ISOLATION, GENE EXPRESSION AND STRUCTURAL ANALYSIS OF FRUIT BROMELAIN FROM *Ananas comosus* CULTIVAR MD 2

PANG WEI CHENG

MASTER OF SCIENCE

UNIVERSITI MALAYSIA PAHANG

UNIVERSITI MALAYSIA PAHANG

DECLARATION OF THESIS AND COPYRIGHT

Author's Full Name : <u>PANG WEI CHENG</u>

Date of Birth : <u>21 APRIL 1993</u>

Title : <u>ISOLATION, GENE EXPRESSION AND STRUCTURAL</u>
<u>ANALYSIS OF FRUIT BROMELAIN FROM *Ananas comosus*</u>
<u>CULTIVAR MD 2</u>

Academic Session : <u>Semester 1 2019/2020</u>

I declare that this thesis is classified as:

☐ CONFIDENTIAL (Contains confidential information under the Official Secret Act 1997)*

☐ RESTRICTED (Contains restricted information as specified by the organization where research was done)*

☑ OPEN ACCESS I agree that my thesis to be published as online open access (Full Text)

I acknowledge that Universiti Malaysia Pahang reserves the following rights:

1. The Thesis is the Property of Universiti Malaysia Pahang
2. The Library of Universiti Malaysia Pahang has the right to make copies of the thesis for the purpose of research only.
3. The Library has the right to make copies of the thesis for academic exchange.

Certified by:

_____
(Student's Signature)

930421-01-6387
_____
New IC/Passport Number
Date: 6 January 2020

_____
(Supervisor's Signature)

Aizi Nor Mazila binti Ramli
_____
Name of Supervisor
Date: 6 January 2020

NOTE : * If the thesis is CONFIDENTIAL or RESTRICTED, please attach a thesis declaration letter.

SUPERVISOR'S DECLARATION

I hereby declare that I have checked this thesis and in my opinion, this thesis is adequate in terms of scope and quality for the award of the degree of Master of Science.

_____

(Supervisor's Signature)

Full Name   : AIZI NOR MAZILA BINTI RAMLI

Position   : SENIOR LECTURER

Date   : 6 JANUARY 2020

STUDENT'S DECLARATION

I hereby declare that the work in this thesis is based on my original work except for quotations and citations which have been duly acknowledged. I also declare that it has not been previously or concurrently submitted for any other degree at Universiti Malaysia Pahang or any other institutions.

_____

(Student's Signature)

Full Name   : PANG WEI CHENG

ID Number   : MKT17006

Date   : 6 JANUARY 2020

ISOLATION, GENE EXPRESSION AND STRUCTURAL ANALYSIS OF FRUIT
BROMELAIN FROM *Ananas comosus* CULTIVAR MD 2

PANG WEI CHENG

Thesis submitted in fulfillment of the requirements

for the award of the degree of

Master of Science

Faculty of Industrial Sciences & Technology

UNIVERSITI MALAYSIA PAHANG

JANUARY 2020

ACKNOWLEDGEMENTS

ABSTRAK

Ananas comosus atau nanas adalah salah satu buah tropika yang popular di dunia. Terdapat pelbagai kultivar nanas dengan ciri-ciri yang berbeza. *A. comosus* kultivar MD 2 adalah hibrida nanas yang paling berjaya dengan warna emasnya yang menarik, rasa manis dan kualiti terunggul. Nanas kaya dengan enzim proteolitik seperti bromelain buah. Dalam kajian ini, bromelain buah telah dicirikan dan dibandingkan di antara *A. comosus* kultivar MD 2 yang ranum dan muda. Selain itu, transkrip bromelain buah juga diperolehi melalui teknik "conventional Polymerase Chain Reaction" dan pengklonan. Maklumat jujukan gen bromelain buah telah digunakan untuk pemodelan 3D struktur dan diikuti dengan analisis termostabiliti. Untuk mengaji perbezaan tahap ekspresi gen dan aktiviti proteolitik buah bromelain antara *A. comosus* kultivar MD 2 yang ranum dan muda, "quantitative Polymerase Chain Reaction" dan ujian enzim kasein telah dilaksanakan. Maklumat jujukan gen bromelain buah telah dianalisis menggunakan beberapa program bioinformatik seperti "BioEdit", "BLAST", "InterProScan" dan "ProtParam". 3D struktur bromelain buah telah dimodelkan menggunakan "MODELLER" dan simulasi molekul dinamik. Selanjutnya, termostabiliti bromelain buah juga telah dinilai menggunakan ujian enzim kasein dan simulasi molekul dinamik. Hasil kajian menunjukkan bahawa pengekspresan bromelain buah telah diturunkan sebanyak 90 % dalam *A. comosus* kultivar MD 2 yang ranum. Ia juga didapati bahawa *A. comosus* kultivar MD 2 yang muda mempunyai aktiviti yang lebih tinggi (1.91 ± 0.08 U/mL) berbanding dengan *A. comosus* kultivar MD 2 yang ranum (1.13 ± 0.09 U/mL). Di samping itu, tiga jujukan bromelain buah iaitu OAY62650.1, OAY68270.1 dan OAY85858.1 telah dipilih untuk diklon. Keputusan penjujukan menunjukkan perubahan nukleotida. Selain itu, "Verify 3D", "ERRAT" dan "PROCHECK" menunjukkan model bromelain buah yang dihasilkan mempunyai kualiti yang tinggi. Analisis struktur mengungkapkan interaksi antara "pro-domain" dan "catalytic domain". "Binding subsites" bromelain buah termasuk His160, Trp183, Glu19 dan Asn159 juga telah dikenal pasti. Tambahan pula, profil haba bromelain buah telah dibina. Suhu optimum bromelain buah telah dikenal pasti pada 60 °C. Apabila suhu meningkat, kestabilan bromelain buah telah terganggu. Hasil simulasi menunjukkan bahawa pemusnahan struktur bromelain buah mungkin disebabkan gelung sekunder yang fleksibel serta asid amino tanpa cas dan hidrofilik. Maklumat yang diperoleh daripada kajian ini dijangka dapat menambahbaik penggunaan bromelain buah pada masa akan datang.

ABSTRACT

Ananas comosus, or commonly known as pineapple, is one of the most popular tropical fruits in the world. There are various pineapples cultivars with distinct characteristics. *A. comosus* cultivar MD 2 is the most successful pineapple hybrid with its attractive gold colour, super sweet taste and superior quality. Pineapple is rich in proteolytic enzymes such as fruit bromelain. In this study, fruit bromelain was characterised and compared between ripe and unripe *A. comosus* cultivar MD 2. Besides that, fruit bromelain transcripts were also isolated by conventional PCR and cloning. The extracted fruit bromelain sequence information was used in tertiary structure modelling followed by thermostability analysis. To investigate the gene expression level and proteolytic activity of fruit bromelain of *A. comosus* cultivar MD 2, quantitative Polymerase Chain Reaction (PCR) and casein enzymatic assay were conducted. Meanwhile, the sequence information of fruit bromelain was analysed using several bioinformatic tools including BioEdit, BLAST, InterProScan and ProtParam. The tertiary structure of the selected fruit bromelain sequences was modelled using MODELLER and refined via molecular dynamics simulation. Subsequently, the thermostability of fruit bromelain was evaluated using casein enzymatic assay and molecular dynamics simulation. The result revealed that fruit bromelain was down-regulated by 90 % in ripe *A. comosus* cultivar MD 2. It was also found that unripe *A. comosus* cultivar MD 2 has higher fruit bromelain activity (1.91 ± 0.08 U/mL) than ripe *A. comosus* cultivar MD 2 (1.13 ± 0.09 U/mL). On the other hand, three fruit bromelain sequences namely OAY62650.1, OAY68270.1 and OAY85858.1 were selected as sequences of interest to be isolated and cloned. The sequencing result demonstrated several nucleotides alteration in the isolated sequences. Moreover, Verify 3D, ERRAT and PROCHECK showed that the generated fruit bromelain models have high stereochemical quality. Structural analysis revealed interactions between the pro-domain and catalytic domain. Binding subsites of fruit bromelain His160, S2 Trp183, Glu19 and Asn159 were also identified. The His160, Trp183 and Glu19 were found conserved between fruit bromelain and papain. In addition, the thermal profile of MD 2 fruit bromelain was constructed. The optimum temperature of fruit bromelain was identified at 60 °C which is within the expected range. As the temperature raises, the stability of fruit bromelain was disrupted. The simulation result showed that fruit bromelain structures destruction may be due to flexible secondary loops as well as the presence of non-charged and hydrophilic amino acids residues. The information obtained from this study is expected to improve the applications of fruit bromelain in the future.

TABLE OF CONTENT

LIST OF TABLES

LIST OF FIGURES

xi

LIST OF SYMBOLS

%	Percent
α	Alpha
Å	Angstrom
β	Beta
°C	Degree celcius
R^2	Coefficient of determination
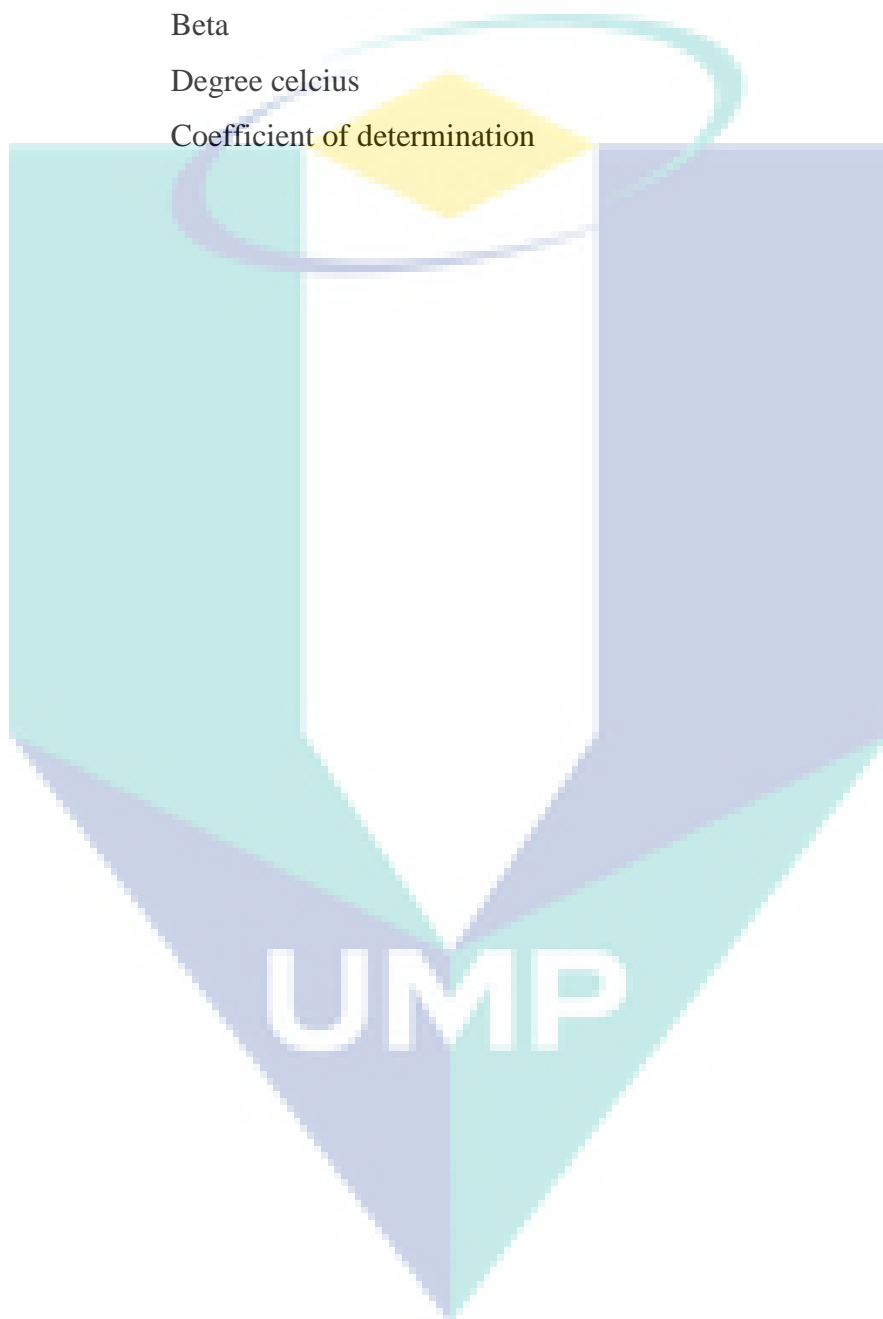
LIST OF ABBREVIATIONS

3D	Three-dimensional
A	Absorbance
Ala	Alanine
Arg	Arginine
Asp	Aspartate
Asn	Asparagine
BLAST	Basic Local Alignment Tool
bp	Base pair
cDNA	Complementary deoxyribonucleic acid
C_q	Quantification cycle
Cys	Cysteine
dNTP	Deoxynucleotide
DOPE	Discrete optimized protein energy
EC	Enzyme Commission
ETP	Economic Transformation Programme
FAO	Food and Agriculture Organization of the United Nations
F-C	Folin Ciocalteu
fs	Femtosecond
g	gram
GC	Guanine cytosine
Gly	Glycine
Gln	Glutamine
Glu	Glutamate
GROMACS	GROningen MAchine for Chemical Simulations
His	Histidine
I29	Cathepsin pro-peptide inhibitor
Ile	Isoleucine
LB	Luria Bertani
Leu	Leucine
kDa	Kilodalton
kg	Kilogram

LINCS	LINear Constraint Solver
Lys	Lysine
Met	Methionine
min	Minute
miRNA	Micro Ribonucleic Acid
mg	Milligram
mL	Millilitre
MPIB	Malaysian Pineapple Industrial Board
mRNA	Messsenger Ribonucleic Acid
N. A	Not available
NCBI	National Center for Biotechnology Information
ng	Nanogram
NKEA	National Key Economic Area
nm	Nanometer
NMR	Nuclear Magnetic Resonance
NPT	Constant number of particles, volume and temperature
NVT	Constant number of particles, pressure and temperature
ns	Nanosecond
PCR	Polymerase Chain Reaction
PDB	Protein Database Bank
Phe	Phenylalanine
pI	Isoelectric point
PLCE	Papain-like cysteine endopeptidase
PME	Particle mesh Ewald method
pNA	p-nitroalanine
Pro	Proline
ps	Picosecond
qPCR	Quantitative polymerase chain reaction
Rg	Radius of gyration
RMSD	root-mean-square deviation
RMSF	root-mean-square fluctuation
RNA	Ribonucleic acid
rpm	Revolutions per minute

rRNA	Ribosomal ribonucleic acid
RT-PCR	Reverse transcription polymerase chain reaction
SASA	Solvent accessible surface area
SAVES	Structure Analysis and Verification Server
SD	Standard deviation
SE	Standard error
Sec	Second
Ser	Serine
snRNA	Small nuclear ribonucleic acid
SPDBV	Swiss PDB Viewer
TA	Thymine adenine
Thr	Threonine
Tm	Melting temperature
Trp	Tryptophan
tsRNA	tRNA-derived small ribonucleic acid
Tyr	Tyrosine
μL	Microliter
μM	Micrometer
μmol	Micromoles
Val	Valine

LIST OF APPENDICES

CHAPTER 1

INTRODUCTION

1.1 Background of study

Ananas comosus or more commonly known as pineapple is a tropical fruit with a pinecone shape. It is a non-climacteric, herbaceous and monocot perennial plant endemic in coastal and tropical regions (Asim et al., 2015; Moyle, Fairbairn, Ripi, Crowe, & Botella, 2005; Wali, 2019). With a fascinating taste combination of sweet and sour, it is one of the most popular tropical fruits around the world. Currently, nine major pineapple cultivars are planted in Malaysia including Gandul, Josapine, N36, Masapine, Moris, Moris Gajah, MD2, Sarawak and Yankee. Among these cultivars, MD2 has the best quality in terms of its sweetness as well as consistency and uniformity in size and ripeness (Thalip, Tong, & Ng, 2015). Besides that, it is also suitable for cold storage and is less likely to be affected by blackheart disease (Raimbault, Zuily-Fodil, Soler, Mora, & de Carvalho, 2013). Thus, *A. comosus* cultivar MD 2 has the largest commercial value compare to the other cultivars available in the market.

The pineapple fruit serves as a rich source of proteolytic enzymes including bromelain, ananain and comosain (Larocca, Rossano, Santamaria, & Riccio, 2010). Among them, bromelain has received the most attention. Bromelain was first separated from fruit via fermentation by a Venezuelan chemist, Vicente Marcano, in 1891 (Srujana & Narayana, 2017). It was reported that the commercial value of bromelain can cost around 2400 USD/kg (Ketnawa, Chaiwut, & Rawdkuen, 2012). This proteolytic enzyme is highly recognised for its industrial and therapeutic applications due to its high catalytic efficiency, simple extraction methods and higher yield (Han et al., 2018). In the industry, it is used in meat tenderisation, as additives in detergent, leather processing, and in

1

baking. On the other hand, its proteolytic activity also plays an important role in several therapeutic applications such as anti-inflammatory, anti-cancer and anti-biotics.

Despite proteases being used routinely in our daily lives, the information on their expression rate, function and regulation still remain to be investigated. For instance, proteases are expressed in multiple highly homologous isoforms (Butts et al., 2016). The slight difference between these isoforms may carry different physiological properties. In the MD 2 pineapple genome sequences, up to 17 fruit bromelain sequences are found (Redwan, Saidin, & Kumar, 2016). Hence, a proper analysis of these fruit bromelain isoforms is required to understand the existence of fruit bromelain in nature. Besides that, it is believed that proteases are responsible for plant defence. Owing to the advancement in technology, the use of advanced molecular biology techniques such as quantitative PCR (qPCR) can provide information on the expression of gene of interest (Taylor, Wakem, Dijkman, Alsarraj, & Nguyen, 2010). An investigation on the differences of fruit bromelain expression as the pineapple fruit ripens allows speculation of its possible role during fruit development.

It is now commonly accepted that protein sequences from the same family share certain sequence features and adopt similar folding and structure. The papain-like family is constituted by a large group of plant cysteine enzymes including papain, ananain, caricain and ficain (Ramli, Manas, Hamid, Hamid, & Illias, 2018). To date, the crystal structure of fruit bromelain is not available. As one of the papain-like family members, fruit bromelain is expected to contain similar features and its structure should closely resemble other papain-like enzymes. Thus, the predicted structure of fruit bromelain can be built using a highly homologous papain-like protein as a template via comparative modelling. In addition, the protein structure is flexible, and the structural changes are always induced by variations in the surrounding environment. Therefore, the conformational changes of fruit bromelain under the influence of temperature is studied using molecular dynamics simulation.

1.2   Problem statement

Even though biochemical properties of fruit bromelain have been investigated intensively, information of fruit bromelain gene expression and proteolytic activity between ripe and unripe pineapple fruits is still limited. In order to have a broader exploration on bromelain, investigation on fruit bromelain expression during pineapple fruit development should gain attention since gene expression is regulated during the fruit ripening process (Janssen et al., 2008). Besides that, plant cysteine proteases are usually encoded in multiple highly homologous isoforms. It was found that several isoforms of fruit bromelain is present in *A. comosus* cultivar MD 2 genome (Redwan et al., 2016). The slight differences in amino acids may affect the function and structure to these fruit bromelain sequences (Betts & Russel, 2003). In addition, the structural information of fruit bromelain is still limited because its three-dimensional (3D) structure is still not available. Nevertheless, the predicted structure of fruit bromelain can be modelled using the computational method.

1.3   Objective of study

1. To determine fruit bromelain gene expression between unripe and ripe *A. comosus* cultivar MD 2.

2. To analyse the fruit bromelain sequences from *A. comosus* cultivar MD 2.

3. To model the tertiary structure of fruit bromelain.

4. To investigate the structural stability at different temperatures.

1.4   Scope of study

This study was initiated by analysed the expression of fruit bromelain between unripe and ripe *A. comosus* cultivar MD 2 via quantitative PCR (qPCR) and enzymatic assay. The fruit bromelain transcripts OAY62650.1, OAY68270.1 and OAY 85858.1 were isolated via Reverse Transcription Polymerase Chain Reaction (RT-PCR) and conventional PCR. Domain and physicochemical properties analysis of the isolated fruit bromelain sequences were conducted via bioinformatic tools using BioEdit, BLAST, InterProScan, ProtParam and SignalP. This was followed by prediction of the structure

of fruit bromelain by using MODELLER and structure refinement via GROMACS. The predicted fruit bromelain structures were assessed via ERRAT, PROCHECK and Verify3D. The structural information of the fruit bromelain structures was interpreted. Lastly, elucidation of structural stability of fruit bromelain at different temperatures via molecular dynamics simulation and proteolytic assay.

1.5   Significance of study

The present study focusses on expression, sequence and structural study of fruit bromelain. The expression information is necessary to elucidate the biological significance of fruit bromelain in pineapple fruit Besides that, the characteristics of these fruit bromelain isoforms can be investigated by interpreting their amino acid sequences using different bioinformatic tools without performing a laborious experiment. The structural analysis allows the investigation of binding residues and interactions exist within the fruit bromelain structure. Moreover, the predicted fruit bromelain structure can also be simulated to illustrate its conformational changes at an elevated temperature. This provides explicit details of the fruit bromelain structure which is needed for better control of enzyme usage.

CHAPTER 2

LITERATURE REVIEW

2.1   *Ananas comosus*

A. comosus belongs to the Bromeliaceae family and Bromeliodeae subfamily. It was discovered by Christopher Columbus on 4th November 1493 in Guadeloupe Island (Botella & Smith, 2008). It is called 'nanas' or 'ananas' which means 'excellent fruit' in the Guarani language by the people of South America (Joy & Anjana, 2015). The pineapple fruit has an unique sweet-sour taste and is rich in vitamins and minerals such as vitamin C, calcium, potassium as well as fibre (Hossain, 2016). The pineapple was spread worldwide during the naval exploration by the Portuguese and Spanish in the 16th century (Dawson, 2016). Due to its rareness in the early age, it was the symbol of wealth in America (Asim et al., 2015). The pineapple was introduced in Malaysia in 1921 and expanded rapidly at Gambut, Johor (MPIB, 2019). Currently, the development and marketing of the pineapple industry in Malaysia is regulated by the Malaysian Pineapple Industrial Board (MPIB); a statutory body established in 1957.

There are around 30 pineapple cultivars of *A. comosus* and these cultivars are further classified into four classes, namely Red Spanish, Pernambuco, Queen, as well as Smooth Cayenne (Wali, 2019). Among these cultivars, Smooth Cayenne or the Hawaiian pineapple dominated 70 % of the world market (Moretti-Almeida, 2018). This group is characterised by its cylindrical shape, small amount of spines, good organoleptic qualities and high yields (Dawson, 2016). They are named Kew, Sarawak, Esmeralda, Claire, Typhoon and Saint Michel in different regions (Chan, d'Eeckenbrugge, & Sanewski, 2003). The Red Spanish is also known as Network Spanish; it is a plant of medium size, vigorous, with dark green leaves, and has small and short spines. Several cultivars namely

Espanola roja, Singapore Spanish, Nanas Merah and Masmerah are grouped in this category (Joy & Anjana, 2015). Furthermore, Queen is mostly cultivated in the southern hemisphere such as Australia and South Africa. It is small and vigorous, with silvery leaves and with the occurrence of dense spines. In addition, Pernambuco, or Perola, has a soft, white and juicy flesh and is prevalent in Latin America.

The pineapple plant can grow up to between 0.75 and 1.50 m and 0.90-1.20 m wide (Upadhyay, Lama, & Tawata, 2010). The axis of the pineapple plant is a stem which is surrounded by a sheath of thick and lanceolate leaves (Dawson, 2016). According to Dhukani (2013), the pineapple stem will elongate near to the apex during blooming time. After that, it is followed by the development of the crown and fruit. It usually takes four months for the pineapple plant to fruit after flower induction (Rosnah, Daud, Takrif, & Hassan, 2009). It is said that the pineapple fruit which ripens during the summer will have the best quality compared to those that mature during the winter season by having a stronger fragrance (Liu & Liu, 2017). Interestingly, the hexagonal sections on outer layer of the pineapple fruit are always equal and independent on its size or shape (Asim et al., 2015). In addition, the pineapple is reproduced via propagation using its crowns, slip, suckers or seeds (Hossain, 2016). Figure 2.1 illustrates the morphological structure of a pineapple plant (Leal & d'Eeckenbrugge, 2002).
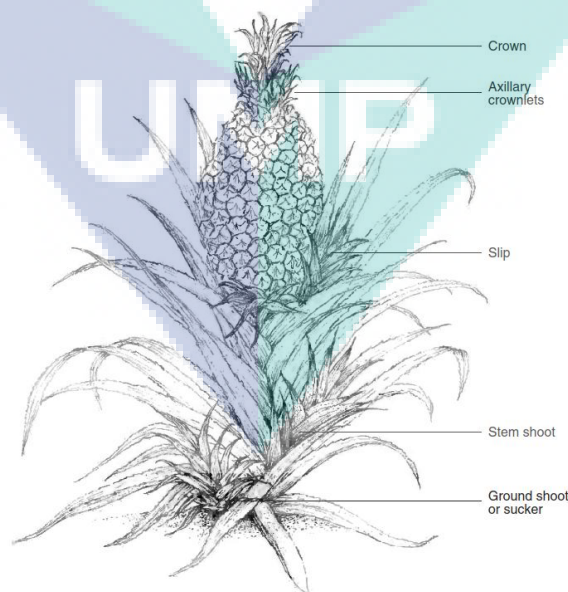
Figure 2.1   Morphological structure of pineapple plant.

Source: Leal & d'Eeckenbrugge (2002)

According to the Food and Agriculture Organization of the United Nations (FAO, 2019), during 2013-2017, Costa Rica has the highest average pineapple production of 2871285 tonnes, followed by Philippines (2566502 tonnes), Brazil (2546569.4 tonnes), Thailand (1924006.2 tonnes), India (1815348 tonnes), Indonesia (1728007 tonnes), Nigeria (1527873.6 tonnes), China (1488367 tonnes), Mexico (850188 tonnes) and Columbia (828237.8 tonnes) (Figure 2.2). Meanwhile, Malaysia is lagging behind in pineapple production. Malaysia has recorded an increment in pineapple production in 2013-2017. In 2013, the total pineapple production in Malaysia was 244353 tonnes which steadily increased to 335725 tonnes and 452021 tonnes in 2014 and 2015, respectively (Figure 2.3). After that, the pineapple production attracted less interest. The pineapple production showed a decreased trend to 391714 tonnes and 340722 tonnes in 2016 and 2017, respectively.
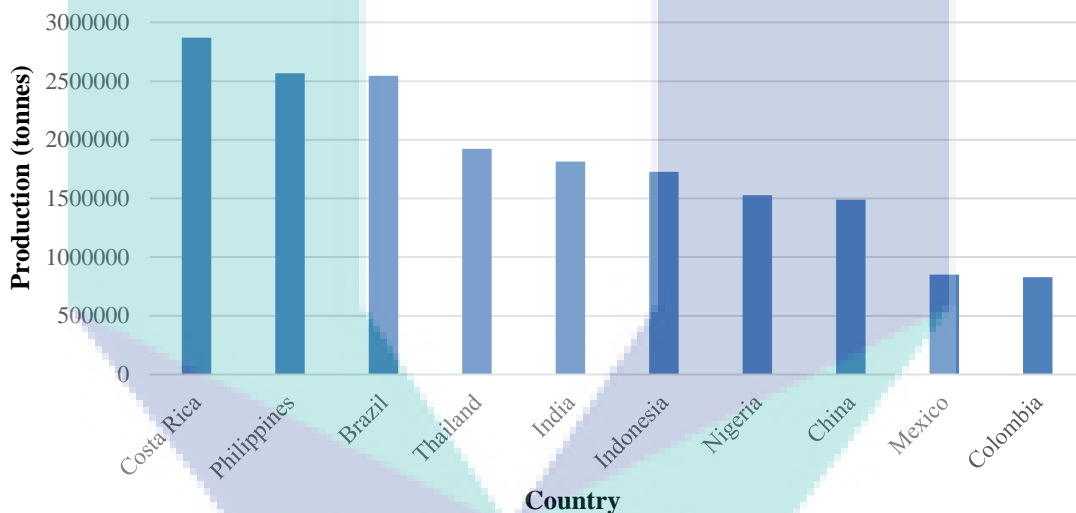
Figure 2.2   Top 10 pineapple producers during 2013-2017.

Source: FAO (2019)
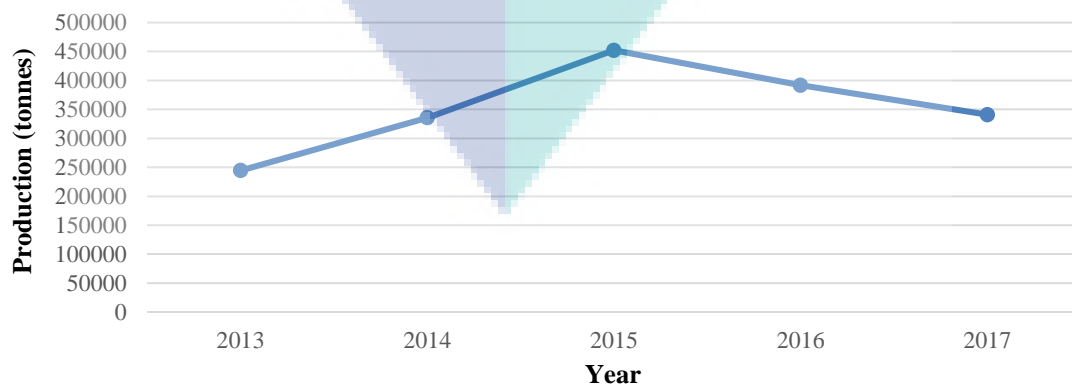
Figure 2.3   Production quantities of pineapples in Malaysia during 2013-2017.

Source: FAO (2019)

In order to penetrate global markets and to ensure national food security, *A. comosus* cultivar MD 2 was selected as a key crop under the National Key Economic Area (NKEA) of the Economic Transformation Programme (ETP) (Thalip et al., 2015). The *A. comosus* cultivar MD 2 (Golden Ripe, Super Sweet, Rompine or Gold) is a hybrid consisting of Smooth Cayenne, Smooth Guatamelan, Queen and Pernambuco developed by Del Monte which was first introduced to the European and US market in 1996 (Hidayat, Chandrika, Izana, Azman, & Alina, 2013; Joy & Anjana, 2015). The *A. comosus* cultivar MD 2 has transformed the pineapple industry with its bright gold colour, higher vitamin C, sweeter taste, lower fibre and acidity, thinner skin, and longer shelf life (Dawson, 2016; Thalip et al., 2015). With the effort of promoting *A. comosus* cultivar MD 2 in Malaysia, the production of pineapple in our country should be increased in the future.

2.2 Proteases

A. comosus contain a large amount of protein hydrolases which is known as proteases. Proteases are degradative enzymes that can hydrolyse peptide bonds of proteins into amino acid and smaller peptides (Jisha et al., 2013; Mahajan & Badgujar, 2010). Proteases are ubiquitous in nature and are widely distributed in plants, animals and microorganisms, which is important in their physiological processes (Chew, Toh, & Ismail, 2018). Besides that, there are more than 2% of the functional genes encoded for proteolytic enzymes found in the human genome (Nair & Jayachandran, 2019). On top of that, plant proteases (43.85 %) are the most found protease in the biosphere, followed by bacteria (18.09%), fungi (15.08 %), animals (11.15 %), algae (7.42 %); virus proteases is the least, which only occupies around 4.41 % (Mahajan & Badgujar, 2010). Proteases have dominated 60% of the total commercialised enzymes sales as they have been recognised through their importance in laboratory, clinical and industrial processes (Mahajan & Badgujar, 2010; Wu, Ng, Sun, & Lan, 2017).

