THE UNIVERSITY
of EDINBURGH

A SYNTHETIC BIOLOGY APPROACH FOR GREEN MACROALGAL BIOMASS DEPOLYMERIZATION

ALEJANDRO ANDRÉS SALINAS VACCARO

THE UNIVERSITY *of* EDINBURGH

Doctor of Philosophy

The University of Edinburgh

2017

Lay summary

The finite nature and instability of fossil fuel supplies, as well as the high levels of carbon dioxide emissions generated due to the use of these kind of fuels, show the importance of alternative and more sustainable energy sources. An interesting option is to produce biofuels like ethanol or butanol from biomass. The production of biofuels from crops such as corn or sugarcane is well established. However, this requires large amounts of cultivable land, leading to higher prices for food due to the competition for land between crops for food and the ones used for the production of biofuels. In this scenario, the use of green algae as an alternative source of biomass seems to be an interesting alternative. They do not require arable land, fertilizers or fresh water as with land plants, and they grow considerably faster.

Green algal biomass contains different types of sugar chains, including cellulose and ulvan. For the production of biofuels, these sugar chains need to be shortened into single sugars, which can be converted into biofuels using microorganisms. Different microorganisms are able to degrade different types of sugar chains. Synthetic biology offers the possibility to import these degradative capabilities from one microorganism into another. This study explored the possibility of modifying the bacterium *Escherichia coli* adding to it the degradative capabilities from other microorganisms so that it would be able to degrade the main sugar chains present in green algae, cellulose and ulvan. In addition, the ulvan degrading capability of the microorganism *Formosa agariphila* is also explored.

Abstract

Green macroalgae represent an attractive source of renewable carbon. Conversion of algal biomass to useful products requires depolymerization of the cell wall polysaccharides cellulose and ulvan. Cellulose saccharification has been widely studied and involves synergistic action of endoglucanases, exoglucanases, and β-glucosidases. The enzymatic depolymerization of ulvan has not received the same attention and additional studies are required in order to fully understand the mechanisms involved in its biodegradation.

Synthetic biology offers the possibility of importing modules such as biomass-degrading systems and biofuel producing pathways from different organisms into a genetically tractable host such as *Escherichia coli*. In this study it was shown that *E. coli* expressing the glycosidase CHU2268 of *Cytophaga hutchinsonii* grows well on cello-oligosaccharides such as cellohexaose, and co-expression with the endoglucanase CenA of *Cellulomonas fimi* allows growth on untreated crystalline cellulose. Moreover, a model for ulvan utilization was built for the first time based on a polysaccharide utilization locus from the alga-associated flavobacterium *Formosa agariphila*. It was also shown that *F. agariphila*, is able to grow using biomass from the green macroalga *Ulva lactuca* as its sole carbon source, and enzymes with ulvanase activity are induced by the presence of this alga in the culture medium. Enzymes for ulvan depolymerization from *F. agariphila*, including an ulvan lyase, xylanases and rhamnosidases, were cloned using the PaperClip DNA assembly method and expressed in active form in *E. coli*. Furthermore, a secretion system based on the use of the Antigen 43 was successfully used to secrete an active ulvan lyase using *E. coli* and ribosome binding sites of different strengths were studied and used to optimize the system.

These results represent a first step for the design of a microorganism capable of utilizing green macroalgal biomass for the production of biofuels and other valuable bio-products.

Declaration

I declare that this thesis was composed by myself and all the results presented in it are my own, unless otherwise stated. This work has not been submitted for any other degree or personal qualification.

Alejandro Andrés Salinas Vaccaro

Acknowledgements

I would like to express my sincere gratitude…

…to my supervisor, Professor Chris French, for giving me the freedom to follow my own ideas but at the same time having an open door to give me advice and guidance when I needed it.

…to the members of my committee, Dr Louise Horsfall, Professor Stephen Fry and Dr Jon Marles-Wright, for their helpful suggestions and advice.

…to Anita who kindly proof read all my work during my PhD. To Chao-Kuo for the stimulating discussions. To Mai-Britt for always encouraging me to continue. To Jan for always being helpful in a very peculiar way. To Joe, Maryia, Dave, Kwabena, Craig, Lina, Chris N., and many others, for not being only great colleagues but also good friends, making my time in the French lab unforgettable.

…to my friends who even though they are scattered all over the world, they are always close. To Hannah for her patience, support and for always believing in me.

…to my parents, Sergio and Gabriela, and my sisters, Gabriela and Macarena, for their unconditional love and support.

Finally, I would like to thank the Becas Chile scholarship program from the Chilean National Commission for Scientific and Technological Research (CONICYT) for the funding that allowed me to do this research.

Table of contents

List of figures

List of tables

Abbreviations

Abbreviation	Meaning
Ag43	Antigen 43
AIR	Alcohol insoluble residue
BMCC	Bacterial microcrystalline cellulose
BSA	Bovine serum albumin
CAZyme	carbohydrate-active enzyme
CBD	Cellulose binding domain
CBM	Carbohydrate-binding modules
CBP	Consolidate bioprocessing
CDD	Conserved domain database
CMC	Carboxymethyl cellulose
CML	Chloramphenicol
DF	Downstream forward
DHAP	Dihydroxyacetone phosphate
DKI	5-dehydro-4-deoxy-D-glucuronate
DKII	3-deoxy-D-glycero-2,5-hexodiulosonate
DNS	3,5-Dinitrosalicylic acid
DR	Downstream reverse
ED	Entner–Doudoroff
EDTA	Ethylenediaminetetraacetic acid
EMP	Embden–Meyerhof–Parnas
EPPAW	Ethyl acetate, pyridine, propanol, acetic acid and water
FGly	Formylglycine
GAG	Glycosaminoglycan
GAP	D-glyceraldehyde 3-phosphate
GalA	Galacturonic acid
GFP	Green fluorescent protein
GH	Glycoside hydrolase
GlcA	Glucuronic acid
HTCS	Hybrid two-component system
IduA	Iduronic acid
IPTG	Isopropyl β-D-1-thiogalactopyranoside
KDG	2-dehydro-3-deoxy-D-gluconate
KduI	4-deoxy-L-threo-5-hexosulose-uronate ketol-isomerase
KduD	2-dehydro-3-deoxy-D-gluconate 5-dehydrogenase
LB	Lysogeny broth
MABEL	Mutagenesis with blunt-end ligation

Abbreviation	Meaning
MB	Marine broth
MMS	Master mix standard
MSF	Major facilitator superfamily
MU	4-Methylumbelliferone
MUC	4-Methylumbelliferyl β-D-cellobioside
MUG	4-Methylumbelliferyl β-D-glucopyranoside
MUGlcU	4-Methylumbelliferyl-β-D-glucuronide
MUX	4-Methylumbelliferyl-β-D-xylopyranoside
OD	Optical density
PASC	Phosphoric acid-swollen cellulose
PBS	Phosphate buffered saline
PCR	Polymerase chain reaction
PL	Polysaccharide lyase
pNP	4-Nitrophenol
pNPR	4-Nitrophenyl α-L-rhamnopyranoside
PP	Pentose phosphate
PUL	Polysaccharide utilization loci
Rha3S	Rhamnose-3-sulfate
RBB	remazol brilliant blue R
RBS	Ribosome binding site
SDS-PAGE	Sodium dodecyl sulphate polyacrylamide gel electrophoresis
SHF	Separate hydrolysis and fermentation
SSF	Simultaneous saccharification and fermentation
TBDR	TonB-dependent receptor
TFA	Trifluoroacetic acid
TLC	Thin layer chromatography
TR	Transcription regulator
U	Unit of activity
UF	Upstream forward
UGL	Unsaturated glucuronyl hydrolase
UR	Upstream reverse
UV	Ultraviolet
Xyl	Xylose
X-gal	5-Bromo-4-chloro-3-indolyl-β-D-galactoside

1 Introduction

1.1 General background

Finding energy sources that are renewable and less damaging to the environment is essential. The possibility of producing biofuels such as ethanol and n-butanol from green macroalgal biomass seems to be an attractive alternative. This is due to a number of desirable characteristics such as no competition for arable land, fast growth rates, absence of structural biopolymers such as lignin, and the capability to grow in saline water, amongst others (John et al., 2011). Additionally, this kind of biomass can be used to produce other useful co-products. Ulvan, one of its main cell-wall polysaccharides, is a source of the rare sugar rhamnose, which can be used as a precursor for the synthesis of fine chemicals such as fragrances (Lahaye & Robic, 2007) or the commodity chemical 1,2-propanediol (Saxena et al., 2010).

In order to convert green macroalgal biomass into useful products the depolymerization of its two main cell wall polysaccharides, cellulose and ulvan, is required. The hydrolysis of cellulose has been widely studied and the most accepted mechanism for its degradation involves synergistic action of enzymes from three classes: endoglucanases, exoglucanases, and β-glucosidases (Lynd et al., 2002). However, knowledge about the enzymatic saccharification of ulvan is scarce and the mechanisms involved in its biodegradation still have to be identified.

To design an ideal microorganism capable of producing useful products from biomass in an economically viable process, the use of synthetic biology seems to be a suitable approach, offering the possibility of importing modules such as biomass-degrading systems and biofuel producing pathways from different organisms into a genetically tractable host such as *Escherichia coli* (Alper & Stephanopoulos, 2009). In this context, synthetic biology allows the testing of multiple combinations of enzymes (e.g. cellulases) in order to determine which of these have better performance using different cellulosic substrates (French, 2009). In addition, it is possible to study the

presence of synergistic patterns between different enzymes (inter and intra class), which could be exploited to improve the hydrolysis process.

Considering the background information presented, this project will be focused on the use of synthetic biology to study the enzymatic degradation of the two main polysaccharides present in green macroalgal biomass; cellulose and ulvan.

1.2 Biofuels from macroalgal biomass

Interest in biofuels has been increasing for two main reasons: the insecurity caused by the imminent depletion of fossil fuel reserves and the need to reduce carbon dioxide emissions caused by fossil fuel usage in order to mitigate climate change (Solomon et al., 2007).

Industrial production of biofuels from sugar/starch crops (e.g. sugar cane, sugar beet, corn and wheat), known as first generation biofuels, is well established and can address these issues. However, the cost of the raw materials is high and their production requires substantial amounts of cultivable land, leading to an increase in the price of food due to the competition for available land (John et al., 2011).

In this scenario, new sources of raw materials should be studied. Low-cost feedstocks such as lignocellulosic residues and algal biomass (second generation biofuels) are attractive alternatives. However, hydrolysis of the carbohydrates present in the biomass is necessary to obtain fermentable sugars that can be used for biofuel production (John et al., 2011), leading to higher production costs compared with first generation biofuels (Tan et al., 2008).

In the case of lignocellulosic feedstocks a pre-treatment step is necessary to remove lignin and make the carbohydrates accessible for enzymatic hydrolysis (Klinke et al., 2004). Lignin is a structural biopolymer that is very difficult to degrade biologically (Harun et al., 2010). Potential fermentation inhibitors, such as phenols, furans, carboxylic acids and inorganic salts, which have a negative effect on cell membrane function, growth, and glycolysis, are formed during the pre-treatment (Klinke et al., 2004).

On the other hand, algal cells do not contain lignin, which is essential for growth of higher plants in a terrestrial environment. Therefore, the pre-treatment step is not necessary for this kind of biomass simplifying the overall conversion process of biomass to useful products (John et al., 2011). Furthermore, macroalgae are able to grow at considerably faster rates than land plants, because their entire surface is able to absorb nutrients and the energy requirements for the production of supporting tissues are lower (Adams et al., 2008). In addition, macroalgae do not require arable land, fertilizer or fresh water as in the case of land plants (Lahaye & Robic, 2007; Wargacki et al., 2012).

The reasons presented above make macroalgal biomass an excellent candidate to be used as the raw material for the production of biofuels and other valuable co-products.

1.3 Green macroalgal biomass structure

Macroalgae can be divided into three main classes according to their pigmentation: green, red and brown algae, with a carbohydrate content of 25-50%, 30-60%, and 30-50% dry weight, respectively (Jung et al., 2013). These classes produce a variety of different carbohydrates with specific properties and structures (John et al., 2011). Considering the heterogeneity among the different classes of macroalgae it is crucial to know the composition of the biomass to be used, in order to effectively utilize their carbohydrates as a feedstock for the production of biofuels and other bioproducts.

Among the different classes of macroalgae, green algae have been the least studied (Alves et al., 2013); however, their fast growth rates and broad tolerance to environmental conditions such as irradiation, temperature and salinity makes them an attractive source of biomass that can be used as a feedstock for the production of biomolecules with important commercial value (Taylor et al., 2001).

Green macroalgal cell wall polysaccharides belong to four families: two major ones, water-soluble ulvan and insoluble cellulose, and two minor ones, alkali-soluble xyloglucan and glucuronan (Lahaye & Robic, 2007). In addition, the presence of small amounts of starch and lipids have been also reported (1-4% and 0-6% dry wt., respectively) (Bruton et al., 2009).

In particular, ulvan is composed of a number of different repeating oligosaccharides made up of rhamnose, glucuronic acid, iduronic acid, xylose and sulphate (Lahaye & Robic, 2007). The main repeating disaccharides are made up of rhamnose 3-sulfate (Rha3S) linked with either glucuronic acid (GlcA), iduronic acid (iduA), or xylose (Xyl), resulting in GlcA-Rha3S (ulvanobiouronic acid A), IduA-Rha3S (ulvanobiouronic acid B) and Xyl-Rha3S (ulvanobiose 3-sulfate), respectively (Collen et al., 2011) (Figure 1-1). Additionally, minor repeating units containing xylose 2-sulfate or branches of glucuronic acid on C-2 of rhamnose 3-sulfate have also been reported (Lahaye & Robic, 2007).

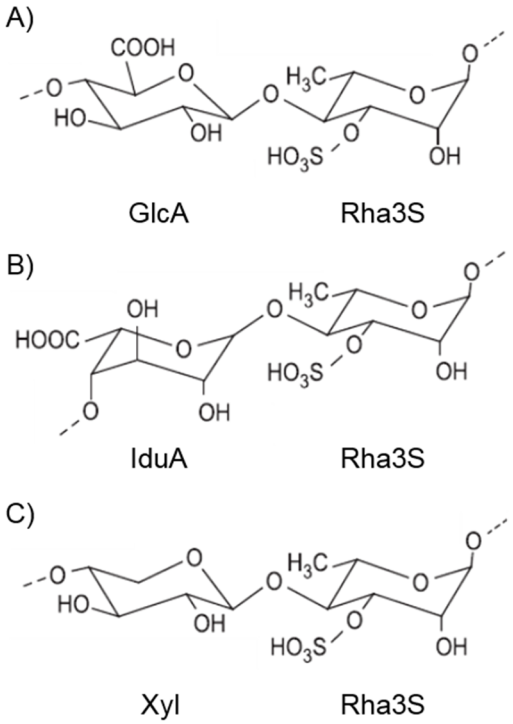

Figure 1-1 Main repeating disaccharides found in ulvan. A: ulvanobiouronic acid A, B: ulvanobiouronic acid B, C: ulvanobiose 3-sulfate. Rha3S: rhamnose 3-sulfate, GlcA: glucuronic acid, IduA: iduronic acid, Xyl: xylose. Adapted from Collen et al. (2011).

The other major polysaccharide in green macroalgae is cellulose, a homopolysaccharide composed of glucose monomers (D-glucopyranose) linked by β-1,4-glycosidic bonds (OSullivan, 1997). The crystalline structure of algal cellulose contains predominantly cellulose Iα (triclinic crystalline form) instead of cellulose Iβ (monoclinic crystalline form) as in the case of plants. Iα cellulose contains weaker

hydrogen bonds which provide easier access for endoglucanases during the saccharification process (Daroch et al., 2013). As in the case of cellulose, algal xyloglucan structure also differs from that of higher plants, having a linear trisaccharide structure (β-D-xylopyranosyl-(1→4)-β-D-glucopyranosyl-(1→4)-glucopyranose) instead of a cellulose-like backbone chain with side chain xylose residues attached to position C6 (Fry, 1988; Lahaye et al., 1994). Finally, algal glucuronan is composed of β-1,4-D-polyglucuronic acids (Elboutachfaiti et al., 2011). Schematic representations of the cell wall polysaccharides from green macroalgae are shown in Figure 1-2.

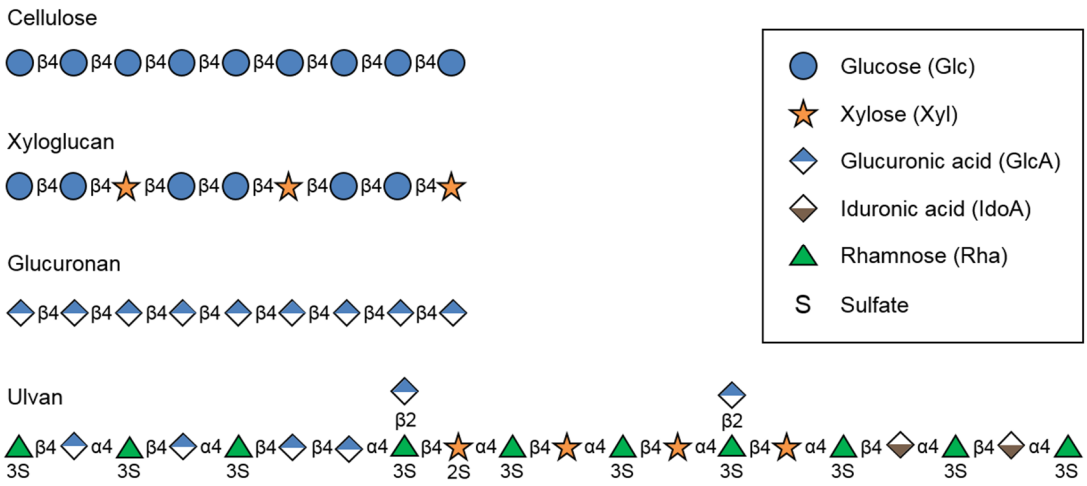

Figure 1-2 Structure of cell-wall polysaccharides of green macroalgae. Ulvan and xyloglucan model fragments were constructed based on the structures summarized by Lahaye and Robic (2007) and Lahaye et al. (1994), respectively. All glycosidic linkages are assumed to originate from C1. Monosaccharide symbols follow the Symbol Nomenclature for Glycans (SNFG) system (Varki et al., 2015).

The main sugars present in green macroalgae are glucose, rhamnose, xylose and uronic acids. Table 1-1 contains the sugar and sulfate composition of samples of the algae *Ulva* sp. (Bobin-Dubigeon et al., 1997) and *Ulva lactuca* (van der Wal et al., 2013).

Table 1-1 Sugar and sulfate composition of the algae *Ulva* sp. and *Ulva lactuca* (Bobin-Dubigeon et al., 1997; van der Wal et al., 2013).

	Ulva* sp.	***Ulva lactuca****
Sulfate	30.8	17.9
Uronic acid	18.2	11
Rhamnose	20.3	24.1
Xylose	7.3	15.5
Mannose	0.7	0
Galactose	1.4	3.4
Glucose	21.3	28.2

*: % of total sugars + sulfate.

1.4 Green macroalgal biomass production

Macroalgal biomass can be obtained from a broad range of sources, including natural stocks, land-based tanks and ponds production systems (Neveux et al., 2014). The majority of the global seaweed resource is a result of cultivation activities and only 6% is harvested from natural stocks (Bruton et al., 2009).

In particular, the cosmopolitan green macroalgae *U. lactuca* can be grown in a range of climates from tropical to polar (Ben-Ari et al., 2014; Bruhn et al., 2011), with an optimal growth temperature of 15°C (Fortes & Lüning, 1980). Green macroalgae are only cultivated commercially in brackish waters and estuaries in Asia, for food consumption (Ohno & Triet, 1997), and in raceway ponds in South Africa (Figure 1-3), for abalone feed (Bolton et al., 2009). However, pilot land-based systems to grow *Ulva* species have been tested in various locations around the world, including Australia, Denmark and Tanzania, where the algae are cultivated alone (Bruhn et al., 2011; Msuya & Neori, 2008; Neveux et al., 2014), and others in Saudi Arabia, Israel, Spain, Chile, USA and Denmark, where the algae is used as a biofilter to remove nutrients from manure or aquaculture effluents (Al-Hafedh et al., 2015; Ben-Ari et al., 2014; Hernández et al., 2005; Macchiavello & Bulboa, 2014; Neori et al., 1996; Nielsen et al., 2012).

Figure 1-3 Commercial seaweed-abalone pond farm, Haga Haga, on South Africa's southeast cost (Neori et al., 2007). *Ulva* is grown in paddle raceways utilizing abalone effluent and fed back to the abalone farm (left).

Aeration is an energy intensive process that could constitute about 85% of the green macroalgae cultivation cost (Ben-Ari et al., 2014). However, it has been reported that the main role of water agitation, and aeration in particular, is to accelerate the rate of diffusion of nutrients into the algae, and it may not be required if the concentration of nutrients in the culture is high enough (Msuya & Neori, 2008). In order to reach high algal biomass yields, the use of aquaculture effluents from the fish and mollusc industries (Al-Hafedh et al., 2015; Ben-Ari et al., 2014; Bolton et al., 2009; Hernández et al., 2005; Macchiavello & Bulboa, 2014) or manure (Nielsen et al., 2012) have emerged as cost-effective solutions to supply the necessary nutrients to the cultures. At the same time the algae act as a biofilter which may help solve the environmental problems caused by the discharge of nutrient loads from these industries into coastal waters, generating a win-win situation. *U. lactuca* has been proposed as an ideal candidate for simultaneous bioremediation and biomass production due to its high photosynthetic rates, high ability to uptake dissolved nitrogen and high pollution tolerance compared with most macroalgae (Al-Hafedh et al., 2015; Ben-Ari et al., 2014).

Biomass yields ranging from 72 to 95 dry t/ha/y have been achieved by cultivation of *Ulva* in raceway ponds utilizing energy-efficient paddles to provide water agitation

(Bolton et al., 2009). These levels of production are similar to the 74 dry t/ha/y obtained by the US Aquatic Species Programme when *U. lactuca* was cultivated using an energy intensive aeration system, and considerable higher than the 27 dry t/ha/y produced with a non-energy intensive system (Ryther et al., 1984). Additionally, it has been reported that *U. lactuca* can be cultivated as far north as Denmark with an estimated yield of 45 dry t/ha/y (Bruhn et al., 2011), which is three times higher than the one obtained using the brown algae *Laminaria* in temperate waters (Kelly & Dworjanyn, 2008). These yields are also higher than those reported when conventional terrestrial energy crops are grown (Aylott et al., 2008; Barbanti et al., 2006; Cadoux et al., 2012; Gallego et al., 2011; Heaton et al., 2004; Romanelli et al., 2012; Seppälä et al., 2008; Sims et al., 2010) and *Ulva* is collected from natural stocks (Bruton et al., 2009). The biomass yields of the macroalgae grown under different conditions and a list of common energy crops are summarized in Table 1-2.

Table 1-2 Biomass yields of different macroalgae and crops.

	Growth yield [dry t/ha/y]	**References**
Macroalgae		
Ulva – raceway system	72-95	(Bolton et al., 2009)
U. lactuca – energy intensive	74	(Ryther et al., 1984)
U. lactuca – non-energy intensive	27	(Ryther et al., 1984)
U. lactuca – cultivation tanks	45	(Bruhn et al., 2011)
Ulva sp.– natural stock	2	(Bruton et al., 2009)
Laminaria	15	(Kelly & Dworjanyn, 2008)
Crops		
Willow	5-11	(Aylott et al., 2008)
Poplar	2-10	(Aylott et al., 2008)
Eucalyptus	10-12	(Romanelli et al., 2012)
Miscanthus	5-43	(Cadoux et al., 2012)
Switchgrass	5-19	(Heaton et al., 2004)
Reed canary grass	2-10	(Sims et al., 2010)
Alfalfa	1-17	(Gallego et al., 2011)
Fibre sorghum	16-43	(Barbanti et al., 2006)
Maize	9-18	(Seppälä et al., 2008)

Another possible source of *U. lactuca* biomass are green macroalgal blooms, also known as “green tides”. Green tides have increased in number and magnitude all over the world due to coastal eutrophication (Smetacek & Zingone, 2013). However, the delivery of biomass fluctuates seasonally and it is relatively limited (Bruhn et al., 2011; Bruton et al., 2009), making it difficult to build an industry based on this resource. It has been proposed that this kind of biomass could be integrated into a process already utilizing other types of biomass as the feedstock (Bruton et al., 2009).

Nowadays most of the produced *U. lactuca* biomass is a resource with no industrial application (Bruton et al., 2009; Chiellini & Morelli, 2011). The possibility of using this renewable biomass as a feedstock for sustainable energy has attracted increasing attention in recent years (Chiellini & Morelli, 2011). The high content of ash and alkali components in green macroalgal biomass make it unsuitable for conversion technologies such as combustion or pyrolysis (Bruhn et al., 2011). On the other hand, wet processing techniques such as fermentation or anaerobic digestion could take advantage of the intrinsic wet nature of the macroalgal biomass (McKendry, 2002).

1.5 Production of second generation biofuels

The production of second generation biofuels usually involves three main stages: (i) pre-treatment of the biomass, (ii) saccharification of polysaccharides, and (iii) fermentation of the released simple sugars to biofuels (Taherzadeh & Karimi, 2007).

1.5.1 Pre-treatment

The main objective of the pre-treatment stage is to transform the biomass from its native form into another, on which the enzymatic saccharification is more effective. In the case of lignocellulosic biomass this is reached through the solubilization and/or redistribution of lignin, which leads to an increase in the surface area accessible to cellulase enzymes (Lynd et al., 2002). On the other hand, macroalgal biomass does not contain lignin and seems to be less recalcitrant to hydrolysis than lignocellulosic biomass. It has been reported that the pre-treatment process is not required for this kind of biomass, and efficient saccharification has been achieved from untreated

macroalgae biomass (Adams et al., 2008; Yanagisawa et al., 2011). The use of untreated biomass would reduce the number of steps for its bioconversion lowering the overall cost of the process (Adams et al., 2008). Nonetheless, the pre-treatment of green macroalgal biomass under mild conditions has also been reported (van der Wal et al., 2013).

It is important to note that as a result of the pre-treatment, depending on both the kind of biomass and the pre-treatment applied, potential fermentation inhibitors such as phenols, furans, carboxylic acids and inorganic salts can be produced, affecting the overall feasibility of the bioconversion process (Klinke et al., 2004).

1.5.2 Saccharification of polysaccharides

This step's main aim is the depolymerization of the biomass in order to obtain simple sugars that can be easily fermented. The products of this process are usually reducing sugars including glucose. Due to the lower cost in comparison with acid and alkaline hydrolysis, this process is usually mediated by enzymes, which can catalyse the hydrolysis of the biomass under mild conditions (pH 4.8 and temperatures between 45 and 50°C), preventing corrosion problems (Sun & Cheng, 2002). However, this process also has some disadvantages such as the high cost of the enzymes and the long period of time necessary for the reaction to complete (Taherzadeh & Karimi, 2007).

It should be noted that the major factors which affect the saccharification process are: the concentration and quality of the substrate, the pre-treatment used, the activity of the enzymes, and the conditions used for the hydrolysis of the biomass; such as the temperature, pH and agitation (Taherzadeh & Karimi, 2007). In addition, cellulose accessibility has been reported as a key factor for successful enzymatic conversion of biomass into fermentable sugars (Jeoh et al., 2007).

1.5.3 Fermentation to biofuels and other bioproducts

The simple sugars generated from the saccharification process have to be fermented to biofuel through the use of microorganisms. The most commonly used microorganism for this process at an industrial level is the yeast *Saccharomyces cerevisiae*, which has

proven to be very robust and suitable for the fermentation of sugars obtained from the saccharification of biomass (Galbe & Zacchi, 2002). However, *S. cerevisiae* is unable to use pentose sugars, a disadvantage when the biomass is composed of a wide range of sugars, as in the case of lignocellulosic or macroalgal biomass. In contrast, *E. coli* is capable of making better use of these types of biomass, consuming both pentose and hexose sugars, and it is also highly amenable to metabolic engineering and synthetic biology techniques used to improve the biofuel production process (Huffer et al., 2012).

Among the different biofuels that can be produced from the sugars obtained from the saccharification of biomass, ethanol has attracted the most attention (French, 2009). Ethanol is produced from pyruvate requiring only two enzymes: a pyruvate decarboxylase which converts pyruvate to acetaldehyde and CO_2, and an alcohol dehydrogenase which reduces the acetaldehyde to ethanol. This pathway has mainly been studied in *S. cerevisiae,* the bacterium *Zymomonas mobilis* or recombinant *E. coli* (Fischer et al., 2008). Another interesting biofuel that can be obtained by microbial production is n-butanol. This biofuel is produced along with acetone and ethanol in a process known as 'ABE' fermentation used by *Clostridium acetobutylicum* and *Clostridium beijerinckii* (van der Wal et al., 2013). Butanol is produced through the dimerization of two molecules of acetyl-CoA into acetoacetyl-CoA, which is subsequently reduced and dehydrated to butyric acid. Finally, the butyrate is re-metabolized via a four-electron reduction to n-butanol (Fischer et al., 2008). This pathway was functionally expressed in recombinant *E. coli* by Atsumi et al. (2008). Butanol has several advantages over ethanol: it can be used in petrol engines without modification, it has a higher energy density, it is less corrosive, and it is not fully miscible with water, thus it is possible to transport in pipelines (French, 2009). However, it has been reported that it is far more toxic to most microorganisms (Fischer et al., 2008).

Moreover, rhamnose one of the main sugars present in green macroalgal biomass, can be fermented to the commodity chemical 1,2-propanediol (Boronat & Aguilar, 1981). This diol is currently produced by chemical synthesis from the non-renewable petrochemical derivative propylene oxide, and has applications in a broad spectrum of

areas, including the synthesis of biodegradable plastics and polymer resins (Saxena et al., 2010). The production of 1,2-propanediol through the fermentation of rhamnose has been reported in several microorganisms including *E. coli* and various yeasts (Saxena et al., 2010). 1,2-Propanediol biosynthesis involves several steps. First, L-rhamnose is isomerized to L-rhamnulose by a rhamnose isomerase. L-Rhamnulose is then phosphorylated to L-rhamnulose-1-phosphate by a rhamnulose kinase. The phosphorylated sugar is subsequently cleaved by an aldolase to produce dihydroxyacetonephosphate and L-lactaldehyde. Finally, under anaerobic conditions the lactaldehyde is reduced to 1,2-propanediol by a NAD dependent oxidoreductase (Bennett & San, 2001). However, this process is not commercially feasible due to the high price of rhamnose (Altaras & Cameron, 1999). The possibility of obtaining this sugar at lower prices through the saccharification of macroalgal biomass makes it an attractive alternative to the traditional chemical synthesis of 1,2-propanediol.

The conversion of biomass to biofuels can be divided into three main biologically-mediated stages: (i) production of degradative enzymes, (ii) saccharification of cellulose and other polysaccharides, and (iii) product formation from the released sugars (e.g., fermentation to biofuels) (Lynd et al., 2002). Based on these stages it is possible to characterize a process according to its degree of consolidation. When the enzyme production, saccharification and fermentation are carried out separately, the process involves three discrete steps and it is called separate hydrolysis and fermentation (SHF). The simultaneous saccharification and fermentation (SSF) consolidates the hydrolysis and fermentation into a single one, while the enzyme production is performed separately. Finally, in the case of consolidated bioprocessing (CBP), all the stages are coupled in a single process (Lynd et al., 2002). A schematic summary of the different strategies for the biomass conversion process is shown in Figure 1-4. A larger number of stages allows the use of specific conditions for each of the processes; however, more reactors are required increasing the overall cost of the process (Lynd et al., 2002). Considering the above, it has been proposed that the CBP configuration has the potential to provide the lowest cost route for the production of biofuels and other high value bioproducts from biomass (Lynd et al., 2005).

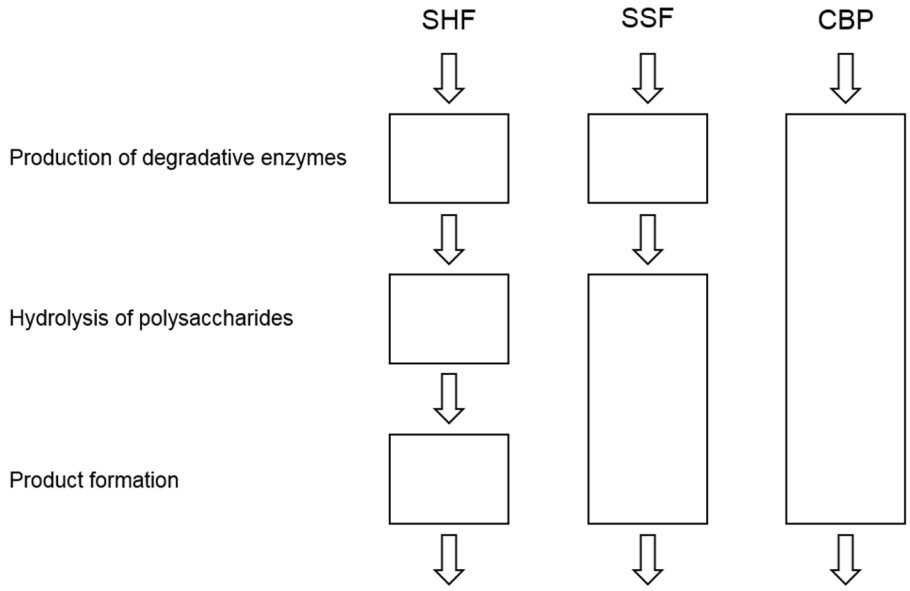

Figure 1-4 Different configurations for the biomass conversion process. On the left of the figure are the biologically mediated stages. Each box represents a bioreactor, these are not to scale. SHF: separate hydrolysis and fermentation; SSF: simultaneous saccharification and fermentation; CBP: consolidated bioprocessing.

1.6 Enzymes involved in biomass saccharification

In order to depolymerise the different cell wall polysaccharides present in green macroalgal biomass, enzymes with specific activities for each of them are required. The mechanism involved in the enzymatic hydrolysis of cellulose has been extensively studied (Lynd et al., 2002; Zhang et al., 2006). However, the enzymatic depolymerization of the other cell wall polysaccharides present in green macroalgal biomass has not received the same attention and only a few enzymes involved in their degradation have been identified.

1.6.1 Cellulose hydrolysis

The most accepted mechanism for cellulose hydrolysis involves synergistic action of endoglucanases (EC 3.2.1.4), exoglucanases, including cellobiohydrolases (EC 3.2.1.91) and cellodextrinases (EC 3.2.1.74), and β-glucosidases (EC 3.2.1.21) (Lynd et al., 2002). Endoglucanases randomly hydrolyse β-1,4-glycosidic bonds from the amorphous regions of the polysaccharide, creating new chain ends; exoglucanases act on these reducing and non-reducing cellulose chain ends, releasing cellobiose and

other short oligosaccharides; which are finally hydrolysed to glucose by β-glucosidases (Zhang et al., 2006). Figure 1-5 shows a diagram of cellulose hydrolysis by the enzymatic mechanism mentioned above (Lynd et al., 2002).

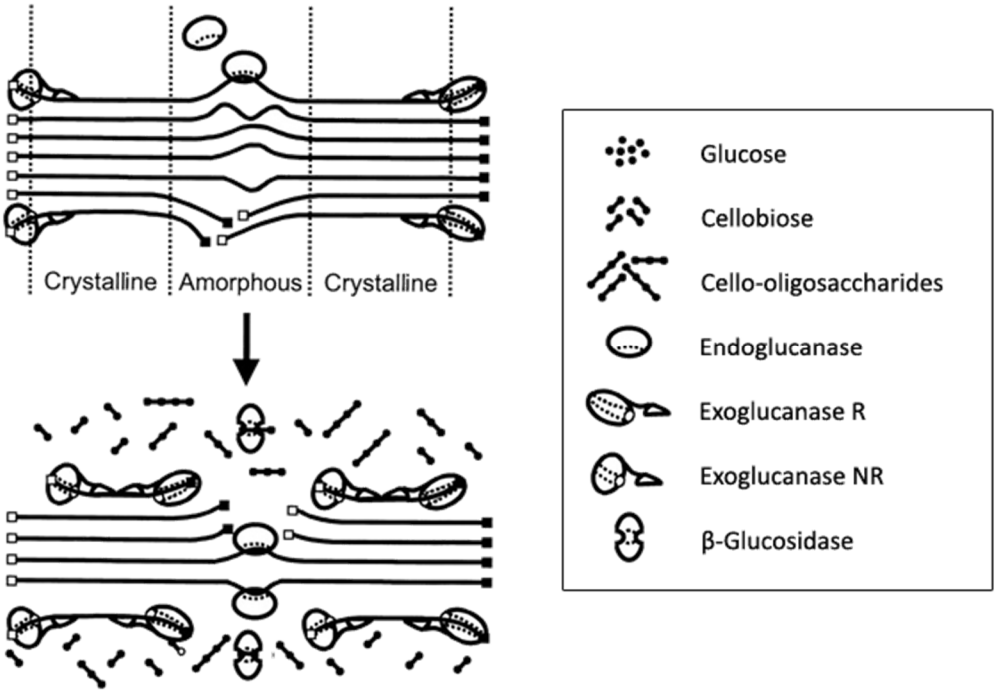

Figure 1-5 Schematic representation of enzymatic hydrolysis of cellulose. The reducing and non-reducing ends are represented by black and white squares, respectively. The exoglucanases R and NR hydrolyses from reducing and non-reducing ends, respectively. Amorphous and crystalline regions are indicated. Adapted from Lynd et al. (2002).

Both endoglucanases and exoglucanases hydrolyse β-1,4-glycosidic bonds and the specificity seems to be determined by the topology of the active site. Whereas endoglucanases have open active site clefts, exoglucanases active sites are located in long looped tunnels in the protein structure (Linder & Teeri, 1997).

In general cellulases are bimodular enzymes, with an extensive catalytic domain and a small cellulose binding domain (CBD). The two domains are linked by a highly glycosylated protein sequence rich in proline and hydroxy amino acid residues such as threonine and serine (Tomme et al., 1995). The specific role of the linker is not entirely clear, but it has been attributed functions associated with the disruption of regular packing of the cellulose fibrils and/or the increase of the local concentration of the catalytic domain to which they are attached (Brun et al., 1995). The CBD enables an

increase in the enzyme concentration at a particular site on the surface of the cellulose, leading the enzyme to specific locations on the substrate, besides making the substrate more accessible by neutralizing structural non-covalent associations (Tomme et al., 1995).

Several organisms contain efficient cellulose degradation systems, including the filamentous fungus *Trichoderma reesei*, the aerobic Gram-positive bacterium *Cellulomonas fimi*, the aerobic Gram-negative bacterium *Cytophaga hutchinsonii*, and the anaerobic Gram-positive bacterium *Clostridium thermocellum* (French, 2009). The majority of commercial cellulases are produced by *Trichoderma* and have been the most widely studied and best characterized (Taherzadeh & Karimi, 2007). In particular, analysis of the genomic sequences of the less studied cellulolytic bacterium *Cytophaga hutchinsonii*, which has the ability to rapidly digest crystalline cellulose, has revealed no obvious homologs of known cellobiohydrolases and it has been suggested that this organism may use an unusual method of cellulose utilization (Xie et al., 2007).

1.6.2 Saccharification of other polysaccharides

In the case of ulvan, the other major polysaccharide in green macroalgae biomass, only a few studies about the utilization of enzymes for its depolymerization have been reported. In most of these studies the enzymes were used as a tool for the better understanding of the chemical structure of this polysaccharide, rather than for depolymerisation to utilize the released sugars for the production of high-value bioproducts (Alves et al., 2013).

The first enzyme with ulvan lyase activity was reported by Lahaye et al. (1997). The ulvan lyase was obtained from a Gram-negative bacterium isolated from muds containing decomposing *Ulva* sp., and it was shown to cleave the linkage between Rha3S and GlcA, generating an unsaturated uronic acid at the newly created non-reducing end. However, no amino acid sequence of the enzyme or further characterization of the bacterium that produces it was provided. *Nonlabens ulvanivorans*, a marine Bacteroidetes isolated from the faeces of the sea slug *Aplysia punctata* fed with *Ulva* sp., was shown to be an efficient ulvan degrader (Barbeyron et

al., 2011). An ulvan lyase from *N. ulvanivorans* was purified, sequenced, and heterologously expressed in *E. coli* by Collen et al. (2011). Biochemical characterization of the ulvan lyase showed that it was an endolytic enzyme able to cleave ulvan between Rha3S and GlcA or IduA via a β-elimination mechanism. The released oligosaccharides had a degree of polymerization higher than two and contained an unsaturated uronyl residue at the non-reducing end. Considering that no characterized homologs were found in the databases, the enzyme was designated as the first member of a novel polysaccharide lyase family. Recently, a new family of ulvan lyases, showing no homology with ulvan lyase from *N. ulvanivorans*, was identified in three bacteria from the Alteromonadales order by Kopel et al. (2016). The novel ulvan lyases were heterologously expressed in *E. coli* and characterized to determine their mode of action. In contrast with the ulvan lyase from *N. ulvanivorans*, these enzymes only cleave between the residues Rha3S and GlcA.

In addition, an unsaturated glucuronyl hydrolase (UGL) from the glycoside hydrolase family 105 (GH105) was found in the genomic environment of the ulvan lyase from *N. ulvanivorans* (Collen et al., 2014). The UGL was shown to specifically cleave the unsaturated uronic residues generated by the ulvan lyase. Considering the proximity within the genome of the ulvan lyase and the UGL from *N. ulvanivorans*, it has been proposed that they might be part of a polysaccharide utilization locus (Kopel et al., 2016).

A diagrammatic scheme of the enzymatic depolymerisation of ulvan by the ulvan lyase and the UGL is shown in Figure 1-6. Considering the complex structure of ulvan, it is clear that several other enzymes such as other lyases, rhamnosidases, xylanases and sulfatases must be involved and their identification is crucial in order to understand the entire enzymatic system which leads to the saccharification of this biopolymer (Collen et al., 2011).

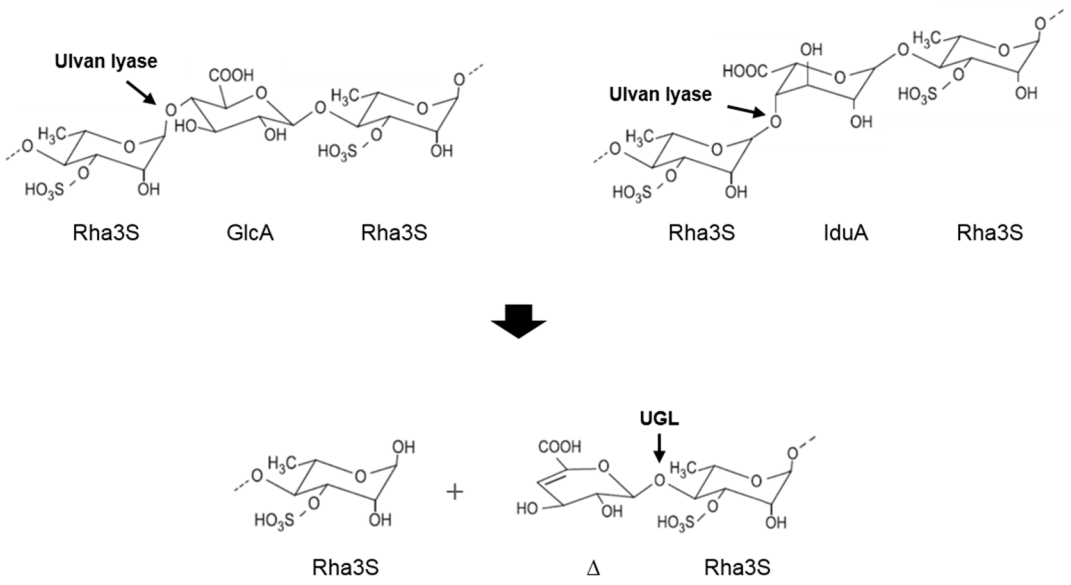

Figure 1-6 Enzymes reported to be involved in the saccharification of ulvan. First, an ulvan lyase cleaves via a β-elimination mechanism between rhamnose 3-sulfate and either glucuronic acid or iduronic acid, releasing a product with an unsaturated uronic acid at the non-reducing end of the chain, which is subsequently cleaved by an unsaturated β-glucuronyl hydrolase (UGL) from GH105. Cleavage sites are indicated by arrows. Rha3S: rhamnose 3-sulfate, GlcA: glucuronic acid, Δ: unsaturated uronic acid.

Moreover, it has been reported that the particular linear structure of green macroalgal xyloglucan can be depolymerized to glucose, xylose and cellobiose using the commercially available enzyme preparation Celluclast® (Lahaye et al., 1994). Celluclast® main activity is cellulase; however, it has been shown that this enzyme preparation also contains β-xylosidase side activity (Sorensen et al., 2003). Considering the structure of this kind of xyloglucan, its complete saccharification is probably achieved through the joint action of exo and endo-cellulases, β-(1→4)-glucosidases, and β-(1→4)-xylosidases, enzyme classes that have been widely studied. Furthermore, several glucuronan lyases produced by both bacteria and fungi have been identified and employed for glucuronan depolymerization (Redouan et al., 2009). In addition, partial digestion of ulvan by a glucuronan lyase has also been reported, revealing the occurrence of consecutive glucuronic acid moieties in this polysaccharide (Delattre et al., 2006).

1.7 Use of synthetic biology for biofuel production

Synthetic biology is an emerging discipline that aims to adopt engineering concepts, such as standardization of parts and assembly methods, in order to more easily construct novel biological systems (Ellis et al., 2011). Current efforts focus on the use of modular genetic parts, which can be easily assembled through the use of assembly methods to generate more complex multi-part systems, which are then introduced into a host organism or 'chassis' (Figure 1-7). The modular parts include promoters, ribosome binding sites (RBSs), coding DNA sequences (CDSs), and terminators. The selection of an appropriate chassis organism able to support the new introduced pathways and provide background processes such as replication and protein synthesis, is also required. Well-characterized microorganisms, such as *S. cerevisiae*, *E. coli* and *Bacillus subtilis* have been widely used for these purposes (French, 2009).

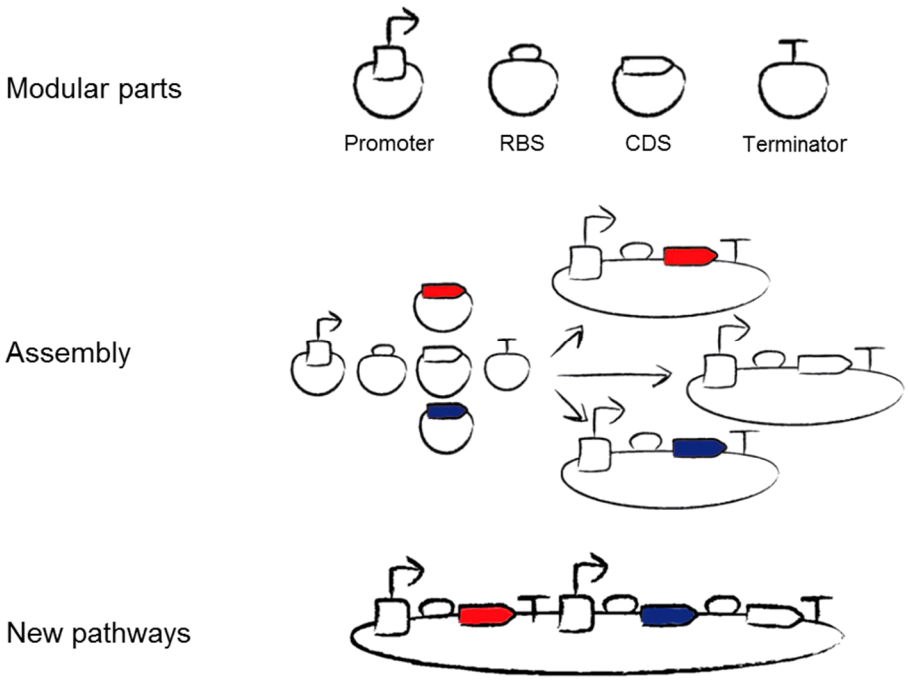

Figure 1-7 Generation of new pathways using modular DNA parts. Modular parts include promoters, ribosome binding sites (RBSs), DNA coding sequences (CDSs), and terminators. These parts are combined using DNA assembly methods to generate new constructs than can then be characterized. Finally, completely new pathways can be generated through combination of the characterized constructs. Figure adapted from Ellis et al. (2011).

The use of synthetic biology seems to be an attractive approach in order to reach an effective conversion of biomass to biofuels or other valuable bioproducts. In this sense,

several genetic modules from biomass-degrading systems and biofuel producing pathways from different organisms can be imported into a genetically tractable host, in order to enhance its biofuel production capabilities (Alper & Stephanopoulos, 2009). The modular and interchangeable nature of the parts allows the generation of organisms able to express a variety of different combinations of biomass-degrading enzymes. These can then be screened for effective degradation of different biomass substrates. These organisms can be also used for empirical determination of synergies between the different enzymes (inter and intra class), which could be exploited in order to reduce the amount of total protein required for effective saccharification (French, 2009). Libraries of biomass-degrading enzymes from the bacteria *C. fimi* and *C. hutchinsonii* are already in development for these purposes in the French lab, and would be used for the generation of a cellulose degradation module in this study (French, 2009).

DNA assembly is a limiting factor for the efficient construction of gene expression systems (Ellis et al., 2011). To overcome this problem several DNA assembly methods have been developed. They can be divided into two categories: restriction endonuclease based methods and homology based methods. Restriction endonuclease methods rely on the use of restriction sites for the assembly of the different parts and include BioBrick (Knight, 2003), Golden Gate (Engler et al., 2008), GoldenBraid (Sarrion-Perdigones et al., 2011), modular cloning (MoClo) system (Weber et al., 2011), and modular overlap-directed assembly with linkers (MODAL) (Casini et al., 2014). Homology based methods rely on homology between the ends of neighbouring parts and include Gibson assembly (Gibson et al., 2009), circular polymerase extension cloning (CPEC) (Quan & Tian, 2009), seamless ligation cloning extract (SLiCE) (Zhang et al., 2012), and the PaperClip assembly (Trubitsyna et al., 2014). In this study the BioBrick and PaperClip assembly methods were used to generate the different constructs.

1.7.1 BioBrick assembly

The BioBrick assembly standard was introduced by Knight (2003) to address the lack of standardization in assembly techniques for DNA sequences. Each DNA part, or

BioBrick, is flanked by standard prefix and suffix sequences. The prefix sequence includes EcoRI and XbaI restriction sites, while the suffix sequence includes SpeI and PstI restriction sites. Different compatible overhangs are generated depending on the combinations of restriction enzymes used to digest the BioBricks. This allows any BioBrick to be inserted upstream or downstream of any other BioBrick. A DNA 'scar' is generated after the ligation of two BioBricks, corresponding to a mixed SpeI/XbaI site. After each assembly a new BioBrick with the same physical format is generated, which can be used to generate more complex systems with several genetic components. A diagrammatic explanation of the BioBrick assembly standard is shown in Figure 1-8 (French, 2009). Several thousand BioBricks including promoters, ribosome binding sites, coding sequences and terminators, are freely available for the design and construction of genetic systems at the Registry of Standard Biological Parts (partsregistry.org).

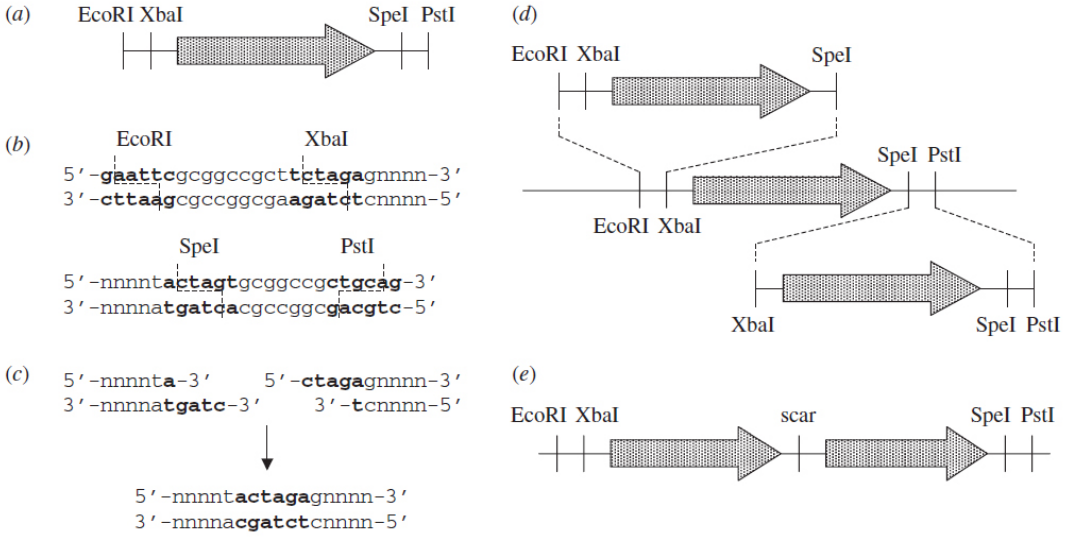

Figure 1-8 Schematic representation of the BioBrick 1.0 assembly standard. (a) Each BioBrick consists of a genetic component, such as promoters, ribosome binding sites, coding sequences, terminators or any combination of them, flanked on the upstream end by EcoRI and XbaI restriction sites, and on the downstream end by SpeI and PstI. (b) Standard prefix and suffix sequences for BioBricks. Recognition sites are in bold and dashed lines indicate the cuts made by each enzyme. (c) Example of 'scar' generation after the ligation of two BioBricks cut by a specific pair of restriction enzymes. The scar is not recognized by the enzymes used for the generation of the cohesive ends. (d) Assembly of two BioBricks. Depending on the restriction enzymes used, any BioBrick can be inserted upstream or downstream of any other BioBrick. (e) The result of the assembly of any two BioBricks is always another BioBrick, which can be used for other assemblies, allowing the generation of complex systems composed of several genetic components. Figure from French (2009).

The BioBrick assembly standard is a highly flexible method; however, it does not allow the assembly of more than two BioBricks in a single reaction, which makes it laborious for the assembly of systems involving several genetic components. In addition, it requires the removal of any EcoRI, XbaI, SpeI and PstI restriction sites within the BioBricks prior to assembly. Moreover, the scar generated after the ligation encodes an in-frame stop codon, thus this method can not be used for the generation of fusion proteins.

1.7.2 PaperClip assembly

PaperClip assembly is a recently developed homology based method (Trubitsyna et al., 2014), which use double stranded bridging oligonucleotides or 'clips' to enable the assembly of multiple DNA sequences in one reaction. Unlike other homology based methods, the use of these clips allows the assembly of a given set of parts in any order without requiring order-specific oligonucleotides. In addition, PaperClip does not use restriction enzymes, thus no mutagenesis of forbidden restriction sites is required.

