Characterisation of carbon fixation proteins in the macroalgal biomass feedstock, Ulva spp.

ROMERO-LOPEZ, ITZEL,FABIOLA

Characterisation of carbon fixation proteins in the macroalgal biomass feedstock, *Ulva* spp.

Itzel Fabiola Romero-López

Submitted for the qualification of Master of Science (MSc)

Department of Biosciences, Durham University

April 2020

Abstract

Ulva spp. is a macroalgae widely distributed and abundant in nature, however, its exploitation as a biofuel feedstock has been scarce due the lack of information about its metabolic functioning. On the other hand, microalgae have been extensively studied allowing understanding of the complexity of these organisms and, at the same time, providing a guide which could be extrapolated to macroalgal metabolic systems.

As an essential metabolic process, the study of carbon dioxide fixation and associated intracellular structures is fundamental in order to improve and engineer changes in algal biomass yield. As part of these intracellular components, the pyrenoid is a microcompartment where carbon dioxide is fixed by maintaining a CO_2 rich environment around Rubisco. Although Rubisco is the main constituent of the pyrenoid matrix, it is not the only one and it is not efficient enough to carry out CO_2 fixation by itself. Previous work in the model microalga *Chlamydomonas reinhardtii* has revealed the existence of a linker protein called EPYC1, whose main role is to bind Rubisco together to form a complex avoiding CO_2 leakage.

To study pyrenoid in *Ulva*, a chloroplast isolation method was developed. Bioinformatics searches were performed in order to find a putative protein in the *Ulva* proteome with similar physicochemical properties to EPYC1. A single target candidate which fulfilled all physicochemical properties was identified. Finding of a putative of *Ulva* EPYC1-like protein allows for further studies.

Table of contents

List of figures

List of tables

List of abbreviations

Common abbreviations have been used for temperature (°C), mass (µg, g), volume (µl, ml), area (mm^2), distance (m), molarity (µM, mM, M), rotational speed (rpm, g, mot/min), time (s, min, h), voltage (V), wavelength nanometre (nm).

Supplementary abbreviations are listed below:

ACN	Acetonitrile
ASW	Artificial seawater
ATP	Adenosine triphosphate
BCA	Bicinchoninic acid
BLAST	Basic Local Alignment Search Tool
BSA	Bovine serum albumin
$CaCl_2$	Calcium chloride
CCM	Carbon concentration mechanism
CHAPS	3-((3-cholamidopropyl) dimethylammonio)-1-propanesulfonate
Ci	Inorganic carbon
CO_2	Carbon dioxide
CO_3	Carbon trioxide
DTT	Dithiothreitol
EB	Extraction buffer
EDTA	Ethylenediaminetetraacetic acid
EPYC1	Essential Pyrenoid Component 1
epyc1	Essential Pyrenoid Component 1 mutant
ExPASy	Expert Protein Analysis System
GAP	Glyceraldehyde 3-phosphate
GAP-DH	GAP-dehydrogenase
HC	High carbon
HCl	Hydrochloric acid
HCO_3	Bicarbonate
ID	Identifier
iTRAQ	isobaric Tagging for Relative and Absolute Quantitation
kDa	Kilodalton
K_m	Michaelis constant

LC	Low carbon
LC-MS	Liquid chromatography-Mass spectrometry
Lys	Lysine
MES	2-(N-morpholino) ethanesulfonic acid
min seq	Minimum sequence
MOPS	3-(N-morpholino) propanesulfonic acid
MQ	Milli-Q
Mw	Molecular weight
NaCl	Sodium chloride
$NADP^+$	Nicotinamide adenine dinucleotide phosphate
$NaHCO_3$	Sodium bicarbonate
NaOH	Sodium hydroxide
NCBI	National Center for Biotechnology Information
NOAA	National Oceanic and Atmospheric Administration
OrcAE	Online resource for community Annotation of Eukaryotes
PEB	Protoplast extraction buffer
PGA	3-phosphoglycerate
pH	Potential for hydrogen
pI	Isoelectric point
Pondr	Predictor of Natural Disordered Regions
$ppCO_2$	CO_2 partial pressure
rbcL	Rubisco large subunit
rbcS	Rubisco small subunit
Rubisco	Ribulose-1,5-biphosphate carboxylase/oxygenase
RuBP	Ribulose 1,5-biphosphate
SDS	Sodium dodecyl sulfate
SDS-PAGE	Sodium dodecyl sulfate polyacrylamide gel electrophoresis
TAGs	Triacylglicerols
TMD	Transmembrane domain
TR	Tandem repeats
TRIS	Tris(hydroxymethyl)aminomethane
v/v	Volume/volume
w/v	Weight/volume
w/w	Weight/weight

XSTREAM	Variable sequence tandem repeats extraction and architecture modelling

Acknowledgments

I would first like to express my gratitude to my supervisor, Dr John Bothwell, for his guidance, support and always optimistic point of view. Your advice and patience have been essential for this project to go forward, despite the difficulties.

I am also thankful to people in Lab 5, who always were there for an advice, or only to chat to look how things were going, specially to Dr Steve Chivasa for providing me assistance and ideas for my experimental work, Annabel for her help in the late stage of my lab work, and Colleen Turnbull for being patient and always maintaining the lab in the best conditions. Thanks also to my lab mates Alex Goodridge and Akanksha Agrawal for arranging collection days and being there for support.

Thanks to Biosciences department and its staff for being so welcoming and always willing to help during this process, specially to Dr Adrian Brown for letting me work in his lab and for analysing my samples.

To all the amazing people who accompanied me during my stay in Durham: Diana (my labmate and adventure partner since 2009), Ana Cristina, Martha, Diego, Balder, David, Karen and Alexandra; thank you for the laughs, the celebrations, the support when things got stressful, and for making me feel at home. And last but not least, to Jeroen, for always cheering me up and being there for me no matter where and how.

I am forever thankful to my parents, Faviola and Octavio, and my second family Julieta and Iván, for unconditionally encourage me to reach my goals and for always being so loving and caring. Gracias, les quiero.

Finally, I would like to thank the Mexican funding institutions CONACyT-SENER, without which my studies at Durham University wouldn’t had been possible.

1. Introduction

1.1. The growing need for Biofuels

Energy is a fundamental requirement for human activity. However, the fossil fuel use of the past century is no longer sustainable and, in addition, its use contributes to climate change: over three-quarters of CO_2 emissions are the consequence of the burning of fossil fuels such as oil and coal, mainly for the electricity generation, transportation, heating, and cooling needed to support industrial activities (Dow & Downing, 2011). The Carbon Dioxide Information Analysis Center estimates that fossil fuel emissions between 2011 and 2012 increased from 9.4 to 9.6 billion metric tons of carbon, and these increases are thought to drive the constant rises in atmospheric CO_2 concentration such as those reported by the US National Oceanic and Atmospheric Administration (NOAA) over marine surface sites (Figure 1.1).

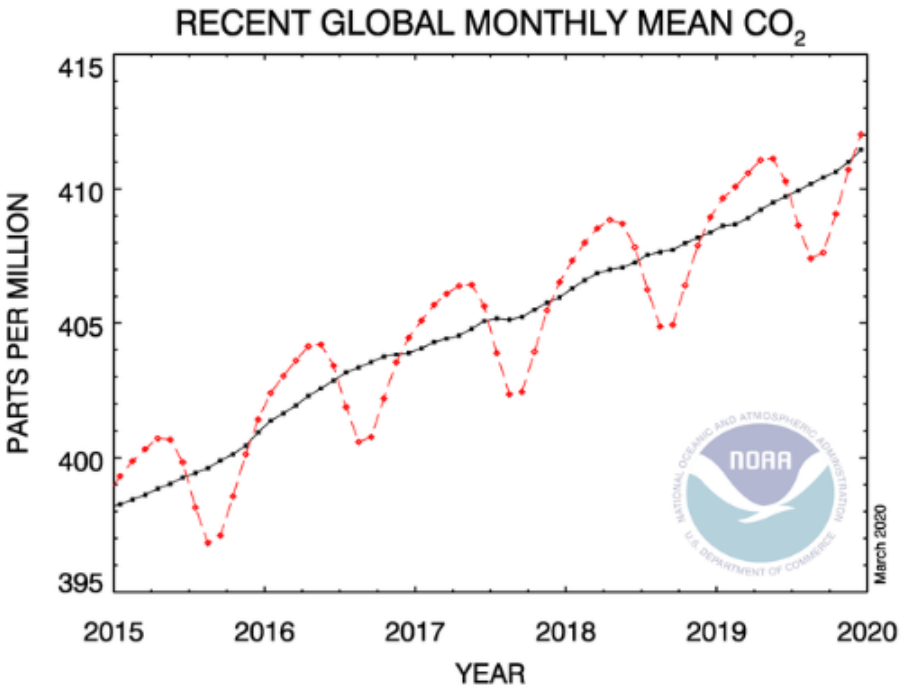

Figure 1.1 *Recent global mean of CO_2 over marine surface sites.*
The red line represents the monthly mean values and the black one represents the same data after correction for the average seasonal cycle. Retrieved from https://www.esrl.noaa.gov/gmd/ccgg/trends/global.html

Due to energy demand and supply inequalities, environmental and social issues, and the costs that conventional sources of energy involve, bioenergy is a promising possible way to help meet the challenges posed by a transition to clean and sustainable energy (Kumar et al., 2016). Biomass production can contribute to a

wide range of energy technologies, such as electricity generation, heat generation and fuel production (Aslani et al., 2018). Accordingly, biofuels have real potential to replace, or at least reduce, the use and effect of fossil fuels. A variety of renewable forms of biomass are being explored and some of them have been more exploited than others. For example, interest in agricultural residues such as wheat straw, olive residue, and corn stover as feedstocks to supply the production of energy and liquid fuels, and to reduce the environmental effect caused by the use of fossil fuels has been increasing (Arvelakis & Koukios, 2013).

Biofuels are often classified by “generations”: first generation biofuels are derived from crop plants and, currently, more than 95% of biodiesel sources are first generation agricultural edible crop oils, such as soybean, corn and sugar cane (Gui et al., 2008; Demirbas, 2011; Rawat et al., 2013). Unfortunately, this often means that first generation biofuels compromise food availability, which helps to explain why second generation biofuels were developed: these fuels can be generated from non-crop sources such as waste cooking oil, animal fats and jatropha oil, as well as from non-edible crops, such as low-priced wood and switch grass (Antizar-Ladislao & Turrion-Gomez, 2008; Alalwan et al., 2019). However, even though second generation biofuels overcome many limitations of the first generation, their efficiency is compromised because non-edible plants often have high lignin contents, requiring expensive enzymatic processes to depolymerize and release sugar. And, in addition, they can have a negative impact on food security due to land use (Aro, 2016; Alalwan et al., 2019).

Finally, third generation biofuels are those extracted from algal biomass, solving the land use and food availability issues that accompany second generation biofuels (Alaswad et al., 2015). To these three major biofuel generations, a fourth, more speculative, one has been introduced in the last 12 years: fourth generation biofuels focus on systems-level metabolic engineering of oxygenic photosynthetic organisms, mainly microalgae (Lü et al., 2011).

1.2. Algal biomass as an energy resource

Algae may be grown as third and fourth-generation biofuels. These algal-based biofuels offer particular versatility, sustainability and economic advantages that makes them a promisingly important component in the ongoing fight to avoid future

energetic and environmental crises. Specifically, the use of algal biomass as a biofuel feedstock avoids certain limitations with land crop biomass: algae grow in water, so do not compromise food security and land availability (Ashokkumar et al., 2017); they can be converted into solid, liquid and gaseous biofuels for bioenergy generation (Vassilev & Vassileva, 2016); they can grow around an order of magnitude faster than many conventional terrestrial crops (Balina et al., 2017), and they also have greater potential for CO_2 remediation because of their high photon conversion efficiency (Abomohra et al., 2018), which averages around 6–8%, rather than the 1.8–2.2% of terrestrial biomass (Aresta et al., 2005).

1.2.1. Macroalgae and biofuel production

To date, the bulk of interest in algal bioenergy has focussed on the oil-rich microalgae, for which genetic modification tools already exist. However, there is now increasing interest in the macroalgae, or seaweeds, which are the green, red and brown algae that, at some stage of their life cycle, form large multicellular thalli (Hurd et al., 2014). Seaweeds have the distinct advantage that they are easier to grow and harvest than microalgae; indeed, seaweed industries already exist in a number of countries worldwide. Accordingly, they are attracting attention as potential sunlight-driven factories for CO_2 uptake and biomass generation (Milledge et al., 2014).

However, biofuel yields from seaweeds are currently relatively low, because their carbohydrate-rich nature means that they are primarily fermented into bioethanol or digested into biogas (Felix et al., 2018). This means that farming seaweeds for biofuel is only on the borderline of being economically viable. At the moment, the main approach to improving this viability looks at biorefinery approaches, in which value is derived from both fuel and non-fuel components. The evolutionary breadth of seaweeds means that they produce an impressive range of metabolites, with over 3,000 compounds reported, often as a consequence of unique evolutionary adaptations to living in a harsh environment (Belghit et al., 2017). Utilising this range of metabolites and their composition (Table 1.1) will be critical if the many advantages of seaweeds are to be realised. In particular, the content of carbohydrates, addressing biogas, biobutanol, and bioethanol production is key for biofuel production success, additionally, the content of triacylglicerols (TAGs), suitable for biodiesel production and their relative lack of phosphorus, sulphur and nitrogen that means that they digest and ferment well (Suutari et al., 2015).

Table 1.1 *Macroalgae composition*

	Green algae	Red algae	Brown algae
Polysaccharide	Mannan Ulvan Starch Cellulose	Carrageenan Agar Cellulose Lignin	Laminarin Mannitol Alginate Glucan Cellulose
Monosaccharide	Glucose Mannose Uronic acid	Glucose Galactose Agarose	Glucose Galactose Uronic acid
Representative species	*Ulva lactuca*	*Gelidium amansii*	*Laminaria japonica*
Carbohydrates (% w/w)*	54.3	83.6	59.5
Lipid (% w/w)*	6.2	0.9	1.5
Protein (%w/w)*	20.6	12.2	30.9
Ash (%w/w)*	18.9	3.3	8.1

** Dry biomass*
(Chen et al., 2015)

1.2.2. *Ulva* spp. as biofuel feedstocks

For any given biomass to be considered as a strong biofuel prospect, its composition is not the only consideration: a fast growth rate and the ability to grow in diverse geo-climatic conditions are also important. Accordingly, this project will focus on the cosmopolitan green seaweed, *Ulva* spp., whose distribution includes tropical and temperate coasts, whose growth rate increases in response to nutrient availability (Smetacek & Zingone, 2013) and which is already being investigated for its biorefinery potential (Trivedi et al., 2016).

However, there is also another reason to study *Ulva*, which is increasing importance in coastal ecosystem management because its development drives green tides in shallow environments (Smetacek & Zingone, 2013; Wichard et al., 2015). Consequently, these “green blooms” cause damage in tourism and local economies, as well as to larger environments: even though *Ulva* itself is not toxic, its decay produces toxic hydrogen sulfide gas (De Clerck et al., 2018). These nuisance blooms arise, at least in part, because of *Ulva’s* unusual subcellular architecture: *Ulva* chloroplasts contain pyrenoids that give it a higher affinity for inorganic carbon ($Ci = CO_2 + HCO_3^- + CO_3^{-2}$) than terrestrial C3 plants (Mukherjee & Moroney, 2019), enhancing its growing rate and CO_2 uptake. This mechanism opens up a possible way to improve *Ulva* cultivation.

1.2.3. Algae cultivation and maintenance in the industry

The emerging demand for seaweeds as industrial feedstocks has led to their large scale cultivation, in addition to harvesting their natural populations. This cultivation has traditionally been in the natural environment: in lagoons or sheltered bays to receive nutrients directly from the sea. Such seaweed farms can be simple rock-based ones, where seaweed explants are attached to a rock; fixed-off bottom farming, in which seedlings are tied to a 10-20 m length rope in shallow waters, allowing near-shore cultivation but needing more intensive maintenance; or floating raft farming, where the seaweed is suspended by floats or wooden rafts near the sea surface, enabling farming in deeper water sites (Sudhakar et al., 2018). This cultivation is still being explored using, for example, novel concepts for increasing the yield per unit area such as mixing the crops in order to exploit natural photon capture and carbon fixation rates (Golberg & Liberzon, 2015; Buschmann et al., 2017).

For all of these farms, yields depend on various factors such as the location, water quality, depth, current, fauna presence, temperature and light (Sudhakar et al., 2018). These two last factors are of primary importance because they feed directly into photosynthesis, which drives growth and lifecycle progression as light induces spore release in seaweeds, making their reproduction possible and increasing the amount of carbon sequestered by populations (Castelar et al., 2014; Buschmann et al., 2017).

1.3. Importance of carbon dioxide fixation in algal biomass generation

Biological carbon fixation is essential in the global carbon cycle because at the same time as plants and other organisms produce resources such as food and fuels, the atmospheric composition is regulated. It was estimated that, every year, about 10^{11} metric tons of CO_2 are converted into organic material by photosynthesis which supports much of the life on Earth (Cleland et al., 1998). Around a third of this global carbon fixation occurs in the algae (Mackinder et al., 2016). Algal growth depends on various factors, such as pH, temperature, nutrient availability, light and CO_2 concentration; it is therefore important to understand the mechanisms that underlie each of these if we are to improve algal biomass yields.

1.3.1. Mechanism of CO_2 handling in seaweeds

Carbon metabolism is a well-studied process in higher plants, with much work having been done on its enzymatic machinery and mechanisms. While these core mechanisms are conserved in seaweeds, there are some differences related to carbon concentration and acquisition, biochemical strategies to avoid the oxygen/CO_2 interference of Rubisco in photorespiration, and C4 metabolism (Gomez & Huovinen, 2012). These differences confer on seaweeds the ability to operate under particular limitations, such as low CO_2, light and nutrient availability and in changing environments.

The basic seaweed carbon fixation pathway is shown in figure 1.2 and starts with inorganic carbon acquisition of CO_2, HCO_3^- and CO_3^{2-}, which are in equilibrium as part of the water's carbonated buffer systems (Hurd et al., 2014). This inorganic carbon is then fed into the chloroplast Calvin-Benson cycle, which consists of three main reactions: carboxylation, reduction and regeneration.

