An *in silico* analysis, purification and partial kinetic characterisation of a serine protease from *Gelidium pristoides*

University of Fort Hare

Together in Excellence

A dissertation

submitted in fulfilment of the requirement for the degree of

Master of Science (MSc) Biochemistry

by

Zolani Ntsata

Supervisor: Prof. Graeme Bradley

2020

Department of Biochemistry and Microbiology

University of Fort Hare
Together in Excellence

Declaration

I, Zolani Ntsata (201106067), declare that this dissertation, entitled 'An in silico analysis and kinetic characterisation of proteases from red algae' submitted to the University of Fort Hare for the Master's degree (Biochemistry) award, is my original work and has NOT been submitted to any other university.

Signature: 

I, Zolani Ntsata (201106067), declare that I am highly cognisant of the University of Fort Hare policy on plagiarism and I have been careful to comply with these regulations. Furthermore, all the sources of information have been cited as indicated in the bibliography.

Signature: 

I, Zolani Ntsata (201106067), 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 these regulations. There was no need for ethical clearance.

Signature: 

University of Fort Hare
Together in Excellence

Dedication

I dedicate this work to my grandmother, Nyameka Mabi.

Acknowledgements

Above all things, I would like to give thanks to God for the opportunity to do this project and for the extraordinary strength to persevere in spite of the challenges that came along.

I am thankful to my family, especially my grandmother, for her endless support.

I would also like to acknowledge Prof Graeme Bradley for his supervision and guidance.

Thanks to my friends and colleagues, especially Yanga Gogela and Ntombekhaya Nqumla, and the plant stress group for their help and support.

I would like to acknowledge the GMRDC, NRF, and SAIAB Phuhlisa programme for funding this project, and lastly, the University of Fort Hare for providing the facilities to do this project.

Abstract

The aim of this study was to characterize the protease enzyme (s) from red algae. An *in silico* analysis of red algae genomes was used to identify gene coding for protease. Protease sequences identified from these genomes were examined for conserved domains, active site and structures. The domain search revealed that the identified sequences were from the five classes of protease enzymes. For function inference, the red algae sequences were aligned to identify the catalytic sites, and the tertiary structures were predicted using homology modelling.

An *in silico* analysis provides an indication of the class and potential functions of the enzymes. However, it cannot predict whether the gene is constitutively expressed in the red algae or under which conditions it may be induced, and it cannot determine the kinetic efficiency of an enzyme against various substrate, or the optimum conditions for the protein activity.

Attempts to clone and recombinantly express selected red algae proteases, proved unsuccessful, as the available genomes where from red algae species found mainly in Asia, and the designed primers, therefore, did not amplify a corresponding PCR product from the red algae harvested in South Africa.

Crude extracts of red algae collected from Kenton-on-Sea, along the East Coast of South Africa, were screened for protease activity using Benzoyl-Arginine-pNitroAnilide (BApNA) as substrate. The proteases detected in the crude extract were purified using ammonium sulphate precipitation and HiPrep DEAE FF 16/10; CM FF 16/10, and HiPrep Q FF 16/10 columns for ion-exchange chromatography. The HiPrep Q FF 16/10 column yielded active protein, which revealed two bands of 11kDa and 17kDa on SDS-PAGE. It was assumed that these bands represented two subunits of the purified protease.

Kinetic characterisation of the purified protease revealed a pH optimum of 9, using BApNA as substrate, a temperature optimum at 60ºC, and sensitivity to temperature when stored above 4ºC. The protease activity was inhibited by Ferric chloride (32%), induced by calcium chloride (156%), no inhibition by magnesium chloride (97%) and slight inhibition by potassium chloride (77%) and manganese chloride (70%). Phenylmethylsulfonyl fluoride (PMSF), a serine protease inhibitor, almost totally inhibited the protease activity, indicating that the protease from red algae was most likely a serine protease. The K_m and k_{cat} values were 1.96 µM, and 0.364 s^{-1}, respectively using BApNA as the substrate.

This study revealed that the red algae genome contains numerous genes that encode for proteases from almost all the classes of proteases. A serine protease from the red algae *Gelidium pristoides* was partially purified and kinetically characterised, confirming that red algae found along the Eastern Coast of South Africa contain genes that express active proteases that may be of medical or industrial interest. Further studies, however, are required to recombinantly express, purify and characterise the numerous proteases encoded by the genes identified in the *in silico* analysis of this study.

University of Fort Hare
Together in Excellence

Table of content

University of Fort Hare
Together in Excellence

University of Fort Hare
Together in Excellence

University of Fort Hare
Together in Excellence

University of Fort Hare
Together in Excellence

Table of Figures

University of Fort Hare
Together in Excellence

University of Fort Hare
Together in Excellence

List of tables

University of Fort Hare
Together in Excellence

University of Fort Hare
Together in Excellence

Chapter 1
Literature Review

1. INTRODUCTION

Proteases are enzymes that catalyse the hydrolysis of proteins into peptides and or amino acids. They are also referred to as 'proteinases' or 'peptidases' (Devi and HemaLatha, 2014). However, the term 'protease' is used interchangeably with the term 'proteolytic enzyme'. Proteolytic enzyme encompasses even those enzymes that do not act on peptide bonds. For instance, some lyases are proteolytic, catalysing an elimination reaction instead of hydrolysis (Rawlings *et al.*, 2011).

Proteases constitute one of the major groups of industrial enzymes and represent about 60% of the global enzyme market (Anwar and Saleemuddin, 1998; Joshi and Satyanarayana, 2013; Rao *et al.*, 1998). Protease study is endorsed by its fundamental function in various areas of industry and therapy (Mótyán *et al.*, 2013). The vast diversity of proteases and their ability to conduct a very specific chemical transformation has promoted them as a good tool in such areas (Gohel and Sing, 2012; Li *et al.*, 2012). Proteases have a wide range of applications including in detergent manufacture, leather manufacture, different sectors of the food industry, and for bioremediation (including wastewater treatment) (de Castro *et al.*, 2014; Devi and HemaLatha, 2014; Vijayaraghavan *et al.*, 2014).

Although protease application shows great potential in the industry, it still has insufficient ability to resist adverse industrial conditions for commercial success (Gohel and Singh, 2012). Various approaches have attempted to overcome such challenges. These include protein engineering, together with biotechnology. However, since there is a high demand for proteases, a search for new, potent sources with promising proteases is ongoing (Anbu, 2013). It is believed that the evolutionary complexity of living organisms provides an enormous range of different proteases with a large repertoire and specificity. Furthermore, different proteases catalysing the same reaction permit flexibility in choice since they can have different features (stability, pH, and temperature optima) (Abidi *et al.*, 2011). This review highlights the protease enzyme classes, sources, applications, and production.

University of Fort Hare
Together in Excellence

1.1 Classification of proteases

The International Union of Biochemistry and Molecular Biology has categorised all enzymes into six classes: Oxidoreductases, Transferases, Hydrolases, Lyases, Isomerases, and Ligases, according to the type of reaction they catalyse (Boyce and Tipton, 2001). Hydrolases are classified as the third class, with 13 subclasses. Proteases are grouped in subclass 4 (EC 3.4) of hydrolases (hydrolysing peptide bonds) following carbohydrates (EC 3.3) (Jegannathan and Nielsen, 2013; Boyce and Tipton, 2001; Rao *et al.*, 1998). Although it is not difficult to classify enzymes generally, proteases do not always comply with the enzyme international system for classification and nomenclature, their vast diversity in action and structure (Rao *et al.*, 1998; Tavano, 2013). However, three major standards/measures are considered in the classification of protease: (i) type of reaction, (ii) catalytic site residues, and (iii) phylogenic relation with regards to the structure (Rao *et al.*, 1998).

Broadly, proteases can firstly be grouped, based on the site of action, as endo-/exo-peptidases. Endopeptidases cut within the peptide chain away from the termini. In contrast, exopeptidases catalye the cleavage of peptide bonds towards the N-/C- termini (Mótyán *et al.*, 2013; Rawlings *et al.*, 2012). Those that act proximally to the N-terminus are referred to

as 'aminopeptidase' and those that act proximally to the carboxyl terminus are referred to as 'carboxypeptidase' (Sawant and Negendran, 2014). In the case of cathepsins, both peptidase forms (N-/C-) can be observed, since their structure can present elements that are positively charged (cathepsin X) to bind to the negatively charged terminus of the protein, or negatively charged (cathepsin H) to bind to the positively charged terminus on the protein.

Development in the characterisation of the active site and structure prediction has revealed that proteases can be grouped into clans and families (Di Cera, 2009; Rao *et al.*, 1998; Rawlings, 2016). The clans contain all the information about the active site (Antão and Malcata, 2005). Proteases are classified into six categories based on the residues within the catalytic site: serine protease, aspartic protease, a cysteine protease (thiol), metalloprotease, glutamic protease, and threonine (Chi *et al.*, 2009; Li *et al.*, 2013; Mienda *et al.*, 2014). The name of each group is based on the presence of nucleophilic residue in the active site. The seventh clan of proteolytic enzymes, asparaginase, was recently identified in 2011 (Rawlings *et al.*, 2016; 2011).

The arrangement of these residues defines the enzyme-substrate interaction as well as its specificity. The latter determines the position at which the protease will hydrolyse the peptide bond (Tavano, 2013). In addition, protease specificity provides an option for a better choice of a protease. For example, different proteases can produce different hydrolysates from the same protein substrate. Kamnerdpetch *et al.* (2007) hydrolysed potato pulp using four different proteases (exopeptidase corolase, endopeptidase alcalase, Novo Pro-D, and exopeptidase mix-flavourzyme) and produced completely different hydrolysates.

Proteases can also be classified on the basis of pH optimum: acidic (pepsins), basic (alkaline), or neutral (e.g. calpains) (Abraham *et al.*, 2014; Gupta *et al.*, 2002), substrate specificity (e.g. collagenases, keratinases, elastase) (Inácio *et al.*, 2014; Rao *et al.*, 1998), as well as the source (e.g. scytalidoglutamic peptidase from Scytalidium lignicola).

University of Fort Hare
Together in Excellence

1.1.1 Serine protease

The arrangement of residues in the serine protease active site allows these enzymes to be grouped into 16 clans with 46 families. Clans PA, SB, SC and SF share a common catalytic triad, with serine (nucleophile), aspartic (electrolyte) and histidine (base), to hydrolyse peptide bonds (Ekici *et al.*, 2008). The amino acid sequences of these clans have shown no similarity. However, the geometric arrangement of the folded tertiary structure employs a similar catalytic triad. This provides evidence that there are four evolutionary origins of the serine proteases (Rawlings *et al.*, 2016; Rawlings and Barrett, 1994a).

In contrast, the other clans prefer a catalytic dyad in which serine residue pairs with Hys/Lys residues. Two clans that possess a catalytic triad different from the ones above include clans SH (His, Ser, His) and SS (Ser, Glu, His). Examples are cytomegalovirus assembling and L, D-carboxypeptidase. Another unique feature of serine proteases is the conservation of glycine residues in the vicinity of the catalytic serine (Gly-Xaa-Ser-Yaa-Gly) (Di Cera, 2009; Rawlings *et al.*, 2016).

Serine proteases usually display a maximum activity at neutral or highly alkaline pH, with an optimum pH between 7 and 11, and are irreversibly inhibited by PMSF, DFP, and TLCK. Their isoelectric point ranges from 4-6 and the molecular weight from 18-35kda, except for the case

of *Blakeslea trispora* which has a serine protease with a molecular weight of 126kda (Gupta *et al.*, 2005; Mienda *et al.*, 2014; Rao *et al.*, 1998).

1.1.2 Aspartic protease

Aspartic proteases are referred to as 'acidic' proteases as they are maximally active at low pH (pH3-4). Their catalysis relies on the two highly conserved aspartic residues which actuate a water molecule to perform a nucleophilic attack. Fifteen families are recognised and distributed within seven clans (Rawlings *et al.*, 2016). Their isoelectric point ranges from pH 3 - 4.5 and the molecular weight from 30 – 45 kDa. Pepsins are the most studied aspartic proteases in the A1 family. They possess a bilobal structure with an active site cleft located between the lobes. The active site is usually found within the motif Asp-Xaa-Gly, in which Xaa can be serine or threonine (Rao *et al.*, 1998).

Aspartic proteases are inhibited by pepstatin and are also susceptible to diazo ketone compounds such as DAN (diazo acetyl-DL-norleucine methyl ester) and 1.2-epoxy-3 (p-nitro phenoxy) propane (EPNP) in the presence of copper ions (Rawlings and Barrett, 1995).

1.1.3 Cysteine protease

Cysteine proteases are also referred to as 'thiol' proteases. They contain histidine and cysteine, constituting a catalytic dyad. The order of Cys and His (Cys-His and His-Cys) residues differ among the families (Barrett, 1994). According to Rawlings *et al.* (2016), about 72 families, distributed among 13 clans, are recognised. Broadly, they are divided into four groups: (1) papain-like, (2) trypsin-like (specific to arginine), (3) specific to glutamic acid, and (4) others. Papain is an example of well-studied cysteine protease.

University of Fort Hare
Together in Excellence

Cysteine proteases are optimally active at neutral pH, even though some have displayed maximum activity at acidic pH, for example, lysosomal protease. They are sensitive to sulfhydryl agents such as PCMB, but are unaffected by DFP and metal-chelating agents such as EDTA (Rao *et al.*, 1998; Rawlings and Barrett, 1994b).

1.1.4 Metalloprotease

This is the most diverse group in terms of catalytic type, with enzymes from a wide range of origins. Similar to aspartic protease, a nucleophilic attack is by water molecule which is activated by a divalent metal, usually zinc ion (other metals include copper ions, cobalt, nickel, and iron) (Rawlings *et al.*, 2016; 2011; Rawlings and Barret, 1995). Metalloproteases are grouped on the basis of their amino acid nature which constitutes the metal binding site; for example, Clan MA has a unique HEXXH-E motif feature that differentiates it from the rest of the clans. This feature includes thermolysins, mycolysins, seralysins, and so on. Clan MB corresponds to the motif HEXXH-H, e.g. metzincin). Currently, 69 families, grouped into 16 clans, are recognised, including 12 metalloproteases with unknown functions (Rawlings *et al.*, 2016). Broadly, they are categorised based on their specificity:

Neutral- specific to hydrophobic amino acids

Alkaline – Very broad in specificity

Myxobacter 1- Small amino acids residues on either side of cleavage bond

Myxobacter 2 - Specific for lysine residue on the amino side of the peptide bond.

Metalloproteases are inhibited by EDTA but are unaffected by sulfhydryl agents or DFP (Rao *et al.*, 1998).

1.1.5 Threonine protease

Threonine proteases were discovered as the fifth catalytic type in 1997 after the proteasome structure was solved. It was found that threonine residue at the N-terminal acts as a nucleophile to perform catalysis (Buller and Townsend, 2013; Oda, 2011). Catalysis occurs in a two-step reaction:

a. The secondary alcohol (nucleophile) attacks the peptide to form an acyl-enzyme intermediate.
b. Water molecules hydrolyse the intermediate, releasing the free enzyme and the final product. In the case of ornithine acyltransferase, the substrate executes the second nucleophilic approach and leaves with the acyl group (Rawlings *et al.*, 2011).

Five families categorised into two superfamilies (clans) are recognised. These evidently represent two independent, convergent evolutions of the same catalytic site (Rawlings *et al.*, 2016).

1.1.6 Glutamic protease

Glutamic protease discovery began around 1972 with the work of Murao and co-workers (Murao *et al.*, 1972) whose purification and characterisation studies resulted in the classification of these enzymes as the A4 family of aspartic proteases (Murao *et al.*, 1972; Oda and Murao, 1974; Sims *et al.*, 2004). However, the structure and catalytic mechanism analysis showed this type of proteases was a novel protease family because of the presence of nucleophilic glutamate residue (Fujinaga *et al.*, 2004; Sasaki *et al.*, 2012; Sims *et al.*, 2004).

University of Fort Hare
Together in Excellence

Two families, G1 and G2, have been recognised by MEROPS. The G1 family has a catalytic dyad consisting of glutamic and glutamine residues, whereas the G2 family catalytic diad consisting of glutamic acid and aspartic acid (Rawlings *et al.*, 2016). These proteases have been shown to be optimally active at pH 2 when casein is used as a substrate. They are insensitive towards pepstatin, S-PI (acyl pepstatin), and DAN, but inhibited by EPNP (Jensen *et al.*, 2010; Murao *et al.*, 1992, Oda *et al.*, 1987, Oda and Murao, 1974).

1.1.7 Asparagine peptide lyase (proteolytic enzyme but not protease)

The existence of this seventh class of proteolytic enzymes was shown after the crystal structure of autotransporter from Escherichia coli revealed asparagine as a nucleophile. It was reported that the ability of asparagines to form a five-membered ring structure could result in the breaking of its own peptide bond by means other than hydrolysis (Rawlings *et al.*, 2011; Tajima *et al.*, 2010). Rawlings *et al.* (2011) classified this type, based on the enzyme nomenclature, as lyases (EC 4.3) as their action resembles that of amide lyases (C-C, C-O, and C-N). However, IUBMB has currently grouped asparagine-containing proteolytic enzymes as EC 3.5 (Boyce and Tipson, 2001). This includes those enzymes specific for linear and cyclic amides as well as amidines.

1.2 Applications of protease

Proteases have been applied in a variety of sectors such as the development of therapeutics, in the detergent, food and detailing industries (leather), and in silver recovery. Commercial

proteases can be grouped as industrial and/or therapeutic enzymes based on their application (Truter, 2014).

1.2.1 Therapy

Modern drug development aims to identify two main macromolecules, namely enzymes and G-protein coupled receptors (Rengasamy *et al.*, 2014). Because pathogenic organisms are resistant to synthetic drugs, enzymes have received much attention as they are able to conduct a specific reaction (Li *et al.*, 2012). They are studied to explore their biochemical features and structure-function relationship to evaluate their potential as therapeutics (Craik *et al.*, 2011; Rani *et al.*, 2012).

Protease enzymes are used in the treatment of many disorders, such as cancer, infectious diseases, inflammation, cardiovascular conditions, necrotic wounds, and so on (Sawant and Nagendran, 2014; Truter, 2014). Craik and co-workers (2011) reported about twelve commercially protease drugs, including urokinase, alteplase, thrombin, activated protein C, botulinum toxin (A and B), and zenpep. Many reports have focused on proteases with fibrinolytic activity (thrombolytic), especially from the Bacillus genus. These include nattokinase (NK), subtilisin DFE and DFE DJ-14 (Cheng *et al.*, 2015; Inácio *et al.*, 2015; Li *et al.*, 2013). Other proteases are extensively used in the pharmaceutical industry in ointments for wound debridement. A thermostable serine protease identified from a medicinal plant, *Wrightia tinctoria*, was shown to be directly involved in the wound-healing process (Yariswamy *et al.*, 2013).

University of Fort Hare
Together in Excellence

1.2.2 Detergent

Proteases have been used in detergents since 1913 when pancreatic extracts and sodium carbonate were incorporated in the first detergent, "Burnus" (Gupta *et al.*, 2002; Inácio *et al.*, 2015; Rao *et al.*, 1998). Protease application in detergents showed substantial growth after the first microbial alkaline protease, Carlsberg (Raval *et al.*, 2014). Currently, detergent proteases have taken a major share of the total industrial enzyme market (Feijoo-Siota and Villa, 2011). There are two features that make an enzyme useful in the detergent industry, namely, the alkaline pH, and its compatibility with current detergents (Anwar and Saleemuddin, 1998; Kirk *et al.*, 2002).

All enzymes used as detergent additives in the market are from Bacillus sp. These include *Bacillus licheniformis* and *Bacillus amyloliquefaciens*. The search for new proteases with more potential is ongoing, especially from the unexplored environment. Elhoul *et al.* (2015) isolated a novel detergent-stable and solvent-tolerant serine thiol alkaline protease from *Streptomyces kayongensis* TN650. An extracellular thermostable serine alkaline protease from *Caldicoprobacter guelmensis*, isolated from a hydrothermal hot spring, was found to be optimally active at 70°C and pH 10. The enzyme showed compatibility with commercial detergents and stability at 80°C and 90°C for 180 and 60 mins (Bouacem *et al.*, 2015). A thermo- and halo-stable alkaline protease isolated from *Bacillus alveayuensis* CAS5 showed potential for application in the detergent industry. Mokashe *et al.* (2015) suggested that protease from halotolerant *Jeotgalicoccus* sp. could be a better bio-additive agent for the detergent industry.

1.2.3 Dairy industry

The major application of protease in this field is in cheese production. Usually, cheese is made by curdling the milk protein (casein), using protease enzymes from calf rennet. Calf rennet is a complex of two acid proteases, chymosin and pepsin, isolated from the fourth stomach. Chymosin is often found in a higher ratio than pepsin. Various factors, such as animal rights and the high cost of this enzyme complex, have led to the search for new protease enzymes from plants and microorganisms (Chaplin and Bucke, 1990).

Aspartic proteases with a milk-clotting activity have been reported from several plant species, such as oryzasin (from rice seeds) and cardosin A (from cardoon) which resemble calf rennet coagulant (González-Rábade *et al.*, 2011). Cucumisin has also shown more potential compared to other cysteine proteases, such as papain which produced bitter-tasting peptides (Asif-Ullah *et al.*, 2006; Gupta *et al.*, 2002; Sewant and Nagendran, 2014). There have been challenges with microbial proteases such as low yield, non-specificity, and heat-stability, which led to cheese becoming bitter during storage. Proteases from *Mucor miehei* exhibited activity during cheese maturation, producing bitter off-flavours. Use of oxidizing agents, such as hydrogen peroxide and peracids, modified methionine amino acids into their sulfoxides, reducing temperature stability by 10°C (Chaplin and Bucke, 1990). It has been reported that proteases from *Pseudomonas flourescens* R098 that hydrolyse the peptides responsible for the bitter taste in cheese are used as debittering agents (Sumantha *et al.*, 2006).

1.2.4 Baking industry

In the baking industry, proteases are applied in the hydrolysis of the insoluble protein, gluten, which influences the properties of the dough. Bromelain has been used to reduce hypersensitivity reactions to wheat flour (Feijoo-Siota and Villa, 2011). Proteases (exo-and endo-) from *Aspergillus oryzae* have been shown to modify wheat gluten by limited proteolysis (Sewant and Nagendran, 2014). Neutrase is another commercial protease used to degrade flour proteins for baking biscuits, cookies and crackers (Sumantha *et al.*, 2006). Enzymatic treatment of dough is associated with reduced time for mixing, improvement of dough quality, and increased loaf volumes.

University of Fort Hare
Together in Excellence

1.2.5 Meat tenderisation

The frequently used protease enzymes in the meat tenderisation process are of plant origin. These include bromelain (from *Ananas comosus*), papain (*Carica papaya*), and the less employed ficin (*Ficus carica*) (Maiti *et al.*, 2008). Search for other proteases continues, with zingibain (from ginger) and actinidin (from kiwi) having been identified recently (Bekhit *et al.*, 2013). Other protease sources have been reported, including bacteria and fungi. *Aspergillus oryzae* aspartic protease and Bacillus subtilis alkaline elastase were approved as GRAS (Generally recognized as safe). However, the *Aspergillus oryzae* protease showed poor hydrolysis of collagen, whereas elastase exhibited poor hydrolysis of myofibrillar proteins (Calkins and Sullivan, 2007).

1.2.6 Hydrolysis of protein

Proteases can hydrolyse proteins from various sources such as fish, plants, or animals to obtain hydrolysates with a wide range of applications. Hydrolysates produced by a commercial protease, alcalase, from casein, whey protein, and soy protein have been applied

in infant formulas. They are also used as nutrient supplements in juices and therapeutic diets (Ellaiah *et al.*, 2002). Some commercial proteases produced by Novozymes include Molsin F, Kojizyme, and Flavourzyme (Sumantha *et al.*, 2006).

1.2.7 Silver recovery

Previously, silver was recovered from x-rays by burning the films, followed by electrolytic stripping of the gelatin layer, a process which has a negative environmental impact since the base film could not be recycled. The introduction of proteases has played a crucial role in recovering the silver: these enzymes hydrolyse the gelatin layer embedded with silver. Alkaline protease from *Bacillus subtilis* hydrolysed gelatin within 30 min at 50-60°C, liberating silver (Gupta *et al.*, 2002; Sawant and Nagendran, 2014). Rathod and Pathak (2014) characterised alkaline protease from *Bacillus alcalophilus* LW8. This protease was effective in the decomposition of the gelatin layer in x-ray film.

1.2.8 Dehairing industry

Conventionally, animal hair in the leather industry is removed by a process called 'liming' which produces a hazardous effluent. In the liming process, hair is degraded by a saturated solution of lime and sodium sulfide, an unpleasant and expensive process expensive. Proteases can be employed as a substitute for the sulfides (Anwar and Saleemuddin, 1998; Jegannathan and Nielsen, 2013). Abraham *et al.* (2014) showed the efficacy of alkaline protease in dehairing cowhide, and Najafi *et al.* (2005) isolated a protease from *Pseudomonas aeruginosa* PD100, with potential applications in the leather industry.

1.3 Sources of protease enzymes

Because of their physiological significance, proteases are naturally ubiquitous and are found in all cellular life kingdoms, from viruses to humans (Cheng *et al.*, 2015; Di Cera, 2009; Mótyán; 2013; Rao *et al.*, 1998).

University of Fort Hare
Together in Excellence

1.3.1 Plant proteases

Plant proteases have been extracted from a variety of plant tissues such as seeds (*Arabidopsis thaliana*, rice, barley, hemp seed, cucumber and squash), potato tubers and leaves, tomato plant leaves, maize pollen, and from the latexes of several plant families such as *Asteraceae, Asclepiadaceae, Apocynaceae, Caricaceae, Moraceae,* and *Euphorbiaceae* (González-Rábade *et al.*, 2011).

Proteases isolated from plants are attractive because they have wide substrate specificity and are active at a range of temperatures and pH levels (González-Rábade *et al.*, 2011). Some plant proteases employed industrially include papain from pawpaw (*Carica papaya*), bromelain from pineapple plants (*Ananas comosus*), and ficsin from *Ficus racemosa*, *Ficus glabrata*, and *Ficus carica*. Papain from *Pachyrizhus erosus* has shown activity higher than *Carica papain* (González-Rábade *et al.*, 2011). Cucumisin extracted from the melon plant (*Cucumis melo*) has been the best serine protease, with its homolog from Kachri fruit (*Cucumisin trigons* Roxburgh) used as a meat tenderizer (Antão and Malcata, 2005; Asif-Ullah *et al.*, 2006).

Proteases are significant in plant physiological processes and development such as protein turnover, enzyme modification and regulation of gene expression, defence, senescence, and

programmed cell death, as well as proteolysis of misfolded proteins (Iketani *et al.*, 2013; Raimbault *et al.*, 2013). Aspartic proteases have been reported to be involved in the drought resistance of the plant cultivar in common beans and cowpeas (González-Rábade *et al.*, 2011). Overexpression of leucine aminopeptidase in some plants has been associated with stress conditions such as osmotic stress and infections. Cysteine protease from corn callus was shown to be associated with the inhibition of larval growth of fall armyworm (Jiang *et al.*, 1995).

Challenges in using plants as producers are land availability, climate fluctuation and time; growing, reaping and preparing the plants is a time-consuming process (Moo-Young, 2011; Rao *et al.*, 1998).

1.3.2 Animal protease

The best-studied animal proteases are those of the digestive tract: trypsin, chymotrypsin, pepsin, and rennin. Trypsin is an intestinal enzyme used to develop bio-control of insects. Pure chymotrypsin is one of the most expensive enzymes and as a result, it has limited applications and is used only in analytical and diagnostic processes. Rennin is usually isolated from the stomach of a calf. It is a pepsin-like enzyme used to de-allergize milk (Singh *et al.*, 2016; Rani *et al.*, 2012; Rao *et al.*, 1998). Other proteases of animal origin include thrombin, Factor Xa, enteropeptidase, elastase, carboxypeptidase (A&B), and cathepsins (Sawant and Nagendran, 2014). The major hurdle with regard to animal proteases is the issue of ethical clearance, and the limited number of animals available for slaughter.

1.3.3 Microbial protease

Commercially, most proteases used are from microorganisms (Inácio *et al.*, 2015) with microbial proteases accounting for approximately 40% of the total enzyme market (Rani *et al.*, 2012). The key advantages of microorganisms as a source include their diversity and ubiquity, their short cultivation time, and the higher number of desired proteases (de Castro *et al.*, 2015). Furthermore, their production is inexpensive, and they can be genetically manipulated. The protease enzymes are extracellular, thereby minimizing the downstream processing (Sawant and Nagendran, 2014). Additionally, microbial proteases are diverse with regard to specificity and are active at a wide range of pH levels (Rani *et al.*, 2012).

The popular, productive genera include *Bacillus* species, *Aspergillus* species, and *Saccharomyces* (also called *Kluyveromyces*) (Chaplin and Burke, 1990). Bacteria are recognized as the key source of alkaline proteases, accounting. They account for about 60% of detergent enzyme sales, with *Bacillus* species being dominant. Proteases (mainly subtilisins) used as additives in household detergents are from *Bacillus* sp. with *Bacillus licheniformis, Bacillus pumilis* and *Bacillus subtilis* being the primary sources for the Carlsberg-type enzyme (Mienda *et al.*, 2014).

In contrast, fungal proteases are produced in variety (basic, neutral and acidic) as seen in the case of *Penicillium* species and *Aspergillus oryzae*. However, they are less thermostable and have a lower reaction rate than bacterial proteases (Inácio *et al.*, 2015; Rani *et al.*, 2012).

University of Fort Hare
Together in Excellence

1.3.4 Seaweeds: a potential source of protease enzymes

About 70% of the earth's surface is covered by the ocean, which contains a variety of organisms ranging from unicellular to multicellular (Kulshreshtha *et al.*, 2015). However, compared to terrestrial habitats, the marine environment is little exploited and therefore, provides an opportunity to discover natural products that can be beneficial to humans (Rengasamy *et al.*, 2014). Screening organisms that live in unexplored habitats can greatly facilitate the search for protease enzymes (Raval *et al.*, 2014).

