao, Z., & Mohnen, D. (2013). Evolving views of pectin biosynthesis. *Annual Review of Plant Biology*, 64, 747–79. doi:10.1146/annurev-arplant-042811-105534
- Baneyx, F. (1999). Recombinant protein expression in *Escherichia coli*. *Current Opinion in Biotechnology*, 10(5), 411–421. doi:10.1016/S0958-1669(99)00003-8
- Barral, P., Suárez, C., Batanero, E., Alfonso, C., Alché, J. D. D., Rodríguez-García, M. I., ... Rodríguez, R. (2005). An olive pollen protein with allergenic activity, Ole e 10, defines a novel family of carbohydrate-binding modules and is potentially implicated in pollen germination. *The Biochemical Journal*, 390(Pt 1), 77–84. doi:10.1042/BJ20050456
- Bayer, E. A., Belaich, J.-P., Shoham, Y., & Lamed, R. (2004). The Cellulosomes: Multienzyme Machines for Degradation of Plant Cell Wall Polysaccharides. *Annual Review of Microbiology*, 58(1), 521–554. doi:10.1146/annurev.micro.57.030502.091022
- Bayer, E. A., Chanzy, H., Lamed, R., & Shoham, Y. (1998). Cellulose, cellulases and cellulosomes. *Current Opinion in Structural Biology*. doi:10.1016/S0959-440X(98)80143-7
- Bayer, E. a., Lamed, R., White, B. a., & Flints, H. J. (2008). From cellulosomes to cellulosomics. *Chemical Record*, 8(6), 364–377. doi:10.1002/tcr.20160
- Bayer, E., Lamed, R., & Himmel, M. (2007). The potential of cellulases and cellulosomes for cellulosic waste management. *Current Opinion in Biotechnology*, 18(3), 237–45. doi:10.1016/j.copbio.2007.04.004
- Béguin, P., & Alzari, P. (1998). The cellulosome of *Clostridium cellulolyticum*. *Biochem Soc Trans.*, 26(2), 178–85.
- Bhalekar, M., Sonawane, S., & Shimpi, S. (2013). Synthesis and characterization of a cysteine xyloglucan conjugate as mucoadhesive polymer. *Brazilian Journal of Pharmaceutical Sciences*, 49(2), 285–292. doi:10.1590/S1984-82502013000200010
- Bharathiraja, B., Chakravarthy, M., Kumar, R. R., Yuvaraj, D., Jayamuthunagai, J., Kumar, R. P., & Palani, S. (2014). Biodiesel production using chemical and biological methods - A review of process, catalyst, acyl acceptor, source and process variables. *Renewable and Sustainable Energy Reviews*, 38, 368–382. doi:10.1016/j.rser.2014.05.084
- Blake, A. W., McCartney, L., Flint, J. E., Bolam, D. N., Boraston, A. B., Gilbert, H. J., & Knox, J. P. (2006). Understanding the biological rationale for the diversity of cellulose-directed carbohydrate-binding modules in prokaryotic enzymes. *The Journal of Biological Chemistry*, 281(39), 29321–9. doi:10.1074/jbc.M605903200
- Bolam, D. N., Xie, H., White, P., Simpson, P. J., Hancock, S. M., Williamson, M. P., & Gilbert, H.

- J. (2001). Evidence for Synergy between Family 2b Carbohydrate Binding Modules in *Cellulomonas fimi* Xylanase 11A. *Biochemistry*, 40(8), 2468–2477. doi:10.1021/bi002564l
- Boraston, A. B., Bolam, D. N., Gilbert, H. J., & Davies, G. J. (2004). Carbohydrate-binding modules: fine-tuning polysaccharide recognition. *The Biochemical Journal*, 382(Pt 3), 769–781. doi:10.1042/BJ20040892
- Boraston, A. B., Healey, M., Klassen, J., Ficko-Blean, E., Van Bueren, A. L., & Law, V. (2006). A structural and functional analysis of ??-glucan recognition by family 25 and 26 carbohydrate-binding modules reveals a conserved mode of starch recognition. *Journal of Biological Chemistry*, 281, 587–598. doi:10.1074/jbc.M509958200
- Boraston, A. B., Kwan, E., Chiu, P., Warren, R. A. J., & Kilburn, D. G. (2003). Recognition and hydrolysis of noncrystalline cellulose. *Journal of Biological Chemistry*, 278(8), 6120–6127. doi:10.1074/jbc.M209554200
- Boraston, A. B., McLean, B. W., Chen, G., Li, A., Warren, R. A. J., & Kilburn, D. G. (2002). Co-operative binding of triplicate carbohydrate-binding modules from a thermophilic xylanase. *Molecular Microbiology*, 43(1), 187–194. doi:10.1046/j.1365-2958.2002.02730.x
- Bottcher, H. (IIASA). (2010). Biomass Energy Europe - Summary Report on Illustration Cases. D6.1, (Final Report), 1–20.
- Boudet, A. M., Kajita, S., Grima-Pettenati, J., & Goffner, D. (2003). Lignins and lignocellulosics: a better control of synthesis for new and improved uses. *Trends in Plant Science*, 8(12), 576–81. doi:10.1016/j.tplants.2003.10.001
- Brás, J. L. A. (Faculdade de M. V. (2012). *Structure and function relationships in novel cellulosomal enzymes and cohesin-dockerin complexes*.
- Brett, C. T., Waldren, K., & Brown, E. (1996). *Physiology and Biochemistry of Plant Cell Walls. Topics in Plant Functional Biology*. (E. (Black, M. & Charlewood, B., Ed.) (Chapman an.). London.
- Brown, G. D., & Gordon, S. (2001). Immune recognition: A new receptor for [beta]-glucans. *Nature*, 413(6851), 36–37. Retrieved from <http://dx.doi.org/10.1038/35092620>
- Burton, R., & Fincher, G. (2014). Plant cell wall engineering: applications in biofuel production and improved human health. *Current Opinion in Biotechnology*, 26, 79–84. doi:10.1016/j.copbio.2013.10.007
- Cameron, K. (2014). *Structure and function relationships in novel cohesin- dockerin complexes*.
- Campbell, J. A., Davies, G. J., Bulone, V., & Henrissat, B. (1997). A classification of nucleotide-diphospho-sugar glycosyltransferases based on amino acid sequence similarities. *Biochemical Journal*, 326(Pt 3), 929–939. Retrieved from <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1218753/>
- Cantarel, B. L., Coutinho, P. M., Rancurel, C., Bernard, T., Lombard, V., & Henrissat, B. (2009). The Carbohydrate-Active EnZymes database (CAZy): an expert resource for Glycogenomics. *Nucleic Acids Research*, 37(Database issue), D233–D238. doi:10.1093/nar/gkn663
- Carere, C. R., Sparling, R., Cicek, N., & Levin, D. B. (2008). Third generation biofuels via direct cellulose fermentation. *International Journal of Molecular Sciences*, 9(7), 1342–1360. doi:10.3390/ijms9071342
- Carvalho, A. L., Dias, F. M. V., Nagy, T., Prates, J. A. M., Proctor, M. R., Smith, N., ... Gilbert, H. J. (2007). Evidence for a dual binding mode of dockerin modules to cohesins. *Proceedings of the National Academy of Sciences of the United States of America*, 104(9), 3089–3094. doi:10.1073/pnas.0611173104
- Carvalho, A. L., Dias, F. M. V., Prates, J. A. M., Nagy, T., Gilbert, H. J., Davies, G. J., ... Fontes, C. M. G. A. (2003). Cellulosome assembly revealed by the crystal structure of the cohesin-dockerin complex. *Proceedings of the National Academy of Sciences of the United States of America*, 100(24), 13809–13814. doi:10.1073/pnas.1936124100
- Carvalho, A. L., Pires, V. M. R., Gloster, T. M., Turkenburg, J. P., Prates, J. A. M., Ferreira, L. M.

- A., ... Gilbert, H. J. (2005). Insights into the structural determinants of cohesin - Dockerin specificity revealed by the crystal structure of the type II cohesin from *Clostridium thermocellum* SdbA. *Journal of Molecular Biology*, 349(5), 909–915. doi:10.1016/j.jmb.2005.04.037
- Choi, O. K., Song, J. S., Cha, D. K., & Lee, J. W. (2014). Biodiesel production from wet municipal sludge: Evaluation of in situ transesterification using xylene as a cosolvent. *Bioresource Technology*, 166, 51–56. doi:10.1016/j.biortech.2014.05.001
- Chun, S.-J., Choi, E.-S., Lee, E.-H., Kim, J. H., Lee, S.-Y., & Lee, S.-Y. (2012). Eco-friendly cellulose nanofiber paper-derived separator membranes featuring tunable nanoporous network channels for lithium-ion batteries. *J. Mater. Chem.*, 22(32), 16618–16626. doi:10.1039/C2JM32415F
- Corma, A., Iborra, S., & Velty, A. (2007). Chemical routes for the transformation of biomass into chemicals. *Chemical Reviews*, 107(6), 2411–502. doi:10.1021/cr050989d
- Correia, M. (2009). *Structural and functional insights into the role of Carbohydrate Esterases and Carbohydrate-Binding Modules in plant cell wall hydrolysis*. Faculdade de Medicina Veterinária. Retrieved from <http://www.repository.utl.pt/handle/10400.5/2067>
- Cosgrove, D. J. (2005). Growth of the plant cell wall. *Nature Reviews. Molecular Cell Biology*, 6, 850–861. doi:10.1038/nrm1746
- Dassa, B., Borovok, I., Ruimy-Israeli, V., & Lamed, R. (2014). Rumen cellulosomics: Divergent fiber-degrading strategies revealed by comparative genome-wide analysis of six ruminococcal strains. *PloS One*, 9(7). doi:10.1371/journal.pone.0099221
- Davies, G. J., Gloster, T. M., & Henrissat, B. (2005). Recent structural insights into the expanding world of carbohydrate-active enzymes. *Current Opinion in Structural Biology*. doi:10.1016/j.sbi.2005.10.008
- Demirbas, A. (2001). Biomass resource facilities and biomass conversion processing for fuels and chemicals, 42, 1357–1378.
- Din, N., Gilkes, N. R., Tekant, B., Miller, R. C., Warren, R. A. J., & Kilburn, D. G. (1991). Non-Hydrolytic Disruption of Cellulose Fibres by the Binding Domain of a Bacterial Cellulase. *Nat Biotech*, 9(11), 1096–1099. Retrieved from <http://dx.doi.org/10.1038/nbt1191-1096>
- Ding, S. Y., Rincon, M. T., Lamed, R., Martin, J. C., McCrae, S. I., Aurilia, V., ... Flint, H. J. (2001). Cellulosomal scaffoldin-like proteins from *Ruminococcus flavefaciens*. *Journal of Bacteriology*, 183, 1945–1953. doi:10.1128/JB.183.6.1945-1953.2001
- Ding, S. Y., Xu, Q., Crowley, M., Zeng, Y., Nimlos, M., Lamed, R., ... Himmel, M. E. (2008). A biophysical perspective on the cellulosome: new opportunities for biomass conversion. *Current Opinion in Biotechnology*. doi:10.1016/j.copbio.2008.04.008
- Doi, R. H. (2008). Cellulases of mesophilic microorganisms: Cellulosome and noncellulosome producers. In *Annals of the New York Academy of Sciences* (Vol. 1125, pp. 267–279). doi:10.1196/annals.1419.002
- Doi, R. H., & Kosugi, A. (2004). Cellulosomes: plant-cell-wall-degrading enzyme complexes. *Nature Reviews. Microbiology*, 2(7), 541–51. doi:10.1038/nrmicro925
- El Khoury, D., Cuda, C., Luhovyy, B. L., & Anderson, G. H. (2012). Beta glucan: Health benefits in obesity and metabolic syndrome. *Journal of Nutrition and Metabolism*. doi:10.1155/2012/851362
- Elkins, J. G., Raman, B., & Keller, M. (2010). Engineered microbial systems for enhanced conversion of lignocellulosic biomass. *Current Opinion in Biotechnology*, 21(5), 657–62. doi:10.1016/j.copbio.2010.05.008
- Faaij, A. P. C. (2006). Bio-energy in Europe: changing technology choices. *Energy Policy*, 34(3), 322–342. doi:10.1016/j.enpol.2004.03.026
- FAO, F. and A. O. of T. U. N. (2010). Global Forest Resources Assessment 2010 Main Report. *FAO Forestry Paper*, 163. doi:ISBN 978-92-5-106654-6