For each part to be assembled four oligonucleotides of about 40 bases have to be designed. Two, upstream forward (UF) and upstream reverse (UR), match with the upstream end of the part, while the other two, downstream forward (DF) and downstream reverse (DR), match with the downstream end of the part (Figure 1-9 (A)). Forward and reverse oligonucleotides are annealed together to form two double stranded DNA half-clips. The upstream half-clip contains a GCC overhang at the 5' of the UF oligonucleotide, while the downstream half-clip contains a GGC at the 5' end of the DR oligonucleotide. The half-clips can be stored and used in any assembly involving the part. In order to assemble two parts, the downstream half-clip of one part has to be ligated with the upstream half-clip of the next part, generating a full clip (Figure 1-9 (B)). These clips guide the order of the parts in the final assembly, and considering that all the half-clips have identical sticky ends, any part can be assembled after any other part. Parts can be amplified by PCR using the UF and DR oligonucleotides as primers; alternatively, they can be introduced into the assembly as linearized or circular plasmids containing the desired part. Finally, a single assembly

reaction is performed with all the clips and parts generated using either PCR or cell extracts expressing recombination proteins (Figure 1-9 (C)).

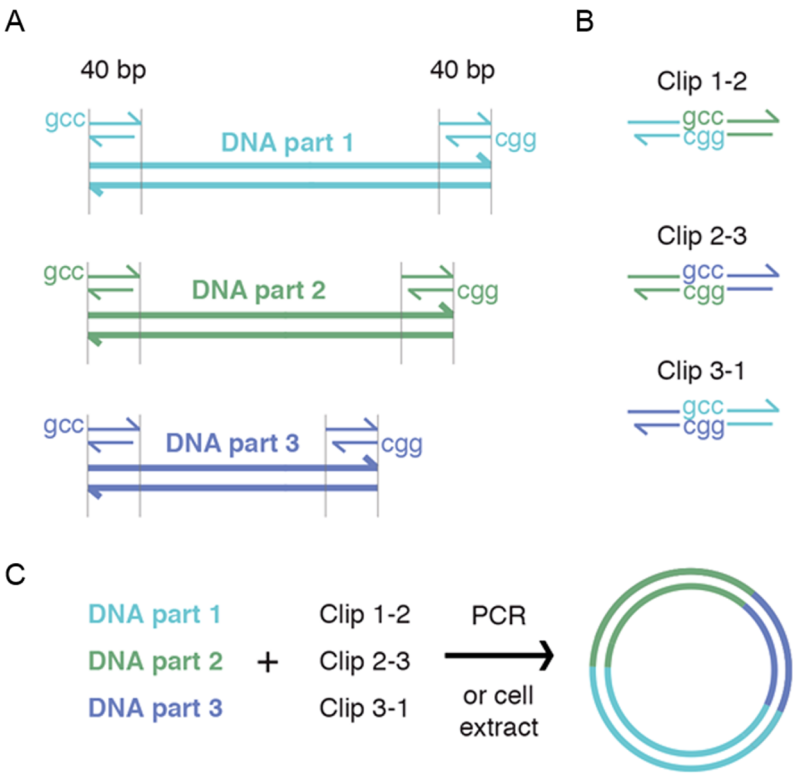

Figure 1-9 Schematic representation of the PaperClip assembly method. (A) For each part two pairs of complementary oligonucleotides of ~40 bases are designed. The pair matching the upstream end contains a GCC overhang at the 5' end of the forward oligonucleotide, while the pair matching the downstream end contains a GGC overhang at the 5' end of the reverse oligonucleotide. (B) The complementary oligonucleotides are annealed and phosphorylated, generating two half-clips for each part. The downstream half-clip from one part is ligated with the upstream half-clip of the next part generating a full clip, which will give the order to the assembly. (C) A PCR or cell extract assembly reaction is performed using all the clips and parts to get the final product. Figure from Trubitsyna et al. (2016).

In addition, PaperClip offer the possibility of inserting short intervening sequences of 35 to 200 bp between the DNA parts (Figure 1-10). These sequence can be linkers, ribosome binding sites or protein tags, among others. The sequence to be inserted has to be obtained as two pair of complementary oligonucleotides. As for the case of the half-clips, the upstream pair should contain a GCC overhang on the forward oligonucleotide, while the downstream pair has a GGC overhang on the reverse oligonucleotide. Additionally, 5 base complementary sticky ends are used to restore the sequence of interest. After annealing and phosphorylation of the two pairs of

oligonucleotides, these are inserted between two half-clips using a two-step ligation reaction. First, two separate reactions are performed, the downstream half-clip from part 1 is ligated with the upstream part of the intervening sequence, and the upstream half-clip of part 2 is ligated with the downstream part of the intervening sequence. Finally, an expanded clip containing the intervening sequence is generated by mixing the two previous ligations. This method allows the fast design of both fusion proteins containing different linker sequences and constructs containing RBSs of different strengths.

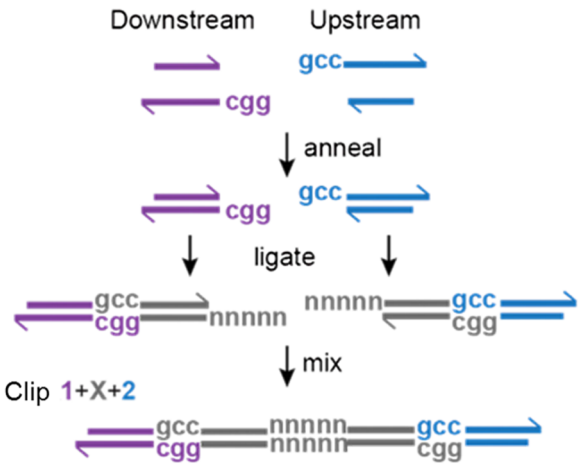

Figure 1-10 Addition of intervening sequences between two PaperClip DNA parts. Short intervening sequences (35 to 200 bp) such as linkers, ribosome binding sites and protein tags can be added between two parts. The intervening region is obtained as two pairs of complementary oligonucleotides. The upstream pair contains a GCC overhang at the 5' end of the forward oligonucleotide and a 5 base overhang at the 5' end of the reverse oligonucleotide, while the downstream pair contains a GGC overhang at the 5' end of the reverse oligonucleotide and 5 base overhang complementary to the one from the upstream reverse oligonucleotide at the 5' end of the forward oligonucleotide. The two pairs of oligonucleotides are annealed, phosphorylated and a two-step ligation performed to get the clip. First, the downstream half-clip from part 1 is ligated with the upstream half of the intervening sequence and the downstream half of the intervening sequence ligated with the upstream half-clip from part 2. A final ligation is performed mixing the results of the two previous ligations, generating an expanded clip containing the intervening sequence. Figure adapted from Trubitsyna et al. (2014).

1.8 Aim and objectives

1.8.1 Aim

To study the enzymatic depolymerization of green macroalgal biomass.

1.8.2 Specific objectives

- To design a module for the depolymerization of cellulosic biomass.
- To identify new enzymes involved in the depolymerization of ulvan.
- To design new DNA parts potentially useful for the construction of an ulvan depolymerization module.

2 Materials and Methods

2.1 Materials

2.1.1 Chemicals and reagents

Table 2-1 summarises the main chemicals and reagents used during this study.

Table 2-1 Chemicals and reagents used during the study

Supplier	Chemical or reagent
Alfa Aesar	Cellobiose
Calbiochem	4-Methylumbelliferyl-β-D-glucuronide (MUGlcU)
Megazyme	Azo-CM-Cellulose Cellohexaose Cellopentaose Cellotriose
Melford Laboratories	5-bromo-4-chloro-3-indoyl-β-D-galactoside (X-gal) Agarose Chloramphenicol (CML) Isopropyl β-D-1-thiogalactopyranoside (IPTG)
NBS Biologicals	SafeView Nucleic Acid Stain
New England Biolabs	1 kb and 50 bp DNA ladders Q5® High-Fidelity DNA Polymerase Restriction enzymes: EcoRI, XbaI, SpeI, PstI and DpnI T4 DNA ligase T4 Polynucleotide Kinase
Novagen	KOD Hot start DNA Polymerase
Promega	Kit Taq DNA Polymerase
QIAGEN	QIAquick PCR Purification Kit QIAprep Spin Miniprep Kit QIAquick Gel Extraction Kit
Sigma-Aldrich	3,5-Dinitrosalicylic acid (DNS) 4-Methylumbelliferyl-β-D-cellobioside (MUC) 4-Methylumbelliferyl-β-D-glucopyranoside (MUG) 4-Methylumbelliferyl-β-D-xylopyranoside (MUX) 4-Nitrophenyl-α-L-rhamnopyranoside (pNPR) Carboxymethyl cellulose (CMC)

Supplier	Chemical or reagent
Sigma-Aldrich	Congo red
	Glucose
	Remazol brilliant Blue R (RBB)
	Sea salts
	Yeast extract
Thermo Scientific	BigDye Terminator v3.1 Cycle Sequencing Kit
	Coomassie (Bradford) protein assay kit
VWR	Cetylpyridinium chloride monohydrate

2.1.2 Bacterial strain and expression vectors

DNA manipulation and cloning was carried out using *E. coli* JM109 and *E. coli* TOP10. *E. coli* JM109, *E. coli* TOP10, *E. coli* BL21 and *Formosa agariphila* KMM 3901 were used for growth assay and protein expression. More details about the strains used during the study are summarised in Table 2-2.

Table 2-2 List of strains used during the study

Strain	Genotype	Supplier
E. coli JM109	F´ *traD36 proA*$^{+}$*B*$^{+}$ *lacI*^{q} *Δ(lacZ)M15/ Δ(lac-proAB) glnV44 e14*$^{-}$ *gyrA96 recA1 relA1 endA1 thi hsdR17*	New England Biolabs
E. coli TOP10	F^{-} *mcr*A Δ(*mrr-hsd*RMS-*mcr*BC) Φ80*lac*ZΔM15 Δ*lac*X74 *rec*A1 *ara*D139 Δ(*ara leu*) 7697 *gal*U *gal*K *rps*L (StrR) *end*A1 *nup*G	Invitrogen
E. coli BL21 (DE3)	*E. coli* str. B F^{-} *ompT gal dcm lon* *hsdS*$_{B}$($r_{B}^{-}m_{B}^{-}$) λ(DE3 [*lacI lacUV5-T7 gene 1 ind1 sam7 nin5*]) [*malB*$^{+}$]$_{K-12}$(λS)	New England Biolabs
F. agariphila KMM 3901	Native strain	DSMZ

Construction and expression of the parts were carried out using the plasmid pSB1C3 obtained from the Registry of Standard Biological Parts (partsregistry.org). pSB1C3 is a high copy number plasmid (100-300 copies per cell) carrying a pMB1 replication origin, a chloramphenicol resistance gene and terminators bracketing its multiple cloning site (Figure 2-1).

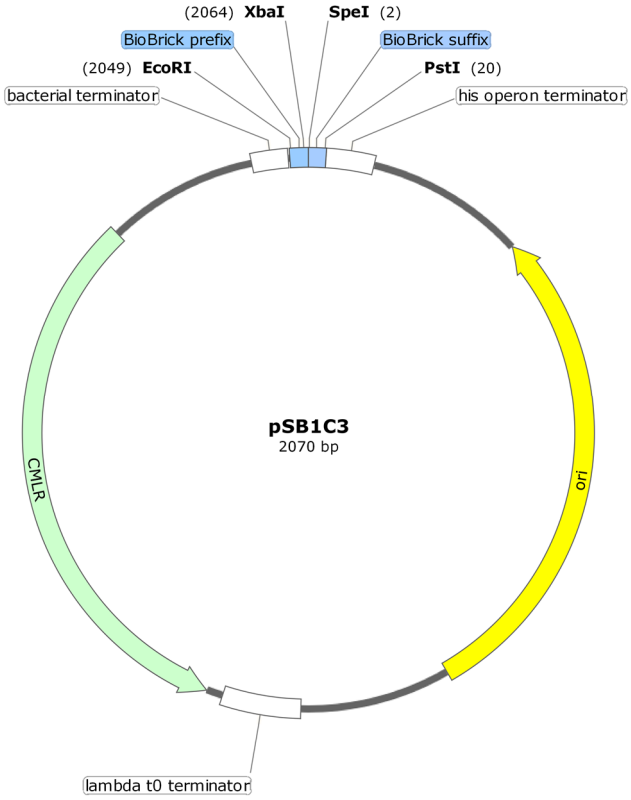

Figure 2-1 Plasmid map of pSB1C3 and its sequence reference points. CMLR: Chloramphenicol resistance gene; ori: pMB1 replication origin. Plasmid map generated using SnapGene® software (GSL Biotech).

2.1.3 Restriction enzymes

Table 2-3 shows the different restriction enzymes used during this study.

Table 2-3 Restriction enzymes used during the study

Enzyme	Restriction site (5'→3')*
EcoRI-HF	G▼AATTC CTTAA▲G
XbaI	T▼CTAGA AGATC▲T
SpeI-HF	A▼CTAGT TGATC▲A
PstI-HF	C▼TGCAG GACGT▲C
DpnI	Gm6A▼TC CT▲m6AG

* The black triangles indicate the cleavage site of the enzyme. m6A: N6-Methyladenine.

2.1.4 Primers and oligonucleotides

Oligonucleotides for PaperClip DNA assembly method were designed according to Trubitsyna et al. (2014). For each part or intervening sequence to be assembled four oligonucleotides were designed: upstream forward (UF), upstream reverse (UR), downstream forward (DF) and downstream reverse (DR).

Two divergent non-overlapping primers, one forward and one reverse, were designed to delete or insert DNA sequences using MABEL (Mutagenesis with blunt-end ligation). To delete a DNA fragment from a construct the divergent primers were designed flanking the unwanted region. To insert a short DNA fragment into a construct a tail at the 5' end of one of the primers containing the desired sequence was added.

VF2 and VR primers flanked the BioBrick cloning site and were used for sequencing constructs in pSB1C3 and routine colony PCR to check the lengths of the assemblies. Additionally, sequence specific primers were designed to sequence internal regions of the constructs.

All the primers and oligonucleotides were ordered from Integrated DNA Technologies. Primer and oligonucleotide sequences are listed in Appendix 8.1.1.

2.1.5 DNA parts

Table 2-4 summarizes the DNA parts used during the study; including the type, length and origin of each of them.

Table 2-4 DNA parts used during this study

Name	Type	Origin	Length [bp]
pSB1C3	Backbone	iGEM Registry	2059
BBa_J33207	Promoter	iGEM Registry	600
P_{lacZ}*	Promoter	*E. coli*	379
CHU2268**	CDS	*C. hutchinsonii*	2280
CHU2268ΔSP	CDS	*C. hutchinsonii*	2226
CHU2268HT	CDS	*C. hutchinsonii*	2244
CenA***	CDS	*C. fimi*	1350

Name	Type	Origin	Length [bp]
CenAΔSC	CDS	*C. fimi*	1347
Cex***	CDS	*C. fimi*	1461
Cfbglu***	CDS	*C. fimi*	1455
OsmY	CDS	*E. coli*	603
Ag43SP	Signal peptide	*E. coli*	156
Ag43N455	CDS	*E. coli*	1758
FA2190	CDS	*F. agariphila*	1146
FA2194	CDS	*F. agariphila*	843
FA2195	CDS	*F. agariphila*	795
FA2197	CDS	*F. agariphila*	2475
FA2204	CDS	*F. agariphila*	3435
FA2205	CDS	*F. agariphila*	2976
FA2206	CDS	*F. agariphila*	2946
FA2209	CDS	*F. agariphila*	2712
FA2213	CDS	*F. agariphila*	2118
FA2216	CDS	*F. agariphila*	1716
FA2217	CDS	*F. agariphila*	2796
FA2219	CDS	*F. agariphila*	1551
tFA2219	CDS	*F. agariphila*	951
FA2219ΔSP	CDS	*F. agariphila*	1473
tFA2219ΔSP	CDS	*F. agariphila*	873
tFA2219ΔSPΔSC	CDS	*F. agariphila*	870
FA2220	CDS	*F. agariphila*	1374
FA2222	CDS	*F. agariphila*	1059
FA2223	CDS	*F. agariphila*	2211
FA2225	CDS	*F. agariphila*	3465

* Designed by Maryia Trubistsyna, it contains P_{lac} + *lacZ'*. ** First cloned into pSB1A2 by Chao-Kuo Liu. *** First cloned into pSB1A2 by Damian Barnard.

Table 2-5 shows the ribosome binding sites (RBSs) used during this study and their respective strengths.

Table 2-5 List of RBSs

Name	Sequence (5'→3')	Strength*
BBa_J15001**	TACTAGAGCTCAAGGAGGTACTAG	13670.10

Name	Sequence (5'→3')	Strength*
CHU2268 RBS	GGCAAGTAGTAGTCGCAAACGACAGGGCGCGGGGTCCCCC	9536.96
CHU2268ΔSP RBS	CTCCTTTCAGAGCAAAGTCAGCTACAGGCTCGGAGGATCG	9864.36
CHU2268HT RBS	CGCTCGTTAACGAACGCGAAATTCTAGGCCGGAGTTCTCG	9862.91
CenA RBS	TGGTCTAACGTCGCCGCCCGAGGAGATAAAAACATAGGTAGCC	10889.42
FA2190 RBS	GCCTTCAACGGCTACACGAATATAATCACACCCGGTACGA	9514.13
FA2197 RBS	AAAAACCACCATCAGGGTTTCGATACCAAGGGGCGTTACT	10410.21
FA2204 RBS	CGACGAACAACGAGTTCACAGAATCTTAGGGGGACAGCTC	9545.55
FA2205 RBS	GTCGTAAGAGGGCCACCGGGAAGGTAATAGGAACAGGAAA	10915.55
FA2206 RBS	TGGGTCACCCGGCCAACGGTAAATATCACGGATTAAAAGA	9536.96
FA2209 RBS	GCATGGCTGGGTTAGACCCCATCCTTACATAGGTTACTGT	10029.99
FA2213 RBS	ATAACAACGAGGTATCTCATAGTTAATCGGAGGGCCACGA	10974.66
FA2216 RBS	TTTAGGCAAGCAAATTTATATCGAAGAGAGACAGGCAGTT	10193.82
FA2217 RBS	AGGCCCACCAAGGAACAAGCTCAAGGAACACCCTAGTAAA	9117.27
FA2219 RBS	GGAGGCTATCCCAGGACGAAGCCGTAAGTACGAGGAGGCAGCC	10889.42
FA2219ΔSP RBS	CAATTCTCTAACACTACGTCCAACGCAGGAAGCGAGGAGCGCC	10915.55
FA2220 RBS	TGTCTCGCAGCTATTAGGACACTCAAAGCAGGAGCATAAT	9952.09
FA2222 RBS	CGGTAAAAACGTTTCCACACGCATTATAAGGACGGAGAGG	9952.09
FA2223 RBS	CCAACGACCGTCGCACGTACATTCAGCAAGAGGTCTAGAA	10491.7
FA2225 RBS	GTTTCCAAAAAACTCAAGACATAACAAGGAGCAAACGCCT	11053.97
OsmY RBS	ACATAGCAAGAATTAAGGCGAAACGGCAATAAGAGGATCGGCC	10029.99
Ag43 RBS	CGGACACCTCGCTAAATCGATCACAACCCAAGAGAATTTCGCC	9117.27
Ag43 1k RBS	AAGAAAAAGCTCTCCATAA	1055.96

Name	Sequence (5'→3')	Strength*
Ag43 5k RBS	TGGCCATCTCTAAGTAATCC	4437.53

* The strength of the RBSs was calculated using the RBS calculator (Salis et al., 2009). ** CHU2268 was used as a reference to calculate the strength of this RBS.

Table 2-6 shows a list of linkers used in this study with their respective nucleic acid and amino acid sequences.

Table 2-6 List of linkers

Name	Nucleic acid sequence (5'→3')	Amino acid sequence
FlexL	GGAGGCGGAGGAAGCGGCGGTGGTGGCTCTGGT GGAGGTGGCTCGGGTGGCGGCGGTTCG	$[GGGGS]_4$
RigidL	GAAGCAGCCGCAAAGGAGGCTGCAGCAAAG	$[EAAAK]_2$
CexL	GGCGCGAGCCCGACGCCGACGCCCACCACGCCG ACCCCGACGCCCACGACGCCGACGCCGACCCCG ACGTCCGGTCCGGCCGGG	GASPTPTPTTPTPTP TTPTPTPTSGPAG
OmpTL	GGTGGTCGTCGTTCTCGTCGTGTTGGTACC	GGRRSRRVGT
OmpTfL	GGTGGTGGTGGTTCTGGTGGTCGTCGTTCTCGT CGTGTTGGTACTGGTGGTGGTGGTTCT	GGGGSGGRRSRRVGT GGGGS

2.2 Methods

2.2.1 DNA manipulation

2.2.1.1 *PCR conditions*

PCR was performed in a reaction mixture containing 0.2 mM dNTPs, 0.5 µM forward primer, 0.5 µM reverse primer, 1X Q5 Reaction Buffer, 1 U/ µl Q5® High-Fidelity DNA Polymerase and either 0.25 µl of plasmid DNA or a small amount of colony as the template in a final volume of 50 µl. 1X Q5 High GC Enhancer was used for sequences with high GC content. The amplification conditions were: 1 cycle at 98°C for 30 s; 30 cycles of 98°C for 10 s, annealing temperature optimized for each primer set for 30 s and 72°C for 30 s/kb; and a final extension cycle at 72°C for 2 min.

Colony PCR was performed in a reaction mixture containing 0.2 mM dNTPs, 0.5 µM reverse primer, 0.5 µM forward primer, 1X Green GoTaq® Reaction Buffer, 0.025 U/µl GoTaq® DNA Polymerase and a small amount of colony in a final volume of 15 µl. The amplification conditions were: 1 cycle at 95°C for 2 min; 30 cycles of 95°C for 0.5 min, annealing temperature optimized for each primer set for 0.5 min and 72°C for 1 min/kb; and 1 final additional cycle at 72°C for 10 min.

PCR products were analyzed by agarose gel electrophoresis.

2.2.1.2 *Agarose gel electrophoresis*

Standard agarose gel electrophoresis was carried out using 8 g/l agarose gels in 1X TAE buffer at 90 V and for 30-40 min. 20 g/l agarose gels run at 60 V were used for small DNA fragments (e.g. clips). New England Biolabs 1 kb and 50 bp DNA ladders were used as the markers. Gels were stained using 0.1 µl/ml SafeView Nucleic Acid Stain and visualized under blue light.

2.2.1.3 *DNA purification*

DNA fragments amplified by PCR were purified using the QIAquick PCR Purification Kit or the QIAquick Gel Extraction Kit according to the manufacturer's instructions.

2.2.1.4 *BioBrick assembly*

Two digestions were performed in parallel to obtain the insert and the acceptor. SpeI-HF and PstI-HF were used to digest the acceptor, while Xbal and PstI-HF were used to digest the insert. The digestions were carried out at 37°C for 2 h in a reaction mixture containing: 1X CutSmart buffer, 1 U/µl enzyme 1, 1 U/µl enzyme 2 and 3 µl of plasmid DNA in a final volume of 20 µl. The digested samples were checked by agarose gel electrophoresis and purified using the QIAquick PCR Purification Kit according to the manufacturer's instructions. The purified fragments were then ligated at 16°C overnight in a mixture containing: 1X T4 DNA ligase, 40 U/µl of T4 DNA ligase 3.5 µl of the acceptor and 10.5 µl of the insert in a final volume of 20 µl. 5 µl of the assembly reaction were used for transformations.

2.2.1.5 *PaperClip assembly*

DNA parts were amplified by PCR from plasmid DNA or colonies using their corresponding UF and DR oligonucleotides as primers (section 2.1.4). When plasmid DNA was used as the template, 1 µl of DpnI was added after the PCR and the sample incubated for 1 h at 37°C to digest the parental plasmid and reduce background colonies after transformation. All samples were purified using the QIAquick PCR Purification Kit according to the manufacturer's instructions. The concentration of the DNA parts were measured on a Nanodrop 2000 spectrophotometer (Thermo Scientific, US). All parts were stored at -20˚C until use.

Upstream and downstream half-clips were prepared mixing a final concentration of 40 µM of each of the corresponding forward and reverse oligonucleotides in a final volume of 50 µl. The oligonucleotides were aligned together by slow cooling down (1˚C/s) from 95˚C to 4˚C. All half-clips were stored at -20˚C until use.

Clips were prepared by ligation of the downstream half-clips associated to the part 1 with the upstream half-clip associated to the part 2. Upstream and downstream half-clips were first phosphorylated for 30 min at 37°C in a reaction mixture containing 14 µM of each half-clip, 1X T4 ligase buffer and 1 U T4 polynucleotide kinase in a final volume of 10 µl. After the phosphorylation 1 U T4 ligase was added and the samples incubated at 16˚C for 1 h. When short intervening sequences were added (e.g. RBSs or protein linkers) a two step ligation was used to generate the clips. In the first step, in two different tubes the downstream oligonucleotide associated to part 1 and the upstream oligonucleotide associated to part 2 were phosphorylated and ligated as above with the upstream half of the intervening sequence and the downstream half of the intervening sequence, respectively. In the second step the two tubes were mixed together and incubated for an additional hour at 16 ˚C resulting in the complete clip. All clips were checked by agarose gel electrophoresis.

The assembly reaction contained: 4 ng/µl of backbone, 2 ng/µl of each other part, 60 nM of each clip, 1xKOD Hot Start DNA polymerase buffer, 0.2 mM dNTPs, 1.5 mM $MgSO_4$ and 0.02 U/µl KOD Hot Start DNA polymerase in a final volume of 50 µl. 1X Q5 High GC Enhancer was used when parts with high GC content were assembled. The assembly conditions were: 1 cycle at 95°C for 2 min; 5 cycles of 95°C for 20 s and 70°C for 25 s/kb. 5 µl of the assembly reaction were used for transformation.

2.2.1.6 *MABEL (Mutagenesis with blunt-end ligation)*

MABEL was used to insert or delete DNA fragments from constructs that were already cloned into pSB1C3. PCR was performed using two divergent non-overlapping primers designed as described in section 2.1.4 and plasmid DNA as the template. The PCR product was treated with 1 µl of DpnI for 1 h at 37°C to destroy the parental plasmid and then purified using the QIAquick PCR Purification Kit according to the manufacturer's instructions. The products were phosphorylated at 37°C for 30 min in a reaction mixture containing: 1X T4 DNA ligase reaction buffer, 1 U/µl T4 Polynucleotide Kinase and 10 µl DNA in a final volume of 20 µl. The samples were then self-ligated at 16 ˚C overnight using 40 U/µl of T4 DNA ligase. The enzymes

were heat inactivated for 10 min at 65˚C. 5 µl of the final products were used for transformation.

2.2.1.7 *DNA sequencing*

All the constructs were sequence verified by Gene Pool, the sequencing service of the University of Edinburgh. The samples were prepared in a reaction mixture containing: 2 µl 5X sequencing buffer, 1 µl BigDye, 1 µl primer (forward or reverse), 2 µl of plasmid DNA and 4 µl distilled H_2O. The sequencing conditions were: 50 cycles of 95°C for 30 s, 50°C for 20 s and 60°C for 4 min, and a cycle at 60°C for 1.25 min. Primers used for sequencing are described in section in section 2.1.4.

2.2.2 Transformation of *E. coli* and DNA plasmid minipreps

Competent *E. coli* cells were prepared and transformed as described by Chung et al. (1989). An additional heat shock step of 90 s at 42°C followed by 90 s on ice was performed before adding the Lysogeny broth (LB) for recovery. To transform cells, 1 µl of plasmid DNA or 5 µl of assembly were added to the competent cells

Plasmid DNA minipreps were prepared from 2 ml of a 5 ml overnight *E. coli* culture using the QIAprep Spin Miniprep Kit according the manufacturer's instructions.

2.2.3 Media and culture conditions

LB or LB agar were the chosen media used for routine *E. coli* culturing. 40 µg/ml chloramphenicol (CML) was used for selection when the plasmid pSB1C3 was used. 90 µg/ml IPTG and 40 µg/ml X-gal were used for induction and selection when the constructs contained P_{lacZ}. All liquid cultures were incubated at 37°C with shaking at 200 rpm.

Marine broth (MB) or MB agar were used for routine *F. agariphila* culturing. The composition of this culture medium is a modification of the Marine Broth 2216 from Difco™, using Sea Salts to substitute all the minerals. The medium also contains peptone and yeast extract that provide a good source of nutrients. Liquid cultures were

incubated at 21°C with shaking at 200 rpm. Solid culture were incubated at room temperature.

Culture media recipes are given in Appendix 8.1.1.

2.2.3.1 *E. coli growth on cellodextrins*

M9 medium agar plates supplemented with a 5 g/l carbon source, 0.2 g/l yeast extract 40 μg/ml CML and 90 μg/ml IPTG were used for growth assays on solid medium. Glucose and cellobiose were used as the carbon source. Glucose, cellobiose and yeast extract were filter-sterilized and added to the medium after autoclaving the mineral base.

When liquid medium was used single colonies were used to inoculate 5 ml of LB supplemented with 40 μg/ml CML and incubated at 37°C with shaking at 200 rpm overnight. In order to remove the LB the overnight cultures were centrifuged for 10 min at 16000 g, washed twice and re-suspended in 1ml of M9 medium. 96-well plates (Greiner Bio-One, Germany) were loaded with 200 μl of M9 medium supplemented with 40 μg/ml CML, 90 μg/ml IPTG, 0.2 g/l yeast extract, 4 g/l of the different carbon sources and the overnight cultures diluted to OD 600nm 0.02. Glucose, cellobiose, cellotriose, cellotetraose and cellohexaose were used as the carbon sources. The cultures were incubated at 37°C in a Sunrise Blue microplate reader (Tecan Ltd, UK) for 72 h with shaking at 150 rpm. OD 595 nm was recorded every 15 min. The experiments were done in triplicate.

2.2.3.2 *E. coli growth on cellulose paper*

Single colonies were used to inoculate 5 ml of LB supplemented with 40 μg/ml CML and incubated at 37°C with shaking at 200 rpm overnight. Cultures were centrifuged for 10 min at 16000 g, washed twice and re-suspended in 1 ml of M9 medium. The overnight cultures were diluted to OD 600nm 0.5 in a glass universals containing 5 ml of M9 medium supplemented with 40 μg/ml CML, 90 μg/ml IPTG, 0.2 g/l yeast extract and 3 pieces of 1 cm^2 of pure cellulose blotting paper (Whatman® GB003, 0.8 mm,

300 gsm, Sigma-Aldrich). The universals were incubated at 37°C with shaking at 200 rpm and the total protein assay to determine growth (see section 2.2.8).

2.2.3.3 *E. coli growth on enzymatically pretreated ulvan*

Single colonies were used to inoculate 5 ml of LB supplemented with 40 μg/ml CML and incubated at 37°C with shaking at 200 rpm overnight. Cultures were centrifuged for 10 min at 16000 g and resuspended in 1 ml of M9 medium. 96-well plates (Greiner Bio-One, Germany) were loaded with 200 μl of overnight culture diluted to OD 600nm 0.02 in enzymatically pretreated solubilized ulvan (see 2.2.17). The cultures were incubated at 37°C in a SPECTROstar Nano microplate reader (BMG Labtech, Germany) for 18 h with shaking at 700 rpm. OD 600nm was recorded every 6 min. The experiment was done in triplicate.

2.2.3.4 *F. agariphila growth assays*

F. agariphila colonies grown on MB plates for 3 days at room temperature were used to inoculate flasks containing 50 ml of a 35 g/l sea salts preparation supplemented with 1 g/l yeast extract and different carbon sources. The carbon sources used were 5 g/l peptone, 4 g/l glucose and 10 g/l *U. lactuca*. The flasks were incubated at 21°C with shaking at 200 rpm and OD 600nm recorded to determine growth.

2.2.4 Heterologous protein expression

E. coli colonies with the desired constructs, checked by sequencing, were used to inoculate vials containing 5 ml of LB with 40 μg/ml CML. Each vial was incubated overnight at 37°C with shaking at 200 rpm and 1 ml transferred to a flask containing 50 ml LB. The new culture was incubated under the same conditions until OD 600nm ~0.6 was reached and protein expression was induced with 90 μg/ml IPTG for 4 h or overnight.

2.2.5 Crude lysate preparation

Cells with the expressed proteins were transferred to 50 ml disposable centrifuge tubes and harvested by centrifugation at 5000 g for 30 min at 4°C. The pellet was re-suspended in 1 ml of phosphate buffered saline (PBS) and kept on ice until use. The cells were lysed by sonication using 10 cycles of 10 s bursts (12 μm amplitude) and 20 s of rest on ice slurry. The sonicated cells were centrifuged at 16000 g for 5 min and the supernatant saved as the cell lysate. The pellet was washed twice with 1ml of PBS and the last resuspension saved as the cell debris.

2.2.6 Cell fractionation

The cells were harvested by centrifugation (5000 g for 30 min at 4°C) and the supernatant saved as the extracellular fraction. To separate cytoplasmic and periplasmic fractions an osmotic shock was carried out as described by Rutter et al. (2013). The pellets were resuspended in 3 ml of shock buffer (0.5 mM sodium ethylenediaminetetraacetic acid (EDTA), 100 mM Tris–HCl, pH 8.0, phenylmethylsulfonyl fluoride and 0.5 M sucrose). 1 ml of this suspension was incubated on ice for 5 min and then centrifuged at 16000 g for 5 min. The pellets were warmed up to room temperature and resuspended in 1.5 ml of ice-cold water. After 1 min on ice, 85 µl of 20 mM $MgCl_2$ was added. Cells were centrifuged at 16000 g for 5 min and the supernatant saved as the periplasmic fraction. The pellets were resuspended in 1.5 ml of PBS and lysed by ultrasonication using the same conditions as for the total cell lysates. The samples were centrifuged for 5 min at 16000 g and the supernatant saved as the cytoplasmic fraction. The pellets were washed twice, resuspended in 1.5 ml of PBS and saved as the debris.

2.2.7 Protein quantification

The protein concentration of the different samples was determined using the Bradford Assay Reagent. For each sample, a cuvette with 1 ml of Bradford Assay Reagent was mixed with an appropriate amount of sample (1-2 µl), covered with parafilm and mixed several times by inversion. A standard curve was constructed by mixing 20 µl samples of different concentration of Bovine Serum Albumin (BSA) with 1 ml of Bradford

Assay Reagent (see Appendix 8.1.3). The absorbance of each sample or standard was measured at OD 595nm. All samples were measured in triplicate.

2.2.8 Total protein assay

The total protein assay was used when the OD results weren't reliable due to the presence of insoluble substrates in the media. 100 µl of culture was mixed with 900 µl of Bradford Assay Reagent in a cuvette and the mixture incubated at 65°C for 1 hour to lyse the cells. The mixture was allowed to cool down for 1 h at room temperature and the absorbance of the sample measured at 595nm. A standard curve was constructed measuring OD 600nm and total protein of different concentrations of an *E. coli* culture grown on M9 medium supplemented with 5 g/l glucose, 0.2 g/l yeast extract and 40 µg/ml CML (see Appendix 8.1.3).

2.2.9 SDS PAGE

SDS-PAGE (Sodium dodecyl sulphate polyacrylamide gel electrophoresis) was carried out as described by Sambrook and Russell (2001) using the mini-PROTEAN kit from Bio-Rad, with a stacking and resolving gel of 5 and 12% (w/v) of polyacrylamide, respectively. Electrophoresis was performed at 80 V until all the sample was at the end of the stacking gel and then the voltage was increased to 180 V for 45 min. The gel was stained with Coomassie Brilliant Blue.

2.2.10 His tag purification

His tag purification was carried out using HisTrap HP columns (GE Healthcare Lifesciences, US) according to the manufacturer's recommendations. Binding buffer consisted of 25mM TrisHCl (pH 8), 20 mM imidazole, 0.3 M NaCl and 10% (v/v) glycerol and it was also used for the lysis of the cells. An imidazole step gradient of 0.1, 0.2, 0.3 and 1 M was used for elution.

2.2.11 Buffer exchange

Buffer exchange was performed using PD-10 desalting columns (GE Healthcare Lifesciences, US) according to the manufacturer's instructions. PBS was used as the elution buffer.

2.2.12 Source of *Ulva lactuca* samples

Dry *U. lactuca* biomass was obtained from La Herradura de Guayacán bay, Chile, and was kindly provided by Dr. Cristian Bulboa from Universidad Andrés Bello, Chile.

2.2.13 Preparation of alcohol insoluble residue (AIR)

Alcohol insoluble residue (AIR) of *U. lactuca* was obtained as described by Fry (1988). 20 g of *U. lactuca* was mixed with 160 ml of 70 % (v/v) ethanol and the biomass milled using a blender. The suspension was stirred using a magnetic stirrer bar for 24 h and the biomass collected by filtration through Miracloth (Calbiochem, Germany). The previous step was repeated several times washing with 200 ml of 70% (v/v) ethanol until the filtrate was colourless and free of low molecular weight compounds. The final material was air-dried at room temperature and stored in an air tight bottle to prevent moisture damaging the samples

2.2.14 Ulvan purification

Ulvan was extracted from the *U. lactuca* AIR using ammonium oxalate as a chelating agent (Fry, 1988). A suspension containing 500 mg of AIR in 5 ml 0.2M ammonium oxalate (pH 4 adjusted with formic acid) was incubated for 2.5 h at 85^0C and then stirred overnight at 37°C. The sample was centrifuged at 4500 g for 30 min and the supernatant saved. The pellet was washed with 5 ml of deionized water and centrifuged at 4500 g for 30 min. The supernatants were combined and dialyzed against deionized water using a 3.5K molecular-weight cut off dialysis cassette (Thermo Fisher Scientific, US) and then freeze-dried.

2.2.15 Preparation of unsaturated ulvan oligosaccharides

A stock containing 10 g/l of the purified ulvan in PBS (pH 7.0±0.2) was prepared. 1.5 ml of the stock were mixed with 20 µl of cell lysate containing the ulvan lyase FA2219ΔSP from *F. agariphila* and incubated at 37°C overnight. The enzyme was inactivated for 20 min at 80^0C and the oligosaccharide mixture filtered (0.2 µm). Formation of unsaturated ends was confirmed by detection of C=C double bonds by increase of absorbance at 235nm.

2.2.16 Ulvan solubilisation

U. lactuca was milled and resuspended at 10 g/l in water. A pure hydrothermal pre-treatment was carried out by autoclaving the sample at 121^0C for 15 min or incubating it at 85^0C for 4h. The pre-treated samples were centrifuged for 30 min at 4500 g and the supernatant containing the solubilised ulvan saved and analysed by thin layer chromatography (TLC). Autoclaving was used as the standard pre-treatment and the biomass resuspended in different media for the solubilisation.

2.2.17 Enzymatic saccharification of soluble ulvan

Different combinations of enzyme extracts were used for the enzymatic saccharification of soluble ulvan. The reaction was carried out in a reaction mixture containing 1.9 ml of ulvan solubilized in M9 medium and 20 µl of each enzyme extract. The samples were incubated at 37°C overnight.

2.2.18 Activity assays

2.2.18.1 *Congo Red activity assay*

The Congo Red activity assay described by Teather and Wood (1982) was used as qualitative assay for endoglucanase activity detection. Activity assays were performed either using live cells or enzyme extracts. When live cells were used, 5 ml of LB supplemented with 40 µg/ml CML was inoculated with a single colony and the culture incubated at 37°C with shaking at 200 rpm overnight. The culture was centrifuged for

10 min at 16000 g and the pellet resuspended in water. The cells were diluted to OD 600nm 1 and a 10 µl drop used to inoculate a 2 g/l CMC-LB-agar plate containing 90 µg/ml IPTG and 40 µg/ml CML. When enzyme extracts were used, a 3 µl aliquot was deposited on a 2 g/l CMC-PBS-agar plate. The plates were incubated at 37°C overnight and then flooded using 5 ml of 1 mg/ml Congo Red solution for 20 min at room temperature. Finally, Congo Red was removed and the plate washed with 10 ml of 1 M sodium chloride for 15 min at room temperature. Unstained halos correspond to the areas where the CMC has been degraded.

2.2.18.2 *Azo-CM-Cellulose activity assay*

For each sample 10 µl of enzyme extract were mixed with 20 µl Azo-CM-Cellulose substrate solution and 470 µl of PBS (pH 7.0±0.2) in a 2 ml tube. The tube was incubated for 1 h at 37°C and the reaction terminated by adding 1 ml of absolute ethanol and mixing vigorously for 10 s. The sample was centrifuged for 5 min at 16000 g and 1 ml of the supernatant transferred into a cuvette to measure absorbance at 590nm. Each sample was tested in triplicate. A standard curve was prepared measuring absorbance at 590nm of different concentrations of Remazol brilliant Blue R (RBB) (see Appendix 8.1.3). A unit of enzyme activity [U] was defined as the amount of the enzyme that catalyzes the release of 1 mmole of RBB per minute at 37°C.

2.2.18.3 *4-Methylumbelliferone based activity assays*

For each sample 20 µl of enzyme extract was mixed in a cuvette with 960 µl of PBS (pH 7.0±0.2) and 20 µl of 10 mM 4-methylumbelliferyl-β-D-glucopyranoside (MUG), 10 mM 4-methylumbelliferyl-β-D-cellobioside (MUC), 10 mM 4-Methylumbelliferyl-β-D-xylopyranoside (MUX) or 10 mM 4-Methylumbelliferyl-β-D-glucuronide (MUGlcU). The cuvette was incubated at 37°C for 1 h and fluorescence recorded with the Modulus™ Single Tube Multimode Reader (Turner BioSystems, US). Each sample was tested in triplicate. A standard curve was prepared measuring the fluorescence of different concentrations of 4-methylumbelliferone (MU) (see

Appendix 8.1.3). A unit of enzyme activity [U] was defined as the amount of the enzyme that catalyzes the release of 1 µmole of MU per minute at 37°C.

2.2.18.4 *4-Nitrophenyl-α-L-rhamnopyranoside activity assay*

For each sample 20 µl of crude lysate was mixed with 960 µl of PBS (pH 7.0±0.2) and 20 µl of 10 mM p-Nitrophenyl-α-L-rhamnopyranoside (pNPR) in a cuvette. The cuvette was incubated at 37°C for 1 h and OD 410nm recorded with a SPECTROstar Nano reader (BMG Labtech, Germany). Each sample was tested in triplicate. A standard curve was prepared measuring the OD 410nm of different concentrations of p-Nitrophenol (pNP) (see Appendix 8.1.3). A unit of enzyme activity [U] was defined as the amount of the enzyme that catalyzes the release of 1 µmole of pNP per minute at 37°C.

2.2.18.5 *Cetyl pyridinium chloride activity assay*

The plate assay described by (Gacesa & Wusteman, 1990) was used as qualitative assay for ulvan lyase activity detection. Activity assays were performed either using live cells or enzyme extracts. When live cells were used, 5 ml of LB supplemented with 40 µg/ml CML were inoculated with a single colony and the culture incubated at 37°C with shaking at 200 rpm overnight. The culture was centrifuged for 10 min at 16000 g and the pellet resuspended in water. The cells were diluted to OD 600nm 1 and a 10 µl drop was used to inoculate a 1 g/l ulvan-LB-agar plate with 90 µg/ml IPTG and 40 µg/ml CML. When enzyme extracts were used, a 3 µl aliquot was deposited on a 1 g/l ulvan-PBS-agar plate. The plates were incubated at 37°C overnight and then flooded using 5 ml of 10% (w/v) cetyl pyridinium chloride solution for 20 min at room temperature. Unstained halos correspond to the areas where the ulvan has been degraded. The same results were obtained when using either purified or solubilized ulvan for the plates. Plates containing solubilized ulvan were prepared from a starting solution containing 10 g/l *U. lactuca*.

2.2.18.6 *DNS activity assay*

Reducing sugar equivalents were quantified by the method of Miller (1959) with some modifications. For each sample 10 to 40 µl of enzyme extract were mixed with 160 µl ulvan solubilized in PBS (pH 7.0±0.2) (see 2.2.16) and PBS (pH 7.0±0.2) up to 200 µl. Specific blanks were used for each sample containing the same mixture but without the ulvan. Samples were incubated at 37°C for 4h or overnight. PCR tubes containing 90 µl of DNS solution (10 g/l 3,5-dinitrosalicylic acid, 0.5 g/l sodium sulphite, 10 g/l sodium hydroxide and 300 g/l potassium sodium tartrate) were mixed with 90 µl of the reaction. Samples were incubated at 90°C for 12 min and cooled on ice for 5 min to stop the reaction. 100 µl were transferred into a 96 well plate and absorbance at 590nm recorded with a SPECTROstar Nano reader (BMG Labtech, Germany). Each sample was tested in triplicate. A standard curve was prepared measuring absorbance 590nm of different concentrations of rhamnose (see Appendix 8.1.3). A unit of enzyme activity [U] was defined as the amount of the enzyme that catalyzes the release of 1 µmole of reducing sugar per minute at 37°C.

2.2.18.7 *Unsaturated β-glucuronyl hydrolase activity assay*

Enzyme activity was measured by monitoring the decrease in absorbance at 235nm, corresponding to the loss of C=C double bonds in the substrate. A mix of unsaturated oligosaccharides obtained by degradation of purified ulvan by an ulvan lyase was used as the substrate (see 2.2.15). Reactions were performed using UV-Star 96-well plates (Greiner, Austria) in a temperature-controlled SPECTROstar Nano plate reader (BMG Labtech, Germany) set at 37°C. The reaction mixture consisted of 0.5 µl of enzyme extract diluted 50 times and 100 µl of 200 µM of unsaturated oligosaccharides in PBS buffer (pH 7.0±0.2). The concentration of the substrate was calculated using the extinction coefficient of the double bond (4800 $M^{-1}cm^{-1}$) (Macmillan et al., 1964). Each sample was tested in triplicate.

2.2.19 TLC (Thin layer chromatography)

Trifluoroacetic acid (TFA) was used to hydrolyse the samples. When insoluble biomass was used (e.g. *U. lactuca* or *U. lactuca* AIR), 0.01 mg was used and mixed

with 1 ml of 2M TFA. In the case of soluble polysaccharides the samples were centrifuged for 5 min 16000 g and 500 µl of the supernatant mixed with 500 µl of 4M TFA. All samples were incubated at 100^0C for 1 h and cooled down on ice. The samples were centrifuged for 5 min 16000 g and the supernatants transferred into new tubes to be dried at 80^0C for 1 day.

Each sample was resuspended in 100 µl of 0.5% (v/v) chlorobutanol and 3 µl loaded on a TLC silica plate (Merck, Germany). The same amount of a 1 g/l master mix standard (MMS) containing some of the most common cell wall monosaccharides (galacturonic acid (GalA), galactose glucose, mannose, arabinose xylose and rhamnose) and 1/g l Glucuronic acid (GlcA) were loaded as reference. The samples were allowed to dry at room temperature and the plate run in a TLC chamber containing 10 ml of the solvent mixture EPPAW (ethyl acetate, pyridine, propanol, acetic acid and water) in a ratio 4:2:2:1:1. The plate was run for 3 h and air-dried overnight. For visualization the plate was stained with a thymol solution (5 mg thymol, 95 ml ethanol and 5 ml sulfuric acid), allowed to dry at room temperature for 30 min and wrapped in aluminium foil before being heated for 5 min at 105°C for colour development.

2.2.20 Bioinformatics

LipoP (Juncker et al., 2003) and Lipo v1.0 (Berven et al., 2006) were used to predict lipoprotein signal peptides. TatP (Bendtsen et al., 2005) was used to predict Tat signal peptides. TMHMM v2.0 (Sonnhammer et al., 1998) was used to predict transmembrane helices. Phobius (Kall et al., 2004) was used to predict both transmembrane spans and signal peptides. SignalP v4.1 (Petersen et al., 2011) was used to predict Sec signal peptides. BOMP (Berven et al., 2004) was used to predict beta barrel spans. MetaCyc (Caspi et al., 2014) and KEGG (Kanehisa et al., 2016) databases were used as references for the reconstruction of sugar utilization pathways.

3 Development of a cellulose degradation system

3.1 Introduction

In order to develop a heterologous system for cellulose degradation a proper selection of the enzymes that comprise it is crucial. A typical noncomplexed cellulose degrading system includes endoglucanases (EC 3.2.1.4); two kind of exoglucanases, cellodextrinases (EC 3.2.1.74) and cellobiohydrolases (EC 3.2.1.91); and β-glucosidases (EC 3.2.1.21) (Lynd et al., 2002). In particular, β-glucosidases have an essential role in the biomass conversion process hydrolysing cellobiose to glucose, a sugar necessary for cell growth. A library of cellulases encoding BioBrick parts from bacteria *C. fimi* and *C. hutchinsonii*, originally designed in the French lab, will be used to obtain the different enzymes.

The complete genome sequence of *C. hutchinsonii*, a highly active aerobic cellulose degrader, was reported by Xie et al. (2007). Surprisingly, no genes encoding recognizable exoglucanases were detected, leading to the suggestion that this microorganism might either employ novel exoglucanases or has an unusual mechanism for cellulose degradation (Wilson, 2009; Xie et al., 2007).

However, a recent study reported that CHU2268 from *C. hutchinsonii*, a predicted β-glucosidase, possesses activity against the exoglucanase model substrate MUC and it is not able to hydrolyse the typical β-glucosidase model substrate MUG, suggesting that this enzyme may be the "missing" exoglucanase of the *C. hutchinsonii* cellulose degradation system (Liu, 2012).

CHU2268 seems to be an interesting candidate for the development of a heterologous cellulose degradation system; nonetheless, further analysis of the enzyme is required. In order to validate the above results, a His tag was added to CHU2268 for purification and the enzyme assayed for β-glucosidase and exoglucanase activity. In addition, growth analysis using *E. coli* JM109 cells transformed with this enzyme using different cellodextrins as a carbon source, was performed. Finally, constructs under the control of a single inducible *lac* promoter containing the gene encoding for the enzyme

CHU2268 and the endoglucanase CenA from *C. fimi* were assembled in order to obtain a microorganism able to grow on pure cellulose paper as a sole carbon source.

3.2 Characterization and purification of the enzyme CHU2268 from *C. hutchinsonii*

3.2.1 Cloning and expression of CHU2268

CHU_2268 from *C. hutchinsonii* encodes a 758 amino acid protein (CHU2268), with a predicted molecular weight of 83.7 kDa. CHU2268, a candidate periplasmic β-glucosidase (Xie et al., 2007), contains an N-terminal GH3-family domain, a C-terminal fibronectin type III domain of unknown function, and no recognizable carbohydrate-binding module (CBM). Conserved domains were identified using the CD-search tool from the NCBI Conserved Domain Database (CDD) (Marchler-Bauer et al., 2007). Additionally, a 19 amino acid N-terminal lipoprotein signal peptide was predicted using the software LipoP 1.0 Server (Juncker et al., 2003).

CHU2268 was obtained from a construct already cloned in pSB1A2 (Registry of Standard Biological Parts) containing BBa_J33207 and BBa_J15001 upstream of the gene to add a *lac* promoter and a strong RBS. The whole construct was re-cloned into pSB1C3 (Registry of Standard Biological Parts), and maintained using 40 mg/l CML (Figure 3-1).

In order to facilitate purification and characterisation of CHU2268, the putative signal peptide of the enzyme was replaced with a 6xHis tag using MABEL mutagenesis. PCR was performed using the divergent non-overlapping primers CHU2268HisF1 and HisTagRI and the CHU2268 construct in pSB1C3 as the template. The PCR product was purified, self-ligated, and the new plasmid used to transform *E. coli* JM109 cells. The new construct was validated by sequencing and the resulting protein designated CHU2268HT (Figure 3-1).

E. coli JM109 cells were transformed with both constructs and cultivated in liquid medium with the purpose of obtaining enzyme extracts for activity assays. For each construct 5 ml LB vials containing 40 mg/l CML were incubated with shaking

overnight at 37°C, and then used to inoculate 50 ml LB flasks. The flasks were incubated under the same conditions until OD_{600} ~0.6 was reached and protein expression induced with 90 μg/ml IPTG for 20 h. Cells were harvested by centrifugation, re-suspended in PBS and lysed by sonication. The broken cells were centrifuged and the resulting supernatant retained for further analysis. Cultures containing the construct pSB1C3-P_{lac}-*lacZ'* were used as negative control.

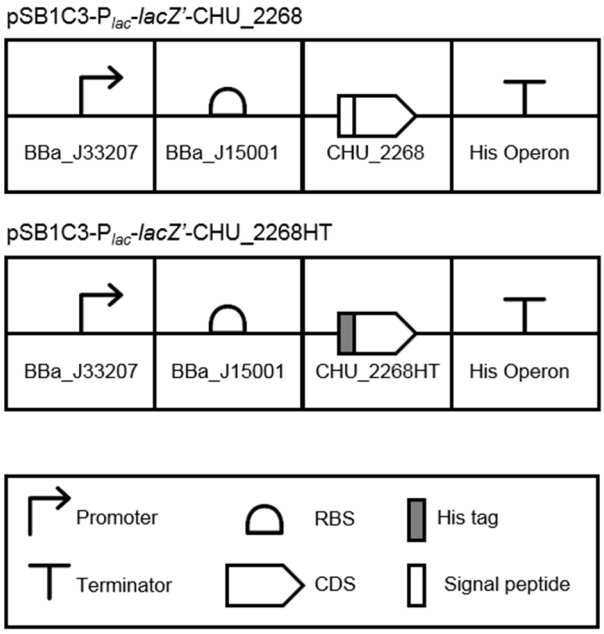

Figure 3-1 CHU2268 and CHU2268HT constructs for expression in *E. coli*. The construct pSB1C3-P_{lac}-*lacZ'*-CHU2268 was assembled using the BioBrick assembly method. A C-terminal His tag was added to CHU2268 and its signal peptide removed using MABEL mutagenesis. CHU2268: β-glucosidase from *C. hutchinsonii*, BBa_J33207: P_{lac}-*lacZ'*, BBa_J15001: strong RBS. The constructs are described according to the Synthetic Biology Open Language (SBOL). RBS: ribosome binding site, CDS: coding DNA sequence.

MUC and MUG were used as model substrates to determine exoglucanase and β-glucosidase activity, respectively. Crude lysate samples were mixed with each of the substrates, and the reactions carried out at 37°C. The results obtained are shown in Figure 3-2 (A). Surprisingly, no activity was observed when the sample containing the CHU2268 construct was used. In contrast, CHU2268HT showed activity against both substrates, activity against MUG being over 2-fold higher than activity against MUC. None of the negative controls showed activity.

The activity assays were repeated using the same conditions but with samples obtained from cultures induced at 30°C instead of 37°C (Figure 3-2 (B)). This time the CHU2268 construct showed clear activity against MUG and very little activity against MUC. Higher levels of activity were observed in the CHU2268HT samples, with a similar MUG/MUC activity ratio to the one obtained when the cultures were induced at 37°C. This results suggest that CHU2268 might be not only a β-glucosidase but also a cellodextrinase.