Seaweeds (also known as marine algae) are plant-like, multicellular, macroscopic algae found on rocky shores. They are classified on the basis of pigmentation into three main phyla: Phaeophyta (brown), Chlorophyta (green), and Rhodophyta (red) (Dawes, 1997). Usually, seaweeds are found attached to rocks and serve as shelter and food for marine organisms like fish and sea urchins (Mumtaj *et al.*, 2015). They are affected by the physical features of their habitat and rely on continuous water movement for nutrients and gases. Sometimes they are exposed to diseases caused by opportunistic microorganisms (*Salmonella* and *Vibrio* species) where fishes are known to be the carriers (Pal *et al.*, 2014; Thanigaivel *et al.*, 2015).

Seaweeds found in the intertidal zone are likely to be subjected to adverse conditions, such as hypo-osmotic shock, desiccation, salt stress, high levels of solar radiation and ultraviolet rays (Collen *et al.*, 2006). However, seaweeds have adapted to various abiotic stresses from their environment (Pal *et al.*, 2014): they have developed defence mechanisms against these conditions and certain diseases. Such defence mechanisms are mediated by specific proteins and a variety of structurally diverse secondary metabolites including polysaccharides, vitamins, minerals, and polyphenols (Gupta and Ab-Ghannam, 2011; Hu *et al.*, 2012; Rengasamy *et al.*, 2014). Mumtaj (2015) reported a high concentration of enzymes, nitrate reductase and glutamine synthetase in *Bryopsis, Turpanaria gonaidae* and *Gracilaria folifera*, respectively. There has been less information on proteases from algae. Although proteases have been described in some green micro- and macro-algae (Kadokami *et al.*, 1990; Matsubara *et al.*, 1998, 1999, 2000, and 2002), only three studies have focused on red seaweeds (El-shora *et al.*, 2016; Pérez-Lloréns *et al.*, 2003; Wang *et al.*, 1998).

University of Fort Hare
Together in Excellence

In the evolutionary timeline, seaweeds existed before terrestrial plants, which are recognized as a source of copious bioactive secondary metabolites. Novel compounds from plants have been applied in pharmaceuticals and industry. Seaweeds are referred to as lower cryptogam, seedless plant-like organisms, and have structurally diverse bioactive compounds, certain of which have not been found in higher plants. *Laminariales* and *Ceramiales* have shown some high enzymatic inhibition activity because of the presence of phenols and terpenoids (Rengasamy *et al.*, 2014). Therefore, marine macroalgae have the potential to be an alternative source of proteases with unique features.

1.4 Production by recombinant technology

Recombinant technology refers to the combination of DNA molecules from different sources into a host organism. It has played an important role in the growth of therapy and industry as it is now possible to produce pure enzymes on a large scale. For instance, the direct use of seaweeds as a source of enzymes is hampered by factors such as low production (if the enzyme is expressed), the impurity of the product, time consumption, and cost. For these reasons, cloning and expression of protease in mesophilic hosts is easier and more cost-

effective in terms of high production (Toplak *et al.*, 2013). For example, bromelain from the pineapple plant has numerous applications in industry and has been extracted in low efficiency and is often denatured by oxidation. The use of yeast (*Pichia pastoris*) as an expression system has resulted in high-yield, stable Bromelain (Spohner *et al.*, 2015).

Almost all FDA-approved therapeutic enzymes are recombinant. Extensive development of recombinant technology has been ongoing and has produced different promoters for enhanced production and different expression systems (*Erichia coli, Bacillus subtilis*, and others). There have also been some plant expression systems, such as tobacco and edible rice. Immunoglobulins have been expressed in tobacco plants and used to treat Ebola in Africa (Yao *et al.*, 2015).

1.5 Problem statement

With a 60% share of the global enzyme market, proteases constitute one of the major groups of industrial enzymes, and their application has great potential in various industrial processes. However, the issue of stability for commercial success has been a major hurdle because most proteases are optimally active at physiological conditions. Because of the high cost of current enzymes and environmental issues (in the case of chemicals), there is still a need for new protease enzymes which possess characteristics of interest for industrial application. Screening the organisms that live in unexplored habitats can greatly facilitate the search for new protease enzymes.

Kenton-on-Sea seaweeds are exposed to varying habitat conditions that affect growth and survival: light, since the coast is exposed; a temperature which changes seasonally, with high rainfall during winter; desiccation; a lack of available nutrients during low tide, and salinity. However, these seaweeds have adapted themselves to survive such stresses by developing defence mechanisms mediated by specific proteins and enzymes.

University of Fort Hare
Together in Excellence

1.6 Research hypothesis

Red algae found along the Eastern coastline of South Africa express an active serine protease.

1.7 Aim and objectives

1.7.1 Aim

To analyse hypothetical protease genes in silico and kinetically characterise selected purified protease enzymes from red algae.

1.7.2 Objectives

1.7.2.1 To search available red algae genomes for genes encoding protease enzymes.

1.7.2.2 To clone and express a selected protease gene.

1.7.2.3 To isolate and purify a selected protease enzyme(s) from red seaweed.

1.7.2.4 To characterise the biochemical and kinetic properties of the purified protease.

Chapter 2
In silico analysis of red algae genomes

2. INTRODUCTION

Protein science has entered a new era that promises to elucidate the full functional mechanism of the cell (UniProt: The Universal Protein Knowledgebase). Genome sequencing projects have led to an increase in the number of protein sequences in databases such as UniProt, National Centre for Biotechnology Information (NCBI), and Protein Data Bank (Boutet *et al.*, 2016; NCBI database; Berman *et al.*, 2000). The UniProt database contained about 60 million protein sequences in 2016, a number that increased in 2017 to more than 80 million, with InterPro62 covering about 80% of the UniProt database, and gene ontology terms for more than 48 million proteins predicted (Finn *et al.*, 2016). The Enzyme Function Initiative-Enzyme Similarity Tool has been updated to utilise 85 827 304 UniProt sequences in 2017 (TrEMBL=85 272 789 and Swiss-Prot=554 515) (Gerlt, 2017).

Although the number of protein sequences in databases is rapidly increasing (Figure 2.1), the major challenge is assigning the biochemical and cellular functions of the gene products (Mills *et al.*, 2015); at present, most gene entries retain only an ambiguous functional annotation (putative, hypothetical, or unknown) (Hawkins and Kihara, 2007). More than 48 million hypothetical sequences were labelled by GenBank in NCBI (Yandell and Ence, 2012). *Homo sapiens* contain about 40% of unknown proteins, whereas bacteria and yeast consist of 16 and 30% hypothetical proteins, respectively. Model plant systems, *Arabidopsis thaliana* and *Oryza sativa* (rice) consist of about 30-40% protein of unknown functional domain (Dhanyalakshmi *et al.*, 2016; Martinez, 2013).

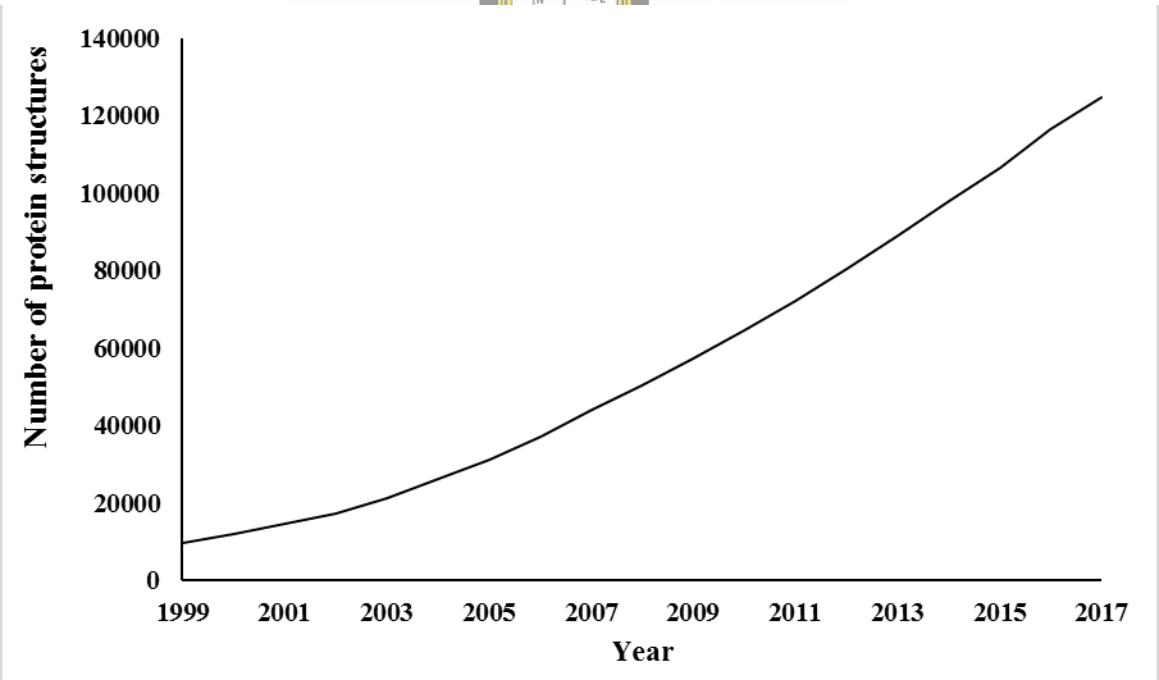


Figure 2.1: The growth of protein structures per year (Graph was constructed in excel based on data obtained from PDB database (Rose et al., 2017).

The Protein Data Bank (PDB) database contains about 125 116 (as of August 2017) protein structures from structural genomics projects such as Protein Structural Initiative (PSI) (Gabanyi *et al.*, 2013; Rose *et al.*, 2017).

Therefore, predicting unknown protein function has become an essential tool in bioinformatics (Hawkins and Kihara, 2007, Juncker *et al.*, 2009). Bioinformatics is the application of computational, statistical, and mathematical tools to an automated biomolecular information processing and analysis process, particularly in genomics. The availability of accessible genome sequences and utilisation of bioinformatics tools has made it possible to facilitate computational approaches and elucidate protein function(s) effectively (Xu *et al.*, 2003). One of the ultimate goals of bioinformatics is to deduce functional data for the gene entries generated by genome sequencing projects (Xu *et al.*, 2003). Computational prediction has contributed greatly in biology, not only in the annotation process, but also in the analysis of experimental information (e.g. gene expression by microarray). Prediction of a protein function depends on the number of factors, such as protein sequence, co-occurrence of genes across multiple genomes, protein structure, proteomics-experiment based, associated-based, and combined methods (Hawkins and Kihara, 2007).

The most frequently used computational approach for function assignment is comparative (or homology) modelling which can be based on sequence, structure and/or sequence-structure methods:

- Sequenced-based method – uses sequence alignments to determine the function;
- Structure-based methods – depends solely on the structure fold for functional data;
- Sequence-structure-based method – utilises the information from the sequence similarity and catalytic site residues.

2.1 Sequence-based approach

The sequence-based approach employs sequence alignments to transfer functional information from protein sequences with high similarity and also to trace their phylogeny. Primarily, the FASTA and Basic Local Alignment Sequence Tool (BLAST) suite (BLAST and PSI-BLAST), provided by the NCBI, the European Molecular Biology Laboratory-European Bioinformatics Institute (EMBL-EBI) and UniProt, are used as gateways for recognising homology between query protein sequences and known protein sequences from various databases (Dhanyalakshmi *et al.*, 2016). There are also functional motif tools such as SMART (Letunic *et al.*, 2017), PRINTS (Attwood, 2000), Pfam (Finn *et al.*, 2016), Superfam (Gough *et al.,* 2001), Gene3D, and others. Using different sequence-based methods can result in better annotation (Hawkins and Keihara, 2007). Tools that can incorporate all of these resources are being developed, for an example, Interpro from EMBL-EBI (Jones *et al.*, 2014).

The fundamental advantage of sequence-based methods is the ability to predict the functional data in proteins that may not be related with regard to folding (Dukka *et al.*, 2013). Based on various sequence comparisons, it has been proposed that sequences with 40% similarity can have the same function by transferability (Lee *et al.,* 2007). However, there are exceptions; for instance, viral capsid protein (1PIV:1) and glycosyltransferase (1HMP: A) share 44% sequence identity but the function and the structure are different (Figure 2.2).

University of Fort Hare
Together in Excellence

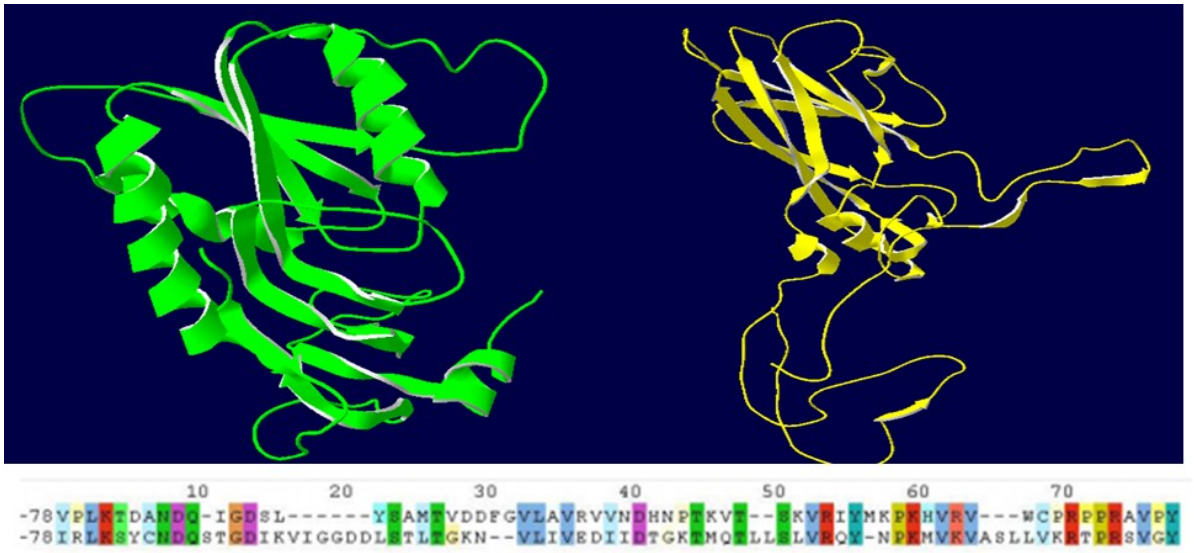


Figure 2.2: Comparison between viral capsid protein and human Glycosyltransferase

If sequence identity dips below that percent, the sequence-based approach becomes ineffective. This is often observed in cases where two proteins have convergently evolved: a catalytic site becomes more conserved than global sequence, for example, *V. stercoraria* (bacteria, 1VHB) and *P. marinus* (eukaryote, 2LHB) haemoglobins share 8% sequence identity.

2.2 Structure-based approach

Structure-based prediction methods have been developed to the extent that functional information can be obtained efficaciously (Mills *et al.*, 2015; Hawkins and Kihara, 2007). The protein function is usually inherently associated with its structure (Hawkins and Kihara, 2007). Structure-based methods can broadly be categorised as global (fold-based) or local (based on substrate-binding sites) based methods.

University of Fort Hare
Together in Excellence

2.2.1 Global structure-based method

For uncharacterised protein (lacking function and experimental structure), a structure predicted by computational methods can be helpful in that it possesses some signatures for predicting function (Hawkins and Kihara, 2007). The growth of available protein structures in databases like PDB (Rose *et al.*, 2017), and the development of advanced protein structure prediction tools have improved the prediction of protein structure: a protein structure can be predicted from the alignments (Dukka, 2013). Proteins sharing 30% similarity are considered to have a similar fold and it has been observed that protein folds are more conserved than sequence (Hawkins and Kihara, 2007). Methods for protein structure prediction are becoming more readily available and accurate and include homology modelling, threading, and *ab initio* predictions. The most frequently used software for homology modelling includes Modeller (Eswar *et al.*, 2006) and Swiss pdbViewer (SPDBV). Various web servers are available for template-based prediction, including Primo, Phyre2, Swiss-model, and others.

The major challenge for predicting function from the structure is the lack of experimentally affirmed protein structures. Available experimentally proven protein structures with reliable

function occupy about 2% of the total number of structures; the remaining proteins are computationally predicted (Tiwari *et al.,* 2012).

Although protein fold is more conserved than sequence, various protein families (TIM barrel, Ferredoxin and Rossman folds, etc.) have shown that global fold similarity does not always correlate with functional similarity. Other most common motifs include β-meanders, which are observed in a variety of diverse folds (Lee *et al.*, 2007). Owing to these super folds, it is unlikely to be able to assign a function for a novel protease structure since similar fold can have a different function. An example is seen in signal transduction protein cheY (1YMV), electron transport flavodoxin (1FLA), and mannose transporter (1PDO) which have similar folds but different functions (Figure 2.3).

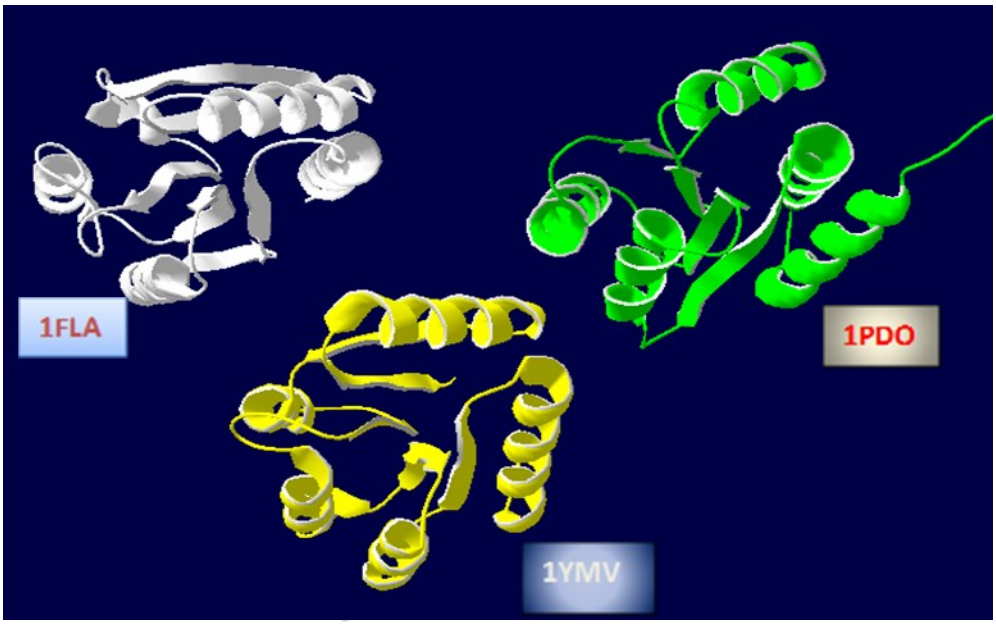


Figure 2.3: The common fold of Flavodoxin in different protein structures with different functions

Global structure alignments do not necessarily transfer functional data since point mutations in an active/binding site may result in divergent evolution. Baird *et al.* (2006) converted a serine protease into a functional threonine protease by mutating a nucleophilic serine195 residue into a threonine195 residue.

2.2.2 Local structure-based method

Local structure prediction methods are based on the fact that enzymatic catalysis is conducted by a few residues that constitute an active site (Lee *et al.*, 2007). Therefore, for function inference through evolution, catalytic residues must be conserved. Several methods have been developed such that specific enzymatic function can be retrieved efficiently from the local-fold similarity. One of the common methods includes aligning the query structure with the template with known catalytic residues. This method is sometimes followed by docking.

When predicting, mutual distances and angles of catalysis amino acid residues are taken into account (Hawkins and Kihara, 2007). These methods can be geometric-based and/or energetic-based. The geometric-based method predicts the catalytic residues by searching for pockets/ cavities in a protein structure. These include, but not are not limited to: CASTp, SURFnet, Pocket-Finder, LIGSITE, FPOP and POCKPICKER. The energetic-based approach

identifies functional residues by employing a variety of interaction energies (Dukka, 2013) such as Van der Waals, and electrostatic interactions. These include eF-Site, Q-SiteFinder, Pocket-Finder, MSDsite.

2.2.3 Combined methods (sequence-structure-based methods)

The combined-method approach has emerged from the strengths of each approach (global and local-fold prediction methods) to improve the functional assignment of unknown proteins (Mills *et al.*, 2015). There is an increasing number of methods that use both sequence and structure properties. These include ConCavity, ProFunc, and ProKnow. ProFunc utilises BLAST and HMM searches to predict the function of protein with known 3D structures. It uses both sequence and structure. ProKnow assigns a function, based on the 3D fold, motifs, and sequence, via ProKnow knowledgebase of features. ConCavity scans protein surfaces for binding pockets and locates them by evolutionary trace (Lee *et al.,* 2007).

2.3 Problem statement

Red algae have been identified as a potential source of bioactive compounds and are, therefore, of high scientific interest. About eight red algae species have been sequenced: however, the function of most of the protease genes remains equivocal.

2.4 Aim

To identify hypothetical protease sequences from the red algae and predict their function.

2.4.1 Objectives

2.4.1.1 To search for protease sequences

2.4.1.2 To analyse in silico the identified protease sequences

University of Fort Hare
Together in Excellence

2.5 Methods

The following experimental design (Figure 2.4), was used to identify the genes encoding for proteases in the published genomes sequences of red algae.

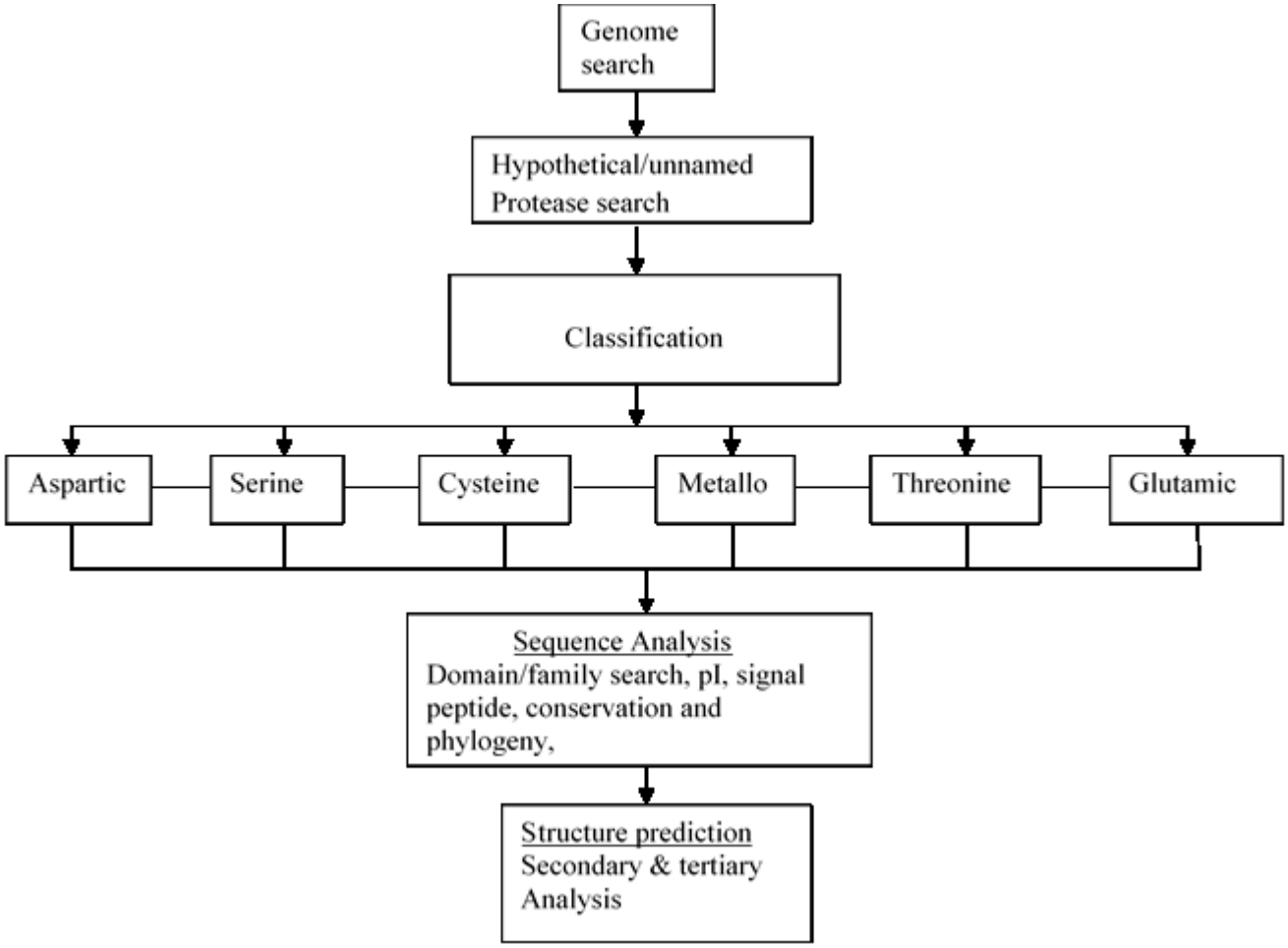


Figure 2.4: The highlights of the overall methods for protease identification and function prediction.

2.5.1 Genome search

The red algae genome sequences were searched using the literature and various databases, including NCBI, and specific red algae databases.

2.5.1.1 National Centre for Biotechnology Information (NCBI) genome search

The National Centre for Biotechnology Information (www.ncbi.nlm.nih.gov) is part of the United States National Library of Medicine (NLM), which is a department within the National Institute of Health (NIH), and is a resource for bioinformatics tools. It consists of a number of databases, including GenBank, which is connected to the DNA Data Bank of Japan (DDBJ), the European Molecular Biology Laboratory (EMBL), the protein databank, and the Conserved protein Domain Database (CDD). All these databases are accessed through a search engine called Entrez. Genome sequences were searched by typing “red algae genomes”, “seaweed genome”, and “red algae protease genes”.

2.5.1.2 Literature search

Different search engines were used to search for red algae genomes from various literature databases and journals.

2.5.1.2.1 Connecting Repositories (CORE) (https://core.ac.uk/search)

The Connecting Repositories system is a search engine that provides open access to over 10 million research papers and connects a huge number of repositories (Knoth and Zdrahal, 2012). For this study, we text-searched three journals (including PLOS ONE), and 14 repositories (including Springer, Elsevier, Frontiers, NORA (Norwegian Open Research Archives)), and others. The terms used were "red algae genome sequence", and "red algae sequences".

2.5.1.2.2 Bielefeld Academic Search Engine (BASE) (https://www.base-search.net/)

Bielefeld Academic Search Engine is one of the copious search engines and provides access to more than 100 million papers from more than 5000 sources. We used the same terms mentioned above to search for publications on algal genomes.

2.5.2 Protease sequence search

Various approaches were used to search protease sequences from the identified red algae genomes.

2.5.2.1 NCBI protease search

The protease sequences were searched for by typing each protease's class name into the search box, for example, "serine protease" or "aspartic protease from red algae". The search was filtered to focus on protein sequences only.

University of Fort Hare
Together in Excellence

2.5.2.2 Kyoto Encyclopedia of Genes and Genomes (KEGG) database

The Kyoto Encyclopedia of Genes and Genomes (KEGG) (http://www.genome.jp/kegg/) database is a database resource for functionalities of biological systems from the molecular level. At first, genome annotation was searched by organism. Having entered the annotation, hypothetical protease sequences were searched via text search. Terms such as "hypothetical protease/peptidase" and "peptidase" were typed on the search box.

2.5.2.2.1 MEROPS database

MEROPS (http://merops.sanger.ac.uk/) is a peptidase and inhibitor database. The information on peptidase can be accessed by searching the organism, MEROPS identifier, or its name. We searched by the organisms we identified as having a full genome sequence (refer to 2.1). The database groups the sequences of each organism according to their hypothetical class.

2.5.2.2.2 Basic Local Alignment Sequence Tool (BLAST) search

Basic Local Alignment Sequence Tool (BLAST) searches (https://blast.ncbi.nlm.nih.gov/Blast.cgi) for similarity regions between the sequences. It enables comparison between query and database sequences.

2.5.2.2.3 MegaBLAST tool

MegaBLAST was used to scan genomic scaffolds by downloading the nucleotide sequences and scanning them against the database. This approach was time consuming and often produced undesired results and was therefore discontinued.

2.5.2.2.4 BLASTp

Putative protease sequences obtained from KEGG or MEROPS were used as a query to search for more similar red algae sequences using the BLASTp programme at NCBI.

2.5.2.2.5 ViroBLAST

The ViroBLAST tool was used to scan protein contigs available in databases other than NCBI and UniProt. Protease sequences identified from the databases were used as a query. ViroBLAST home is https://indra.mullins.microbiol.washington.edu/viroblast/viroblast.php.

2.5.2.2.6 Alga-PrAS

Algal protein annotation suite (Alga-PrAS) is an analytical proteome database for algae. The similar protease sequences were searched from other algae by copying and pasting a query sequence (the hypothetical protease sequence) under "sequence search". The programme used was BLASTp.

2.5.3 Sequence analysis

Following identification of the protease sequence, a representative protease sequence from each class was selected for further analysis.

2.5.3.1 Domain/family search

The protease sequences were then submitted to various databases for domain and family searches (Table 2.1). The accession numbers for domains or families were used as evidence and to group the sequences that belonged to the same family.

University of Fort Hare
Together in Excellence

Table 2.1: Databases for functional motifs and families

Database	Available site
Pfam	http://pfam.xfam.org/
InterPro	https://www.ebi.ac.uk/interpro/
NCBI CDD	https://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi
PRINTS	http://130.88.97.239/cgi-bin/dbbrowser/PRINTS/printsBLAST.cgi
Prosite	https://prosite.expasy.org/scanprosite/
SUPERFAM	http://supfam.org/SUPERFAMILY/hmm.html
GENE3D	http://gene3d.biochem.ucl.ac.uk/Gene3D/

2.5.3.2 Physical characteristics

Each protease sequence was analysed by Protparam (available at ExPASy bioinformatics resource portal: http://web.expasy.org/protparam/) for molecular weight and isoelectric

point. The Grand Average of Hypathicity (GRAVY) score was predicted using a sequence multiple suite available at http://www.bioinformatics.org/sms2/protein_gravy.html.

2.5.3.3 Signal peptide prediction

The signal peptide sequences and their position were predicted using the PrediSi tool available at http://www.predisi.de.

2.5.3.4 Subcellular localisation

The site of action within the cell was predicted using Wolf-Psort prediction tool at GenScript (https://www.genscript.com/wolf-psort.html). The prediction was limited to plants only.

2.5.3.5 Functional site prediction

The representative proteases were used as query to BLAST search similar protein sequences from the NCBI database. These sequences were then aligned using Clustal Omega v1.2.4 (https://www.ebi.ac.uk/Tools/msa/clustalo/). The putative functional sites were deduced from the alignment following the conservation.