During the carboxylation phase, carbon passes through the active site of Ribulose-1,5-biphosphate carboxylase/oxygenase (Rubisco) (Cleland et al., 1998). Rubisco is a bifunctional enzyme that is present in all photosynthetic eukaryotes, including algae (Moritz et al., 2012; Atkinson et al., 2017). Rubisco has a significant role in the fixation of atmospheric CO_2 using Mg^{2+} and inorganic carbon as cofactors (Linthwaite et al., 2018) and catalyses the incorporation of inorganic carbon through the carboxylation of ribulose 1,5-biphosphate (RuBP) (Keown et al., 2013) forming an unstable intermediate, enediol. Finally, two molecules of 3-phosphoglycerate (PGA) are formed as the first stable product of photosynthesis by Rubisco-catalyzed carboxylation of RuBP (Gomez & Huovinen, 2012).

In the reduction phase, PGA is converted into glyceraldehyde 3-phosphate (GAP), also called triose phosphate, using ATP and NADPH formed in the photochemical reactions by GAP-dehydrogenase (GAP-DH). Finally, the regeneration phase allows the continuation of the cycle operation by resupplying RuBP from triose phosphate molecule.

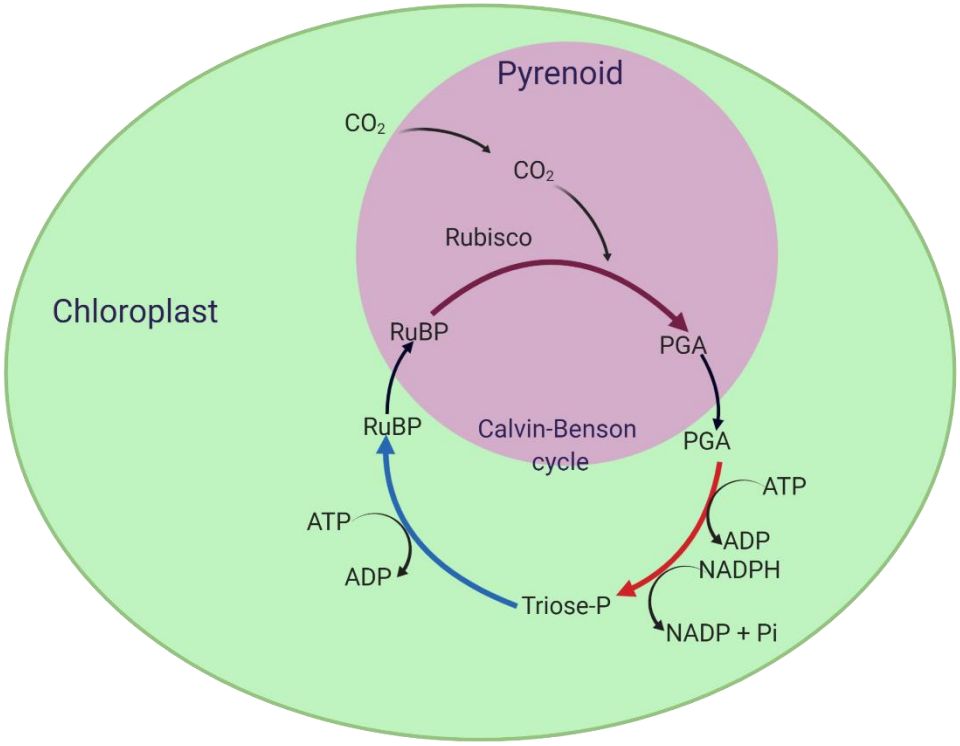

Figure 1.2 *Calvin-Benson cycle in the chloroplast*
Carboxylation (purple arrow) takes place in the pyrenoid, where RuBP conversion into PGA is catalysed by Rubisco and requires a supply of CO_2. RuBP diffuses passively from the chloroplast stroma into the pyrenoid, while PGA formed flows from the pyrenoid into the stroma. Reduction (red arrow) and regeneration (blue arrow) take place in the chloroplast stroma. Figure created with BioRender.com

.

1.3.2. Oxygenation: Rubisco carbon fixation obstacle

The central role of Rubisco in carbon fixation might make it look like an ideal enzyme to engineer to improve algal biomass yields or to mitigate atmospheric carbon increases. However, Rubisco has two catalytic activities, one wanted and one unwanted, which makes its engineering for carbon fixation harder. Rubisco catalysis begins with carbamylation of an uncharged ε-amino group of a Lysine residue by non-substrate CO_2 followed by coordination of Mg^{2+} at Rubisco's active site to one of the carbonyl oxygens of the carbamate (Figure 1.3), the result of which is an inactivated enzyme (Lorimer & Miziorko, 1980). When RuBP bonds with this carbamate, it can either be combined with CO_2 to form two molecules of 3-phophoglycerate (the desired carboxylation reaction) or with O_2 to form one molecule of 3-phosphoglycerate and another of 2-phosphoglycolate, in an undesired oxygenation reaction called photorespiration. As a consequence of the oxygenase activity of Rubisco, the cell expends a significant amount of energy which inhibits

biomass formation of around 50% (Giordano et al., 2005; Kumar et al., 2011). Moreover, the 2-phosphoglycolate produced in this step is useless and, during its breakdown, liberates CO_2 which has been previously fixed by the carboxylase activity, limiting photosynthetic productivity (Kumar et al., 2011; Atkinson et al., 2017).

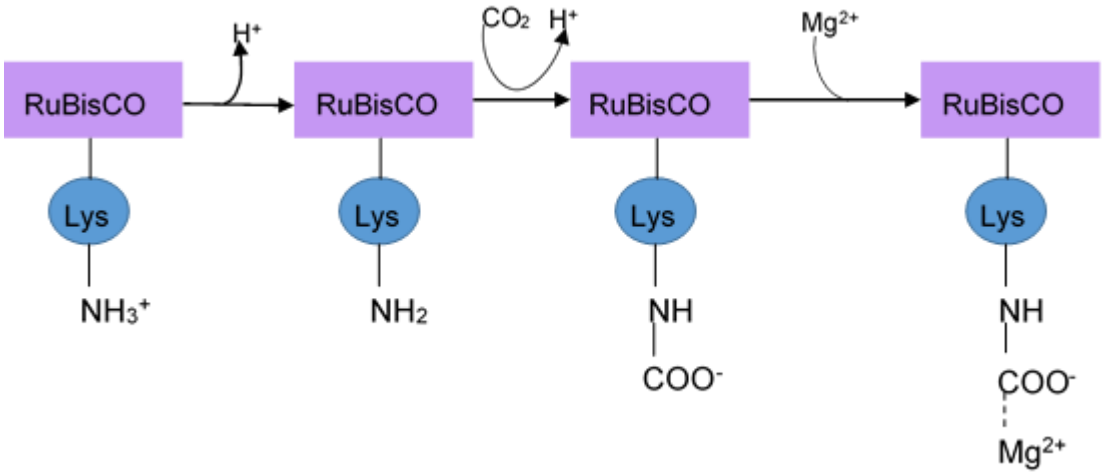

Figure 1.3 *Formation of Rubisco carbamate*

The catalytic reactions carried out by Rubisco are unique to photosynthetic CO_2 fixation as a Ci source, therefore, imperfections during this photosynthetic process are mainly those of a catalytic turnover per active site and its oxygenase reaction exposed above (Badger et al., 1998). Also, another existing limitation for marine organisms' CO_2 uptake is physicochemical-related: even when $ppCO_2$ in water is about the same as in the air, CO_2 diffusion coefficient in seawater is lower than in air ($1.9x10^{-5}$ cm^2s^{-1} and 0.16 cm^2s^{-1} respectively (Goldman & Dennett, 1983)), hence the need for an optimizing-performance strategy such as a carbon concentration mechanism (CCM).

Because of the low CO_2 affinity of Rubisco (K_m=48.9 ± 2.8 µM; (Boller et al., 2015)), its efficiency can be enhanced by elevating the CO_2 concentration in the chloroplast through the CCM (Atkinson et al., 2017). This CCM activates at low dissolved carbon concentration and the concentration at which CCMs are active depends on light, pH and cell adaptation. (Kumar et al., 2011). As in the atmosphere, CO_2 and O_2 compete at Rubisco's catalytic site, the CCM increases the CO_2 concentration around Rubisco thus the carbon fixation is improved (Mackinder et al., 2016).

1.4. The pyrenoid: a CO_2 fixation centre

Algae have a unique organelle associated with their CCM: the pyrenoid (International Scientific Vocabulary, from New Latin *pyrena:* stone of a fruit). The pyrenoid is a subcellular, membrane-less, proteinaceous microcompartment buried in the chloroplast stroma of many eukaryotic algae (Tanaka et al., 2007; Rochaix, 2017), and which sequesters the primary carboxylase, Rubisco (Moritz et al., 2012). Up to 90 percent of the pyrenoid consists of Rubisco and its chaperone, Rubisco activase (Villarreal & Renner, 2012; Mackinder et al., 2016). This microcompartment plays an important role in the activation of any CCM, as CO_2 levels near the pyrenoid are up to 180 times above normal CO_2 concentrations in the rest of the cell, enhancing the efficiency of carbon assimilation and countering the diffusion limitations inherent in aquatic photosynthesis (Moritz et al., 2012; Villarreal & Renner, 2012).

However, even though Rubisco is the main constituent of pyrenoids, it is not the only one: they also contain a starch deposit and a matrix where CO_2 is concentrated along with Rubisco (Rochaix, 2017). Previous studies carried out in *Chlamydomonas reinhardtii* have revealed that the pyrenoid structure and composition is more complex than previously assumed (Moritz et al., 2012; Mackinder et al., 2016; Zhan et al., 2018), with recent discoveries showing that Rubisco accumulation was possible through the action of a disordered repeat protein, called Essential Pyrenoid Component 1 (EPYC1) (Mackinder et al., 2016).

1.4.1. Importance of EPYC1 in *Chlamydomonas reinhardtii* pyrenoid constitution

Recent investigations carried out in the green microalga, *Chlamydomonas reinhardtii*, have provided information about the structure and composition of the pyrenoid, which seems to be an organelle with liquid-like properties, instead of a solid or crystalline structure (Mackinder et al., 2016; Freeman Rosenzweig et al., 2017; Atkinson et al., 2019).

The component that maintains many of the *Chlamydomonas* pyrenoid characteristics, such as phase-separated liquid behaviour, normal size, number, morphology, Rubisco content, and efficient carbon fixation at low CO_2 conditions, is the protein EPYC1 (Rochaix, 2017). This protein is intrinsically disordered and

contains four repeats in its amino acid sequence; it is predicted to bind Rubisco through multiple low affinity interactions that stimulate phase transitions, allowing dynamic internal reorganization within the pyrenoid (Atkinson et al., 2019).

Mackinder et al., 2016, proposed two possible arrangements between EPYC1 and Rubisco (Figure 1.4), illustrating how EPYC1 builds a network together with Rubisco in which each EPYC1 binds four Rubiscos, acting as a “molecular glue” (Rochaix, 2017). The exact nature of this interaction remains unknown; however, if the interactions are weak, then the EPYC1-Rubisco complex could explain both the structure and the fluidity of the pyrenoidal matrix (Freeman Rosenzweig et al., 2017).

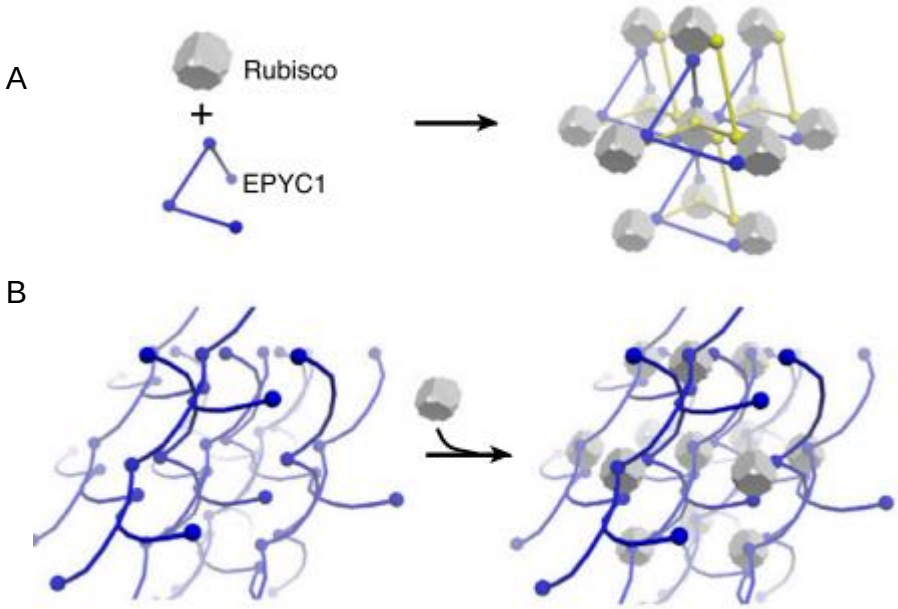

Figure 1.4 *Rubisco-EPYC1 proposed arrangements.*
(A)Each EPYC1 binds four Rubisco holoenzymes forming a co-dependent network. (B) A scaffold of EPYC1 could bind Rubisco holoenzymes. Both arrangements could expand indefinitely in every direction. Each sphere throughout EPYC1 structure represents an amino acid repeat in its sequence. Figure from Mackinder et al., (2016).

Further studies have described how the lack of EPYC1 in the pyrenoid affects its normal functioning and structure: an *epyc1* mutant with reduced levels of EPYC1 required higher CO_2 concentrations for photoautotrophic growth, and is unable to concentrate CO_2 effectively. Moreover, the pyrenoid decreased in size, having a reduced matrix density and seeing Rubisco relocalized to the chloroplast stroma (Mackinder, 2018; Wunder et al., 2018). Because the existence of such a protein, acting like a scaffold, allows the normal operation of the pyrenoid (Figure 1.5), a well formed pyrenoid reduces the leakage of CO_2, conversely, if the pyrenoid is not correctly formed, then CO_2 will leak out of the cell (Mukherjee & Moroney, 2019).

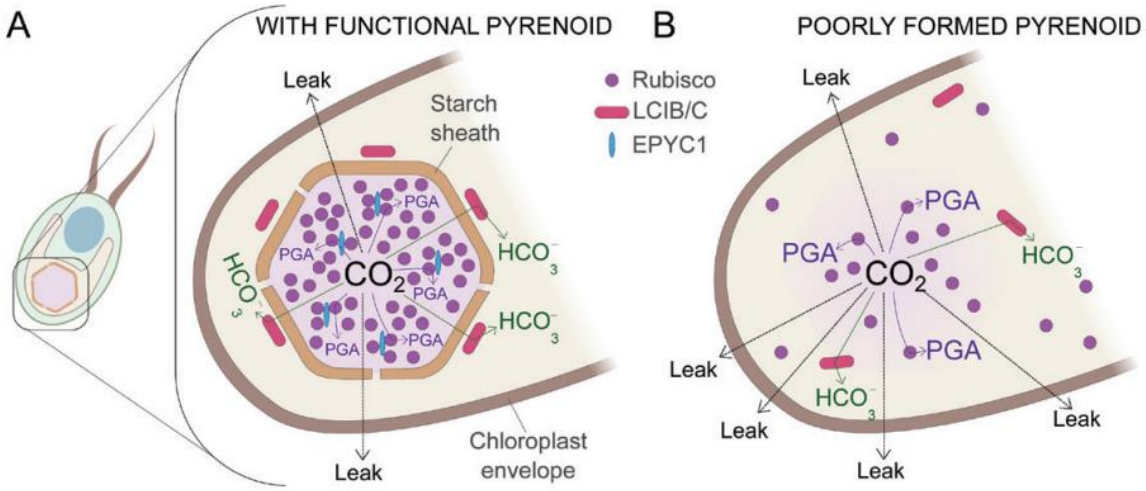

Figure 1.5 *The pyrenoid structure: its effect in CO_2 trapping.*
(A) In a functional pyrenoid Rubisco interacts with EPYC, acting like a scaffold and aiding in the formation of the pyrenoid. The Rubisco product glycerate-3-phosphate (PGA) is formed when CO_2 concentrates in the pyrenoid. Leaked CO_2 is sometimes recaptured and converted to HCO_3^- by LCIB/C, a carbonic anhydrase in the stroma, or remains in the chloroplast. (B) Absence of EPYC1 does not allow the formation of the pyrenoid, so accumulated HCO_3^- once converted to CO_2 easily leaks out and is not recaptured. Figure from Mukherjee & Moroney,(2019).

Additionally, it has been demonstrated that EPYC1 is responsive to CO_2 concentrations, with the level of EPYC1 in the pyrenoid increasing by ~12-fold after algae shift from high to low CO_2, comparable to the changes seen in Rubisco (Mackinder et al., 2016). Accordingly, knowledge of how the presence of EPYC1 improves the efficiency of algal CO_2 metabolism is important if we are to engineer changes in biomass yields, because EPYC1 is essential for pyrenoid formation and CCM functioning, allowing autotrophic growth as well as CO_2 mitigation.

2. Aims and hypotheses

Previous work has elucidated the cellular components implicated in carbon fixation across a range of algal photosynthetic pathways, such as those in *Chlamydomonas* and some cyanobacteria (Atkinson et al., 2017; Linthwaite et al., 2018). However, these species were all unicellular microalgae and amenable to genetic transformation techniques, allowing pyrenoid mutants to be generated and used to identify pathway components (Keown et al., 2013; Mackinder et al., 2016).

My project focusses on carbon concentration mechanisms in the related multicellular green macroalgae. There are known to be metabolic differences between micro- and macro-algae, not least because the macroalgae are much more exposed to tidal changes in their daily environment. To look at macroalgal carbon concentration, I chose to work on the green macroalgal species, *Ulva* spp*. Ulva* spp*.* are extensively distributed along temperate coasts, penetrating freshwater streams as well (De Clerck et al., 2018). *Ulva* can cause “green tides”, covering hundreds of miles of coastline; while not directly toxic, the biomass rots to produce dangerous hydrogen sulfide (De Clerck et al., 2018). However, this rapid growth also makes *Ulva* biomass a promising feedstock for biofuels.

Accordingly, my project began by addressing the following hypotheses:

2.1. Hypotheses

- The carbon concentration mechanism in *Ulva* involves CO_2-binding proteins that are specific to the green macroalgae.
- Photosynthetic efficiency may be improved by modification of these CO_2-binding proteins.

Due to the 2019-2020 global coronavirus pandemic, this work is focussed on the first of these hypotheses.