- Fontes, C. M. G. A., & Gilbert, H. J. (2010). Cellulosomes : Highly Efficient Nanomachines Designed to Deconstruct Plant Cell Wall Complex Carbohydrates. doi:10.1146/annurev-biochem-091208-085603
- Fry, S. C. (1989). Cellulases, hemicelluloses and auxin-stimulated growth: a possible relationship. *Physiologia Plantarum*, 75, 532–536. doi:10.1111/j.1399-3054.1989.tb05620.x
- Gilbert, H. J. (2007). Cellulosomes: Microbial nanomachines that display plasticity in quaternary structure. *Molecular Microbiology*. doi:10.1111/j.1365-2958.2007.05640.x
- Gilbert, H. J., Stalbrand, H., & Brumer, H. (2008). How the walls come crumbling down: recent structural biochemistry of plant polysaccharide degradation. *Current Opinion in Plant Biology*, 11(3), 338–348. doi:10.1016/j.pbi.2008.03.004
- Gourlay, K. I. (2014). The Role of Amorphogenesis in the Enzymatic Deconstruction of Lignocellulosic Biomass by, (October).
- Guillén, D., & Sánchez, S. (2010). Carbohydrate-binding domains : multiplicity of biological roles, 1241–1249. doi:10.1007/s00253-009-2331-y
- Ha, M. A., Apperley, D. C., & Jarvis, M. C. (1997). Molecular Rigidity in Dry and Hydrated Onion Cell Walls. *Plant Physiology*, 115(2), 593–598. Retrieved from <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC158519/>
- Haimovitz, R., Barak, Y., Morag, E., Voronov-Goldman, M., Shoham, Y., Lamed, R., & Bayer, E. A. (2008). Cohesin-dockerin microarray: Diverse specificities between two complementary families of interacting protein modules. *Proteomics*, 8(5), 968–979. doi:10.1002/pmic.200700486
- Hammel, M., Fierobe, H. P., Czjzek, M., Kurkal, V., Smith, J. C., Bayer, E. A., ... Receveur-Bréchet, V. (2005). Structural basis of cellulosome efficiency explored by small angle x-ray scattering. *Journal of Biological Chemistry*, 280(46), 38562–38568. doi:10.1074/jbc.M503168200
- Hammel, M., Fierobe, H.-P., Czjzek, M., Finet, S., & Receveur-Bréchet, V. (2004). Structural insights into the mechanism of formation of cellulosomes probed by small angle X-ray scattering. *The Journal of Biological Chemistry*, 279(53), 55985–55994. doi:10.1074/jbc.M408979200
- Hashimoto, H. (2006). Recent structural studies of carbohydrate-binding modules. *Cellular and Molecular Life Sciences CMLS*, 63(24), 2954–2967. doi:10.1007/s00018-006-6195-3
- Henrissat, B. (1991). A classification of glycosyl hydrolases based sequence similarities amino acid. *Biochemical Journal*, 280((Pt 2)), 309–316. doi:10.1007/s007920050009
- Hervé, C., Rogowski, A., Blake, A. W., Marcus, S. E., Gilbert, H. J., & Knox, J. P. (2010). Carbohydrate-binding modules promote the enzymatic deconstruction of intact plant cell walls by targeting and proximity effects. *Proceedings of the National Academy of Sciences of the United States of America*, 107(34), 15293–8. doi:10.1073/pnas.1005732107
- Hyeon, J. E., Jeon, S. D., & Han, S. O. (2013). complexes for advanced biotechnology tool development : Advances and applications. *Biotechnology Advances*, 31(6), 936–944. doi:10.1016/j.biotechadv.2013.03.009
- Jamal, S., Nurizzo, D., Boraston, A. B., & Davies, G. J. (2004). X-ray crystal structure of a non-crystalline cellulose-specific carbohydrate-binding module: CBM28. *Journal of Molecular Biology*, 339(2), 253–258. doi:10.1016/j.jmb.2004.03.069
- Jervis, E. J., Haynes, C. a., & Kilburn, D. G. (1997). Surface diffusion of cellulases and their isolated binding domains on cellulose. *Journal of Biological Chemistry*, 272(38), 24016–24023. doi:10.1074/jbc.272.38.24016
- Jindou, S., Soda, A., Karita, S., Kajino, T., Béguin, P., Wu, J. H. D., ... Ohmiya, K. (2004). Cohesin-dockerin interactions within and between clostridium josui and clostridium thermocellum: Binding selectivity between cognate dockerin and cohesin domains and species specificity. *Journal of Biological Chemistry*, 279(11), 9867–9874.

- doi:10.1074/jbc.M308673200
- Keegstra, K., Talmadge, K. W., Bauer, W. D., & Albersheim, P. (1973). The Structure of Plant Cell Walls: III. A Model of the Walls of Suspension-cultured Sycamore Cells Based on the Interconnections of the Macromolecular Components. *Plant Physiology*, 51, 188–197. doi:10.1104/pp.51.1.188
- Khan, T. M. Y., Atabani, A. E., Badruddin, I. A., Badarudin, A., Khayoon, M. S., & Triwahyono, S. (2014). Recent scenario and technologies to utilize non-edible oils for biodiesel production. *Renewable and Sustainable Energy Reviews*, 37, 840–851. doi:10.1016/j.rser.2014.05.064
- Knudsen, K. E. B. (1997). Carbohydrate and lignin contents of plant materials used in animal feeding. *Animal Feed Science and Technology*. doi:10.1016/S0377-8401(97)00009-6
- Kraulis, J. . C. G. M. . N. M. . J. T. A. . P. G. . K. J. . G. A. M. (1989). Determination of the three-dimensional solution structure of the C-terminal domain of cellobiohydrolase I from *Trichoderma reesei*. A study using nuclear magnetic resonance and hybrid distance geometry-dynamical simulated annealing. *Biochemistry*, 28, 7241–7257. doi:2554967
- Langer, G. G., Cohen, S. X., Lamzin, V. S., & Perrakis, A. (2008). Automated macromolecular model building for X-ray crystallography using ARP/wARP version 7. *Nature Protocols*, 3(7), 1171–1179. doi:10.1038/nprot.2008.91
- Lehtiö, J., Sugiyama, J., Gustavsson, M., Fransson, L., Linder, M., & Teeri, T. T. (2003). The binding specificity and affinity determinants of family 1 and family 3 cellulose binding modules. *Proceedings of the National Academy of Sciences of the United States of America*, 100(2), 484–489. doi:10.1073/pnas.212651999
- Lombard, V., Bernard, T., Rancurel, C., Brumer, H., Coutinho, P. M., & Henrissat, B. (2010). A hierarchical classification of polysaccharide lyases for glycogenomics. *The Biochemical Journal*, 432(3), 437–444. doi:10.1042/BJ20101185
- Lombard, V., Golaconda Ramulu, H., Drula, E., Coutinho, P. M., & Henrissat, B. (2014). The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Research*, 42(D1), 490–495. doi:10.1093/nar/gkt1178
- Lucia, L. A., Argyropoulos, D. S., Adamopoulos, L., & Gaspar, A. R. (2006). Chemicals and energy from biomass. *Canadian Journal of Chemistry*, 84(7), 960–970. doi:10.1139/v06-117
- Lynd, L. R., Weimer, P. J., Zyl, W. H. Van, & Pretorius, I. S. (2002). Microbial Cellulose Utilization : Fundamentals and Biotechnology, 66(3), 506–577. doi:10.1128/MMBR.66.3.506
- Lytle, B. L., Volkman, B. F., Westler, W. M., Heckman, M. P., & Wu, J. H. (2001). Solution structure of a type I dockerin domain, a novel prokaryotic, extracellular calcium-binding domain. *Journal of Molecular Biology*, 307(3), 745–753. doi:10.1006/jmbi.2001.4522
- Macías-Sánchez, M. D., Robles-Medina, A., Hita-Peña, E., Jiménez-Callejón, M. J., Estéban-Cerdán, L., González-Moreno, P. A., & Molina-Grima, E. (2015). Biodiesel production from wet microalgal biomass by direct transesterification. *Fuel*, 150, 14–20. doi:10.1016/j.fuel.2015.01.106
- Mateos, E., & González, J. (2007). Biomass: Potential source of useful energy. *Cellulose*, (1), 1–6. Retrieved from <http://www.icrepq.com/icrepq07/335-mateos.pdf>
- Matthews, B. W. (1968). Solvent content of protein crystals. *Journal of Molecular Biology*, 33(2), 491–497. doi:10.1016/0022-2836(68)90205-2
- Maza, S. (2012). The Cultural Origins of the French Revolution. In *A Companion to the French Revolution* (pp. 42–56). doi:10.1002/9781118316399.ch3
- McCoy, A. J. (2004). Liking likelihood. *Acta Crystallographica Section D*, 60(12 Part 1), 2169–2183. doi:10.1107/S0907444904016038
- McCoy, A. J. (2007). Solving structures of protein complexes by molecular replacement with {\it Phaser}. *Acta Crystallographica Section D*, 63(1), 32–41. doi:10.1107/S0907444906045975
- Mechaly, A., Fierobe, H. P., Belaich, A., Belaich, J. P., Lamed, R., Shoham, Y., & Bayer, E. A. (2001). Cohesin-dockerin interaction in cellulosome assembly: A single hydroxyl group of a

- dockerin domain distinguishes between nonrecognition and high affinity recognition. *Journal of Biological Chemistry*, 276(13), 9883–9888. doi:10.1074/jbc.M009237200
- Menon, V., & Rao, M. (2012). Trends in bioconversion of lignocellulose: biofuels, platform chemicals & biorefinery concept. *Progress in Energy and Combustion Science*, 38(4), 522–550. doi:10.1016/j.pecs.2012.02.002
- Miras, I., Schaeffer, F., Béguin, P., & Alzari, P. M. (2002). Mapping by Site-Directed Mutagenesis of the Region Responsible for Cohesin–Dockerin Interaction on the Surface of the Seventh Cohesin Domain of *Clostridium thermocellum* CipA†. *Biochemistry*, 41(7), 2115–2119. doi:10.1021/bi011854e
- Murshudov, G. N., Skubák, P., Lebedev, A. A., Pannu, N. S., Steiner, R. A., Nicholls, R. A., ... Vagin, A. A. (2011). {REFMAC}5 for the refinement of macromolecular crystal structures. *Acta Crystallographica Section D*, 67(4), 355–367. doi:10.1107/S0907444911001314
- Noach, I., Frolow, F., Alber, O., Lamed, R., Shimon, L. J. W., & Bayer, E. A. (2009). Intermodular Linker Flexibility Revealed from Crystal Structures of Adjacent Cellulosomal Cohesins of *Acetivibrio cellulolyticus*. *Journal of Molecular Biology*, 391(1), 86–97. doi:10.1016/j.jmb.2009.06.006
- Noach, I., Frolow, F., Jakoby, H., Rosenheck, S., Shimon, L. J. W., Lamed, R., & Bayer, E. A. (2005). Crystal structure of a type-II cohesin module from the *Bacteroides cellulosolvens* cellulosome reveals novel and distinctive secondary structural elements. *Journal of Molecular Biology*, 348(1), 1–12. doi:10.1016/j.jmb.2005.02.024
- Noach, I., Levy-Assaraf, M., Lamed, R., Shimon, L. J. W., Frolow, F., & Bayer, E. A. (2010). Modular arrangement of a cellulosomal scaffoldin subunit revealed from the crystal structure of a cohesin dyad. *Journal of Molecular Biology*, 399(2), 294–305. doi:10.1016/j.jmb.2010.04.013
- Notenboom, V. . B. A. B. . W. S. J. . K. D. G. . R. D. R. (2002). High-resolution crystal structures of the lectin-like xylan binding domain from *Streptomyces lividans* xylanase 10A with bound substrates reveal a novel mode of xylan binding. *Biochemistry*, 41, 4246–4254. doi:10.1021/bi011854e
- Pagès, S., Bélaïch, A., Bélaïch, J.-P., Morag, E., Lamed, R., Shoham, Y., & Bayer, E. A. (1997). Species-specificity of the cohesin-dockerin interaction between *Clostridium thermocellum* and *Clostridium cellulolyticum*: Prediction of specificity determinants of the dockerin domain. *Proteins: Structure, Function, and Bioinformatics*, 29(4), 517–527. doi:10.1002/(SICI)1097-0134(199712)29:4<517::AID-PROT11>3.0.CO;2-P
- Pape, T., & Schneider, T. R. (2004). HKL2MAP: a graphical user interface for macromolecular phasing with SHELX programs. *Journal of Applied Crystallography*, 37(5), 843–844. doi:10.1107/S0021889804018047
- Raschka, A., & Carus, M. (n.d.). Industrial material use of biomass Basic data for Germany , Europe and the world Industrial material use of biomass Basic data for Germany , Europe and the world.
- Rincón, L. E., Jaramillo, J. J., & Cardona, C. a. (2014). Comparison of feedstocks and technologies for biodiesel production: An environmental and techno-economic evaluation. *Renewable Energy*, 69, 479–487. doi:10.1016/j.renene.2014.03.058
- Rubin, E. M. (2008). Genomics of cellulosic biofuels. *Nature*, 454(7206), 841–845. doi:10.1038/nature07190
- Schaeffer, F., Matuschek, M., Guglielmi, G., Miras, I., Alzari, P. M., & Béguin, P. (2002). Duplicated Dockerin Subdomains of *Clostridium thermocellum* Endoglucanase CelD Bind to a Cohesin Domain of the Scaffolding Protein CipA with Distinct Thermodynamic Parameters and a Negative Cooperativity†. *Biochemistry*, 41(7), 2106–2114. doi:10.1021/bi011853m
- Shen, L., Worrell, E., & Patel, M. (2010). Present and future development in plastics from biomass. *Biofuels, Bioproducts and Biorefining*, 4(1), 25–40. doi:10.1002/bbb.189
- Shimon, L. J. W., Frolow, F., Yaron, S., Bayer, E. A., Lamed, R., Morag, E., & Shoham, Y. (1997).