2.5.3.6 Phylogeny analysis

Similar protein sequences were searched using different database resources. The NCBI BLAST search was filtered to match sequences from plants (taxid: 3193), animals (taxid: 33208), red algae (taxid: 2763), and bacteria (taxid: 2). We also accessed other resources (such as SWISS-PROT and TrEMBL) through the Alga-PrAS database search. The resulting, similar sequences were selected on the basis of query cover, identity, expect value, and whether they were experimentally annotated or predicted. The phylogenetic tree was constructed using an online server available at http://www.phylogeny.fr/ and MEGA 7 software. The sequences were aligned with muscle, and the alignment were curated using Gblocks.

University of Fort Hare
Together in Excellence

2.5.4 Structure prediction

2.5.4.1 Secondary structure

The secondary structures of the proteases were predicted using SOPMA https://npsa-prabi.ibcp.fr/cgi-bin/npsa_automat.pl?page=/NPSA/npsa_sopma.html. The parameters were set to default (conformation number=4; threshold = 8; and windows = 17).

2.5.4.2 Tertiary structure prediction

The tertiary structures were predicted on the basis of homology using Primo (https://primo.rubi.ru.ac.za). The templates were searched by BLASTp programme, and the alignment was made using muscle or T-coffee and in pseudo-expresso in 3D mode. The alignments were manually edited before modelling, and the programme was set to generate four models. The best models were chosen based on low DOPE-Z score and high RMSD.

The predicted structures were submitted to the ProFunc server (http://www.ebi.ac.uk/thornton-srv/databases/cgi-bin/pdbsum/GetPage.pl?pdbcode=index.html) for function inference using templates and GO terms. The ProFunc is also connected to PDBsum (http://www.ebi.ac.uk/thornton-srv/databases/cgi-bin/pdbsum/GetPage.pl?pdbcode=index.html) quality assessment and topology analysis.

2.6 Results and Discussion

2.6.1 Genome search

The red algae species with full genome sequences were identified from two subclasses of Rhodophyceae, Florideophyceae and Bangiophyceae (Table 2.2). *Laurencia dendroidea* genome sequence was not available, but its estimated genome size was reported by de Oliveira *et al.* (2012). The *Kappaphycus alvarezii* draft genome sequence has not been published but is accessible through the NCBI database. *Porphyridium. purpureum* protein contigs were downloaded from http://cyanophora.rutgers.edu/porphyridium/ and *Pyropia. yezoensis* at http://nrifs.fra.affrc.go.jp/ResearchCenter/5_AG/genomes/nori/.

Table 2.2: Comparison of genomes from various species of red algae

Species/strain	Class	Genome size	No. of genes	GC%	Accession numbers
C. merolae10D	Bangiophyceae	16.5Mbp	5 331	55%	Matsuzaki *et al.*, 2004
G. sulphuraria 017W **G. sulphuraria 107.79*	Bangiophyceae Bangiophyceae	13.7Mbp 12Mbp	6723 _____	37.9% 41.3%	Schönknecht *et al.*, 2013 (Unpublished)
P. purpureum	Bangiophyceae	19.7Mbp	8 355	55.5%	Bhattacharya *et al.*, 2013
P. yezoensis	Bangiophyceae	43Mbp	10 327	63.6%	Nakamura *et al.*, 2013 Not available @NCBI
P. umbilicalis	Bangiophyceae	87.7Mbp	13 125	65.8%	Brawley *et al.*, 2017
C. crispus	Florideophyceae	105Mbp	9 606	52.9%	Collén *et al.*, 2013
***L. dendroidea*	Florideophyceae	853Mbp	_____	____	de Oliveira *et al.*, 2012 (Not available)

*The sequence for *G. sulphuraria*107.79 strain is available on NCBI site but not published in a journal.
** *L. dendroidea* has a draft genome and is not available on databases.

University of Fort Hare
Together in Excellence

Since our aim was focused on protease sequences, nothing much was done on the whole genome sequences. However, the information presented here was gathered from the literature. It was noted that most of the red algae possess few and short introns; for example, for *C. crispus*, the genome size is about 105Mbp, but the number of introns is very low. This is very rare when considering other eukaryotes of the same size. Recent genomic studies in red algae are expected to provide great insights into the evolution of eukaryotes since they are recognised as primitive eukaryotic organisms.

2.6.2 Protease sequence search

Total protease sequences were first identified from the *Chondrus crispus* and *Cyanodioschyzon merolae* sequences. All the identified protease sequences were analysed and grouped, on the basis of conserved domains, into five protease classes: aspartic (Asp), cysteine (Cys), metalloprotease (Met), serine (Ser), and threonine (Threo) proteases. The other two classes, asparagine (Aspar) and glutamic (Glu) proteases, did not seem to be present. The glutamic proteases have only been reported from selected filamentous fungi, such as *Aspergillus niger* and *Scytalidium lignocolum* (O'Donoghue *et al.,* 2008; Srilakshmi *et al.,* 2015). Asparagine proteases have mostly been found to be distributed among viruses (N1, N2, N8, N7 and N5) as viral coat proteins, in bacteria as autotransporter (N6 and N4), and in fungi as intein-containing proteins (N9, N10, and N11) (Rawlings *et al.*, 2011).

2.6.2.1 Aspartic protease

2.6.2.1.1 Sequence analysis

2.6.2.1.1.1 Protease sequence search

Seven red algae sequences were obtained from the Genbank database when using the *Chondus crispus* sequence as the query sequence in the BLASTp search engine (Table 2.3). Sequence similarities ranging from 56 – 100% were found in these protein sequences to the query sequence.

Table 2.3: Similar aspartic proteases from red algae. The sequences were aligned with BLASTp programme at NCBI.

Organisms	Max Score	Query Cover	E-Value	Identity	Accession
Choc	1361	100%	0.0	100%	XP005710223
Choc	385	70%	7e-122	42.20%	XP005716566
Cyme	291	56%	1e-86	40.78%	XP005537329
Gasu	373	73%	6e-118	40.11.5	PX005704310
Gasu	305	58%	2e-90	38.77%	XP005708387
Gracho	736	98%	0.0	59.21%	PXF48559
Gracho	411	69%	2e-132	45.24%	PXF47168
Gracho	347	65%	3e-108	41.08%	PXF42676
Pumb	462	65%	2e-150	52.18%	OSX77654
Pumb	359	58%	1e-111	46.25%	OSX69083
Purp	325	85%	4e-98	36.36%	KAA8496942
Pyezo*	463	87%	1e-129	43%	contig_13873_g3338

**Pyropia yezoensis* sequence was obtained, using a BLASTp programme, from Alga-PrAS database. Choc- *Chondrus crispus*; Cyme - *Cyanidioschyzon merolae strain 10D*; Gasu - *Galdieria sulphuraria*; Gracho - *Gracilariopsis chorda*; Pumb - *Porphyra umbilicalis*; Purp - *Porphyridium purpureum*; Pyezo - *Pyropia yezoensis*.

University of Fort Hare
Together in Excellence

2.6.2.1.1.2 Family and domain search

Several domains were identified from online domain databases (Table 2.4), in the identified red algae protein sequences. The domains identified included the Pepsin-like domain, Aspartic peptidase A1 family and Aspartyl (acid) proteases, which confirmed that the red algae sequences are aspartic proteases.

Table 2.4: Families/domains found within the red algae sequences.

	Choc	Cyme	Gasu	Pumb	Pyezo	Purp
NCBI CDD: cd05471	106-465	103-467	24-524	190-550	154-514	153-518
INTERPRO IPR001461	91-538	79-511	24-524	161-595	126-577	109-573
PFAM: PF00026	105-465	104-467	123-484	189-550	153-514	236-519
SUPERFAM: SSF50630	195-472	98-469	122-177 211-485	186-219 279-557	152-183 243-519	328-573
PANTHER: PTHR13683	37-466	47-467	124-492	146-550	___	122-518
GENE3D: G3DSA:2.40.70.10	278-538	276-511	294-524	161-595	126-514	328-573

University of Fort Hare
Together in Excellence

2.6.2.1.1.3 Sequence properties

The identified red algae aspartic proteases showed variations in sequence lengths ranging from 532 to 726 amino acids (Table 2.5). However, all contained a signal peptide and were located in either the plastid or chloroplast.

Table 2.5:Predicted sequence properties for aspartic protease from red algae

	Choc	Cyme	Gasu	Pumb	Pyezo	Purp
Length (aa)	661	564	532	726	694	698
Signal Peptide	26	20	64	62	25	41
Subcellular Localisation	Plastid	Plastid	Plastid	Chloroplast	Chloroplast	Plastid
Biological activity	GO: 0006508	GO: 0006508	GO: 0006508	GO: 0006508	GO: 0006508	GO: 0006508
Molecular activity	GO: 0004190	GO: 0004190	GO: 0004190	GO: 0004190	GO: 0004190	GO: 0004190
Solubility	insoluble	insoluble	Insoluble	Insoluble	Insoluble	Insoluble

2.6.2.1.1.4 Sequence alignment

The protein sequence alignment of the aspartic-like proteases (Figure 2.5) identified in the red algae genomes, showed a high degree of similarity, especially around the two catalytic aspartic residues. All the red algae were grouped in one clad on the phylogenetic tree (Figure 2.6), and recently diverged from the aspartic proteases from plants and bacteria.

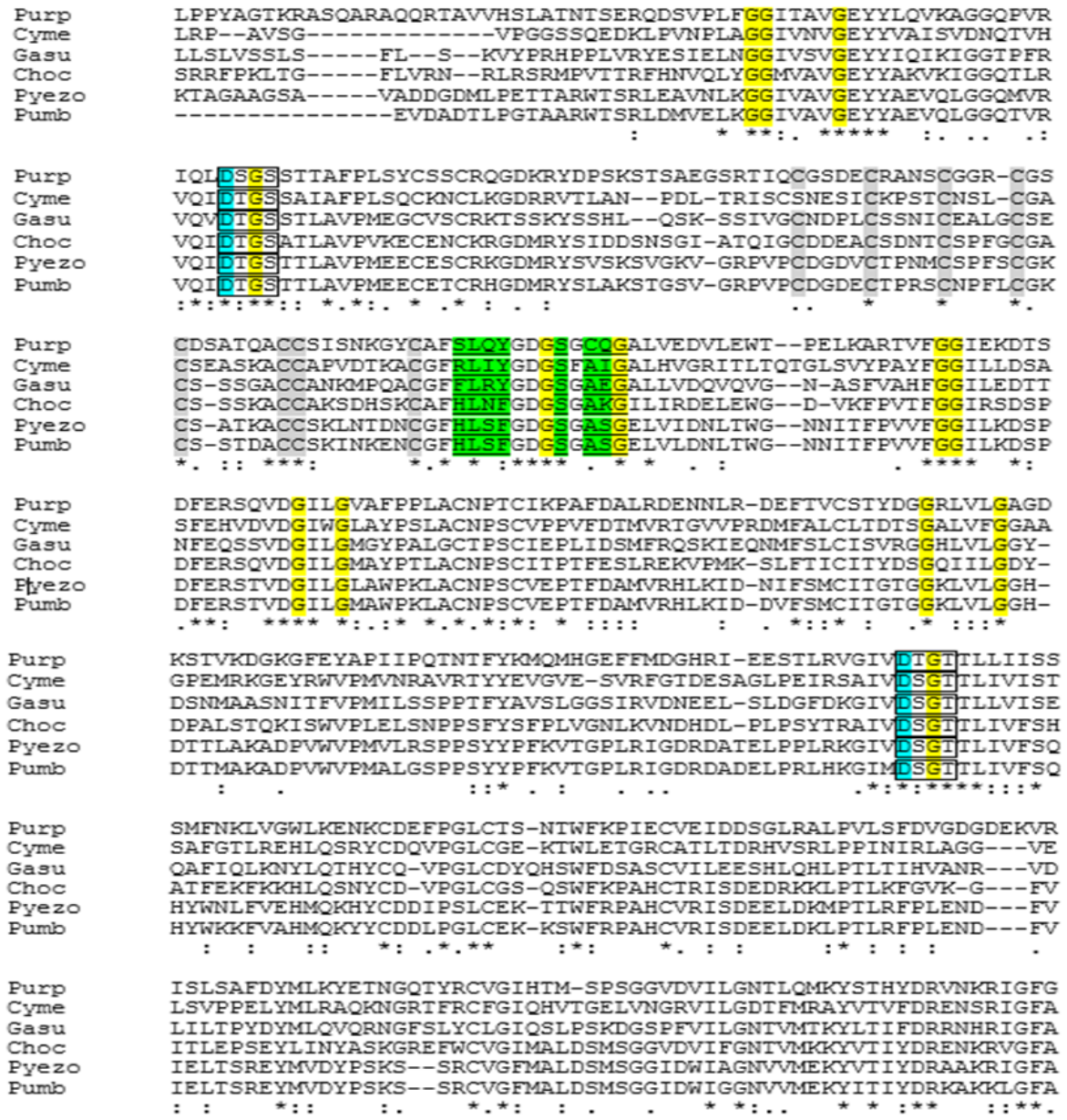
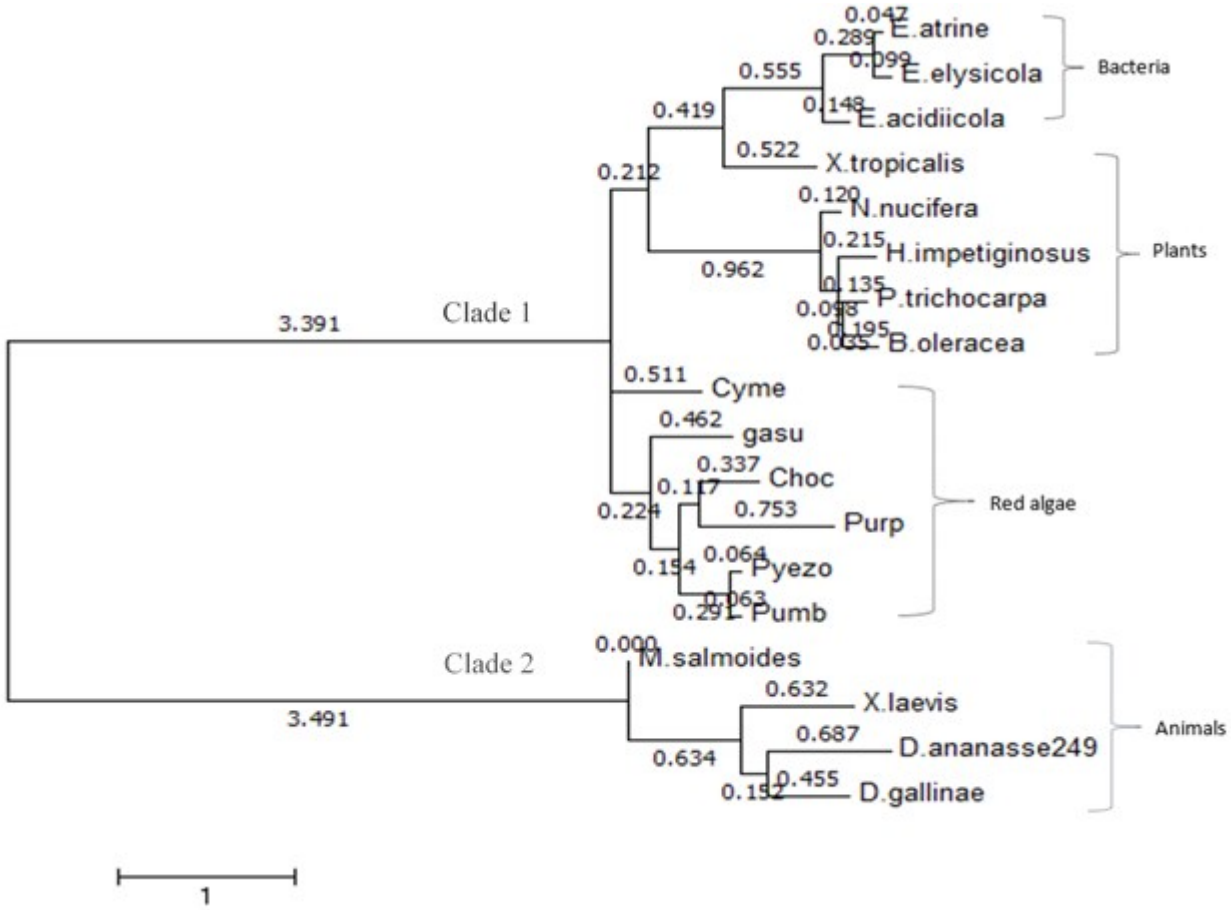


Figure 2.5: Alignment of aspartic proteases from red algae.

The overhanging gaps from the N- and C-termini were truncated manually. Choc- *Chondrus crispus*; Cyme - *Cyanidioschyzon merolae strain 10D*; Gasu - *Galdieria sulphuraria*; Gracho - *Gracilariopsis chorda*; Pumb - *Porphyra umbilicalis*; Purp - *Porphyridium purpureum*; Pyezo - *Pyropia yezoensis*.

2.6.2.1.1.5 Phylogenetic tree



University of Fort Hare
Together in Excellence

Figure 2.6: Phylogenic analysis of aspartic protease sequences from red algae.
Choc- *Chondrus crispus*; Cyme - *Cyanidioschyzon merolae strain 10D*; Gasu - *Galdieria sulphuraria*; Gracho - *Gracilariopsis chorda*; Pumb - *Porphyra umbilicalis*; Purp - *Porphyridium purpureum*; Pyezo - *Pyropia yezoensis*.

2.6.2.1.2 Structure prediction

2.6.2.1.2.1 Secondary structure prediction

Analysis of the secondary structure of the aspartic-like proteases from the various red algae, showed a high degree of variation, with some sequences having a high percentage of β-sheets, while others showed a high percentage of α-helices (Figure 2.7 and Table 2.6). However, the predicted tertiary structures (Figure 2.8) were very similar, which is also indicated by the high convervation of the structure of the active site pocket (Figure 2.9).

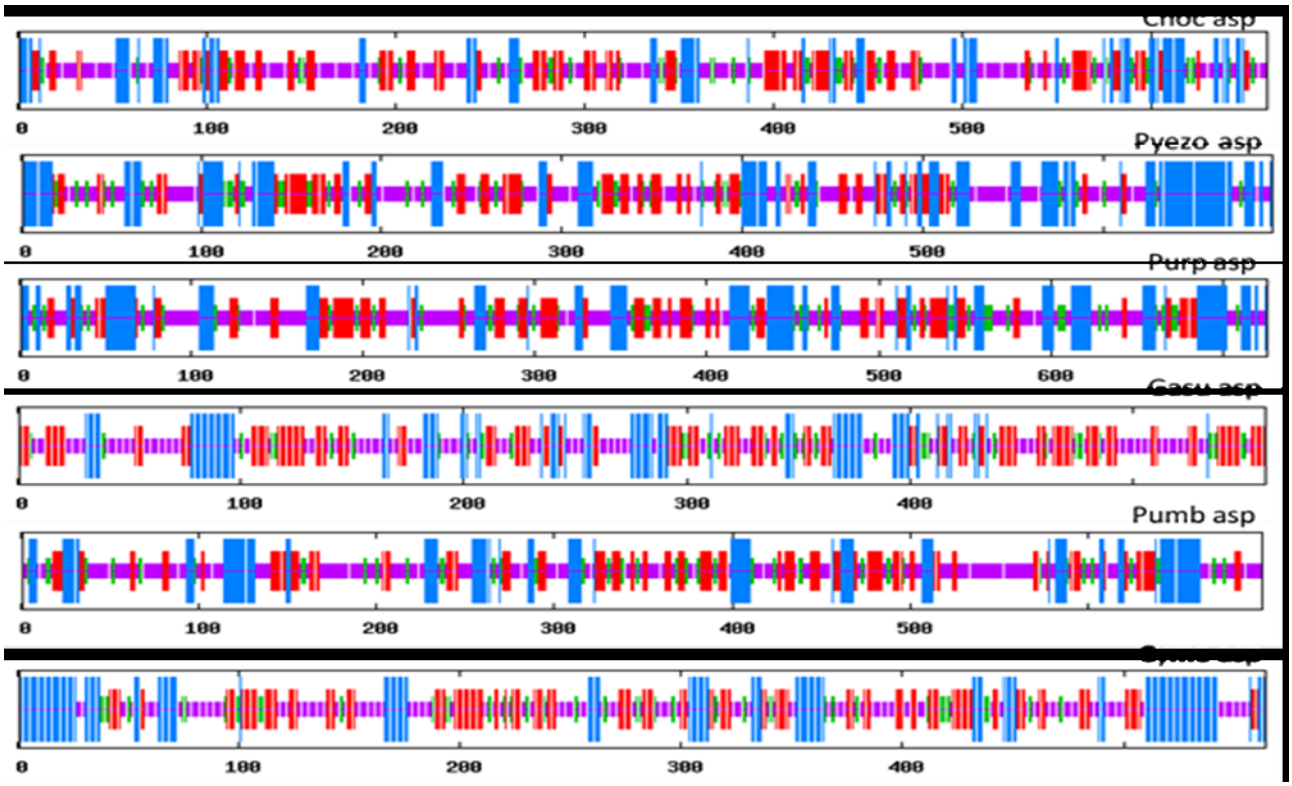


Figure 2.7: Secondary structures of aspartic protease sequences from red algae.

Table 2.6: Summary of secondary structures found in aspartic acid sequences.

Secondary Structure Element	***Choc***	***Purp***	***Cyme***	***Pumb***	***Pyezo***	***Gasu***
Alpha helix	17.85%	19.91%	25.18%	25.07%	29.39%	18.25%
Extended strand	25.57%	23.07%	26.6%	22.73%	20.61%	32.38%
Coil	47.5%	46.85%	37.77%	41.32%	39.63%	39.53%
Beta turn	9.08%	10.17%	10.46%	10.88%	10.37%	9.84%

2.6.2.1.2.2 Tertiary structure prediction

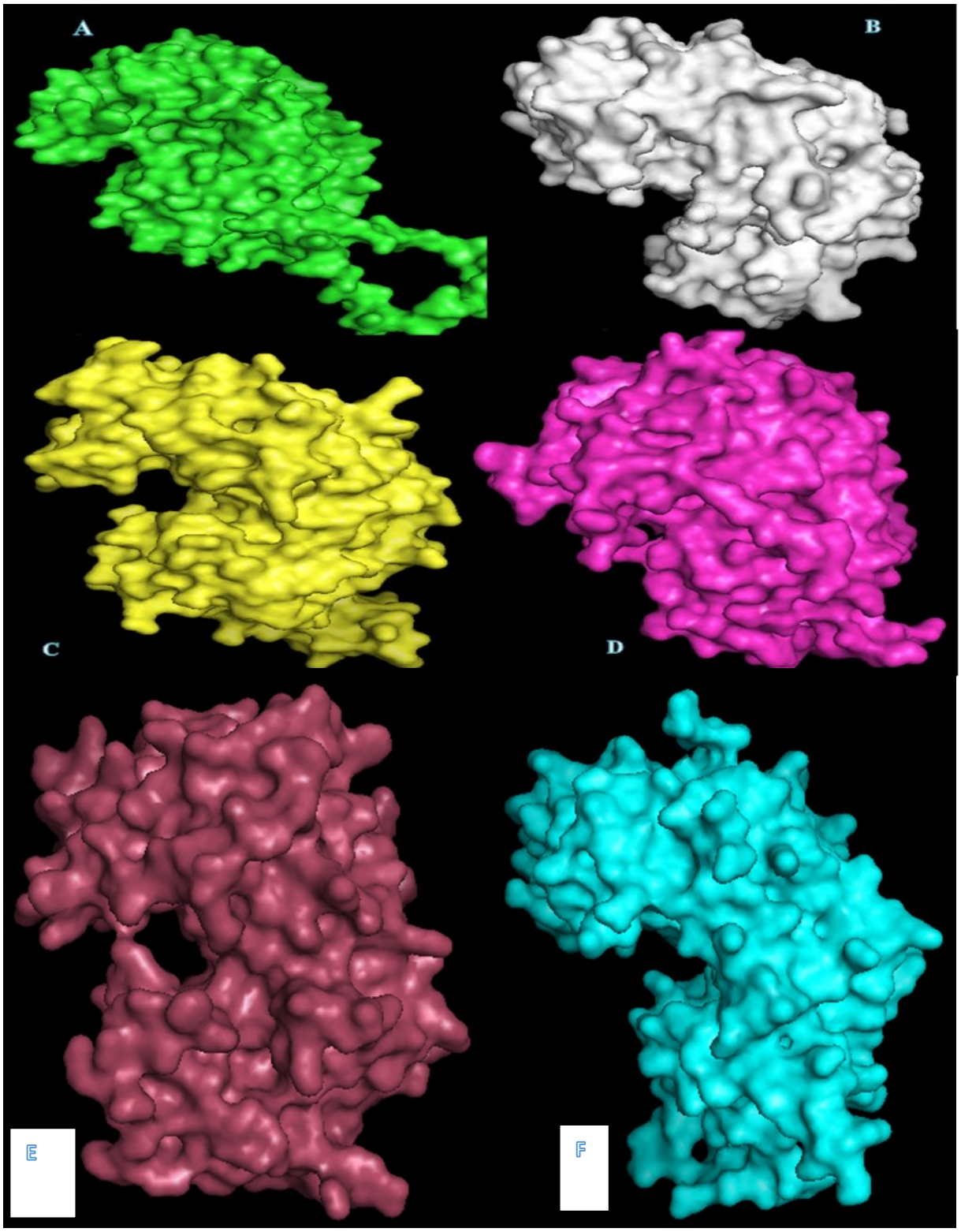


Figure 2.8: The predicted global structure of red algae aspartic proteases viewed in PyMol as surface.

Global structure: A-Choc; B-Cyme; C-Gasu; D-Purp; E-Purp and F-Pyezo. Choc- *Chondrus crispus*; Cyme - *Cyanidioschyzon merolae strain 10D*; Gasu - *Galdieria sulphuraria*; Gracho - *Gracilariopsis chorda*; Pumb - *Porphyra umbilicalis*; Purp - *Porphyridium purpureum*; Pyezo - *Pyropia yezoensis*.

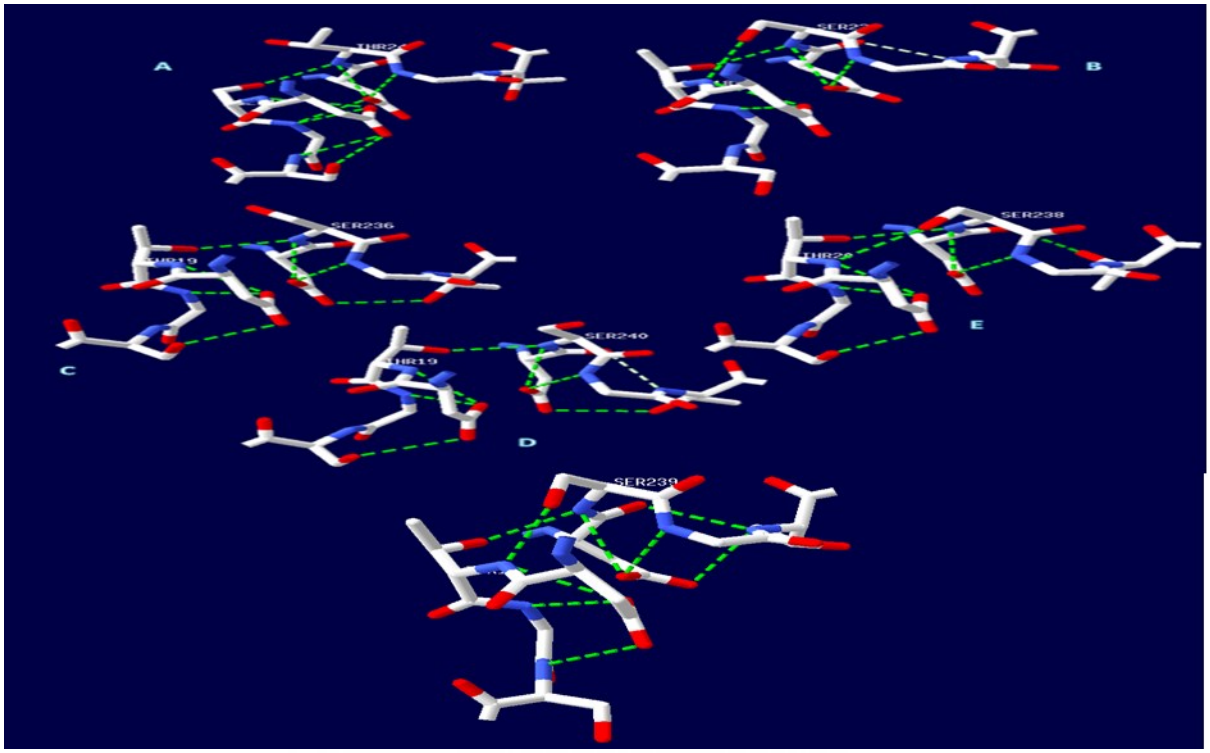


Figure 2.9: Catalytic motifs from the aspartic proteases identified in red algae.

Catalytic motif from A-Pyezo; B-Purp; C- Gasu; D-Cyme; E-Choc; F-Pumb. Choc- *Chondrus crispus*; Cyme - *Cyanidioschyzon merolae strain 10D*; Gasu - *Galdieria sulphuraria*; Gracho - *Gracilariopsis chorda*; Pumb - *Porphyra umbilicalis*; Purp - *Porphyridium purpureum*; Pyezo - *Pyropia yezoensis*.

University of Fort Hare
Together in Excellence

2.6.2.2 Metalloprotease

2.6.2.2.1 Sequence analysis

Aminopeptidases are known for hydrolysing the N-terminus residues in peptides and proteins, and are categorised on the basis of substrate specificity towards the basic, acidic, or neutral amino acids (Sherwin *et al.,*). In this research, we identified an aspartyl aminopeptidase from the red algae. This enzyme is a member of M18 metalloproteases (Rawlings *et al.*, 2012) and has been previously reported in various organisms, including yeast, fungi, and certain species of mammals (Cai *et al.,* 2010; Yokoyami *et al.*, 2006).