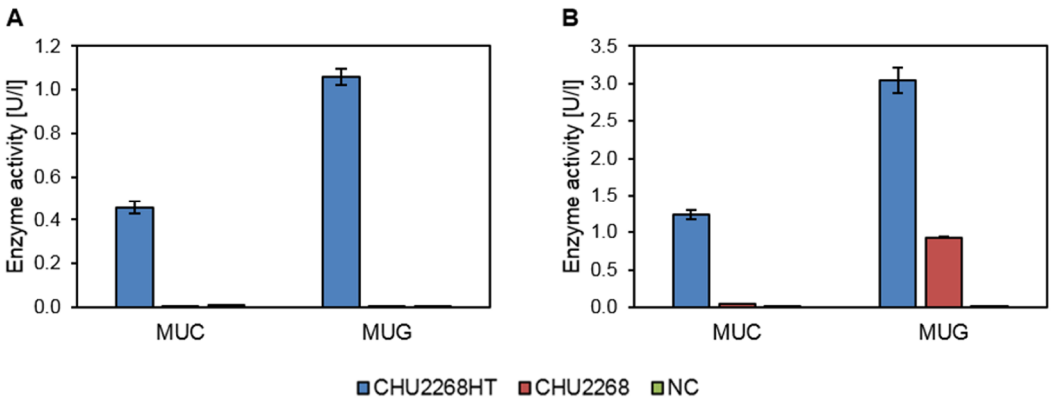

Figure 3-2 MUC and MUG activity assays for CHU2268HT and CHU2268. Cell lysates were prepared from cultures grown at 37°C (A) and 30°C (B). Activity assays were performed using 0.2 mM of MUC or MUG as the substrate. Samples were incubated at 37°C for 1 h. Experiments were done in triplicate. Activities are expressed in units of activity per litre of culture volume. Error bars indicate one standard error of the mean. NC: pSB1C3-P_{lac}-*lacZ'*.

The above results suggest that the modifications to the enzyme, replacing the signal peptide with a His tag, had an important effects on its expression. In order to determine whether this change was due to the absence of the signal peptide new constructs were designed, including one containing a version of CHU2268 with neither a signal peptide nor a His tag designated CHU2268ΔSP (Figure 3-3). The assemblies were done using the PaperClip assembly method (Trubitsyna et al., 2014). All parts were amplified by PCR using their respective UF and DR primers and clips prepared to connect the different parts. Parts included the backbone (pSB1C3), promoter (P_{lac}), and coding DNA sequence. RBS were added as intervening sequences in the middle of the clips connecting the promoters with the coding sequences. Considering the high influence of the sequence following the start codon in the protein translation initiation rates (Salis et al., 2009), specific RBSs were designed using the RBS calculator (Salis et al.,

2009) to have similar translation rates for all the coding sequences. After the assembly reaction, the constructs were used to transform *E. coli* JM109 cells and the sequences validated by sequencing.

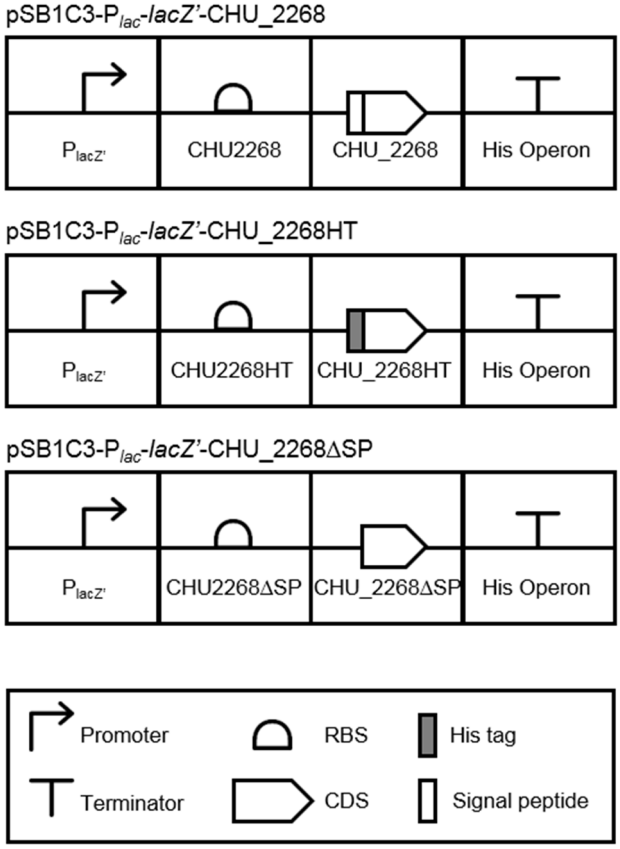

Figure 3-3 CHU2268 constructs containing specific RBSs for expression in *E. coli*. All constructs were assembled using the PaperClip assembly (Trubitsyna et al., 2014). RBSs of similar strength were designed using the RBS calculator (Salis et al., 2009). CHU2268: β-glucosidase from *C. hutchinsonii*, $P_{lacZ'}$: P_{lac}-*lacZ'*. The constructs are described according to the Synthetic Biology Open Language (SBOL). RBS: ribosome binding site, CDS: coding DNA sequence.

E. coli JM109 cells containing the new constructs were cultivated in liquid media using the same conditions as described above; however, induction was for 20 h at 37°C. Cell lysates were prepared and used for activity assays against MUC and MUG (Figure 3-4). Both CHU2268HT and CHU2268ΔSP showed activity against the two substrates, presenting a similar profile, while CHU2268 showed little or no activity. These results confirm that the signal sequence is deleterious to expression of active enzyme CHU2268.

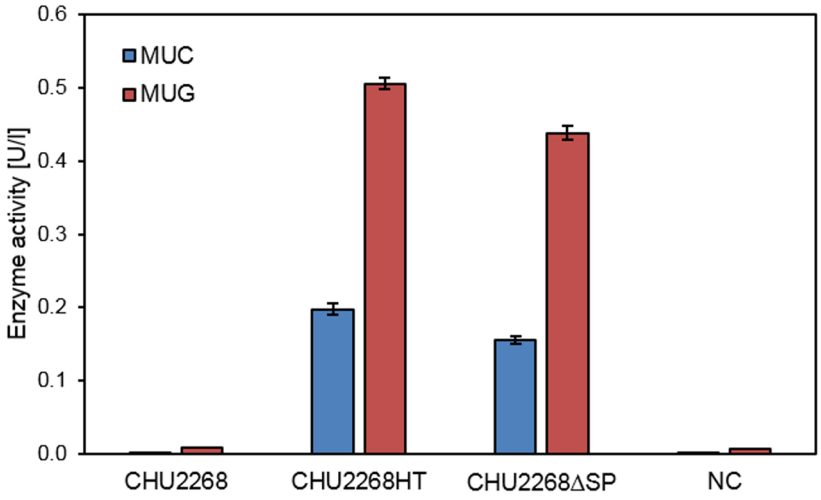

Figure 3-4 MUC and MUG activity assays for CHU2268 constructs containing specific RBSs. Activity assays were carried out using cell lysates and 0.2 mM of MUC or MUG as the substrate. Samples were incubated at 37°C for 1 h. Experiments were done in triplicate. Activities are expressed in units of activity per litre of culture volume. Error bars indicate one standard error of the mean. NC: pSB1C3-P_{lac}-*lacZ'*.

3.2.2 Growth of recombinant *E. coli* on cellodextrin-M9 media

Having demonstrated that the enzymes CHU2268 and CHU2268HT are active against the model substrates MUC and MUG, cell growth experiments were performed to determine whether *E. coli* cells expressing these enzymes were able to utilize the β-glucosidase substrate cellobiose. To do that, *E. coli* JM109 cells transformed with the constructs pSB1C3-P_{lac}-lacZ'-CHU2268 and pSB1C3-P_{lac}-lacZ'-CHU2268HT were grown on plates containing cellobiose as the sole carbon source. M9-agar plates containing 5 g/l cellobiose, 0.2 g/l yeast extract, 90 μg/ml IPTG and 40 μg/ml CML were used for the growth assay. Additionally, plates containing 5 g/l glucose as the carbon source and plates with no sugar were used as positive and negative controls for the medium. The plates were streaked with colonies containing the different constructs and incubated at 37°C overnight. Cells transformed with the construct pSB1C3-P_{lac}-*lacZ'* were used as a negative control for the cells. The results are shown in Figure 3-5. As expected all plates had growth when glucose was used as the carbon source. Poor growth was observed when the plates were supplemented only with yeast extract. In the case of the plates supplemented with cellobiose, only the cells expressing the enzymes CHU2268 and CHU2268HT were able to grow. These results indicate that

CHU2268 is capable of hydrolysing cellobiose to glucose at a sufficient rate for cell growth and its use could be key for the development of a cellulose degradation system.

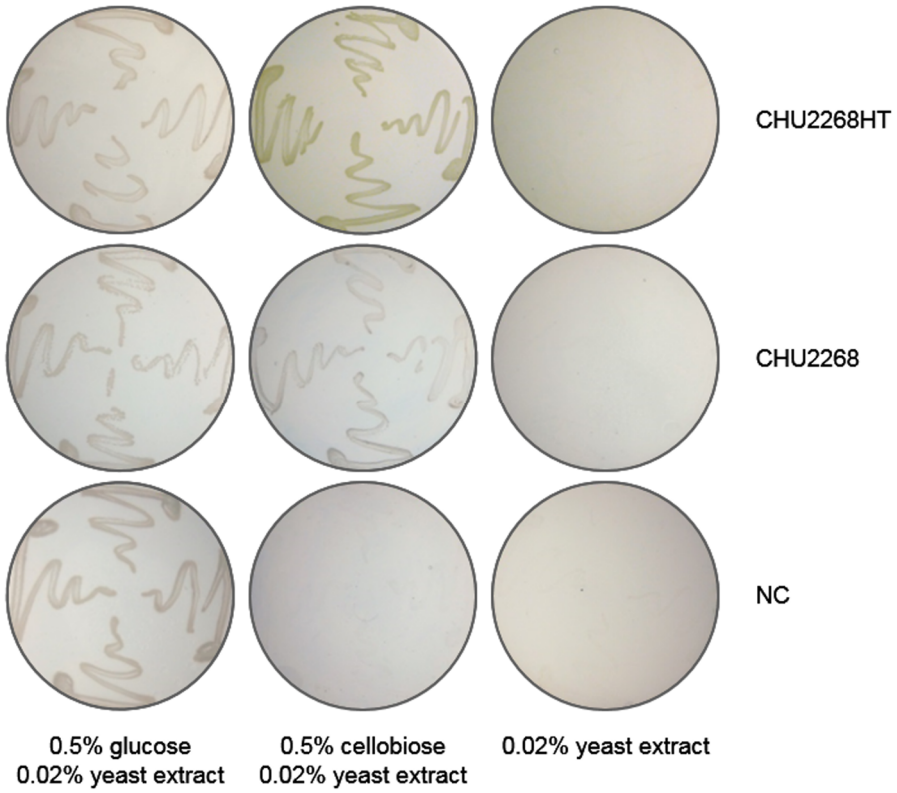

Figure 3-5 Growth of recombinant *E. coli* JM109 on 5 g/l cellobiose M9 medium. Glucose and cellobiose in combination with yeast extract were used as the carbon source. *E. coli* cells transformed with constructs containing the enzyme CHU2268 with and without the His tag were used for the assay. Plates were induced with 90 μg/ml IPTG and incubated at 37°C overnight. NC: pSB1C3-P_{lac}-*lacZ'*.

In order to confirm the exoglucanase activity of CHU2268, a growth analysis using cellodextrins of different degree of polymerisation as sole carbon source was done. *E. coli* JM109 cells transformed with the constructs for the expression of CHU2268, CHU2268HT and Cfbglu (locus tag Celf_2783), a 'typical' β-glucosidase from *C. fimi* (Wakarchuk et al., 1984), were used to inoculate M9 medium containing 4 g/l cellodextrin, 0.2 g/l yeast extract, 90 μg/ml IPTG and 40 μg/ml CML. Cells containing the construct pSB1C3-P_{lac}-*lacZ'* were used as a negative control. The cultures were grown for 72 h at 37°C and cell growth recorded. All cultures, including the negative control, were able to grow when glucose was used as the carbon source; however, the cells expressing CHU2268 grew at a considerably lower rate than those containing the

other constructs, probably due to toxicity of its signal peptide (Figure 3-6 (A)). Confirming the results obtained from the growth on plates, *E. coli* JM109 cells expressing CHU2268 or CHU2268HT were able to grow well utilizing cellobiose as the sole carbon source, with superior growth than the one observed in cells expressing Cfbglu (Figure 3-6 (B)).

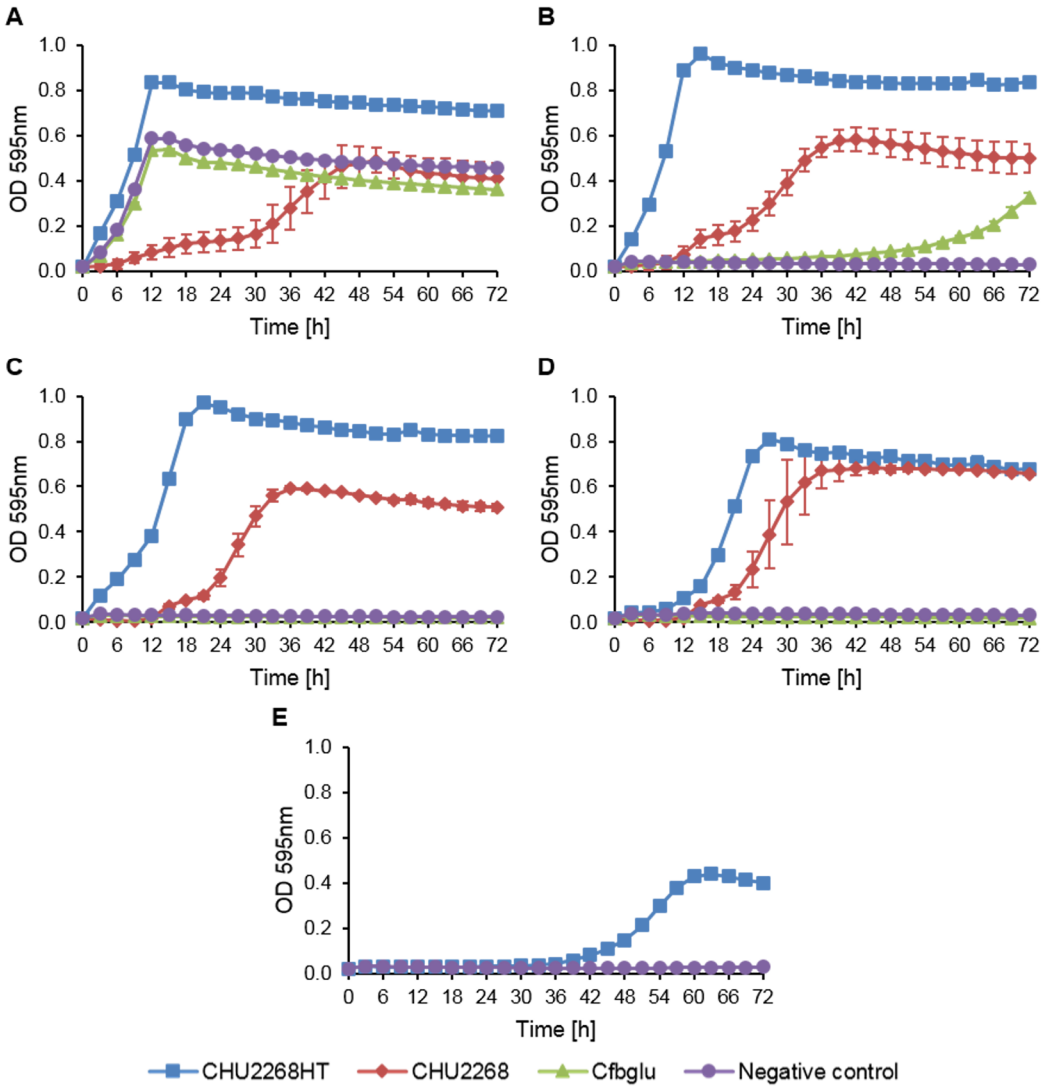

Figure 3-6 Growth of *E. coli* JM019 expressing CHU2268, CHU2268HT or Cfbglu, on various cellodextrins. In all cases, substrates are provided at 4 g/l in M9 medium with 0.2 g/l yeast extract. A negative control strain (pSB1C3-P_{lac}-*lacZ'*) is included in all experiments. A: glucose, B: cellobiose, C: cellotriose, D: cellotetraose, and E: cellohexaose. All experiments included three biological replicates. Error bars indicate one standard error of the mean.

Moreover, cells expressing either CHU2268 or CHU2268HT were also able to grow in minimal medium containing cellotriose or cellotetraose, whereas such growth was not observed in cells expressing Cfbglu (Figure 3-6 (C-D)). *E. coli* expressing CHU2268HT was even able to grow using cellohexaose (Figure 3-6 (E)). However, except for cellobiose, where the cells had a similar growth rate as on glucose, the growth rate slowed down as the degree of polymerisation of the cello-oligosaccharides increased (Figure 3-7). These results suggest that CHU2268 has both β-glucosidase (EC 3.2.1.21) and cellodextrinase activity (EC 3.2.1.74).

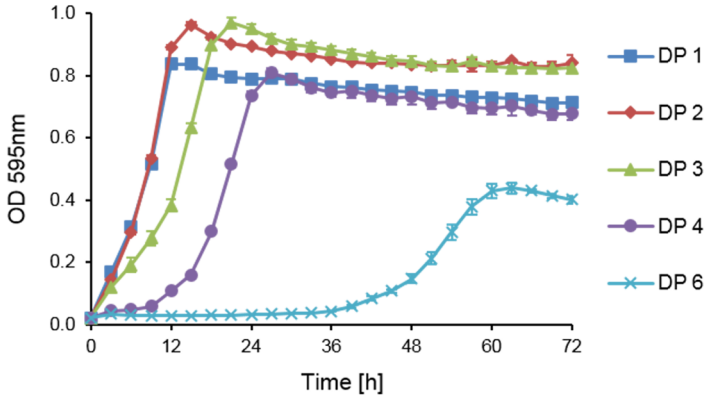

Figure 3-7 *E. coli* JM109 expressing CHU2268HT grown on cellodextrins of different degree of polymerization. DP: degree of polymerization. All experiments included three biological replicates. Error bars indicate one standard error of the mean.

In order to study the subcellular localization of the proteins in *E. coli*, cultures induced for 20 h at 30°C were used for cell fractionation. Cells containing the construct pSB1C3-P_{lac}-*lacZ'* were used as the negative control. The different sub-cellular fractions were analyzed using MUC and MUG as the substrates. Activity assays were carried out for 1 h at 37°C. The negative control did not show activity in any of the fractions. Both CHU2268 and CHU2268HT samples showed activity only in the cytoplasm when MUC was used as the substrate (Figure 3-8 (A)). The same results were obtained for CHU2268HT when MUG was used. Surprisingly, when MUG was used CHU2268 activity was observed not only in the cytoplasm fraction but also in the extracellular fraction, indicating that some enzyme was secreted (Figure 3-8 (B)).

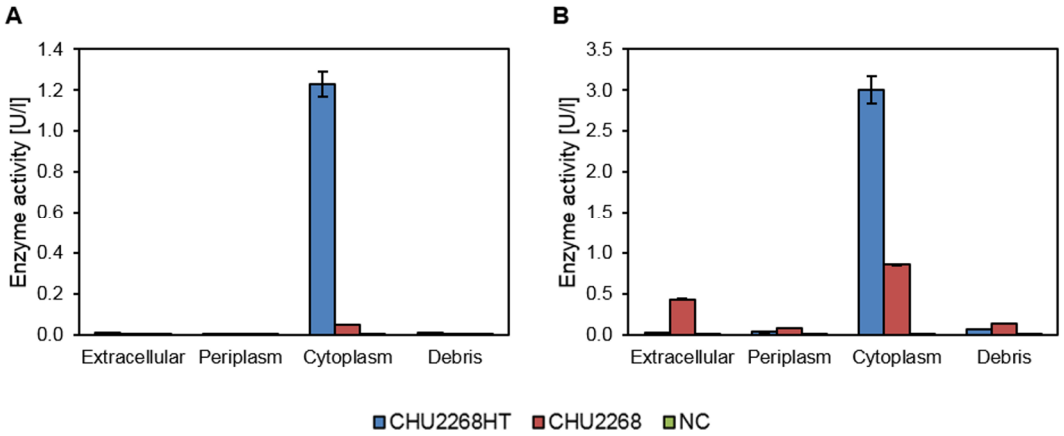

Figure 3-8 Localization of CHU2268 and CHU2268HT in *E. coli* JM109. Activity assays were performed using samples from the different fraction of the cell and 0.2 mM of MUC (A) or 0.2 mM of MUG (B) used as the substrates. Samples were incubated at 37°C for 1 h. Experiments were done in triplicate. Activities are expressed in units of activity per litre of culture volume. Error bars indicate one standard error of the mean. NC: pSB1C3-P_{lac}-*lacZ'*.

3.2.3 CHU2268 purification

Having confirmed that the enzyme CHU2268 does not lose its activity due to the addition of a 6xHis tag, CHU2268HT was purified from the cell lysate using HisTrap HP columns. The enzyme was eluted with an imidazole gradient of 0.1, 0.2, 0.3 and 1 M. SDS-PAGE of the purified samples was performed to check in which of the different fractions the enzyme was eluted and the degree of purity (Figure 3-9). A strong single band appeared in the fraction corresponding to the elution with 0.2 M imidazole; however, the band appears slightly smaller than the expected molecular weight for the enzyme (82.5 kDa).

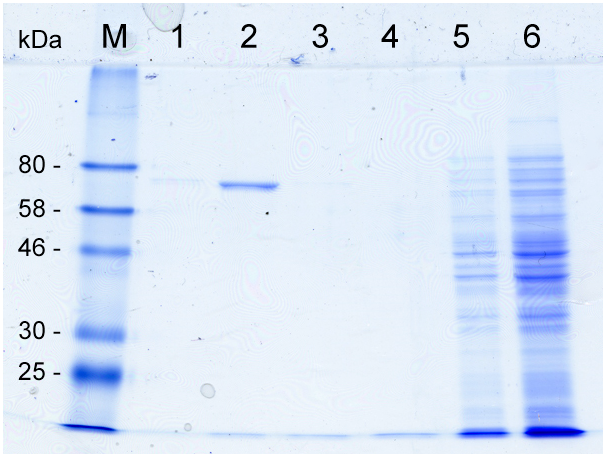

Figure 3-9 SDS-PAGE of the purified CHU2268. The gel shows an elution gradient with different concentrations of imidazole. Lane M: protein marker (7-175 kDa), lane 1: 0.1M imidazole; lane 2: 0.2 M imidazole; lane 3: 0.3 M imidazole; lane 4: 1 M imidazole; lane 5: 0.03 mg of cell lysate; and lane 6: 0.1 mg of cell lysate. A 12% (w/v) acrylamide gel stained with Coomassie Brilliant Blue was used.

In order to ensure that the purified enzyme is actually the desired one, the different fractions were assayed for MUC and MUG activity immediately after the purification (Figure 3-10), confirming that the purification process was successful and most of the enzyme is present in the fraction eluted with 0.2 M imidazole.

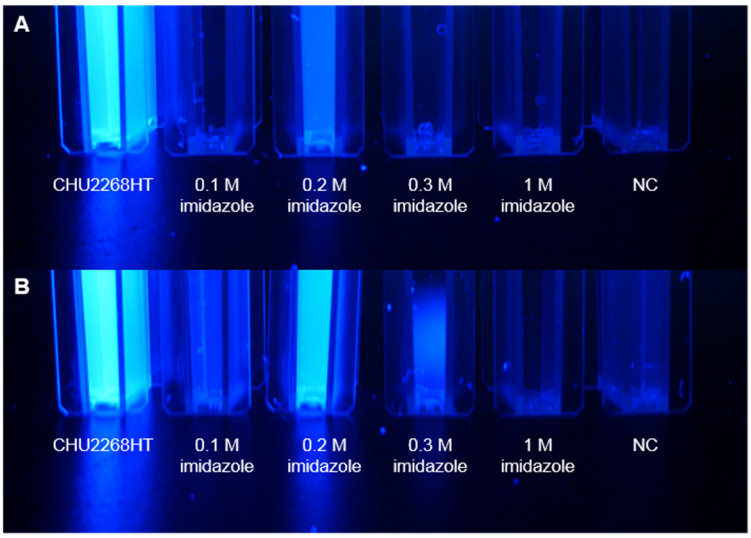

Figure 3-10 MUC and MUG activity assays of the purified CHU2268. Activity assays were performed using 0.2 mM of MUC (A) or 0.2 mM of MUG (B) as the substrate. Samples were incubated at 37°C for 1 h. CHU2268HT: cell lysate before the purification. NC: pSB1C3-P_{lac}-*lacZ'*. Cuvettes were visualized under UV light.

The activity assay was repeated using the fraction eluted with 0.2 M imidazole and MUC and MUG as the substrates 3 to 4 h after the previous one. The results show a complete lack of activity in the samples containing the purified enzyme (Figure 3-11), suggesting that the enzyme is rapidly inactivated by the buffer used for the elution process.

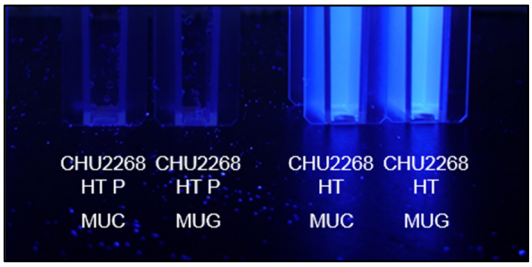

Figure 3-11 MUC and MUG activity assays after CHU2268 purification. Activity assays were performed using 0.2 mM of MUC or MUG as the substrate. Samples were incubated at 37°C for 1 h. CHU2268HT: cell lysate before the purification; CHU2268HT P: 0.2M imidazole elution. Cuvettes were visualized under UV light.

In order to solve the inactivation problem a buffer exchange to PBS using PD-10 columns was performed immediately after the purification process. The samples with the new buffer were tested for activity against MUC and MUG and the results are shown in Figure 3-12. Although the purified enzyme has activity against both substrates, it is considerably lower than the activity of the sample which was not purified. This result can be explained by several factors such as a partial inactivation of the enzyme during the time that it was exposed to the elution buffer, an increase in the dilution degree of the sample after the buffer exchange or the loss of enzyme in other elution fractions during the purification process, amongst other possibilities.

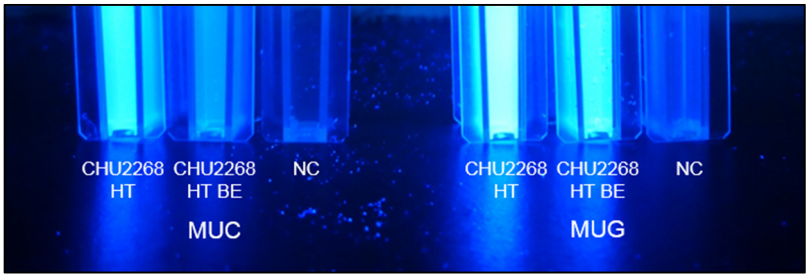

Figure 3-12 MUC and MUG activity assays after buffer exchange. Activity assays were performed using 0.2 mM of MUC or MUG as the substrate. Samples were incubated at 37°C for 1 h. CHU2268 HT: cell lysate before the purification; CHU2268 BE: purified sample after buffer exchange; NC: pSB1C3-P_{lac}-*lacZ'*. Cuvettes were visualized under UV light.

Finally, purification tables were prepared using MUC and MUG as the substrates. The results are shown in Table 3-1 and Table 3-2, respectively. The results were consistent, having purifications yields of 39% and 38%.

Table 3-1 Purification of CHU2268, table prepared using MUC as the substrate

Step	Volume [ml]	Total activity [U]*	Total protein [mg]	Specific activity [U/mg]	Yield [%]	Purification factor
Enzyme extract	2	0.217	16.3	0.013	100	
His-tag purification	3.5	0.084	0.6	0.139	39	10.4

* A unit of enzyme activity (U) is defined as the amount of the enzyme that catalyses the release of 1 µmole of MU per minute.

Table 3-2 Purification of CHU2268, table prepared using MUG as the substrate

Step	Volume [ml]	Total activity [U]*	Total protein [mg]	Specific activity [U/mg]	Yield [%]	Purification factor
Enzyme extract	2	0.244	16.3	0.015	100	
His-tag purification	3.5	0.093	0.6	0.156	38	10.4

* A unit of enzyme activity (U) is defined as the amount of the enzyme that catalyses the release of 1 µmole of MU per minute.

3.3 Cellulose degradation system

3.3.1 Heterologous secretion of the endoglucanase CenA from *C. fimi*

After having confirmed the dual activity of CHU2268, being able to hydrolyse both cellobiose (EC 3.2.1.21) and cellodextrins (EC 3.2.1.74), an endoglucanase (EC 3.2.1.4) is required to complete the cellulose degradation system. One of the main requisites for this enzyme is that it has to be secreted by *E. coli*, and in this way be able to hydrolyse the cellulose in the culture medium into smaller cello-oligosaccharides that can be taken up by *E.coli*.

The endoglucanases from *C. hutchinsonii* have been reported to be highly toxic or inactive when expression by *E.coli* was attempted (Liu, 2012). Thus, these were discarded as options for the system. However, CenA a well characterized

endoglucanase from *C. fimi* has been functionally secreted/leaked into the culture medium by *E. coli* (Guo et al., 1988), and it seems to be an interesting alternative to build up the cellulose degradation system.

The gene *cenA* encodes a 449 amino acid protein, with a predicted molecular weight of 46.7 kDa. CenA is composed of a 31 amino acid signal peptide, an N-terminal family 2 CBM and a C-terminal GH6-family domain. Conserved domains were identified using the CD-search tool from the NCBI CDD (Marchler-Bauer et al., 2007) and the signal peptide predicted using the software SignalP 4.1 Server (Petersen et al., 2011).

CenA was obtained from a construct already cloned in pSB1A2 (Registry of Standard Biological Parts) containing BBa_J33207 and BBa_J15001 upstream of the gene to add a *lac* promoter and a strong RBS. The whole construct was re-cloned into pSB1C3 (Registry of Standard Biological Parts), and maintained using 40 mg/l CML (Figure 3-13).

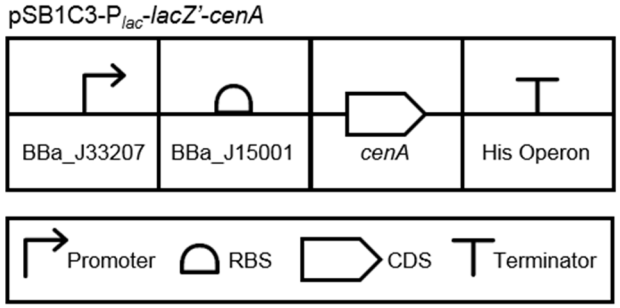

Figure 3-13 CenA construct for expression in *E. coli*. The construct pSB1C3-P$_{lac}$-*lacZ'-cenA* was assembled using the BioBrick assembly method. CenA: endoglucanase from *C. fimi,* BBa_J33207: P$_{lac}$-*lacZ'*, BBa_J15001: strong RBS. The constructs are described according to the Synthetic Biology Open Language (SBOL). RBS: ribosome binding site, CDS: coding DNA sequence.

In order to study the subcellular localization of CenA in *E. coli* JM109, cultures containing the construct pSB1C3-P$_{lac}$-*lacZ'-cenA* were induced for 20 h at 37°C and used for cell fractionation. Cells containing the construct pSB1C3-P$_{lac}$-*lacZ'* were used as the negative control. The different sub-cellular fractions were analyzed using the specific substrate for endoglucanase activity Azo-CM-cellulose. The samples were mixed with the substrate and incubated at 37°C for 1 h. Endoglucanase activity was

detected in all the samples, with the periplasmic and cytoplasmic fractions showing the highest levels of activity (Figure 3-14 (A)). A considerable amount of activity was also observed in the extracellular samples, indicating that *E. coli* is able to secrete CenA. No activity was detected when the fractions from the negative control were examined. Additionally, an activity assay using agar plates containing CMC was performed. Samples from the different subcellular locations were loaded on the plate and incubated at 37°C overnight. The plate was then stained with Congo Red and washed with NaCl. Similar results to the ones from the Azo-CM-Cellulose activity assay were obtained, showing an accumulation of CenA in the periplasm and confirming the secretion of the enzyme into the extracellular space (Figure 3-14 (B)).

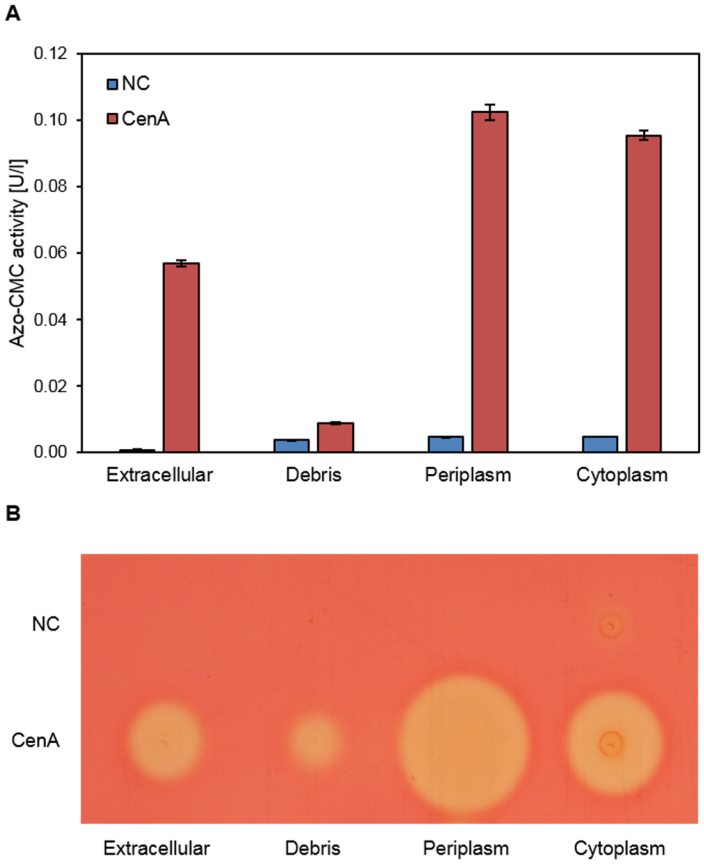

Figure 3-14 Localization of the endoglucanase CenA in *E. coli* JM109. (A) Azo-CM-Cellulose activity assay. Enzyme extracts were mixed with Azo-CM-Cellulose substrate solution and incubated at 37°C for 1 h. Experiments were done in triplicate. Activities are expressed in units of activity per litre of culture volume. Error bars indicate one standard error of the mean. (B) Congo Red activity assay. Enzyme extracts were deposited on a 2 g/l CMC-PBS-agar plate and incubated at 37°C overnight. The plate was stained with Congo Red and washed with NaCl. Unstained halos correspond to areas where CMC has been hydrolysed. NC: pSB1C3-P_{lac}-*lacZ'*.

3.3.2 Assembly of the degradation system

Having demonstrated that the endoglucanase CenA from *C. fimi* is secreted/leaked by *E. coli*, new constructs containing a β-glucosidase and CenA under the control of a single promoter were assembled using BioBrick standard assembly (Figure 3-15). CHU2268, CHU2268HT and Cfbglu were used as the β-glucosidase. The plasmid pSB1C3 was used as the backbone. Successful assembly of the constructs was validated by sequencing. The new constructs were used to transform *E. coli* JM109 cells and maintained using 40 mg/l CML.

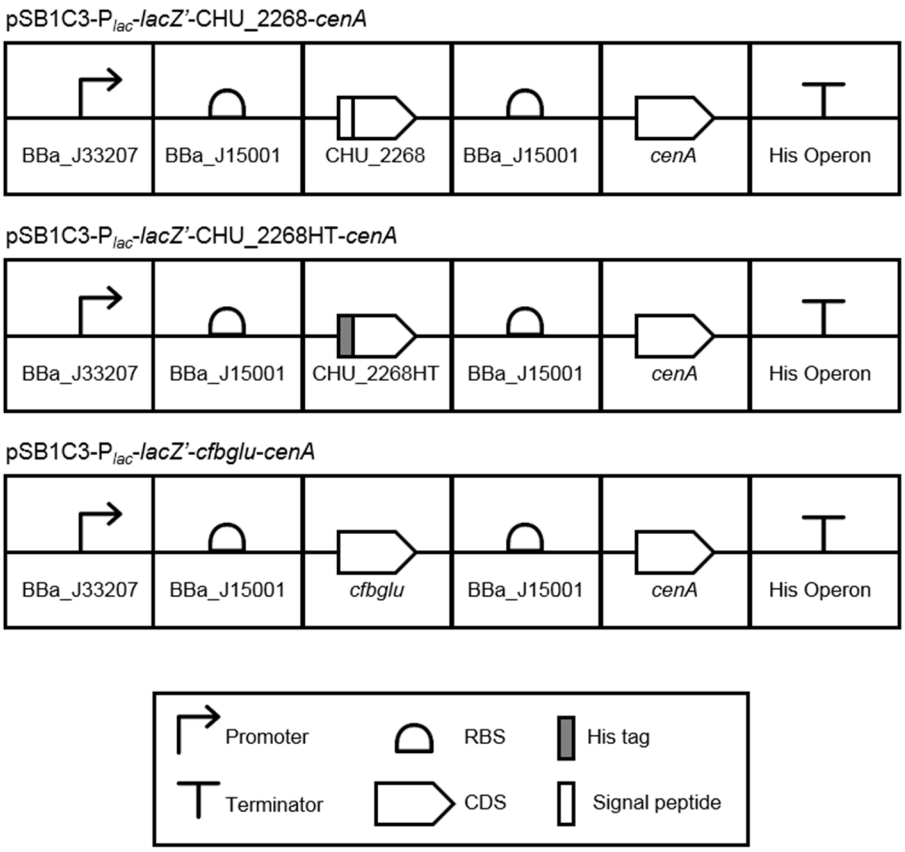

Figure 3-15 Cellulose degradation constructs for expression in *E. coli*. All constructs were assembled using the BioBrick assembly method. CHU2268: β-glucosidase from *C. hutchinsonii*, Cfbglu: β-glucosidase from *C. fimi*, CenA: endoglucanase from *C. fimi*, BBa_J33207: P$_{lac}$-*lacZ'*, BBa_J15001: strong RBS. The constructs are described according to the Synthetic Biology Open Language (SBOL). RBS: ribosome binding site, CDS: coding DNA sequence.

A Congo Red activity assay using live cells containing the different constructs was performed in order to check whether the endoglucanase CenA was being secreted.

Cells containing the construct pSB1C3-P_{lac}-*lacZ'* were used as the negative control. A LB-agar plate containing 2 g/l CMC and 90 μg/ml IPTG for induction was inoculated with cells from overnight cultures containing the different constructs and the plate then incubated at 37°C overnight. The plates were stained with Congo Red and washed with NaCl. Results are shown in Figure 3-16. Unstained halos surrounding the cells correspond to the areas where CMC has been degraded. Activity was observed surrounding all the cells with constructs expressing CenA, confirming that the enzyme is secreted into the medium. Neither the negative control nor the cells expressing only CHU2268 or CHU2268HT were able to degrade the CMC.

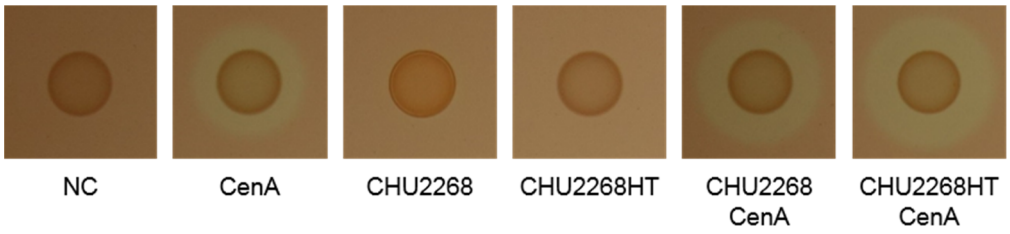

Figure 3-16 CMC activity assay using live cells expressing the cellulose degradation constructs. An LB-agar plate containing 2 g/l CMC and 90 μg/ml IPTG was used for the assay. Cells containing the different constructs were grown overnight at 37°C. The plate was stained with Congo Red and washed with NaCl. Unstained halos correspond to areas where CMC has been hydrolysed. NC: pSB1C3-P_{lac}-*lacZ'*.

3.3.3 Cellulose paper deconstruction and utilization

In order to determine whether the *E.coli* JM109 strain expressing the cellulose degrading systems was able to grow on pure cellulose paper as the sole carbon source, cells containing the different constructs from section 3.3.2 were used to inoculate M9 medium supplemented with 40 μg/ml CML, 90 μg/ml IPTG, 0.2 g/l yeast extract and 9 g/l of pure cellulose blotting paper. The cultures were incubated at 37°C with shaking for 8 days, and growth was monitored every day using the total protein assay. The total protein assay was used because of the presence of insoluble material making OD 600 readings unreliable. The results are shown in Figure 3-17 (A). Total protein readings were converted to OD600 readings using the standard curve from 8.1.3. Only the cells expressing CHU2268HT and CenA were able to grow on cellulose paper. No clear growth was detected for the negative control or the other constructs. The growth was

confirmed by the clear cellulose paper deconstruction which was observed when CHU2268HT and CenA were expressed by *E. coli* (Figure 3-17 (B)).

These results demonstrate that the heterologous expression of the two enzyme degradation system containing: CHU2268HT, a cellodextrinase expressed in the cytoplasm, and CenA, a secreted endoglucanase, is sufficient to confer on *E. coli* JM109 the ability to grow on pure cellulose paper.

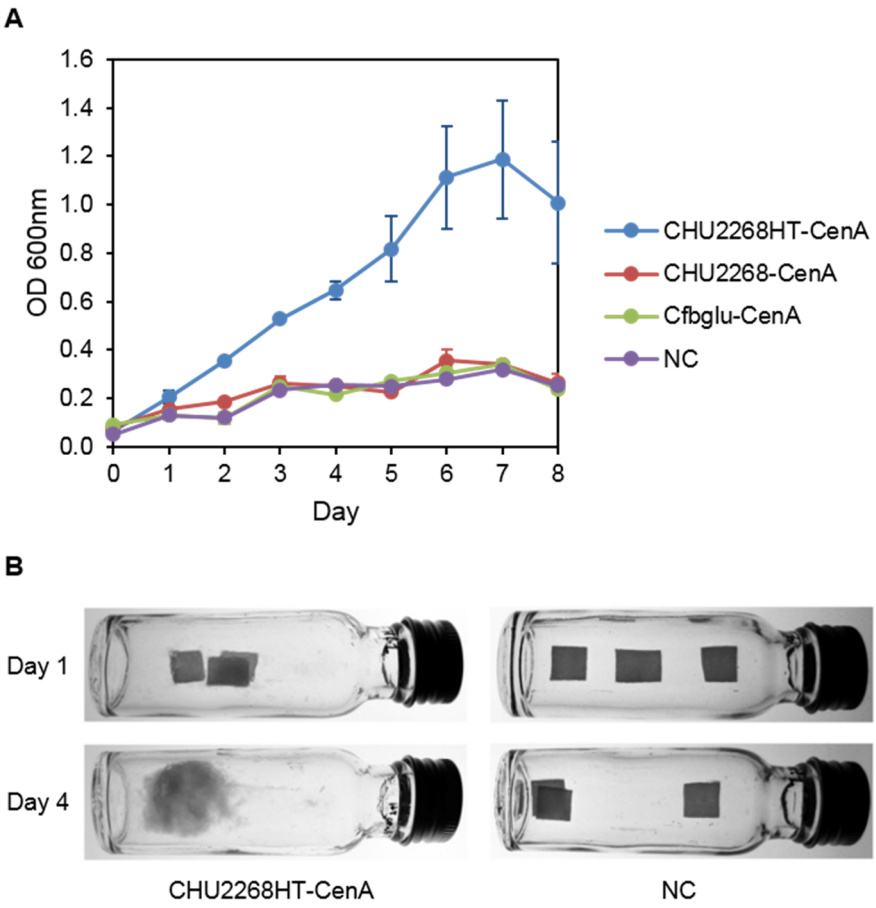

Figure 3-17 Growth of *E. coli* JM019 on cellulose paper expressing cellulose degrading modules. (A) M9 supplemented with 9 g/l cellulose paper and 0.2 g/l yeast extract was used as the medium. The samples were incubated at 37°C. All experiments included three biological replicates. Error bars indicate one standard error of the mean. (B) Cellulose paper deconstruction by co-expression of CHU2268HT and CenA by *E. coli* JM109. NC: pSB1C3-P_{lac}-*lacZ'*.

3.4 Discussion

C. hutchinsonii, an aerobic soil bacterium of the phylum Bacteroidetes, has been reported as an efficient degrader of cellulosic biomass (Zhu et al., 2010). Candidate endoglucanases and β-glucosidases were identified in the *C. hutchinsonii* genome sequence; however, no recognizable exoglucanases were detected, suggesting an unusual method of cellulose utilization (Xie et al., 2007). Recently, it was proposed that CHU2268, a predicted β-glucosidase of *C. hutchinsonii* might be the "missing" exoglucanase of *C. hutchinsonii*, showing activity against MUC, a model substrate for exoglucanases and not against the cellobiose analogue MUG (Liu, 2012).

In this study CHU2268 and CHU2268HT, a version of the enzyme where the secretion signal was replaced by a 6xHis tag, were expressed using *E. coli* JM109 and the activity against MUC and MUG analysed. Surprisingly, only CHU2268HT showed activity and not only against MUC but also against MUG, contradicting the results presented by Liu (2012). A reduction of the induction temperature from 37 to 30°C was required to functionally express the native version of CHU2268. This approach has been commonly used to improve protein folding; however, the drawback of this strategy is a reduction in productivity (Baneyx & Mujacic, 2004). The use of a lower temperature also emulates the conditions at which CHU2268 is expressed in *C. hutchinsonii*, a microorganism that grows at temperatures between 20^0C and 30^0C (Stanier, 1942). However, CHU2268HT did not lose its activity when the cells were induced at 37°C suggesting that the signal peptide might be the reason for the low levels of activity. It has been previously shown that signal peptides can modulate protein stability and folding in *E. coli* (Beena et al., 2004; Singh et al., 2013). In order to check that the increase of activity was due the removal of the signal peptide and not the addition of the His tag new constructs were designed including a version of the enzyme without the signal peptide but no His tag (CHU2268ΔSP). Specific RBSs were used in order to obtain similar protein translation initiation rates. CHU2268ΔSP showed similar levels of activity to CHU2268HT against MUC and MUG, confirming the importance of the signal peptide.

E. coli JM109 expressing either CHU2268 or CHU2268HT was able to grow on plates containing cellobiose as sole carbon source, confirming that the enzyme has β-

glucosidase activity. Additionally, cells expressing CHU2268HT were able to grow well in minimal medium with cellotriose, cellotetraose or cellohexaose, showing that the enzyme is able to release glucose from cellodextrins. CHU2268 also allowed *E. coli* to grow on cellodextrins; however, clear toxicity was observed. Cfbglu, a β-glucosidase from *C. fimi*, also allowed growth on cellobiose but at a considerably lower speed and no growth was observed on any of the other cellodextrins. CHU2268 was not detected in the periplasm but was detected in the cytoplasm and the extracellular fractions to a lesser extent, probably due to cell death. As expected CHU2268HT was found only in the cytoplasm, indicating that *E. coli* is able to take up cello-oligosaccharides at least as large as cellohexaose. Growth rate decreased as the cello-oligosaccharides increased in length, presumably due to slower diffusion of large cellodextrin molecules through the outer membrane (Rutter et al., 2013). These results contradict those presented by Rutter et al. (2013), where *E. coli* growth was only achieved when the cellodextrinase was expressed in the periplasm. The cytoplasmic expression of a β-glucosidase in *E. coli* to allow growth on cellobiose has been also reported by Shin et al. (2014). It has been demonstrated that LacY, a lactose permease, is responsible for cellobiose uptake in *E. coli* (Sekar et al., 2012). A lactose permease has also been utilized for the uptake of cellobiose in *S. cerevisiae* (Sadie et al., 2011). It can be hypothesized that LacY might also facilitate the transport of longer cellodextrins into the cytoplasmic space in *E. coli*. Cellodextrin transport systems including cellodextrin permeases and β-glucosidases have been reported to be essential for optimal fungal growth on cellulose (Galazka et al., 2010). This highlights the importance of these two types of enzymes for the effective utilization of cellulosic biomass. In order to check if LacY actually acts as a cellodextrin permease in *E. coli*, an analogous approach to the one reported by Sekar et al. (2012) to check whether this permease was responsible for cellobiose uptake could be used. This would involve testing the growth on cellodextrins of an *E. coli* strain where the gene *lacY* has been knocked out. The permeases from the cryptic operons *chb* and *asc* from *E. coli* have been used to confer *E. coli* a cellobiose-utilization phenotype (Vinuselvi & Lee, 2011) and could also be analyzed as potential cellodextrin permeases.

It would also be interesting to see if growth is affected by changing the subcellular localization of the enzyme CHU2268. Naturally occurring signal peptides from *E. coli*

(e.g., those of PelB, OmpA, MalE or PhoA) could be used to target the enzyme to the periplasmic space (Baneyx & Mujacic, 2004). Other β-glucosidases have been already secreted (Gupta et al., 2013) or displayed on the cell surface of *E. coli* (Tanaka et al., 2011); however, in the latter case no activity was observed when they tried to use CHU2268.

CHU2268HT was successfully purified from the cell lysate using a Ni Sepharose column. An imidazole step gradient of 0.1, 0.2, 0.3 and 1 M was used for the elution. Most of the enzyme was obtained in the 0.2 M fraction. Total inactivation was observed just a few hours after the purification process. A buffer exchange step was required to retain the enzyme activity, achieving a purification yield of 38%. The high concentration of imidazole in the elution buffer might be the reason for the inactivation. Imidazole has been previously reported as a β-glucosidase inhibitor interfering in the general acid-base catalysis employed by this enzyme and its presence in the buffer could be the reason for this inactivation (Field et al., 1991). The use of buffer with a low pH for the elution has been reported and these conditions might be an attractive alternative to improve the purification process (Terpe, 2003). However, this conditions might exert a negative effect on the protein activity. Further characterization of the enzyme, including the construction of a pH activity profile, will be required in order to check whether the enzyme activity is affected under those conditions.

The endoglucanase CenA from *C. fimi* was expressed using *E. coli* JM109 and the protein location analyzed. Endoglucanase activity was observed in all fractions of the cell, indicating that the native signal peptide of CenA is recognized by *E. coli* and the enzyme translocated into the periplasm. However, the mechanism utilized by *E. coli* to secrete the enzyme into the medium is not entirely understood and non-specific leakage from the periplasm when the enzyme is expressed at high levels has been proposed (Guo et al., 1988). In this study, accumulation of CenA in the periplasm was also observed and this could support this hypothesis. Clear degradation of CMC in the extracellular space was achieved when live cells were used, suggesting that CenA might be used as the first step of a cellulose degradation system.

In order to design a cellulose degradation system, constructs containing a β-glucosidase and CenA were assembled. *E. coli* JM109 co-expressing CHU2268HT and CenA was the only strain able to grow on pure cellulose paper as the sole carbon source. Moreover, clear deconstruction of the substrate was also observed. These results indicate that CenA is secreted at a sufficient level in *E. coli* and acts as a source of cello-oligosaccharides for CHU2268HT that subsequently release glucose allowing cell growth. Stalbrand et al. (1998) reported 25% solubilization of the highly crystalline bacterial microcrystalline cellulose (BMCC) and 61% solubilization of the low crystalline phosphoric acid-swollen cellulose (PASC) after incubation with CenA. In both cases the soluble products were mainly cellobiose and only traces of cellotriose were observed (Stalbrand et al., 1998). Cellobiose is the expected product for cellobiohydrolases but not necessarily for endoglucanases (Stalbrand et al., 1998). The capacity of CenA to generate oligosaccharides as short as cellobiose is probably the reason why this cellulose degradation system allows *E. coli* to grown on pure cellulose not requiring enzymes with cellobiohydrolases activity. However, the levels of solubilization and final products generated by CenA after the hydrolysis of cellulose paper have yet to be studied.

Cellulose is one of the main polysaccharides present in green macroalgal biomass (Lahaye & Robic, 2007), thus a potential source of sugars to be used by this system. However, it should be noted that different type of cellulosic biomass require different enzymes for their degradation and other/additional enzymes might be required. Considering that, in order to design a microorganism able to utilize green algal biomass the cellulose degrading system designed in this study has to be tested using cellulose obtained from the algae. Enzymatic hydrolysis of *U. lactuca* has been already reported using a commercial cellulase cocktail (GC220); however, it was not optimized requiring a relatively high load of enzymes to ensure polysaccharide degradation (van der Wal et al., 2013).

In summary, it was possible to generate a cellulose degradation system for *E coli* using only two enzymes; CHU2268HT from *C. hutchinsonii,* an enzyme expressed in the cytoplasm showing both β-glucosidase and cellodextrinase activity, and CenA from *C. fimi,* a secreted endoglucanase. The expression of CHU2268HT allows cells to utilise

cellodextrins and cellobiose, expanding its substrate specificity in comparison to a normal β-glucosidase. This system could be potentially used for the design of a consolidated bioprocess for the production of biofuels or other bio-products from green macroalgal biomass or other cellulosic materials.

4 Enzymatic ulvan depolymerisation

4.1 Introduction

Species of the genus *Ulva* are green macroalgae found worldwide (Lahaye & Robic, 2007). *Ulva* are grown or collected for food consumption (Anupama & Ravindra, 2000) and responsible for green tides (Smetacek & Zingone, 2013), an increasing economic and environmental problem. Currently the generated biomass is of little value; however, up to 54% of algae's dry weight corresponds to cell wall polysaccharides (Lahaye & Robic, 2007), representing an interesting source of renewable biomass to be explored.

The main carbohydrate of *Ulva* is ulvan, a water-soluble cell wall polysaccharide made up mainly of rhamnose, glucuronic acid, iduronic acid, xylose and sulphate (Lahaye & Robic, 2007). The predominant repeating disaccharides in ulvan are Rha3S-GlcA (ulvanobiouronic acid A), Rha3S-IduA (ulvanobiouronic acid B) and Rha3S-Xyl (Lahaye & Robic, 2007).

Compared to the enzymatic depolymerisation of carbohydrates from brown and red algae, the saccharification of ulvan has been investigated less extensively. An extracellular ulvan lyase from *Nonlabens ulvanivorans* was characterized and its catalytic domain heterologously expressed in *E. coli* (Collen et al., 2011). This enzyme showed endolytic activity cleaving the glycosidic bond between the sulfated rhamnose and a glucuronic or iduronic acid by β-elimination mechanism generating an unsaturated uronic acid at the non-reducing end of the chain (Collen et al., 2011). Additionally, an unsaturated glucuronyl hydrolase (UGL) from *N. ulvanivorans* was shown to cleave specifically the unsaturated non-reducing end produced by the ulvan lyase (Collen et al., 2014). Considering the complexity of ulvan other enzymes required for its complete saccharification such as other lyases, rhamnosidases, xylanases, and sulfatases remain to be identified.

Many seaweed-associated bacteria are able to enzymatically decompose algal cell walls (Hollants et al., 2013), thus are an interesting source of ulvanolytic enzymes. Recently, the complete genome of *Formosa agariphila*, a Bacteroidetes isolated from

green alga *Acrosiphonia sonderi*, was reported by Mann et al. (2013), revealing a broad potential for algal polysaccharides degradation. In particular, a polysaccharide utilization locus (PUL) described in the study seems to have all the enzymes required for the complete saccharification of ulvan.

