Bioinformatic analysis, isolation and kinetic characterisation of red algae (*Gelidium capense*) dehydrogenases

University of Fort Hare
Together in Excellence

A dissertation submitted in fulfilment of the requirements for the degree of

MASTER OF SCIENCE (MSc) BIOCHEMISTRY

BY

GOGELA YANGA (201010137)

DEPARTMENT OF BIOCHEMISTRY AND MICROBIOLOGY

UNIVERSITY OF FORT HARE

SUPERVISOR: PROF. GRAEME BRADLEY

ABSTRACT

Lactate and alcohol dehydrogenases have attracted much attention in various industries and scientific research for their ability to produce chirally pure compounds and be assayed for activity using more straightforward and reproducible assay methods. These enzymes have been previously isolated and purified from various plants, animals and microorganisms. So far, the molecular and biochemical properties of enzymes from these dehydrogenase families in red algae are mostly unknown. Red macroalgae have been used for centuries for the treatment of various diseases and as a source of ingredients in the food industry. The aim of this study was to identify genes in the sequenced red algae genomes that encode dehydrogenases, to use bioinformatic tools to confirm that the proteins encoded are dehydrogenases and to isolate and kinetically purify alcohol or lactate dehydrogenase from red algae species found along the coastline of the Eastern Cape Province. A combination of bioinformatics tools, molecular and biochemical techniques were used to identify, purify, and characterise ADH and LDH enzymes.

Bioinformatics analysis revealed two alcohol dehydrogenase genes and two hypothetical genes encoding functional domains similar to D-lactate dehydrogenases from other species. The ADH and LDH-like genes shared low sequence identity at the protein level with medium-chain dehydrogenases/reductases (MDRs) and 2-hydroxy acid dehydrogenases, respectively. These two dehydrogenase genes showed a highly conserved NAD-binding motif (Rossmann-fold) similar to many other NAD-dependent dehydrogenases. The ADH and LDH proteins contained no signal peptides and may be located in the cytoplasm. The phylogenetic tree analysis showed that the two ADH genes belonged to cinnamyl and class III alcohol dehydrogenases, whereas the LDH-like genes were grouped with D-lactate dehydrogenases from other organisms. The ADH and LDH gene family showed *cis*-acting regulatory elements that are mostly involved in stress response and hormonal response. Structural analysis showed that the dehydrogenases 3D structure predicted models comprise of two domains, namely the substrate binding and the coenzyme binding domains that are rich in beta-strands secondary structure elements.

The LDH from red algae was purified approximately 4-fold with a specific activity of 0.044 U/mg. The purified LDH enzyme had a molecular weight of approximately 37kDa. The LDH was active across a broad pH range from 5-9 with a pH optimum observed at 7.5. The LDH

enzyme in red algae exhibits a temperature optimum of 40 ºC and heat stability up to 40 ºC. Above 50 °C the LDH activity rapidly decreased showing that the LDH in red algae is not thermostable. The LDH enzyme showed a K_m value of 0.8 mM and V_{max} of 0.0067 $mM.min^{-1}$ when using sodium pyruvate as a substrate.

Keywords: dehydrogenases, red algae, bioinformatics, seaweed, characterisation, *Chondrus crispus*, genome**,** *Gelidium capense*, purification

DECLARATIONS

I, Yanga Gogela do hereby declare that this dissertation titled "Bioinformatics analysis, isolation and kinetic characterisation of red algae (*Gelidium capense*) dehydrogenases" submitted for the award of the MSc (Biochemistry) degree in the Faculty of Science and Agriculture, at the University of Fort Hare, is my own work and has never been submitted for any other degree at this university or any other university.

…….. …….

Yanga Gogela Date

I, Yanga Gogela (student#: 201010137) hereby declare that I am fully aware of the University of Fort Hare's policy on plagiarism and I have taken every precaution to comply with the regulations.

…… ……

Signature Date

I, Yanga Gogela (student#: 201010137) do hereby declare that I am fully aware of the University of Fort Hare's policy on Research Ethics and I have taken every precaution to comply with the regulations. There was no ethical clearance needed for this study.

…….. ……...……

Signature Date

DEDICATION

This work is dedicated to my family, especially to my aunty Cwayita Gogela who played a mother figure in my life. Thank you for keep on challenging me to improve and make a difference in other people's lives so that they can look at my story and see it as an inspiration. I also dedicate this dissertation to my late mother and grandfather, may your souls rest in peace.

ACKNOWLEDGEMENTS

- Obviously there have been challenges and struggles along the way, but I choose not to see it in that light. Instead I see it as a journey.
- I would love to give praise to God Lord Almighty for the wisdom, knowledge, and strength. It has been a blessing, God is really amazing. I am truly blessed in so many ways with people that are around me who helped me for the past few years.
- My gratitude goes to my supervisor professor Graeme Bradley for his courage, expertise and being patient with me.
- I would like to acknowledge the department of Biochemistry and Microbiology staff members, postgraduates, plant stress response and marine science students for their support, ideas, and laughter.
- I would also like to say thanks to Dr. Ntombekhaya Nqumla for being a good friend and mentor for all these years.
- A special mention goes to my best friend and colleague Mr. Zolani Ntsata; we've been in this journey together my brother and thank you for advice, prayers and for being there through the struggle.

- Aknowlegde the National Research Foundation (NRF) and Govan Mbeki Research and Development Centre (GMRDC) at the University of Fort Hare, ACEP Phuhlisa programme and South African Institute and Aquatic Biodiversity (SAIAB) for funding this research.
- Most importantly, I would love to say thank you to my family for unconditional love, spiritual, emotional and financial support and most importantly for their unconditional love.
- GOD BLESSES YOU ALL!!!

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

ABA	Abscisic acid
ABRE	Abscisic acid Responsive Element
ACE	ACGT-containing element
ADH	Alcohol Dehydrogenase
APS	Ammonium persulphate
ARE	Anaerobic Responsive Element
BLAST	Basic Local Alignment Search Tool
BLASTp	Basic Local Alignment Search Tool protein
CAD	Cinnamyl Alcohol Dehydrogenase
CAREs	*cis*-acting regulatory elements
CcADH	Chondrus crispus Alcohol dehydrogenase
CDD	Conserved Domain Database
CmADH	Cyanidioschyzon merolae Alcohol Dehydrogenase
DNA	Deoxyribonucleic acid
GAR	Gibberellin acid responsive element
GsADH	*Galdieria sulphuraria* Alcohol Dehydrogenase
kDa	kilo Daltons
LDH	Lactate Dehydrogenase
LDR	Long chain dehydrogenases/reductases
LRE	Light Responsive Element
LTR	Low-Temperature responsiveness

MBS	MBY binding site
MDRs	Medium chain dehydrogenases/reductases
MEGA	Molecular Evolutionary Genetics Analysis
MeJA	Methyl jasmonate
NAD	Nicotinamide adenine dinucleotide
NADH	Nicotinamide adenine dinucleotide hydrogen
NCBI	National Center for Biotechnology Information
PCR	Polymerase Chain Reaction
pI	Isoelectric point
PMSF	Phenylmethanesulfonylfluoride
PuADH	*Porphyra umbilicalis* Alcohol Dehydrogenase
RNA	Ribonucleic acid
SA	Salicylic acid
SDR	Short-chain dehydrogenases/reductases
SDS	Sodium Dodecyl Sulphate
SDS-PAGE	Sodium Dodecyl Sulphate Polyacrylamide Gel Electrophoresis
TEMED	Tetramethylethylenediamine
TFs	Transcription factors
TRIS-HCl	Tris Hydroxy Methyl Amino Methane Hydrochloric Acid
TSS	Transcription start site

CHAPTER 1: Marine algae: An overview

1.1. Introduction

Algae are plant-like organisms that are found mostly in aquatic environments. These plant-like organisms exist as unicellular species commonly known as microalgae and multicellular species known as macroalgae or seaweeds (Raven and Giordano, 2014). Algae are divided into three broad groups namely Rhodophyceae (red algae), Chlorophyceae (green algae) and Phaeophyceae (brown algae). The red and green algae are members of the Archaeplastida that belong to kingdom Plantae (De Clerck *et al*., 2012). The brown algae belong to kingdom Chromista, a group also referred to as Stramenopiles (or Heterokonts), and this group has been reported to evolve independently of the animal, plants and fungal lineages (Maneveldt *et al*., 1997; Cock *et al*., 2011). Among these three groups of algae, the red algae have been found to be the most essential type because of their biological and economic benefits (El Gamal, 2010).

1.1.1 Red algae genomes

The genomic studies in higher plants and microalgae are well-studied compared to red macroalgae (Ho *et al*., 2017). So far, eight genomes of red algae have been fully sequenced, i.e. *Chondrus crispus* (Collen *et al*., 2013), *Galdieria sulphuraria* (Schonknecht *et al*., 2013), *Cyanidioschyzon merolae* (Matsuzaki *et al*., 2004), *Pyropia yezoensis* (Nakamura *et al*., 2013), *Porphyridium purpureum* (Bhattacharya *et al*., 2013), *Kappaphycus alvarezii* (Yee *et al*., 2017), *Porphyra umbilicalis* (Brawley *et al*., 2017), and *Laurancia dendroidea* (de Oliveira *et al*., 2012). Among these species, only *Laurancia dendroidea* has no genome sequence available in the public domain databases such as the National Centre for Biotechnological Information (NCBI). The availability of these genomes is vital to accelerate functional studies of individual genes and provide information about the evolutionary relationships of the red algae.

1.1.2 Ecological and commercial role of seaweeds

Seaweeds play an ecological role in the marine ecosystem as primary sources of food to many marine organisms, and they supply approximately 50% oxygen to the atmosphere (Chapman, 2013). Commercially, seaweed extracts have been used as food additives, medicinal/pharmaceutical, cosmetic, brewing and chemical components in many industries

(Renn, 1997; Dhargalkar and Pereira, 2005). In addition, about 10% of the marine production is obtained from seaweeds (Battacharyya *et al*., 2015).

The red algae (Rhodophyta) species attracted considerable interest in various fields of scientific research and biotechnology because they contain a wide range of bioactive compounds and enzymes with commercial value. Red algae species from the genera *Gelidium*, *Gracilaria*, and C*hondrus* are essential sources of raw material for extraction of agar and agarose (Rhein-Knudsen *et al*., 2015). The commercial use of algae in developing countries such as South Africa is growing and reported in the literature (Anderson *et al*., 1989; Maneveldt *et al*., 2006). There is a need to expand the use of red algae into other modern areas of application in order to explore the riches of these marine plants. The red algae species are among the most essential macroalgae with anticancer, antiviral, antibacterial and antimicrobial compounds (Seenivasan *et al*., 2010; Zandi *et al*., 2010; Wang *et al*., 2012; Oumaskour *et al*., 2013). Red algae of the genera *Gelidium* have been used in the treatment of kidney, bladder and lung diseases.

Red algae are an untapped source of enzymes, in particular dehydrogenases which have attracted considerable interest for various medical and industrial applications. The lactate dehydrogenases and alcohol dehydrogenases are two crucial enzymes used for various industrial and research purposes worldwide. These dehydrogenase enzymes, although abundant in nature, are mostly underexplored in red algae. Although red algae have been used for various purposes such as the production of agar and carrageenans, molecular and biochemical properties of their dehydrogenases, such as alcohol and lactate dehydrogenases, remain unknown.

Several dehydrogenases have been identified in some of the red algae species, including those with fully sequenced genomes such *Chondrus crispus, Galdieria sulphuraria, Cyanidioschyzon merolae* (Liaud *et al.*, 1993, 1994; Gross and Meyer, 2003). The genome analysis using computational approaches revealed several genes encoding dehydrogenase proteins. The protein sequences encoded by the identified dehydrogenase coding regions obtained in the NCBI database can be used to find similar sequences in the GenBank non-redundancy (nr) protein database. These findings will provide the necessary background information for dehydrogenases at the protein level and also provide insight into experimental design.

1.2. *Chondrus crispus* as a model organism for red algae

Chondrus crispus is a marine red algae species that is found mostly attached to the rocks of the Atlantic coast of Europe and North America. This alga is considered as the model organism for red algae, and its genome is fully sequenced (Collen *et al*., 2013, 2014). According to Collen *et al.* (2013) gene annotation of *C.crispus* is predicted to 9606 genes. Although the use of red algae in scientific research is still far behind compared to land plants and animals, progress has been made over the years to explore the richness of this vital marine plant. *Chondrus crispus* is a marine macroalgae, and easily accessible and therefore there are several studies reported on it in literature.

The genomic information about *C. crispus* is available on the NCBI public domain database (www.ncbi.nlm.nih.gov/genome/). According to Collen *et al*. (2013), the *C. crispus* genome was sequenced using the Sanger technology approach using DNA purified from a clonally unialgal culture of a gametophyte of *C. crispus*. The genome of this alga is made up of 1266 scaffolds that form a 105Mbp size genome (Collen *et al*., 2013). Although the *C. crispus* genome has many different genes, the function and structure of these genes are still unclear. The 9606 genes annotated for *Chondrus crispus* give insight in the evolution of the eukaryotic genomes (Collen *et al*., 2013).

Analysis of the protein-coding genes gives some insight into the function and cellular location of the encoded protein. In the red algae, however, it is difficult to accurately determine the role of these protein-coding genes because most of the proteins in the *C. crispus* are not experimentally validated and are based on the computational analysis. According to Pruitt *et al*., (2014), the predicted proteins are denoted by the prefix accession formats 'XP for proteins and XM for nucleic acid'.

This research study focussed on oxidoreductases, with emphasis on lactate and alcohol dehydrogenases. These enzymes are known for catalysing the oxidation-reduction reactions and are found within the first class of enzymes. These enzymes are some of the most important enzymes in scientific research and for industrial applications.

1.3 Dehydrogenase enzymes

1.3.1 Background

Dehydrogenases are enzymes that belong to class 1 oxidoreductases. Dehydrogenase enzymes carry out most of their reactions using NADH/NAD as a cofactor, and these cofactors are necessary because they activate the biocatalytic function of the enzyme. The dehydrogenases are widely distributed in nature, and several studies have been carried out with these biocatalysts (Valvona *et al*., 2016)

The oxidoreductases are a group of enzymes classified as EC 1 in the EC number enzyme classification. This group comprises the largest families of enzymes that are involved in the catalysis of the oxidation-reduction reactions. The enzymes from this group have the potential to be used for many industrial processes such as in food, cosmetics, textile, polymers, organic synthesis, and pharmaceuticals and in diagnostics (Drent *et al*., 1996; Singh *et al*., 2016). However, microbial enzymes have been favoured over their plant and animal counterparts because of their simplicity in production, stability, catalytic mechanism, and optimization (Singh *et al*., 2016). And therefore, given their ability to catalyze a wide range of biological reactions, the oxidoreductase enzymes attract much attention for use in many biotechnology processes. So far, many industrial uses of the oxidoreductase class 1 enzymes have been reported (Sing *et al*., 2016; Martinez *et al*., 2017).

The oxidoreductases include families of enzymes such as dehydrogenases, oxygenases, reductases, and oxidases. The present study focused on lactate and alcohol dehydrogenase enzymes, which have been proven to be useful for many industrial processes (Ravot *et al*., 2003; Beck *et al*., 2017). Among the industrially important dehydrogenases, alcohol and lactate dehydrogenases attracted the most interest. Alcohol dehydrogenases have been used in the pharmaceutical, food and fine chemical industries for the production of chiral alcohols, aldehydes and ketones (Schoemaker *et al*., 2003; Kroutil *et al*., 2004).

The lactate dehydrogenases play a significant role in the production of different pharmaceutical products (Karaguler *et al*., 2007). The LDH enzyme has been isolated from many different organisms, but the fungal species such as *Saccharomyces cerevisiae*, *Rhizopus oryzae*, *Pichia*

stipites and the *Candida* genera have been used to produce high yield of L-lactate (Skory, 2003; Ilmen *et al*., 2007; 2013; Skory e*t al*., 2007; Ikushima e*t al*., 2009). Lactate dehydrogenases have also been shown to have potential to be a therapeutic target for the treatment of cancer (Valvona *et al*., 2015). The marine environment possesses many important organisms which produce a variety of enzymes that can be used for both scientific research and industrial applications. The marine red algae have been used for the discovery of new knowledge using both computational methods and molecular biology approaches to isolate, purify, and express enzymes.
This allows the characterization of different enzymes from a wide range of marine algal species.

1.3.2 Classification of dehydrogenases

The dehydrogenase enzymes are grouped into three large superfamilies of proteins, namely longchain dehydrogenases/reductases (LDRs), medium-chain dehydrogenases/reductases (MDRs) and short-chain dehydrogenases/reductases (SDRs) (Persson *et al*., 1991; Nordling *et al*., 1994; Jornvall *et al*., 1995). The dehydrogenases/reductases groups are differentiated from each other by their protein size, the structural organization, and catalytic mechanisms. Although, these dehydrogenases/reductases differ from each other, common traits among each other have been noticed, such as the presence of Rossman-fold coenzyme binding motif (Rossman *et al*., 1975; Kallberg and Persson, 2006) and their ability to interconvert substrates containing hydroxyl or Oxo-groups.

1.3.2.1 Short-chain dehydrogenases/reductases (SDRs)

The SDRs are by far the largest group of proteins with more than 46000 protein sequences found in the sequence databases and with more than 300 structures submitted to the protein data bank (PDB) (Persson *et al*., 2009). However, the SDR superfamily of proteins has a relatively low sequence identity at the protein level, with only 15-30% amino acid sequence identity (Jornvall *et al*., 1995). The SDR proteins are present in almost all organisms as NAD(P)(H)-dependent enzymes which catalyse the conversion of alcohols to aldehydes or ketones using nicotinamide adenine dinucleotide phosphate as a cofactor (Gani *et al*., 2008). The SDR enzymes have been used in drug developmental research which indicates their potential use in pharmacology (Oppermann *et al*., 2003).

The SDR proteins have highly conserved amino acid residues such as serine (S), tyrosine (Y), lysine (K) and asparagine (N) that are important for their catalytic activity (Filling *et al*., 2002; Gani *et al*., 2008). Alcohol dehydrogenases are part of the SDR superfamily of proteins and have been isolated and characterized from many organisms. The human SDRs consist of more than 70 genes and have been used for annotation and reference purposes for many other SDR proteins (Oppermann e*t al*., 2001; Bray *et al*., 2009).

1.3.2.2 Medium-chain dehydrogenases/reductases (MDRs)

The MDRs are protein dimers or tetramers that form two-domain subunits, one in the catalytic binding domain and the other in the coenzyme-binding domain (Jornvall and Bergman, 2013). The superfamily of MDRs has more than 15 000 members available in the sequence databases and share around 25% sequence identity (Hedlund *et al*., 2010). The MDRs are present in most organisms, and they are about 350 amino acids in length (Jornvall *et al*., 1999). The subunits of MDR proteins bind two zinc ions, with one zinc ion essential for functional and catalytic activity, whereas the other zinc ion is responsible for structural stability (Branden *et al*., 1975; Jornvall and Bergman, 2013). The zinc-ion containing MDR proteins are denoted as dehydrogenases, whereas non-zinc-ion containing MDRs are denoted as reductases (Hedlund *et al*., 2010).

The MDRs have various enzyme activities of which alcohol dehydrogenases are the most studied (Nordling *et al*., 2002). ADHs from the MDRs superfamily are responsible for catalysing the reversible oxidation of primary and secondary alcohols to their corresponding aldehydes or ketones using the cofactors [NAD(H) or NADP(H)] in the hydride transfer process (Jornvall and Bergman, 2013). The MDRs alcohol dehydrogenases have been isolated and characterized in almost all organisms (Ammendola *et al*., 1992; Danielsson and Jornvall, 1992; Magonet *et al*., 1992; Esposito *et al*., 2002; Sibout *et al*., 2005; Youn *et al*., 2006).

1.3.2.3 Long-chain dehydrogenases/reductases (LDRs)

The LDR protein superfamily is the smallest of the three dehydrogenases/reductases super families with only 66 members that have been classified based on the amino acid sequence similarity (Dean *et al*., 1993; Schneider *et al*., 1993; Adams *et al*., 1994; Pirrung *et al*., 1994; Rowland *et al*., 1994; Liu *et al*., 2000; Ruzheinikov *et al*., 2001; Karsten *et al*., 2002; Klimacek

et al., 2003). The LDR proteins vary from 350 to 560 amino acids in length, and they are mostly found in prokaryotic organisms (Persson *et al*., 1991; Klimacek *et al*., 2003). Mannitol dehydrogenase is the most studied enzyme in the LDR superfamily (Klimacek and Nidetzky, 2002; Bubner *et al*., 2008). The long-chain dehydrogenases contain two domains (catalytic and coenzyme-binding) just like the MDRs family members, with the active site found in the cleft between these domains but employing the lysine-based catalytic mechanism (Kavanagh *et al*., 2002; Klimacek and Nidetzky, 2002). The LDR enzymes have been purified and characterized from several organisms for medical and biotechnological applications (Schneider and Giffhorn, 1989; Diamandis *et al*., 1994; Slatner e*t al*., 1998).

1.4 Alcohol dehydrogenases

Alcohol dehydrogenase is one of the well-investigated families of dehydrogenase enzymes. The ADH enzymes are found in both the SDR and MDR superfamily. Alcohol dehydrogenases (ADH; EC 1.1.1.1), are oxidoreductase enzymes that are widely distributed in all organisms and represent the most abundant classes of enzymes. They are involved in many oxidation or reduction reactions that include inter-conversion between alcohols and ketones or aldehydes using NAD^+, NAD $(P)^+$ or NADH, NADH (P). The ADHs in humans and animals are involved in many processes such as the breakdown of alcohols, or generating useful ketones, aldehyde, or alcohol groups during biosynthesis of different metabolites. However, ADHs in plants, yeast and bacteria plays important role in the catalysis of reverse reaction as part of fermentation to make sure that there is a continuous source of NAD^+.

ADHs related to the MDRs superfamily have been reported in plants (Grima-Pettenati *et* al., 1994; Blanco-Portales *et* al., 2002; Strommer, 2011). The SDR enzymes (generally 250 amino acid residues) have also been studied in plants, but their function is still unclear in plants (Strommer, 2011). The ADH enzyme class 1 (ADH1) was characterized in mammalian organisms in the early 1970s (Jornvall, 1970). Ever since then, many ADH enzymes have been isolated, purified and characterized from many other organisms (Danielsson *et al*., 1994; Jornvall *et al*., 1995; Wilkin *et al*., 1999; Blanco-Portales *et al*., 2002; Larroy *et al*., 2003; Vidal *et al*., 2009). The ADH enzymes are now grouped into 8 different classes in vertebrates (Duester *et al*., 1999; Peralba *et al*., 1999).

The MDRs alcohol dehydrogenase is responsible for the catalysis of alcohols to aldehydes or ketones using NAD^+ as electron acceptor with a wide range of substrates. However, the ADH class III enzymes have shown a specific function in glutathione-dependent formaldehyde dehydrogenase (Staab *et al*., 2008). Other members of the MDRs superfamily of ADH enzymes are described as cinnamyl alcohol dehydrogenases (CAD) and glutathione-dependent formaldehyde dehydrogenases (Dolferus *et al*., 1997; Blanco-Portales *et al*., 2002). Studies have shown that ADH class I-IV enzymes comprised of two zinc atoms that are involved in the catalytic activity and structural stability. The zinc atom that is involved in the catalytic activity is bound by two cysteine and one histidine residues in the ADH sequence (Eklund *et al*., 1976). The human ADH class III enzymes have also been shown to have an extra glutamate that interacts with the catalytic zinc active site (Sanghani *et al*., 2006). The zinc atom that is responsible for structural stability in mammalian ADH is organized by four cysteine residues (Eklund *et al*., 1976; Jelokova *et al*., 1994). The ADH enzymes are well investigated in many organisms, which indicate their essential function in various life forms.

1.4.1. Classes of alcohol dehydrogenases

The class 1 ADH is one of the first enzymes to be crystallized and is now well characterized in the liver and has dehydrogenase activity with ethanol (Jornvall e*t al*., 1993). This class of enzyme exists as homo or heterodimer *in vitro* and converts alcohol to acetaldehyde. The class II ADHs has also been studied and has a different catalytic mechanism to class I ADH. This class of ADH enzymes is highly expressed in the liver and also in the intestine (Estonius *et al*., 1993). ADH class III are enzymes with specific activity with glutathione-dependent formaldehyde dehydrogenase

(GSH-FDH), and it is believed to be the ancestral form of ADH enzymes (Danielsson *et al*., 1994; Persson *et al*., 2008). Studies have shown that ADH III have lower activity with ethanol, which suggests that ADH III and glutathione-dependent formaldehyde dehydrogenase are identical enzymes. However, a recent study has also shown that a glutathione-dependent formaldehyde dehydrogenase homolog from *Bacillus subtilis* had activity with propanol, rather than other common glutathione-dependent formaldehyde dehydrogenase substrates (Ashraf *et al*., 2017).

The class IV ADH is an enzyme that is highly expressed in epithelial cells in the ventricle and gastrointestinal tract (Pares *et al*., 1992; Moreno *et al*., 1996). And studies have suggested that ADH IV play a crucial role in retinoid metabolism, especially in foetal development (Duester *et al*., 1997). The ADH V is one of the least studied classes of ADH enzymes, and it has only been identified as the gene and mRNA with no functional role illustrated (Stromberg and Hoog, 2000). The unsuccessful attempts to isolate and characterize the ADH V protein have been linked to being unstable in a non-cellular environment as compared to other ADH enzymes (Ostberg *et al*., 2016). The ADH class VI has been identified in rat and deer mouse (Zheng *et al*., 1993; Hoog and Brandt, 1995). The ADH VII is the enzyme that is found in humans, and it metabolizes a wide range of substrates (Satre *et al*., 1994). The ADH VII is highly expressed in the stomach than the liver. The ADH VIII is the only class of the ADH enzymes in vertebrates that is specific for NADP(H) and is found in the gastric tissues (Rosell *et al*., 2003; Borras *et al*., 2014).

1.5. Lactate dehydrogenases (LDH)

The lactate dehydrogenase is an enzyme found in animal tissues, plants and in prokaryotes. Under anaerobic glycolysis, the LDH is involved in the sequence of reactions that promote the breakdown of glucose to lactate, and therefore it is vital for the production of ATP and also in the generation of NAD^+ essential in the glycolysis pathway (Kopperschlager and Kirchberger, 1996). The LDH plays an essential role in the last step of the metabolic chain of anaerobic glycolysis by catalyzing the conversion of pyruvate to lactate (Smuts *et al*., 2014). In the Cori cycle, the lactate dehydrogenases present in the liver catalyses the reverse reaction, whereby lactate is converted to pyruvate. The equilibrium, however, strongly favours the catalysis of pyruvate to lactate reaction (Drent *et al*., 1996).

Pyruvate + NADH + H => L-(or D-) lactate + NAD^+ (Drent *et al*., 1996)

The LDH is an enzyme found in body tissues which include skeletal muscle, heart, liver, kidney, lungs, brain, and blood cells. The LDH enzymes have received much attention in many research studies because of their important role in glycolytic metabolism and because it displays different isozymes.

1.5.1. Lactate dehydrogenase source

Lactate dehydrogenase is widely distributed among invertebrates. They are commercially extracted from different sources which include chicken, bovine heart, rabbit muscle, porcine heart and muscle, as well as human heart and erythrocyte. The LDH genes exist as LDH-A, LDH-B, LDH-C, and LDH-D, in which LDH A, B, and C are the L-isomers, while the LDH-D is the D-isomer (Valvona *et al*., 2015). The L-isomer of lactate in all higher animals is the product of the reduction of pyruvate. In bacteria and lower animals, catalysis of D-lactate isomer from pyruvate occurs often. Structurally, the L- isomers and D-isomers are not evolutionary related, and their products belong to the L-specific NAD-dependent dehydrogenases and the D-isomer specific 2-hydroxy acid dehydrogenases respectively (Cristescu *et al*., 2008).

The L-lactate dehydrogenases have been purified to homogeneity from different animal sources, and their physical, kinetic and catalytic properties are reported in the literature (Ike *et al*., 1992; Javed *et al*., 1997; Savijoki and Palva, 1997). Although the D-lactate specific LDHs have (been) purified to some extent (Kochhars *et al*., 1992), there is little knowledge on the structure, function, and kinetic properties of the D-lactate dehydrogenase (Flick and Konieczny, 2002). Despite the several literature studies on LDH in many vertebrates, the occurrence of the LDH in some plants is relatively low (O'carra and Mulcary, 1996). Interestingly, the study done by O'carra and Mulcary (1996) showed that there was no activity in red and brown algae species tested, which shows that the activity of LDH in plant-like organisms is very low.

1.5.2 Lactate dehydrogenase Isozymes

Lactate dehydrogenase exists as five isozymes that are made from the H and M subunits (Valvona *et al*., 2015). The isozymes that are made from the H and M subunits are; LDH1 (HHHH), LDH2 (HHHM), LDH3 (HHMM), LDH4 (HMMM) and LDH5 (MMMM) (Valvona *et al*., 2015). The M subunit is found in the skeletal muscle, and the H form is found in the heart. The H and M subunits have a different electrophoretic mobility and whereby the H (H4) subunit moves in a gel quicker than the M (M_4) subunit (Koolman and Roehm, 2005).

1.5.3 Structural characterization of Lactate dehydrogenase (LDH)

The lactate dehydrogenases are tetrameric enzymes with four subunits of 36 kDa mass each that gives a molecular weight of 144 kDa (Koolman and Roehm, 2005). The L-lactate (EC 1.1.1.27) and the D-lactate (EC 1.1.1. 28) dehydrogenases are two types of dehydrogenases that has been found with different stereo-specificities (Taguchi and Ohta, 1993). Analysis of the primary structure of these dehydrogenases showed that the genes encoding the D-lactate and L-lactate dehydrogenases are evolutionarily different (Taguchi and Ohta, 1991; Cristescu *et al*., 2008). The four LDH subunits have their own active site and a peptide sequence or chain of 334 amino acids (Koolman and Roehm, 2005). The make-up of the tertiary structure of the H and M isozymes of LDH showed some differences in the catalytic properties (Koolman and Roehm, 2005). The LDH1 (H subunit) and LDH5 (M subunit) are characterized by similar active site regions, but the only difference is in 81 out of 332 amino acid positions that are mostly situated in the first 22 and last 38 residues, and they pose a reduced effect on the overall structure (Read *et al*., 2001; Valvona *et al*., 2015). Studies suggested that the LDHA (M subunit) has a higher affinity for pyruvate and they convert pyruvate to lactate with concomitant conversion of NADH to NAD^+, whereas the LDHB (H subunit) has higher affinity for lactate and convert lactate and NAD^+ to pyruvate and NADH respectively (Kopperschlager and Kirchberger, 1996; Read e*t al*., 2001).

1.5.4 Importance of Lactate dehydrogenases

Clinically, lactate dehydrogenase is a critical enzyme that is used to determine whether a particular tissue has been damaged or not and it can serve as a marker for various disease and cell problems (Kopperschlager and Kirchberger, 1996). The LDH activity is found in the cytoplasm in almost all body tissues. The activity of LDH isoenzymes is measured to check its diagnostic value. For example in a study done by Cobben *et al*. (1997) showed that high level of LDH in serum is associated with lung tissue injury. The LDH enzymes have many applications in clinical biochemistry, pharmaceuticals and nanotoxicology. The LDHs have received much attention in research studies due to their importance in carbohydrate metabolism, as a diagnostic tool, and their significance as a tool in the enzyme analysis of other enzymes and biological compounds

(Kopperschlager and Kirchberger, 1996). The present study explores the use of dehydrogenase enzymes, particular from lactate and alcohol dehydrogenase family from the red algae species.

1.6 The proposal for the present study

The considerable interest in the marine red algae is driven by the ever-improving world of biotechnology and the development of new essential products from algal sources. The research studies on the marine organisms in developing countries like South Africa has been done and most of the algal species reported are used for the production of hydrocolloids, abalone farming and in aquaculture industry (Troell *et al*., 2006; Amosu *et al*., 2013). One of the challenges facing the South African macroalgae sector is the identification and availability of sequenced genomes of red algae species. Studies on DNA barcoding (Saunders, 2005; Robba *et al*., 2006; Kogame *et al*., 2017) and genome sequencing are particularly needed in order to identify and sequence genomes of various red algae species that are available in the South African oceans.

Although sequenced red algae genomes have not been reported for species found in South Africa, progress has been made to discover new information using various computational and molecular biology techniques in other developing countries. One of the essential commercial red algae species in Southern Africa belongs to the genera of *Gelidium* and *Gracilaria* which are used for the production of agar (Anderson *et al*., 1989). Based on the information in the literature, most of the red algae species are unexplored and little is known about the enzymes from these sources, especially enzymes that belong to the dehydrogenase family.

Enzymes are biological catalysts that are responsible for catalysing many biochemical reactions in living organisms. Enzymes are unstable in nature and are often isolated under certain environmental conditions such as temperature, pH and ionic strength (Dako *et al*., 2012). Although enzymes are proteins and they display the same isolation and purification methods, however, they differ at some point because enzyme activity must be determined at each purification step in order to get the purified form of the desired enzyme (Reymond and Sicard, 2006). Dehydrogenases are one of the prominent families of enzymes that have attracted a lot of industrial processes and have been isolated and well characterized in many organisms. These enzymes have been purified from a variety of sources such as microbial and yeast sources which

are used for industrial processes (Hummel, 1999; Vieille and Zeikus, 2001; Radianingtyas and Wright, 2003; Schoemaker *et al*., 2003; Karaguler *et al*., 2007; Ilmen *et al*., 2013). The dehydrogenase enzymes from the red algae organisms are still scattered and poorly documented. One of the challenges in the red algae is the difficulty to find suitable methods for protein isolation, and nucleic acid extraction of high purity for biochemical, molecular and cell biology studies (Ho *et al*., 1996; Hu *et al*., 2004; Barbarino and Lourenco, 2005). In addition, some of the plant-specific nucleic acids commercial kits are not reliable, and they result in reduced yields when used in algae (Falcao *et al*., 2008). The red algae are considered as marine plants, and several well-known methods used in other plant research studies can be optimized and used to obtain purification steps and to achieve high yields of quality nucleic acids for biochemical and molecular biology applications respectively.

The present study focuses on enzymes from the dehydrogenase family with emphasis on lactate and alcohol dehydrogenases. These two enzymes have received much attention in research studies because of their importance in scientific research and industrial processes. This thesis will discuss mostly *in-silico* analysis of LDH and ADH enzymes from the red alga *Chondrus crispus* and the extraction, purification and characterization studies will be solely based on lactate dehydrogenase enzyme from *Gelidium capense*. Also in this dissertation, several attempts were made to isolate and clone alcohol dehydrogenase enzymes as discussed in chapter 3.

1.7 Problem statement

The dehydrogenases are perhaps one of the least studied groups of enzymes in the marine red algae species, and yet they are believed to have essential properties for industrial applications. The little knowledge on the marine algae dehydrogenases is perhaps caused by the scarcity of well-commercialized tools that can be used to obtain and purify algal enzymes for both laboratory and industrial processes. The use of dehydrogenase enzymes has attracted many biotechnological processes such as bioremediation, sensors and biofuel cells. Despite the several literature studies on the biochemical and molecular biology of dehydrogenase enzymes in many prokaryotes and eukaryotes, there are few studies reported on the red algae species. The role of the dehydrogenases in plants and more importantly in red algae is still unknown.

1.8 Hypothesis

The red algae species are a new source of dehydrogenase enzymes.

1.9 Aim

The present research is aimed at identifying, isolating, purifying, and kinetically characterising the LDH enzyme from red algae harvested along the coastline of the Eastern Cape.

1.10 Specific objectives to

1.10.1 Identify the red algae genomes and analysis of dehydrogenase genes using bioinformatics tools.

1.10.2 Extract and validate the alcohol dehydrogenase insert from red algae by PCR

1.10.3 Isolate and purify lactate dehydrogenase enzymes from the red algae

1.10.4 Kinetically characterize lactate dehydrogenase from the red algae

1.11 References

Adams, M.J., Ford, G.C., Koekoek, R., Lentz, P.J., McPherson, A., Jr., Rossmann, M.G., Smiley, I.E., Schevitz, R.W., and Wonacott, J.J., 1970. Structure of lactate dehydrogenase at 2-8 A resolution. *Nature. 227,* pp. *1098-1103.*

Adrio, J.L., and Demain, A.L., 2010. Recombinant organisms for production of industrial products. *Bioengineering Bugs.1,* pp. *116-131.*

Adrio, J.L., and Demain, A.L., 2014. Microbial enzymes: Tools for biotechnological processes. *Biomolecules. 4,* pp. *117-139.*

Anderson, R.J., Simons, R.H., and Jarman, N.G., 1989. Commercial seaweeds in Southern Africa: a review of utilization and research. *South African Journal of Marine Science. 8(1),* pp. *277-299.*

Avezov, K., Reznick, A.Z., and Aizenbud, D., 2014. LDH enzyme activity in human saliva: Effect of exposure to cigarette smoke and its different component. *Archives of Oral Biology 59,* pp. *142148.*

Bajpai, V.K., 2014. Antimicrobial bioactive compounds from marine algae: *Indian Journal of Geo-Marine Sciences. 45(9),* pp. *1076-1085.*

Barbarino, E., and Lourenco, S.O., 2005. An evaluation of methods for extraction and quantification of protein from marine macro-and microalgae. *Journal of Applied Phycology. 17,* pp. *447-460.*

Barbosa, M., Valentao, P., and Andrade, P.B., 2014. Bioactive compounds from macroalgae in the new millennium: implications for neurodegenerative diseases. *Marine Drugs. 12,* pp. *4934-4972.*

Battacharyya, D., Babgohari, M.Z, Rathor, P., and Prithiviraj, B., 2015. Seaweed extracts as biostimulants in horticulture. *Scientia Horticulturae. 196,* pp. *39-48.*

Beck, K.R., Kaserer, T., Schuster, D., Odermatt, A., 2017. Virtual screening applications in shortchain dehydrogenase/reductases research. *Journal of Steroid Biochemistry and Molecular Biology. 171,* pp. *157-177.*

Benkovic, S.J., and Hammes-Schiffer, S., 2003. A perspective on enzyme catalysis. *Science. 301(5637),* pp. *1196-1202.*

Bixler, H.J., and Porse, H., 2010. A decade of change in the seaweed hydrocolloids industry. *Journal of Applied Phycology. 23,* pp. *321-335.*

Blanco-Portales, R., Medina-Escobar, N., Lopez-Raez, J.A., Gonzalez-Reyes, J.A., Vaillalba, J.M., Moyano, E. Caballero, J.L., and Munoz-Blanco, J., 2002. Cloning, expression and immunolocalization pattern of a cinnamyl alcohol dehydrogenase gene from strawberry (*fragaria* X *ananassa* cv. Chandler). *Journal of Experimental Botany. 53(375),* pp. *1723-1734.*

Blunt J.W., Copp, B.R., Hu, W.P., Munro, M.H., Northcote, P.T., and Prinsep, M.R., 2007 *Marine natural products. Natural Products Reports. 24(1),* pp. *31-86*.

Bonnichsen, R.K., and Wassen, A.M., 1948. Crystalline alcohol dehydrogenase from horse liver. *Archives of Biochemistry. 18(2),* pp. *361-363.*

Branden, C.L., Eklund, H., Nordstrom, B., Boiwe, T., Soderlund, G., Zeppezauer, E., Ohlsson, I., and Akeson, A., 1973 Structure of liver alcohol dehydrogenase at 2.9-angstrom resolution. *Proceedings of the National Academy of Sciences of the United States of America. 70(8),* pp. *2439-2442.*

Buchner, M., Ford, G.C., Moras, D. Olsen, K.W., and Rossmann, M.G., 1974 Structural determination of crystalline lobster D-glyceraldehyde 3-phosphate. *Journal of Molecular Biology. 82,* pp. *563-585*.

Carlucci, M.J., Pujol, C.A., Ciancia, M., Noseda, M.D., Matulewicz, M.C., Damonte, E.B., and Cerezo, A.S., 1997. Antiherpetic and anticoagulant properties of carrageenans from red seaweed *Gigartina skottsbergii* and their cyclized derivatives: correlation between structure and biological activity. *International Journal of Biology and Macromolecules. 20,* pp. *97-105.*

Cavalier-Smith, T., 2007. Evolution and relationships of algae; major branches of the tree of life in unravelling the algae: The past, present and future of algal systematics, ed., J. Brodie and J. Lewis, pp. 22-25., CRC Press, Boca Raton, London and New York.