Currently, there are about 4000 proteases which are found and deposited in the MEROPS database available at https://www.ebi.ac.uk/merops/index.shtml (Rawlings et al., 2018). These proteases are well organised into 62 clans and 268 families based on structural similarity or sequence features. In general, proteases are categorised based the functional group found on their catalytic domain, namely cysteine protease, serine protease, threonine protease, aspartic protease, glutamic protease and metalloprotease

respectively. They can also be classified based on the basis of their mechanism of action. For instance, endopeptidases act on internal peptide bonds while exopeptidases hydrolyse the proteins at the terminal end of the peptide chains. In addition, proteases are also categorised according to their optimum pH namely alkaline proteases (pH 8 to 13), acidic proteases (pH 2 to 6) and neutral proteases (pH 6 to 8) (Ali & Muhammad, 2017; Jisha et al., 2013). Information on proteases classification is shown in Table 2.1.

Table 2.1   Classification of protease

Protease	Mechanism
Exopeptidases	Cleave the peptide bond proximal to the amino or carboxy termini of the substrate
Endopeptidases	Cleave internal bonds in polypeptide chains
Aspartic protease	An aspartic acid residue for their catalytic activity
Cysteine protease	Possesses a cysteine in the active centre
Glutamic protease	Containing a glutamic acid residue within the active site
Serine protease	Endopeptidases have an active centre serine involved in the catalytic process
Threonine protease	Harbouring a threonine (Thr) residue within the active site
Metalloprotease	Use a metal ion in the catalytic mechanism

Adapted from: Jisha et al. (2013) and Ali and Muhammad (2017)

2.2.1   Plant cysteine protease

Plant proteases represent the major fraction of the proteolytic enzymes found in nature. According to van der Hoorn, Leeuwenburgh, Bogyo, Joosten and Peck (2004), plant genomes encode hundreds of proteases. In plants, cysteine protease is the most abundant plant protease which accounts up to 34.92 % (Mahajan & Badgujar, 2010). Latex and fruit are considered as the most important source for these proteolytic enzymes (Kwon et al., 2015; Lin, Burns, & Gardner, 1993). The cysteine protease perform a vast array of cellular functions depending on their cellular distribution and intracellular localisation (Dubey, Pande, Singh, & Jagannadham, 2007). They are responsible for plant cellular housekeeping tasks such as intracellular protein recycling, defensive functions, leaf senescence and immune system cascade amplification as well as nutrient digestions in carnivorous plants (Butts et al., 2016; Díaz-Mendoza, Velasco-Arroyo, González-Melendi, Martínez, & Díaz, 2014; Li et al., 2013; Misas-villamil, Hoorn, & Doehlemann, 2016).

The most studied plant cysteine protease is papain (Ramli et al., 2018). Papain is isolated and purified from the latex of papaya (Sharma & Chatterjee, 2017). It is classified in clan CA and family C1 in the MEROPS database, which constitutes the largest cysteine protease superfamily (Verma, Dixit, & Pandey, 2016). The other plant proteolytic enzymes that are also classified in the same family are stem bromelain, fruit bromelain, comosain, ananain, actinidain, caricain and ficin. These plant cysteine proteinase usually have higher activities in unripe fruits (Butts et al., 2016). According to Amri and Mamboya (2012) and Nishiyama (2007), higher papain activity is found in green papaya and papain is reduced to trace the amount during maturation. This is also applied to actinidain, a cysteine protease found in kiwifruit. As reported by Nieuwenhuizen et al. (2007) and Afshar-Mohammadian et al. (2011), the mRNA level of actinidain is the highest in the mature harvest stage, and its activity and concentration increases until it reaches the harvest stage

Despite there are 4000 proteases are sequenced and stored in MEROPS database, there are only 1222 cysteine endopeptidase structure are deposited in RCSC PDB database. Since papain is the earliest plant cysteine protease determined protein 3D structures, any proteins that resemble papain are classified in the papain superfamily. The other available plant cysteine protease structure including actinidain from kiwi, ananain from pineapple, and ficain from fig (Figure 2.4). In general, plant cysteine proteases are depicted as two domains, L and R, respectively. The L domain is mainly α-helical, consisting of three α-helices and the R-domain is constituted by an antiparallel β-sheet structure.
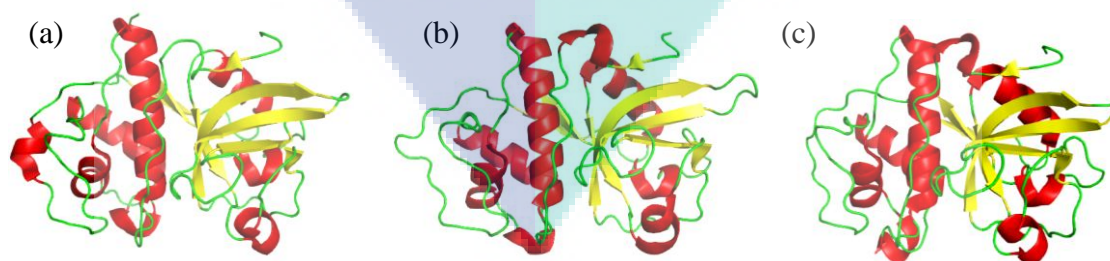
Figure 2.4   3D structure of (a) actinidain (PDB ID: 3P5U), (b) ananain (PDB ID: 6OKJ) and (c) ficain (PDB ID: 4YYQ). The alpha-helices, β-sheets and loops are coloured in red, yellow and green respectively.

Berger and Schechter (1970) defined the active site of papain constituted by seven subsites namely S1-His159; S2-Trp-177; S3-Gln19; S4-Gly23; S2'-Asp158 and S3'-Asn64 (Figure 2.5). No key residue was determined for S1' subsites. Later, it was found that S1' subsites were regulated by Gln135, Ala136, Ala137 and Gln142 lying on the R-subdomain and does not contribute to the substrate specificity to papain (Cordara et al., 2016; Menard et al., 1993). After that, Turk, Guncar, Podobnik, and Turk (1998) excluded S4 and S3' as binding subsites due to non-conservation among the papain-like family members.
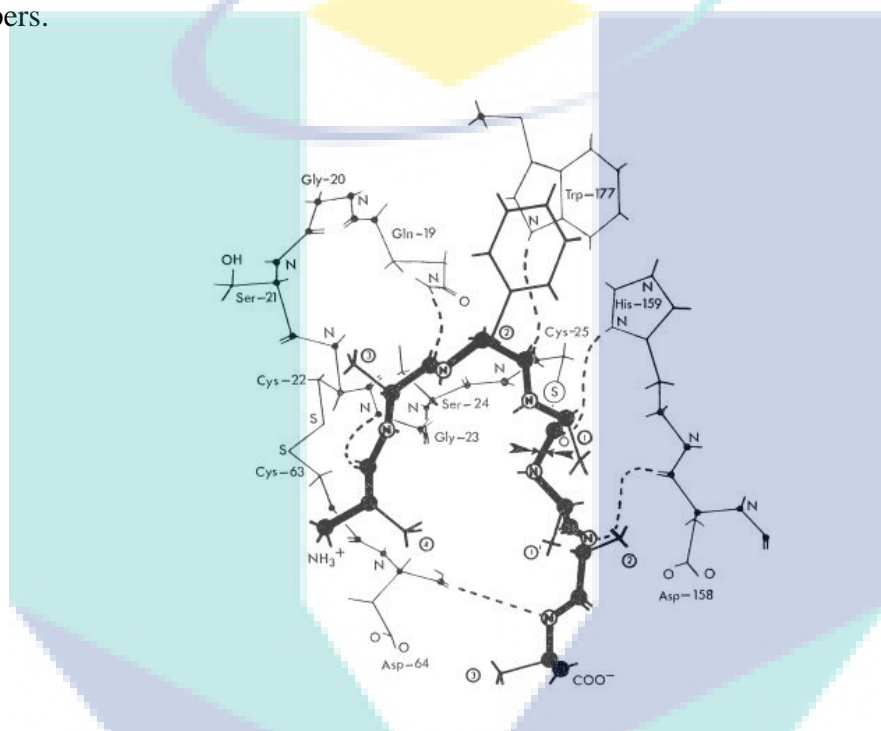
Figure 2.5   Projection of papain subsites.
Source: Berger and Schechter (1970)

On the other hands, the catalytic triad (Cys25-His159-Asn175) are found within the V-cleft (Figure 2.6). The catalytic residues Cys25 and His159 exist in the zwitterion form in which the thiol group of Cys25 was ionised while the imidazole ring of His159 was protonated (Beveridge, 1996). During catalysis, the thiol group of Cys25 performed a nucleophilic attack to the carbonyl carbon in the backbone of the substrate (Amri & Mamboya, 2012). After that, the Cys 25 was deprotonated by His 159 to allow the completion of the reaction. The deprotonation was aided by Asn175 to ensure the proper positioning of the imidazole ring of His159 by forming a hydrogen bond. Aside from the Cys-His-Asn triad, Gln19 was anticipated in the formation of an oxyanion hole which assists in the stabilisation of the transition-state complex (Menard et al., 1991).
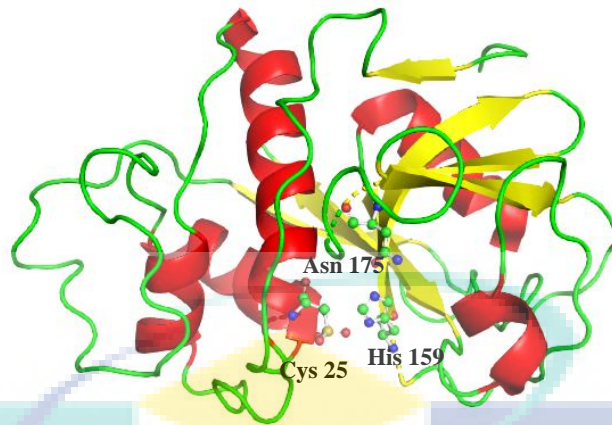
Figure 2.6   3D structure of papain (PDB ID: 9PAP). The catalytic residues Cys 25, His 159 and Asn 175 are labelled and represented as balls and sticks.

Plant cysteine proteases are produced in the zymogen form which is an inactive precursor regulated by pro-peptide at N-terminal (Figure 2.7). This pro-segment consists of three α-helices and one short beta-strand. This α-helical domain blocks the catalytic cleft with an extended part across between the pro-domain and mature part of the papain (Roy, Choudhury, Aich, Dattagupta, & Biswas, 2012). The zymogen is activated in an acidic condition in which the pro-peptide is removed to yield a mature and fully active papain enzyme. Interestingly, this pro-domain is also envisaged to store important structural information in a natural event termed as "protein memory" to act as intramolecular chaperones which is crucial in protein folding (Demidyuk, Shubin, Gasanov, & Kostrov, 2010; Shinde, Liu, & Inouye, 1997). According to Roy et al. (2012), this pro-domain can fold itself independently and direct the folding of the mature part of the papain into a native state by lowering the free-energy barrier of folding.
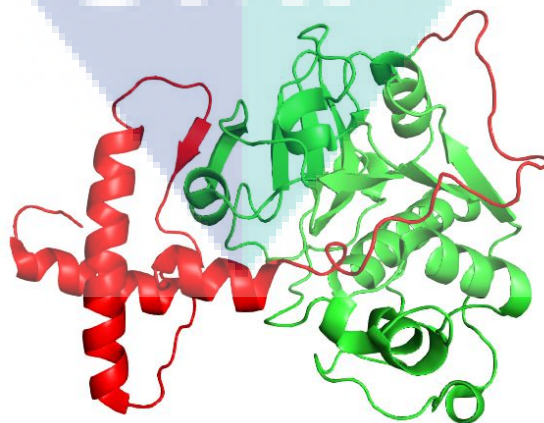
Figure 2.7   3D structure of plant cysteine protease using pro-papain (PDB ID: 3TNX) as reference. The red region indicated pro-peptide while the green region marked the mature part of the plant cysteine protease.

A full-length papain protein sequence (UniProt Accession No: P00784) is displayed in Figure 2.8. According to the MEROPS database, this sequence has a length of 345 amino acid residues which constituted the signal peptide at amino acid position 1-18; amino acids position 19-133 forming pro-peptide and the active chain containing amino acids 134-345. The catalytic residues cysteine, histidine and asparagine are located at amino acids position 158, 292 and 308 respectively. The sequence characteristics of papain are well characterised. Amino acids residues EKIYRFEIFKDNLKYIDETN form a helix core in the pro-domain while GLNVFAD is located at the kink of beta-sheet (Groves et al., 1996; Ramli et al., 2018). Conserved evolution of these motifs was found in other C1A proteases in a pattern of $EX_3RX_3FX_2NX_3IX_3N$ (ERFNIN) and GXNXFXD (GNFD) such as bromelain, cathepsin L and procaricain (Groves et al., 1996; Lee et al., 2012; Ramli et al., 2018). Besides that, other conserved motifs including CGSCWAF, HA and NSW that carry the catalytic residues are also well conserved in other C1A proteases. Thus, these conserved motifs serve as a basis to identify papain-like protease.


```
            10        20        30        40        50        60
    ....|....|....|....|....|....|....|....|....|....|....|....|
    MAMIPSISKLLFVAICLFVYMGLSFGDFSIVGYSQNDLTSTERLIQLFESWMLKHNKIYK

            70        80        90       100       110       120
    ....|....|....|....|....|....|....|....|....|....|....|....|
    NIDEKIYRFEIFKDNLKYIDETNKKNNSYWLGLNVFADMSNDEFKEKYTGSIAGNYTTTE

           130       140       150       160       170       180
    ....|....|....|....|....|....|....|....|....|....|....|....|
    LSYEEVLNDGDVNIPEYVDWRQKGAVTPVKNQGSCGSCWAFSAVVTIEGIIKIRTGNLNE

           190       200       210       220       230       240
    ....|....|....|....|....|....|....|....|....|....|....|....|
    YSEQELLDCDRRSYGCNGGYPWSALQLVAQYGIHYRNTYPYEGVQRYCRSREKGPYAAKT

           250       260       270       280       290       300
    ....|....|....|....|....|....|....|....|....|....|....|....|
    DGVRQVQPYNEGALLYSIANQPVSVVLEAAGKDFQLYRGGIFVGPCGNKVDHAVAAVGYG

           310       320       330       340
    ....|....|....|....|....|....|....|....|
    PNYILIKNSWGTGWGENGYIRIKRGTGNSYGVCGLYTSSFYPVKN
```

Figure 2.8   Full length protein sequence of papain (UniProt Accession No: P00784). The conserved motifs are highlighted in the red box.

2.3   Bromelain

Bromelain is a general term to describe the crude extract with endopeptidase activity obtained from family Bomeliacea (Ramalingam et al., 2012). There are four major proteolytic fractions which constitute bromelain including ananain, comosain, fruit and stem bromelain. Stem bromelain (EC 3.4.22.32) and fruit bromelain (EC3.4.22.33) are present as the most significant fractions and can be isolated from the stem and pineapple fruit, respectively. Besides that, bromelain can also be obtained from the leaves and peel of the pineapple (Bresolin, Bresolin, Silveira, Tambourgi, & Mazzola, 2013). The main component of bromelain is sulfhydryl proteolytic fraction while non-enzymatic components are escharase, acid phosphatases, glycosidases, peroxidases, ribonucleases, cellulases, glycoproteins, carbohydrates and protease inhibitors (da Silva López, 2017).

Despite the fact that the chemical composition of bromelain has been revealed, there are some remaining ambiguities to be investigated. First, bromelain is present as multiple isoforms in the pineapple crude extract. Ota and Muta (1985) reported an isolation of six isoforms of stem bromelain and two isotypes of fruit bromelain. However, Larocca et al. (2010) separated at least six isoforms of fruit bromelain using 2-D zymography. It is suspected that these bromelain isoforms are the result of splicing and/or post-translational modifications. Intriguingly, it was also found that stem bromelain and fruit bromelain co-exist in the pineapple plant. According to Raimbault et al. (2013), both fruit and stem bromelain can be found in the pineapple fruit. Furthermore, Amid et al. (2011) reported the isolation of fruit bromelain (NCBI Genbank Accession Number: BAA21849) from the pineapple stem.

Bromelain has a wide substrate spectrum, from synthetic low molecular mass amides and dipeptides up to high molecular substrates. It can act on various substrates such as fibrin, albumin, casein, angiotensin II, bradykinin, on their glycyl, alanyl and leucyl bonds (Maurer, 2001). Among the small molecule substrates, stem bromelain is highly specific on Z-Arg-Arg-I-NHMec while fruit bromelain is preferable to cleave on Bz-Phe-Val-Arg-l-NHMec (Ramalingam et al., 2012). In general, both stem and fruit exhibited optimum activity at pH 7-8; 40-60 °C (Table 2.2).

Table 2.2   Optimum pH and temperature of bromelain towards different substrate

Stem bromelain					
Species (variety)	**Substrate**	**Optimum pH**	**Optimum temperature (°C)**	**Activity**	**Reference**
A. comosus (L. Merr.)	Casein	7	45	555.13 U/mg	Kothare, Pardhi, Chivte, Muley, and Shanbhag (2017)
A. comosus (L. Nang Lae)	Casein	7	50	36,111 ± 1.62 U	
A. comosus (L. Phu Lae)	Casein	7	60	42,482 ± 2.22 U	
Fruit bromelain					
Species (variety)	**Substrate**	**Optimum pH**	**Optimum temperature (°C)**	**Activity**	**Reference**
A. comosus	Azocasein	8	50	9.82 U/mg	Ramalingam et al. 2012
A. comosus (L. Merr.)	Casein	7	45	1477.45 U/mg	Kothare et al. (2017)
A. comosus	Azocasein	6.5	50	N. A	Corzo, Waliszewski, and Welti-Chanes (2012)
	Azoalbumin	7.5	55	N. A	
	Casein	7.7	59	N. A	
	Sodium caseinate	6.5	59	N. A	
	Haemoglobin	2.9	37	N. A	

Furthermore, temperature affects the stability of bromelain. Ramalingam et al. (2012) reported that bromelain is able to keep its activity after incubation at 50 °C for 15 mins and complete inactivation was observed after incubating at 80 °C for the same duration. Besides that, Jutamongkon and Charoenrein (2010) also reported a similar finding in which bromelain isolated from Smooth Cayenne showed no loss of proteolytic activity at 40 °C up to one hour; 83 % of activity is preserved at 50 °C; and the activity was totally lost after heating at 80°C for 8 mins. Liang, Li, Shi, Liao, and Wu (2012) found that bromelain underwent significant deterioration when the temperature exceeded 60 °C. Bromelain isolated from the Josapine pineapple displayed a different thermal profile. The Josapine bromelain was still able to retain 60 % of its activity after 30 mins incubation at 85 °C and 20 % activity remained after heating at 105 °C.

2.3.1 Expression and function of fruit bromelain

Currently, information about the expression and function of fruit bromelain is still fragmentary (Martinez, Cambra, Gonzalez-Melendi, E. Santamarıa, & Diaz, 2012). It is now known that the expression of fruit bromelain is different to papain, which does not disappear as the fruit ripens (Upadhyay et al., 2010). Findings via microarray and Northern blot analysis found that the fruit bromelain transcripts are strongly down-regulated during the ripening of the pineapple fruit (Koia, Moyle, & Botella, 2012; Moyle et al., 2005). According to Bresolin et al. (2013), the concentration of bromelain in matured fruit is relatively high despite its proteolytic activity declining with the increase of fruit maturation. It seems like the fruit bromelain is highly expressed in the early stages of pineapple fruit development and its level increased along its maturation. However, there was no explanation on why the activity of fruit bromelain was reduced in mature fruits.

According to Misas-villamil et al. (2016), CA1 proteinases play a role in plant defence. Thus, fruit bromelain may act as defence protein to eschew animal and insect feeding since fruit bromelain can produce a sore feeling in the mouth as a result of the reaction between bromelain and proteins present on the tongue tissue (George, Bhasker, Madhav, Nair, & Chinnamma, 2014; Srujana & Narayana, 2017). Besides that, pineapple plants are susceptible to the attack of insects and fungus such as *Cyanophora paradoxa*, *Diaspis bromeliea* and mealybugs which act as vector for diseases or infections that can affect the quality of pineapple fruits (Dawson, 2016; Rohrbach & Johnson, 2003). Hence, higher fruit bromelain expression in early pineapple fruit development is needed to protect the pineapple fruits from these bugs and pathogens. This hypothesis was further supported by pathogen resistance observed in fruit bromelain overexpressed transgenic *Arabidopsis thaliana* and *Brassica rapa* (Jung et al., 2008; Wang et al., 2014).

2.3.2 Applications of bromelain

Bromelain is reputable for its therapeutic properties, which has been used in traditional medicine for various health related issues in native cultures such as the Philippines and Hawaii (Muhammad & Ahmad, 2017; Rathnavelu, Alitheen, Sohila, Kanagesan, & Ramesh, 2016). It was introduced to be used as a therapeutic compound

by Heinicke in 1957 (Kelly, 1996). At present, it is used as a complementary or alternative medication to glucocorticoids, nonsteroidal antirheumatics and immunomodulatory agents (Maurer, 2001). Therapeutic applications of bromelain include being an anti-inflammatory agent, anti-cancer agent, or anti-biotics; improved gastrointestinal tract related issues; inhibited the formation of thrombus; and is used in dermatological disorder treatments as shown in Table 2.3 (Manzoor, Nawaz, Mukhtar, & Haq, 2016; Muhammad & Ahmad, 2017). The use of bromelain for therapeutic purposes is efficient because it is easily absorbed in the gastrointestinal tract and high dosage (12g/day) is allowed without side effects (Wali, 2019).

Table 2.3   Role of bromelain as therapeutic agent

Application	Description	References
Anti-inflammatory and analgesic	The exact mechanism is still unclear; it was hypothesised three possible pathways which bromelain modulate the action of different biomolecules and/or secretion of hormones: a. Kallikrein-kinin pathway b. Arachidonic acid pathway c. Cell Mediated Immunity	Muhammad & Ahmad (2017)
Anti-biotics	Bromelain is capable to increase the permeability of organ and tissue towards antibiotics which increase the level of antibiotics in organism	Bhattacharyya (2008)
Avoid platelet Aggregation/ formation of thrombus	Due to fibrinolytic activity of bromelain.	Kelly (1996)
Anti-cancer	Modification of key pathways which allow the growth of tumor cells.	Pavan, Jain, Shraddha, and Kumar (2012)
Debridement	Hydrolyse fibrin clot, collagen, elastin, laminin, fibronectin and other damaged components of extracellular matrix.	da Silva López (2017)

Apart from that, bromelain also has impressive characteristics as a meat tenderisation enzyme to break down the collagen fibres of the meat, which is accounted for approximately 95 % of the meat tenderising enzyme in the United States (Han et al., 2018; Manohar, Gayathri, & Vishnupriya, 2016). Besides that, bromelain was also used in baking to degrade gluten in the flour to improve the digestibility, flavour, nutritional value and texture of the product (Heredia-Sandoval, Valencia-Tapia, de la Barca, & Islas-

Rubio, 2016; Nair & Jayachandran, 2019). In the dairy industry, fruit bromelain was used to replace rennet in cheese production to allow milk clotting to occur by hydrolysing specific peptide bonds in k-casein protein (Ismail et al., 2019). Likewise, it also facilitates cheese ripening and whey hydrolysis. In addition, its potential to be applied in different industrial sectors also cannot be neglected such as beverages production, leather processing and detergents (Table 2.4).

Table 2.4   Applications of bromelain in industry

Industry	Application
Leather	Bating of leathers
	Dehairing and dewooling of skins
Food processing	Modification of protein rich material such as soy protein or wheat gluten
Baking	Dough conditioners
Dairy	Coagulation of milk protein, production of enzyme modified cheese; whey processing
Detergent	Removal of protein stain
Meat	Meat tenderization
Beverage	Removal of turbidity
Confectionary	Reverse hydrolysis in aspartame synthesis

Adapted from Mahajan & Badgujar (2010)

2.4   Protein structure prediction

Protein is constituted as a chain of amino acids which is referred to as the protein's primary structure. Due to the chemical and physical interactions within amino acids and its environment, the amino acids start to fold periodically, forming its secondary structure which eventually leads to the formation of a tertiary structure and/or quaternary structure (combination of two or more polypeptide chains). Current approaches on the protein 3D structure investigation rely on protein crystallography (X-ray diffraction), electron microscopy and nuclear magnetic resonance (NMR) (Dorn, Silva, Buriol, & Lamb, 2014). However, the cost is extremely high and laborious (Wang, Eickholt, & Cheng, 2010). Thus, the discovery of the protein structure is still lagging behind in contrast to the exponential increment of known protein sequences (Lee, Freddolino, & Zhang, 2017). In this regard, the use of the *in silico* protein structure prediction method can serve as an alternative to protein structure related studies.

Protein structure prediction is divided into two parts: free modelling (*ab initio*) and template-based modelling (Kryshtafovych & Fidelis, 2009). The *ab initio* method is solely based on physico-chemical principles. The predicted protein structure is in conformations which are thermodynamically and stereochemically favourable (Pavlopoulou & Michalopoulos, 2011). The protein is built as its lowest energy state by considering torsion angles and the position of atoms in the protein structure based on the first principle (Dorn et al., 2014). However, *ab initio* method may result in a new protein fold which is not true in reality since it is not driven by a fundamental understanding of the mechanisms and structure formation (Mihăşan, 2010). Besides that, *ab initio* is too computationally demanding which may only be useful for a very short protein fragment and when a suitable template is not available. Hence, this method has been replaced by template-based modelling.

Template-based modelling is further classified into homology modelling/ comparative modelling and fold recognition where the protein structure is modelled based on the known structure available in PDB on the basis of similar sequences sharing similar structures or different sequences sharing a similar protein fold (Mihăşan, 2010). According to Pavlopoulou and Michalopoulos (2011), comparative modelling is the most reliable approach for protein structure prediction. In general, comparative modelling always starts by searching the best homologous template of a known protein structure, followed by optimising the alignment between the target sequence and the chosen template to build the protein structure of interest that resembles the template protein structure (Meier & Söding, 2015). This method can produce protein models with high quality which is comparable to low-resolution X-ray crystallography or medium-resolution NMR solution structures (Fiser, 2010).

2.4.1 Structure prediction of bromelain

To date, there is no bromelain crystal structure deposited in the PDB database. Herein, a current discussion is focused on the prediction of the bromelain structure via computational modelling. According to Das and Bhattacharyya (2018), a theoretical model of stem bromelain (UniProtKB - P14518) was deposited in the PDB database (PDB_ID:1W0Q) in 1990. They made a comparison between comparative stem bromelain structure using ervatamin B as a template via the Swiss Model software with

1W0Q which showed that there was an 85 % identical arrangement of backbone structures and non-overlapping segments form either β-sheets or random coils. However, this model is currently not accessible because of PDB's new policy which separates the theoretical model coordinates from its main achievements.

Besides that, a comparative model of stem bromelain (Gene Bank Accession Numbers: ADY68475.1) was built using 1PCI as a template (Tap, Majid, & Khairudin, 2016). The predicted stem bromelain has a similar structure as papain and is constituted by 40 % of alpha-helix, 11 % beta-strands and 39 % coils. Later, another comparative structure of ADY68475.1 was generated using 1YAL as a template (Tap, Majid, & Khairudin, 2017). This model was reported to contain 7 alpha-helix and 2 beta-strands. Recently, comparative stem bromelain (Gene Bank Accession Numbers: CAA08861.1) and fruit bromelain (Gene Bank Accession Numbers: BAA21848.1) structures were reported by Ramli et al. (2018). Both models were constructed using 1PCI as a template comprising 10 α-helices and β-strands.

It is noticeable that the current available predicted bromelain structures are built using a single template. Since the chosen template may not necessarily cover the full length of the bromelain sequence, multiple templates can be used in such a way as to cover the complete sequence of bromelain to be modelled (Ganugapati & Akash, 2017). Besides that, a combination of multiple templates can also cover the weakness of one template by another template. For instance, the absence of structural information in one template can be provided by other templates; a fragment of the target which shows low structural similarity with one of the templates may exhibit higher structural similarity with the other templates (Chakravarty, Godbole, Zhang, Berger, & Sanchez, 2008)

2.5   Molecular dynamics simulation

The function of biological molecules are always the result of interaction and motion since they are highly dynamic in reality (Dror, Dirks, Grossman, Xu, & Shaw, 2012). Despite there being a large number of biomolecule 3D structures available from PDB that can be used for structural and biochemical studies, these static structures do not represent their true behaviour in nature. This makes the study of behaviour and interaction of these macromolecules difficult. With the rapid advent in computing technology,

molecular dynamics simulation is created to study the physical movements of atoms and molecules in order to understand the dynamics and structure during the motion of individual atoms (Chen, Huang, Miao, Feng, & Campanella, 2019). The molecular dynamics simulation acts as a toolbox to mimic the natural environment and permit the molecules of interest to interact for a period of time to allow the investigation of the time-dependent behaviour of a molecular system (Feng et al., 2015).

2.5.1 Molecular dynamics simulation for protein structure refinement

One of the applications of molecular dynamics simulation is to refine the predicted protein structure. The differences of amino acids due to insertions and gaps between the target and template as well as the interaction with the other biomolecules present in the template can cause errors in the protein structure prediction (Feig, 2017; Ishitani, Terada, & Shimizu, 2008; Park, Ovchinnikov, Kim, DiMaio, & Baker, 2018). This is achieved by simulating the predicted protein structure with a force field; a complex mathematical expression that defines molecular features such as bond length and angles, bond rotations, Lennard–Jones potentials, van der Waals and electrostatic interactions (Gelpi, Hospital, Goñi, & Orozco, 2015). Throughout the simulation, a trajectory of conformation is generated to allow the protein structure to adopt to its sequence and reach its favourable thermodynamics and stereochemical state (Gelpi et al., 2015). This allows improvement of the predicted structures comparable to experimental determined structures.

2.5.2 Investigation on protein structural changes using molecular dynamics simulation

Due to dynamics and flexibility, protein undergo conformational changes from one to another under the influence of physiological conditions such as temperature (Dror et al., 2012; Moree et al., 2015). This affects the protein activity and even leads to protein denaturation when exposed to a high temperature for a long period of time (Fields, Dong, Meng, & Somero, 2015). Thus, it is interesting to investigate the relationship between protein function and its conformation at a condition of interest. This is particularly important if the protein is used in harsh conditions in certain industrial processing steps. In contrast to the crystallography method which only provides the static information of a protein treated at one point, molecular dynamics simulation captures the molecular resolution that evolved with time to disclose the stability of protein under a selected

temperature throughout the simulation (Childers & Daggett, 2017; Marchand et al., 2018). This reveals important information on the changes in integrity and stability of protein under the influence of temperature and is useful for protein engineering of a thermostable enzyme (Zeiske, Stafford, & Palmer III, 2016)

2.6   Summary of literature review

As a summary, the literature review covers the background of *A. comosus*, plant cysteine proteases, and fruit bromelain. Indeed, expression analysis of fruit bromelain during the growth cycle of pineapple fruits is very limited. Previous studies on cultivar Smooth Cayenne revealed the expression of fruit bromelain is down-regulated in mature pineapples. A similar result is expected to be found in cultivar MD 2. Besides that, the only available fruit bromelain structure prediction study is constructed using single template. The accuracy and quality of the fruit bromelain model will be addressed by the use of multiple templates in the present study. Since papain is the earliest studied enzyme and present as the model enzyme in the plant cysteine protease family, the generated fruit bromelain structures will be compared with the available papain structures to illustrate the structural information of fruit bromelain. Last but not least, simulation of fruit bromelain at different temperature to elucidate its conformation changes represent the novelty of this study and to serve as a starting point to improve thermal stability of fruit bromelain for industrial applications.

METHODOLOGY

3.1 Flow chart

The overall experimental procedure in this research is shown in Figure 3.1.

Objective 1:To determine fruit bromelain gene expression between unripe and ripe *A. comosus* **cultivar MD 2**

-Total RNA isolation

-Relative gene expression analysis via qPCR

-Crude fruit bromelain extraction

-Enzymatic assay using casein as substrate

Objective 2:To analyse the fruit bromelain sequences from *A. comosus* **cultivar MD 2**

-Retrieve fruit bromelain sequences from NCBI genbank

-Sequence alignment using ClustalW

-Full-length amplification of selected fruit bromelain transcripts

-Data mining using BLAST, SignalP 5.0, InterProScan and ProtParam

Objective 3:To model the tertiary structure of fruit bromelain

-Template selection via BLAST against PDB database

-Structure generation using MODELLER 9.20

-Structure refinement via molecular dynamics simulation

Objective 4: To investigate the structural stability at different temperatures

-Enzymatic analysis of fruit bromelain at different temperatures

-Simulate fruit bromelain structures at several target temperatures

Figure 3.1   The flowchart of overall experimental procedure.