In order to better understand the enzymatic saccharification of ulvan, *F. agariphila*'s capability to utilize this polysaccharide was investigated. To do that, the potential PUL for ulvan depolymerization was manually curated and analysed in-silico by comparative genomics. Additionally, the capability of *F. agariphila* to grow using the green alga *U. lactuca* as the sole carbon source was assayed and the induction of enzymes with ulvanase activity when this alga is present in the culture medium studied. Finally, the genes of the PUL identified as carbohydrate-active enzymes (CAZymes) were cloned into *E. coli* and heterologously expressed in order to validate their predicted activities.

4.2 *F. agariphila* for ulvan depolymerisation

4.2.1 *F. agariphila*'s PUL for ulvan utilization

The glycan degradation and import machinery in Bacteroidetes is usually arranged in clusters of co-regulated genes known as polysaccharide utilization loci (PUL) (Terrapon et al., 2015). A total of 13 PULs were found in *F. agariphila*'s genome (Mann et al., 2013). These were identified using the presence of at least one pair of TonB-dependent receptor (TBDR)/SusD-like protein associated with polysaccharide utilization genes as the delimiting criterion (Sonnenburg et al., 2010). *F. agariphila*'s PULs were named by Mann et al. (2013) from A to M and the most probable substrates for each of them deduced based on in-depth CAZyme analysis. PUL H, a 77309 bp region composed of 39 predicted genes was presumed to degrade sulfated (decorated) rhamnogalacturonans (Mann et al., 2013). However, the presence of a putative ulvan lyase, not reported by Mann et al. (2013), might suggest that the substrate for this PUL is actually the green macroalgal polysaccharide ulvan.

In order to elucidate whether ulvan is the substrate for this PUL an exhaustive analysis of the genes that comprise the PUL was performed. Taking into account the fact that it is difficult to determine the boundaries of a PUL without experimental data, the flanking regions of the PUL were analysed to ensure that no gene that might be associated with the utilization of the polysaccharide is excluded from the analysis. Two additional genes immediately downstream of the sequence were identified as potential transcription factors (locus tags BN863_22290 and BN863_22300), thus added to the analysis. Considering this, the PUL was redefined going from BN863_21900 to BN863_22300.

The 41 genes composing the PUL were subjected to manual curation to discard possible errors arising from the automated prediction and annotation of the genes. A total of 9 genes were shown to be incomplete, missing important nucleotides at the 5'-end of the sequences. The corrected sequences are given as protein sequences in the supplementary results 8.2.1. Once all the genes were defined, the associated proteins were named as FA and the first four numbers of the locus tag for simplification (e.g. FA2190 for the protein codified by BN864_21900).

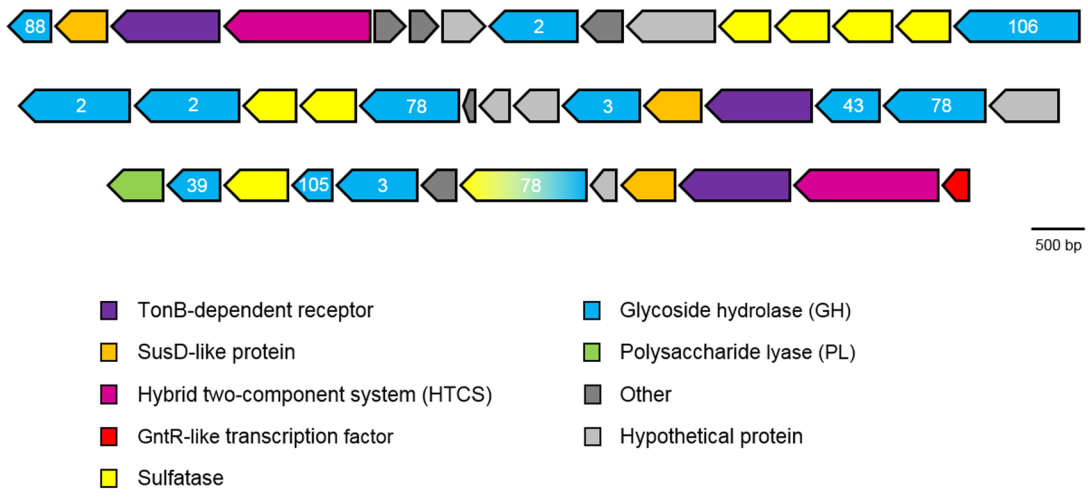

Figure 4-1 Tentative polysaccharide utilization loci for ulvan utilization of *F. agariphila*. Gene prediction and annotation were manually curated. Each gene is represented by an arrow that reflects the relative size and orientation of the coding region. Putative functions are color-coded. Glycoside hydrolase families are indicated by numbers. Intergenic regions are not considered in the diagram. A 500 bp scale bar is supplied.

Protein functions were predicted by similarity using BlastP (Altschul et al., 1990) and HMMER hmmscan (Eddy, 2011) against the Pfam (Finn et al., 2016) and TGRFAM (Haft et al., 2003) databases. The results showed that the PUL includes 13 CAZymes, 7 sulfatases, 3 TBDR, 3 SusD-like proteins, 2 hybrid two-component systems (HTCSs), and a GntR transcription factor, among other enzymes. A list containing all the proteins of the PUL and their respective predicted functions is supplied in Table 8-3. Additionally, a schematic representation of the PUL including size, orientation and functionality of the different genes is shown in Figure 4-1.

Protein subcellular localization was predicted in-silico based on the decision tree developed by Romine (2011). Less common signal peptides associated with lipoproteins and twin-arginine translocation (Tat) substrates were checked first, followed by the prediction of transmembrane spans and Sec signal peptides. Domains that consistently occurred only in proteins found in specific subcellular localizations were also analyzed; however, none of these domains were detected in the proteins of the PUL. Predicted localizations of the different proteins are shown in Table 4-1. For more details of the tools utilized and the results obtained from each of them see supplementary Table 8-3.

Table 4-1 Enzymes from *F. agariphila*'s ulvan utilization PUL

Protein[a]	Locus tag	Function[b]	Localization[c]
FA2190	BN864_21900	GH88	OM
FA2191	BN864_21910	SusD-like protein	OM
FA2192	BN864_21920	TonB-dependent receptor	OM
FA2193	BN864_21930	HTCS	IM
FA2194	BN864_21940	kduI	CP
FA2195	BN864_21950	kduD	CP
FA2196	BN864_21960	Conserved hypothetical protein	OM
FA2197	BN864_21970	GH2	CP
FA2198	BN864_21980	6-phosphogluconolactonase	CP
FA2199	BN864_21990	Conserved hypothetical protein	OM
FA2200	BN864_22000	Sulfatase	OM
FA2201	BN864_22010	Sulfatase	OM
FA2202	BN864_22020	Sulfatase	PP
FA2203	BN864_22030	Sulfatase	PP
FA2204	BN864_22040	GH106	PP

Protein[a]	Locus tag	Function[b]	Localization[c]
FA2205	BN864_22050	GH2	CP
FA2206	BN864_22060	GH2	CP
FA2207	BN864_22070	Sulfatase	OM
FA2208	BN864_22080	Sulfatase	OM
FA2209	BN864_22090	GH78	OM
FA2210	BN864_22100	L-rhamnose mutarotase	CP
FA2211	BN864_22110	Conserved hypothetical protein	PP
FA2212	BN864_22120	Conserved hypothetical protein	PP
FA2213	BN864_22130	GH3	PP
FA2214	BN864_22140	SusD-like protein	OM
FA2215	BN864_22150	TonB-dependent receptor	OM
FA2216	BN864_22160	GH43	OM
FA2217	BN864_22170	GH78	OM
FA2218	BN864_22180	Conserved hypothetical protein	PP
FA2219	BN864_22190	Ulvan lyase	EC
FA2220	BN864_22200	GH39	PP
FA2221	BN864_22210	Sulfatase	OM
FA2222	BN864_22220	GH105	PP
FA2223	BN864_22230	GH3	OM
FA2224	BN864_22240	Oxidoreductase	CP
FA2225	BN864_22250	GH78/sulfatase	PP
FA2226	BN864_22260	Hypothetical protein	PP
FA2227	BN864_22270	SusD-like protein	OM
FA2228	BN864_22280	TonB-dependent receptor	OM
FA2229	BN864_22290	HTCS	IM
FA2230	BN864_22300	GntR-like transcriptional regulator	CP

[a] Protein names were designated in this study. [b] Protein functions were predicted using HMMER hmmscan (Eddy, 2011) against the Pfam database (Finn et al., 2016). [c] Protein localization were predicted according to Romine (2011). CP: cytoplasm; IM: inner membrane; PP: periplasm; OM: outer membrane; EC: extracellular.

In order to predict the different regions within the CAZymes from the PUL, including catalytic domains and carbohydrate-binding modules (CBM), the annotation tool dbCAN HMMs v5.0 (Yin et al., 2012) based on the CAZy database (Lombard et al., 2014) from 07/15/2016 was used. Additionally, HMMER hmmscan (Eddy, 2011) against the Pfam (Finn et al., 2016) and TGRFAM (Haft et al., 2003) databases was used. Domains of unknown function were not considered. The obtained results

indicated the presence of 13 glycoside hydrolases and 1 polysaccharide lyase. The domain organization of the different enzymes is shown in Figure 4-2; predicted signal peptides are also included in the figure.

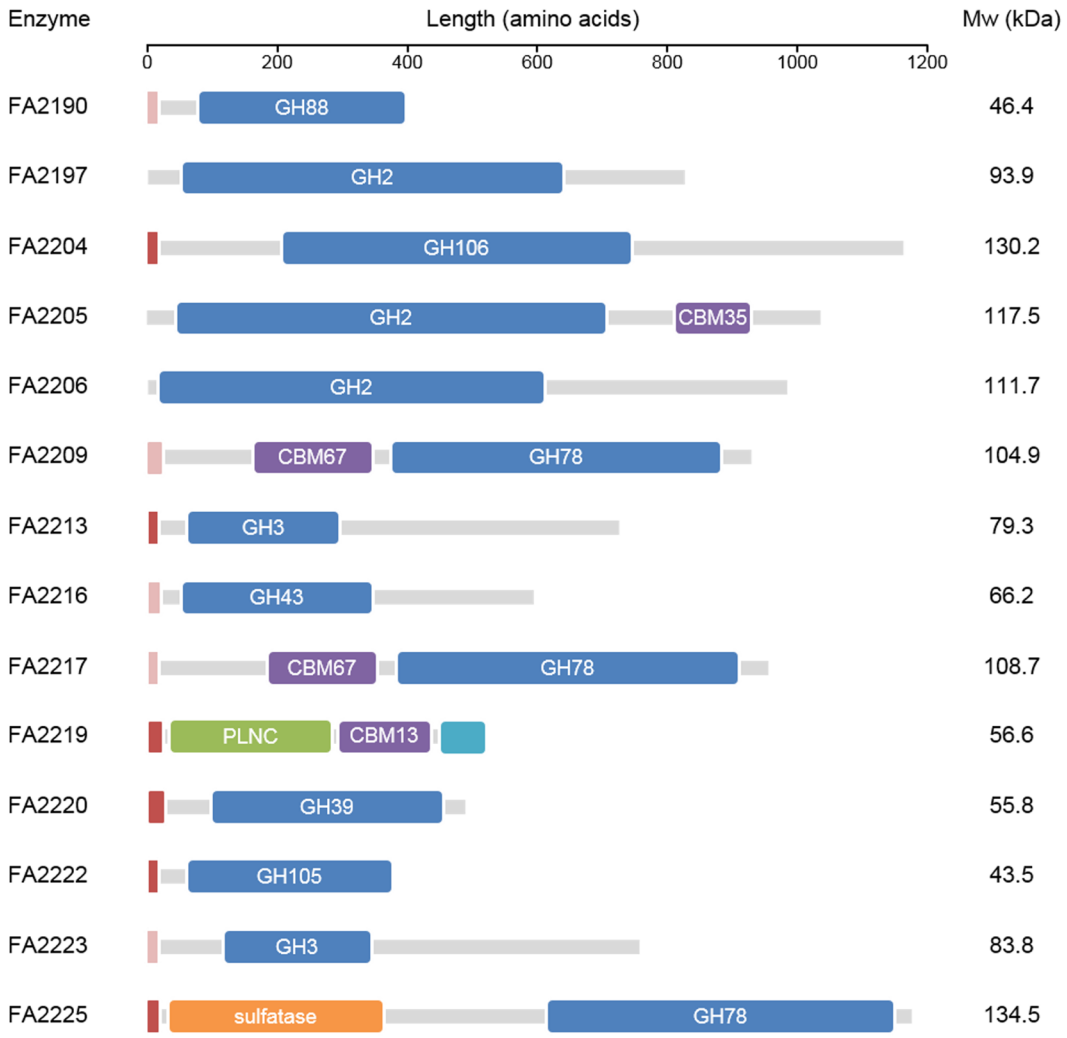

Figure 4-2 Domain organization of the CAZymes from *F. agariphila*'s ulvan utilization PUL. Cazy domains were predicted by dbCAN (Yin et al., 2012) based on the CAZy database (Lombard et al., 2014) from 07/15/2016. Additional domains were predicted using HMMER hmmscan (Eddy, 2011) against the Pfam (Finn et al., 2016) and TGRFAM (Haft et al., 2003) databases. Signal peptides (red) and lipoprotein signal peptide (light red) were predicted using SignalP v4.1 (Petersen et al., 2011) and LipoP v1.0 (Juncker et al., 2003), respectively. A Por_Secre_tail domain is indicated in turquoise. Glycoside hydrolase (GH) and carbohydrate-binding module (CBM) family numbers are indicated. PLNC: polysaccharide lyase "non classified".

The glycoside hydrolases belong to 8 different families: 3 GH2, 2 GH3, 1 GH39, 1 GH43, 3 GH78, 1 GH88, 1 GH105 and 1 GH106, and 3 of them also showed CBMs

from families 35 (FA2205) and 67 (FA2209 and FA2217). Surprisingly, FA2225 not only showed a GH78 domain but also a sulfatase domain, this architecture is very rare and only 4 other proteins containing it have been identified in the Pfam database (Finn et al., 2016), none of which have been characterized. No polysaccharide lyase (PL) domain was identified for the predicted ulvan lyase FA2219; however, it appears in the CAZy database as a polysaccharide lyase "non classified" (PLNC). Two additional domains were found in FA2219, a CBM of the family 13 and a C-terminal Por_Secre_tail domain. Por_Secre_tail are associated with the recently discovered Por Secretion System (PorSS) or following the naming convention, Type IX Secretion System (T9SS) (Sato et al., 2010). FA2219's PL domain position was determined based on homology with the ulvan lyase from *N. ulvanivorans* (accession number G8G2V6), this being the only characterized ulvan lyase belonging to the same group of unclassified PL.

In order to predict the activity of the CAZymes from the PUL the different families they belong to were analyzed. In some of the families more than one activity was found. In those cases a phylogenetic study using only the catalytic domains of characterized enzymes from bacteria was performed. The families GH78 and GH106 contains only α-L-rhamnosidases (EC 3.2.1.40), suggesting that FA2204, FA2209, FA2217 and FA2225 are also rhamnosidases. Additionally, FA2209 and FA2217 contained CMBs from family 68 which have only been identified associated with rhamnosidases. GH88 only contains characterized UGLs (EC 3.2.1.-) suggesting that FA2190 is also a UGL. GH105 is composed by three enzymes: 1 UGL and 2 unsaturated rhamnogalacturonyl hydrolases (EC 3.2.1.172); however, only the UGL from *N. ulvanivorans* (accession number L7P9J4) showed a high level of sequence identity with FA2222 (76% amino acid identity), indicating that this enzyme might also be a UGL. Finally, as mentioned before FA2219 is categorized as a PLNC and shows a high level of sequence identity with an ulvan lyase from *N. ulvanivorans* (76% amino acid identity). Levels of amino acid identity were obtained with BlastP. All the other families contained several experimentally characterized enzymes with different activities, thus were analyzed using a phylogenetic approach.

Bacterial GH2 includes β-galactosidases (EC 3.2.1.23), β-glucuronidases (EC 3.2.1.31), endo-β-mannosidases (EC 3.2.1.152), and exo-β-glucosaminidases (EC 3.2.1.165). Up to 5 sequences of experimentally characterized enzymes of each activity were selected and aligned with the GH2 from the PUL using MUSCLE (Edgar, 2004) and a maximum likelihood phylogenetic tree was constructed using MEGA7 (Kumar et al., 2016) (Figure 4-3). Clear clusters containing enzymes of the same activity were observed. All the enzymes from the PUL were localized next to the cluster containing the β-glucuronidases, suggesting that FA2197, FA2205 and FA2206 have the same activity. Considering the structure of ulvan, where glucuronic acid is one of the main building blocks, these results are consistent.

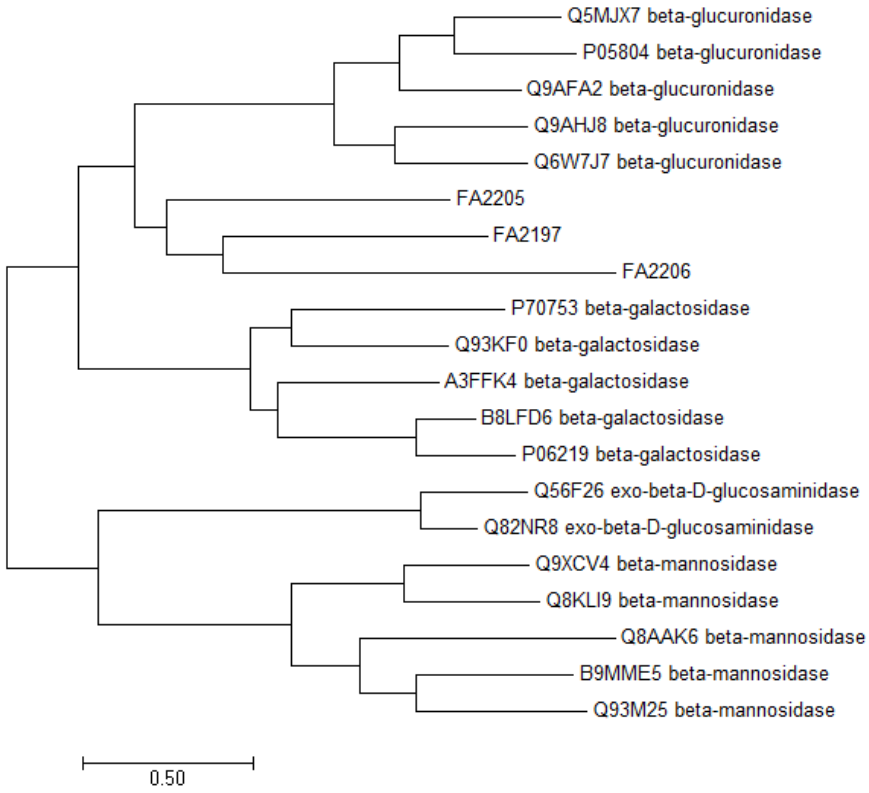

Figure 4-3 Phylogenetic analysis of the catalytic domains of the GH2 enzymes from *F. agariphila*'s ulvan utilization PUL. Protein sequences were obtained from UniProt (UniProt, 2015) and catalytic domains identified using dbCAN (Yin et al., 2012). Catalytic domains were aligned using MUSCLE (Edgar, 2004) and a maximum likelihood phylogenetic tree was constructed using MEGA7 (Kumar et al., 2016). UniProt accession number and activity of the enzymes are indicated. The branch lengths indicate the divergence among the amino acid sequences. Bar, 0.5 substitutions per site.

Bacterial GH3 includes β-glucosidases (EC 3.2.1.21), cellodextrinases (EC 3.2.1.74), β-xylosidases (EC 3.2.1.37), and β-acetylhexosaminidase (EC 3.2.1.52). A phylogenetic tree was built using the same criteria as the one for the GH2 family. CHU2268 from *C. hutchinsonii* was also added to this analysis. This time only two clear clusters were observed, one containing the β-acetylhexosaminidase and the other all the other enzymes analyzed (Figure 4-4). Bifunctional proteins are common in this family (β-glucosidases/β-xylosidases and β-glucosidases/cellodextrinases) and might indicate no clear structural differences between the catalytic domains of its members; however, considering that ulvan does not contain glucose but xylose it is highly likely that FA2213 and FA2223 correspond to β-xylosidases.

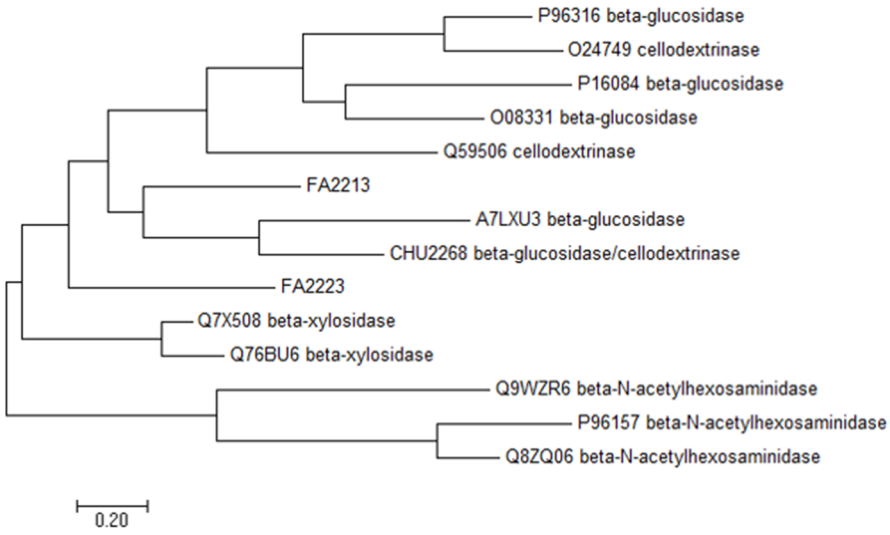

Figure 4-4 Phylogenetic analysis of the catalytic domains of the GH3 enzymes from *F. agariphila*'s ulvan utilization PUL. Protein sequences were obtained from UniProt (UniProt, 2015) and catalytic domains identified using dbCAN (Yin et al., 2012). Catalytic domains were aligned using MUSCLE (Edgar, 2004) and a maximum likelihood phylogenetic tree was constructed using MEGA7 (Kumar et al., 2016). UniProt accession number and activity of the enzymes are indicated. The branch lengths indicate the divergence among the amino acid sequences. Bar, 0.2 substitutions per site.

GH39 is a small family and not many enzymes have been functionally characterized. GH39 contains β-xylosidases (EC 3.2.1.37) and α-iduronidases (EC 3.2.1.76). No α-iduronidases have been reported in bacteria; therefore, characterized α-iduronidases from Eukaryotes were included in the analysis. FA2220 was found to be phylogenetically remote to all the other enzymes of the tree (Figure 4-5) and did not share significant levels of homology with any characterized enzyme after a BlastP

search. FA2220 appears in the CAZy database as a “non classified” GH. No conclusive results can be inferred from this results; however, considering the presence of iduronic acid in ulvan this enzyme might be a novel α-iduronidase from bacteria.

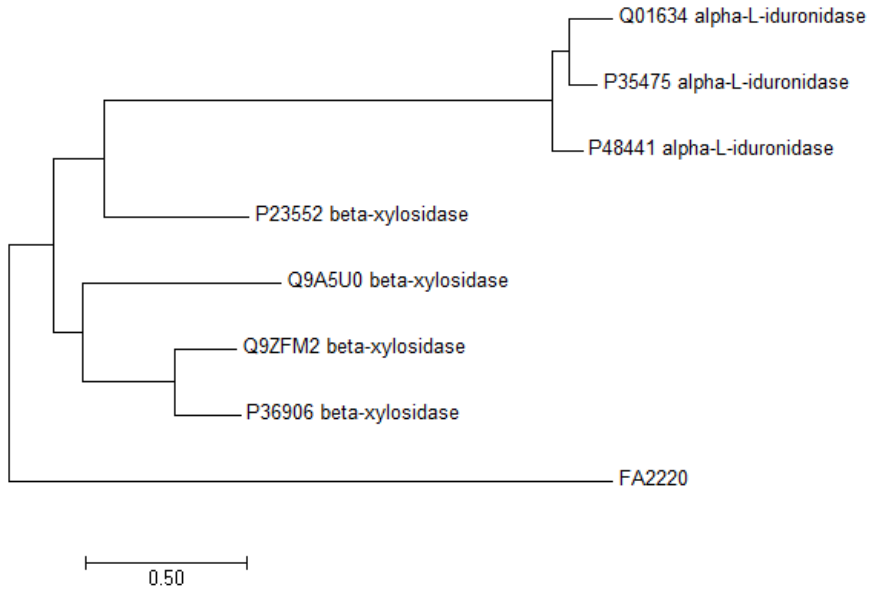

Figure 4-5 Phylogenetic analysis of the catalytic domains of the GH39 from *F. agariphila*'s ulvan utilization PUL. Protein sequences were obtained from UniProt (UniProt, 2015) and catalytic domains identified using dbCAN (Yin et al., 2012). Catalytic domains were aligned using MUSCLE (Edgar, 2004) and a maximum likelihood phylogenetic tree was constructed using MEGA7 (Kumar et al., 2016). Uniprot accession number and activity of the enzymes are indicated. The branch lengths indicate the divergence among the amino acid sequences. Bar, 0.5 substitutions per site.

Bacterial GH43 includes arabinanases (EC 3.2.1.99), β-xylosidases (EC 3.2.1.37), and α-arabinofuranosidases (EC 3.2.1.55). The GH43 family has been recently divided into sub-families for more detailed characterization (Mewis et al., 2016). FA2216 belongs to the sub-family 10 which contains only β-xylosidases and α-arabinofuranosidases The phylogenetic tree was built as for the other families. Clear clusters containing enzymes of the same activity were observed (Figure 4-6). FA2216 was found in the cluster containing the β-xylosidases, suggesting that the enzyme possesses the same activity.

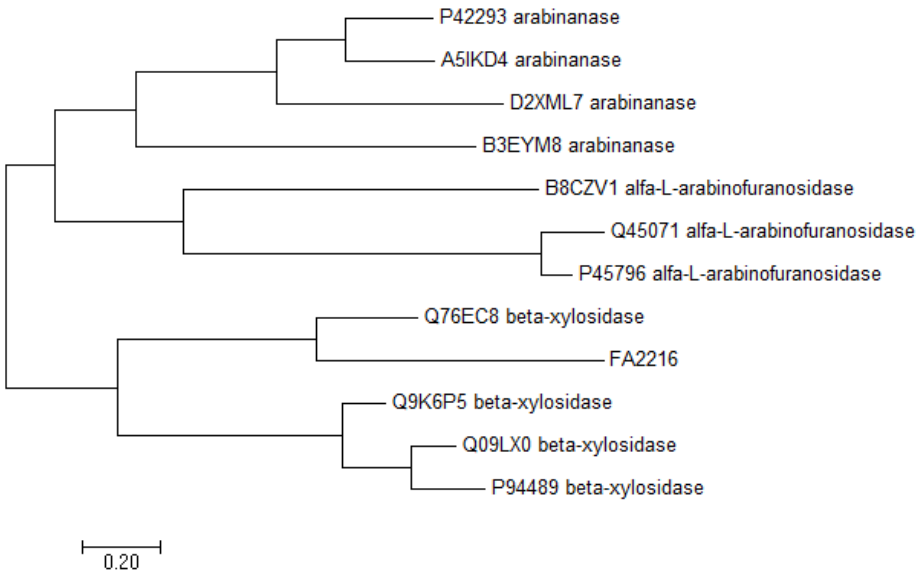

Figure 4-6 Phylogenetic analysis of the catalytic domains of the GH43 enzymes from *F. agariphila*'s ulvan utilization PUL. Protein sequences were obtained from UniProt (UniProt, 2015) and catalytic domains identified using dbCAN (Yin et al., 2012). Catalytic domains were aligned using MUSCLE (Edgar, 2004) and a maximum likelihood phylogenetic tree was constructed using MEGA7 (Kumar et al., 2016). UniProt accession number and activity of the enzymes are indicated. The branch lengths indicate the divergence among the amino acid sequences. Bar, 0.2 substitutions per site.

Table 4-2 shows a summary of the expected activity of the CAZymes of *F. agariphila*'s PUL for ulvan utilization.

Table 4-2 Expected activity of the CAZymes

Enzyme	Expected activity
FA2190	unsaturated β-glucuronyl hydrolase (EC 3.2.1.-)
FA2197	β-glucuronidase (EC 3.2.1.31)
FA2204	α-L-rhamnosidase (EC 3.2.1.40)
FA2205	β-glucuronidase (EC 3.2.1.31)
FA2206	β-glucuronidase (EC 3.2.1.31)
FA2209	α-L-rhamnosidase (EC 3.2.1.40)
FA2213	β-xylosidase (EC 3.2.1.37)
FA2216	β-xylosidase (EC 3.2.1.37)
FA2217	α-L-rhamnosidase (EC 3.2.1.40)
FA2219	ulvan lyase (EC 4.2.2.-)
FA2220	α-iduronidase (EC 3.2.1.76)
FA2222	unsaturated β-glucuronyl hydrolase (EC 3.2.1.-)
FA2223	β-xylosidase (EC 3.2.1.37)
FA2225	α-L-rhamnosidase (EC 3.2.1.40)

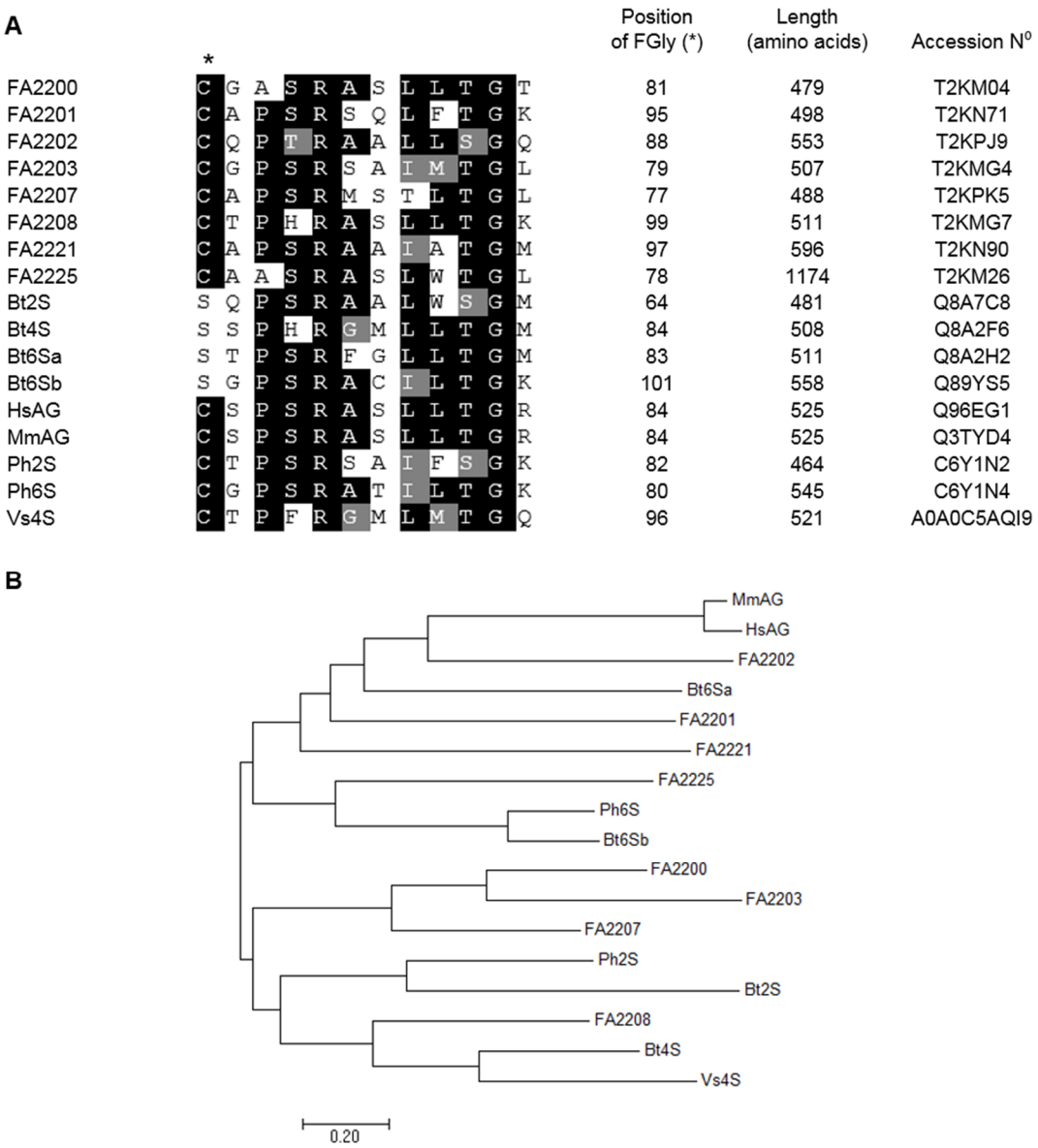

Figure 4-7 Analysis of sulfatases from *F. agariphila*'s ulvan utilization PUL. (A) Alignment of signature sequences of sulfatases. The signature sequence is critical for directing the post-translational modification of the initial C/S into the catalytic residue FGly. Protein sequences were obtained from UniProt (UniProt, 2015) and signature sequences aligned using MUSCLE (Edgar, 2004). Black and grey boxes indicate identical and similar residues, respectively. FGly position, protein length and UniProt accession number are indicated. Enzymes are named after the organisms that express them and substrate specificity. Bt: *Bacteroides thetaiotaomicron*, Hs: *Homo sapiens*, Mm: *Mus musculus*, Ph: *Pedobacter heparinus*, and Vs: *Vibrio sp.*. 2S: 2-O-sulfatase, 4S: 4-O-endosulfatase, 6S: 6-O-sulfatase, and AG: arylsulfatase G. (B) Phylogenetic analysis of sulfatases catalytic domain by maximum likelihood method. Catalytic domains were identified using HMMER hmmscan (Eddy, 2011) against the Pfam database (Finn et al., 2016). Catalytic domains were aligned using MUSCLE and a maximum likelihood phylogenetic tree was constructed using MEGA7 (Kumar et al., 2016). The branch lengths indicate the divergence among the amino acid sequences. Bar, 0.2 substitutions per site.

Considering that ulvan is a highly sulfated polysaccharide, containing rhamnose 3-sulfate and xylose 2-sulfate among its building blocks, sulfatases are likely to have an important role in ulvan saccharification. A total of 8 sulfatases were found in the PUL confirming the role of this class of enzyme in the system. Glycosaminoglycan (GAG) degradation systems including sulfatases have been described by Hanson et al. (2004). The degradation is initiated by specific lyases which cleave the glycans into smaller oligosaccharides that are then desulfated by sulfatases before they are cleaved into monosaccharides by GHs. Several bacterial sulfatases involved in these kinds of systems have been characterized by their residue-specificity, including 6-O-sulfatases, endo-4-O-sulfatases, and 2-O-sulfatases (Hanson et al., 2004). A 3-O-sulfatase from *Pedobacter heparinum* involved in heparin degradation has been described by Bruce et al. (1985). However, the gene has not been cloned and no nucleotide or amino acid sequence associated with the enzyme is available. Considering the structural similarities between GAGs and ulvan the characterized glycosaminoglycan-degrading sulfatases will be used as a starting point for the analysis of the sulfatases from the PUL.

Sulfatases contain a very conserved N-terminal signature sequence comprising 12 amino acids (C/S-X-P-S/X-R-X-X-X-L/X-T/X-G/X-R/X), where the initial cysteine or serine residue is post-translationally modified into the catalytically active residue formylglycine (FGly) (Hanson et al., 2004). In order to study the signature sequences of the sulfatases of the PUL, their amino acids sequences were aligned with those of the characterized glycosaminoglycan-degrading sulfatases using MUSCLE. Additionally, two eukaryotic arylsulfatases G that showed 3-O-sulfatase activity were included in the analysis (Kowalewski et al., 2012). The results of the alignment are shown in Figure 4-7 (A). All the sulfatases from the PUL showed a cysteine as the first amino acids of the signature sequence, thus are categorized as Cys-type sulfatases. Moreover, a maximum likelihood phylogenetic tree of the catalytic domains of the sulfatases was constructed using MEGA7 Figure 4-7 (B). A small cluster containing enzymes having the same substrate specificity was found; however, the sulfatases from the PUL were found distributed all over the tree. These results showed an amazing diversity at the sequence level, but it is difficult to predict the substrate specificity of

these enzymes as no characterized sulfatase acting against the sulfated residues found in ulvan has been reported.

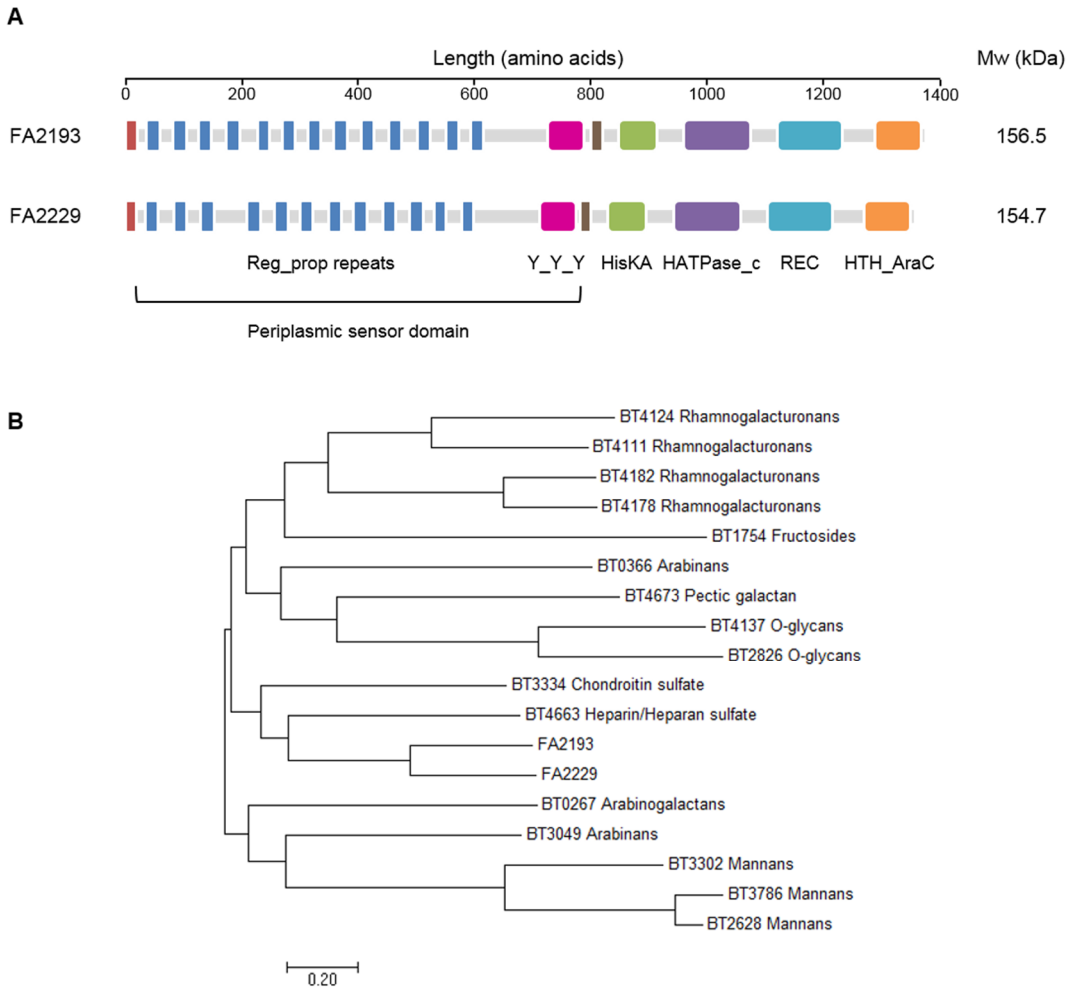

Figure 4-8 Analysis of hybrid two-component system proteins from *F. agariphila*'s ulvan utilization PUL. (A) Domain organization of the hybrid two-component system (HTCS) proteins. Reg_prop repeats, Y_Y_Y, phosphoacceptor (HisKA), ATPase (HATPase_c), receiver (REC), and DNA binding (HTH_AraC) domains were identified using HMMER hmmscan (Eddy, 2011) against the Pfam database (Finn et al., 2016). Signal peptides in red and transmembrane segments in brown were identified using Phobius (Kall et al., 2004). (B) Phylogenetic analysis of HTCS by maximum likelihood method. Protein sequences were obtained from UniProt (UniProt, 2015), aligned using MUSCLE (Edgar, 2004) and a phylogenetic tree constructed using MEGA7 (Kumar et al., 2016). Locus tag and predicted polysaccharide utilization pathway that is controlled by the HTCS of *Bacteroides thetaiotaomicron* are indicated. The branch lengths indicate the divergence among the amino acid sequences. Bar, 0.2 substitutions per site.

HTCSs are a key part of Bacteroidetes' ability to sense and degrade complex polysaccharides. Most glycan PULs are activated through the recognition of specific

signature oligosaccharides by their associated HTCS (Martens et al., 2011). HTCSs are chimeric proteins composed of an N-terminal periplasmic sensor domain comprising 14 Reg_prop repeats (Pfam 07495) and a Y_Y_Y domain (Pfam 07495), and the cytoplasmic components of a "classical" two-component systems: phosphoacceptor (Pfam HisKA), ATPase (Pfam HATPase_c), receiver (Pfam REC), and DNA binding (Pfam HTH_AraC) domains (Lowe et al., 2012). Figure 4-8 (A) shows the domain organization of the two HTCSs found in the ulvan utilization PUL. Domains were identified using HMMER hmmscan (Eddy, 2011) against the Pfam database (Finn et al., 2016). FA2193 and FA2229 showed all the components of a HTCS; however, only 13 repeats of the 14 Reg_prop were identified for FA2193 and 12 for FA2229. The observed differences might indicate the recognition of different signature oligosaccharides. Additionally, a search using the CD-search tool showed that both proteins contain sensor domains from the family COG3292, known to bind oligosaccharides (Ravcheev et al., 2013).

An extensive study of the HTCSs from *Bacteroides thetaiotaomicron* using comparative genetics was reported by Ravcheev et al. (2013). A total of 16 HTCS were associated with the utilization of different polysaccharides. The two HTCS from the PUL were aligned with the ones from *B. thetaiotaomicron* using MUSCLE, and a maximum likelihood phylogenetic tree constructed using MEGA7. FA2192 and FA2229 were found close together sharing a cluster with *B. thetaiotaomicron*'s HTCS associated with the utilization of chondroitin sulfate and heparin/heparan sulfate. This result might suggest that the sulfate groups are of great importance for the recognition of the signature oligosaccharides. Reinforcing this idea, it has been shown that BT4663 bound specifically to unsaturated heparin and HS-derived disaccharides with a sulfated glucosamine component (Lowe et al., 2012).

In addition, FA2230 was identified as a transcription regulator (TR) from the GntR family. FA2230 showed a high level of sequence identity with an annotated transcriptional regulator of rhamnose utilization from *Jejuia pallidilutea* (76% amino acid identity); however, this has not been characterized. No published GntR-like TR controlling the rhamnose utilization pathway was found. A conserved domain search using the CD-search tool showed a ligand-binding domain of the negative

transcriptional regulator FucR, a member of the GntR family. FucR from *B. thetaiotaomicron* regulates the metabolic pathways of L-fucose, a methyl pentose closely related to L-rhamnose (Hooper et al., 1999). Rhamnose utilization in *B. thetaiotaomicron* is positively regulated by RhaR, a TR from the AraC family (Patel et al., 2008); however, no homologues to RhaR were identified in *F. agariphila*. Similar results were obtained when the rhamnose regulators RhaR and RhaS from *E. coli*, that are also part of the AraC family but not homologous to RhaR from *B. thetaiotaomicron*, were analyzed. These results suggest that FA2230 might be the first member of a novel subfamily of the GntR family of TRs.

The activity of 5 other enzymes of the PUL were recognized from their amino acid sequences. FA2210 was identified as an L-rhamnose mutarotase. Enzymes of this class catalyze the α-β conversion of rhamnose allowing the cells to utilize this carbon source more efficiently (Ryu et al., 2005). FA2194 and FA2195 were identified as a 4-deoxy-L-threo-5-hexosulose-uronate ketol-isomerase (KduI) and 2-dehydro-3-deoxy-D-gluconate 5-dehydrogenase (KduD), respectively. This enzyme cluster converts 5-dehydro-4-deoxy-D-glucuronate (DKI) via 3-deoxy-D-glycero-2,5-hexodiulosonate (DKII) to 2-dehydro-3-deoxy-D-gluconate (KDG). FA2198 and FA2224 were identified as a 6-phosphogluconolactonase and an oxidoreductase, respectively; however, their roles in the system are unknown. Additionally, 6 hypothetical proteins were also recognized, among them FA2199 showed 27% amino acid identity with a heparinase II/III-like protein from *Bacteroides dorei*, suggesting that this enzyme might be part of a novel family of ulvan lyases.

Putative pathways for the utilization of the monosaccharides released from ulvan were reconstructed using comparative genomics (Figure 4-9). Pathways obtained from MetaCyc (Caspi et al., 2014) and KEGG (Kanehisa et al., 2016) were used as reference for the reconstruction. Rhamnose is transported across the inner membrane by FA2161, a rhamnose permease (RhaT). Once in the cytoplasm rhamnose is α-β interconverted by the rhamnose mutarotose FA2210 and subsequently converted to lactaldehyde and dihydroxyacetone phosphate (DHAP) by a "classic" rhamnose utilization pathway, composed of a rhamnose isomerase (FA2162), rhamnulose kinase (FA3058), and a rhamnulose aldolase (FA3060). Lactaldehyde is then oxidized to pyruvate by an aldehyde dehydrogenase (FA3393) and an L-lactate dehydrogenase (FA2185), while DHAP is interconverted to D-glyceraldehyde 3-phosphate (GAP) by a triose-phosphate isomerase (FA2719). GAP enters the Entner–Doudoroff (ED) pathway. Xylose is translocated into the cytoplasm by a xylose permease (FA18), where is sequentially converted by a xylose isomerase (FA17) and a xylulose kinase (FA16) to xylulose 5-phosphate, which enters to the non-oxidative part of the pentose phosphate (PP) pathway. The unsaturated uronic acids rearrange spontaneously to DKI, which is converted by the KduI FA2194 to DKII and subsequently transformed to KDG by the KduD FA2195. KdgT and KdgX have been reported as DKI permeases (Rodionov et al., 2004). However, no homologues to these proteins were found in *F. agariphila*, suggesting that other permeases might facilitate the diffusion of DKI in this microorganism. Glucuronic acid is transported into the cytoplasm by the glucuronic acid permease (FA2278). In the cytoplasm glucuronic acid is also converted to KDG by a three-step pathway composed by a glucuronate isomerase (FA9430), an altronate oxidoreductase (FA2082), and an altronate hydrolase (FA2083). KDG is metabolized to pyruvate and GAP by a 2-dehydro-3-deoxygluconate kinase (FA980) and a 2-dehydro-3-deoxyphosphogluconate aldolase (FA982). The pathway for iduronic acid was not reconstructed due to lack of information associated with the utilization of that sugar.

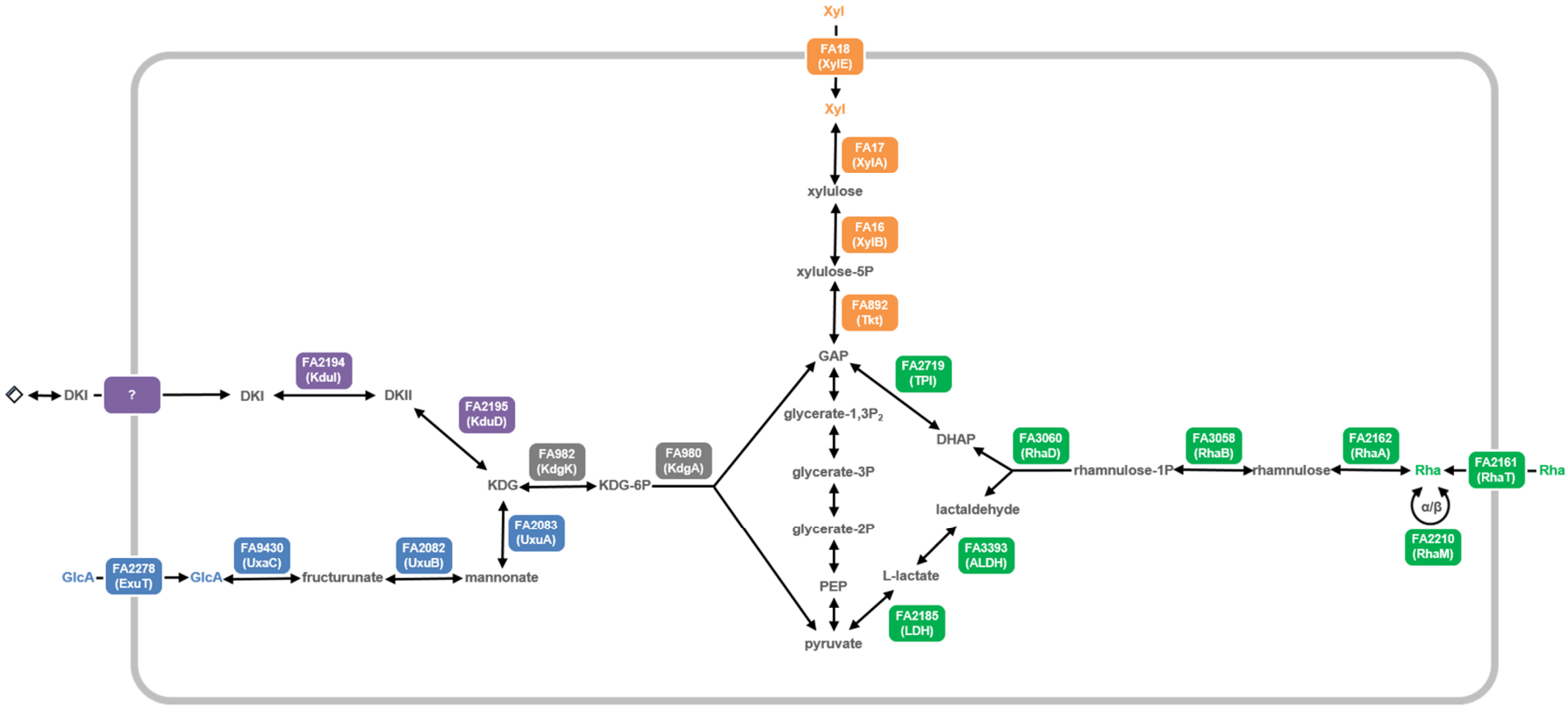

Figure 4-9 Reconstructed pathways for monosaccharide utilization in *F. agariphila*. Reference pathways were obtained from MetaCyc (Caspi et al., 2014) and KEGG (Kanehisa et al., 2016). Identified enzymes are indicated in boxes and the names of their homologues given in brackets. No DKI permease was identified. Rha: rhamnose, Xyl: xylose, GlcA: glucuronic acid, and rhombus: unsaturated uronic acid.

Finally, a tentative model for the ulvan utilization pathway was built based on the predicted activities and localization of the enzymes of the PUL (Figure 4-10). As a first step, the extracellular ulvan lyase degrades ulvan via a β-elimination reaction into unsaturated oligosaccharides. Subsequently, the CBM67 of the GH78 rhamnosidases bound to these oligosaccharides helping to their further depolymerization by the outer membrane enzymes of the system, including glycoside hydrolases and sulfatases. The resulting carbohydrates are delivered by the SusD-like proteins to a TBDR that is energized by the TonB-ExbBD complex to transport them into the periplasm. In the periplasm the HTCSs recognize the unsaturated signature glycans activating the ulvan depolymerization PUL. In addition, the unsaturated uronic acids at the non-reducing end of the oligosaccharides are cleaved by a GH105, and the resulting oligosaccharide depolymerized to monomeric sugars by the periplasmic glycoside hydrolases and sulfatases. The simple sugars and some small oligosaccharides are then transported across the inner membrane to the cytoplasm by sugar permeases from the major facilitator superfamily (MSF). In the cytoplasm a GntR-like transcription regulator is bound by rhamnose, allowing the expression of the pathway to convert rhamnose to L-lactaldehyde. Rhamnose is degraded via L-lactaldehyde to pyruvate, while xylose is converted via the non-oxidative part of the PP pathway to α-D-glucose-6-phosphate, which enters the Embden–Meyerhof–Parnas (EMP) pathway. The unsaturated uronic acids spontaneously rearrange to DKI that is converted to DKII by a KduI. DKII is converted by a KduD to KDG, which enters the ED pathway. The remaining glucuronic acid dimers are cleaved by glucuronidases and the released products converted to KDG by a three-step pathway composed by UxaC, UxuB, and UxuA.