Chan, C.-X., Ho, C.-L., and Phang, S.-M., 2006. Trends in seaweed research. *Trends in Plant Science. 11(4),* pp. *165-166.*

Chapman, R.L., 2013. Algae: the world's most important "plants" – an introduction. *Mitig. Adapt. Strategy. Glob. Change. 18,* pp. *5-12.*

Cheleski, J., Freitas, R.F., Wiggers, H.J., Rocha, J.R., de Araujo, A.P.U., and Montanari, C.A., 2011. Expression, purification and kinetic characterization of His-tagged glyceraldehyde-3phosphate dehydrogenase from *Trypanosoma cruzi. Protein Expression and Purification. 76,* pp. *190-196.*

Cheng, F., Hu, T., An, Y., Huang, J., and Xu, Y., 2013. Purification and enzymatic characterization of alcohol dehydrogenase from *Arabidopsis thaliana*. *Protein Expression and Purification. 90,* pp. *74-77.*

Collen, J., Cornish, M.L., Craigie, J., Ficko-Bean, E., Herve, C., Krueger-Hadfield, S.A., *et al*. (2014) "*Chondrus crispus*. A present and historical model organism for red seaweeds," in *Advances in Botanical Research. 71, Sea Plants,* ed. N. Bourgougnon (London: Elsevier), 53-89.

Collen, J., Porcel, B., Carre, W., Ball., S.G., Chaparro, C., Tonon, T., Barbeyron, T., Michel, G., Noel, B., Valentin, K. and Elias, M., 2013. Genomes structure and metabolic features in the red seaweed *Chondrus crispus* shed light on evolution of the Archaeplastida. *Proceedings of the National Academy of Sciences, 110*(13), pp. 5247-5252.

Cristescu, M., Innes, D.J., Stillman, J.H., and Crease, T., 2008. D- and L- lactate dehydrogenases during invertebrate evolution. *BCM Evolutionary Biology. 8,* p. *268.*

Cuesta, S.M., Rahman, S.A., Furnham, N., and Thornton, J.M., 2015. The classification and evolution of enzyme function. *Biophysical Journal. 109(6),* pp. *1082-1086.*

DAFF, 2013, Description of the commercial seaweed sector, accessed at http://www.nda.agric.za/doaDev/sideMenu/fisheries/03_areasofwork/Resources%20Research/Seaweed%20main%20web%20page.pdf

Dako, E, Bernier, A-M., Dadie, A.T., and Jankowski, C.K., 2012. The problems associated with enzyme purification, Chemical Biology, Prof. Deniz Ekinci (ed.), ISBN:978-953-51-0049-2, InTech, accessed at http://www.intechopen.com/books/chemical-biology/the-problemsassociated-with-enzyme-purification.

Datta, D., Talapatra, S.N., and Swarnakar, S., 2015, Bioactive compounds from marine invertebrates for potential medicines- An overview. *International Letters of Natural Sciences. 34,* pp. *42-61.*

Dave, K., and Punekar, N.S. (2015) Expression of lactate dehydrogenase in A*spergillus niger* for L-lactic acid production. *PLoS ONE 10(12),* p.*e0145459.*

Dhargalkar, V.K., and Pereira, N., 2005. Seaweed: Promising plant of the millennium. *Science and Culture, 71,* pp. *3-4.*

Dolferus, R., Osterman, J.C., Peacock, W.J., and Dennis, E.S., 1997. Cloning of the *Arabidopsis* and *rice* formaldehyde dehydrogenase genes: implications for the origin of plant ADH enzymes. *Genetics. 146,* pp. *1131-1141.*

Drent, M., Cobben, N.A.M., Henderson, R.F., Wouters, E.F.M., and Dieijen-Visser, M., 1996. Usefulness of lactate dehydrogenase and its isoenzymes as indicators of lung damage or inflammation. *European Respiratory Journal. 9,* pp. *1736-1742.*

Eklund, H., Branden, C.I., and Jornvall, H., 1976. Structural comparisons of mammalian, yeast and bacillary alcohol dehydrogenase. *Journal of Molecular Biology. 102(1),* pp. *61-73.*

Eklund, H., Nordstrom, B., Zeppezauer, E., Soderlund, G., Ohlsson, I., Bolwe, T., et al., 1976. Three-dimensional structure of horse liver alcohol dehydrogenase at 2.4 Å resolution. *Journal of. Molecular Biology. 102,* pp. *27-59.*

El Gamal, A.A., 2010. Biological importance of marine algae. *Saudi Pharmaceutical Journal. 18(1),* pp. *1-25.*

Ermis, E., 2017. Halal status of enzymes used in the food industry. *Trends in Food Science and Technology. 64,* pp. *69-73.*

Falcao, V.D.R., Tonon, A.P., Oliveira, M.C., and Colepicolo, P., 2008. RNA Isolation method for polysaccharide-rich algae: agar producing *Gracilaria tenuistipitata* (Rhodophyta). *Journal of Applied Phycology. 20(1),* pp. *9-12.*

Fedorov, S.N., Ermakova, S.P., Zvyagintseva, T.N., and Stonik, V.A., 2013 Anticancer and cancer preventive properties of marine polysaccharides: Some results and prospects. *Marine Drug. 11,* pp. *4876-4901.*

Filling, C., Berndt, K.D., Benach, J., Knapp, S., Prozorovski, T., Nordling, E., Ladenstein, R., Jornvall, H., and Oppermann, U., 2002. Critical residues for structure and catalysis in short-chain dehydrogenases/reductases. *Journal of Biological Chemistry. 277,* pp. *25677-25684.*

Flick, M.J., and Konieczny, S.F., 2002. Identification of putative mammalian D-lactate dehydrogenase enzymes. *Biochemica Biophysica Research Communications. 295,* pp. *910-916.*

Gade, R., Tulasi, M.S., and Bhai, V.A., 2013. Seaweeds: a novel biomaterial. *International Journal of Pharmacy and Pharmaceutical Sciences. 5(2),* pp. *40-44.*

Gross, W., and Meyer, A. 2003. Distribution of myo-inositol dehydrogenase in algae. *European Journal of Phycology. 38(3),* pp. *191-194.*

Guiry, M.D., 2012. How many species of algae are there? *Journal of Phycology. 48,* pp. *10571063.*

Ho, C.L., Phang, S.M., Sinnappah, N.D., and Pang, T., 1996. Molecular approaches in the taxonomy of the red and brown seaweeds. In: Chaudhary BR and Agrawal SB (Eds.) Cytology, Genetics and Molecular Biology of Algae. Amsterdam: SPD Academic Publishing. pp. 351-362.

Holdt, S.L., and Kraan, S., 2011. Bioactive compounds in seaweed: functional food applications and legislation. *Journal of Applied Phycology. 23(3),* pp. *543-597.*

Hood, E.E., 2002, From green plants to industrial enzymes. *Enzyme and Microbial Technology. 30,* pp. *279-283.*

Hoog, J-O., Herdberg, J.J., Stromberg, P., and Svensson, S., 2001. Mammalian alcohol dehydrogenases- functional and structural implications. Jo*urnal of Biomedical Sciences. 8,* pp. *71-*
6.

Hu, Z., Zeng, X., Wang, A., Shi, C., and Duan, D., 2004. An efficient method for DNA isolation from red algae. *Journal of Applied Phycology. 16(3),* pp. *161-166.*

Huang, J., Huang, Y., Wu, X., Du, W., Yu, X., Hu, X., 2009. Identification, expression, characterization, and immunolocalization of lactate dehydrogenase from *Taenia asiatica*. *Parasitology Research. 104,* p. *287.*

Hummel, W., 1999. Large-scale applications of NAD(P)-dependent oxidoreductases: recent developments. *Trends in Biotechnology. 17,* pp. *487-492.*

Huo, J., Shi, H., Yao, Q., Chen, H., Wang, L., and Chen, K., 2010. Cloning and purification of recombinant silkworm dihydrolipoamide dehydrogenase expressed in *Escherichia coli*. *Protein Expression and Purification. 72(1),* pp. *95-100.*

Ike, J., Sangan, P., and Gunasekaran, M., 1992. Purification and properties of lactate dehydrogenase from Nocardia asteroids. *Microbios. 69(279),* pp. *119-127.*

Ilmen, M., Koivuranta, K., Ruohonen, L., Rajgarhia, V., Suominen, P., and Penttila, M., 2013. Production of L-lactic acid by the yeast *Candida sonorensis* expressing heterologous bacterial and fungal lactate dehydrogenases. *Microbial Cell Factories. 12,* p.*53.*

Javed, M.H., Azimuddin, S.M.I., Hussain, A.N., Ahmed, A., and Ishaq, M., 1997. Purification and characterization of lactate dehydrogenase from Varanus liver. *Experimental and Molecular Medicine. 29(1),* pp. *25-30.*

Jelokova, J., Karisson, C., Estonius M., Jornvall, H., and Hoog, J-O., 1994. Features of structural zinc in mammalian alcohol dehydrogenase. Site-directed mutagenesis of the zinc ligands. *European Journal of Biochemistry. 225,* p.*1015-9.*

Jornvall, H., and Harris, J.I., 1970. Horse liver alcohol dehydrogenase. On the primary structure of the ethanol-active isoenzyme. *European Journal of Biochemistry. 13(3),* pp. *565-576.*

Jornvall, H., Landreh, M., and Ostberg, L.J., 2015. Alcohol dehydrogenase, SDR and MDR structural stages, present update and altered era. *Chemico-Biological Interactions. 234,* pp. *75-9.*

Jornvall, H., Persson, B., Krook, M., Atrian, S., Gonzalez-Duarte, R., Jeffery, J., and Gosh, D., 1995. Short-chain dehydrogenases/reductases (SDR), *Biochemistry. 34,* pp. *6003-6013.*

Jornvall, H., Shafqat, J., and Persson, B., 2001. Variation and constant patterns in eukaryotic MDR enzymes conclusions from novel structures and characterized genomes. *Chemico-Biological Interactions. 131-132,* pp. *491-498.*

Kallberg, Y., and Persson, B., 2006. Prediction of coenzyme specificity in dehydrogenases/reductases: a hidden Markov model-based method and its application on complete genomes. *FEBS Journal. 273,* pp. *1177-1184.*

Kallberg, Y., Oppermann, U., and Persson, B., 2010. Classification of the short-chain dehydrogenase/reductases superfamily using hidden Markov models. *The FEBS journal. 277(10),* pp. *2375-2386.*

Karaguler, N.G., Sessions, R.B., Binay, B., Ordu, E.B., and Clarke, A.R., 2007. Protein engineering applications of industrially exploitable enzymes: Geobacillus stearothermophilus LDH and Candida methylica FDH. *Biochemistry Society. Transactions. 35(6),* pp. *1610-1615.*

Kavanagh, K.L., Jornvall, H., Persson, B., and Oppermann, U., 2008. Functional and structural diversity within the short-chain dehydrogenase/reductases (SDR) superfamily. *Cellular and Molecular Life Sciences. 65,* pp. *3895-3906.*

Kavanagh, K.L., Klimacek, M., Nidetzky, B., and Wilson, D.K., 2002. Crystal structure of *Pseudomonas fluorescens* Mannitol 2-dehydrogenase binary and ternary complexes: specificity and catalytic mechanism. *Journal of Biologial Chemistry. 277,* pp. *43433-43442.*

Kharkwal, H., Joshi, DD., Panthari, P., Pant, M.K., and Kharkwal, A.C., 2012. Algae as future drugs. *Asian Journal of Pharmaceutical and Clinical Research. 5(4),* pp.*1-4.*

Kim, S.-K., and Dewapriya, P., 2012. Bioactive compounds from marine sponges and their symbiotic microbes: A potential source of nutraceuticals. *Advances in Food and Nutrition Research. 65,* pp.*137-151.*

Kirk, O., Borchert, T.V., and Fuglsang, C.C., 2002. Industrial enzyme applications. C*urrent Opinions in Biotechnology. 13(4),* pp.*345-351.*

Klimacek, M., and Nidetzky, B., 2002. A catalytic consensus motif for D-mannitol 2dehydrogenase, a member of a polyol-specific long dehydrogenase family, revealed by kinetic characterization of site-directed mutants of the enzyme from Pseudomonas fluorescens. *Biochemistry Journal. 367,* pp. *13-18.*

Kochhars, S., Hunziker, P.E., Leong-Morgenthaler, P., and Hottinger, H., 1992. Primary structure, physiochemical properties, and chemical modification of NAD(+)-dependent D-lactate dehydrogenase. Evidence for the presence of Arg-235, His-303, Tyr-101, and Trp-19 at or near the active site. *Journal of Biological Chemistry. 267,* pp. *8499-8513.*

Kogame, K., Uwai, S., Anderson, R.J., Choi, H.-G., and Bolton, J.J., 2017. DNA barcoding of South African geniculate coralline red algae (Corallinales, Rhodophyta). *South African Journal of Botany. 108,* pp. *337-341.*

Koolman, J., and Roehm, K.M., 2005. Color atlas of biochemistry, 2nd ed, Georg Thieme Verlag.

Kopperschlager, G., and Kirchberger, J., 1996. Methods for the separation of lactate dehydrogenases and clinical significance of the enzyme. *Journal of Chromatography B Biomedical Sciences and Applications. 684,* pp. *25-49.*

Kroutil, W., Mang, H., Edegger, K., and Faber, K., 2004. Biocatalytic oxidation of primary and secondary alcohols. *Advanced Synthesis & Catalysis. 346(2-3),* pp. *125-142.*

Kulshreshtha, G., Borza, T., Rathgeber, B., Stratton, G.S., Thomas, N.A., Critchley, A., Hafting, J., and Prithiviraj, B. (2016) Red seaweeds *Sarcodiotheca gaudichaudii* and *Chondrus crispus* down regulate virulence factors of *Salmonella* Enteritidis and induce immune response in *Caenorhabditis elegans*. *Frontiers in Microbiology. 7,* p. *421.*

Leskovac, V., Trivic, S., and Pericin, D., 2002. The tree zinc-containing alcohol dehydrogenases from baker's yeast, Saccharomyces cerevisiae. *FEMS Yeast Research, 2(4),* pp. *481-494.*

Lewis, L.A., and McCourt, R.M., 2004. Green algae and the origin of land plants. *American Journal of Botany. 91,* pp. *1535-1556.*

Li, S., Yang, X., Yang, S., Zhu, M., and Wang, X., 2012. Technology prospecting on enzymes: Application, marketing and engineering. *Computational and Structural Biotechnology Journal. 2(3),* p. *e201209017.*

Li, Y., Ogola, H.J.O. and Sawa, Y., 2012. L-Aspartate dehydrogenase: features and applications. *Applied Microbiology and Biotechnology. 93(2),* pp. *503-516.*

Liaud, M.F., Valentin, C., Brandt, U., Bouget, F.Y., Kloareg, B., and Cerff, R., 1993. The GAPDH gene system of the red alga *Chondrus crispus*: promoter structures, intron/exon organization, genomic complexity and differential expression of genes. *Plant Molecular Biology. 23(5),* pp.*981994.*

Liaud, M.F., Valentin, C., Brandt, U., Bouget, F.Y., Kloareg, B., and Cerff, R., 1994. The evolutionary origin of red algae as deduced from the nuclear genes encoding cytosolic and

chloroplast glyceraldehyde-3-phosphate dehydrogenases from *Chondrus crispus*. *Journal of Molecular Evolution. 38(4),* pp. *319-327.*

Lin, C-F., Jiang, R-H., Jiang, L-Z., Qian, X-Y., Attia, K., and Yang, J-S. (2004) Cloning, characterization and prokaryotic expression of cytosolic malate dehydrogenase from *Oryza sativa*. *DNA Sequence. 15(4),* pp. *314-318.*

Lin, Q. and Qin, S., 2014, Tipping points in seaweed genetic engineering: scaling up opportunities in the next decade. *Marine Drugs. 12,* pp. *3025-3045.*

Liu, Z., Weis, R., and Glieder, A., 2004. Enzymes from higher eukaryotes for industrial biocatalysts. *Food Technology and Biotechnology. 42(4),* pp. *237-249.*

Machielsen, R., Uria, A.R., Kengen, S.W.M., and van der Oost, J., 2006. Production and characterization of a thermostable alcohol dehydrogenase that belongs to the aldo-keto reductases superfamily. *Applied and Environmental Microbiology. 72(1),* pp. *233-238.*

Martinez-Rivas, J.M., and Vega, J.M. (1998) Purification and Characterization of NAD-Isocitrate dehydrogenase from *Chlamydomonas reinhardtii*. *Plant Physiology. 118,* pp. *249-255.*

Matsukazi, M., Misumi, O., Shin-i, T., Maruyama, S., Takahara, M., Miyagishima, S., Mori, T., Nishida, K., Yagisawa, F., Nishida, K., Yoshida, Y., Nishimura, Y., Nakao, S.,Kobayashi, T., Momoyama, Y., Higashiyama, T., Minoda, A., Sano, M., Nomoto, H., Oishi, K., Hayashi, H., Ohta, F., Nishizaka, S., Haga, S., Miura, S., Morishita, T., Kabeya, Y., Terasawa, K., Suzuki, Y., Ishii, Y., Asakawa, S., Takano, H., Ohta, N., Kuroiwa, H., Tanaka, K., Shimizu, N., Sugano, S., Sato, N., Nozaki, H., Ogasawara, N., Kohara Y., and Kuroiwa, T., 2004. Genome sequence of the ultrasmall unicellular red alga Cyanidioschyzon merolae 10D. *Nature. 428,* pp. *653-657.*

Mayer, A.M.S., and Lehmann, V.K.B., 2000. Marine pharmacology in 1998: Marine compounds with antibacterial, anticoagulant, antifungal, anti-inflammatory, anthelmintic, antiplatelet, antiprotozoal, and antiviral activities; with actions on the cardiovascular, endocrine, immune, and nervous systems; and other miscellaneous mechanisms of action. *Pharmacologist. 42,* pp. *62-69.*

McHugh, D. J., 2003. A guide to the seaweed industry, FAO Fisheries technical papers Rome, Italy, 9-25104-958-0 (441), 105.

McHugh, D.J., 1987. Production and utilization of products from commercial seaweeds. 288, 1189. FAO Fisheries technical paper no. 288, pp.1-189.

Milledge, J.J., Smith, B., Dyer, P.W., and Harvey, P., 2014. Macroalgae-derived biofuel: a review of methods of energy extraction from seaweed biomass. *Energies. 7,* pp. *7194-7222.*

Moran, J.F., Sun, Z., Sarath, G., Arredondo-Peter, R., James, E.K., Becana, M., and Klucas, R.V., 2002. Molecular cloning, functional characterization, and subcellular localization of soybean nodule dihydrolipoamide reductases. *Plant Physiology. 128(1),* pp. *300-313.*

Moummou, H., Kallberg, Y., Tonfack, L.B., Persson, B., and van der Rest, B., 2012. The plant short-chain dehydrogenase (SDR) superfamily:genome-wide inventory and diversification patterns. *BMC Plant Biology. 12,* p. *219.*

Musumeci, V., Cherubini, P., Zuppi, C., et al (1993) Aminotransferases and lactate dehydrogenase in saliva of diabetic patients. *Journal of Oral Pathology. 22,* pp. *73-76.*

Neori, A., Shpigel, M., and Ben-Ezra, D. (2000) A sustainable integrated system for culture of fish, seaweed and abalone. *Aquaculture. 186,* pp. *279-291.*

Nordling, E., Jornvall, H., and Persson, B., 2002. Medium-chain dehydrogenases/reductases (MDRs): family characterizations including genome comparisons and active site modeling. *Eur. Journal of Biochemistry. 269,* pp. *4267-4276.*

O'carra, P., and Mulcary, P., 1996. Lactate dehydrogenase in plants: distribution and function. *Phytochemistry. 42(3),* pp. *581-587.*

Oppermann, U., Filling, C., Hult, M., Shafqat, N., Wu, X., Lindh, M., Shafqat, J., Nordling, E., Kallberg, Y., Persson, B. et al., 2003. Short-chain dehydrogenases/reductases (SDR): the 2002 update. *Chemico-Biological Interactions. 143-144,* pp. *247-253.*

Oumaskour, K., Boujaber, N., Etahiri, S., and Assobhei, O., 2013. Anti-inflammatory and antimicrobial activities of twenty-three marine red algae from the coast of Sidi Bouzid (El JadidaMorocco). *International Journal of Pharmacy and Pharmaceutical Sciences. 5(3),* pp. *145-149.*

Perez-Llorens, J.L., Benitez, E., Vergara, J.J. and Berges, J.A., 2003. Characterization of proteolytic enzyme activities in macroalgae. *Journal of Phycology., 38,* pp *55-64.*

Persson, B., and Kallberg, Y. (2013) Classification and nomenclature of the superfamily of shortchain dehydrogenases/reductases (SDRs). *Chemico-Biological Interactions. 202,* pp. *111-115.*

Persson, B., Krook, M., and Jornvall, H. (1991) Characteristics of short-chain alcohol dehydrogenases and related enzymes. *European Journal of Biochemistry. 200,* pp. *537-543.*

Pruitt, K.D., Brown, G.R., Hiatt, S.M., Thibaud-Nissen, F., Astashyn, A., Ermolaeva, O., Farrell, C.M., Hart, J., Landrum, M.J., McGarvey, K.M., Murphy, M.R., O'Leary, N.A., Pujar, S., Rajput, B., Rangwala, S.H., Riddick, L.D., Shkeda, A., Sun, H., Tamez, P., Tully, R.E., Wallin, C., Webb, D., Weber, J., Wu, W., DiCuccio, M., Kitts, P., Maglott, D.R., Murphy, T.D., and Ostell, J.M. (2014) RefSeq: An update on mammalian reference sequences. *Nucleic Acids Research. 42(Database issue),* pp. *D756-D763.*

Radianingtyas, H. and Wright, P.C. (2003) Alcohol dehydrogenases from thermophilic and hyperthermophilic archaea and bacteria. *FEMS Microbiol Rev. 27: 593-616.*

Ravot, G., Wahler, D., Favre-Bulle, O., Cilla, V., and Lefevre, F., 2003. High throughput discovery of alcohol dehydrogenases for industrial biocatalysis. *Advanced Synthesis & Catalysis. 345(6-7),* pp. *691-694.*

Read, J.A., Winter V.J., Eszes, C.M., Sessions, R.B., and Brady, R.L., 2001. Structural basis for altered activity of M- and H- isoenzyme forms of human lactate dehydrogenase. *Proteins. 43,* pp. *175-185.*

Rengasamy, K.R., Kulkarni, M.G., Stirk, W.A., and Van Staden, J., 2014. Advances in algal drug research with emphasis on enzyme inhibitors. *Biotechnology Advances. 32(8),* pp. *1364-1381.*

Renn, D., 1997. Biotechnology and the red seaweed polysaccharide industry: status needs and prospects. *Trends in Biotechnology. 15,* pp. *9-14.*

Rhein-Knudsen, N., Ale, M.T., and Meyer, A.S., 2015. Seaweed hydrocolloid production: an update to enzyme assisted extraction and modification technologies. *Marine Drugs, review. 13* pp. *3340-3359.*

Robba, L., Russell, S.J., Barker, G.L., and Brodie, J., 2006. Assessing the use of the mitochondrial *cox1* marker for use in DNA barcoding of red algae (Rhodophyta). *American Journal of Botany. 93(8),* pp. *1101-1108.*

Rossmann, M.G, Liljas, A., Branden, C.I., and Banaszk, L., 1975. Evolutionary and structural relationships among dehydrogenases. In: The enzymes, 3rd edn., 2:61-102, Boyer, P.D. (ed.), Academic Press, NY.

Rossmann, M.G., Moras, D., and Oslen, K.W., 1974. Chemical and biological evolution of the nucleotide-binding protein. *Nature. 250,* pp. *194-199.*

Saha, D. and Bhattachyra, S., 2010. Hydrocolloids as thickening and gelling agents in food: a critical review. *Journal of Food Science and Technology. 47(6),* pp. *587-597.*

Samarakoon, K. and Jeon, Y-J., 2012. Bio-functionalities of proteins derived from marine algae-A review. *Food Research International. 48,* pp. *948-960.*

Sanghani, P.C., Davis, W.I., Zha, L., and Robinson, H., 2006. Structure-function relationships in human glutathione-dependent formaldehyde dehydrogenase. Role of Glu-67 and Arg-368 in the catalytic mechanism. *Biochemistry. 45,* pp. *4819-30.*

Saunders, G.W., 2005. Applying DNA barcoding to red macroalgae: a preliminary appraisal holds promise for future applications. *Philos. Trans. R. Soc. Lond. B Biol Sci. 360(1462),* pp. *1879-1888.*

Savijoki, K., and Palva, A., 1997. Molecular genetic characterization of the L-lactate dehydrogenase gene (ldhL) of *Lactobacillus helveticus* and biochemical characterization of the enzyme. *Applied Environmental Microbiology 63(7),* pp. *2850-2856.*

Schoemaker, H.E., Mink, D., and Wubbolts, M.G., 2003. Dispelling the myths-biocatalysis in industrial synthesis. *Science. 299,* pp. *1694-1697.*

Schönknecht, G., Chen,W.-H., Ternes, C.M., Barbier, G.G., Shrestha, R.P., Stanke, M., Bräutigam, A., Baker, B.J., Banfield, J.F., Garavito, M.,Carr, K., Wilkerson, C., Rensing, S.A., Gagneul, D., Dickenson, N.E., Oesterhelt, C., Lercher, M.J. Weber, A.P.M., 2013. Gene transfer from bacteria and archaea facilitated evolution of an extremophilic eukaryote. *Science. 339,* pp.*1207-1210.*

Seenivasan, R., Indu, H., Archana, G., and Geetha, S., 2010. The antibacterial activity of some marine algae from South East coast of India. *American Eurasian Journal of Agriculture. & Environmental Science. 9(5),* pp. *480-489.*

Shanmugum, M. and Mody, K.H., 2000. Heparinoid-active sulphated polysaccharides from marine algae as potential blood anticoagulant agents. *Current Science. 79(12),* pp. *1672-1683.*

Shao, Z., Zhang, P., Li, Q., Wang, X., and Duan, D., 2014. Characterization of Mannitol-2dehydrogenase in *Saccharina japonica*: evidence for new polyol-specific long-chain dehydrogenases/reductases. *PLoS ONE 9(5)* p.*e97935.*

Gross, W., and Meyer, A., 2003. Distribution of Myo-inositol dehydrogenase in algae. *European Journal of Phycology. 38(3),* pp. *191-194.*

Martinez-Rivas, J.M., and Vega, J.M., 1998. Purification and Characterization of NAD-Isocitrate dehydrogenase from Chlamydomonas reinhardtii. *Plant Physiol. 118,* pp. *249-255.*

Christopher, M.E., and Good, A.G., 1996. Characterization of hypoxically inducible lactate dehydrogenase in maize. *Plant Physiology. 112,* pp. *1015-1022.*

Shao, Z., Zhang, P., Li, Q., Wang, X., and Duan, D., 2014. Characterization of Mannitol-2dehydrogenase in *Saccharina japonica*: evidence for new polyol-specific long-chain dehydrogenases/reductases. *PLoS ONE 9(5),* p.*e97935.*

Singh, R., Kumar, M., Mittal, A., and Mehta, P.K.. 2016. Microbial enzymes: Industrial progress in the 21st century. *Biotechnology. 6,* p. *174.*

Amosu, A.O., Robertson-Andersson, D.V., Maneveldt, C.W., Anderson, R.J., and Bolton, J.J., 2013. South African seaweed aquaculture: A sustainable development example for other African coastal countries. *African Journal of Agricultural Research. 8(43),* pp. *5268-5279.*

Skory, C.D., 2000. Isolation and expression of lactate dehydrogenase genes from *Rhizopus oryzae*. *Applied Environmental Microbiology. 66(6),* pp.*2343-2348.*

Smuts, C., Millis, J., Myles, R., and Gaal, T., 2016. Lactate dehydrogenase activity in abdominal fluid from horses with colic. *Journal of Equine Veterinary Science. 36,* pp. *58-62.*

Spieler, R., 2002. Seaweed compound's anti-HIV efficacy will be tested in Southern Africa. *Lancet. 359(9318),* p. *1675.*

Staab, C.A., Hellgren, M., and Hoog, J-O., 2008. Medium- and short-chain dehydrogenase/reductases gene and protein families. Dual functions of alcohol dehydrogenase 3: implications with a focus on formaldehyde dehydrogenase and S-nitrosoglutathione reductases activities. *Cell & Molecular Life Science. 65,* pp.*3950-60.*

Strommer, J., 2011. The plant ADH gene family. The plant genome: an evolutionary view of structure and function. *The Plant Journal. 66,* pp. *128-142.*

Taguchi, H., and Ohta, T., 1991. D-lactate dehydrogenase is a member of the D-isomer-specific 2-hydroxyacid dehydrogenase family. Cloning, sequencing, and expression in Escherichia coli of the D-lactate dehydrogenase gene of *Lactobacillus plantarum*. *The Journal of Biological Chemistry. 266,* pp. *12588-12594.*

Taguchi, H., and Ohta, T., 1993. Histidine 296 is essential for the catalysis in *Lactobacillus plantarum* D-lactate dehydrogenase. *The Journal of Biological Chemistry. 268(24),* pp. *1803018034.*

Thomas, N.V. and Kim, S-K., 2013. Beneficial effects of marine algal compounds in cosmeceuticals. *Marine Drug., 11,* pp. *146-164.*

Torres, F.A.E., Passalacqua, T.G., Velasquez, A.M.A., de Souza, R.A., Colepicolo, P., and Graminha, M.A.S., 2014. New drugs with antiprotozoal activity from marine algae: *Rev. Bras. Farmacogn. 24,* pp. *265-276.*

Trincone, A., 2011. Marine biocatalysts: enzymatic features and application. *Marine Drugs. 9,* pp. *478-499.*

Troell, M., Robertson-Andersson, D., Anderson, R.J., Bolton, J.J., Maneveldt, G., Halling, C., and Probyn, T., 2006. Abalone farming in South Africa: an overview with perspectives on kelp resources, abalone feed, the potential for on-farm seaweed production and socio-economic importance. *Aquaculture. 257,* pp. *266-281.*

Truniger, V. and Boos, W., 1994. Mapping and cloning of *gldA*, the structural gene of the *Escherichia coli* glycerol dehydrogenase. *Journal of Bacteriology. 176(6),* pp. *1796-1800.*

Valvona, C.J., Fillmore, H.L., Nunn, P.B., and Pilkington, G.J., 2015. The regulation and function of lactate dehydrogenase A: Therapeutic potential in brain tumor. *Brain Pathology. 26(1),* pp. *317.*

Vieille, C., and Zeikus, G.J., 2001. Hyperthermophilic enzymes: sources, uses, and molecular mechanisms for thermostability. *Microbiology and Molecular Biology. Rev. 65,* pp. *1-43.*

Wang, N., Shi, H., Yao, Q., Zhou, Y., Kang, L., Chen, H., and Chen, K., 2011. Cloning, expression and characterization of alcohol dehydrogenases in the silkworm *Bombyx mori*. *Genetics and Molecular Biology. 34(2),* pp. *240-243.*

Wang, W., Wang, S-X., and Gaun, H-S., 2012. The antiviral Activities and Mechanisms of Marine Polysaccharides: An overview. *Marine Drugs. 10,* pp. *2795-2816.*

Zandi, K., Tajbakhsh, S., Nabipour, I., Rastian, Z., Yousefi, F., Sharafian, S., and Sartavi, K., 2010. *In vitro* antitumor activity of *Gracilaria corticata* (a red alga) against Jurkat and molt-4 human cancer cell lines. *African Journal of Biotechnology. 9(40),* pp. *6787-6790.*

Zeng, T., Liu, S., Luo, R., Gong, P., Zhao, D., and Fang, X., 2011. Cloning and expression of an Alcohol Dehydrogenase from *Lotus japonicus* and Characterization of LjADH1. *Legume Genomics and Genetics. 2(2),* pp. *6-13.*

Zhang, C., and Kim, S.-K., 2010. Research and application of marine microbial enzymes: Status and prospects. *Marine Drugs. 8(6),* pp. *1920-1934.*

Zheng, Y., Guo, S., and Wang, X., 2004. Effects of N-terminal deletion mutation on rabbit muscle lactate dehydrogenase. *Biochemistry (Mosc). 69,* pp. *401-406.*

CHAPTER 2: Bioinformatics analysis of alcohol and lactate dehydrogenase gene families from red algae genomes

2.1 Introduction

Genomic studies in algae are very scarce and limited as compared to other terrestrial plants. The limited information on algal genomes makes the functional analysis of genes from red algae difficult, considering low similarities between algae and other organisms, including terrestrial plants (Chan *et al*., 2006; Tirichine *et al*., 2011). The use of bioinformatics tools to analyse red algae genomes should accelerate and contribute to the ongoing advances in molecular and computational approaches for red algae research.

So far *Chondrus crispus* is one of the very few red seaweed species with a fully sequenced and annotated genome (Collen *et al*., 2013). Several dehydrogenase sequences are available in the genome of this red seaweed. However, the functional characteristics of these gene encoding sequences in *C. crispus* has not been reported. The alcohol and lactate dehydrogenases are widely distributed and characterized in several organisms. Many members of alcohol dehydrogenases identified and characterized at the gene level in plants belong to the MDRs superfamily (Chase, 1999).

2.2 Subcellular localization and Biophysical properties

Protein subcellular localization play important role in prediction of a protein function. Many proteins in a cell are encoded in the nuclear DNA (Hoglund *et al*, 2005). ADH and LDH in plant species have been found in the cytoplasm (Wang *et al*., 2009). Several computational prediction methods for subcellular localization available include SignalP, TargetP, and WoLF-PSORT. SignalP target proteins for translocation in the plasma membrane and endoplasmic reticulum in prokaryotes and eukaryotes respectively (Emanuelsson *et al*., 2007). WoLF-PSORT is also a good prediction method that coverts protein amino acid sequences into numerical features based on functional motifs (Horton *et al*., 2007). LoctTree3 is fast tool used to predict subcellular in domains of life. Protparam from ExPAsy tool at the SIB bioinformatics resource portal is used to analyse physiochemical properties in proteins sequence (Gasteiger *et al*., 2003).

2.3 Sequence and Structure analysis

Structure-based alignment plays important role in predicting and improving accuracy of the function of protein (Branden and Tooze, 1999; Kim and Lee, 2007). The 3D structures of LDH and ADH proteins have been predicted and well characterized in other organisms (Kim *et al*., 2014; Raj *et al*., 2014). Most LDH proteins possess a beta-strand rich structure and consist of two conserved substrate binding and cofactor binding domains (Stoll *et al*., 1996; Kim *et al*., 2014; Zhu *et al*., 2015). The ADH protein structures possess similar overall fold containing catalytic binding and coenzyme binding with active site residues located in the interdomain cleft (Yang *et al*., 1997; Esposito *et al*., 2002).

Basic Local Alignment Search Tool (BLAST) searches proteins and nucleotide sequence identity similarity in the non-redundant database (GenBank) (Atschul *et al*., 1997). These sequences are then used in alignment using tools like Clustal Omega to predict possible function through sequence comparison of known function (Larkin *et al*., 2007). The sequences with similarities especially in the active site residues are expected to share common structure and function.

2.4 Regulation of gene expression

Initiation and regulation of transcription is coordinated by transcription factors. The *cis*-acting regulatory elements in the promoter control transcriptional regulation (Hughes *et al*., 1999; Hernandez-Garcia and Finer, 2014). Plant-based databases such as PlantCARE and PLACE are used to scan the 5′ upstream region for regulatory elements (Hugo *et al*., 1999; Lescot *et al*., 2002).

To know how the gene expression is regulated is essential to fully understand the function of a genome and search for therapeutic compounds for industrial and medical applications. Gene expression regulation is a much more complex network in eukaryotes compared to prokaryotes (Griffiths *et al*., 2000; Mathews *et al*., 2000). Regulatory elements are often present in regions where complex protein interaction occurs. Genes with common *cis*-acting elements in the 5′ upstream region suggest similar expression pattern.

2.5 Methods

2.5.1 Red algae genome search from NCBI Database

The National Centre for Biotechnology Information (NCBI) internet database (http://www.ncbi.nlm.gov) was used to search for available red algae genomes. The keywords "red algae" were typed in the NCBI search box and BioProject was selected from the Entrez pull-down menu, to find sequenced red algae genomes, including those in progress.

2.5.2 Search for dehydrogenase-like genes from the selected red algae genomes

The keywords dehydrogenase AND "red algae" were used to retrieve all the dehydrogenase-like genes from red algae genomes in the NCBI database. All the dehydrogenase-like genes from the selected red algae genomes were analysed. The genes that encode alcohol dehydrogenase-like and lactate dehydrogenase-like proteins were then selected from the red algae genomes and used for further Bioinformatics analysis.

The red alga, *Chondrus crispus*, genome was selected as a model organism for the study and genes encoding alcohol and lactate dehydrogenase-like proteins from this genome were obtained from the NCBI database. The keywords "lactate dehydrogenase" OR "alcohol dehydrogenase" AND "*Chondrus crispus*" were used in the search box and "All databses" in NCBI was selected for the search. The genes tab in the resultant results window was selected to provide the list of all lactate/alcohol dehydrogenase genes from the *Chondrus crispus* genome.

Each Gene ID entry was individually selected and in the resultant window, the Protein tab on the right side of the results window was selected to provide a list of the encoded proteins. The protein sequences encoded by the identified alcohol and lactate dehydrogenase coding regions were used in a BLAST search to retrieve similar sequences in the GenBank non-redundancy (nr) protein database. All the protein sequences with a low E value and a high score (>200) were selected and saved in a FASTA format in a text file. The default settings were used for all the parameters in the BLAST search.

2.5.3 Sequence Alignments

The LDH and ADH protein sequences obtained from the above BLAST searches were used in the Clustal Omega tool (Larkin *et al*., 2007), to generate a multisequence alignment of the known LDHs and ADHs from different organisms. The default settings were used for all the sequence alignments.

2.5.4 Phylogenetic tree analysis

Phylogenetic and molecular evolutionary analyses for all known LDHs and ADHs proteins from other organisms were conducted using MEGA version 6 (Tamura *et* al., 2013), using the bootstrapping method and the nodes of bootstrap replications were set to 1000.

2.5.5 Properties, signal peptide and subcellular localization predictions

The ExPASy tools such ProtParam were used to predict the molecular weight and theoretical isoelectric point (pI) of LDHs and ADHs proteins using the website (http://www.expasy.org?tools/protparam.html). The signal peptide was predicted by SignalP web server (http://www.cbs.dtu.dk/services/SignalP/) and TargetP (http://www.cbs.dtu.dk/services/TargetP/) (Emanuelsson *et al*., 2007). The subcellular localization predictions were performed by the WoLF PSORT (http://genscript.com/wolf-psort.html) and LOCTREE 3 (Goldberg *et al*., 2012; 2014). The default settings were used in all the predictions.

2.5.6 Domains, families and functional sites

The InterProScan tool (http://www.ebi.ac.uk/tools/pfa/iprscan) was used to identify the presence of the functional domains for LDHs and ADHs proteins. The NCBI's Conserved Domain Database (CDD) (http://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi) and the Pfam database (http://pfam.xfam.org) was used to check the domain organization in the LDH and ADH protein sequences, and PROSITE (http://prosite.expasy.org) was used to check the signature patterns for the LDH and ADH proteins to confirm the family to which they belong.

2.5.7 Secondary structure analysis

The secondary structures of LDH and ADH proteins were predicted by the Phyre2 server (http://www.sbg.bio.ic.ac.uk/phyre2) (Kelley *et al*., 2015) and the default settings were used. The PROMALS3D web server (http://prodata.swmed.edu/promals3d/promals3d.php) (Pei *et al*., 2008; Pei and Grishin, 2014) was used to identify homologs with the known 3D structure for the input LDH and ADH protein sequences to construct consistency-based multiple sequence alignments. The default settings were used to perform multiple sequence alignments with local structure and 3D structures.