2.6.2.2.1.1 Protease sequence search

The *Chondrus crispus* amino acid sequence was used as a query to search for similar sequences. The *C. crispus* sequence was found to be homologous to baking yeast (AJ28087), mouse (NP_001104301), and human (QAULA01), sharing 43%, 45%, and 44% identity, respectively. Similar sequences from other red algae were also found to be above 40%, except in the case of *Pyropia yezoensis* (Pyezo), which was approximately half the other sequences (118aa) (Table 2.7).

Table 2.7: Metalloprotease identified across red algae species.

	Max Score	**Query Cover**	**E-Value**	**Identity**	**Accession**
CHOC	1003	100%	0.0	100%	XP_005710409
CYME	391	97%	2e-131	43%	XP_005535853
GASU	_____	_____	_____	_____	_____
PUMB	407	95%	2e-137	50%	OSX75570
PURP	468	85%	5e-161	50%	Contig_2030.16
PYEZO	96.3	81%	6e-19	63%	Contig_40872_G9353

University of Fort Hare
Together in Excellence

Note: *Chondrus crispus* sequence was used as a query during BLASTp search. Choc- *Chondrus crispus*; Cyme - *Cyanidioschyzon merolae strain 10D*; Gasu - *Galdieria sulphuraria*; Gracho - *Gracilariopsis chorda*; Pumb - *Porphyra umbilicalis*; Purp - *Porphyridium purpureum*; Pyezo - *Pyropia yezoensis*.

2.6.2.2.1.2 Domain/family search

Several domains were identified from online domain databases (Table 2.8), in the identified red algae protein sequences. The domains identified included the Peptidase M18 family, Aminopeptidase I zinc metalloprotease (M18) family, aspartyl aminopeptidases, metal binding site [ion binding site], which confirmed that the red algae sequences are metallopeptidases.

Table 2.8: Domain/family identifiers from various databases found within the red algae metalloprotease sequences.

Functional domain database	*Choc*	*Cyme*	*Pumb*	*Pyezo*	*Purp*
NCBI CDD: cd05658	13-461	30-509	46-512	2-78	77-531
INTERPRO: IPR001948	98-249	118-245	133-228 259-297	___	164-306
PFAM: PF02127	21-461	36-507	52-530	15-72	83-530
SPERFAM: SSF101821	98-249	118-245	45-132 297-531	9-78	164-306
PANTHER: PTHR28570	9-473	26-517	45-541	12-79	66-538
GENE3D G3DSA:3.40.630.10	8-148 176-468	22-127 268-511	310-530	4-88	72-208 211-536

University of Fort Hare
Together in Excellence

2.6.2.2.1.3 Sequence properties

The nucleotide sequence of Pyezo was not included in the further analysis as it was assumed that the gene coding for this protease had been disrupted or deleted. The analysis of the other predicted red algae aspartyl aminopeptidases, were found to be cytosolic and they did not appear to have any signal peptide. Cai *et al.* (2010) reported aspartyl aminopeptidase, expressed by human and mouse islet cells, to be cytosolic. The predicted pI was found to be close to neutral (Table 2.9). A neutral optimum pH was reported in yeast (Yokoyami *et al.*, 2006).

Table 2.9: Predicted sequence properties of red algae metalloprotease.

	Choc	*Cyme*	*Pumb*	*Pyezo*	*Purp*
Length (aa)	485	524	549	118	546
pI	6.0	6.2	7.7	4.2	6.7
Gravy	-0.12	-0.16	0.19	0.39	-0.20
Signal Peptide	None	None	None	None	None
Subcellular Localisation	Cyto	Cyto	Cyto	Cyto	Cyto
Biological Activity	GO:0006508	GO:0006508	GO:0006508	GO:0006508	GO:0006508
Molecular Activity	GO:0004177 GO:0008270	GO:0004177 GO:0008270	GO:0004177 GO:0008270	GO:0004177 GO:0008270	GO:0004177 GO:0008270
Solubility	Insoluble	Insoluble	Insoluble	Insoluble	insoluble

University of Fort Hare
Together in Excellence

Choc- *Chondrus crispus*; Cyme - *Cyanidioschyzon merolae strain 10D*; Gasu - *Galdieria sulphuraria*; Gracho - *Gracilariopsis chorda*; Pumb - *Porphyra umbilicalis*; Purp - *Porphyridium purpureum*; Pyezo - *Pyropia yezoensis*.

2.6.2.2.1.4 Sequence alignment

The alignment of red algae aspartyl aminopeptidase (Figure 2.10) showed that these sequences lack the sequence motif (HExxH + E) found in another aminopeptidase, such as glutamic aminopeptidase. However, a signature (H.D.E.D.H) that has been reported as strictly conserved among aspartyl aminopeptidases was observed. Chaikuad *et al.* (2012) reported human aspartyl aminopeptidase as a binuclear metalloprotease with Zn 1 interacting with glu302 and H440, whereas Zn 2 interacts H94 and asp346. The asp 264 acts as a connection between these pairs (hys94-**D264**-glu302-asp346-hys440).

Figure 2.10: Aligned metalloprotease sequences from the red algae.

The catalytic motif is represented by the blue-coloured residues. The residues with outside borders constitute the metal binding site.

2.6.2.2.1.5 Phylogeny analysis

Phylogenetic analysis of red algae sequences resulted in an unrooted tree with two main clades (Figure 2.11). The DNPEP in red algae were shown to be closely related as they appeared in the same cluster. However, Cyanodioschyzon merolae (Cyme) appears outside the cluster, suggesting that it is closely related to the ancestor.

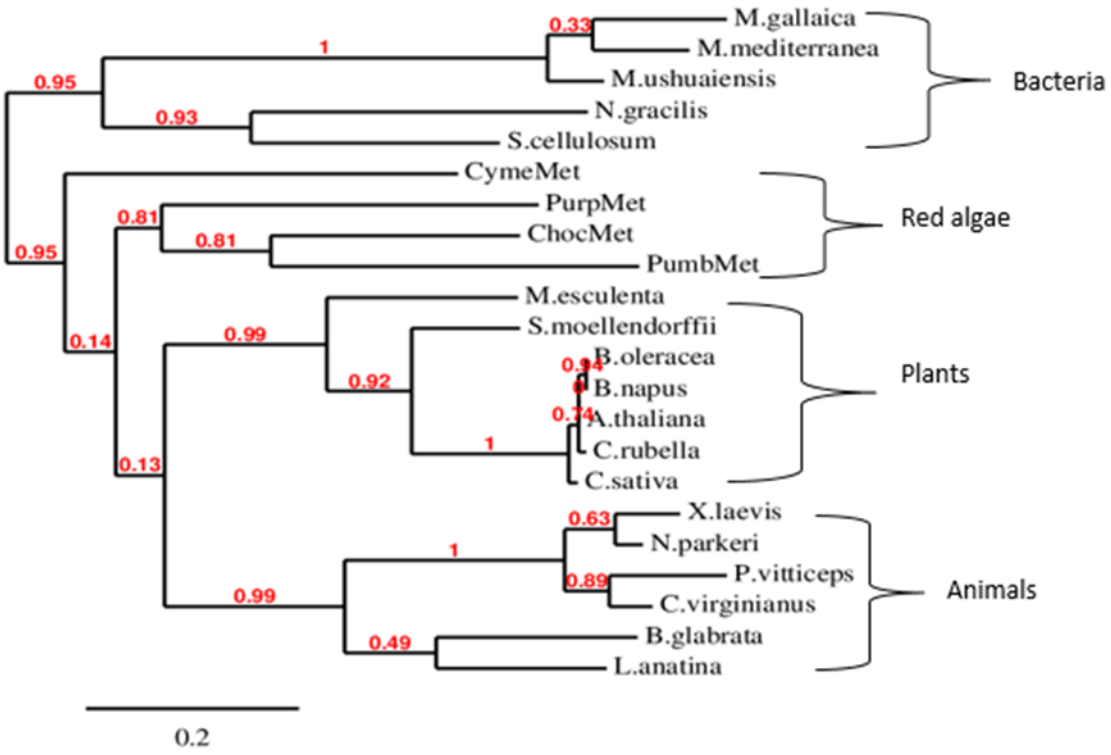


Figure 2.11: Phylogenetic analysis of metalloproteases from the red algae.

The phylogenetic tree was constructed using an online server, phylogeny.fr. B.glabrata (XP_013061743), X.laevis (OCT63383), N.parkeri (XP_018410420), P.vitticeps (XP_020647787), L.anatina (XP_013413903), C.virginianus (OXB77088), N.gracilis (WP_005009050), M.gallaica (WP_067036087), M.mediterranea (WP_013662134), M.ushuaiensis (WP_036160278), S.cellulosum (KYF49526), C.rubella (XP_006287486), C.sativa (XP_010452373), B.oleracea (XP_013606163), B.napus (XP_022566017), M.esculenta (XP_021615595), S.moellendorffii (XP_002975004), A.thaliana (NP_196091), ChocMet (XP_005710409), Pumb (OSX75570), Cyme (XP_005535853), Purp (Contig_2030.16), Pumb (Contig_40872_G9353).

2.6.2.2.2 Structure prediction

2.6.2.2.2.1 Secondary structure prediction

The secondary structure analysis (Figure 2.12 and Table 2.10) revealed that the DNPEP from Cyme and Purp are dominated by coils, whereas Choc and Pumb are dominated by alpha helices. Because of these secondary structure elements, Choc and Pumb tertiary structures appeared to be more stable on the Ramachandran plot with over 90 residues found in the most favoured region. Cyme and Purp had 86 and 85% of residues in most favoured regions with 0.9 and 0.7% in the unflavoured region, respectively.

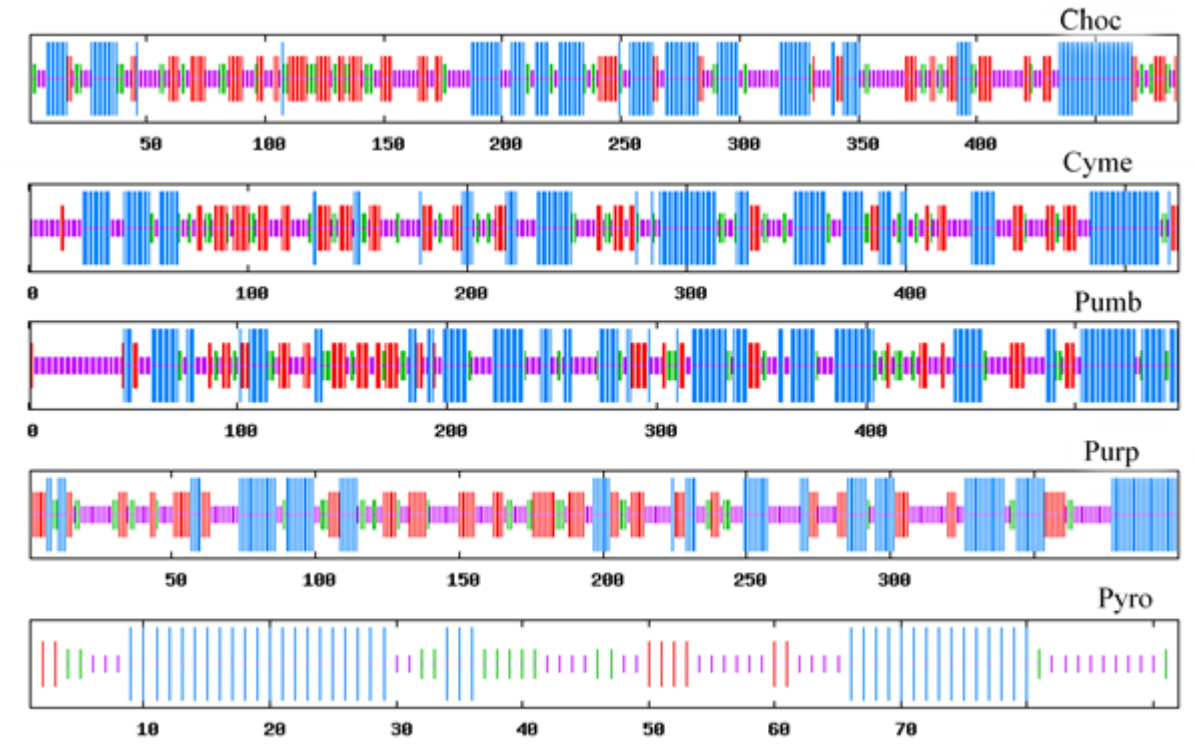


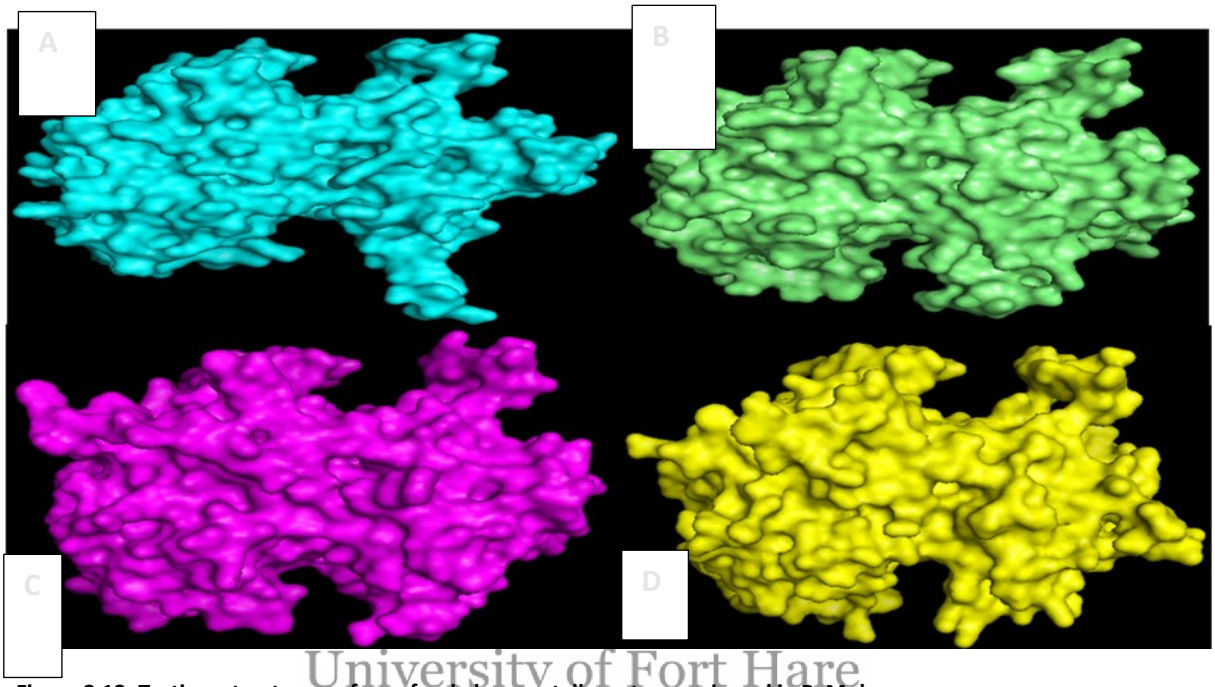
Figure 2.12: The sequence-based secondary structure prediction of metalloprotease from red algae.
The blue colour = alpha helix, red = extended strands, green = beta green, purple = random coils, and the numbers are the sequence length.

Table 2.10: The estimated percent of secondary structure elements of metalloproteases from the red algae metalloproteases.

SS element	CYME	PURP	CHOC	PUMB	PYEZO
Alpha helix	33.97%	30.75%	31.34%	38.07%	42.39%
Extended strand	19.66%	24%	22.27%	16.39%	10.87%
Coil	35%	35.35%	34.02%	35.34%	32.61%
Beta turn	10.38%	9.75%	12.37%	10.2%	14.13%

2.6.2.2.2.2 Tertiary structure prediction

The tertiary structure was conserved in all of the species and consisted of two domains, proteolytic and dimerization domains, forming a butterfly-like shape (Figure 2.13). The active site was situated within the concave groove formed by the domains (Figure 2.14). Chaikuad *et al.* (2012) reported a similar structure as a subunit of the dodecameric aspartyl aminopeptidase from humans (Chaikuad *et al.,* 2012). These structures adopt an alpha/beta/alpha topology with dominating beta strands sandwiched between several alpha helices.



University of Fort Hare
Together in Excellence

Figure 2.13: Tertiary structure surface of red algae metalloproteases viewed in PyMol.
A-Cyme (cyan); B-Choc (lime); C- Pumb (magenta); D- Purp (yellow).

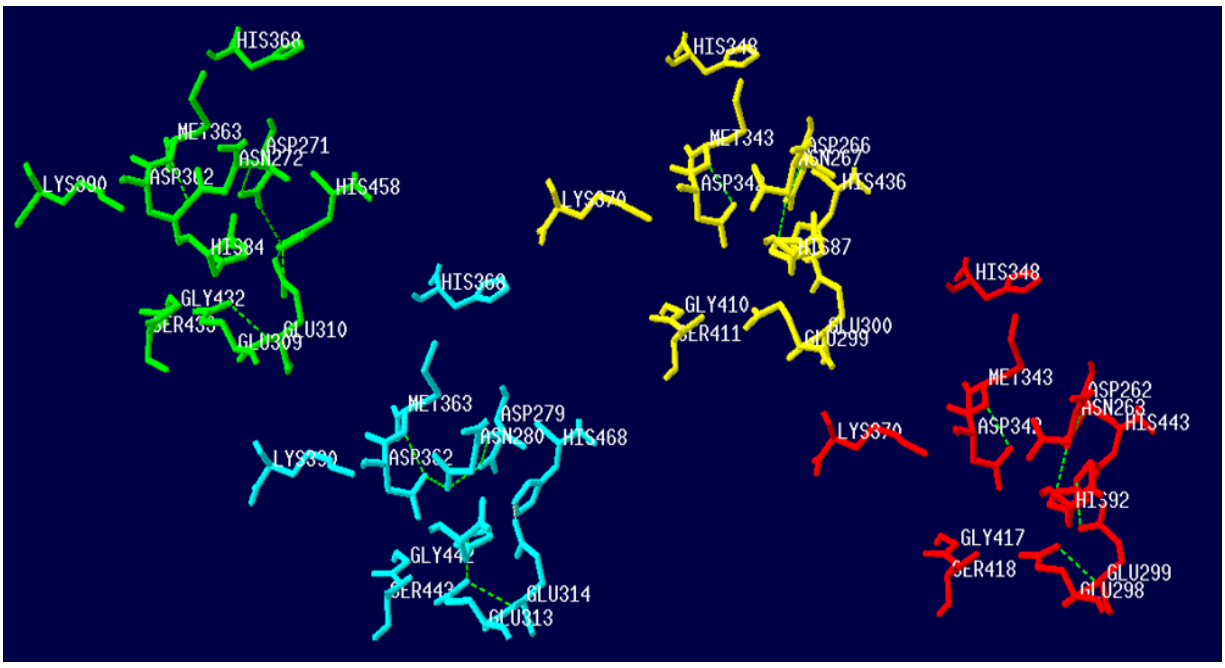


Figure 2.14: The catalytic site of the red algae aspartyl metalloproteases.
Cyme is coloured green; Choc is coloured yellow; Pumb is coloured cyan; Purp is coloured red.

2.6.2.3 Threonine protease

An asparaginase protease family was selected as a representative from the N-terminal nucleophile (NTN) hydrolase clan of the threonine protease class. An asparaginase is an enzyme that catalyses the conversion of asparagine to aspartate.

The study of asparaginase enzyme is justified by its involvement in cancer therapy. Its major sources are *Erwinia crysanthemi (now known as Dickeya dadantii)* (Crysantaspase, Erwinase) and *Escherichia coli* (Colaspase, Elspar). The latter has been the predominant source. However, it has been noted that the enzymes of bacterial origin possess many side effects and attempts have been made to develop drugs with human-based enzymes. Nomme *et al.* (2012) reported a human asparaginase structure as an alternative to develop a drug with human-based enzymes. The major hurdle was the inefficient catalysis. Here, we have identified the asparaginase enzymes from the red algae.

2.6.2.3.1 Sequence analysis

2.6.2.3.1.1 Protease sequence search

Chondrus crispus sequence was used as a query to search similar sequences from the other red algae. The results are summarised in Table 2.11. The sequences shown in the table are those that appeared to have the highest identity, highest maximum number, and highest query coverage. All the asparaginase sequences presented here were above 40% identical to *C. crispus*.

Table 2.11: Similar threonine proteases identified from different red algae genomes.

	MAX SCORE	QUERY COVER	E-VALUE	IDENTITY	ACCESSION
CHOC	627	100%	0.0	100%	XP_005713659
CYME	248	98%	5e-79	45%	XP_005536214
GASU	248	95%	1e-80	48%	XP_005704015
PUMB	176	59%	8e-52	46%	OSX75919
PURP	258	80%	9e-84	48%	Contig_779.3
PYEZO	106	88%	2e-13	44%	Contig_6972_g1614

Choc- *Chondrus crispus*; Cyme - *Cyanidioschyzon merolae strain 10D*; Gasu - *Galdieria sulphuraria*; Gracho - *Gracilariopsis chorda*; Pumb - *Porphyra umbilicalis*; Purp - *Porphyridium purpureum*; Pyezo - *Pyropia yezoensis*.

University of Fort Hare
Together in Excellence

2.6.2.3.1.2 Domain/family search

Sequence analysis revealed that those sequences consisted of one functional conserved domain (asparaginase: *PF01112*) (Table 2.12). However, *Galdieria sulphuraria* (Gasu) appeared to contain another domain, the CN-hydrolase domain. This domain is found in many enzymes, but chiefly in nitrilases (Piotrowski *et al.*, 2003).

Table 2.12: Domain/family identifiers from various databases found within red algae metalloprotease sequences.

	Choc	*Cyme*	*Gasu*	*Pumb*	*Pyezo*	*Purp*
NCBI CDD: *Cd04702*	9-298	84-433	24-323	84-342	cd04701	2-306
INTERPRO *IPR000246*	9-307	85-241 307-435	24-322	75-354	4-356	4-315
PFAM: *PF01112*	9-300	86-433	24-322	75-354	4-356	2-308
SUPERFAM: *PSF56235*	0-301	85-241 307-435	24-322	75-202 251-362	2-166 224-356	50-312
PANTHER: *PTHR10188*	9-307	87-443	25-321	75-354	3-358	4-315
GENE3D: *G3DSA:3.60.20.30*	_	310-441	202-328	253-376	227-360	192-318

2.6.2.3.1.3 Sequence properties

The predicted red algae threonine protease sequences were found to not have any signal peptides and their pIs were predicted to be close to pH 5, with the exception of Pumb and Cyme which had pI readings of 11.5 and 8, respectively (Table 2.13). The subcellular location differed with organisms: Choc, Pyezo and Purp were predicted to be cytoplasmic, whereas Pumb and Gasu were located in the chloroplast.

Members of NTN hydrolases are produced as an inactive precursor which then undergoes a post-translational modification, cleaving the peptide into two subunits (alpha and beta), exposing the N-terminal threonine as a nucleophile (Nomme *et al.*, 2012). This cleavage has been reported to be different from that of pro-enzymes (e.g. pepsinogen and trypsinogen). Although cleavage takes place in the middle of the sequence, the subunits remain together, connected by a loop, constituting one functional structure.

Table 2.13: Predicted sequence properties of red algae asparaginase (threonine protease).

	Choc	**Cyme**	**Gasu**	**Pumb**	**Pyezo**	**Purp**
Length (aa)	312	459	454	416	360	381
PI	4.8	8	5.5	11.5	5.2	4.8
GRAVY	0.14	-0.27	-0.04	-0.135	0.23	-0.04
Signal Peptide	None	None	None	None	None	None
Subcellular Localisation	Cyto	Nucl	Chlo	Chlo	Cyto	Cyto
Asparaginase activity	GO:0004067	__	GO:0004067	GO:0004067	__	GO:0004067
Hydrolase activity		GO:0016787	GO:00016810	GO:00016810	GO:0016787	GO:0016787
Solubility	insoluble	insoluble	insoluble	insoluble	soluble (0.505633)	insoluble

University of Fort Hare
Together in Excellence

2.6.2.3.1.4 Sequence alignment

Sequence alignment revealed cleavage occurs between Asp175 and Thr176 (Figure 2.15, bold black coloured arrow), according to *C. crispus* numbering. This residue has been shown to be critical for auto-cleavage and catalysis (Nomme *et al.*, 2014). Other residues involved in catalysis and substrate binding are coloured in cyan (Figure 2.15).

The amino acids that antecede the threonine nucleophile (coloured yellow) have been observed to be highly diverse and non-conserved. The residues in this region were predicted to form a loop structure (Figure 2.18). Similar observation has been reported and the region was referred to as a "variable loop" or flexible loop (Kerckhoven *et al.*, 2017; Maggi *et al.*, 2015).

All the red algae threonine protease did not group in one clad on the phylogenetic tree (Figure 2.16), and but rather grouped with either animal, plant or bacteria threonine proteases.

University of Fort Hare
Together in Excellence

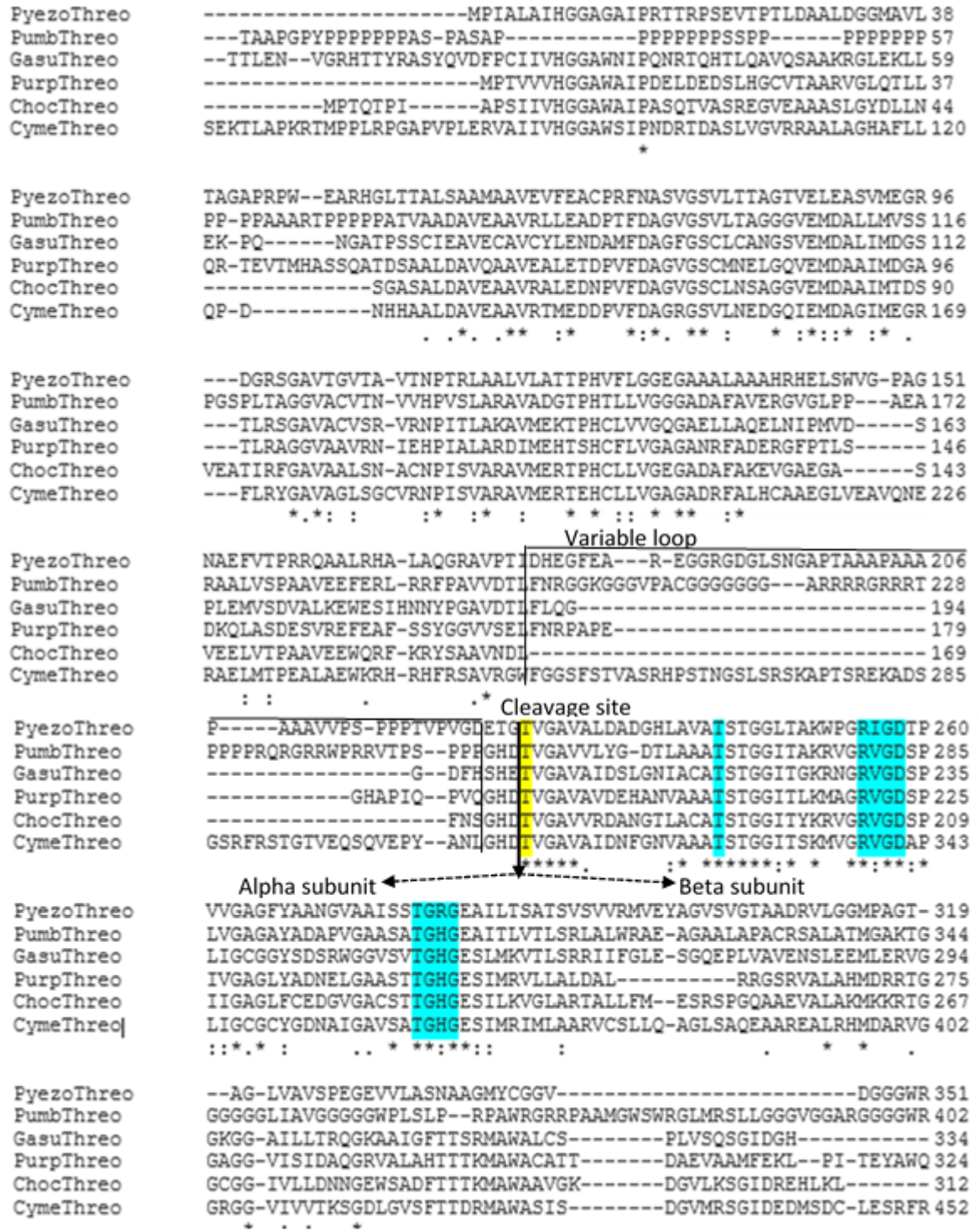


Figure 2.15: Alignment of red algae threonine sequences.

The yellow coloured residues are the threonine nucleophile and the turquoise coloured residues are the binding sites. Choc- *Chondrus crispus*; Cyme - *Cyanidioschyzon merolae strain 10D*; Gasu - *Galdieria sulphuraria*; Gracho - *Gracilariopsis chorda*; Pumb - *Porphyra umbilicalis*; Purp - *Porphyridium purpureum*; Pyezo - *Pyropia yezoensis*.