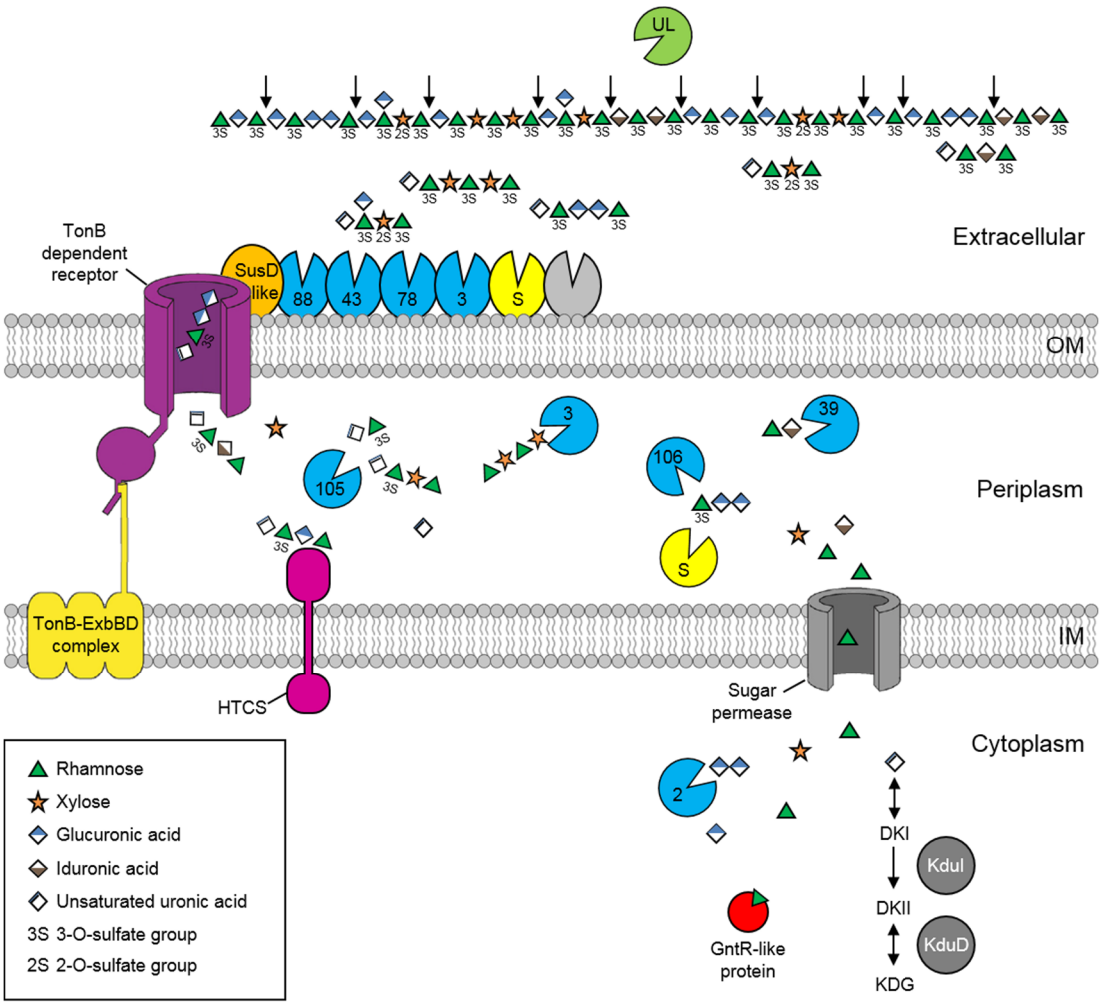

Figure 4-10 Tentative model of the ulvan utilization pathways of *F. agariphila*. Prediction of protein localization was carried out in silico according to Romine (2011). Ulvan is first degraded into unsaturated oligosaccharides by an extracellular ulvan lyase. The generated oligosaccharides are further depolymerized by outer membrane enzymes, including glycoside hydrolases and sulfatases, and bound by a SusD-like protein that facilitates its delivery to a TonB-dependent receptor (TBDR). The oligosaccharides are transported into the periplasm by the TBDR that is energized by a TonB-ExbBD complex. Unsaturated signature glycans are recognized by the periplasmic sensor domain of a hybrid two-component system protein (HTCS) activating the ulvan depolymerization PUL. The oligosaccharides are further processed by periplasmic glycoside hydrolases and sulfatases, releasing monomeric sugars. In particular, the unsaturated uronic acids at the non-reducing end of the oligosaccharides are cleaved by a GH105. Monomers and small oligosaccharides are transported into the cytoplasm by sugar permeases. In the cytoplasm rhamnose binds the GntR-like transcription regulator allowing the rhamnose utilization pathway to be expressed. Rhamnose is likely to be degraded via L-lactaldehyde to pyruvate, while xylose is converted via the nonoxidative part of the PP pathway to α-D-glucose-6-phosphate, which enters the EMP pathway. Glucuronic acid dimers are cleaved by GH2 enzymes. Unsaturated uronic acids spontaneously rearrange to 5-dehydro-4-deoxy-D-glucuronate (DKI) which is converted to 3-deoxy-D-glycero-2,5-hexodiulosonate (DKII) by a KduI. DKII is converted by a KduD to 2-dehydro-3-deoxy-D-gluconate (KDG), which enters the ED pathway. Glucuronic acid is also converted to KDG via a three-step pathway by UxaC, UxuB, and UxuA. Glycoside hydrolase families are indicated by numbers. S: sulfatase; OM: Outer membrane; IM: Inner membrane.

4.2.2 *F. agariphila* growth on *U. lactuca*

A growth assay was performed in order to determine whether *F. agariphila* is able to grow on green macroalgal biomass as the sole carbon source. *F. agariphila* cells were used to inoculate a solution containing 35 g/l of sea salts supplemented with 1 g/l yeast extract and 10 g/l *U. lactuca* biomass. The sea salts are an artificial mix of salts resembling the composition of ocean water. As a positive control the algal biomass was replaced with 4 g/l of glucose. As a negative control the sea salts, supplemented with only yeast extract, was used as the growth medium. The cultures were incubated at 21°C with shaking for 6 days, and growth was monitored every day using OD 600nm readings. The results are shown in Figure 4-11. Only the cultures containing *U. lactuca* biomass or glucose as the carbon source showed cell growth. No clear growth was detected for the negative control. These results demonstrate that *F. agariphila* is able to utilize *U. lactuca* biomass as the main source of carbon to grow.

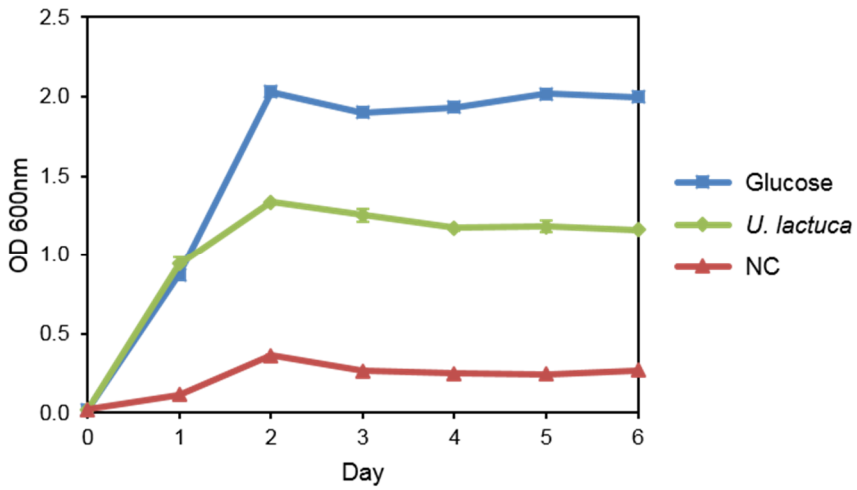

Figure 4-11 Growth of *F. agariphila* on *U. lactuca* biomass. A 35 g/l sea salts preparation supplemented with 1 g/l yeast extract and either 4 g/l glucose or 10 g/l *U. lactuca* was used as the medium. Medium containing only sea salts and yeast extract was used as the negative control (NC). The cultures were incubated at 21°C. All experiments included three biological replicates. Error bars indicate one standard error of the mean.

4.2.3 Activity screening for *F. agariphila*

The influence of *U. lactuca* biomass on the expression of *F. agariphila*'s enzymes with ulvanolytic activity was studied. *F. agariphila* cells were used to inoculate marine broth (MB) medium with and without 10 g/l *U. lactuca* biomass. Additionally, *F. agariphila* was grown on 35 g/l of sea salts supplemented with 1 g/l yeast extract and 10 g/l *U. lactuca* biomass. The cultures were incubated at 21°C with shaking for 3 days, and growth was monitored every day measuring OD 600nm. Good growth was observed under the three conditions (Figure 4-12 (A)). The highest levels of *F. agariphila* growth were obtained when MB supplemented with *U. lactuca* biomass was used as the medium. At the same time ulvan lyase activity was measured using extracellular samples from the cultures. Samples from the different cultures were obtained every day, the cells removed by centrifugation and the supernatant used for the assay. A PBS-agarose plate containing solubilized ulvan was loaded with the extracellular samples and incubated at 37°C overnight. Ulvan was solubilized by autoclaving a sample containing 10 g/l *U. lactuca*. For details about the pre-treatment see supplementary results 8.2.3. A 10% (w/v) solution of cetyl pyridinium chloride was used to stain the plate. The results are shown in Figure 4-12 (B). Halo formation indicates areas were ulvan has been degraded. Only the samples coming from cultures grown on media containing *U. lactuca* biomass showed activity. These results suggest that the presence of *U. lactuca* in the culture medium induce the production of an extracellular ulvan lyase. The samples from the MB + *U. lactuca* biomass cultures produced more defined halos. This can be explained by the higher levels of growth reached by those cultures.

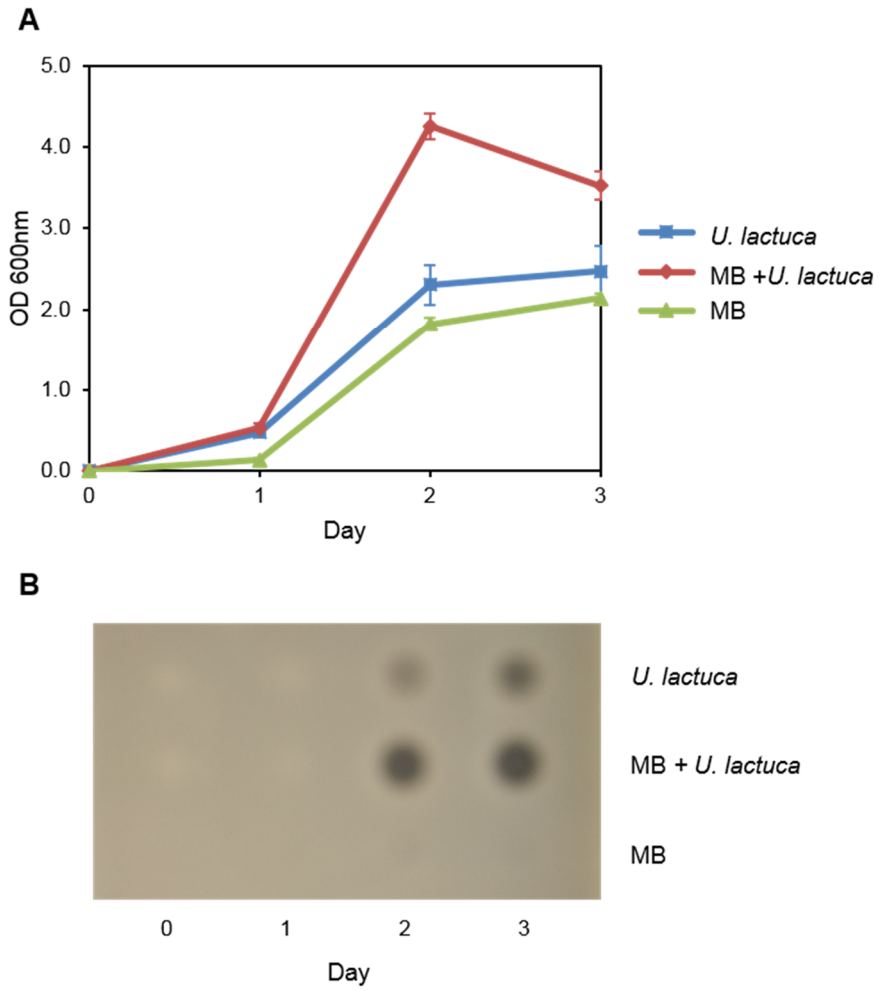

Figure 4-12 *F. agariphila* ulvan lyase induction by *U. lactuca* biomass. (A) *F. agariphila* growth on different media. A 35 g/l sea salts preparation supplemented with 1 g/l yeast extract and 10 g/l *U. lactuca* biomass, Marine Broth (MB), and MB supplemented with 10 g/l of *U. lactuca* biomass were used as the medium. The cultures were incubated at 21°C. All experiments included three biological replicates. Error bars indicate one standard error of the mean. (B) Ulvan lyase activity assay. Extracellular samples from the different cultures were deposited on a 10 g/l solubilized ulvan-PBS-agar plate and incubated at 37°C overnight. The plate was stained with 10% (w/v) cetyl pyridinium chloride. Unstained halos correspond to the areas where the ulvan has been degraded.

In order to screen for other activities, cells from the fourth day of growth were fractionated and the obtained enzyme extracts analyzed against different substrates. Samples coming from an overnight culture of *E. coli* were also fractionated and used as a control. A general ulvanase activity assay using ulvan as the substrate was performed. The samples were mixed with the substrate and incubated at 37°C for 16 h. The reducing sugars released were quantified using the DNS method. Clear activity was only detected in the extracellular samples coming from the cultures grown on

media containing *U. lactuca* biomass (Figure 4-13 (A)). As in the plate assay, the MB + *U. lactuca* samples showed approximately double the activity of those with *U. lactuca* alone. It is important to note that other extracellular enzymes might also be generating reducing sugars in this assay. The presence of rhamnosidases was analyzed using the model substrate pNPR. The samples were mixed with the substrate and incubated at 37°C for 1 h. Activity was detected only in the samples obtained from *F. agariphila* cells grown on media containing *U. lactuca* biomass (Figure 4-13 (B)). Most of the activity was observed in the extracellular space. Interestingly, the samples grown on MB + *U. lactuca* showed considerably higher levels of activity in comparison with the ones grown only on the algal biomass. Xylosidase activity was studied using the substrate MUX. The reaction was carried out at 37°C for 1 h. In this case the activity was observed in both intra- and extracellular samples, and as in the case of the other substrates the activity levels were considerably higher when *U. lactuca* biomass was present in the culture media (Figure 4-13 (C)). This time the cell lysate sample from the culture grown in *U. lactuca* showed significantly higher levels of activity than the other sub-cellular fractions. In the case of the samples obtained from the cultures grown on MB + *U. lactuca* similar levels of activity were obtained for the cell lysate and extracellular samples. Glucuronidase activity was assayed using MUGlcA as the substrate. The samples were incubated with the substrate at 37°C for 1 h. Surprisingly, no glucuronidase activity was observed in any of the samples obtained from *F. agariphila* cultures (Figure 4-13 (D)). In contrast, the cell lysate sample obtained from *E. coli* cultures showed clear activity against this substrate. These results confirm that not only an ulvan lyase is induced by the presence of *U. lactuca* biomass in the culture medium, but also rhamnosidases and xylosidases. However, contrary to what was expected no glucuronidase activity was observed, suggesting the lack of this kind of enzymes in the system. Conversely, *E. coli* cells only showed activity against MUGlcA.

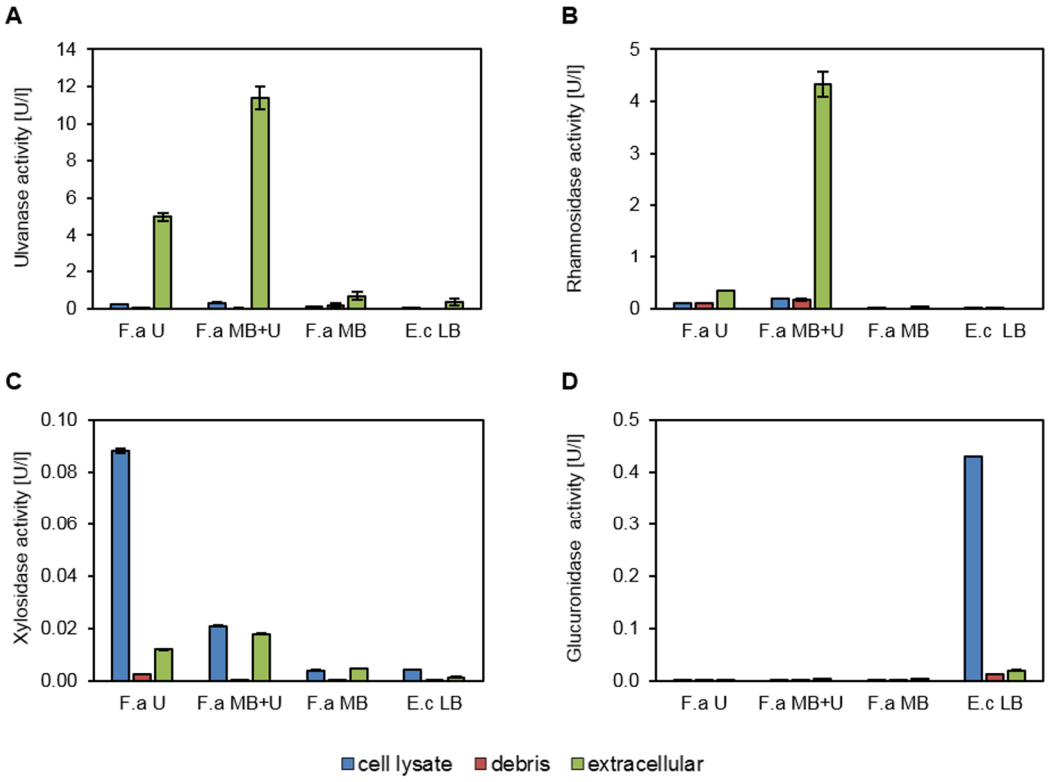

Figure 4-13 Activity screening of different sub-cellular fractions of *F. agariphila*. (A) DNS activity assay using *U. lactuca* biomass as the substrate. (B) Rhamnosidase activity assay using pNPR as the substrate. (C) Xylosidase activity assay using MUX as the substrate. (D) Glucuronidase activity assay using MUGlcA as the substrate. Activity assays were carried out using samples from different fractions of the cell. The samples were incubated with the substrate at 37°C. Key: F. a U: *F. agariphila* grown using *U. lactuca* biomass as the sole carbon source; F. a. MB+U: *F.agariphila* grown on MB medium supplemented with *U. lactuca* biomass; F. a. MB: *F. agariphila* grown on MB medium; and E.c LB: *E. coli* grown on LB medium. Experiments were done in triplicate. Activities are expressed in units of activity per litre of culture volume. Error bars indicate one standard error of the mean.

4.3 Heterologous expression of *F. agariphila*'s CAZY enzymes

4.3.1 Ulvan lyase from *F. agariphila*

In order to experimentally validate the activity of the ulvan lyase from the ulvan utilization PUL, four constructs containing the gene were designed (Figure 4-14). The first two, pSB1C3-P$_{lac}$-*lacZ'*-FA2219 and pSB1C3-P$_{lac}$-*lacZ'*-tFA2219, contain the wild-type enzyme (FA2219) and a truncated version of this one only containing the catalytic domain (tFA2219). Additionally, two other constructs containing versions of these enzymes without the signal peptide were assembled: pSB1C3-P$_{lac}$-*lacZ'*-FA2219ΔSP and pSB1C3-P$_{lac}$-*lacZ'*-tFA2219ΔSP. All the assemblies were done using

the PaperClip assembly method. All parts were amplified by PCR using their respective UF and DR primers and clips prepared to connect the different parts. Parts included the backbone (pSB1C3), a promoter (P_{lac}), and a coding DNA sequence. Specific RBSs designed using the RBS calculator were added as intervening sequences in the middle of the clips connecting the promoters with the coding sequences. After the assembly reaction, the constructs were used to transform *E. coli* JM109 cells and the sequences validated by sequencing.

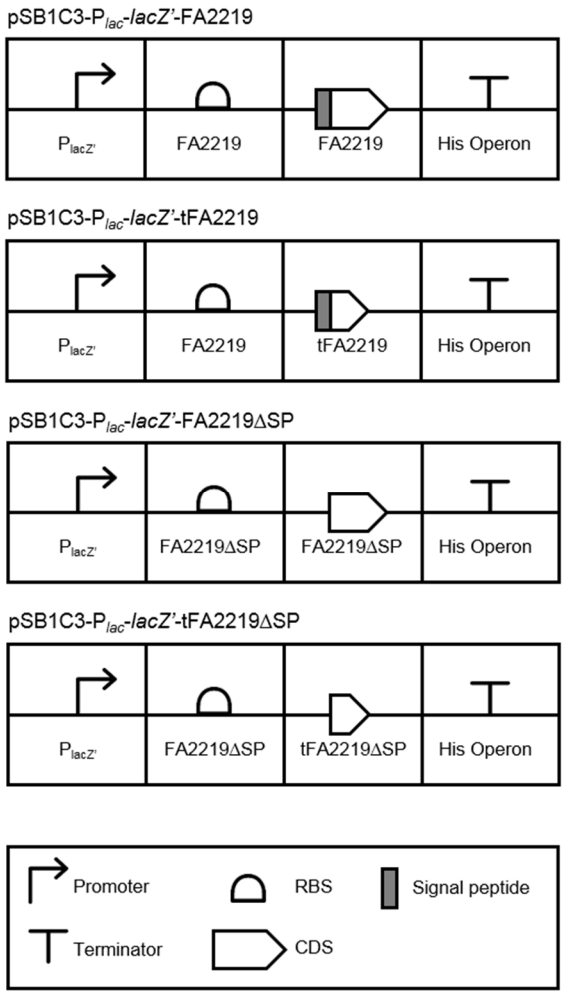

Figure 4-14 Constructs for the expression of the different versions of FA2219. All constructs were assembled using the PaperClip assembly (Trubitsyna et al., 2014). RBSs of similar strength were designed using the RBS calculator (Salis et al., 2009). FA2219: ulvan lyase from *F. agariphila*, tFA2219: truncated FA2219, FA2219ΔSP: FA2219 without signal peptide, tFA2219ΔSP: tFA2219 without signal peptide, $P_{lacZ'}$: P_{lac}-*lacZ'*. The constructs are described according to the Synthetic Biology Open Language (SBOL). RBS: ribosome binding site, CDS: coding DNA sequence.

An ulvan lyase activity assay was performed using cell lysates. *E. coli* JM109 cells carrying the four different constructs were each cultured in LB medium at 37°C and induced for 20 h with IPTG. The construct pSB1C3-P_{lac}-*lacZ'* was used as the negative control. Cells were harvested by centrifugation, re-suspended in PBS and lysed by sonication. Ulvan-PBS-agarose plates were loaded with the cell lysates and incubated overnight at 37°C. Cetyl pyridinium chloride was used to stain the plate after the incubation period. The results are shown in Figure 4-14 (A). Only the construct containing the enzymes with no signal peptide showed clear halos of activity. In addition, an activity assay using live cells was done to check for enzyme secretion. Overnight *E. coli* JM109 liquid cultures carrying the same constructs were used to inoculate ulvan-LB-agar plates containing CML and IPTG. The plates were incubated at 37°C overnight and then stained with cetyl pyridinium chloride. None of the cultures showed halos of activity (Figure 4-15 (B)). The use of pSB1C3-P_{lac}-*lacZ'*-FA2219 and pSB1C3-P_{lac}-*lacZ'*-tFA2219 in both assays produced no observable activity, suggesting that the signal peptide has a deleterious effect on the expression of the enzyme. However, FA2219ΔSP and tFA2219ΔSP were actively expressed, confirming the enzyme as an ulvan lyase.

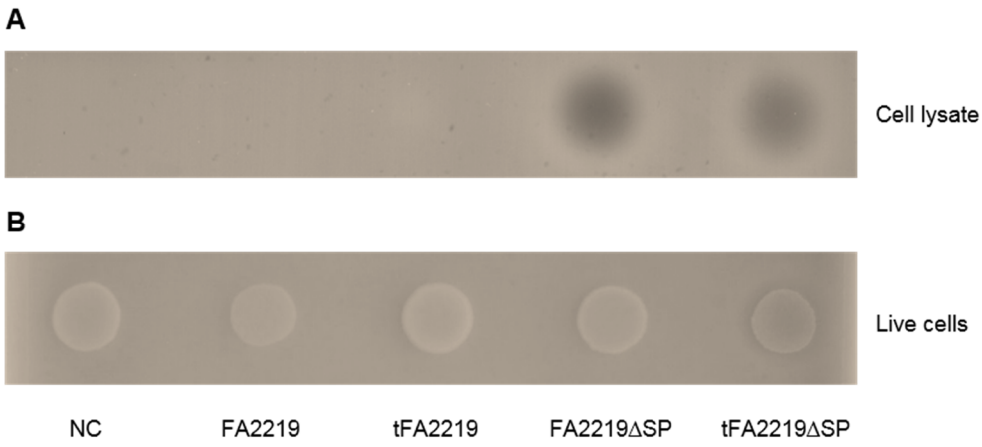

Figure 4-15 Ulvan lyase activity assay for different versions of FA2219 expressed by *E. coli*. Cell lysate (A) and live cells (B) were used for the assays. Cell lysate samples were deposited on a 10 g/l solubilized ulvan-PBS-agarose plate. *E. coli* cells carrying the different constructs were used to inoculate an LB-agarose plate containing 10g/l solubilized ulvan and 90 μg/ml IPTG. Both plates were incubated at 37°C overnight and stained with 10% (w/v) cetyl pyridinium chloride. Unstained halos correspond to the areas where the ulvan has been degraded. NC: pSB1C3-P_{lac}-*lacZ'*.

4.3.2 Characterization of the glycoside hydrolases

New constructs for the expression of the 13 identified glycoside hydrolases from the ulvan utilization PUL were assembled using PaperClip (Figure 4-16). The vector pSB1C3 was used as the backbone and a *lac* promoter added upstream of all the genes. Specific PaperClip oligonucleotides were used to amplify the different parts and prepare the clips. RBSs were added as intervening sequences between the promoter and the glycoside hydrolase genes. Considering the previous results, all identified putative signal peptides were removed by PCR. Assemblies were validated by sequencing and used to transform *E. coli* JM109 cells.

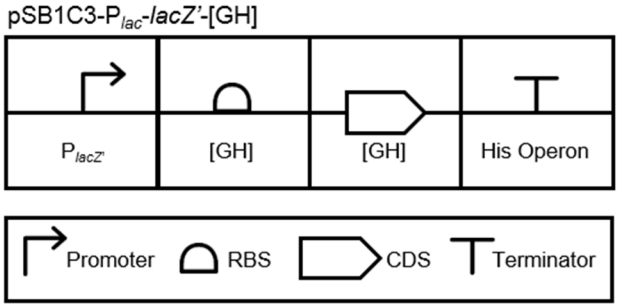

Figure 4-16 Constructs for the expression of *F. agariphila* glycoside hydrolases in *E. coli*. All constructs were assembled using the PaperClip assembly (Trubitsyna et al., 2014). RBSs of similar strength were designed using the RBS calculator (Salis et al., 2009). GH: Glycoside hydrolase from *F. agariphila*, P$_{lacZ'}$: P$_{lac}$-*lacZ'*. The constructs are described according to the Synthetic Biology Open Language (SBOL). RBS: ribosome binding site, CDS: coding DNA sequence.

In order to validate the activity of the predicted unsaturated β-glucuronyl hydrolases, a mixture of unsaturated oligosaccharides to be used as the substrate was generated from ulvan using the ulvan lyase FA2219. Cell lysates were obtained from *E. coli* JM109 cells carrying the constructs pSB1C3-P$_{lac}$-*lacZ'*-FA2190 and pSB1C3-P$_{lac}$-*lacZ'*-FA2222 in an analogous way to the ones from the ulvan lyase. The construct pSB1C3-P$_{lac}$-*lacZ'* was used as the negative control. Cell lysates were mixed with the unsaturated oligosaccharides and the reaction carried out 37°C. Loss of C=C double bonds was monitored by decrease in absorbance at 235nm (Figure 4-17). Only FA2222 glycoside hydrolase from the family 105 showed activity.

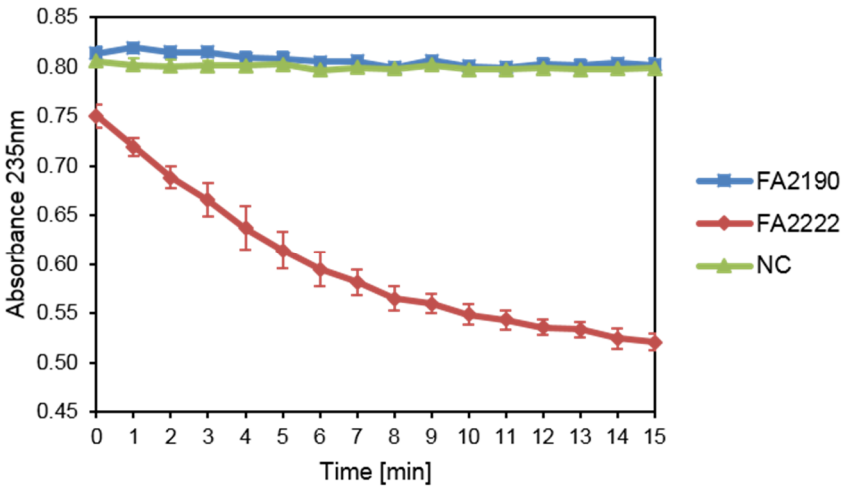

Figure 4-17 Unsaturated β-glucuronyl hydrolase activity assay. A mixture of unsaturated oligosaccharides obtained by degradation of ulvan by FA2219 from *F. agariphila* was used as the substrate. Cell lysate samples were mixed with the substrate and incubated at 37°C. Loss of C=C double bonds was monitored by the decrease in absorbance at 235nm**.** NC: pSB1C3-P_{lac}-*lacZ'*. Experiments were carried out in triplicate. Error bars indicate one standard error of the mean.

All the other glycoside hydrolases were assayed against the substrates MUX and pNPR, to check for xylosidase and rhamnosidase activities, respectively. *E. coli* JM109 cells carrying the different constructs were cultured in LB and induced with IPTG for 20 h. Cell lysates were obtained by sonication of the cells. The substrates were mixed with the different cell lysates and incubated at 37°C. The results of the xylosidase and rhamnosidase activity assays are shown in Figure 4-18 (A) and (B), respectively. FA2213, a predicted xylosidase from GH3, was the only enzyme that showed activity against MUX. FA2204 and FA2209, predicted rhamnosidases belonging to GH106 and GH78, presented higher activity against pNPR, the activity of FA2209 being approximately three times higher than that of FA2204. The activities obtained for these three enzymes coincide with the in-silico prediction.

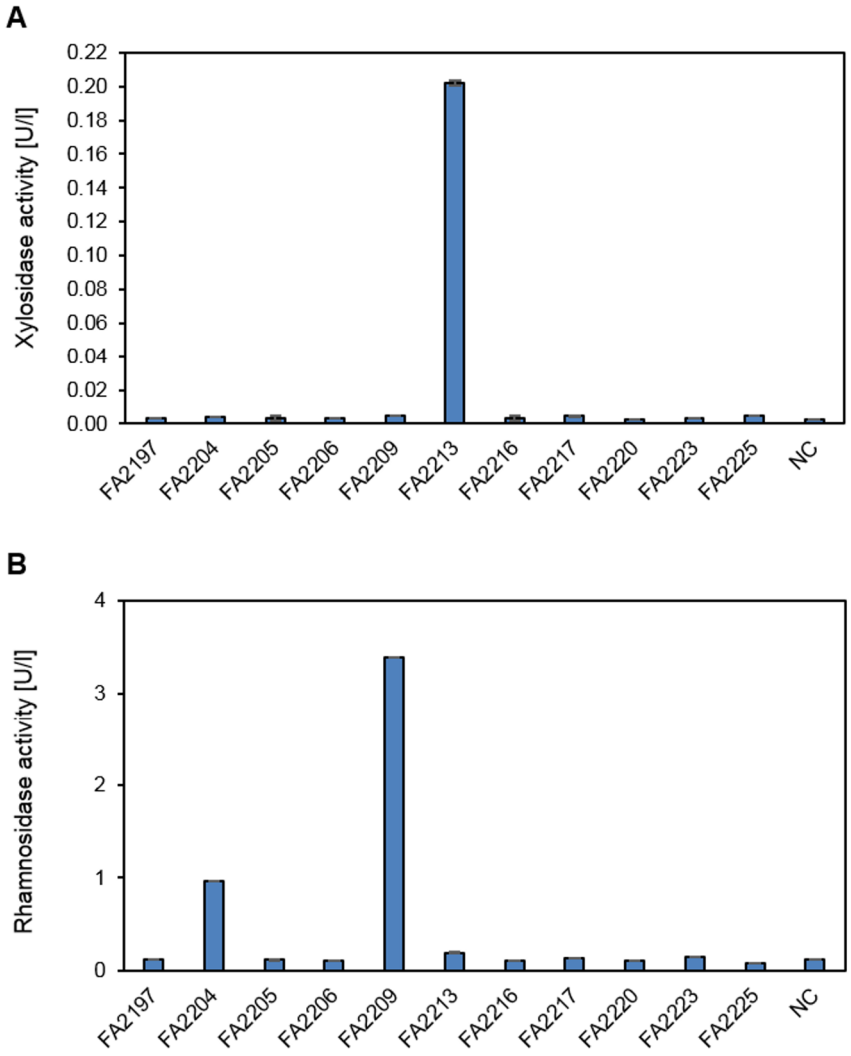

Figure 4-18 Activity screening for *F. agariphila* glycoside hydrolases. (A) Xylosidase activity assay using MUX as the substrate. (B) Rhamnosidase activity assay using pNPR as the substrate. Cell extracts were incubated with the substrates for 1h at 37°C. NC: pSB1C3-P_{lac}-*lacZ'*. Experiments were done in triplicate. Activities are expressed in units of activity per litre of culture volume. Error bars indicate one standard error of the mean.

4.3.3 *F. agariphila* enzymes for ulvan degradation and utilization

A construct containing the genes coding for the enzymes FA2194 and FA2195, which are responsible for the utilization of the unsaturated uronic acid released by the UGL, was assembled using the PaperClip assembly method (Figure 4-19). The genes were cloned into pSB1C3 and a single *lac* promoter was added. Parts and clips were prepared using specific PaperClip oligonucleotides. RBSs were added upstream of the

genes as intervening sequences. Assemblies were used to transform *E. coli* JM109 cells and the sequence validated by sequencing.

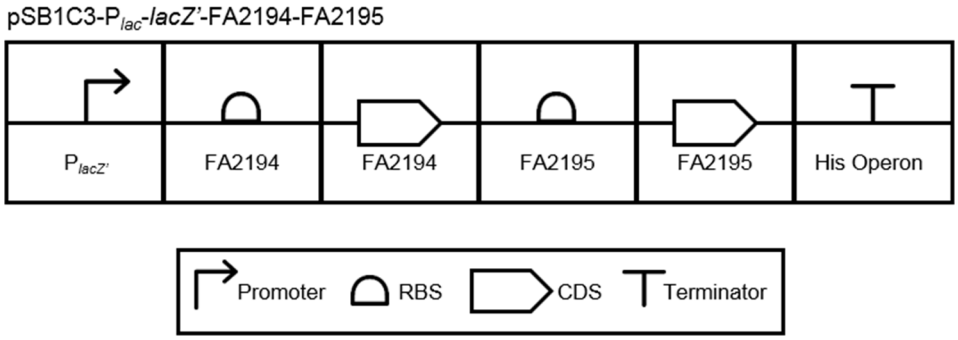

Figure 4-19 Constructs for the expression of FA2194 and FA2195 from *F. agariphila* in *E. coli*. The construct was assembled using the PaperClip assembly (Trubitsyna et al., 2014). RBSs of similar strength were designed using the RBS calculator (Salis et al., 2009). FA2194: 4-deoxy-L-threo-5-hexosulose-uronate ketol-isomerase (KduI) from *F. agariphila*, FA2295: 2-dehydro-3-deoxy-D-gluconate 5-dehydrogenase (KduD), P$_{lacZ'}$: P$_{lac}$-*lacZ'*. The constructs are described according to the Synthetic Biology Open Language (SBOL). RBS: ribosome binding site, CDS: coding DNA sequence.

In order to enzymatically pretreat ulvan the use of different combinations of heterologous expressed CAZymes was studied. M9 medium supplemented with ulvan was prepared and used as the substrate for the reaction. Cell lysates were obtained using the construct for the production of the rhamnosidase FA2209, the xylosidase FA2213, the ulvan lyase FA2219, and the UGL FA2222. Different combinations of the cell lysates were mixed with the ulvan-containing medium and the samples incubated at 37°C for 16 h. A cell lysate from *E. coli* cells carrying the construct pSB1C3-P$_{lac}$-*lacZ'* was used as a negative control. The DNS method was used to quantify the reducing sugars released by the enzymatic pre-treatments. The results are shown in Figure 4-20 (A). The pre-treatment containing only FA2219 showed considerably higher activity than the other single enzyme pre-treatments. This can be explained by the fact that FA2219 is the only enzyme having endolytic activity, highlighting its importance in the ulvan degradation system. No clear synergies were observed between the different enzymes; this might be attributed to the lack of sulfatases in the pre-treatment reactions.

A growth assay using *E. coli* JM109 cells and the enzymatically pre-treated media was performed. Cells were grown for 18 h at 37°C and the results are shown in Figure 4-20

(B). Surprisingly, an improvement in growth was observed when FA2219 was used alone for the pre-treatment, but no major changes when the other enzymes were added. In addition, the use of construct pSB1C3-P$_{lac}$-*lacZ'*-FA2194-FA2195, containing the enzymes for the utilization of the unsaturated uronic acid released by the UGL was studied. *E. coli* JM109 cells carrying the construct were grown using the same conditions as described for the previous experiment. Similar levels of growth were achieved for all the pre-treatments; however, with overall slower growth rates (Figure 4-20 (C)). These results suggest that the construct has a negative effect on *E. coli* growth and no apparent benefit.

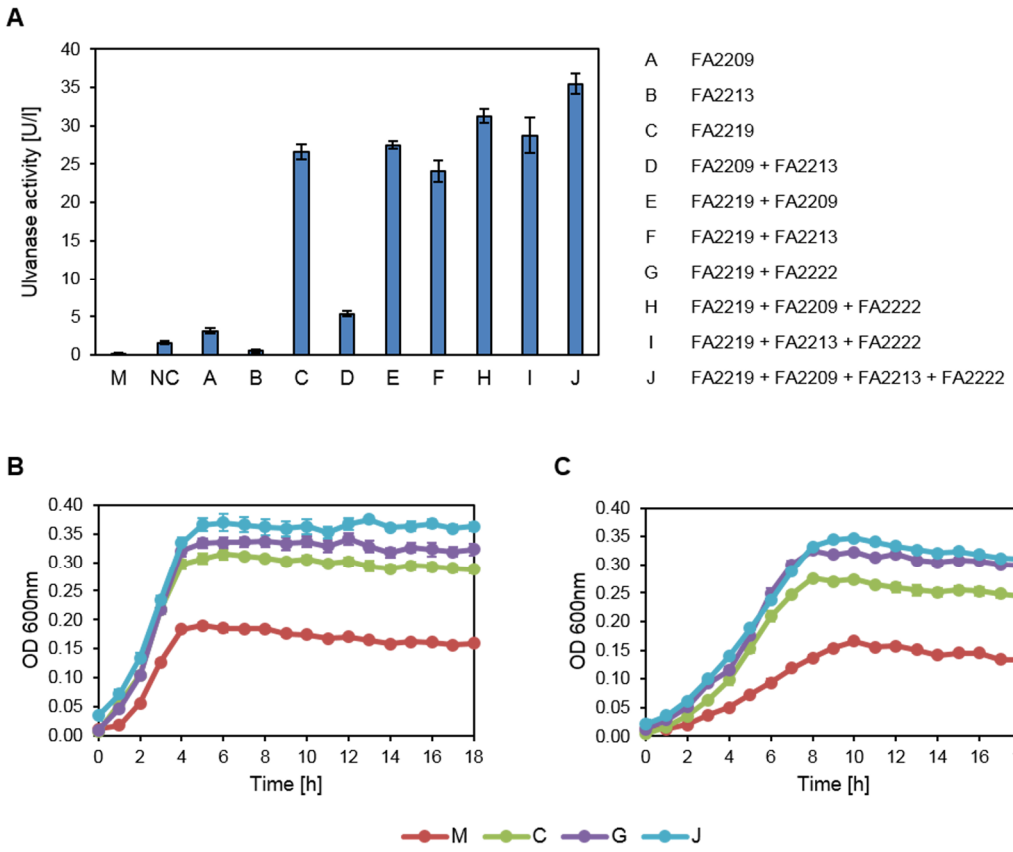

Figure 4-20 Growth of *E. coli* JM109 on enzymatically pre-treated ulvan. (A) DNS activity assay using ulvan in M9 medium supplemented with 0.2 g/l yeast extract as the substrate. Different combinations of enzyme extracts were mixed with the substrate and incubated at 37°C for 16 h. (B) Growth of *E. coli* JM019 on enzymatically pretreated ulvan. (C) Growth of *E. coli* JM019 carrying the construct pSB1C3-P$_{lac}$-*lacZ'*-FA2194-FA2195 on enzymatically pretreated ulvan. All cultures were grown at 37°C and IPTG used for induction. M: not pre-treated medium. NC: Enzyme extract obtained from cell carrying the construct pSB1C3-P$_{lac}$-*lacZ'*. Experiments were done in triplicate. Error bars indicate one standard error of the mean.

To verify whether pSB1C3-P_{lac}-*lacZ'*-FA2194-FA2195 has a negative effect on *E. coli* growth, a new analysis was done this time using M9 supplemented with glucose as the medium. *E. coli* JM109 cells with and without the construct were cultured at 37°C for 18 h. The results confirmed a decrease in the growth rate when the construct was used (Figure 4-21).

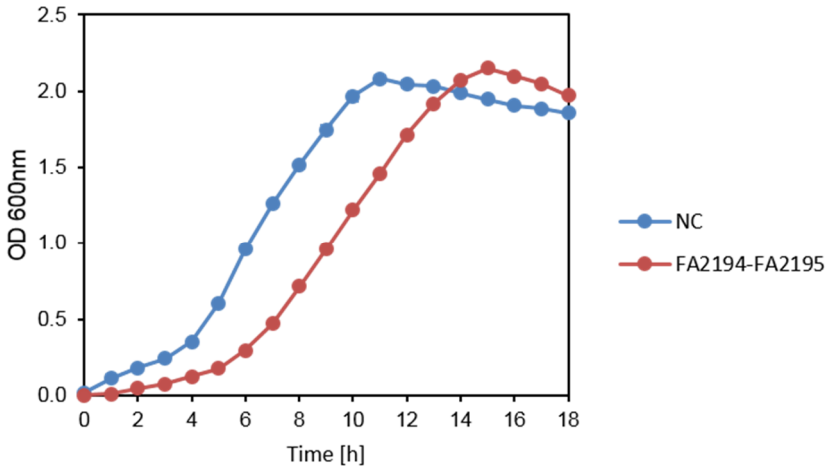

Figure 4-21 Growth of *E. coli* JM019 expressing FA2194 and FA2195 on M9-glucose medium. M9 supplemented with 0.2 g/l yeast extract and 4 g/l glucose. FA2194-FA2195: cells carrying the construct pSB1C3-P_{lac}-*lacZ'*-FA2194-FA2195. NC: cells carrying the construct pSB1C3-P_{lac}-*lacZ'*. All cultures were grown at 37°C and IPTG used for induction. Experiments were carried out in triplicate. Error bars indicate one standard error of the mean.

4.4 Discussion

A potential PUL for ulvan utilization was detected in the genome of the seaweed-associated bacteria *F. agariphila*. A total of 13 PULs were identified in *F. agariphil*a's genome (Mann et al., 2013). The PUL for ulvan utilization is the longest one, which can be explained by the high complexity and variability of the substrate. The activity and cell localization of the proteins encoded by the PUL were predicted in-silico and used to build a tentative model for ulvan utilization. The PUL includes an extracellular ulvan lyase and a UGL, enzymes that have been already reported to be involved in ulvan depolymerization (Collen et al., 2014; Collen et al., 2011). However, other enzymes such as rhamnosidases, xylanases, β-glucuronidases and sulfatases were also identified as part of the saccharolytic machinery. The PUL also contains the enzymes KduI and KduD, which participate in the utilization of the unsaturated uronic acids

released by the UGL. Additionally, two HTCSs and a GntR-like protein were found. HTCSs are specific regulatory systems from Bacteroidetes involved in the transcriptional regulation of the genes contained in the PULs, while the GntR-like proteins control specific monosaccharide utilization pathways (Ravcheev et al., 2013).

The activities of the CAZymes from the PUL were predicted based on the CAZY families they belong to and phylogenetic studies of the characterized members of the family when this possessed more than one activity. The enzymes belonging to GH3 were clustered with enzymes possessing different activities, and the structure of ulvan was considered as an additional criterion for the assignation of the activity. In particular, FA2220, a member of GH39, did not show high levels of homology with any of the enzymes of the family; however, the presence of iduronic acid in ulvan might suggest that FA2220 is a novel α-iduronidase. So far no bacterial α-iduronidase has been reported and all of the eukaryotic enzymes with that activity belong to GH39. Overall one ulvan lyase (EC 4.2.2.-), two UGL (EC 3.2.1.-), four α-L-rhamnosidases (EC 3.2.1.40), three β-xylosidases (EC 3.2.1.37), three β-glucuronidases (EC 3.2.1.31), and one α-iduronidase (EC 3.2.1.76) were predicted. Nevertheless, experimental validation of the activities is required. Additionally, CBMs were found to be associated with four of the enzymes, providing extra insights about the activities of their cognate catalytic domains. Members of CBM35, a family of CBM identified in one of the predicted cytoplasmic β-glucuronidases, have shown ligand specificity for glucuronic acid (Montanier et al., 2009). Likewise, the two predicted outer membrane rhamnosidases contain a CBM67. This family of CBMs has been shown to enhance the catalytic efficiency of the enzymes by binding terminal rhamnose residues from oligosaccharides and polysaccharides (Fujimoto et al., 2013). Finally, a CBM13 was found in the ulvan lyase. CBM13s have been reported to improve catalytic efficiency of both glycoside hydrolases and alginate lyases (Fujimoto, 2013; Li et al., 2015). Moreover, a predicted outer membrane hypothetical protein showed a low level of sequence identity (27% amino acid identity) with a heparinase II/III-like protein, suggesting that it might be part of a novel ulvan lyase family.

Ulvan is a highly sulfated polysaccharide (Lahaye & Robic, 2007). Sulfate groups confer resistance to saccharification by bacterial glycoside hydrolases (Ulmer et al.,

2014). Considering that, desulfation of the oligosaccharides by sulfatases is crucial for the release of monosaccharides units by the different glycoside hydrolases. Several sulfatases involved in the degradation of GAGs have been characterized according their substrate specificity (Hanson et al., 2004). However, no sulfatase involved in ulvan utilization has been reported and thus far, no amino acid sequence associated with a bacterial 3-O-sulfatase is available. Nevertheless, the characterized GAG degrading sulfatases were used for a phylogenetic analysis of the 8 sulfatases found in the PUL. Although it was not possible to determine the substrate specificity of the enzymes, the results showed high levels of sequence diversity. Moreover, all the sulfatases of the PUL were categorized as Cys-type after the identification of their conserved signature sequence. This facilitates their heterologous expression in a host like *E. coli*, not requiring the presence of an anaerobic sulfatase-maturating enzyme (anSME) as in the case of Ser-type sulfatases (Ulmer et al., 2014). The characterization of these enzymes is of great importance to fully understand the biodegradation of ulvan.

Three types of specific regulatory systems have been reported for the regulation of PULs from Bacteroidetes: SusR-like regulators, HTCSs and extracytoplasmic function (ECF) sigma/anti sigma factors (Ravcheev et al., 2013). In the case of the PUL for ulvan utilization from *F. agariphila* two HTCS were identified. Both HTCSs possess a COG3292 sensor domain, thus are predicted to be activated by oligosaccharides. A phylogenetic analysis showed that both proteins are closely related to HTCSs associated with the utilization of chondroitin sulfate and heparin/heparan sulfate in *B. thetaiotaomicron*, suggesting that the sulfate groups might be important for the recognition of the signature oligosaccharides. However, the signature oligosaccharides that activate the PUL still remain unknown and further characterization of the proteins is required. Isothermal titration calorimetry (ITC) has been used to study the binding of different oligosaccharides to the sensor domain of a HTCS (Lowe et al., 2012), and it might be useful to determine the signature oligosaccharide which induces this system.

Moreover, FA2230, a TR that belongs to the GntR-like family, was identified as part of the PUL. Interestingly, FA2230 showed high homology with a conserved domain

of the sub-family FucR, which regulates the utilization of fucose. However, no fucose is found in the structure of ulvan. Considering that fucose is closely related to rhamnose and no TRs homologous to known rhamnose associated TRs were identified in *F. agariphila*, FA2230 is likely to be the first member of a novel subfamily of GntR-like TRs. Additionally, the ulvan utilization PUL is the only PUL in *F. agariphila* that is associated with the degradation of rhamnose-based polysaccharides and the presence of a TR associated with the utilization of its main final product is logical. FA2230 is likely to function in an analogous way to FucR but instead of having fucose binding to the FucR repressor to induce the expression of the fucose utilization pathway (Hooper et al., 1999), rhamnose binds FA2230 to induce the expression of enzymes for the utilization of rhamnose.

The pathways for the utilization of the released monosaccharides in *F. agariphila* were also reconstructed by comparative genomics. The reconstructed pathways considered all the enzymes involved in the transformation of the sugar to an intermediate of the central metabolism, including cytoplasmic catabolic enzymes and inner membrane sugar transporters. The pathways for the utilization of rhamnose, xylose and glucuronic acid were fully reconstructed. Additionally, the pathway for the utilization of the unsaturated uronic acid released by the UGL was partially reconstructed including two of the enzymes from the PUL (FA2194 and FA2195). However, no proteins homologous to reported permeases were identified. No utilization pathway for iduronic acid was found in the literature. The reconstructed pathways serve as bridges between the ulvan degradation machinery and the central metabolism of *F. agariphila*; however, they need to be expanded and refined with experimental data.

F. agariphila was able to grow on *U. lactuca* biomass as the sole carbon source, suggesting that it possess all the pathway to degrade and utilize the algal biomass. Considering that *F. agariphila* is not able to grow on cellulose (Mann et al., 2013), the growth might be associated with the expression of the enzymes of the ulvan utilization PUL. In support of this theory, it was shown that enzymes with ulvanase activities were induced when the algal biomass was in the culture medium. Ulvan lyase, rhamnosidase and xylosidase activities were detected when media containing *U. lactuca* biomass were utilized. These results are consistent with the predicted enzymes

observed in the PUL. The ulvan lyase activity was observed mainly in the culture medium, confirming that is an extracellular enzyme. Interestingly, differences were observed in rhamnosidase and xylosidase activities associated with the different subcellular fractions when *F. agariphila* was grown on *U. lactuca* or MB + *U. lactuca*. This might suggest that the peptone in the medium might affect somehow the regulation of the PUL. Surprisingly, no activity was observed when the model substrate for β-glucuronidases MUGlcA was used. A transcriptome analysis of the PUL might be useful to elucidate how the expression of these enzymes is affected by the presence of the algal biomass in the culture medium.

To validate the activity of the CAZymes identified in the PUL, these were cloned into *E. coli* and heterologously expressed. FA2219 was confirmed as an ulvan lyase. Additionally, a truncated version of FA2219 containing only the catalytic domain was also functionally expressed. However, the ulvan lyase activity was observed only when the enzymes were expressed without their putative signal peptides. These results suggest that, as in the case of CHU2268, the signal peptides have a deleterious effect on the expression of the enzyme. Considering this, all the other enzymes were expressed without their signal peptides. From the predicted UGL, only FA2222 showed activity when unsaturated oligosaccharides produced by the ulvan lyase were utilized as the substrate. The remaining glycoside hydrolases were tested for xylosidase and rhamnosidase activity. FA2204 and FA2209 showed rhamnosidase activity, while FA2213 showed xylosidase activity. These results confirm their predicted activities obtained in-silico. MUGlcA was not tested as a substrate due to background activity generated by a β-glucuronidase from *E. coli*. Considering this, purification of the enzymes might be required. Moreover, different growth conditions and additional substrates have to be tested to fully characterize the CAZymes from the PUL.

In order to study possible synergies between the heterologous expressed CAZymes, different combination of these enzymes were used to degrade ulvan. However, no synergies were detected and most of the degradation was generated by the ulvan lyase. These results might be explained by the presence of sulfate groups preventing the action of the glycoside hydrolase, implying that the sulfatases are essential for the

complete depolymerization of ulvan. Additionally, the enzymatically pre-treated ulvan was used as the culture medium for *E. coli*. Surprisingly, a considerable growth increase was observed when ulvan was pretreated only with the ulvan lyase, suggesting that *E. coli* is able to utilize the unsaturated oligosaccharides generated by this enzyme. A GH105 found in *E. coli's* genome (locus tag WP_001767851) showed 34% amino acid identity with the UGL from *F. agariphila* and might explain these results. However, this enzyme has not been characterized and its activity against the unsaturated oligosaccharides has to be experimentally validated. Furthermore, homologues to KduI (locus tag WP_000383256) and KduD (locus tag WP_000345753) showing 55% and 56% amino acid identity with the ones from *F. agariphila* were also found in *E. coli*. Moreover, *E. coli* growing in the same pretreated media expressing KduI and KduD from *F. agariphila* was also analyzed; however, only a decrease in the growth rate was observed. These results suggest that the use of *E. coli*'s native enzymes is sufficient to allow growth from the unsaturated residues generated by the ulvan lyase. Nevertheless, the levels of growth achieved were very poor and the system has to be optimized.

Overall, it was possible to confirm *F. agariphila*'s capability to utilize green algal biomass by three lines of evidence: (1) analysis of a PUL found in its genome containing all the potential genes required for ulvan utilization, (2) growth on *U. lactuca* as the sole carbon source with clear induction of ulvanase enzymes, and (3) the heterologous expression of the CAZymes from the PUL validating their predicted activities. Moreover, it was possible to reconstruct for the first time a pathway for ulvan utilization based on the enzymes from the PUL. Furthermore, ulvan lyase, rhamnosidase, and xylosidase activities were found to be induced by the presence of *U. lactuca* biomass in the culture medium. The heterologous expression in *E. coli* of an ulvan lyase, 2 rhamnosidases and a xylanase from the ulvan utilization PUL, confirmed the importance of these enzymes for the depolymerization of ulvan. This is the first time that rhamnosidases and xylosidases associated with the depolymerization of ulvan have been identified. Nevertheless, this is just a first step to completely understand the machinery required for the utilization of ulvan and further work regarding the identification of the enzymes and regulation of the pathway is necessary.

5 Secretion of the ulvan lyase from *F. agariphila* in *E. coli*

5.1 Introduction

Consolidated bioprocessing (CBP) has been proposed as a cost effective alternative for biofuel production (Lynd et al., 2005), combining enzyme generation, biomass saccharification and biofuel/bio-product production in a single bioreactor. This can be achieved using genetic engineering, introducing both biomass-degrading and biofuel-producing pathways into a single microorganism, avoiding the costs associated with a dedicated enzyme generation step (Bokinsky et al., 2011).

This approach has been already successfully used for the production of ethanol from brown macroalgal biomass using *E. coli* as the host (Wargacki et al., 2012). However, no CBPs have been reported for the bioconversion of green macroalgal biomass into biofuels or other valuable products.

E. coli's unparalleled genetic tractability and its natural capability to assimilate rhamnose, xylose and glucuronic acid, the main building blocks of ulvan, make it an ideal host for the design of a CBP for the utilization of green macroalgal biomass. Nevertheless, a heterologous ulvan degradation system still has to be designed and incorporated into *E. coli*. In particular, the secretion of an endolytic ulvan lyase is crucial in order to generate smaller oligosaccharides from ulvan that can be uptaken by the cells.

E. coli is known as a poor secretor of proteins (Ni & Chen, 2009). Several strategies have been studied to promote the extracellular expression of recombinant enzymes, including the use of outer membrane leaky strains, engineered secretion apparatus from *E. coli* and secretion partners with unknown translocation mechanisms (Ni & Chen, 2009). Successful secretion of specific enzymes using these systems is difficult to predict and usually involves much trial and error.

In chapter 4 it was shown that FA2219 from *F. agariphila* expressed by *E. coli* JM109 was active against ulvan, making it an interesting candidate for a heterologous ulvan

degradation system. However, this strain was not able to secrete the enzyme, probably due to the lack of the type IX secretion system (T9SS) which is usually only found in Bacteroidetes (Abby et al., 2016). In order to secrete this enzyme, and use it as the first step of an ulvan degradation pathway, the use of three different system were analyzed: (1) the secretion partner OsmY from *E. coli* (Qian et al., 2008), (2) an antigen 43 (Ag43) autotransporter based system (Wargacki et al., 2012), and (3) the endoglucanase CenA from *C. fimi* used as secretion partner.