2.5.8 Tertiary structure analysis

The tertiary structure of LDHs and ADHs were predicted by using PRIMO (http://primo.rubi.ru.ac.za/) (Hatherley *et al*., 2016) and SWISS-MODEL (http://swissmodel.expasy.org/) (Arnold *et al*., 2006). The template identification and target template sequence alignment were done by PRIMO with only minor manual edits to the alignment. After the modelling of LDHs and ADHs was completed the models were evaluated by DOPE Zscore and RMSD. The Ramachandran Plot was also used to test the quality of the LDHs and ADHs models. The superposition of LDHs and ADHs models with selected templates were analyzed by the Swiss-PDBViewer v4.1 (Guex *et al*., 2012).

2.5.9 Analysis of putative *cis*-acting regulatory elements

The analysis of the 1500bp 5´ upstream region from the transcriptional start site of the *Chondrus crispus* ADH genes was achieved by scanning for the presence of putative *cis*-acting regulatory elements using plant-specific database PlantCARE (http://bioinformatics.psb.ugent.be/webtools/plantcare/html/) (Rombauts *et al*., 1999; Lescot *et al*., 2002) and the default settings were used.

2.6 Results and Discussion

2.6.1 Red algae genome analysis

The search for red algae species with fully sequenced genomes from NCBI database showed that *Chondrus crispus*, *Cyanidioschyzon merolae* strain 10D, *Galdieria sulphuraria*, *Porphyra umbilicalis*, *Porphyridium purpureum*, *Kappaphycus alvarezii*, *Gracilariopsis lemaneiformis*, *Gracilariopsis chorda*, and *Pyropia yezoensis* have complete genome sequences (**Table 2.1).** The information on these red algae sequenced genomes were obtained using the keywords "red algae" in the NCBI website **(**https://www.ncbi.nlm.nih.gov/assembly/?term=red%20algae**).** However, the pyropia yezoensis genome did not appear on the list of algae genomes when the keywords "red algae" used in the NCBI search box. According to Nakamura et al. (2013) the draft genome assembly of pyropia yezeonsis was determined. The search for this genome was done separately and showed a genome length of 43 Mbp.

The genome analysis in the NCBI database gives insight on genome size variation and showed a relatively small number of predicted genes. This result is similar to the observation showed by Lee et al. (2018) on draft genome analysis of Gracilariopsis chorda. The purpose of this section was to check the available red algae genomes from the NCBI database. The genomic information on red algae is scarce; about 9 red algae species with full genome sequences were obtained in the NCBI database for this study. The genomic information is important for functional studies of individual genes and provides insight into the evolutionary relationships of red algae.

Table 2. 1: Identified and fully sequenced red algae genomes from the NCBI database

Organism	**Class**	**Genome size**	**Assembly level**	**GenBank assembly accession number**	**Reference**	**Gene annotation**
Chondrus crispus	Florideophyceae	105 Mbp	Scaffold	GCA_000350225.2	Collen *et al*., 2013	9606
Cyanidioschyzon merolae Strain 10D	Bangiophyceae	16.7 Mbp	Complete genome	GCA_000091205.1	Matsuzaki *et al*., 2004	5678
Galdieria sulphuraria	Bangiophyceae	13.7 Mbp	Scaffold	GCA_000341285.1	Schonknecht *et al*., 2013	6623
Porphyridium purpureum	Bangiophyceae	19.7 Mbp	Contig	GCA_000397085.1	Bhattacharya *et al*., 2013	8355
Pyropia yezoensis	Bangiophyceae	43 Mbp	Contig	Not found	Nakamura *et al*., 2013	10327
Porphyra umbilicalis	Bangiophyceae	87.7 Mbp	Scaffold	GCA_002049455.2	Brawley *et al*., 2017	13125
Kappaphycus alvarezii	Florideophyceae	349.5 Mbp	Contig	GCA_002205965.2	Liu *et al*., 2017	234
Gracilariopsis chorda	Florideophyceae	92.1 Mbp	Contig	GCA_003194525.1	Lee *et al*., 2018	10938
Gracilariopsis lemaneiformis	Florideophyceae	88.69 Mbp	Scaffold	GCA_003346895.1	Not published	9281

2.6.2 Genome selection and identification dehydrogenases

To our knowledge, the *C. crispus* is one of the only red macroalgae species with a fully sequenced genome and annotated genes. This genomic information encouraged the determination and identification of genes used in this study. Several genes encoding dehydrogenase-like proteins were obtained in the NCBI database from each of the 7 red algae genomes. The NCBI database search for dehydrogenase-like genes showed that *Galdieria sulphuraria* contained 215 dehydrogenase-like genes, followed by *Cyanidioschyzon merolae* with 179 genes, *Chondrus crispus* with 155 genes, *Kappaphycus alvarezii* with 12 genes, *Pyropia yezoensis* and *Porphyra umbilicalis* with 11 genes each, and *Porphyridium purpureum* with 2 genes. The protein sequences encoded by the identified dehydrogenase coding regions were obtained and used in a BLAST search against the GenBank non-redundancy protein database and similar sequences were obtained, indicating that the identified genes may belong to several dehydrogenase families.

Since many dehydrogenase-like genes were obtained from the red algae genomes available in the NCBI database, it was decided to focus only on the alcohol dehydrogenase (ADH) and lactate dehydrogenase (LDH) family. These two enzymes are well-investigated and serve as important enzymes for both scientific research and commercial applications (Hassan *et al*., 2013; Liu *et al*., 2014).

All the LDH and ADH genes in the red algae genomes were collected from the NCBI GenBank database and compared with other similar genes from other organisms. The search for lactate dehydrogenases revealed 11 LDH-like genes, including 2 hypothetical genes from *Chondrus crispus* genome. Six of these 11 LDH-like genes were obtained from the *C. merolae* genome, and 3 LDH-like genes were obtained from *G. sulphuraria* genome. The search for alcohol dehydrogenases showed 12 ADH genes which include 8 genes from *G. sulphuraria* and 2 genes from *C. crispus* and *C. merolae*, each. All the predicted alcohol dehydrogenases and lactate dehydrogenase-like genes were analysed using ExPASy Bioinformatics Resource Portals (www.expasy.org/tools) as shown in **Table 2.2**.

Table 2. 2: The alcohol dehydrogenases and lactate dehydrogenase-like gene family in red algae genomes using the NCBI GenBank database (http://www.ncbi.nlm.nih.gov/gene/). The analysis was done through ExPASy Bioinformatics Resource Portals (www.expasy.org/tools)

Gene name	Aliases/gene ID	Location	Nucleotide length (bp)	Protein accession number	Protein length (aa)	Molecular weight (kDa)	Isoelectric point (pI)
Alcohol dehydrogenase (Zn-containing)	Gasu_04790	179708..181012 complement	1305	XP_005708911	398	43.49	8.85
Alcohol dehydrogenase (Zn-containing)	Gasu_42440	42391..43601	1161	XP_005704762	351	36.81	8.73
Alcohol dehydrogenase class III	CYME_CMS125C	Chromosome 19 (321089.322237 complement)	1149	XP_005538792	382	40.63	6.34
Probable alcohol dehydrogenase	CYME_CMQ301C	Chromosome 17 (777042..778136 complement)	1095	XP_005538212	364	40.28	8.36
Alcohol dehydrogenase (Zn-containing)	Gasu_38040	85311..86566	1,205	XP_005705275	375	40.24	6.75
Alcohol dehydrogenase	Gasu_45570	46931..48081	1151	XP_005704413	341	36.26	8.06
Alcohol dehydrogenase (NADP+)	Gasu_53790	48626..49860 complement	1235	XP_005703564	341	37.46	8.53

Alcohol dehydrogenase	Gasu_57970	47361..48475	1115	XP_005703083	338	36.62	7.56
Alcohol dehydrogenase	CHC_T00009372001	182742..184011 complement	1184	XP_005714659	377	40.89	5.79
Alcohol dehydrogenase	CHC_T00010066001	27096..28641	1426	XP_005711469	377	40.01	6.02
Zn-containing alcohol dehydrogenase	Gasu_23390	176381..178146	1614	XP_005706957	372	41.34	5.74

Alcohol dehydrogenase (Zn-containing)	Gasu_54710	44691..45650 complement	960	XP_005703418	244	25.36	5.95
L-lactate dehydrogenase	CYME_CMK006C	Chromosome 11 (14006..15070 compliment)	1065	XP_005536444	354	37.91	5.84
L-lactate dehydrogenase	CYME_CMI306C	Chromosome 9 (804566..805630)	1065	XP_005536442	354	37.89	5.84

L-lactate dehydrogenase	CYME_CMC188C	Chromosome 3 (476222..477286)	1065	XP_005535386	354	37.92	5.84
L-lactate dehydrogenase	CYME_CMA145C	Chromosome 1 (417033..418097)	1065	XP_005535115	354	37.89	5.84
L-lactate dehydrogenase	CYME_CMJ002C	Chromosome 10 (4509..5339 complement)	831	XP_005534764	276	29.94	6.66
Probable D-lactate dehydrogenase	CYME_CMQ347C	Chromosome 17 (887379..889571 compliment)	2193	XP_005538239	730	79.51	5.87

D-lactate dehydrogenase	Gasu_43520	105761..106996	1097	XP_005704707	340	36.93	6.66
D-lactate dehydrogenase (cytochrome)	Gasu_38460	34691..36641	1849	XP_005705157	563	62.32	6.17

D-lactate dehydrogenase (cytochrome)	Gasu_04450	114636..116590 complement	1893	XP_005708873	504	55.94	5.94
Hypothetical gene	CHC_T00003974001	160412..161585 complement	1174	XP_005715286	338	36.95	6.16
Hypothetical gene	CHC_T00003490001	125045..126088	1044	XP_005714861	347	36.98	9.02

The ADH and LDH-like genes from *C. crispus* were then selected and used for further study. The *C. crispus* was used as a model organism for this study based on the interest in macroalgal studies and the availability of relevant information about genes in the literature made this alga a suitable choice for this study (Collen *et al*., 2006, 2007, 2013; Kowalczyk *et al*., 2014; Gao *et al*., 2016). The *C. crispus* genome has useful scientific research background information, and it has been used as a model species for red algae (Kowalczyk *et al*., 2014).

In *C. crispus*, 2 alcohol dehydrogenases and 2 hypothetical genes were identified. These alcohol dehydrogenase proteins encoded by these genes, are referred to in this study as the cinnamyl alcohol dehydrogenase (*CAD1*) and the class III alcohol dehydrogenase (*ADH3*), because of the presence of conserved putative domains with cinnamyl alcohol dehydrogenases and class III alcohol dehydrogenase, respectively. The proteins encoded by the two hypothetical genes were identified as D-lactate dehydrogenase-like (D-LDH-like) proteins, because of the conserved NAD and catalytic binding domains identified in these sequences during BLASTp search against the GenBank non-redundancy protein database. The BLAST search for ADH and LDH-like proteins showed low protein identity among CAD1 coding region. The ADH3 coding region in several other organisms showed more than 50% protein identity among the class III alcohol dehydrogenases from other species.

The protein BLASTp search, using the hypothetical protein sequences as a query, showed protein identity with D-lactate dehydrogenases, D-isomer specific and 2-hydroxyacid dehydrogenases, formate and D-3-phosphoglycerate dehydrogenases from other species (**See Appendix A**). The coding nucleotide sequence lengths for alcohol dehydrogenases (ADH) were 1184, and 1426 bp and that of the hypothetical genes (LDH-like) were 1044 and 1174 bp, respectively. The proteins encoded by these genes have a predicted molecular weight ranging from 36.95 to 40.97 kDa and isoelectric points of 5.79 to 9.02, respectively. The information of ADH and LDH-like gene family in *C. crispus* is summarised in **Table 2.3**.

Table 2. 3: The alcohol dehydrogenases and hypothetical genes encoding LDH-like proteins from NCBI database

Gene name	Gene symbol	Amino acid	Protein accession number	mRNA sequence length	Nucleotide accession number	Molecular weight	Location
Alcohol dehydrogenase	CHC_T00009372001	377 aa	XP_005714659	1184 bp	XM_005714602	40.89 kDa	Scaffold_18
Alcohol dehydrogenase	CHC_T00010066001	377 aa	XP_005711469	1426 bp	XM_005711412	40.01 kDa	Scaffold_82
Hypothetical gene	CHC_T00003490001	347 aa	XP_005714861	1044 bp	XM_005714804	36.98 kDa	Scaffold_191
Hypothetical gene	CHC_T00003974001	338 aa	XP_005715286	1174 bp	XM_005715229	36.95 kDa	Scaffold_208

2.6.3 Domains, families and functional and structural sites in dehydrogenases

Several protein databases were used to check for the presence of functional domains in the *C. crispus* ADH and hypothetical LDH proteins. The alcohol dehydrogenase (XP_005711469) showed putative conserved domains for class III alcohol dehydrogenases. These proteins are identical to members of the glutathione-dependent formaldehyde dehydrogenases (FDH) and belong to the MDRs protein superfamily as shown in the NCBI Conserved Domain Database http://www.ncbi.nlm.nih.gov/Structure/ccd/cddsrv.cgi).

The *C. crispus* alcohol dehydrogenase and glutathione-dependent formaldehyde dehydrogenase genes reported in the literature (Dolferus *et al*., 1997; Jin *et al*., 2016) have the same functional domains (**Appendix A**). The analysis of NCBI Conserved Domain and InterPro (www.ebi.ac.uk/interpro/) databases confirmed that this protein is a member of the zinc dependent/ medium-chain alcohol dehydrogenases.

The other ADH protein (XP_005714659) was named as *C. crispus* cinnamyl alcohol dehydrogenase (*CcCAD1*) based on the functional domains similar to other plant CAD proteins. The cinnamyl alcohol dehydrogenases are also members of the Medium-Chain Dehydrogenases/Reductases (MDRs) superfamily, and they reduce cinnamyl aldehydes to cinnamyl alcohols in the last step of monolignol metabolism in plant cell walls. In non-plant species such as the yeast, CAD proteins are involved in the aldehyde reductase activity. The NCBI CDD search showed that *CcCAD1* protein also has a specific hit for D-arabinose 1-dehydrogenase.

The proteins in this group are zinc-dependent alcohol dehydrogenases that are involved in the carbohydrate transport and metabolism. The InterPro database analysis revealed that the N-terminal (IPRO13154) and C-terminal (IPRO13149) regions in alcohol dehydrogenase proteins are catalytic and zinc-binding domains, respectively. These conserved domain regions contain residues essential for catalytic activity in many alcohol dehydrogenases (Murzin, 1996; Taneja and Mande, 1999).

The hypothetical protein (XP_005715286) was named as *C. crispus* D-lactate dehydrogenase (*CcD-LDH*) based on the presence of the same functional domain with D-lactate and related dehydrogenases. This domain had a sequence length of 335 amino acid residues.

The NCBI Conserved Domain Database showed that this protein catalyses the conversion of pyruvate and lactate, and is a member of the 2-hydroxyacid dehydrogenase family. The other hypothetical protein (XP_005714861) was named as *CcD-LDH-like* due to the specific BLAST hits related to D-isomer specific 2-hydroxyacid dehydrogenase and NAD-binding domains. Most members of D-isomer specific 2-hydroxyacid dehydrogenase family are NAD-dependent proteins and the present of these functional domains suggest that the hypothetical protein in *C.crispus* may belong to this family.

The InterPro database also showed that the catalytic domain is larger in most D-isomer specific 2-hydroxyacid dehydrogenases and has been reported to have a number of conserved charged residues which may be involved in the catalytic mechanism (Dengler *et al*., 1997). The NAD-binding domain contains a glycine-rich region that is located in the central section of many NAD- dependent D-isomer specific 2-hydroxyacid dehydrogenases (Dengler *et al*., 1997).

The NAD-binding domain is inserted into the catalytic domain, and these domains which are located in the C-terminal and N-terminal regions, respectively of many NAD-dependent 2hydroxyacid dehydrogenases has been shown to be functionally and structurally similar (Dengler *et al*., 1997). In members of the 2-hydroxyacid dehydrogenase, the N-terminal (IPR006139) and C-terminal (IPR006140) regions are located in the catalytic and NAD-binding domains, respectively.

Table 2. 4:Analysis of the InterPro database results of the *CcADH* and *CcD-LDH* showing conserved domains, protein family and gene ontology term prediction (See Appendix C for the full result of the InterPro database analysis)

Gene	Family	N-terminal domain position	C-terminal domain position	Biological process	Molecular function
CcADH3	glutathione-dependent formaldehyde dehydrogenase	33-157	202-333	Ethanol oxidation Oxidationreduction process	Zinc-ion binding Oxidoreductase activity, S(hydroxymethyl) glutathione dehydrogenase activity.
CcCAD1	cinnamyl alcohol dehydrogenase	48-166	209-334	Oxidationreduction process	Oxidoreductase activity
CcD-LDH	2-hydroxyacid dehydrogenase	5-330	114-301	Metabolic process Oxidationreduction process	Oxidoreductase activity, acting on the CH-OH group of donors, NAD or NADP as acceptor NAD binding
CcD-LDH-like	2-hydroxyacid dehydrogenase	19-342	128-313	Metabolic process Oxidationreduction process	Oxidoreductase activity, acting on the CH-OH group of donors, NAD or NADP as acceptor NAD binding

Members of the D-isomer specific 2-hydroxyacid dehydrogenase family have been shown to be functionally and structurally related (Taguchi and Ohta, 1991). The SIB Bioinformatics Resource Portal (http://prosite.expasy.org) was also used to check the protein domains, family and functional sites in ADH and LDH proteins. The signature patterns for the *CcD-LDH* and *CcD-LDH-like* proteins showed a consensus pattern that corresponds to the NAD-binding and the catalytic domains. The ScanProsite results obtained from the entry PS00670 for both D-lactate dehydrogenase-like (*CcD-LDH* and *CcD-LDH-like*) proteins revealed a consensus pattern [LF(X)2SDII(X)LH(X)P(X)4T(X)2M(X)D] and [LL(X)2SDII(X)LH(X)P(X)4T(X)2M(X)G] which corresponds to the NAD-binding domain, respectively. The signature patterns for *CcD-LDH* [MK(X)G(X)ILIN(X)SRG(X)LMD] and *CcD-LDH-like* [CK(X)G(X)LLIN(X)SRG(X)LLD] were obtained from the PROSITE entry PS00671, which corresponds to the catalytic domain. The catalytic domain has essential amino acid residues involved in the catalytic mechanism of many NAD-dependent 2-hydroxyacid dehydrogenases. The signature patterns for *CcADH3* protein showed a zinc-containing alcohol dehydrogenase signature pattern [GHE(X)AGV(X)4G(X)2V] which corresponds to the highly conserved zinc-binding signature in most MDR proteins (Jin *et al*., 2016). There were no hits found for all PROSITE motifs on the ADH (*CcCAD1*) amino acid sequence.

2.6.4 Sequence alignment

The sequence alignments of the two alcohol dehydrogenase proteins in *C. crispus* showed 23% identity similarity with each other. The alcohol dehydrogenases of the MDRs superfamily have been reported in the literature to have low sequence identity at the protein level (Knoll and Pleiss, 2008), supporting our observation in this study. The low sequence identity similarity between these two ADH proteins may indicate that these proteins are members of two different families within the MDRs superfamily of proteins. When these two proteins were used in sequence alignments with other known ADH proteins from other species they showed varied homology among the cinnamyl alcohol and class III alcohol dehydrogenases, also known as glutathione-dependent formaldehyde dehydrogenases ranging from 22 to 68% sequence identity similarity (**See Appendix D**).

The *CcCAD1* amino acid sequence when used as a query to find regions of similarity with known or putative cinnamyl alcohol dehydrogenases from other species including plants showed low

sequence identity similarity (about 36%) in other plant CAD proteins. This low similarity identity among these amino acid sequences was by no means outside the range of some of the similar proteins reported in the literature (Bukh *et al*., 2012; Jin *et al*., 2014). The *CcADH3* was used as a query sequence with other known or putative ADH sequences from other species and revealed identity ranging between 62-66%. These proteins are known as glutathione-dependent formaldehyde dehydrogenases and showed highly conserved regions among plant species which have been reported in the literature (Dolferus *et al*., 1997; Jin *et al*., 2016). In this study, the *CcADH3* showed high similarity with zebrafish (*Danio rerio*) and soil-living amoeba (*Dictyostelium discoideum* AX4) with 60% sequence identity similarity. The CAD and ADH3 amino acid sequences used in the multiple sequence alignments showed sequence identity ranging from 31 to 66%, as shown in **Table 2.5**.

Table 2. 5: Percentage protein identity obtained from the Clustal Omega multiple sequence alignment of *C. crispus* class III alcohol dehydrogenase and cinnamyl alcohol dehydrogenases with other species.

Plant CAD Proteins	*CcCAD1*	ADH3 Proteins	*CcADH3*
AtCAD	33	GsADH	64
AtCAD1	36	CmADH3	65
AtCAD4	32	PuADH-like	64
SbCAD4	34	AtADH3	62
ZmCAD1	31	DrADH3	66
TaCAD1	31	HsADH3	65
FaCAD1	31	SoADH	64
NtCAD1	31	DdADH	66
LpCAD1	35	DmADH	64
PtCAD1	31		
AcCAD1	32		

The analysis of CAD proteins among higher plants has shown that some CAD proteins have high sequence identity similarity among each other (Ma, 2010). The multiple sequence alignments revealed that the CAD genes showed lower sequence identity at the protein level with each other. Several plant CAD sequences were used in Clustal omega multiple sequence alignments with *C. crispus* CAD sequence and showed conserved regions (**Figure 2.1**). The CAD proteins in MDRs superfamily revealed highly conserved catalytic Zn binding amino acid residues which consist of two highly conserved cysteine residues and a histidine.

These amino acid residues in *C. crispus* CAD protein sequence were located at Cys-58, His-80 and Cys-183. The structural zinc binding site consisted of four identical cysteine residues at position 111, 114, 117 and 125 in the *C. crispus* protein. The substrate binding motifs in the proteins used in the alignment showed highly conserved residues including Cys-58, Thr-60, His80, and Cys-183. The conserved binding motif [GXG(X)2G] (residues Gly-206 to Gly-212) in the NADPH-binding domain of the CAD sequences, used in the alignments, showed that these proteins are members of the MDRs and may be zinc-dependent alcohol dehydrogenases as reported in literature (Rossman *et al*., 1974; Knoll and Pleiss, 2008).


```
XP_005714659CcCAD1      LCHTCDIHMRDNDWGISNYPLLAGHEGVAHYTRVGSAVRNIRPDDRIAITWIRDSCRACDS  116
AAP40269AtCAD1          VCYADVIWSRNQHGDSKYPLVPGHEIAGIVTKVGPNVQRFKVGDHVGVGTYVNSCRECEY  106
CAA06687ZmCAD1          ICHTCDIHQAKNHLGASKYPMVPGHEVVGEVVEVGPEVAKYGVGDVVGVGVIVGCCRECSP  105
ADI59734TaCAD1          ICHTCDVHQVKNDLGASKYPMVPGHEVVGEVVEVGPEVSKFRAGDVVGVGVIVGCCRDCRP  106
AAK97808FaCAD1          ICHTCDLHQTKNHLGASKYPMVPGHEVVGEVVEVGPEVSKYSVGDVVGVGVIVGCCRDCRP  106
AAP59434AtCAD4          ICHTCDIHQIKNDLGMSNYPMVPGHEVVGEVLEVGSDVSKFTVGDVVGVGVVVGCCGSCKP  106
CAA44216NtCAD1          LCHTCDLHQVKNDLGMSNYPLVPGHEVVGEVVEVGPDVSKFKVGDTVGVGLLVGSCRNCGP  105
BAA03099AcCAD1          ICHTCDIHQIKNDLGASNYPMVPGHEVVGEVVEVGSDVTKFKVGDCVGDGTIVGCCKTCRP  105
AAF43140PtCAD1          ICHTCDIHQIKNDLGMSHYPMVPGHEVVGEVVEVGSDVTKFKAGDVVGVGVIVGSCKNCHP  105
NP_179765AtCAD          VCHSDLHTIKNHWGFSRYPIIPGHEIVGIATKVGKNVTKFKEGDRVGVGVIIGSCQSCES  102
XP_002462348SbCAD4      ICHTCDLHVIKNEWGNAMYPVVPGHEVVGVVTDVGHGVTKFKAGDTVGVGYFVDSCRTCES  105
AAL99535LpCAD1          ICHTCDLHIIKNDWGNALYPIVPGHEIVGVVASVGSGVSSFKAGDTVGVGYFLDSCRTCYS  102
                        :*::*:    *. * : **:: *** .. .  **  *     .* :.     ..*  *

XP_005714659CcCAD1      CLAGRENICEDSY---QGTYLGDSAGAWGSSKFHYNENGGCFSKIQRIEERFAIKIPDGV  173
AAP40269AtCAD1          CNEGQEVNCAKGVFTFNGIDH------------DGSVTKGGYSSHIVVHERYCYKIPVDY  154
CAA06687ZmCAD1          CKANVEQYCNKKIWSYNDVYT------------DGRPTQGGFASTMVVDQKFVVKIPAGL  153
ADI59734TaCAD1          CKANVEQYCNKKIWSYNDVYT------------DGKPTQGGFASAMVVDQKFVVKIPAGL  154
AAK97808FaCAD1          CKANVEQYCNKKIWSYNDVYT------------DGKPTQGGFASAMVVDQKFVVKIPAGL  154
AAP59434AtCAD4          CSSELEQYCNKRIWSYNDVYT------------DGKPTQGGFADTMIVNQKFVVKIPEGM  154
CAA44216NtCAD1          CKRDIEQYCNKKIWNCNDVYT------------DGKPTQGGFAKSMVVDQKFVVKIPEGM  153
```

```
BAA03099AcCAD1          CKADVEQYCNKKIWSYNDVYT------------DGKPTQGGFSGHMVVDQKFVVKIPDGM  153
AAF43140PtCAD1          CKSELEQYCNKKIWSYNDVYT------------DGKPTQGGFAESMVVDQKFVVRIPDGM  153
NP_179765AtCAD          CNQDLENYCPKVVFTYNSRSS-----------DGTSRNQGGYSDVIVVDHRFVLSIPDGL  151
XP_002462348SbCAD4      CSTGHENYCPDLVLTSNGVDHH----------HHGATTKGGFSDVLVVSQDFVVRVPESL  155
AAL99535LpCAD1          CSKGYENFCPTLTLTSNGVDG------------GGATTQGGFSDVLVVNKDYVIRVPDNL  150
                        *    *  *       :.                  . * ::    : . :   :*
XP_005714659CcCAD1      PSEIACPLLCGGGTVYEPICDYA---GPNVKVGVAGVGGLGTAAIKLAKLRGCIITAFST  230
AAP40269AtCAD1          PLESAAPLLCAGITVYAPMMRHNMNQ-PGKSLGVIGLGGLGHMAVKFGKAFGLSVTVFST  213
CAA06687ZmCAD1          APEQAAPLLCAGVTVYSPLKHFGLTT-PGLRGGILGLGGVGHMGVKVAKAMGHHVTVISS  212
ADI59734TaCAD1          APEQAAPLLCAGVTVYSPLKHFGLMT-PGLRGGILGLGGVGHMGVKVAKSMGHHVTVISS  213
AAK97808FaCAD1          APEQAAPLLCAGVTVYSPLKHFGLMT-PGLRGGILGLGGVGHMGVKVAKSMGHHVTVISS  213
AAP59434AtCAD4          AVEQAAPLLCAGVTVYSPLSHFGLMA-SGLKGGILGLGGVGHMGVKIAKAMGHHVTVISS  213
CAA44216NtCAD1          APEQAAPLLCAGITVYSPLNHFGFKQ-SGLRGGILGLGGVGHMGVKIAKAMGHHVTVISS  212
BAA03099AcCAD1          APEQAAPLLCAGVTVYSPLTHFGLKEISGLRGGILGLGGVGHMGVKLAKAMGHHVTVISS  213
AAF43140PtCAD1          SPEQAAPLLCAGLTVYSPLKHFGLKQ-SGLRGGILGLGGVGHMGVKIAKAMGHHVTVISS  212
NP_179765AtCAD          PSDSGAPLLCAGITVYSPMKYYGMTKESGKRLGVNGLGGLGHIAVKIGKAFGLRVTVISR  211
XP_002462348SbCAD4      PLDGAAPLLCAGVTVYSPMAQYALN-EPGKHLGVVGLGGLGHMAVKFAKAFGMTVTVISS  214
AAL99535LpCAD1          PLAGAAPLLCAGVTVYSPMVEYGLN-APGKHLGVVGLGGLGHVAVKFGKAFGMTVTVISS  209
                            ..****.* *** *:  .      .   *: *:**:*  .:*..*  *  :*.:*

XP_005714659CcCAD1      SPHKKEAALG-AGAFEFVNMSDQAQVDAHGGCLDVLLDTTPVESNIDKYMALLKINGTYV  289
AAP40269AtCAD1          SISKKEEALNLLGAENFVISSDHDQMKALEKSLDFLVDTASGDHAFDPYMSLLKIAGTYV  273
CAA06687ZmCAD1          SSKKRAEAMDHLGADAYLVSSDAAAMGPAADSLDYIIDTVPVHHPLEPYLALLKLDGKLV  272
ADI59734TaCAD1          SNKKRAEAMDDLGADAYLVSSDTDQMAAAADSLDYIIDTVPAKHPLEPYLALLKMDGKLV  273
AAK97808FaCAD1          SNKKRAEAMDDLGADAYLVSSDEAQMAAAMDSLDYIIDTVPVKHPLEPYLALLKMDGKLV  273
AAP59434AtCAD4          SDKKKEEAIEHLGADDYVVSSDPAEMQRLADSLDYIIDTVPVFHPLDPYLACLKLDGKLI  273
CAA44216NtCAD1          SNKKRQEALEHLGADDYLVSSDTDKMQEASDSLDYIIDTVPVGHPLEPYLSLLKIDGKLI  272
BAA03099AcCAD1          SDKKKEEAIDHLGADAYLVSSDATQMQEAADSLDYIIDTVPVFHPLEPYLSLLKLDGKLI  273
AAF43140PtCAD1          SDKKREEAMEHLGADEYLVSSDVESMQKAADQLDYIIDTVPVVHPLEPYLSLLKLDGKLI  272
NP_179765AtCAD          SSEKEREAIDRLGADSFLVTTDSQKMKEAVGTMDFIIDTVSAEHALLPLFSLLKVNGKLV  271
XP_002462348SbCAD4      SPGKRDEALGRLGADAFLVSHDAAQMKAAAATLDGIIDTVSAGHQIVPLLALLKPMGQMV  274
AAL99535LpCAD1          SDRKRDEALGRLGADAFLVSSDPEQMKAAAGTMDGIIDTVSAGHPIVPLLDLLKPMGQMV  269

                        *  *.  *:   **  ::   *   :      :* ::**.    :   :  **  *  :
```

```
XP_005714659CcCAD1      KIGIPAMSQQTFSYNFSSLIFQQKKIVGTVVTGTRRMKDMLDLVAKNIDFMKDTDAWKTE  349
AAP40269AtCAD1          LVGFPSEIK-IS---PANLNLGMRMLAGSVTGGTKITQQMLDFCAAHK------IYPNIE  323
CAA06687ZmCAD1          LLGVIGEPLSFV---SPMVMLGRKAITGSFIGSIDETAEVLQFCVDKG------LTSQIE  323
ADI59734TaCAD1          LMGVIAEPLSFV---SPMVMLGRKTITGSFIGSMDETEEVLQFCVDKG------LTSQIE  324
AAK97808FaCAD1          LMGVIAEPLSFV---SPMVMLGRKTITGSFIGSIEETEEVLRFCVEKG------LTSQIE  324
AAP59434AtCAD4          LMGVINTPLQFV---TPLVILGRKVISGSFIGSIKETEEVLAFCKEKG------LTSTIE  324
CAA44216NtCAD1          LMGVINTPLQFI---SPMVMLGRKSITGSFIGSMKETEEMLDFCKEKG------VTSQIE  323
BAA03099AcCAD1          LMGVINTPLQFI---SPMVMLGRKAITGSFIGSMKETEEMLDFCNEKG------ITSTIE  324
AAF43140PtCAD1          LMGVINTPLQFV---SPMVMLGRKSITGSFIGSMKETEEMLEFCKEKG------LASMIE  323
NP_179765AtCAD          ALGLPEKPLDLP---IFSLVLGRKMVGGSQIGGMKETQEMLEFCAKHK------IVSDIE  322
XP_002462348SbCAD4      VVGAPSTPLELP---AYAIITGGKRVAGNGVGSVADCQAMLDFAGEHG------VTADIE  325
AAL99535LpCAD1          VVGAPSKPLELP---AFAIIGGGKRLAGSGTGSVAHCQAMLDFAGKHG------ITADVE  320
                         :*              :    : : *.   .      :* :   :           *
```

Figure 2. 1: Multiple sequence alignment of CcCAD1 with several validated and putative plants CAD-like proteins, using the Clustal Omega tool.

The GenBank accession numbers for each CAD protein was marked by the turquoise background colour. The conserved important amino acid residues were marked as follows: the red rectangular box showed the catalytic zinc binding amino acid residues. The conserved amino acid residues in the NAD(P) binding site were marked using a red font. The green marked amino acid residues showed putative substrate binding site residues. The structural Zn binding residues were denoted by the yellow marked amino acid residues. The ADH proteins showed a highly conserved region for the substrate, catalytic zinc, structural zinc and NADPH-binding amino acid residues.

The amino acid residues in the catalytic Zn binding and structural binding sites of CAD proteins used in multiple sequence alignment are depicted in **Figure 2.1**. The amino acid residues involved in the Zn binding sites were marked with capital letters and identical amino acids among the organisms used in **Table 2.6** were marked. These amino acid residues were identical for all the sequences, and the difference was the only position located in each organism.

Table 2. 6: The amino acid residues involved in catalytic Zn and a structural binding site for all the CAD proteins used in the sequence alignment.

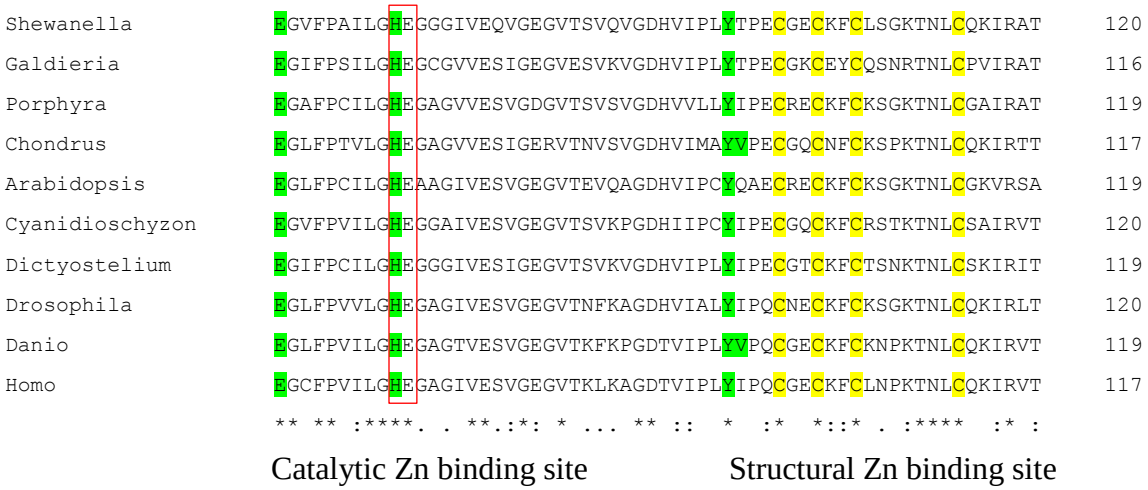
Organism	**Catalytic Zn binding site**			**Structural Zn binding site**			
	C	**H**	**C**	**C**	**C**	**C**	**C**
Chondrus crispus	√	√	√	√	√	√	√
Arabidopsis thaliana 1	√	√	√	√	√	√	√
Zea mays	√	√	√	√	√	√	√
Sorghum bicolor	√	√	√	√	√	√	√
Tristicum aestivum	√	√	√	√	√	√	√
Aralia cordata	√	√	√	√	√	√	√
Festuca arundinacea	√	√	√	√	√	√	√
Populous tremuloides	√	√	√	√	√	√	√
Lolium perenne	√	√	√	√	√	√	√
Arabidopsis thaliana 4	√	√	√	√	√	√	√
Nicotiana tabacum	√	√	√	√	√	√	√
Arabidopsis thaliana homolog 2	√	√	√	√	√	√	√

In **Table 2.7**, the amino acids that are known to be involved in the substrate and NAD binding sites were indicated by capital letters. The amino acid residues that were not marked with a tick showed that they have a different amino acid compared to that of *CcCAD* sequence.