2.6.2.3.1.5 Phylogenetic analysis

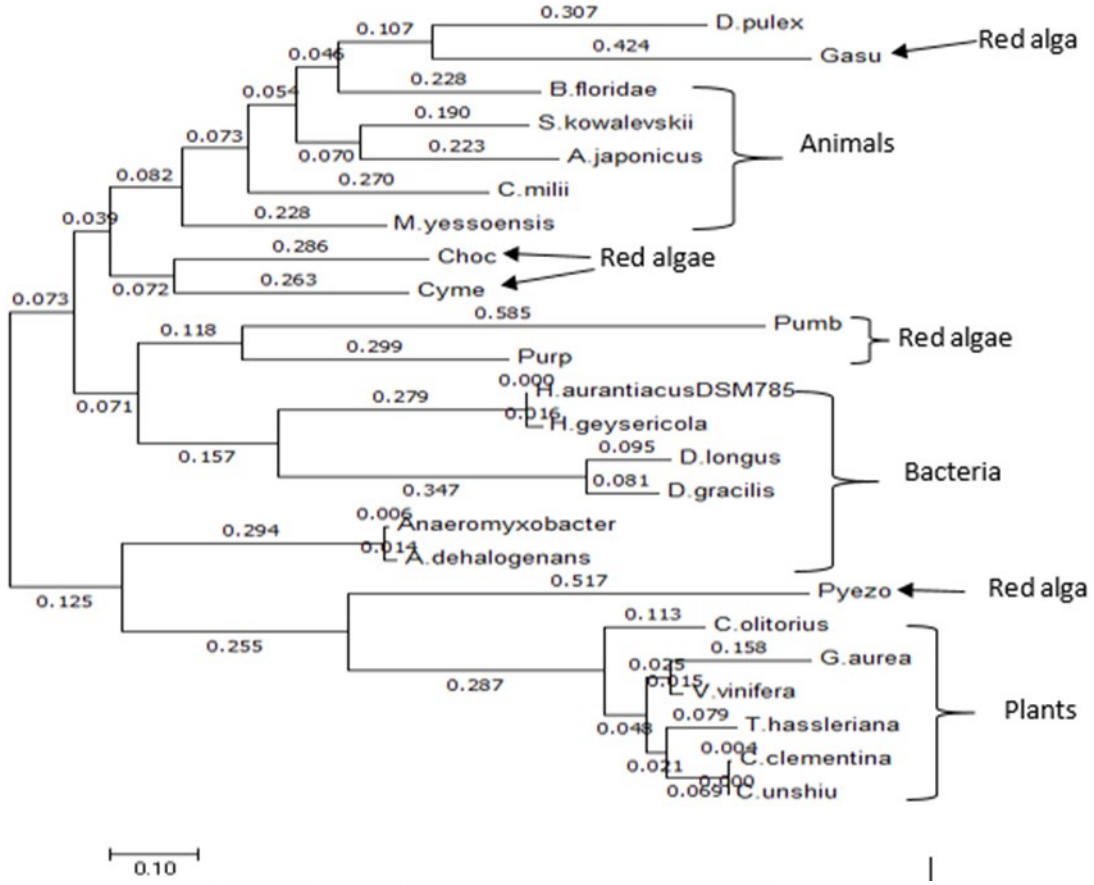


Figure 2.16: Phylogenetic analysis of threonine proteases by maximum likelihood, based on the JTT matrix-based model. The sequences were aligned using clustal W in MEGA 6 software and manually adjusted. Choc- *Chondrus crispus*; Cyme - *Cyanidioschyzon merolae strain 10D*; Gasu - *Galdieria sulphuraria*; Gracho - *Gracilariopsis chorda*; Pumb - *Porphyra umbilicalis*; Purp - *Porphyridium purpureum*; Pyezo - *Pyropia yezoensis*.

University of Fort Hare
Together in Excellence

2.6.2.3.2 Structure prediction

2.6.2.3.2.1 Secondary structure prediction

Analysis of the secondary structure of the threonine-like proteases from the various red algae, showed a high degree of variation, with some sequences having a high percentage of β-sheets, while others showed a high percentage of α-helices (Figure 2.17 and Table 2.14). However, the predicted tertiary structures (Figure 2.18) were very similar.

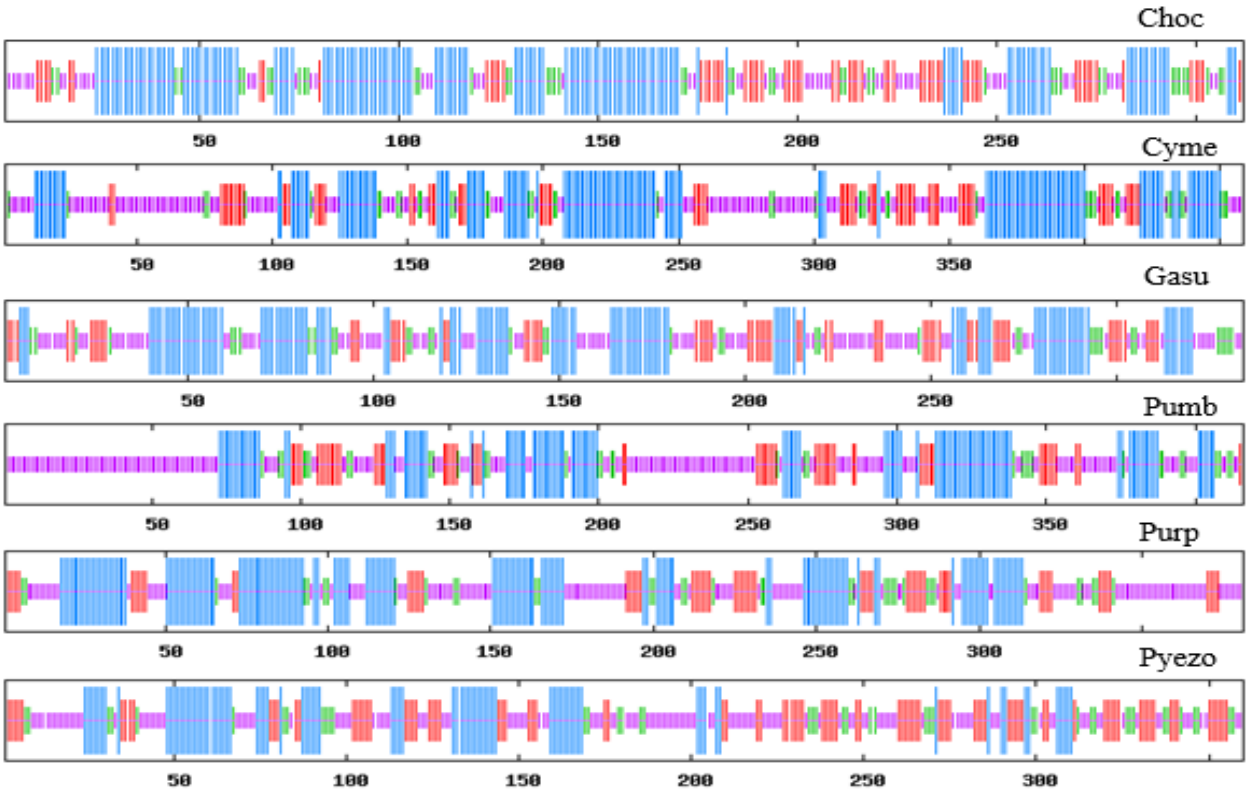


Figure 2.17: The sequence-based secondary structure prediction of threonine protease (asparaginase) from red algae. The blue colour = alpha helix, red = extended strands, green = beta green, purple = random coils, and the numbers are the sequence length.

Table 2.14: The estimated percent of secondary structure elements of metalloproteases from the red algae metalloproteases.

SS ELEMENT	CYME	GASU	PURP	CHOC	PUMB	PYEZO
ALPHA HELIX	35.51%	34.73%	35.96%	44.87%	27.16%	22.5%
EXTENDED STRAND	15.69%	20.06%	15.75%	19.87%	13.46%	26.39%
COIL	39.22%	34.13%	38.85%	21.15%	50.96%	39.17%
BETA TURN	9.59%	11.08%	9.45%	14.10%	8.41%	11.94%

2.6.2.3.2.2 Tertiary structure prediction

Homology modelling was used to predict the uncleaved asparaginase tertiary structures. The multi-template approach was followed to cover the gaps. The structures were predicted to have an alpha-beta fold with approximately 12 strands sandwiched between 6 and 2 helices. Procheck revealed that the structures were stable, ranging from 84% up to 92% of the residues in the most favoured region. The overall average of the G-factors ranged from -0.08 to -0.34, which is higher than -0.5 normalisation.

The N-terminal nucleophile was found to be in between two structures, a flexible loop and a stable hairpin-like structure (Figure 2.18). The flexible loop has an unclear function due to non-conservation of the amino acids that constitute the loop (Kerckhoven *et al.*, 2017). In contrast, Maggi *et al.* (2015) suggested that the loop is involved in the regulation of the binding site accessibility of the active enzyme (auto-cleaved).

We analysed the repeats and some unusual amino acids in the loop. Pumb showed an array of glycine and arginine residues followed by a short stretch of proline residues. Pyezo was rich in alanine residues, whereas Cyme showed a sporadic distribution of serine-arginine residues. Analysis of these amino acids suggested that they are involved in the acceleration of the peptide cleavage. Glycine residues have been shown to accelerate the peptide cleavage.

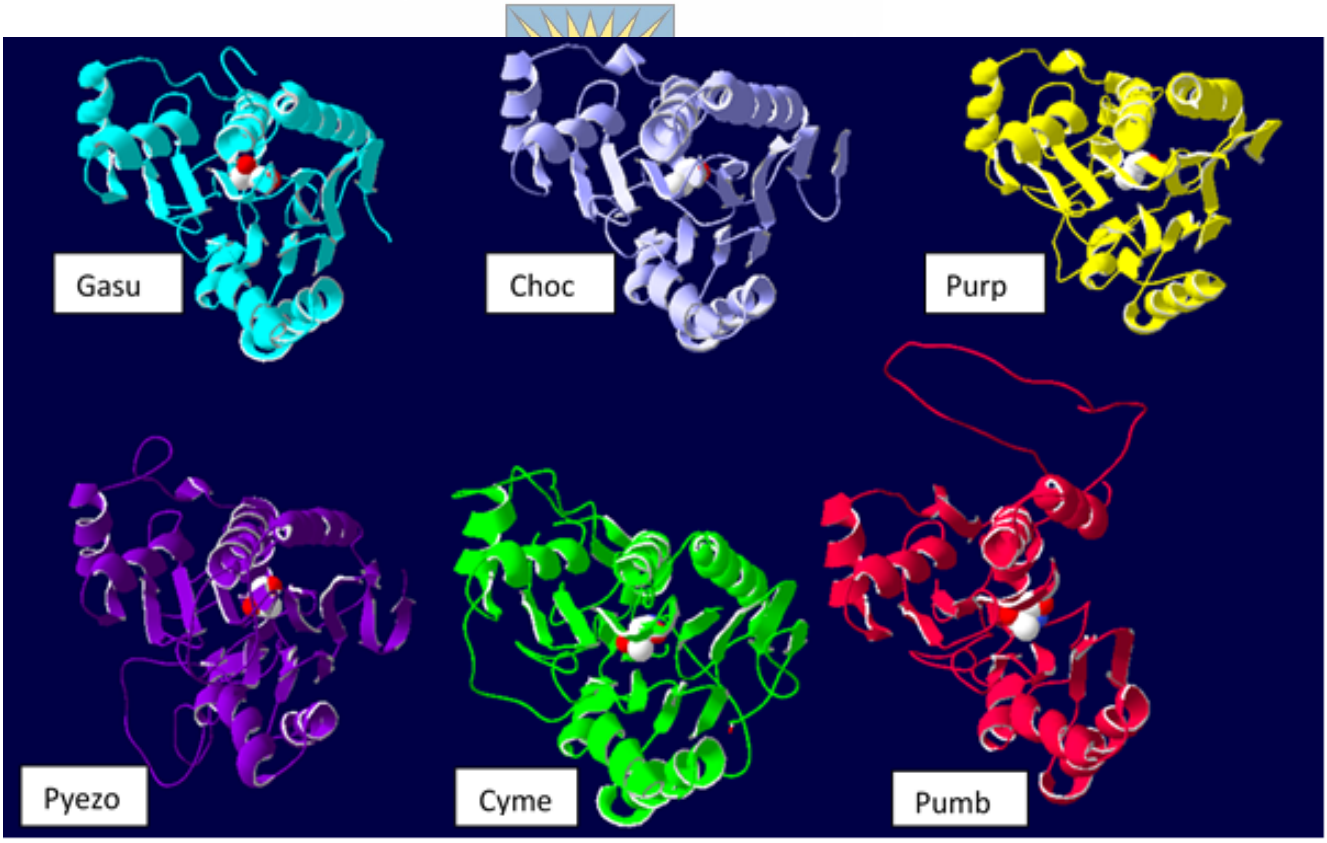


Figure 2.18: Uncleaved tertiary structures of red algae threonine protease (asparaginase) viewed as ribbon using spdbviewer.

Threonine nucleophile is shown as spheres coloured in cpk. Gasu (cyan)-G. suphuraria, Choc (light blue)- C. crispus, Purp (yellow)- P. purpureum, Pyezo (purple) – P. yezoensis, Cyme (green) – C. merolae, and Pumb (maroon red) P. umbilicalis

2.6.2.4 Serine protease

Trypsin is a serine endopeptidase which plays a crucial role in digestion of proteins. This enzyme is ubiquitous in all living organisms and is specific to peptide bonds at the arginine and lysine carboxyl termini.

Apart from its physiological role, it is utilized in basic research, therapy and industry (Khandagale *et al.*, 2015). Several studies have characterized trypsin from various fish species since the fish display diverse feeding habits and the enzyme is similar to that of mammals with regard to molecular (22-30kDa), amino acid composition, and susceptibility to similar inhibitors (Khandagale *et al.*, 2015; Klomklao *et al.*, 2006; Silva *et al.*, 2006; Marcuschi *et al.*, 2010; Temiz *et al.*, 2013; Ásgeirsson *et al.*, 1988).

However, since the seaweeds (particularly red seaweeds) have adapted to survive the varying environmental factors, studying the enzymes from such organisms represent the initial phase in assessing their potential for biotechnology application.

2.6.2.4.1 Sequence analysis

2.6.2.4.1.1 Protease sequence analysis

Serine protease encoding seqeunces were identified in all six red algae genomes (Table 2.15), when using the *Chondrus cripus* serine protease sequence as a query sequence in the BLASTp search engine.

Table 2.15: Red algae serine proteases identified by BLASTp programme.

	MAX SCORE	QUERY COVER	E-VALUE	IDENTITY	ACCESSION
CHOC	463	100%	0.0	100%	XP_005716713
CYME	400	74%	2e-135	57%	XP_005536821
GASU	428	74%	8e-146	59%	XP_005705279
PUMB	396	77%	2e-133	58%	OSX70935
PURP	439	96%	8e-153	62%	Contig_814.1
PYEZO	413	62%	1e-139	58%	Contig_12937_g3083

University of Fort Hare
Together in Excellence

2.6.2.4.1.2 Family and domain search

Several domains were identified from online domain databases (Table 2.16), in the identified red algae protein sequences. The domains identified included the PDZ domain of trypsin-like serine proteases, domain found in peptidase family S1, Trypsin-like peptidase domain, which confirmed that the red algae sequences are serine proteases.

Table 2.16: Domain identifiers and their position within the serine proteases from red algae.

Database resource	***Choc***	***Cyme***	***Gasu***	***Pumb***	***Pyezo***	***Purp***
NCBI CDD: **Cd00987**	242-439	Cd00987	440-538	405-497	403-495	71-210
INTERPRO **IPR009003** **IPR036034**	111-333 315-432	142-361 352-470	211-428 439-534	173-391 389-496	171-389 387-494	15-236 245-345
PFAM: **PF13365** **PF13180**	167-306 343-436	196-335 371-469	265-404 440-538	230-369 406-497	228-367 404-495	71-210 247-344
SUPERFAM: **SSF50494** **SSF50156**	11-333 315-432	142-361 352-470	211-428 439-534	173-391 389-496	171-389 387-494	15-336 245-345
PANTHER: **PTHR22939** **PTHR22939:SF109**	67-434 67-434	116-469 116-469	185-540 185-540	165-505 165-505	155-505 155-505	4-348 4-348
PRINTS: **PR00834**	390-402	205-217 226-246 268-292 306-323 328-345 419-431	274-286 295-315 337-361 375-392 397-414 488-500	239-251 260-280 302-326 340-357 362-379 453-465	237-249 258-278 300-324 338-355 360-377 451-463	294-306
SMART: SM00228	342-432	371-461	440-530	405-495	403-493	246-336
GENE3D: G3DSA:.30.42.10 G3DSA:2.40.10.10	235-332 340-447	264-361 366-484	438-553 334-430	298-395 403-512	296-393 401-510	244-358 140-358

University of Port Harcourt
Together in Excellence

2.6.2.4.1.3 Sequence properties

All the red algae sequences showed similar protein lengths (463 to 567 aa), except for Purp, which had a much shorter sequence (364 aa). The pI values were within the pH 10 area, except for Purp sequence which had a more acidic pI of 5.7 (Table 2.17). No signal peptide was found in any of the sequences and all of the sequences were located to the chloroplast.

Table 2.17: Predicted sequence properties of red algae trypsin-like protease.

	Choc	*Cyme*	*Gasu*	*Pumb*	*Pyezo*	*Purp*
Length (aa)	463	510	558	552	567	364
PI	10.63	9.55	9.54	9.08	9.4	5.7
Gravy	-0.091	-0.239	-0192	0.032	0.072	-0.071
Signal Peptide	None	None	None	None	None	None
Subcellular Localisation	Chlo	Chlo	Chlo	Chlo	Chlo	Chlo
Biological Activity	GO:0006508	GO:0006508	GO:0006508	GO:0006508	GO:0006508	GO:0006508
Molecular Activity	GO:0004252 GO:0005515	GO:0004252 GO:0005515	GO:0004252 GO:0005515	GO:0004252 GO:0005515	GO:0004252 GO:0005515	GO:0004252 GO:0005515
Solubility	Soluble 0.677	Soluble 0.807	Soluble 0.68		Soluble 0.689	Soluble 0.971

University of Fort Hare
Together in Excellence

2.6.2.4.1.4 Sequence alignment

The sequence alignment of the predicted serine protease should a high degree of sequence similarity and that the catalytic triad, His, Asp, Ser, were conserved in all the sequences (Figure 2.19). The red algae sequences grouped into a single clad, which was separate from that of the green algae and plant serine proteases (Figure 2.20), but showed higher similarity to the green algae sequences than to the plant sequences.

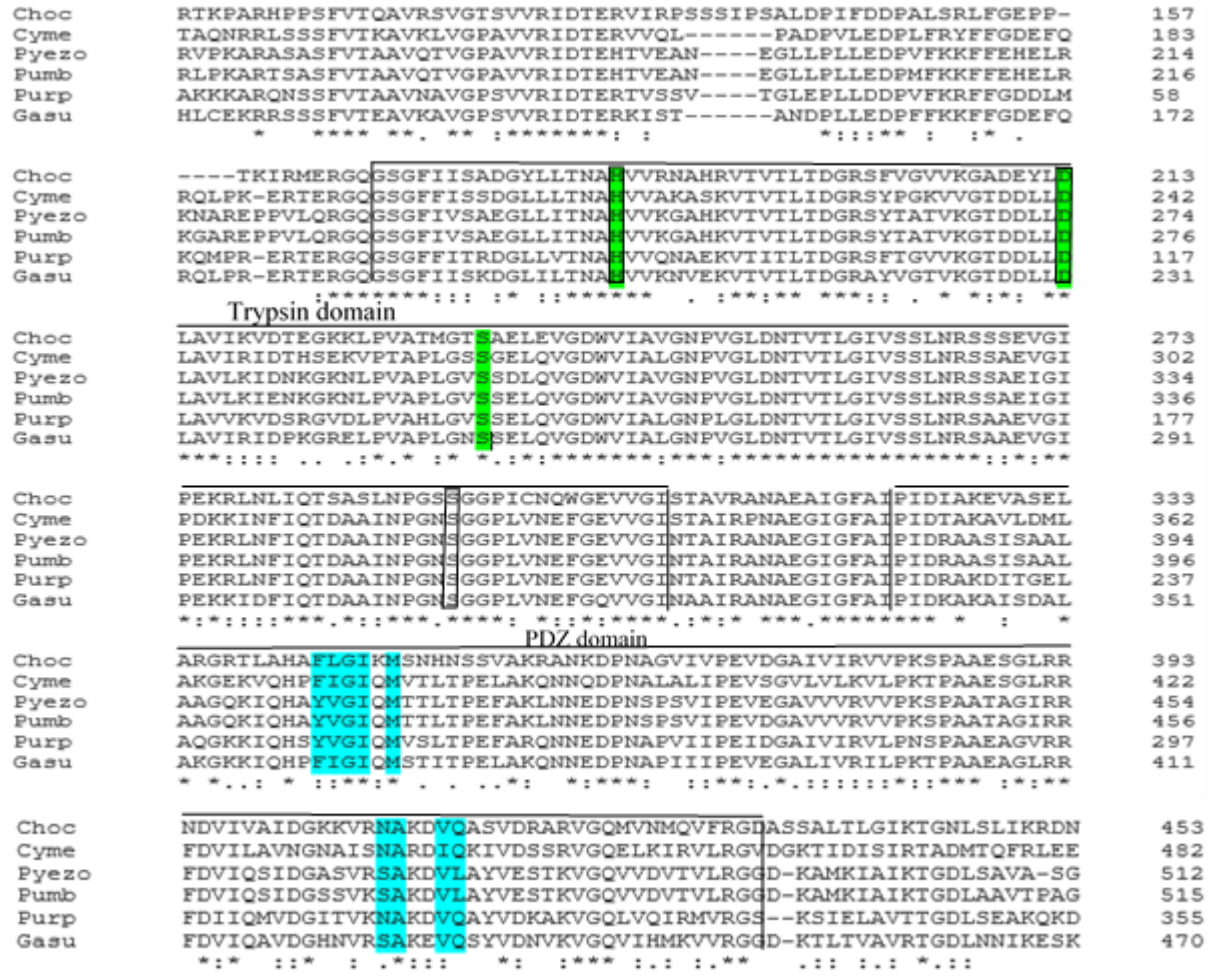


Figure 2.19: An alignment of serine protease sequences from red algae.

The cyan coloured residues are the binding sites in PDZ domain and the bright green are active site residues in trypsin domain. The non-conserved (gaps) were manually truncated.

2.6.2.4.1.5 Phylogenetic analysis

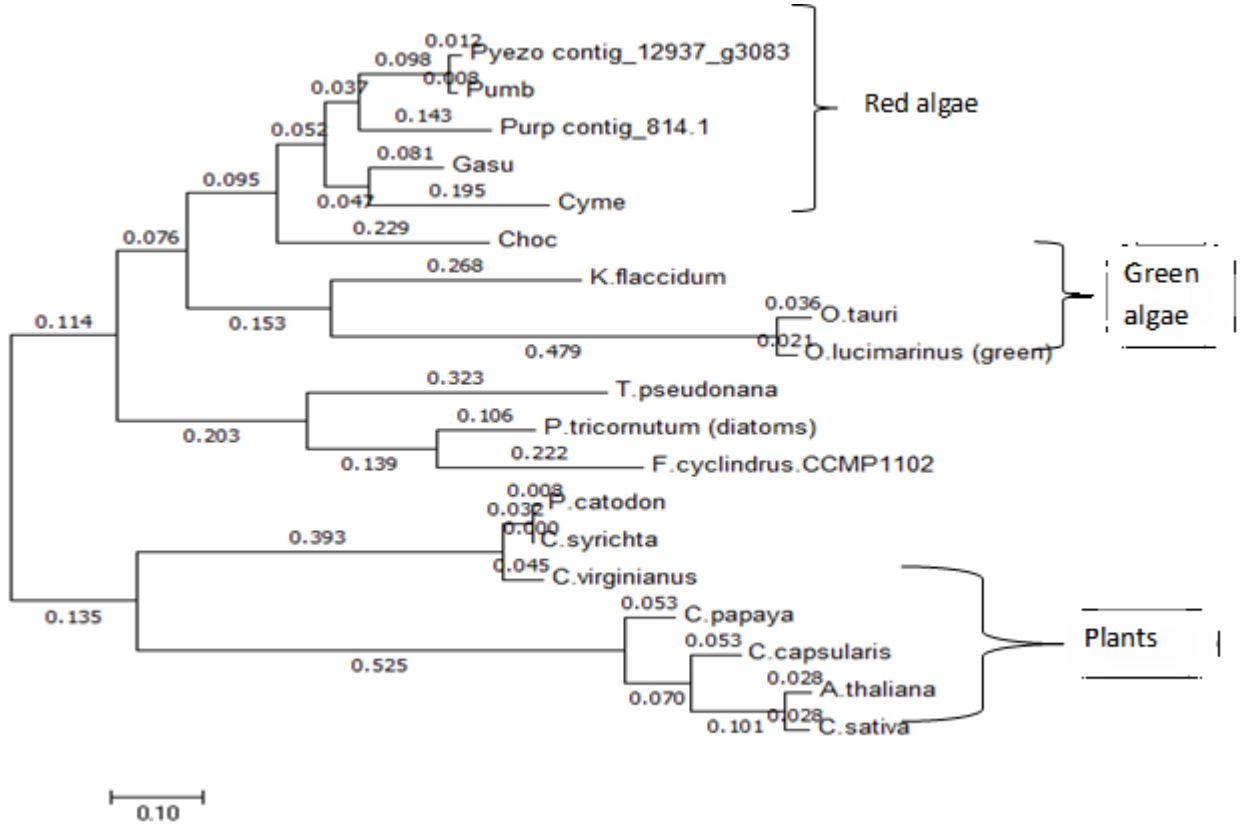


Figure 2.20: Phylogenetic analysis of threonine proteases by maximum likelihood based on the JTT matrix-based model. The sequences were aligned using clustal W in mega 6 software. The overhang gaps were cleaved off.

University of Fort Hare
Together in Excellence

2.6.2.4.2 Structure analysis

2.6.2.4.2.1 Secondary structure prediction

Analysis of the secondary structure of the serine protease-like sequences from the various red algae, showed a high degree of variation, with some sequences having a high percentage of β-sheets, while others showed a high percentage of α-helices (Figure 2.21 and Table 2.18). However, the predicted tertiary structures (Figure 2.22) were very similar, which is also indicated by the high convervation of the structure of the active site pocket (Figure 2.23)

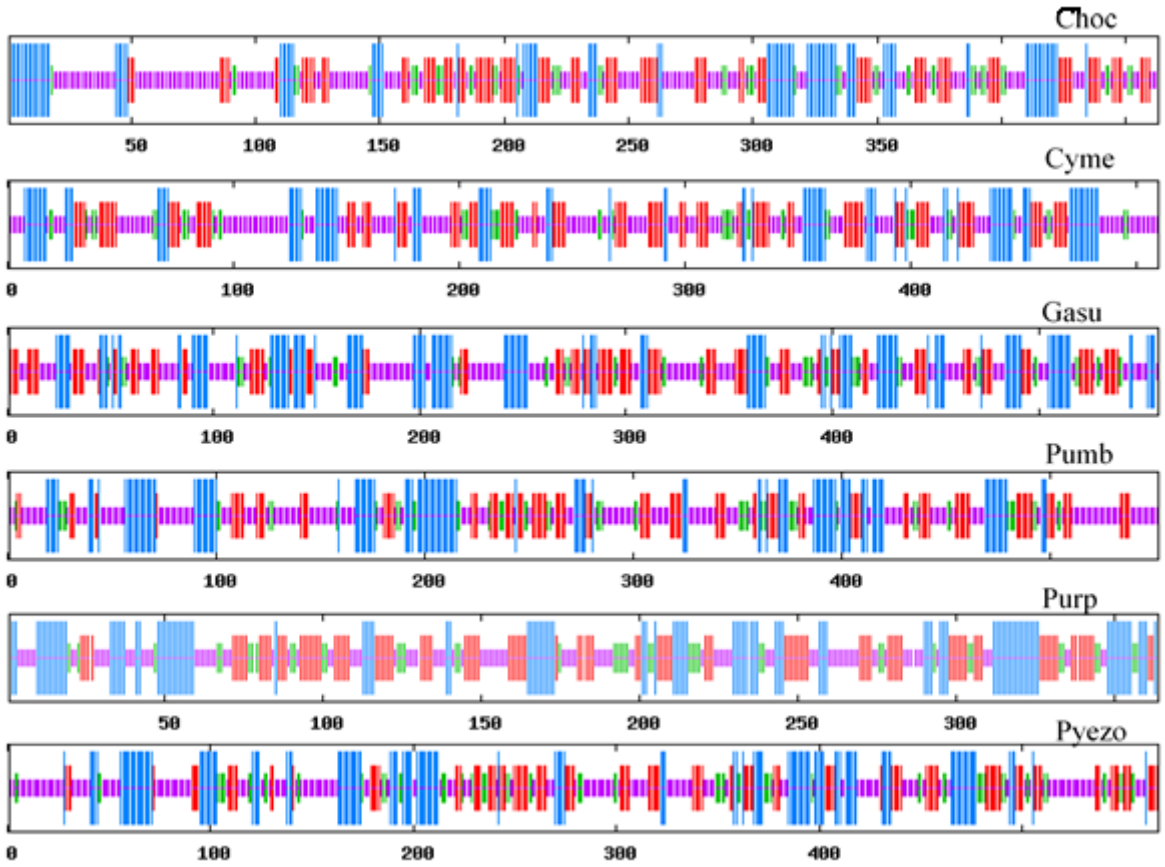


Figure 2.21: The secondary structures of red algae serine proteases predicted based from the sequence..
The blue colour = alpha helix, red = extended strands, green = beta green, purple = random coils, and the numbers are the sequence length.

Table 2.18: Estimated percent of predicted secondary structure elements from the red algae aspartic proteases.