5.2 Secretion of the ulvan lyase using different systems

5.2.1 Secretion using the fusion partner OsmY

OsmY, a small (18.2 kDa) osmotically-inducible periplasmic protein from *E. coli*, has been reported as one of the most efficient fusion partners for the secretion of heterologous proteins of different sizes and origins using *E. coli* (Qian et al., 2008). This secretion system has been already utilized for the design of CBPs for the production of bio-products from different types of polysaccharides (Bokinsky et al., 2011; Zheng et al., 2012). Therefore, it is an interesting alternative for the secretion of the ulvan lyase from *F. agariphila*.

Three fusion proteins were designed in order to study the secretion of the ulvan lyase FA2219 using OsmY as a secretion partner (Figure 5-1 (A)). Considering that a smaller protein requires less energy and carbon flux diverted toward its synthesis and that it has been reported that larger proteins are more difficult to pass though the membrane (Ni & Chen, 2009), only the catalytic domain of FA2219 without the signal peptide (tFA2219) will be secreted. Linkers with different properties were used between OsmY and tFA2219, including standard flexible and rigid linkers (FlexL and RigidL, respectively) (Chen et al., 2013), and a proline-threonine linker from the exoglucanase Cex from *C. fimi* (CexL) (Poon et al., 2007) (see Table 2-6 for the sequences).

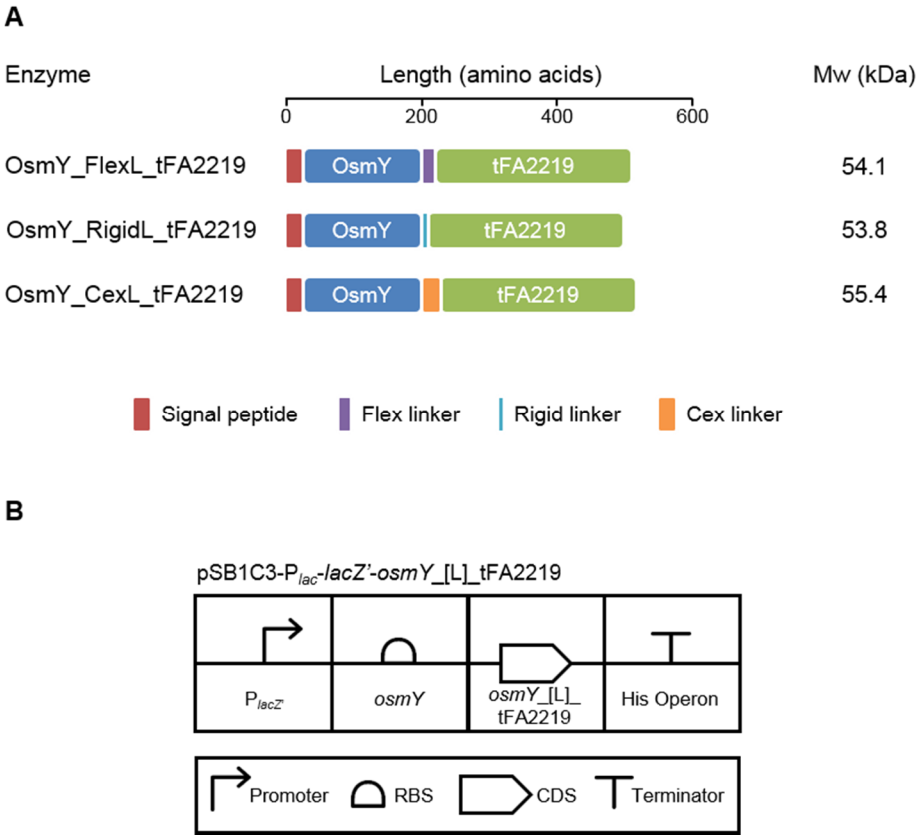

Figure 5-1 OsmY secretion system for *F. agariphila*'s ulvan lyase. (A) Design of the OsmY-tFA2219 fusion proteins. Three different linkers were used to connect the enzymes. The signal peptide from OsmY is also indicated. (B) Constructs for the secretion of the ulvan lyase using the fusion partner OsmY. All constructs were assembled using the PaperClip assembly (Trubitsyna et al., 2014). OsmY: osmotically inducible protein from *E. coli*, tFA2219: truncated ulvan lyase without signal peptide from *F. agariphila*, $P_{lacZ'}$: P_{lac}-*lacZ'*, [L]: linker. The constructs are described according to the Synthetic Biology Open Language (SBOL). RBS: ribosome binding site, CDS: coding DNA sequence.

The different constructs were assembled using PaperClip assembly (Figure 5-1 (B)). All parts were amplified by PCR using their respective UF and DR primers. In particular, *osmY* without its stop codon was obtained using *E. coli* JM109 cells as the template. A specific RBS for OsmY with a target translation initiation rate of 10,000 on the RBS calculator proportional scale, designed using the RBS calculator (see Table 2-5 for the sequence), and the linker sequences were added as intervening sequences in the middle of the clips connecting $P_{lacZ'}$-*osmY* and *osmY*-tFA2219, respectively. Four part assembly reactions were performed using the parts pSB1C3, $P_{lacZ'}$, *osmY* and tFA2219 and the respective clips to connect the parts. The results of the assembly reactions were used to transform *E. coli* JM109 cells and constructs validated by sequencing.

An ulvan lyase activity assay using live cells was performed to check whether *E. coli* is able to secrete the OsmY-tFA2219 fusion proteins. It has been reported that *E. coli* BL21 (DE3) cells exhibit considerably higher levels of secretion compared with *E. coli* K12 strains using this system (Qian et al., 2008), thus it was selected to be used for the assay. *E. coli* BL21 (DE3) cells were transformed with the different constructs and overnight cultures used to inoculate an ulvan-LB-agarose plate. Cells carrying the construct pSB1C3-P_{lac}-*lacZ'* were used as a negative control. The plate was incubated at 37°C overnight and then stained with cetyl pyridinium chloride. Halos indicating degradation of ulvan were observed only when the cells were carrying the constructs containing the flexible linkers, pSB1C3-P_{lac}-*lacZ'-osmY*_FlexL_tFA2219 and pSB1C3-P_{lac}-*lacZ'-osmY*_CexL_tFA2219 were used (Figure 5-2). These results suggest that the linkers between the two domains are the great importance, affecting the secretion and/or activity of the ulvan lyase.

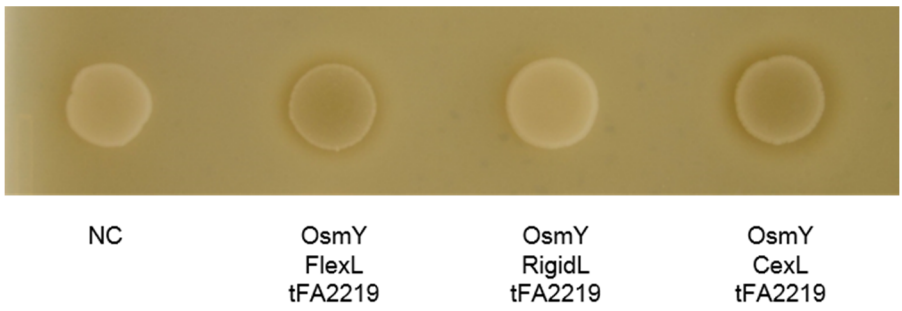

Figure 5-2 Ulvan lyase activity assay for the different OsmY secretion constructs expressed by *E. coli*. *E. coli* BL21 (DE3) cells carrying the different constructs were used to inoculate an LB-agarose plate containing 10g/l solubilized ulvan and 90 µg/ml IPTG. The plate was incubated at 37°C overnight and stained with 10% (w/v) cetyl pyridinium chloride. Unstained halos correspond to the areas where the ulvan has been degraded. NC: pSB1C3-P_{lac}-*lacZ'*.

5.2.2 Secretion using CenA as a fusion partner

Considering that the endoglucanase CenA from *C. fimi* is secreted by *E. coli* (Guo et al., 1988), the possibility of using this enzyme as a secretion partner and at the same time generating a fusion enzyme able to degrade both ulvan and cellulose was also explored. An enzyme containing these two activities would be highly useful considering that ulvan and cellulose are the main polysaccharides from green algal biomass. The use of an endoglucanase from *Bacillus subtilis* as a fusion partner for

secretion of heterologous proteins in *E. coli* has been recently reported (Gao et al., 2015), supporting the idea of using CenA as a secretion partner too.

The constructs containing the CenA-tFA2219 fusion proteins were designed and assembled in an analogous way to the ones where OsmY was used, utilizing the same three different linkers between the two enzymes (Figure 5-3). The results of the three assemblies were used to transform E. coli *JM109* and the constructs validated by sequencing.

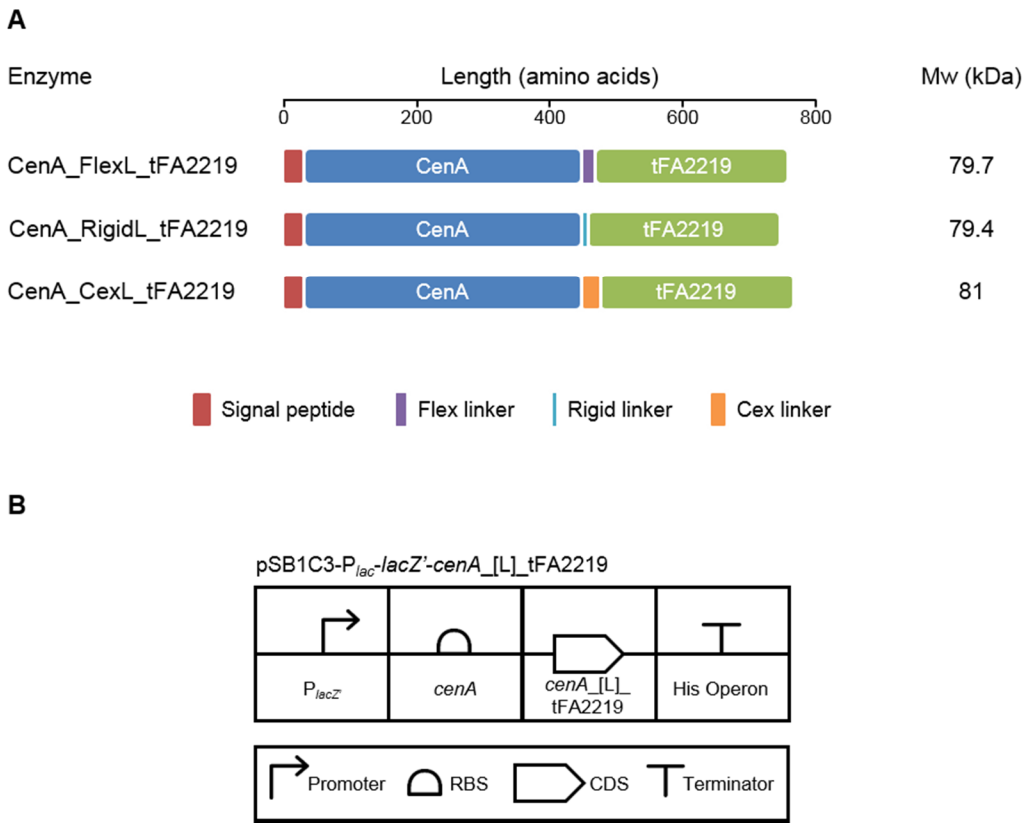

Figure 5-3 CenA secretion system for *F. agariphila*'s ulvan lyase. (A) Design of the CenA-tFA2219 fusion proteins. Three different linkers were used to connect the enzymes. The signal peptide from CenA is also indicated. (B) Constructs for the secretion of the ulvan lyase using CenA as a fusion partner. All constructs were assembled using the PaperClip assembly (Trubitsyna et al., 2014). CenA: endoglucanase from *C. fimi*, tFA2219: truncated ulvan lyase without signal peptide from *F. agariphila*, P$_{lacZ'}$: P$_{lac}$-*lacZ'*, [L]: linker. The constructs are described according to the Synthetic Biology Open Language (SBOL). RBS: ribosome binding site, CDS: coding DNA sequence.

In order to check whether *E. coli* is able to secrete the different CenA-tFA2219 fusion proteins, activity assays for endoglucanase and ulvan lyase activity using live cells were performed. To do so, *E. coli* BL21 (DE3) cells were transformed with the

different constructs and overnight cultures used to inoculate plates containing CMC-LB-agar and ulvan-LB-agarose. Cells carrying the constructs pSB1C3-P_{lac}-*lacZ'* and pSB1C3- P_{lac}-*lacZ'-cenA* were used as a negative control and positive control for the endoglucanase activity assay, respectively. The plates were incubated at 37°C overnight. The plate containing CMC was stained using Congo Red and washed with NaCl. The results are shown in Figure 5-4 (A), where the unstained halos correspond to the areas where the CMC has been degraded. The cells expressing CenA alone were the ones that generated the biggest halo, following by the ones expressing CenA_FlexL_tFA2219, CenA_CexL_tFA2219 and CenA_RigidL_tFA2219. No clear activity was observed for the negative control. The presence of ulvan lyase activity was studied after staining the plate containing ulvan with cetyl pyridinium chloride. None of the samples showed halos of activity (Figure 5-4 (B)). This indicates that no enzyme with ulvan lyase activity was secreted.

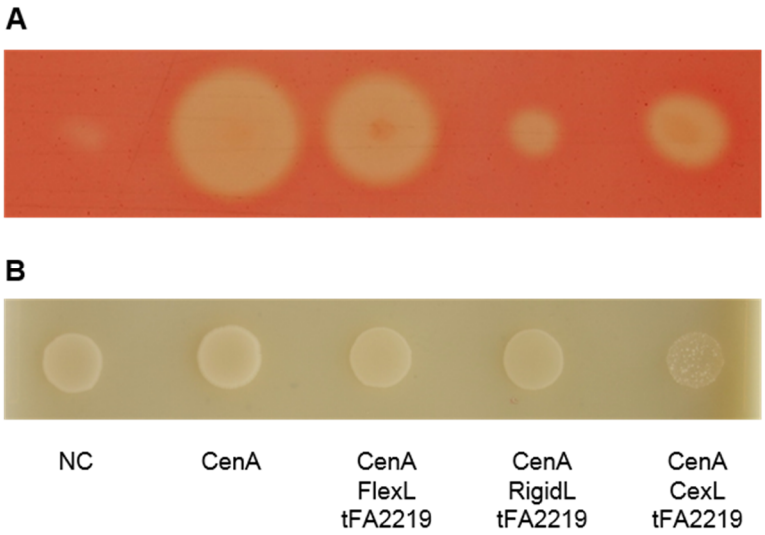

Figure 5-4 Live cell activity assays for the CenA secretion constructs expressed by *E. coli*. (A) CMC activity assay for endoglucanase activity. An LB-agar plate containing 2 g/l CMC and 90 µg/ml IPTG was used for the assay. *E. coli* BL21 (DE3) cells carrying the constructs were used to inoculate an LB-agar plate containing 2 g/l CMC and 90 µg/ml IPTG. The plate was incubated at 37°C overnight, stained with Congo Red and washed with NaCl. Unstained halos correspond to areas where CMC has been hydrolysed. (B) Ulvan lyase activity assay. *E. coli* BL21 (DE3) cells carrying the constructs were used to inoculate an LB-agarose plate containing 10g/l solubilized ulvan and 90 µg/ml IPTG. The plate was incubated at 37°C overnight and stained with 10% (w/v) cetyl pyridinium chloride. Unstained halos correspond to the areas where the ulvan has been degraded. NC: pSB1C3-P_{lac}-*lacZ'*.

An additional endoglucanase activity assay was performed using enzyme extracts from different sub-cellular localization in *E. coli* BL21 (DE3) cell carrying the CenA-tFA2219 constructs. The cells were obtained from LB cultures grown at 37°C overnight. The samples were deposited on a CMC-PBS-agar plate and incubated at 37°C overnight. The plate was stained with Congo Red and washed with NaCl. The results are shown in Figure 5-5. The samples containing CenA were the ones that showed the biggest halos in all the different fractions. A similar activity profile to the one obtained using the live cells was observed when the extracellular samples were assayed. Interestingly, a different activity profile was observed for the periplasmic samples, where similar levels of activity were observed from the construct containing the flexible and rigid linkers. None of the fractions associated with the negative control showed activity.

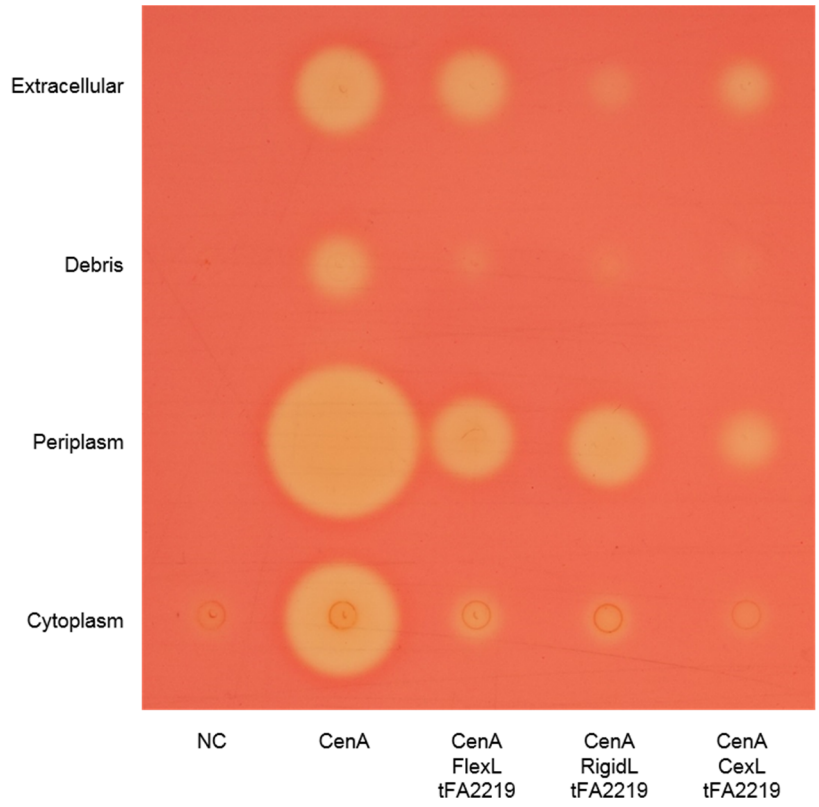

Figure 5-5 CMC activity assay of the different sub-cellular fractions of *E. coli* cells carrying the CenA secretion constructs. *E. coli* BL21 (DE3) cells carrying the different constructs were grown in LB medium with 90 µg/ml IPTG at 37°C overnight. Enzyme extracts belonging to the different sub cellular fractions obtained and deposited on a 2 g/l CMC-PBS-agar plate and incubated at 37°C overnight. The plate was stained with Congo Red and washed with NaCl. Unstained halos correspond to areas where CMC has been hydrolysed. NC: pSB1C3-P_{lac}-*lacZ'*.

5.2.3 Secretion using the autotransporter protein Ag43

Antigen 43 (Ag43), is an autotransporter protein found in most *E. coli* strains that contains within itself all the components required for transport to the outer membrane and extracellular secretion (Kjaergaard et al., 2002). Ag43 consists of an N-terminal signal peptide that directs the protein to the periplasmic space, an α-passenger domain, and a C-terminal outer membrane translocator or β-carrier domain (Henderson & Owen, 1999). The α-passenger domain is transported outside the cell by the outer membrane β-domain. Once translocated, the passenger domain is cleaved, presumably by an autoproteolytic mechanism, and anchored to the cell surface by noncovalent interaction with the β-domain (Henderson & Owen, 1999).

This system has been mainly used as a tool to display epitopes and protein domains on bacteria cell surfaces (Kjaergaard et al., 2002). However, Wargacki et al. (2012) re-engineered the system implementing two design modifications in order to secrete an alginate lyase. First, a substantial portion from the α-domain was removed to prevent surface attachment (from A52 to N455) and replaced by the catalytic domain of the alginate lyase. Second, an aspartyl protease active site from the removed portion of the α-domain was re-inserted between the alginate lyase domain and the remaining Ag43.

An Ag43-tFA2219 secretion system was designed based on the specification given by Wargacki et al. (2012) (Figure 5-6 (A)). A construct containing the Ag43-tFA2219 secretion system was assembled using PaperClip assembly (Figure 5-6 (B)). A 5 part assembly was performed using pSB1C3, $P_{lacZ'}$, Ag43 signal peptide, tFA2219 and the Ag43N455 as the parts. Ag43 signal peptide and Ag43N455 were amplified by PCR using DNA obtained from *E. coli* JM109 cells as the template. Additionally, a specific RBS with a target translation initiation rate of 10,000 on the RBS calculator proportional scale, designed using the RBS calculator (see Table 2-5 for the sequence), and the aspartyl protease active site were added as intervening sequences in the middle of the clips used to connect $P_{lacZ'}$-Ag43 signal peptide and tFA2219-Ag43N455, respectively. *E. coli* JM109 cells were transformed with the result of the assembly and the sequence confirmed by sequencing.

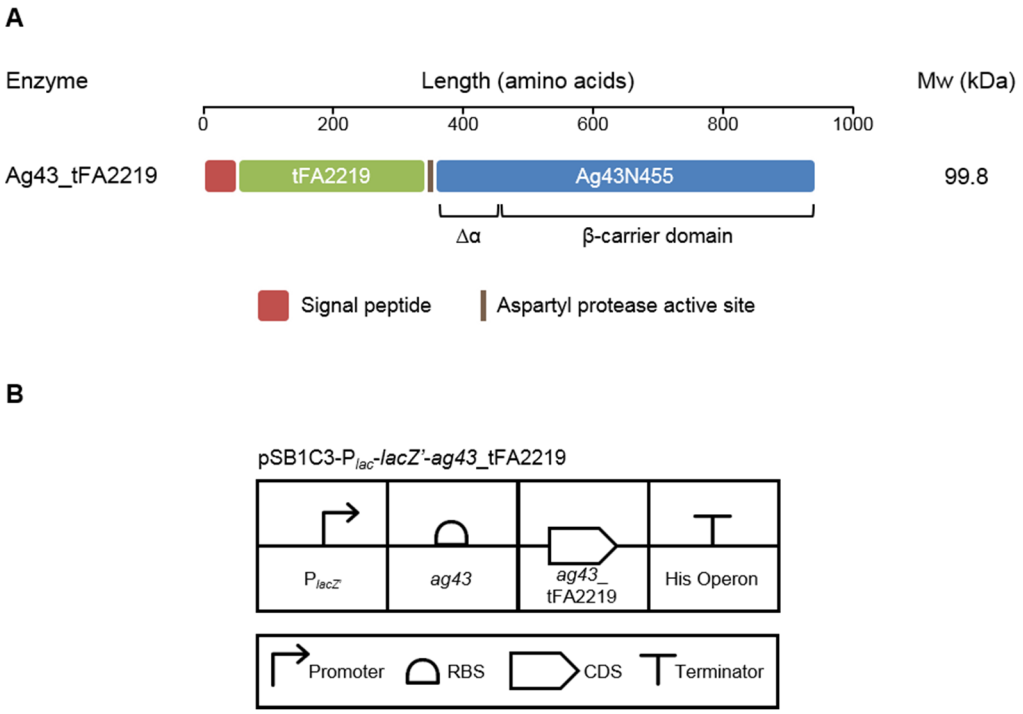

Figure 5-6 Ag43 secretion system for *F. agariphila*'s ulvan lyase. (A) Design of the Ag43-tFA2219 secretion system. The truncated ulvan lyase tFA2219 was introduced as a passenger between the signal peptide from Ag43 and the Ag43N455 domain. Additionally, an aspartyl protease active site from the original passage domain of Ag43 was added as a linker between the two domains. The Ag43N455 domain is composed by the β-carrier domain and a small portion of the α-passenger domain (Δα). (B) Construct for the secretion of the ulvan lyase using the Ag43 system. The construct was assembled using the PaperClip assembly (Trubitsyna et al., 2014). Ag43: antigen 43 from *E. coli*, tFA2219: truncated ulvan lyase without signal peptide from *F. agariphila*, P$_{lacZ'}$: P$_{lac}$-*lacZ'*. The constructs are described according to the Synthetic Biology Open Language (SBOL). RBS: ribosome binding site, CDS: coding DNA sequence.

In order to check whether the Ag43-based secretion system was working an ulvan lyase activity assay using live cells was performed. An ulvan-LB-agarose plate was inoculated with *E. coli* BL21 (DE3) cells carrying the construct pSB1C3-P$_{lac}$-*lacZ'*-*ag43*_tFA2219 and incubated at 37°C overnight. The construct pSB1C3-P$_{lac}$-*lacZ'* was used as a negative control. The plate was stained with cetyl pyridinium chloride to visualize ulvan degradation. A clear halo of activity was observed when the Ag43-based secretion system was used (Figure 5-7), indicating that the ulvan lyase is secreted. However, the cell growth seems to be adversely affected by the expression of the secretion system. This might be explained by the destabilization of the outer membrane due to the high level of insertions of the β-carrier domain (Narayanan & Chou, 2008).

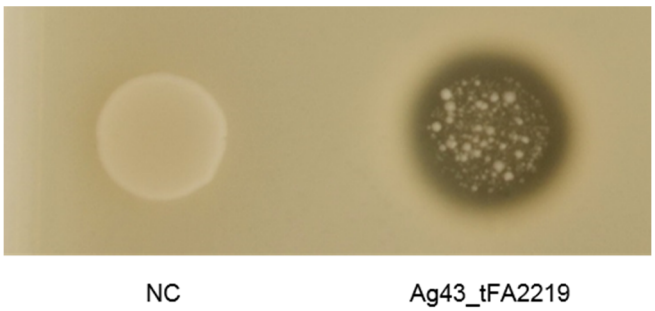

Figure 5-7 Ulvan lyase activity assay for the Ag43 secretion construct expressed by *E. coli*. *E. coli* BL21 (DE3) cells carrying the construct were used to inoculate an LB-agar plate containing 10g/l solubilized ulvan and 90 μg/ml IPTG. The plate was incubated at 37°C overnight and stained with 10% (w/v) cetyl pyridinium chloride. Unstained halos correspond to the areas where the ulvan has been degraded. NC: pSB1C3-P_{lac}-*lacZ'*.

5.2.4 Comparison of the different secretion systems

A total of three different constructs showed clear ulvanase activity: pSB1C3-P_{lac}-*lacZ'-osmY*_FlexL_tFA2219, pSB1C3-P_{lac}-*lacZ'-osmY*_CexL_tFA2219, and pSB1C3-P_{lac}-*lacZ'-ag43*_tFA2219. In order to compare the different constructs and determine the sub-cellular localization of the enzymes a quantitative activity assay using ulvan as the substrate was performed. *E. coli* BL21 (DE3) cells carrying the different constructs were used to inoculate 50 ml LB medium and the flasks incubated at 37°C with shaking until OD_{600} ~0.6 was reached. Protein expression was induced with 90 μg/ml IPTG under the same conditions for 20 h. Cells were harvested by centrifugation and the different sub-cellular fractions obtained. The protein extracts were mixed with solubilized ulvan in PBS and the reaction incubated at 37°C for 4 h. The DNS method was used to quantify the reduced sugar released. The results are shown in Figure 5-8. The samples from the extracellular fraction showed considerably higher activity than the ones from the other sub-cellular fractions when the three construct were used. These results indicate that the major fraction of the enzyme is efficiently secreted into the medium. Additionally, the levels of activity exhibited by the Ag43-based secretion system were markedly higher than the ones from the OsmY-based systems, confirming the results obtained from the plate activity assay. Moreover, the activity obtained from the extracellular sample of the culture expressing the construct pSB1C3-P_{lac}-*lacZ'-osmY*_CexL_tFA2219 was significantly higher than the sample expressing pSB1C3-P_{lac}-*lacZ'-osmY*_FlexL_tFA2219.

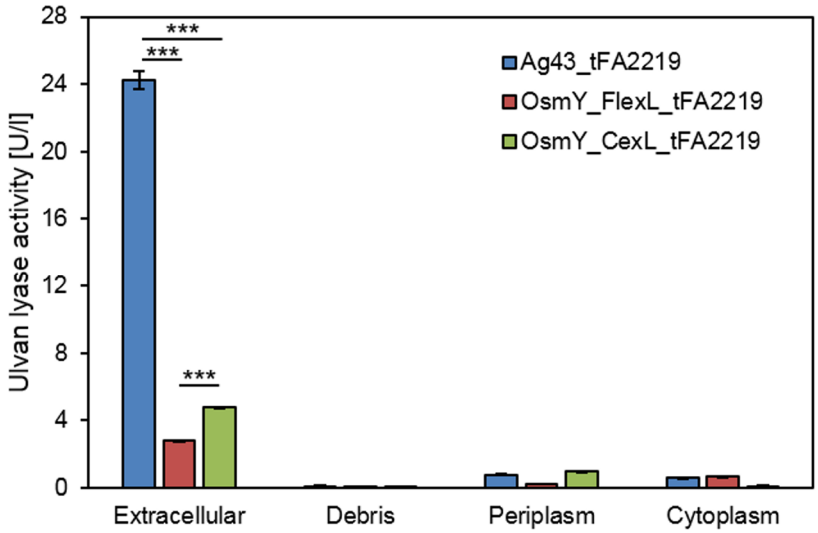

Figure 5-8 Ulvan lyase activity sub-cellular distribution profiles of the different secretion system. *E. coli* BL21 (DE3) cells carrying the different constructs were grown in LB at 37°C and used to obtain the sub-cellular fractions. The enzyme extracts were mixed with solubilized ulvan and the reactions carried out at 37°C. The DNS method was used to quantify the reducing sugars generated. Experiments were done in triplicate. Error bars indicate one standard error of the mean. *** $p < 0.001$.

In addition, 48 h time-course profiles of the growth and extracellular ulvan lyase activity were obtained using cultures carrying the constructs pSB1C3-P_{lac}-*lacZ'-osmY*_CexL_tFA2219 and pSB1C3-P_{lac}-*lacZ'-ag43*_tFA2219. *E. coli* BL21 (DE3) overnight cultures were used to inoculate 50 ml LB medium, the cultures were incubated at 37°C with shaking and 90 µg/ml IPTG used for induction after 3 h. The cultures were maintained under the same conditions for additional 45 h and samples collected after 6, 12, 24 and 48 h from the start. The growth curves are shown in Figure 5-10 (A). The cells carrying the OsmY-based secretion system showed a dramatically higher level of growth after 6 h of incubation, however, similar levels of growth were observed from the 12 h time point onward. To measure the ulvan lyase activity, the extracellular samples were mixed with solubilized ulvan and incubated at 37°C for 4 h. The DNS method was used to quantify the activity. Congruently with the scarce growth observed at the start of the culture, the Ag43-based secretion system showed less ulvan lyase activity than the OsmY-based system after 6 h of growth. Nevertheless, considerably higher levels of activity were reached using the Ag43-based secretion system in all the other time points, reaching a maximum difference of over 5-fold after 24 h of growth. These results suggest that even considering the

diminished growth obtained by the cells carrying the Ag43-based secretion system at the start of the culture, these can adapt, reaching normal levels of growth and considerably higher levels of ulvan lyase activity than when the OsmY-based system is used.

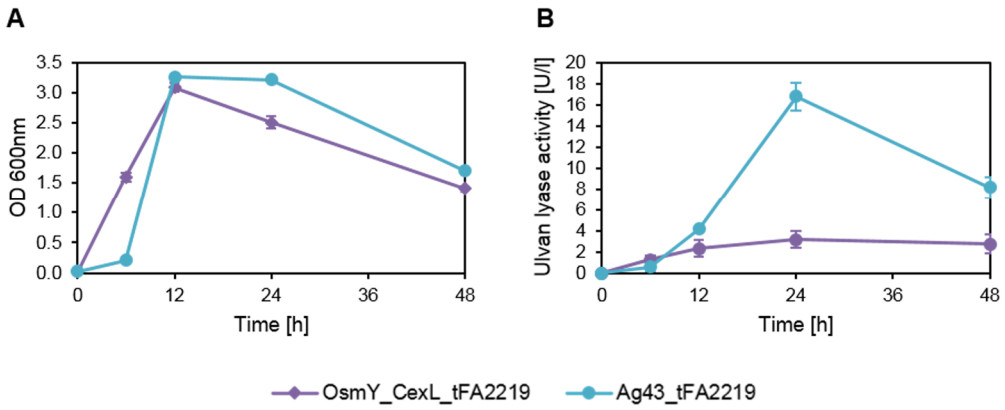

Figure 5-9 Comparison of Ag43 and OsmY ulvan lyase secretion systems in *E. coli*. (A) Growth curve over 48 h. (B) Time-course profile of extracellular ulvan lyase activity monitored over 48 h. Activity assays were carried out at for 4h at 37°C using extracellular and solubilized ulvan in PBS. DNS method was used to quantify the reducing sugars generated. *E. coli* BL21 (DE3) cells carrying the constructs were cultured in LB at 37°C. Experiments were done in triplicate. Error bars indicate one standard error of the mean.

5.3 Optimization of the secretion systems

5.3.1 Cleavage of the fusion proteins using a protease cleavage site

The utilization of different linkers is shown to have a great influence in the activity of the secreted fusion proteins, suggesting that physical interaction between the domains might be affecting their activity. In order to overcome this problem a system where the fusion proteins are cleaved in-vivo after the secretion without reliance on further treatment is required. The use of proteases naturally secreted by *E. coli* seems to be an interesting strategy to explore. A similar approach has been already reported by Fleetwood et al. (2014), where an autotransporter-based expression system was engineered for the secretion of Affibody molecules using a cleavage site for the outer membrane protease OmpT.

New secretion systems were designed using linkers containing an OmpT cleavage site between the tFA2219 and the secretion partners OsmY and CenA (Figure 5-10). The design of the linker containing the OmpT recognition site (OmpTL) was based on the peptide GGRRSRRVGT, identified by McCarter et al. (2004), which was shown to be readily cleaved by OmpT. An additional linker was designed using the same cleavage site but this time flanked by GGGGS flexible regions (OmpTfL). The constructs for the expression of these secretion systems were obtained using the PaperClip assembly. One part assemblies containing either the part tFA2219-pSB1C3-P$_{lac}$-*lacZ'-cenA* or the part tFA2219-pSB1C3-P$_{lac}$-*lacZ'-cenA* and a clip containing one of the OmpT-cleavable linkers as an intervening sequence were performed. The results of the assemblies were transformed into *E. coli* JM109 cells and the constructs validated by sequencing.

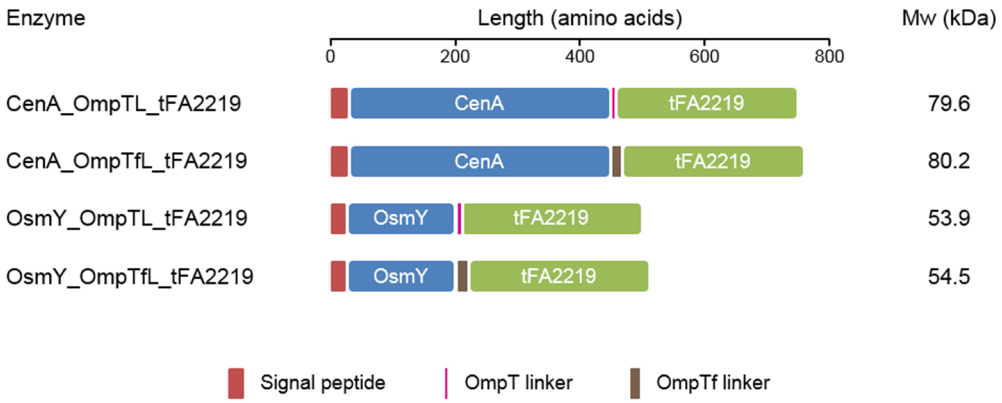

Figure 5-10 Design of fusion proteins for secretion of tFA2219 with linkers containing a protease cleavage site. Linkers were added as intervening regions between the domains using the PaperClip assembly (Trubitsyna et al., 2014). OmpT linker: linker containing an OmpT cleavage site, OmpTf linker: OmpT linker flanked by flexible regions, CenA: endoglucanase from *C. fimi*, OsmY: osmotically induced protein from *E. coli*, and tFA2219: truncated ulvan lyase without signal peptide from *F. agariphila*.

An ulvan lyase activity assay using live cells was performed in order to check whether the ulvan lyase is secreted by the new OsmY-based systems containing the OmpT recognition sequence. Considering that *E. coli* BL21 (DE3) is an OmpT$^-$ strain, *E. coli* TOP10 and *E. coli* JM109 were also used for the assay. The three strains were transformed with the constructs pSB1C3-P$_{lac}$-*lacZ'-osmY*_OmpTL_tFA2219 and pSB1C3-P$_{lac}$-*lacZ'-osmY*_OmpTfL_tFA2219 and overnight cultures used to inoculate an ulvan-LB-agarose plate. The construct pSB1C3-P$_{lac}$-*lacZ'-osmY*_FlexL_tFA2219

was used as a positive control. The plate was incubated at 37°C overnight and then stained with cetyl pyridinium chloride to reveal activity. The results are shown in Figure 5-11. Only two samples show halos of activity, the *E. coli* BL21 (DE3) cells expressing the positive control and when the same strain was used to express the OsmY_OmpTL_tFA2219 fusion protein. None of the *E. coli* $OmpT^+$ strains showed clear degradation of ulvan. These results confirm that *E. coli* BL21 (DE3) is a better chassis for the expression of the OsmY-based secretion system than the K12 strains. However, the apparent lack of secretion using the $OmpT^+$ strains prevented the study of the functionality of the OmpT-based cleavage system.

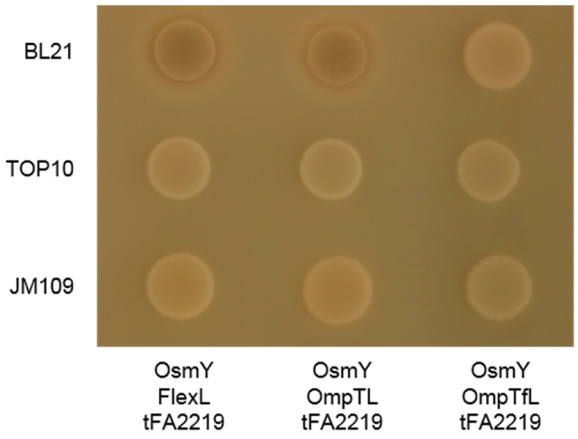

Figure 5-11 Ulvan lyase activity assay for the different OsmY secretion constructs containing the OmpT cleavage site. *E. coli* BL21 (DE3), TOP10 and JM109 cells carrying the different constructs were used to inoculate an LB-agar plate containing 10g/l solubilized ulvan and 90 μg/ml IPTG. The plate was incubated at 37°C overnight and stained with 10% (w/v) cetyl pyridinium chloride. The construct pSB1C3-P_{lac}-*lacZ'-osmY*_FlexL_tFA2219 was used as a positive control. Unstained halos correspond to the areas where the ulvan has been degraded.

The CenA-based secretion systems containing the OmpT cleavage site were also studied. Ulvan lyase and endoglucanase activity assays were performed. The same three *E. coli* strains were used for the expression of the constructs. The construct pSB1C3-P_{lac}-*lacZ'-cenA*_FlexL_tFA2219 was used as a positive control for the endoglucanase activity assay. Overnight cultures containing the different constructs were used to inoculate a CMC-LB-agar plate and the plate was then incubated at 37°C overnight. The plate was stained with Congo Red and washed with NaCl to reveal degradation of the CMC. The results are shown in Figure 5-12 (A). Clear halos of

degradation were observed associated with all the samples. Considerably higher levels of activity were obtained when *E. coli* BL21 (DE3) and *E. coli* TOP10 strains were used, suggesting that the secretion of CenA using these two strains is more efficient than with *E. coli* JM109. Additionally, the size of the halos obtained when the secretion system containing the OmpT linkers were expressed by $OmpT^{+}$ strains seem to be slightly bigger than the positive control, while the ones expressed in *E. coli* BL21 (DE3) appear smaller than the control. However, these results have to be validated with quantitative methods. The same procedure was performed for the ulvan lyase activity assay but this time using an ulvan-LB-agarose plate and cetyl pyridinium chloride to reveal the results. None of the cultures showed ulvan lyase activity (Figure 5-12 (B)), suggesting that the ulvan lyase domain is somehow affected when it is expressed as a fusion enzyme with endoglucanase CenA.

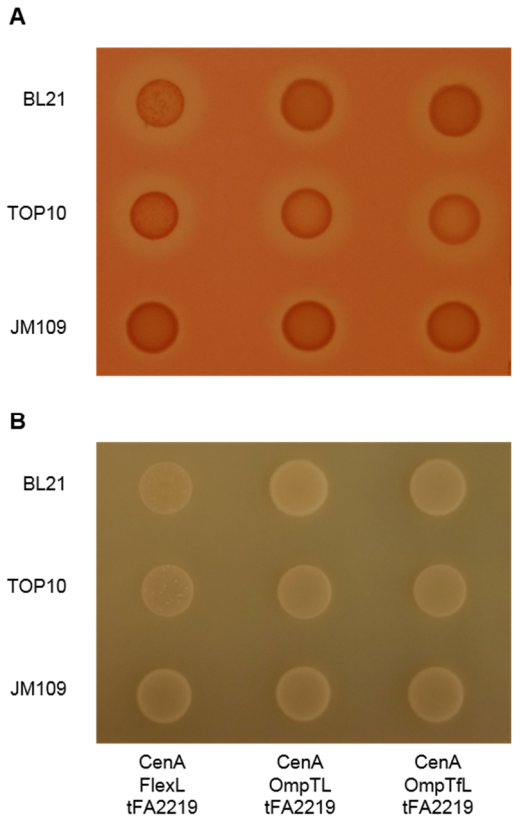

Figure 5-12 Activity assays for the different CenA secretion constructs containing the OmpT cleavage site. (A) CMC activity assay for endoglucanase activity. *E. coli* BL21 (DE3), TOP10 and JM109 cells carrying the constructs were used to inoculate an LB-agar plate containing 2 g/l CMC and 90 µg/ml IPTG. The plate was incubated at 37°C overnight, stained with Congo Red and washed with NaCl. Unstained halos correspond to areas where CMC has been hydrolysed. (B) Ulvan lyase activity assay. *E. coli* BL21 (DE3), TOP10 and JM109 cells

carrying the constructs were used to inoculate an LB-agar plate containing 10g/l solubilized ulvan and 90 μg/ml IPTG. The plate was incubated at 37°C overnight and stained with 10% (w/v) cetyl pyridinium chloride. Unstained halos correspond to the areas where the ulvan has been degraded.

5.3.2 RBSs of different strengths to improve Ag43 secretion system

In order to reduce the destabilization of the outer membrane caused by the over expression of the Ag43-based secretion system, the use of RBSs for lower transcription initiation rates was explored. Two additional RBSs were designed using the RBS calculator to be approximately 2 (Ag43 5k RBS) and 10 times (Ag43 1k RBS) weaker than the original one designed for the system (see Table 2-5 for the sequences). The new constructs are shown in Figure 5-13. The original RBS was replaced by the weaker ones using MABLE mutagenesis. The new vectors were used to transform *E. coli* JM109 cells and the construct validated by sequencing.

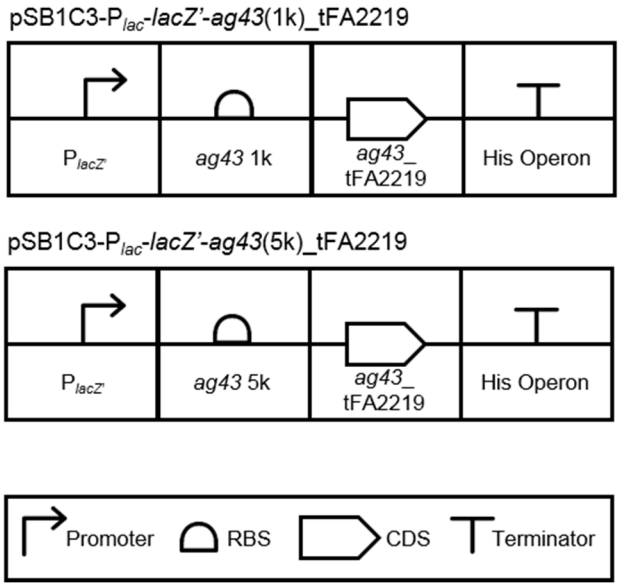

Figure 5-13 Ag43 constructs containing RBSs of different strengths for expression in *E. coli*. RBSs approximately 10 times (ag43 1k) and 2 times (ag43) weaker than the original RBS for the expression of Ag43_FA2219 were designed using the RBS calculator (Salis et al., 2009). The new constructs were generated replacing the original RBS using MABEL mutagenesis. $P_{lacZ'}$: P_{lac}-*lacZ'*. The constructs are described according to the Synthetic Biology Open Language (SBOL). RBS: ribosome binding site, CDS: coding DNA sequence.

An ulvan lyase activity assay using live cells was used to compare the constructs containing the RBSs of different strength. *E. coli* BL21 (DE3) cells were transformed

with the constructs for the three Ag43-based secretion systems. The construct pSB1C3-P_{lac}-*lacZ'* was used as a negative control. An ulvan-LB-agarose plate was inoculated with overnight cultures containing the different systems and incubated at 37°C overnight. The plate was stained with cetyl pyridinium chloride to observe ulvan degradation. Activity halos were observed associated with the cells expressing the three Ag43-based secretion systems (Figure 5-14). A negative correlation was found between growth levels and RBS's strength, with the highest level of growth observed in the cells containing the constructs with the weakest RBS. Interestingly, the activity levels did not show the same pattern. The strongest activity was observed in the cells expressing the construct containing the weakest RBS, whereas the cells with the medium RBS showed the lowest activity. Taking together, there seems to be a trade-off between growth and activity depending on the strength of the RBS. With the current data the weakest RBS is likely to be the best candidate for the secretion of the ulvan lyase.

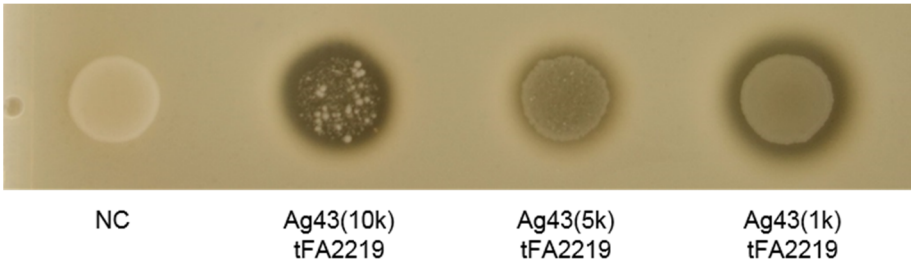

Figure 5-14 Ulvan lyase activity assay for the Ag43 secretion constructs using RBSs of different strength expressed by *E. coli*. *E. coli* BL21 (DE3) cells carrying the different constructs were used to inoculate an LB-agar plate containing 10g/l solubilized ulvan and 90 μg/ml IPTG. The plate was incubated at 37°C overnight and stained with 10% (w/v) cetyl pyridinium chloride. Unstained halos correspond to the areas where the ulvan has been degraded. Ag43 (10k) tFA2219: original Ag43-based secretion system, Ag43 (5k) tFA2219: construct containing an RBS two times weaker that the original one, Ag43 (1k) tFA2219: construct containing an RBS ten times weaker that the original one, NC: pSB1C3-P_{lac}-*lacZ'*.

5.4 Discussion

Consolidated bioprocessing of green macroalgal biomass require the secretion of an ulvan lyase to start the saccharification of the extracellular substrate. In this study three different strategies were analysed for the secretion of the ulvan lyase FA2219 from *F. agariphila* using *E. coli*. Two of them successfully secreted an active ulvan lyase, the

OsmY secretion partner and an engineered Ag43-based secretion system, whereas no extracellular ulvan lyase activity was achieved when the ulvan lyase was fused to the endoglucanase CenA from *C. fimi*.

The use of linker regions with different properties was explored for the design of the fusion partner secretion systems. In the case of the OsmY-based secretion systems, the two flexible linkers lead to higher levels of extracellular ulvan lyase activity than the construct containing the rigid linker, which showed no extracellular activity. These results could suggest that the flexible linkers facilitate the translocation of the fusion protein. Another possible explanation for the higher activity is that the ulvan lyase domain might require the degree of movement given by the presence of a flexible linker to maintain its activity. In the case of the CenA-based secretion systems, regardless of the linker used no extracellular ulvan lyase activity was detected. However, extracellular endoglucanase activity was observed when the systems were expressed, suggesting that the fusion proteins are actually secreted into the culture medium. Strong interaction between the two domains composing the fusion proteins preventing the proper folding and display of the catalytic domain of the ulvan lyase might be the cause of the lack of activity. As in the case of the OsmY-based systems, higher levels of endoglucanase activity were observed when the flexible linkers were utilized. When the subcellular fractions of the cells were analyzed similar levels of periplasmic activity were observed for the samples containing the flexible and rigid linkers. However, the extracellular activity was much higher when the flexible linker was used. These results support the idea that the use of a flexible linker might facilitate the extracellular translocation of the fusion enzymes.

The activity of the functional domain of a fusion protein can be diminished by steric hindrance or interference between the domains that it is composed of (Chen et al., 2013). In order to overcome the possible interactions between the domains of the secretion partner systems, linkers containing a cleavage site for the outer membrane protease OmpT were designed. No extracellular ulvan lyase activity was observed when the $OmpT^{+}$ strains *E. coli* JM109 and *E. coli* TOP10 were used for the secretion of OsmY-tFA2219, thus it was not possible to see whether the cleavage system was working. Low secretion of the OsmY fusion partner using *E. coli* K12 strains has been

already reported by Qian et al. (2008). No ulvan lyase activity was detected when the CenA-tFA2219 fusion proteins containing the cleavage linkers were expressed. However, similar levels of extracellular endoglucanase activity were obtained when CenA-based secretion system were expressed by *E. coli* BL21 (DE3) and *E. coli* TOP10, suggesting that the secretion mechanism used by CenA is different to the one used by OsmY. In spite of the negative results obtained using the endoglucanase CenA for the secretion of FA2219, the use of this enzyme as a secretion partner for other enzymes still has to be studied.

As a third strategy an engineered Ag43-based secretion system was successfully used for the secretion of the ulvan lyase FA2219. Nevertheless, low levels of growth were observed when *E. coli* BL21 (DE3) cells expressing this system were grown in solid medium. This was confirmed using liquid cultures, obtaining considerably lower growth rates at the start of the culture compared with the ones observed from cells expressing the OsmY-based systems. However, similar levels of growth were achieved later in the growth culture and extracellular ulvan lyase activity was more than 5 times higher than with any other of the analysed secretion systems. Additionally, the sub-cellular localization of the ulvan lyase was studied, showing that the major fraction of the enzyme was efficiently secreted into the medium. Finally, the utilization of weaker RBSs in order to reduce a possible destabilization of the outer membrane due to the high levels of insertions of the β-carrier domain was also analyzed. A RBS approximately 10 times weaker than the one originally used for the secretion system showed both better growth and higher levels of ulvan lyase secretion. Overall, the Ag43-based secretion system appeared to be the best strategy for the secretion of *F. agariphila's* ulvan lyase FA2219 using *E. coli*.

Considering *E. coli*'s apparent capability to utilize unsaturated uronic acids shown in chapter 4, the utilization of this ulvan lyase secretion system might be enough to allow *E. coli* cells to grow using ulvan as the sole carbon source. Nevertheless, this still has to be demonstrated.

The heterologous secretion of an ulvan lyase is just the first step for the design of an ulvan utilization module. Other enzymes such as rhamnosidases, xylosidases and sulfatases still have to be added to the system. Ulvan is the main polysaccharide from

green macroalgal biomass (Lahaye & Robic, 2007), and its complete saccharification is crucial to enable consolidated bioprocessing for production of biofuels or other high value bioproducts.

6 Conclusions and future prospects

6.1 Conclusions

The current study investigated the enzymatic depolymerization of green macroalgal biomass. To start with, a novel module for the saccharification of cellulose using *E. coli* was designed. The module includes only two enzymes: the cellodextrinase CHU2268 from *C. hutchinsonii* without its putative signal peptide, and the endoglucanase CenA from *C. fimi*. This module was able to degrade pure cellulose paper and could be potentially used for the degradation of the cellulose contained in green macroalgae or other cellulosic materials.

The Bacteroidetes *F. agariphila* was identified as an effective degrader of green macroalgal biomass. The induction of enzymes with ulvan lyase, rhamnosidase and xylosidase activities was demonstrated when green macroalgal biomass was present in the culture medium. An ulvan utilization PUL was found in *F. agariphila*'s genome; this allowed the identification of enzymes potentially useful for the degradation of ulvan. Moreover, a first model for the utilization of ulvan was built based on the enzymes comprising the ulvan utilization PUL from *F. agariphila*.

The identification of the ulvan utilization PUL of *F. agariphila* allowed the design of novel DNA parts for the depolymerisation of ulvan. These enzymes were expressed in *E. coli* and their activity characterized. An ulvan lyase (FA2219) and an unsaturated β-glucuronyl hydrolase (FA2222), enzymes already reported to be involved in the saccharification of ulvan, were found to be part of the PUL and their activities confirmed. Additionally, the activity of two rhamnosidases (FA2204 and FA2209) and a xylosidase (FA2213) was validated. This way it was possible to confirm, for the first time, the involvement of enzymes with rhamnosidase and xylosidase activities in the depolymerisation of ulvan. Possible synergies between the expressed enzymes were also studied; however, no synergies were observed when the enzymes were combined for the degradation of ulvan, suggesting that the sulfatases are crucial to achieve complete depolymerisation of this polysaccharide.

An engineered secretion system based on the use of the autotransporter Ag43 was used to successfully secrete the ulvan lyase FA2219 from *F. agariphila* in *E. coli*. Secretion was also achieved using the fusion partner OsmY; however, the levels of extracellular ulvan lyase activity were considerably lower than those obtained when the Ag43-based system was utilized. Moreover, it was possible to improve the Ag43-based system obtaining higher levels of growth and activity using RBSs of different strengths.

In conclusion, the results of this study increase our understanding of the processes involved in the biodegradation of green macroalgal biomass. This is an important step forward to be able to design a consolidated bioprocessing system for the bioconversion of this kind of biomass into valuable products like biofuels or other bioproducts.

6.2 Future prospects

The use of a two-enzyme cellulose degradation system designed in this study allowed *E. coli* to successfully utilize pure cellulose paper as its sole carbon source. However, it would be interesting to see if this system could also be used for the degradation of cellulosic materials with the potential to be utilized at an industrial scale. In particular, algal cellulose contains weaker hydrogen bonding than plant cellulose providing easier access for endoglucanases (Daroch et al., 2013), making it an attractive feedstock to be tested. In order to test this type of biomass as a possible substrate, cellulose purification from *U. lactuca* biomass would be required. Cellulases exhibit preferences toward different forms of cellulases (Daroch et al., 2013), thus the activity of the endoglucanase CenA against this form of cellulose will have to be checked. The levels of solubilization and the final products generated after incubation with CenA could be compared with those obtained after the hydrolysis of pure cellulose paper. In the event that CenA can effectively hydrolase the cellulose from *U. lactu*ca, a growth analysis using *E. coli* expressing the degradation system and green macroalgal cellulose as the sole carbon source could be performed.