Table 2. 7: The amino acid residues in the substrate and NAD-binding site of all the CAD sequences used in the sequence alignment (See Figure 2.1)

Organism	**Substrate binding site**						**Putative NAD(P)-binding site**																						
	C	**T**	**H**	**W**	**C**	**V**	**C**	**H**	**T**	**H**	**C**	**T**	**G**	**V**	**G**	**G**	**L**	**G**	**S**	**T**	**K**	**M**	**T**	**T**	**V**	**I**	**G**	**T**	**V**	**V**
C. crispus	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√
A. thaliana 1	√	A	√	T	√	T	√	Y	A	I	√	√	√	L	√	√	√	√	√	√	√	S	√	A	G	V	√	S	√	T
Z. mays	√	√	√	V	√	I	√	√	√	√	√	√	√	L	√	√	V	√	√	S	√	S	√	V	√	L	√	S	F	I
S. bicolor	√	√	√		√		√	√	√	√	√	√	√	L	√	√	L	√	√	S	√	S	√	V	A	V	√	N	G	√
T. aestivum	√	√	√		√		√	√	√	√	√	√	√	L	√	√	V	√	√	S	√	S	√	V	A	M	√	S	F	I
A. cordata	√	√	√		√		√	√	√	√	√	√	√	L	√	√	V	√	√	S	√	S	√	V	√	M	√	S	F	I
F. arundinacea	√	√	√		√		√	√	√	√	√	√	√	L	√	√	V	√	√	S	√	S	√	V	√	M	√	S	F	I
P. tremuloides	√	√	√		√		√	√	√	√	√	√	√	L	√	√	V	√	√	S	√	S	√	V	√	M	√	S	F	I
L. perenne	√	√	√		√		√	√	√	√	√	√	√	L	√	√	L	√	√	S	√	S	√	V	A	V	√	S	G	T
A. thaliana 4	√	√	√		√		√	√	√	√	√	√	√	L	√	√	V	√	√	S	√	S	√	V	√	M	√	S	F	I
N. tabacum	√	√	√		√		√	√	√	√	√	√	√	L	√	√	V	√	√	S	√	S	√	V	√	M	√	S	F	I
A. thaliana	√	S	√		√		√	√	S	√	√	√	√	L	√	√	L	√	√	R	√	T	√	V	A	L	√	S	Q	I

The Clustal omega multiple sequence alignment of *CcADH3* sequence with related sequences from other species (**Figure 2.2**) showed highly conserved amino acid residues among the class III alcohol dehydrogenases. These highly conserved regions include the catalytic Zn amino acid residues Cys-45, His-67, Glu-68 and Cys-174. The four cysteine amino acid residues 97, 100, 103 and 111 showed in *CcADH3* sequence have a role in the structural stability of zinc-dependent alcohol dehydrogenases (Sanghani *et al*., 2002; Auld and Bergman, 2008; Eklund and Ramaswamy, 2008), and were conserved among all the species used in the sequence alignment. The class III alcohol dehydrogenases used in the alignment also showed the highly conserved amino acid residues in the substrate binding site, suggesting that the members of this group have a particular substrate for activity. The amino acid residues in the NADPH-binding site were found to be highly conserved, suggesting that these proteins have a similar function among the MDRs superfamily.



```
Shewanella          QGKGLM-PDGTTRF--YKDGQPIFHYMGCSTFSEYTVLPEISLAKVNKTAPLKEICLLGC    177
Galdieria           QGKGYM-PDGTVRF--HCKGKDIHHYMGTSTFSEYTVLPEISVAKVDPSVPFERACLFGC    173
Porphyra            QGKGVM-PDGTSRLTATCSDRGLLHFMGTSTFCEYAVLPAIAVATVDAAAPLTSVCLLGC    178
Chondrus            QGAGVM-PDGTGRF--TCKGKDLYHYMGTSTFAEYTVVADISVVRIREDAPLDKVCLMAC    174
Arabidopsis         TGVGIMMNDRKSRF--SVNGKPIYHFMGTSTFSQYTVVHDVSVAKIDPTAPLDKVCLLGC    177
Cyanidioschyzon     QGQGLM-PDRTTRY--SCNGRSLFHYMGCSCFSQYIVLPEIAVAKIRQDAPLDRVCLLGC    177
Dictyostelium       QGKGQM-PDGTTRF--KCKGKEIFHFMGTSTFSQYTVLPEISCCVVREDAPLDKVCLLGC    176
Drosophila          QGAGVM-PEGTSRL--SCKGQQLFHFMGTSTFAEYTVVADISLTKINEKAPLEKVCLLGC    177
Danio               QGQGLM-PDNTSRF--TCKGKQLFHFMGTSTFSEYTVVAEISLAKVDEHAPLDKVCLLGC    176
Homo                QGKGLM-PDGTSRF--TCKGKTILHYMGTSTFSEYTVVADISVAKIDPLAPLDKVCLLGC    174
                     * * *  : . *     ..: : *:** * *.:* *:  ::   :   .*:   **:.*

Shewanella          GVTTGMGAVMNTAKVEAGATVAIFGLGGIGLSAIIGATMAKASRIIAIDINESKFELARK    237
Galdieria           GVTTGIGAVLNTCKVERGKTVGVFGLGAVGLSAIQGARLAGASRIVAIDINEGKFELAKK    233
Porphyra            GIPTGVGAVRNTCKVEEGATVAVFGLGGVGLSVVQGAVLAGASRIIGVDIDAGKFANATK    238
Chondrus            GVTTGIGAVRNNCKVEPGSTVAVFGLGGVGLSAIQGAKMAGAKAIYGIDTNPEKFAMAET    234
Arabidopsis         GVPTGLGAVWNTAKVEPGSNVAIFGLGTVGLAVAEGAKTAGASRIIGIDIDSKKYETAKK    237
Cyanidioschyzon     GITTGIGAVLNTAKVEQGSTVAVFGLGGVGLSVVQGARIAGASRIIGVDTNESKFPLAKQ    237
Dictyostelium       GITTGFGAAKITAKVEEGSTVAIFGLGAVGLSVAQGAVDCGAKRIIGIDNNETKFGPGKD    236
Drosophila          GISTGYGAALNTAKVEAGSTCAVWGLGAVGLAVGLGCKKAGAGKIYGIDINPDKFELAKK    237
Danio               GISTGYGAAINTAKVEAGSTCAVFGLGAVGLAVVMGCKSAGATRIIGIDVNPDKFEIAKK    236
Homo                GISTGYGAAVNTAKLEPGSVCAVFGLGGVGLAVIMGCKVAGASRIIGVDINKDKFARAKE    234
                    *: ** **.  ..*:* *   .::*** :***.  *.  . *  * .:* :  *:  .

                                NAD binding site

Shewanella          LGATDCINPKN---FDKPIQEVIVEMTDGGVDYSFECIGNVNVMRSALECCHKGWGESVI    294
Galdieria           MGATDCVNPNK---YDKPIQQVLIEMTNGGLDYTFEAVGNVKLMRAALEACHRGWGESCI    290
Porphyra            FGATECVNPKDY--GDKPIQEVLVEMTDGGLDYTFEAIGRTATMRSALEAAHKGWGMSCI    296
Chondrus            LGATVCVNPKD---HSKPIQQVLVELTGGGFDYTFECIGKVETMRAALESCHKGWGESCI    291
Arabidopsis         FGVNEFVNPKD---HDKPIQEVIVDLTDGGVDYSFECIGNVSVMRAALECCHKGWGTSVI    294
Cyanidioschyzon     LGATECINPLKF--GEKPIQQVLIDMTDGGPDYTFEAIGNVKTMRAALEASHKGWGVSVI    295
Dictyostelium       FGCTEFINPSKDLPEGKTIQQHLVDITDGGVDYSFECIGNVNVMRAALECCHKGWGVSTI    296
Drosophila          FGFTDFVNPK-DVADKGSIQNYLIDLTDGGFDYTFECIGNVNTMRSALEATHKGWGTSVV    296
Danio               FGATEFVNPK-DHS--KPIQEVLVELTDGGVDYSFECIGNVGIMRAALEACHKGWGTSVI    293
Homo                FGATECINPQ-DFS--KPIQEVLIEMTDGGVDYSFECIGNVKVMRAALEACHKGWGVSVV    291
                    :* .  :**         **: ::::*.** **:**.:*..  **:***. *:*** * :
```

```
Shewanella          IGVAGAGQEISTRPFQLVTGRVWRGSAFGGVKGRSQLPKIVEQYLAGEFKLDDFITHTMG    354
Galdieria           IGVAGAGEEISTRPFQLVTGRVWRGSAFGGVKGRSQLGGFLERYKQGEIFVDDLVTGELP    350
Porphyra            IGVAGAGEMIQTRPFQLVTGRRWVGTAFGGTKGRTEVPLLVKEYMDGKLKIDECITKKYS    356
Chondrus            IGVAGSGQEISTRPFQLVTGRVWRGSAFGGVKSRSQLPGMIDEYMCGDIKIDEMISKTYP    351
Arabidopsis         VGVAASGQEISTRPFQLVTGRVWKGTAFGGFKSRTQVPWLVEKYMNKEIKVDEYITHNLS    354
Cyanidioschyzon     IGVAASGEEISTRPFQLVTGRTWKGTAFGGAKSRTQLPELVDMYMKGVINIDDYVTGTYK    355
Dictyostelium       VGVAPAGAEISTRPFQLVTGRVWKGSAFGGVKSRSQLPSIIDKYMDKKLKVDEYVTFTYP    356
Drosophila          IGVAGAGQEISTRPFQLVVGRVWKGSAFGGWRSVSDVPKLVEDYLKKDLLVDEFITHELP    356
Danio               IGVAGAGQEISTRPFQLVTGRTWKGTAFGGWKSVESVPKLVNDYMNKKLMVDEFVTHTLP    353
Homo                VGVAASGEEIATRPFQLVTGRTWKGTAFGGWKSVESVPKLVSEYMSKKIKVDEFVTHNLS    351
                    :*** :*  * *******.** * *:**** :.  .:  ::. *    : :*: ::

Shewanella          LEQVNEAFDLMHEGKSIRSVIHFDK--          379
Galdieria           FESIQQAFDDLHHGKAIRTVLKYNQ--          375
Porphyra            LDQINDAFADMHDGKLIRGVIVF----          379
Chondrus            LKDINQAFEDMHSGKNIRGVVLFEAK-          377
Arabidopsis         LGEINKAFDLLHEGTCLRCVLDTSK--          379
Cyanidioschyzon     LDDINRAFEEMHNGRSIRSIILMDDDA          382
Dictyostelium       LNEINTAFDVMHEGKSLRSVVNL----          379
Drosophila          LSQINEAFDLMHKGESIRSIIKY----          379
Danio               FAQINEAFDLMHAGKSIRAVLQF----          376
Homo                FDEINKAFELMHSGKSIRTVVKI----          374
                    : .:: **  :* *  :* ::
```


Figure 2. 2: Multiple sequence alignment of class III alcohol dehydrogenases from different organisms using *Chondrus crispus ADH* as a query protein.

The Clustal omega tool was used to align amino acid sequences with GenBank accession numbers: *Chondrus crispus* (XP_005711469), *Galdieria sulphuraria* (XP_005705275), *Cyanidioschyzon merolae* (XP_005538792), *Porphyra umbilicalis* (OSX74245), *Arabidopsis thaliana* (CAA57973), *Danio rerio* (NP_571924), *Shewanella oneidensis* MR-1 (NP_717657), *Drosophila melanogaster* (NP_524310), *Dictyostelium discoideum* AX4 (XP_640467), and *Homo sapiens* (NP_000662). Conserved amino acid residues are shown as follows: solid black box showed catalytic zinc ion coordinating residues, yellow shaded residues showed amino acids involved in the structural zinc binding, amino acid residues with red background showed essential residues in the NAD-binding site, and the green highlighted residues represent the amino acids for putative substrate binding. The identical amino acid residues are marked with an asterisk (*), conserved amino acid residues marked with a colon (:) and semi-conserved residues are marked with a full stop (.).

The essential conserved catalytic Zn and structural Zn binding site in all class III ADH sequences used in sequence alignments and *CcADH3* amino acid residues were shown in **Table 2.8** as capitalized letters for each amino acid symbol. The tick marks were used to show homologous amino acid sequences in each organism.

Table 2. 8: The catalytic and structural Zn binding site amino acid residues in class III alcohol dehydrogenases (See Figure 2.2 for Clustal omega sequence alignment)

Organisms	**Catalytic Zn binding site**				**Structural Zn binding site**			
	C	**H**	**E**	**C**	**C**	**C**	**C**	**C**
Chondrus crispus	√	√	√	√	√	√	√	√
Shewanella oneidensis	√	√	√	√	√	√	√	√
Galdieria sulphuraria	√	√	√	√	√	√	√	√
Porphyra umbilicalis	√	√	√	√	√	√	√	√
Arabidopsis thaliana	√	√	√	√	√	√	√	√
Cyanidioschyzon merolae	√	√	√	√	√	√	√	√
Dictyostelium discoideum	√	√	√	√	√	√	√	√
Danio rerio	√	√	√	√	√	√	√	√
Homo sapiens	√	√	√	√	√	√	√	√
Drosophila melanogaster	√	√	√	√	√	√	√	√

The amino acid residues involved in the substrate and NAD-binding site in class III alcohol dehydrogenases from different organisms are indicated by ticks and the capital letters indicate different amino acid residues (**Table 2.9**).

Table 2. 9: The amino acid residues involved in the substrate and NAD-binding site of class III alcohol dehydrogenases in different species using the *C.crispus* amino acid residues as reference (bold capitalized letters)

Organism	**Substrate binding site**											**NAD-binding site**																					
	C	**T**	**D**	**P**	**E**	**H**	**Y**	**V**	**C**	**V**	**A**	**H**	**T**	**C**	**T**	**G**	**L**	**G**	**G**	**V**	**D**	**T**	**K**	**C**	**I**	**E**	**T**	**I**	**G**	**S**	**A**	**F**	**R**
C. crispus	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√
S. oneidensis	√	√	√	√	√	√	√	T	√	√	√	√	√	√	√	√	√	√	A	I	√	I	√	√	√	N	V	√	√	√	√	√	√
G. sulphuraria	√	√	√	√	√	√	√	T	√	√	√	√	√	√	√	√	√	√	√	√	√	I	√	A	V	K	L	√	√	√	√	√	√
P. umbilicalis	√	√	√	√	√	√	√	I	√	√	√	√	√	√	√	√	√	√	√	√	√	I	√	A	√	A	√	√	√	T	√	√	√
A. thaliana	√	√	√	√	√	√	√	Q	√	√	√	√	√	√	√	√	√	√	T	√	√	I	√		√	S	V	V	√	T	√	√	√
C. merolae	√	√	√	√	√	√	√	I	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	A	√	K	√	√	√		√	√	√
D. discoideum	√	√	√	√	√	√	√	I	√	√	√	√	√	√	√	√	√	√	A	√	√	N	√	√	√	N	V	V	√	√	√	√	√
D. rerio	√	√	√	√	√	√	√	I	√	√	√	√	√	√	√	√	√	√	A	√	√	V	√	√	√	G	I	√	√	T	√	√	√
H. sapiens	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	I	√	√	√	K	V	V	√	T	√	√	√
D. melanogaster	√	√	√	√	√	√	√	I	√	√	√	√	√	√	√	√	√	√	A	√	√	I	√	√	√	N	√		√	√	√	√	√

The *Cc*D-LDH sequences were used in alignment to find regions of similarity, and it was found that these two LDH-like proteins in *C. crispus* share 43% identity similarity with each other. When they aligned with other LDH-like proteins in other red algae species, it was found to be most similar to D-LDH in *G. sulphuraria* with 48% identity. These LDH-like proteins were further aligned with plant species, and they showed 46% identity similarity with D-LDH from Q*uercus suber* and lower identity with D-3-phosphoglycerate and formate dehydrogenase with both monocot and dicot plant species. High protein identity was found in 2-hydroxyacid dehydrogenase from bacterial species with a 55% identity cyanobacteria *Nodularia spumigena*. The protein identity similarity among the members of the 2-hydroxyacid dehydrogenase family is outlined in **Table 2.10**.

Table 2. 10: Sequences that produces significant sequence alignments with *CcD-LDH*

Protein	*Cc*D-LDH
*Cc*D-LDH-like	43%
*Gs*D-LDH	46%
*Ps*D-LDH	49%
*Nm*D-LDH	50%
*So*D-LDH	48%
*Ec*D-LDH	47%
*S*_2-HACID	49%
*Ns*_2-HACID	55%
*Qs*D-LDH	46%
*Dd*D-LDH	43%
*Ob*D-3-PGH	34%
*Os*D-3-PGH	34%
*Br*FDH	29%
*At*FDH	29%

The D-lactate dehydrogenase sequences from *C. crispus* were aligned with similar amino acids from other species, and the multiple sequence alignment revealed lower identity similarity among these sequences. The critical binding motifs of these amino acid sequences from different organisms were highly conserved, suggesting that these proteins were members of the same family. The amino acid residues in the active site in these sequences were highly conserved, suggesting that these proteins share the same functional properties.

In the multiple sequence alignments, the amino acids involved in the active site were marked with a rectangular box outlined with the red colour. The amino acid residues identified in the sequence alignment were highly similar or even identical to the ones reported for members of the 2hydroxyacid dehydrogenase family (Taguchi and Ohta, 1991). The amino acid residues essential for catalysis of D-lactate dehydrogenase were located at Arg-238, Glu-267 and His-299. These three amino acid residues have been reported in the literature as critical for substrate binding and catalytic activity of D-lactate dehydrogenases (Razeto *et al*., 2002).

The amino acid sequences in the putative ligand binding or substrate binding site were also highly conserved in all the sequences used in alignment. The amino acid resides Cys-80, Ala-81, Gly-82, Tyr-104, Arg-238, and His-299 were present in D-lactate dehydrogenase, suggesting a wide range of substrate specificity among the D-LDHs. The NAD-dependent 2-hydroxyacid dehydrogenases have been shown to have D- isomers for their substrates (Dengler *et al*., 1997).

The sequence alignment also revealed a NAD-binding site motif [GXGXXGX17D] (Gly-155, Gly-156, Lys-158, Ile-159 and Gly-160) and this region had been shown to be the coenzyme binding domain (Razeto *et al*., 2002). This binding motif site further confirmed that the D-lactate dehydrogenase gene in *C. crispus* is the member of the 2-hydroxy acid dehydrogenase family because of the presence of this highly conserved nucleotide binding domain in all the sequences used in the alignment. These highly conserved amino acid residues in the multiple sequence alignments are also outlined in **Table 2.11** and **2.12**.

```
XP_005714861CcD-LDH-like      MADSFNAAGQSQRLRVIMFGARQYEVAAFRNEMKLSPETQLLDIVFVSAPLDKSTATLAA   60
XP_640819DdD-LDH              -------------MKITLFSSKPYWVKWFNELNK-----FSYEINYVTSACDIKSVNEAK   42
XP_005715286CcD-LDH           -------------MKVTVFSSKRYERDFLRQHPTLTRVTPHIDFEFLAMPLSEGTAPLAA   47
NP_716597SoD-LDH              -------------MRIGFFSAKHYDMQHFNRTNA----AFDAQIEYFDYRLCMQTVKLAE   43
NP_274689NmD-LDH              -------------MKIAIYGTKSYDREHFTRANR----HFGFELEFFDFMLDAKTAKMAE   43
NP_415898EcD-LDH              -------------MKLAVYSTKQYDKKYLQQVNE----SFGFELEFFDFLLTEKTAKTAN   43
WP_010874176S_2-HACID         -------------MKIAFFSSKAYDRQFFQQANH----PHQREMVFFDAQLNLDTAILAE   43
NP_249618 PaD-LDH             -------------MRILFFSSQAYDSESFQASNH----RHGFELHFQQAHLQADTAVLAQ   43
                                           ::: .:.:: *    :            :: :        :.  *

XP_005714861CcD-LDH-like      GAKAVSLFANDYADAAILQILHASGVELITLRYAGVPAVDRNKVASLKMRLAPAPAHAPT   120
XP_640819DdD-LDH              GSEAVCCFVNDDLSKEVIETLHSNGTKVILMRCAGFNKVDLDTANKLGIPVLRVPAYSPN   102
XP_005715286CcD-LDH           GSDAVCAFVNDALDAPVIARLAEGAVRCILLRCAGFNHVDLAAASKHRIKVLRVPAYSPF   107
NP_716597SoD-LDH              GFEVVCAFVNDSLCEEVLVELAKGGTKIIAMRCAGFNNVDLVAAKRLGMQVVNVPAYSPE   103
NP_274689NmD-LDH              GAEAVCIFVNDDGSRPVLEKLAQIGVKTVALRCAGFNNVDLKAAEELGLKVVRVPAYSPE   103
NP_415898EcD-LDH              GCEAVCIFVNDDGSRPVLEELKKHGVKYIALRCAGFNNVDLDAAKELGLKVVRVPAYDPE   103
WP_010874176S_2-HACID         DCPVICLFVNDQAPAPVLEKLAAQGTKLIALRSAGYNNVDLKTAADLGLKVVHVPSYSPH   103
NP_249618.PaD-LDH             GFEVVCAFVNDDLSRPVLERLAAGGTRLVALRSAGYNHVDLAAAEALGLPVVHVPAYSPH   103
                              .  .:. *.**     ::  *   ... : :* **   **   .    : :  .*:: *

XP_005714861CcD-LDH-like      SIAEYTVLLILSLNRKLHLAYNRVREGNMTMHGLVGFDMKGKVVGLIGTGKVGTTVARIL   180
XP_640819DdD-LDH              AVSEYALSLIMALNRKTHKAHDRVRDANFEINGMEGFNMVSKVYGIVGTGNIGEQLCRVL   162
XP_005715286CcD-LDH           AVAEFAVALLLTVGRKTHKAYNRTRESNFSLAGLMGFDVNGKTIGVCGTGKIGRLFAKIM   167
NP_716597SoD-LDH              SVAEHTVALMLTLNRKIHKAYQRTRDANFSLEGLVGFNMFGKTVGVIGTGKIGVATIKVL   163
NP_274689NmD-LDH              SVAEHTVGLMLTLNRRIHKAYQRTRDANFSLEGLTGFNMYGKTAGVIGTGKIGIATMRIL   163
NP_415898EcD-LDH              AVAEHAIGMMMTLNRRIHRAYQRTRDANFSLEGLTGFTMYGKTAGVIGTGKIGVAMLRIL   163
WP_010874176S_2-HACID         AVAEHTVGLILALNRKLYRAYNRVRDDNFSLEGLLGFDLHGTTVGVIGTGKIGLAFAQIM   163
NP_249618.PaD-LDH             AVAEHAVGLILTLNRRLHRAYNRTREGDFSLHGLTGFDLHGKRVGVIGTGQIGETFARIM   163
                              :::*.:: ::::::.*: : *::*.*: :: : *: ** : ..  *: ***::*    :::

XP_005714861CcD-LDH-like      -SGFGCNVIAFDMIESPDIKHMGI-KYVSINTLLATSDIISLHAPLVSGTRHMIGAQSIP   238
XP_640819DdD-LDH              KLGFGAKVIAYDIIENKAVTDIGIEYVKTLDEIWKQCDVISLHTPLNSQTKYMVNSESIE   222
XP_005715286CcD-LDH           -LAFGTTVLAYDVYRNPEAEAMGV-EYVSKDDLFARSDIISLHCPLLPSTRHMIDAEAIR   225
NP_716597SoD-LDH              -LGFGCKVIAFDPYPNPAVEALDV-EYQDLDTIYATSDIISLHCPLTPDNHHLLNKDSFA   221
NP_274689NmD-LDH              -KGFGMNLLAYDPFCNPEAEKIGG-KYVDLDELYARSDIITLHCPATPENHYMLNEAAFD   221
NP_415898EcD-LDH              -KGFGMRLLAFDPYPSAAALELGV-EYVDLPTLFSESDVISLHCPLTPENYHLLNEAAFE   221
WP_010874176S_2-HACID         -NGFGCHLLGYDAFPNDKFTAIGQALYVSLNELLAHSDIISLHCPLLPETHYLINTNTIA   222
NP_249618.PaD-LDH             -AGFGCELLAYDPYPNPRIQALGG-RYLALDALLAESDIVSLHCPLTADTRHLIDAQRLA   221
                                .**  ::.:*   .    :.         :   .*:::** *    . :::.   :
```

Figure 2. 3: Multiple sequence alignment of D-isomer specific 2-hydroxyacid dehydrogenase family proteins from seven different organisms.

The amino acid sequences of D-LDHs and 2-hydroxyacid dehydrogenases were aligned from: *C. crispus* (CcD-LDH and CcD-LDH-like), *Dictyostelium discoideum* AX4 (DdD-LDH), *Shewanella oneidensis* MR-1 (SoD-LDH), *Neisseria meningitidis* MC58 (NmD-LDH), *Escherichia coli* MG1655 (EaD-LDH), *Synechocystis* (S_2-HACID), and *Pseudomonas aeruginosa* POA1 (PaD-LDH). The GenBank accession numbers of each protein are marked by the turquoise colour behind to the corresponding protein. The conserved important amino acid residues are indicated by the red rectangular box for active site residues. The green coloured background residues denoted amino acid residues involved in the ligand binding/substrate binding site. The yellow coloured amino acid residues were shown to be involved in the NAD binding site.

The amino acid residues Arg-238, Glu-267 and His-299, were highly conserved in the D-lactate dehydrogenases sequences used in the alignments. These amino acid residues have been well investigated and shown to be critical for substrate binding and catalysis of many D-isomer specific 2-hydroxyacid dehydrogenases (Razeto *et al*., 2002) (**Table 2.11**).

Table 2. 11: The critical amino acid residues in the active site and substrate binding site in *C. crispus* D-lactate dehydrogenase and several other organisms used in sequence alignment (Figure 2.3).

Organism	**Active site**			**Putative ligand binding site**						
	R	**E**	**H**	**C**	**A**	**G**	**Y**	**R**	**H**	**F**
Chondrus crispus	√	√	√	√	√	√	√	√	√	√
Dictyostelium discoideum	√	√	√	√	√	√	√	√	√	W
Shewanella oneidensis	√	√	√	√	√	√	√	√	√	√
Neisseria meningitidis	√	√	√	√	√	√	√	√	√	√
Escherichia coli	√	√	√	√	√	√	√	√	√	√
Synechocystis	√	√	√	S	√	√	√	√	√	√
Pseudomonas aeruginosa	√	√	√	S	√	√	√	√	√	√
Chondrus crispus LDH-like	√	√	√	Y	√	√	H	√	√	A

Table 2. 12: The amino acid residues found in the NAD-binding site of the 2-hydroxyacid dehydrogenases protein sequences used in the Clustal omega sequence alignment (Figure 2.3)

Organism	**Putative NAD-binding site amino acid residues**																							
	Y	**V**	**G**	**G**	**K**	**I**	**G**	**Y**	**D**	**V**	**H**	**C**	**P**	**L**	**T**	**M**	**T**	**S**	**R**	**D**	**V**	**H**	**A**	**F**
Chondrus crispus	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√
Dictyostelium discoideum	√	√	√	√	N	√	√	√	√	I	√	T	√	√	√	√	V	√	√	√	√	√	√	W
Shewanella oneidensis	√	√	√	√	√	√	√	F	√	P	√	√	√	√	N	L	√	√	√	√	√	√	√	√
Neisseria meningitidis	√	√	√	√	√	√	√	√	√	P	√	√	√	A	N	√	√	√	√	√	√	√	√	√
Escherichia coli	√	√	√	√	√	√	√	F	√	P	√	√	√	√	N	L	√	√	√	√	√	√	√	√
Synechocystis	√	√	√	√	√	√	√	√	√	A	√	√	√	√	√	L	√	√	√	√	√	√	G	√
Pseudomonas aeruginosa	√	√	√	√	Q	√	√	√	√	A	√	√	√	√	√	L	√	√	√	√	√	√	√	√
Chondrus crispus LDHlike	H	I	√	√	√	V	√	F	√	M	√	A	√	√	√	√	√	√	√	√	S	√	√	A

2.6.5 Signal peptide, transmembrane and subcellular localization prediction of the ADH and LDH proteins

The analysis of the ADH and LDH protein sequences by SignalP 3.0 and 4.0 Server (Emanuelsson *et al*., 2007; Petersen *et al*., 2011), showed no signal peptides. These proteins were deemed as nonsecretory proteins, and no typical transmembrane domain was observed by TMHMM 2.0 server. The subcellular localization prediction by TargetP 1.1 Server and WoLF PSORT showed that these proteins are most likely situated in the cytoplasm.

2.6.6 Phylogenetic analysis of dehydrogenases

To have a better understanding of the functional divergence of the ADH and LDH genes in *C. crispus*, the amino acid sequences of these two family members, including red algae, plants, human, bacteria, insect, and zebrafish, were analyzed by the Neighbor-joining (NJ) phylogenetic tree using the MEGA6 program. The two alcohol dehydrogenases were used to build a neighbourjoining phylogenetic tree along with cinnamyl alcohol dehydrogenase and class III alcohol dehydrogenases from other species.

The phylogenetic tree showed these two ADH proteins belong to different groups, suggesting that they may have a different function. It has been reported that most of the CAD proteins in plant cell walls are involved in the last step of monolignol metabolism (Eudes *et al*., 2006; Ma, 2010). Members of the CAD proteins were found to be close to CAD1 protein found in *C. crispus*. The other ADH protein was found in the group with class III alcohol dehydrogenases from other species.

The family members of these proteins have been reported to evolve from the glutathione-dependent formaldehyde dehydrogenases (Strommer, 2011). All the ADH proteins in *C. crispus* were grouped as medium-chain dehydrogenases/reductases based on the amino acid length and presence of zinc ion which has been found to be present in most MDRs superfamily proteins. The difference in these families may also indicate that these ADH proteins have different substrate specificities among the *C. crispus* genome.

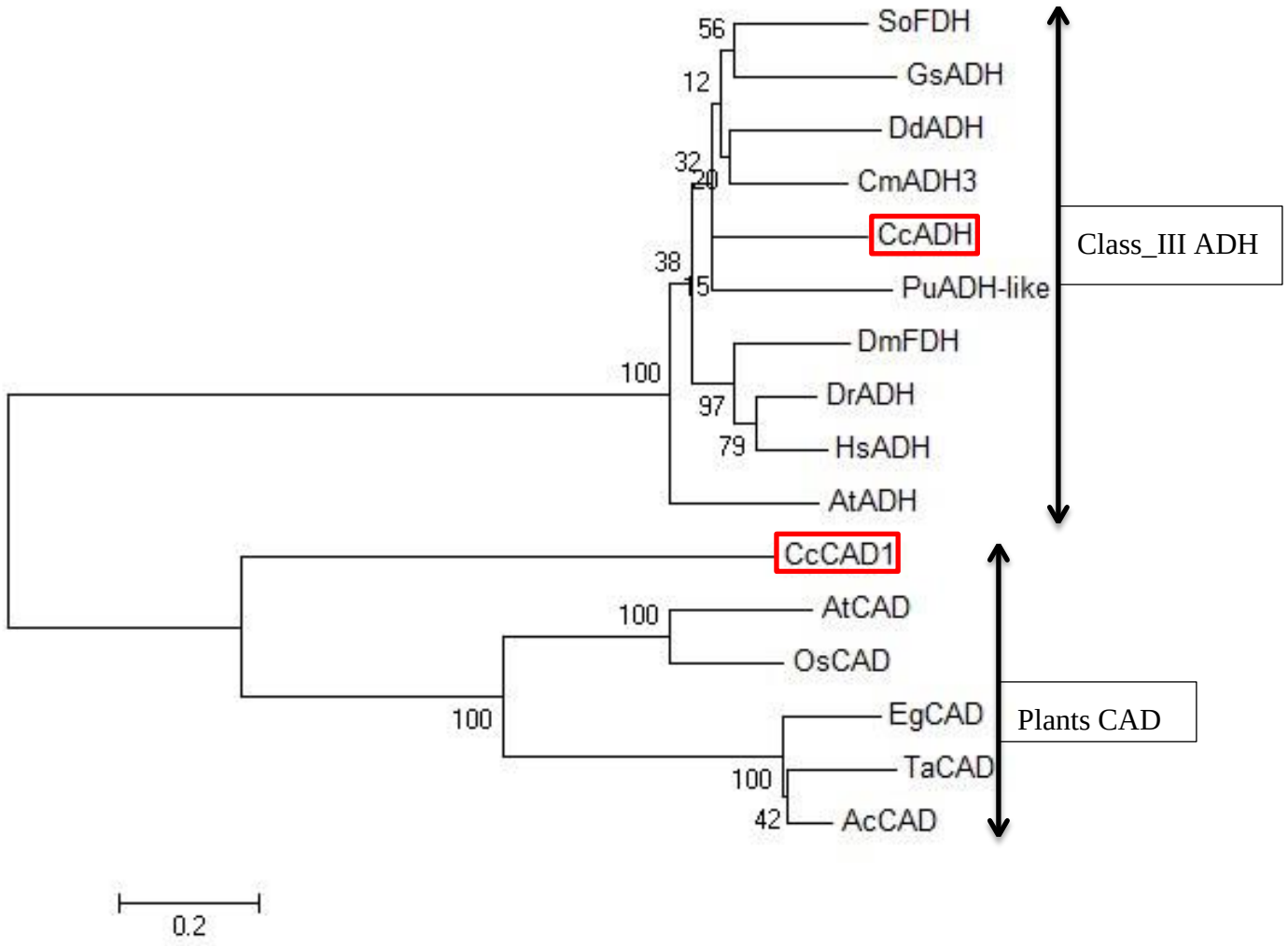


Figure 2. 4: The phylogenetic relationship of class III alcohol dehydrogenase and cinnamyl alcohol dehydrogenase proteins of *C.crispus* (in red rectangular box) with similar proteins from other organisms using MEGA6 program.

The numbers below the branches represent the bootstrap values of mostly 50% or more support. The scale bar showed the levels of sequence divergence. The CAD and class III ADH sequences were shown to belong to two different groups in the phylogeny tree. The plant's CAD proteins were grouped together and most of these protein groups showed a bootstrap value of 100% showing that the estimation of these groups has a high probability. The ADH III proteins were shown to be the descendants of the glutathione-dependent formaldehyde dehydrogenases.

The phylogenetic tree was constructed using the Neighbour-joining (NJ) method from MEGA6 program and several lactate dehydrogenases from different organisms were used to check the

evolutionary relationship of these proteins. The tree showed several groups of 2-hydroxyacid dehydrogenase families from different organisms. The LDH-like proteins from red algae are grouped among the D-lactate dehydrogenases from other species such as plants (*Quercus suber*) and bacteria (mostly cyanobacteria). This showed that the LDH proteins obtained in *C. crispus* have functional divergence from D-lactate dehydrogenases. Compared to other 2-hydroxyacid dehydrogenases, the D-lactate dehydrogenases have D-isomers for their substrate, and they may slightly differ with other 2-hydroxyacid dehydrogenases in plants and other bacterium species.

Most plants species were grouped together, and they are reported to have conserved regions with formate/glycerate dehydrogenases. The 2-hydroxyacid dehydrogenase proteins from fungi and bacteria identified were grouped together. Proteins which have been found to lack the D-isomers for their substrates have been grouped together, including L-LDH isomers which are mostly common in many vertebrates. The phylogenetic analysis revealed that the hypothetical genes in *C. crispus* have the same evolutionary origin with D-lactate dehydrogenase from other species. This may indicate that these proteins share the same functional characteristics and are members of the D-isomer specific 2-hydroxyacid dehydrogenases (**see Appendix B**).

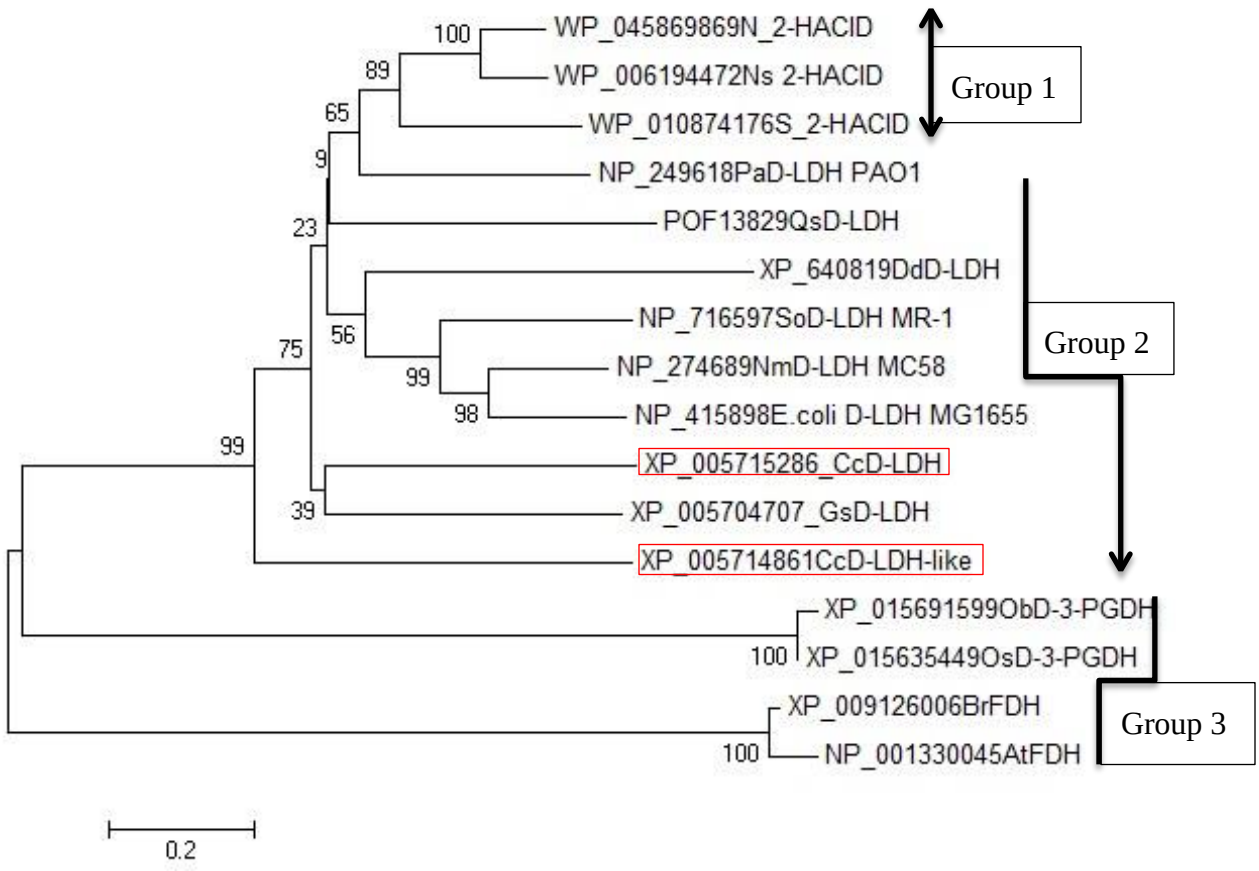


Figure 2. 5: Phylogenetic tree of two *C. crispus D-LDH* proteins (marked in rectangular red box) and other proteins members of 2-hydroxyacid dehydrogenase gene family from different species using MEGA6 program.

The Neighbour-joining (NJ) tree was constructed using the MEGA6 program, and the bootstrap of 1000 replicates was employed in the analysis. The GenBank accession number of each protein was written next to the corresponding proteins. The following species were used to construct the phylogeny: *C. crispus*, *Galdieria sulphuraria*, *Quercus suber*, *Nostocales*, *Nodularia spumigena*, *Dictyostelium discoideum*, *Synechocystis*, *pseudomonas aeruginosa* PAO1, *Neisseria meningitidis* MC58, *Escherichia coli* str.K-12 substr. MG1655, S*hewanella oneidensis* MR-1, *Brassica rapa*, *Arabidopsis thaliana*, *Oryza sativa* japonica group, and *Oryza branchyantha*.

2.6.7 Computation analysis of the *cis*-acting regulatory elements in ADH and LDH gene families from red alga (*Chondrus crispus)*.

The genomic sequences obtained in the 1.5 kbp 5′ upstream from the translational start site of alcohol dehydrogenase and lactate dehydrogenase genes in *C. crispus* were analysed by the PlantCARE Database (Rombauts *et al*., 1999; Lescot *et al*., 2002) and several *cis*-acting regulatory elements (CAREs) were identified in the upstream regions of these dehydrogenase genes. The PlantCARE database revealed 88 and 84 *cis*-acting regulatory elements with different sequence lengths between 4-12 bp for both LDH and ADH genes, respectively. The identified *cis*-acting regulatory elements were grouped by their motif name, sequence, position and functions (**see Appendix E**).

The names of the identified CAREs included TATA-box, CCAT and GC-box, which are common in the upstream region of the translation start site. Most of the CAREs in the gene promoters were identified to be involved in light response, hormonal response, abiotic and biotic stress, as well as in growth and development. The highest number of *cis*-acting elements in these genes was associated with light responses, and hormonal responses and the stress-responsive elements were abundant in the 1.5kbp 5′ regulatory regions of these dehydrogenase gene family. This may indicate that the expression pattern of these genes could be regulated by various environmental factors.

The light responsive elements were spread across the 5′ upstream regions of these gene families. Light is a vital source of energy for many organisms including plants, and therefore it is one of the critical environmental factors and plants have complex transcriptional network that mediates developmental changes in response to light (Jiao *et al*., 2007). developed a Red algae have also been shown to utilize light for various physiological and developmental processes, and it has been reported that marine algae and plants have the same phytochrome signalling system (Duanmu *et al*., 2014).

The light response-associated *cis*-acting regulatory elements such as ACE, CATT_motif, GT1_motif, G-Box, GAG_motif, MNF1, Box 4, Sp1, AE-box, and l-box were identified in 1.5kbp 5′ upstream regulatory region in dehydrogenase gene sequence. Plant hormones such methyl jasmonate (MeJA), abscisic acid (ABA) and GARE-motifs were also identify in the upstream region of these dehydrogenase genes.

These CAREs play essential roles in the development and cellular processes in plants. MeJA is a plant hormone which is regulated during various developmental processes and in *C. crispus*; this plant hormone plays a physiological role as a stress hormone in red algae (Collen *et al*., 2006). The cis-acting regulatory elements which are involved in MeJA responsiveness were identified including TGACG_motif and CGTCA_motif. The CAREs such as salicylic acid (SA), gibberellin (GA), and auxin responsive elements were identified, and they play crucial roles in various plant developmental and cellular processes (Hagen and Guilfoyle, 2002)

Other CAREs such as circadian, Sk_1_motif, and O2-site are also regulated during various stages of plant growth and developmental stages. Although most of the CAREs in *C. crispus* gene promoter are involved in stress responsiveness, the mechanism to which they respond in these conditions is unknown. Studies need to be done in future using experimental evidence from other sources to try to expand this knowledge by investigating the role of red algae genes in these response mechanisms. The use of computational approach is helpful in fast-tracking this knowledge, and given the theoretical and structural background of several genes from other sources, this might be a fruitful consideration for future studies.

The illustration and distribution of the identified *cis*-acting regulatory elements in ADH and LDH gene family in *C. crispus* genome were revealed by a plant-based promoter sequence analysis PlantCARE database. The putative *cis*-acting regulatory elements identified in the 5′ upstream regulatory region in the ADH3 gene sequence showed a high number of sequence motifs involved in stress responsiveness (**Figure 2.6**).

The CAREs in the ADH genes showed similar sequence motifs in the 5′ regulatory regions. In **Figure 2.7**, the CcCAD gene showed a high number of light responsive elements in the 5′ upstream region. The LDH gene family showed a high number of CAREs that are activated during the stress response. There were also several CAREs that are involved in hormonal regulation. The number of putative motifs in the 5′ regulatory region in *C. crispus* LDH gene revealed *cis*-acting regulatory elements that are involved in stress response, hormonal change and growth as well as developmental processes (**Figure 2.8**).