SS Element	*Choc*	*Purp*	*Cyme*	*Pumb*	*Pyezo*	*Gasu*
Alpha helix	20.52%	28.57%	19.80%	23.15%	23.63%	26.16%
Extended strand	25.27%	29.12%	25.29%	22.46%	22.57%	24.91%
Coil	44.28%	30.77%	45.49%	44.57%	41.80%	40.5%
Beta turn	9.94%	11.54%	9.41%	9.78%	11.99%	8.42%

2.6.2.4.2.2 Tertiary structure prediction

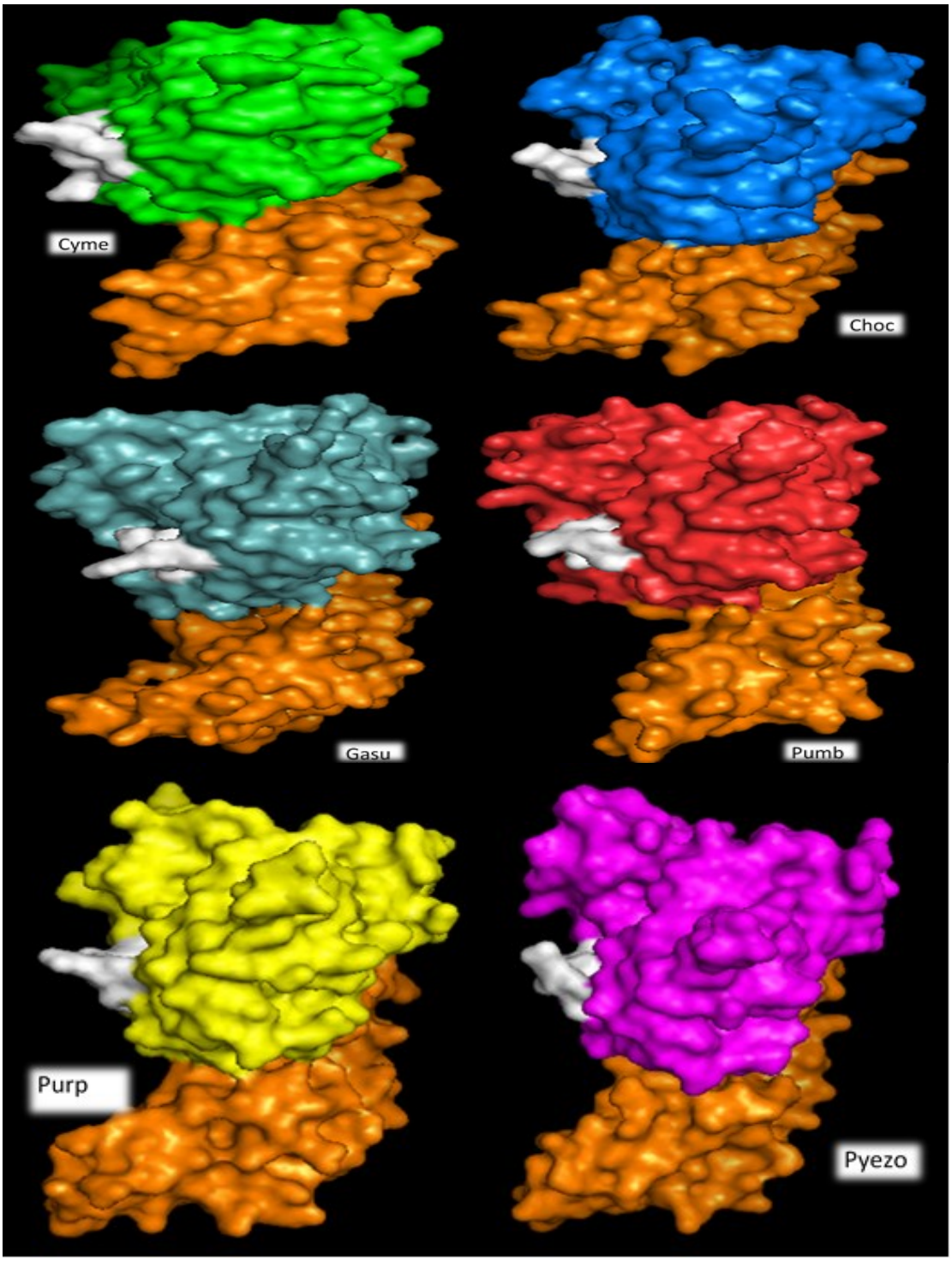


Figure 2.22: The two domain structure of red algae trypsin-like protease viewed as surface in PyMol.
The trypsin domain is coloured green (cyme), blue (Choc), cyan (Gasu), Red (Pumb), yellow (Purp) and Purple (Pyezo). The PDZ domain and a connecting loop are coloured brown and white, respectively.

2.6.2.4.2.3 Active site prediction

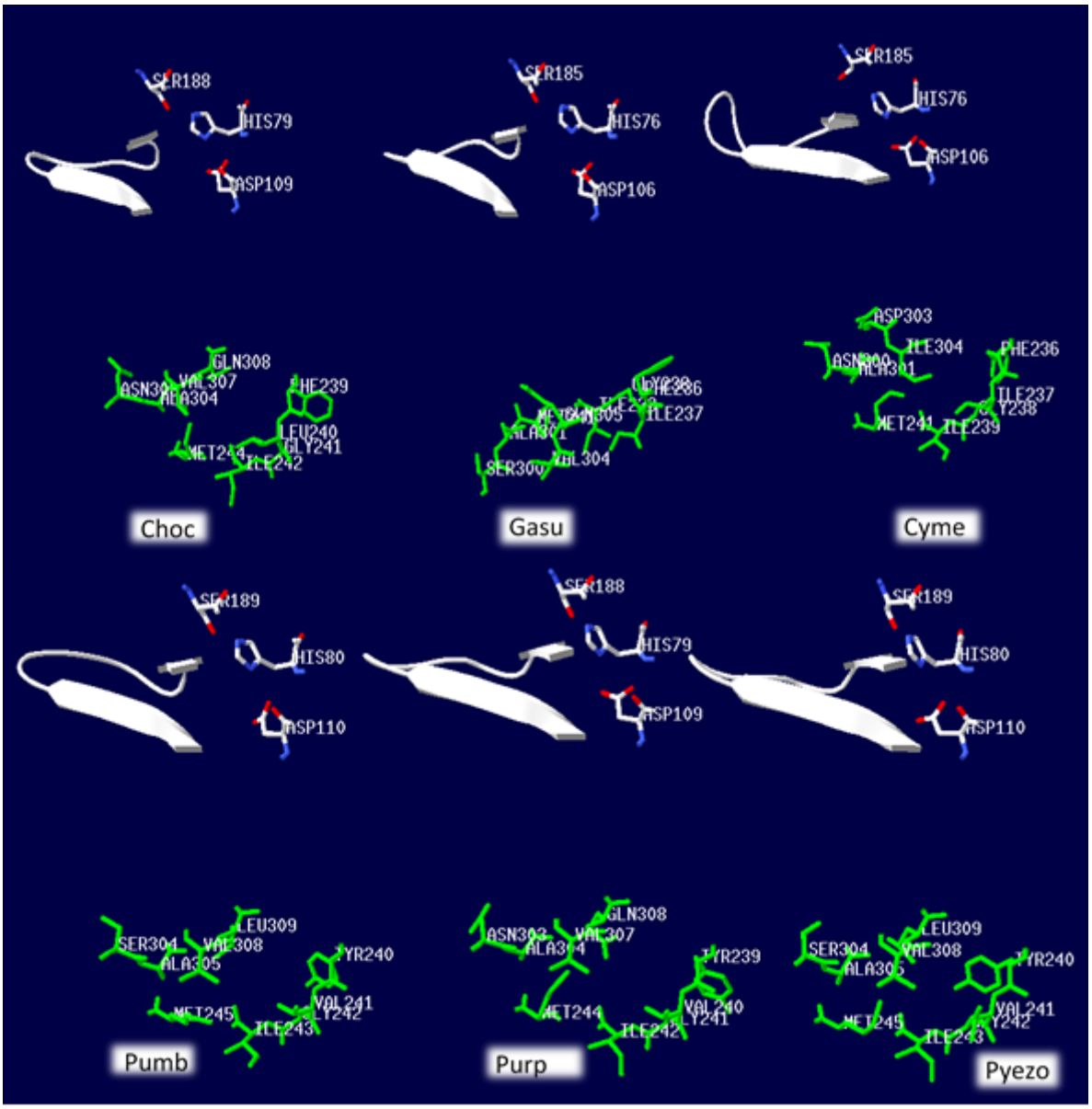


Figure 2.23: The catalytic sites from the red algae trypsin-like proteases viewed in spdbViewer.
The active site residues (in trypsin domain) are coloured in cpk and substrate binding sites (in PDZ domain) are coloured in green. The connecting loop is coloured white and viewed as ribbon.

2.7 Conclusion

Protease sequences from all six protease classes were identified in the six red algae genomes that have been sequenced and are publically available. The serine proteases were identified as an important group of algae proteases to isolated and characterise, due to their medical and biotechonology applications. Chapter 3 describes the attempt to clone and express a serine protease and Chapter 4 the isolation, purification and kinetic characterisation of the *Gelidium pristoides* serine protease.

Chapter 3
Cloning and Expression of *Gelidium pristoide* Serine Protease

University of Fort Hare
Together in Excellence

3. INTRODUCTION

Developments in DNA sequencing methods have led to an increase in genomic information which awaits annotating. Because the number of genome sequences is increasing rapidly, the most frequent approach for annotation is computer-based (see Chapter 2). Although *in silico* analysis methods have been successful, serious concerns have been raised because the number of computationally predicted genes is far greater than the experimentally validated genes with a known function. Potential errors in predicting a gene function may be transferred from one annotation to the next (Schnoes *et al.*, 2010). Therefore, precise function assignment, exclusively based on the computational approach, is not yet possible. Furthermore, it has been noted that:

- About 20–40% of the genes predicted computationally still have no functional data, including model organisms such as *E. coli* k12, *A. thaliana* (Dhanyalakshmi *et al.*, 2016; Martinez, 2013).
- Some genes with putative function have not been experimentally confirmed.
- Owing to the high volume of information in the databases, mis-annotation is possible.
- The computational approach cannot prove whether a predicted gene is active, and its biochemical properties cannot be determined (Shoemaker *et al.*, 2001).

These problems raise the need for experimental validation of the predicted genes. Computational methods are important in that they provide clues to the gene structure and function and assist in determining the choice of assays for enzymes (Bkhaira *et al.*, 2016). If a high level of confidence is obtained from computational predictions, a variety of experimental methods, such as PCR-based methods with subsequent sequencing, are undertaken to validate those predictions (Yandell *et al.,* 2005). In this chapter, we attempt to amplify the previously identified protease gene using reverse transcription polymerase chain reaction (RT-PCR) with the aim of sequencing and cloning the amplicons.

University of Fort Hare
Together in Excellence

RT-PCR is the synthesis of cDNA molecules from the mRNA. It is gene-specific and takes advantage of spliced mRNA (Gardner and Slazek, 2015). This method has proved to be successful in validating gene prediction, analysing gene expression, and in target-gene cloning (Porcar and Juárez-Pérez, 2003). Previous studies have shown its efficiency, and early human gene predictions have been validated using this method (Tenney *et al.*, 2004). A disadvantage of the approach lies in the fact that it depends on the expression, and genes expressed at very low levels are usually missed (Brzoska *et al.*, 2006). Usually, this method is more effective when it is coupled with sequencing.

For gene activity analysis, this method is joined together with cloning to produce recombinant proteins for biochemical characterisation (Yang *et al.,* 2009). We aimed to obtain active full-length transcripts; thus, the primers used were target-gene specific and were designed on the basis of open reading frame (ORF) from start and stop codons.

3.1 Aim and Objectives

3.1.1 Aim

- To experimentally clone and express the predicted protease gene(s) from the red seaweeds found at Kenton-On-Sea, Eastern Cape.

3.1.2 Objectives

3.1.2.1 To collect red seaweed from Kenton-On-Sea, Eastern Cape

3.1.2.2 To extract total RNA from the collected red seaweeds

3.1.2.3 To design primers based on predicted protease gene(s)

3.1.2.4 To amplify the identified protease gene(s)

3.1.2.5 To clone and express the protease gene(s)

3.1.2.6 To characterise the recombinant protease(s)

3.2 Methods

3.2.1 Algal sample preparation

Figure 3.1: Satellite view of Kenton-on-Sea beach [Map data©2018 AfriGIS (Pty) Ltd, Google].
The map was zoomed to 650 m. Latittude: 33º41'44.32" S and Longitude: 26º40'048" S

The algae samples were collected by scraping them off the semi-submerged rocks at Kenton-On-Sea, adjacent to the Bushman's River mouth (Figure 3.1). The algae samples were kept on ice and transported to the lab in plastic bags. Before storing at -80°C, they were washed with distilled water to remove salt and remaining rock debris. The epiphytes were removed by a blade, wiped with 100% alcohol.

3.2.2 RNA isolation

The red seaweed was frozen in liquid nitrogen and then ground into powder using a pestle and mortar. The RNA was isolated from the resultant powdered algae using an SV total RNA extraction kit from Promega. The kit was optimised by using different masses of the powered algae. The manufacturer's protocol was followed and DNases 1 was used to remove the DNA molecules. The presence of RNA was analysed by agarose gel (1%) electrophoresis.

3.2.3 Primer design

The primers (Table 3.1) were designed on the basis of ORF (Open reading frame) using the gene runner software version 5.1.06 beta.

Table 3.1: Primer sequences and their melting temperatures.

Primers:	**Forward**	**Reverse**
Cysteine-like protease		
Tm max/min	63.47/63.47	62.86/62.86
sequence	5'-ATGGGCTTAGGCAGCGTGCTG -3'	5' -TCACGTGTAGGCTAATAAGGCATC -3'
Serine-like protease		
Tm max/min	58.02/58.02	62.86/62.86
Sequence	5'-ATGTTCGTCATTAAGAATGCAAAG-3	5'-TTACGCATGTGCAGAAGAAGCTAC-3'

3.2.4 Reverse Transcriptase PCR

The cDNA was synthesised in one reaction tube with PCR using the Qiagen One-Step RT-PCR kit. The kit contained enzymes developed specifically for both reverse transcription and PCR. These enzymes included Omniscript (for RNA amounts>50ng) and sensiscript (for small RNA amounts) reverse transcriptases, and hot start DNA polymerase. These are activated by a change in temperature as described by the protocol. The reaction components are shown in Table 3.2. and the thermal cycler conditions are shown in Table 3.3.

University of Fort Hare
Together in Excellence

Table 3.2: Reaction components for One-step RT-PCR.

Components	Volume/reaction	Final conc.
RNAse-free water	10 µl	—
5x Qiagen one-step buffer	5.0 µl	1x
dntp mix (containing 10 mm of each dntp)	1.0 µl	400 µM of each dNTP
Rev Primer	1.5 µl	0.6 µM
Fwd Primer	1.5 µl	0.6 µM
Qiagen one-step RT-PCR enzyme mix	1.0 µl	___
Template RNA, added at step 4	**	1 pg – 2 µg/reaction
Total volume	25 µl	

Table 3.3: Thermal cycler conditions.

Cycle steps	***Temp.***	***Time***	***No. Of cycle***
Reverse transcription	50°C	30 mins	1
Initial PCR activation step	95°C	15 mins	1
3-STEP CYCLING: ➢ ***Denaturation*** ➢ ***Annealing*** ➢ ***Extension***	 94°C 55°C 72°C	 1 min 1 min 1 min	40X
Final extension:	72°C	10 min	1

3.2.5 DNA isolation

DNA was isolated as an alternative because of numerous unsuccessful attempts at amplification. As with RNA isolation, DNA was isolated from the powdered algae using ZR plant/seed DNA prep ™ kit. RNase 1 was used to remove residual DNA fragments, following the manufacturer's protocol. Some of the DNA was provided by SAIAB. All the DNA samples were analysed by agarose gel (0.8%) electrophoresis.

3.2.6 Standard PCR

The isolated DNA was used as a template to amplify the protease gene of interest. The standard PCR was performed using a PCR kit from Takara Bio Inc. The preparation of reaction mix is summarised in Table 3.4 and the thermocycler conditions were set according to Table 3.5. The resultant amplicons were visualised in agarose gel (1%).

University of Fort Hare
Together in Excellence

Table 3.4: The standard PCR reaction components.

Components	Volume
Takara EX taq	0.125 µl
10x EX taq buffer	2.5 µl
dntps	2 µl
Rev. Primer	0.25 µl
Fwd. Primer	0.25 µl
Template	**
Nuclease-free water	variable
Total	25 µl

Table 3.5: Thermal cycle conditions.

Cycle steps	Temp.	Time	No. Of cycles
Initial PCR activation step	94°C	30 s	1
3-step cycling: ➢ Denaturation ➢ Annealing ➢ Extension	 94°C 55°C 72°C	 30 s 30 s 3 min	 x40
Final extension	72°C	10 mins	1

3.2.7 Gradient PCR

The gradient PCR was optimised to reduce the primer dimer formation using a PCR kit from Thermo Scientific. The preparation of reaction mix is summarised in Table 3.6 and the thermocycler conditions were set according to Table 3.7. The amplicons were visualised on the agarose gel (1%) electrophoresis.

University of Fort Hare
Together in Excellence

Table 3.6: PCR reaction components.

Components	Volume
PCR master mix	12.5 µl
Rev. Primer	0.25 µl
Fwd. Primer	0.25 µl
Template	5 µl
Nuclease-free water	7 ul
Total	25 µl

Table 3.7: Thermal cycler conditions

CYCLE STEPS	TEMP.	TIME	NO. OF CYCLES
Initial PCR activation step	95°C	15 mins	1
3-STEP CYCLING: ➢ Denaturation ➢ Annealing ➢ Extension	 94°C 50-65°C 72°C	 1 min 1 min 1 min	 40X
Final extension	72°C	10 mins	1

3.2.8 Gradient PCR with additives

Bioinformatics analysis revealed that red seaweeds are rich in GC content, and as a result, designed primers had secondary structures. To reduce the formation of such undesired structures, additives were included in the reaction (Table 3.8 and 3.9). The results were analysed using (1%) agarose gel electrophoresis.

Table 3.8: PCR reaction components.

COMPONENTS	***VOLUME***
PCR master mix	12.5 µl
Rev. primer	0.25 µl
Fwd. Primer	0.25 µl
Template	5 µl
DMSO	2%
Glycerol (0.5%)	1%
Nuclease-free water	6 ul
Total	25 µl

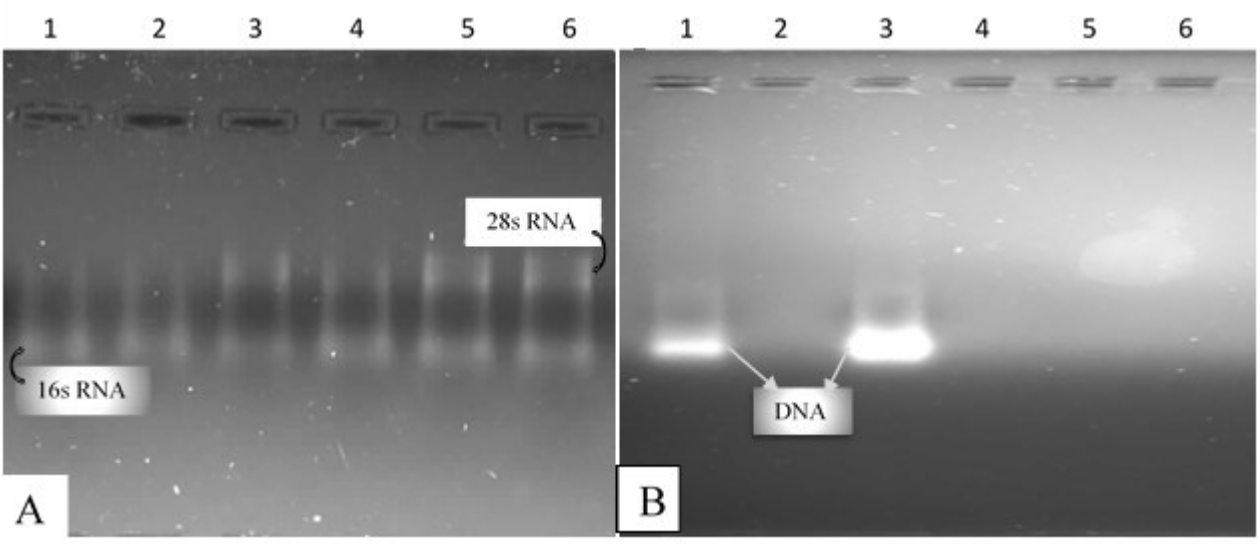
University of Fort Hare
Together in Excellence

Table 3.9: Thermocycler conditions.

Cycle steps	***Temp.***	***Time***	***No. of cycles***
Initial PCR activation step	95°C	15 min	1
3-step cycling:			
➢ *Denaturation*	94°C	1 min	
➢ *Annealing*	45-68°C	1 min	40X
➢ *Extension*	72°C	1 min	
Final extension	72°C	10 min	1

3.3 Results and Discussion

Earlier hypothetical protease genes were identified from different red algae genomes, using bioinformatics tools (Chapter 2). The bioinformatics analysis provided clues regarding the function of these genes. In this chapter, we aimed to identify and clone protease genes, through the molecular techniques, from the South African seaweeds. The primers were designed on the basis of the ORF to obtain cDNA molecules from the active full-length transcripts (**Table 3.1**). Owing to the lack of genomic sequences for the sampled red seaweeds, primers were designed based on the *Chondrus crispus* genome sequence. This is the first macro-alga species to have a complete genomic sequence and has been used as a model species for florideophytes. This approach has previously been used successfully, regardless of its disadvantage of specificity (Eakin *et al.*, 1990; Sakanari *et al.*, 1989).



Together in Excellence

Figure 3.2: Nucleic acid isolation from Gelidium pristoides and Plocamium corallorhiza sampled from Kenton-on-Sea.
Figure 3.2A: The total RNA isolation. Lanes 1 to 3 show the optimisation (sample = 30 mg, 60 mg, 100 mg) for *Gelidium pristoides*. Lanes 4 to 6 show optimisation (as shown above) for *Plocamium corallorhiza*. Figure 3.2B: The isolated DNA from *Gelidium pristoides* (lane 1) and *Plocamium corallorhiza* (lane 3).

Subsequently, upon successful total RNA isolation (Figure 3.2A), a reverse transcription polymerase chain reaction (RT-PCR) was performed to obtain an active cDNA using primers specific for cysteine protease gene. These primers were tested for *Gelidium pristoides* and *Plocamium corallorhiza*. However, the obtained results were negative as the agarose gel showed no bands (Figure 3.3A). At this stage, a number of possible causes were identified: the specificity of primers, a low level of expression, and/or the gene itself being absent in the genome. Contrastingly, bioinformatics analysis (Chapter 2) showed that the gene was present in all the genome sequences analysed. Therefore, either the primers were not specific, or the level of expression was low.

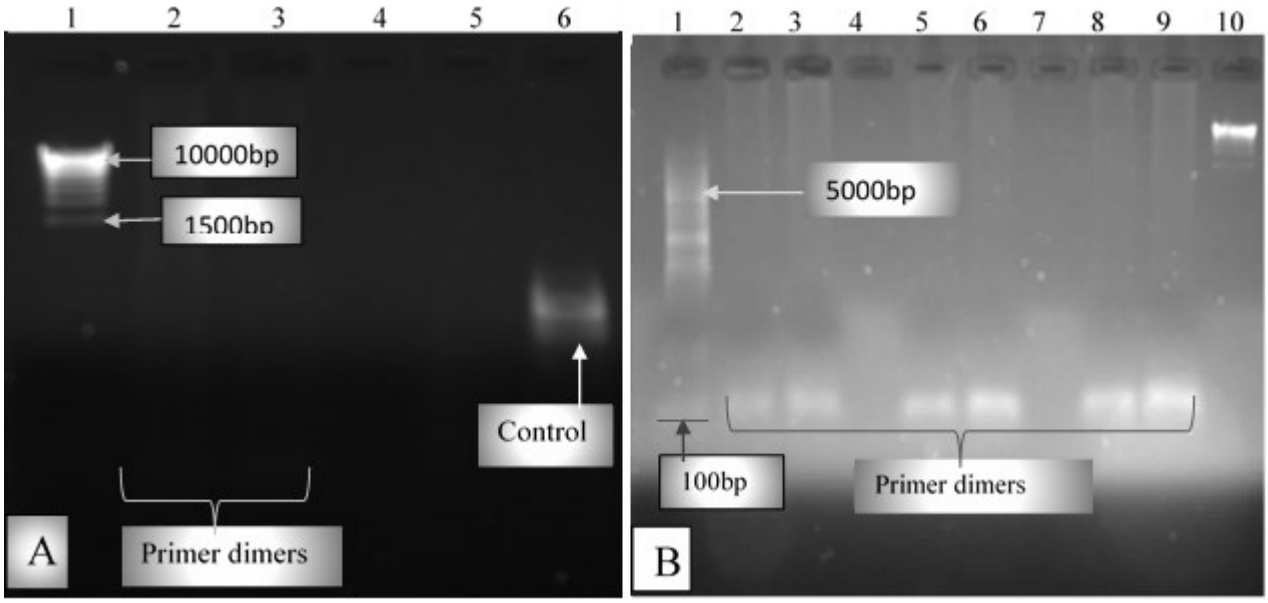


Figure 3.3: First attempt to amplify the cysteine protease gene.
Figure3.3A: Reverse transcriptase polymerase chain reaction (RT-PCR) results for *Gelidium pristoides* (lanes 2 and 3) and *Plocamium corallorhiza* (lanes 4 and 5). High range ladder was loaded in lane 1 and a control in lane 6. Figure3.3B: Conventional PCR results (from DNA) for *Plocamium corallorhiza* (lanes 2 and 3) and *Gelidium pristoides* (lanes 5 and 6). Lanes 1 and 10 were loaded with middle range and high range DNA ladder, respectively. Lanes 8 and 9 show results for alcohol dehydrogenase.

Consequently, we isolated the DNA from the same species (Figure 3.2B) and performed a conventional PCR. The expected size was 1863 base pairs, unexpectedly the results showed bands that corresponded to 100bp of the middle range ladder (100bp–5000bp). These were run on agarose gel, together with dehydrogenase results (from the Yanga Gogela's study) (Figure 3.3B). It was noticed that the bands were also of the same range, and thus were considered as the primer dimers. Even if these bands were the gene of interest, this would mean that the cDNA was incompletely amplified. As a result, further processing would be difficult and would lead to incorrect assumptions.

The cause of the dimer formation was suspected to be the presence of primer secondary structures. This resulted in the use of another variant of PCR (temperature gradient PCR), and the design of new primers. New primers, specific for serine protease gene, were designed in the same manner and were analysed for secondary structures (Table 3.1). The results showed

no signs of amplification (Figure 3.3B). The gradient PCR results for cysteine protease were similar to those previously obtained from traditional PCR (Figure 3.4A and Figure 3.4B). Owing to the persistence of the primer dimer formation, we modified the method by adding PCR additives, such as glycerol and dimethyl sulfoxide (DMSO) (Table 3.8).

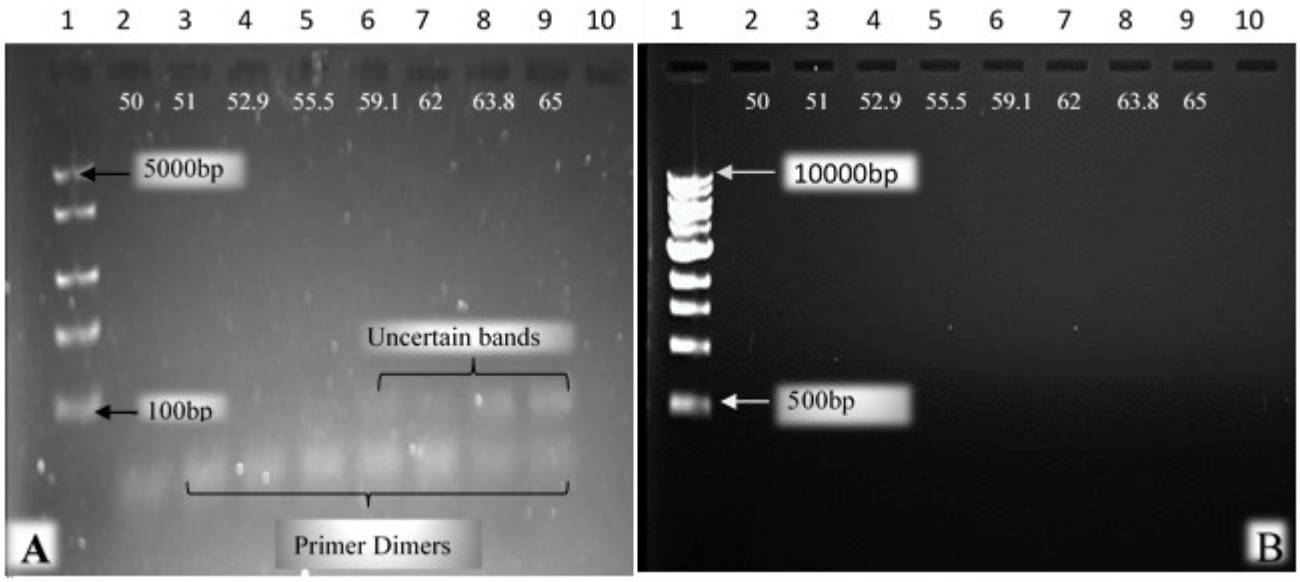


Figure 3.4: Gradient PCR results for *Plocamium corallorhiza*.
Figure 3.4A: Amplification using primers specific for cysteine protease gene. Figure3.4B: Serine protease gene amplification using gene specific primers. Amplicons were loaded from lanes 2 to 9 and annealing temperatures (°C) are shown below the lane wells. A 1kb ladder (500 to 10000bp) was used for both gels.

University of Fort Hare
Together in Excellence

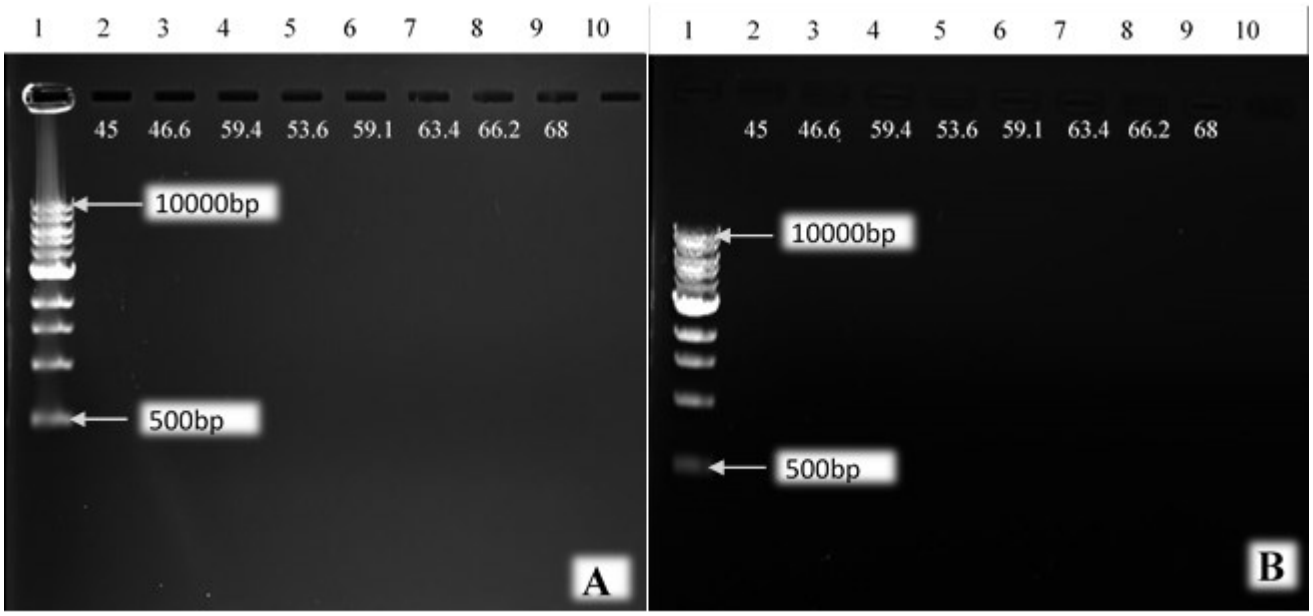


Figure 3.5: Gradient PCR results for Plocamium corallorhiza.
Figure 3.5A: Amplification using primers specific for cysteine protease gene. Figure3.5B: Serine protease gene amplification using gene specific primers. Amplicons were loaded from lane 2 to 9 and annealing temperatures (°C) are shown below the lane wells. We used 1kb ladder (500 to 10000bp) for both gels.