Cellodextrin permeases are crucial for effective cellulose utilization in fungi and co-expression with a β-glucosidase can promote rapid growth of the yeast *S. cerevisiae* on cellodextrins (Galazka et al., 2010). In this study it was shown that *E. coli* is able to uptake cellodextrins at least as large as cellohexaose; however, the mechanism used

to do so remain unknown. It was proposed that the lactose permease LacY, responsible for cellobiose uptake in *E. coli* (Sekar et al., 2012), might also be involved in the uptake of longer cellodextrins. LacY capability to uptake cellodextrins could be checked using *E. coli* strains where the *lacY* gene has been knocked out. Considering the importance that cellodextrin permeases might have for the design of an efficient cellulose degradation system using *E. coli*, other permeases could also be assayed to check their cellodextrin uptake capability. Other potential cellodextrin permeases that could be studied are the ones from the cryptic cellobiose utilization operons *chb* and *asc* from *E. coli*.

In order to fully understand the enzymatic degradation of ulvan, characterization of several enzymes of the PUL is still required. These include all the glycoside hydrolases that did not show activity against the model substrates pNPR and MUX. Different conditions for their expression and other substrates have to be tested. Purification of the enzymes might be required to avoid background activity coming from the host used for the expression (e.g. β-glucuronidase activity from *E. coli*). In particular, desulfation of ulvan seems to be crucial for its complete saccharification. A total of 8 sulfatases were identified in the PUL, confirming the importance of this class of enzyme. These enzymes showed a high level of sequence diversity; however, it was not possible to infer their substrate specificities, and heterologous expression will be required for their full characterization. Moreover, activity assays to check whether the hypothetical protein FA2199 is part of a novel family of ulvan lyases have to be conducted.

The bioconversion of green macroalgal biomass into valuable bioproducts requires depolymerization of ulvan, its main polysaccharide. In order to do so, an ulvan depolymerisation module has to be designed. *E. coli* was proposed to be used as the host due to its ability to utilize all the monosaccharides that comprise ulvan. In this study, the secretion of an ulvan lyase using *E. coli* was demonstrated. This is the first step toward the design of an ulvan degradation module. Additionally, it was shown that *E. coli* is able to utilize the unsaturated β-glucuronyl residues generated from the cleavage of ulvan by the ulvan lyase, presumably through the action of a UGL (locus tag WP_001767851) classified as a GH105. Cloning and expression of this enzyme would be required to experimentally validate its activity against unsaturated β-

glucuronyl residues. Moreover, unsaturated β-glucuronyl hydrolase induction when *E. coli* is grown in presence of ulvan or unsaturated β-glucuronyl could also be tested. Taking into account *E. coli*'s capability to utilize unsaturated β-glucuronyl residues, the possibility of using *E. coli* expressing the ulvan lyase secretion system to grow on ulvan as the sole carbon source has to be explored. In addition, based on the studies of the degradation of ulvan, enzymes such as rhamnosidases, xylosidases and sulfatases have to be added to the ulvan degradation module in order to achieve complete depolymerization of this polysaccharide.

Finally, in order to complete the design of a CBP for the bioconversion of green macroalgal biomass, a pathway for the production of a valuable product has to be added to the system. Pathways for the production of biofuels like ethanol and n-butanol have already been incorporated into *E. coli* (Atsumi et al., 2008; Ingram et al., 1987; Inui et al., 2008), and can prove to be interesting alternatives for the system. Additionally, rhamnose, the main sugar present in ulvan, is a precursor for the chemical commodity 1,2-propanediol (Altaras & Cameron, 1999). Currently, 1,2-propanediol is produced using hazardous chemicals from propylene, a non-renewable resource (Saxena et al., 2010; Siebert & Wendisch, 2015). However, this process is not commercially feasible due to the price of rhamnose (Altaras & Cameron, 1999). Production of this diol by direct fermentation of rhamnose using *E. coli* has been already reported (Boronat & Aguilar, 1981). Considering this, the possibility of producing 1,2-propanediol from green macroalgal biomass may be an attractive alternative. The use of this kind of biomass for the production of 1,2-propanediol has been already proposed by van der Wal et al. (2013).

7 References

Abby, S.S., Cury, J., Guglielmini, J., Neron, B., Touchon, M., Rocha, E.P. 2016. Identification of protein secretion systems in bacterial genomes. *Sci Rep*, **6**, 23080.

Adams, J.M., Gallagher, J.A., Donnison, I.S. 2008. Fermentation study on *Saccharina latissima* for bioethanol production considering variable pre-treatments. *Journal of Applied Phycology*, **21**(5), 569-574.

Al-Hafedh, Y.S., Alam, A., Buschmann, A.H. 2015. Bioremediation potential, growth and biomass yield of the green seaweed, *Ulva lactuca* in an integrated marine aquaculture system at the Red Sea coast of Saudi Arabia at different stocking densities and effluent flow rates. *Reviews in Aquaculture*, **7**(3), 161-171.

Alper, H., Stephanopoulos, G. 2009. Engineering for biofuels: exploiting innate microbial capacity or importing biosynthetic potential? *Nat Rev Microbiol*, **7**(10), 715-23.

Altaras, N.E., Cameron, D.C. 1999. Metabolic engineering of a 1,2-propanediol pathway in *Escherichia coli*. *Appl Environ Microbiol*, **65**(3), 1180-5.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., Lipman, D.J. 1990. Basic local alignment search tool. *J Mol Biol*, **215**(3), 403-10.

Alves, A., Sousa, R.A., Reis, R.L. 2013. A practical perspective on ulvan extracted from green algae. *Journal of Applied Phycology*, **25**(2), 407-424.

Anupama, Ravindra, P. 2000. Value-added food: single cell protein. *Biotechnol Adv*, **18**(6), 459-79.

Atsumi, S., Cann, A.F., Connor, M.R., Shen, C.R., Smith, K.M., Brynildsen, M.P., Chou, K.J., Hanai, T., Liao, J.C. 2008. Metabolic engineering of *Escherichia coli* for 1-butanol production. *Metab Eng*, **10**(6), 305-11.

Aylott, M.J., Casella, E., Tubby, I., Street, N.R., Smith, P., Taylor, G. 2008. Yield and spatial supply of bioenergy poplar and willow short-rotation coppice in the UK. *New Phytol*, **178**(2), 358-70.

Baneyx, F., Mujacic, M. 2004. Recombinant protein folding and misfolding in *Escherichia coli*. *Nat Biotechnol*, **22**(11), 1399-408.

Barbanti, L., Grandi, S., Vecchi, A., Venturi, G. 2006. Sweet and fibre sorghum (Sorghum bicolor (L.) Moench), energy crops in the frame of environmental protection from excessive nitrogen loads. *European Journal of Agronomy*, **25**(1), 30-39.

Barbeyron, T., Lerat, Y., Sassi, J.-F., Le Panse, S., Helbert, W., Collén, P.N. 2011. *Persicivirga ulvanivorans* sp. nov., a marine member of the family Flavobacteriaceae that degrades ulvan from green algae. *Int J Syst Evol Microbiol*, **61**(8), 1899-1905.

Beena, K., Udgaonkar, J.B., Varadarajan, R. 2004. Effect of signal peptide on the stability and folding kinetics of maltose binding protein. *Biochemistry*, **43**(12), 3608-19.

Ben-Ari, T., Neori, A., Ben-Ezra, D., Shauli, L., Odintsov, V., Shpigel, M. 2014. Management of *Ulva lactuca* as a biofilter of mariculture effluents in IMTA system. *Aquaculture*, **434**, 493-498.

Bendtsen, J.D., Nielsen, H., Widdick, D., Palmer, T., Brunak, S. 2005. Prediction of twin-arginine signal peptides. *BMC Bioinformatics*, **6**, 167.

Bennett, G.N., San, K.Y. 2001. Microbial formation, biotechnological production and applications of 1,2-propanediol. *Appl Microbiol Biotechnol*, **55**(1), 1-9.

Berven, F.S., Flikka, K., Jensen, H.B., Eidhammer, I. 2004. BOMP: a program to predict integral beta-barrel outer membrane proteins encoded within genomes of Gram-negative bacteria. *Nucleic Acids Res*, **32**(Web Server issue), W394-9.

Berven, F.S., Karlsen, O.A., Straume, A.H., Flikka, K., Murrell, J.C., Fjellbirkeland, A., Lillehaug, J.R., Eidhammer, I., Jensen, H.B. 2006. Analysing the outer membrane subproteome of Methylococcus capsulatus (Bath) using proteomics and novel biocomputing tools. *Arch Microbiol*, **184**(6), 362-77.

Bobin-Dubigeon, C., Lahaye, M., Guillon, F., Barry, J.-L., Gallant, D.J. 1997. Factors limiting the biodegradation of *Ulva* sp cell-wall polysaccharides. *Journal of the Science of Food and Agriculture*, **75**(3), 341-351.

Bokinsky, G., Peralta-Yahya, P.P., George, A., Holmes, B.M., Steen, E.J., Dietrich, J., Lee, T.S., Tullman-Ercek, D., Voigt, C.A., Simmons, B.A., Keasling, J.D. 2011. Synthesis of three advanced biofuels from ionic liquid-pretreated switchgrass using engineered *Escherichia coli*. *Proc Natl Acad Sci U S A*, **108**(50), 19949-54.

Bolton, J.J., Robertson-Andersson, D.V., Shuuluka, D., Kandjengo, L. 2009. Growing Ulva (Chlorophyta) in integrated systems as a commercial crop for abalone feed in South Africa: a SWOT analysis. *Journal of Applied Phycology*, **21**(5), 575-583.

Boronat, A., Aguilar, J. 1981. Metabolism of L-fucose and L-rhamnose in *Escherichia coli*: differences in induction of propanediol oxidoreductase. *J Bacteriol*, **147**(1), 181-5.

Bruce, J.S., McLean, M.W., Long, W.F., Williamson, F.B. 1985. *Flavobacterium heparinum* 3-O-sulphatase for N-substituted glucosamine 3-O-sulphate. *Eur J Biochem*, **148**(2), 359-65.

Bruhn, A., Dahl, J., Nielsen, H.B., Nikolaisen, L., Rasmussen, M.B., Markager, S., Olesen, B., Arias, C., Jensen, P.D. 2011. Bioenergy potential of *Ulva lactuca*: biomass yield, methane production and combustion. *Bioresour Technol*, **102**(3), 2595-604.

Brun, E., Gans, P., Marion, D., Barras, F. 1995. Overproduction, purification and characterization of the cellulose-binding domain of the *Erwinia chrysanthemi* secreted endoglucanase EGZ. *Eur J Biochem*, **231**(1), 142-8.

Bruton, T., Lyons, H., Lerat, Y., Stanley, M., Rasmussen, M.B. 2009. A review of the potential of marine algae as a source of biofuel in Ireland. Sustainable Energy Ireland.

Cadoux, S., Riche, A.B., Yates, N.E., Machet, J.-M. 2012. Nutrient requirements of *Miscanthus* x *giganteus*: Conclusions from a review of published studies. *Biomass and Bioenergy*, **38**, 14-22.

Casini, A., MacDonald, J.T., De Jonghe, J., Christodoulou, G., Freemont, P.S., Baldwin, G.S., Ellis, T. 2014. One-pot DNA construction for synthetic biology: the Modular Overlap-Directed Assembly with Linkers (MODAL) strategy. *Nucleic Acids Res*, **42**(1), e7.

Caspi, R., Altman, T., Billington, R., Dreher, K., Foerster, H., Fulcher, C.A., Holland, T.A., Keseler, I.M., Kothari, A., Kubo, A., Krummenacker, M., Latendresse, M., Mueller, L.A., Ong, Q., Paley, S., Subhraveti, P., Weaver, D.S., Weerasinghe, D., Zhang, P., Karp, P.D. 2014. The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of Pathway/Genome Databases. *Nucleic Acids Res*, **42**(Database issue), D459-71.

Chen, X., Zaro, J.L., Shen, W.C. 2013. Fusion protein linkers: property, design and functionality. *Adv Drug Deliv Rev*, **65**(10), 1357-69.

Chiellini, F., Morelli, A. 2011. *Ulvan: a versatile platform of biomaterials from renewable resources*. INTECH Open Access Publisher.

Chung, C.T., Niemela, S.L., Miller, R.H. 1989. One-step preparation of competent *Escherichia coli*: transformation and storage of bacterial cells in the same solution. *Proc Natl Acad Sci U S A*, **86**(7), 2172-5.

Collen, P.N., Jeudy, A., Sassi, J.F., Groisillier, A., Czjzek, M., Coutinho, P.M., Helbert, W. 2014. A novel unsaturated beta-glucuronyl hydrolase involved in ulvan degradation unveils the versatility of stereochemistry requirements in family GH105. *J Biol Chem*, **289**(9), 6199-211.

Collen, P.N., Sassi, J.F., Rogniaux, H., Marfaing, H., Helbert, W. 2011. Ulvan lyases isolated from the *Flavobacteria Persicivirga* ulvanivorans are the first members of a new polysaccharide lyase family. *J Biol Chem*, **286**(49), 42063-71.

Daroch, M., Geng, S., Wang, G.Y. 2013. Recent advances in liquid biofuel production from algal feedstocks. *Applied Energy*, **102**, 1371-1381.

Delattre, C., Michaud, P., Keller, C., Elboutachfaiti, R., Beven, L., Courtois, B., Courtois, J. 2006. Purification and characterization of a novel glucuronan lyase from *Trichoderma* sp. GL2. *Appl Microbiol Biotechnol*, **70**(4), 437-43.

Eddy, S.R. 2011. Accelerated Profile HMM Searches. *PLoS Comput Biol*, **7**(10), e1002195.

Edgar, R.C. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res*, **32**(5), 1792-7.

Elboutachfaiti, R., Delattre, C., Petit, E., Michaud, P. 2011. Polyglucuronic acids: Structures, functions and degrading enzymes. *Carbohydrate Polymers*, **84**(1), 1-13.

Ellis, T., Adie, T., Baldwin, G.S. 2011. DNA assembly for synthetic biology: from parts to pathways and beyond. *Integr Biol (Camb)*, **3**(2), 109-18.

Engler, C., Kandzia, R., Marillonnet, S. 2008. A one pot, one step, precision cloning method with high throughput capability. *PLoS One*, **3**(11), e3647.

Field, R.A., Haines, A.H., Chrystal, E.J., Luszniak, M.C. 1991. Histidines, histamines and imidazoles as glycosidase inhibitors. *Biochem J*, **274 (Pt 3)**, 885-9.

Finn, R.D., Coggill, P., Eberhardt, R.Y., Eddy, S.R., Mistry, J., Mitchell, A.L., Potter, S.C., Punta, M., Qureshi, M., Sangrador-Vegas, A., Salazar, G.A., Tate, J., Bateman, A. 2016. The Pfam protein families database: towards a more sustainable future. *Nucleic Acids Res*, **44**(D1), D279-85.

Fischer, C.R., Klein-Marcuschamer, D., Stephanopoulos, G. 2008. Selection and optimization of microbial hosts for biofuels production. *Metab Eng*, **10**(6), 295-304.

Fleetwood, F., Andersson, K.G., Stahl, S., Lofblom, J. 2014. An engineered autotransporter-based surface expression vector enables efficient display of Affibody molecules on OmpT-negative *E. coli* as well as protease-mediated secretion in OmpT-positive strains. *Microb Cell Fact*, **13**, 179.

Fortes, M.D., Lüning, K. 1980. Growth rates of North Sea macroalgae in relation to temperature, irradiance and photoperiod. *Helgoländer Meeresuntersuchungen*, **34**(1), 15-29.

French, C.E. 2009. Synthetic biology and biomass conversion: a match made in heaven? *J R Soc Interface*, **6 Suppl 4**, S547-58.

Fry, S.C. 1988. *The growing plant cell wall: chemical and metabolic analysis*. Longman Scientific & Technical; Wiley, Harlow, Essex, England; New York.

Fujimoto, Z. 2013. Structure and function of carbohydrate-binding module families 13 and 42 of glycoside hydrolases, comprising a beta-trefoil fold. *Biosci Biotechnol Biochem*, **77**(7), 1363-71.

Fujimoto, Z., Jackson, A., Michikawa, M., Maehara, T., Momma, M., Henrissat, B., Gilbert, H.J., Kaneko, S. 2013. The structure of a *Streptomyces avermitilis* alpha-L-rhamnosidase reveals a novel carbohydrate-binding module CBM67 within the six-domain arrangement. *J Biol Chem*, **288**(17), 12376-85.

Gacesa, P., Wusteman, F.S. 1990. Plate assay for simultaneous detection of alginate lyases and determination of substrate specificity. *Appl Environ Microbiol*, **56**(7), 2265-7.

Galazka, J.M., Tian, C., Beeson, W.T., Martinez, B., Glass, N.L., Cate, J.H. 2010. Cellodextrin transport in yeast for improved biofuel production. *Science*, **330**(6000), 84-6.

Galbe, M., Zacchi, G. 2002. A review of the production of ethanol from softwood. *Appl Microbiol Biotechnol*, **59**(6), 618-28.

Gallego, A., Hospido, A., Moreira, M.T., Feijoo, G. 2011. Environmental assessment of dehydrated alfalfa production in Spain. *Resources, Conservation and Recycling*, **55**(11), 1005-1012.

Gao, D., Wang, S., Li, H., Yu, H., Qi, Q. 2015. Identification of a heterologous cellulase and its N-terminus that can guide recombinant proteins out of *Escherichia coli*. *Microb Cell Fact*, **14**, 49.

Gibson, D.G., Young, L., Chuang, R.Y., Venter, J.C., Hutchison, C.A., Smith, H.O. 2009. Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nature Methods*, **6**(5), 343-U41.

Guo, Z., Arfman, N., Ong, E., Gilkes, N.R., Kilburn, D.G., Warren, R.A.J., Miller, R.C. 1988. Leakage of Cellulomonas fimi cellulases from *Escherichia coli*. *FEMS Microbiology Letters*, **49**(2), 279-283.

Gupta, S., Adlakha, N., Yazdani, S.S. 2013. Efficient extracellular secretion of an endoglucanase and a beta-glucosidase in *E. coli*. *Protein Expression and Purification*, **88**(1), 20-25.

Haft, D.H., Selengut, J.D., White, O. 2003. The TIGRFAMs database of protein families. *Nucleic Acids Res*, **31**(1), 371-3.

Hanson, S.R., Best, M.D., Wong, C.H. 2004. Sulfatases: structure, mechanism, biological activity, inhibition, and synthetic utility. *Angew Chem Int Ed Engl*, **43**(43), 5736-63.

Harun, R., Danquah, M.K., Forde, G.M. 2010. Microalgal biomass as a fermentation feedstock for bioethanol production. *Journal of Chemical Technology & Biotechnology*, **85**(2), 199-203.

Heaton, E., Voigt, T., Long, S.P. 2004. A quantitative review comparing the yields of two candidate C4 perennial biomass crops in relation to nitrogen, temperature and water. *Biomass and Bioenergy*, **27**(1), 21-30.

Henderson, I.R., Owen, P. 1999. The major phase-variable outer membrane protein of *Escherichia coli* structurally resembles the immunoglobulin A1 protease class of exported protein and is regulated by a novel mechanism involving Dam and oxyR. *J Bacteriol*, **181**(7), 2132-41.

Hernández, I., Fernández-Engo, M.A., Pérez-Lloréns, J.L., Vergara, J.J. 2005. Integrated outdoor culture of two estuarine macroalgae as biofilters for dissolved nutrients from *Sparus auratus* waste waters. *Journal of Applied Phycology*, **17**(6), 557-567.

Hollants, J., Leliaert, F., De Clerck, O., Willems, A. 2013. What we can learn from sushi: a review on seaweed-bacterial associations. *FEMS Microbiol Ecol*, **83**(1), 1-16.

Hooper, L.V., Xu, J., Falk, P.G., Midtvedt, T., Gordon, J.I. 1999. A molecular sensor that allows a gut commensal to control its nutrient foundation in a competitive ecosystem. *Proc Natl Acad Sci U S A*, **96**(17), 9833-8.

Huffer, S., Roche, C.M., Blanch, H.W., Clark, D.S. 2012. *Escherichia coli* for biofuel production: bridging the gap from promise to practice. *Trends Biotechnol*, **30**(10), 538-45.

Ingram, L.O., Conway, T., Clark, D.P., Sewell, G.W., Preston, J.F. 1987. Genetic engineering of ethanol production in *Escherichia coli*. *Appl Environ Microbiol*, **53**(10), 2420-5.

Inui, M., Suda, M., Kimura, S., Yasuda, K., Suzuki, H., Toda, H., Yamamoto, S., Okino, S., Suzuki, N., Yukawa, H. 2008. Expression of *Clostridium acetobutylicum* butanol synthetic genes in *Escherichia coli*. *Appl Microbiol Biotechnol*, **77**(6), 1305-16.

Jeoh, T., Ishizawa, C.I., Davis, M.F., Himmel, M.E., Adney, W.S., Johnson, D.K. 2007. Cellulase digestibility of pretreated biomass is limited by cellulose accessibility. *Biotechnol Bioeng*, **98**(1), 112-22.

John, R.P., Anisha, G.S., Nampoothiri, K.M., Pandey, A. 2011. Micro and macroalgal biomass: a renewable source for bioethanol. *Bioresour Technol*, **102**(1), 186-93.

Juncker, A.S., Willenbrock, H., Von Heijne, G., Brunak, S., Nielsen, H., Krogh, A. 2003. Prediction of lipoprotein signal peptides in Gram-negative bacteria. *Protein Sci*, **12**(8), 1652-62.

Jung, K.A., Lim, S.-R., Kim, Y., Park, J.M. 2013. Potentials of macroalgae as feedstocks for biorefinery. *Bioresource Technology*, **135**, 182-190.

Kall, L., Krogh, A., Sonnhammer, E.L. 2004. A combined transmembrane topology and signal peptide prediction method. *J Mol Biol*, **338**(5), 1027-36.

Kanehisa, M., Sato, Y., Kawashima, M., Furumichi, M., Tanabe, M. 2016. KEGG as a reference resource for gene and protein annotation. *Nucleic Acids Res*, **44**(D1), D457-62.

Kelly, M., Dworjanyn, S. 2008. The potential of marine biomass for anaerobic biogas production: a feasibility study with recommendations for further research. *Scotland: Scottish Association for Marine Science.*

Kjaergaard, K., Hasman, H., Schembri, M.A., Klemm, P. 2002. Antigen 43-mediated autotransporter display, a versatile bacterial cell surface presentation system. *J Bacteriol*, **184**(15), 4197-204.

Klinke, H.B., Thomsen, A.B., Ahring, B.K. 2004. Inhibition of ethanol-producing yeast and bacteria by degradation products produced during pre-treatment of biomass. *Appl Microbiol Biotechnol*, **66**(1), 10-26.

Knight, T. 2003. Idempotent Vector Design for Standard Assembly of Biobricks. in: *MIT Synthetic Biology Working Group*, Massachusetts Institute of Technology.

Kopel, M., Helbert, W., Belnik, Y., Buravenkov, V., Herman, A., Banin, E. 2016. New Family of Ulvan Lyases Identified in Three Isolates from the Alteromonadales Order. *J Biol Chem*, **291**(11), 5871-8.

Kowalewski, B., Lamanna, W.C., Lawrence, R., Damme, M., Stroobants, S., Padva, M., Kalus, I., Frese, M.A., Lubke, T., Lullmann-Rauch, R., D'Hooge, R., Esko, J.D., Dierks, T. 2012. Arylsulfatase G inactivation causes loss of heparan sulfate 3-O-sulfatase activity and mucopolysaccharidosis in mice. *Proc Natl Acad Sci U S A*, **109**(26), 10310-5.

Kumar, S., Stecher, G., Tamura, K. 2016. MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. *Mol Biol Evol*, **33**(7), 1870-4.

Lahaye, M., Brunel, M., Bonnin, E. 1997. Fine chemical structure analysis of oligosaccharides produced by an ulvan-lyase degradation of the water-soluble cell-wall polysaccharides from *Ulva* sp, (Ulvales, Chlorophyta). *Carbohydr Res*, **304**(3-4), 325-33.

Lahaye, M., Jegou, D., Buleon, A. 1994. Chemical characteristics of insoluble glucans from the cell wall of the marine green alga *Ulva lactuca* (L.) Thuret. *Carbohydrate Research*, **262**(1), 115-125.

Lahaye, M., Robic, A. 2007. Structure and functional properties of ulvan, a polysaccharide from green seaweeds. *Biomacromolecules*, **8**(6), 1765-74.

Li, S., Yang, X., Bao, M., Wu, Y., Yu, W., Han, F. 2015. Family 13 carbohydrate-binding module of alginate lyase from *Agarivorans* sp. L11 enhances its catalytic efficiency and thermostability, and alters its substrate preference and product distribution. *FEMS Microbiol Lett*, **362**(10).

Linder, M., Teeri, T.T. 1997. The roles and function of cellulose-binding domains. *Journal of Biotechnology*, **57**(1), 15-28.

Liu, C.-K. 2012. Cellulose degradation system of *Cytophaga hutchinsonii*. in: *School of Biological Sciences*, Vol. PhD, University of Edinburgh. UK.

Lombard, V., Golaconda Ramulu, H., Drula, E., Coutinho, P.M., Henrissat, B. 2014. The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Res*, **42**(Database issue), D490-5.

Lowe, E.C., Basle, A., Czjzek, M., Firbank, S.J., Bolam, D.N. 2012. A scissor blade-like closing mechanism implicated in transmembrane signaling in a Bacteroides hybrid two-component system. *Proc Natl Acad Sci U S A*, **109**(19), 7298-303.

Lynd, L.R., van Zyl, W.H., McBride, J.E., Laser, M. 2005. Consolidated bioprocessing of cellulosic biomass: an update. *Curr Opin Biotechnol*, **16**(5), 577-83.

Lynd, L.R., Weimer, P.J., van Zyl, W.H., Pretorius, I.S. 2002. Microbial cellulose utilization: fundamentals and biotechnology. *Microbiol Mol Biol Rev*, **66**(3), 506-77, table of contents.

Macchiavello, J., Bulboa, C. 2014. Nutrient uptake efficiency of *Gracilaria chilensis* and *Ulva lactuca* in an IMTA system with the red abalone *Haliotis rufescens*. *Lat. Am. J. Aquat. Res.*, **42**(3), 523-533.

Macmillan, J.D., Phaff, H.J., Vaughn, R.H. 1964. The Pattern of Action of an Exopolygalacturonic Acid-Trans-Eliminase from *Clostridium Multifermentans*. *Biochemistry*, **3**, 572-8.

Mann, A.J., Hahnke, R.L., Huang, S., Werner, J., Xing, P., Barbeyron, T., Huettel, B., Stuber, K., Reinhardt, R., Harder, J., Glockner, F.O., Amann, R.I., Teeling, H. 2013. The genome of the alga-associated marine flavobacterium *Formosa agariphila* KMM 3901T reveals a broad potential for degradation of algal polysaccharides. *Appl Environ Microbiol*, **79**(21), 6813-22.

Marchler-Bauer, A., Anderson, J.B., Derbyshire, M.K., DeWeese-Scott, C., Gonzales, N.R., Gwadz, M., Hao, L., He, S., Hurwitz, D.I., Jackson, J.D., Ke, Z., Krylov, D., Lanczycki, C.J., Liebert, C.A., Liu, C., Lu, F., Lu, S., Marchler, G.H., Mullokandov, M., Song, J.S., Thanki, N., Yamashita, R.A., Yin, J.J., Zhang, D., Bryant, S.H. 2007. CDD: a conserved domain database for interactive domain family analysis. *Nucleic Acids Res*, **35**(Database issue), D237-40.

Martens, E.C., Lowe, E.C., Chiang, H., Pudlo, N.A., Wu, M., McNulty, N.P., Abbott, D.W., Henrissat, B., Gilbert, H.J., Bolam, D.N., Gordon, J.I. 2011. Recognition and degradation of plant cell wall polysaccharides by two human gut symbionts. *PLoS Biol*, **9**(12), e1001221.

McCarter, J.D., Stephens, D., Shoemaker, K., Rosenberg, S., Kirsch, J.F., Georgiou, G. 2004. Substrate specificity of the *Escherichia coli* outer membrane protease OmpT. *J Bacteriol*, **186**(17), 5919-25.

McKendry, P. 2002. Energy production from biomass (Part 1): Overview of biomass. *Bioresour Technol*, **83**(1), 37-46.

Mewis, K., Lenfant, N., Lombard, V., Henrissat, B. 2016. Dividing the Large Glycoside Hydrolase Family 43 into Subfamilies: a Motivation for Detailed Enzyme Characterization. *Appl Environ Microbiol*, **82**(6), 1686-92.

Miller, G.L. 1959. Use of Dinitrosalicylic Acid Reagent for Determination of Reducing Sugar. *Analytical Chemistry*, **31**(3), 426-428.

Montanier, C., van Bueren, A.L., Dumon, C., Flint, J.E., Correia, M.A., Prates, J.A., Firbank, S.J., Lewis, R.J., Grondin, G.G., Ghinet, M.G., Gloster, T.M., Herve, C., Knox, J.P., Talbot, B.G., Turkenburg, J.P., Kerovuo, J., Brzezinski, R., Fontes, C.M., Davies, G.J., Boraston, A.B., Gilbert, H.J. 2009. Evidence that family 35 carbohydrate binding modules display conserved specificity but divergent function. *Proc Natl Acad Sci U S A*, **106**(9), 3065-70.

Msuya, F.E., Neori, A. 2008. Effect of water aeration and nutrient load level on biomass yield, N uptake and protein content of the seaweed *Ulva lactuca* cultured in seawater tanks. *Journal of Applied Phycology*, **20**(6), 1021-1031.

Narayanan, N., Chou, C.P. 2008. Physiological improvement to enhance *Escherichia coli* cell-surface display via reducing extracytoplasmic stress. *Biotechnol Prog*, **24**(2), 293-301.

Neori, A., Krom, M.D., Ellner, S.P., Boyd, C.E., Popper, D., Rabinovitch, R., Davison, P.J., Dvir, O., Zuber, D., Ucko, M., Angel, D., Gordin, H. 1996. Seaweed biofilters as regulators of water quality in integrated fish-seaweed culture units. *Aquaculture*, **141**(3), 183-199.

Neori, A., Troell, M., Chopin, T., Yarish, C., Critchley, A., Buschmann, A.H. 2007. The Need for a Balanced Ecosystem Approach to Blue Revolution Aquaculture. *Environment: Science and Policy for Sustainable Development*, **49**(3), 36-43.

Neveux, N., Yuen, A.K., Jazrawi, C., Magnusson, M., Haynes, B.S., Masters, A.F., Montoya, A., Paul, N.A., Maschmeyer, T., de Nys, R. 2014. Biocrude yield and productivity from the hydrothermal liquefaction of marine and freshwater green macroalgae. *Bioresour Technol*, **155**, 334-41.

Ni, Y., Chen, R. 2009. Extracellular recombinant protein production from *Escherichia coli*. *Biotechnol Lett*, **31**(11), 1661-70.

Nielsen, M.M., Bruhn, A., Rasmussen, M.B., Olesen, B., Larsen, M.M., Møller, H.B. 2012. Cultivation of *Ulva lactuca* with manure for simultaneous bioremediation and biomass production. *Journal of Applied Phycology*, **24**(3), 449-458.

Ohno, M., Triet, V.D. 1997. Artificial seeding of the green seaweed *Monostroma* for cultivation. *Journal of Applied Phycology*, **9**(5), 417-423.

OSullivan, A.C. 1997. Cellulose: the structure slowly unravels. *Cellulose*, **4**(3), 173-207.

Patel, E.H., Paul, L.V., Patrick, S., Abratt, V.R. 2008. Rhamnose catabolism in *Bacteroides thetaiotaomicron* is controlled by the positive transcriptional regulator RhaR. *Res Microbiol*, **159**(9-10), 678-84.

Petersen, T.N., Brunak, S., von Heijne, G., Nielsen, H. 2011. SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nat Methods*, **8**(10), 785-6.

Poon, D.K., Withers, S.G., McIntosh, L.P. 2007. Direct demonstration of the flexibility of the glycosylated proline-threonine linker in the *Cellulomonas fimi* Xylanase Cex through NMR spectroscopic analysis. *J Biol Chem*, **282**(3), 2091-100.

Qian, Z.G., Xia, X.X., Choi, J.H., Lee, S.Y. 2008. Proteome-based identification of fusion partner for high-level extracellular production of recombinant proteins in *Escherichia coli*. *Biotechnol Bioeng*, **101**(3), 587-601.

Quan, J., Tian, J. 2009. Circular polymerase extension cloning of complex gene libraries and pathways. *PLoS One*, **4**(7), e6441.

Ravcheev, D.A., Godzik, A., Osterman, A.L., Rodionov, D.A. 2013. Polysaccharides utilization in human gut bacterium *Bacteroides thetaiotaomicron*: comparative genomics reconstruction of metabolic and regulatory networks. *BMC Genomics*, **14**, 873.

Redouan, E., Cedric, D., Emmanuel, P., Mohamed el, G., Bernard, C., Philippe, M., Cherkaoui el, M., Josiane, C. 2009. Improved isolation of glucuronan from algae and the production of glucuronic acid oligosaccharides using a glucuronan lyase. *Carbohydr Res*, **344**(13), 1670-5.

Rodionov, D.A., Gelfand, M.S., Hugouvieux-Cotte-Pattat, N. 2004. Comparative genomics of the KdgR regulon in *Erwinia chrysanthemi* 3937 and other gamma-proteobacteria. *Microbiology*, **150**(Pt 11), 3571-90.

Romanelli, T.L., Milan, M., Tieppo, R.C. 2012. Energy-based evaluations on eucalyptus biomass production. *International Journal of Forestry Research*, **2012**.

Romine, M.F. 2011. Genome-wide protein localization prediction strategies for gram negative bacteria. *BMC Genomics*, **12 Suppl 1**, S1.

Rutter, C., Mao, Z., Chen, R. 2013. Periplasmic expression of a *Saccharophagus* cellodextrinase enables *E. coli* to ferment cellodextrin. *Appl Microbiol Biotechnol*, **97**(18), 8129-38.

Ryther, J., DeBusk, T., Blakeslee, M. 1984. Cultivation and conversion of marine macroalgae. *Final Subcontract Report*.

Ryu, K.S., Kim, J.I., Cho, S.J., Park, D., Park, C., Cheong, H.K., Lee, J.O., Choi, B.S. 2005. Structural insights into the monosaccharide specificity of *Escherichia coli* rhamnose mutarotase. *J Mol Biol*, **349**(1), 153-62.

Sadie, C.J., Rose, S.H., den Haan, R., van Zyl, W.H. 2011. Co-expression of a cellobiose phosphorylase and lactose permease enables intracellular cellobiose utilisation by *Saccharomyces cerevisiae. Appl Microbiol Biotechnol*, **90**(4), 1373-80.

Salis, H.M., Mirsky, E.A., Voigt, C.A. 2009. Automated design of synthetic ribosome binding sites to control protein expression. *Nat Biotechnol*, **27**(10), 946-50.

Sambrook, J., Russell, D.W. 2001. *Molecular cloning : a laboratory manual*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.

Sarrion-Perdigones, A., Falconi, E.E., Zandalinas, S.I., Juarez, P., Fernandez-del-Carmen, A., Granell, A., Orzaez, D. 2011. GoldenBraid: an iterative cloning system for standardized assembly of reusable genetic modules. *PLoS One*, **6**(7), e21622.

Sato, K., Naito, M., Yukitake, H., Hirakawa, H., Shoji, M., McBride, M.J., Rhodes, R.G., Nakayama, K. 2010. A protein secretion system linked to bacteroidete gliding motility and pathogenesis. *Proc Natl Acad Sci U S A*, **107**(1), 276-81.

Saxena, R.K., Anand, P., Saran, S., Isar, J., Agarwal, L. 2010. Microbial production and applications of 1,2-propanediol. *Indian J Microbiol*, **50**(1), 2-11.

Sekar, R., Shin, H.-D., Chen, R. 2012. Engineering Escherichia coli cells for cellobiose assimilation through a phosphorolytic mechanism. *Applied and environmental microbiology*, **78**(5), 1611-1614.

Seppälä, M., Paavola, T., Lehtomäki, A., Pakarinen, O., Rintala, J. 2008. Biogas from energy crops—optimal pre-treatments and storage, co-digestion and energy balance in boreal conditions. *Water Science and Technology*, **58**(9), 1857-1863.

Shin, H.D., Wu, J., Chen, R. 2014. Comparative engineering of *Escherichia coli* for cellobiose utilization: Hydrolysis versus phosphorolysis. *Metab Eng*, **24**, 9-17.

Sims, R.E., Mabee, W., Saddler, J.N., Taylor, M. 2010. An overview of second generation biofuel technologies. *Bioresour Technol*, **101**(6), 1570-80.

Singh, P., Sharma, L., Kulothungan, S.R., Adkar, B.V., Prajapati, R.S., Ali, P.S., Krishnan, B., Varadarajan, R. 2013. Effect of signal peptide on stability and folding of *Escherichia coli* thioredoxin. *PLoS One*, **8**(5), e63442.

Smetacek, V., Zingone, A. 2013. Green and golden seaweed tides on the rise. *Nature*, **504**(7478), 84-8.

Solomon, B.D., Barnes, J.R., Halvorsen, K.E. 2007. Grain and cellulosic ethanol: History, economics, and energy policy. *Biomass and Bioenergy*, **31**(6), 416-425.

Sonnenburg, E.D., Zheng, H., Joglekar, P., Higginbottom, S.K., Firbank, S.J., Bolam, D.N., Sonnenburg, J.L. 2010. Specificity of polysaccharide use in intestinal bacteroides species determines diet-induced microbiota alterations. *Cell*, **141**(7), 1241-52.

Sonnhammer, E.L., von Heijne, G., Krogh, A. 1998. A hidden Markov model for predicting transmembrane helices in protein sequences. *Proc Int Conf Intell Syst Mol Biol*, **6**, 175-82.

Sorensen, H.R., Meyer, A.S., Pedersen, S. 2003. Enzymatic hydrolysis of water-soluble wheat arabinoxylan. 1. Synergy between alpha-L-arabinofuranosidases, endo-1,4-beta-xylanases, and beta-xylosidase activities. *Biotechnol Bioeng*, **81**(6), 726-31.

Stalbrand, H., Mansfield, S.D., Saddler, J.N., Kilburn, D.G., Warren, R.A., Gilkes, N.R. 1998. Analysis of molecular size distributions of cellulose molecules during hydrolysis of cellulose by recombinant *Cellulomonas fimi* beta-1,4-glucanases. *Appl Environ Microbiol*, **64**(7), 2374-9.

Stanier, R.Y. 1942. The Cytophaga Group: A Contribution to the Biology of Myxobacteria. *Bacteriol Rev*, **6**(3), 143-96.

Sun, Y., Cheng, J. 2002. Hydrolysis of lignocellulosic materials for ethanol production: a review. *Bioresour Technol*, **83**(1), 1-11.

Taherzadeh, M.J., Karimi, K. 2007. Enzymatic-based hydrolysis processes for ethanol from lignocellulosic materials: A review. *BioResources*, **2**(4), 707-738.

Tan, K.T., Lee, K.T., Mohamed, A.R. 2008. Role of energy policy in renewable energy accomplishment: The case of second-generation bioethanol. *Energy Policy*, **36**(9), 3360-3365.

Tanaka, T., Kawabata, H., Ogino, C., Kondo, A. 2011. Creation of a cellooligosaccharide-assimilating *Escherichia coli* strain by displaying active beta-glucosidase on the cell surface via a novel anchor protein. *Appl Environ Microbiol*, **77**(17), 6265-70.

Taylor, R., Fletcher, R.L., Raven, J.A. 2001. Preliminary studies on the growth of selected 'Green tide' algae in laboratory culture: Effects of irradiance, temperature, salinity and nutrients on growth rate. *Botanica Marina*, **44**(4), 327-336.

Teather, R.M., Wood, P.J. 1982. Use of Congo red-polysaccharide interactions in enumeration and characterization of cellulolytic bacteria from the bovine rumen. *Appl Environ Microbiol*, **43**(4), 777-80.

Terpe, K. 2003. Overview of tag protein fusions: from molecular and biochemical fundamentals to commercial systems. *Appl Microbiol Biotechnol*, **60**(5), 523-33.

Terrapon, N., Lombard, V., Gilbert, H.J., Henrissat, B. 2015. Automatic prediction of polysaccharide utilization loci in Bacteroidetes species. *Bioinformatics*, **31**(5), 647-55.

Tomme, P., Warren, R.A., Gilkes, N.R. 1995. Cellulose hydrolysis by bacteria and fungi. *Adv Microb Physiol*, **37**, 1-81.

Trubitsyna, M., Liu, C.K., Salinas, A., Elfick, A., French, C.E. 2016. PaperClip: a simple method for flexible multi-part DNA in: *Synthetic DNA: Methods and Protocols*, Humana Press.

Trubitsyna, M., Michlewski, G., Cai, Y., Elfick, A., French, C.E. 2014. PaperClip: rapid multi-part DNA assembly from existing libraries. *Nucleic Acids Res*, **42**(20), e154.

Ulmer, J.E., Vilen, E.M., Namburi, R.B., Benjdia, A., Beneteau, J., Malleron, A., Bonnaffe, D., Driguez, P.A., Descroix, K., Lassalle, G., Le Narvor, C., Sandstrom, C., Spillmann, D., Berteau, O. 2014. Characterization of glycosaminoglycan (GAG) sulfatases from the human gut symbiont *Bacteroides thetaiotaomicron* reveals the first GAG-specific bacterial endosulfatase. *J Biol Chem*, **289**(35), 24289-303.

UniProt, C. 2015. UniProt: a hub for protein information. *Nucleic Acids Res*, **43**(Database issue), D204-12.

van der Wal, H., Sperber, B.L., Houweling-Tan, B., Bakker, R.R., Brandenburg, W., Lopez-Contreras, A.M. 2013. Production of acetone, butanol, and ethanol from biomass of the green seaweed *Ulva lactuca*. *Bioresour Technol*, **128**, 431-7.

Varki, A., Cummings, R.D., Aebi, M., Packer, N.H., Seeberger, P.H., Esko, J.D., Stanley, P., Hart, G., Darvill, A., Kinoshita, T., Prestegard, J.J., Schnaar, R.L., Freeze, H.H., Marth, J.D., Bertozzi, C.R., Etzler, M.E., Frank, M., Vliegenthart, J.F., Lutteke, T., Perez, S., Bolton, E., Rudd, P., Paulson, J., Kanehisa, M., Toukach, P., Aoki-Kinoshita, K.F., Dell, A., Narimatsu, H., York, W., Taniguchi, N., Kornfeld, S. 2015. Symbol Nomenclature for Graphical Representations of Glycans. *Glycobiology*, **25**(12), 1323-4.

Vinuselvi, P., Lee, S.K. 2011. Engineering *Escherichia coli* for efficient cellobiose utilization. *Appl Microbiol Biotechnol*, **92**(1), 125-32.

Wakarchuk, W., Kilburn, D., Miller Jr, R., Warren, R. 1984. The preliminary characterization of the β-glucosidases of *Cellulomonas fimi*. *Microbiology*, **130**(6), 1385-1389.

Wargacki, A.J., Leonard, E., Win, M.N., Regitsky, D.D., Santos, C.N., Kim, P.B., Cooper, S.R., Raisner, R.M., Herman, A., Sivitz, A.B., Lakshmanaswamy, A., Kashiyama, Y., Baker, D., Yoshikuni, Y. 2012. An engineered microbial platform for direct biofuel production from brown macroalgae. *Science*, **335**(6066), 308-13.

Weber, E., Engler, C., Gruetzner, R., Werner, S., Marillonnet, S. 2011. A modular cloning system for standardized assembly of multigene constructs. *PLoS One*, **6**(2), e16765.

Wilson, D.B. 2009. Evidence for a novel mechanism of microbial cellulose degradation. *Cellulose*, **16**(4), 723-727.

Xie, G., Bruce, D.C., Challacombe, J.F., Chertkov, O., Detter, J.C., Gilna, P., Han, C.S., Lucas, S., Misra, M., Myers, G.L., Richardson, P., Tapia, R., Thayer, N., Thompson, L.S., Brettin, T.S., Henrissat, B., Wilson, D.B., McBride, M.J. 2007. Genome sequence of the cellulolytic gliding bacterium *Cytophaga hutchinsonii*. *Appl Environ Microbiol*, **73**(11), 3536-46.

Yanagisawa, M., Nakamura, K., Ariga, O., Nakasaki, K. 2011. Production of high concentrations of bioethanol from seaweeds that contain easily hydrolyzable polysaccharides. *Process Biochemistry*, **46**(11), 2111-2116.

Yin, Y., Mao, X., Yang, J., Chen, X., Mao, F., Xu, Y. 2012. dbCAN: a web resource for automated carbohydrate-active enzyme annotation. *Nucleic Acids Res*, **40**(Web Server issue), W445-51.

Zhang, Y., Werling, U., Edelmann, W. 2012. SLiCE: a novel bacterial cell extract-based DNA cloning method. *Nucleic Acids Res*, **40**(8), e55.

Zhang, Y.H., Himmel, M.E., Mielenz, J.R. 2006. Outlook for cellulase improvement: screening and selection strategies. *Biotechnol Adv*, **24**(5), 452-81.

Zheng, Z., Chen, T., Zhao, M., Wang, Z., Zhao, X. 2012. Engineering *Escherichia coli* for succinate production from hemicellulose via consolidated bioprocessing. *Microb Cell Fact*, **11**, 37.

Zhu, Y., Li, H., Zhou, H., Chen, G., Liu, W. 2010. Cellulose and cellodextrin utilization by the cellulolytic bacterium *Cytophaga hutchisonii*. *Bioresour Technol*, **101**(16), 6432-7.

8 Appendix

8.1 Additional methodology

8.1.1 List of primers and oligonucleotides

Table 8-1 List of primers and oligonucleotides used for PaperClip

Name	Target*	Sequence (5'→3')
Used to get the parts		
PC_pSB1C3md1_UF	pSB1C3	GCCACTAGTAAATTTATCTGCAGTCTATCAAAAAAGGG
PC_pSB1C3md1_UR	pSB1C3	TTTTTTGATAGACTGCAGATAAATTTACTAGT
PC_pSB1C3md1_DF	pSB1C3	TTAGCTTTCGCTAAGGATGATTTCTGGAATTC
PC_pSB1C3md1_DR	pSB1C3	GGCGAATTCCAGAAATCATCCTTAGCGAAAGC
PC_lacZ_UF	P_{lac} and LacZ'	GCCTCTAGATTCGGAGTGAGCGCAACGCAATTAATGTGAGTTAGCTCAC
PC_lacZ_UR	P_{lac} and LacZ'	AGCTAACTCACATTAATTGCGTTGCGCTCACTCCGAATCTAGA
PC_lacZb_DF	P_{lac} and LacZ'	CCAGAAGCGGTGGCCGGAAAGCTGGCTGGATGAGCCTGGATCC
PC_lacZb_DR	P_{lac} and LacZ'	GGCGGATCCAGGCTCATCCAGCCAGCTTTCCGGCCACCGCTTC
PC_FA-2190_UF	FA-2190	GCCATGAAGTCTGAAACTAAACAGAATAAAAAAGAAGAACAG
PC_FA-2190_UR	FA-2190	TTCTTCTTTTTTATTCTGTTTAGTTTCAGACTTCAT
PC_FA-2190_DF	FA-2190	GAAGCATTAAAAAGAAAACAAAATCTAATTTTAAAATAA
PC_FA-2190_DR	FA-2190	GGCTTATTTTAAAATTAGATTTTGTTTTCTTTTTAATGC
PC_FA-2194_UF	FA-2194	GCCATGAGCACAAAGTATGAGAGTCGTTATGCCTCAAGC
PC_FA-2194_UR	FA-2194	TGAGGCATAACGACTCTCATACTTTGTGCTCAT
PC_FA-2194_DF	FA-2194	TAACGATATGGATGTTGCAAAAATAACAGAATTAAGATAA
PC_FA-2194_DR	FA-2194	GGCTTATCTTAATTCTGTTATTTTTGCAACATCCATATCG
PC_FA-2195_UF	FA-2195	GCCATGAGTGTAGATTTATTTGATGTAAAAGGTAAAATAGCC

Name	Target*	Sequence (5'→3')
PC_FA-2195_UR	FA-2195	TATTTTACCTTTTACATCAAATAAATCTACACTCAT
PC_FA-2195_DF	FA-2195	TTAGCAACCATAGGTAAGCCTTCAAACGAAGAGTAA
PC_FA-2195_DR	FA-2195	GGCTTACTCTTCGTTTGAAGGCTTACCTATGGTTGC
PC_FA-2197_UF	FA-2197	GCCATGGGTTTTTGTATGAAAGATAGTAAACAATATTATAAATC
PC_FA-2197_UR	FA-2197	TTATAATATTGTTTACTATCTTTCATACAAAAACCCAT
PC_FA-2197_DF	FA-2197	AATCAAAATTTTAAAGGTACCTATTTAACCATAGAATAG
PC_FA-2197_DR	FA-2197	GGCCTATTCTATGGTTAAATAGGTACCTTTAAAATTTTG
PC_FA-2204_UF	FA-2204	GCCATGCAAAACTTAGTACCAGAAGTAACTGAGTTAGAATCGG
PC_FA-2204_UR	FA-2204	ATTCTAACTCAGTTACTTCTGGTACTAAGTTTTGCAT
PC_FA-2204_DF	FA-2204	GGGCCGGTTAAAATAAATGTGTTAGTAGAAAAATCGTTATAA
PC_FA-2204_DR	FA-2204	GGCTTATAACGATTTTTCTACTAACACATTTATTTTAACCGG
PC_FA-2205_UF	FA-2205	GCCATGCAACGTATAGAAACCAATTTCAACAATAATTGGC
PC_FA-2205_UR	FA-2205	AATTATTGTTGAAATTGGTTTCTATACGTTGCAT
PC_FA-2205_DF	FA-2205	GAGAGAGTGGGCCTTTTATTGATGAACTATTTATAGATTAA
PC_FA-2205_DR	FA-2205	GGCTTAATCTATAAATAGTTCATCAATAAAAGGCCCACTCT
PC_FA-2206_UF	FA-2206	GCCATGACCATTTTATGTTCTTGTGCACAGCAGCAGCAG
PC_FA-2206_UR	FA-2206	CTGCTGCTGTGCACAAGAACATAAAATGGTCAT
PC_FA-2206_DF	FA-2206	ATTGTAGAATGTATACCGTTATGGTCAAGGAATTTAAATAG
PC_FA-2206_DR	FA-2206	GGCCTATTTAAATTCCTTGACCATAACGGTATACATTCTAC
PC_FA-2209_UF	FA-2209	GCCATGTCTTGCGAAAATTCTTCTGTTTTACAAGAGGCAAAAC
PC_FA-2209_UR	FA-2209	TTGCCTCTTGTAAAACAGAAGAATTTTCGCAAGACAT
PC_FA-2209_DF	FA-2209	CAACCGGGTACCTATGAATTTCAAGCAAAATATAGTTTATAA
PC_FA-2209_DR	FA-2209	GGCTTATAAACTATATTTTGCTTGAAATTCATAGGTACCCGG

Name	Target*	Sequence (5'→3')
PC_FA-2213b_UF	FA-2213	GCCATGCAAAATAATGCGCAAACAAAATCTAATTCTGATGAAG
PC_FA-2213b_UR	FA-2213	CATCAGAATTAGATTTTGTTTGCGCATTATTTTGCAT
PC_FA-2213b_DF	FA-2213	AGCTGATGATTTAAAAACAGTGTCGTTAACTGTAAAATAA
PC_FA-2213b_DR	FA-2213	GGCTTATTTTACAGTTAACGACACTGTTTTTAAATCATCA
PC_FA-2216_UF	FA-2216	GCCATGGAGCCCGAAAAAAAACAAGCGACACCTTCAAAAGAAATAG
PC_FA-2216_UR	FA-2216	TTTCTTTTGAAGGTGTCGCTTGTTTTTTTTCGGGCTCCAT
PC_FA-2216_DF	FA-2216	TACCTATGATTATGACGGGCCAAAGGCCGCTAATCAATAA
PC_FA-2216_DR	FA-2216	GGCTTATTGATTAGCGGCCTTTGGCCCGTCATAATCATAG
PC_FA-2217b_UF	FA-2217	GCCATGGAACAAAAGTTACAAACATCATTAGATTTTGACAAG
PC_FA-2217b_UR	FA-2217	GTCAAAATCTAATGATGTTTGTAACTTTTGTTCCAT
PC_FA-2217b_DF	FA-2217	TGGGTTCTGGTAAGTATAAAATTCTATACAATAAAAGATAA
PC_FA-2217b_DR	FA-2217	GGCTTATCTTTTATTGTATAGAATTTTATACTTACCAGAAC
PC_FA2219SP_UF	FA-2219	GCCATGCTAGAAAAAACTACTCTTAAAAATATTATTTTAATAC
PC_FA2219SP_UR	FA-2219	TTAAAATAATATTTTTAAGAGTAGTTTTTTCTAGCAT
PC_FA2219woSP_UF	FA2219ΔSP	GCCGCGCCCGATGAAGATACATCTGCTATTACAAGATGTAC
PC_FA2219woSP_UR	FA2219ΔSP	CATCTTGTAATAGCAGATGTATCTTCATCGGGCGC
PC_FA2219woSPSC_UF	FA2219ΔSP::SC	GCCATGGCGCCCGATGAAGATACATCTGCTATTACAAGATG
PC_FA2219woSPSC_UR	FA2219ΔSP::SC	CTTGTAATAGCAGATGTATCTTCATCGGGCGCCAT
PC_FA-2219b_DF	FA-2219	TAACAACAAGTATCATAAGAAATTGATAGTAAAATAA
PC_FA-2219b_DR	FA-2219	GGCTTATTTTACTATCAATTTCTTATGATACTTGTTG
PC_FA-2219c_DF	tFA-2219	TATCAGATTCGGGAGTTAGTGCTTCAGCAGTAATTTAA
PC_FA-2219c_DR	tFA-2219	GGCTTAAATTACTGCTGAAGCACTAACTCCCGAATCTG
PC_FA-2219bSS_DF	FA-2219ΔSC	TAACAACAAGTATCATAAGAAATTGATAGTAAAA