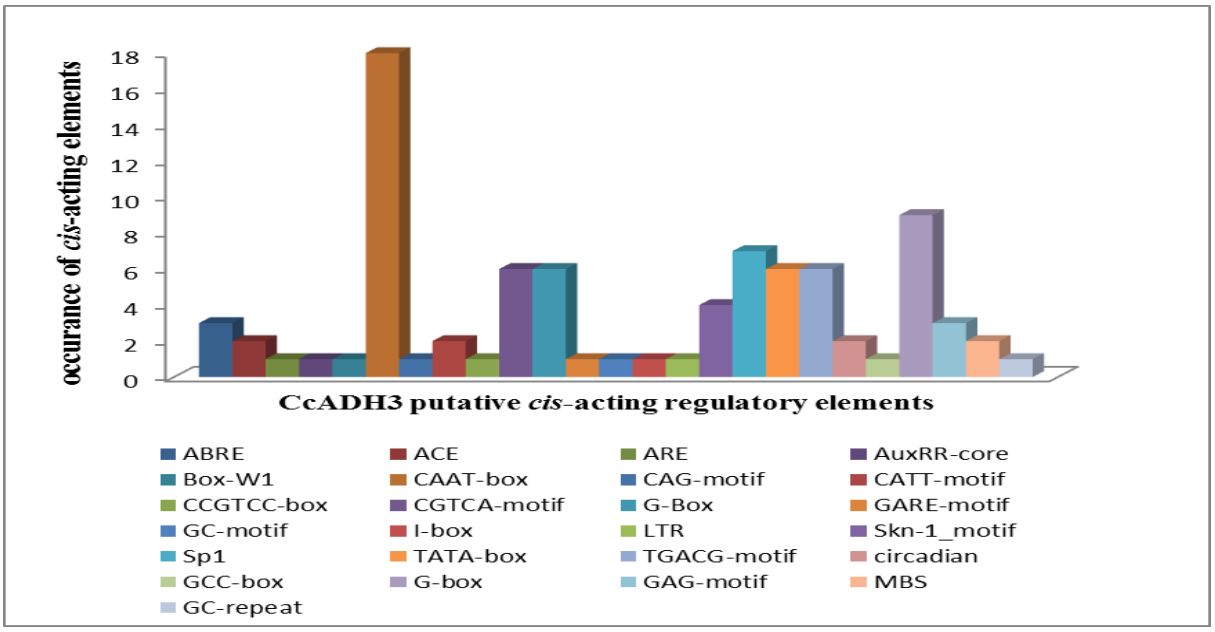


Figure 2. 6: Illustration of the putative *cis*-acting regulatory elements in the 5′ upstream region of the *Cc*ADH3 gene family.

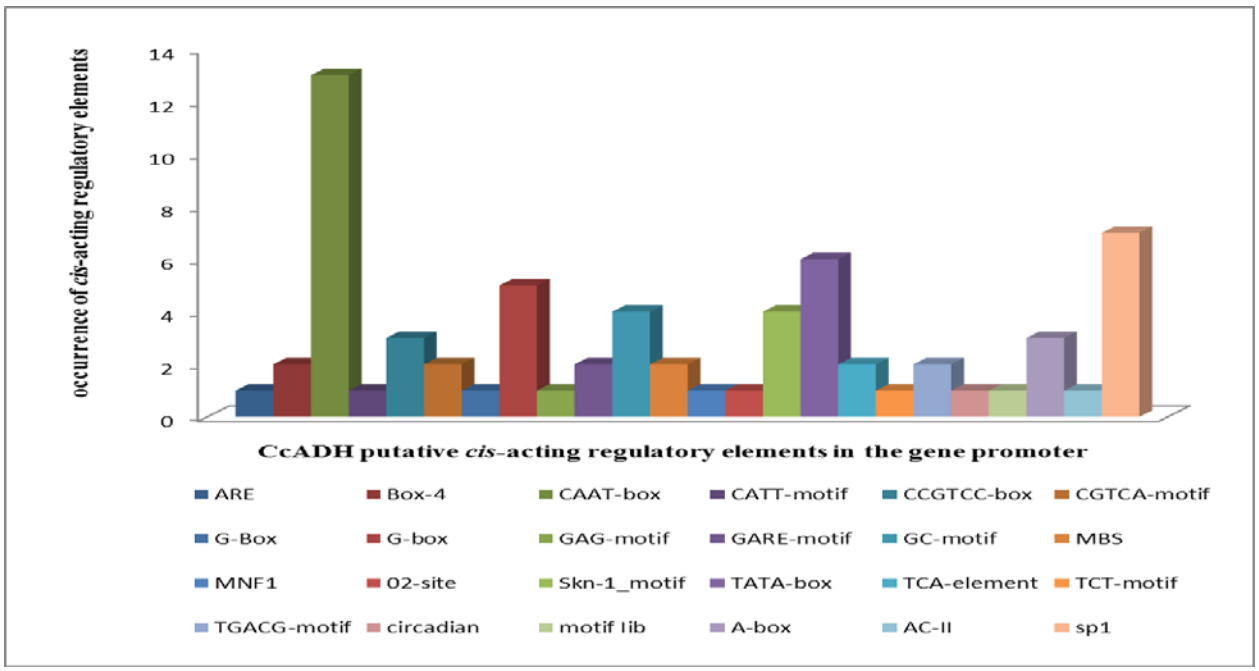
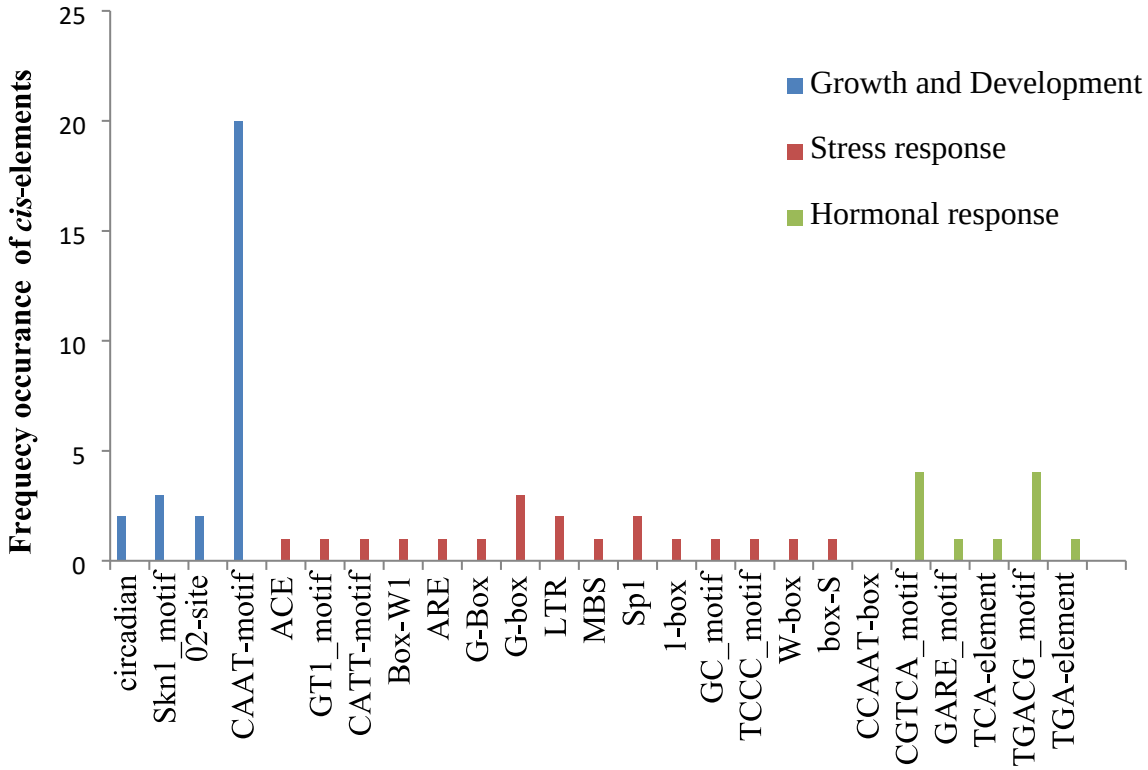


Figure 2. 7: Illustration of the putative cis-acting regulatory elements in the 5′ upstream regulatory region of *Cc*CAD gene family.



Putative *cis*-acting regulatory elements in the promoter region of *C. crispus* gene

Figure 2. 8: Illustration of the occurrence of the cis-acting regulatory elements in the 5′ upstream region of *C. crispus* LDH gene family.

2.6.8 Secondary structure analysis

The secondary structure analysis of *C. crispus* alcohol dehydrogenase class III (A) showed 14 (29%) alpha helices and 19 (28%) beta strands. The predicted secondary structure by the 3Phyre[2] server was based on the template of *Arabidopsis thaliana* s-nitrosoglutathione reductase (PDB code: 3UKO), which showed 62% identity. The secondary structure analysis of cinnamyl alcohol dehydrogenase (B) showed 12 (25%) alpha helices and 19 (29%) beta strands. Transmembrane helices were also identified in the secondary structures A (168-210) and B (176-219) in the N- to C-terminal regions, respectively (**see Appendix F**). The secondary structure confidence key is indicated by the coloured amino acid residues, with red coloured residues showing high confidence and purple coloured residues showing low confidence as shown in **Figure 2.9**.

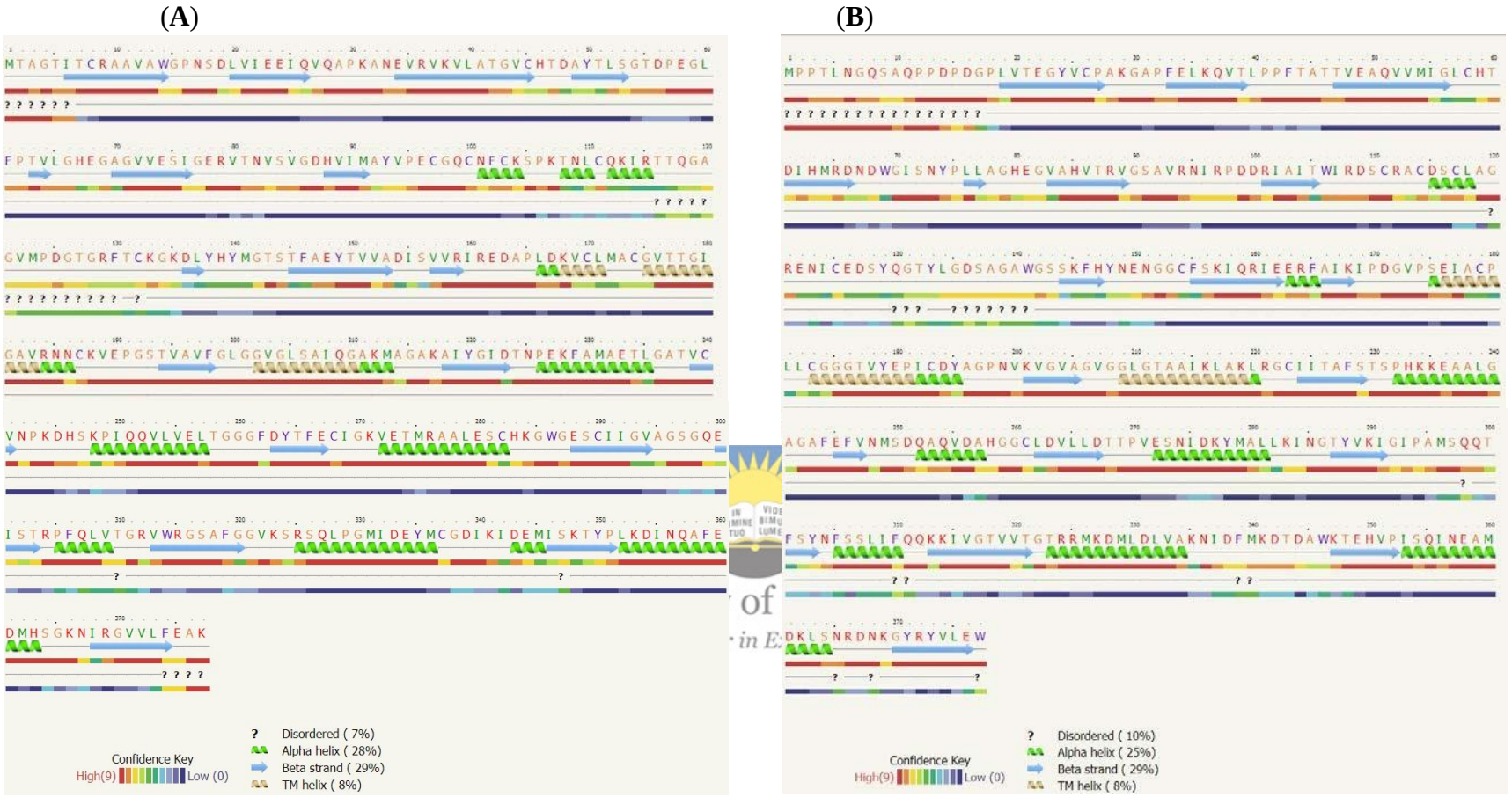


Figure 2. 9: Secondary structure of *Cc*ADH predicted by Phyre[2] server showing the alpha helix, beta strands and transmembrane region in the two *C. crispus* alcohol dehydrogenase amino acid sequences. (A) *C. crispus ADH* class III alcohol dehydrogenases; (B) *C. crispus* cinnamyl alcohol dehydrogenase (CcCAD1).

The class III ADH sequences showed over 50% sequence identity when aligned with the query protein from *C. crispus*. The structure-based alignment was performed by the PROMALS3D tool server using *C. crispus* class III ADH (*Cc*ADH) sequence and *Homo sapiens* 3D structures (PDB code: 1MC5, 3QJ5, 1TEH, and 2FZW) and *Arabidopsis thaliana* 3D structure (PDB code: 3UKO). The *C. crispus* sequence showed 65% identity with *H. sapiens* 3D structure sequences and 62% identity with *A. thaliana* 3D structure sequence.

The numbers at the top of each amino acid residues in the structure-based alignment indicates the conservation index for the aligned sequences. The high conservation index number 9 was observed in all the alignments, suggesting that *C. crispus* ADH sequence produces high similarities with class III ADH and reductase proteins from *H. sapiens* and *A. thaliana*. The predicted secondary structure elements showed 19 beta-strands and 14 alpha helices marked by red amino acid residues for alpha helices and blue coloured residues for beta-strands.

The conserved amino acids in the alignment are indicated by bold and uppercase letters. The consensus amino acid sequences (consensus aa) in the alignment showed symbols for different classes or groups of amino acids such as aliphatic, aromatic, hydrophobic, alcohol, and polar residues, small, bulky residues, positively charged, negatively charged and charged residues. The structure-based sequence alignments of class III ADH in *C. crispus* and similar proteins from *H. sapiens* and *A. thaliana* with known 3D structures are shown in **Figure 2.10**.

The other *C. crispus* ADH protein (*Cc*CAD) was also used in alignment with proteins from other organisms with known 3D structures in the PDB. These sequences showed high similarities and most of the secondary structure elements were predicted as beta-strands. The structure-based sequence alignment was performed using the structure from *Populus tremuloides* (PDB code: 1YQD), and *Methylobacterium extorquens* AM1 (PDB code: 4GI2). The conservation index for the aligned sequences was 9 for most of the conserved amino acid residues, and most of the consensus amino acid sequences were hydrophobic residues (**Figure 2.11**). The sequence identity similarity among these proteins was low, being 35% and 18%, respectively.

Figure 2. 10: Illustration of structure-based sequence alignments of *C. crispus* alcohol dehydrogenase proteins and other related proteins from different organisms with known 3 dimensional structures in the Protein Data Bank. (A) CcADH class III structure-based alignments with other related proteins from other organisms; (B) CcCAD structure-based alignments with other related proteins from other organisms.

The secondary structure of LDH proteins in *C. crispus* was predicted using the Phyre^2 server (**Figure 2.10**). The amino acid sequence analysis showed that D-LDH-like protein shown in **Figure 2.10A** contains 15 (39%) alpha helices and 12 (19%) beta strands. The secondary structure prediction analysis of LDH-like protein was based on the template of the *Chlamydomonas reinhardtii* protein structure (PDB ID: 4ZGS), which showed the highest identity similarity of 44%. Secondary structure analysis of D-LDH structure shown in Figure 2.10B revealed 16 (40%) alpha helices and 13 (19%) beta strands. The amino acid residues were coloured by their secondary structure confidence key, with red coloured residues showing highest confidence and purple coloured residues showing lowest confidence. The secondary structure analysis of *Cc*D-LDH was based on the template of *Pseudomonas aeruginosa* (PDB ID: 3WWZ), which showed 49% identity.

The structure-based amino acid sequence alignment was performed since the structure is better conserved than the amino acid sequence and the accurate sequence alignment is essential for homology searches and for building protein 3D structures (Kim and Lee, 2007). The structure-based sequence alignment between the *Cc*D-LDH sequence and two D-lactate dehydrogenase 3D structures from other organisms showed high conservation index number 9 as shown in Figure 2.11A. The D-lactate dehydrogenases from *Pseudomonas aeruginosa* (PDB code: 3WWZ) and *Salmonella typhi* (PDB code: 4CUJ) shared 49% and 48% identity, respectively with *C. crispus* protein sequence. The high structure-based alignment similarities among these three sequences indicated that *Cc*D-LDH has similar structure and function with these two experimentally validated crystal structures. The consensus amino acid sequences in the alignment were mainly hydrophobic residues, and the predicted secondary structure elements showed many alpha helices (13) than beta-strands (11).

The other protein shown in Figure 2.11B showed conservation index above 4, indicating high similarity among the aligned *Chondrus crispus* proteins with *Salmonella typhi* (PDB code: 4CUJ), *Pseudomonas aeruginosa* (PDB code: 3WWZ), *Escherichia coli* (PDB code: 4CUK), and *Chlamydomonas reinhardtii* (PDB code: 4ZGS). These proteins showed between 35 to 43% identity among each other. The secondary structure elements analysis showed 12 beta-strands and 15 alpha helices.

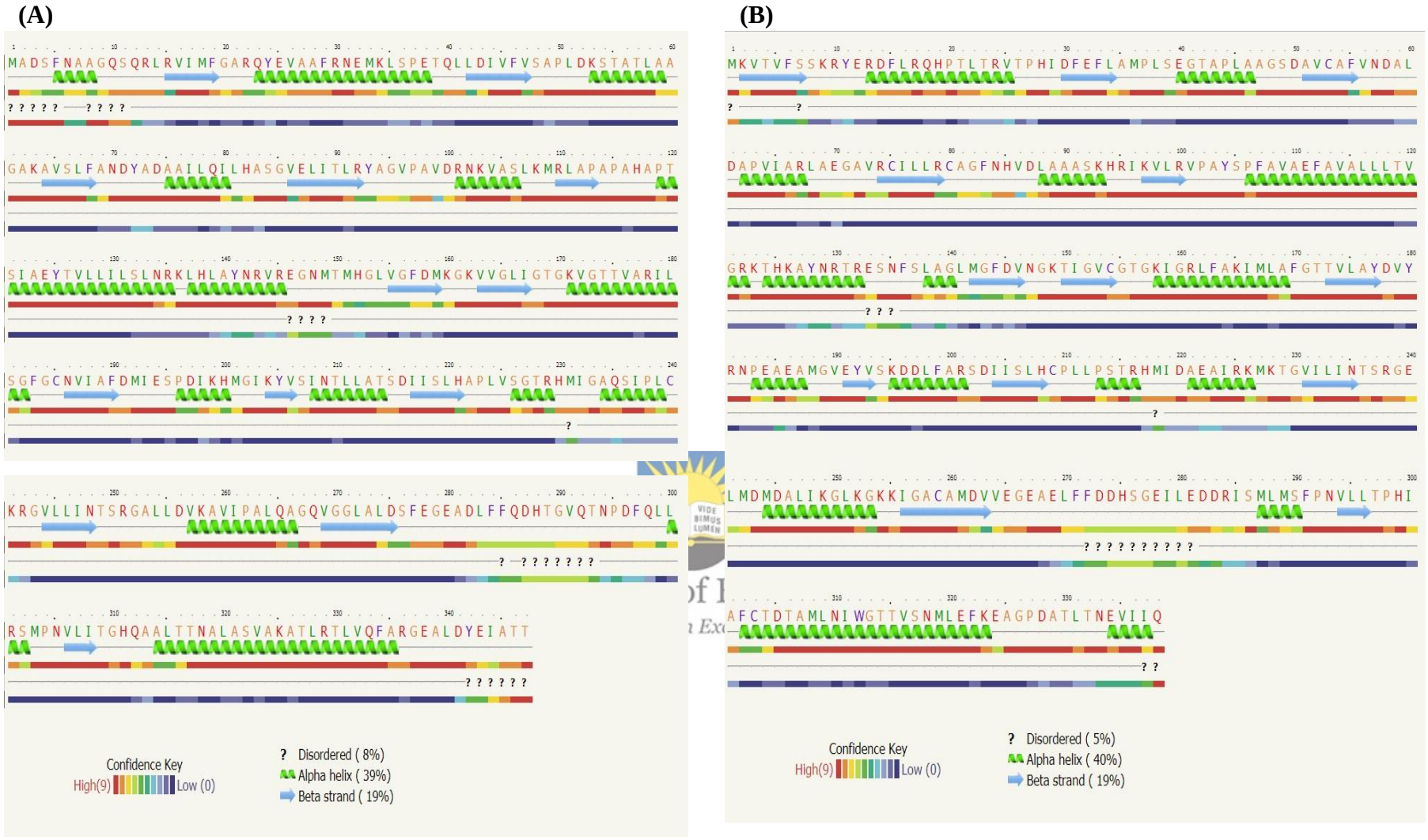


Figure 2. 11: Secondary structure prediction analysis of *Cc*D-LDH proteins showing α-helices (green) and β-sheets (blue arrow) by the Phyre[2] server. (A) Secondary structure prediction of C. crispus D-lactate dehydrogenase-like protein; (B) Secondary structure prediction of C. crispus D-lactate dehydrogenase

Figure 2. 12: Structure-based sequence alignments of *C. crispus* D-lactate dehydrogenases with other related proteins form other organisms with known 3D structures. (A) Structure-based alignment of CcD-LDH with other related proteins; (B) Structure-based alignment of CcD-LDH –like protein with D-lactate dehydrogenases from other organisms.

2.6.9 Tertiary structure analysis

The 3D structures of *C. crispus* proteins were built by the Swiss-PDB Viewer (Deep View). The multiple templates were modelled based on the closely related ADH and LDH proteins using PRIMO and SWISS-MODEL. The BLAST analysis showed that alcohol and lactate dehydrogenases from *Homo sapiens* (1MC5, 1TEH, 2FZW, 3QJ5), *Populus tremuloides* (1YQD), *Escherichia coli* K-12 (1UUF), *Catharanthus roseus* (5H83), *Pseudomonas aeruginosa* (3WWZ), *Salmonella typhi* (4CUJ), *Escherichia coli* (4CUK) were closely related to the *Chondrus crispus* proteins and these templates generated good models. The three-dimensional models of *C. crispus* proteins were generated by PRIMO and SWISS-MODEL, these two databases showed homologous models. The best-modelled structures were then selected by analysing the least DOPE-Z-score and accessed with Ramachandran plot.

The tertiary structure analysis of class III ADH modelled with alcohol dehydrogenase template structure from *H. sapiens* (PDB code: 1MC5) which showed the highest identity. The 3D ribbon structure model was rich in β-strands than alpha helices, and similar type of structures has been identified from other studies in literature (Amelia *et al*., 2015). The quality analysis of the modelled structure was evaluated by the Ramachandran plot showing that 95.1% (355 amino acid residues) were found in the most favourable region, 4.5% (17 amino acids) in the additional allowed region and 1% (4 amino acids) in the generously allowed region. There were no amino acid residues detected in the disallowed region, suggesting that the *C. crispus* 3D structure model has a stable conformation. The templates from *H. sapiens* were also used to generate a 3D model with ADH class III-like protein from other red algae genomes, and these structures showed similar overall fold, indicating that they might share a common function (**see Appendix G**).

The other *C. crispus* alcohol dehydrogenase which is related to the cinnamyl dehydrogenases was modelled using the template structure from *E. coli* (1UUF) sharing more than 30% protein identity similarity. The Ramachandran plot showed 83.6% (260 amino acids) of residues falling in the most favoured regions, 13.5% (42 amino acids) found in the additional allowed region, 1.9% (6 amino acids) in the generously allowed regions and 1% (3 amino acid) residues in the disallowed regions. An illustration of the tertiary structure and validation of *C. crispus* alcohol dehydrogenases showing a successful model of the target sequence using very slow refinement is shown in **Figure 2.13**.

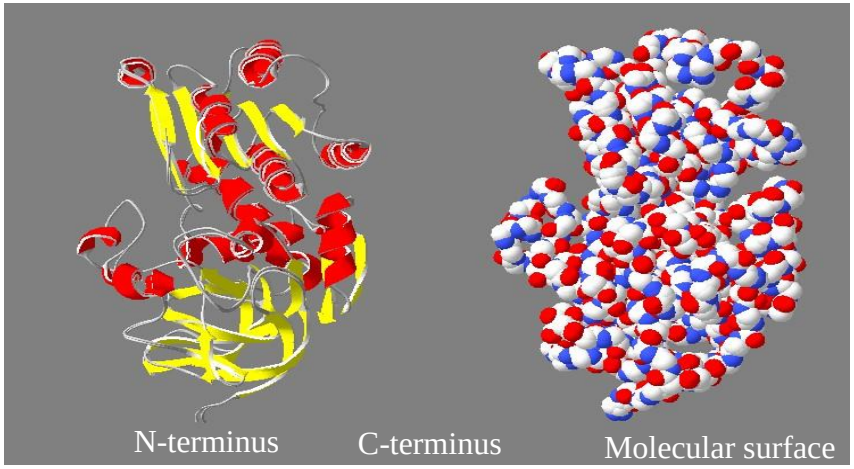


Figure 2.13A: Homology modeled 3D structure of *C. crispus* class III ADH showing the N- and C-terminal regions and also the molecular surface.

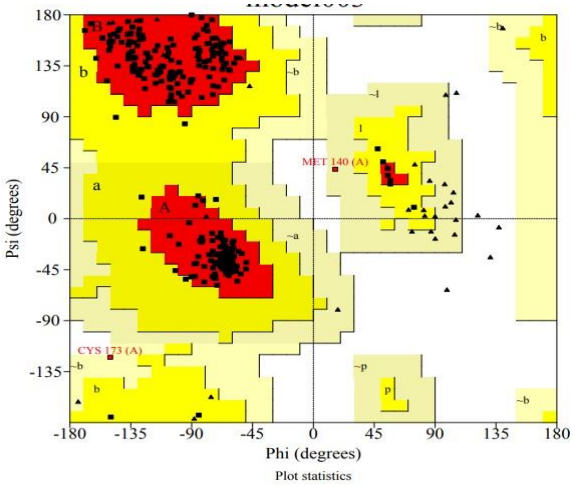


Figure 2.13B: Ramachandran plot showing the phi/psi values of the modeled class III ADH a protein. Red color: most favorable regions, yellow color- additional allowed region, pale yellow: generously allowed region and white color: disallowed regions.

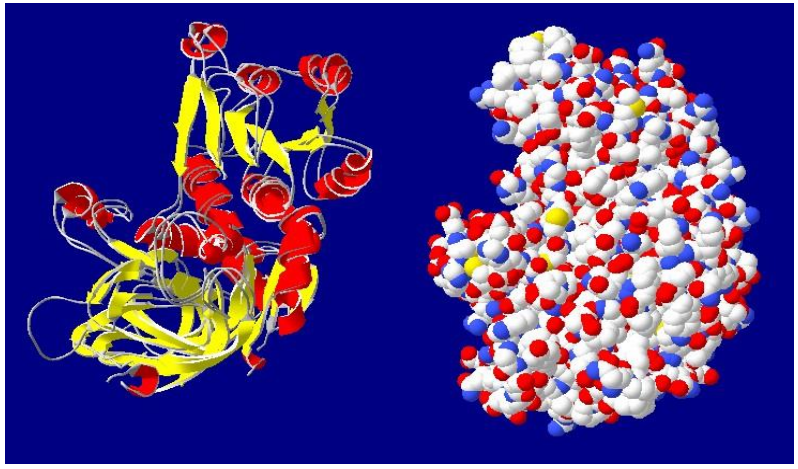


Figure 2.13C: Homology modeled 3D structure of *C. crispus* cinnamily alcohol dehydrogenase showing the N- and C-terminal regions and the molecular surface.

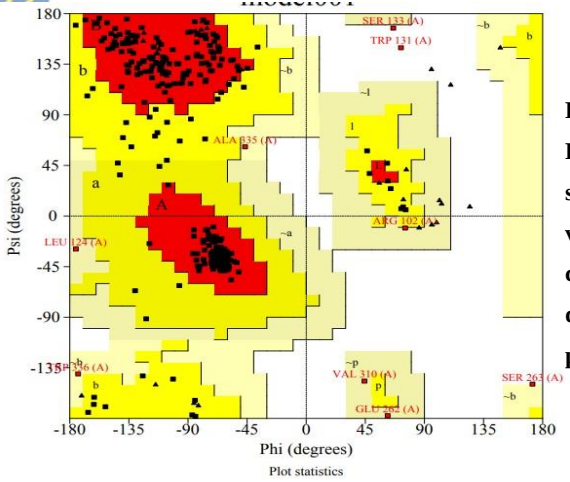


Figure 2.13D: Ramachandran plot showing the phi/psi values of the modeled cinnamyl alcohol dehydrogenase protein.

Figure 2. 13: Tertiary structure analysis of *C. crispus* alcohol dehydrogenases showing the N- and C-terminal regions. The 3D structures in Fig.2.12A and Fig. 2.12C showed ribbon models colored by secondary structure elements such as beta-strand (red), alpha helices (yellow) and random coils (gray). The Ramachandran plots showed in Fig. 2.12B and Fig. 2.12D represent the quality and validation of the model generated.

The overall structures of D-LDH consist of a coenzyme binding domain and substrate binding domains. These two domains were observed in the *C. crispus* D-LDH models and are separated by a deep cleft as observed in other studies (Zhu *et al*., 2015; Kim *et al*., 2014). These two domains were joined together by the linker region which may function as a joint that alters the structural conformation and flexibility during substrate binding (Razeto *et al*., 2002; Kim *et al*., 2014). The 3D ribbon structures of *C. crispus* contained a coenzyme binding domain common in most NAD-dependent dehydrogenase that consisted of six parallel strands surrounded by alpha helices.

This topology was similar to the ones reported for other D-LDHs from other organisms (Kim *et al*., 2014; Zhu *et al*., 2015). The substrate binding domain consisted of five beta strands flanked by five alpha helices and D-LDH from *Sporolactobacillus inulinus* revealed similar topology (Zhu *et al*., 2015). Three amino acid residues His297, Arg237 and Glu266 that have been reported in the literature to be involved in the substrate and catalysis of D-lactate dehydrogenases were located in the centre of the modelled structure. The tertiary structure of *C. crispus* LDH showed an open conformation due to the absence of substrate and cofactor binding residues in the catalytic site. The secondary structure elements in the 3D structure were coloured in yellow for beta-strands, red for alpha helices and grey for random coils as shown in Figure 2.13A.

The quality and validation of the templates and models were analysed by PROCHECK in PRIMO database. The analysis showed that 95% (267 amino acid residues) were found in the most favoured regions, 4.6% (13 amino acid) residues found in the additional allowed region, 4% (1 residue) in generously allowed regions and there were amino acid residues in the disallowed regions. This showed that the modelled 3D structure has stable confirmations. The other LDH protein structure showed that 91.8% (267 amino acids) were found in the most favoured region, 7.9% (23 residues) in the additional allowed region and no amino acid residues were found in the generously allowed regions. One amino acid residue, Glu264 was found in the disallowed region. The Ramachandran plots for the two LDH protein structures for *C. crispus* are shown in Figure 2.13B. The Ramachandran plots determine protein structure and position of amino acids. The quality of the structure modelled is validated using Ramachandran plot and shows allowed and disallowed regions of amino acid showing stability of protein.

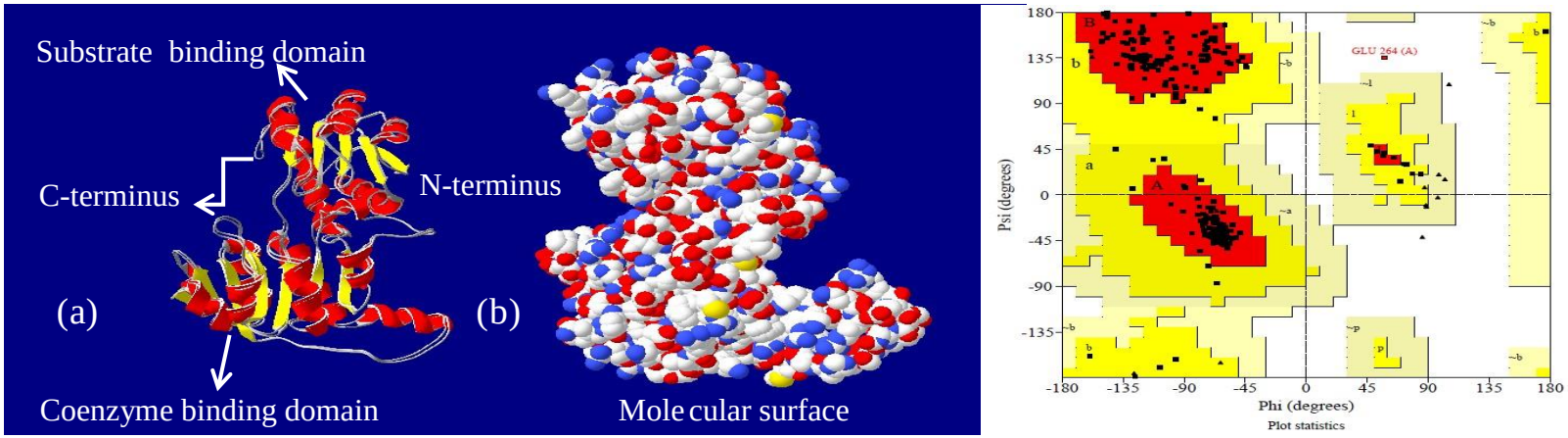


Figure 2. 14: Homology modelling of *C. crispus* LDH proteins using best template from *P. aeruginosa* (3WWZ). The Ramachandran plot showed phi/psi value for the D-LDH protein shown in (a) and (b). Red colour: most favourable regions, yellow colour: additional allowed region, pale yellow: generously allowed region and white colour: disallowed regions.

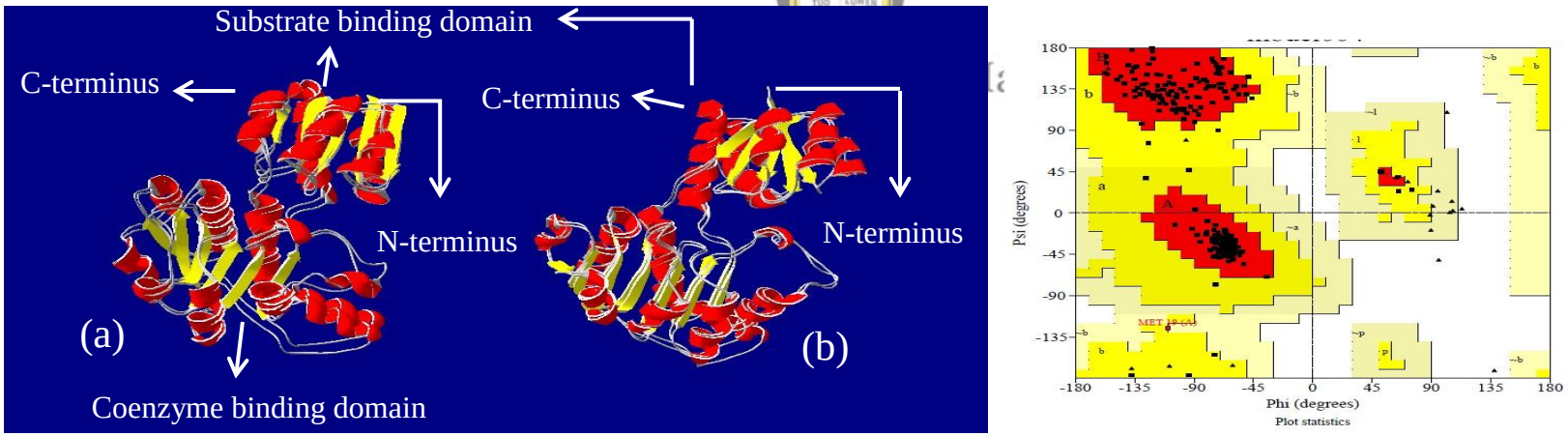


Figure 2. 15: Homology modeling *C. crispus* D-LDH protein using templates from *C. reinhardtii* (4ZGS) for structure A and *P. aeruginosa* (3WWZ). The phi/psi values were determined by the Ramachandran plot to check the validation and quality of the model in (a) and (b).

2.3 Concluding remarks

Two alcohol dehydrogenases and two hypothetical genes that code for D-lactate dehydrogenases were identified in the *Chondrus crispus* genome. The alcohol dehydrogenase proteins belong to the medium-chain dehydrogenase/reductases (MDRs) superfamily, whereas the D-lactate dehydrogenases showed similar characteristics with the members of the D-isomer specific to hydroxyacid dehydrogenase family. The conserved Zinc-binding motifs among the ADH proteins in the catalytic activity showed that these two *C. crispus* are Zinc-dependent alcohol dehydrogenases.

The *C. crispus* LDH proteins showed a highly conserved Rossman-fold binding motif which is present in most NAD-dependent dehydrogenases (Stoll *et al*., 1996). The amino acid residues that are critical for substrate binding and catalysis in ADH formed a catalytic triad (C-H-C) and those involved in lactate dehydrogenase formed (R-E-H) catalytic triad. The phylogenetic tree analysis showed that *C. crispus* ADH proteins belong to class III and cinnamyl alcohol dehydrogenases. The *C. crispus* LDH proteins showed specificity among the D-isomers and showed functional characteristics to D-lactate dehydrogenases.

Many *cis*-acting regulatory elements of the dehydrogenase gene family in *C. crispus* were associated with abiotic stress and hormonal response. The beta-strand rich secondary structure elements were observed in the *C. crispus* dehydrogenase proteins. The structure-based sequence alignments were highly conserved among the ADH and LDH proteins. The 3D structure analysis suggests that *C. crispus* ADH is analogous to *H. sapiens* class III ADH proteins. Homology modelling of *C. crispus* D-LDH 3D structure has substrate binding and cofactor binding domains surrounding a deep cleft containing the active site. These similarities among the dehydrogenase proteins in *C. crispus* suggest similar function and stable structural conformation.

2.4 References

Altschul, S.F., Madden, T.L., Schäffer, A.A., Zhang, J., Zhang, Z., Miller, W. and Lipman, D.J., 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic acids research*, *25*(17), pp.3389-3402.

Arnold, K., Bordoli, L., Kopp, J. and Schwede, T., 2006. The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling. *Bioinformatics*, *22*(2), pp.195-201.

Yang, Z.N., Bosron, W.F. and Hurley, T.D., 1997. Structure of human χχ alcohol dehydrogenase: a glutathione-dependent formaldehyde dehydrogenase. *Journal of molecular biology*, *265*(3), pp.330-343.

Brändén, C.I. and Tooze, J., 1999. *Introduction to protein structure*. Taylor & Francis.

Bukh, C., Nord-Larsen, P.H. and Rasmussen, S.K., 2012. Phylogeny and structure of the cinnamyl alcohol dehydrogenase gene family in Branchypodium distachyon. *Journal of Experimental Botany, 63 (17)*, pp. 6223-6236.

Chan, C.-X., Ho, C.-L., and Phang, S.-M., 2006. Trends in seaweed research. *Trends In Plant Science. 11(4)*, pp. *165-166.*

Chase, T., 1999. Alcohol dehydrogenases: identification and names for gene families. *Plant Molecular Biology Reporter*, *17*(4), pp.333-350.