2.2. Approaches

Genetic transformation techniques are only slowly being developed in the macroalgae, which rules out the use of a mutagen in *Ulva* spp. Accordingly, to know which subcellular components (i.e. proteins) are implicated in CO_2 metabolism in *Ulva,* I chose the following approaches:

1. Develop a protocol for the preparation of pyrenoid/chloroplast fractions from *Ulva* spp. to allow biochemical and biophysical comparisons between samples. This would represent one of the first attempts to fractionate macroalgal cell components in this way
2. Bioinformatic searches for candidate proteins using the *Ulva* genome sequence
3. Manipulation of CO_2 concentration in *Ulva* culture media to cause changes in carbon concentrating proteins
4. Proteomic analysis to assess changes in candidate proteins

2.3. Outcomes

The immediate intended outcomes from my work are:

- Identification of protein(s) enriched in the pyrenoid as being possible candidates in green macroalgal carbon concentrating mechanisms
- Characterisation of protein responses induced by changes in CO_2 concentration, to assess the dynamic responses of candidate proteins

Data obtained from my project will therefore help to elucidate how the carbon concentrating mechanism in *Ulva* works and in which instances *Ulva* is sensitive to inorganic carbon changes. This will identify targets for study and engineering to improve macroalgal biomass yields for biofuel generation, as well as having implications for the mitigation of environmental impacts on marine ecosystems.

3. Materials and methods

All solutions used in this project were prepared using MQ water.

3.1. Sample collection

Ulva spp. biomass samples were collected in Seaham Harbour in county Durham. The timing of biomass collections depended on weather and tidal conditions because *Ulva* is best reached when low tide is around about 0.5 m above datum. The biomass collected was taken to the laboratory in containers with natural seawater in order to reduce any stress.

Once in the laboratory, samples were cleaned and rinsed, in order to remove diverse debris and epiphytes, and were then immersed in a preparation of artificial seawater (ASW) for maintenance at 8°C ± 2°C with constant agitation (80 rpm) and a 16:8 h light:dark cycle; ASW was replaced once a week to ensure the good condition of the material.

For the preparation of ASW for biomass maintenance, the salts were dissolved in the concentrations shown below with a stirring plate (IKA, RCT basic).

Table 3.1 *Artificial sea water preparation*

Chemical	**Concentration (g/l)**
Sea salts mix	35
Sodium nitrate	0.070
B-glycerophosphate disodium salt hydrate	0.010

All chemicals were from SIGMA

3.2. Protoplast isolation

Protoplast preparation was adapted from previous methods (Reddy et al., 2006), with some modifications; the overall protocol is shown diagrammatically in figure 3.1.

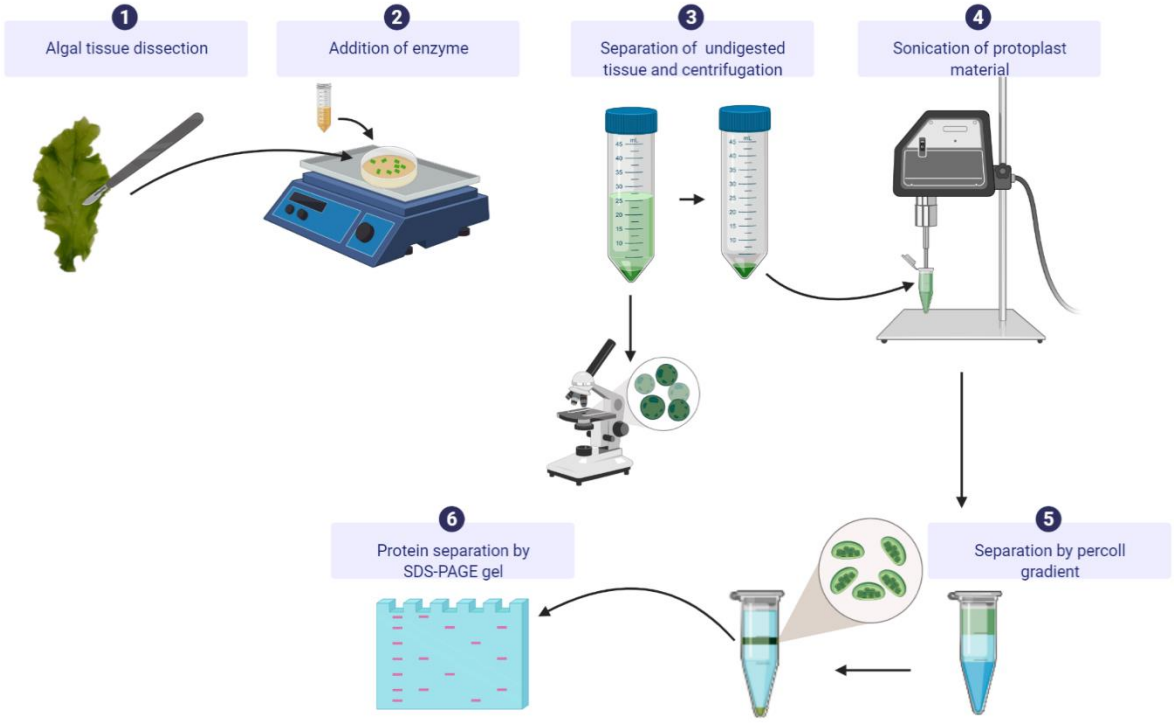

Figure 3.1 *Overview of chloroplast fraction preparation.*
Steps 1-3 correspond to protoplast isolation. Steps 4 and 5 correspond to chloroplast enriched fraction separation. Figure created with BioRender.com.

Fresh *Ulva* thalli were washed with ASW and then with autoclaved MQ water, the tissue was chopped into ~1 mm^2 pieces with a scalpel and these were washed with cold autoclaved MQ water.

Cell wall digestion then began by treating these fragments with a 2% enzymatic solution prepared dissolving 0.2 g of cellulase Onozuka R-10 (Duchefa Biochemie) in 20 ml of protoplast extraction buffer (PEB), pH 6 (Table 3.2). The solution was cleared by centrifugation (Avanti 30, Beckman Coulter, USA) at 10,000 g, 4°C for 20 minutes. Enzyme solution was added to the chopped tissue and incubated in the dark at 20°C for 2 h in a rotary shaker (KS501, IKA, Germany) at 40-50 mot/min.

Table 3.2 *Protoplast extraction buffer*

Chemical	**Concentration**
Dextran sulphate	0.5%
NaCl	1%
MES	25 mM
$CaCl_2$	1 mM
Mannitol	0.8M

The resulting material was filtered through miracloth (pore size 22-25 µm) and the undigested debris was discarded. The filtrate was centrifuged at 200 g for 5 min, and the resulting pellet was twice washed by replacing half of its supernatant with ice cold PEB and by re-centrifuging. Finally, all the supernatant was discarded and the pellet was resuspended in 200 µl of ice cold PEB with a protease inhibitor cocktail (cOmplete ULTRA tablets, Mini, EASY pack. Roche) according to the supplier's instructions.

3.3. Chloroplast enrichment

The protoplast preparation was centrifuged at 200 g for 5 min and the pellet was resuspended in 1.5 ml ice cold protein extraction buffer (EB) (Table 3.3). Protoplasts were then lysed by sonication (Soniprep 150, MSE UK, London, UK) at 6 x 30 s bursts of 20 microns amplitude, with samples being placed on ice for 15 s between bursts. Lysate was clarified by centrifugation at 15,000 g, 4°C for 20 min and lysis was checked under the light microscope by confirming that protoplasts had been ruptured and chloroplasts released.

Table 3.3 *Protein extraction buffer*

Chemical	Concentration
Bicine	50 mM
$NaHCO_3$	10 mM
$MgCl_2$	10 mM
DTT	1 mM

pH 8

To enrich chloroplasts from this lysate, the lysate was spun down on a percoll (Sigma-aldrich) gradient. This gradient consisted of 500 µl 40% percoll layered on top of 500 µl 80% percoll, percoll was ice-cold and diluted in 0.25 M sucrose. Cell lysate was carefully placed on the upper layer of the gradient to avoid mixture of the fractions and centrifuged at 25,000 g, 4°C for 15 min.

3.4. Protein quantification and fraction separation

The chloroplast enriched fraction, placed between 40% and 80% percoll layers, as well as the pellet were each removed with a syringe and resuspended in 100 µl EB. The protein content was determined using the Pierce BCA Protein Assay Kit (Thermo Fisher Scientific) in microplates using 2 µl and 5 µl of both resuspensions per well.

The soluble proteins in each fraction were also separated on a precast Expedeon RunBlue SDS 4-20% gel; loading was normalised with 10 µg of protein per lane and using 3 µl of colour protein standard (P7719G, BioLabs). The gel was run for 40 min at 170 V in 1X running buffer (Table 3.4).

Table 3.4 *20x running buffer*

Chemical	Concentration % (w/v)
MOPS	12.56
TRIS	14.52
SDS	2
Sodium disulfate	0.52

3.5. Confocal fluorescence imaging

Confocal imaging of the protoplast preparation, as well as of the chloroplast fraction, was carried out with a Leica SP5 Confocal Laser Scanning Microscope, using a Leica 63x HCX PL APO CS Water lens and an Argon laser (488 nm excitation and emission between 650-700 nm).

3.6. Protein extraction from intact algal tissue

In order to further investigate enrichment in the chloroplast fraction, as well as for the CO_2 conditions experiment (see p.20), a protein extraction of the intact tissue was carried out as follows (Figure 3.2).

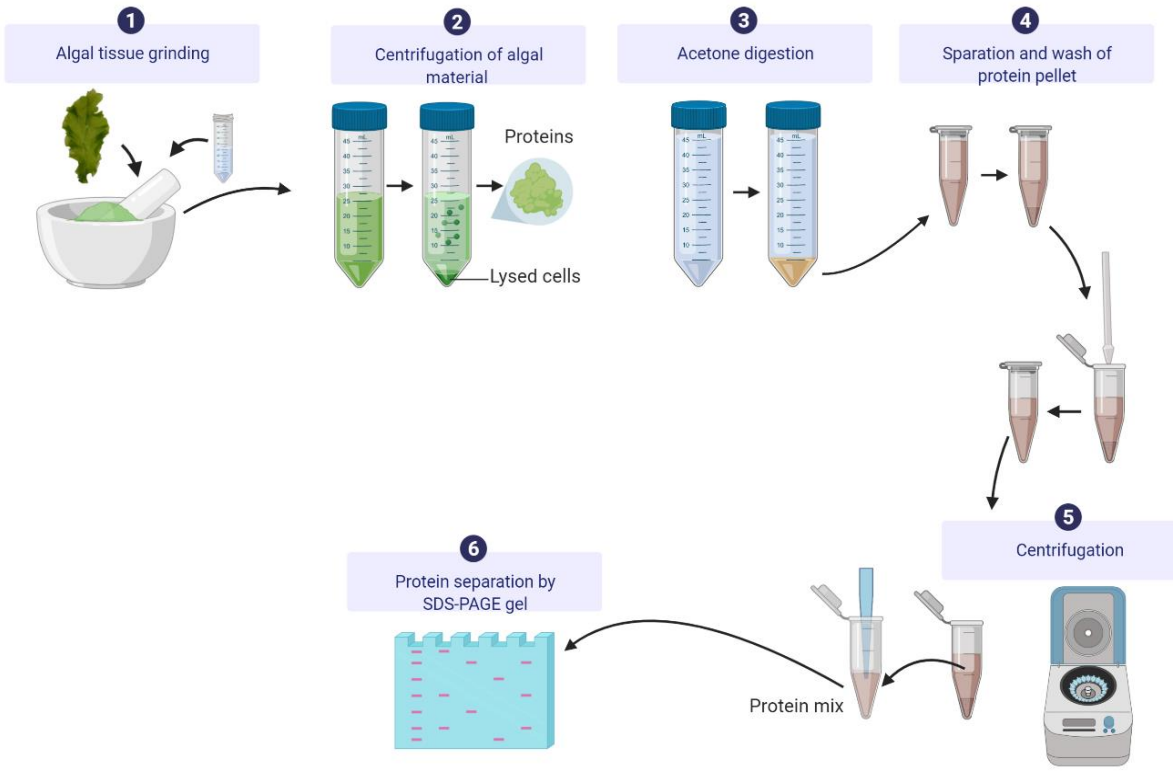

Figure 3.2 *Overview of protein extraction from intact Ulva spp. tissue.*
Protein extraction was carried out from the whole tissue until protein separation in SDS-PAGE gel.
Figure created with BioRender.com.

Initially, 2.5 g of fresh algal tissue was washed with sterile MQ water to remove excess seawater and dried with paper tissue. Next, the algal tissue was ground to a fine powder under liquid nitrogen and resuspended in ~5 ml of grinding buffer (10 mM TRIS, 1 mM EDTA, 1% v/v 2-Mercaptoethanol) in the mortar. The homogenate was transferred to a falcon tube; Triton X-100 was added to a concentration of 1% and the tube was vortex mixed for 2 minutes and clarified by centrifugation for 10 min at 3000 g. The pellet was discarded and 100% acetone was added to the supernatant in a 1:4 proportion and samples allowed to precipitate overnight at -20°C.

The following morning, the precipitate was pelleted by centrifugation for 10 min at 3,000 g and the supernatant discarded almost completely; the pellet was left with ~0.5 ml of supernatant remaining to ease its transfer from the falcon tube to an Eppendorf tube. The pellet was then centrifuged again (10 min at 17,000 g) to remove the residual acetone and resuspended in 200 µl lysis buffer (9 M urea, 2 M thiourea, 4% CHAPS).

3.7. Protein quantification and separation from intact algal tissue

A Bradford Assay was used to quantify proteins in the samples using bovine serum albumin (BSA) as a standard and using 3 µl of sample.

Soluble proteins were also separated on a precast Expedeon RunBlue 12% Bis-Tris Gel, loading was normalised by 10 µg of protein per lane and using 2 µl of SDS7 protein standard (SIGMA-ALDRICH). The gel was run for 45 min at 200 V in 1X NuPAGE MOPS buffer (Invitrogen).

3.8. Bioinformatic searches for candidate proteins

In order to find proteins involved in carbon metabolism, a series of predictions were carried out using a range of specific prediction programs (Table 3.5). The whole proteome of *Ulva mutabilis* was downloaded from OrcAE (https://bioinformatics.psb.ugent.be/orcae/, retrieved on 15 august 2019). The molecular weight (Mw) and isoelectric point (pI) were batch predicted using ExPASy, followed by subcellular localization prediction using DeepLoc-1.0 predictor.

The second wave of predictions were focused on finding proteins in *Ulva* with physicochemical similarities to EPYC1. To achieve this, tandem repeat motifs (TR) were looked in the whole proteome to narrow down the list of possible candidates. This analysis was carried out in XSTREAM with default settings, except for the following: min period, 40; max period, 80; min copy number, 2.5; min TR domain, 75; min seq. content 0.7. These setting modifications were made in order to find a putative protein with similar EPYC1 TR characteristics, where the period corresponds to the repeat length, copy number to the number of repeats in the sequence, Min TR domain to the number of characters in the whole TR, and min seq. content is the minimum percentage of the input sequence that should be covered by TR.

Next, natural disordered profiles were obtained from those proteins with repeat motifs using the VLTX program in PONDR and, finally, disordered protein sequences were analysed in TMHMM Server v. 2.0 to look for transmembrane domains (TMD).

Table 3.5 *Programs used to carry out protein physicochemical properties predictions*

Program	Webpage	Citation
pI/Mw tool	https://web.expasy.org/compute_pi/	
DeepLoc-1.0: Eukaryotic protein subcellular localization predictor	http://www.cbs.dtu.dk/services/DeepLoc/	(Almagro Armenteros et al., 2017)
XSTREAM: Variable sequence tandem repeats extraction and architecture modelling	https://amnewmanlab.stanford.edu/xstream/	(Newman & Cooper, 2007)
PONDR: Predictor of Natural Disordered Regions	http://www.pondr.com/	
TMHMM Server v. 2.0: Prediction of transmembrane helices in proteins	http://www.cbs.dtu.dk/services/TMHMM/	(Krogh et al., 2001)

3.9. CO_2 conditions

To start the experiment, 5 g of *Ulva* spp. thalli were transferred to each of two flasks with 500 ml of ASW in an incubator at 20°C and constant light (Figure 3.3). There were 2 sealed flasks bubbled with pure CO_2 until a pH of 7.5 was reached. The pH was monitored and measured every 4 to 8 hours; as the pH increased to 8, the culture was bubbled with pure CO_2 until the pH dropped to 7.5 again.

For the low carbon controls, two more flasks were set with the same amount of seaweed and the same volume of ASW and they were bubbled with compressed air until the culmination of the experiment. As with the CO_2 treatment, pH was measured regularly, maintaining a pH of 8.5 ± 0.2 without any pH adjustment.

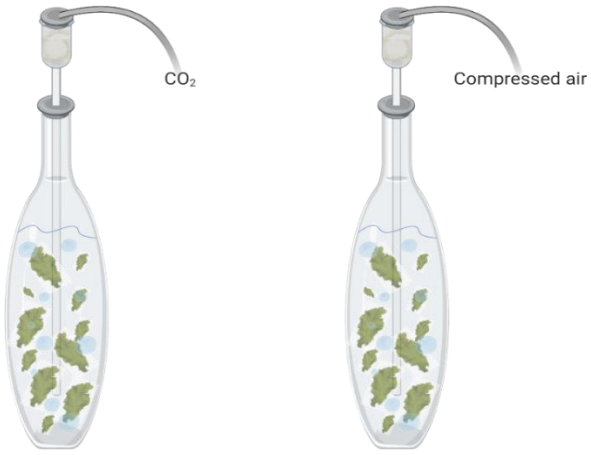

Figure 3.3 *High and low CO_2 experiment setting.*
Figure created with BioRender.com.

The harvesting took place after 5 days (120 h) and the seaweed samples were washed with sterile MQ water and excessive moisture was removed with paper tissue. Next, the content of each flask was split in two (2.5 g of fresh seaweed each) to get 4 replicates of high carbon treated thalli (HC) and 4 low carbon controls (LC). Finally, the samples were frozen in liquid nitrogen and maintained at -80°C until protein extraction (see p.17).

3.10. Protein digestion and isobaric tagging for relative and absolute quantitation (iTRAQ) labelling

Protein extracted from the whole tissue, treated with CO_2 as well as the controls, were taken to carry out an iTRAQ labelling as follows.

A volume containing 20 µg of total protein was taken from each sample, then 5 µl of TRIS-1.5 M HCl pH 8.8 was added, and volume made up to 100 µl with MQ water in an Eppendorf tube. The protein preparation was briefly vortexed and 400 µl of 100% acetone were added to the samples, which were then left for 1 h at room temperature. The tubes were then centrifuged for 10 min at 17,000 g and the supernatant was discarded.