3.2   Chemicals

All chemicals used in this study were purchased from Bioline (US), First Base (Malaysia), Fisher Scientific (UK), Merck (US), Oxoid (UK) and R&M (Malaysia) unless stated otherwise.

3.3   Bioinformatics analysis

The software and programmes for bioinformatics and structural analysis used in this study are BioEdit (Hall, 1999), Basic Local Alignment Search Tool (BLAST), available at https://blast.ncbi.nlm.nih.gov/Blast.cgi (Altschup, Gish, Miller, Myers, & Lipman, 1990) GROMACS 5.14 (Abraham et al., 2015; Feig, 2016), locPREFMD available at http://feig.bch.msu.edu/locprefmd/ (Feig, 2016), MODELLER version 9.20 (Sali & Blundell, 1993), PyMOL version 2.2.2, SignalP 5.0 available at http://www.cbs.dtu.dk/services/SignalP/ (Armenteros et al., 2019), InterProScan available at https://www.ebi.ac.uk/interpro/search/sequence-search (Jones et al., 2014), ProtParam available at https://web.expasy.org/protparam/ (Gasteiger et al., 2005), Swiss-PdbViewer (Guex & Peitsch, 1997) and structure stereochemical quality checking tools including ERRAT, PROCHECK and Verify3D at The Structure Analysis and Verification Server version 5.0 (SAVES 5.0) available at http://servicesn.mbi.ucla.edu/SAVES.

3.4   Expression analysis of fruit bromelain between ripe and unripe *A. comosus* cultivar MD 2

The gene expression differences of fruit bromelain between ripe and unripe *A. comosus* cultivar MD 2 was studied via relative gene expression analysis using qPCR. The expression level of fruit bromelain between ripe and unripe *A. comosus* cultivar MD was normalised using actin as a reference gene. Besides that, the proteolytic activity between ripe and unripe *A. comosus* cultivar MD 2 was investigated via enzymatic assay using casein as a substrate.

3.4.1   Total RNA isolation

A. comosus cultivar MD 2 was provided by the Malaysian Pineapple Industry Board (Figure 3.2). The pineapple flesh was grinded with liquid nitrogen using mortar

and pestle. After that, total RNA was extracted from the grinded powder using TransZol Up Plus RNA Kit (TransGen Biotech, China) according to the manufacturer's instruction. The concentration and purity of RNA was quantified by OPTIZEN NanoQ (Mecasys, Korea) and RNA's integrity was examined on 1.0 % (w/v) agarose gel.

(a) (b)
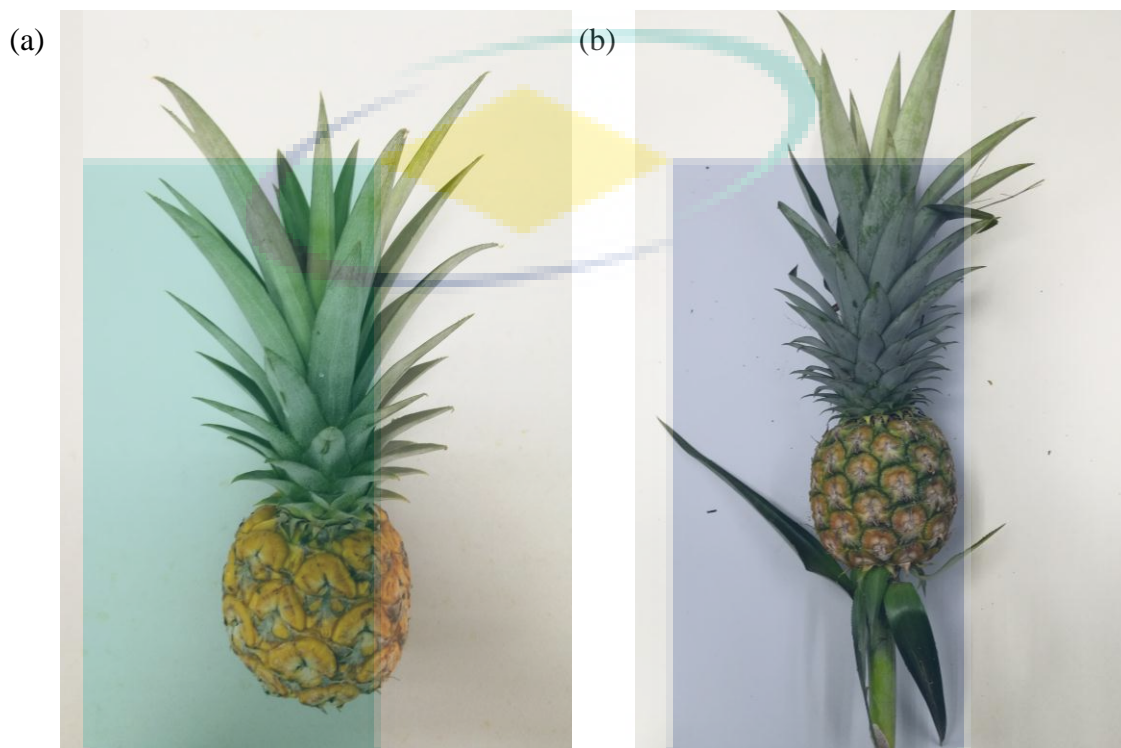
Figure 3.2   (a) Ripe *A. comosus* cultivar MD 2. (b) Unripe *A. comosus* cultivar MD 2.

3.4.2   TAE buffer preparation and agarose gel electrophoresis

TAE buffer (50×) constituted 242 g of Tris base, 57.1 mL of glacial acetic acid and 100 mL of 0.05 M ethylenediaminetetraacetic acid (EDTA) (pH 8.0) was made to 1 L with distilled water (Sambrook & Russell, 2006). The TAE buffer was diluted to 1× with distilled water prior to use. To prepare the agarose gel, agarose powder was mixed with 1× TAE buffer and was heated in a microwave until it dissolved. After that, the agarose gel solution was mixed with GelRed solution (0.02 µl/mL) and poured into a gel cast to allow solidification. The solidified agarose gel was transferred to an electrophoresis tank. Subsequently, 5 µL of sample was mixed with 1 µL of 6× loading dye. The electrophoresis system was run at 110 volts for 40 mins. Lastly, the agarose gel was observed under Amersham Imager 680 (GE Healthcare, US).

3.4.3 First-stand cDNA synthesis

The extracted total RNA was used to synthesise the first-strand complementary DNA (cDNA) via RT-PCR using QuantiNovaTM Reverse Transcription Kit (Qiagen, Germany) following the manufacturer's instruction. Firstly, 1 µg of total RNA, 2 µL of gDNA Removal Mix and RNase-free water were mixed and incubated at 45 °C for 2 mins to remove genomic DNA. Then, the reaction mixture was added with 4 µL of Reverse Transcription Mix and 1 µL of Reverse Transcription Enzyme. After that, oligo-dT and random primers were allowed to anneal to mRNA at 25 °C for 3 mins. Subsequently, the reverse transcription process was conducted at 45 °C for 10 mins followed by inactivation at 85°C for 5 mins.

3.4.4 Primer design

The primer design for both target gene fruit bromelain and reference gene actin were followed Raimbault et al. (2013) as shown in Table 3.1. The primers have reasonable GC content (percentage of guanine and cytosine) of ~50 % and melting temperature (Tm) of 53-56 °C (Dieffenbach et al., 1993).

Table 3.1   Primer design for fruit bromelain and actin

Primer name	Primer sequence (5'→ 3')	GC content (%)	Tm (°C)
Fruit bromelain (Forward)	CAA GGA CAA CGA CGA GAA GA	50.0	54.5
Fruit bromelain (Reverse)	CAA ATG CAC CAC TGG CTC	52.6	53.9
Actin (Forward)	GTG GCA CTT GAC TTT GAG CA	50.0	55.9
Actin (Reverse)	CTT CCT GAT ATC CAC ATC GC	50.0	52.9

3.4.5 Relative gene expression analysis of fruit bromelain using qPCR

The qPCR reaction mixture was prepared in a total volume of 20 µL containing 1× QuantiNova SYBR Green PCR Master Mix, 0.7 µM of forward and reverse primers, 1 µL of first-strand cDNA and RNase-free water. The PCR amplification was carried out in the Mastercycler ® ep Realplex Real-time PCR (Eppendorf, Germany) with cycling parameters: 95 °C for 2 mins; 40 cycles of 95 °C for 30 s, annealing for 30 s and extension

at 72 °C for 30 s. The optimum annealing temperature was investigated by running the amplification at five different annealing temperatures: 55 °C, 57 °C, 60 °C, 62 °C and 65 °C. Subsequently, melting curve analysis was conducted at 55–95°C with 0.5 °C increments for 15 s each. The amplifications were run triplicates. The annealing temperature which produced the lowest quantitation cycle (C_q) value was chosen as the optimum temperature for subsequent amplifications.

After that, the efficiency of primers anneal to the target region was examined. The first-strand cDNA was diluted in a series of 5-folds (5^{-1}, 5^{-2}, 5^{-3}, 5^{-4}, 5^{-5}) to be used as a template. To determine the efficiency of amplification, a standard of curve of C_q values against the cDNA concentration was plotted and calculated as in Eq 3.1:

$$\text{Efficiency} = 10^{\,(-1/\text{slope})} - 1 \qquad\qquad 3.1$$

where the slope is the gradient of linear equation generated from the standard curve.

Subsequently, the relative changes of the expression level of fruit bromelain between ripe and unripe *A. comosus* cultivar MD was expressed in fold changes via the $2^{-\Delta\Delta C_q}$ method by using the unripe *A. comosus* cultivar MD as a calibrator (Livak & Schmittgen, 2001):

$$\text{Fold changes} = 2^{-\Delta\Delta C_q} \qquad\qquad 3.2$$

where $^{\Delta\Delta C_q}$ is the differences in C_q values of fruit bromelain of ripe and unripe *A. comosus* cultivar MD 2 normalised by actin.

3.4.6 Enzymatic assay buffer preparation

Potassium phosphate buffer was prepared by mixing potassium dihydrogen phosphate and dipotassium hydrogen phosphate according to the Henderson-Hasselbalch equation as shown in Eq 3.3:

$$\text{pH} = \text{pKa} + \log\,[\text{OH}^-]/[\text{H}] \qquad\qquad 3.3$$

where [OH$^-$] is concentration of base; [H] is concentration of acid; pKa is the negative log of the weak acid dissociation constant and log means the base ten logarithm. A total volume of 0.65 % (w/v) casein (pH 7) was dissolved in 50 mM potassium phosphate buffer (pH 7.5). Subsequently, tricholoroacetic acid (TCA) and Folin and Ciocalteu's (F-

C) phenol reagent were diluted from their stock solution to a final concentration of 110 mM and 0.5 M respectively. Sodium carbonate (Na_2CO_3) buffer was prepared by adding 53 g of anhydrous Na_2CO_3 in 1 L of distilled water to a final concentration of 500 mM. L-tyrosine (1.1 mM) was prepared by dissolving 0.2 mg of L-tyrosine 1 L distilled water under low heat. The pH was adjusted with 1 M hydrochloric acid (HCl) or 1 M sodium hydroxide (NaOH) wherever necessary.

3.4.7   Fruit bromelain extraction and enzymatic assay

The fruit bromelain was extracted following the method reported by Al-Sa'ady, Al-Hadban and Al-Zubaidy (2016) with some modifications. Firstly, a whole pineapple was cleaned and unwanted parts including the crown and skin were discarded. After that, the pineapple fruit was cut into pieces. Next, approximately 50 g of pineapple flesh was homogenised with 100 mL of cold 0.1 M potassium phosphate buffer (pH 7.0) in a blender for 2 mins. The obtained crude extract was filtered using a muslin cloth and further clarified by centrifugation at 6000 rpm for 20 mins at 4 °C. The supernatant obtained was stored at -20 °C and used as crude enzyme in the subsequent experiment.

The enzymatic analysis was conducted as described by Cupp-enyard and Sigma-Aldrich (2008) with some modifications. Firstly, 5 mL of 0.65 % (w/v) of casein was added to both the "test" vial and "blank" vial, respectively, followed by equilibration at 37 °C in a water bath. After that, 1 mL of the crude bromelain was added to "test" vial, mixed and incubated at 37 °C in a water bath for 10 mins. The reaction was terminated by adding 5 mL of 110 mM TCA solution into the "test" vial and "blank" vial respectively. Subsequently, 1ml of the crude bromelain was added to the "blank" vial. The reaction mixture was mixed and centrifuged at 6000 rpm for 10 mins. Next, 2 mL of the supernatant was mixed with 5 mL of 500 mM Na_2CO_3 solution and 1mL of 0.5 mM F-C phenol reagent. The reaction mixture was mixed and left standing for 30 mins at room temperature prior to absorbance reading at 660 nm.

On the other hand, a tyrosine standard curve was generated by measuring the absorbance of tyrosine of different concentrations at 660 nm. One unit of enzyme activity is defined by the amount of crude bromelain required to hydrolyse casein into 1 µM of tyrosine in one minute in the experimental condition per mL as explained in Eq 3.4:

$$\text{Enzymatic activity (U/mL)} = \frac{(\text{μmol tyrosine equivalents released}) \times (1.1)}{(1) \times (10) \times (2)} \qquad 3.4$$

where 1.1 is the total volume of assay in mL; 10 is the duration of assay in mins; 1 is the volume of crude fruit bromelain in mL and 2 is the volume of sample used in colorimetric determination in mL. The enzymatic assay was conducted with three biological replicates in triplicates.

3.5 Fruit bromelain sequences analysis and isolation

The *A. comosus* cultivar MD2 fruit bromelain sequences were retrieved and analysed. Three fruit bromelain sequences of interest with accession no. OAY62650.1, OAY68270.1 and OAY85858.1 were isolated through PCR amplification and cloning methods. A detailed study on the isolated fruit bromelain sequences was performed using several bioinformatic tools.

3.5.1 Sequence analysis

A. comosus cultivar MD 2 fruit bromelain protein sequences were retrieved from NCBI genbank. The retrieved protein sequences were analysed based on their length of amino acid were aligned using ClustalW built in BioEdit.

3.5.2 Full-length amplification of fruit bromelain transcripts

Forward and reverse primers were designed manually to amplify the selected full-length fruit bromelain sequences (Table 3.2). The primers were synthesised by Integrated DNA Technologies Pte. Ltd, Singapore.

Table 3.2   Primers design of OAY62650.1, OAY68270.1 and OAY 85858.1

Primer name	Primer sequence (5'→ 3')	GC content (%)	Tm (°C)
OAY62650.1 Forward	ATG GCT TCC AAA TTT CAA CTA GTG	37.5	53.9
OAY62650.1 Reverse	TCA AGT TTT AGA AAC AAT CTT AAT AAC TTC GG	28.1	54.6
OAY68270.1 Forward	ATG ATG AAG CGG TTT GAA GAA TGG ATG	40.7	57.8
OAY68270.1 Reverse	TCA AGT TTC AGA AAC CAT CTT AAT AAG TTC GGC ATT AGC	35.9	61.1

Table 3.2 Continued

Primer name	Primer sequence (5'→ 3')	GC content (%)	Tm (°C)
OAY85858.1 Forward	ATG GCT TCC AAA GTT CAA CTC G	45.5	55.8
OAY85858.1 Reverse	TCA AGT TTC AGA AAC CAT CTT AAT AAC TGC	33.3	55.9

The fruit bromelain transcripts were amplified using TopTaq Master Mix Kit (Qiagen, Germany) following the manufacturer's instruction. The reaction mixture contained 1× PCR buffer, 1× CoralLoad Concentrate, 1 µL of first-strand cDNA template, 0.25 of µM forward and reverse primers, respectively, as well as RNAse free water in a total volume of 25 µL. PCR was carried out by pre-denaturation at 95 ℃ for 3 min, followed by 35 cycles of amplification (denaturation at 94 ℃ for 30 sec; annealing at 60º C for 30 sec; extension at 72 ℃ for 1 min) and final extension at 72℃ for 10 mins in Mastercycler Pro S (Eppendorf, Germany). The PCR product was then electrophoresed on 1.0 % (w/v) agarose gel together with 1kb DNA ladder as a marker (Promega, US).

3.5.3 Bacterial growth media preparation

The Luria Bertani (LB) medium was prepared by mixing 10 g of tryptone, 5 g of yeast extract, 10 g of sodium chloride (NaCl) and distilled water in a total volume of 1 L. When necessary, bacteriological agar powder (15 g/L) was added for agar plate preparation. The LB medium was autoclaved at 15 psi, 121 °C for 15 mins and kept at 4°C. Ampicillin (100 µg/mL) and 1 M isopropylthio-β-D-galactoside (IPTG) were dissolved in distilled water and filter sterilised by using a 0.22 µm syringe filter. 5-bromo-4-chloro-3-indolyl β-D-galactopyranoside (X-gal, 20mg/mL) was dissolved in dimethylformamide solvent. Ampicillin, IPTG and X-gal stock solution were stored at -20 °C.

3.5.4 *Escherichia coli* DH 5α competence cells preparation

Single colony of *E. coli* DH 5α was inoculated into 5 mL LB broth and grown overnight at 250 rpm, 37 °C. On the next day, 1 mL of the bacterial culture was transferred into 10 mL of fresh LB broth, incubated at 250 rpm, 37 °C for 2 hours. After that, the bacterial culture was centrifuged at 1000 ×g at 4 °C for 5 mins. The bacterial pellet was

then resuspended with 5 mL of ice-cold 75 mM calcium chloride solution ($CaCl_2$), followed by incubation on ice for 20 mins. The $CaCl_2$ treated bacterial cells were centrifuged at 1000 ×g at 4°C for 5 mins and resuspended with 1 mL of ice-cold 75 mM $CaCl_2$. Subsequently, 50 % (v/v) sterile glycerol solution was added to the mixture to a final concentration of 15 % (v/v). The bacteria cells were mixed by pipetting and aliquot into a sterile microcentrifuge tube.

3.5.5 Purification, ligation and transformation of PCR amplified fruit bromelain transcripts

The amplified PCR product was purified using QIAquick PCR Purification Kit (Qiagen, Germany) following the manufacturer's instruction. After that, the purified fruit bromelain transcripts were inserted into pGEM-T Easy vector (Promega, US). The ligation mixture was prepared as shown in Table 3.3 and incubated overnight at 4 °C.

Table 3.3　　　　Reagents composition of the ligation mixture

Reagents (µL)	Volume (µL)
2X Rapid Ligation Buffer, T4 DNA Ligase	5
pGEM®-T Easy Vector	1
Control Insert DNA	2
T4 DNA Ligase (3 Weiss units/µl)	1
Deionized water	1
Total volume	10

Transformation was performed by adding 2 µL of ligation mixture into prepared 50 µL of *E. coli* competent cells. The reaction mixture was mixed by flicking the tube a few times and placed on ice for 30 mins. Subsequently, the bacterial mixture was heat shocked at 42 °C in a water bath for 30 sec followed by ice incubation for 5 mins. Next, 950 µl of LB broth was added to the heat-shocked competent cells, incubated at 37 °C with continuous shaking at 150 rpm for 1 hour. This was followed by spreading 25 µL, 50 µL and 100 µL of transformed *E. coli* cells on fresh LB agar plate containing 100µg/mL of ampicillin and 20 mg/mL of IPTG. Subsequently, the transformation agar plate was incubated at 37 °C overnight.

After overnight incubation, colony PCR was conducted to identify the successful transformant harbouring the recombinant plasmid. White bacteria colony was picked randomly to be used as the DNA template. The colony PCR was performed using TopTaq

Mastermix Kit as described in section 3.4.2. The PCR products were then electrophoresed on 1.0 % (w/v) agarose gel.

3.5.6   Plasmid extraction

Successful transformants were selected and grown in 5 mL of LB ampicillin broth (100 µg/mL of ampicillin) at 37°C with continuous shaking at 150 rpm overnight. The next day, the overnight culture was subject to plasmid extraction using QIAprep Spin Miniprep Kit (Qiagen, Germany) following the manufacturer's instruction. The extracted plasmids were subjected to the single pass sequencing service provided by Apical Scientific Sdn Bhd.

3.5.7   Data mining

The fruit bromelain sequences were edited using BioEdit. The vector sequences were removed and fruit bromelain DNA sequences were translated into their respective amino acid sequences. Both DNA and amino acid sequences of fruit bromelain was aligned to their original reference sequences using ClustalW. The fruit bromelain amino acid sequence was further analysed BLAST; the presence of signal peptide in the fruit bromelain sequences was examined using SignalP 5.0; domain analysis via InterProScan and the physicochemical properties of fruit bromelain was investigated using.

3.6   Model development and evaluation

The fruit bromelain amino acid sequences of FB_1, FB_2 and FB_3 were BLAST against the PDB database to find suitable templates. Single and multiple templates selected for modelling were based on length of coverage, sequence identity and gaps between target and template. The fruit bromelain models were generated using MODELLER. From a hundred models generated, the best model was chosen based on the lowest discrete optimized protein energy (DOPE) score which was then subjected to model refinement via molecular dynamics simulation.

The refinement method was adopted from Heo and Feig (2018b) with some modifications. First, the local stereochemistry of the selected fruit bromelain models was refined by using locPREFMD server prior molecular dynamics simulation using GROMACS. The simulation was conducted in a simple cubic box with at least 10 Å from

the box edge with CHARMM36m force field (Huang et al., 2016). The system was solvated with TIP3P water molecules and neutralised by an appropriate number of sodium ions. After that, the ensemble was energy minimised by 5000 steps of the steepest descent algorithm and equilibrated to 298 K at NVT phase (constant Number of particles, Volume and Temperature) and 1 bar pressure at NPT phase (constant Number of particles, Pressure and Temperature) for 100 ps respectively. Finally, the simulation was performed at 298 K and 1 bar pressure for 50 ns. LINear Constraint Solver (LINCS) was used to constrain the bond length while electrostatic interactions were evaluated by the particle mesh Ewald method (PME) with a 12 Å cut-off for both coulomb and van der Walls interactions. The integration time step was 2 fs and a snapshot of the structures was recorded every 1 ps during the simulation. The stability of the trajectory was evaluated by root mean square deviation (RMSD) and radius of gyration (Rg) using the GROMACS utility. The stabilised structure was extracted and submitted to locPREFMD for the final round of local stereochemistry refinement.

The stereochemical quality of the fruit bromelain models were evaluated by Verify3D to determine the number of residues in an atomic model that is compatible with its own amino acid sequence with a passing score of 80 % (Bowie, Ltcy, & Eisenberg, 1991), ERRAT to analyse the statistics of non-bonded interactions between different atom types (Colovos & Yeates, 1993) and PROCHECK to evaluate the stereochemical quality of a protein structure in favoured, allowed and outlier regions (Laskowski, MacArthur, Moss, & Thornton, 2012) through SAVES 5.0 server. The energy level of fruit bromelain models was calculated via Swiss-PdbViewe. The tertiary structure of fruit bromelain was illustrated via PyMOL.

3.7 Thermostability of fruit bromelain

The thermal profile of fruit bromelain including its optimum temperature and thermostability was generated using enzymatic assay while the structural stability of fruit bromelain at different temperatures was investigated via molecular dynamics simulation.

3.7.1 Thermal profile using enzymatic analysis

The crude fruit bromelain sample was prepared as described in section 3.3.7. To determine the optimum temperature of fruit bromelain, the crude enzyme was allowed to hydrolyse casein at 40 °C, 50 °C, 60 °C, 70 °C and 80 °C, respectively for 10 mins. The

reaction was terminated by TCA and the reaction mixture was centrifuged. After that, the supernatant was mixed with Na_2CO_3 and F-C phenol reagent for colour development. On the other hand, for fruit bromelain thermostability study, the crude enzyme was incubated at five different temperatures of 40 °C, 50 °C, 60 °C, 70 °C and 80 °C without substrate for 1 hour. After that, the hydrolysis was conducted at 37 °C in a similar manner as described above. A plot of fruit bromelain activity against different temperatures was generated.

3.7.2 Dynamics behaviour and conformational study using molecular dynamics simulation

The refined fruit bromelain models were assigned to molecular dynamics simulation to observe dynamics and conformation changes at the selected temperature. In order to study the behaviour of the mature protein under the influence of temperature, the pro-peptide of fruit bromelain was removed prior the simulation. Three different systems were set up for each fruit bromelain model, solvated with TIP3P water molecules and neutralised by an appropriate number of sodium ions. After that, the ensemble was energy minimised by 5000 steps of steepest descent algorithm and equilibrated at 313 K, 333 K and 353 K via NVT and NPT, respectively. Subsequently, the simulation was conducted for 100 ns. LINCS was used to constrain the bond length while electrostatic interactions were evaluated by PME with a 12 Å cut-off for both coulomb and van der Walls interactions. The integration time step was 2 fs and snapshot of the structures was recorded every 1 ps during the simulation. The dynamic behaviour and conformational changes of fruit bromelain was studied by analysing RMSD, root mean square fluctuation (RMSF), Rg, solvent accessible surface area (SASA) and number of hydrogen bonds.

CHAPTER 4

RESULTS AND DISCUSSION

4.1 Fruit bromelain gene expression and proteolytic activity analysis

In this section, the total RNA and fruit bromelain was extracted from ripe and unripe *A. comosus* cultivar MD 2. The concentration and integrity of the isolated total RNA were quantified and examined. This is followed by relative gene expression analysis of fruit bromelain between ripe and unripe *A. comosus* cultivar MD 2. Subsequently, the proteolytic activity of fruit bromelain was determined by enzymatic assay using casein as a substrate.

4.1.1 Total RNA extraction from unripe and ripe *A. comosus* cultivar MD 2

Total RNA was extracted from three samples of unripe and ripe *A. comosus* cultivar MD 2, respectively. The extracted RNA was evaluated based on their yield, purity and integrity. The concentration of total RNA from extracted unripe *A. comosus* cultivar MD 2 are 82.3 ng/µL, 77.6 ng/µL and 88.3 ng/µL respectively (Table 4.1). In contrast, the concentrations of the total RNA obtained from ripe *A. comosus* cultivar MD 2 were lower, which were only 45.6 ng/µL, 47.6 ng/µL and 50.2 ng/µL respectively. Apparently, the RNA concentrations from unripe *A. comosus* cultivar MD 2 were much higher compared to the ripe *A. comosus* cultivar MD 2. This always occurs in mature plant tissues due to the increase of secondary metabolites and elevation of RNase activity during the ripening process (Brasil, Lourdes, Otoch, & Costa, 2008). All the isolated total RNA samples have A_{260}/A_{280} ratio of 1.9-2.2 indicating good RNA purity (Taylor et al., 2010).

Table 4.1   Concentration and purity readings of the isolated total RNA

Sample	Concentration (ng/µL)	A_{260}/A_{280}
Unripe 1	82.3	1.92
Unripe 2	77.6	1.97
Unripe 3	88.3	2.10
Ripe 1	45.6	1.90
Ripe 2	47.6	1.95
Ripe 3	50.2	2.21

Furthermore, the integrity of RNA was examined based on the intactness and intensity of the 28S and 18S ribosomal RNA (rRNA) on the agarose gel. Despite that the sample Unripe 2 has a higher concentration of 77.6 ng/µL, the intensity of 28S and 18S rRNA is much lower compared to Unripe 1 and Unripe 3. This is because the measured concentration is the sum of mRNA, microRNA (miRNA), small nuclear RNA (snRNA), tRNA-derived sRNAs (tsRNA), small nucleolar RNAs (snoRNA) and small interfering RNAs (siRNAs) instead of rRNA alone (Wu et al., 2014). Thus, Figure 4.1 demonstrates that Unripe 2 has a lower concentration of rRNA. On top of that, the 28S and 18S rRNA were present as intact bands and no smearing was observed. Thus, the extracted total RNA can be used for subsequent downstream applications such as cDNA synthesis (Brasil et al., 2008).

Figure 4.1   Electrophoresis of RNA on 1.0 % (w/v) agarose gel. Lane 1-3 represent the RNA isolated from unripe samples whereas lane 4-6 represent the RNA isolated from ripe samples of *A. comosus* cultivar MD 2, respectively. The electrophoresis was conducted at 110 volts for 40 mins.

4.1.2   qPCR analysis of fruit bromelain from ripe and unripe *A. comosus* cultivar MD 2

The qPCR allows the comparison of the gene transcript level of fruit bromelain between unripe and ripe in *A. comosus* cultivar MD 2. In contrast to conventional PCR in which the PCR product is analysed at the end of the PCR process, qPCR monitors and quantifies the accumulation of amplicons as the PCR cycling progresses. This is achieved by measuring the number of cycles which is needed to exceed the fluorescent background. This threshold detection level is termed as quantification cycle (C_q). The C_q value is correlates with the amount of target present in the initial experiment. A sample with a higher amount of target gene will have a low C_q value because it can produce a quantifiable signal in the early cycling stage. On the contrary, a sample with a lower amount of target gene will have a higher C_q value because more amplification cycles are needed to produce a sufficient detectable fluorescent signal (Wong & Medrano, 2005).

In this study, actin was used as a reference gene to normalise variations in RNA extraction, reverse-transcription as well as the PCR amplification efficiency to allow comparisons of the gene transcript level across different samples (Bustin et al., 2009). This is because actin has been reported with a good consistency as a reference gene by several previous studies (Koia et al., 2012; Raimbault et al., 2013). Designing and testing the primers that is suitable to for qPCR analysis is generally time consuming and involves laborious work. Thus, Bustin et al. (2010) suggested the utilization of primer sequences that has been successfully used in previous studies. In the present study, the primers used were similar to that reported by Raimbault et al. (2013). According to them, the primers for fruit bromelain is designed to amplify a consensus region between fruit bromelain isoforms. This is due to the limitation in designing qPCR primers that are able to distinguish each fruit bromelain isoforms due to their high similarity. Moreover, the primers for actin were designed based on actin sequence (contig_42) from PineappleDB: The Online Pineapple Bioinformatics Resource.

The optimum annealing temperature for each fruit bromelain and actin primers was determined by running a gradient analysis at several temperatures ranging from 55-65 °C. The C_q value of fruit bromelain did not show a significant difference at 55 °C, 57 °C and 60 °C which are 28.03 ± 0.04, 27.31 ± 0.30, 27.73 ± 0.19, respectively (Figure

4.2). At 62 °C, the C_q value went slightly higher to 30.71 ± 0.03. The annealing temperature analysis for actin showed a similar pattern to fruit bromelain where the resulting Cq values are 24.24 ± 0.37, 24.26 ± 0.09, 25.23 ± 0.25, 28.73 ± 0.06 at 55 °C, 57 °C, 60 °C, and 62 °C, respectively (Figure 4.2). It was observed that no C_q value was recorded for both fruit bromelain and actin at 65°C which indicated that this temperature is too high for amplification to occur. The optimum annealing temperature for fruit bromelain and actin were identified at 57 °C and was used as the annealing temperature in the subsequent qPCR cycling.
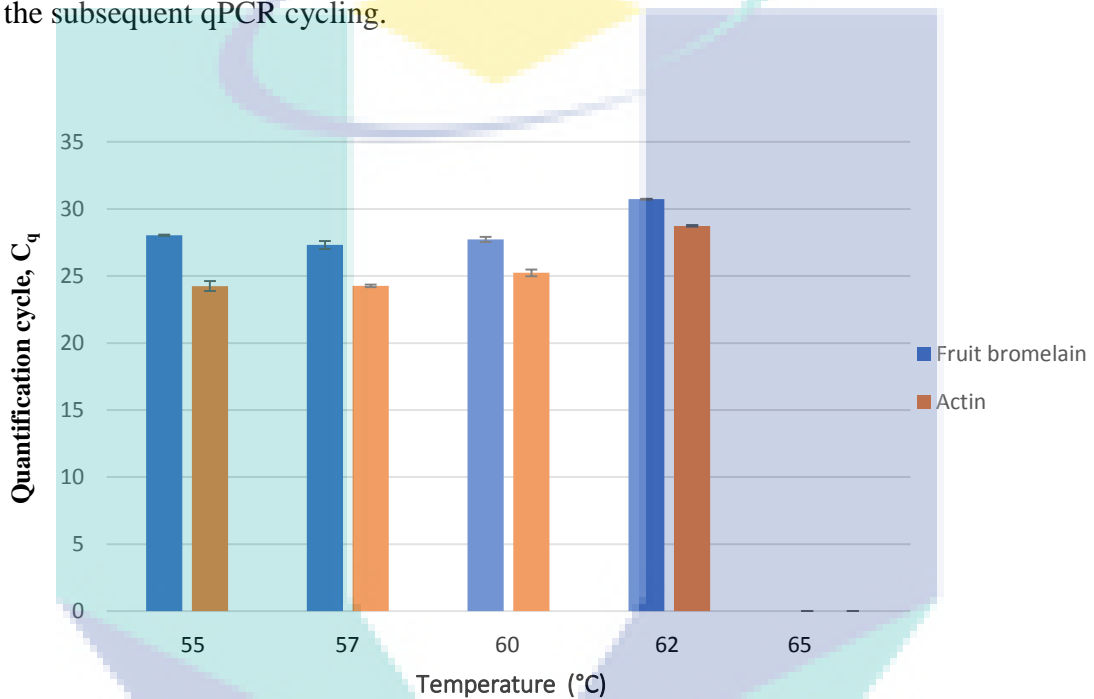
Figure 4.2   Determination of optimum annealing temperature for fruit bromelain primers and actin primers.