Collen, J., Herve, C., Guisle-Marsollier, I., Legar, J.J. and Boyen, C., 2006. Expression profiling of Chondrus crispus (Rhodophyta) after exposure to methyl jasmonate. *Journal of Experimental Botany, 57(*14), pp. 3869-3881.

Collen, J., Porcel, B., Carre, W., Ball., S.G., Chaparro, C., Tonon, T., Barbeyron, T., Michel, G., Noel, B., Valentin, K. and Elias, M., 2013. Genomes structure and metabolic features in the red seaweed Chondrus crispus shed light on the evolution of the Archaeplastida. *Proceedings of the National Academy of Sciences, 110*(13), pp. 5247-5252.

Dengler, U., Niefind, K., Kieβ, M. and Schomburg, D., 1997. Crystal structure of a ternary complex of D-2-hydroxyisocaproate dehydrogenase from Lactobacillus casei, NAD+ and 2-oxoisocaproate at 1.9 Å resolution. *Journal of molecular biology*, *267*(3), pp.640-660.

Dolferus, R., Osterman, J.C., Peacock, W.J. and Dennis, E.S., 1997. Cloning of the Arabidopsis and rice formaldehyde dehydrogenase genes: implications for the origin of plant ADH enzymes *Genetics*, *146(3)*, pp. 1131-1141.

Emanuelsson, O., Brunak, S., Von Heijne, G. and Nielsen, H., 2007. Locating proteins in the cell using TargetP, SignalP and related tools. *Nature protocols*, *2*(4), p.953.

Esposito, L., Sica, F., Raia, C.A., Giordano, A., Rossi, M., Mazzarella, L. and Zagari, A., 2002. Crystal structure of the alcohol dehydrogenase from the hyperthermophilic archaeon Sulfolobus solfataricus at 1.85 Å resolution. *Journal of molecular biology*, *318*(2), pp.463-477.

Evolutionary Genetics Analysis version 6.0. *Molecular Biology and Evolution*:30, pp. 2725-2729.

Gasteiger, E., Hoogland, C., Gattiker, A., Wilkins, M.R., Appel, R.D. and Bairoch, A., 2005. Protein identification and analysis tools on the ExPASy server. In *The proteomics protocols handbook* (pp. 571-607). Humana press.

Goldberg, T., Hamp, T. and Rost, B., 2012. LocTree2 predicts localization for all domains of life. *Bioinformatics*, *28*(18), pp.i458-i465.

Goldberg, T., Hecht, M., Hamp, T., Karl, T., Yachdav, G., Ahmed, N., Altermann, U., Angerer, P., Ansorge, S., Balasz, K. and Bernhofer, M., 2014. LocTree3 prediction of localization. *Nucleic acids research*, *42*(W1), pp.W350-W355.

Hagen, G. and Guilfoyle, T., 2002. Auxin-responsive gene expression: gene, promoters and regulatory factors. *Plant Molecular Biology*, *49(3-4)*, pp. 373-385.

Hatherley, R., Brown, D.K., Glenister, M. and Bishop, Ö.T., 2016. PRIMO: an interactive homology modeling pipeline. *PloS one*, *11*(11), p.e0166698.

Hernandez-Garcia, C.M. and Finer, J.J., 2014. Identification and validation of promoters and cis-acting regulatory elements. *Plant Science*, *217*, pp.109-119.

Höglund, A., Dönnes, P., Adolph, H.W. and Kohlbacher, O., 2005. From prediction of subcellular localization to functional classification: discrimination of DNA-packing and other nuclear proteins. *Online J. Bioinformatics*, *6*, pp.51-64.

Horton, P., Park, K.J., Obayashi, T., Fujita, N., Harada, H., Adams-Collier, C.J. and Nakai, K., 2007. WoLF PSORT: protein localization predictor. *Nucleic acids research*, *35*(suppl_2), pp.W585-W587.

Jiao, Y., Lau, O.S. and Deng, X.W., 2007. Light-regulated transcriptional networks in higher plants. *Nature Reviews Genetics*, *8*(3), p.217.

Jin, Y., Zhang, C., Liu, W., Qi, H., Chen, H. and Cao, S., 2014. The cinnamyl alcohol dehydrogenase gene family in melon (Cucumis melo L.): bioinformatics analysis and expression patterns. *PloS ONE, 9(7)*, p.e101730.

Jin, Y., Zhang, C., Liu, W., Tang, Y., Qi, H., Chen, H. and Cao, S., 2016. The alcohol dehydrogenase gene family in melon (Cucumis melo L.): bioinformatics analysis and expression patterns. *Frontiers in Plants Science, 7*, pp.670.

Johansson, M.U., Zoete, V., Michielin, O. and Guex, N., 2012. Defining and searching for structural motifs using DeepView/Swiss-PdbViewer. *BMC bioinformatics*, *13*(1), p.173.

Kelley, L.A., Mezulis, S., Yates, C.M., Wass, M.N. and Sternberg, M.J., 2015. The Phyre2 web portal for protein modeling, prediction and analysis. *Nature protocols*, *10*(6), p.845.

Kim, C. and Lee, B., 2007. Accuracy of structure-based sequence alignment of automatic methods. *BMC bioinformatics*, *8*(1), p.355.

Kim, S., Gu, S.A., Kim, Y.H. and Kim, K.J., 2014. Crystal structure and thermodynamic properties of D-lactate dehydrogenase from Lactobacillus jensenii. *International Journal of Biological Macromolecules, 68,* pp. 151-157.

Knoll, M. and Pleiss, J., 2008. "The Medium Chain Dehydrogenase/Reductase Engineering Database: A systematic analysis of a diverse protein family to understand sequence-structurefunction relationship" *Protein Science,* 17(10), pp. 1689-1697.

Larkin, M.A., Blackshields, G., Brown, N.P., Chenna, R., McGettigan, P.A., McWilliam, H., Valentin, F., Wallace, I.M., Wilm, A., Lopez, R. and Thompson, J.D., 2007. Clustal W and Clustal X version 2.0. *bioinformatics*, *23*(21), pp.2947-2948.

Lescot, M., Déhais, P., Thijs, G., Marchal, K., Moreau, Y., Van de Peer, Y., Rouzé, P. and Rombauts, S., 2002. PlantCARE, a database of plant cis-acting regulatory elements and a portal to tools for in silico analysis of promoter sequences. *Nucleic acids research*, *30*(1), pp.325-327.

Ma, Q.H., 2010. Functional analysis of a cinnamyl alcohol dehydrogenase involved in lignin biosynthesis in wheat. *Journal of Experimental Botany*, *61*(10), pp.2735-2744.

Nakamura, Y., Sasaki, N., Kobayashi, M., Ojima, N., Yasuike, M., Shigenobu, Y., Satomi, M., Fukuma, Y., Shiwaku, K., Tsujimoto, A. and Kobayashi, T., 2013. The first symbiont-free genome sequence of marine red alga, Susabi-nori (Pyropia yezoensis). *PloS one*, *8*(3), p.e57122.

Pei, J. and Grishin, N.V., 2014. PROMALS3D: multiple protein sequence alignment enhanced with evolutionary and three-dimensional structural information. In *Multiple Sequence Alignment Methods* (pp. 263-271). Humana Press, Totowa, NJ.

Pei, J., Kim, B.H. and Grishin, N.V., 2008. PROMALS3D: a tool for multiple protein sequence and structure alignments. *Nucleic acids research*, *36*(7), pp.2295-2300.

Petersen, T.N., Brunak, S., von Heijne, G. and Nielsen, H., 2011. SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nature methods*, *8*(10), p.785.

Raj, S.B., Ramaswamy, S. and Plapp, B.V., 2014. Yeast alcohol dehydrogenase structure and catalysis. *Biochemistry*, *53*(36), pp.5791-5803.

Razeto, A., Kochhar, S., Hottinger, H., Dauter, M., Wilson, K.S. and Lamzin, V.S., 2002. Domain closure, substrate specificity and catalysis of D-lactate dehydrogenase from Lactobacillus bulgaricus. *Journal of Molecular Biology, 318(1),* pp. 109-119.

Rombauts, S., Déhais, P., Van Montagu, M. and Rouzé, P., 1999. PlantCARE, a plant cis-acting regulatory element database. *Nucleic acids research*, *27*(1), pp.295-296.

Rossmann, M.G., Moras, D. and Olsen, K.W., 1974. Chemical and biological evolution of a nucleotide-binding protein. *Nature, 250(5463),* pp.194.

Sanghani, P.C., Robinson, H., Bosron, W.F. and Hurley, T.D., 2002. Human glutathione-dependent formaldehyde dehydrogenase. Structures of apo, binary, and inhibitory ternary complexes. *Biochemistry*, *41*(35), pp.10778-10786.

Stoll, V.S., Kimber, M.S. and Pai, E.F., 1996. Insights into substrate binding by D-2-ketoacid dehydrogenases from the structure of Lactobacillus pentosus D-lactate dehydrogenase. *Structure,4*(4), pp. 437-447.

Strommer, J., 2011. The plant ADH gene family. *The Plant Journal*, *66*(1), pp.128-142.

Taguchi, H. and Ohta, T., 1991. D-lactate dehydrogenase is a member of the D-isomer-specific 2hydroxyacid dehydrogenase family. Cloning, sequencing, and expression in Escherichia coli of the D-lactate dehydrogenase gene of Lactobacillus plantarum. *Journal of Biological Chemistry, 266(19)*, pp.12588-12594.

Tamura, K., Stecher, G., Peterson, D., Filipski, A. and Kumar, S. (2013) MEGA6: Molecular

Tirichine, L. and Bowler, C., 2011. Decoding algal genomes: tracing back the history of photosynthetic life on Earth. *The Plant Journal*, *66*(1), pp.45-57.

Wilks, H.M., Halsall, D.J., Atkinson, T., Chia, W.N., Clarke, A.R. and Holbrook, J.J., 1990. Designs for broad substrate specificity keto acid dehydrogenase. *Biochemistry*, *29*(37), pp. 85878591.

Zhu, L., Xu, X., Wang, L., Dong, H. and Yu, B. 2015. The D-lactate dehydrogenase from Sporolactobacillus insulinus also possessing reversible deamination activity. *PloS ONE, 10(9)*, p.e0139066.

CHAPTER 3: Molecular characterization of alcohol dehydrogenases (ADHs) from red algae

3.1 Introduction

The molecular characterization of proteins requires the use of many techniques, including extraction of high quality DNA or RNA, gene amplification by PCR or RT-PCR, cloning and expression (Wink, 2006). The integrity of the extracted nucleic acid is essential because it increase the chances of a successful downstream application (Cseke *et al*., 2004). The RNA, compared to DNA, is very sensitive and easily exposed to degradation because it is a single-stranded molecule (Kojima and Ozawa, 2002). The polymerase chain reaction (PCR) is a technique used to amplify and replicate DNA copies. The reverse-transcriptase polymerase chain reaction (RT-PCR) is used to reverse transcribe RNA to its complementary strand using reverse transcriptase enzyme. Gel electrophoresis is a technique that is used to separate samples based on their charge and molecular weight.

The success of any molecular biology lab experiment depends on the quality of the nucleic acid of the desired product. The nucleic acid extracted from the marine algal sources has been reported to be of poor quality because of various secondary metabolites that interfere with the extraction (Ho e*t al*., 1996; Gehrig *et al*., 2000). Eukaryotic genes contain both introns and exons and it is difficult to splice the introns without the messenger RNA (mRNA). Amplification of most eukaryotic genes requires extraction of full-length mRNA and then generates a cDNA for other downstream applications.

The primary interest in studying alcohol dehydrogenases in red algae is that the study of ADH gene-enzymes is still new in the red algae species, and the molecular and functional characteristics of these ADH genes are mostly unknown. The candidate gene for ADH has been annotated in the *C. crispus* (CHC_T00010066001); however, the functional characteristic of the predicted protein is unknown. If the gene encoding the predicted protein has functional ADH, the structure, kinetics and biological functions need to be described. Thus, we isolated the mRNA for molecular characterization of the ADH gene in *C. crispus*.

The aim of this chapter is to evaluate various DNA and RNA extraction methods to find the optimal method for amplification of the alcohol dehydrogenase (ADH) gene, and further analyse its functional and molecular characterization properties.

3.2 Methods

3.2.1 Sample collection

Red algae samples were collected during 2017 at Kenton-On-Sea (33°41'44.4"S 26°40'00.3"E

-33.695657, 26.666738). Samples were collected by hand and washed with distilled water and stored at -80 ºC freezer until nucleic acid extraction.

3.2.2 DNA and RNA extraction

The red algae *(Gelidium capense*) material was isolated using Promega SV total RNA isolation kit following manufacturer's instructions. Sample of 180 mg, 250 mg, 280 mg and 300 mg were frozen in liquid nitrogen and ground to a fine powder using a mortar and pestle. The DNA was isolated using the ZR Plant/Seed DNA MiniPrep Kit (The Epigenetics Company) according to manufacturer's instructions. The red algae sample of 300 mg frozen in liquid nitrogen was ground to a fine powder using mortar and pestle and placed in a ZR bashing bead™. The integrity of the nucleic acid was determined by running 5 µl red algae sample on 1% agarose gel.

3.2.3 PCR amplification of the ADH gene using RT-PCR

Amplification of the ADH gene was performed using the Qiagen® OneStep Ahead RT-PCR kit (Qiagen).

3.2.2.1 Oligonucleotide primers

A set of primers (forward and reverse) was designed by the gene runner tool to amplify the region of 1184 bp that encodes alcohol dehydrogenase in *Chondrus crispus* (**Table 3.1**). The 10 µM working solution of the forward and reverse primers were prepared by diluting the 100 µM stock solution with nuclease-free water as indicated in the Inqaba Biotech sheet supplied with the primers.

Table 3. 1: RT-PCR oligonucleotide primers (ordered from Inqaba Biotechnical Industries)

Primer	Sequence	Tm (°C)	Primer length	GC %
Forward	ATGCCACCAACTCTGAAC	57.62	18 bp	50
Reverse	CTACCACTCTAGCACATATC	58.35	20 bp	45

3.2.2.2 Reaction mixture

The reaction mixture was prepared as shown in Table 3.2 below in a thin-walled PCR tube.

Table 3. 2: Reaction components and setup for OneStep RT-PCR

Component	Volume/reaction	Final concentration
Reaction mix	10 µl	1X
OneStep ahead RT-PCR	10 µl	_
Master mix	10 µl	_
OneStep ahead RT-PCR mix	1 µl	1X
Primer (forward)	1.25 µl	0.5 µM
Primer (reverse)	1.25 µl	0.5 µM
RNase-free water	Variable	_
Template RNA	variable	0.1 pg-1µg
Total reaction volume	25 µl	1 reaction

In order to make the working solution, the volumes per reaction in Table 3.2 were multiplied by 3. Three different concentrations of the template were amplified by using 0 µl, 5µl or 10 µl of the RNA template in the reaction mix, as summarized in **Table 3.3**.

Table 3. 3: RT-PCR setup using different concentrations of the sample

Components	Concentrations		
	0 µl	5 µl	10 µl

Master mix (43.5µl)	13.5 µl	13.5 µl	13.5 µl
RNase-free water	11.5 µl	6.5 µl	1.5 µl
RNA template	0 µl	5 µl	10 µl
Total volume/reaction	25 µl	25 µl	25 µl

3.2.2.3 Temperature cycling conditions

The RT-PCR reactions were performed using the parameters indicated in **Table 3.4** below, using a MyCycler thermal cycler from Bio-Rad.

Table 3. 4:Thermal cycling conditions for RT-PCR amplification

Step	time	temperature	Number of cycles
Reverse transcription	15 minutes	45 °C	1
Initial denaturation	5 minutes	95 °C	1
3-step cycling			
denaturation	10 seconds	95 °C	40
Annealing	15 seconds	55 °C	40
Extension	1 minute	72 °C	40
Final extension	2 minute	72 °C	1
Keep	indefinite	4 °C	1

3.2.2.4 Standard PCR procedure

The standard PCR reaction mixture (**Table 3.5**) was prepared for amplification of the ADH gene from red algae using the Takara Kit (Takara Bio Inc.).

Table 3. 5: Reaction setups for standard PCR

General reaction mixture	50 µl total volume	25 µl total volume	4 x 25 µl rxn

TakaRa EX Taq (5units/µl)	0.25 µl	0.125 µl	0.5 µl
10X EX Taq buffer	5 µl	2.5 µl	10 µl
dNTP mixture (2.5mM)	4 µl	2 µl	8 µl
Template	<500 ng	0 µl, 5 µl, and 10 µl	
Forward primer	0.2-1.0 µM	0.25 µl	1 µl
Reverse primer	0.2-1 µM	0.25 µl	1 µl

Volumes of water to be added in each of the volume/reaction (50µl and 25µl)

DNA to be added with 25µl total volume

0µl DNA = 19.875µl

5µl DNA = 14.875µl

10µl DNA = 9.875µl

Primer volume to be added calculations

C1V1 =C2V2

C1= 100 µM, C2=1 µM, V2=25µl

V1= C2V2/C2 = 1X25/100

V1= 0.25 µl

Table 3. 6:Thermal cycling conditions for standard PCR

Step	Temperature	Time	Cycle
Initial denaturation	98 °C	1 minute	1
Denaturation	98 °C	30 seconds	30
Annealing	55 °C	1 minute	

Extension	72 °C	1 minute	
Final extension	72 °C	1 minute	1

3.2.2.5 The gradient PCR

The PCR components mixture was prepared for gradient PCR using the reaction mixture as follows; Master Mix = 8 µl

Forward primer = 1 µl

Reverse primer = 1 µl

DNA sample = 1 µl dH20

= 4 µl

Total volume = 15µl

Gradient PCR was performed using the PCR conditions in **Table 3.7**.

Table 3. 7: Gradient PCR conditions

Step	Temperature	Time	Cycle
Initial denaturation	95 °C	3 minutes	1X
Denaturation	95 °C	1 minute	40X
Annealing	45-65 °C	30 seconds	
Extension	72 °C	1 minute	
Final extension	72 °C	10 minutes	1
Standby (Hold)	4 °C	indefinite	1

3.2.4 Agarose gel electrophoresis of the RT-PCR products

Agarose gel electrophoresis was run from the RT-PCR, standard PCR and gradient PCR product results. The PCR products were resolved at 100 V for 45 minutes on a 1% agarose gel prepared from TAE buffer (Tris-acetic acid and EDTA). About 0.5 µg/ml ethidium bromide was used in the gel mixture to stain. The mixture of 2 µl of 6X DNA gel loading buffer and 10 µl sample was visualized in the transilluminator system.

3.3 Results and Discussion

3.3.1 RNA and DNA isolation

The various extraction conditions were used to optimize the yield of ADH extracts from the red algae samples. The total RNA was isolated using the Promega SV total RNA isolation kit. The 1% agarose gel showed two faint bands (**Figure 3.1**). These two bands are observed only in the starting material in algae sample of 300 mg. The present of these two faint bands may indicate lower yield of the RNA sample in the red algae. Although the marker was not used in the gel, most intact RNA samples reported in algae are indicated by two separate bands of 18S and 28S (Chan *et al*., 2002; Pearson *et al*., 2006).The poor yield of RNA in algae is reported in literature (Ho *et al*., 1996; Kim *et al*., 1997 Gehrig *et al*., 2000). The good quality RNA from algae is reported when CTAB method and LiCl precipitation are used in the RNA extraction (Chan *et al*., 2002; Pearson *et al*., 2006).

The cell wall of most seaweed is rich in secondary metabolites and it is hard to breakdown the cell wall because most of thes metabolites interfere with extraction (Falcao *et al*., 2008). The other challenge in the extraction of RNA in seaweed is that some methods used can be strain-and species-specific (Chan *et al*., 2004). The use of CTAB method could have been explored to reduce or overcome the impact of some of the secondary metabolites. However, the chemicals for this method are expensive and time consuming (Lee and Lee, 2003).

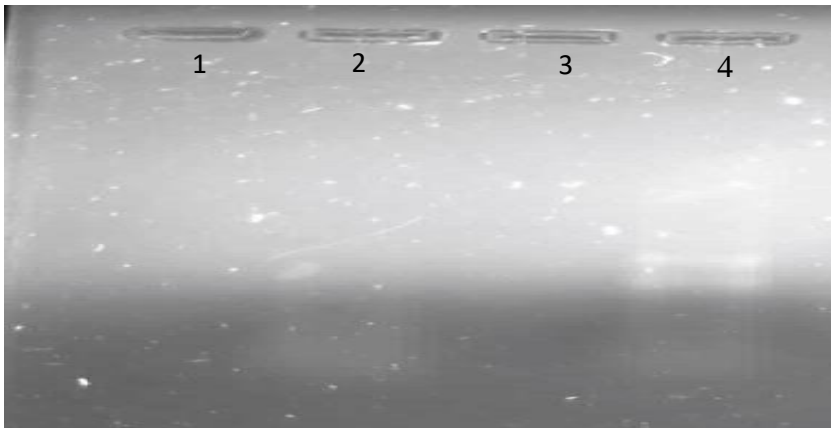


Figure 3. 1: 1% agarose gel showing band for Red algae RNA sample extracted from Gelidium species.

The RNA samples were loaded on a 1% agarose gel as follows: lane 1:180 mg RNA sample, lane 2: 250 mg RNA sample, and lane 3: 280 mg and lane 4:300 mg RNA sample.

The DNA sample was used in an attempt to amplify alcohol dehydrogenase and two bright bands were obtained from 1% agarose gel electrophoresis (**Figure 3.2**). These two bright, intact DNA bands that appeared on 1% agarose showed that DNA isolation in red algae is more stable than that of RNA. The DNA marker was not used and the bands showed are from one species. The two bands indicated that the DNA isolation in algae is much more stable than RNA. The DNA was used for standard PCR to amplify the region of the desired gene. A high yield of pure DNA is required for molecular biology in red algae. The high-quality of DNA from algae has been reported in the literature (Lee and Lee, 2003). The DNA samples were used in the analysis for standard PCR.

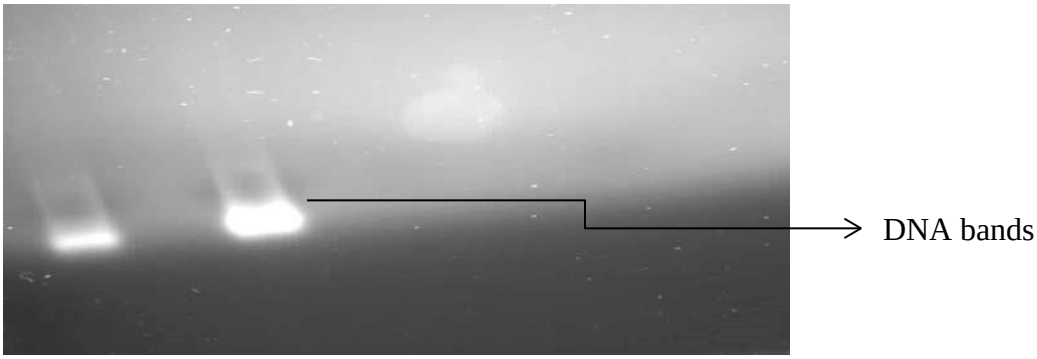


Figure 3. 2: DNA sample extracted from red algae material and visualized on 1% agarose gel electrophoresis. The DNA bands observed in the gel were loaded in lane 1 and lane 3 using the same sample.

An attempt to amplify the region of 1184 bp, encoding alcohol dehydrogenase (ADH) from the red algae, was performed using the RT-PCR. A set of primers was designed based on the open

reading frame (ORF) of the alcohol dehydrogenase gene. From the results shown in **Figure 3.3**, the attempt to amplify the ADH-gene in red algae was unsuccessful. There were no detectable bands observed after running 1% agarose gel electrophoresis. There seem to be a problem with the DNA ladder as shown in Fig. 3.3. The ladder may be overloaded and it is hard to separate. The amplification of ADH- gene was not achieved and the gel showed no detactable bands.

The primer sets used in the amplification of ADH-gene were designed from the genome of red algae, *Chondrus crispus* and were used to amplify the ADH-gene from other red algae species found in the Algoa Bay Region. The various conditions were optimized to evaluate the optimal PCR conditions that can be used for the amplification of ADH-gene in *C. crispus*. The ADH-gene was not detected in the RT-PCR analysis of the total RNA isolated from red algae species.

The universal DNA ladder was used as a marker, but it appears as if there was a problem with the separation of the DNA marker used. The reason it did not separate may be the high volume of the DNA marker loaded and as a result, the time used to run the samples on 1% agarose gel might not have been enough for the marker to completely separate. Molecular techniques such as RT-PCR are becoming a challenge in analyzing marine algae-derived enzymes, as shown by the various unsuccessful attempts to amplify the ADH gene from the red algae samples used in this study.

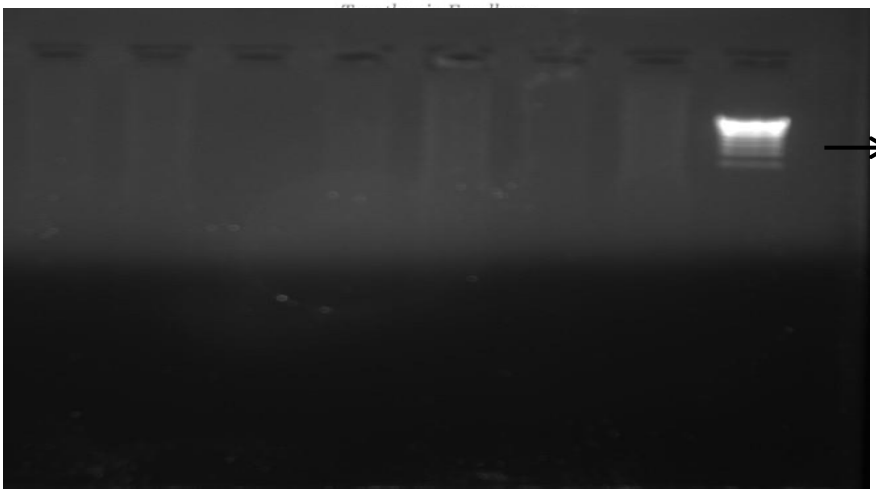


Figure 3. 3: The analysis of the RT-PCR amplification by 1% agarose gel electrophoresis.

The standard PCR was then used as an alternative route to try and amplify the ADH-gene using the DNA samples as shown in **Figure 3.4**. The attempted amplification using standard PCR method was unsuccessful, and only dimers appeared on 1% agarose gel.

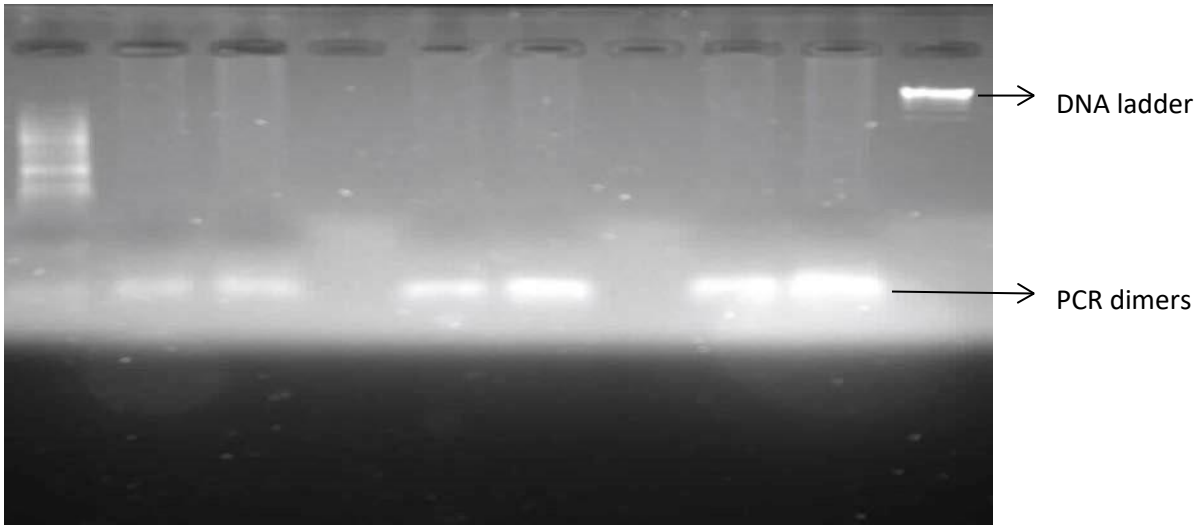


Figure 3. 4: Visualization of standard PCR products on 1% agarose gel electrophoresis.

The red algae samples laoded using different concentration of the DNA showed in lane 2 and 3 are 0 µl DNA template, lane 5 and 6 its 5 µl DNA template and lane 8 and 9 is 10 µl.

The ADH-gene from *Chondrus crispus* (1184 bp) could not be amplified from the red algal species obtained from Kenton-On-Sea. The ADH-genes have been amplified and characterized in many other organisms; however, the use of the red algae species for such a relatively important investigation did not produce any significant results due to unspecific primer sets used. The primers that were used in this experiment were designed based on the open reading frame of the *C. crispus* genome, and that may be one of the reasons the ADH-gene could not be amplified when tested from other red algae species.

There is a considerable lack of genome information on the local species of red algae, and as a result, the *C. crispus* was an ideal organism for this research using bioinformatics to analyse and interpret the ADH-like genes, including the one used for amplification in this study. The use of the primers, (designed based on the conserved amino acid regions among several red algal species), could have been used to amplify different DNA fragments. This approach has been successfully used with ADH-like genes in several yeast organisms according to Lertwattanasakul *et al*. (2014). This approach, although it is expensive, is reproducible and it could provide insight into organisms with relatively similar amino acids residues. The gradient PCR for single sample perfomed using different annealing temperatures did not produce the desired product of 1184 bp length. The sets of primers used in this section were designed using Chondrus crispus sequence and used to try to amplify ADH gene in Gelidium species. The results showed some faint band

but the size of the bands in the gel did not correspond to the ADH gene size. The non-binding of the primers could be the result of the difference in codon usage. However, work need to be done to find evidence of codon usage between red algae species.

Figure 3. 5: The 1% agarose gel electrophoresis visualizing the products of the gradient PCR.

From right to left in the diagram showing gradient PCR products (lane 1: DNA ladder; lane 2-9 is the samples at different annealing temperatures 63, 62, 60.4, 57.7, 54.8, 52.5, 50.9, and 50.0 °C respectively).

3.4 Concluding remarks

The ADH gene could not be amplified due to the unspecific primers tested in the red algae samples collected along the Kenton-On-Sea region, in Port Alfred. The red algae model organism *C. crispus* ADH coding region was used to design the primers and the non-binding of these primers could be due to the difference in the codon usage or DNA sequence of the *Gelidium capense* genome as compared to the *C. crispus* genome. We therefore undertook a study to directly extract the protein from red algae (*Gelidium capense*) as our wet bench approach to investigate whether these dehydrogenase proteins have activity in red algae. However, the alcohol dehydrogenase did not show any activity when primary alcohol like ethanol, propanol, methanol and butanol were used as substrate. It was therefore decided to directly extract and purify the lactate dehydrogenase enzyme from red algae using biochemical techniques to understand the

structure and function and to apply several analytical methods to check the presence and purity of the LDH (Chapter 4).

3.5 References

Chan, C.X., Ho, C.L., Othman, R.Y. and Phang, S.M., 2002, January. Total RNA extraction for the red seaweed Gracilaria changii (Gracilariales, Rhodophyta). In *Asia-Pacific Conference in Marine Science and Technology* (Vol. 5, pp. 12-16).

Chan, C.X., Teo, S.S., Ho, C.L., Othman, R.Y. and Phang, S.M., 2004. Optimisation of RNA extraction from Gracilaria changii (Gracilariales, Rhodophyta). *Journal of Applied Phycology*, *16*(4), pp.297-301.

Cobben, N.A.M., Drent, M., Schols, A.M.W.J., Lamers, R.J.S., Wouters, E.F.M. and Van Dieijen-Visser, M.P., 1997. Serum lactate dehydrogenase and its isoenzyme pattern in ex-coalminers. *Respiratory medicine*, *91*(10), pp.616-623.

Collen, J., Cornish, M.L., Craigier, J., Ficko-Bean, E., Herve, C. Krueger-Hadfield, S.A., *et al*., 2014. "*Chondrus crispus*. A present and historical model organism for red seaweeds," in *Advances in Botanical Research, 71, Sea Plants,* ed. Bourgougnon N., editor. (London: Elsevier ;), 53-89.

Collen, J., Porcel, B., Carre, W., Ball, S.G., Chaparro, C., Tonon, T., *et al.* (2013) Genome structure and metabolic features in the red seaweed *Chondrus crispus* shed light on the evolution of the Archeaplastida. *Proceedings of the National Academy of Sciences. U.S.A. 26,* pp. *52475252*.

Cseke, L.J., Kaufman, P.B., Podila, G.K., Tsai, C.J., 2004. Handbook of molecular and cellular methods in biology and medicine. 2nd ed. Boca Raton: CRC Press.

Falcao, V.D.R., Tonon, A.P., Oliveira, M.C., and Colepicolo, P., 2008. RNA isolation method for polysaccharide rich algae: agar producing *Gracilaria tenuistipitata* (Rhodophyta). *Journal of Applied Phycology. 20,* pp. *9-12.*

Gehrig, H.H., Winter K., Cuishman J., Borland, A., and Taybi, T., 2000. An improved RNA isolation method for succulent plant species rich in polyphenols and polysaccharides. *Plant Molecular Biology Reporter. 18,* pp.*369-376.*

Ho C.L., Phang, S.M., Sinnappah, N.D., and Pang, T., 1996. Molecular approaches in the taxonomy of the red and brown seaweeds. In: Chaudhary BR and Agrawal SB (Eds.). Cytology, Genetics and Molecular Biology of algae. Amsterdam: SPB Academic Publishing. p.351-362.

Kim, S.H., Lee, Y.K., Hong, C.B. and Lee, I.K., 1997. A simple method for RNA extraction from marine macroalgae. *Algae*, *12*, pp.53-56.

Kojima, K., and Ozawa, S., 2002. Method for isolating and purifying nucleic acids. United States patent: US 2002/0192667

Lee, J., Yang, E.C., Graf, L., Yang, J.H., Qiu, H., Zelzion, U., Chan, C.X., Stephens, T.G., Weber, A.P., Boo, G.H. and Boo, S.M., 2018. Analysis of the draft genome of the red seaweed Gracilariopsis chorda provides insights into genome size evolution in Rhodophyta. *Molecular biology and evolution*, *35*(8), pp.1869-1886.

Lee, Y.G. and Lee, H.G., 2003. A simple method for DNA extraction from red algae. *Algae*, *18*(1), pp.65-69.

Lee, Y.K. and Lee, H.K., 2003. A Simple Method for DNA Extraction from Red Algae. *Algae, 18(1),* pp. *65-69.*

Lertwattanasakul, N., Sootsuwan, K., Limtong, S., Thanonkeo, P. and Yamada, M., 2014. Comparison of the gene expression patterns of alcohol dehydrogenase isozymes in the thermotolerant yeast *Kluyveromyces marxianus* and their physiological functions. *Bioscience, Biotechnology, and Biochemistry, 71(5),* pp. *1170-1182.*

Maes, M. and Messens, E., 1992. Phenol as grinding material in RNA preparations. *Nucleic acids research*, *20*(16), p.4374.

Wink, M., 2006. An introduction to molecular biotechnology: molecular fundamentals, methods and applications in modern biotechnology. Weinheim: Wiley-VCH.

CHAPTER 4: Isolation, purification and kinetic characterization of lactate dehydrogenase (LDH) from red algae collected at Kenton-On-Sea region.

4.1 Introduction

The lactate dehydrogenase is one of the essential enzymes occurring in most organisms. LDH is mostly found in the cytoplasm and is made up of one or two of five isoenzymes (Vanderlinde, 1985; Javed *et al*., 1997). This enzyme catalyzes the reversible reaction between pyruvate and lactate using NADH or NAD^+ as coenzymes (Azmat *et al*., 2013). LDH assay is measured using spectrophotometry by observing a change in absorbance of NADH at 340 nm over time. The lactate-to-pyruvate reaction is produced by an increase in NADH at 340 nm, and the pyruvate-to-lactate direction produces a decrease in NADH at 340 nm (Vanderlinde, 1985). The kinetic assay systems for the measurement of LDH reaction in both directions are well investigated, and this study focused on red algae as a source of LDH enzyme.

LDH enzyme from animals, fungi and bacteria has been published in the literature (Sommer *et al*., 1985; Javed *et al*., 1997; Nadeem *et al*., 2011). The LDH enzyme exists as a tetramer with subunit molecular weight of 35 to 36 kDa (Javed *et al*., 1997; Al-Jassabi, 2002). The LDH enzyme in these sources has advantages which include its abundance, high specific activity, well-investigated kinetics, inexpensive and straightforward assays (Wolf, 1988). In contrast, very few studies have been reported in plants because alcohol is the main product of plants under anaerobic conditions (Oba *et al*., 1977; Hassan *et al*., 2013). The lactate dehydrogenase activity in plants has been investigated. However, the function and other properties of this enzyme in plants remain poorly understood as compared to bacteria, animals and fungi (O'Carra and Mulcahy, 1996).

The value of the red algae as a source of a protein depends on the yields and functional properties. It has also been shown in the literature that in order to achieve nutraceuticals, pharmaceutical and cosmeceutical properties it is essential to preserve protein function (Polikovsky *et al*., 2016). Marine red algae are the untapped source of enzymes, particularly lactate dehydrogenases, so a broader exploration of the potential of the red algae is considered necessary. Lactate dehydrogenase attracted considerable interest since it may be used as a valid therapeutic target for diseases (Hassan *et al*., 2013). Proteins are purificed from a variety of species based on their size, solubility, charge, and binding capacity (Berg e*t al*., 2002). A mixture of proteins is exposed to a

series of separations to obtain pure protein. Affinity chromatography is one of the most effective techniques used to purify proteins. Affinity chromatography works by passing a sample through a column and proteins with binding affinity for specific chemical group will bind and retained, while the unbound proteins will be washed off the column. This techniques works well for several dehydrogenase proteins such as lactate dehydrogenase as reported in literature (O'Carra and Barry, 1972; Javed *et al*., 1997; Karamanos, 2014). This chapter deals with the extraction, purification and characterization of lactate dehydrogenase from red algae found in the Kenton-On-Sea region, Port Alfred in the Eastern Cape of South Africa. Several important biochemical methods including spectrophotometry, chromatography, centrifugation and electrophoresis will be used to describe the biochemical properties of LDH from red algae.

4.2 Methods

4.2.1 Sample preparation and extraction of crude protein

Gelidium capense samples were collected by hand from submerged rocks on the coast near the Kenton-on-Sea beach. After collection, the samples were washed with sea water before packing them into plastic bags. The samples were then transported to the laboratory in plastic bags stored in big containers filled with ice. In the laboratory, epiphytes and rock debris were removed from the samples before rinsing them with distilled water multiple times to remove surface salts. The samples were then stored in a -80 ℃ freezer until the day of extraction.

For extraction of proteins, 75 g of *Gelidium capense* was weighed out and ground to a fine powder with a mortar and pestle containing liquid nitrogen. The powder was added to 150 ml extraction buffer in (100 mM Tris-HCl, pH 7.5 containing 0.2 mM $MgCl_2$, 2 mM DTT, and 10% glycerol) and on ice for 1 hour with occasional shaking and then centrifuged at 4600 x g for 20 minutes at 4 °C. The supernatant (crude extract) was retrieved, and the total volume was recorded. A 5 ml sample was removed for total protein, LDH activity assay and SDS-PAGE analysis and the balance of the crude extract processed further.

4.2.2 Ammonium sulphate precipitation

The crude protein sample was concentrated by adding ammonium sulphate [$(NH_4)_2SO_4$]. The precipitation was done by slowly adding finely ground ammonium sulphate to reach 70% saturation, with continuous stirring at 4 ℃ for 30 minutes. The 70% ammonium sulphate solution

was allowed to stand on ice for 1 hour, with continuous stirring. The precipitate was collected by centrifugation in a Beckman Coulter, Avanti® J-E Centrifuge at 4600 x g for 20 min. at 4 °C using an SS-34 Fixed Angle Rotor. The supernatant was decanted into a clean tube, and the total volume recorded before 5 ml was removed for total protein, LDH activity assay and SDS-PAGE analysis.

The remaining supernatant was dialysed as described in the next section.