For the following steps, iTRAQ Reagent 8 Plex Buffer kit (AB Sciex, P/N: 4381664) was used. 2 µl of denaturant (2% SDS) were added to the pellet before incubation at 60°C for1 h. Afterwards, the sample was pulse-centrifuged, followed by the addition of 48 µl of dissolution buffer, and the tube was placed in a shaker for 20 min. Then, 2 µl of reducing reagent were added and the samples were incubated for 1 h at 6°C. After this incubation, 1 µl of cysteine blocking reagent was added, followed by a brief vortex mixing and 10 min rest at room temperature. At this point, the samples were ready for the trypsin digestion, which was carried out as explained below.

The digestion was made with Trypsin Gold, mass spectrometry grade (Promega, USA). To start with, the trypsin (100 µg, dry weight), was resuspended in 200 µl of buffer composed of 4 µl of 0.115 M HCl, 22 µl of acetonitrile (ACN) and 206 µl of MQ water. 4 µl of trypsin preparation were added to each sample and these were incubated overnight at 37°C. Next, the samples were freeze dried and the resulting powder was resuspended in 50 µl of MQ water and placed in a vortex shaker.

The tags were diluted with isopropanol (included in iTRAQ Reagent 8 Plex Buffer kit), and 50 µl of them were added in order to label four replicates of HC and four replicates of LC as shown below:

Table 3.6 *iTRAQ tags used for each sample*

Sample number	Sample ID	Tag ID
1	HC1	113
2	HC2	114
3	HC3	115
4	HC4	116
5	LC1	117
6	LC2	118
7	LC3	119
8	LC4	121

To ensure the correct labelling of the samples, the pH was checked with pH test strips 4.5-10.0 (Sigma, USA) to ensure the samples had a pH of 7.5. After 2 h incubation of the tags at room temperature, all samples were mixed together in one Eppendorf tube and later divided into 3 tubes.

Finally, the tubes were freeze dried and sent to be analysed by LC-MS.

4. Results

4.1. Intact protoplasts can reliably be isolated from thalli

The isolation of protoplasts was a preliminary step to purify the pyrenoid and avoid endogenous enzymatic activity. The first attempts to isolate protoplasts from *Ulva* spp. were done as described by Reddy et al., (2006) with their modified enzyme preparation. Subcellular material was obtained, but this material did not always form obvious protoplasts; instead there was often significant damaged material (Figure 4.1).

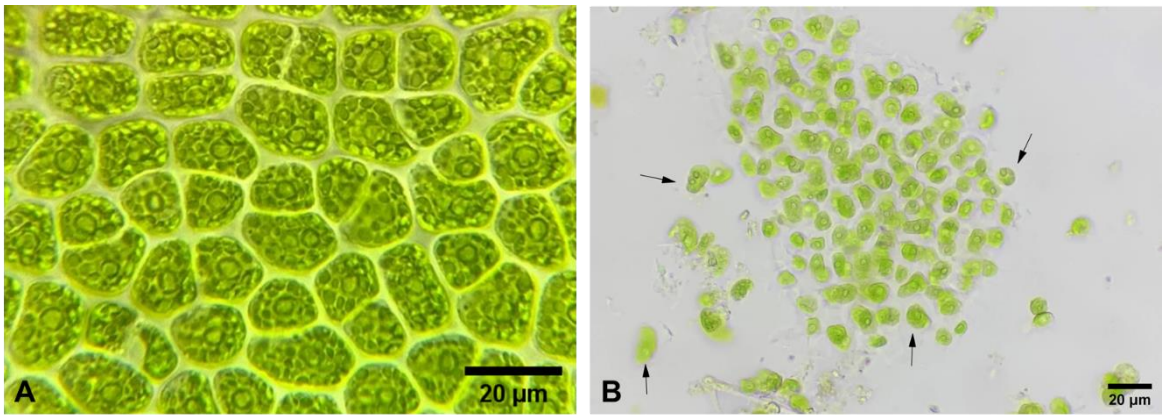

Figure 4.1 *Ulva cell organisation before and after enzymatic treatment*
(A) Normal tissue organisation of Ulva thalli before cell wall enzymatic digestion. (B) Cellular material obtained before $CaCl_2$ addition to the enzyme preparation. Disruption of cellular membrane is obvious (arrows).

To improve the protoplast yield, two main modifications were made: $CaCl_2$ was added as a membrane stabilizer to the enzyme preparation (Saure, 2005) and the protoplast pellet was washed with PEB instead of MQ water (Table 4.1). A range of $CaCl_2$ concentrations were tested by adding 1 mM, 3 mM and 5 mM $CaCl_2$ to the enzyme preparation. Results showed an optimum $CaCl_2$ of 1 mM $CaCl_2$, which gave improved protoplast yield: more intact protoplasts were seen with their cell membrane and spherical shape, as shown by light and confocal microscopy (Figure 4.2). Interestingly, higher concentrations of $CaCl_2$ (3 mM and 5 mM) did not show significant improvement compared to 1mM $CaCl_2$; instead, cell disruption was similar to disruption with no added $CaCl_2$.

Table 4.1 *Different variations in protoplast isolation*

$CaCl_2$ concentration	**Incubation time**	**Protoplast washing solution**	**observations**
1 mM	2 h	ASW	Cell disruption was observed under light microscope
		PEB	Spherical shaped protoplasts were observed under light microscope. Incubation time is the same as proposed in Reddy et. al., (2016)
	3 h	ASW	Protoplast quantity was lower than in the 2 h incubation
		PEB	
3 mM	2 h	ASW	Cell disruption was similar to non-$CaCl_2$ addition
		PEB	
	3 h	ASW	
		PEB	
5 mM	2 h	ASW	
		PEB	
	3 h	ASW	
		PEB	

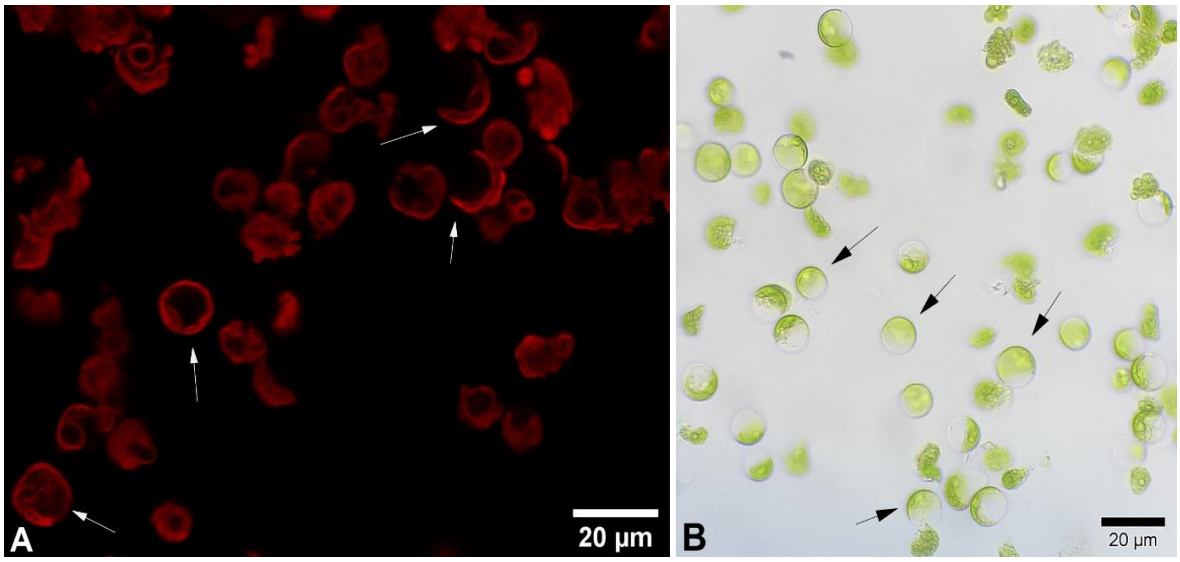

Figure 4.2 *Protoplasts isolated from Ulva spp. tissue.*
(A) Confocal microscopy of protoplast, intact protoplasts showing autofluorescence in red are indicated. (B) Light microscopy Intact protoplasts showing their cellular content surrounded by cell membrane.

4.1.1. Centrifugation through percoll gives chloroplast enrichment

As a second move towards pyrenoid purification, the protoplast preparation was lysed to allow further fractionation. The lysate obtained by sonication was inspected by confocal microscopy to check out the autofluorescence of chlorophyll and confirm the absence of intact cells. Figure 4.3 shows cellular debris that include both, red autofluorescence bits and some non-autofluorescence debris. Subsequently, percoll gradient aimed to separate non-autofluorescence debris to be discarded.

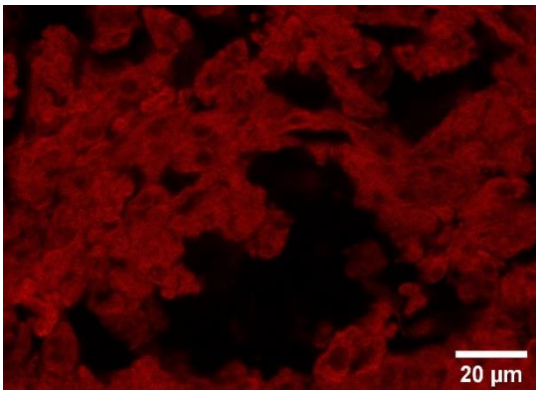

Figure 4.3 *Confocal microscopy of Ulva's cell lysate.*
Red autofluorescence of chlorophyll observed at 650-700 nm.

After protoplast lysis, two different separation gradients were tested: firstly, the percoll was diluted in MQ water, however, since the upper layer with the lysate did not separate completely, was decided to test diluting the percoll with 0.25 M sucrose and a defined green fraction was obtained as well as a small pellet (Figure 4.4).

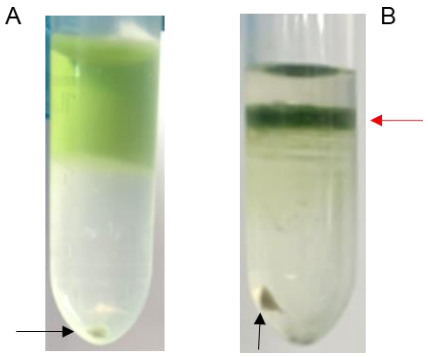

Figure 4.4 *Cell lysate percoll gradient.*
(A) Gradient with percoll diluted in MQ water with a visible pellet separation (black arrow). (B) Gradient with percoll diluted in 0.25 M sucrose, separation can be appreciated in a pellet (black arrow) and a green fraction between 40% and 80% percoll (red arrow).

4.1.2. The chloroplast enriched fraction contains proteins that correspond in size to Rubisco and EPYC1

In order to test whether proteins could be reliably extracted from the chloroplast-enriched fraction, a separation by SDS-PAGE gel was carried out. Multiple bands were visible (Figure 4.5) and two bands obtained around 24 and 30 kDa were of particular interest because their Mw were similar to EPYC1 in *Chlamydomonas reinhardtii* (Mackinder et al., 2016). Also, two bands, one around 55 kDa and other around 20 kDa were placed where rbcL and rbcS should respectively be.

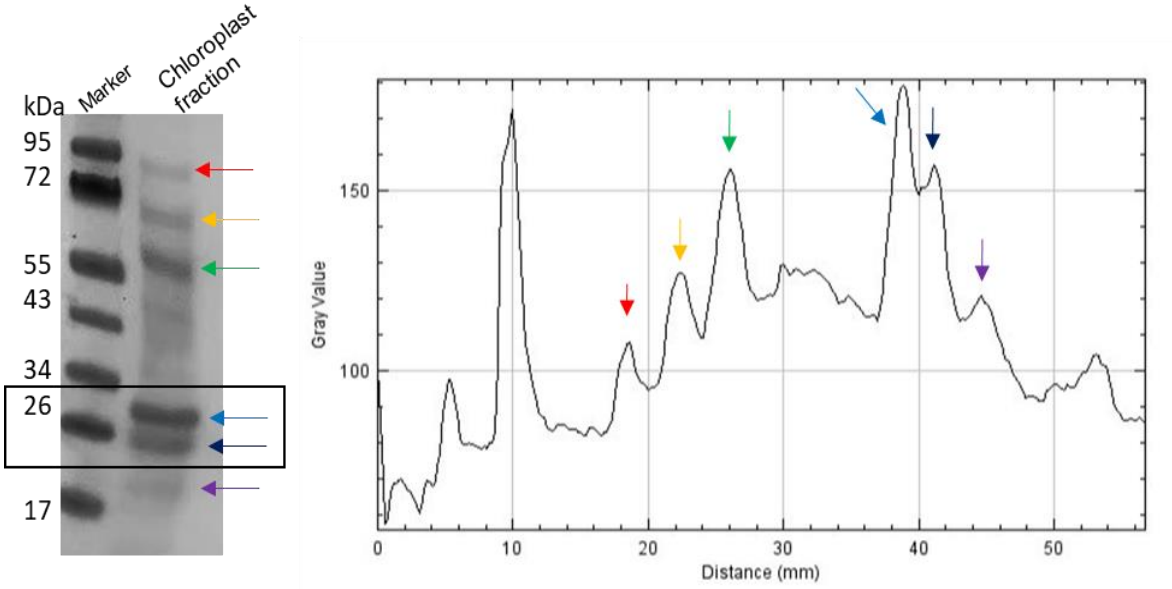

Figure 4.5 *SDS-PAGE of chloroplastic fraction.*
Each band is indicated with a different coloured arrow. Bands in the box are those with similar Mw as EPYC1 in Chlamydomonas reinhardtii. Each peak in the profile plot corresponds to one band in the gel, where the x axis indicates the band mobility through the gel in distance (mm), and y axis to the grey value of each band. Plot generated with Fiji (Schindelin et al., 2012).

4.1.3. The protein complement of chloroplast-enriched fractions differs from that of whole algal tissue

The qualitative comparison between the protein extract from the chloroplast-enriched fraction and that from whole *Ulva* tissue showed strong differences in protein complements, representative separation of 3 runs is show in figure 4.6, and it exemplifies the strong chloroplast band placed at 27 kDa in figure 4.5.

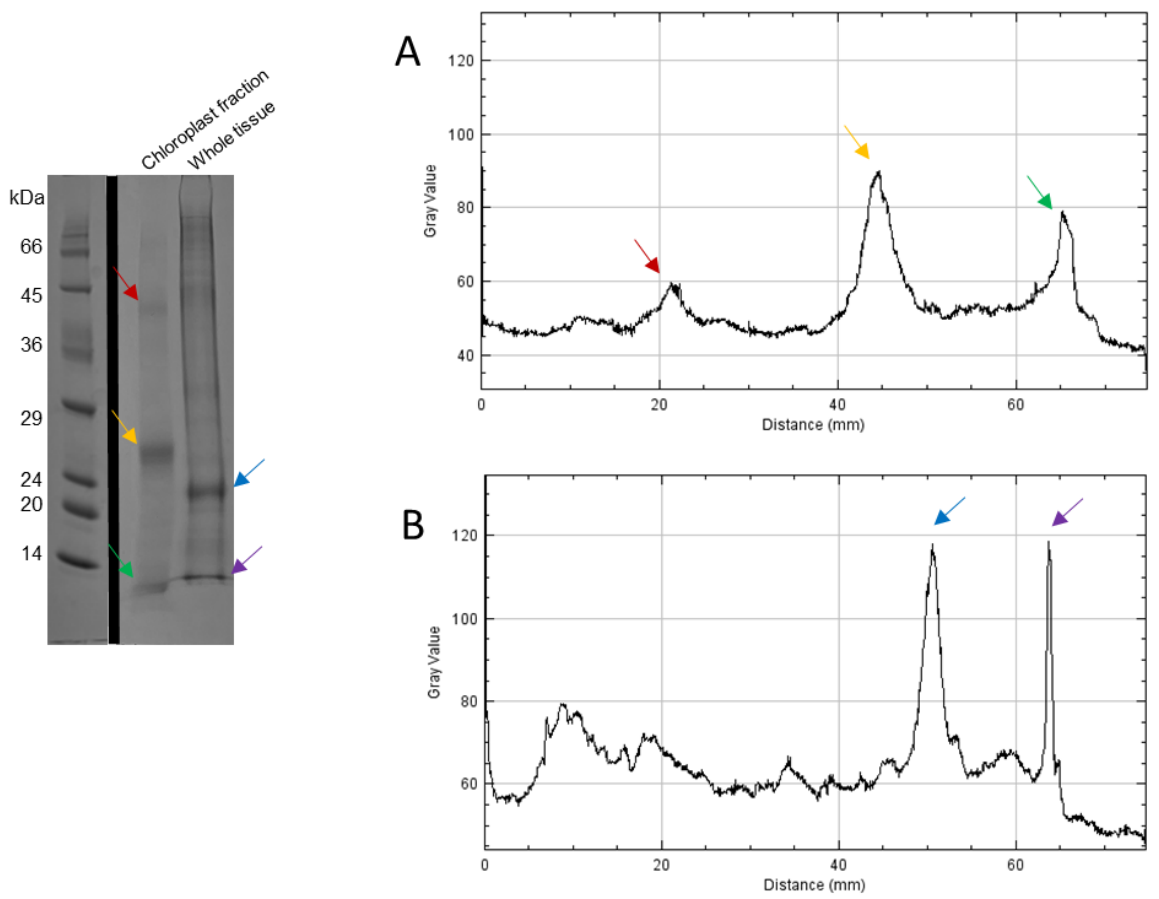

Figure 4.6 *SDS-PAGE of enriched fraction and whole tissue.*
Each band is indicated with a different coloured arrow. Each peak in the profile plot corresponds to one band in the gel. (A) Chloroplast fraction profile plot where three bands can be seen, one band at ~27 kDa (yellow arrow) corresponds to the band obtained in the previous experiment. (B) Whole tissue profile plot, with two main bands (blue and purple arrows). Each peak in the profile plot corresponds to one band in the gel, where the x axis indicates the band mobility through the gel in distance, and y axis to the grey value of each band. Plots generated with Fiji (Schindelin et al., 2012).

4.2. Bioinformatic analysis identifies a range of candidate genes for *Ulva* chloroplast proteins

To start the second strand of my research, bioinformatic analyses were carried out so that both the qualitative separation of the proteins present in the fractions and the actual presence of putative proteins with appropriate physicochemical characteristics in *Ulva* could be corroborated.

To achieve this, the complete proteome of *Ulva mutabilis* from OrcAE (Sterck et al., 2012), was used to search for proteins involved in CO_2 metabolism. I began with prediction of Mw in order to further analyse predicted proteins with similar Mw to EPYC1 in *Chlamydomonas reinhardtii*, followed by the prediction of pI, because of its influence in determining protein-protein interactions. At the same time,

subcellular localization was predicted in order to find candidate chloroplast-localised proteins.