- Crystallization and preliminary X-ray analysis of a cohesin domain of the cellulosome from *Clostridium thermocellum*. *Acta Crystallographica Section D: Biological Crystallography*, 53(1), 114–115. doi:10.1107/S090744499601164X
- Shoseyov, O., Shani, Z., & Levy, I. (2006). Carbohydrate binding modules: biochemical properties and novel applications. *Microbiology and Molecular Biology Reviews : MMBR*, 70(2), 283–95. doi:10.1128/MMBR.00028-05
- Simpson, P. J., Xie, H., Bolam, D. N., Gilbert, H. J., & Williamson, M. P. (2000). The structural basis for the ligand specificity of family 2 carbohydrate-binding modules. *Journal of Biological Chemistry*, 275(52), 41137–41142. doi:10.1074/jbc.M006948200
- Singh, K. M., Reddy, B., Patel, D., Patel, A. K., Parmar, N., Patel, A., ... Joshi, C. G. (2014). High Potential Source for Biomass Degradation Enzyme Discovery and Environmental Aspects Revealed through Metagenomics of Indian Buffalo Rumen, 2014.
- Sinnott, M. L. (1990). Catalytic mechanism of enzymic glycosyl transfer. *Chemical Reviews*, 90(7), 1171–1202. doi:10.1021/cr00105a006
- Smith, S. P., & Bayer, E. A. (2013). Insights into cellulosome assembly and dynamics : from dissection to reconstruction of the supramolecular enzyme complex. *Current Opinion in Structural Biology*, 23(5), 686–694. doi:10.1016/j.sbi.2013.09.002
- Somerville, C., Bauer, S., Brininstool, G., Facette, M., Hamann, T., Milne, J., ... Youngs, H. (2004). Toward a systems approach to understanding plant cell walls. *Science*, 306, 2206–2211. doi:10.1126/science.1102765
- Southall, S. M., Simpson, P. J., Gilbert, H. J., Williamson, G., & Williamson, M. P. (1999). The starch-binding domain from glucoamylase disrupts the structure of starch. *FEBS Letters*, 447(1), 58–60. doi:10.1016/S0014-5793(99)00263-X
- Sticklen, M. B. (2008a). Plant genetic engineering for biofuel production: towards affordable cellulosic ethanol. *Nature Reviews. Genetics*, 9(6), 433–43. doi:10.1038/nrg2336
- Sticklen, M. B. (2008b). Plant genetic engineering for biofuel production: towards affordable cellulosic ethanol. *Nature Reviews. Genetics*, 9(6), 433–443. doi:10.1038/nrg2336
- Sullivan, A. C. O. (1997). Cellulose : the structure slowly unravels, 173–207.
- Talbott, L. D., & Ray, P. M. (1992). Changes in molecular size of previously deposited and newly synthesized pea cell wall matrix polysaccharides : effects of auxin and turgor. *Plant Physiology*, 98, 369–379. doi:10.1104/pp.98.1.369
- Thygesen, A., Oddershede, J., Lilholt, H., Thomsen, A. B., & Ståhl, K. (2005). On the determination of crystallinity and cellulose content in plant fibres. *Cellulose*, 12, 563–576. doi:10.1007/s10570-005-9001-8
- Tomme, P., Warren, R. A., & Gilkes, N. R. (1995). Cellulose hydrolysis by bacteria and fungi. *Advances in Microbial Physiology*, 37, 1–81.
- Van Bueren, A. L., Higgins, M., Wang, D., Burke, R. D., & Boraston, A. B. (2007). Identification and structural basis of binding to host lung glycogen by streptococcal virulence factors. *Nat Struct Mol Biol*, 14(1), 76–84. Retrieved from <http://dx.doi.org/10.1038/nsmb1187>
- Vazana, Y., Barak, Y., Unger, T., Peleg, Y., Shamshoum, M., Ben-Yehezkel, T., ... Bayer, E. a. (2013). A synthetic biology approach for evaluating the functional contribution of designer cellulosome components to deconstruction of cellulosic substrates. *Biotechnology for Biofuels*, 6(1), 182. doi:10.1186/1754-6834-6-182
- Venditto, I., Centeno, M. S. J., Ferreira, L. M. A., Fontes, C. M. G. A., & Najmudin, S. (2014). Expression, purification and crystallization of a novel carbohydrate-binding module from the *Vit Ruminococcus flavefaciens* cellulosome. *Acta Crystallographica Section F*, 70(12), 1653–1656. doi:10.1107/S2053230X14024248
- Venditto, I., Najmudin, S., Luis, A. S., Ferreira, L. M. a., Sakka, K., Knox, J. P., ... Fontes, C. M. G. a. (2015). Family 46 Carbohydrate-Binding Modules contribute to the enzymatic hydrolysis of xyloglucan and β -1,3-1,4-glucans through distinct mechanisms. *Journal of*

- Biological Chemistry*, jbc.M115.637827. doi:10.1074/jbc.M115.637827
- Wallace, I. S., & Somerville, C. R. (2014). A Blueprint for Cellulose Biosynthesis, Deposition, and Regulation in Plants. In *Plant Cell Wall Patterning and Cell Shape* (pp. 65–95). John Wiley & Sons, Inc. doi:10.1002/9781118647363.ch3
- Warren, R. A. J. (1996). MICROBIAL HYDROLYSIS OF POLYSACCHARIDES, 183–212.
- Winn, M. D., Ballard, C. C., Cowtan, K. D., Dodson, E. J., Emsley, P., Evans, P. R., ... Wilson, K. S. (2011). Overview of the CCP 4 suite and current developments. *Acta Crystallographica Section D Biological Crystallography*, 67(4), 235–242. doi:10.1107/S0907444910045749
- Xu, Q., Bayer, E. A., Goldman, M., Kenig, R., Shoham, Y., & Lamed, R. (2004). Architecture of the *Bacteroides cellulosolvens* Cellulosome: Description of a Cell Surface-Anchoring Scaffoldin and a Family 48 Cellulase. *Journal of Bacteriology*, 186, 968–977. doi:10.1128/JB.186.4.968-977.2004
- Xu, Q., Gao, W., Ding, S., Kenig, R., Shoham, Y., Bayer, E. a, & Lamed, R. (2003). The Cellulosome System of *Acetivibrio cellulolyticus* Includes a Novel Type of Adaptor Protein and a Cell Surface Anchoring Protein The Cellulosome System of *Acetivibrio cellulolyticus* Includes a Novel Type of Adaptor Protein and a Cell Surface Anchoring P. *Journal of Bacteriology*, 185(15), 4548–4557. doi:10.1128/JB.185.15.4548
- Yang, B., & Wyman, C. E. (2008). Pretreatment : the key to unlocking low-cost cellulosic ethanol, 26–40. doi:10.1002/bbb
- Zhang, K., Cowtan, K., & Main, P. (1997). Combining constraints for electron-density modification. *Methods Enzymology*, 277, 53–64.
- Ziolkowska, J. R. (2014). Optimizing biofuels production in an uncertain decision environment: Conventional vs. advanced technologies. *Applied Energy*, 114, 366–376. doi:10.1016/j.apenergy.2013.09.060

Appendixes

Appendix A - Reagents and Solutions

1) Water

All the water used was certified for laboratory purposes. Distilled water was obtained from Elix®-5 Water Purification System (Millipore Corporation), showing a resistivity greater than 5 MΩ cm and a conductivity higher than 0.2 μ S/cm at 25°C. The bi-distilled water was obtained from Milli-Q® Water Purification System (Millipore Corporation), showing a resistivity higher than 18.2 MΩ cm and a conductivity higher than 0.05 μ S/cm.

2) Reagents

All reagents used are listed below in alphabetical order according to each supplier:

AppliChem

Cobalt chloride

Iron chloride

Nickel chloride

Biokar Diagnostics

Agar

Yeast extract

Sodium sulfite

Tryptone

Fluka

Hydroxyethylcellulose (HEC)

GE Healthcare

TEMED

Megazyme

Beta-Glucan (Barley)

Merk

Zinc sulphate heptahydrate

NZYTEch

Acrylamide

Panreak Química

Chloridric Acid

Sigma-Aldrich

Bovine serum albumin (BSA)

Ampicillin

Coomassie Bright blue

Glycerol

Sodium hydroxide

Imidazole

SDS

USB

Hepes

VWR International

Acetic acid

Citric acid

Ammonium chloride

Magnesium chloride

Manganese chloride tetrahydrate

Potassium chloride

Phenol

Potassium phosphate monobasic

Glucose

Methanol

Magnesium sulfate

Nickel sulfate

Sodium sulfate

3) Growth Medium

Growth Medium	Composition	
*LB (<i>Luria Bertani</i>)	Tryptone	10 g/l
	Yeast extract	5 g/l
	NaCl	10 g/l
LB agar	2 % (w/v) of agar added to LB medium	
**SOB	Tryptone	5 g/l
	Yeast extract	20 g/l
	NaCl	10 mmol/l
	KCl	2,5 mmol/l
	MgCl ₂	10 mmol/l
	MgSO ₄	10 mmol/l
SOC	Sterilized SOB	1 l
	2 % (w/v) Glucose solution	1 mol/l
	sterilized by membrane filtration 0,20 µm Gelman	

* pH corrected to 7,5 with NaOH solution 1 mol/l

** pH corrected to 7,0 with NaOH solution 1 mol/l

4) Autoinduction Growth Medium (LBE) for 1 l

Tryptone	10 g
Yeast extract	5 g
Glycerol	5 ml
Glucose	0,5 g
Lactose	2 g
Na ₂ SO ₄	0,7 g
NH ₄ Cl	2,5 g

Add distilled water to reach 900 ml final volume, mix and sterilize.

Later add 1 ml of sterilized solution of 2 M MgSO₄, 1 ml of metal mix 1000x (see below) and 100 ml of a mix solution of potassium phosphate sterilized and filtered (10 ml 1 M KH₂PO₄, 40 ml, 1 M K₂HPO₄ and 50 ml distilled water).

Metal Mix 1000x

0,1 M FeCl ₃ -6H ₂ O dissolved in 0,1 M HCl	50 ml
1 M MnCl ₂ -4H ₂ O	1 ml
1 M ZnSO ₄ -7H ₂ O	1 ml
0,2 M CoCl ₂ -6H ₂ O	1 ml
0,2 M NiCl ₂ -6H ₂ O	1 ml
Distilled water	46 ml

5) Buffers

Buffers	Composition	
10 mM Imidazol*	10 mM Imidazol	0.68 g/l
	50 mM NaHepes	11.916 g/l
	1 M NaCl	58.44 g/l
	5mM CaCl ₂	1 ml/l
60 mM Imidazol*	60 mM Imidazol	4.085 g/l
	50 mM NaHepes	11.916 g/l
	1 M NaCl	58.44 g/l
	5 mM CaCl ₂	1 ml/l
300 mM Imidazol*	300 mM Imidazol	20.424 g/l
	50 mM NaHepes	11.916 g/l
	1 M NaCl	58.44 g/l
	5mM CaCl ₂	1 ml/l
Cleaning **	NaCl	5.84 g/l
	NaPO ₄	2.4 g/l
	1 M EDTA (pH8)	100 ml/l
Gel Filtration*	50 mM NaHepes	11.916 g/l
	1 M NaCl	58.44 g/l
	5 mM CaCl ₂	1 ml/l

*Correct pH to 7.5 and filter with 0.2 µM filter.

** Correct pH to 7.4 and filter with 0.2 µM filter.

Filter and degas all solutions used for FPLC/Gel Filtration/ Ion exchange.

Appendix B - Gel Support Materials

1) SDS PAGE

Resolving Gel	Vt=5 ml		Stacking Gel	Vt=2.5 ml	
	14 %			4 %	
	Water	0.60 ml		Water	1.5 ml
	40 % Acrilamide/bisacrilamide	1.75 ml		40 % Acrilamide/bisacrilamide	0.25 ml
	50 % Glycerol	0.50 ml		Stacking Buffer (pH 6.8)	0.50 ml
	Resolution Buffer (pH 8.8)	0.50 ml		10 % APSO ₄	25 µl
	10 % APSO ₄	50 µl		TEMED	5µl
	TEMED	5µl			

2) Native

Vt= 5 ml	10 % Native Gel
Water	1.545 ml
40 % Acrilamide/bisacri.	1.25 ml
Glycerol	0.50 ml
Buffer	1.65 ml
CaCl ₂	25 µl
10 % APSO ₄	50 µl
Temed	5 µl

3) AG buffers

Buffers	Composition	
Sample Denaturing SDS buffer (for 20 ml)	SDS	2 g
	125 mM Tris (pH 6.8)	10 ml
	Glycerol 100 %	4 ml
	β -mercaptoethanol	6 ml
	Bromophenol Blue	2/3 grains
Sample Native (for 20 ml)	125 mM Tris (pH 6.8)	2.5 ml
	1.25 M Glycine	1.87 g
	Glycerol 100%	4 ml
	β -mercaptoethanol	6 ml
	Bromophenol Blue	2/3 grains
10x SDS Electrophoresis (1 l)	125 mM Tris (pH 6.8)	30.27 ml
	1.25 M Glycine	144.13 g
	SDS	10 g
10x Native Electrophoresis (1 l)	125 mM Tris (pH 6.8)	30.26 ml
	1.25 M Glycine	187.66 g