Agarose gel electrophoresis showed no signs of amplification and there was no formation of the secondary structures (Figure 3.5). We concluded that the reason for unsuccessful amplification lay in the primers being not specific, and/or the absence of the gene(s) of interest. In the following chapter, we describe the test for the presence of the native protease enzyme(s).

University of Fort Hare
Together in Excellence

The unavailability of genomic sequences in red algae, together with the lack of annotation for the sequenced genome, have been the major hurdle in research; within these genomes there are numerous bioactive compounds which could be utilised in industry and therapy. However, there are various approaches which can be used to overcome the unavailability of genomic sequences: synthesis of the gene (usually mRNA to avoid introns), purification and characterisation of the enzyme (most effectively with subsequent sequencing), and lastly, coupling the genome sequencing with annotation. In Chapter 4, we attempt to identify, purify, and characterise the protease enzyme from the red algae.

Chapter 4
Purification and kinetic characterisation of Serine Protease from *Geldium pristoides*

University of Fort Hare
Together in Excellence

4. INTRODUCTION

In the previous chapter, we attempted to identify genes from the red seaweeds found at Kenton-on-Sea. However, amplification was unsuccessful because of the possible absence of the gene of interest, or because primers were not specific. In this chapter, we attempt to identify and characterise serine-like and cysteine-like proteases in order to determine whether the primers were unspecific (if the protease enzyme(s) of interest is present), the gene(s) of interest was absent (if the enzyme was absent), or the gene was not expressed.

In general, seaweeds are found attached to rocks, and serve as shelter and a food source for marine organisms such as fish and sea urchins (Mumtaj *et al.*, 2015). Seaweeds are affected by the physical features of their habitat; they rely on continuous water movement for nutrients and gases. Sometimes they are exposed to diseases caused by opportunistic microorganisms (*Salmonella* and *Vibrio* species) of which fish are known to be carriers (Pal *et al.*, 2014; Thanigaivel *et al.*, 2015).

The sampled red seaweeds were found in the temperate intertidal zone of Kenton-on-Sea. The red algae in this area are usually subjected to varying habitat factors that affect growth and survival: light, as the coast is exposed; temperature, which changes seasonally with high rainfall during winter; desiccation and lack of nutrients during low tide, and salinity (Anderson *et al.*, 2009). Seaweeds adapt to survive such stresses by developing defence mechanisms mediated by specific proteins and enzymes (Pal *et al.*, 2014). Berges and Falkowski (1998) reported the induction of chlorophyte and diatom proteases in response to nitrogen or light limitation.

In comparison to higher plants, there is a lack of information on proteases in algae. Although proteases have been described in some green micro- and macroalgae (Kadokami *et al.*, 1990; Matsubara *et al.*, 1998, 1999, 2000, and 2002), very few studies have focused on red seaweed peptidases (El-shora *et al.*, 2016; Pérez-Lloréns *et al.*, 2003; Wang *et al.*, 1998).

University of Fort Hare
Together in Excellence

4.1 Material and Methods

4.1.1 Algal collection and preparation

The algae samples (*Gelidium pristoides*) were collected by scraping them off the semi-submerged rocks found along the Kenton-on-Sea coast. The algae samples were kept on ice and transported to the lab in plastic bags. Before storing at -80°C, they were washed with running tap water to remove salts and remaining rock debris. Visible epiphytes were removed before the seaweed was ground in liquid nitrogen.

4.1.2 Protease extract preparation

The algal sample (50 g) was ground in liquid nitrogen using a pestle and mortar. The liquid nitrogen was allowed to evaporate before suspending the resultant powder in 150 mL of extraction buffer (0.1M Tris-HCl, pH 8, 1% Triton X 100). The resultant homogenate was centrifuged (Heraeus Multifuge 3SR+ centrifuge, Thermo Scientific) at 4 600 rpm for 10 min at 4°C.

4.1.3 Specific activity with chromogenic substrate

The enzymes were screened and assayed using Nα-Benzoyl-L-arginine p-nitroanilide (BApNA). BApNA is a chromogenic substrate that, when hydrolysed at the arginine site, releases p-nitroanilide which gives a yellowish colour. P-Nitroanilide can be read at 405 nm (Silva *et al.*, 2011) and 410 nm (Shamsi *et al.*, 2016). BApNA has been used successfully in cysteine (Yoneda *et al.*, 2012; Raśković *et al.*, 2015) and serine protease assays (Khandgale *et al.*, 2015).

4.1.3.1 Trypsin

The trypsin-like protease activity assay described by Silva *et al.* (2011) was followed with slight modifications. BApNA (8mM) was prepared in dimethyl sulphoxide (DMSO) and added to the final concentration of 1.2 mM. The reaction mixture contained 30 µl of algal sample (or commercial trypsin for standard) and 140 µl of the buffer (0.1 M Tris-HCl, 8 mM $CaCl_2$, pH 8). The enzyme activity was followed at 25ºC for 10 min and read at 405 nm. One unit of enzyme activity was defined as the amount of enzyme capable of hydrolysing 1 µmol of BApNA per min under the conditions, using a molar coefficient of 9100 $mM^{-1}cm^{-1}$.

Equation 1 (Activity): $U = [(Slope \times Vf \times 1000)df] \div (9\,100 \times EV)$

Equation 2 (Relative activity): % RU = $(U_{test} \div U_{untreated})$ X 100

Definitions

Slope = A_{405}/min

Vf = final volume of the reaction

df = dilution factor based on unit definition

1000 = change in slope based on unit definition

EV = amount of the enzyme used

RU = relative activity

U_{test} = units of the treated enzyme

$U_{untreated}$ = units of the untreated

University of Fort Hare
Together in Excellence

4.1.3.2 Cysteine protease

Cysteine protease activity was measured according to the assay described by Potempa *et al.* (1998) with slight modifications. Briefly, the reaction consisted of 140 µl of reaction buffer (0.1 M Tris-HCl, 0.02 M NaCl, 8 mM $CaCl_2$, 2 mM soybean trypsin inhibitor, 10 mM L-cysteine, pH 7.6), and 30 µl of algal sample. The reaction was initiated by adding BApNA (8 mM) to the final concentration of 1.2 mM. The release of p-nitroanilide was monitored at 37ºC for 10 min and read at 405 nm. One unit of enzyme activity was defined as the amount of enzyme capable of hydrolysing one µmol of BApNA per min under the above-mentioned conditions, using a molar coefficient of 9100 mM^{-1} cm^{-1}.

4.1.4 Enzyme Purification

The crude extract was filtered and subjected to ion exchange chromatography. The columns used were DEAE FF 16/10, CM FF 16/10, and HiPrep^{TM} Q FF 16/10 equilibrated with 5 column volumes (1 Volume = 20 mL) of 0.1 M Tris-HCl, pH 8 at 5 ml/min. Unbound proteins were washed with 2 column volumes of the same buffer and collected in 2 ml. The bound proteins were eluted with 5 volumes of 1M NaCl in equilibration buffer, and 2 ml was collected.

4.1.4.1 Optimisation

The powdered algal sample was increased to 100 g per 250 mL of the extraction buffer (0.1 M Tris-HCl, pH 5.5, 1% Triton X 100) and incubated at 4°C (in the fridge for 2 hours while stirring. The sequent mixture was centrifuged at 4 600 rpm for 15 minutes at 4°C.

The total proteins were concentrated by ammonium sulphate fractionation (0-30% and 30-70%). Subsequently, centrifugation was performed at 4600 rpm for 15 min at 4°C. The pellets were dissolved in 20 mL of 0.02 M Tris-HCl, pH 5.5. The fractions with activity were pooled and dialysed in 0.02 M Tris-HCl, pH 5.5 containing 8 mM $CaCl_2$ for 16 hours with two subsequent buffer changes.

The dialysate was purified using ion exchange chromatography. The column was HiPrep DEAE FF 16/10, CM FF 16/10, and HiPrep Q FF 16/10 equilibrated with 5 column volumes of 0.1 M Tris-HCl, pH 8 at 5 ml/min. Unbound proteins were washed with 2 column volumes of the same buffer and collected in 2 ml fractions. The bound proteins were eluted with 5 volumes of 1 M NaCl in equilibration buffer, and 2 ml fractions were collected.

4.1.5 Temperature optima and stability

The temperature optimum was evaluated by incubating the substrate and the buffer at various temperatures (15°C, 25°C, 37°C, 50°C, 65°C, and 80°C) for one hour. Thermal stability of the enzymes was investigated in a similar fashion but with the enzymes incubated instead of a substrate.

University of Fort Hare
Together in Excellence

4.1.6 pH optimum

For pH optimum, the reaction was assayed over a range of pH 4 – 11 (0.1 M sodium acetate, pH 4-5; 0.1 M phosphate buffer 6-7; 0.1 M Tris-HCl, pH 7.5-8.5; 0.1 M glycine-NaOH, pH 9-11).

4.1.7 Effect of metal ions

The effect of metal ions was studied, according to Silva *et al.* (2011). The metal ions KCl, $MgCl_2$, $MnCl_2$ and $FeCl_3$ were incubated with the enzyme sample for 30 mins and were added to the final concentration of 1 mM. Subsequently, 110 µl of Tris-HCl buffer (pH 8) and the substrate were added.

4.1.8 Inhibition studies

The inhibition studies were done using phenylmethylsulphonyl fluoride (PMSF) and soybean trypsin activity. The partially purified enzyme (30 µl) was incubated with the inhibitor (0-2 mM) for 30 mins in a microtiter plate.

4.1.9 Determination of kinetic parameters

The serine protease activity was assayed at different concentrations of BApNA (0-4 mM). The initial velocity used in the Lineweaver-Burke plot was determined from the linear portion of

the slopes of the progress curve plots of absorbance 405 nm versus time. The K_m and V_{max} were determined using a Lineweaver-Burk plot. The k_{cat} was calculated from the total enzyme concentration and V_{max}, using the following equation: $k_{cat} = V_{max}/[E]t$, where [E]t is the concentration of the enzyme used.

4.1.10 SDS-PAGE

The size and the purity of the fractions were determined using SDS-PAGE prepared, according to Laemmli (1979). The gel was 4% and 12% for stacking and resolving, respectively.

University of Fort Hare
Together in Excellence

4.2 Results and Discussion

Earlier, protease genes were identified from various red algae genomes. Their amplification was unsuccessful, possibly due to non-specificity of primers. Here, we aimed to purify and characterise cysteine and serine-like protease enzyme. BApNA was used as a chromogenic substrate for both protease enzyme classes as it was convenient for the assay system used. Although a number of studies have used BApNA in the purification of trypsin-like enzymes, gingipain-R, a cysteine protease, has been shown to have the same specificity (Arginine-X) as trypsin-like enzyme (Chen *et al.*, 1992; Potempa *et al.*, 1998).

The amidase activity on BApNA was screened on crude extract. Although the activity obtained was not much, it was not affected by the addition of cysteine. However, it decreased when PSMF and soybean trypsin inhibitors were added. This implied that a serine trypsin-like protease was present but a cysteine protease was absent or not specific for BApNA. This did not come as a surprise as the latter has only been reported on *Porphyromonas gingivalis*, an oral bacterium, but not on seaweed or any other multicellular organisms (Wang *et al.*, 1999; Tsutsui *et al.*, 1987, 1998).

After attempting to purify the trypsin-like enzyme, the activity was often lost completely. It was assumed that this was due to low expression of the enzyme, together with autolysis. Therefore, purification was optimised (refer to 4.2.4.1). Previously, trypsin has been reported to be stable in the presence of calcium chloride, and its optimum activation/activity was achieved at pH between 7 and 10 (Castillo-Yañez *et al.,* 2005; Granier, 1995; Green and Neurath, 1953; Kunitz and Northrop, 1993; Silver *et al.,* 2011). The purification results are summarised in Table 4.1.

University of Fort Hare
Together in Excellence

Table 4.1: Seaweed trypsin-like enzyme purification summary.

Fraction	Volume (ml)	Total activity (Units)	Total Protein (mg)	Specific activity (Units/mg)	Fold	Yield (%)
Crude lysate	300	2.81	220.37	0.01	1	100
Crude extract	240	3.38	102.21	0.03	1	100
AM1	20	0.3	16.43	0.02	0.55	8.85
AM2	20	0.31	11.97	0.03	0.79	9.20
Dialysis	21	0.69	9.32	0.07	2.24	20.45
IEX	12	0.9	0.95	0.18	31.7	26.63

After grinding the seaweed, trypsin-like enzyme amidase activity was immediately tested. This fraction was called crude lysate (CL). It was found that CL activity (2.81 units) was low compared to the subsequent step, crude extract (3.38 units) (Table 4.1). Cell debris or other dense material, possibly interfered with the assay. It is clear that these were removed by centrifuge. We therefore treated crude extract as a base or initial step.

The crude extract was fractionated through a saline precipitation (ammonium sulphate). Although many proteins (16.43 mg) precipitated at 30% ammonium sulphate (AM1), the highest activity (0.31 units) was obtained at 70% ammonium sulphate (AM2). Klomklao *et al.* (2006) also precipitated trypsin enzyme at 30–70% ammonium sulphate and obtained the highest activity. After desalting through dialysis, the activity doubled, regardless of the few proteins that were lost (Table 4.1). This could be due to the change in the ionic strength of the buffer. Furthermore, about 20.45% of the enzyme was recovered with the purification fold of 2.24, the dialysate was then subjected to ion exchange chromatography (IEX).

After passing through DEAE FF 16/10 column, two peaks were obtained and the large amount of proteins was found in the fall-through peak (Figure 4.1A). However, no activity was obtained in either fall-through or eluted peaks.

University of Fort Hare
Together in Excellence

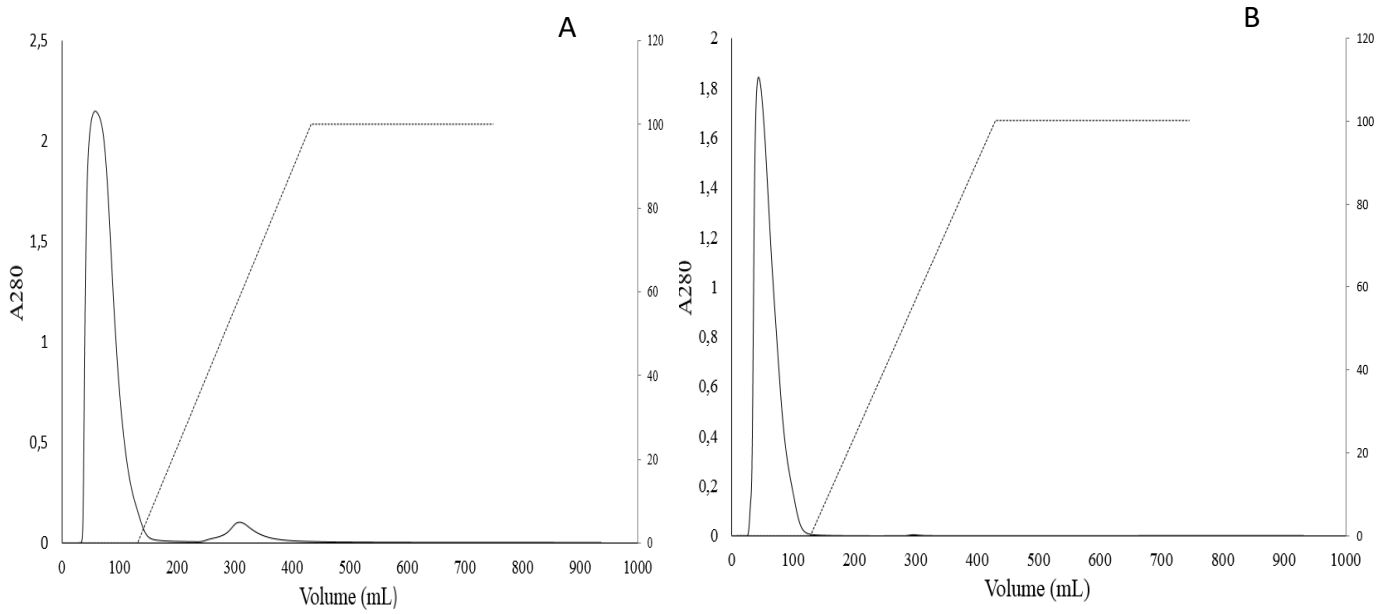


Figure 4.1: Partial purification profile of serine-like protease using ion exchange chromatography.
Figure 4.1A: Purification using DEAE FF 16/10 column. B); Purification using CM FF 16/10 column.

We also tested CM FF 16/10 column but no separation or activity was achieved (Figure 4.1B). We went back to anion chromatography, and changed the column to HiPrep^{TM} Q 16/10. Separation was obtained when 0.5 ml was loaded into the column. Three major peaks were observed with the activity on the unbound peak (Figure 4.2A). The tubes that showed activity were pooled. About 26.63% was recovered with the purification fold of 31.7.

University of Fort Hare
Together in Excellence

Purity and size of the enzyme were determined on the SDS-PAGE. It was noticed that all the fractions had a doublet, two protein bands that moved together. The bands corresponded to 17 and 11 kDa on the standard (Fig 4.2B). It was assumed that the two bands were the result of the domain dissociation. The bioinformatics analysis showed that serine proteases, belonging to the HtrA family, consist of a PDZ domain. Lipinska *et al.* (1989, 1990) reported an unstable serine protease which dissociated on SDS-PAGE into trypsin-like and PDZ domains. Furthermore, trypsin-like enzyme from *Lactobacillus plantarum* was found to consist of three domains and presented four bands on the SDS-PAGE gel. The bands appeared to be 47.4kDa, 38.4kDa. Upon sequencing these bands, it was found that two bands (47.4 and 38.4kDa) were of the active domain and PDZ domain, respectively (Margono *et al.*, 2014).

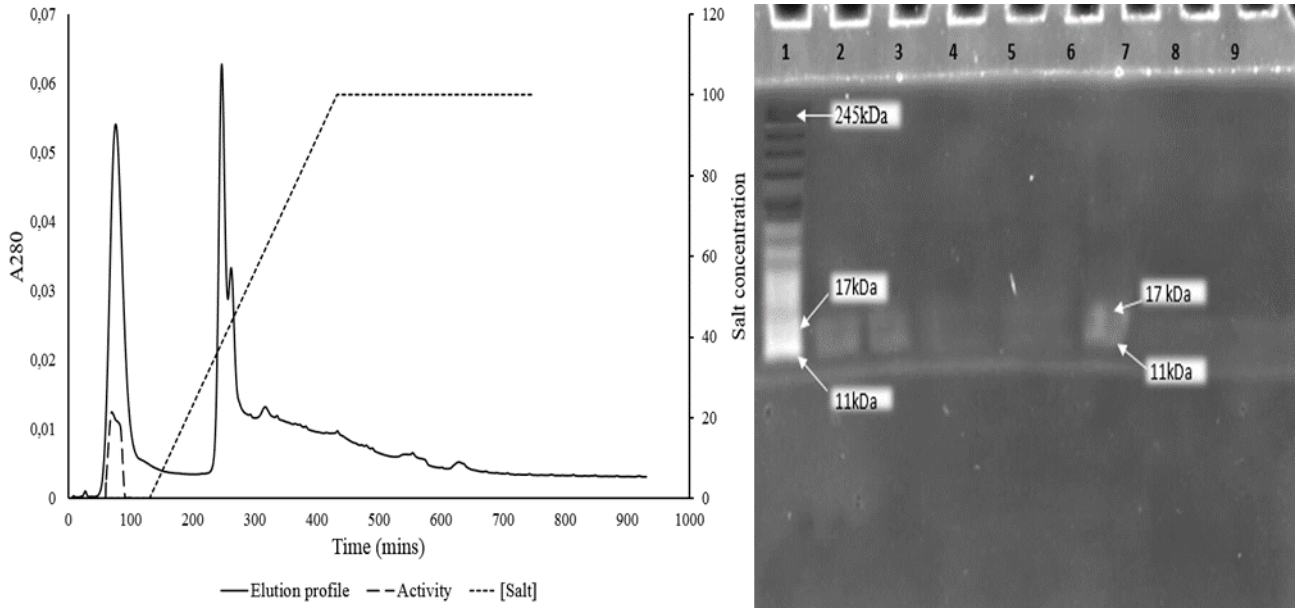


Figure 4.2: Purification of serine-like protease by anion exchange chromatography using HiPrepTM Q 16/10 column.
A 0.5ml was loaded into the column. B) Molecular determination of the partially purified serine-like protease. Well 1- ladder; well 2 – crude lysate; well 3 – crude extract; well 4 – Ammonium sulphate (0-40%); well 5 – ammonium sulphate (30-70%); well 5 – dialysate; well 6 – fall-through peak fraction.

Numerous studies have reported that various factors, such as temperature, pH, metal ions, and inhibitors can influence the protease activity and stability. Having achieved purification, the effect of such factors was studied. Previously, different inhibitors have been used to classify proteases (Klomklao *et el*., 2006). The phenylmethanesulfonyl fluoride (PMSF) inhibitor was used to confirm that the purified protease belonged to the serine group. Usually, inhibition by PMSF can be obtained when the concentration is between 0.1-5 mM (sigma) depending, mostly, on the source. At 0.5 mM PMSF concentration, more than 80% inhibition was obtained (Figure 4.3B). The *Luphiosilurus alexandri* trypsin enzyme was completely inhibited by 1 mM PMSF concentration (dos Santos *et al.*, 2016). Marcuschi *et al.* (2010) obtained about 86% inhibition with 1mM PMSF concentration.

University of Fort Hare
Together in Excellence

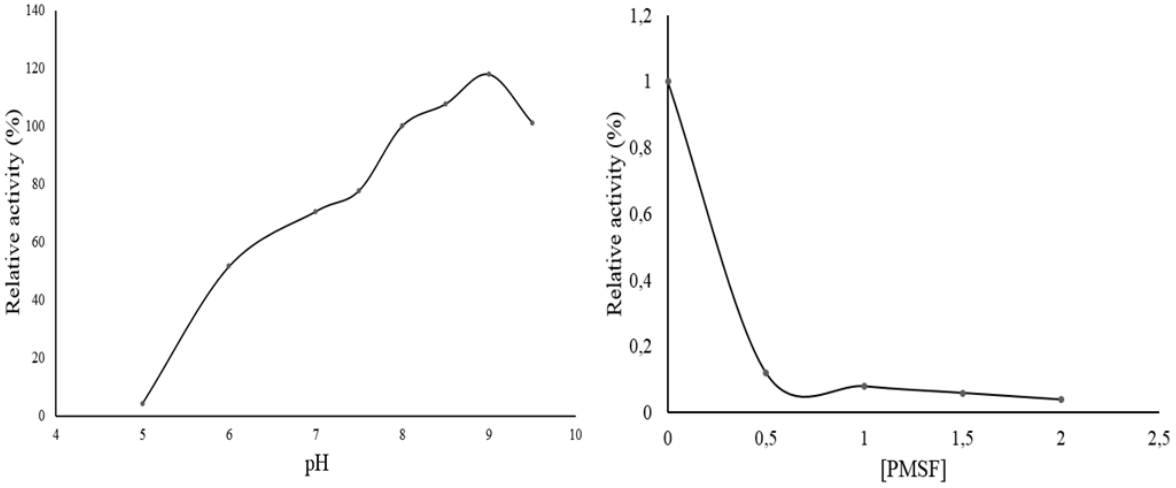


Figure 4.3: The pH effect on enzyme activity and inhibition by PMSF inhibitor.
Figure 4.3A Trypsin-like enzyme activity at different pH. Figure 4.3B Inhibition of a serine protease with different concentrations of PMSF inhibitor. The pH 8 was treated as the control (100%). All the reactions were carried in the presence of calcium chloride.

The partially purified enzyme showed a maximum activity towards BApNA at pH 9 (Figure 4.3A) and 50°C. Similarly, trypsin from *Rastralliger kanagurta* (Khandagale *et al.,* 2015) and *Luphiosilurus alexandri* (dos Santos *et al.*, 2016) exhibited the same optimum pH and temperature when BApNA was used as a substrate. Furthermore, trypsin activity from tambaqui obtained optimum pH and temperature of pH 9 and 50°C when z-FR-MCA (carbobenzoxy-Phe-Arg-7-amido-4-methylcoumarin) was used as a substrate (Marcuschi *et al.*, 2010). Additionally, thermal stability was also evaluated (Figure 4.4). The enzyme was incubated at different temperatures (15–80°C). A gradual decrease, between 4 and 40°C was observed, which was assumed to be due to autolysis. However, more than 70% of the activity within this range was retained. A significant loss of activity (about 60%) was observed above 40°C with some negligible activity at 80°C (Figure 4.4A).
It has been noted from previous reports that the stability of an enzyme varies with species. Kishimura *et al.* (2008) reported a correlation between thermo-stability of trypsin and species habitat. This suggested that variation in stability was due to the natural environmental conditions.

By contrast, the thermal stability observed in red algae (*Gracilaria beckeri*) trypsin-like thermal stability (Figure 4.4B) was similar to that of Frigid Zone trypsins; brown hakeling, walleye pollock, threadfin hakeling, and elkhorn sculpin (Kishimura *et al.*, 2006a; 2007; 2008; 2010). The red algae in this study were sampled from the Agulhas Marine Region (Kenton-on-Sea), which is classified as a temperate region with cool rainy winters (classified, based on Köppen-Geiger, as Cfb) (Conradie, 2012). Anderson *et al.* (2009) reported that the mean annual temperature in Agulhas Marine Region is between 17.2 to 18.2°C.

University of Fort Hare

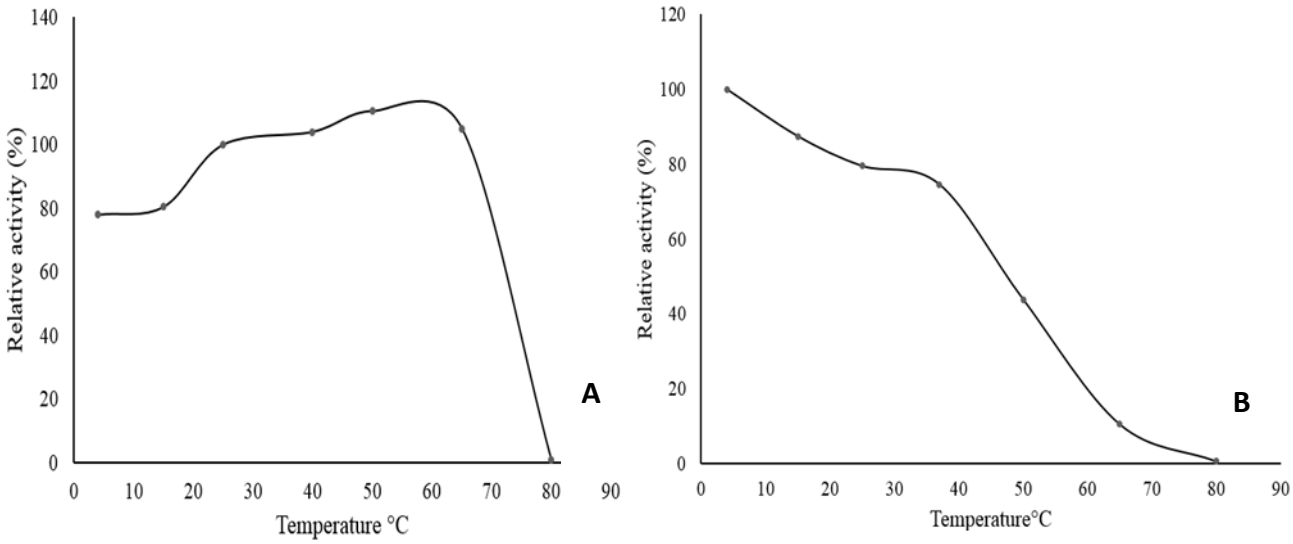


Figure 4.4: The effect of temperature on serine-like enzyme.
Figure 4A shows the effect of temperature on activity of the partially purified enzyme. For comparison in the temperature opimum study, the reaction read at 25°C was treated as 100%. Figure 4.4 B shows the thermo-stability of the partially purified enzyme. The enzyme kept on ice was treated as control (100%)

The effect of metal ions on the enzyme stability was evaluated. The results are summarised in Table 4.2. All the ions were tested in the presence of 1.2 mM of calcium chloride from the buffer. Ferric chloride ($FeCl_3$=32%) possessed an inhibitory effect whereas calcium chloride ($CaCl_2$=156%) appeared to have an inducive effect on the enzyme. Enzyme activity was retained by 77 and 70 when an enzyme was incubated with potassium chloride (KCl) and manganese chloride, respectively.

Table 4.2: The effect of metal ions on the stability of the serine-like protease.

Ions (2mM)	Relative activity (%)
Control (1.2mM $CaCl_2$)	100
$FeCl_3$	32
$CaCl_2$	156
$MgCl_2$	97
$MnCl_2$	70
KCl	77

University of Fort Hare
Together in Excellence

However, these were less intensive since more than 70% of the activity was retained. The magnesium chloride ($MgCl_2$=97) seemed to have no effect on the enzyme. Previous reports have shown that the effect of metal ions varies with species. For instance, trypsin activity from *Bacillus licheniformis* (Gong *et al.*, 2015), Nile tilapia (Bazerra *et al.*, 2005), and spotted goatfish (Souza *et al.*, 2007) were shown to be inhibited by the presence of calcium ions, whereas trypsin activity from red algae (in this study), Indian mackerel (Khandagale *et al.*, 2015), and silver mojarra (Silva *et al.*, 2011) was increased. In *Octopus vulgaris* (Liu *et al.*, 2013), sailfin catfish (Villalba-Villalba *et al.*, 2013) and shrimp (Gates and Travis, 1969) trypsin appeared to be unaffected by the presence of calcium. Similarly, a less intensive inhibitory effect of potassium and manganese was also observed in Black Sea anchovy, sailfin fish, and spotted goatfish (Temiz *et al.*, 2013; Villalba-Villalba *et al.*, 2013; Souza *et al.*, 2007). However, silver mojarra trypsin appeared to be activated by these ions (Silva *et al.,* 2011).