4.2.3 Dialysis

The pellet was re-suspended into 20 ml dialysis buffer (100 mM Tris-HCl pH 7.5 containing 1 mM 2-mercaptoethanol) and placed into a dialysis bag and dialyzed two times in 1 L of dialysis buffer for 16-24 hours at 4 °C. The dialysate was transferred into a clean tube, and 2 ml sample was removed for total protein, LDH activity assay and SDS-PAGE analysis and the remaining total volume was recorded.

4.2.4 Affinity chromatography

The HiTrap™ 5 ml Blue HP column was connected, at 4 °C, to the AKTA FPLC system which is controlled by the Unicorn software version 5.1. The preservatives were removed by equilibrating the column with 5-10 column volumes of binding buffer (20 mM Tris-HCl, pH 8.6) at a flow rate of 5 ml/minute. The dialysate extract of different volumes (0.2-2 ml) was applied onto the affinity column at a flow rate of 1 ml/minute and washed with 5-10 column volumes of equilibration buffer. The fall-through peak was collected in a beaker, and the eluate obtained using a salt gradient over five column volumes of elution buffer (20 mM Tris-HCl, pH 8.6 containing 2M NaCl). The eluate was collected in 2 ml fractions. The collected fractions were assayed for LDH activity, and those containing activity were pooled and labelled appropriately. A sample (2 ml) was removed for total protein and LDH activity assay and SDS-PAGE analysis, and the balance stored at -20 °C until used.

4.2.5 LDH activity assay

The reaction velocity of each purification step was determined by the decrease in absorbance of NADH at 340 nm over time. The enzyme unit was defined as 1 unit that causes the oxidation of 1 µmole of NADH per minute at 25 °C and pH 7.3. The LDH activity assay was performed using the protocol found on the website: (http://www.worthington-biochem.com/ldh/assay.html) with slight modifications. The reaction mixture contained 280 µl of 0.2 M Tris-HCl pH 7.3, 10 µl of 30

mM sodium pyruvate and 10 µl of 6.6 mM NADH. The mixture was then incubated for 4-5 minutes at room temperature. After this incubation period, 10 µl of enzyme extract was added to the mixture and read the change in absorbance at 340 nm in the spectrophotometer for 6 minutes from the initial linear portion. The enzyme activity assay was done in triplicates.

4.2.6 Protein quantification

The total protein content of the sample was determined using the microtiter plate method as described by Redinbaugh and Campbell (1985) and Bradford method (Hammond *et al*., 1988) using BSA as the standard. The assay components were mixed by pipetting 10 µl of protein sample into each well before adding 200 µl of the dye reagent, and the absorbance was measured at 595 nm.

4.2.7 SDS-PAGE analysis

The purity of the LDH enzyme extract in each purification step was determined using SDS-PAGE (Laemmli, 1970) on fractions exhibiting LDH activity. Samples (25 µl) from each fraction were mixed with 50 µl of sample buffer, before applying 25 µl to a 12% resolving and 4% stacking SDS-PAGe and electrophoresed at 200 V for 45 min. A broad range rainbow coloured pre-stained protein standard, (5 µl, New England Biolabs, Inc.) was used to calibrate the gel for Molecular Weight Determination.

4.2.8 Effect of pH on LDH activity

The activity of the purified LDH was determined at different pH values ranging from 5-10 using the assay procedure described in section 4.2.5 with changing buffer. The following buffers were used to determine the optimum pH for LDH activity: sodium acetate buffer pH 5, sodium phosphate buffer pH 6, and Tris-HCl buffer pH 7 to 8.5, and Glycine-NaOH buffer pH 9-10. Triplicate absorbance readings were taken.

4.2.9 Effect of temperature on LDH activity

The temperature profile for purified LDH was determined over a temperature range of 20 to 80 °C at the optimum pH using the assay method in section 4.2.5, and the enzyme activity was done in triplicates.

4.2.10 Thermal stability

The purified LDH was used to establish the optimal stability by incubating the enzyme at 20, 25, 30, 37, 40, 50, 60, 70, and 80 °C. The samples were prepared in triplicate and enzyme activity was measured as described in section 4.2.5. The thermal stability was tested to check the activity of LDH exposed to heat.

4.2.11 Kinetic studies

The kinetic studies were performed at 25 °C with pyruvate substrate concentrations ranging from 0.5 to 5 mM final concentration as described in section 4.2.5 and the kinetic parameters Km and Vmax were determined from the Lineweaver-Burk plot and Michaelis-Menten plot.

4.3 Results and Discussion

The activity of lactate dehydrogenase was determined from the red alga *Gelidium capensa*. The LDH activity in the crude extract was low and the extract concentrated in ammonium sulphate precipitation increased the activity. The ammonium sulphate precipitation is important for fractionation and concentrating proteins. However, it is observed that there is a loss of activity after dialysis, suggesting that protein is exposed to denaturation during dialysis. The purification of LDH on HiTrap™ Blue 5ml column showed very low peak when 2 ml sample was loaded. A flow-through peak was observed (labelled 1-A and 1-B), followed by a tiny peak (labelled 1-C). The LDH activity was measured for the tiny peak but there was no detectable LDH activity observed.

The non-binding of the LDH sample to the column could mean that there were contaminating non-LDH proteins in the column. The non-binding could also reflect that there were some salts remaining from the ammonium sulphate precipitation step, even though the dialysis was used to remove the salts, the unbound LDH may be caused by the remaining salts in a sample which competed with LDH for binding to the column. In the graph shown in **Figure 4.1**, the purification profile of 2 ml sample injection on affinity chromatography was analysed using HiTrap ^{TM} Blue 5 ml column. As a result, when this low peak used in the measurement of LDH activity, there was no activity at this stage it seems like nothing bound to the column.

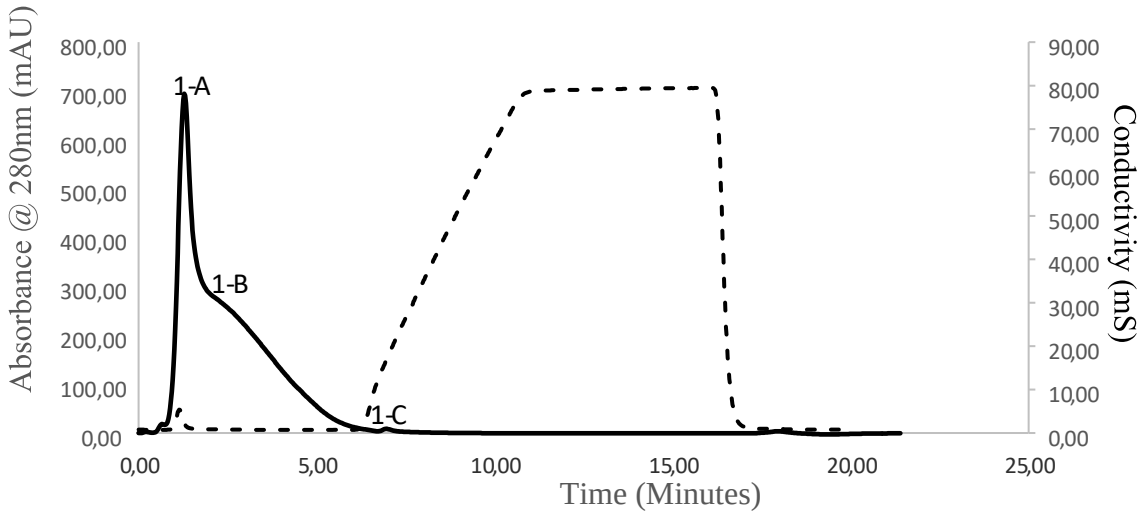


Figure 4. 1: The purification of LDH enzyme from red algae on a HiTrap™ Blue 5 ml column.

The small dotted region showed a salt gradient in equilibration buffer and the black solid line showed the protein profile at A280 nm. The fractions used for LDH activity assay were labelled 1-C and used in the SDS-PAGE to check the purity.

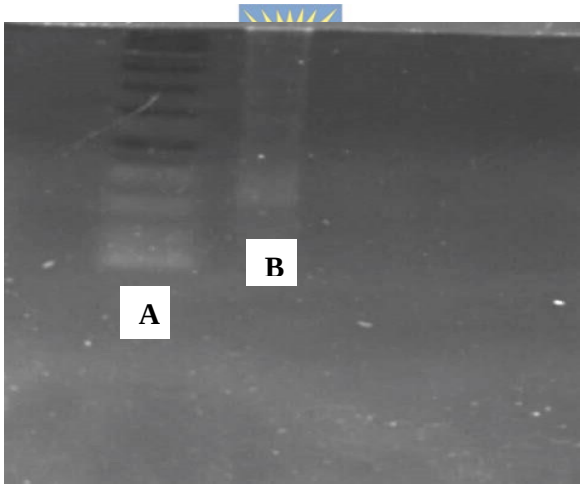


Figure 4. 2: SDS-PAGE profile showing no desired bands for LDH fractions eluted in the A_{280} profile as shown in figure 4.2. (A) Protein marker (B) sample 1-C.

The sample injection of 0.5 ml was loaded to the affinity column and several peaks were obtained. The SDS-PAGE was used to confirm the activity from the active peaks observed on the collected fractions. The LDH activity was found in four of the six fractions tested showing varied activity between 0.010 and 0.015 Units/mg. LDH protein of approximately 37 kDa in size was

detactable in the SDS-PAGE gel. The flow-through peaks were also observed, but these peaks prior to salt gradient are considered as unbound proteins and they are washed off the column. The affinity chromatography techniques are based on the separation of proteins based on the affinity charge or group, and the proteins that are unbound to the affinity column may well have no similar binding affinity.

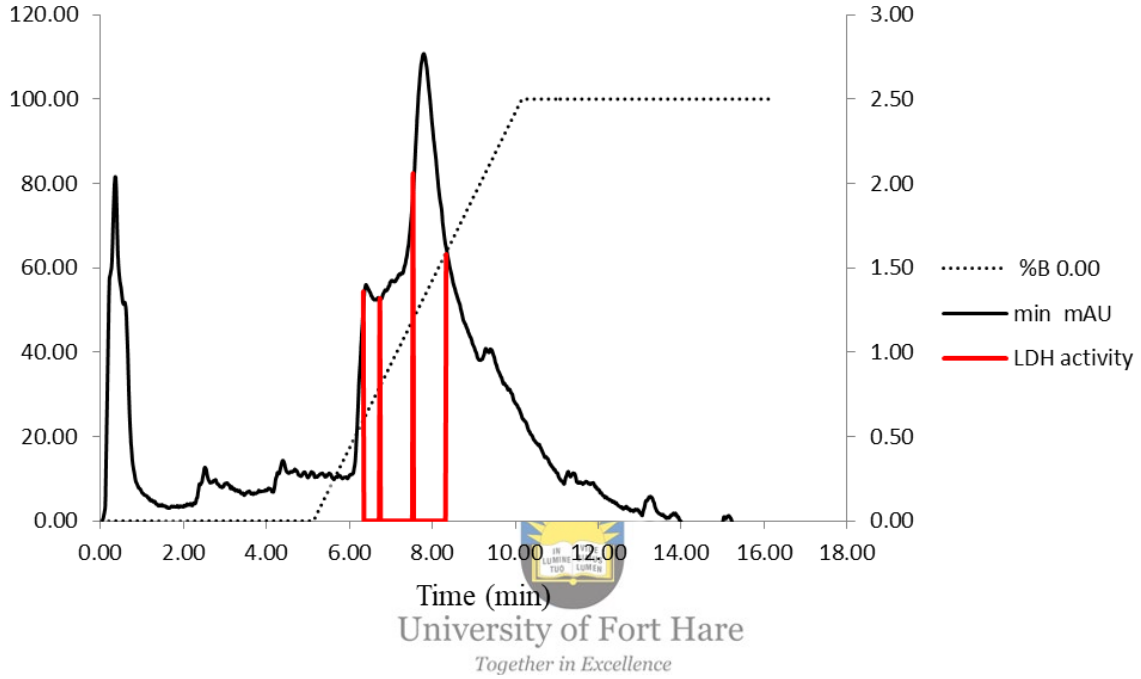


Figure 4. 3: Affinity chromatography profile of proteins at 280 nm showing active peaks with LDH activity, salt gradient and time.

The purified LDH protein from affinity chromatography showed several bands of high purity when analyzed by the SDS-PAGE. The major bands obtained from SDS-PAGE analysis showed molecular weight of approximately 37 kDa. The SDS-PAGE analysis showed very low integrity for crude extracts and dialysis samples, indicating some degradation in the proteins. The molecular weight of LDH protein obtained in **Figure 2.5** correlates to the one predicted by computational approaches shown in **Chapter 2** (**Table 2.2**).

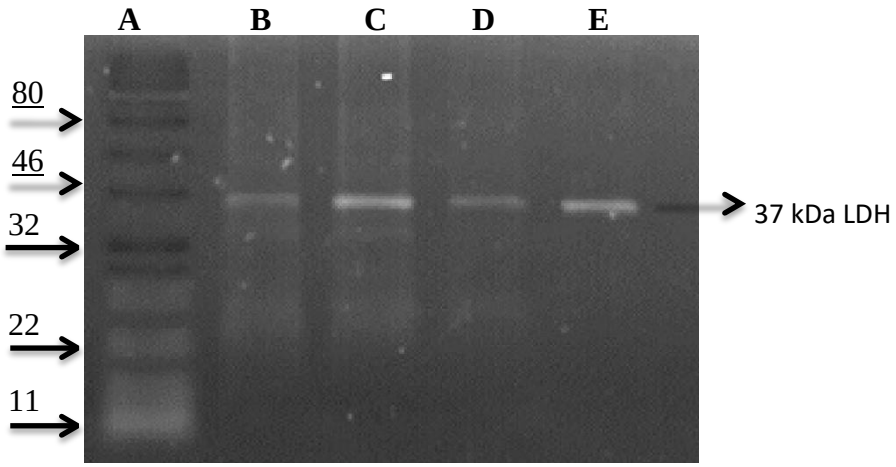


Figure 4. 4: SDS-PAGE analysis of LDH fractions from red algae after affinity chromatography. Lane A showed protein standards, lane B to E showed the LDH band(s) corresponding to each fraction from crude extract, fractionation, and dialysis and also the affinity chromatography, respectively.

The purification steps of LDH protein was carried out as shown in **Table 4.1**. The first step to enzyme purification is to analyze the tissue samples that contain the desired protein and break the cell wall to obtain the protein of the interest. The purification analysis of LDH at each step showed the highest protein concentration, total protein and total activity in the crude extract. This high value in these components was no surprise because the crude extract contains a lot of different proteins at this stage. However, the SDS-PAGE gel showed very little bands and smears for other proteins, suggesting that LDH in red algae maybe the dominant protein in these steps.

These high values decreases at each step because some of the proteins now were removed as the LDH was purified. One of the important steps in a purification of proteins is the specific activity, percentage yield and the fold purification. These steps showed how effective is the purification of the protein of interest. The purification table showed that more than 44% LDH activity was lost during the ammonium precipitation as shown by the 55.76 percentage yield recovery.

Although the recovered LDH from the step was not high, it showed that even after precipitating the proteins, more than 50% LDH was recovered and this step worked well in removing contaminating, non-LDH proteins. It was noticed in the purification table that the most effective step was the affinity chromatography because of the higher specific activity obtained, however the percentage recovery was low and this may indicate contamination which causes the LDH protein to denature. This step produces approximately 4-fold purification of LDH. The

purification of LDH was effective based on the observed purification table analysis shown in **Table 4.1**.

Table 4. 1: Purification of LDH protein from red algae by several steps including crude extract, ammonium sulphate precipitation, dialysis and affinity chromatography techniques

Purification step	[protein] (mg/ml)	Total volume (ml)	Total protein (mg)	Activity (Units/ml)	Total activity (Units)	Specific activity (Units/mg)	% Yield	Fold
Crude extract	0.891	90	80.19	0.011	0.990	0.012	100	1
Ammonium sulphate	0.657	60	39.44	0.014	0.552	0.014	55.76	1.67
Dialysis	0.560	20	11.2	0.015	0.309	0.028	31.21	2.33
Affinity chromatography	0.213	8	1.704	0.010	0.076	0.044	7.071	3.67

4.2.8 Effect of pH on LDH activity

LDH enzyme is well characterized and has been found in different organism with slightly different pH optimum. The pH on enzyme activity is important because it determines the stability and the rate of a chemical reaction. The LDH activity assay was determined at different pH values between pH 5 to pH 10 and it was found that LDH showed high activity at pH 5 to pH 9. The LDH showed highest activity at pH 7.5, suggesting that LDH is most active at pH optimum of 7.5. The stability of enzymes is important, especially for industrial applications and it was observed in this study that LDH in red algae showed activity for broad range of pH 5 to pH 9. Extremely high pH values result in loss of activity for most enzymes, and at pH 9.5 to pH 10 there was high loss of LDH activity. The observations in this study are similar to other findings from other organisms reported in literature.

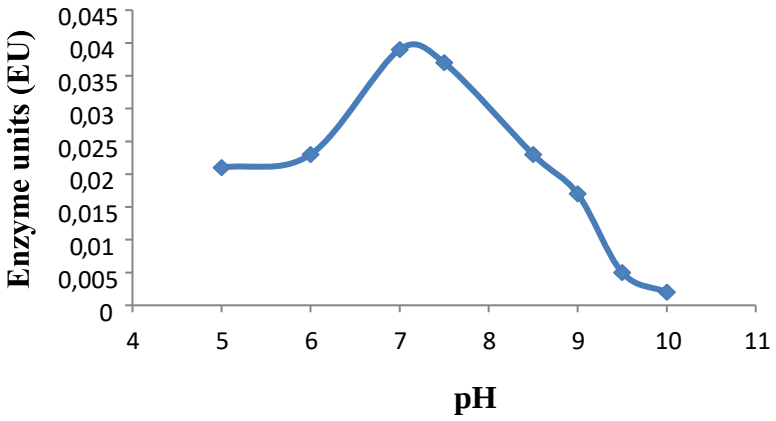


Figure 4. 5: Effect of pH on LDH activity showing activity assay at pH 5 to pH 10

4.2.9 Effect of temperature on LDH activity

The LDH activity assay was performed at different temperature to determine the optimum temperature. Enzymes are inactive at low temperatures, but as the temperature increases to room temperature, the enzymes become active. The LDH optimum temperature was found at 40 ºC. The LDH was also observed to be active over the range 20 to 40 ºC. The marine red algae are exposed to different environmental conditions, including temperature, and it was observed that at 50 to 70 ºC LDH activity dropped and then significantly decreases at 80 ºC. The loss of LDH activity may indicate the instability and denaturation.

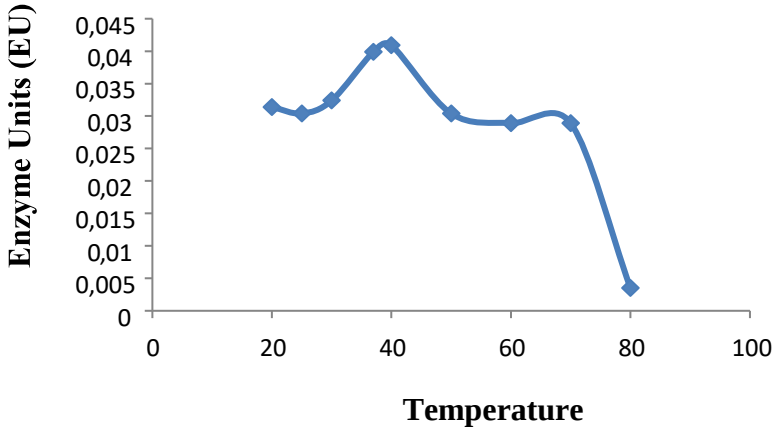


Figure 4. 6: The effect of temperature on LDH activity.

4.2.10 Thermal stability of LDH

The temperature stability of the purified LDH enzyme was tested by exposing the enzyme at different temperatures as follows; 20, 25, 30, 37, 40, 50, 60, 70 and 80 ºC. The LDH enzyme showed thermal stability from 20 to 40 ºC. The maximum activity observed at 40 ºC suggests that LDH to perform well requires temperature above the room temperature. The LDH activity assay above 40 ºC usually led to loss of activity and a complete loss of LDH activity was observed between 60 to 80 ºC. The loss of activity at higher temperature showed that LDH enzyme in red algae cannot function at extremetely higher temperature. The stability of enzymes is important especially for industrial application because it shows that the enzyme can active in a solution even if it expose at temperature above room temperation. It is also interesting to note that there is a considerable difference in the activity shown in Fig. 4.6 and Fig. 4.7compared to the Fig. 4.8. The low activity values in Fig.4.8 may indicate that storage condition and time caused this decrease in activity. LDH enzymes are sensitive to denaturation and working in ice or at temperature below 4 ºC can reduce exposure of the enzymes to denaturation.

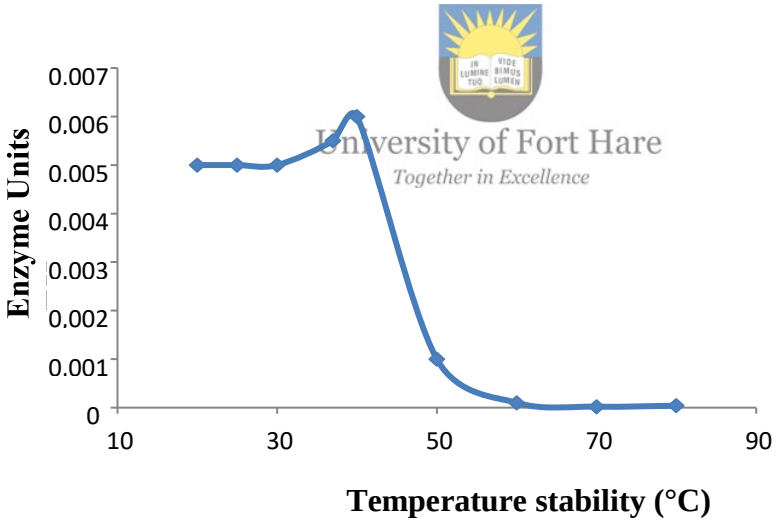


Figure 4. 7: Thermal stability of LDH enzyme after exposure at different temperature.

4.2.11 Estimation of kinetic parameters

The LDH enzyme is well characterized from different organisms and kinetic parameters are important for determining the rate at which substrate is converted to product. The graph in **Figure 4.9** showed velocity determined in mM per minute against sodium pyruvate concentration in mM.

The conversion of pyruvate to lactate using Lineweaver-Burk plot reciprocal plot showed a linear relationship between the velocity and sodium pyruvate substrate concentration.

The maximum velocity (Vmax) for LDH was reached at 0.0067 mM/ min and the enzyme at this point was saturated with substrate and the velocity could not increase, unless more enzymes were added. The sodium pyruvate concentration when initial velocity was equal to one half Vmax showed a Km value of 0.8 mM. At very low substrate concentration it is reported in the literature that the value of Km because very small.

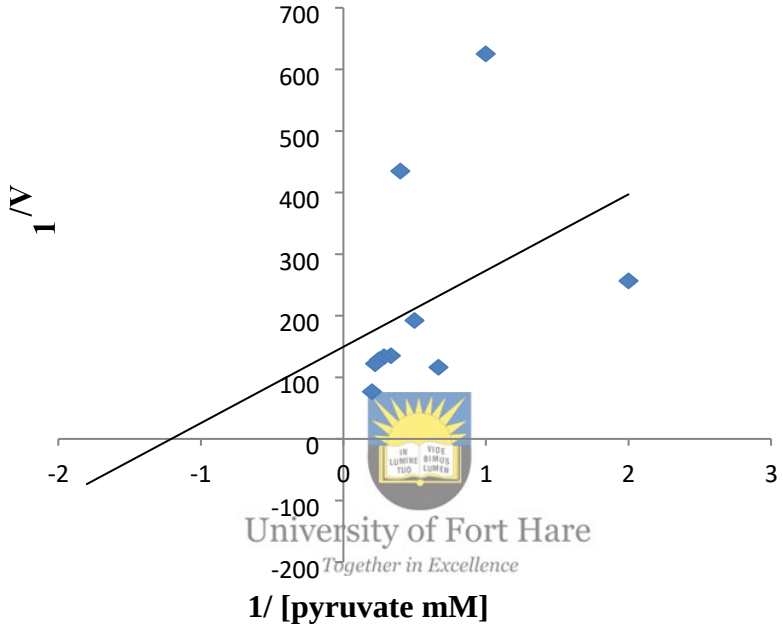


Figure 4. 8: Lineweaver-Burk plot of LDH activity determined at different substrate of pyruvate concentrations.

4.4 Concluding remarks

The LDH activity was observed in each purification step which was purified from the red algae sample using different four purification steps. The affinity chromatography step showed high specific LDH activity with percentage recovery and fold purification of 7.071 and 3.67, respectively. The SDS-PAGE analysis showed the molecular weight of LDH to be approximately 37 kDa. The optimum pH was 7.5 and the optimum temperature was observed at 40 °C. The LDH enzyme was stable for up to 40 °C, suggesting that LDH in red algae performs well at temperature above room temperature. The higher temperatures result in complete loss of activity. The kinetic parameters of LDH showed Km 0.8 mM and Vmax 0.0067 mM.min^{-1}.

4.5 References

Al-Jassabi, S., 2002. Purification and kinetic properties of skeletal muscle lactate dehydrogenase from the lizard Agama stellio. *Biochemistry*, *67*(7), pp. 786-789.

Azmat, R., Hayat, A., Aziz, F. and Qadri, M., 2013. Function of enzyme lactate dehydrogenase in relation with lipid and glucose in biotic and abiotic stresses in the seedlings of Vigna radiate. *African Journal of Microbiology Research, 7*(4), pp. 290-297.

Berg, J.M., Tymoczko, J.L. and Stryer, L., 2002. The purification of proteins is an essential first step in understanding their function. *Biochemistry*, *5*, p.20.

Hassan, S.W., Verma, S., Srivastava, S.K., Luqman, S., Gupta, U. and Masood, N., 2013. Activity guided isolation and characterization of antioxidant and antibacterial agents from some local Nigerian plants. *African Journal of Biotechnology, 12*(44), p. 635.

Javed, M.H., Azimuddin, S.M.I., Hussain, A.N., Ahmed, A. and Ishaq, M., 1997. Purification and characterization of lactate dehydrogenase from Varanus liver. *Experimental and Molecular Medicine*,*29*(1), p. 25.

Nadeem, M.S., Moran, J., Murtaza, B.N., Muhammad, K. and Ahmad, H., 2014. Cloning, E.coli expression, and Characterization of Heart Lactate Dehydrogenase B From River Buffalo (Bubalus bubalis). *Animal Biotechnology, 25*(1), pp. 23-34.

O'carra, P. and Mulcahy, P., 1996. Lactate dehydrogenase in plants: distribution and function. *Phytochemistry*, *42*(3), pp. 581-587.

Oba, K., Murakami, S. and Uritani, I., 1977. Partial purification and characterization of L-lactate Dehydrogenase Isozymes from Sweet potatoes roots. *The Journal of Biochemistry*, *81*(5), pp. 1193-1201.

O'carra, P. and Barry, S., 1972. Affinity chromatography of lactate dehydrogenase Model studies demonstrating the potential of the technique in the mechanistic investigation as well as in the purification of multi-substrate enzymes. *FEBS letters*, *21*(3), pp.281-285.

Polikovsky, M., Fernand, F., Sack, M., Frey, W., Muller, G. and Goldberg, A., 2016. Towards marine biorefineries: Selective proteins extractions from marine macroalgae Ulva with pulse electric fields *Innovative Food Science and Emerging Technologies*, 37, pp. 194-200.

Rabiu, S., Umar, R.A., Hassan, S.W. and Umaru, M.A., 2013. Characterization and inhibition studies of lactate dehydrogenase isolated from chicken muscles by aqueous and ethanolic leaves extracts of allium sativum and gossypium arboreturn. *International Journal of Chemical Technology Research, 5*(1), pp. 322-341.

Sommer. P., Klein, J.P., Scholler, M. and Frank, R.M., 1985. Lactate Dehydrogenase from Streptococcus mutans: Purification, Characterization, and Crossed Antigenicity with lactate dehydrogenases from Lactobacillus casei, actinomyces viscous, and Streptococcus sanguis. *Infection and Immunity*, *47*(2), pp. 489-495.

Vanderlinde, R.E., 1985. Measurement of total lactate dehydrogenase activity. *Annals of Clinical and Laboratory Science 15*(1), pp. 13-31.

Wolf, E.C., 1988. The partial purification and characterization of lactate dehydrogenase. *Biochemistry and Molecular Biology Education, 16(*4), pp.231-234.

CHAPTER 5: General Discussion and Future studies

5.1 General discussion

Marine red algae (seaweed) are untapped sources of raw materials for extraction of enzymes, particularly dehydrogenases which have attracted considerable interest from various applications in medicine and pharmaceutical industries. Lactate and alcohol dehydrogenases are important enzymes abundant in nature and yet underexplored in red seaweed. Molecular and biochemical properties of the dehydrogenase gene family in red algae remain largely unknown. To identify the candidate dehydrogenase genes, the *C. crispus* genome was analysed and two alcohol dehydrogenase two hypothetical genes encoding nearly-full length D-lactate dehydrogenases were found in this red alga genome.

The potential of the marine algal species has been utilized in the aim to explore and discover enzymes from the dehydrogenase gene family using a combination of bioinformatics tools, molecular and biochemical techniques. The bioinformatics analysis showed two alcohol dehydrogenase and two hypothetical genes encoding proteins belonging to the D-lactate dehydrogenases. The database search for alcohol dehydrogenases in the non-redundant protein sequence database using BLASTp (protein-protein BLAST) program showed low identity at protein levels as observed in other studies (Knoll and Pleiss, 2008; Bukh *et al*., 2012; Jin *et al*., 2014, 2016).

The two alcohol dehydrogenases in *C. crispus* belong to the MDRs superfamily and revealed conserved domains with cinnamyl alcohol dehydrogenase and class III alcohol dehydrogenase (also known as glutathione-dependent formaldehyde dehydrogenase) studied in other plant organisms (Dolferus *et al*., 1997; Bukh *et al*., 2012; Jin *et al*., 2014). The alcohol dehydrogenase proteins possess a Zn binding motif, suggesting these proteins are zinc-containing alcohol dehydrogenases. The presence of zinc amino acid residues in the active site of these proteins proves that the identified proteins are members of the MDRs superfamily, not SDR which lack the zinc binding characteristics. The catalytic domain of these ADH proteins showed conserved amino acids in the active site and also contain conserved binding motif [GXG(X)2G], which is common in most NAD-dependent dehydrogenases (Rossman *et al*., 1974; Knoll and Pleiss, 2008).

The phylogenetic analysis of alcohol dehydrogenases showed that they belong to two different groups and they possess different substrate specificity. The cis-acting regulatory elements in alcohol dehydrogenase gene family revealed stress and hormonal response regulatory elements. Most of the cis-acting regulatory elements were associated in light response, suggesting that ADHs are regulated during abiotic stress. Structural analysis of alcohol dehydrogenases showed beta-strand rich regions and these secondary structural elements were similar to the ones reported in literature for *H. sapiens* and *A. thaliana* proteins for ADH class III and *P. tremuloides* and *M. extorquens* for cinnamyl alcohol dehydrogenases. The tertiary structure analysis showed two domains surrounding the active site. The catalytic triad formed by two cysteine amino acid residues and one histidine. The position of these conserved residues differs in each species but the function is the same in most organisms.

The lactate dehydrogenases proteins in *C. crispus* showed conserved domains with members of the D-isomer specific 2-hydroxyacid dehydrogenase (Taguchi and Ohta, 1991; Razeto *et al*., 2002). The amino acid residues arginine and histidine are crucial for substrate binding and catalysis of D-lactate dehydrogenases (Razeto *et al*., 2002; Kim *et al*., 2014). The phylogenetic analysis showed that D-lactate dehydrogenases belong to the 2-hydroxyacid dehydrogenase family and they shared low similarities with formate and glycerate dehydrogenases from terrestrial plants. Although members of this family shared structural and sometimes functional analysis, the NAD-dependent dehydrogenases have D-isomers for their substrates. The D-isomer and L-isomer specific lactate dehydrogenases belong to two different evolutionary origins as shown in literature (Taguchi and Ohta, 1991).

The CAREs in lactate dehydrogenase were associated with growth and development, stress response and hormonal response. Most of the cis-acting regulatory elements obtained in 5′ upstream regulatory region of lactate dehydrogenase gene family were likely to be involved in light response, suggesting that expression pattern of LDH genes could be regulated by environmental factors. Several plant hormones were identified in the upstream sequence of *C. crispus* and these hormones have been reported to play important roles in developmental and cellular processes (Hagen and Guilfoyle, 2002; Collen *et al*., 2006). The mechanisms through which these regulatory elements respond to these various stimuli are still unknown.

The structural analysis of red algal LDH showed highly conserved structure-based sequence alignment with the green alga *Chlamydomonas reinhardtii*, *Pseudomonas aeruginosa*, and

Salmonella typhi D-lactate dehydrogenases. These highly conserved regions are important for 3D structure analysis as structure-based alignment has been shown to better conservation than sequence based alignment (Kim and Lee, 2007). The 3D structure of D-lactate dehydrogenases comprises a coenzyme binding domain and substrate binding domains that surround the active site amino acid residues (Kim *et al*., 2014; Zhu *et al*., 2015). The six parallel beta-strand and five alpha helices is conserved in most D-LDH protein structures (Stoll *et al*., 1996; Zhu *et al*., 2015). The conserved structural topology and secondary structure elements indicate that these proteins shared similar functions.

The amplification of the identified ADH-genes from *C. crispus* in red algae was however, unsuccessful because the suitable expression could not be found after numerous attempts using RT-PCR, Standard PCR and Gradient PCR techniques. The availability of the red algae genome could improve chances of success, as it would be easier to express proteins using the fully sequence genomic information from local red algae species. The *C. crispus* genome, as a model species for the genomic studies in red algae, provides insight on different genes and proteins (Collen *et al*., 2013).

The affinity chromatography HiTrap™ Blue HP 5ml column was used to separate LDH extracted from red algae. Crude extraction yielded a low LDH activity because at this step most of the proteins are still active and LDH may be exposed to denaturation. The purification steps of LDH, including crude extraction, ammonium sulphate, dialysis, and affinity chromatography yielded a 3.67-fold purification increase. Several attempts to purify the LDH enzyme were performed and initially the failure to adsorb to the affinity might suggest that the coenzyme-binding domain in red algae LDH is organized differently in D-LDH than L-LDH. The studies of Kochhar *et al*. (1992) showed similar challenge whereby the enzyme does not bind to the column. According to the study of Wilks *et al*. (1990), which was reported by Kochhar *et al*. (1992), the poor adsorption of the enzyme to the matrix was due to the increased hydrophobicity in the "jaw" region which caused coenzyme binding loop and helix G to stick together, and therefore preventing the effective interactions with the affinity (Wilks *et al*., 1990; Kochhar *et al*., 1992). The LDH activity from red algae was measured by the spectrophotometer in the pyruvate to lactate direction by observing the pyruvate-dependent NADH oxidation at 340 nm.

The SDS-PAGE analysis showed LDH-like protein with a molecular weight between 35 to 40 kDa. This band was distinct and bands with similar size were presence for crude extract,

ammonium sulphate and dialysis samples, suggesting that they are LDH proteins bands. The crude extract showed lower LDH activity. The low percentage yield in the affinity suggested that there was a loss of activity. Studies have shown that during the dialysis step, the protein might lose activity because the enzyme might be more diluted or the dialysis step result in loss of cofactors that are required for enzyme activity.

The LDH enzyme activity was determined within a broad range of pH values from pH 5 to pH 9; however, at pH 10 there was a complete loss of activity. This may indicate that at higher pH range, there is a loss enzyme activity. The highest activity was observed above neutral pH, indicating that LHD in red algae perform well at an optimum pH 7.5. The temperature optimum of LDH was determined between 37 °C and 40 °C. The LDH activity was found to be between 25 °C to 40 °C. These results were observed in other studies and it has been reported that most enzymes performed well between 37 and 40 °C. There was no LDH activity at about 50 °C, suggesting that red algae proteins are stable up to 40 °C. At high temperatures the activity of the enzymes was completely lost and this showed that proteins at high temperatures denature. Other studies have shown that LDH from other organisms can adapt to more than 85 °C (Coquelle *et al*., 2007). The Lineweaver-Burk plot drawn showed a Km of 0.8 mM, Vmax 0.0067 mM.min^{-1}.

5.2 Future studies

In this research study we have used bioinformatics tools to identify and characterize the structural properties of genes encoding dehydrogenase family proteins in red algae. We then described purification and kinetic characterization of LDH like enzyme. The LDH in red algae is considered NAD-dependent dehydrogenase that used D-isomers for their substrates. However, their role and catalytic mechanism remain unknown in algae and therefore, further investigation is needed to determine the function of these enzymes in algae. Red algae species contain a lot of dehydrogenases-like genes and studies need to be done in future to confirm that most of the hypothetical genes in most red algae species encode dehydrogenase proteins. The *C. crispus* LDH and ADH 3D models were analogous to crystalized 3D structure from other organisms. However, further studies to be done to using the experimental validated route to confirm the predicted models.

The regulation of gene expression of dehydrogenase gene family has not been characterized and experimentally validated in red algae isolates. However, it was recognized in some plant species

that most of the NAD-dependent dehydrogenases are involved in plant development and stress response and these similar result obtained in this study might help to accelerate these studies and validate the *cis*-acting regulatory elements obtained by *in-silico* analysis. The alcohol dehydrogenase and lactate dehydrogenase found in red algae still need to be analysed by several molecular and biochemical techniques to have a better understanding of the roles in algae and their ability to be used for commercial purposes. The lack of fully sequenced genomes in South Africa makes the task of studying genes and experimentally validate because of the absence of sample materials to extract proteins and other important compounds. Further studies using PCR, cloning, expression and characterization methods are required to investigate more about these important industrial enzymes. Inhibition studies are also needed to test the type of inhibition on these proteins and also to test their roles in medicinal and cancer research.

5.3 References

Bukh, C., Nord-Larsen, P.H. and Rasmussen, S.K., 2012. Phylogeny and structure of the cinnamyl alcohol dehydrogenase gene family in Branchypodium distachyon. *Journal of experimental botany, 63 (17)*, pp. 6223-6236.

Collen, J., Herve, C., Guisle-Marsollier, I., Legar, J.J. and Boyen, C., 2006. Expression profiling of Chondrus crispus (Rhodophyta) after exposure to methyl jasmonate. *Journal of Experimental Botany, 57(*14), pp. 3869-3881.

Collen, J., Porcel, B., Carre, W., Ball., S.G., Chaparro, C., Tonon, T., Barbeyron, T., Michel, G., Noel, B., Valentin, K. and Elias, M., 2013. Genomes structure and metabolic features in the red seaweed Chondrus crispus shed light on evolution of the Archaeplastida. *Proceedings of the National Academy of Sciences, 110*(13), pp. 5247-5252.

Coquelle, N., Fioravanti, E., Weik, M., Vellier, F. and Madern, D., 2007. Activity, stability and structural studies of lactate dehydrogenases adapted to extreme thermal environments. *Journal of Molecular Biology, 374,* pp. 547-562.

Dolferus, R., Osterman, J.C., Peacock, W.J. and Dennis, E.S., 1997. Cloning of the Arabidopsis and rice formaldehyde dehydrogenase genes: implications for the origin of plant ADH enzymes *Genetics, 146(3)*, pp. 1131-1141.

Hagen, G. and Guilfoyle, T., 2002. Auxin-responsive gene expression: genes, promoters and regulatory factors. *Plant molecular biology*, *49*(3-4), pp.373-385.

Hagen, G. and Guilfoyle, T., 2002. Auxin-responsive gene expression: gene, promoters and regulatory factors. *Plant Molecular Biology, 49(3-4),* pp. 373-385.

Hammond, J.B. and Kruger, N.J., 1988. The bradford method for protein quantitation. In *New Protein Techniques* (pp. 25-32). Humana Press.

Jin, Y., Zhang, C., Liu, W., Qi, H., Chen, H. and Cao, S., 2014. The cinnamyl alcohol dehydrogenase gene family in melon (Cucumis melo L.): bioinformatics analysis and expression patterns. *PloS ONE, 9(7),* p.e101730.