Subsequently, PCR efficiency was evaluated to ensure the robustness and preciseness of qPCR (Bustin et al., 2009). According to Taylor et al. (2010), a PCR efficiency of 90-110 % is acceptable. In this study, the PCR efficiency for both fruit bromelain and actin were evaluated in a serial of 5-fold dilution. A standard curve for each dilution was generated. The linear equation of fruit bromelain is y = -3.162 + 18.41 while actin is y = -3.131 + 21.433 (Appendix A). Meanwhile, the coefficient of determination (R^2) for fruit bromelain and actin are 0.9863 and 0.9844, respectively. Since the experiment was run in triplicates, a higher R^2 indicates a good result with high consistency and correlation without significant pipetting error between each replicate

(Mukaka, 2012). By using Eq 4.1, the calculated PCR efficiency for fruit bromelain and actin were 107.14 % and 108.6 %, respectively which fell within the acceptable range.

Melting curve analysis was used to determine the specificity of the PCR products for both fruit bromelain and actin to avoid a false-positive result that can be caused by primer-dimers (Hui & Feng, 2013). The PCR products were heated with an elevated temperature. At a certain temperature, the PCR products were dissociated and lead to the decrease of signal due to the release of a fluorescent dye. This sudden decrease of the signal will result in a peak formation in the melting curve. Therefore, the number of peaks formed in the melting curve can be used to determine the number of PCR bands and the presence of primer-dimers in the reaction. It can be seen that the melting curve of fruit bromelain and actin showed a single peak at ~80 °C and ~83 °C respectively (Figure 4.3). The actin (200 bp) has a higher melting temperature due to its larger product size than fruit bromelain (170 bp). The presence of the single peak in the melting curve analysis indicates that both fruit bromelain and actin primers are highly specific and the PCR condition is well-optimised.
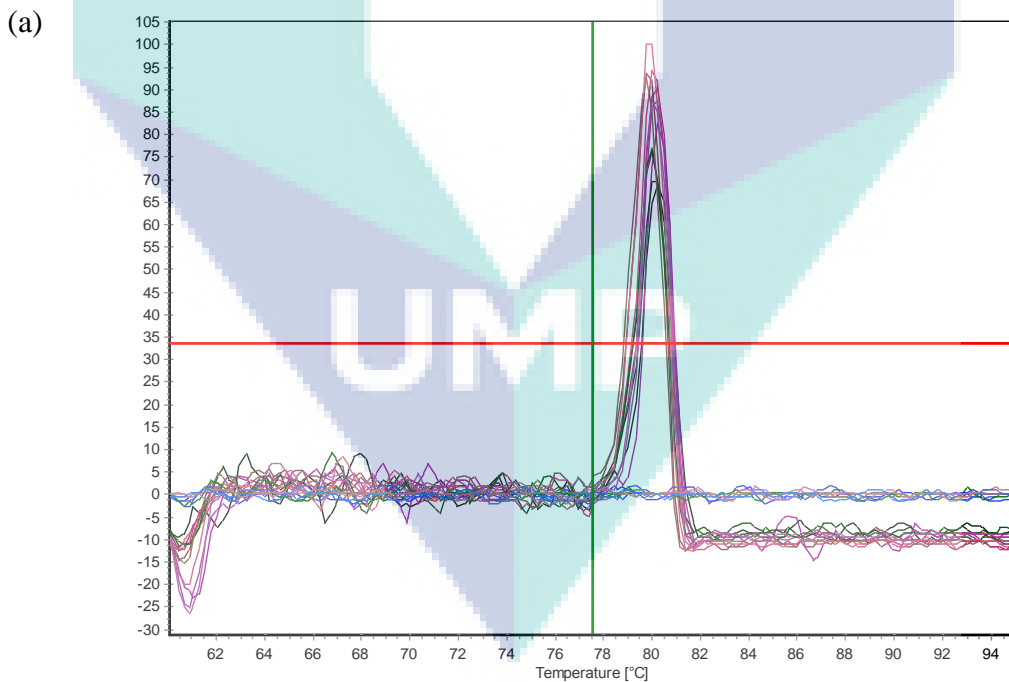
Figure 4.3   Melting curve analysis of (a) fruit bromelain (b) actin. A single peak was observed indicated the PCR products are highly specific without the formation of primer-dimers.

(b)
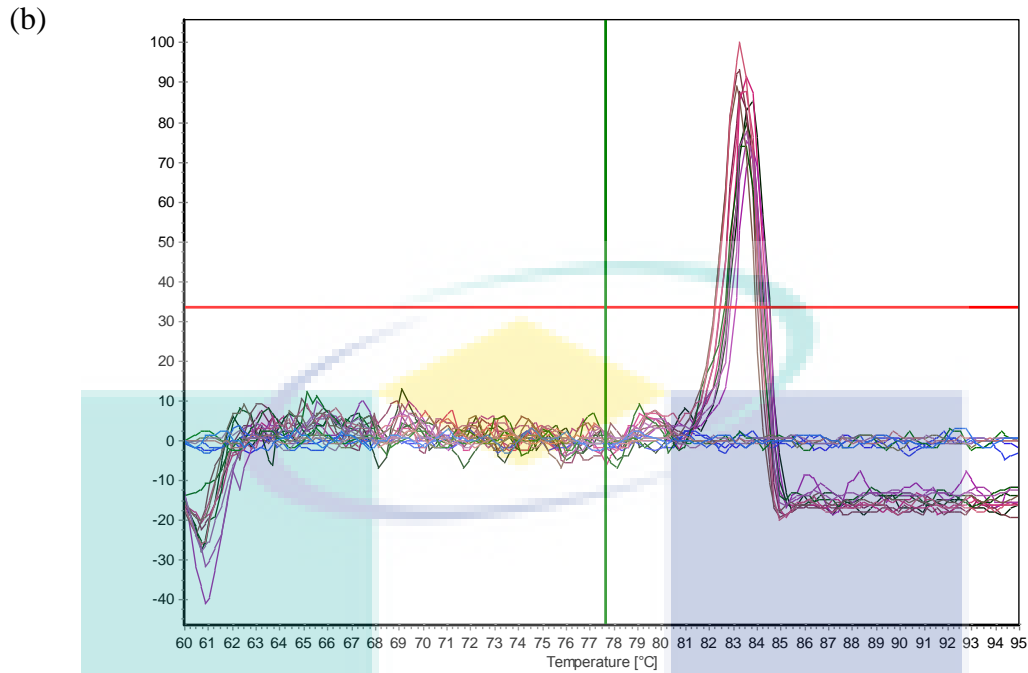
Figure 4.3   Continued.

In this study, the $2^{-\Delta\Delta Cq}$ method was used to investigate the relative changes of fruit bromelain expression in different ripening stages. Unripe *A. comosus* cultivar MD 2 was selected as a calibrator to contrast the fruit bromelain transcript level in ripe *A. comosus* cultivar MD 2. The obtained C_q values were computed in the $2^{-\Delta\Delta Cq}$ method and found that the expression fruit bromelain was 10-fold down-regulated in ripe *A. comosus* cultivar MD 2 (Figure 4.4). This result was affirmed with an early study in which fruit bromelain was found to be down-regulated in microarray and Northern analysis as the Smooth Cayenne pineapple fruits ripen (Koia et al., 2012; Moyle et al., 2005). Indeed, a higher gene expression at a young fruit stage is a unique feature in the CA1 family (Butts et al., 2016). A similar expression pattern was observed in the actinidain of kiwifruit. Actinidain was reported with a higher mRNA expression level before the kiwifruits ripen (Nieuwenhuizen et al., 2007).
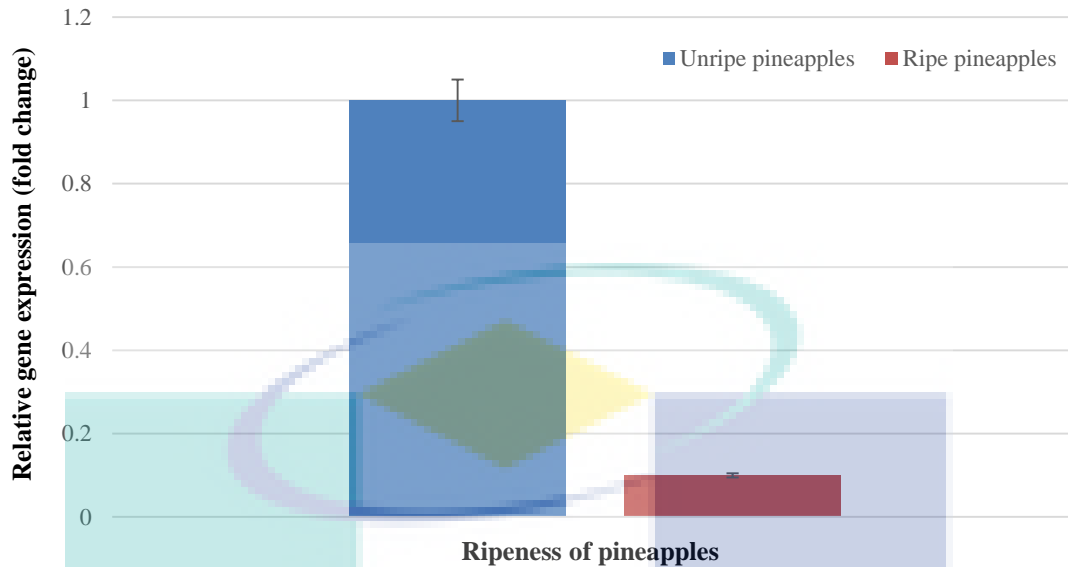
Figure 4.4   Relative gene expression of fruit bromelain in unripe and ripe *A. comosus* cultivar MD 2 normalised by actin.

The specific role of fruit bromelain in pineapple fruit is undefined (Martinez et al., 2012). However, higher fruit bromelain expression in unripe *A. comosus* cultivar MD 2 suggests that fruit bromelain is required for pineapple fruit ripening, e.g. to act as a defence protein (Misas-villamil et al., 2016; Nieuwenhuizen et al., 2012). It was found that the vacuolar cysteine protease and papain which have a typical CA1 protease expression pattern contribute to pathogen resistance in tomato and papaya fruits, respectively (Malek et al., 2016; Wang et al., 2017). Since pineapple is vulnerable to several fungi, nematodes, bacteria and virus associated diseases, higher fruit bromelain expression is postulated to protect the unripe pineapple fruit during its development in a similar manner (Dawson, 2016; Rohrbach & Johnson, 2003). This hypothesis is supported by the evidence of overexpression of fruit bromelain which results in an increase of pest resistance in transgenic *A. thaliana* and *B. rapa* (Jung et al., 2008; Wang et al., 2014).

4.1.3   Enzymatic analysis of **fruit bromelain from ripe and unripe *A. comosus* cultivar MD 2**

In order to relate the fruit bromelain transcript level detected by qPCR, fruit bromelain from ripe and unripe samples of *A. comosus* cultivar MD2 were extracted and used in casein enzymatic assay. According to Cupp-enyard and Sigma-Aldrich (2008), tyrosine is liberated when casein is digested. The interaction between tyrosine and F-C phenol reagent produces a quantifiable blue chromophore and the intensity of the blue

chromophore is directly proportional to the amount of the tyrosine. Thus, protease with higher enzymatic activity is always associated with higher chromophore intensity since a large amount of tyrosine is produced. To quantify the enzymatic activity of the crude fruit bromelain, a tyrosine standard curve was generated which is expressed in a linear equation $y = 1.511x + 0.0178$ with R^2 value of 0.9997 (Appendix B).

The determined proteolytic activity of the fruit bromelain is 1.91 ± 0.08 U/mL for unripe *A. comosus* cultivar MD 2 and 1.13 ± 0.09 U/mL for ripe *A.comosus* cultivar MD 2 (Figure 4.5). This result demonstrated that unripe *A. comosus* cultivar MD 2 have higher enzymatic activity than unripe *A.comosus* cultivar MD 2. This is in line with a previous study in which reduced proteolytic activity was observed in matured pineapple fruits (Bresolin et al., 2013). A similar observation was observed in ficain whereby the specific activity of ficain decreased upon the ripening of fig fruits (Raskovic, Lazic, & Polovic, 2016). Moreover, actinidain and papain were also found to be more active in greener fruits (Afshar-Mohammadian et al., 2010; Amri & Mamboya, 2012; Sharma & Chatterjee, 2017). It was conjectured that higher proteolytic activity is needed to produce an undesirable sensation on tongue tissues to avoid unripe pineapple fruits from animal consumption as well as to enhance anti-pathogen properties (George et al., 2014; Raskovic et al., 2016; Srujana & Narayana, 2017).
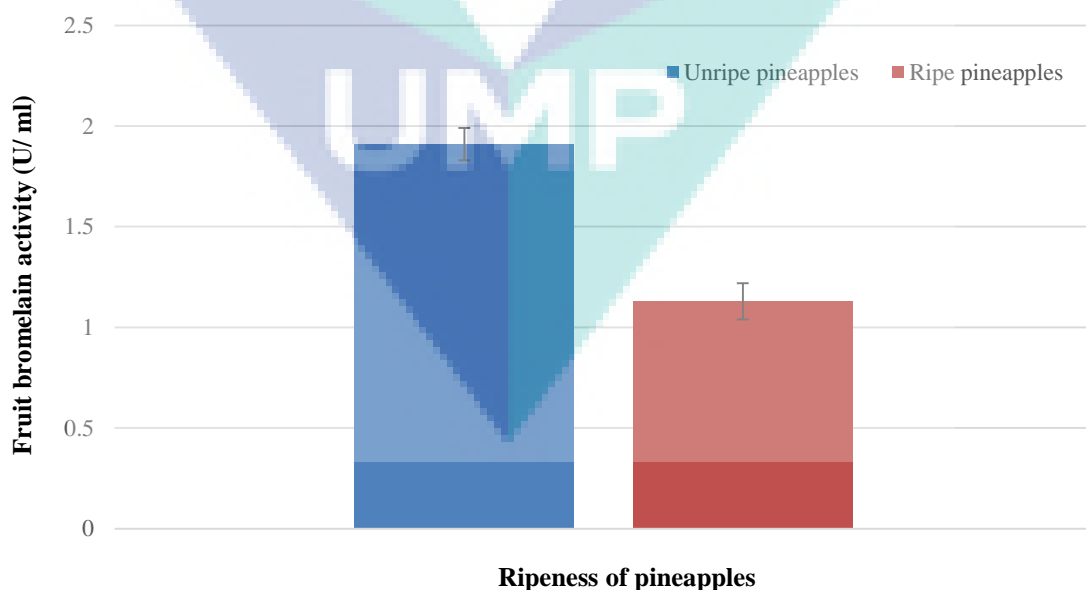
Figure 4.5   Enzymatic activity of the crude fruit bromelain from unripe and ripe *A.comosus* cultivar MD 2.

In conclusion, in the present study, gene expression level and the proteolytic activity of fruit bromelain in ripe and unripe *A. comosus* cultivar were characterised. The results showed that the expression level and proteolytic activity of fruit bromelain is higher in the unripe stage during ripening of pineapple fruits. Based on this result, we speculate that fruit bromelain has an important protection role against phytopathogens including *Cyanophora paradoxa*, *Diaspis bromeliea* and mealybugs during pineapple fruit development.

4.2   Sequence analysis and isolation of fruit bromelain transcripts

In this section, fruit bromelain sequences of *A. comosus* cultivar MD 2 were retrieved from the NCBI genbank. These sequences were analysed based on length and were aligned with a reference sequence retrieved from the MEROPS database. Specific primers were designed to isolate the selected fruit bromelain sequences for further study using several bioinformatic tools.

4.2.1   Sequence analysis

A total of 17 *A. comosus* cultivar MD 2 fruit bromelain protein sequences with lengths ranging from 150 to 1924 amino acids were retrieved from the NCBI genbank (Figure 4.6). Fruit bromelain usually have a read length between 300-350 amino acid residues, containing cathepsin pro-peptide inhibitor (I29) and papain-like cysteine endopeptidase (PLCE) domains (Ramli et al., 2018). The I29 and PLCE domains form the pro-peptide and catalytic region of fruit bromelain, respectively. The retrieved fruit bromelain sequences were compared with fruit bromelain reference sequence (Accession number: O23791) retrieved from the MEROPS database which is an online database that stores information of proteases (Rawlings et al., 2018). Among the retrieved sequences, OAY67114.1, OAY68387.1, OAY80099.1 and OAY85828.1 have incomplete sequence information at N-terminal and C-terminal due to their short length. In contrast, OAY76881.1, OAY80102.1, OAY83410.1, OAY85856.1 and OAY85857.1 contain more than one fruit bromelain protein sequences. In addition, OAY85826.1 shares 100% sequence identity with BAA21848.1 which has been investigated in a previous study (Ramli et al., 2018). These sequences were excluded for further analysis.
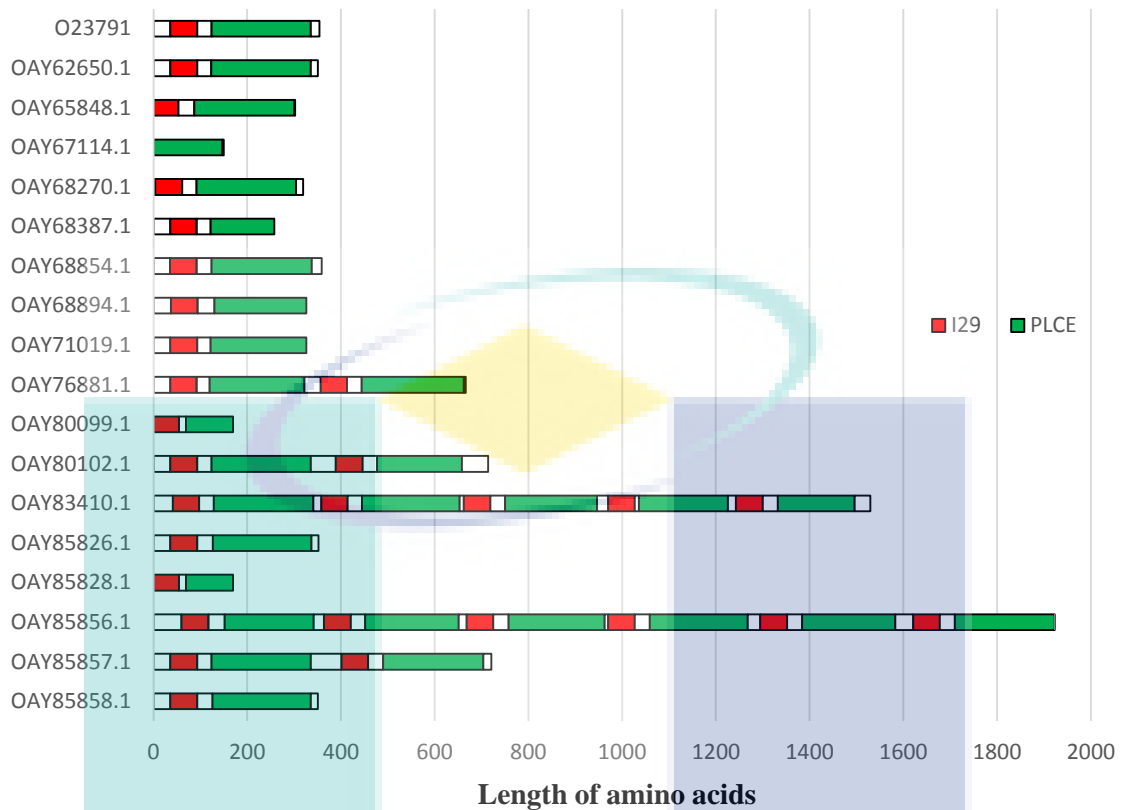
Figure 4.6   Domain organisation of fruit bromelain retrieved from NCBI genbank. The I29 and PLCE domains are labelled in the red and green box, respectively. The I29 and PLCE domains form the pro-peptide and catalytic region of fruit bromelain, respectively.

The remaining fruit bromelain sequences were aligned fruit bromelain reference sequence (Accession number: O23791). According to the MEROPS database, the first 24 amino acid residues of O23791 amino acids are signal peptide, followed by 97 and 240 amino acid residues of pro-peptide and the mature protein, respectively. The signal peptide is a small part of protein located at the N-terminus which plays a role in the protein secretory pathway to ensure the protein enters the endoplasmic reticulum membrane and is cut off by signal peptidase I upon protein maturation (Armenteros et al., 2019; Zou, Huang, Xie, & Yang, 2018). Among these sequences, OAY68270.1 does not contain signal peptide, suggesting that it may have an alternate route to bypass the Golgi apparatus (Bellucci, De Marchis, & Pompa, 2018) (Figure 4.7). Furthermore, OAY65848.1, OAY68894.1 and OAY71019.1 have incomplete N-terminal or C-terminal end and were eliminated for subsequent analysis. In addition, OAY68854.1 was also removed from this study because it has a dissimilar C-terminal end in comparison to other fruit bromelain sequences. Hence, fruit bromelain sequences OAY62650.1, OAY68270.1 and OAY85858.1 were selected for further analysis in the subsequent sections.

```
                10        20        30        40        50        60        70        80
          ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
O23791    MASKVQLVFLFLFLCAMWASPSAASRDEPND-PMMKRFEEWMAEYGRVYKDDDEKMRRFQIFKNNVKHIETFNSRNENSY
OAY62650.1 MASKFQLVFLFLFLCMMWASPSAASRGEPSD-PMMKTFEAWMAQYGRVYSDDNEKMRRFQIFKNNVNYIETFNNGSRNSY
OAY65848.1 ----------------------------------------MAKYNRTYRDEGEKWRRFEIFKENVGFIDAFNQGGVQSY
OAY68270.1 ----------------------------MMKRFEEWMAEYGRVYKDNDEKMRRFQIFKNNVNHIETFNSRNGNSY
OAY68854.1 MASKFQILFLFLFLSVMWAS-SLASRGEADD-SMMKRFEEWMADFGRVYSDDAEKMRRFQIFKDNVNRIEAFNRRGGNSY
OAY68894.1 MALTFQFLLTGSSLILLCFSFWFSSWNTRIDGPGMTMFEYWMSEHGRVYANAKEKQKRYEIFENNVKYIEGFNKVGGRSY
OAY71019.1 MAYSFHYIILIISAVALHWPQHAASRGEPHW-SMAKRHEEWAAEHGRVYGDADEKQRRFEIFSDNVRYIDSFNMERKYNY
OAY85858.1 MASKVQLVFLFLFLCVMWASPSAASRDEPSD-PMMKRFEEWMAEYGRVYKDNDEKMRRFQIFKNNVNHIETFNNRNGNSY

                90       100       110       120       130       140       150       160
          ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
O23791    TLGINQFTDMTKSEFVAQYTGVSLPLNIE-REPVVS--FDDVNISAVPQSIDWRDYGAVNEVKNQNPCGSCWSFAAIATV
OAY62650.1 TLGTNQFTDLTNNEFIAQHTG-ALPLNIK-REPVVL--FDDVNISAVPQSIDWRYYGAVTPIKDQGSCGSCWAFSAIATV
OAY65848.1 TLAVNQFADLTNKEFVATYTG-AKPSNSSRPSPPSPPPMRYASPRGPPPSIDWRERGAVTDVKYQGPCGSCWAFATVAAI
OAY68270.1 TLGINQFTDMTKSEFVAQYTGVSLPLNIE-REPVVS--FDDVNISAVPQSIDWRDYGAVNEVKNQNPCGSCWSFAAIATV
OAY68854.1 TLGINQFTDMTNNEIVAQHVGLSLPLNMTNLEPSVS--FEDVNMSAIPQSIDWRDYGAVTPVKNQGSCGSCWAFSSIATV
OAY68894.1 TLGVNQFSDLTNEEFTNTYAVGIAEQDIPTDIETAPMDSEYEN-AALRPSVDWR-----------TEGCCWAFSTVATV
OAY71019.1 KLGINQFADMTNEEFVATHTG-SRPRNGP--RSMTE--FQYARVSDLPSSIDWR-----------ITGSCWAFSAIAAV
OAY85858.1 TLGINKFTDMTNNEFVAQYTGVSLPLNFK-REPVVS--FDDVNISAVGQSIDWRDYGAVTEVKDQNPCGSCWAFSAIATV

               170       180       190       200       210       220       230       240
          ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
O23791    EGIYKIKTGYLVSLSEQEVLDCAVS-YGCKGGWVNKAYDFIISNNGVTTEENYPYLAYQGTCNANSFP--NSAYITGYSY
OAY62650.1 EGIYKIKTGFLLSLSEQKVLDCAVS-NGCNGGQVNKAYDFIISNNGVTSTVFYPYKGNQGTCAANRVP--NSAYITGYSY
OAY65848.1 EGIFKIKKKQLISLSEQELVDCVAANFGCGGGFLDRAYNFVVLNQGLNTEHMYPYTAVQSSCDFNKMAY-KAASVSNFHY
OAY68270.1 EGIYKIKTGYLVSLSEQEVLDCAVS-YGCKGGWVNKAYDFIISNNGVTTEENYPYQAYQGTCNANSFP--NSAYITGYSY
OAY68854.1 EGIYKIKTGQLISLSEQEVLDCTVS-NGCTSGWVHKAYEFIIANKGVTAQANYPYVGYKGTCAANSKP--NAAYITGYQQ
OAY68894.1 ESIYKIKKGQLKSLSEQEVLDCSNG-GDCTGDSRYEAFDFIVANQGLTTEANYPYTASKGACDQTKLP-DHAAYITGYRM
OAY71019.1 EGIVQIKTGQLISLSEQELVDCDYTDNGCNWGWVDKTFDFIARFGGVSSEAEYPFTGAKDNCNITKLSQNKAASIGGYEY
OAY85858.1 EGIYKIVTGYLVSLSEQEVLDCAVS-NGCDGGFVDNAYDFIISNNGVASEADYPYQAYQGDCAANSWP--NSAYITGYSY

               250       260       270       280       290       300       310       320
          ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
O23791    VR-RNDERSMMYAVSNQPIAALIDAS-ENFQYYNGGVFSGPCGTSLNHAITIIGYGQDSS-GTKYWIVRNSWGSSWGEGG
OAY62650.1 VP-RNDERSMMYAASNQPIAALIDASGNNFRSYQGGVFSGPCGTSLDHVITIIGYGQDIS-GTKYWIVKNSWGMSWGEGG
OAY65848.1 VP-KNDERALKKAVANQPVSVYVEAVGSFFQFYSGGVFKGPCGTAHNHAIAIVGYGEDNT-GTKYWIGKNSWGSDWGDHG
OAY68270.1 VR-RNDERSMMYAVSNQPIAALIDAS-ENFQYYNGGVFSGPCGTSLNHAITIIGYGQDSS-GTKYWIVRNSWGSSWGEGG
OAY68854.1 VQPSYNERAIMYAVANQPTVIAIDASSYYFNHYNGGIFKGPCGTNIFHAVTVVGYGQDSSTGNKYWIIKNSWGNTWGENG
OAY68894.1 VT-QNNEAELMKAVNDQPVAVAVNAR--PWQQYTGGIFDQDCSPDVTHAVVVVGYGEESG--KKYWIILKNSYGPNWGESG
OAY71019.1 VP-SNDETMLMKAVANQPVSVNIDGSGSPFQHYSGGIFDGPCNTTMNHYVTVVGYGEDER-GTKYWIAKNSWGATWGENG
OAY85858.1 VR-SNDESSMKYAVWNQPIAAAIDASGDNFQYYNGGVFSGPCGTSLNHAITIIGYGQDSS-GTQYWIVKNSWGSSWGERG

               330       340       350       360
          ....|....|....|....|....|....|....|....|.....|..
O23791    YVRMARGVSSSSGVCGIAMAPLFPTLQSG-ANAEVIKMVSET-----
OAY62650.1 YIRMARDVSSSAGLCGIAMAPLFPTLRSA-ANAEVIKIVSKT-----
OAY65848.1 YILLERDIQAKEGLCGLAMDAVYPIIV---------------------
OAY68270.1 YVRMARGVSSSSGVCGIAMSPLFPTLQSG-ANAELIKMVSET-----
OAY68854.1 YVRMLRDTSYP-GLCGLASYGIYPTLVSSQRTTEPSDMGSDDRVSSM
OAY68894.1 YMRIAKDVAAKEGLCGIATAPLYPYI-----------------------
OAY71019.1 YIRFKKDVEAKEGVCGLAMYPSYPTF------------------------
OAY85858.1 YVRMARGVSSS-GLCGIAMDPLYPTLQSG-ANVAVIKMVSET-----
```

Figure 4.7   Sequence alignment of fruit bromelain. Dashes denoted gaps between sequences. The dissimilar C-terminal end of OAY68854.1 with other fruit bromelain sequences is indicated in red box.

4.2.2   Amplification of fruit bromelain transcripts

The single-stranded cDNA of unripe sample 3 was used as a template to amplify fruit bromelain transcripts since it has the highest total RNA $A_{260/280}$ purity of 2.10. The amplified PCR product was electrophoresed on agarose gel together with 1kb ladder. The 1kb ladder was used as a marker to estimate the size of PCR products. As shown in Figure 4.8, the 1kb ladder was separated into several fragments with sizes ranging from 250 bp to 10000 bp on the agarose gel. Besides that, the result indicated that the amplicon size of the amplified fruit transcripts of OAY62650.1, OAY68270.1 and OAY85858.1 were observed at approximately 1000 bp, which corresponds to their expected size of 1053 bp,

960 bp and 1053 bp, respectively. Single and intact bands were observed which indicated the specificity of the PCR products.
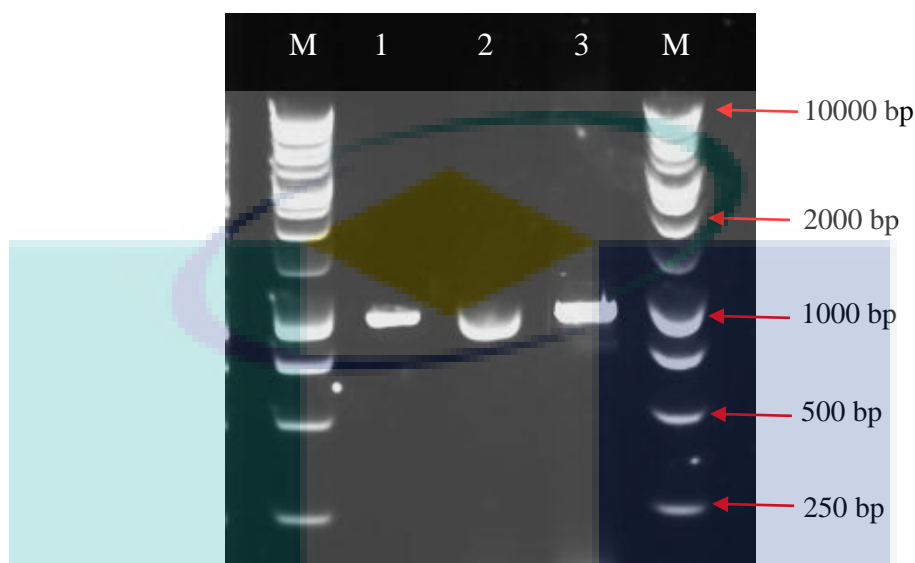
Figure 4.8　　DNA Electrophoresis of PCR products on 1.0 % (w/v) agarose gel. M represent 1kb ladder (Promega, US) whereas lane 1-3 represent the fruit bromelain transcripts of OAY62650.1, OAY68270.1 and OAY85858.1, respectively. The electrophoresis was conducted at 110 volts for 40 minutes.