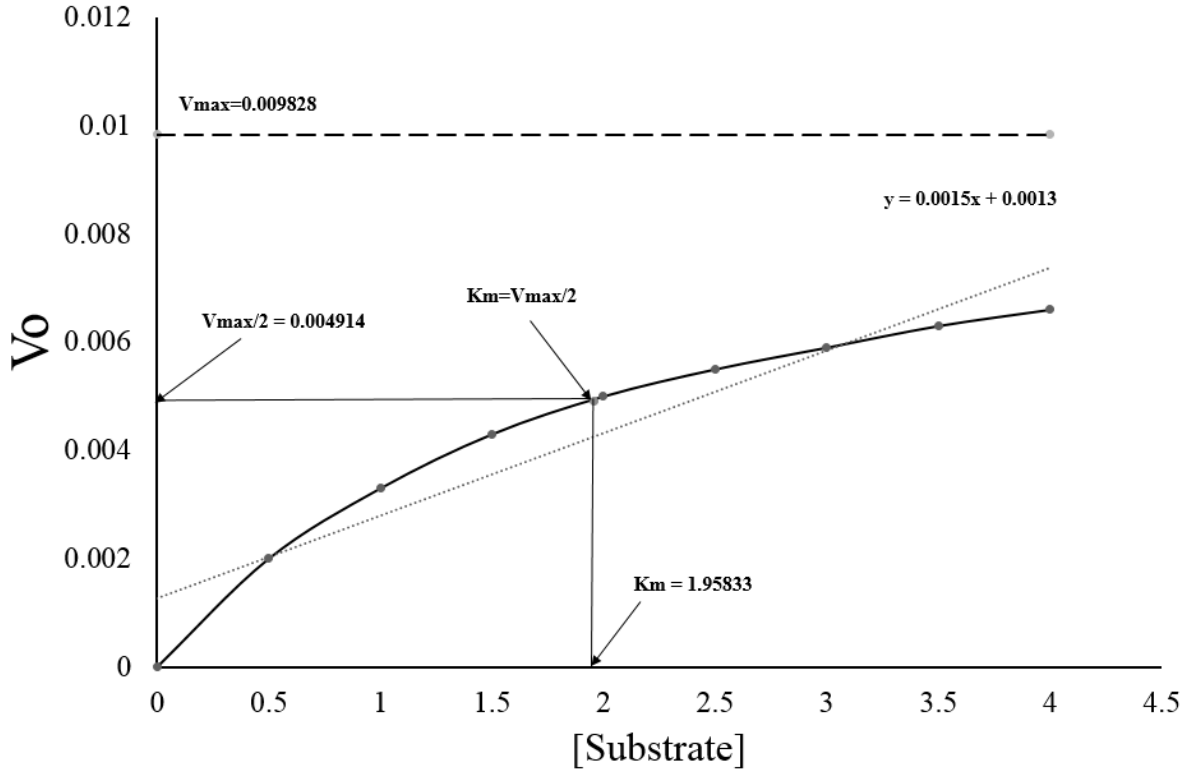
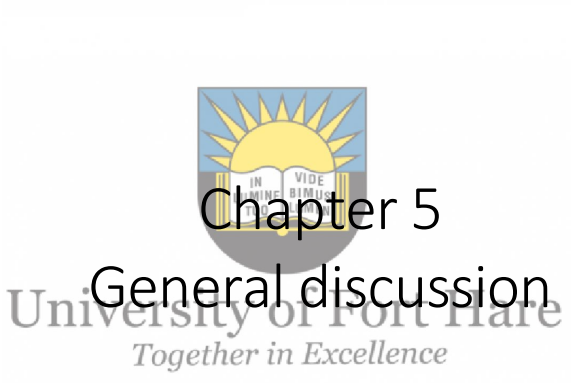


Figure 4.5: Michaelis-Menten plot for serine-like enzyme at different substrate concentration.

The kinetic parameters were evaluated by following amidase activity at different concentrations of BApNA. The Michaelis constant (Km) indicates how much an enzyme binds to its substrate, and the catalytic constant (or turnover number) gives a number of molecules of substrate converted per unit of time (min^{-1} or s^{-1}). The K_m, V_{max}, k_{cat}, and catalytic efficiency for the purified red algae trypsin-like enzyme were determined and found to be 1.95833 µM, 0.009828 µmol. min^{-1}, 0.364 s^{-1}, and 0.186 $s^{-1}mM^{-1}$, respectively (Figure 4.5).

University of Fort Hare
Together in Excellence



Chapter 5
General discussion

5. DISCUSSION

Proteases constitute one of the major groups of industrial enzymes with a 60% share of the global enzyme market, and their application has great potential in various industrial processes. However, the issue of stability for commercial success has been a major hurdle because most proteases are optimally active at physiological conditions. Owing to the high cost of the current enzymes and to environmental issues (in the case of chemicals), there is still a need for new protease enzymes which possess characteristics of interest for industrial application. Screening the organisms that live in unexplored habitats can greatly facilitate the search for new protease enzymes.

Seaweed, among the red algae, is the least explored in terms of available protease information. The bioinformatics tools were employed in search for protease genes. All the sequences identified were from two subclasses of Rhodophyceae, Florideophyceae and Bangiophyceae. The protease sequences were first identified from the *Chondrus crispus* and *Cyanodioschyzon merolae* sequences. *Chondrus crispus* sequences were used as query during *in silico* analysis. All the identified protease sequences were analysed and grouped, on the basis of conserved domains, into five protease classes: aspartic (Asp), cysteine (Cys), metalloprotease (Met), serine (Ser), and threonine (Threo) proteases.

The other two classes, asparagine (Aspar) and glutamic (Glu) proteases did not seem to be present. The glutamic proteases have only been reported from the selected filamentous fungi, such as *Aspergillus niger* and *Scytalidium lignocolum* (O'Donoghue *et al.,* 2008; Srilakshmi *et al.,* 2015). Asparagine protease have mostly been found to be distributed among viruses (N1, N2, N8, N7 and N5) as viral coat proteins, in bacteria as an autotransporter (N6 and N4), and in fungi as intein-containing proteins (N9, N10, and N11) (Rawlings *et al.*, 2011).

University of Fort Hare
Together in Excellence

Although a number of sequences were identified computationally, gene amplification was unsuccessful due to unspecific primers used. The primers were designed on the based on *Chondrus crispus* genome sequence because our red seaweed did not have the available sequence, but belongs to the same subclass, Florideophyceae.

Because amplification failed, we had to isolate and characterise the protease enzyme. We attempted to isolate cysteine and serine protease enzyme because of convenience in the assay system. BApNA was used as a substrate. Although a number of studies have utilised BApNA in the purification of trypsin-like enzymes, gingipain-R, a cysteine protease, has shown to have the same specificity (Arginine-X) as trypsin-like enzyme (Chen *et al.,* 1992; Potempa *et al.,* 1998).

However, only serine protease was positive during screening. This enzyme was screened in the presence of PMSF and soybean trypsin to see if the trypsin-like enzyme was present. Subsequently, purification was performed in a three-step process: ammonium sulphate precipitation, dialysis, and ion exchange chromatography. The enzyme was purified 31.7-fold with the recovery of 26.63%. The SDS-PAGE showed a double with two bands corresponding to 17 kDa and 11 kDa. The bioinformatics analysis showed that serine proteases in red algae belong to the HtrA family and consist of two domains, trypsin-like (dominant) and PDZ. Lipinska *et al.* (1989, 1990) reported that the proteins of this family are produced as an

unstable precursor which dissociates into more stable subunits. Margono *et al.* (2014) sequenced the 47.4kDa and 38.4kDa subunits and found that they are of the same protein.

The enzyme showed optimal activity at 50ºC and pH 9, and retained more than 70% in the presence of K^+, and Mn^{2+}, with Mg^{2+} having almost no effect.

The Michaelis constant, K_m of 1.96 µM, was higher by at least an order for the *Gelidium pristoides* serine protease, compared to the various K_m values (0.033 – 0.43 mM) reported for trypsin from various species, using BApNA as substrate (Ahsan and Watabe, 2001; Asgeirsson and Cekan, 2006; Chen *et al*., 2009; . The k_{cat} value ($0.364s^{-1}$) was lower than the reported values for trypsin, $1.33s^{-1}$ for human trypsin (Chen *et al*., 2009); $1.13s^{-1}$ for bovine trypsin and $3,44s^{-1}$ for the Atlantic cod trypsin (Asgeirsson and Cekan, 2006) and 1,94 – $3,7s^{-1}$ for Japanese anchovy (Ahsan and Watabe, 2001). Further studies using different substrates are therefore needed to establish the substrate preference of the *Gelidium pristoides* serine protease.

In conclusion, it is still believed that red algae have the potential to become a source of many important compounds that could be employed in industry and therapy. Although amplification was unsuccessful, *in silico* analysis did give clues about the function of the protease we worked on, for instance, the serine proteases from the HtrA family on the SDS-PAGE.

The unavailability of the genome sequence has made it very difficult to study some of the protease enzymes further. These difficulties include the purified enzyme for which there is no information regarding its structure, sequence, and the actual mode of action.

University of Fort Hare
Together in Excellence

6. REFERENCES

Abidi F., Chobert J., Haertle T., and Marzouki M.N. (2011). Purification and biochemical characterization of stable alkaline protease Prot-2 from *Botrytis cinerea*. *Process Biochemistry*, 46, p. 2301-2310.

Abraham J., Gea T., and Sáchez, A. (2014). Substitution of chemical dehairing by proteases from solid-state fermentation of hair wastes. *Journal of Cleaner Production,* 74, p. 191-198.

Ahsan, M.N., Watabe, S., J. (2001). Kinetic and structural properties of two isoforms of trypsin isolated from the viscera of Japanese anchovy, Engraulis japonicus. *Protein Chemistry* 20, 49-58

Anbu P. (2013). Characterization of solvent stable extracellular protease from *Bacillus koreensis* (BK-P21A). *International Journal of Biological Macromolecules*, 56, p. 162-168.

Antão C. M. and Malcata F.X. (2005). Plant serine proteases: biochemical, physiological and molecular features. *Plant Physiology Biochemistry*, 43, p. 637– 650.

Anwar A. and Saleemuddin M. (1998). Alkaline protease: A review. *Bioresource Technology,* 64, p. 175-183.

Asgeirsson, B., Cekan, P. (2006). Microscopic rate-constants for substrate binding and acylation in cold-adaptation of trypsin I from Atlantic cod. *FEBS Letters* 580, 4639-4644

Asif-Ullah M., Kim K., and Yu, Y.G. (2006). Purification and characterization of a serine protease from *Cucumis trigonus* Roxburghi.*Phytochemistry,* 67, p. 870-875.

Barrett A. J. (1994). Classification of peptidases. *Methods in Enzymology*, 244, 1-15.

Bekhit A.E.A., Hopkins D.L., Geesink G., Bekhit A.A., and Franks, P. (2013): Exogenous proteases for meat tenderization. *Critical Reviews in Food Science and Nutrition*. DOI:10.1080/10408398.2011.623247

Bouacem K, Bouanane-Darenfed A., Laribi-Habchi H., Elhoul M.B., Hmida-Sayari A, Hacene H., Ollivier B., Fardeau M., Jaouadi B., and Bejar S. (2015). Biochemical characterization of a detergent-stable serine alkaline protease from *Caldicoprobacter guelmensis. International Journal of Biology Macromolecules*, 81, p. 299-307.

Boutet, E., Lieberherr, D., Tognolli, M., Schneider, M., Bansal, P., Bridge, A., Poux, S., Bougueleret, L. and Xenarios, I. (2016). UniProtKB/Swiss-Prot, the Manually Annotated Section of the UniProt KnowledgeBase: How to Use the Entry View. *Plant Bioinformatics*, pp.23-54.

Boyce, S. and Tipton, K. (2001). Enzyme Classification and Nomenclature. *Encyclopedia of Life Sciences*. https://doi.org/10.1038/npg.els.0000710

University of Fort Hare

Buller A. R., and Townsend C.T. (2013). Intrinsic evolutionary constraints on protease structure, enzyme acylation, and the identity of the catalytic triad. *Proceedings of the national academy of sciences*, 110(8), p. 653-661.

Calkins C. R., and Sullivan G. (2007). Adding enzymes to improve beef tenderness. *National Cattlemen's Beef Association*, Web. 16 Mar. 2016. Beef Facts Product Enhancement.

Chaplin M. F, and Bucke C. (1990). *Enzyme Technology*. 1st Ed. Cambridge: Cambridge University Press.

Chen, N., Zou, J., Wang, S., Ye, Y., Huang, Y., Gadda, G., Yang, J.J. (2009). Designing protease sensors for real-time imaging of trypsin activation in pancreatic cancer cells. *Biochemistry* 48, 3519-3526

Cheng Q., Xu, F. Hu N., Liu X., Liu, Z. (2015). A novel Ca^{2+}- dependent alkaline serine-protease (Bvsp) from *Bacillus sp.* With high fibrinolytic activity. *Journal of Molecular Catalysis B: Enzymatic,* 117, p. 69-74.

Chi Z., Chi Z., Zhang T., Liu G., Li J., and Wang X. (2009). Production, characterization and gene cloning of the extracellular enzymes from the marine-derived yeasts and their potential applications". *Biotechnology Advances* 27.3 (2009): p. 236-255.

Choi J., Han S., and Kim H. (2015). Industrial applications of enzyme biocatalysis: Current status and future aspects. *Biotechnology Advances: doi.org/10.1016/j.biotechadv.2015.02.014*

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

Craik C.S., Page M.J., and Madison E.L. (2011). Proteases as therapeutics. *Biochemistry Journal*, 435, p. 1-16.

Database Resources of the National Center for Biotechnology Information. (2016). *Nucleic Acids Research,* 45(D1), pp. D12-D17.

Dawes, C. J. (1997). *Marine Botany*. 1st Ed. New York: John Wiley.

de Castro R.J.S., Ohara A., Nishide T.G., Albernaz J.R.M., Soares M.H., and Sato H.H. (2015). A new approach for proteases production by *Aspergillus niger* based on the kinetic and thermodynamic parameters of the enzymes obtained. *Biocatalysis and Agricultural Biotechnology*, 4, p. 199-207.

Devi B.G. and HemaLatha K.P.J. (2014). Isolation, partial purification and characterization of alkaline serine protease from seeds of *Cucumis Melo VarAgrestis*. *International Journal of Research in Engineering and Technology*, 03, p. 88-96.

Di Cera E. (2009). Serine proteases. *IUBMB Life*, 61(5), pp. 510-515

University of Fort Hare
Together in Excellence

Ekici Ö. D., Paetzel M. And Dalbey R. E. (2008). Unconventional serine proteases: variations on the catalytic ser/his/asp triad configuration. *Protein Science* 17(12): p. 2023-2037.

Elhoul M.B., Jaouadi N.Z., Rekik H., Bejar W., Touioui S.B., Hmidi M., Badis A., Bejar S., and Jaouadi B. (2015). A novel detergent-stable solvent-tolerant serine thiol alkaline protease from *Streptomyces koyangensis* TN650. *International Journal of Biological Macromolecules,* 79, p. 871-882.

Ellaiah P., Srinivasulu B., and Adinarayana K. (2002). A review on microbial proteases. *Journal of Scientific and Industrial Research,* 61, p.690-704.

El-Shora, H., Abou-El-Wafa, G. and Kadhim, A. (2016). Purification and Biochemical Characteristic of Protease from the Red Seaweed Petrocladia capillacea. *International Journal of Current Microbiology and Applied Sciences*, 5(2), pp.297-308.

Feijoo-Siota L., and Villa T.G. (2011).Native and biotechnologically engineered plant proteases with industrial applications. *Food Bioprocess Technology, 4,* pp. 1066-1088.

Fujinaga M., Cherney M. M., Oyama H., Oda K., and James M.N.G. (2004). The molecular structure and catalytic mechanism of a novel carboxyl peptidase from Scytalidium lignicolum. *Proceedings of the National Academy of Sciences* 101 (10), p. 3364-3369.

Gohel S.D., and Singh S.P. (2012). Cloning and expression of alkaline protease genes from two salt-tolerant alkaliphilic actinomycetes in *E.coli. International Journal of Biological Macromolecules*, 50, p. 664-671.

University of Fort Hare
Together in Excellence

González-Rábade N., Badillo-Corona J.A., Aranda-Barradas J.S., and Oliver-Salvador D.C. (2011). Production of plant proteases in vivo and in vitro-A review. *Biotechnology Advances,* 29, p. 983-996.

Gupta A., Roy I., Khare S.K., and Gupta M.N. (2005). Purification and characterization of solvent-stable protease from *Pseudomonas aeruginosa* PseA. *Journal of chromatography A,* 1069(2), p. 155-161.

Gupta R., Beg Q.K., and Lorenz P. (2002). Bacterial alkaline proteases: molecular approaches and industrial applications. *Applied Microbiology and Biotechnology,* 59(1), p. 15-32.

Gupta S. and Abu-Ghannam N. (2011). Recent developments in the application of seaweeds or seaweed extracts as a means for enhancing the safety and quality attributes of foods. *Innovative Food Science and Emerging Technology*, 12, p. 600-609.

Hu J, Yang B, Lin X, Zhou XF, Yang XW, Liu Y. (2012).Bioactive metabolites from seaweeds. Handbook of Marine Macroalgae: *Biotechnology and Applied Phycology*, p. 262–84.

Iketani A., Namura M., Suzuki Y., Awai K., and Shioi Y. (2013). A novel serine protease with caspase- and legumain-like activities from edible basidiomycete Flammulina velutipes. *Fungal Biology*, 117, p. 173-181.

Inácio F.D., Ferreira R.O., Vaz de Araujo C.A., Brugnari T., Castoldi R., Peralta R.M., de Souza C.G.M. (2015). Proteases of wood rot fungi with emphasis on the genus *Pleurotus. BioMed Research International*. doi:10.1155/2015/290161.

Jegannathan, K.R., and Nielsen, P.H. (2013). Environmental assessment of enzyme use in industrial production- a literature review. *Journal of Cleaner Production*, 42, p. 228-240.

Jensen K., Oestergaard P. R., Wilting R., and Lassen S. F. (2010). Identification and characterization of a bacterial glutamic peptidase. *BMC Biochemistry,*11(47), p.1-12.

Jiang B., Sirega U., Willeford K.O., Luthe D.S., and Williams P. (1995). Association of a 33-kilodalton cysteine protease found in corn callus with the inhibition of fall armyworm larval growth. *Plant Physiology*, 108, p. 1631-1640.

Joshi S. and Satyanarayana T. (2013). Characteristics and applications of a recombinant alkaline serine protease from a novel bacterium *Bacillus Lehensis*. *Bioresource Technology*, 131, p. 76-85.

Kadokami K., Yoshida N., Mizusaki K., Noda K., and Makisumi S. (1990). Some properties of trypsin-like proteases extracted from the seaweed *Codium fragile* and their purification. *Marine Biology* 107, 513-517.

Kamnerdpetch C., Weiss M., Kasper C., and Scheper T. (2007). An improvement of potato pulp protein hydrolyzation process by the combination of protease enzyme systems. *Enzyme and Microbial Technology,* 40(4), pp. 508-514.

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

University of Fort Hare
Together in Excellence

Kulshreshtha G., Burlot A.S., Marty C., Critchley A., Hafting J., Bedoux G., Bourgougnon N., and Prithiviraj B. (2015). "Enzyme-Assisted Extraction of Bioactive Material from *Chondrus crispus* and *Codium fragile* and its Effect on Herpes simplex Virus (HSV-1). *Marine Drugs,*13, 558-580.

Li Q., Yi L., Marek P., and Iverson B.L. (2013). Commercial proteases: Present and future. *FEBS Letter,* 587, 1155-1163.

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

Maiti A.K., Ahlawat S.S., Sharma D.P., and Khanna, N. (2008). Application of natural tenderizers in meat-A review. *Agricultural Reviews,* 29,(3), pp. 226-230.

Matsubara K, Hori K, Matsuura Y, Miyazawa K. (1999). A fibrinolytic enzyme from a marine green alga, *Codium latum*. *Phytochemistry,* 52, pp. 993-999.

Matsubara K., Hori K, Matsuura Y, Miyazawa K. (2000). Purification and characterization of a fibrinolytic enzyme and identification of fibrinogen clotting enzyme in a marine green alga, *Codium divaricatum*. *Comparative Biochemistry and Physiology,* 125B: pp. 133-139.

Matsubara K., Matsuura Y., Sumi H., Hori K., and Miyazawa K. (2002). A fibrinolytic enzyme from the green alga *Codium latum* activates plasminogen. *Fisheries Science* 68, pp. 455-457.

Matsubara K., Sumi H., Hori K., and Miyazawa K. (1998). Purification and Characterization of two fibronolytic enzymes from a marine green alga, *Codium intricatum*.

Mienda B.S., Yahya A., Galadima I.A., and Shamsir M.S. (2014). An overview of microbial proteases for industrial application. *Research Journal of Pharmaceutical, Biology and Chemical Sciences*, 5(1), pp. 388-396.

Mokashe N., Chaudhari A., and Patil U. (2015). Optimal production and characterization of alkaline protease from newly isolated halotolerant *Jeotgalicoccus sp*. *Biocatalysis and Agricultural Biotechnology,*4, pp. 235-243.

Moo-Young M. (2011). Comprehensive Biotechnology. 2nd Ed. Amsterdam: Elsevier, pp. 555-569.

Mótyán J.A., Tóth F., and Tözser J. (2013). Research applications of proteolytic enzymes in molecular biology. *Biomolecules,* 3, pp. 923-942.

Mumtaj S. (2015). Enzymes concentrations of selected seaweeds from Mandapam coast, Rameswaram, Tamil Nadu, India. *International Journal of Pure and Applied Zoology,* 3 (4), pp.293-296.

University of Fort Hare
Together in Excellence

Murao S., Oda, K., and Matsushita Y. (1972). Isolation and identification of a microorganism which produces none *Streptomyces* pepsin inhibitor and N-diazoacetyl-DL-norleucine methyl ester sensitive acid protease. *Agricultural and biological chemistry,* 37 (6), pp.1421-10973.

Murao S., Ohkuni K., Nagao M., Hirayama K., Fukuhara K., Oda K., Oyama H., and Shin T. (1992). Purification and Characterization of Kumamolysin, a Novel Thermostable Pepstatin-insensitive Carboxyl Proteinase from *Bacillus* novosp. MN-32. *The Journal of Biological Chemistry*, 268 (1), pp. 3349-355.

Najafi, M.F., Deobagkar, D., and Deobagkar, D. (2005). Potential application of protease isolated from *Pseudomonas aeruginosa* PD100. *Electronic Journal of Biotechnology,* 8(2), pp. 197-203.

Oda K. (2011). New families of carboxyl peptidases: serine-carboxyl peptidases and glutamic peptidases. *Journal of Biochemistry,* 151(1), pp. 13-25.

Oda K. and Murao S. (1974). Purification and some enzymatic properties of acid protease A and B of *Scytalidium lignicolumn* ATCC 24568. *Agricultural and Biological Chemical*, 38(12), pp. 2435-2444.

Oda K., Nakazima T., Terashita T., Suzuki K., and Murao S. (1987). Purification and properties of an S-PI (Pepstatin AC)-insensitive carboxyl proteinase from *Xanthomonas sp* bacterium. *Agricultural and Biology Chemistry*, 51(11), pp. 3073-3080.

O'Donoghue, A., Mahon, C., Goetz, D., O'Malley, J., Gallagher, D., Zhou, M., Murray, P., Craik, C. and Tuohy, M. (2008). Inhibition of a Secreted Glutamic Peptidase Prevents Growth of the FungusTalaromyces emersonii. *Journal of Biological Chemistry*, 283(43), pp.29186-29195.

Okoniewska, M., Tanaka, T. and Yada, R. (2000). The pepsin residue glycine-76 contributes to active-site loop flexibility and participates in catalysis. *Biochemical Journal*, 349(1), p.169.

Pal A., Kamthania M.C. and Kumar A. (2014). Bioactive Compounds and Properties of Seaweeds—A Review. *Open Access Library Journal*, 1: pp752.

Patil, S.S. and Rebecca, L.J. (2014). Isolation and characterization of protease from marine algae*. International Journal of Pharmaceutical Sciences Review and Research*, 27: 188-190.

Pérez-Lloréns J. L., Benítez E., Vergara J. J., and Berges J. A. (2003). Characterization of proteolytic enzyme activities in macroalgae. *European Journal of Phycology*, 38, pp. 55-64.

Raimbault A., Zuily-Fodil Y., Soler A., and Cruz de Carvalho M.H. (2013). A novel aspartic acid protease gene from pineapple fruit (Ananas comosus): Cloning, characterization and relation to postharvest chilling stress resistance. *Journal of Plant Physiology*, 170, pp. 1536-1540.

University of Fort Hare
Together in Excellence

Rani K., Rana R., and Datt S. (2012). Review on latest overview of proteases. *International Journal of Current Life Sciences,* 2 (1), pp. 12-18.

Rao M. B., Tanksale A.M., Ghatge M.S., and Deshpande V.V. (1998). A molecular and Biotechnological aspect of microbial proteases. *Microbiology and Molecular Biology Reviews*, 62(3), pp. 597-635.

Rathod M.G., and Pathak A.P. (2014). Wealth from waste: Optimized alkaline protease production from agro-industrial residues by *Bacillus alcalophilus* LW8 and its biotechnological applications. *Journal of Taibah University for Science*, 8, pp. 307-314.

Raval V.H., Pillai S., Rawal C.M., Singh S.P. (2014). Biochemical and structural characterization of a detergent-stable serine alkaline protease from seawater haloalkaliphilic bacteria. *Process Biochemistry,* 49, pp. 955-962.

Rawlings N. D. and Barrett A.J. (1994a). Families of serine peptidases. *Methods in Enzymology*, 244, pp. 19-61.

Rawlings N. D., and Barrett A.J. (1994b). Families of cysteine peptidases. *Methods in Enzymology* 244, 461-486.

Rawlings N. D., and Barrett A.J. (1995). Families of aspartic peptidases, and those of unknown catalytic mechanism. *Proteolytic Enzymes: Aspartic and Metallo Peptidases,* 248, pp. 105-120.

Rawlings N. D., Barret A.J., and Bateman A. (2011). Asparagine peptide lyases: A seventh catalytic type of proteolytic enzymes. *The Journal of Biological Chemistry,* 286 (44), pp.38321-38328.

Rawlings N.D., Barrett A.J. & Finn, R.D. (2016). Twenty years of the MEROPS database of proteolytic enzymes, their substrates and inhibitors. *Nucleic Acids Research*, 44, pp.343-350.

Rawlings N.D., Barrett A.J., Bateman A. (2012). MEROPS: The database of proteolytic enzymes, their substrates and inhibitors. *Nucleic Acids Research*,40, pp.343–350.

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

Sasaki H., Kubota K., Lee W. C., Ohtsuka J., Kojima M., Iwata S., Nakagawa A., Takahashi K., and Tanokura, M. (2012). The crystal structure of an intermediate dimer of aspergilloglutamic peptidase that mimics the enzyme-activation product complex produced upon autoproteolysis. *Journal of Biochemistry,* 152(1), pp. 45-52.

Sawant R., and Nagendran A. (2014). Protease: An enzyme with multiple industrial applications. *World Journal of Pharmacy and Pharmaceutical* Science 3 (6), pp.568-579.

University of Fort Hare
Together in Excellence

Sims A.H., Dunn-Coleman N.S., Robson G.D., and Oliver S.G. (2004). Glutamic protease distribution is limited to filamentous fungus. *FEMS Microbiology Letter,* 239, pp. 95-101.

Singh R., Mittal A., Kumar M., and Mehta P.K. (2016). Microbial Proteases in Commercial Applications*. Journal of Pharmaceutical, Chemical, and Biological Sciences*, 4 (3), pp. 365-374.

Spohner S.C., Müller H., Quitmann H., Czermak P. (2015). Expression of enzymes for the usage in food and feed industry with *Pichia pastoris. Journal of Biotechnology*, 202, pp.118-134.

Sumantha A., Larroche C., and Pandey A. (2006). Microbiology and industrial biotechnology of food-grade protease: A perspective. *Food Technology and Biotechnology,* 44(2), pp. 211-220.

Tajima N., Kawai F., Park S. And Tame J. R. (2010). A novel intein-like autoproteolytic mechanism in autotransporter proteins. *Journal of Molecular Biology,* 402(4), pp. 645-656.

Tavano O.L. (2013). Protein hydrolysis using proteases: An important tool for food biotechnology. *Journal of Molecular Catalysis B: Enzymatic,* 90, 1-11.

Thanigaivel S., Chandrasekaran N., Mukherjee A., and Thomas J. (2015). Investigation of seaweed extracts as a source of treatment against bacterial fish pathogen. *Aquaculture,* 448, pp.82-86.

Toplak A., Wu B., Fusetti F., Quaedflieng P.J.L.M., and Janssen, D.B.(2013). Proteolysin, a novel highly thermostable and cosolvent-compatible protease from the thermophilic bacterium *Coprothermobacter proteolyticus. American Society for Microbiology,* 79(18), pp. 5625-5632.

Truter, I. (2014). Prescription patterns of enzyme-containing products in South Africa over a 2-year period. *South African Journal of Science*, 110 (9), pp.1-5.

Vijayaraghavan P., Lazarus S., and Vincent S.G.P.(2014). De-hairing protease production by an isolated *Bacillus cereus* strain AT under solid-state fermentation using cow dung: Biosynthesis and properties. *Saudi Journal of Biological Sciences,* 21, pp. 27-34.

Wang, Y., Li, X.Y., Xu, E.S. & Dai, J.X. (1998). Spectral profiles on the protease activities in Laminaria. *Undaria and Porphyra. Acta Bot. Sin*.40: 939-946.

Yariswamy M., Shivaprasad H.V., Joshi V., Nanjaraj U.A.N., Nataraju, A., and Vishwanath, B.S. (2013). Topical application of serine proteases from *Wrightia tinctoria* R.Br. (Apocynaceae) latex augments healing of experimentally induced excision wound in mice. *Journal of Ethnopharmacology*, 149, 377-383.

University of Fort Hare
Together in Excellence