Based on the size of the strong bands in the SDS-PAGE gels (Figure 4.5 and Figure 4.6), a longlist of all *Ulva mutabilis* proteins that met the following criteria was made:

- Predicted localization in the plastid, where CO_2 metabolism takes place;
- Mw between 20 kDa and 40 kDa, similar to EPYC1 in *Chlamydomonas reinhardtii* (33.3 kDa) which also corresponded to the band in the chloroplast fraction gel (Figure 4.5); and
- pI≥8, similar to EPYC1 high pI

As a result of the subcellular localization and Mw predictions (Figure 4.7), 1,129 proteins out of the total 12,924 proteins were predicted to be localised in the plastid, and from those, 407 proteins had a Mw between 20 kDa and 40 kDa.

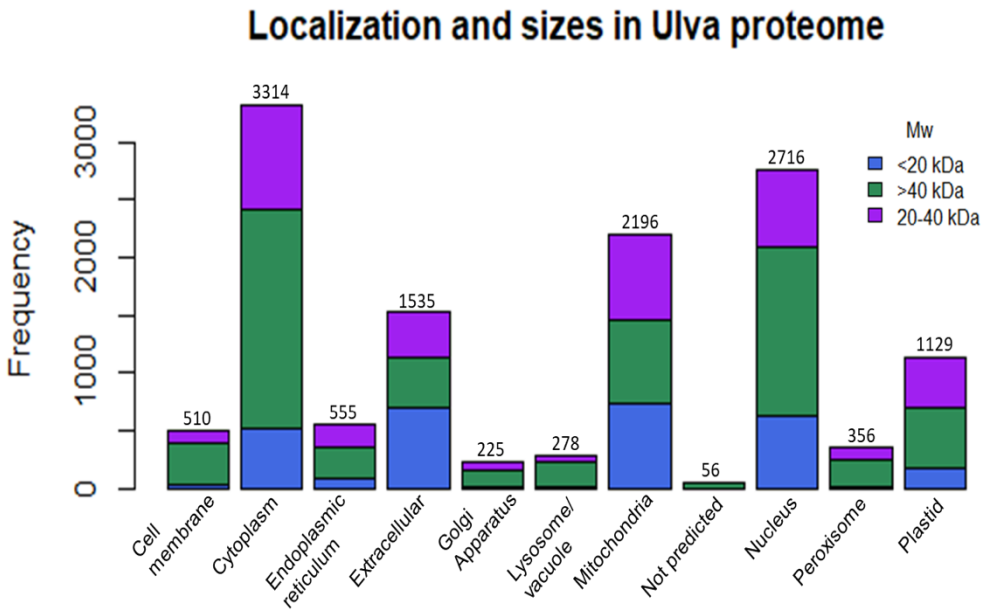

Figure 4.7 *Summary of localization predictions of Ulva mutabilis proteome.*
The colours indicate the molecular weight range of the proteins and the number over the bars corresponds to the quantity of proteins predicted to be in each localization.

Finally, taking into consideration a pI value ≥8, 180 proteins were left in the longlist and were BLASTed (Altschul et al., 1997) against the *Chlamydomonas reinhardtii* proteome. Proteins that resulted to have sequence similarity and are involved in CO_2 metabolism are listed in table 4.2:

Table 4.2 *Proteins involved in CO_2 metabolism*

Protein identifier (from OrcAE)	pI	Mw (kDa)	*C. reinhardtii* accession (NCBI)	Description/biological process
UM149_0031.1(248);mRNA;r:101697-102726	9.54	26.64	XP_001694126.1	Oxygen evolving enhancer protein 2
UM023_0013.1(282);mRNA;f:118668-120530	9.48	30.06	XP_001698344.1	Low CO_2 inducible protein
UM018_0004.1(357);mRNA;f:42411-45285	9.27	37.65	PNW72088.1	Chloroplast organization
UM042_0106.1(267);mRNA;f:491270-495376	9.2	29.38	AAB65498.1	Carbonic anhydrase
UM014_0218.1(338);mRNA;f:1165706-1167693	8.55	36.40	XP_001702409.1	Rubisco small subunit

Proteins listed above are those which met the following criteria: plastid localization, Mw= 20-40 kDa and pI≥8. 5 out of 180 proteins were found.

A supplementary table with proteins that did not meet the pI criteria (pI≥8), but did meet the Mw and predicted localization criteria, is shown below.

Table 4.3 *Proteins involved in CO_2 metabolism with sequence similarity proteins in C. reinhardtii **

Protein identifier (from OrcAE)	pI	Mw (kDa)	*C. reinhardtii* accession (NCBI)	Description/biological process
UM003_0220.1(206);mRNA;f:1339169-1341885	6.23	22.58	XP_001689751.1	Chloroplast-Rubisco complex assembly
UM005_0050.1(255);mRNA;r:252750-254771	7.1	28.05	PNW83629.1	Primary metabolic process
UM005_0087.1(240);mRNA;f:378452-380146	6.1	26.26	XP_001698115.1	Oxidation-reduction process
UM007_0115.1(378);mRNA;r:606660-609795	6.17	40.47	XP_001694768.1	Low CO_2-induced aldose reductase activity.
UM014_0011.1(361);mRNA;r:63734-67874	5.4	38.80	PNW82793.1	Oxido-reductase activity
UM015_0242.1(350);mRNA;f:1033436-1035325	6.75	37.40	XP_001691801.1	Oxidation-reduction process
UM038_0118.1(302);mRNA;f:470240-472422	7.13	32.73	XP_001695919.1	Low CO_2-induced protein
UM060_0084.1(370);mRNA;r:313712-315781	6.21	39.34	XP_001698344.1	Low CO_2-induced protein

**Proteins which met the criteria: Plastid localization and Mw= 20-40 kDa*

Concurrently and since the previous approach did not show up any EPYC1-like protein, a second approach was carried out in which *Ulva* proteome was screened, using the protocol in Mackinder et al. (2016). First, starting from the complete proteome, a search for proteins with tandem repeats (TR) in their sequence was carried out, because of the important role TRs play in protein architecture and linker functions. This search gave, as a result, a list of 11 proteins. This shortlist was screened to detect transmembrane domains, discarding those proteins with

signalling and transport activity. Finally, the last screening consisted in predicting disorder profiles, discarding those with non-disordered profiles.

As a result, a second, and final, list of candidate proteins was obtained (Table 4.4), and based on these predicted physicochemical properties, it can be seen that only two predicted *Ulva* proteins fulfilled EPYC1-like characteristics. However, one of these two (UM120_0037.1) had a suggestive Mw of 31.39 kDa, while its fellow candidate (UM048_0068.1) had a much larger Mw of 219.25 kDa.

The *Ulva* genome's most promising candidate (UM120_0037.1) amino acid sequence is shown below, with each TR in a different colour:

>UM120_0037.1(292);mRNA;f:179461-181230

MLSLRAPVSSARRAVVLNARRPQPDYAARRSGGREARRSGAATNRQPIQRRAS
VDYSSRRSGGEEENNRRRSGGYSSPTPSRSSSYSPPQRNSSPAPANRLVSQAV
LKRRSVDYAARRSGGREAQRSGSSPPARRASSSSSYSAPERNTSPAPANRLISQ
AVLKRRSVDYAARRSGGREAQRSGSSGRPSSSSSYSSPQRSSSPAPADRLVSQ
AVLKRRSVDYAARRSGGRAGSNDSRRRSSSFSSPTPSRSSSPAPAPANRLVSP
AVRKRRSVDYSARRSGSQGRRNSRR

Table 4.4 *Ulva protein candidates based on physicochemical properties.*

Protein identifier (from OrcAE)	pI	Mw (kDa)	Period	Repeat copy	Repeat sequence	Transmembrane domains	Oscillating disorder profile*
UM010_0052.1(222);mRNA;f:448889-449554	3.32	20.94	66	3.17	AVGDLAEV-D-V--A--VVGTAAVTAVGVMAVADLAEEDA-VVAGTAAVTAVGVMAVADLAE-E-A-VVAGTAAVT	5	No
UM113_0011.1(493);mRNA;r:266061-269025	3.39	50.16	57	6.49	ACDETNDACIVEAINEGGSCADGLFCNGAEICTAGVCAAADTDPCDDFSPVTACSRP	0	No
UM010_0048.1(361);mRNA;f:445655-446737	3.4	33.84	48	7.21	MAVA-DRVEE-AWV-VVGTAAVTAVGVMAVGD-LAEVDVAVV-GTAAVTAVGV	10	No
UM010_0051.1(206);mRNA;f:448144-448761	3.74	19.60	48	3.92	GDLAEVDAAVAGTAAVTAVGVMAVADRVEE-AWV-VVGTAAVTAVGVMAV	6	No
UM015_0170.1(276);mRNA;f:753949-754776	3.89	26.30	54	5.09	MLCSCCKDEENGGASDASDGLQPLPGAALGSCCNESGLGWAGSASDGLRPLAGA	0	Yes
UM010_0050.1(187);mRNA;f:447412-447972	4.09	18.42	49	3.45	MAVEDLAEEDVAVVAGTAAVTAVGVMAAADRVEEAWV-VVGTAAVTAVGV	5	No
UM008_0183.1(610);mRNA;f:1194948-1197936	7.13	68.20	76	8.01	MQIFVKTLTGKTITLEVESSDTIDNVKAKIQDKEGIPPDQQRLIFAGKQLEDGRTLADYNIQKESTLHLVLRLRGG	0	Yes
UM086_0041.1(156);mRNA;f:275077-275544	8.48	15.64	49	2.63	LRLLLLHCGCCGCCCCTAAAAAALRLLRLLLL----HCGCCCCTAAAAAAAAA	5	Yes
UM002_0039.1(245);mRNA;r:349540-350828	8.68	24.62	67	2.61	MAGDNRLGILAVVFLPVIGWVLFNIAGPALNQVNNMADKNKSLIAGAGLGAALLSAGQADAAEEMMT	6	Yes
UM048_0068.1(2115);mRNA;f:466349-472693	11.81	219.25	59	34.64	SAPHWGPQVPASHSRVPVHRGASA-QHASPLSPQS-GVGPQVPMSQVKPVSHTLFEQHGSV	0	Yes
UM120_0037.1(292);mRNA;f:179461-181230	12.11	31.39	53	4.32	SVDYAARRSG--G-RE--AQR-S-G-SS---RRPSSS-SSYSAPQRNSS--PAPANRLVSQAVLKRR	0	Yes

**Oscillating disorder profiles can be found below (Figures 4.8, 4.9 and 4.10).*
The list is sorted by ascending pI value. "Period" corresponds to the repeat length and "repeat copy" to the number of repeats predicted in the sequence. The shaded protein resulted to be the protein with similar physicochemical properties to EPYC1 in Chlamydomonas reinhardtii.

Besides the Mw, the disorder profile was an important characteristic to consider discarding UM048_0068.1 as a candidate. The profiles of the two proteins showed disordered regions, but the difference is that the discarded protein showed equally ordered and disordered regions, whereas the selected candidate was highly disordered (Figure 4.10).

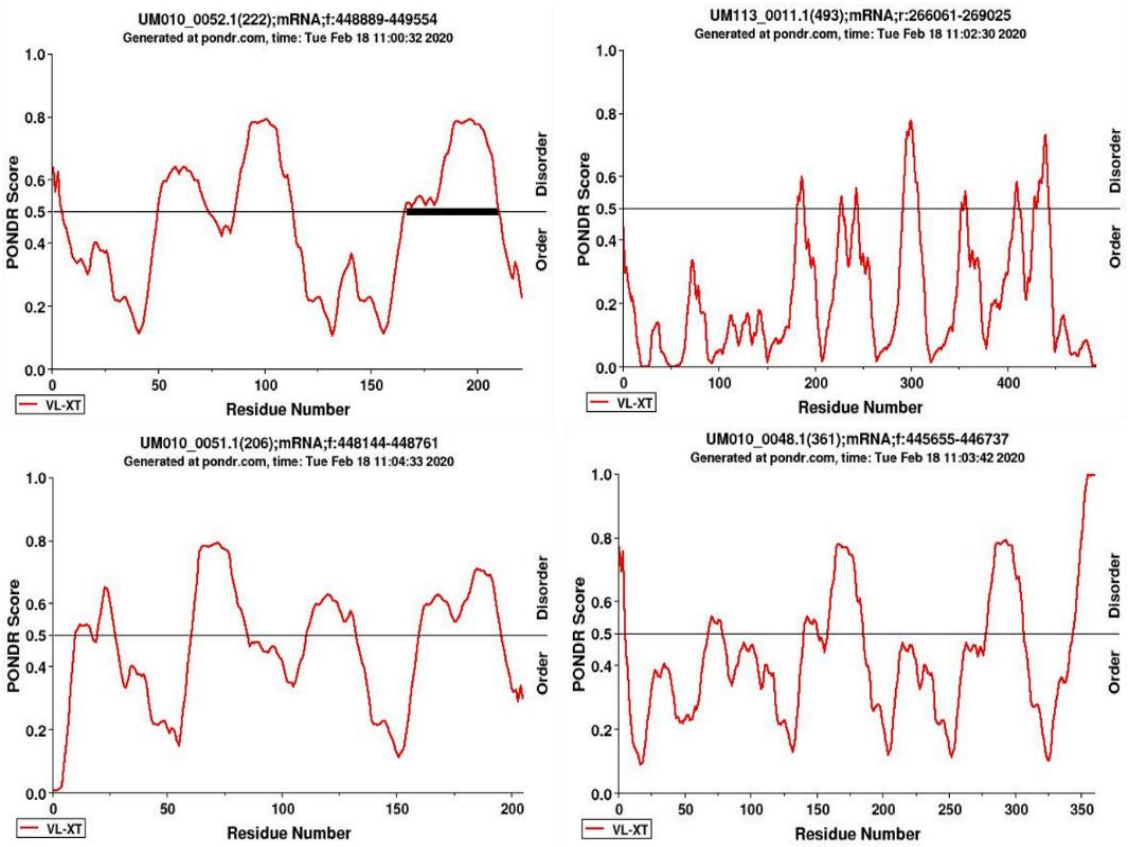

Figure 4.8 *Disorder profile of EPYC1-like protein candidates in Ulva mutabilis.*
X axis represents the residue number and Y axis =0 stands for ordered and =1 disordered. Plots generated at pondr.com with VLXT algorithm.

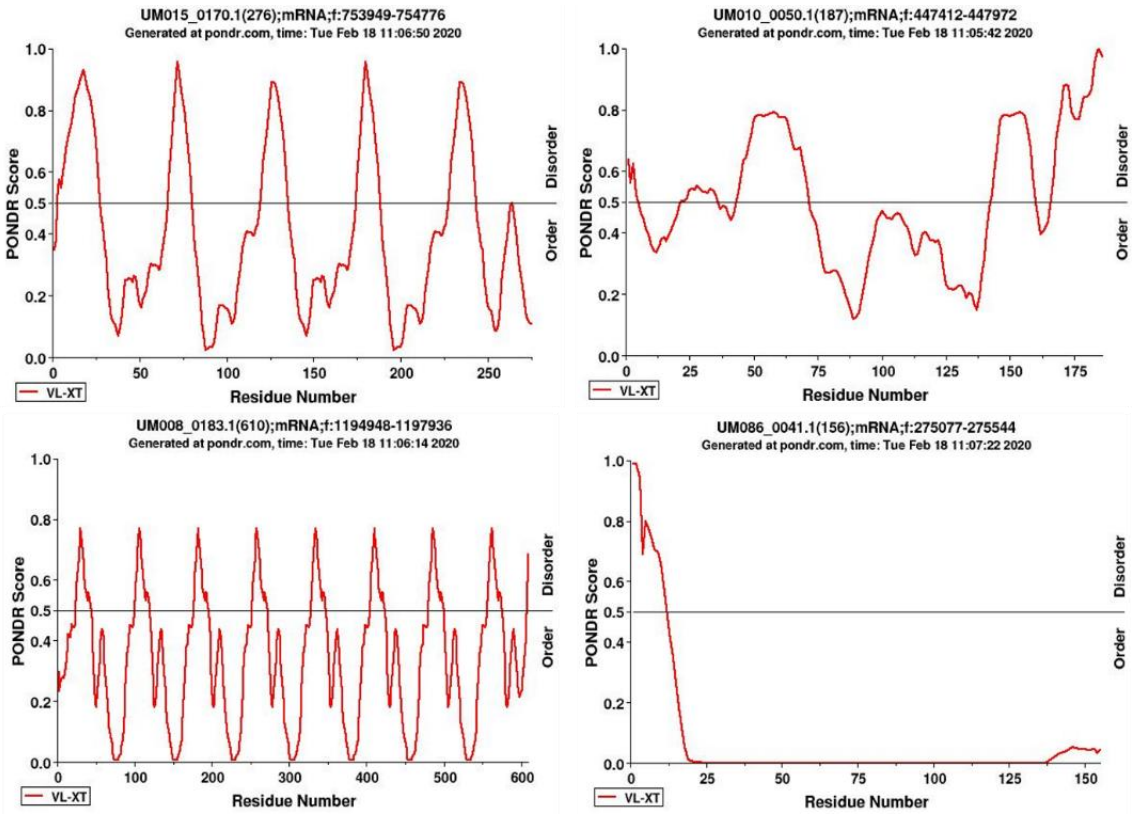

Figure 4.9 *Continuation. Disorder profile of EPYC1-like protein candidates in Ulva mutabilis. X axis represents the residue number and Y axis =0 stands for ordered and =1 disordered. Plots generated at pondr.com with VLXT algorithm.*

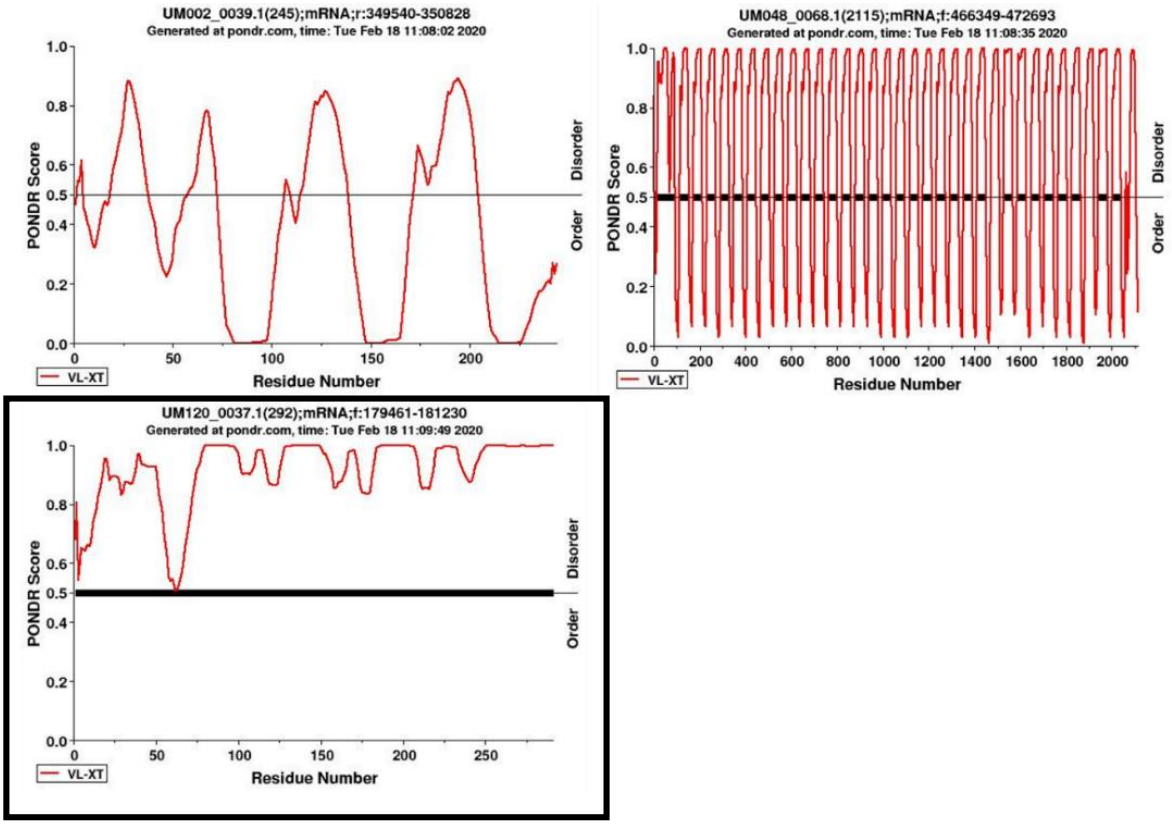

Figure 4.10 *Continuation. Disorder profile of EPYC1-like protein candidates in Ulva mutabilis. X axis represents the residue number and Y axis =0 stands for ordered and =1 disordered. Plots generated at pondr.com. Framed plot corresponds to the protein with more similarity to EPYC1 (shaded in table 4.2).*

4.2.1. Sequence comparisons between the putative protein in *Ulva mutabilis* and EPYC1 show similar patterns

Additionally, further comparison between EPYC1-like proteins in other algae was made complementing the existing study made by Mackinder et al. (2016) with the inclusion of putative proteins in *Ulva mutabilis* retrieved in this study (Table 4.5). From this table can be recognized similar values for the physicochemical properties between EPYC1 and the putative proteins, as well as the fact that same phylum algae can show differences in the presence of EPYC1-like proteins even if a pyrenoid structure exists, such is the case of *Chlorella variabilis*.