Jin, Y., Zhang, C., Liu, W., Tang, Y., Qi, H., Chen, H. and Cao, S., 2016. The alcohol dehydrogenase gene family in melon (Cucumis melo L.): bioinformatics analysis and expression patterns. *Frontiers in Plants Science, 7,* p.670.

Kim, C. and Lee, B., 2007. Accuracy of structure-based sequence alignment of automatic methods. *BMC Bioinformatics, 8*(1), p. 355.

Kim, S., Gu, S.A., Kim, Y.H. and Kim, K.J., 2014. Crystal structure and thermodyanamic properties of d-lactate dehydrogenase from Lactobacillus jensenii. *International Journal of Biological Macromolecules, 68,* pp. 151-157.

Knoll, M. and Pleiss, J., 2008. "The Medium Chain Dehydrogenase/Reductase Engineering Database: A systematic analysis of a diverse protein family to understand sequence-structurefunction relationship" *Protein Science*, 17(10), pp. 1689-1697.

Kochhar, S., Hunziker, P.E., Leong-Morgenthaler, P. and Hottinger, H., 1992. Primary structure, physiochemical properties, and chemical modification of NAD (+)-dependent D-lactate dehydrogenase. evidence for the presence of Arg-235, His-303, Tyr-101, and Trp-19 at or near the active site. *Journal of Biological Chemistry*, *267*(12), pp. 8499-8513.

Razeto, A., Kochhar, S., Hottinger, H., Dauter, M., Wilson, K.S. and Lamzin, V.S., 2002. Domain closure, substrate specificity and catalysis of D-lactate dehydrogenase from Lactobacillus bulgaricus. *Journal of Molecular Biology, 318(1),* pp. 109-119.

Rossmann, M.G., Moras, D. and Olsen, K.W., 1974. Chemical and biological evolution of a nucleotide-binding protein. *Nature, 250(5463),* p.194.

Stoll, V.S., Kimber, M.S. and Pai, E.F., 1996. Insights into substrate binding by D-2-ketoacid dehydrogenases from the structure of Lactobacillus pentosus D-lactate dehydrogenase. *Structure,4*(4), pp. 437-447.

Taguchi,H. and Ohta, T., 1991. D-lactate dehydrogenase is a member of the D-isomer-specific 2hydroxyacid dehydrogenase family. Cloning, sequencing, and expression in Escherichia coli of the D-lactate dehydrogenase gene of Lactobacillus plantarum. *Journal of Biological Chemistry, 266(19)*, pp.12588-12594.

Wilks, H.M., Halsall, D.J., Atkinson, T., Chia, W.N., Clarke, A.R. and Holbrook, J.J., 1990. Designs for broad substrate specificity keto acid dehydrogenase. *biochemistry*, *29*(37), pp. 85878591.

Zhu, L., Xu, X., Wang, L., Dong, H. and Yu, B. 2015. The D-lactate dehydrogenase from Sporolactobacillus insulinus also possessing reversible deamination activity. *PloS one, 10(9)*, p.e0139066.

Websites used

http://www.ncbi.nlm.nih.gov/genome/ http://www.intechopen.com/books/chemical-biology/the-problems-associated-with-enzymepurification

http://www.expasy.org?tools/protparam.html http://www.cbs.dtu.dk/services/SignalP/

http://www.cbs.dtu.dk/services/TargetP/

http://genscript.com/wolf-psort.html

http://www.ebi.ac.uk/tools/pfa/iprscan

http://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi

http://pfam.xfam.org http://prosite.expasy.org

http://www.sbg.bio.ic.ac.uk/phyre2

http://prodata.swmed.edu/promals3d/promals3d.php

http://primo.rubi.ru.ac.za/ http://swissmodel.expasy.org/

http://bioinformatics.psb.ugent.be/webtools/plantcare/html/

www.expasy.org/tools http://www.ncbi.nlm.nih.gov/gene/

www.ebi.ac.uk/interpro/

http://blast.ncbi.nlm.nih.gov/Blast.cgi

http://www.ebi.ac.uk/Tools/msa/clustalo/

http://plants.ensembl.org/index.html http://www.rcsb.org/

Databases and Software/programs

Clustal Omega - https://www.ebi.ac.uk/Tools/msa/clustalo/

EnsemblPlants - http://plants.ensembl.org/index.html

InterPro - https://www.ebi.ac.uk/interpro/

LocTree 3 - https://rostlab.org/services/loctree3/

MEGA 6 - https://www.megasoftware.net/

NCBI – www.ncbi.nlm.nih.gov

Pfam - https://pfam.xfam.org/

PHYRE^2 - http://www.sbg.bio.ic.ac.uk/~phyre2/html/page.cgi?id=index

PlantCARE - http://bioinformatics.psb.ugent.be/webtools/plantcare/html/

PRIMO (Protein Interactive Modeling) - https://primo.rubi.ru.ac.za/

PROMALS3D - http://prodata.swmed.edu/promals3d/promals3d.php

ProtParam - https://web.expasy.org/protparam/

RCSB PDB - https://www.rcsb.org/

ScanProsite - https://prosite.expasy.org/scanprosite/

SignalP 4.1 server - http://www.cbs.dtu.dk/services/SignalP/

SPDBV4.1 - https://spdbv.vital-it.ch/

SWISS-MODEL - https://swissmodel.expasy.org/

TargetP - http://www.cbs.dtu.dk/services/TargetP/

TMHMM server ver 2.0 - http://www.cbs.dtu.dk/services/TMHMM/ WoLF

PSORT - https://www.genscript.com/wolf-psort.html

APPENDICES

Appendix A: Detailed summary of proteins from different organisms that showed significant blast hits with *C. crispus* hypothetical protein (accession number: XP_005715286)

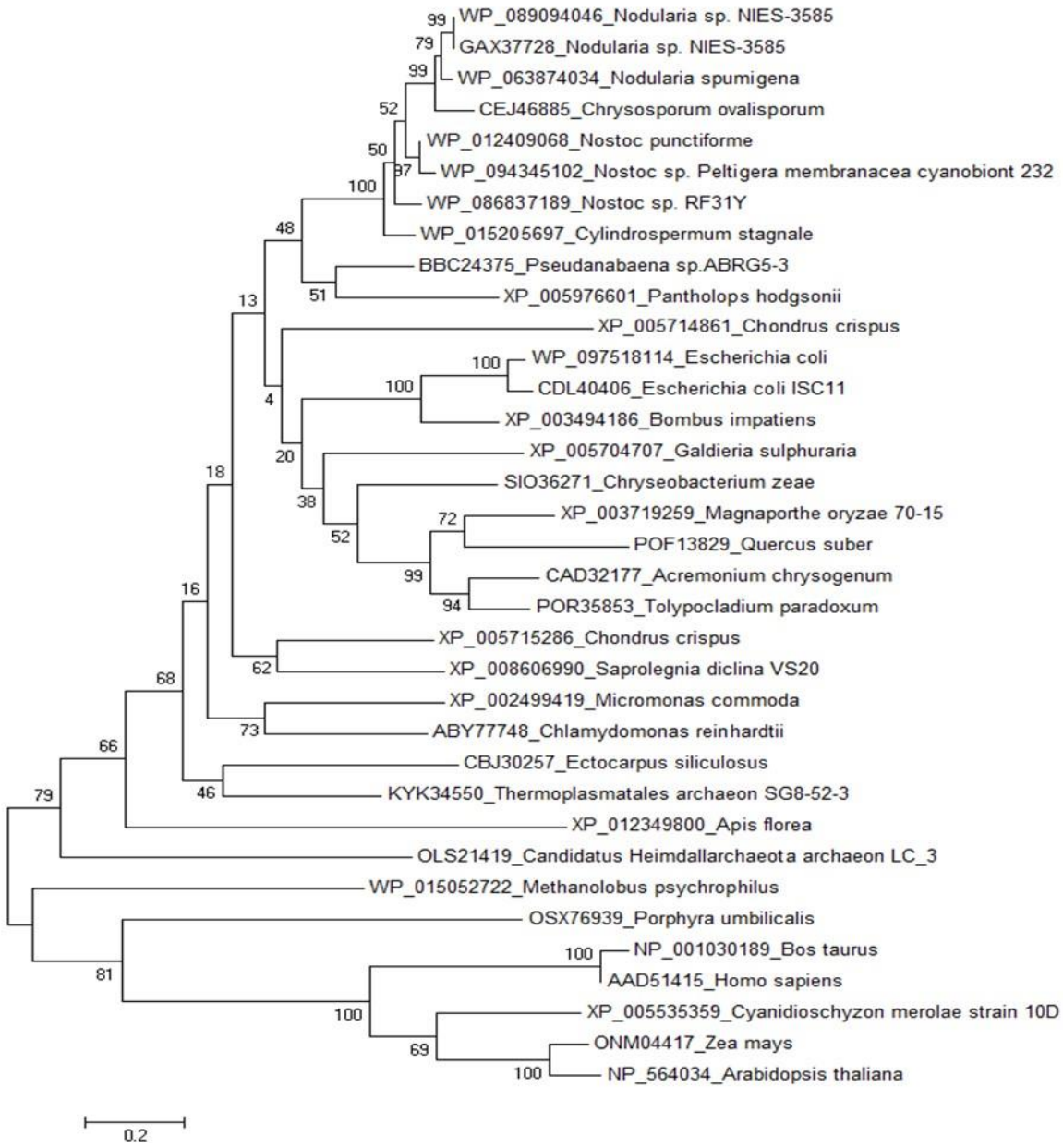
Organism name	Protein name	Accession number	Max score	Query cover	E-value	Percent identity
Chondrus crispus	Unnamed protein	XP_005714861	271	95%	6e-89	43%
Galdieria sulphuraria	D-lactate dehydrogenase	XP_005704707	312	99%	3e-105	48%
Porphyra umbilicalis	Hypothetical protein BU14 0167s0014	OSX76939	130	85%	1e-34	31%
Cyanidioschyzon merolae	Phosphoglycerate dehydrogenase	XP_005535359	127	68%	4e-32	35%
Nodularia spumigena	2-hydroxyacid dehydrogenase	WP_063874034	360	100%	6e-121	54%
Nodularia sp.NIES-3585	2-hydroxyacid dehydrogenase	WP_089094046	360	100%	1e-120	54%
Nodularia sp.NIES-3585	D-isomer specific 2-hydroxyacid dehydrogenase	GAX37728	359	100%	3e-120	54%
Nostoc punctiforme	2-hydroxyacid dehydrogenase	WP_012409068	355	100%	4e-119	54%
Nostoc sp.	2-hydroxyacid dehydrogenase	WP_094345102	353	100%	4e-118	53%
Cylindrospermum stagnale	2-hydroxyacid dehydrogenase	WP_015205697	353	100%	4e-118	53%
Chrysosporum ovalisporum	D-lactate dehydrogenase	CEJ46885	351	96%	2e-117	54%

Pseudanabaena sp.ABRG5-3	D-lactate dehydrogenase	BBC24375	347	99%	1e-115	53%
Chryseobacterium zeae	D-lactate dehydrogenase	SIO36271	346	97%	2e-115	53%

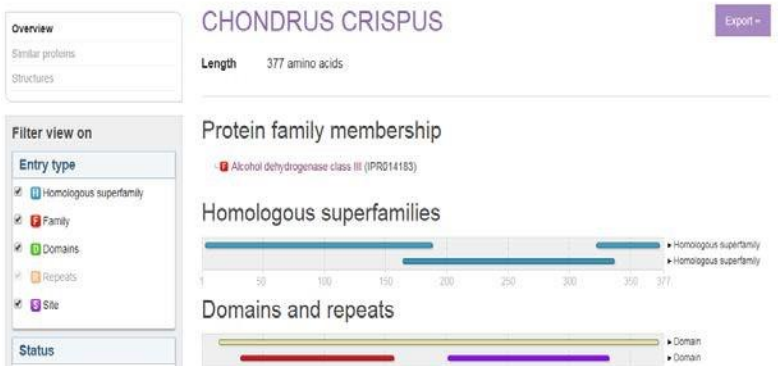
Saprolegnia diclina VS20	D-lactate dehydrogenase	XP_008606990	345	98%	8e-115	52%
Quercus suber	D-lactate dehydrogenase	POF13829	290	96%	2e-94	46%
Arabidopsis thaliana	D-3-phosphoglycerate dehydrogenase	NP_564034	127	71%	2e-30	32%
Zea mays	D-3-phosphoglycerate dehydrogenase	ONM04417	130	69%	2e-32	34%
Triticum aestivum	NAD-dependent formate dehydrogenase	ANW82830	103	78%	1e-23	29%
Zostera marina	Formate dehydrogenase	KMZ5638	106	75%	1e-24	30%
Pantholops hodgsonii	PREDICTED:2-hydroxyacid dehydrogenase homolog 1-like	XP_005976601	331	96%	1e-114	50%
Bos taurus	D-3-phosphoglycerate dehydrogenase	NP_001030189	120	66%	8e-30	30%
Chlamydomonas reinhardtii	Putative-D-lactate dehydrogenase	ABY77748	292	98%	9e-96	48%
Micromonas commoda	D-lactate dehydrogenase	XP_002499419	276	90%	2e-90	48%
Ectocarpus siliculosus	2-hydroxyacid dehydrogenase	CBJ30257	245	99%	2e-77	41%
Bombus impatiens	PREDICTED: 2-hydroxyacid dehydrogenase	XP_003494186	308	96%	7e-102	48%
Apis florea	PREDICTED: 2-hydroxyacid dehydrogenase homolog 1-like	XP_012349800	210	96%	6e-62	36%
Magnaporthe oryzae 70-50	D-lactate dehydrogenase	XP_003719259	336	94%	3e-112	51%
Tolypocladium paradoxium	2-hydroxyacid dehydrogenase	POR35853	334	96%	1e-111	52%

Acremonium chrysogenum	Putative D-hydroxyacid dehydrogenase	CAD32177	326	98%	3e-108	52%
Thermoplasmatales archaeon SG8-52-3	Hydroxyacid dehydrogenase	KYK34550	278	90%	2e-90	48%
Candidatus Heimdallarchaeota archaeon LC_3	D-lactate dehydrogenase	OLS21419	198	98%	3e-59	32%
Methanolobus psychrophilus	Glycerate dehydrogenase	WP_015052722	170	78%	2e-48	38%
Homo sapiens	D-3-phosphoglycerate dehydrogenase	AAD51415	121	67%	1e-29	30%
Escherichia coli	2-hydroxy dehydrogenase	WP_097518114	305	97%	4e-101	47%
Escherichia coli ISC11	D-lactate dehydrogenase	CDL40406	302	97%	6e-100	47%

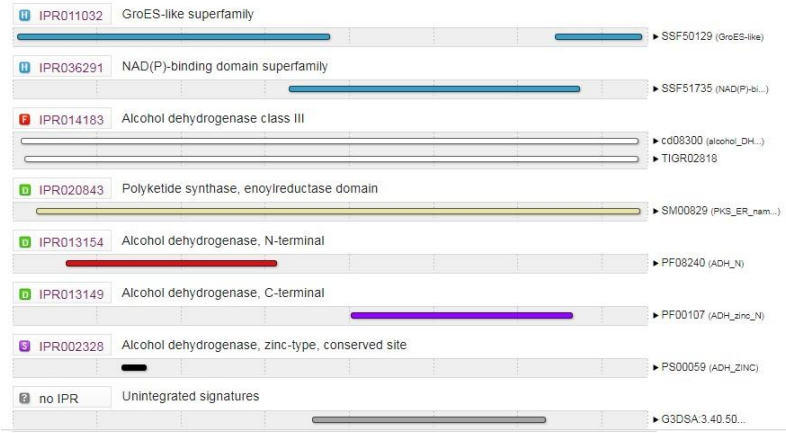
Appendix B: Phylogenetic tree of 2-hydroxyacid dehydrogenase protein family members



Appendix C: InterPro results for alcohol and lactate dehydrogenase classification and conserved domain and functional sites



Detailed signature matches



Residue annotation

GO term prediction

Biological Process

GO:0006069 ethanol oxidation
GO:0055114 oxidation-reduction process

Molecular Function

GO:0008270 zinc ion binding
GO:0016491 oxidoreductase activity
GO:0051903 S-(hydroxymethyl)glutathione dehydrogenase activity

Cellular Component

None predicted.

CAD1_CHONDRUS CRISPUS

Export =

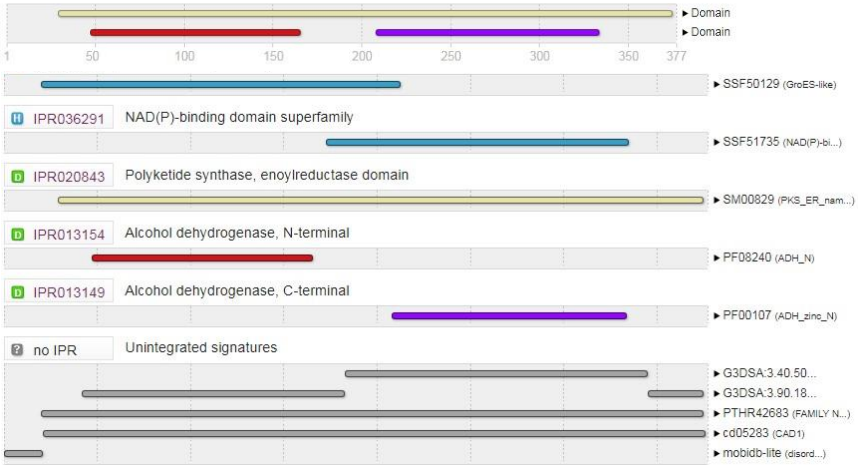
Length 377 amino acids

Protein family membership

None predicted.

Homologous superfamilies

Domains and repeats



Residue annotation

GO term prediction

Biological Process

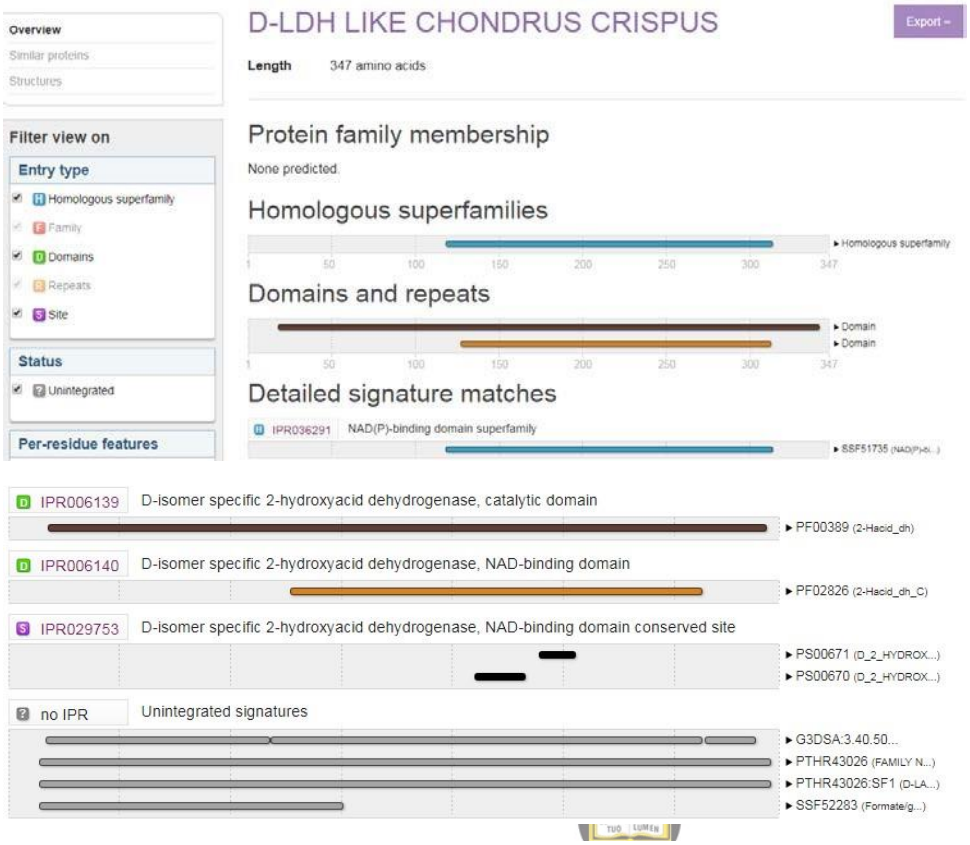
GO:0055114 oxidation-reduction process

Molecular Function

GO:0016491 oxidoreductase activity

Cellular Component

None predicted.



GO term prediction

Biological Process

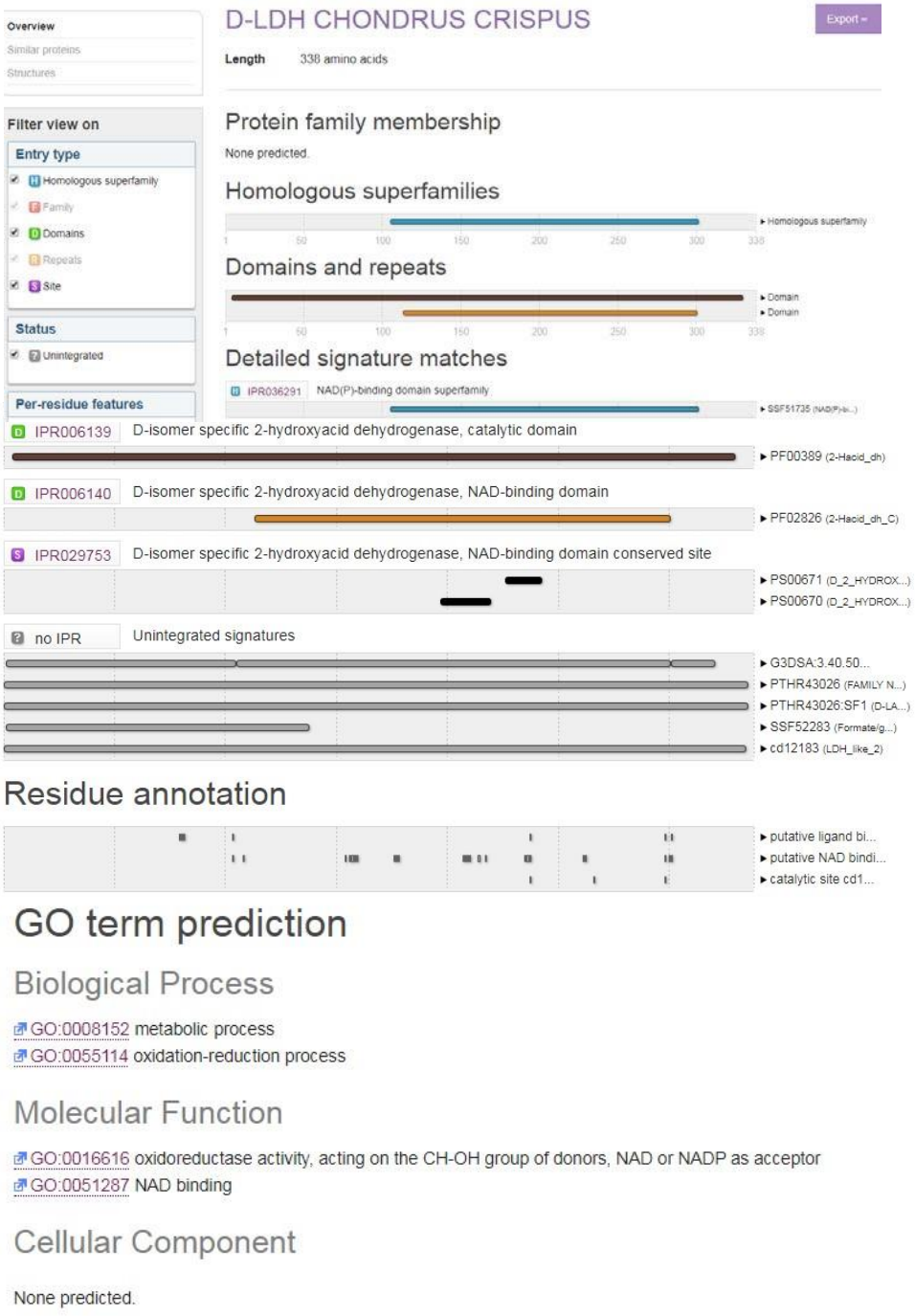
GO:0008152 metabolic process
GO:0055114 oxidation-reduction process

Molecular Function

GO:0016616 oxidoreductase activity, acting on the CH-OH group of donors, NAD or NADP as acceptor
GO:0051287 NAD binding

Cellular Component

None predicted.



Appendix D: Comparison of *C. crispus* ADH3 proteins with proteins from other species

Species name	Accession number	Amino acid	Identity (%)

Danio rerio	NP_571924	376 aa	66
Homo sapiens	NP_000662	374 aa	65
Mus musculus	NP_031436	374 aa	64
Drosophila melanogaster	NP_524310	379 aa	64
Caenorhabditis elegans	NP_001024016	386 aa	62
Schizosaccharomyces pombe 972h-	NP_588247	380 aa	60
Saccharomyces cerevisiae S288C	NP_010113	386 aa	59
Dictyostelium discoideum AX4	XP_640467	379 aa	62
Arabidopsis thaliana	NP_199207	379 aa	66
Glycine max	NP_001341703	379 aa	61
Shewanella oneidensis MR-1	NP_717657	379 aa	64
Pseudomonas aeruginosa PAO1	NP_252319	370 aa	64
Synechocystis sp. PCC 6803	WP_010871793	369 aa	63
Microcystis aeruginosa	WP_012265350	369 aa	61
Escherichia coli str. K-12 substr. MG1655	NP_414890	369 aa	62
Neisseria meningitides MC58	NP_274323	378 aa	60
Zea mays	NP_001105485	381 aa	62
Pogona vitticeps	XP_020658586	374 aa	66
Thamnophis sirtalis	XP_013930436	374 aa	64
Taeniopygia guttata	NP_001232570	374 aa	66
Sturnus vulgaris	XP_014730744	374 aa	67
Gallus gallus	NP_001026323	374 aa	64
Solanum lycopersicum	NP_001238796	379 aa	64

Solanum tuberosum	NP_001274960	379 aa	64
Rhincodon typus	XP_020376859	377 aa	66
Callorhinchus milii	NP_001279055	377 aa	65
Auxenochlorella protothecoides	XP_011399595	381 aa	57
Cyanidioschyzon merolae strain 10D	XP_005538792	382 aa	65
Galdieria sulphuraria	XP_005705275	375 aa	64
Porphyra umbilicalis	OSX74245	379 aa	64
Sus scrofa	NP_001231762	385 aa	65
Oryza sativa japonica group	XP_015627169	381 aa	62
Pseudospirillum japonicum	WP_093309558	376 aa	67
Dickeya zeae	WP_023639876	373 aa	67
Xenopus tropicalis	NP_001011502	376 aa	65

Appendix E: Summary of the *cis*-acting regulatory elements identified in the 1500bp 5'upstream promoter region of the *Chondrus crispus* class III ADH gene family

Motif name	Motif sequence	Position	Motif function
A-box	CCGTCC	73	Cis-acting regulatory element
ABRE	CCTACGTGGC TACGTG CACGTG	323; 661; 487	Cis-acting element involved in the abscisic acid responsiveness
AC-I	CCCACCTACC	146; 544	
ACE	ACGTGGA AAAACGTTTA	591; 1310	Cis-acting element involved in light responsiveness
ARE	TGGTTT	1356	Cis-acting element regulatory element for the anaerobic induction
AuxRR-core	GGTCCAT	287	Cis-acting regulatory element involved in auxin responsiveness
Box-W1	TTGACC	53	Fungal elicitor responsive element
CAAT-box	CAAAT CAATT CCAAT CAAT	41; 51; 270; 271;362; 413; 597;637; 657; 689;720; 800; 938;1316; 348; 1350; 1403; 1491	Common cis-acting element in promoter and enhancer regions

CAG-motif	GAAAGGCAGAC	1074	Part of a light response element

CATT-motif	GCATTC	500; 1083	Part of a light response element
CCGTCC-box	CCGTCC	73	Cis-acting regulatory element related to meristem specific activation
CGTCA-motif	CGTCA	97;358;283; 491;244;313	Cis-acting regulatory element involved in the MeJA-responsiveness
G-Box	CACGTG;CACGTT; CACGTA	487;785;601; 997;590;661	Cis-acting regulatory element involved in light responsiveness
G-box	CACGAC;CACGTT CACGTG;TACGTG CACGTGG	380; 785; 601; 1103; 487; 997; 661; 486; 590	Cis-acting regulatory element involved in light responsiveness
GAG-motif	AGAGAGT AGAGATG	131; 1258; 1053	Part of a light responsive element
GARE-motif	TCTGTTG	812	Gibberellin-responsive element
GC-motif	CCCCCG	122	Enhancer-like element involved in anoxic specific inducibility
GC-repeat	GGCCTCGCCACG	317	

GCC box	AGCCGCC	1066	
I-box	TATTATCTAGA	1139	Part of a light responsive element
LTR	CCGAAA	1061	Cis-acting element involved in low-temperature responsiveness
MBS	CAACTG	628; 699	MYB binding site involved in drought-inducibility
Skin-1_motif	GTCAT	20;359;284;464	Cis-acting regulatory element required for endosperm expression
Sp1	CC(G/A)CCC GGGCGG	119;914;150; 1421;141;1065; 544	Light responsive element
TATA-box	TAATA;TTTTA; TTTAAAAA TACAAAA TACATAAA TATAAA;ATATAT	349; 1306; 753; 1307; 605; 1139	Core promoter element around -30 of transcription start
TGACG-motif	TGACG	97;358;283; 491;244;313	Cis-acting regulatory element involved in the MeJA-responsiveness
W box	TTGACC	53	

Circadian	CAANNNNATC	433; 800	Cis-acting regulatory element involved in circadian control

Summary of the predicted cis-acting regulatory elements that have been found in the 1500bp 5' upstream region of *C .crispus* cinnamyl ADH gene

Motif name	Position	Sequence	Function
A-box	113; 1452; 1253	CCGTCC	cis-acting regulatory element
ABRE	907	CACGTG	cis-acting element involved in the abscisic acid responsiveness
AC-II	1480	(C/T)T(T/C)(C/T)(A/C) (A/C)C(A/C)A(A/C)C(C/A) (C/A)C	
AE-box	676	AGAAACTT	part of a module for light response

ARE	774	TGGTT T	cis-acting regulatory element essential for the anaerobic induction
Box 4	1368; 1372	ATTAAT	part of a conserved DNA module involved in light responsiveness
CAAT-box	188; 1322;1092;1415; 643;1381;1143;1366;515; 1414;1115;1058; 1195	CAATT; CAAT	common cis-acting element in promoter and enhancer regions
CATT-motif	633	GCATTC	part of a light responsive element
CCGTCC-box	113; 1453; 1253	CCGTCC	cis-acting regulatory element related to meristem specific activation
CGTCA-motif	69; 91	CGTCA	cis-acting regulatory element involved in the MeJA-responsiveness
G-Box	907	CACGTG	cis-acting regulatory element involved in light responsiveness

G-box	550; 946; 906; 902; 907	CACATGG; CACGAC; ACACGTGT;ACAC GTGTCACC	Cis-acting regulatory element involved in light responsiveness
GAG-motif	1001	AGAGAGT	Part of light responsive element
GARE-motif	386; 681	AAACAGA; TCTGTTG	Gibberellin-responsive element
GC-motif	376; 1424; 819; 1438	GCCCCGG; CCCCG	Enhancer-like element involved in anoxic specific inducibility

MBS	235; 252	TAACTG	MYB binding site involved in drought-inducibility
MNF1	1212	GTGCCC (A/T) (A/T)	Light responsive element
02-site	1291	GATGACATGA	Cis-acting regulatory element involved in zein metabolism regulation
Skn-1_motif	47; 92; 70; 336	GTCAT	Cis-acting regulatory element required for endosperm expression
Sp1	372;1257;820;1449; 792;1437;1249	CC (G/A) CCC	Light responsive element
TATA-box	358;1395;1048;1409;396;1276	TAATA; TATA	Core promoter element around -30 of transcription start
TATCCAT/C-motif	1282	TATCCAT	
TCA-element	1216; 1300	CCATCTTTTT	Cis-acting regulatory element involved in salicylic acid responsiveness
TCT-motif	762	TCTTAC	Part of light responsive element
TGACG-motif	69; 91	TGACG	Cis-acting regulatory element involved in the MeJA- responsiveness
circadian	531; 724	CAANNNNATC	Cis-acting regulatory element involved in circadian
motif IIb	1438	CCGCCGCGCT	Abscisic acid responsive element

Summary of putative *cis*-acting regulatory elements found in the 1500bp 5' upstream region of translational start site of *Chondrus crispus* hypothetical gene (D-LDH) using PlantCARE database

Motif name	Motive sequences	Positions	Motif function
ACE	AAAACGTTTA	1275	Cis-acting element involved in light responsiveness

ARE	TGGTTT	21	Cis-acting regulatory element important for anaerobic induction
Box-W1	TTGACC	275	Fungal elicitor responsive element
CAAT-box	CCAAT, CAAT, CAAT, CCAT, CAATT, CAAT, CAAT, CAAT, CAAT, CAAT, CAAT, CAATT, CCAAT, CAAT, CAAT, CCAAT, gGCAAT, CAAT, CAAAT, CAAT	69, 70, 124, 230, 376, 377, 389, 401, 479, 584, 613, 731, 765, 766, 893, 1025, 1057, 1111, 1113, 1312, 1408	Common cis-acting element in promoter enhancer
CATT-motif	GCATTC	301	Part of a light responsive element
CCAAT-box	CAACGG	159	MYBv1 binding site
CGTCA-motif	CGTCA, CGTCA, CGTCA, CGTCA	332, 1062, 394, 1468	Cis-acting element involved in the MejA-responsiveness

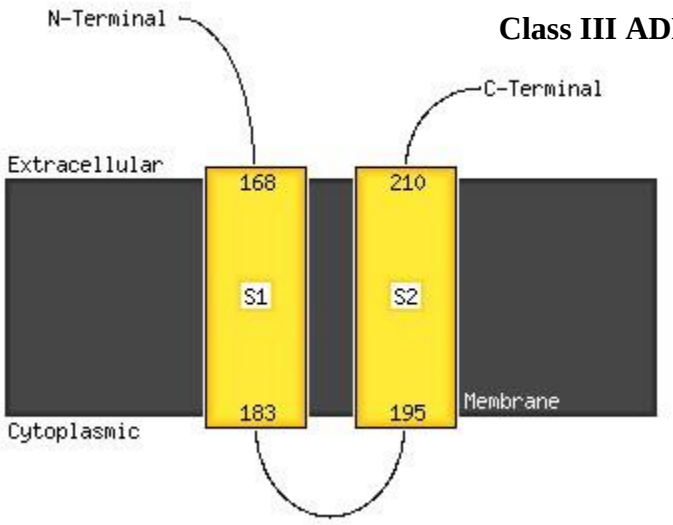
G-Box	CACGTT	1237	Cis-acting regulatory element involved in light responsiveness
G-box	CACGAC, CACGTC, CACGTT	735, 1466, 1237	Cis-acting regulatory element involved in light responsiveness

GARE-motif	TCTGTTG	271	Gibberellin-responsive element
GC-motif	CCCCCG	1377	Enhancer-like element involved in anoxic specific inducibility
GT1-motif	GGTTAA	1433	Light responsive element
I-box	GATATGG	1344	Part of a light responsive element
LTR	CCGAAA, CCGAAA	24, 1244	Cis-acting element involved in low-temperature responsiveness
MBS	CAACTG	574	MYB binding site involved in drought inducibility
O2-site	GATGATGTGG, GATGATATGG	844 1344	Cis-acting regulatory involved in zein metabolism
Skn-1_motif	GTCAT, GTCAT, GTCAT	393, 1469, 750	Cis-acting regulatory element required for endosperm expression
Sp1	CC (G/A) CCC, CC (G/A) CCC	197, 1376	Light responsive elements
TATA-box	TTTTA, TATA, TAATA, TAATA, TACAAAA	246, 1398, 595, 592, 1194	Core promoter element around -30 of transcription start site
TCA-element	GAGAAGAATA	708	Cis-acting element involved in salicylic acid responsiveness

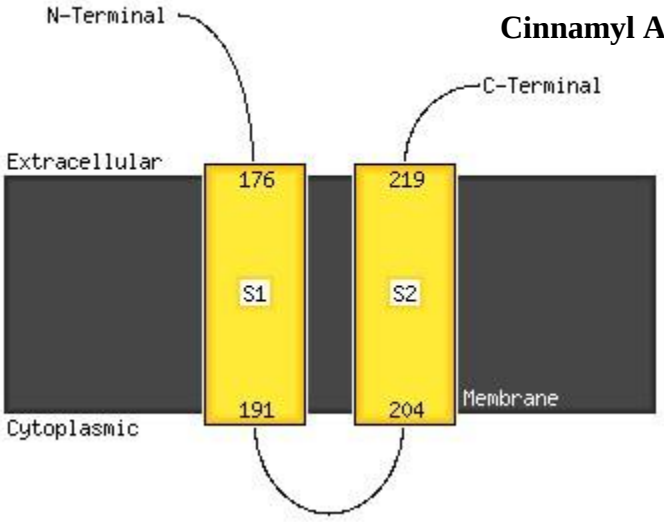
TCCC-motif	TCTCCCT	680	Part of a light responsiveness
TGA-element	AACGAC	1301	Auxin-responsive element
TGACG-motif	TGACG, TGACG, TGACG,TGACG	332, 1062, 394, 1468	Cis-acting regulatory element involved in the MejA-responsiveness
Unnamed_1 Unnamed_3	CGTGG, CGTGG, CGTGG, CGTGG, CGTGG, CGTGG	135, 907, 505, 1236, 259, 701	
Unnamed_4	CTCC, CTCC, CTCC, CTCC, CTCC, CTCC, CTCC, CTCC, CTCC, CTCC, CTCC,CTCC,CTCC	12, 1327, 682, 1489, 617, 1387, 818, 1374, 552, 1447, 707, 634, 847	
W-box	TTGACC	275	
box-S	AGCCACC	779	
Circadian	CAANNNNATC, CAANNNNATC	1065, 1312	*Cis*-acting regulatory element involved in circadian control

Appendix F: Transmembrane topology region in the secondary structures of *C. crispus* Alcohol dehydrogenases

Class III ADH



Cinnamyl ADH



are

Appendix G: Overall-fold of class III ADH protein structures in red algae predicted proteins using *Homo sapiens* templates (PDB code: 1MC5, 2FZW, 1TEH, and 3QJ5)

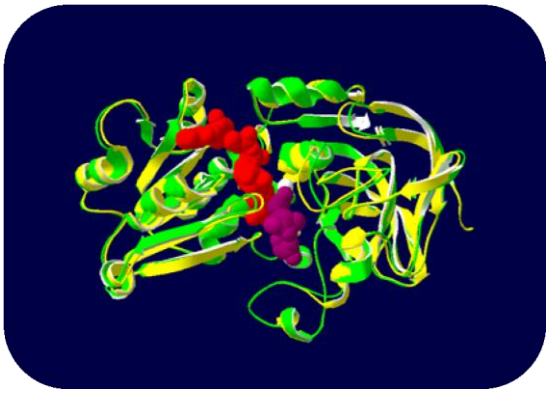


Figure 1:3D-structure model of ADH from *C. crispus*

Figure 2:3D-structure model of ADH_3 from *C. merolae*

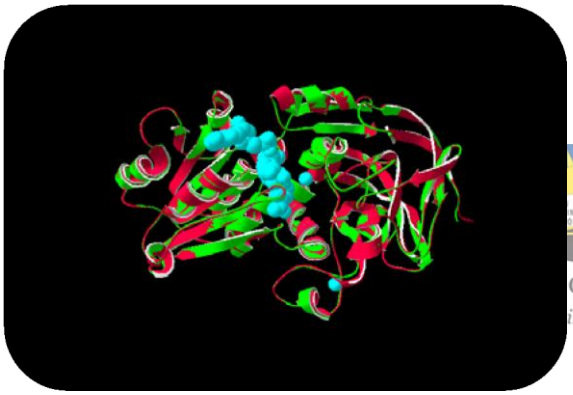


Figure 3:3D-structure model of ADH (zinc-containing) from *G. sulphuraria*

Figure 4:3D-structure model of hypothetical protein BU14 from *P. umbilicalis*