4.2.3　Cloning of fruit bromelain transcript into pGEM-T Easy vectors

The amplified fruit bromelain transcripts were ligated into pGEM-T Easy Vectors via TA cloning. According to Yao, Hart, & An (2016), TA cloning is a simpler and more efficient cloning method in comparison to blunt-end and cohesive-end cloning. The DNA fragments of fruit bromelain of OAY62650.1, OAY68270.1 and OAY85858.1 amplified by *Taq* DNA polymerase contain an overhanging adenine sequence at 3'-terminal ends. Meanwhile, pGEM-T Easy vector is a linearised vector designed to carry 3´-terminal thymidine. This creates a compatible overhang to allow amplified fruit bromelain fragments to be inserted into the pGEM-T Easy Vectors. The ligated plasmids were then transformed into DH5α *E. coli* competent cells. The DH5α *E. coli* cells are endA and recA mutated, ensuring high stability and yield of the plasmids (Padmanabhan, Banerjee, & Mandi, 2011).

The pGEM-T Easy Vectors containing ampicillin-resistant gene, which only allow successful transformed bacteria cells to multiply in the presence of ampicillin. To

further confirm the transformed bacteria cells harbouring the fruit bromelain transcript, blue-white screening was performed. The pGEM-T Easy Vectors consist of α-peptide coding region of the enzyme β-galactosidase. The expression of lac Z gene is induced by IPTG, a non-metabolisable analog of galactose. The bacteria cells harbouring the self-annealed pGEM-T Easy vector produced β-galactosidase which hydrolysed X-gal into 5-bromo-4-chloro-indoxyl and spontaneously dimerised to form an insoluble blue pigment of 5,5'-dibromo-4,4'-dichloro-indigo (Padmanabhan et al., 2011). On the other hand, the recombinant vectors with disrupted lac Z operon due to insertion of the fruit bromelain transcript appeared as white colonies on the agar plate (Figure 4.9).
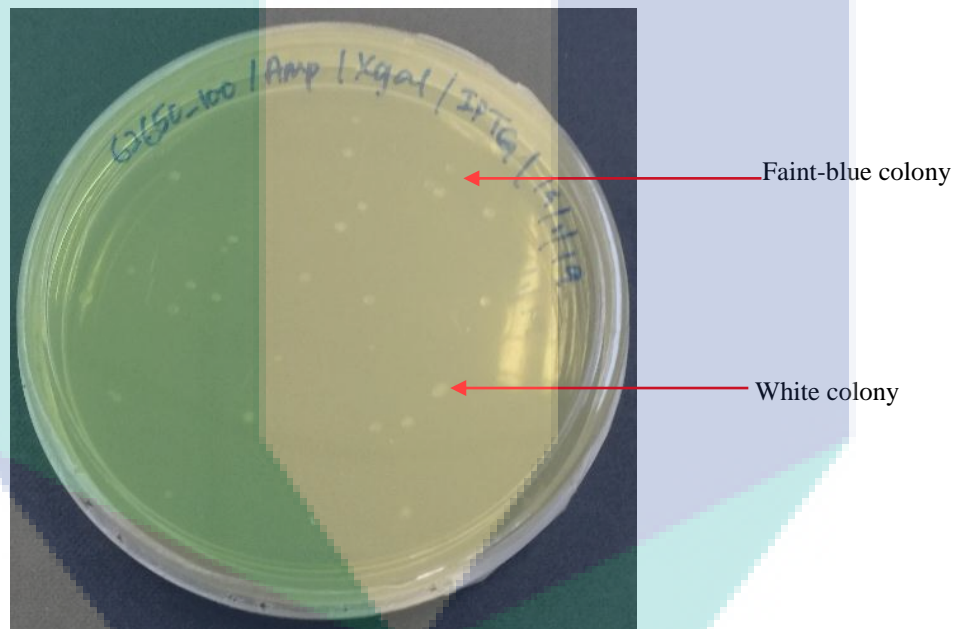
Figure 4.9   Blue-white screening of the transformed colonies. White colonies are successful transformants of fruit bromelain while false-positive colonies are showed in faint-blue.

However, in some cases, the non-recombinant may lose their blue colour caused by an instability of chemicals e.g. X-gal and IPTG (Banerjee, Kumar, Apte-Deshpande, & Padmanabhan, 2010). Hence, the successful transformants were further verified via colony PCR amplification. A total number of eight colonies were selected for colony PCR amplification for each recombinant harbouring OAY62650.1, OAY68270.1 and OAY85858.1 transcripts. All colonies were integrated with the inserts (Figure 4.10 (a)) except colony 4 and colony 8 for OAY62650.1 (Figure 4.10 (b)) as well as colony 2 for OAY68270.1 (Figure 4.10 (c)). The absence of the PCR band for these colonies may be

due to insufficient or excessive amounts of bacteria used in colony PCR. Hence, not enough template was provided for PCR amplification or inhibition caused by bacterial cell materials such as the cell membrane (Woodman, Savage, Arnold, & Stevenson, 2018).

Figure 4.10   Electrophoresis of colony PCR products on 1.0 % (w/v) agarose gel. M represent 1kb ladder (Promega, US). (a) Lane1-3 and Lane 5-7 showed positive amplification for insert OAY62650.1 (b) Lane 1-8 showed positive amplification for insert OAY68270.1 (c) Lane 1-8 showed positive amplification for insert OAY85858.1. The electrophoresis was conducted at 110 volts for 40 minutes.

4.2.4 Sequencing of fruit bromelain

The extracted plasmids containing fruit bromelain genes of OAY62650.1, OAY68270.1 and OAY85858.1 were subjected to DNA sequencing. The isolated fruit bromelain sequences were annotated as FB_1, FB_2 and FB_3, which correspond to their original reference sequences with accession no. of OAY62650.1, OAY68270.1 and OAY85858.1, respectively. Nucleotides substitution, deletion and insertion were observed in the isolated sequences (Appendix C). The alteration resulted in the substitution, insertion and deletion of amino acid residues.

FB_1 displayed amino acid substitution at 65 amino acid locations. During pairwise alignment with reference sequence of OAY62650.1, these substitutions were observed at alignment position 16 (M→A), 27 (G→D), 30 (S→N), 36 (T→R), 39 (A→E), 43 (Q→E), 49 (S→K), 51-52 (DN→ND), 66-67 (NY→KH), 73-76 (NGSR→SRNG), 83 (T→I), 89 (L→M), 91-92 (NN→KS), 95 (L→V), 98 (H→Y), 102 (A→S), 107 (K→E), 114 (L→S), 131 (Y→D), 136-138 (TPI→NEV), 140 (D→N), 142-143 (GS→NP), 151 (S→A), 166 (F→Y), 168 (L→V), 174 (K→E), 182 (N→Y), 185 (N→K), 188 (Q→W), 204-207 (STVF→TEEN), 212-213 (GN→AY), 218 (A→N), 221-222 (RV→SF), 235 (P→R), 246 (A→V), 260 (N→E), 263-264 (RS→QY), 266 (Q→N), 279 (D→N), 281 (V→A), 291 (I→S), 300 (K→R), 305 (M→S), 313 (I→V), 318 (D→G), 323 (A→S), 325 (L→V), 331 (A→S), 339 (R→Q) and 340 (A→G) including the insertion of valine at alignment position 101 and deletion glycine at alignment position 259 (Figure 4.11 (a)).

Meanwhile, FB_2 and FB_3 only displayed minor alterations compared to the original sequences of OAY68270.1 and OAY85858.1, respectively. Single amino acid substitution of serine to alanine was observed at alignment position 298 in FB_2 (Figure 4.11 (b)) whereas, FB_3 showed amino acid substitutions at alignment position 102 (V→I), 104 (L→R), 118-110 (FKR→IEK) and 314 (V→I) as well as the insertion of glycine at alignment position 101 (Figure 4.11 (c)). In addition, the length of the translated amino acid sequence of FB_1, FB_2 and FB_3 are 351, 319 and 352, respectively.

(a)

```
                  10        20        30        40        50        60
           ....|....|....|....|....|....|....|....|....|....|....|....|
OAY62650.1 MASKFQLVFLFLFLCMMWASPSAASRGEFSDPMMRTFEAWMAQYGRVYSDDNEKMRRFQI
FB_1       MASKFQLVFLFLFLCAMWASPSAASRDEPNDPMMRRFEEWMAEYGRVYKDNDEKMRRFQI

                  70        80        90       100       110       120
           ....|....|....|....|....|....|....|....|....|....|....|....|
OAY62650.1 FKNNVNYIETFNNGSRNSYTLGTNQFTDLINNEFIAQHIG-ALPLNIKREPVVLFDDVNI
FB_1       FKNNVKHIETFNSRNGNSYTLGINQFTDMTKSEFVAQYTGVSLPLNIEREPVVSFDDVNI

                 130       140       150       160       170       180
           ....|....|....|....|....|....|....|....|....|....|....|....|
OAY62650.1 SAVPQSIDWRYYGAVTPIKDQGSCGSCWAFSAIATVEGIYKIKTGFILSLSEQKVLDCAV
FB_1       SAVPQSIDWRDYGAVNEVKNQNPCGSCWAFAAIATVEGIYKIKTGYIVSLSEQEVLDCAV

                 190       200       210       220       230       240
           ....|....|....|....|....|....|....|....|....|....|....|....|
OAY62650.1 SNGCNGGQVNKAYDFIISNNGVTSTVFYPYKGNQGTCAANRVPNSAYITGYSYVPRNDER
FB_1       SYGCKGGWVNKAYDFIISNNGVTTEENYPYKAYQGTCNANSFPNSAYITGYSYVRRNDER

                 250       260       270       280       290       300
           ....|....|....|....|....|....|....|....|....|....|....|....|
OAY62650.1 SMMYAASNQPIAALIDASGNNFRSYQGGVFSGPCGTSLDHVITIIGYGQDISGTKYWIVK
FB_1       SMMYAVSNQPIAALIDAS-ENFQYYNGGVFSGPCGTSLNHAITIIGYGQDSSGTKYWIVR

                 310       320       330       340       350
           ....|....|....|....|....|....|....|....|....|....|....|..
OAY62650.1 NSWGMSWGEGGYIRMARDVSSSAGLCGIAMAPLFPTLRSAANAEVIKIVSKT
FB_1       NSWGSSWGEGGYVRMARGVSSSSGVCGIAMSPLFPTLQSGANAEVIKIVSKT
```

(b)

```
                  10        20        30        40        50        60
           ....|....|....|....|....|....|....|....|....|....|....|....|
OAY68270.1 MMKRFEEWMAEYGRVYKDNDEKMRRFQIFKNNVNHIETFNSRNGNSYTLGINQFTDMTKS
FB_2       MMKRFEEWMAEYGRVYKDNDEKMRRFQIFKNNVNHIETFNSRNGNSYTLGINQFTDMTKS

                  70        80        90       100       110       120
           ....|....|....|....|....|....|....|....|....|....|....|....|
OAY68270.1 EFVAQYTGVSLPLNIEREPVVSFDDVNISAVPQSIDWRDYGAVNEVKNQNPCGSCWSFAA
FB_2       EFVAQYTGVSLPLNIEREPVVSFDDVNISAVPQSIDWRDYGAVNEVKNQNPCGSCWSFAA

                 130       140       150       160       170       180
           ....|....|....|....|....|....|....|....|....|....|....|....|
OAY68270.1 IATVEGIYKIKTGYLVSLSEQEVLDCAVSYGCKGGWVNKAYDFIISNNGVTTEENYPYQA
FB_2       IATVEGIYKIKTGYLVSLSEQEVLDCAVSYGCKGGWVNKAYDFIISNNGVTTEENYPYQA

                 190       200       210       220       230       240
           ....|....|....|....|....|....|....|....|....|....|....|....|
OAY68270.1 YQGTCNANSFPNSAYITGYSYVRRNDERSMMYAVSNQPIAALIDASENFQYYNGGVFSGP
FB_2       YQGTCNANSFPNSAYITGYSYVRRNDERSMMYAVSNQPIAALIDASENFQYYNGGVFSGP

                 250       260       270       280       290       300
           ....|....|....|....|....|....|....|....|....|....|....|....|
OAY68270.1 CGTSLNHAITIIGYGQDSSGTKYWIVRNSWGSSWGEGGYVRMARGVSSSSGVCGIAMSPL
FB_2       CGTSLNHAITIIGYGQDSSGTKYWIVRNSWGSSWGEGGYVRMARGVSSSSGVCGIAMAPL

                 310
           ....|....|....|....
OAY68270.1 FPTLQSGANAELIKMVSET
FB_2       FPTLQSGANAELIKMVSET
```

Figure 4.11 Sequence alignment of (a) FB_1 and OAY62650.1 (b) FB_2 and OAY68270.1 and (c) FB_3 and OAY85858.1. The amino acid residues substitution, deletion and insertion are indicated in the red box.

(c)

```
                  10        20        30        40        50        60
         ....|....|....|....|....|....|....|....|....|....|....|....|
OAY85858.1  MASKVQLVFLFLFLCVMWASPSAASRDEPSDPMMKRFEEWMAEYGRVYKDNDEKMRRFQI
FB_3        MASKVQLVFLFLFLCVMWASPSAASRDEPSDPMMKRFEEWMAEYGRVYKDNDEKMRRFQI

                  70        80        90        100       110       120
         ....|....|....|....|....|....|....|....|....|....|....|....|
OAY85858.1  FKNNVNHIETFNNRNGNSYTLGINKFTDMTNNEFVAQYTG-VSLPLNFKREPVVSFDDVN
FB_3        FKNNVNHIETFNNRNGNSYTLGINKFTDMTNNEFVAQYTGGISRPLNIEKEPVVSFDDVN

                  130       140       150       160       170       180
         ....|....|....|....|....|....|....|....|....|....|....|....|
OAY85858.1  ISAVGQSIDWRDYGAVTEVKDQNPCGSCWAFSAIATVEGIYKIVTGYLVSLSEQEVLDCA
FB_3        ISAVGQSIDWRDYGAVTEVKDQNPCGSCWAFSAIATVEGIYKIVTGYLVSLSEQEVLDCA

                  190       200       210       220       230       240
         ....|....|....|....|....|....|....|....|....|....|....|....|
OAY85858.1  VSNGCDGGFVDNAYDFIISNNGVASEADYPYQAYQGDCAANSWPNSAYITGYSYVRSNDE
FB_3        VSNGCDGGFVDNAYDFIISNNGVASEADYPYQAYQGDCAANSWPNSAYITGYSYVRSNDE

                  250       260       270       280       290       300
         ....|....|....|....|....|....|....|....|....|....|....|....|
OAY85858.1  SSMKYAVWNQPIAAAIDASGDNFQYYNGGVFSGPCGTSLNHAITIIGYGQDSSGTQYWIV
FB_3        SSMKYAVWNQPIAAAIDASGDNFQYYNGGVFSGPCGTSLNHAITIIGYGQDSSGTQYWIV

                  310       320       330       340       350
         ....|....|....|....|....|....|....|....|....|....|..
OAY85858.1  KNSWGSSWGERGYVRMARGVSSSGLCGIAMDPLYPTLQSGANVAVIKMVSET
FB_3        KNSWGSSWGERGYIRMARGVSSSGLCGIAMDPLYPTLQSGANVAVIKMVSET
```

Figure 4.11   Continued.

This observation can be explained by the use of different starting materials for sequencing. The targeted fruit bromelain reference sequences are annotated from genomic library which does not consider the ability of gene to produce different transcripts through post-transcription modification and splicing. Instead, mRNA which is the end product of gene transcription was used in the present study. Besides that, higher error rate (at least 15%) of the PacBio sequencing method used in assembling the *A. comosus* cultivar MD2 genome library (Quail et al., 2012; Rhoads & Au, 2015; Salmela, Walve, Rivals, Ukkonen, & Sahinalp, 2017). In contrast, the Sanger sequencing technique was used in this study which is more reliable because the Sanger method has a lower error rate than the next generation sequencing (NGS) method which makes it the "gold standard" for DNA sequencing (Beck, Mullikiin, & Biesecker, 2016; Zhu, Wang, Peng, & Shete, 2016). This conventional sequencing method was reported to have fidelity with an accuracy of up to 99.999% (Shendure & Ji, 2008). The errors present in the assemble library caused the failure in gene annotation software to distinguish and recognise introns and exons, respectively.

4.2.5 Data mining

Since the isolated fruit bromelain transcripts FB_1, FB_2 and FB_3 were different to their original reference sequences OAY62650.1, OAY68270.1, and OAY85858.1 respectively, a further analysis of these sequences was required. Sequence validation was performed through the NCBI BLAST database. The BLAST result of fruit bromelain protein sequences against available sequences in the database showed that FB_1, FB_2 and FB_3 are highly homologous to fruit bromelain sequence O23791.1 (98 %), OAY68270.1 (99 %) and BAA21848.1 (99 %), respectively with a 100% query coverage (Table 4.2). Subsequently, the isolated fruit bromelain sequences were subjected to signal peptide and domain analysis. It was found that signal peptide is present only in FB_1 and FB_3, with a length of 24 amino acid residues starting from the N-terminal site (Figure 4.12). Furthermore, InterProScan revealed that the FB_1, FB_2 and FB_3 are classified in the C1A family with domain I29 and PLCE. The I29 domain constituted 58 amino acid residues while the PLCE domain constituted 213 amino acid residues.

Table 4.2   Protein BLAST analysis of the isolated fruit bromelain amino acid sequences

Sequence	Accession number of the identical sequences	Description	Query cover (%)	Identity (%)
FB_1	O23791.1	Fruit bromelain	100	98
FB_2	OAY68270.1	Fruit bromelain	100	99
FB_3	BAA21848.1	Fruit bromelain	100	99
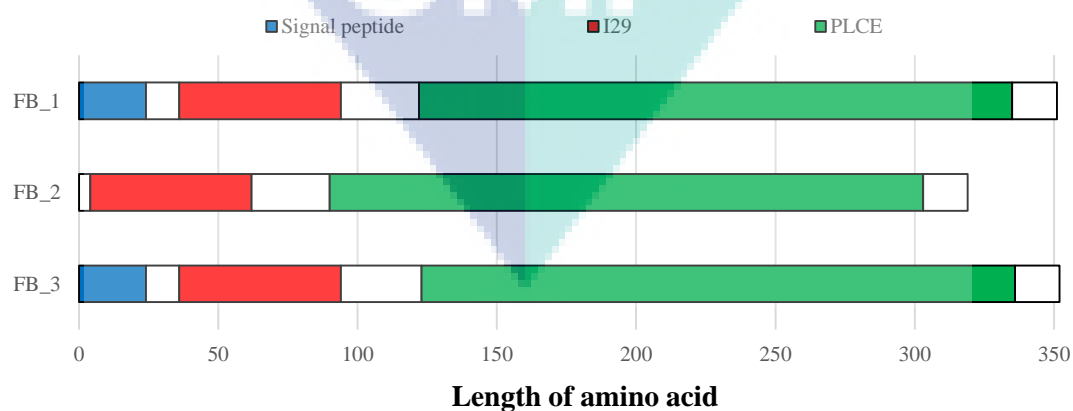
Figure 4.12   Domain organisation of fruit bromelain. The blue box is signal peptide, the red box is I29 domain and the green box is PLCE domain.

In order to illustrate the sequence features of the fruit bromelain, amino acid sequences of FB_1, FB_2 and FB_3 were aligned. This allows the discovery of conserved sequence motifs which are sets of highly conserved amino acid residues. These protein motifs are important for protein structure formation and therefore are unlikely to be involved in evolution (Mohamed, Elloumi, & Thompson, 2016). Several highly conserved C1A protease motifs were found in the alignment, indicating that FB_1, FB_2 and FB_3 shared common papain-like family sequence features. It can be seen that ERFNIN and GNFD motifs were found in the pro-peptide region while CGSCWAF, HA and NSW motifs which contain the catalytic residues cysteine, histidine and asparagine were found in the PLCE domain (Figure 4.13). The structural significance of these protein domains, conserved motifs and catalytic residues will be further discussed in the section 4.3.3. Furthermore, sequence alignment between mature protein of ananain with FB_1, FB_2 and FB_3 showed that the mature part of FB_1, FB_2 and FB_3 have a size of 215 residues.

```
                  10        20        30        40        50        60        70
         ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
Ananain  ---------------------------------------------------------------------
FB_1     SRDEPNDPMMKRFEEWMAEYGRVYKDNDEKMRRFQIFKNNVKHIETFNSRNGNSYTLGINQFTDMTKSEF
FB_2     --------MMKRFEEWMAEYGRVYKDNDEKMRRFQIFKNNVNHIETFNSRNGNSYTLGINQFTDMTKSEF
FB_3     SRDEPSDPMMKRFEEWMAEYGRVYKDNDEKMRRFQIFKNNVNHIETFNNRNGNSYTLGINKFTDMTNNEF

                  80        90       100       110       120       130       140
         ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
Ananain  -----------------------------VPQSIDWRDSGAVTSVKNQGRCGSCWAFASIATVESIYKIK
FB_1     VAQYTG-VSLPLNIEREPVVSFDDVNISAVPQSIDWRDYGAVNEVKNQNFCGSCWAFAAIATVEGIYKIK
FB_2     VAQYTG-VSLPLNIEREPVVSFDDVNISAVPQSIDWRDYGAVNEVKNQNFCGSCWSFAAIATVEGIYKIK
FB_3     VAQYTGGISRPLNIEKEPVVSFDDVNISAVGQSIDWRDYGAVTEVKDQNFCGSCWAFSAIATVEGIYKIV

                 150       160       170       180       190       200       210
         ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
Ananain  RGNLVSLSEQQVLDCAVSYGCKGGWINKAYSFIISNKGVASAAIYPYKAAKGTCKTNGVPNSAYITRYTY
FB_1     TGYLVSLSEQEVLDCAVSYGCKGGWVNKAYDFIISNNGVTTEENYPYKAYQGTCNANSFPNSAYITGYSY
FB_2     TGYLVSLSEQEVLDCAVSYGCKGGWVNKAYDFIISNNGVTTEENYPYQAYQGTCNANSFPNSAYITGYSY
FB_3     TGYLVSLSEQEVLDCAVSNGCDGGFVDNAYDFIISNNGVASEADYPYQAYQGDCAANSWPNSAYITGYSY

                 220       230       240       250       260       270       280
         ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
Ananain  VQRNNERNMMYAVSNQPIAAALDAS-GNFQHYKRGVFTGPCGTRLNHAIVIIGYGQDSSGKKFWIVRNSW
FB_1     VRRNDERSMMYAVSNQPIAAALIDAS-ENFQYYNGGVFSGPCGTSLNHAITIIGYGQDSSGTKYWIVRNSW
FB_2     VRRNDERSMMYAVSNQPIAAALIDAS-ENFQYYNGGVFSGPCGTSLNHAITIIGYGQDSSGTKYWIVRNSW
FB_3     VRSNDESSMKYAVWNQPIAAAIDASGDNFQYYNGGVFSGPCGTSLNHAITIIGYGQDSSGTQYWIVKNSW

                 290       300       310       320
         ....|....|....|....|....|....|....|....|....|....|------------
Ananain  GAGWGEGGYIRLARDVSSSFGLCGIAMDPLYPTLQ--------------
FB_1     GSSWGEGGYVRMARGVSSSSGVCGIAMSPLFPTLQSGANAEVIKIVSKT
FB_2     GSSWGEGGYVRMARGVSSSSGVCGIAMAPLFPTLQSGANAELIKMVSET
FB_3     GSSWGERGYIRMARGVSSS-GLCGIAMDPLYPTLQSGANVAVIKMVSET
```

Figure 4.13   Sequence alignment of ananain, FB_1, FB_2 and FB_3. The conserved motifs are highlight in the red boxes while the catalytic residues are indicated by green triangles. The mature part of proteins is separated by blue lines. Dashes denoted gaps between sequences.

53

It is possible to deduce the physical and chemical properties of fruit bromelain based on its amino acid sequence. The signal peptide and I29 domain were removed prior to the analysis to examine the mature part of fruit bromelain. From the Protparam analysis, the molecular weight of the mature protein of FB_1, FB_2 and FB_3 was estimated around 23.40 kDa, 23.40 kDa and 23.15 kDa, respectively (Table 4.3). This corresponds to the molecular weight of purified fruit bromelain (Kothare et al., 2017; Maurer, 2001). Besides that, the pI value of FB_1, FB_2 and FB_3 are 5.40, 5.13 and 4.15 respectively suggesting that they are acidic proteins (Table 4.3). This allows the fruit bromelain to exhibit activity in the acidic environment of pineapple juice (Ramsaroop & Saulo, 2007). FB_1 has a total number of 16 acidic amino acid residues and 14 basic amino acid residues; FB_2 has same number of aspartate and glutamate as well as one less lysine residue than FB_1, making it slightly acidic than FB_1; FB_3 has the greatest number of acidic amino acid residues (14 aspartate and 7 glutamate) and the least number of basic amino acid residues (5 arginine and 4 lysine) which explains why it is the most acidic among the fruit bromelain (Figure 4.14).

Table 4.3   Physicochemical properties of the fruit bromelain

Sequence	Molecular weight (kDa)	(pI)
FB_1	23.40	5.40
FB_2	23.40	5.13
FB_3	23.15	4.15
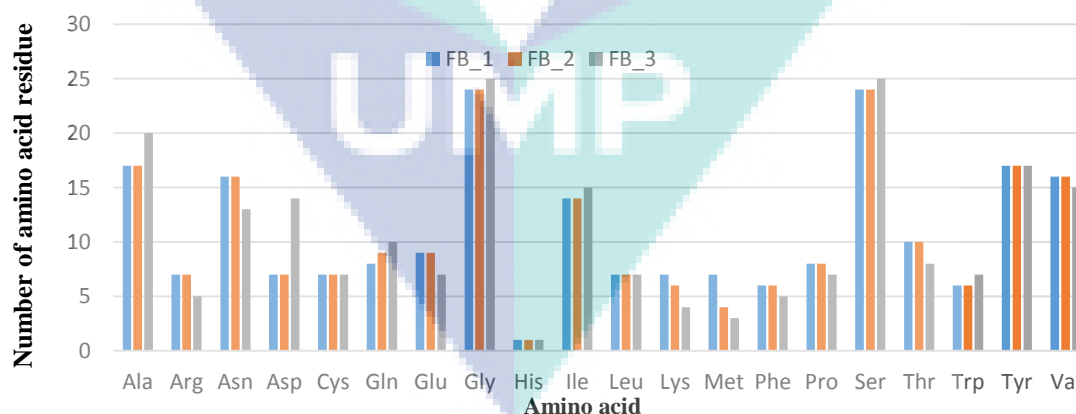
Figure 4.14   Amino acid composition of FB_1, FB_2 and FB_3. The abbreviation of amino acid: Ala (Alanine), Arg (Arginine) Asn (Asparagine), Asp: (Aspartic acid), Cys (Cysteine), Gln (Glutamine), Glu (Glutamic acid) Gly (Glycine), His (Histidine), Ile (Isoleucine), Leu (Leucine), Lys (Lysine), Met (Methionine), Phe (Phenylalanine), Pro (Proline), Ser (Serine), Thr (Threonine), Trp (Tryptophan), Tyr (Tyrosine) and Val (Valine).

4.3   Structural analysis and comparison of target fruit bromelain

The similarities and differences between FB_1, FB_2 and FB_3 are also determined in terms of their secondary and tertiary structures. Since there is no fruit bromelain crystal structure available, a comparison between FB_1, FB_2 and FB_3 with the available crystal structure of papain will also disclose important structural information of fruit bromelain including the interaction between pro-peptide and mature protein as well as binding residues in the fruit bromelain.

4.3.1   Comparative modelling of fruit bromelain using Modeller

In order to search for a suitable template to model fruit bromelain, the isolated fruit bromelain sequences were BLAST against PDB database. The BLAST result revealed that FB_1, FB_2 and FB_3 are highly identical to 6MIR with percent identity of 77.21 %, 76.28 % and 73.15 %, respectively (Table 4.4). However, the query coverage of FB_1, FB_2 and FB_3 by 6MIR was only 68-70 % (Figure 4.15). On the other hand, FB_1 and FB_2 showed the longest query coverage of 96-99 % with 1PCI, 4QTG, 3TNX and 4QRV with percent identity of ~42 %, whereas FB_3 demonstrated 95-97 % query coverage with 5EF4, 1PCI, 3TNX and 4QRG with percent identity between 40-42 %. Furthermore, FB_1, FB_2 and FB_3 have no gaps with 6MIR and the least gap percentage with 1PCI indicate a good alignment between the sequences. (Note: 6MIR structure is obsolete and has been replaced by 6OKJ. However, the information is not update in BLAST tool. To avoid confusion, the 6MIR is used throughout this thesis.)

Table 4.4   Top 5 protein templates that shared the highest similarity with fruit bromelain sequences ranked based on E-value

FB_1					
Template	PDB Description	Query Coverage (%)	E-value	Identity (%)	Gaps (%)
6MIR	Ananain	68	3e-123	77.21	0
1PCI	Procaricain	96	3e-84	42.48	1
4QRG	Papain	96	8e-83	43.09	6
3TNX	Papain	96	3e-83	42.77	6
4QRV	Papain	96	4e-83	42.26	5

Table 4.4   Continued

FB_2					
Template	**PDB Description**	**Query Coverage (%)**	**E-value**	**Identity (%)**	**Gaps (%)**
6MIR	Ananain	70	4e-122	76.28	0
1PCI	Procaricain	99	6e-84	41.83	1
4QRG	Papain	99	2e-82	42.12	6
3TNX	Papain	99	7e-82	41.80	6
4QRV	Papain	99	1e-81	41.29	5
FB_3					
Template	**PDB Description**	**Query Coverage (%)**	**E-value**	**Identity (%)**	**Gaps (%)**
6MIR	Ananain	68	2e-109	73.15	0
5EF4	Amb A 11 Cysteine Protease	95	1e-83	40.89	4
1PCI	Procaricain	97	2e-81	42.58	2
3TNX	Papain	97	3e-81	41.85	5
4QRG	Papain	97	6e-80	41.53	5

(a)

(b)
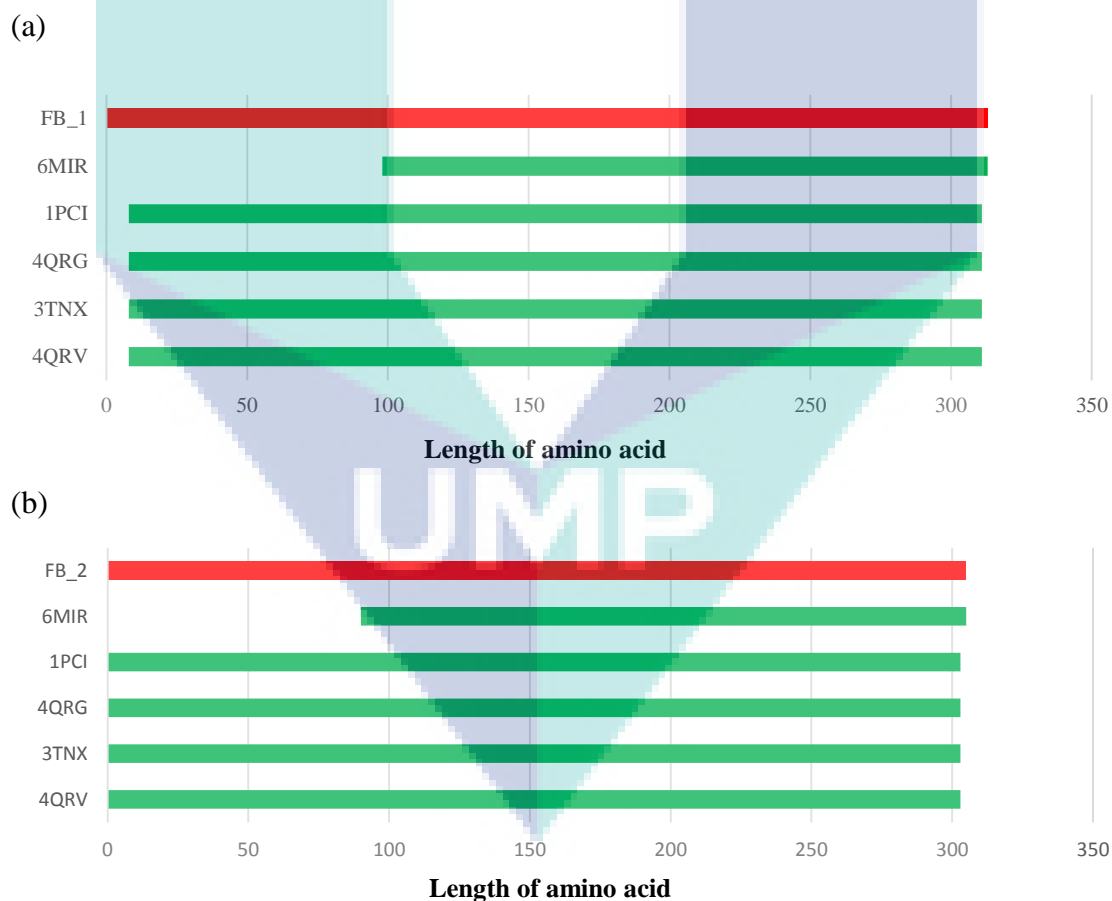
Figure 4.15   Sequence coverage of different templates to (a) FB_1 (b) FB_2 and (c) FB_3. Red bar indicated the length of target sequence while green bar indicated the region cover by the template.