Table 4.5 *Comparison between algae with EPYC1-like proteins*

		Number of proteins with:					Protein characteristics					
Species (Phylum)	Pyrenoid	≥3 repeats (40-80 aa length)	and pI ≥8	And an oscillating disorder profile*	And no TMD	Protein ID (OrcAE, Uniprot or phytozo)	Length (amino acids)	pI	Mw (kDa)	Period	Repeat copy	Repeat sequence
Ulva mutabilis (Chlorophyta)	Yes	9	2	2	2	UM048_0068.1(2115);mRNA;f:466349-472693	2115	11.81	219.25	59	34.64	SAPHWGPQVPASHSRVP VHRGASA-QHASPLSPQS-GVGPQVPMSQVKPVSHTL FEQHGSV
						UM120_0037.1(292);mRNA;f:179461-181230	292	12.11	31.39	53	4.32	SVDYAARRSG--G-RE--AQR-S-G-SS---RRPSSS-SSYSAPQRNSS--PAPANRLVSQAVLKRR
Chlamydomonas reinhardtii (Chlorophyta)	Yes	18	8	1	1	Cre10.g436550 (EPYC1)	318	11.8	32.31	61	3.84	VTPSRSALPSNWKQELES LRSSSPAPASSAPAPARS SSASWRDAAPASSAPARS SSASKKA
Thalassiosira pseudonana (Heterokontophyta)	yes	4	1	1	1	B8CF53_THAPS	376	9.1	38.46	53	6.21	LSSKPSSAPFVRSEKPSSA PSDSPSASVAPTLETSFSP SSSGQPSPMTSESPS
Phaeodactylum tricornutum (Heterokontophyta)	Yes	12	1	1	1	B7GDW7_PHATC	380	9.9	37.17	46	7.17	TGPSMTGPSDSDDRRLRS PSSTGPSLTGPSMTGPSA TGPSMTGPSM
Chlorella variabilis (Chlorophyta)	Yes	3	2	1	0							
Chlorella protothecoides (Chlorophyta)	No	1	0	0	0							

**Oscillating disorder profiles can be found below (Figure 4.11).*

"Period" corresponds to the repeat length and "repeat copy" to the number of repeats predicted in the sequence. Protein in bold resulted to be the protein in Ulva mutabilis with similar physicochemical properties to EPYC1 in Chlamydomonas reinhardtii.

Shaded protein corresponds to Ulva's EPYC1-like selected candidate

The analysis of Chlamydomonas reinhardtii and the last 4 algae in the table was retrieved from Mackinder et al. (2016)

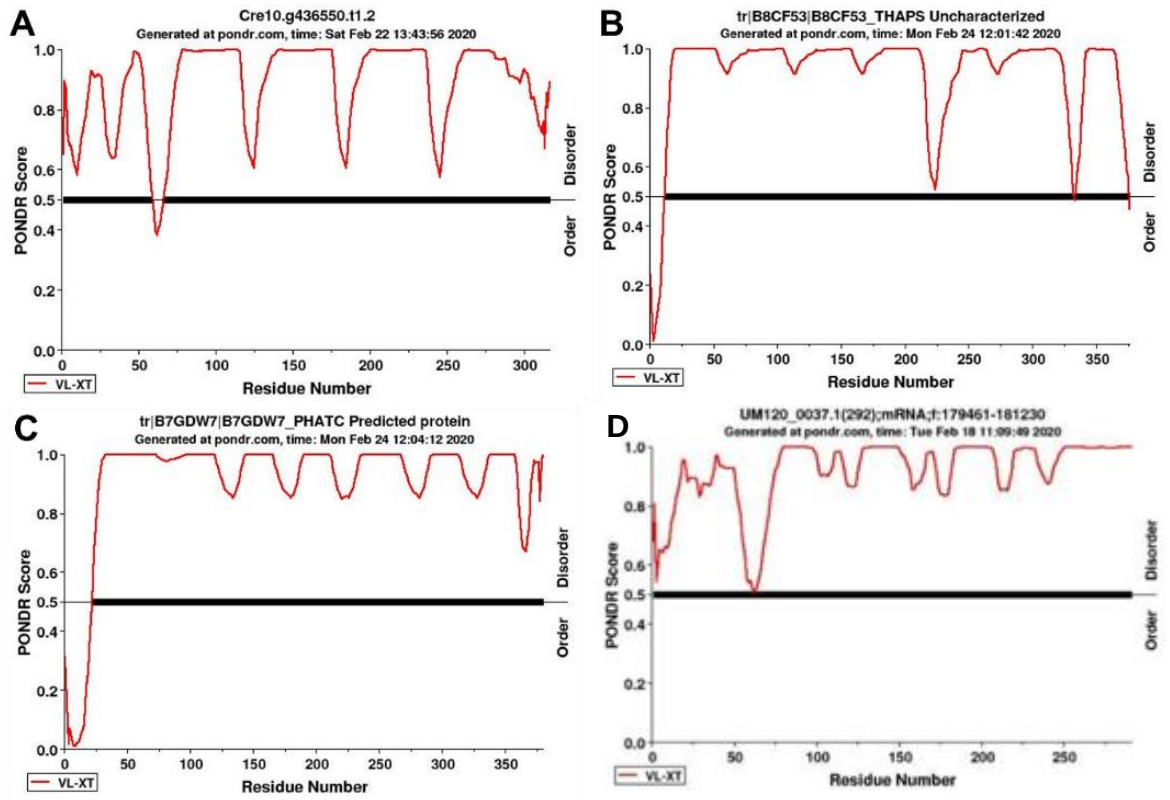

Figure 4.11 *Disorder profiles of EPYC1 and EPYC1-like proteins present in other algae. (A) EPYC1 protein from Chlamydomonas reinhardtii. (B) EPYC1-like protein from Thalassiosira pseudonana. (C) EPYC1-like protein from Phaeodactylum tricornutum. (D) EPYC1-like protein predicted in Ulva mutabilis. Plots generated in pondr.com. Data retrieved from Mackinder et al. (2016)*

As part of these comparisons, a general amino acid sequence similarity analysis between EPYC1 and the putative protein in *Ulva mutabilis* (UM120_0037.1) was carried out, including the putative protein in *Thalassiosira pseudonana* as a reference. For both comparisons, despite the similarity in their physicochemical properties, the amino acid sequences only show a similarity of 20% and 30% in both cases (Figure 4.12).

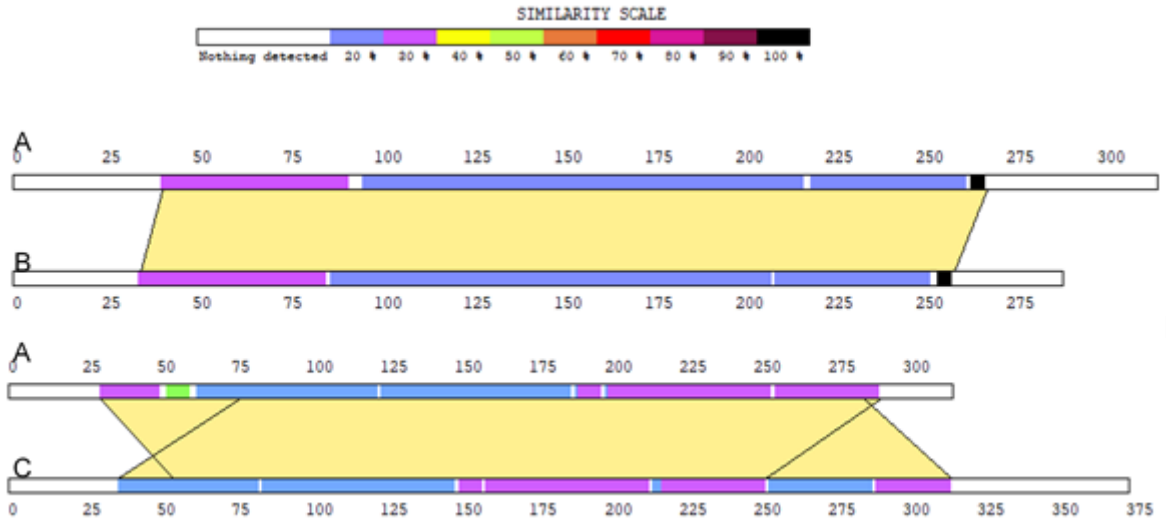

Figure 4.12 *Sequence similarity between EPYC1 (A), UM120_0037.1 (B) and (C) Putative protein in Thalassiosira psudonana, proposed by Mackinder et al. (2016).*
The similarity between the sequences are between 20% and 30%, blue and purple coloured respectively. Figure generated in SIM - Local similarity program (https://web.expasy.org/sim) (Huang & Miller, 1991).

Additionally, a comparison of TR in the sequences was made with the predictions retrieved from the XSTREAM predictor (Newman & Cooper, 2007). However, the analysis of TR in the sequences of putative proteins in *Ulva mutabilis*, *Thalassiosira pseudonana*, and *Chlamydomonas reinhardtii* EPYC1 (Figure 4.13) shows a similar pattern in the repeat lengths, TR quantity and regions where TR occur, so similar structures and function cannot be discarded.

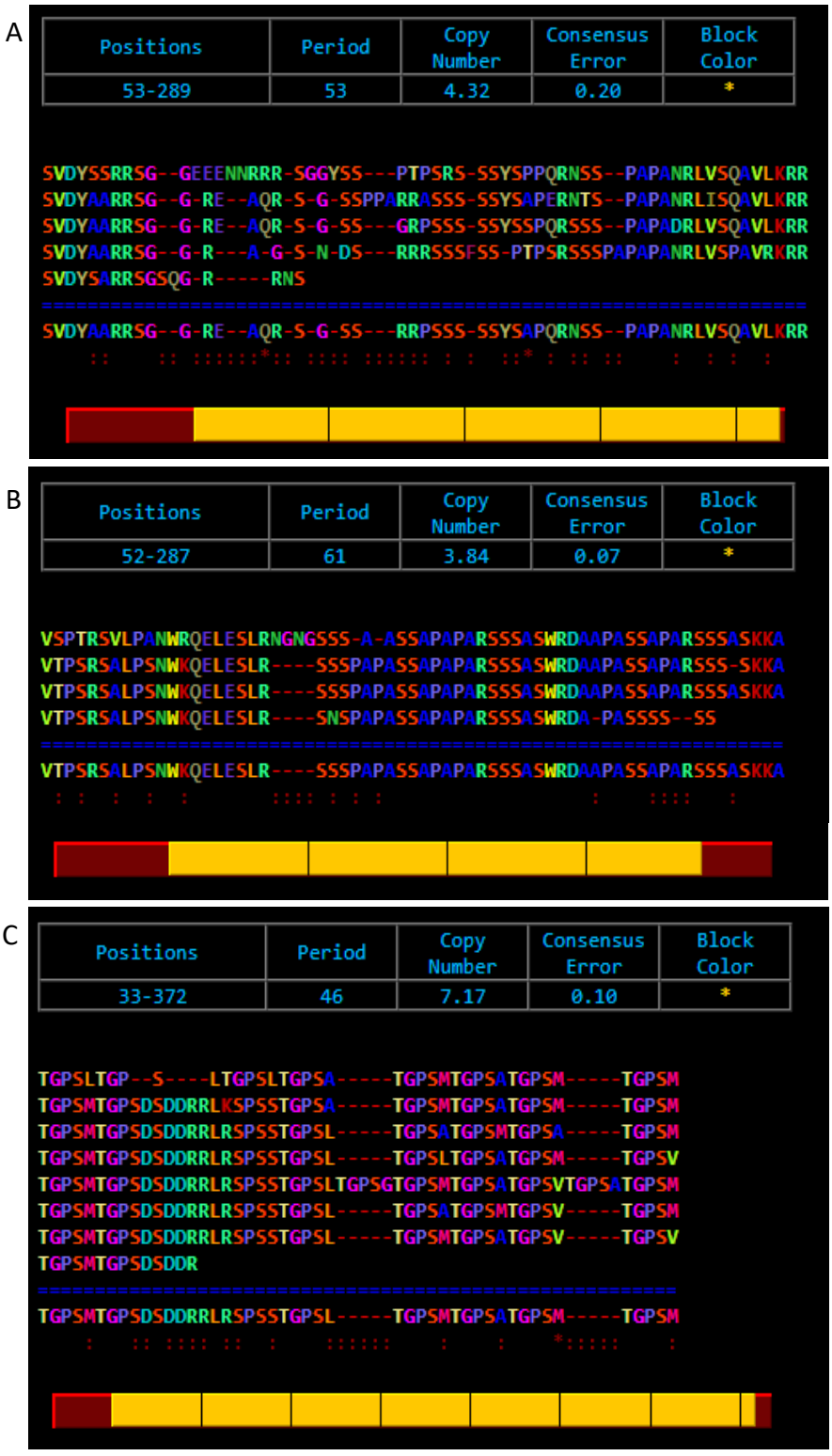

Figure 4.13 *Comparison of tandem repeats in EPYC1 and putative proteins in Ulva mutabilis and Thalassiosira pseudonana amino acid sequences.*
Tables indicate the position of TR, repeat length (period), number of TR in the sequence (copy number), each line of the sequence corresponds to one TR, and the block colour corresponds on the region where the TRs are located in the sequence represented in yellow bars. (A) TR in Ulva mutabilis putative protein (UM120_0037.1), (B) TR in Chlamydomonas reinhardtii EPYC1 and (C) TR in Thalassiosira pseudonana putative protein. Figure generated in XSTREAM (Newman & Cooper, 2007).

Finally, even though a phylogenetic approach was more desirable and appropriate to support the findings of homology in this research, it was not possible due to the very nature of these proteins: the sequences of disordered proteins are known to evolve faster than those of structured proteins, but their physicochemical properties are under selective pressure and are evolutionarily conserved (Mackinder et al., 2016).

4.3. The *Ulva* chloroplast proteome is sensitive to CO_2 conditions.

To begin to understand how the *Ulva* proteome responds to different CO_2 conditions, and thus identify other targets for engineering yields, this experiment aimed to identify CO_2 responsive proteins including a putative EPYC1-like protein in *Ulva*.

The first step of this experiment was to look if the protein extraction from the tissue was achieved, therefore quantification and separation of the proteins were carried out.

Even though this was a preliminary step for the iTRAQ technique, the protein extraction showed qualitative differences in the intensity of the bands obtained between the high CO_2 condition (HC) and the controls (LC) (Figure 4.14). Among the main observations of this preliminary analysis, a change in the grey value of the bands placed at ~26 kDa between the treatments, as well as the bands placed between ~21 kDa and ~13 kDa, indicates an increase of these proteins in the LC treatment implying a sensitive response to CO_2 changes. However, further proteomic analyses are needed in order to get more information about these changes.