4) Bleaching Solution

20% Metanol and 5% Acetic Acid

Appendix C - Protocols Schematics

1) Mutants Construction

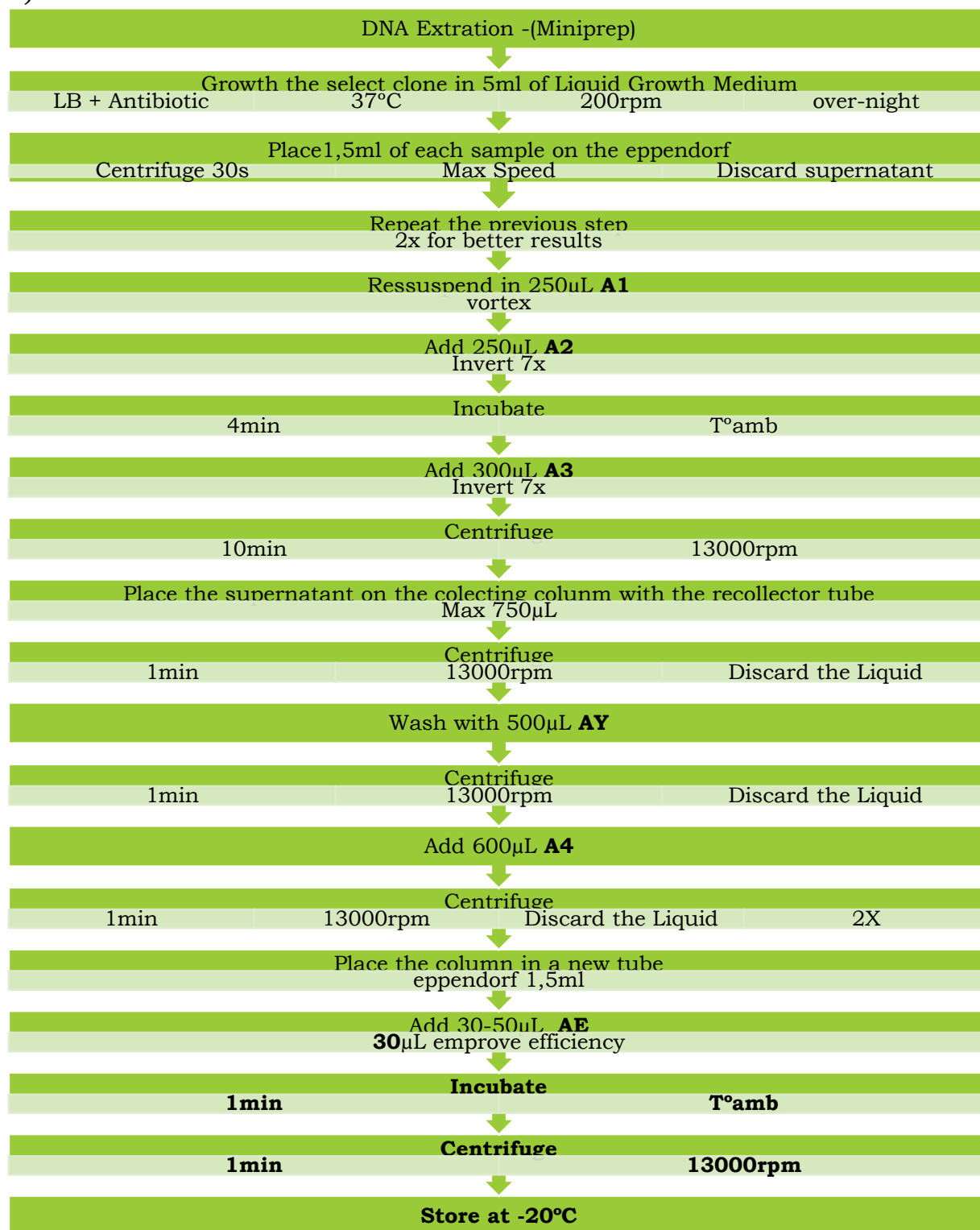


Figure 63- DNA Extraction Procedure

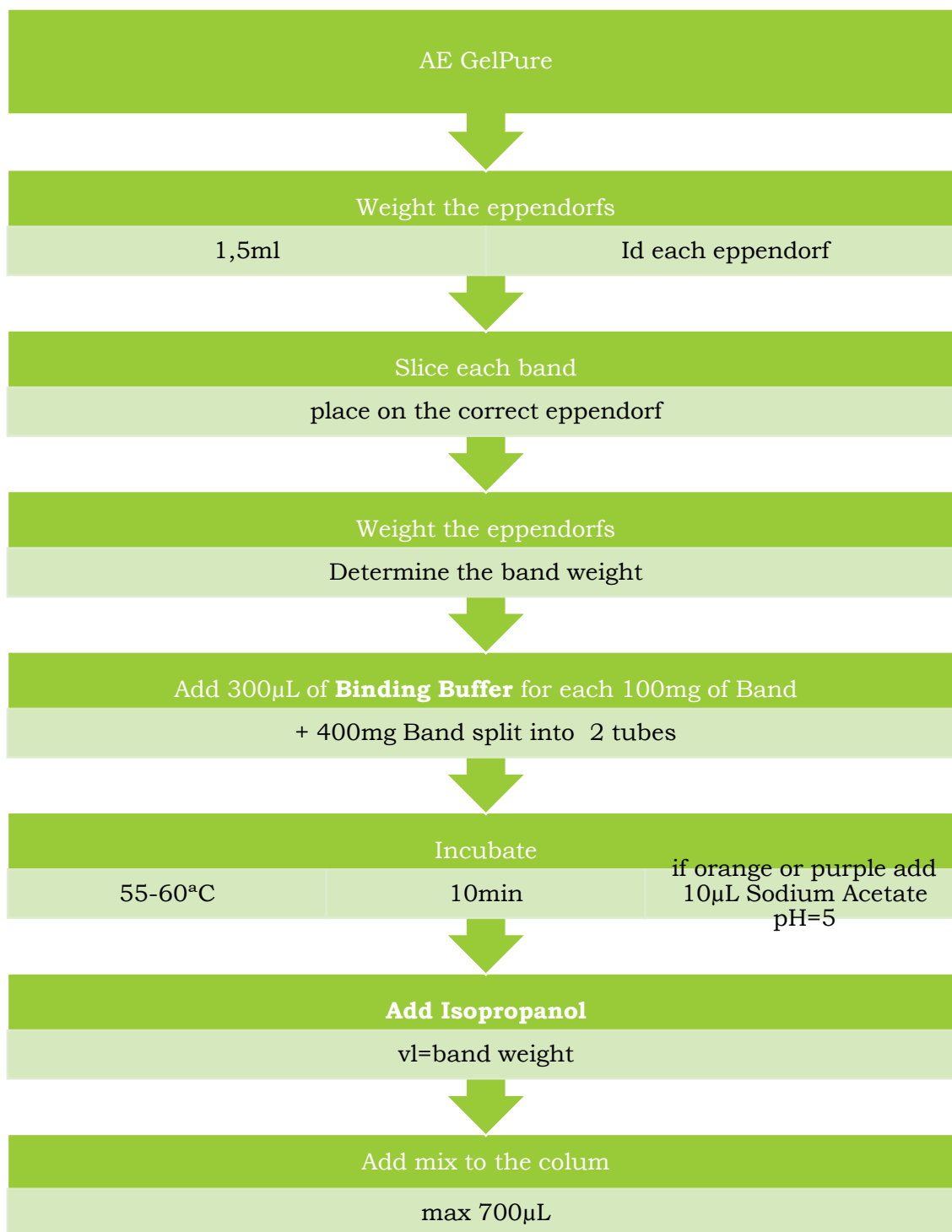


Figure 64 - AE Gel Purification Procedure 1 of 2

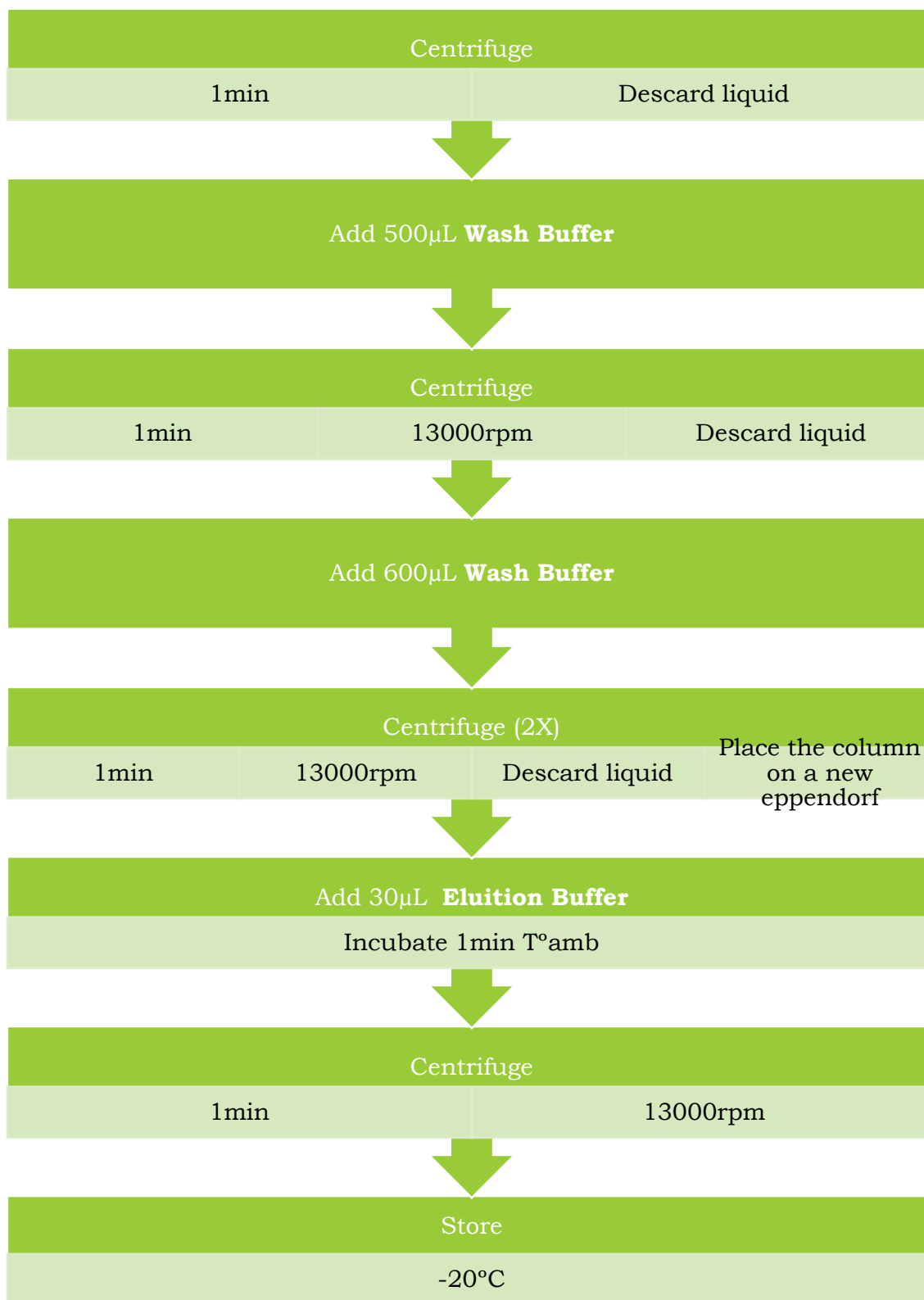


Figure 65- AE Gel Purification Procedure 2 of 2

2) Protein Expression

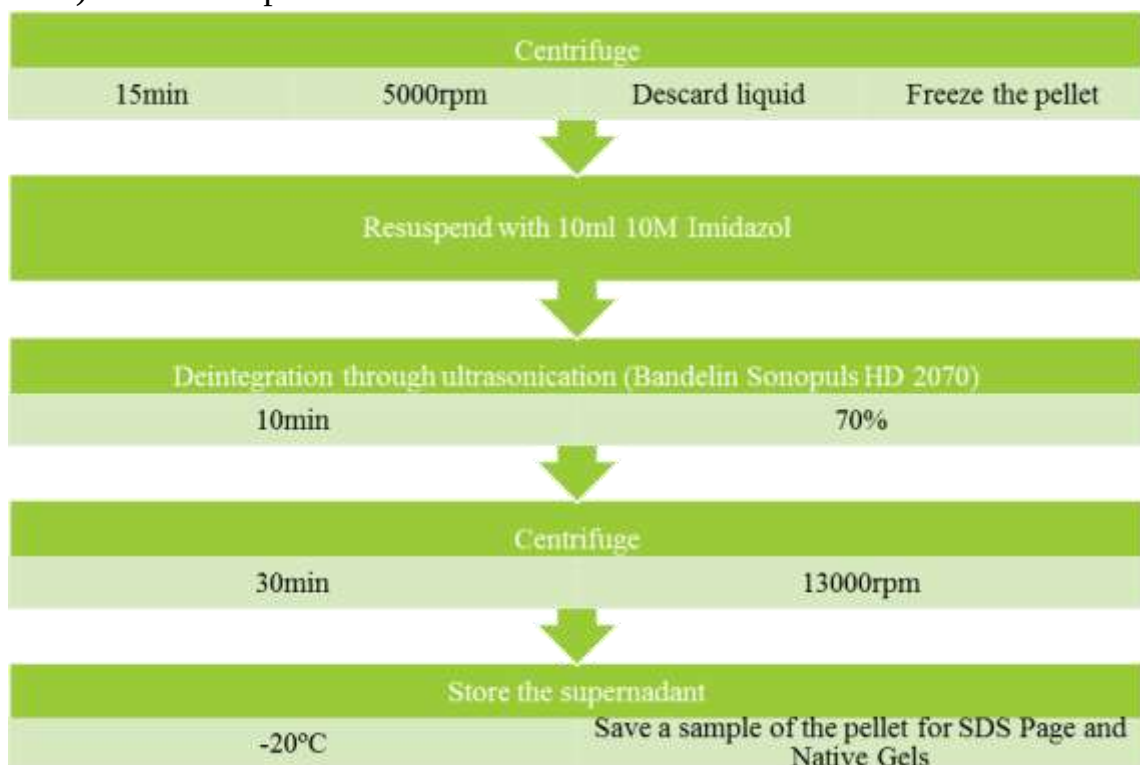


Figure 66- Protein extraction procedure

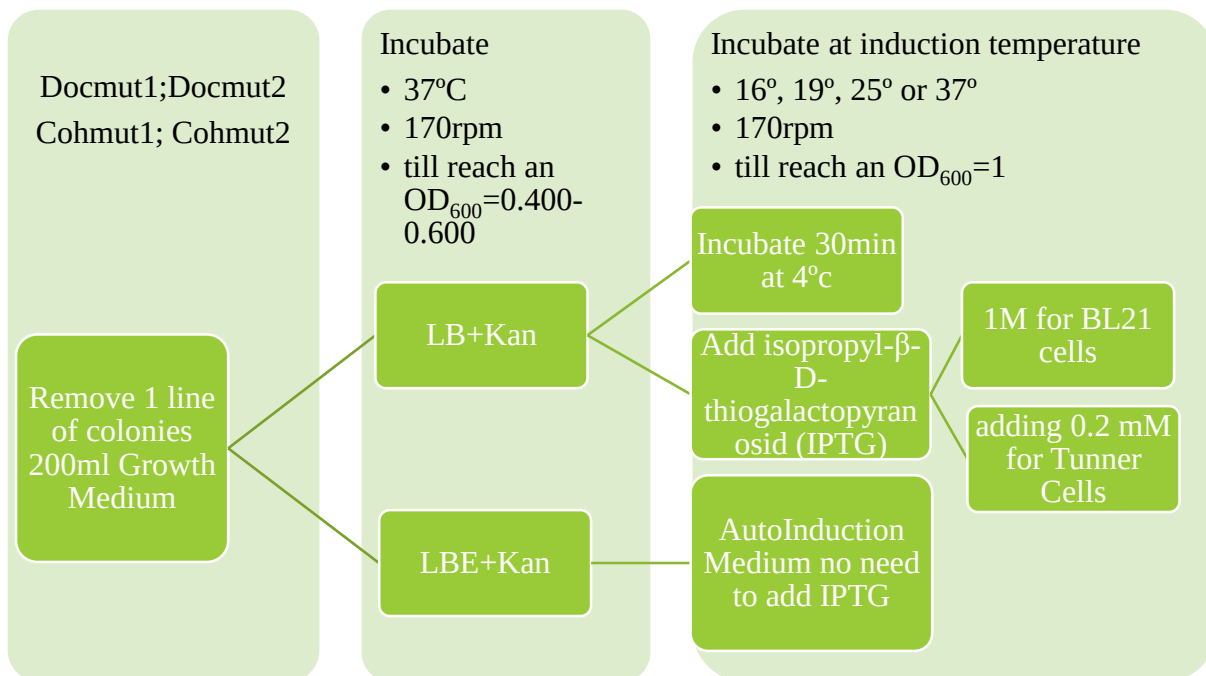


Figure 67- Protein Expression Procedure

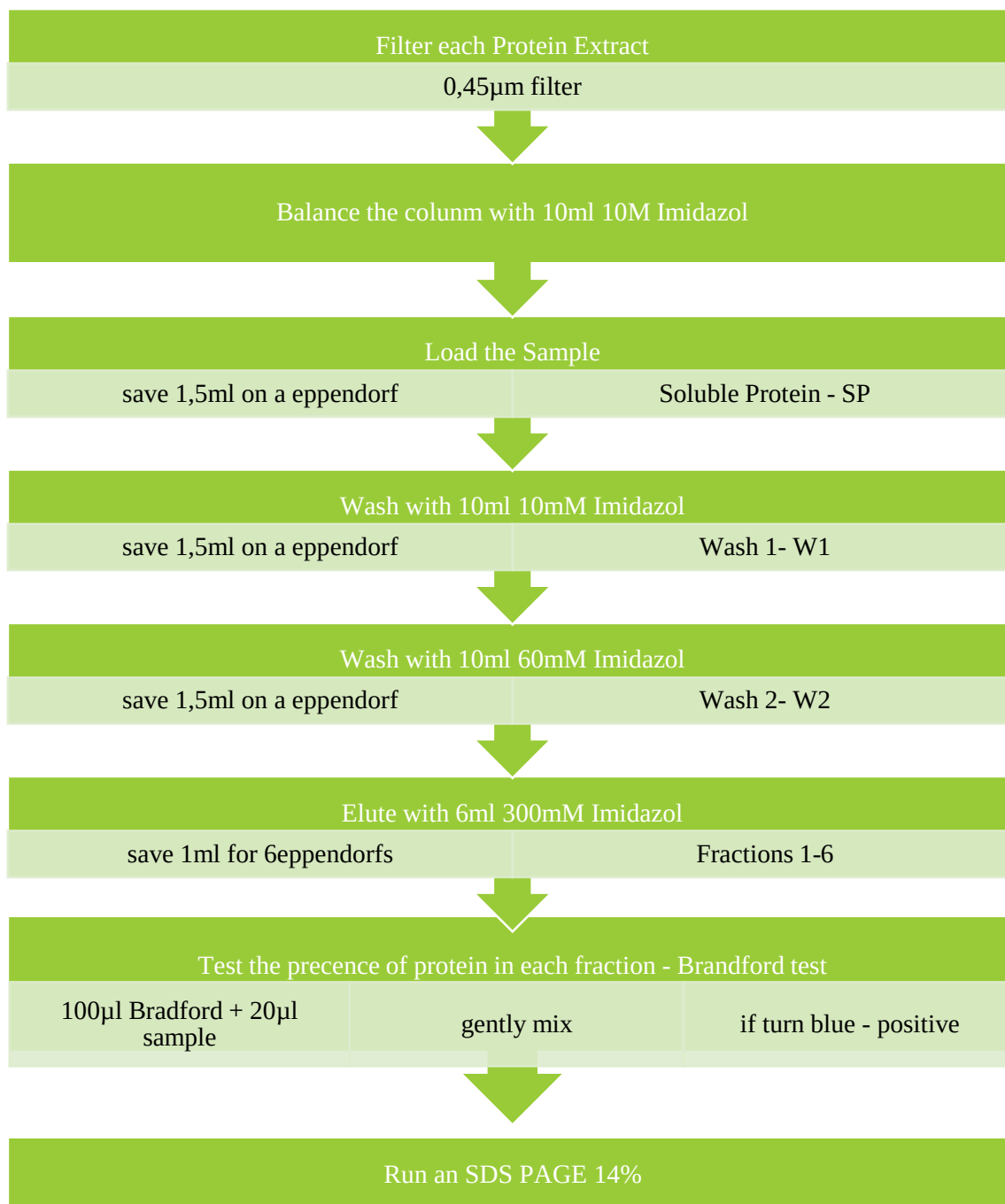


Figure 68 - Protein Purification Procedure

Substrate Affinity Support Data

1) Rf2

Table 23 - HEC AGE results

HEC	%Substrate	BSA(cm)	Rf2	R	R(n)	1/r
	0.000	9.7	5.2	0.54	1.00	1.00
	0.002	8.1	4.3	0.53	0.99	1.01
	0.010	8.5	4.4	0.52	0.97	1.04
	0.020	7.7	3.6	0.47	0.87	1.15
	0.025	7.9	4	0.51	0.94	1.06
	0.025	8.4	4	0.48	0.89	1.13
	0.030	7.7	4.6	0.60	1.11	0.90
	0.050	9.2	2.2	0.24	0.45	2.24
	0.050	9	2.9	0.32	0.60	1.66
	0.075	8.6	2.2	0.26	0.48	2.10
	0.100	9.1	1.5	0.16	0.31	3.25
	0.100	8	2	0.25	0.47	2.14
	0.150	9.9	2.4	0.24	0.45	2.21
	0.200	9.1	0.9	0.10	0.18	5.42
	0.200	10.3	1.6	0.16	0.29	3.45
	0.300	9.5	0.9	0.09	0.18	5.66
		Excluded samples				
	K _d	-0.05816				
	K _a	17.19454				

Table 24 - Xyloglucan AGE results

Xyloglucan	%Substrate	BSA(cm)	Rf2	R	R(n)	1/r
	0.0000	10.3	5.7	0.55	1.00	1.00
	0.0010	9	4.8	0.53	0.96	1.04
	0.0025	8.8	1	0.11	0.21	4.87
	0.0050	8.5	2	0.24	0.43	2.35
	0.0100	8.9	1.5	0.17	0.30	3.28
	0.0250	10.2	1.1	0.11	0.19	5.13
	0.0300	8.5	0.4	0.05	0.09	11.76
	0.0500	8.3	0.3	0.04	0.07	15.31
	0.1000	8.8	0.3	0.03	0.06	16.23
	0.2000	11.8	0.3	0.03	0.05	21.77
	0.3000	11.9	0.2	0.02	0.03	32.93
		Excluded samples				
	K _d	-0.02025				
	K _a	49.37611				

Table 25 - β -Glucan AGE results

B-Glucan	%Substrate	BSA(cm)	Rf2	R	R(n)	1/r
	0.000	8.20	4.50	0.55	1.00	1.00