(c)
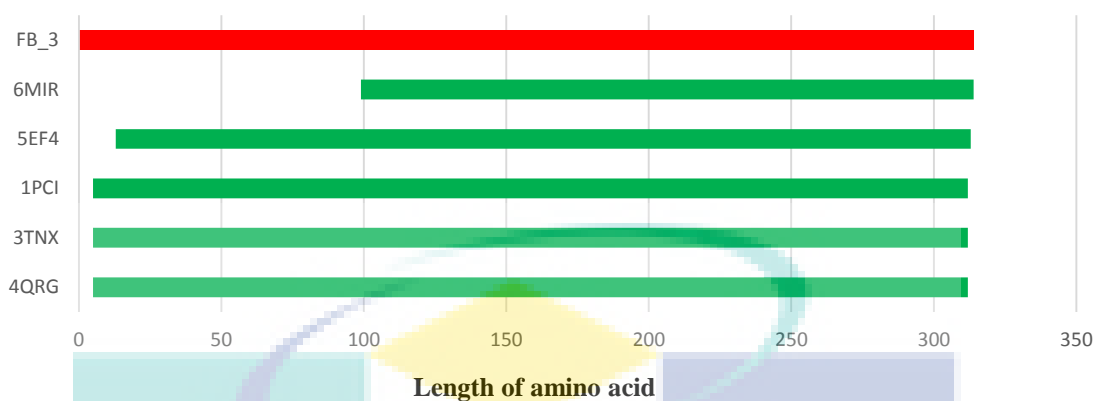
Figure 4.15   Continued.

Initially, 1PCI appeared as the best template because it has the longest coverage (≥96%) and sequence identity of 42 % for modelling. According to Ginalski (2006), protein modelled with a template of 30-50 % identity will not exceed 4 Å RMSD from its native structure. Furthermore, the gap between fruit bromelain sequences FB_1, FB_2 and FB_3 with 1PCI is very low (1-2 %) to avoid misalignment between the target and template sequence which may distort the constructed protein structure (Dorn et al., 2014). In addition, 1PCI was used as a template in previous reported comparative structural bromelain studies which demonstrate it as a suitable template to model fruit bromelain (Ramli et al., 2018; Tap et al., 2016).

In this study, a set of 100 decoys for each fruit bromelain sequences was generated to provide a pool of robust results (Sefidbakht, Ranaei Siadat, & Taheri, 2017). The good model was distinguished from the bad models using the MODELLER in-built assessment method which is the DOPE score. DOPE is a statistical potential assessment that corresponds to non-interacting atoms in a homogeneous sphere with the radius dependent on a sample native structure (Shen & Sali, 2006). A lower DOPE score indicates the structure has a better packing of the atoms and is more accurate at its native conformation (Cloete, Kapp, Joubert, Christoffels, & Malan, 2018). A further evaluation on the stereochemical quality of these selected models was done using the SAVES 5.0 server using different evaluation programs:.As shown in Table 4.5, FB_1, FB_2 and FB_3 passed the Verify 3D test with 86.58 %, 84.59 % and 90.13 %, respectively. Meanwhile, the ERRAT score for FB_1, FB_2 and FB_3 are 73.77, 73.06 and 65.69, respectively. Furthermore, PROCHECK showed FB_1 (99.6 %), FB_2 (99.2 %) and FB_3 (98.9 %)

have acceptable phi and psi dihedral angle distributions of amino acid residues in the modelled structures.

Table 4.5   Models evaluation using Verify 3D, ERRAT and PROCHECK

Model	Template	Verify 3D (%)	ERRAT	PROCEHCK (%)
FB_1	1PCI	86.58	73.77	Favoured region:85.2 Allowed region:14.4 Outlier region:0.4
	1PCI-6MIR	89.46	79.34	Favoured region:88.9 Allowed region:10.7 Outlier region:0.4
	1PCI-6MIR-4QRV	88.18	85.96	Favoured region:90.4 Allowed region:8.8 Outlier region:0.7
FB_2	1PCI	84.59	73.06	Favoured region:86.0 Allowed region:13.2 Outlier region:0.8
	1PCI-6MIR	88.52	81.14	Favoured region:90.9 Allowed region:8.7 Outlier region:0.4
	1PCI-6MIR-4QRV	86.89	74.75	Favoured region:89.8 Allowed region:9.4 Outlier region:0.8
FB_3	1PCI	90.13	65.69	Favoured region:87.5 Allowed region:11.4 Outlier region:1.1
	1PCI-6MIR	89.49	80.72	Favoured region:90.4 Allowed region:9.2 Outlier region:0.4
	1PCI-6MIR-3TNX	89.81	83.99	Favoured region:91.5 Allowed region:7.7 Outlier region:0.7

Since 1PCI only has a sequence identity of ~42 % of the fruit bromelain sequences, it was expected that certain information was missing during the modelling which led to a lower score in ERRAT. The quality of the model was then improved by using multiple templates as an additive mixture of the density functions derived from each individual template to restrain a single target distance on the target model (Meier & Söding, 2015). For instance, the quality of the comparative model of human MCT8 protein was reported to improve when using multiple templates (Shaji., 2017). The protein structure of 6MIR was chosen as the second template because it shared the highest sequence identity of ~73-77 %. A total of 100 models were generated using both 1PCI-6MIR as a template for each FB_1, FB_2 and FB_3. The best models were selected based on the DOPE score and were

subjected to model validation using the SAVES database. From the analysis, FB_1 scored 88.52 % in Verify 3D, 79.34 in ERRAT and has a higher number of amino acid residues (88.9 %) located at a favoured region (Table 4.5). Similarly, FB_2 also has a higher score of 89.51 % in the Verify 3D test, 81.14 in ERRAT and the distribution of amino acid residues at the outlier region was reduced to 0.4 %. Moreover, FB_3 has a Verify 3D score of 89.49 % in Verify 3D, 80.72 in ERRAT and recorded 99.6 % of residues at favoured and allowed regions. As a result, the models of FB_1, FB_2 and FB_3 generated using 1PIC-6MIR showed an improved score in Verify 3D, ERRAT and PROCHECK compared to models generated using a single template (1PCI).

The effect of adding third templates to 1PCI-6MIR was also studied. By considering the length of sequence coverage, sequence identity and gaps between fruit bromelain and templates, 4QRV was selected as the third template to models the targeted fruit bromelains FB_1 and FB_2 and 3TNX to model FB_3. Similarly, 100 models were generated, and the best model selected based on the DOPE score was validated using the SAVES database. The FB_1 scored 86.88% in Verify 3D, 85.96 in ERRAT and have 99.3 % amino acid residues located the reasonable position (Table 4.5). Meanwhile, FB_2 scored 86.89 % in Verify 3D, 74.75 in ERRAT and had 99.2 % amino acid residues distributed within the acceptable location. In addition, FB_3 scored 89.81% in Verify 3D, 83.99 in ERRAT and have 99.3% amino acid residues with an acceptable distribution within the generated structure. It can be seen that FB_1 and FB_3 modelled with the third template system have a slight improvement in ERRAT while there were no significant changes in Verify 3D and PROCHECK. The FB_2 has lower quality scores when modelled with 1PCI-6MIR-4QRV. This indicates that the quality of the generated models are less likely to benefit from the addition of the third template because an undesirable gap may be created in the alignment due to sequence divergence (Li & Cheng, 2016). Therefore, the final selected models are FB_1 (1PCI-6MIR-4QRV), FB_2 (1PCI-6MIR) and FB_3 (1PCI-6MIR-3TNX) based on their overall performance in Verify 3D, ERRAT and PROCHECK. These final models were subsequently used in further analysis.

4.3.2 Model refinement

The comparative protein model may not be always accurate (Ganugapati & Akash, 2017). Although fruit bromelain structures were improved with the use of multiple templates, the structures may still contain errors due to the differences of amino acids

between the target and templates such as insertions and gaps as well as the absence of interaction with the other biomolecules e.g. proteins, nucleic acid and ligands (Feig, 2017; Ishitani et al., 2008; Park et al., 2018). Thus, the last step of comparative modelling is to refine the initial protein structure to achieve better accuracy (Heo & Feig, 2018a). In the present study, the fruit bromelain structures were refined via molecular dynamics simulation.

The refinement method was adopted from Protein REFinment via the Molecular Dynamics (PREFMD) server. PREFMD is a state-of-the-art online server for the refinement of a protein structure in a modified method reported by Feig and Mirjalili (2016), which is ranked as the best refinement method during the 11th Community Wide Experiment on the Critical Assessment of Techniques for Protein Structure Prediction. In this study, the fruit bromelain structures of FB_1, FB_2 and FB_3 were refined in molecular dynamics simulation with a timescale of 50 ns which has been reported as sufficient for structure refinement in previous studies (Ramli et al., 2018). In addition, the fruit bromelain structures were also submitted to the Local Protein Structure Refinement via Molecular Dynamics Simulations (locPREFMD) to correct bonds, angles and torsion angles of the protein structure to improve its stereochemistry properties (Feig, 2016).

The molecular dynamics refined protein structure should form a cluster of conformations that resemble its native state (Raval, Piana, Eastwood, Dror, & Shaw, 2012). Therefore, the stability of the fruit bromelain was analysed as a function of simulation time. The RMSD value of FB_1 was constantly deviated at ~4.0 Å after 40 ns; FB_2 reached a plateau state at around 3.0 Å after 10 ns; FB_3 fluctuated between 3.0-4.0 Å after 30 ns (Figure 4.16 (a)). Based on the RMSD value, the simulation time is sufficient to enable the fruit bromelain to reach a stabilised state. Moreover, the stability of the fruit bromelain model was further analysed via the radius of gyration. Radius of gyration is used to determine the compactness of a protein structure (Lobanov, Bogatyreva, & Galzitskaya, 2008). It was observed that the compactness of FB_1, FB_2 and FB_3 was maintained at ~21 Å, ~22 Å and ~22 Å respectively without significant drift, indicating that the fruit bromelain is able to be preserved in the predicted structure throughout the simulation (Figure 4.16 (b)).
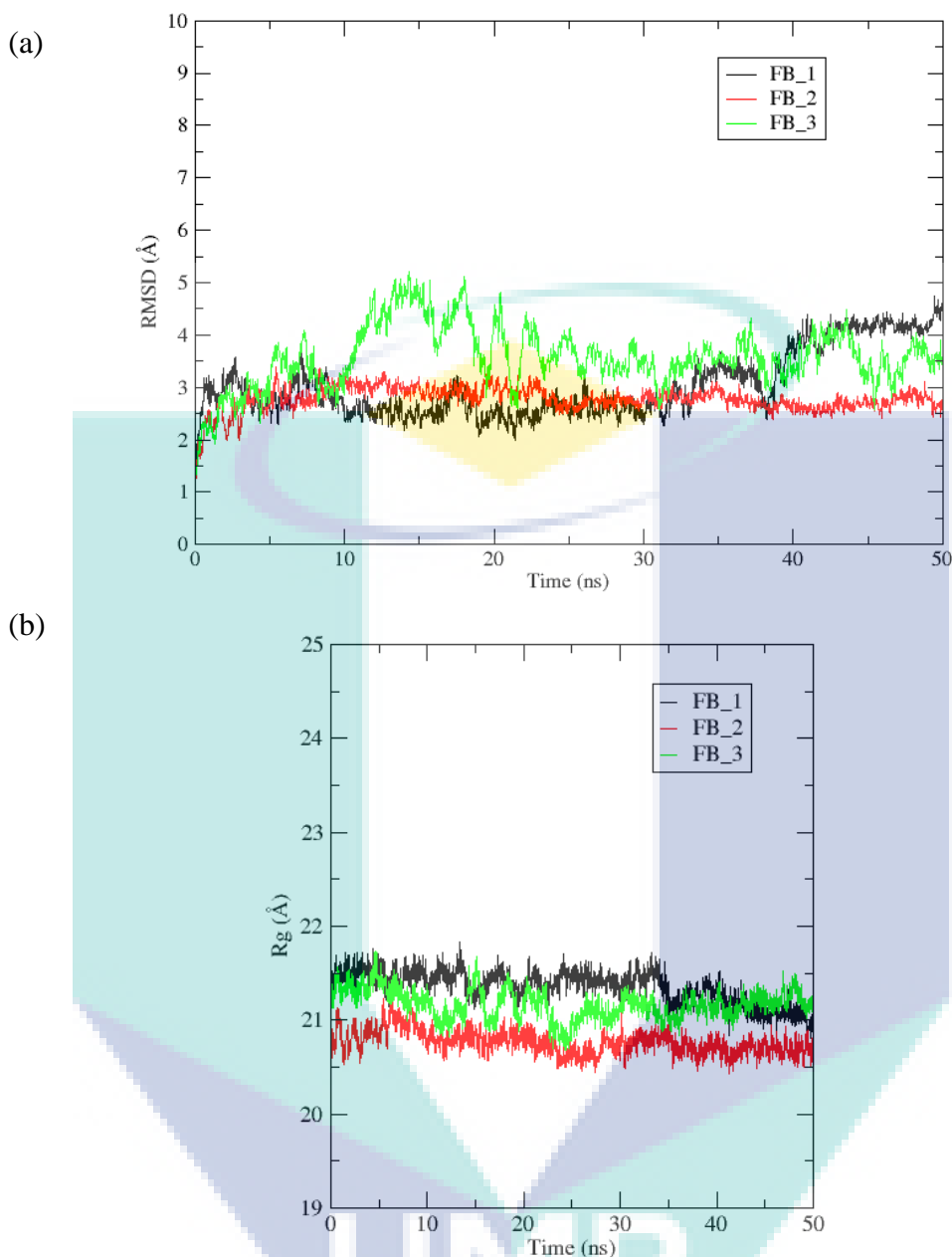
Figure 4.16   (a) RMSD indicate the overall stability of FB_1, FB_2 and FB_3 during the course of simulation. (b) Radius of gyration of represent the overall dimension of FB_1, FB_2 and FB_3 were as a function of time.

The quality of the fruit bromelain structures was found to improve after the refinement process. Verify 3D showed that the percentages of amino acid with correct 3D fold had increased (FB_1: 90.42 %; FB_2: 89.51 %; FB_3: 90.13 %) (Table 4.6). Besides that, ERRAT analysis also indicted an improvement on non-bonded interactions within the fruit bromelain (FB_1: 85.96 %; FB_2: 93.40 %; FB_3: 87.91 %). The Ramachandran plot from PROCHECK suggested that 100% of amino acid residues of FB_1 and FB_2 are located in favoured and allowed regions while FB_3 has 0.4% of residues positioned

at the outlier region. Furthermore, the energy level of FB_1, FB_2 and FB_3 was reduced to -13714.78 kJ/mol, -13468.71 kJ/mol and -12705.32 kJ/mol respectively after refinement. Native protein always folds into conformation with the lowest energy which is the most stable form (Dinner, Sali, Smith, Dobson, & Karplus, 2000; Tsai et al., 2003). This indicates that refined fruit bromelain models are closer and more similar to the native fruit bromelain compared to the unrefined models.

Table 4.6　　　Evaluation of refined models FB_1, FB_2 and FB_3

Model	Energy (kJ/mol)	Verify 3D (%)	ERRAT (%)	PROCEHCK (%)
FB_1 initial	-3570.25	88.18	85.96	Favoured region:90.4 Allowed region:8.8 Outlier region:0.7
FB_1 refined	-13714.78	90.42	86.88	Favoured region:92.6 Allowed region:7.4 Outlier region:0.0
FB_2 initial	-1245.31	88.52	81.14	Favoured region:90.9 Allowed region:8.7 Outlier region:0.4
FB_2 refined	-13468.71	89.51	93.40	Favoured region:94.0 Allowed region:6.1 Outlier region:0.0
FB_3 initial	-4338.05	89.81	83.99	Favoured region:91.5 Allowed region:7.7 Outlier region:0.7
FB_3 refined	-12705.32	90.13	87.91	Favoured region:92.3 Allowed region:7.3 Outlier region:0.4

4.3.3　Secondary structure analysis of fruit bromelain

All the modelled fruit bromelain of FB_1, FB_2 and FB_3 contain 10 α-helices (H1-H10) and six β-strands (E1-E6) as shown in Figure 4.17. This result concurs with the

findings reported by Ramli et al. (2018) in which the fruit bromelain structure with accession no. BAA21848.1 has the same pattern of secondary structures. Despite the differences in the amino acid residues in each fruit bromelain sequence of FB_1, FB_2 and FB_3, the formation of secondary structures showed a consensus between each fruit bromelain models. This is because the protein structure is better conserved compared to the amino acid sequence during evolution (Dong, Pan, Peng, Zhang, & Yang, 2018). On top of that, there is a preference on the selection of amino acid residues forming the secondary structure. It can be seen that amino acid residues M, A, L, E, K are highly prevalent in the α-helix regions (Figure 4.18). These residues are known as helix formers due to their low energetic cost for the helix formation (Haimov & Srebnik, 2016; Pace & Scholtz, 1998). On the other hand, amino acid residues I, V, W, Y, T are predominant in the beta-strands due to the capability of their hydrophobic side chains to stabilise the β structure structure (Merkel, Strutevant, & Regan, 1999).
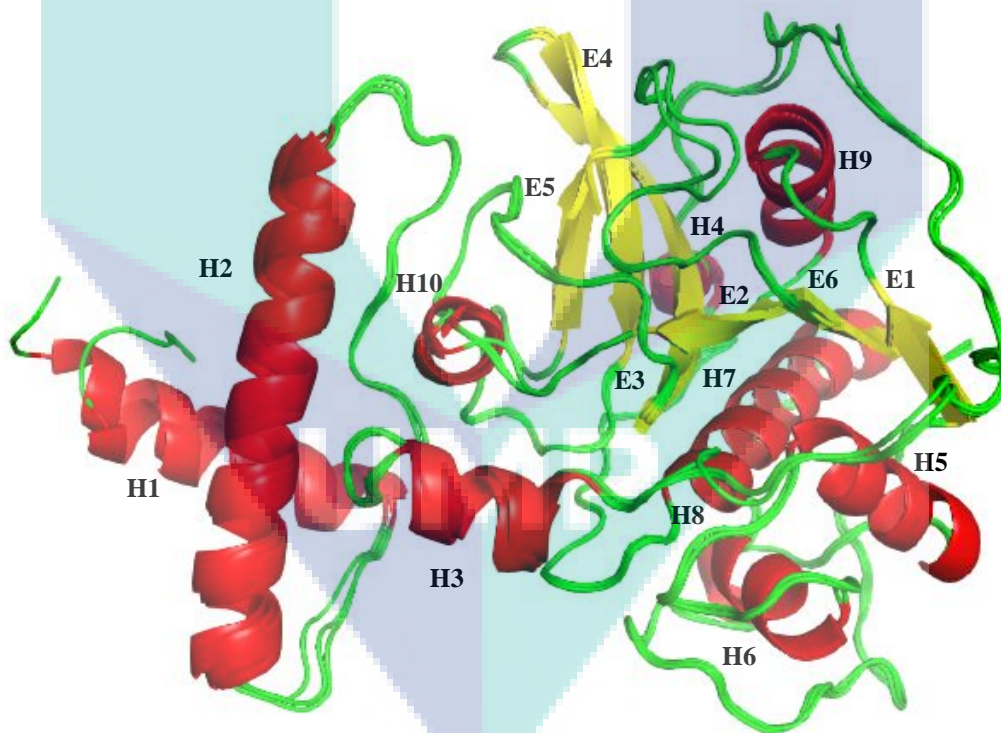
Figure 4.17   Superimposition of the 3D structure of FB_1, FB_2 and FB_3. The α-helices (H1-H10), β-strands (E1-E6) and loops of fruit bromelain are coloured in red, yellow and green, respectively.

```
                10        20        30        40        50        60
        ....|....|....|....|....|....|....|....|....|....|....|....|
FB_1    SRDEPNDPMMKRFEEWMAEYGRVYKDNDEKMRRFQIFKNNVKHIETFNSRNGNSYTLGIN
FB_2    --------MMKRFEEWMAEYGRVYKDNDEKMRRFQIFKNNVNHIETFNSRNGNSYTLGIN
FB_3    SRDEPSDPMMKRFEEWMAEYGRVYKDNDEKMRRFQIFKNNVNHIETFNNRNGNSYTLGIN
                      H1                      H2

                70        80        90       100       110       120
        ....|....|....|....|....|....|....|....|....|....|....|....|
FB_1    QFTDMTKSEFVAQYTG-VSLPLNIEREPVVSFDDVNISAVPQSIDWRDYGAVNEVKNQNP
FB_2    QFTDMTKSEFVAQYTG-VSLPLNIEREPVVSFDDVNISAVPQSIDWRDYGAVNEVKNQNP
FB_3    KFTDMTNNEFVAQYTGGISRPLNIEKEPVVSFDDVNISAVGQSIDWRDYGAVTEVKDQNP
                  H3                                      H4

               130       140       150       160       170       180
        ....|....|....|....|....|....|....|....|....|....|....|....|
FB_1    CGSCWAFAAIATVEGIYKIKTGYLVSLSEQEVLDCAVSYGCKGGWVNKAYDFIISNNGVT
FB_2    CGSCWSFAAIATVEGIYKIKTGYLVSLSEQEVLDCAVSYGCKGGWVNKAYDFIISNNGVT
FB_3    CGSCWAFSAIATVEGIYKIVTGYLVSLSEQEVLDCAVSNGCDGGFVDNAYDFIISNNGVA
                  H5                      H6                H7

               190       200       210       220       230       240
        ....|....|....|....|....|....|....|....|....|....|....|....|
FB_1    TEENYPYKAYQGTCNANSFPNSAYITGYSYVRRNDERSMMYAVSNQPIAALIDAS-ENFQ
FB_2    TEENYPYQAYQGTCNANSFPNSAYITGYSYVRRNDERSMMYAVSNQPIAALIDAS-ENFQ
FB_3    SEADYPYQAYQGDCAANSWPNSAYITGYSYVRSNDESSMKYAVWNQPIAAAIDASGDNFQ
            H8                     E1       H9          E2       H10

               250       260       270       280       290       300
        ....|....|....|....|....|....|....|....|....|....|....|....|
FB_1    YYNGGVFSGPCGTSLNHAITIIGYGQDSSGTKYWIVRNSWGSSWGEGGYVRMARGVSSSS
FB_2    YYNGGVFSGPCGTSLNHAITIIGYGQDSSGTKYWIVRNSWGSSWGEGGYVRMARGVSSSS
FB_3    YYNGGVFSGPCGTSLNHAITIIGYGQDSSGTQYWIVKNSWGSSWGERGYIRMARGVSSS-
        H10              E3          E4               E5

               310
        ....|....|....|....|
FB_1    GVCGIAMSPLFPTLQ
FB_2    GVCGIAMAPLFPTLQ
FB_3    GLCGIAMDPLYPTLQ
                E6
```

Figure 4.18   Alignment of fruit bromelain sequences FB_1, FB_2 and FB_3. The α-helices (H1-H10) are indicated in the red box while β-strands (E1-E6) are indicated in green box, the remaining regions are protein loops connected between α-helices and β-strands.

4.3.4   Tertiary structure analysis of fruit bromelain

The tertiary structure of fruit bromelain constitutes I29 and PLCE domains, respectively (Figure 4.19). The I29 domain is a globular structure consisting of three α-helices. The ERFNIN motif formed the longest α-helix core (H2) while GNFD motif is a protein loop connected between H2 and H3. The α-helices (H1, H2 and H3) in this domain form a cross motif which is similar as seen in pro-papain and pro-caricain (Groves et al., 1996; Roy et al., 2012). The PLCE domain forms the mature part of fruit bromelain. It has a typical papain fold, constituted by L-domain and R-domain of an approximately equal size. The L-domain is an all α-domain. H5 is the longest vertical helix separated by the L-domain from R-domain. Furthermore, the anti-parallel β-strands give the R-domain

a β-barrel structure and is enclosed with an α-helix (H10) which is similar to papain (Turk et al., 2012).
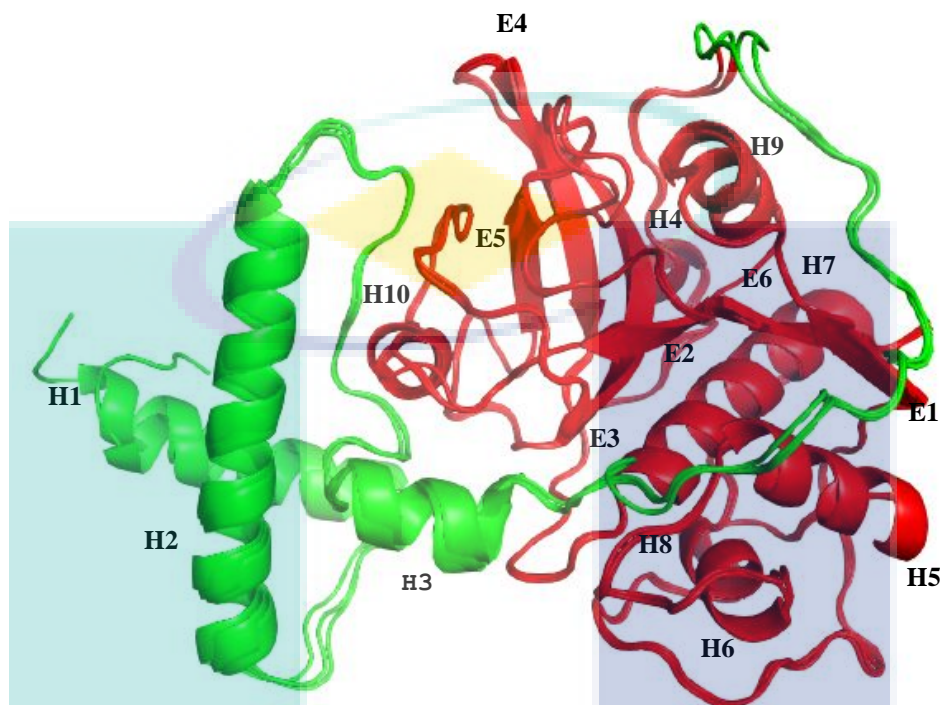
Figure 4.19   Superimposition of the 3D structure of FB_1, FB_2 and FB_3. Green and red regions indicate I29 and PLCE domain, respectively.

Meanwhile, the aforementioned CGSCWAF motif forms a protein loop along the first-turn of H5 in the L-domain. In the R-domain, the HA motif is located at the top of the E3 while the NSW motif is positioned at the protein loop adjacent to the HA motif. The catalytic-triad (Cys-His-Asn) are located at the interface of both subdomains on the top of the molecule. Despite being far apart within the chain, they are positioned in close proximity in the folded structure (Figure 4.20). This conformation is suggested to aid the protonation and deprotonation process during catalysis. First, the ionised thiol group of Cys25 conducts a nucleophilic attack to the substrate (Amri & Mamboya, 2012). This is followed by deprotonation by His159 and is aided by Asn175 to allow the completion of the reaction. Besides that, glutamine which is involved in the formation of the oxyanion hole is also located near the catalytic-triad to stabilise the transition-state complex (Menard et al., 1991).
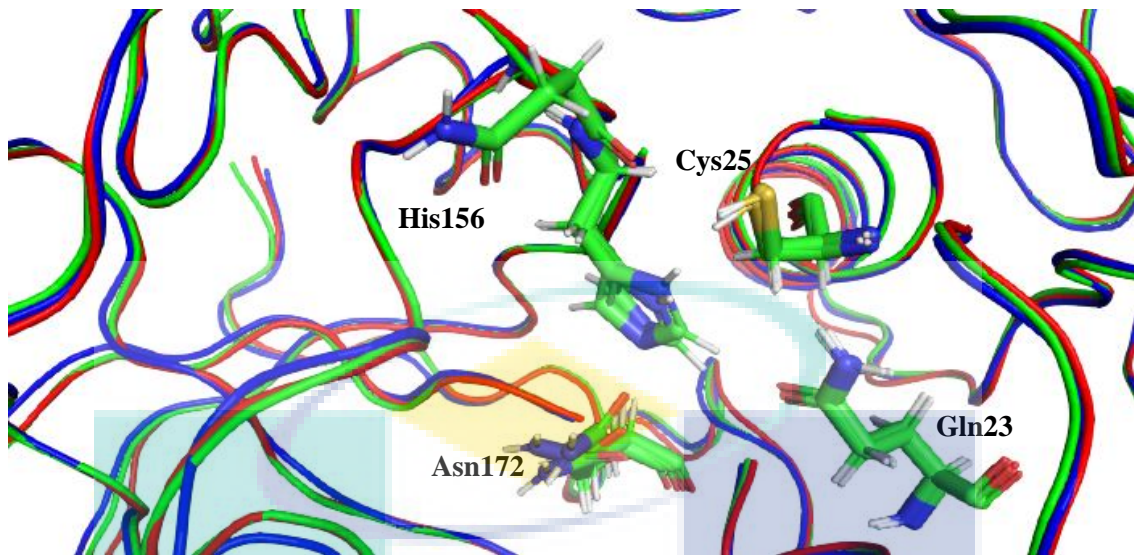
Figure 4.20   Superimposition of FB_1 (red), FB_2 (green) and FB_3 (blue). The fruit bromelain structures are represented in ribbon. The catalytic residues are represented as sticks labelled according papain numbering. Cys25, His156 and Asn172 forming catalytic triad while Gln23 assist in catalysis.

4.3.4.1   Structural stability in pro-domain

The pro-region serves as a scaffold to stabilise the overall zymogen structure of fruit bromelain via a series of electrostatic interactions which are majorly contributed by the ERFNIN and GNFD motifs (Table 4.7) (Coulombe et al., 1996; Roy et al., 2012). By using pro-papain as reference models, a salt-bridge between Glu23 and Lys39 is expected to stabilise the helix structures of H1 and H2 (Figure 4.21 (a)). Besides that, hydrogen bonds between Asp72 with Arg31 and Tyr33 also aid in maintaining the stability of the protein loop connecting H1 and H2 (Figure 4.21 (b)). Conserved substitution of Lys31 in pro-papain with Arg31 in fruit bromelain may increase the stability in this particular region due to arginine's ability to form stronger interactions than lysine (Figure 4.22) (Sokalingam, Raghunathan, Soundrarajan, & Lee, 2012). Besides that, the pro-domain is suggested to be further stabilised by the salt-bridge that formed between Arg42 with Glu38 from H2 and Glu77 from H3 (Figure 4.21 (c)) (Roy et al., 2012).