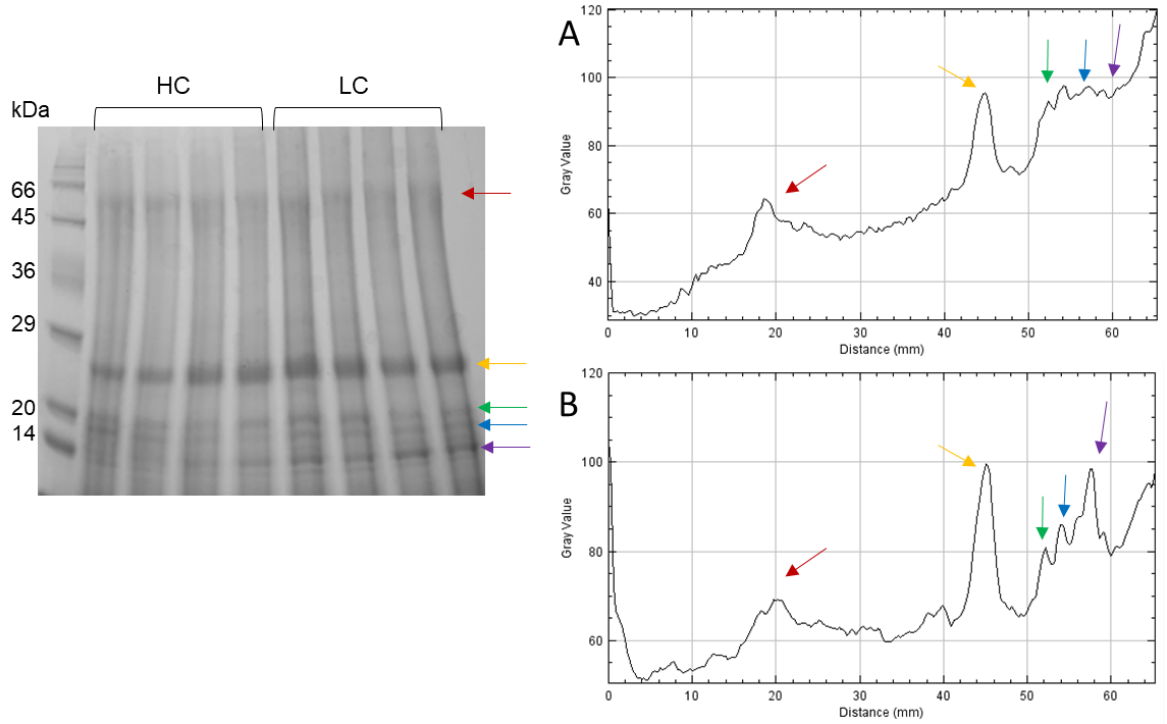

Figure 4.14 *SDS-PAGE gel of protein extraction from algal tissue exposed to different CO_2 conditions.*
Coloured arrows indicate different bands in the protein separation. (A) Is a representative plot of HC condition. (B) Is a representative plot of LC condition, higher relative peak heights in LC sample were obtained compared to HC sample peaks. Each peak in the profile plot corresponds to one band in the gel, where the x axis indicates the band mobility through the gel in distance, and y axis to the grey value of each band. Plots generated with Fiji (Schindelin et al., 2012).

Finally, the same samples were subsequently prepared for an iTRAQ assay as described in the materials and methods section, unfortunately, LC-MS final results were delayed because of COVID-19 pandemic.

5. Discussion

My project aimed to identify candidate proteins involved in the carbon concentration mechanism and CO_2 metabolism in *Ulva*, as well as their response to CO_2 changes. I made three important advances in this field:

First, I developed a protocol for the extraction of chloroplast-enriched material from *Ulva*. This is important because no *Ulva* EPYC1-like proteins are known to exist, therefore, it is necessary to look at what is inside the chloroplast and study it, unlike, for example, the *epyc1* mutants that already exist in the related, but much more widely studied microalgae, *Chlamydomonas reinhardtii*. My chloroplast-enriched fraction protocol now allows targeted study of the biochemistry of *Ulva* CCM.

Second, using *Chlamydomonas reinhardtii* as a reference, I used a range of bioinformatics techniques to identify potential CCM candidate proteins in the *Ulva* genome. This revealed a previously unreported existing orthologue to EPYC1 based exclusively on the similarity to their physicochemical properties.

Third, and finally, I demonstrated changes in the *Ulva* proteome in response to changing CO_2 conditions in the media, and carried out a full iTRAQ labelling and LC-MS sample preparation to correlate protein expression with my bioinformatics predictions. Unfortunately, the global 2020 coronavirus pandemic meant that my departmental analytical facility was shut down for the pandemic before they were able to run the LC-MS results of my final experiment; this will be discussed below, along with each of my findings in more depth.

5.1. Developing a methodology to obtain subcellular material

In previous studies carried out in *Chlamydomonas reinhardtii*, a cell wall-less mutant was used in order to isolate stable subcellular material such as pyrenoids and chloroplasts (Kuchitsu et al., 1988; Zhan et al., 2018). The lack of similar *Ulva* mutants necessitated biochemical protoplast isolation, starting from the removal of the cell wall via enzymatic digestion as a first step to extract the chloroplast and its associated pyrenoid. It is important to note that intact protoplast isolation is needed before chloroplast isolation in order to minimise endogenous enzymatic activity following cell disruption.

First, a modification of the improved protocol used by Reddy et al., (2006) to isolate intact protoplasts was carried out. Originally, this protocol consisted of reducing components and varying concentrations of such components in the enzyme preparations used in previous studies (Reddy et al., 1989; Björk et al., 1992; Uchida et al., 1992). However, I found that it was necessary to maintain an ingredient used in Reddy's "not-improved" mixture: $CaCl_2$. Light and confocal microscopy (Figure 4.2) confirmed that the yield of intact protoplasts increased with the addition of $CaCl_2$, which would be consistent with its activity as a membrane stabilizer evaluated in previous studies carried out in a variety of cells (Boss & Mott, 1980; Davis, 1985; Saure, 2005).

Having developed a protocol that allowed a first isolation of an *Ulva* chloroplast-enriched fraction, I next verified that proteins could be obtained and separated from this fraction (Figure 4.5). A number of major bands in the chloroplast-enriched fraction had molecular masses that were consistent with expected chloroplast proteins in the *Ulva* genome. For example, two major bands in the chloroplast-enriched fraction were found around 55 kDa and 20 kDa. I proposed that these correspond to large and small Rubisco subunits, because their predicted masses are ~52 kDa and ~20 kDa (rbcL, UniProt ID: A0A0E3XIA7_9CHLO; rbcS, UniProt ID: D7EYZ2_9CHL) respectively. Another thing that can be highlighted is the presence of two bands placed at 24 kDa and 30 kDa (Figure 4.5), which have a similar Mw as EPYC1 (~24 kDa) in *Chlamydomonas* (Mackinder et al., 2016; Atkinson et al., 2017).

The next step to follow in order to know if the enrichment was achieved, was a comparison between the proteins obtained in the chloroplast fraction and proteins in the intact *Ulva* tissue. According to the SDS-PAGE analysis, which was carried out after a protein extraction from the intact thallus tissue, a band with a similar Mw (between 29 kDa and 24 kDa) as the bands at 24 kDa and 30 kDa obtained in the previous analysis (Figure 4.5) was obtained from the chloroplast fraction, showing an enrichment indicated with a yellow arrow in the plot (Figure 4.6), compared with the intact tissue. However, due to their abundance, stronger bands were expected at 55 kDa and 20 kDa corresponding to both Rubisco subunits.

These gel runs imply the presence of an EPYC1-sized protein in *Ulva*, assuming that any putative EPYC1-like protein in *Ulva* would have the same, or similar, Mw

as EPYC1 in *Chlamydomonas*. In addition, further experiments such as a gel run with both *Chlamydomonas* EPYC1 and the putative protein in *Ulva* in the same gel, could give further information about their size similarity.

5.2. A promising candidate for *Ulva* EPYC1 has been found

Relative abundance of proteins in *Ulva* chloroplast was predicted in DeepLoc-1.0, resulting in 1,129 plastid localised proteins, later divided in three groups depending on their Mw (Figure 4.7). Those that Mw fitted with EPYC1 Mw (between 20 kDa and 40 kDa) were reciprocally BLASTed against *Chlamydomonas reinhardtii* as well as against EPYC1 amino acid sequence.

BLASTing proteins known to be involved in *Chlamydomonas reinhardtii* CO_2 metabolism against the *Ulva* proteome gave a number of putative *Ulva* proteins whose identities were supported by reciprocal BLAST (Table 4.2). However, there was one obvious exception: extensive BLAST searches (BLAST, BLASTp) of the *Chlamydomonas* EPYC1 sequence against the *Ulva* proteome failed to reveal any *Ulva* EPYC1 homologs.

However, this failure could be because EPYC1, as well as the target protein in *Ulva*, has a low complexity structure, which would mitigate against hits being found by BLAST, as matching hits with low complexity sequences do not guarantee relationship, even with statistically significant scores (Mier et al., 2019). Accordingly, I took a second approach to look for EPYC1-like proteins in the *Ulva* proteome. This resulted in the shortlisting of 9 proteins with 40-80 amino acid length tandem repeats in their amino acid sequences. The presence of tandem repeats is important because of their role in protein modular architecture, being found in important structural proteins such as animal collagens and keratins, insect and spider silks, plant cell wall extensins, and the proteins that form adhesive plaques and byssal threads of bivalve mussels (Newman & Cooper, 2007). The tandem repeats in EPYC1 may help to explain its function as a linker that connects rubisco subunits to confer structure to the pyrenoid.

To narrow down the shortlist, I removed any proteins that contained predicted transmembrane domains, which are related to transport and signalling roles in the cell (Möller et al., 2001; Hildebrand et al., 2004). Since EPYC1 has a linker function

that does not imply any membrane interaction, the total absence of transmembrane domains was required, reducing my shortlist to 5 candidates.

Next, I discarded any candidates that did not have disordered domains, as these should be indicative of more linker proteins because it is known that the structure and function of a protein are highly linked, but not all functional proteins encoded by a diversity of genomes fold into stable globular structures (Wright & Dyson, 1999). However, for this specific linker, the protein, or proteins, which were searched had to also have a high pI (≥8) and only two proteins met these criteria.

Table 4.4 shows a comparison between *Ulva* predicted proteins retrieved in this study. After checking the disorder profile plots (Figure 4.8, Figure 4.9 and Figure 4.10) and molecular weights, UM120_0037.1 was chosen as the most promising candidate over UM048_0068.1 which has a Mw of 219.25 kDa, a considerably higher mass, also, its disorder profile was dispersed between the disorder and ordered regions equally, deferring of the disorder profile of UM120_0037.1, which behaviour shows a mainly disordered sequence (Figure 4.10).

Additionally, in the data presented in Table 4.5 “Comparison between algae with EPYC1-like proteins”, it can be seen that there is no sequence similarity between amino acid sequences of repeats in putative proteins, which suggests that these proteins are not homologous but analogous. A similarity analysis (Huang & Miller, 1991) between EPYC1 and UM120_0037.1, and EPYC1 and the putative protein in *Thalassiosira pseudonana* (Mackinder et al., 2016) is shown in Figure 4.12, where the sequence similarity between these proteins is low, even though their predicted physicochemical properties are similar. However, even with a low sequence similarity, between 20% and 30%, a similar function cannot be dismissed.

Finally, it should be noted that the putative *Ulva* EPYC1-like protein did not possess a predicted chloroplast targeting sequence when assessed using the Deeploc program. However, this is not of real concern, because localisation prediction is often unreliable and EPYC1 itself returns a predicted cytosolic localisation even though it has been shown to occur in the pyrenoid. Finally, an analysis in ChloroP (Emanuelsson et al., 1999) did suggest the presence of chloroplastic transit peptides in EPYC1, as well as in UM120_0037.1 (putative protein in *Ulva mutabilis*). Even when this gives a clearer evidence of the presence of an EPYC1-like protein

in *Ulva*, its actual function as a linker and its activation under limited CO_2 concentration remains unknown.

5.3. Manipulating CO_2 conditions caused changes in the *Ulva* proteome

The preliminary protein analysis by SDS-PAGE of the differences between the proteome of whole thalli exposed to varying CO_2 levels showed visible changes in the proteins present in the crude protein extract between HC and LC samples. Qualitative changes can be appreciated looking at the differences in relative peak height in the plot (Figure 4.14), where the peak corresponding to the band at ~26 kDa was higher in LC than in HC, suggesting a slightly higher amount of proteins expressing in low carbon conditions; on the other hand, proteins with lower mass (~20 kDa and ~17 kDa) were more abundant in high carbon concentration.

To follow this up and to more accurately identify the proteins that changed, I repeated the experiment and prepared samples for iTRAQ analysis. However, the global 2020 coronavirus pandemic has meant that my samples have not yet been run and are unlikely to be run until late 2020.

Comparative proteome analyses have been carried out in *Ulva* and other seaweeds to look at the responses to temperature stress (Fan et al., 2018). Even though proteome analysis has not been carried out to look at *Ulva*'s response to CO_2 levels, these previous studies have shown significant changes in proteomes, so I expect the LC-MS analysis will show a range of proteins that are up-regulated, down-regulated or unchanged in response to CO_2 changes. To speculate, within those differentially expressed proteins (up-regulated and down-regulated) I expect to see mostly chloroplast-related proteins since this is where carbon acquisition reactions take place; I also expect to obtain the most abundant CCM-related proteins such as EPYC1–like proteins, Rubisco subunits and Rubisco activase in the low carbon condition because is under low carbon concentrations that the CCM is more active (Miura et al., 2004; Moritz et al., 2012; Zhou et al., 2020). Moreover, internal carbonic anhydrases should be included in the proteins up-regulated in high carbon condition since these are responsible for maintaining cellular acid-base balance and they catalyse a simple diffusion of CO_2 into the chloroplast (Badger & Price, 1994).

It is, however, possible that the amount and expression of my putative EPYC1-like protein might not vary under different carbon concentrations.

5.4. Future work has been planned

My work has identified a putative pyrenoid-structuring protein in *Ulva*, based on physicochemical properties and proteomic analyses. I have also shown that proteins that correspond to the predicted size of this *Ulva* EPYC1 analog are preferentially found in the chloroplast fraction and are responsive to CO_2 levels. To resolve whether my putative EPYC1 protein was actually changing in response to CO_2 levels, I carried out an iTRAQ and LC-MS sample preparation. Unfortunately, I have not been able to run these samples.

However, when it became apparent that my research was going to be cut short by the COVID-19 pandemic, future work was planned in order to resolve the action of my *Ulva* EPYC1-like gene. The plans are shown below (Figure 5.1).

First, to confirm that my EPYC1-like protein is found in the pyrenoid, the first step would be tagging this gene with GFP and express the recombinant gene in *Ulva*, a recently developed technique by Jonas Blomme's research group in Gent. During the correction of this thesis, I have, indeed, learned that my predicted protein is localised to the pyrenoid.

Second, to confirm that my EPYC1-like protein is able to confer structure on the pyrenoid, complementing a wider studied *epyc1*-deficient *Chlamydomonas reinhardtii* mutant with the predicted protein sequence would be the next step towards the confirmation of my findings in this research. This work is underway in York.

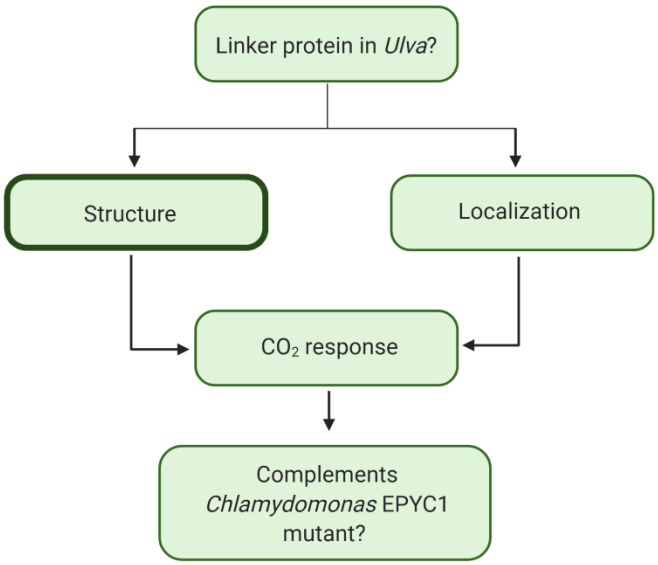

Figure 5.1 *Further work.*
Further work will consist in demonstrate the localization of the predicted protein in this work, its response to CO_2 variable conditions and, as a confirmation, if this protein complements Chlamydomonas EPYC1 mutant.

6. Concluding remarks

To conclude, while the need for fossil fuel alternatives is growing, our knowledge and understanding of the metabolic machinery of common biomass sources remains lacking. Accordingly, their study will contribute to solutions to the current environmental and energetic crisis that society is entering. My work has described the importance of seaweed as a potential biofuel feedstock, and has begun to identify candidate proteins and pathways that will allow the study and improvement of their CO_2 metabolism through consideration of the structure of their characteristic subcellular organelle: the pyrenoid.

The pyrenoid has been thought of as a simple subcellular component; instead, evidence of its structure and protein conformation in other algae shows how complex and essential it is in CO_2 fixation. Accordingly, my search for, and identification of, a candidate pyrenoid-structuring protein in *Ulva* suggest the existence of a good target candidate, as the linker protein responsible for pyrenoid conformation and function.

At the same time, my work has demonstrated the relative lack of methodology available for the study of macroalgal metabolism. To amend this, I have developed new extraction and analytical protocols so that we can match bioinformatic predictions to biochemical measurements. Finally, I have helped to build a network of national and international collaborators and have helped to design the further experiments that will prove or disprove my hypotheses.

7. References

Abomohra, A. E.-F., El-Naggar, A. H., & Baeshen, A. A. (2018). Potential of macroalgae for biodiesel production: Screening and evaluation studies. *Journal of Bioscience and Bioengineering, 125*(2), 231-237. doi:10.1016/j.jbiosc.2017.08.020

Alalwan, H. A., Alminshid, A. H., & Aljaafari, H. A. S. (2019). Promising evolution of biofuel generations. Subject review. *Renewable Energy Focus, 28*, 127-139. doi:10.1016/j.ref.2018.12.006

Alaswad, A., Dassisti, M., Prescott, T., & Olabi, A. G. (2015). Technologies and developments of third generation biofuel production. *Renewable and Sustainable Energy Reviews, 51*, 1446-1460. doi:https://doi.org/10.1016/j.rser.2015.07.058

Almagro Armenteros, J. J., Sønderby, C. K., Sønderby, S. K., Nielsen, H., & Winther, O. (2017). DeepLoc: prediction of protein subcellular localization using deep learning. *Bioinformatics, 33*(21), 3387-3395. doi:10.1093/bioinformatics/btx431

Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W., & Lipman, D. J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res, 25*(17), 3389-3402. doi:10.1093/nar/25.17.3389

Antizar-Ladislao, B., & Turrion-Gomez, J. L. (2008). Second-generation biofuels and local bioenergy systems. *Biofuels, Bioprod. Bioref, 2*(5), 455-469. doi:0.1002/bbb.97

Aresta, M., Dibenedetto, A., & Barberio, G. (2005). Utilization of macro- algae for enhanced CO 2 fixation and biofuels production: Development of a computing software for an LCA study. *Fuel Processing Technology, 86*(14), 1679-1693. doi:10.1016/j.fuproc.2005.01.016

Aro, E.-M. (2016). From first generation biofuels to advanced solar biofuels. *A Journal of the Human Environment, 45*(Supplement 1), 24-31. doi:10.1007/s13280-015-0730-0

Arvelakis, S., & Koukios, E. G. (2013). Critical factors for high temperature processing of biomass from agriculture and energy crops to biofuels and bioenergy. *Wiley Interdisciplinary Reviews: Energy and Environment, 2*(4), 441-455. doi:10.1002/wene.28

Ashokkumar, V., Salim, M. R., Salam, Z., Sivakumar, P., Chong, C. T., Elumalai, S., . . . Ani, F. N. (2017). Production of liquid biofuels (biodiesel and bioethanol) from brown marine macroalgae Padina tetrastromatica. *Energy Conversion and Management, 135*, 351-361. doi:10.1016/j.enconman.2016.12.054

Aslani, A., Mazzuca-Sobczuk, T., Eivazi, S., & Bekhrad, K. (2018). Analysis of bioenergy technologies development based on life cycle and adaptation trends. *Renewable Energy, 127*, 1076-1086. doi:10.1016/j.renene.2018.05.035

Atkinson, N., Leitão, N., Orr, D. J., Meyer, M. T., Carmo-Silva, E., Griffiths, H., . . . McCormick, A. J. (2017). Rubisco small subunits from the unicellular green

alga Chlamydomonas complement Rubisco- deficient mutants of Arabidopsis. *New Phytologist, 214*(2), 655-667. doi:10.1111/nph.14414

Atkinson, N., Velanis, C. N., Wunder, T., Clarke, D. J., Mueller-Cajar, O., & McCormick, A. J. (2019). The pyrenoidal linker protein EPYC1 phase separates with hybrid Arabidopsis- Chlamydomonas Rubisco through interactions with the algal Rubisco small subunit. *Journal of experimental botany, 70*(19), 5271. doi:10.1093/jxb/erz275

Badger, M. R., Ludwig, M., Yellowlees, D., Leggat, W., & Price, G. D. (1998). The diversity and coevolution of Rubisco, plastids, pyrenoids, and chloroplast-based CO2-concentrating mechanisms in algae. Can J Bot 76: 1052-1071. *Can. J. Bot, 76*, 1052-1071.