	0.005	8.00	4.30	0.54	0.98	1.02
	0.020	8.50	3.20	0.38	0.69	1.46
	0.030	9.50	3.60	0.38	0.69	1.45
	0.050	8.80	2.00	0.23	0.41	2.41
	0.100	11.60	3.50	0.30	0.55	1.82
	0.150	9.10	3.00	0.33	0.60	1.66
	0.200	9.20	1.50	0.16	0.30	3.37
	0.300	9.60	1.00	0.10	0.19	5.27
		Excluded samples				
	K _d	-0.08941				
	K _a	11.18439				

Table 26 -Galactomann AGE results

Galactomann	%Substrate	BSA(cm)	Rf2	R	R(n)	1/r
	0.000	8.20	4.50	0.55	1.00	1.00
	0.010	9.60	1.90	0.20	0.36	2.77
	0.025	10.00	1.20	0.12	0.22	4.57
	0.040	9.50	0.90	0.09	0.17	5.79
	0.050	8.90	0.50	0.06	0.10	9.77
	0.075	8.60	0.40	0.05	0.08	11.80
	0.100	8.30	0.30	0.04	0.07	15.18
	0.200	9.50	0.20	0.02	0.04	26.07
	0.300	6.70	0.10	0.01	0.03	36.77
	K _d	-0.01774				
	K _a	56.36702				

2) Rf4

Table 27 - HEC AGE results

HEC	%Substrate	BSA(cm)	Rf2	R	R(n)	1/r
	0.000	8.2	6.6	0.805	1.000	1.000
	0.002	8.1	6.1	0.75	0.936	1.07
	0.010	8.5	6.1	0.72	0.892	1.12
	0.020	7.9	5.3	0.67	0.834	1.20
	0.025	7.9	6.1	0.77	0.959	1.04
	0.025	8.4	6.3	0.75	0.932	1.07
	0.030	7.9	5.1	0.65	0.802	1.25
	0.050	9.4	5.3	0.56	0.701	1.43
	0.050	9	8.5	0.94	1.173	0.85
	0.075	8.6	7.5	0.87	1.084	0.92
	0.100	9.1	6.6	0.73	0.901	1.11
	0.100	8	3.2	0.40	0.497	2.01
	0.150	9.9	4.7	0.47	0.590	1.70
	0.200	9.4	3.4	0.36	0.449	2.23
	0.200	10.1	2.3	0.23	0.283	3.53
	0.300	9.6	1.9	0.20	0.246	4.07
		Excluded samples				
	K _d	-0.09992				
	K _a	10.00801				

Table 28 - Xyloglucan AGE results

Xyloglucan	%Substrate	BSA(cm)	Rf2	R	R(n)	1/r
	0.000	8.2	7.1	0.866	1.076	0.930
	0.0010	9	7.6	0.84	0.98	1.03
	0.0025	8.7	6.1	0.70	0.81	1.23
	0.0050	8.5	6	0.71	0.82	1.23
	0.0100	8.8	6.1	0.69	0.80	1.25
	0.0250	10.2	0.6	0.06	0.07	14.72
	0.0300	8.5	2.9	0.34	0.39	2.54
	0.0500	8.3	1.5	0.18	0.21	4.79
	0.1000	8.7	0.4	0.05	0.05	18.83
	0.1000	9.3	5.5	0.59	0.68	1.46
	0.2000	11.7	0.3	0.03	0.03	33.77
	0.3000	11.9	0.2	0.02	0.02	51.52
		Excluded samples				
	K _d	0.003496				
	K _a	286.0412				