Table 4.7   Interaction of residues in the pro-domain. The amino acids are numbered based on pro-papain position

First residue	Location	amino acid position	Second residue	Location	amino acid position
Glu23	H1	14	Lys39	H2	30
Glu38	H2	29	Arg42	H2	33
Asp72	GNFD motif	64	Lys31/Arg31	Protein loop connecting H1 and H2	22
			Tyr33	Protein loop connecting H1 and H2	24
Glu77	H3	69	Arg42	H2	33

(a)
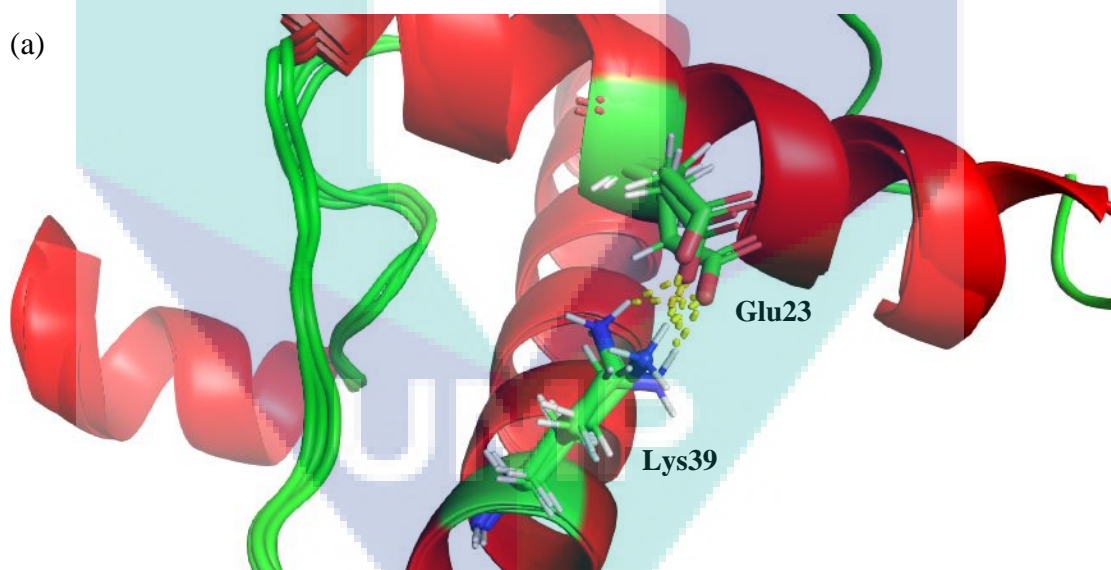
Figure 4.21   Superimposition of pro-domain of FB_1, FB_2 and FB_3 are represented in cartoon while interactions between residue is denoted by yellow dashes. (a) Salt-bridge forming by Glu23 and Lys39 stabilise the helix structures between H1 and H2. (b) Hydrogen bonds forming by Asp72, with Arg31 and Tyr33 are aided in maintaining the stability of protein loop connecting H1 and H2. (c) The pro-domain is further stabilised by salt-bridge forming between Arg42 with Glu38 from H2 and Glu77 from H3.

(b)

Asp72

Arg31

Tyr3r

(c)

Glu33

Glu77

Arg42
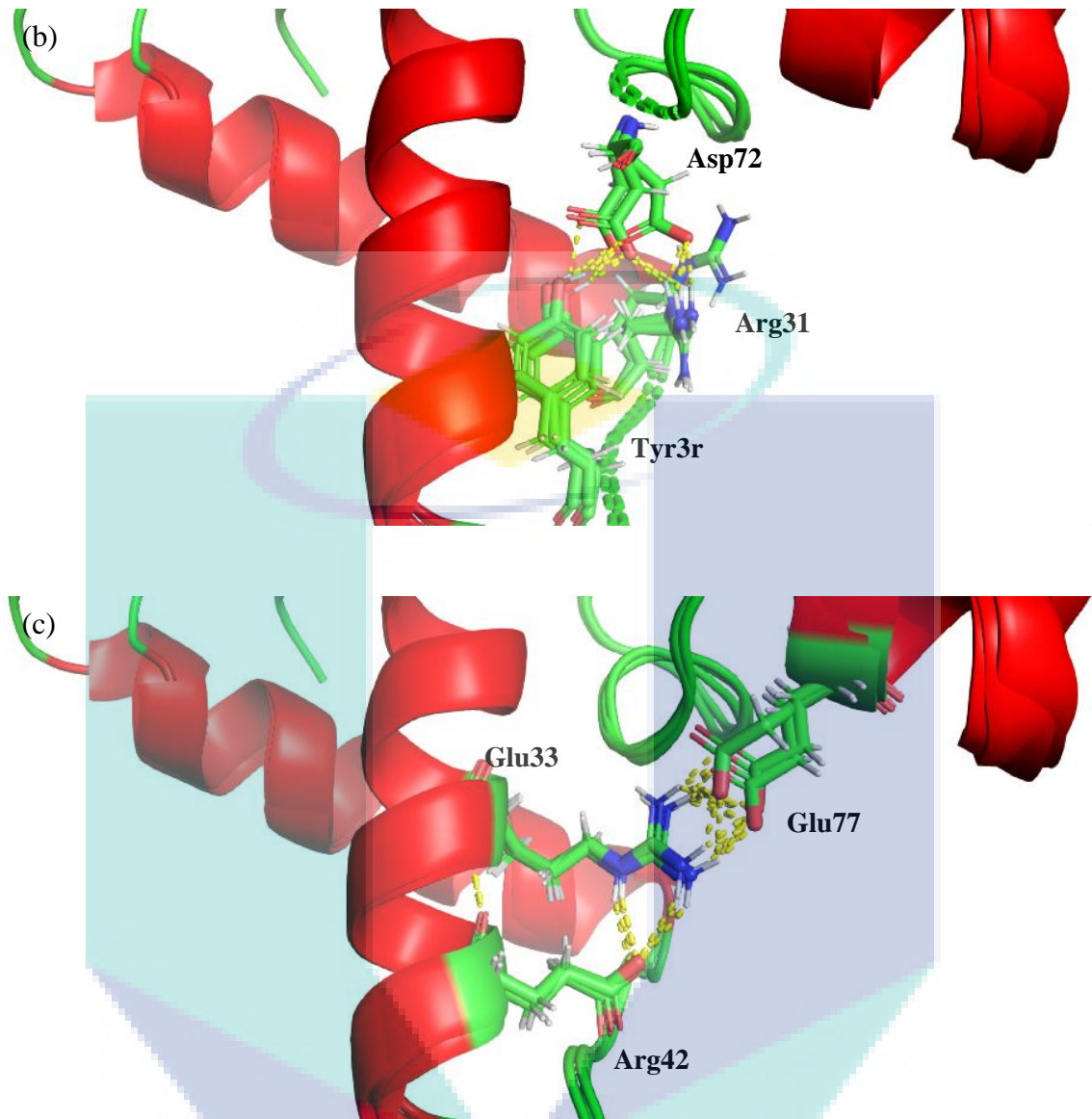
Figure 4.21   Continued.


```
                        10        20        30        40        50
                ....|....|....|....|....|....|....|....|....|....|
Pro-papain      NDLTSTERLIQLFESWMLKHNKIYKNIDEKIYRFEIFKDNLKYIDETNKK
FB_1            SRDEPNDPMMKRFEEWMAEYGRVYKDNDEKMRRFQIFKNNVKHIETFNSR
FB_2            --------MMKRFEEWMAEYGRVYKDNDEKMRRFQIFKNNVNHIETFNSR
FB_3            SRDEPSDPMMKRFEEWMAEYGRVYKDNDEKMRRFQIFKNNVNHIETFNNR
Clustal Consensus    ::: **.** ::.::**: ***: **:***:*:::*:   *.:

                        60        70
                ....|....|....|....|....|
Pro-papain      N-NSYWLGLNVFADMSNDEFKEKYT
FB_1            NGNSYTLGINQFIDMTKSEFVAQYT
FB_2            NGNSYTLGINQFIDMTKSEFVAQYT
FB_3            NGNSYTLGINKFTDMTNNEFVAQYT
Clustal Consensus    * *** **:* *:**::.**  :**
```

Figure 4.22   Sequence alignment of pro-papain and fruit bromelain. The interacting residues are indicated in the red box. Glu23, Tyr33, Glu38, Lys39, Arg42, Asp72 and Glu77 are found conserved between papain and fruit bromelain while Lys31 in papain is replaced to Arg31 in fruit bromelain.

68

4.3.4.2   Role of pro-peptide in blocking the catalytic cleft

Pro-peptide is needed to deter the early activation of the fruit bromelain by blocking its active sites (Butts et al., 2016; Zou et al., 2018). This feature is important to avoid unwanted fruit bromelain degradation (Verma et al., 2016). The pro-peptide hampers the catalytic cleft by forming various interactions with the mature part of fruit bromelain. The first hydrophobic interaction is established between Phe248, Phe256, Cys260, Trp284, Trp288 and Cys307 from the catalytic domain and Phe70, Phe78 and Tyr82 from H3 of the pro-domain (Figure 4.23 (a)). Meanwhile, another hydrophobic interaction is formed between Tyr251, Phe256 and Pro259 from the mature segment with Tyr52 (His52 in fruit bromelain), Trp64 (Thr64 in fruit bromelain) and Leu65 of pro-peptide (Figure 4.23 (b)). Furthermore, hydrogen bonds are also assembled between H2 and H3 with Ala244 (serine in fruit bromelain), Lys246 (glutamic acid in FB_1 and FB_2; aspartate in FB_3), Asp247 (asparagine in fruit bromelain) and Gln249 from the catalytic domain (Figure 4.23 (c)).
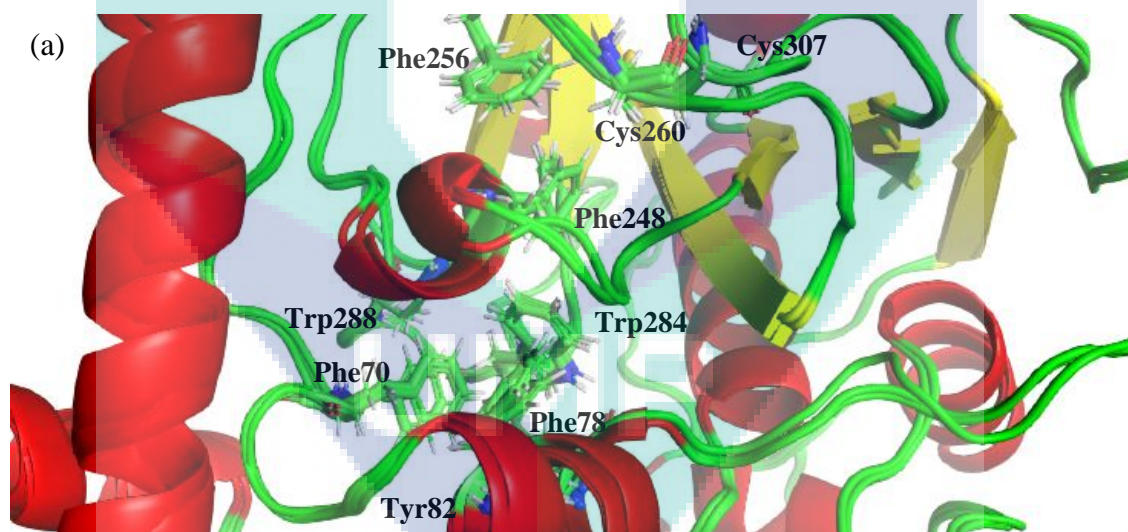
Figure 4.23   Superimposition of of FB_1, FB_2 and FB_3 are represented in cartoon. (a) Hydrophobic interaction is established between Phe248, Phe256, Cys260, Trp284, Trp288 and Cys307 from the catalytic domain and Phe70, Phe78 and Tyr82 from H3 of the pro-domain. (b) Hydrophobic interaction is formed between Tyr251, Phe256 and Pro259 from the mature segment with His52, Thr64 and Leu65 of pro-peptide. (c) Hydrogen bonds are also assembled between H2 and H3 with Ser244, Glu246 in FB_1 and FB_2, Asp246 in FB_3, Asn247 and Gln249 from the catalytic domain
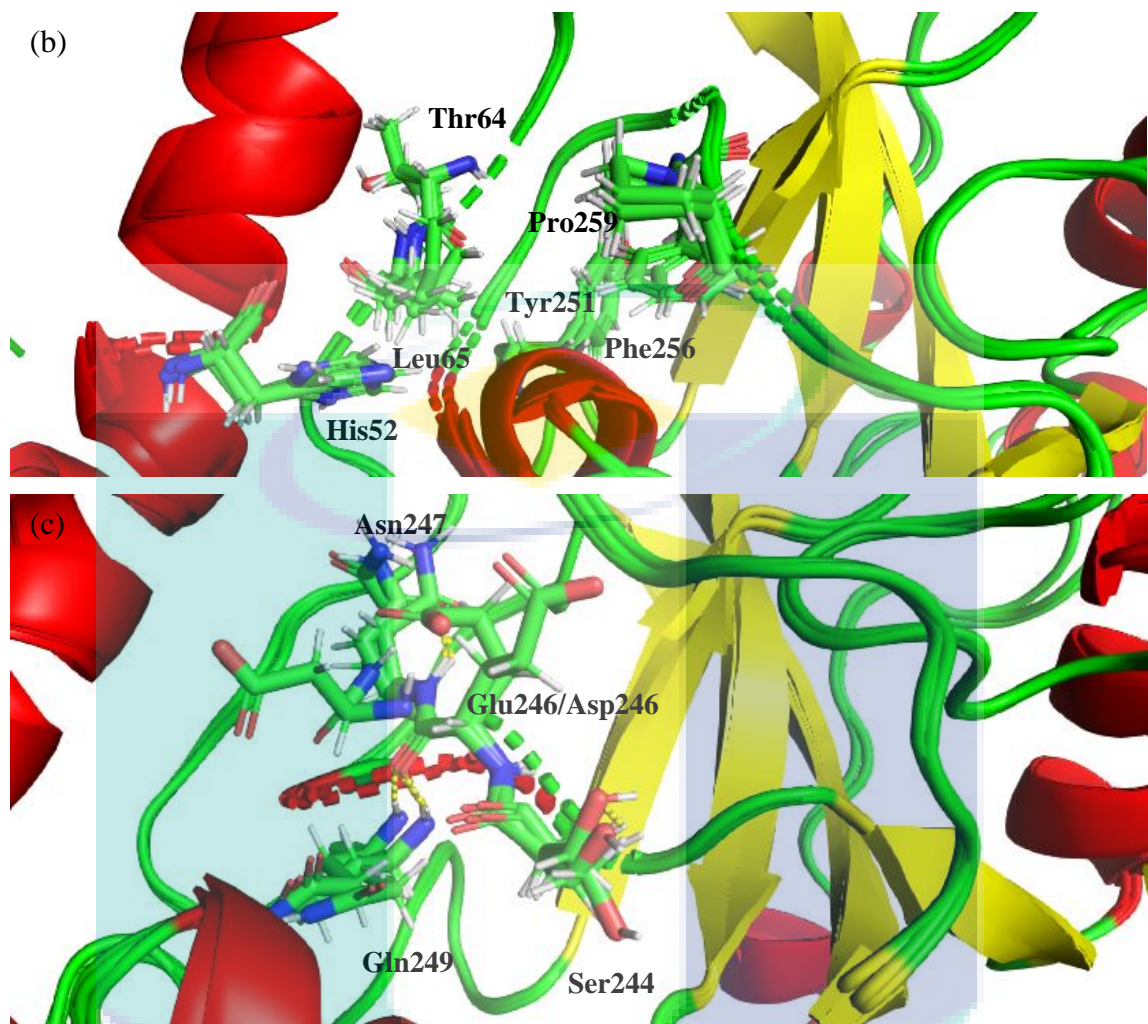
Figure 4.23   Continued.

High conservation of the amino acid residues demonstrated that the pro-peptide of fruit bromelain should pose similar interactions as shown in the papain which is important to maintain the correct position of extended pro-peptide to block the catalytic cleft (Figure 4.24) (Roy et al., 2012). In spite of this, the replacement of Tyr52 (hydrophobic) and Trp64 (hydrophobic) from papain with His52 (amphiphilic) and Thr64 (hydrophilic), respectively, may weaken the hydrophobic interactions in fruit bromelain (Figure 4.24 (b)). Moreover, the substitution of Ala244 (non-polar) with Ser244 (polar), Lys246 (positively-charged) with Glu246/Asp246 (both negative-charged) and Asp247 (negative-charged) with Asn247 (uncharged) respectively as a whole make the number of electrostatic interactions in fruit bromelain different from papain (Figure 4.24 (c)). In addition, the unbinding space of S2 subsites in FB_1 and FB_2 will also increase the replacement of Ile86 with a smaller amino acid residue valine (Figure 4.25). These distinctions may result in a different orientation of the fruit bromelain structure compared

70

to papain and affect the overall interaction strength between pro-peptide and the mature part of fruit bromelain.

(a)

```
                    10        20        30        40        50        60
               ....|....|....|....|....|....|....|....|....|....|....|....|
Pro-papain     NDLTSTERLIQLFESWMLKHNKIYKNIDEKIYRFEIFKDNLKYIDETNKKN-NSYWLGLN
FB_1           SRDEPNDPMMKRFEEWMAEYGRVYKDNDEKMRRFQIFKNNVKHIETFNSRNGNSYTLGIN
FB_2           --------MMKRFEEWMAEYGRVYKDNDEKMRRFQIFKNNVNHIETFNSRNGNSYTLGIN
FB_3           SRDEPSDPMMKRFEEWMAEYGRVYKDNDEKMRRFQIFKNNVNHIETFNNRNGNSYTLGIN
Clustal Consensus  ::: **.** ::.::**: ***: **:***:*:::*:  *.:* *** **:*

                    70        80        90       100       110       120
               ....|....|....|....|....|....|....|....|....|....|....|....|
Pro-papain     VFADMSNDEFKEKYTGSIAGNYTTTELSYEEVLNDGDVNI---PEYVDWRQKGAVTPVKN
FB_1           QFTDMTKSEFVAQYTG-VSLPLN---IEREPVVSFDDVNISAVPQSIDWRDYGAVNEVKN
FB_2           QFTDMTKSEFVAQYTG-VSLPLN---IEREPVVSFDDVNISAVPQSIDWRDYGAVNEVKN
FB_3           KFTDMTNNEFVAQYTGGISRPLN---IEKEPVVSFDDVNISAVGQSIDWRDYGAVTEVKD
Clustal Consensus  *:**::.**  :*** ::   .  :. * *:. .****   : :***: ***. **:

                   130       140       150       160       170       180
               ....|....|....|....|....|....|....|....|....|....|....|....|
Pro-papain     QGSCGSAWAFSAVSTIESIIKIRTGNLNEYSEQELLDCDRRSYGCNGGYPWSALQLVAQY
FB_1           QNPCGSCWAFAAIATVEGIYKIKTGYLVSLSEQEVLDCA-VSYGCKGGWVNKAYDFIISN
FB_2           QNPCGSCWSFAAIATVEGIYKIKTGYLVSLSEQEVLDCA-VSYGCKGGWVNKAYDFIISN
FB_3           QNPCGSCWAFSAIATVEGIYKIVTGYLVSLSEQEVLDCA-VSNGCDGGFVDNAYDFIISN
Clustal Consensus  *..***.*:*:*::*:*.* ** ** *  . ****:***   * **.**:  .* ::: .

                   190       200       210       220       230       240
               ....|....|....|....|....|....|....|....|....|....|....|....|
Pro-papain     -GIHYRNTYPYEGVQRYCRSREKGPYAAKTDGVRQVQPYNEGALLYSIANQPVSVVLEAA
FB_1           NGVTTEENYPYKAYQGTCNAN-SFPNSAYITGYSYVRRNDERSMMYAVSNQPIAALIDAS
FB_2           NGVTTEENYPYQAYQGTCNAN-SFPNSAYITGYSYVRRNDERSMMYAVSNQPIAALIDAS
FB_3           NGVASEADYPYQAYQGDCAAN-SWPNSAYITGYSYVRSNDESSMKYAVWNQPIAAAIDAS
Clustal Consensus  *: .   ***:. *  * * :. . * :*  *   *:  :* :: *:: ***::. .:*:

                   250       260       270       280       290       300
               ....|....|....|....|....|....|....|....|....|....|....|....|
Pro-papain     GKDFQLYRGGIFVGPCGNKVDHAVAAVGYGPN-----YILIRNSWGTGWGENGYIRIKRG
FB_1           -ENFQYYNGGVFSGPCGTSLNHAITIIGYGQDSSGTKYWIVRNSWGSSWGEGGYVRMARG
FB_2           -ENFQYYNGGVFSGPCGTSLNHAITIIGYGQDSSGTKYWIVRNSWGSSWGEGGYVRMARG
FB_3           GDNFQYYNGGVFSGPCGTSLNHAITIIGYGQDSSGTQYWIVKNSWGSSWGERGYIRMARG
Clustal Consensus  .:** *.**:* ****..::**:: :*** :    * :::****:.*** **:*: **

                   310       320
               ....|....|....|....|
Pro-papain     TGNSYGVCGLYTSSFYPVKN
FB_1           VSSSSGVCGIAMSPLFPTLQ
FB_2           VSSSSGVCGIAMAPLFPTLQ
FB_3           VSSS-GLCGIAMDPLYPTLQ
Clustal Consensus  ...* *:**:   .::*. :
```

Figure 4.24   Sequence alignment of pro-papain with FB_1, FB_2 and FB_3. The interactions between pro-region and the mature part of fruit bromelain was identified using pro-papain as reference (a) first hydrophobic interaction involving Phe248, Phe256, Cys260, Trp284, Trp288 and Cys307 from the catalytic domain and Phe70, Phe78 and Tyr82 from the pro-domain (b) residues involving in second hydrophobic interaction are Tyr251, Phe256 and Pro259 from the mature segment with Tyr52 (His52 in FB_1, FB_2 and FB_3), Trp64 (Thr64 in FB_1, FB_2 and FB_3) and Leu65 of pro-peptide (c) the pro-peptide forming hydrogen bonds with Ala244 (Ser244 in FB_1, FB_2 and FB_3), Lys246 (Glu246 in FB_1 and FB_2; Asp246 in FB_3), Asp247 (Asn247 in FB_1, FB_2 and FB_3) and Gln249 from the catalytic domain and H3. The residues involved in the interaction are indicated in the red box.

(b)

```
                      10        20        30        40        50        60
              ....|....|....|....|....|....|....|....|....|....|....|....|
Pro-papain    NDLTSTERLIQLFESWMLKHNKIYKNIDEKIYRFEIFKDNLKYIDETNKKN-NSYWLGLN
FB_1          SRDEPNDPMMKRFEEWMAEYGRVYKDNDEKMRRFQIFKNNVKHIETFNSRNGNSYTLGIN
FB_2          --------MMKRFEEWMAEYGRVYKDNDEKMRRFQIFKNNVNHIETFNSRNGNSYTLGIN
FB_3          SRDEPSDPMMKRFEEWMAEYGRVYKDNDEKMRRFQIFKNNVNHIETFNNRNGNSYTLGIN
Clustal Consensus       ::: **.** ::.::**: ***: **:***:*:::*:  *.:* *** **:*


                      70        80        90        100       110       120
              ....|....|....|....|....|....|....|....|....|....|....|....|
Pro-papain    VFADMSNDEFKEKYTGSIAGNYTTTELSYEEVLNDGDVNI---PEYVDWRQKGAVTPVKN
FB_1          QFTDMTKSEFVAQYTG-VSLPLN---IEREPVVSFDDVNISAVPQSIDWRDYGAVNEVKN
FB_2          QFTDMTKSEFVAQYTG-VSLPLN---IEREPVVSFDDVNISAVPQSIDWRDYGAVNEVKN
FB_3          KFTDMTNNEFVAQYTGGISRPLN---IEKEPVVSFDDVNISAVGQSIDWRDYGAVTEVKD
Clustal Consensus  *:**:.** :*** ::   :. * *:. .****   : :***: ***. **:


                      130       140       150       160       170       180
              ....|....|....|....|....|....|....|....|....|....|....|....|
Pro-papain    QGSCGSAWAFSAVSTIESIIKIRTGNLNEYSEQELLDCDRRSYGCNGGYPWSALQLVAQY
FB_1          QNPCGSCWAFAAIATVEGIYKIKTGYLVSLSEQEVLDCA-VSYGCKGGWVNKAYDFIISN
FB_2          QNPCGSCWSFAAIATVEGIYKIKTGYLVSLSEQEVLDCA-VSYGCKGGWVNKAYDFIISN
FB_3          QNPCGSCWAFSAIATVEGIYKIVTGYLVSLSEQEVLDCA-VSNGCDGGFVDNAYDFIISN
Clustal Consensus *..***.*:.*:*::*:*.* ** ** * . ****:***   * **.**:  .* ::: .


                      190       200       210       220       230       240
              ....|....|....|....|....|....|....|....|....|....|....|....|
Pro-papain    -GIHYRNTYPYEGVQRYCRSREKGPYAAKTDGVRQVQPYNEGALLYSIANQPVSVVLEAA
FB_1          NGVTTEENYPYKAYQGTCNAN-SFPNSAYITGYSYVRRNDERSMMYAVSNQPIAALIDAS
FB_2          NGVTTEENYPYQAYQGTCNAN-SFPNSAYITGYSYVRRNDERSMMYAVSNQPIAALIDAS
FB_3          NGVASEADYPYQAYQGDCAAN-SWPNSAYITGYSYVRSNDESSMKYAVWNQPIAAAIDAS
Clustal Consensus  *:  . ***:. *  * :.. * :*   *   *: :* :: *:: ***::. ::*:


                      250       260       270       280       290       300
              ....|....|....|....|....|....|....|....|....|....|....|....|
Pro-papain    GKDFQLYRGGIFVGPCGNKVDHAVAAVGYGPN-----YILIRNSWGTGWGENGYIRIKRG
FB_1          -ENFQYYNGGVFSGPCGTSLNHAITIIGYGQDSSGTKYWIVRNSWGSSWGEGGYVRMARG
FB_2          -ENFQYYNGGVFSGPCGTSLNHAITIIGYGQDSSGTKYWIVRNSWGSSWGEGGYVRMARG
FB_3          GDNFQYYNGGVFSGPCGTSLNHAITIIGYGQDSSGTQYWIVKNSWGSSWGERGYIRMARG
Clustal Consensus .:** *.**:* ****..::**:: :*** :      * :::****:.*** **:*: **


                      310       320
              ....|....|....|....|
Pro-papain    TGNSYGVCGLYTSSFYPVKN
FB_1          VSSSSGVCGIAMSPLFPTLQ
FB_2          VSSSSGVCGIAMAPLFPTLQ
FB_3          VSSS-GLCGIAMDPLYPTLQ
Clustal Consensus ...* *:**:   .::*. :
```

Figure 4.24   Continued.

72

(c)

```
                         10        20        30        40        50        60
                ....|....|....|....|....|....|....|....|....|....|....|....|
Pro-papain      NDLTSTERLIQLFESWMLKHNKIYKNIDEKIYRFEIFKDNLKYIDETNKKN-NSYWLGLN
FB_1            SRDEPNDPMMKRFEEWMAEYGRVYKDNDEKMRRFQIFKNNVKHIETFNSRNGNSYTLGIN
FB_2            --------MMKRFEEWMAEYGRVYKDNDEKMRRFQIFKNNVNHIETFNSRNGNSYTLGIN
FB_3            SRDEPSDPMMKRFEEWMAEYGRVYKDNDEKMRRFQIFKNNVNHIETFNNRNGNSYTLGIN
Clustal Consensus       ::: **.** ::.::**: ***: **:***:*:::*:  *.:* *** **:*

                         70        80        90        100       110       120
                ....|....|....|....|....|....|....|....|....|....|....|....|
Pro-papain      VFADMSNDEFKEKYTGSIAGNYTTTELSYEEVLNDGDVNI---PEYVDWRQKGAVTPVKN
FB_1            QFTDMTKSEFVAQYTG-VSLPLN---IEREPVVSFDDVNISAVPQSIDWRDYGAVNEVKN
FB_2            QFTDMTKSEFVAQYTG-VSLPLN---IEREPVVSFDDVNISAVPQSIDWRDYGAVNEVKN
FB_3            KFTDMTNNEFVAQYTGGISRPLN---IEKEPVVSFDDVNISAVGQSIDWRDYGAVTEVKD
Clustal Consensus  *:**:.** :*** ::   :. * *:. .****  : :***: ***. **:

                         130       140       150       160       170       180
                ....|....|....|....|....|....|....|....|....|....|....|....|
Pro-papain      QGSCGSAWAFSAVSTIESIIKIRTGNLNEYSEQELLDCDRRSYGCNGGYPWSALQLVAQY
FB_1            QNPCGSCWAFAAIATVEGIYKIKTGYLVSLSEQEVLDCA-VSYGCKGGWVNKAYDFIISN
FB_2            QNPCGSCWAFAAIATVEGIYKIKTGYLVSLSEQEVLDCA-VSYGCKGGWVNKAYDFIISN
FB_3            QNPCGSCWAFAIATVEGIYKIVTGYLVSLSEQEVLDCA-VSNGCDGGFVDNAYDFIISN
Clustal Consensus *..***.*:*:*::*:*.* ** ** * . ****:***  * **.**: .* ::: .

                         190       200       210       220       230       240
                ....|....|....|....|....|....|....|....|....|....|....|....|
Pro-papain      -GIHYRNTYPYEGVQRYCRSREKGPYAAKTDGVRQVQPYNEGALLYSIANQPVSVVLEAA
FB_1            NGVTTEENYPYKAYQGTCNAN-SFPNSAYITGYSYVRRNDERSMMYAVSNQPIAALIDAS
FB_2            NGVTTEENYPYQAYQGTCNAN-SFPNSAYITGYSYVRRNDERSMMYAVSNQPIAALIDAS
FB_3            NGVASEADYPYQAYQGDCAAN-SWPNSAYITGYSYVRSNDESSMKYAVWNQPIAAAIDAS
Clustal Consensus  *:   . ***:. *  * *:. . * :*   *   *:  :* :: *:: ***::. ::*:

                         250       260       270       280       290       300
                ....|....|....|....|....|....|....|....|....|....|....|....|
Pro-papain      GKDEQLYRGGIFVGPCGNKVDHAVAAVGYGPN-----YILIRNSWGTGWGENGYIRIKRG
FB_1            -ENFQYYNGGVFSGPCGTSLNHAITIIGYGQDSSGTKYWIVRNSWGSSWGEGGYVRMARG
FB_2            -ENFQYYNGGVFSGPCGTSLNHAITIIGYGQDSSGTKYWIVRNSWGSSWGEGGYVRMARG
FB_3            GDNFQYYNGGVFSGPCGTSLNHAITIIGYGQDSSGTQYWIVKNSWGSSWGERGYIRMARG
Clustal Consensus .:** *.**:* ****.:::**:: :*** :     * :::****:.*** **:*: **

                         310       320
                ....|....|....|....|
Pro-papain      TGNSYGVCGLYTSSFYPVKN
FB_1            VSSSSGVCGIAMSPLFPTLQ
FB_2            VSSSSGVCGIAMAPLFPTLQ
FB_3            VSSS-GLCGIAMDPLYPTLQ
Clustal Consensus ...* *:**:   .::*. :
```

Figure 4.24   Continued.

```
                         70        80        90        100       110       120
                ....|....|....|....|....|....|....|....|....|....|....|....|
Pro-papain      VFADMSNDEFKEKYTGSIAGNYTTTELSYEEVLNDGDVNI---PEYVDWRQKGAVTPVKN
FB_1            QFTDMTKSEFVAQYTG-VSLPLN---IEREPVVSFDDVNISAVPQSIDWRDYGAVNEVKN
FB_2            QFTDMTKSEFVAQYTG-VSLPLN---IEREPVVSFDDVNISAVPQSIDWRDYGAVNEVKN
FB_3            KFTDMTNNEFVAQYTGGISRPLN---IEKEPVVSFDDVNISAVGQSIDWRDYGAVTEVKD
```

Figure 4.25   Sequence alignment of pro-papain with FB_1, FB_2 and FB_3. The replacement of Ile86 in papain to Val86 in FB_1, FB_2 and FB_3 is indicated in the red box.