Badger, M. R., & Price, G. D. (1994). The role of carbonic anhydrase in photosynthesis. *Annual review of plant biology, 45*(1), 369-392.

Balina, K., Romagnoli, F., Pastare, L., & Blumberga, D. (2017). Use of Macroalgae for Bioenergy Production in Latvia: Review on Potential Availability of Marine Coastline Species. *Energy Procedia, 113*, 403-410. doi:10.1016/j.egypro.2017.04.022

Belghit, I., Rasinger, J. D., Heesch, S., Biancarosa, I., Liland, N., Torstensen, B., . . . Bruckner, C. G. (2017). In- depth metabolic profiling of marine macroalgae confirms strong biochemical differences between brown, red and green algae. *Algal Research, 26*, 240-249. doi:10.1016/j.algal.2017.08.001

Björk, M., Gómez-Pinchetti, J. L., García-Reina, G., & Pedersén, M. (1992). Protoplast isolation from *Ulva rigida* (Chlorophyta). *British Phycological Journal, 27*(4), 401-407. doi:10.1080/00071619200650331

Boller, A., Thomas, P., Cavanaugh, C., & Scott, K. (2015). Isotopic discrimination and kinetic parameters of R ubis CO from the marine bloom-forming diatom, S keletonema costatum. *Geobiology, 13*(1), 33-43.

Boss, W. F., & Mott, R. L. (1980). Effects of divalent cations and polyethylene glycol on the membrane fluidity of protoplast. *Plant Physiology, 66*(5), 835-837.

Buschmann, A. H., Camus, C., Infante, J., Neori, A., Israel, Á., Hernández-González, M. C., . . . Critchley, A. T. (2017). Seaweed production: overview of the global state of exploitation, farming and emerging research activity. *European Journal of Phycology, 52*(4), 391-406. doi:10.1080/09670262.2017.1365175

Castelar, B., Reis, R. P., & dos Santos Calheiros, A. C. (2014). *Ulva lactuca* and *U. flexuosa* (Chlorophyta, Ulvophyceae) cultivation in Brazilian tropical waters: recruitment, growth, and *Ulva*n yield. *Journal of Applied Phycology, 26*(5), 1989-1999. doi:10.1007/s10811-014-0329-z

Chen, H., Zhou, D., Luo, G., Zhang, S., & Chen, J. (2015). Macroalgae for biofuels production: Progress and perspectives. *Renewable and Sustainable Energy Reviews, 47*, 427-437. doi:10.1016/j.rser.2015.03.086

Cleland, W. W., Andrews, T. J., Gutteridge, S., Hartman, F. C., & Lorimer, G. H. (1998). Mechanism of Rubisco: The Carbamate as General Base. *Chemical reviews, 98*(2), 549. Retrieved from https://pubs.acs.org/doi/pdf/10.1021/cr970010r

Davis, B. (1985). Factors influencing protoplast isolation. In *Fungal protoplasts* (pp. 45-71): Marcel Dekker New York.

De Clerck, O., Kao, S.-M., Bogaert, K. A., Blomme, J., Foflonker, F., Kwantes, M., . . . Lattermann, L. (2018). Insights into the Evolution of Multicellularity from the Sea Lettuce Genome. *Current Biology, 28*(18), 2921-2933.e2925. doi:10.1016/j.cub.2018.08.015

Demirbas, M. F. (2011). Biofuels from algae for sustainable development. *Applied Energy, 88*(10), 3473-3480. doi:https://doi.org/10.1016/j.apenergy.2011.01.059

Dow, K., & Downing, T. (2011). Atlas of climate change. Third. Retrieved from Credo Reference: http://ezphost.dur.ac.uk/login?url=https://search.credoreference.com/content/entry/ucpresscc/fossil_fuels/0?institutionId=1856

Emanuelsson, O., Nielsen, H., & von Heijne, G. (1999). ChloroP, a neural network-based method for predicting chloroplast transit peptides and their cleavage sites. *Protein Sci, 8*(5), 978-984. doi:10.1110/ps.8.5.978

Fan, M., Sun, X., Liao, Z., Wang, J., Li, Y., & Xu, N. (2018). Comparative proteomic analysis of *Ulva* prolifera response to high temperature stress. *Proteome Science, 16*(1), 17. doi:10.1186/s12953-018-0145-5

Felix, O., Moses, M., Anders, T., & Francis, K. (2018). Seaweed Bioethanol Production: A Process Selection Review on Hydrolysis and Fermentation. *Fermentation, 4*(4), 99. doi:10.3390/fermentation4040099

Freeman Rosenzweig, E. S., Xu, B., Kuhn Cuellar, L., Martinez-Sanchez, A., Schaffer, M., Strauss, M., . . . Jonikas, M. C. (2017). The Eukaryotic CO2-Concentrating Organelle Is Liquid-like and Exhibits Dynamic Reorganization. *Cell, 171*(1), 148-162.e119. doi:https://doi.org/10.1016/j.cell.2017.08.008

Giordano, M., Beardall, J., & Raven, J. (2005). CO2 concentrating mechanisms in Algae: Mechanisms, environmental modulation, and evolution. *Annual review of plant biology, 56*, 99-131. doi:10.1146/annurev.arplant.56.032604.144052

Golberg, A., & Liberzon, A. (2015). Modeling of smart mixing regimes to improve marine biorefinery productivity and energy efficiency. *Algal Research, 11*, 28-32. doi:https://doi.org/10.1016/j.algal.2015.05.021

Goldman, J. C., & Dennett, M. R. (1983). Carbon Dioxide Exchange Between Air and Seawater: No Evidence for Rate Catalysis. *Science, 220*(4593), 199. doi:10.1126/science.220.4593.199

Gomez, I., & Huovinen, P. (2012). Morpho-functionality of Carbon Metabolism in Seaweeds. In (pp. 25-46).

Gui, M. M., Lee, K. T., & Bhatia, S. (2008). Feasibility of edible oil vs. non-edible oil vs. waste edible oil as biodiesel feedstock. *Energy, 33*(11), 1646-1653. doi:https://doi.org/10.1016/j.energy.2008.06.002

Hildebrand, P. W., Preissner, R., & Frömmel, C. (2004). Structural features of transmembrane helices. *FEBS Letters, 559*(1), 145-151. doi:https://doi.org/10.1016/S0014-5793(04)00061-4

Huang, X., & Miller, W. (1991). A Time-Efficient, Linear-Space Local Similarity Algorithm. *Advances in Applied Mathematics, 12*, 337-357. doi:https://doi.org/10.1016/0196-8858(91)90017-D

Hurd, C. L., Harrison, P. J., Bischof, K., & Lobban, C. S. (2014). *Seaweed Ecology and Physiology*. West Nyack: West Nyack: Cambridge University Press.

Keown, J. R., Griffin, M. D. W., Mertens, H. D. T., & Pearce, F. G. (2013). Small oligomers of ribulose- bisphosphate carboxylase/ oxygenase (Rubisco) activase are required for biological activity. *The Journal of biological chemistry, 288*(28), 20607. doi:10.1074/jbc.M113.466383

Krogh, A., Larsson, B., Von Heijne, G., & Sonnhammer, E. L. L. (2001). Predicting transmembrane protein topology with a hidden markov model: application to complete genomes. *Journal of Molecular Biology, 305*(3), 567-580. doi:10.1006/jmbi.2000.4315

Kuchitsu, K., Tsuzuki, M., & Miyachi, S. (1988). Characterization of the pyrenoid isolated from unicellular green alga Chlamydomonas reinhardtii : Particulate form of RuBisCO protein. *An International Journal of Cell Biology, 144*(1), 17-24. doi:10.1007/BF01320276

Kumar, K., Dasgupta, C. N., Nayak, B., Lindblad, P., & Das, D. (2011). Development of suitable photobioreactors for CO2 sequestration addressing global warming using green algae and cyanobacteria. *Bioresource Technology, 102*(8), 4945-4953. doi:10.1016/j.biortech.2011.01.054

Kumar, K., Ghosh, S., Angelidaki, I., Holdt, S. L., Karakashev, D. B., Morales, M. A., & Das, D. (2016). Recent developments on biofuels production from microalgae and macroalgae. *Renewable and Sustainable Energy Reviews, 65*, 235-249. doi:10.1016/j.rser.2016.06.055

Linthwaite, V. L., Janus, J. M., Brown, A. P., Wong-Pascua, D., O'Donoghue, A. M. C., Porter, A., . . . Cann, M. J. (2018). The identification of carbon dioxide mediated protein post-translational modifications. *Nature Communications, 9*(1). doi:10.1038/s41467-018-05475-z

Lorimer, G. H., & Miziorko, H. M. (1980). Carbamate formation on the epsilon- amino group of a lysyl residue as the basis for the activation of ribulosebisphosphate carboxylase by CO2 and Mg2+. *Biochemistry, 19*(23), 5321. doi:10.1021/bi00564a027

Lü, J., Sheahan, C., & Fu, P. (2011). Metabolic engineering of algae for fourth generation biofuels production. *Energy Environ. Sci., 4*(7), 2451-2466. doi:10.1039/c0ee00593b

Mackinder, L. C. M. (2018). The Chlamydomonas CO2-concentrating mechanism and its potential for engineering photosynthesis in plants. *New Phytologist, 217*(1), 54-61. doi:10.1111/nph.14749

Mackinder, L. C. M., Meyer, M. T., Mettler-Altmann, T., Chen, V. K., Mitchell, M. C., Caspari, O., . . . Jonikas, M. C. (2016). A repeat protein links Rubisco to form the eukaryotic carbon-concentrating organelle. doi:10.17863/CAM.23023

Merriam-Webster, D. "Pyrenoid". Retrieved from https://www.merriam-webster.com/dictionary/pyrenoid

Mier, P., Paladin, L., Tamana, S., Petrosian, S., Hajdu-Soltész, B., Urbanek, A., . . . Andrade-Navarro, M. A. (2019). Disentangling the complexity of low complexity proteins. *Briefings in Bioinformatics*. doi:10.1093/bib/bbz007

Milledge, J. J., Smith, B., Dyer, P. W., & Harvey, P. (2014). Macroalgae-derived biofuel : a review of methods of energy extraction from seaweed biomass. *Energies*. doi:10.3390/en7117194

Miura, K., Yamano, T., Yoshioka, S., Kohinata, T., Inoue, Y., Taniguchi, F., . . . Fukuzawa, H. (2004). Expression Profiling-Based Identification of CO₂-Responsive Genes Regulated by CCM1 Controlling a Carbon-Concentrating Mechanism in Chlamydomonas reinhardtii. *Plant Physiology, 135*(3), 1595-1607. doi:10.1104/pp.104.041400

Möller, S., Croning, M. D. R., & Apweiler, R. (2001). Evaluation of methods for the prediction of membrane spanning regions. *Bioinformatics, 17*(7), 646-653. doi:10.1093/bioinformatics/17.7.646

Moritz, T. M., Todor, G., Jeremy, N. S., Juliette, J., Madeline, C. M., Robert, J. S., & Howard, G. (2012). Rubisco small- subunit α- helices control pyrenoid formation in Chlamydomonas. *Proceedings of the National Academy of Sciences, 109*(47), 19474. doi:10.1073/pnas.1210993109

Mukherjee, A., & Moroney, J. V. (2019). How protein-protein interactions contribute to pyrenoid formation in Chlamydomonas. *Journal of experimental botany, 70*(19), 5033-5035.

Newman, A. M., & Cooper, J. B. (2007). XSTREAM: A practical algorithm for identification and architecture modeling of tandem repeats in protein sequences. *BMC Bioinformatics, 8*, 382-382. doi:10.1186/1471-2105-8-382

Rawat, I., Ranjith Kumar, R., Mutanda, T., & Bux, F. (2013). Biodiesel from microalgae: A critical evaluation from laboratory to large scale production. *Applied Energy, 103*, 444-467. doi:10.1016/j.apenergy.2012.10.004

Reddy, C. R. K., Dipakkore, S., Kumar, G. R., Jha, B., Cheney, D. P., & Fujita, Y. (2006). An improved enzyme preparation for rapid mass production of protoplasts as seed stock for aquaculture of macrophytic marine green algae. *Aquaculture, 260*(1), 290-297. doi:10.1016/j.aquaculture.2006.06.034

Reddy, C. R. K., Migita, S., & Fujita, Y. (1989). Protoplast Isolation and Regeneration of Three Species of *Ulva* in Axenic Culture. *Botanica Marina, 32*(5), 483-490. doi:10.1515/botm.1989.32.5.483

Rochaix, J.-D. (2017). The Pyrenoid: An Overlooked Organelle Comes out of Age. *Cell, 171*(1), 28-29. doi:10.1016/j.cell.2017.09.012

Saure, M. C. (2005). Calcium translocation to fleshy fruit: its mechanism and endogenous control. *Scientia Horticulturae, 105*(1), 65-89. doi:10.1016/j.scienta.2004.10.003

Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., . . . Schmid, B. (2012). Fiji: an open-source platform for biological-image analysis. *Nature methods, 9*(7), 676-682.

Smetacek, V., & Zingone, A. (2013). Green and golden seaweed tides on the rise. *Nature, 504*(7478), 84-88. doi:10.1038/nature12860

Sterck, L., Billiau, K., Abeel, T., Rouzé, P., & Van de Peer, Y. (2012). ORCAE: online resource for community annotation of eukaryotes. Retrieved from https://bioinformatics.psb.ugent.be/orcae/. from Nat. Methods https://bioinformatics.psb.ugent.be/orcae/

Sudhakar, K., Mamat, R., Samykano, M., Azmi, W. H., Ishak, W. F. W., & Yusaf, T. (2018). An overview of marine macroalgae as bioresource. *Renewable and Sustainable Energy Reviews, 91*, 165-179. doi:https://doi.org/10.1016/j.rser.2018.03.100

Suutari, M., Leskinen, E., Fagerstedt, K., Kuparinen, J., Kuuppo, P., & Blomster, J. (2015). Macroalgae in biofuel production. In (Vol. 63, pp. 1-18).

Tanaka, A., Nagasato, C., Uwai, S., Motomura, T., & Kawai, H. (2007). Re-examination of ultrastructures of the stellate chloroplast organization in brown algae: Structure and development of pyrenoids. *Phycological Research, 55*(3), 203-213. doi:10.1111/j.1440-1835.2007.00463.x

Trivedi, N., Baghel, R. S., Bothwell, J., Gupta, V., Reddy, C. R. K., Lali, A. M., & Jha, B. (2016). An integrated process for the extraction of fuel and chemicals from marine macroalgal biomass. *Scientific Reports, 6*(1), 30728. doi:10.1038/srep30728

Uchida, A., Yoshikawa, T., Ishida, Y., & Saga, N. (1992). Stable Protoplast Isolation and Its Regeneration into Thallus of the Marine Green Alga *Ulva* pertusa. *NIPPON SUISAN GAKKAISHI, 58*(1), 153-157. doi:10.2331/suisan.58.153

Vassilev, S. V., & Vassileva, C. G. (2016). Composition, properties and challenges of algae biomass for biofuel application: An overview. *Fuel, 181*, 1-33. doi:10.1016/j.fuel.2016.04.106

Villarreal, J. C., & Renner, S. R. (2012). Hornwort pyrenoids, carbon- concentrating structures, evolved and were lost at least five times during the last 100 million years. *Proceedings of the National Academy of Sciences, 109*(46), 18873. doi:10.1073/pnas.1213498109

Wichard, T., Charrier, B., Mineur, F., Bothwell, J. H., Clerck, O. D., & Coates, J. C. (2015). The green seaweed *Ulva*: a model system to study morphogenesis. *Frontiers in Plant Science, 6*(72). doi:10.3389/fpls.2015.00072

Wright, P. E., & Dyson, H. J. (1999). Intrinsically unstructured proteins: re-assessing the protein structure-function paradigm. *J Mol Biol, 293*(2), 321-331. doi:10.1006/jmbi.1999.3110

Wunder, T., Cheng, S. L. H., Lai, S.-K., Li, H.-Y., & Mueller-Cajar, O. (2018). The phase separation underlying the pyrenoid-based microalgal Rubisco supercharger. *Nature Communications, 9*(1), 5076. doi:10.1038/s41467-018-07624-w

Zhan, Y., Marchand, C. H., Maes, A., Mauries, A., Sun, Y., Dhaliwal, J. S., . . . Zerges, W. (2018). Pyrenoid functions revealed by proteomics in Chlamydomonas reinhardtii. *PLoS ONE, 13*(2). doi:10.1371/journal.pone.0185039

Zhou, L., Gao, S., Huan, L., Wu, S., Wang, G., & Gu, W. (2020). Enzyme activities suggest that the NAD-ME C4 type CCM exist in *Ulva* sp. *Algal Research, 47*, 101809.