Name	Target*	Sequence (5'→3')
PC_FA-2219bSS_DR	FA-2219ΔSC	GGCTTTTACTATCAATTTCTTATGATACTTGTTG
PC_FA-2219cSS_DF	tFA-2219ΔSC	TATCAGATTCGGGAGTTAGTGCTTCAGCAGTAATT
PC_FA-2219cSS_DR	tFA-2219ΔSC	GGCAATTACTGCTGAAGCACTAACTCCCGAATCTG
PC_FA-2220_UF	FA-2220	GCCATGCAAATAGTTCTGCAAACCGACTTTACAGATTCTGAG
PC_FA-2220_UR	FA-2220	AGAATCTGTAAAGTCGGTTTGCAGAACTATTTGCAT
PC_FA-2220_DF	FA-2220	TACCCGTGTTTTCTTTTTTAAAATATGAAGTTAAGGTTCAATAA
PC_FA-2220_DR	FA-2220	GGCTTATTGAACCTTAACTTCATATTTTAAAAAAGAAAACACGG
PC_FA-2222_UF	FA-2222	GCCATGAAAGGGCTAAACCATTCAGAAATTGAAGCTAAAATG
PC_FA-2222_UR	FA-2222	TTTAGCTTCAATTTCTGAATGGTTTAGCCCTTTCAT
PC_FA-2222_DF	FA-2222	AATGGCAGGTAGTGAAGTTTTAAAACTGGAAGAGTAA
PC_FA-2222_DR	FA-2222	GGCTTACTCTTCCAGTTTTAAAACTTCACTACCTGCC
PC_FA-2223_UF	FA-2223	GCCATGACCTATACCCAGTCAAAAGATTATAAAAACGCAAGCC
PC_FA-2223_UR	FA-2223	TTGCGTTTTTATAATCTTTTGACTGGGTATAGGTCAT
PC_FA-2223_DF	FA-2223	GTAGATTATAAAAAAACTTCATTTCAATTAAAAAAATAA
PC_FA-2223_DR	FA-2223	GGCTTATTTTTTTAATTGAAATGAAGTTTTTTTATAATC
PC_FA-2225b_UF	FA-2225	GCCATGTCTCAAACTGTAAAAAAAGAGAAACCAAATATCATTTTC
PC_FA-2225b_UR	FA-2225	AATGATATTTGGTTTCTCTTTTTTTACAGTTTGAGACAT
PC_FA-2225b_DF	FA-2225	CCAGGTAAACATATTATTGAATTAAAAATTAACTCATTTTAA
PC_FA-2225b_DR	FA-2225	GGCTTAAAATGAGTTAATTTTTAATTCAATAATATGTTTACC
PC_Ag43SP_UF	Ag43SP	GCCATGAAACGACATCTGAATACCTGCTACAGGCTGGTATGG
PC_Ag43SP_UR	Ag43SP	TACCAGCCTGTAGCAGGTATTCAGATGTCGTTTCAT
PC_Ag43N455_UF	Ag43N455	GCCAATAACGGCGCCATACTTACCCTTTCCGGGAAGACGGTG
PC_Ag43N455_UR	Ag43N455	CGTCTTCCCGGAAAGGGTAAGTATGGCGCCGTTATT

Name	Target*	Sequence (5'→3')
PC_Ag43N455_DF	Ag43N455	AGGGTATAACGGTCAGGCCACACTGAATGTGACCTTCTGA
PC_Ag43N455_DR	Ag43N455	GGCTCAGAAGGTCACATTCAGTGTGGCCTGACCGTTATAC
PC_OsmY_UF	OsmY	GCCATGACTATGACAAGACTGAAGATTTCGAAAACTCTGCTGG
PC_OsmY_UR	OsmY	GCAGAGTTTTCGAAATCTTCAGTCTTGTCATAGTCAT
PC_OsmY_DF	OsmY	AGATGGTGTGAAAAGCGTTAAAAATGATCTGAAAACTAAG
PC_OsmY_DR	OsmY	GGCCTTAGTTTTCAGATCATTTTTAACGCTTTTCACACCA
PC_CenA_UF	CenA	GCCATGTCCACCCGCAGAACCGCCGCAGCGCTGCTGGCGG
PC_CenA_UR	CenA	CCAGCAGCGCTGCGGCGGTTCTGCGGGTGGACAT
PC_CenAwoSC_DF	CenAΔSC	GCAGGAGATCGCCCTGGAGATGGCGCGCAACGCCAGGTGG
PC_CenAwoSC_DR	CenAΔSC	GGCCCACCTGGCGTTGCGCGCCATCTCCAGGGCGATCTCC
PC_CenA_DF	CenA	TCGCCCTGGAGATGGCGCGCAACGCCAGGTGGTAATAA
PC_CenA_DR	CenA	GGCTTATTACCACCTGGCGTTGCGCGCCATCTCCAGGG
PC_CHU2268_UF	CHU2268	GCCATGAAAAAAATAACCGTATTGATTTCCATCTGGCTCAGTG
PC_CHU2268_UR	CHU2268	TGAGCCAGATGGAAATCAATACGGTTATTTTTTTCAT
PC_CHU2268woSP_UF	CHU2268ΔSP	GCCATGTGTGAGAAAAAAACAGAAGCGGGGAGTAAACCTCAG
PC_CHU2268woSP_UR	CHU2268ΔSP	AGGTTTACTCCCCGCTTCTGTTTTTTTCTCACACAT
PC_CHU2268_DF	CHU2268	GATGATCGGAAATCTCCAGACAGAAATATATTTTAATGAGTAA
PC_CHU2268_DR	CHU2268	GGCTTACTCATTAAAATATATTTCTGTCTGGAGATTTCCGATC
PC_CHU2268HT_UF	CHU2268HT	GCCATGCATCACCACCACCATCACTGTGAGAAAAAAACAGAAG
PC_CHU2268HT_UR	CHU2268HT	CTGTTTTTTTCTCACAGTGATGGTGGTGGTGATGCAT
Used to get the RBSs		
RBS_FA-2190_UF	FA-2190 RBS	GCCGCCTTCAACGGCTACACG
RBS_FA-2190_UR	FA-2190 RBS	ATATTCGTGTAGCCGTTGAAGGC
RBS_FA-2190_DF	FA-2190 RBS	AATATAATCACACCCGGTACGA

Name	Target*	Sequence (5'→3')
RBS_FA-2190_DR	FA-2190 RBS	GGCTCGTACCGGGTGTGATT
RBS_FA-2194_UF	FA-2194 RBS	GCCGCGTCAGTTACTACTTT
RBS_FA-2194_UR	FA-2194 RBS	CCTTAAAAGTAGTAACTGACGC
RBS_FA-2194_DF	FA-2194 RBS	TAAGGACAAGAAGGAGTTTAACA
RBS_FA-2194_DR	FA-2194 RBS	GGCTGTTAAACTCCTTCTTGT
RBS_FA-2195_UF	FA-2195 RBS	GCCTCATAGAAGGAACGAGAC
RBS_FA-2195_UR	FA-2195 RBS	GTCTCGTCTCGTTCCTTCTATGA
RBS_FA-2195_DF	FA-2195 RBS	GAGACAAACGTCGGCCGAGGAA
RBS_FA-2195_DR	FA-2195 RBS	GGCTTCCTCGGCCGACGTTT
RBS_FA-2197_UF	FA-2197 RBS	GCCAAAAACCACCATCAGGG
RBS_FA-2197_UR	FA-2197 RBS	CGAAACCCTGATGGTGGTTTTT
RBS_FA-2197_DF	FA-2197 RBS	TTTCGATACCAAGGGGCGTTACT
RBS_FA-2197_DR	FA-2197 RBS	GGCAGTAACGCCCCTTGGTAT
RBS_FA-2204_UF	FA-2204 RBS	GCCCGACGAACAACGAGTTC
RBS_FA-2204_UR	FA-2204 RBS	TCTGTGAACTCGTTGTTCGTCG
RBS_FA-2204_DF	FA-2204 RBS	ACAGAATCTTAGGGGGACAGCTC
RBS_FA-2204_DR	FA-2204 RBS	GGCGAGCTGTCCCCCTAAGAT
RBS_FA-2205_UF	FA-2205 RBS	GCCGTCGTAAGAGGGCCACCG
RBS_FA-2205_UR	FA-2205 RBS	CTTCCCGGTGGCCCTCTTACGAC
RBS_FA-2205_DF	FA-2205 RBS	GGAAGGTAATAGGAACAGGAAA
RBS_FA-2205_DR	FA-2205 RBS	GGCTTTCCTGTTCCTATTAC
RBS_FA-2206_UF	FA-2206 RBS	GCCTGGGTCACCCGGCCAACG
RBS_FA-2206_UR	FA-2206 RBS	TTTACCGTTGGCCGGGTGACCCA
RBS_FA-2206_DF	FA-2206 RBS	GTAAATATCACGGATTAAAAGA
RBS_FA-2206_DR	FA-2206 RBS	GGCTCTTTTAATCCGTGATA

Name	Target*	Sequence (5'→3')
RBS_FA-2209_UF	FA-2209 RBS	GCCGCATGGCTGGGTTAGACC
RBS_FA-2209_UR	FA-2209 RBS	GATGGGGTCTAACCCAGCCATGC
RBS_FA-2209_DF	FA-2209 RBS	CCATCCTTACATAGGTTACTGT
RBS_FA-2209_DR	FA-2209 RBS	GGCACAGTAACCTATGTAAG
RBS_FA-2213b_UF	FA-2213 RBS	GCCATAACAACGAGGTATCTC
RBS_FA-2213b_UR	FA-2213 RBS	ACTATGAGATACCTCGTTGTTAT
RBS_FA-2213b_DF	FA-2213 RBS	ATAGTTAATCGGAGGGCCACGA
RBS_FA-2213b_DR	FA-2213 RBS	GGCTCGTGGCCCTCCGATTA
RBS_FA-2216_UF	FA-2216 RBS	GCCTTTAGGCAAGCAAATTTA
RBS_FA-2216_UR	FA-2216 RBS	CGATATAAATTTGCTTGCCTAAA
RBS_FA-2216_DF	FA-2216 RBS	TATCGAAGAGAGACAGGCAGTT
RBS_FA-2216_DR	FA-2216 RBS	GGCAACTGCCTGTCTCTCTT
RBS_FA-2217b_UF	FA-2217 RBS	GCCAGGCCCACCAAGGAACAAG
RBS_FA-2217b_UR	FA-2217 RBS	TTGAGCTTGTTCCTTGGTGGGCCT
RBS_FA-2217b_DF	FA-2217 RBS	CTCAAGGAACACCCTAGTAAA
RBS_FA-2217b_DR	FA-2217 RBS	GGCTTTACTAGGGTGTTCC
RBS_FA2219SP_UF	FA2219 RBS	GCCCAATTCTCTAACACTACG
RBS_FA2219SP_UR	FA2219 RBS	TTGGACGTAGTGTTAGAGAATTG
RBS_FA2219SP_DF	FA2219 RBS	TCCAACGCAGGAAGCGAGGAGC
RBS_FA2219SP_DR	FA2219 RBS	GGCGCTCCTCGCTTCCTGCG
RBS_FA2219woSPSC_UF	FA2219ΔSP::SC RBS	GCCGGAGGCTATCCCAGGAC
RBS_FA2219woSPSC_UR	FA2219ΔSP::SC RBS	GCTTCGTCCTGGGATAGCCTCC
RBS_FA2219woSPSC_DF	FA2219ΔSP::SC RBS	GAAGCCGTAAGTACGAGGAGGCA
RBS_FA2219woSPSC_DR	FA2219ΔSP::SC RBS	GGCTGCCTCCTCGTACTTACG
RBS_FA-2220_UF	FA-2220 RBS	GCCTGTCTCGCAGCTATTAGG

Name	Target*	Sequence (5'→3')
RBS_FA-2220_UR	FA-2220 RBS	AGTGTCCTAATAGCTGCGAGACA
RBS_FA-2220_DF	FA-2220 RBS	ACACTCAAAGCAGGAGCATAAT
RBS_FA-2220_DR	FA-2220 RBS	GGCATTATGCTCCTGCTTTG
RBS_FA-2222_UF	FA-2222 RBS	GCCCGGTAAAAACGTTTCCAC
RBS_FA-2222_UR	FA-2222 RBS	TGCGTGTGGAAACGTTTTTACCG
RBS_FA-2222_DF	FA-2222 RBS	ACGCATTATAAGGACGGAGAGG
RBS_FA-2222_DR	FA-2222 RBS	GGCCCTCTCCGTCCTTATAA
RBS_FA-2223_UF	FA-2223 RBS	GCCCCAACGACCGTCGCACG
RBS_FA-2223_UR	FA-2223 RBS	ATGTACGTGCGACGGTCGTTGG
RBS_FA-2223_DF	FA-2223 RBS	TACATTCAGCAAGAGGTCTAGAA
RBS_FA-2223_DR	FA-2223 RBS	GGCTTCTAGACCTCTTGCTGA
RBS_FA-2225_UF	FA-2225 RBS	GCCGTTTCCAAAAAACTCAAG
RBS_FA-2225_UR	FA-2225 RBS	TATGTCTTGAGTTTTTTGGAAAC
RBS_FA-2225_DF	FA-2225 RBS	ACATAACAAGGAGCAAACGCCT
RBS_FA-2225_DR	FA-2225 RBS	GGCAGGCGTTTGCTCCTTGT
RBS_Ag43b_UF	Ag43 RBS	GCCCGGACACCTCGCTAAATC
RBS_Ag43b_UR	Ag43 RBS	TGATCGATTTAGCGAGGTGTCCG
RBS_Ag43b_DF	Ag43 RBS	GATCACAACCCAAGAGAATTTC
RBS_Ag43b_DR	Ag43 RBS	GGCGAAATTCTCTTGGGTTG
RBS_OsmYb_UF	OsmY RBS	GCCACATAGCAAGAATTAAGG
RBS_OsmYb_UR	OsmY RBS	TTTCGCCTTAATTCTTGCTATGT
RBS_OsmYb_DF	OsmY RBS	CGAAACGGCAATAAGAGGATCG
RBS_OsmYb_DR	OsmY RBS	GGCCGATCCTCTTATTGCCG
RBS_cenA_UF	CenA RBS	GCCTGGTCTAACGTCGCCGCC
RBS_cenA_UR	CenA RBS	CCTCGGGCGGCGACGTTAGACCA

Name	Target*	Sequence (5'→3')
RBS_cenA_DF	CenA RBS	CGAGGAGATAAAAACATAGGTA
RBS_cenA_DR	CenA RBS	GGCTACCTATGTTTTTATCT
RBS_CHU2268_UF	CHU2268 RBS	GCCGGCAAGTAGTAGTCGCAA
RBS_CHU2268_UR	CHU2268 RBS	GTCGTTTGCGACTACTACTTGCC
RBS_CHU2268_DF	CHU2268 RBS	ACGACAGGGCGCGGGGTCCCCC
RBS_CHU2268_DR	CHU2268 RBS	GGCGGGGGACCCCGCGCCCT
RBS_CHU2268woSP_UF	CHU2268ΔSP RBS	GCCCTCCTTTCAGAGCAAAG
RBS_CHU2268woSP_UR	CHU2268ΔSP RBS	GCTGACTTTGCTCTGAAAGGAG
RBS_CHU2268woSP_DF	CHU2268ΔSP RBS	TCAGCTACAGGCTCGGAGGATCG
RBS_CHU2268woSP_DR	CHU2268ΔSP RBS	GGCCGATCCTCCGAGCCTGTA
RBS_CHU2268HT_UF	CHU2268HT RBS	GCCCGCTCGTTAACGAACGCG
RBS_CHU2268HT_UR	CHU2268HT RBS	AATTTCGCGTTCGTTAACGAGCG
RBS_CHU2268HT_DF	CHU2268HT RBS	AAATTCTAGGCCGGAGTTCTCG
RBS_CHU2268HT_DR	CHU2268HT RBS	GGCCGAGAACTCCGGCCTAG
Used to get the linkers		
link_Ag43_UF	Ag43 linker	GCCGGCGGTGTACTGCTGGCCGA
link_Ag43_UR	Ag43 linker	CGGAATCGGCCAGCAGTACACCGCC
link_Ag43_DF	Ag43 linker	TTCCGGTGCCGCTGTCAGTGGTACC
link_Ag43_DR	Ag43 linker	GGCGGTACCACTGACAGCGGCAC
L_cex_UF	Cex linker	GCCGGCGCGAGCCCGACGCCGACGCCCA CCACGCCGACCCCG
L_cex_UR	Cex linker	GGCGTCGGGGTCGGCGTGGTGGGCGTCG GCGTCGGGCTCGCGCC
L_cex_DF	Cex linker	ACGCCCACGACGCCGACGCCGACCCCGA CGTCCGGTCCGGCCGGG
L_cex_DR	Cex linker	GGCCCCGGCCGGACCGGACGTCGGGGTC GGCGTCGGCGTCGTG
L_flex_UF	Flex linker	GCCGGAGGCGGAGGAAGCGGCGGTGGTG GC
L_flex_UR	Flex linker	CCAGAGCCACCACCGCCGCTTCCTCCGC CTCC

Name	Target*	Sequence (5'→3')
L_flex_DF	Flex linker	TCTGGTGGAGGTGGCTCGGGTGGCGGCG GTTCG
L_flex_DR	Flex linker	GGCCGAACCGCCGCCACCCGAGCCACCT CCA
L_rigid_UF	Rigid linker	GCCGAAGCAGCCGCAA
L_rigid_UR	Rigid linker	CTCCTTTGCGGCTGCTTC
L_rigid_DF	Rigid linker	AGGAGGCTGCAGCAAAG
L_rigid_DR	Rigid linker	GGCCTTTGCTGCAGC
L_OmpT_UF	OmpT linker	GCCGGTGGTCGTCGT
L_OmpT_UR	OmpT linker	CGAGAACGACGACCACC
L_OmpT_DF	OmpT linker	TCTCGTCGTGTTGGTACC
L_OmpT_DR	OmpT linker	GGCGGTACCAACACGA
L_OmpTflex_UF	OmpTflex linker	GCCGGTGGTGGTGGTTCTGGTGGTCGTC GT
L_OmpTflex_UR	OmpTflex linker	CGAGAACGACGACCACCAGAACCACCAC CACC
L_OmpTflex_DF	OmpTflex linker	TCTCGTCGTGTTGGTACTGGTGGTGGTG GTTCT
L_OmpTflex_DR	OmpTflex linker	GGCAGAACCACCACCACCAGTACCAACA CGA

*ΔSP: without signal peptide; ΔSC: without stop codon; ::SC: start codon added.

Table 8-2 List of primers for MABEL PCR and sequencing

Name	Target	Sequence (5'→3')
Used for MABEL PCR		
Ag43RBS5k_UF	Ag43 RBS 5k	TGGCCATCTCTAAGTAATCCATGAAACGACATCTGAAT ACCTGC
Ag43RBS1k_UF	Ag43 RBS 1k	AAGAAAAAGCTCTCCATAAATGAAACGACATCTGAATA CCTGC
Ag43RBSs_DR	Ag43 RBSs	GGATCCAGGCTCATCCAGCCAGC
CHU2268HisF1	CHU2268HT	TGTGAGAAAAAAACAGAAGCGGG
HisTagRI*	His-tag	GTGATGGTGGTGGTGGTGATGCATAGGTACCTCCTTGA GCTCTAGTAT
FA2190_F	FA2190	ATGAAGTCTGAAACTAAACAGAATAAAAAAGAAG

Name	Target	Sequence (5'→3')
FA2190_R	FA2190	TCGTACCGGGTGTGATTATATTCG
FA2197_F	FA2197	ATGGGTTTTTGTATGAAAGATAGTAAAC
FA2197_R	FA2197	AGTAACGCCCCTTGGTATC
FA2204_F	FA2204	ATGCAAAACTTAGTACCAGAAGTAACTG
FA2204_R	FA2204	GAGCTGTCCCCCTAAGATTC
FA2206_F	FA2206	ATGACCATTTTATGTTCTTGTGCAC
FA2206_R	FA2206	TCTTTTAATCCGTGATATTTACCGTTGG
FA2209_F	FA2209	ATGTCTTGCGAAAATTCTTCTGTTTTACAAG
FA2209_R	FA2209	ACAGTAACCTATGTAAGGATGGG
FA2213_F	FA2213	ATGCAAAATAATGCGCAAACAAAATC
FA2213_R	FA2213	TCGTGGCCCTCCGATTAACTATG
FA2216_F	FA2216	ATGGAGCCCGAAAAAAAACAAGC
FA2216_R	FA2216	AACTGCCTGTCTCTCTTCGATATAAATTTG
FA2217_F	FA2217	ATGGAACAAAAGTTACAAACATCATTAG
FA2217_R	FA2217	TTTACTAGGGTGTTCCTTGAGC
FA2219_F	FA2219	ATGTTTTTAGCTGTAGTAACAGCTCAAAC
FA2219_R	FA2219	GGTCCTCCCCCTTACTTTGTTAG
FA2220_F	FA2220	ATGCAAATAGTTCTGCAAACCGACTTTAC
FA2220_R	FA2220	ATTATGCTCCTGCTTTGAGTGTCC
FA2222_F	FA2222	ATGAAAGGGCTAAACCATTCAGAAATTG
FA2222_R	FA2222	CCTCTCCGTCCTTATAATGCG
FA2223_F	FA2223	ATGACCTATACCCAGTCAAAAGATTATAAAAAC
FA2223_R	FA2223	TTCTAGACCTCTTGCTGAATGTACG
FA2225_F	FA2225	ATGTCTCAAACTGTAAAAAAAGAGAAACC

Name	Target	Sequence (5'→3')
FA2225_R	FA2225	AGGCGTTTGCTCCTTGTTATG
CHU2268_F	CHU2268	ATGAAAAAAATAACCGTATTGATTTCCATCTGGCTCAGTG
CHU2268_R	CHU2268	GGGGGACCCCGCGCCCTGTC
CHU2268woSP_F	CHU2268ΔSP	ATGTGTGAGAAAAAAACAGAAGCGG
CHU2268woSP_R	CHU2268ΔSP	CGATCCTCCGAGCCTGTAG
CHU2268HT_F	CHU2268HT	ATGCATCACCACCACCATC
CHU2268HT_R	CHU2268HT	CGAGAACTCCGGCCTAGAATTTC
Used for sequencing		
VF2	pSB1C3	TGCCACCTGACGTCTAAGAA
VR	pSB1C3	ATTACCGCCTTTGAGTGAGC
S_FA2219linker_R	FA2219 linker	TTAGAATCAAACGTTATGTTATAGACACCC

* Primer designed by Steven Kane.

8.1.2 Culture media and reagents

All media were prepared with distilled water and sterilized by autoclaving at 121°C for 15 minutes or filtering.

- LB

  10 g/l tryptone, 5 g/l yeast extract, 10 g/l NaCl.

- LB-agar

  10 g/l tryptone, 5 g/l yeast extract, 10 g/l NaCl, 15 g/l agar.

- CMC-LB-agar

  10 g/l tryptone, 5 g/l yeast extract, 10 g/l NaCl, 15 g/l agar, 2 g/l CMC.

- M9-agar

  6 g/l Na2HPO4, 3g/l KH2PO4, 0.5 g/l NaCl, 1 g/l NH4Cl, 15 g/l agar.

- MB

  5 g/l peptone, 1 g/l yeast extract, 35 g/l sea salts.

- MB-agar

  5 g/l peptone, 1 g/l yeast extract, 35 g/l sea salts, 15 g/l agar.

- PBS

  0.21 g/l KH_2PO_4, 9 g/l NaCl, 0.726 g Na_2HPO_4.

8.1.3 Standard curves

Figure 8-1 shows the protein assay standard curve using BSA as the standard.

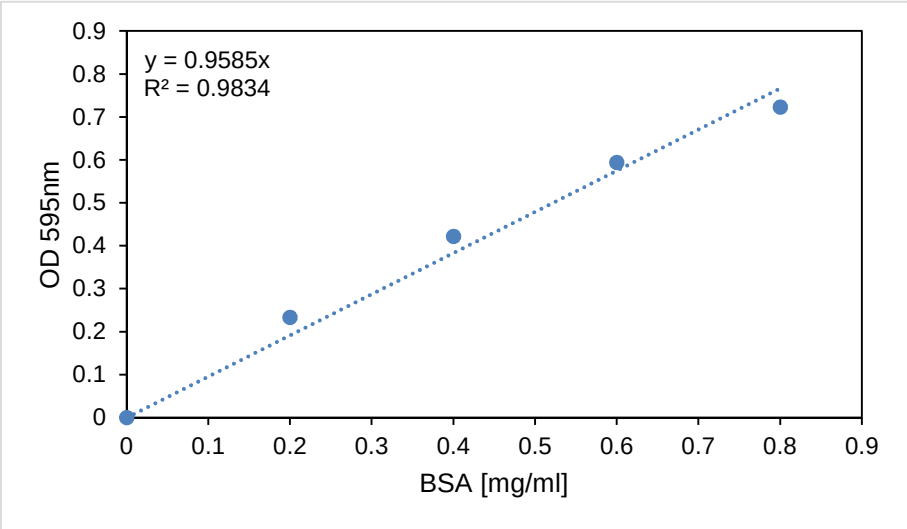

Figure 8-1 BSA standard curve

Figure 8-2 shows the cellular protein assay standard curve, with protein assay result plotted against optical density at 600 nm.

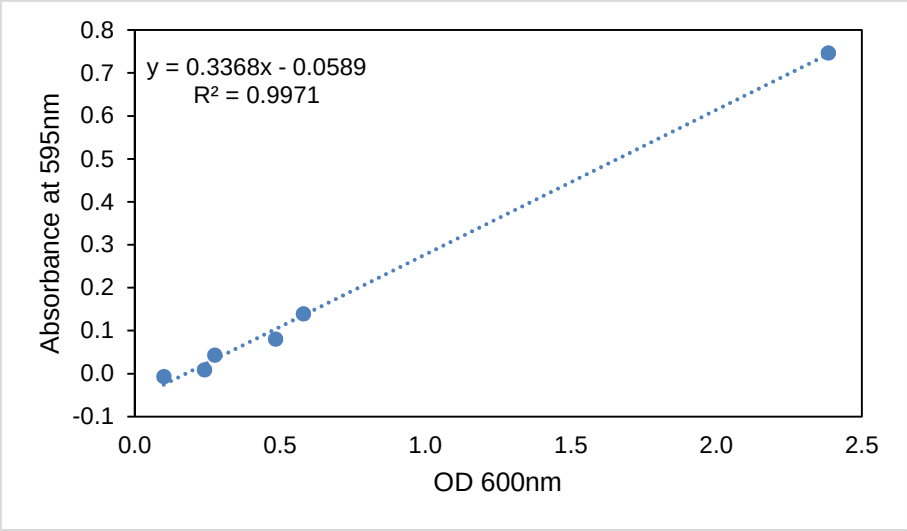

Figure 8-2 Total protein assay standard curve

Figure 8-3 shows a standard curve for the Azo-CM-Cellulose activity assay constructed using different concentrations of RBB.

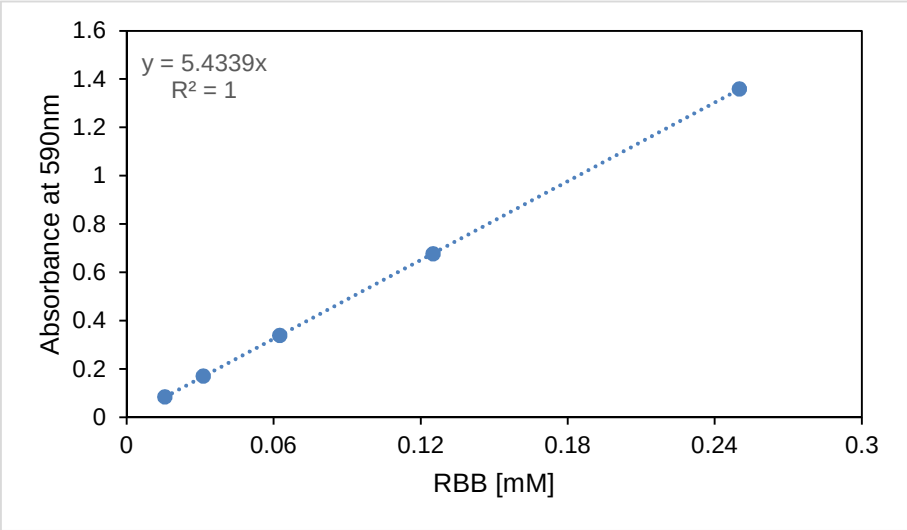

Figure 8-3 RBB standard curve

Figure 8-4 shows the MU standard curve.

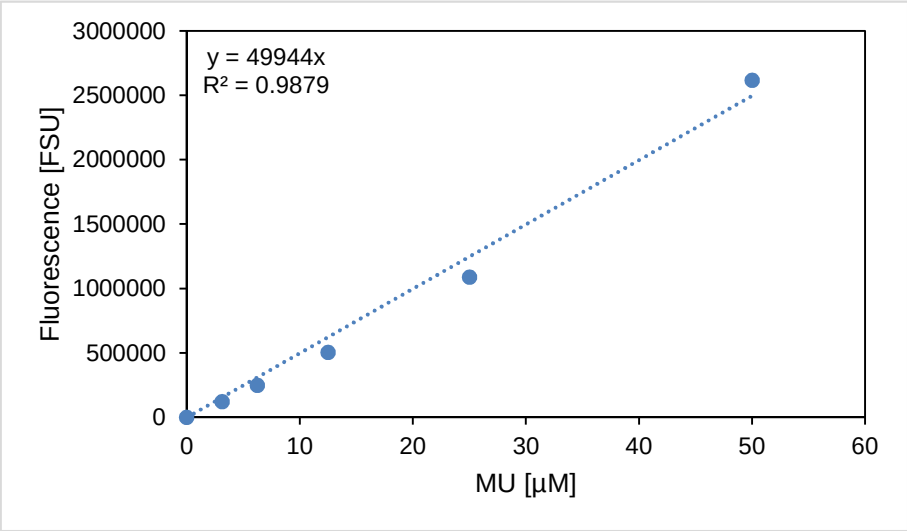

Figure 8-4 MU standard curve

Figure 8-5 shows the pNP standard curve.

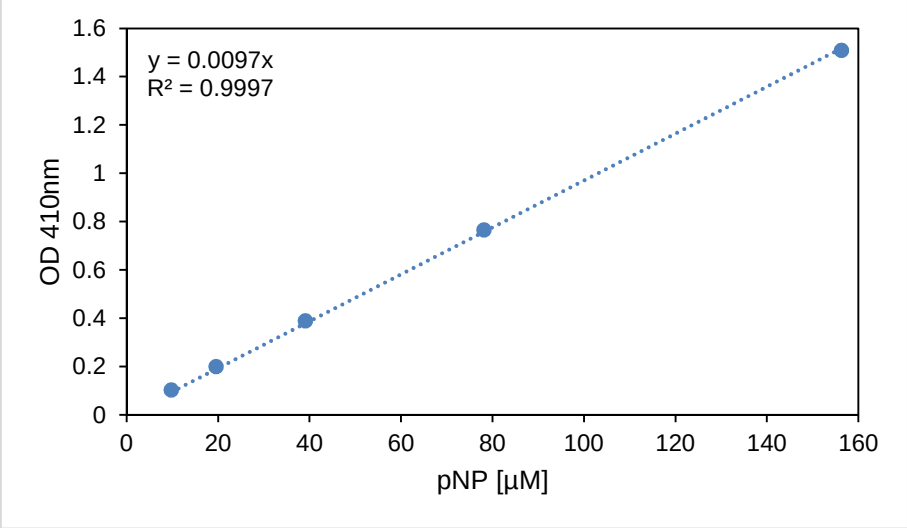

Figure 8-5 pNP standard curve

Figure 8-6 shows a standard curve for the DNS activity assay constructed using different concentrations of rhamnose. Other sugars were also used resulting in similar results.

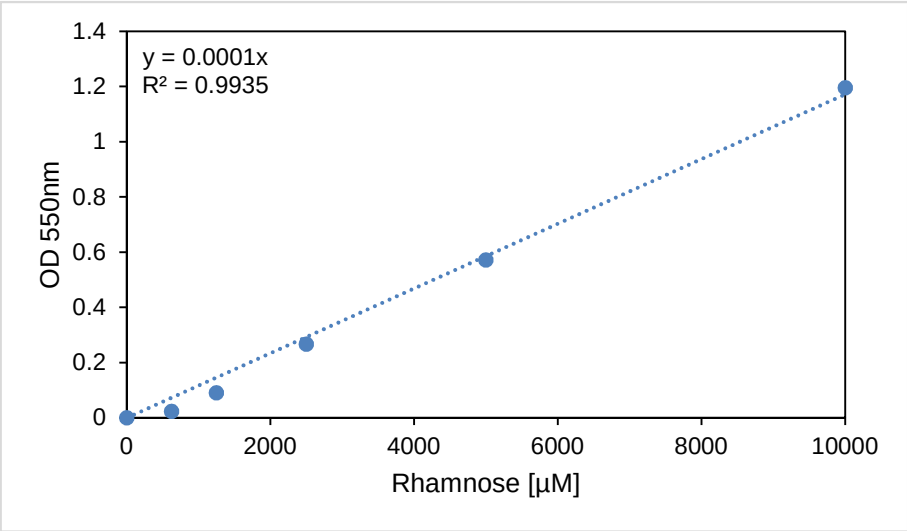

Figure 8-6 Rhamnose standard curve

8.2 Supplementary results

8.2.1 List of manually curated protein sequences

Designated names and locus tag are indicated. Amino acids that were not considered in the original annotation are highlighted.

>FA2196|BN863_21960
MKHVIMLYFIAAATLFSSCAKQDSEHRLITVKNSLDLPRAFETIEISKSDIQLHTGERFEDF
SIQDVATKAILTSQFVDEDQDGTADVLLFQPELNPNSEKQFELVKVDGGVEVDSTVYCYSRF
VPERTDDYTWENNKVAFRTYGPVAQKMVEDSLPGGTLSSGIDAWLKKVEYSIIDNWYAKNDK
DPGYYHIDHGEGLDNFHVGSSRGVGGSAVKVDTSYYISKNFTDYKTITTGPIRTSFILKYAD
WDANEKTISEEKHISLDYGNNFSRFEIHVDGTDELSVGLTLHDNKGEITQNVDQGWIAYWES
EYFDSELGTAIVAPKGVMTASEYYVTSMKDRSNLYAQLNVDNNKVVYYAGFAWKESKQYPTK
ASWEKYIQEFSEKLNTPLEVSIQ

>FA2203|BN863_22030
MKTRYFLLLGICMLSCRTDEKKQVQKEVDKPNVLFIAVDDLNNMISPIANFSNIQTPNFDRL
AAMGVTFTDAHCPAPLCGPSRSAIMTGLRPSTTGIYGMTPDNKIRRDDNEATKDIIFLPEYF
KKNGYHSMGIGKLFHNYAPDGMFDEGGGRVKGFGPFPEKRFVWDGFGTSKSRKGQYGRTNTD
WGAFPESDTLMPDHQAVNWVLERFNKNYKQPFFLALGFQRPHVPLYVPQKWFDLYPLESIQT
PPYQSDDLNDIPPVGLKINDLPMMPSTEWAINSGEWKKIIQAYLACVSFVDYELGRVLDALK
NSPYAKNTIIVLWSDHGYRLGEKGTFAKHALWESATKAPLFFAGPNLPKGKKIDAPVEMLSI
YPTLLELSGLQAYARNEAKSLVRMMQKNEGLKDTYAITTYGKNNHAVKVDGYRYIQYEDGTE
EFYDNASDPNEWINEANNFKFKSKIEALKALLPKTNATWDAESNYTFQPYFVEQKTRGNVNA
AKAVKVIGAER

>FA2205|BN863_22050
MHQKIVECIPLWSRNLNSNVKGMRGEGIACLLIVLCSIIYRTEAQRIETNFNNNWHFILKDS
PDFSKENLDDSSWELLNVPHDWSFEKGVRKGGDQGQGGGYHDGGIGWYRKTFSFSKASLSKT
TYINFDGVYMNSEVWINGNRLGKRPYGYISFRYDISKYLKVGKNTIAVRVDNGLEPSARWYH
SCGIYAPVKLVEVNPTHFKPNTIFIKTPSIEKQQGVVSIDAEIKGAFKGLKYNVELLTANGK
VIATHSEKLASAQPSVQLEVKPPKLWSPESPNLYKAKTQILDGKKVIDEKTTTFGFRTVAWK
TETGFWLNGENVKLKGVCEHWEGGPVGGAWTKPMLRWKLQSLKDMGINAIRPSHNSTPPMFY
DICDEIGLLVMDEIFDGWHKKAPEDYGKQAFDEWWQADVKEWITRDRNHPSIFVWSLGNETH
SDVAPEMVAFGKNLDPTRLFTSGAGNPEDMDIQGVNGGSETKSFIENNKLTKPFISTEAPHT
WQTRGYYRTQTWWRDNELSGTYELPNLTEKEVFFYEGINPKNWKNRKQRFNSSYDNATVRVS
ARKYWEVMRDTPWHSGHFRWTGFDYYGEAGLVHGGLPFNLFMGGALDVAGFKKDLYYFYQSQ
WTEKPMIHMLPHWTHPRMKKGTVIPVWVYANADEVELFLNGISLGKDKPGTVWNEMQCEWLV
PYEEGTLEAVGYINGKVVNRTSFSTAQQPSKLKTSILKLDAEGSFTDSFIVTSESLDTAGHL
YPYGENKVYYHIQGDVKKISMENGNPIDPTSRTKSDFRALFFGKTRTFLRALPEPKEAAVVT
AAILGDKALYTSNLITIDAQHIQLLGKSKTSDLEIRYTTNGENPETHGKLYKDAFMVEDDTT
VKAIVKQNGKTVLSMEETFGKNEGLFWGDEHSADMWIGRGVDISAEEGVLTGAAKPSREAHR
FKGSGFVDFKGGEGSITWYQENDGEPGDYSIRFRYMHNNHGKLHPMKLYVNDEYVRTIEFEP
TGGWEKEWKFVPTIIVLQSGANNIKLET
TGESGPFIDELFID

>FA2207|BN863_22070
MTNMNLKHKIFIMLLLVFCSSKIIAQQSQPNVLVFYVDDLRAELGCYGSKTAITPNIDKLAT
EGVQFNKAYVQQAICAPSRMSTLTGLRPETLGIYSIFTPLRSVHKDVVSVPQLFKENGYKTV
SIGKVYHHGTDDKNQWTNYFTKEPNTYNKPENIALLEQFKKEGKKANGPAFENADVADEAYK
DGRAAKYAVETLKKLKNDKFIMFVGFSKPHLPFNAPKKYWDLYDKNNFEIPERKKPENMYRL
ALTNWGELKGYHGIPNDVEYLDDNLTRDLIHGYHASISYVDAQVGKVMEALEALGLRKNTTV
IFMSDHGYKIGEYGAWCKHSNEEIDVRVPLIVSRETSYKGRVAGKTSDALVENVDIFPTLVE
LCGLEGPKTDGKSILQVIDRPNTPWDQVATAVYARGKNIMGCTATDGEWRYTEWRDAKTQDI
LGAELYEHKNSLLSFKNLSGNTKYKKEEARMKGLLETQFPRNQGPFLQHDTPRN

>FA2209|BN863_22090
MILHKSVFKSYIYVLTYFVFFSVMSCENSSVLQEAKHLTISEGFKNPLGFYDAKPTFSWELP
VVEGVISQSAYQIVVASSPDLLPNNPDLWDSNKQSSSQSVWINYEGKPLVSRQKVFWQVKYW
NQDDKASNWSPVQNFELGLLNNSDWKAKWIGLPTKEEGVLGSQDNIIHRPQYLRKVFELSND
VANARLYITAKGVFDVAINGEDVSDDVMPPGYTPYKKRIETITYDVTDLIESGQNTIGVEVA
AGWHSGRLGWMKSYWSDTESPKILCQLEVTMKDGSKASIISDDTWKATTQGPIRISEIYDGE
TYDAHLEMPHWTTNSFDDKNWKAVQAFPVTSTIKLEPKRHTTVKSKIVLESKEIILKADAAI
FDLQQNMVGVPLLKVPMKMGDTLKIRFAEMLSPDGTFYTDNYRSAQSTDYYIAAKEGTIEWM
PKFTFHGFRYVELSGFDASKTPSKNWVKGVVQYSNFNENGSFTSSHEKLNQLQSNIVWGLRG
NFFDIPTDCPQRDERMGWTGDAQVFGPTSMFNADVYKFWASWMQSVRESQYDNGGIPFVVPD
VLHNGKVSSGWGDVCTIIPWKIYYRTGDVGILEENYDMMKKWVAHHQATSKDFISHMNSFAD
WLQPYPENGNNKGDTSHSLIGTAFFAHSAKLTAKTAEVLGKKEEQATYEALYKSVAKAFENA
FFKNGKVKDVTATQTSYLLALAFDLLSEENKENAKQQLLEKISEADNHLRTGFLGTPLLSEV
LDETGEIDLMYKLLFNETYPSWFYSINQGATTIWERWNSYSKAEGFNPMKMNSLNHYAYGAI
GEWMYERITGIAPLQAGYKIISIAPIPKAPLTSASATLNTPYGEVASSWEIKNETLFLEVVV
PPNTTAEIEIPTDNSESLKVDNENFTNGKNLKLIKNEKRKIKILAQPGTYEFQAKYSL

>FA2218|BN863_22180
MKRRNFIQLSSLATIGMSLPSAGIVNACSSFPEQSLEFKNLTSELLKEWCDGMLKVQINNPS
NLEEHGALRCPSCSHIHGRCMDAVYPFLYMADVSGDEKYIEAAKLVMIWAENNVSQENGAWT
VIPNPKSWKGITIFGAIALAESLHYHSHILDDKTLKAWTNRLARAGQYIYDTFTIDFTNINY
GGTAIYGLDIIGDVLGNGNFKEKSKKMAEEVQAFFTKNDYLLYGECKPEADKLSAKGLHGVD
LGYNVEETLNSLVMYALKNDDQALLQIVTKSLNSHLEFMLPDGGWDNSWGNRMYKWTYWGSR
TCDGSQPAFAMMAHINPAFGTAAVKNTELLKQCTANGLLHGGPHYISAGIPPCVHHTFTHAK
PLAALLDHWKHLPEINKTTALPRVTANGIKHFKDLDVLLFSRGDWRGTVSAYDAEYHYKKDY
RQATGGSLGILYHNKVGLLCAASMAVYNMVEPYNQQPQPGKDIALTPRIETFKEDQWYTNLY
DLTANLEAIDTKEVINLASVVKLKNESRKMVSGTASEFHLTYSCAKEGLTIKVSTQQDILEP
TAFVLPIASPEKEKVEFVNEHEIKISKPGGVVTIKANVPLKLKEYSGTRTFNMVPGLEALPI
ELFFETHIKELVLIVSVV

>FA2219|BN863_22190
MLEKTTLKNIILIHFLMFLAVVTAQTAPDEDTSAITRCTAEGTNPVRETDIPNPVNVGTIDD
RSCYANYKESTVYGKTWGVYNITFDSNDFDTSLQPRIERSLSRSSETGIGSYARLTGVFRIL
EVGDTSGTSQDGTYLAQAKGKHTGGGGSPDPAICLYLAKPVYGTGEDADKQVSFDIYAERIL
YRGGEGDGREIVFLKNVKKDEETNFELEVGFKEDPNDVSKKIQYCNAVIGGDTFNWNIPEPE

RGTESGIRYGAYRVKGGRAQIRWANTTYQKVENVEVTNPGPIGDVYKLKNVATGQYLSDSGV
SASAVIMSDSGEAQNNYWTFVESGSLFNIDNETFGILRAPGAGGPGGAYVVVSTTKEGPSSD
GDKVWTIHYNESNDTYRFESGSSGRFMYQEINGNVTHISAMNTDDRSVWKAIAVESLSVDEN
AILASDVRVFPNPASDSFTISLKTINHVTVNIYDVLGNTIFKSEFNGDTIQIRNKGQFKAGV
YLIQLTDKNNNKYHKKLIVK

>FA2225|BN863_22250
MIKYKAIINLVFIAVFFNNAMSQTVKKEKPNIIFILTDDQRFDAIGYAGNKFVNTPEMDKLA
QQGTYFDHAIVTTPICAASRASLWTGLHERSHNFNFQTGNVREEYMNNAYPKLLKNNGYYTG
FYGKYGVRYDNLESQFDEFESYDRNNRYKDKRGYYYKTINNDTVHLTRYTGQQAIDFIDKNA
TNTQPFMLSLSFSAPHAHDGAPEQYFWQTTTDALLQDTTLPGPDLADEKYFLAQPQAVRDGF
NRLRWTWRYDDPEKYQHSLKGYYRMISGIDLEIKKIRDKLKEKGVDKNTVIIVMGDNGYFLG
ERQLAGKWLMYDNSIRVPLIVFDPRVNKHQDISEMVLNIDVTQTIADLAGVKAPESWQGKSL
LPLVKQETSTISRDTILIEHLWDFENIPPSEGVRTEEWKYFRYVNDKTIEELYNIKKDPKEI
NNLIGKKKYQNVAKALREKLDELIAKNSDEFRAGPSDLTVELIRQPESEVKIFDLKPEFGWT
VPLSSKYQSAYQLLVASSETIINANNGDVWDSGQVRSSQSTNVDFGGKPLKIGETYYWKVRI
WDEENRLVDYSKAQKFTIGESDNYIISTENKFVTDKIKPSKFENRDGVYFIDFGKAAFATME
FNYQAKTPHTLTIRVGEMIDENGNVNRTPPAKSNIRYQELKVEVKPGQTRYRIPIQTDERNT
RPNKAIPLPKGFPPLLPFRYAEIEGAQSSINANDVEQLAYHTFWDEKASSFKSDNNILNQVW
DLSKYSIKATTFNGLYVDGDRERIPYEADAYLNQLSHYTTDREYAMARRTIEYFMKNPTWPT
EWQQHVALLLYADYMYTGNTELVERYYEALKHKSLYELSNEDGLITSTKVDAEFMKKLGFPE
GYKKPLTDIVDWPGANFNGSKTPGERDGFVFQPYNTVINSFFYENMKIMAQFAKILGKTDEV
LDFELRAAKAKKAVNEQMFDKKRGIYVDGIGTDHASLHANMMPLAFGLVPQEHVDTVVEFVK
SRGMACSVYGAQFLLDGLYNVGEADYALDLLASTSERSWYNMIRIGSTITLEAWDNKYKNNL
DWNHAWGAVPANAIPRGLWGIKPKTAGFGIASIKPQMGKLKSSQITVPTVRGAIHATFTHNG
PRSQTYEIEIPGNMVAEFSLDDIDGKDLIHNGQKVPAAFGAVQLSPGKHIIELKINSF

>FA2226|BN863_22260
MKTYNINKRINTLLLLVITMLSFSGCDLEPQEKFRFDPEVDPQFTFGSMTTWEWLQTNPNDE
FGFMIEAIKQTGLQDMYNSKTETYTFFLMKDPNWTNNGPGFFSREFNLKNTADRDPKEVFED
PAVDLDIVRNYLLYLTLPIYVDQGPDHLKTLDLPYTFETLSEDVNNQIMTIARDWNYVMQIN
DSPDLPTGNLGKINVPVGYHNYIFSNGNSVAHIFGLNNNGKMARRYKFGEPKMDF

8.2.2 Localization prediction of proteins from the ulvan utilization PUL

Table 8-3 Localization prediction of proteins of the ulvan utilization PUL

Protein	Localization[a]	Phobius TM	TMHMM	Phobius SP	SignalP	BOMP	TatP	Lipo[a]	LipoP class[b]
FA2190	OM	0	0	Y	0	1	0	0	SpII
FA2191	OM	0	0	Y	0	0	0	OM	SpII
FA2192	OM	0	0	Y	Y	5	0	0	SpI
FA2193	IM	1	1	Y	0	0	0	0	SpI
FA2194	CP	0	0	0	0	0	0	0	CP
FA2195	CP	0	0	0	0	0	0	0	CP
FA2196	OM	0	0	Y	0	0	0	0	SpII
FA2197	CP	1	1	0	0	0	0	0	CP
FA2198	CP	0	0	0	0	0	0	0	CP
FA2199	OM	0	0	Y	Y	2	0	0	SpI
FA2200	OM	0	0	Y	0	0	0	OM	SpII
FA2201	OM	0	0	Y	0	0	0	0	SpII
FA2202	PP	0	0	Y	Y	0	0	0	SpI
FA2203	PP	0	0	Y	0	0	0	0	SpI
FA2204	PP	0	0	Y	Y	0	0	0	SpI
FA2205	CP	0	0	0	0	0	0	0	CP
FA2206	CP	0	0	0	0	0	0	0	CP
FA2207	OM	0	0	Y	Y	0	0	0	SpII
FA2208	OM	0	0	Y	0	0	0	0	SpII
FA2209	OM	1	0	0	0	1	0	0	SpII
FA2210	CP	0	0	0	0	0	0	0	CP
FA2211	PP	0	0	0	Y	0	0	0	CP

Protein	Localization[a]	Phobius TM	TMHMM	Phobius SP	SignalP	BOMP	TatP	Lipo[a]	LipoP class[b]
FA2212	PP	0	1	Y	Y	0	0	0	SpI
FA2213	PP	0	0	Y	Y	0	0	0	SpI
FA2214	OM	0	0	Y	0	0	0	OM	CP
FA2215	OM	0	0	0	0	5	0	0	CP
FA2216	OM	0	0	Y	Y	1	0	OM	SpII
FA2217	OM	0	0	Y	Y	0	0	0	SpII
FA2218	PP	0	0	0	Y	0	Y	OM	CP
FA2219	EC	0	0	Y	Y	0	0	0	SpI
FA2220	PP	1	1	Y	0	0	0	0	SpI
FA2221	OM	0	0	Y	0	0	0	0	SpII
FA2222	PP	0	0	Y	Y	0	0	0	SpI
FA2223	OM	0	0	Y	0	0	0	OM	SpII
FA2224	CP	0	0	0	0	0	0	0	CP
FA2225	PP	0	1	Y	Y	0	0	0	SpI
FA2226	PP	0	0	0	Y	0	0	IM	CP
FA2227	OM	0	0	Y	Y	0	0	OM	SpII
FA2228	OM	0	1	Y	Y	5	0	0	SpI
FA2229	IM	1	2	Y	0	1	0	0	SpI
FA2230	CP	0	0	0	0	0	0	0	CP

LipoP (Juncker et al., 2003) and Lipo v1.0 (Berven et al., 2006) were used to predict lipoprotein signal peptides. TatP (Bendtsen et al., 2005) was used to predict Tat signal peptides. TMHMM v2.0 (Sonnhammer et al., 1998) was used to predict transmembrane helices. Phobius (Kall et al., 2004) was used to predict both transmembrane spans and signal peptides. SignalP v4.1 (Petersen et al., 2011)was used to predict Sec signal peptides. BOMP (Berven et al., 2004) was used to predict beta barrel spans. [a] CP: cytoplasm; IM: inner membrane; PP: periplasm; OM: outer membrane; EC: extracellular. [b] SPI: signal peptide (signal peptidase I); SPII: lipoprotein signal peptide (signal peptidase I).

8.2.3 Ulvan solubilisation

Different pre-treatments were studied in order to solubilise ulvan from *U. lactuca* biomass. The solubilisation of more than 50% of the sugars present in *U. lactuca* was achieved by incubating the samples for 4 h at 85°C and about 70% when 150°C for 10 min were used (van der Wal et al., 2013). In the same study, the addition of sodium hydroxide or sulfuric acid was also analysed; however, this did not improve the levels of solubilisation achieved. In this study a pre-treatment consisting in autoclaving the biomass at 121°C for 15 minutes was also analysed. The new pre-treatment was compared with the one at 85°C and a sample that was not treated. As references samples of *U. lactuca*, *U. lactuca* alcohol insoluble residue (AIR), and purified ulvan using ammonium oxalate were used. All samples were prepared using distilled water.

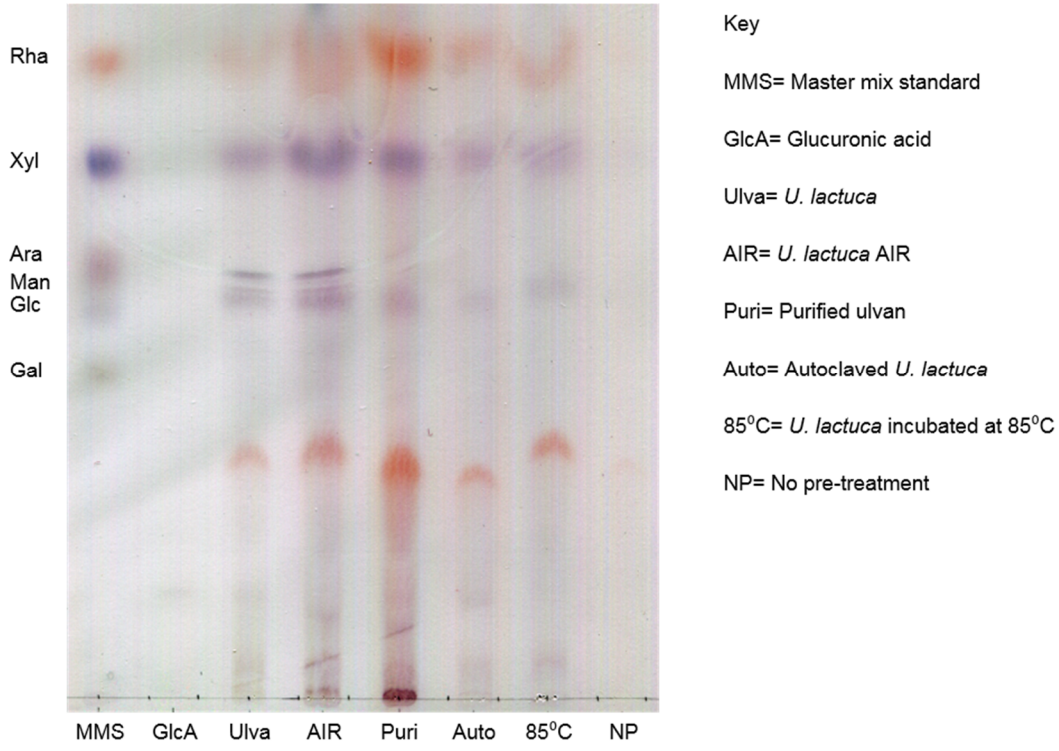

Figure 8-7 TLC analysis of the solubilized ulvan samples. All samples were hydrolysed using TFA. Samples were dried, resuspended in chlorobutanol and loaded on a TLC silica plate. EPPAW (ethyl acetate, pyridine, propanol, acetic acid and water) was used as mobile phase and the plate stained with a thymol solution for visualization. Rha: rhamnose; Xyl: xylose; Ara: Arabinose; Man: mannose; Glc: glucose; Gal: galactose.

All solutions were hydrolysed using TFA and the resulting samples analysed by TLC (Figure 8-7). A master mix standard containing the main monosaccharides found in cell wall polysaccharides was used. Clear bands showing the presence of rhamnose, xylose and glucose were observed in all samples. Weak bands can be observed showing the presence of glucuronic acid. Additionally, a clear band of the same colour of rhamnose appeared between galactose and glucuronic acid in all samples. It has been reported that the linkage between glucuronic acid and rhamnose is recalcitrant to acid hydrolysis (Lahaye & Robic, 2007). Considering this, the extra band probably corresponds to the dimer glucuronic acid – rhamnose. A clear enrichment in rhamnose and this band is observed in the lane containing purified sample. Similar results were observed for the samples that were pre-treated, while almost no sugars were observed when the biomass was not pre-treated. These results demonstrate that the two pre-treatments are effective in solubilizing the ulvan contained in macroalgal biomass. The autoclaving pre-treatment was selected to be used in this study considering its practicality regarding the preparation of culture media.