Table 29- β -Glucan AGE results

B-Glucan	%Substrate	BSA(cm)	Rf2	R	R(n)	1/r
	0.000	8.20	6.60	0.80	1.00	1.00
	0.005	8.00	6.20	0.78	0.96	1.04
	0.010	7.50	0.90	0.12	0.15	6.71
	0.020	8.10	0.90	0.11	0.14	7.24
	0.030	7.80	0.60	0.08	0.10	10.46
	0.050	8.80	2.00	0.23	0.28	3.54
	0.100	8.80	0.50	0.06	0.07	14.17
	0.150	9.70	0.30	0.03	0.04	26.02
	0.200	9.80	0.30	0.03	0.04	26.29
	0.300	8.80	0.20	0.02	0.03	35.41
		Excluded samples				
	K _d	-0.03671				
	K _a	27.24053				

Table 30-Galactomann AGE results

Galactomann	%Substrate	BSA(cm)	Rf2	R	R(n)	1/r
	0.000	8.20	6.60	0.80	1.00	1.00
	0.010	9.60	7.00	0.73	0.91	1.10
	0.025	10.00	7.50	0.75	0.93	1.07
	0.040	9.50	6.60	0.69	0.86	1.16
	0.050	8.70	6.00	0.69	0.86	1.17
	0.075	8.60	5.70	0.66	0.82	1.21
	0.100	8.70	5.60	0.64	0.80	1.25
	0.200	8.50	4.90	0.58	0.72	1.40
	0.300	7.30	4.00	0.55	0.68	1.47
		Excluded samples				
	K _d	-0.7192				
	K _a	1.390434				

Crystal Screen 2™

HR2-112 Reagent Formulation

Tube #	Salt	Tube #	Buffer †	Tube #	Precipitant
1.	2.0 M Sodium chloride	1.	None	1.	10% w/v Polyethylene glycol 6,000
2.	0.5 M Sodium chloride	2.	None	2.	0.01 M Hexadecyltrimethylammonium bromide
3.	0.01 M Magnesium chloride hexahydrate	3.	None	3.	25% v/v Ethylene glycol
4.	None	4.	None	4.	35% v/v 1,4-Dioxane
5.	2.0 M Ammonium sulfate	5.	None	5.	5% v/v 2-Propanol
6.	None	6.	None	6.	1.0 M Imidazole pH 7.0
7.	None	7.	None	7.	10% w/v Polyethylene glycol 1,000
8.	1.5 M Sodium chloride	8.	None	8.	10% w/v Polyethylene glycol 8,000
9.	None	9.	0.1 M Sodium acetate trihydrate pH 4.6	9.	10% v/v Ethanol
10.	0.5 M Sodium chloride	10.	0.1 M Sodium acetate trihydrate pH 4.6	10.	2.0 M Sodium chloride
11.	0.01 M Cobalt(II) chloride hexahydrate	11.	0.1 M Sodium acetate trihydrate pH 4.6	11.	30% v/v (+/-)-2-Methyl-2,4-pentanediol
12.	0.1 M Cadmium chloride hydrate	12.	0.1 M Sodium acetate trihydrate pH 4.6	12.	1.0 M 1,5-Hexanediol
13.	0.2 M Ammonium sulfate	13.	0.1 M Sodium acetate trihydrate pH 4.6	13.	30% v/v Polyethylene glycol 400
14.	0.2 M Potassium sodium tartrate tetrahydrate	14.	0.1 M Sodium citrate tribasic dihydrate pH 5.6	14.	30% w/v Polyethylene glycol monomethyl ether 2,000
15.	0.5 M Ammonium sulfate	15.	0.1 M Sodium citrate tribasic dihydrate pH 5.6	15.	2.0 M Ammonium sulfate
16.	0.5 M Sodium chloride	16.	0.1 M Sodium citrate tribasic dihydrate pH 5.6	16.	1.0 M Lithium sulfate monohydrate
17.	None	17.	0.1 M Sodium citrate tribasic dihydrate pH 5.6	16.	2% v/v Ethylene imine polymer
18.	0.01 M Iron(III) chloride hexahydrate	18.	0.1 M Sodium citrate tribasic dihydrate pH 5.6	17.	35% v/v tert-Butanol
19.	None	19.	0.1 M Sodium citrate tribasic dihydrate pH 5.6	18.	10% v/v Jeffamine® M-600®
20.	None	20.	0.1 M MES monohydrate pH 6.5	19.	2.5 M 1,5-Hexanediol
21.	0.1 M Sodium phosphate monobasic monohydrate	21.	0.1 M MES monohydrate pH 6.5	20.	1.6 M Magnesium sulfate heptahydrate
22.	0.1 M Potassium phosphate monobasic	22.	0.1 M MES monohydrate pH 6.5	21.	2.0 M Sodium chloride
23.	1.6 M Ammonium sulfate	23.	0.1 M MES monohydrate pH 6.5	22.	12% w/v Polyethylene glycol 20,000
24.	0.05 M Cesium chloride	24.	0.1 M MES monohydrate pH 6.5	23.	10% v/v 1,4-Dioxane
25.	0.01 M Cobalt(II) chloride hexahydrate	25.	0.1 M MES monohydrate pH 6.5	24.	30% v/v Jeffamine® M-600®
26.	0.2 M Ammonium sulfate	26.	0.1 M MES monohydrate pH 6.5	25.	1.6 M Ammonium sulfate
27.	0.01 M Zinc sulfate heptahydrate	27.	0.1 M MES monohydrate pH 6.5	26.	30% w/v Polyethylene glycol monomethyl ether 5,000
28.	None	28.	None	27.	25% v/v Polyethylene glycol monomethyl ether 550
29.	0.5 M Ammonium sulfate	29.	0.1 M HEPES pH 7.5	28.	1.6 M Sodium citrate tribasic dihydrate pH 6.5
30.	None	30.	0.1 M HEPES pH 7.5	29.	30% v/v (+/-)-2-Methyl-2,4-pentanediol
31.	None	31.	0.1 M HEPES pH 7.5	30.	10% w/v Polyethylene glycol 6,000
32.	0.1 M Sodium chloride	32.	0.1 M HEPES pH 7.5	31.	5% v/v (+/-)-2-Methyl-2,4-pentanediol
33.	None	33.	0.1 M HEPES pH 7.5	32.	20% v/v Jeffamine® M-600®
34.	0.05 M Cadmium sulfate hydrate	34.	0.1 M HEPES pH 7.5	33.	1.6 M Ammonium sulfate
35.	None	35.	0.1 M HEPES pH 7.5	34.	2.0 M Ammonium formate
36.	None	36.	0.1 M HEPES pH 7.5	35.	1.0 M Sodium acetate trihydrate
37.	None	37.	0.1 M HEPES pH 7.5	36.	70% v/v (+/-)-2-Methyl-2,4-pentanediol
38.	None	38.	0.1 M HEPES pH 7.5	37.	4.3 M Sodium chloride
39.	0.2 M Magnesium chloride hexahydrate	39.	0.1 M Tris pH 8.5	38.	10% w/v Polyethylene glycol 8,000
40.	None	40.	0.1 M Tris pH 8.5	39.	8% v/v Ethylene glycol
41.	0.01 M Nickel(II) chloride hexahydrate	41.	0.1 M Tris pH 8.5	40.	20% w/v Polyethylene glycol 10,000
42.	1.5 M Ammonium sulfate	42.	0.1 M Tris pH 8.5	41.	3.4 M 1,5-Hexanediol
43.	0.2 M Ammonium phosphate monobasic	43.	0.1 M Tris pH 8.5	42.	25% v/v tert-Butanol
44.	None	44.	0.1 M Tris pH 8.5	43.	1.0 M Lithium sulfate monohydrate
45.	0.01 M Nickel(II) chloride hexahydrate	45.	0.1 M Tris pH 8.5	44.	12% v/v Glycerol
46.	0.1 M Sodium chloride	46.	0.1 M BICINE pH 9.0	45.	50% v/v (+/-)-2-Methyl-2,4-pentanediol
47.	None	47.	0.1 M BICINE pH 9.0	46.	20% v/v Ethanol
48.	None	48.	0.1 M BICINE pH 9.0	47.	20% w/v Polyethylene glycol monomethyl ether 2,000
				48.	20% v/v Polyethylene glycol monomethyl ether 550
					2.0 M Magnesium chloride hexahydrate
					2% v/v 1,4-Dioxane
					10% w/v Polyethylene glycol 20,000

† Buffer pH is that of a 1.0 M (0.5 M for MES monohydrate) stock prior to dilution with other reagent components: pH with HCl or NaOH.

Crystal Screen 2 contains forty-eight unique reagents. To determine the formulation of each reagent, simply read across the page.

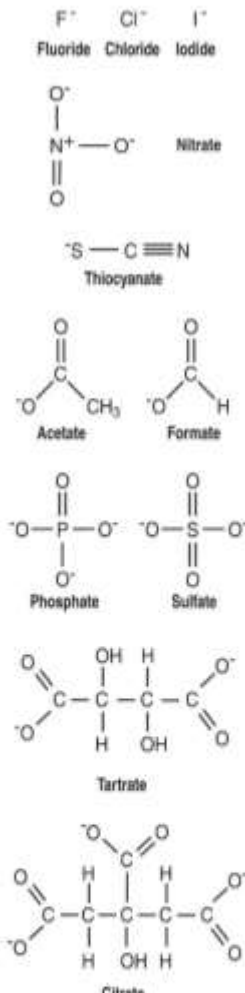
34 Journey
Aliso Viejo, CA 92656-3317 U.S.A.
Tel: (949) 425-1321 • Fax: (949) 425-1611
E-mail: tech@hamptonresearch.com
Website: www.hamptonresearch.com



Solutions for Crystal Growth
© 1991 - 2010 Hampton Research Corp. All rights reserved.
Printed in the United States of America. This guide or
parts thereof may not be reproduced in any form without
the written permission of the publisher.

2 - Peg/Ion I&II

PEG/Ion Screen™			HR2-126 Reagent Formulation		
Tube #	Salt	Tube #	Polymer	Tube #	pH 0
1.	0.2 M Sodium fluoride	1.	20% w/v Polyethylene glycol 3,350	1.	7.3
2.	0.2 M Potassium fluoride	2.	20% w/v Polyethylene glycol 3,350	2.	7.3
3.	0.2 M Ammonium fluoride	3.	20% w/v Polyethylene glycol 3,350	3.	6.2
4.	0.2 M Lithium chloride	4.	20% w/v Polyethylene glycol 3,350	4.	6.8
5.	0.2 M Magnesium chloride hexahydrate	5.	20% w/v Polyethylene glycol 3,350	5.	5.9
6.	0.2 M Sodium chloride	6.	20% w/v Polyethylene glycol 3,350	6.	6.9
7.	0.2 M Calcium chloride dihydrate	7.	20% w/v Polyethylene glycol 3,350	7.	5.1
8.	0.2 M Potassium chloride	8.	20% w/v Polyethylene glycol 3,350	8.	7.0
9.	0.2 M Ammonium chloride	9.	20% w/v Polyethylene glycol 3,350	9.	6.3
10.	0.2 M Sodium iodide	10.	20% w/v Polyethylene glycol 3,350	10.	7.0
11.	0.2 M Potassium iodide	11.	20% w/v Polyethylene glycol 3,350	11.	7.0
12.	0.2 M Ammonium iodide	12.	20% w/v Polyethylene glycol 3,350	12.	6.2
13.	0.2 M Sodium thiocyanate	13.	20% w/v Polyethylene glycol 3,350	13.	6.9
14.	0.2 M Potassium thiocyanate	14.	20% w/v Polyethylene glycol 3,350	14.	7.0
15.	0.2 M Lithium nitrate	15.	20% w/v Polyethylene glycol 3,350	15.	7.1
16.	0.2 M Magnesium nitrate hexahydrate	16.	20% w/v Polyethylene glycol 3,350	16.	5.9
17.	0.2 M Sodium nitrate	17.	20% w/v Polyethylene glycol 3,350	17.	6.8
18.	0.2 M Potassium nitrate	18.	20% w/v Polyethylene glycol 3,350	18.	6.8
19.	0.2 M Ammonium nitrate	19.	20% w/v Polyethylene glycol 3,350	19.	6.2
20.	0.2 M Magnesium formate dihydrate	20.	20% w/v Polyethylene glycol 3,350	20.	7.0
21.	0.2 M Sodium formate	21.	20% w/v Polyethylene glycol 3,350	21.	7.2
22.	0.2 M Potassium formate	22.	20% w/v Polyethylene glycol 3,350	22.	7.3
23.	0.2 M Ammonium formate	23.	20% w/v Polyethylene glycol 3,350	23.	6.6
24.	0.2 M Lithium acetate dihydrate	24.	20% w/v Polyethylene glycol 3,350	24.	7.9
25.	0.2 M Magnesium acetate tetrahydrate	25.	20% w/v Polyethylene glycol 3,350	25.	7.9
26.	0.2 M Zinc acetate dihydrate	26.	20% w/v Polyethylene glycol 3,350	26.	6.4
27.	0.2 M Sodium acetate trihydrate	27.	20% w/v Polyethylene glycol 3,350	27.	8.0
28.	0.2 M Calcium acetate hydrate	28.	20% w/v Polyethylene glycol 3,350	28.	7.5
29.	0.2 M Potassium acetate	29.	20% w/v Polyethylene glycol 3,350	29.	8.1
30.	0.2 M Ammonium acetate	30.	20% w/v Polyethylene glycol 3,350	30.	7.1
31.	0.2 M Lithium sulfate monohydrate	31.	20% w/v Polyethylene glycol 3,350	31.	6.0
32.	0.2 M Magnesium sulfate heptahydrate	32.	20% w/v Polyethylene glycol 3,350	32.	6.0
33.	0.2 M Sodium sulfate decahydrate	33.	20% w/v Polyethylene glycol 3,350	33.	6.7
34.	0.2 M Potassium sulfate	34.	20% w/v Polyethylene glycol 3,350	34.	6.8
35.	0.2 M Ammonium sulfate	35.	20% w/v Polyethylene glycol 3,350	35.	6.0
36.	0.2 M Sodium tartrate dibasic dihydrate	36.	20% w/v Polyethylene glycol 3,350	36.	7.3
37.	0.2 M Potassium sodium tartrate tetrahydrate	37.	20% w/v Polyethylene glycol 3,350	37.	7.4
38.	0.2 M Ammonium tartrate dibasic	38.	20% w/v Polyethylene glycol 3,350	38.	6.6
39.	0.2 M Sodium phosphate monobasic monohydrate	39.	20% w/v Polyethylene glycol 3,350	39.	4.7
40.	0.2 M Sodium phosphate dibasic dihydrate	40.	20% w/v Polyethylene glycol 3,350	40.	9.1
41.	0.2 M Potassium phosphate monobasic	41.	20% w/v Polyethylene glycol 3,350	41.	4.8
42.	0.2 M Potassium phosphate dibasic	42.	20% w/v Polyethylene glycol 3,350	42.	9.2
43.	0.2 M Ammonium phosphate monobasic	43.	20% w/v Polyethylene glycol 3,350	43.	4.6
44.	0.2 M Ammonium phosphate dibasic	44.	20% w/v Polyethylene glycol 3,350	44.	8.0
45.	0.2 M Lithium citrate tribasic tetrahydrate	45.	20% w/v Polyethylene glycol 3,350	45.	8.4
46.	0.2 M Sodium citrate tribasic dihydrate	46.	20% w/v Polyethylene glycol 3,350	46.	8.3
47.	0.2 M Potassium citrate tribasic monohydrate	47.	20% w/v Polyethylene glycol 3,350	47.	8.3
48.	0.2 M Ammonium citrate dibasic	48.	20% w/v Polyethylene glycol 3,350	48.	5.1



0 Measured pH at 25 ° C

PEG/Ion Screen contains forty-eight unique reagents. To determine the formulation of each reagent, simply read across the page.

34 Journey
 Aliso Viejo, CA 92656-3317 U.S.A.
 Tel: (949) 425-1321 • Fax: (949) 425-1611
 E-mail: tech@hmr.com
 Website: www.hamptonresearch.com

HAMPTON
 RESEARCH
Solutions for Crystal Growth

© 1991 - 2015 Hampton Research Corp. All rights reserved.
 Printed in the United States of America. This guide or
 parts thereof may not be reproduced in any form without
 the written permission of the publisher.

PEG/Ion 2 Screen™

HR2-098 Reagent Formulation

Tube #	Salt	Tube #	Buffer ^o	Tube #	Polymer
1.	0.1 M Sodium malonate pH 4.0	1.	None	1.	12% w/v Polyethylene glycol 3,350
2.	0.2 M Sodium malonate pH 4.0	2.	None	2.	20% w/v Polyethylene glycol 3,350
3.	0.1 M Sodium malonate pH 5.0	3.	None	3.	12% w/v Polyethylene glycol 3,350
4.	0.2 M Sodium malonate pH 5.0	4.	None	4.	20% w/v Polyethylene glycol 3,350
5.	0.1 M Sodium malonate pH 6.0	5.	None	5.	12% w/v Polyethylene glycol 3,350
6.	0.2 M Sodium malonate pH 6.0	6.	None	6.	20% w/v Polyethylene glycol 3,350
7.	0.1 M Sodium malonate pH 7.0	7.	None	7.	12% w/v Polyethylene glycol 3,350
8.	0.2 M Sodium malonate pH 7.0	8.	None	8.	20% w/v Polyethylene glycol 3,350
9.	4% v/v Tacsimate™ pH 4.0	9.	None	9.	12% w/v Polyethylene glycol 3,350
10.	8% v/v Tacsimate™ pH 4.0	10.	None	10.	20% w/v Polyethylene glycol 3,350
11.	4% v/v Tacsimate™ pH 5.0	11.	None	11.	12% w/v Polyethylene glycol 3,350
12.	8% v/v Tacsimate™ pH 5.0	12.	None	12.	20% w/v Polyethylene glycol 3,350
13.	4% v/v Tacsimate™ pH 6.0	13.	None	13.	12% w/v Polyethylene glycol 3,350
14.	8% v/v Tacsimate™ pH 6.0	14.	None	14.	20% w/v Polyethylene glycol 3,350
15.	4% v/v Tacsimate™ pH 7.0	15.	None	15.	12% w/v Polyethylene glycol 3,350
16.	8% v/v Tacsimate™ pH 7.0	16.	None	16.	20% w/v Polyethylene glycol 3,350
17.	4% v/v Tacsimate™ pH 8.0	17.	None	17.	12% w/v Polyethylene glycol 3,350
18.	8% v/v Tacsimate™ pH 8.0	18.	None	18.	20% w/v Polyethylene glycol 3,350
19.	0.1 M Succinic acid pH 7.0	19.	None	19.	12% w/v Polyethylene glycol 3,350
20.	0.2 M Succinic acid pH 7.0	20.	None	20.	20% w/v Polyethylene glycol 3,350
21.	0.1 M Ammonium citrate tribasic pH 7.0	21.	None	21.	12% w/v Polyethylene glycol 3,350
22.	0.2 M Ammonium citrate tribasic pH 7.0	22.	None	22.	20% w/v Polyethylene glycol 3,350
23.	0.1 M DL-Malic acid pH 7.0	23.	None	23.	12% w/v Polyethylene glycol 3,350
24.	0.2 M DL-Malic acid pH 7.0	24.	None	24.	20% w/v Polyethylene glycol 3,350
25.	0.1 M Sodium acetate trihydrate pH 7.0	25.	None	25.	12% w/v Polyethylene glycol 3,350
26.	0.2 M Sodium acetate trihydrate pH 7.0	26.	None	26.	20% w/v Polyethylene glycol 3,350
27.	0.1 M Sodium formate pH 7.0	27.	None	27.	12% w/v Polyethylene glycol 3,350
28.	0.2 M Sodium formate pH 7.0	28.	None	28.	20% w/v Polyethylene glycol 3,350
29.	0.1 M Ammonium tartrate dibasic pH 7.0	29.	None	29.	12% w/v Polyethylene glycol 3,350
30.	0.2 M Ammonium tartrate dibasic pH 7.0	30.	None	30.	20% w/v Polyethylene glycol 3,350
31.	2% v/v Tacsimate™ pH 4.0	31.	0.1 M Sodium acetate trihydrate pH 4.6	31.	16% w/v Polyethylene glycol 3,350
32.	2% v/v Tacsimate™ pH 5.0	32.	0.1 M Sodium citrate tribasic dihydrate pH 5.6	32.	16% w/v Polyethylene glycol 3,350
33.	2% v/v Tacsimate™ pH 6.0	33.	0.1 M BIS-TRIS pH 6.5	33.	20% w/v Polyethylene glycol 3,350
34.	2% v/v Tacsimate™ pH 7.0	34.	0.1 M HEPES pH 7.5	34.	20% w/v Polyethylene glycol 3,350
35.	2% v/v Tacsimate™ pH 8.0	35.	0.1 M Tris pH 8.5	35.	16% w/v Polyethylene glycol 3,350
36.	None	36.	0.07 M Citric acid, 0.03 M BIS-TRIS propane / pH 3.4	36.	16% w/v Polyethylene glycol 3,350
37.	None	37.	0.06 M Citric acid, 0.04 M BIS-TRIS propane / pH 4.1	37.	16% w/v Polyethylene glycol 3,350
38.	None	38.	0.05 M Citric acid, 0.05 M BIS-TRIS propane / pH 5.0	38.	16% w/v Polyethylene glycol 3,350
39.	None	39.	0.04 M Citric acid, 0.06 M BIS-TRIS propane / pH 6.4	39.	20% w/v Polyethylene glycol 3,350
40.	None	40.	0.03 M Citric acid, 0.07 M BIS-TRIS propane / pH 7.6	40.	20% w/v Polyethylene glycol 3,350
41.	None	41.	0.02 M Citric acid, 0.08 M BIS-TRIS propane / pH 8.8	41.	16% w/v Polyethylene glycol 3,350
42.	0.02 M Calcium chloride dihydrate, 0.02 M Cadmium chloride hydrate, 0.02 M Cobalt(II) chloride hexahydrate	42.	None	42.	20% w/v Polyethylene glycol 3,350
43.	0.01 M Magnesium chloride hexahydrate 0.005 M Nickel(II) chloride hexahydrate	43.	0.1 M HEPES sodium pH 7.0	43.	15% w/v Polyethylene glycol 3,350
44.	0.02 M Zinc chloride	44.	None	44.	20% w/v Polyethylene glycol 3,350
45.	0.15 M Cesium chloride	45.	None	45.	15% w/v Polyethylene glycol 3,350
46.	0.2 M Sodium bromide	46.	None	46.	20% w/v Polyethylene glycol 3,350
47.	1% w/v Tryptone, 0.001 M Sodium azide	47.	0.05 M HEPES sodium pH 7.0	47.	12% w/v Polyethylene glycol 3,350
48.	1% w/v Tryptone, 0.001 M Sodium azide	48.	0.05 M HEPES sodium pH 7.0	48.	20% w/v Polyethylene glycol 3,350

^o Buffer pH is that of a 1.0 M stock prior to dilution with other reagent components: pH with HCl or NaOH.

PEG/Ion 2 Screen contains forty-eight unique reagents. To determine the formulation of each reagent, simply read across the page.

34 Journey
Aliso Viejo, CA 92656-3317 U.S.A.
Tel: (949) 425-1321 • Fax: (949) 425-1611
E-mail: tech@hamptonresearch.com
Website: www.hamptonresearch.com



Solutions for Crystal Growth

© 1991-2015 Hampton Research Corp. all rights reserved
Printed in the United States of America. This guide or
parts thereof may not be reproduced in any form without
the written permission of the publisher.

3 - 80!

A1 80l 1- 0.2 M CaCl ₂ 0.1 M Act 4.5 30 % MPD	A2 80l 2- 1 M Tart K/Na 0.1M MES 6.5	A3 80l 3- 0.4M F.A.	A4 80l 4- 0.1M TRIS 8.5 3M S.A.	A5 80l 5- 0.2M CitrNa ₃ 0.1MHEPES 7.5 30 % MPD	A6 80l 6- 0.2 M MgCl ₂ 0.1 M Act 4.5 30 % PEG 4K	A7 80l 7- 1.2M CitrNa ₃ 0.1MHEPES 7.5	A8 80l 8-0.2M CitrNa ₃ 2 M (NH ₄) ₂ SO ₄	A9 80l 9- 0.2 M ActNH ₄ 0.1 M Cit 5.5 30 % PEG 400	A10 80l 10- 0.1M Act4.5 1.5 M F.A.	A11 80l 11- 0.2M S.A. 0.1MHEPES 7.5 1.5 M K ₂ HPO ₄ 1.5 M NaH ₂ PO ₄	A12 80l 12- 0.2 M CitrNa ₃ 0.1 M TRIS 8.5 20 % PEG 400
B1 80l 13- 0.2 M CaCl ₂ 0.1MHEPES 7.5 25 % PEG 4K	B2 80l 14-0.1 M MgCl ₂ 0.1 M MES 6.5 30 % PEG 8K	B3 80l 15- 0.2M Li ₂ SO ₄ 0.1 M Cit 5.5 30 % PEG 4K	B4 80l 16- 1 M Li ₂ SO ₄ 0.1 M Act 4.5	B5 80l 17- 0.2 M F.A. 0.1 M TRIS 7.5 30 % MPD	B6 80l 18- 0.2M ActNH ₄ 0.1 M TRIS 7.5 1.5 M K ₂ HPO ₄ 1.5 M NaH ₂ PO ₄	B7 80l 19- 0.1 M S.A. 0.1 M Cit 5.5 30 % PEG 8K	B8 80l 20- 0.1M MES 6.5 30 % MPD	B9 80l 21-0.2 M MgCl ₂ 0.1MHEPES 7.5 30 % PEG 4K	B10 80l 22- 0.2 M ActNa 0.1 M TRIS 8.5 30 % PEG 4K	B11 80l 23- 0.1M TRIS 7.5 1 M Tart K/Na	B12 80l 24- 0.2 M CaCl ₂ 0.1 M TRIS 8.5
C1 80l 25- 0.5M ActNH ₄ 0.1 M Cit 5.5 30 % MPD	C2 80l 26- 2 M ActNa 0.1 M MES 6.5	C3 80l 27- 0.2M Tart K/Na 0.1 M MES 6.5 30 % P 8K	C4 80l 28- 1 M Tart K/Na 0.1MHEPES 7.5	C5 80l 29- 0.2 M S.A. 0.1 M Act 4.5 30 % P 400	C6 80l 30- 0.1 M S.A. 0.1MHEPES 7.5 20 % P 4K	C7 80l 31- 2 M S.A. 0.1 M MES 6.5	C8 80l 32- 0.2 M NaCl 0.1 M MES 6.5 30 % Etanol	C9 80l 33-0.2 M MgCl ₂ 0.1MHEPES 7.5 30 % Etanol	10 80l 34- 0.2 M ActNH ₄ 0.1 M TRIS 8.5 30 % Etanol	C11 80l 35- 0.2 M CaCl ₂ 0.1 M Act 4.5 30 % Etanol	C12 80l 36- 0.2 M ActNa 0.1MHEPES 7.5 30 % Etanol
D1 80l 37-0.2 M MgCl ₂ 0.1MHEPES 7.5 30 % Isopp	D2 80l 38- 0.1 M Cac 6.5 30 % MPD	D3 80l 39- 0.1 M Act 4.5 2 M FormNa	D4 80l 40- 0.2M CitrNa ₃ 0.1 M Cac 6.5 40 % Isopp	D5 80l 41- 0.1MHEPES 7.5 20 % PEG 400 10 % Isopp	D6 80l 42- 0.1MHEPES 7.5 1 M Li ₂ SO ₄	D7 80l 43-0.2 M Li ₂ SO ₄ 0.1 M TRIS 8.5 30 % PEG 4K	D8 80l 44- 0.2 M S.A. 0.1 M Cac 6.5 30 % PEG 6K	D9 80l 45- 0.1M Act 4.5 1.5 M ActNa	D10 80l 46- 0.1M CitrNa ₃ 1 M NH ₄ H ₂ PO ₄	D11 80l 47- 4 M FormNa	D12 80l 48- 0.1MHEPES 7.5 1.2 M CitNa ₃
E1 80l 49- 0.4M Tart K/Na	E2 80l 50-0.2 M MgCl ₂ 0.1 M TRIS 8.5 30 % PEG 4K	E3 80l 51- 0.1 M Cac 6.5 1.4 M ActNa	E4 80l 52- 0.2M ActNH ₄ 0.1 M Cit 5.5 30 % PEG 4K	E5 80l 53- 0.2M ActNH ₄ 0.1 M Act 4.5 30 % PEG 4K	E6 80l 54- 0.2 M CaCl ₂ 0.1MHEPES 7.5 28 % PEG 400	E780l 55-0.2M S.A. 0.1M Cac 6.5 30 % PEG 8K	E8 80l 56- 0.2M ActMg 0.1 M Cac 6.5 20 % PEG 8K	E9 80l 57- 0.2M ActNH ₄ 0.1 M TRIS 8.5 30 % Isopp	E10 80l 58- 0.2 M S.A. 0.1 M Act 4.5 25 % PEG 4K	E11 80l 59-0.2M ActMg 0.1 M Cac 6.5 30 % MPD	E12 80l 60- 0.2 M CaCl ₂ 0.1 M Act 4.5 20 % Isopp
F1 80l 61- 0.1M Imid 7: 1 M ActNa	F2 80l 62- 0.2 M CitrNa ₃ 0.1 M Cac 6.5 20 % Isopp	F3 80l 63- 0.2 M ActNa 0.1 M Cac 6.5 30 % PEG 8K	F4 80l 64- 0.2 M S.A. 30 % PEG 8K	F5 80l 65- 0.2 M S.A. 30 % PEG 4K	F6 80l 66- 0.1MHEPES 7.5 1.6 M K ₂ HPO ₄ 1.6 M NaH ₂ PO ₄	F7 80l 67- 0.1 M TRIS 8.5 8 % PEG 8K	F8 80l 68- 0.1M Act4.5 8 % PEG 4K	F9 80l 69- 0.1MHEPES 7.5 2 % PEG 400 1.8 M F.A.	F10 80l 70- 0.1M Cit 5.5 20 % Isopp 20 % PEG 4K	F11 80l 71- 0.05M K ₂ HPO ₄ 20 % PEG 8K	F12 80l 72- 30 % PEG 1K
G1 80l 73- 0.2 M FormMg	G2 80l 74- 0.2M ActZn 0.1 M Cac 6.5 18 % PEG 8K	G3 80l 75- 0.2M ActCa 0.1 M Cac 6.5 18 % PEG 8K	G4 80l 76- 0.1 M Act 4.5 2 M (NH ₄) ₂ SO ₄	G5 80l 77- 0.1 M TRIS 8.5 2 M (NH ₄) ₂ SO ₄	G6 80l 78- 1 M Li ₂ SO ₄ 2 % PEG 8K	G7 80l 79- 0.5 M Li ₂ SO ₄ 15 % PEG 8K	G8 80l 80- 0.2M ActNH ₄ 0.1 M Cit 5.5 20 % Isopp 20 % PEG 4K	G9 SaltRx1 1-0.1 M BTP 7 1.8 M AcetNa	G10 SaltRx1 4- 0.1 M BTP 7 1.5 M NH ₄ Cl	G11 SaltRx1 6- 0.1 M Act 4.5 3.5 M NH ₄ Cl	G12 SaltRx1 9- 0.1 M Act 4.5 2.2 M NaCl
H1 SaltRx1 14- 0.1 M TRIS 8.5 3.2 M NaCl	H2 SaltRx1 17- 0.1 M BTP 7 1 M Cit(NH ₄) ₂	H3 SaltRx1 19- 0.1 M BTP 7 0.7 M CitNa ₃	H4 SaltRx1 26- 0.1 M BTP 7 0.7M FormMg	H5 SaltRx1 31- 0.1 M BTP 7 3.5 M FormNa	H6 SaltRx1 33- 0.1 M BTP 7 1.2M Malic Ac.	H7 SaltRx1 34- 0.1 M BTP 7 2.2M Malic Ac.	H8 SaltRx1 35- 0.1 M BTP 7 1.4 M MaltNa	H9 SaltRx1 43- 0.1 M Act 4.5 1.5M NaNO ₃	H10 SaltRx1 44- 0.1 M BTP 7 1.5 M NaNO ₂	H11 SaltRx1 46- 0.1 M Act 4.5 4 M NaNaNO ₃	H12 SaltRx1 47- 0.1 M BTP 7 4 M NaNO ₃

4 - JCSG

A1 1-50% P 400 0.1M Act 4.5 0.1M Li_2SO_4	A2 2-20% P 3K 0.1M Cit 5.5	A3 3-20% P 3350 0.2M Cit(NH_4) ₂	A4 4-30% MPD 0.1M Act 4.5 0.02M $CaCl_2$	A5 5-10% P 3350 0.2 FormMg	A6 6-20% P 1000 0.1M FosCit 4.2 0.2M Li_2SO_4	A7 7-20% P 6K 0.1M CHES 9.5	A8 8-20% P 3E 0.2 FormNH ₄	A9 9-20% P 3350 0.2M NH_4Cl	A10 10-20% P 3350 0.2M FormK	A11 11-50% MPD 0.1M TRIS 8.5 0.2M F.A.	A12 12-20% P 3350 0.1M TRIS 8.5 0.2M KNO_3
B1 13-0.1M Cit 4 0.8M S.A.	B2 14-20% P 3350 0.2M NaSCN	B3 15-20% PEG 6K 0.1M Bicine 9	B4 16-10% PEG 8K 8% Etilethanol 0.1M HEPES 7.5	B5 17-40% MPD 0.1M MES 6.5 5% PEG 8K	B6 18-40% Etanol 0.1M FosCit 4.2 5% PEG 1K	B7 19-8% P 4K 0.1M Act 4.5	B8 20-10% P 8K 0.1M TRIS 7 0.2M $MgCl_2$	B9 21-20% P 6K 0.1M Cit 5	B10 22-50% P 200 0.1M MES 6.5 0.2M $MgCl_2$	B11 23- 1.6M CitNa ₃	B12 24-20% P 3350 0.2M CitK
C1 25-20% P 8K 0.1M FosCit 4.2 0.2M NaCl	C2 26-20% P 6K 0.1M Cit 4 1M LiCl	C3 27-20% P 3350 0.2M NH_4NO_3	C4 28-10% P 6K 0.1M HEPES 7.5	C5 29-0.8M $NH_4H_2PO_4$ 0.8M KH_2PO_4 0.1M HEPES 7.5	C6 30-40% P 300 0.1M FosCit 4.2	C7 31-10% P 3K 0.1M Act 4.5 0.2M ActZn	C8 32-20% Etanol 0.1M TRIS 8.5	C9 33-10% Glicerol 25% Propanediol 0.1M Fosf 6.2	C10 34-10% P 20K 2% Dioxano 0.1M Bicina 9	C11 35-2M S.A. 0.1M Act 4.5	C12 36-10% P 1K 10% P 8K
D1 37-24% P 1.5K 20% Glicerol	D2 38-30% P 400 0.1M HEPES 7.5 0.2M $MgCl_2$	D3 39-50% P 200 0.1M Fosf 6.2 0.2M NaCl	D4 40-30% P 8K 0.1M Act 4.5 0.2M Li_2SO_4	D5 41-70% MPD 0.1M HEPES 7.5	D6 42-20% P 8K 0.1M TRIS 8.5 0.2M $MgCl_2$	D7 43-40% P 400 0.1M TRIS 8.5 0.2M Li_2SO_4	D8 44-40% MPD 0.1M TRIS 8.5	D9 45-25.5% P 4K 15% Glicerol 0.17M S.A.	D10 46-40% P 300 0.1M MES 6.5 0.2M ActCa	D11 47-14% Isopp 0.07M Act 4.5 30% Glicerol 0.14M $CaCl_2$	D12 48-16% P 8K 0.04M KH_2PO_4 20% Glicerol
E1 49-1M CitNa ₃ 0.1M MES 6.5	E2 50-2M S.A. 0.1M MES 6.5 0.2M NaCl	E3 51-10% Isopp 0.1M HEPES 7.5 0.2M NaCl	E4 52-1.26M S.A. 0.1M TRIS 8.5 0.2M Li_2SO_4	E5 53-40% MPD 0.1M CAPS 10.5	E6 54-20% P 3K 0.1M Imd 8 0.2M ActZn	E7 55-10% Isopp 0.1M MES 6.5 0.2M ActZn	E8 56- 1M $(NH_4)_2HPO_4$ 0.1M Act 4.5	E9 57-1.6M $MgSO_4$ 0.1M MES 6.5	E10 58-10% P 6K 0.1M Bicina 9	E11 59-14.4% P 8K 30% Glicerol 0.08M MES 6.5 0.16 ActCa	E12 60-10% P 8K 0.1M Imd 8
F1 61-0.05M $CuCl$ 30% Jeff M600 0.1M MES 6.5	F2 62-3.15M S.A. 0.1M Cit 5	F3 63-20% MPD 0.1M TRIS 8	F4 64- 20% Jeff M-600 0.1M HEPES 7.5	F5 65-0.2M $CuCl$ 30% Etilethanol 0.1M TRIS 8.5	F6 66-10% MPD 0.1M Bicina 9	F7 67- 0.8M Succinic 7	F8 68- 2.1M Malic Acid 7	F9 69- 2.4M Malonato 7	F10 70-0.5% Jeff ED- 2001 0.1M HEPES 7.5 1.1M Malonato 7	F11 71-1% Jeff ED- 2001 0.1M HEPES 7.5 1M Succinic 7	F12 72- 30% Jeff M-600 0.1M HEPES 7.5
G1 73-30% Jeff ED- 2001 0.1M HEPES 7.5	G2 74- 22% Polyd S100 0.1M HEPES 7.5 0.2M $MgCl_2$	G3 75- 20% Polyvpk15 0.1M TRIS 8.5 0.1M $CoCl$	G4 76-20% P2UMME 0.1M TRIS 8.5 0.2M TriethylN- oxide	G5 77-12% P 3350 0.1M HEPES 7.5 5mM $CaCl_2$; 5mM GDC; 5mM $MgCl_2$ 5mM NiCl	G6 78-20% P3350 0.2M Malonato 7	G7 79-15% P3350 0.1M Succinic 7	G8 80-20% P3350 0.15 Malic Acid 7	G9 81-30% P2UMME 0.1M KSCN	G10 82-30% P2UMME 0.15M KBr	G11 83-2M S.A. 0.1M Bis-TRIS 5.5	G12 84-3M NaCl 0.1M Bis-TRIS 5.5
H1 85- 0.3M FormMg 0.1M Bis-TRIS 5.5	H2 86-1% P3350 0.1M Bis-TRIS 5.5 1M S.A.	H3 87-25% P3350 0.1M Bis-TRIS 5.5	H4 88-45% MPD 0.1M Bis-TRIS 5.5 0.2M $CaCl_2$	H5 89-45% MPD 0.1M Bis-TRIS 5.5 0.2M ActNH ₄	H6 90-17% P 10K 0.1M Bis-TRIS 5.5 0.1M ActNH ₄	H7 91-25% P3350 0.1M Bis-TRIS 5.5 0.2M S.A.	H8 92-25% P3350 0.1M Bis-TRIS 5.5 0.2M NaCl	H9 93-25% P3350 0.1M Bis-TRIS 5.5 0.2M Li_2SO_4	H10 94-25% P3350 0.1M Bis-TRIS 5.5 0.2M ActNH ₄	H11 95-25% P3350 0.1M Bis-TRIS 5.5 0.2M $MgCl_2$	H12 96-45% MPD 0.1M HEPES 7.5 0.2M ActNH ₄