73

4.3.4.3 Binding subsites of fruit bromelain

It was found that the residues shaping papain subsites at S1 (His160), S2 (Trp183), and S3 (Glu19) are conserved in fruit bromelain (Figure 4.26). Among these subsites, S2 is particularly important because it has a pronounced effect in determining the substrate specificity of cysteine protease (Khouri et al., 1991; Nägler et al., 1999). In this case, both papain and fruit bromelain have tryptophan harbouring an indole side chain which resides in their S2 pocket and interacts preferentially with the substrate containing a hydrophobic side chain such as phenylalanine and valine (Berger & Schechter, 1970). The high similarity in the S2 subsites may explain why fruit bromelain and papain are cleaved more efficiently on Z-Phe-Arg-NMec than Z-Arg-Arg-NMec (Rowan, Buttle, & Barrett, 1990; Tchoupé, Moreau, Gauthier, & Bieth, 1991). Conversely, it was found that glutamine in the S3 subsites carry the carbonyl group in its backbone which prefer to bind with polar aromatic amino acid residues such as tyrosine (Cordara et al., 2016; Taralp et al., 1995). Herein, this explains why the S3 subsites have least affinity with positively-charged histidine (Portaro et al., 2000).

Figure 4.26   Sequence alignment of papain with FB_1, FB_2 and FB_3. The papain subsites are indicated in the red box and labelled. Identical amino acid residues are denoted with "*", conserved substitutions are marked by ":" and semi-conserved substitutions are indicated with ".".

On the contrary, the remained subsites have less-defined specificity. For instance, S1 subsites have a high tolerance in subsites specificity and accept a wide range of amino acid residues (Alves et al., 2001; Fox, Mason, Storer, & Mort, 1995). Nevertheless, several studies demonstrated that the most suitable amino acid residue for this subsite is positively-charged amino acid residues such as arginine (Papamichael, Roustas, & Bieth, 2017). Furthermore, amino acid residue substitutions were observed at primed subsites S2' (Figure 4.26). The inconsistency in S2' is common among CA1 family members due to differences in interaction surface (Cordara et al., 2016). In S2' subsites, papain showed a preference for the substrates with hydrophobic residues including leucine, phenylalanine, tyrosine and tryptophan as well as basic amino acid residues such as histidine and lysine (Portaro et al., 2000). The substitution of Asp→Asn in the S2' subsites may relate to higher proteolytic activity in fruit bromelain than papain (Müller et al., 2016; Vel & Stanley, 2015). Since the papain binding subsites except S2' are conserved in FB_1, FB_2 and FB_3, similar binding conditions in papain should apply to the fruit bromelain structures (Figure 4.27).
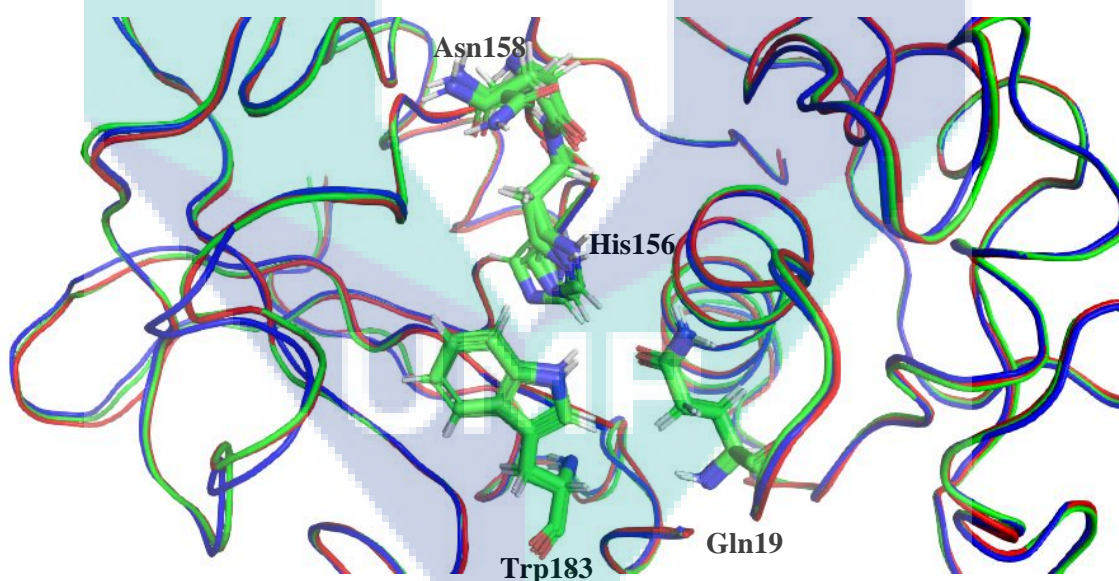
Figure 4.27   Superimposition of FB_1 (red), FB_2 (green) and FB_3 (blue). The fruit bromelain structures are represented in the ribbon. The binding residues His160, Trp183, Glu19 and Asn158 are represented as sticks labelled according papain numbering. Cys25, His156 and Asn172 forming catalytic triad while Gln23 assist in catalysis.

4.3.5   Thermostability of fruit bromelain

The influence of temperature to fruit bromelain was evaluated using the casein enzymatic assay. Optimum temperature of fruit bromelain and ability to retain its activity

after one hour of heat incubation was investigated. Subsequently, the previous generated models FB_1, FB_2 and FB_3 were simulated at an elevated temperature to interpret the conformational changes at the elevated temperature.

4.3.5.1 Influence of temperature to fruit bromelain activity and stability

In order to understand the influence of temperature on fruit bromelain, the optimum temperature of fruit bromelain was determined using casein enzymatic assay. It can be seen that the proteolytic activity of fruit bromelain increased from 2.00 ± 0.06 U/mL at 40 °C to 2.11 ± 0.05 U/mL at 50 °C (Figure 4.28). At 60 °C, fruit bromelain achieved its highest activity which is 2.69 ± 0.01 U/mL. The increase of catalytic activity can be explained by fruit bromelain gaining sufficient kinetic energy to boost its molecular movement and hence is able to conduct catalysis with higher efficiency (Robinson, 2015). However, the continuous raising of temperature results in reduced activity of 2.09 ± 0.10 U/mL at 70 °C and 1.67 ± 0.13 U/mL at 80 °C, respectively. This is because the increase of kinetic energy level at a higher temperature exceeds the thermal threshold and causes a disruption in the fruit bromelain structure structure (Bianco, Iskrov, & Franzese, 2012; R. M. Daniel, Dines, & Petach, 1996). The trend obtained in this study concurs with the findings from Corzo et al. (2012) and Kothare et al. (2017) in which the optimum temperature of fruit bromelain is 60 °C when using casein as a substrate.
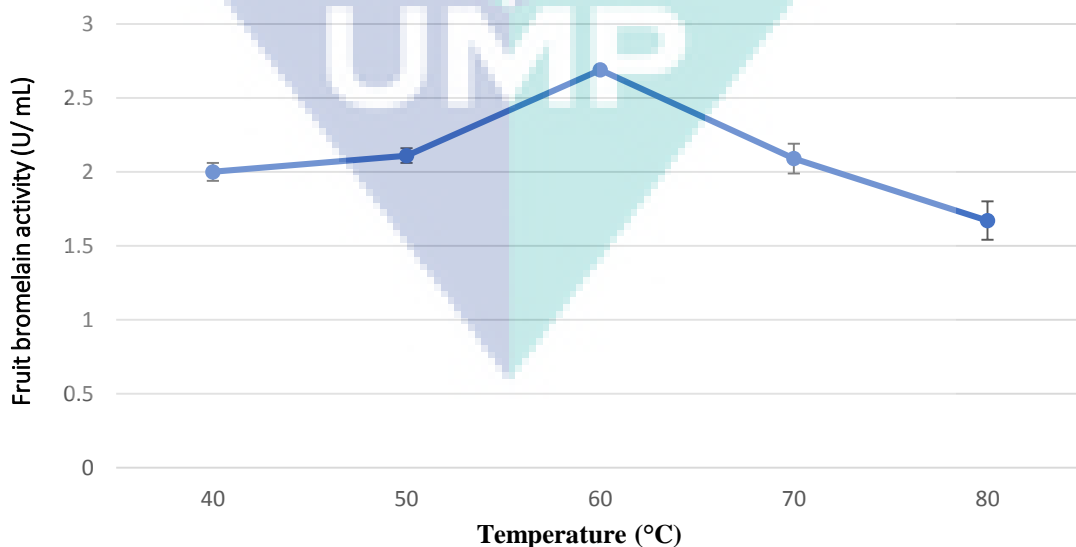
Figure 4.28   Enzymatic activity of the fruit bromelain at different temperatures.

Besides that, it is also important to study the stability of fruit bromelain and how long the fruit bromelain can maintain its activity at an acceptable level under a certain temperature. This information is particularly useful to lengthen the lifespan of fruit bromelain by avoiding from using it at an unsuitable temperature for a fixed duration (e.g. one hour). The fruit bromelain exhibited proteolytic activity of 1.54 ± 0.04 U/mL after one hour of incubation at 40°C (Figure 4.29). After that, the activity of fruit bromelain declined rapidly with the increase of the incubation temperature (50-70 °C) and eventually, catalytic activity was completely lost when the incubation temperature was raised to 80 °C. Apparently, it is not suitable to use fruit bromelain at a high temperature for a long period of time as its hydrolytic activity is largely decreased due to thermal denaturation. This result is consistent with other previous studies in which fruit bromelain is totally inactivated after being incubated at 80 °C (Jutamongkon & Charoenrein, 2010; Ramalingam et al., 2012). In addition, protein thermal denaturation is time dependent (Robinson, 2015). This implies that fruit bromelain is denatured with a faster rate at a higher temperature (80 °C) than a lower temperature (40 °C).
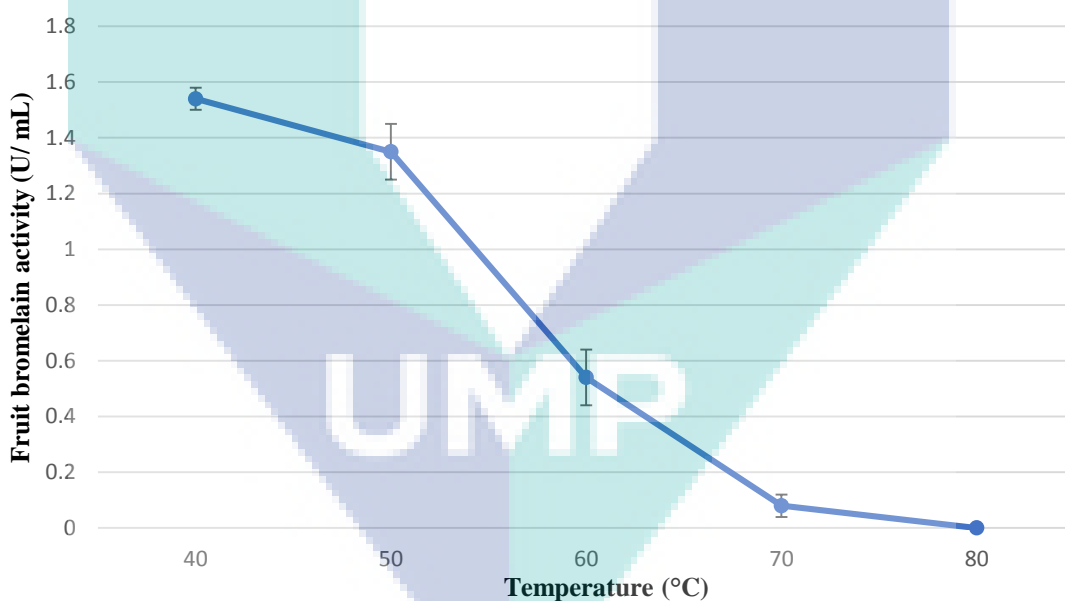
Figure 4.29   Enzymatic activity of the crude fruit bromelain after incubated for one hour at different temperatures.

4.3.5.2 Thermostability analysis of fruit bromelain using molecular dynamics simulation

In section 4.3.5.1, it was shown that fruit bromelain's activity and stability are temperature dependent. The elevation of temperature is considered as two forces which

act on fruit bromelain simultaneously in opposite directions (Robinson, 2015). First, the elevation of temperature causes a rapid conformational change of protein due to the weakening of stabilising interactions within the native folded protein and which lead to the enhancement of enzyme catalysis; however, the protein structure is disrupted once there is no sufficient stabilising interactions to maintain a proper protein structure (Fields et al., 2015). In this section, the dynamic behaviour of fruit bromelain at elevated temperature was explored using molecular dynamics simulation. The fruit bromelain models generated from section 3.4 were simulated at 313 K (equivalent to 40 °C), 333 K (equivalent to 60 °C) and 353 K (equivalent to 80 °C) respectively in molecular dynamics simulation. The pro-peptide of fruit bromelain was excluded in this analysis because it is cleaved off once fruit bromelain reached maturation. This ensures the accuracy of the information obtained by only studying the mature part of fruit bromelain.

In this study, RMSD reflects the thermal motion of fruit bromelain at the investigated temperature. The analysis of FB_1 using molecular dynamics simulation (Figure 4.30 (a)) exhibited that at 313 K FB_1 maintained an RMSD value of 1 Å till 25 ns and was inclined to 1.5 Å with a slight drop to 1 Å observed at 80 ns and kept constant at 1.5 Å again until the end of simulation. At 333 K, the RMSD value of FB_1 was kept around 1.5 Å till 40 ns and raised to 2 Å for the last 60 ns of simulation. Meanwhile, at 353 K, the RMSD value of FB_1 was inclined to ~2 Å throughout the simulation. Furthermore, the analysis of FB_2 (Figure 4.30 (b)) revealed that at 313 K, FB_2 demonstrated RMSD value of ~1.5 till 50 ns and slightly fluctuated until it reached 2 Å at the last 20 ns of simulation. At 333 K, FB_2 exhibited a constant RMSD value of 1.5 Å during 0-70 ns which then escalated to ~2 Å between 70-75 ns and decreased to ~1.5 Å again between 80-100 ns while at 353 K, the RMSD value of FB_2 climbed up sharply to ~3.5 Å after 20 ns of the simulation. In addition, the analysis of FB_3 demonstrated that RMSD values of FB_3 were slightly fluctuated with a similar pattern at all the investigated temperatures and finally converged to 2 Å at the last 10 ns of the simulation (Figure 4.30 (c)).
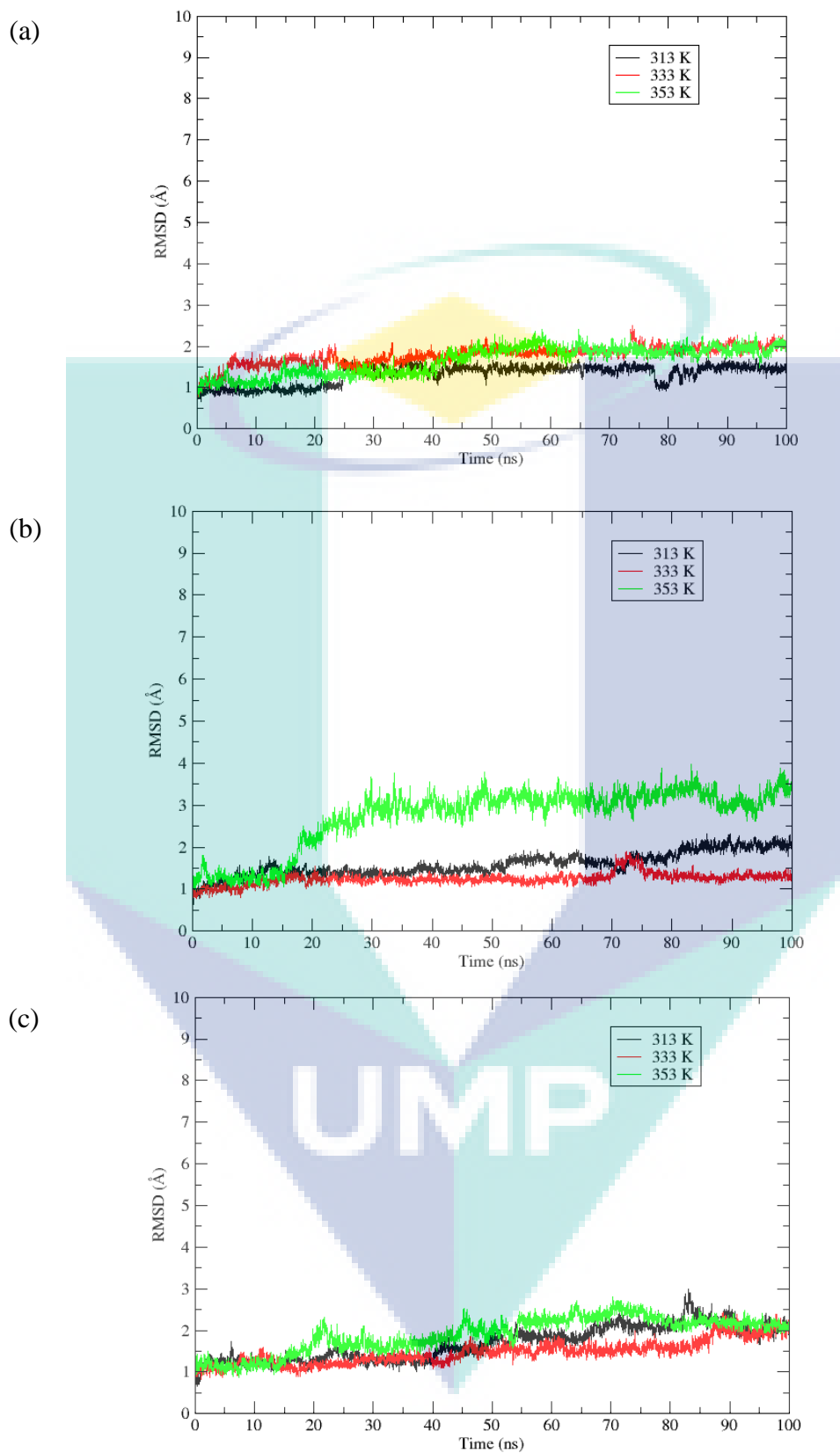
Figure 4.30   RMSD of protein backbone as a function of time at 313 K, 333 K and 353 K of (a) FB_1 (b) FB_2 and (c) FB_3.

The changes of RMSD during the course of simulation indicated that the protein backbone of all fruit bromelain structures was altered as they were moved to a new conformation to retain stability and flexibility as the temperature inclined (Fields et al., 2015; Salleh, Rahim, Rahman, Leow, & Basri, 2012). The overall changes of RMSD (1.5-2.0 Å) is considered small as the depiction of the new fruit bromelain conformations (except FB_2 at 353 K) closely matched to their starting structures (Kato, Nakayoshi, Fukuyoshi, Kurimoto, & Oda, 2017). Therefore, the RMSD result suggests that temperature poses the least effect to FB_1, followed by FB_3 and FB_2 respectively. Moreover, no significant deviation was observed after the fruit bromelain reached RMSD plateau demonstrating that the simulation time of 100 ns is sufficient to allow fruit bromelain to acquire a new stable conformation.

The compactness of protein is another indicator to measure the stability of protein (Paul, Hazra, Barman, & Hazra, 2014). The effect of temperature to the overall dimension of fruit bromelain was gleaned from the Rg analysis. FB_1 displayed a similar dimension of ~16.5 Å at all temperatures (Figure 4.31 (a)). Meanwhile, FB_2 was able to maintain its size of 16.5 Å at 313 K and 333 K respectively (Figure 4.31 (b)). At 353 K, the Rg value of FB_2 was spread to 17 Å between 20-100 ns which corresponds to its RMSD changes that occurred at the same time frame. For FB_3, the Rg value was maintained at a steady value of 16.5 Å at 313 K and 333 K respectively (Figure 4.31 (c)). At 353 K, the Rg value of FB_3 decreased to 16 Å between 50-100 ns. Overall, the Rg of fruit bromelain models remained constant throughout the simulation which implies that the fruit bromelain structures are capable to maintain their original compactness as the temperature rises. Hence, the entire size of fruit bromelain is kept constant during the course of the simulation. Exceptions were observed on FB_2 and FB_3 at 353 K. The changes in Rg suggests that there is a repacking of molecule in response to the changes of temperature due to the loosening of the structural network and/or collapse of the protein system at 353 K (Gu, Tong, Sun, & Lin, 2019; Wu et al., 2015). Similar to RMSD analysis, the compactness analysis of fruit bromelain via Rg also suggests that FB_1 poses the highest stability than FB_2 and FB_3 at a higher temperature.
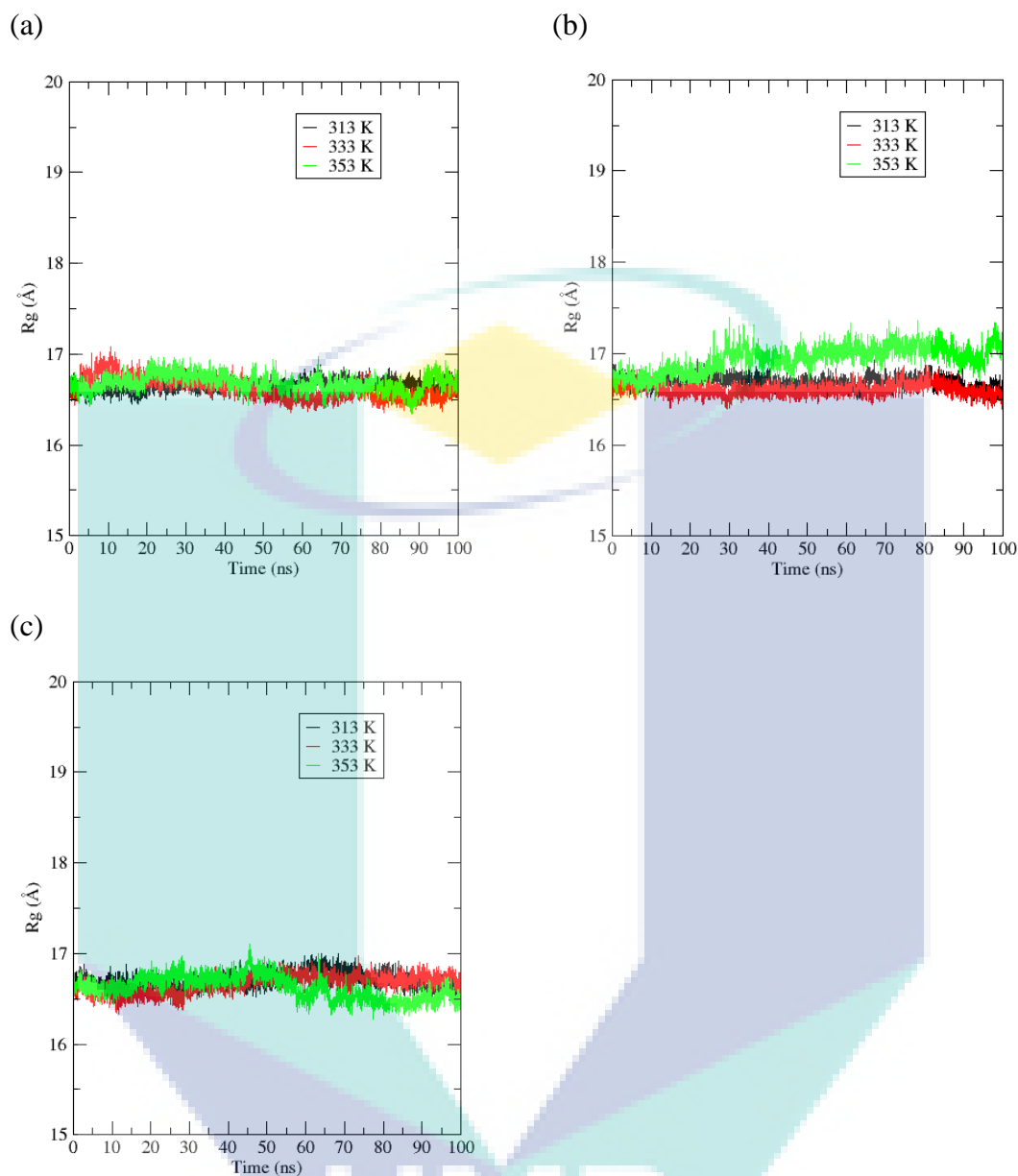
Figure 4.31   Radius of gyration as a function of time at 313 K, 333 K and 353 K of (a) FB_1 (b) FB_2 and (c) FB_3.

Protein folding is driven by hydrophobic effect and is temperature dependent (Camilloni et al., 2016; Pucci & Rooman, 2017). Under normal conditions, the hydrophilic residues are usually on the protein surface while hydrophobic residues are generally buried inside the protein away from the aqueous environment (Ramli et al., 2018). If protein denaturation occurs, the hydrophobic region will be exposed to the solvent (Paul et al., 2014). The influence of temperature to the distribution of hydrophilic and hydrophobic residues of fruit bromelain was examined by the changes of SASA. The SASA value of FB_1 ranges between ~97.5-110 nm^2 at 313K, ~95~110 nm^2 at 333 K and

~97.5-110 nm^2 at 353 K (Figure 4.32 (a)). For FB_2, the SASA value fluctuated between ~97.5-110 nm^2 at 313 K, ~97.5-105 nm^2 at 333K and between ~100-115 nm^2 at 353 K (Figure 4.32 (b)). Meanwhile, the SASA of FB_3 maintained between ~100-110 nm^2 at 313 K, between ~95-105 nm^2 at 333 K; at 353K, FB_3 maintained a stable SASA between 100-110 nm^2 during first 40 ns and decreased to 95-105 nm^2 until the end of simulation (Figure 4.32 (c)).
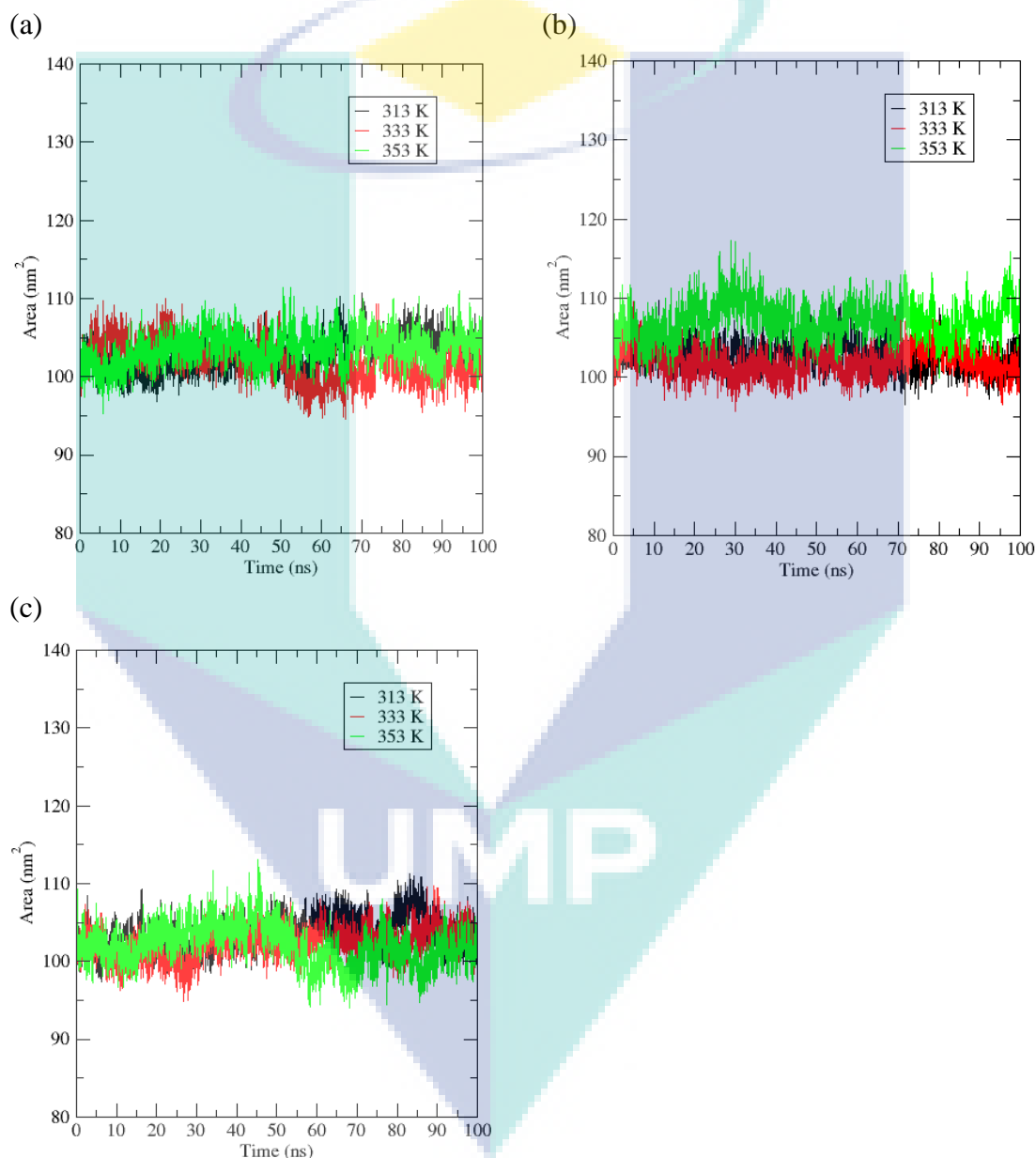
Figure 4.32   SASA of (a) FB_1 (b) FB_2 and (c) FB_3 as a function of time at 313 K, 333 K and 353 K.

The changes of the SASA of fruit bromelain models exhibited a similar pattern as their RMSD and Rg throughout the simulation. The SASA of FB_1 were similar, indicating its stability at all investigated temperatures. This also applies to FB_2 and FB_3 when simulated at 333 K and 353 K respectively. At this point, the protein residues are residing in their native fold without disruption. In contrast, the SASA of FB_2 was enlarged at 353 K as it was destructed or unfolded. As the number of hydrophobic molecules exposed to solvent increases, the surface area of FB_2 exposed to the solvent was also expanded. This is affirmed with the enlargement of Rg as hydrophobic molecules were spread on the surface instead of buried inside the hydrophobic core. In addition, the decrease of SASA of FB_3 is correlated to the overall reduction in size as discussed above. This further confirms the collapse of FB_3 which leads the aggregation at a high temperature as suggested by Rosa, Roberts, & Rodrigues (2017) in their study. During protein aggregation, residues congregate in the interior of the protein from the solvent and therefore results in the reduction of SASA (Mishra, Ranganathan, Jayaram, & Sattar, 2018).

The hydrogen bond is another important temperature dependent interaction in maintaining the stability of protein (Daniel et al., 1996; Pace et al., 2014). In contrast to the hydrophobic effect which directs the folding of the overall protein structure, hydrogen bond is linked to the directionality and specificity of intramolecular interactions within the protein structure including the formation of protein secondary and tertiary structure as well as the selectivity of protein interactions (Gao, Mei, & Zhang, 2015; Mukherjee, Majumdar, & Bhattacharyya, 2005). The starting structures of FB_1, FB_2 and FB_3 have intramolecular hydrogen bonds of 167, 155 and 160 respectively (Figure 4.33). The higher number of intramolecular hydrogen bonds conferred FB_1 and FB_3 higher resistant against heat denaturation (Gu et al., 2019; Paul et al., 2014; Vogt & Argos, 1997). In contrast, the least number of intramolecular hydrogen bonds present in FB_2 implies it has the lowest thermostability among the fruit bromelain models correlated with its distortion observed in RMSD, Rg and SASA at high temperature.
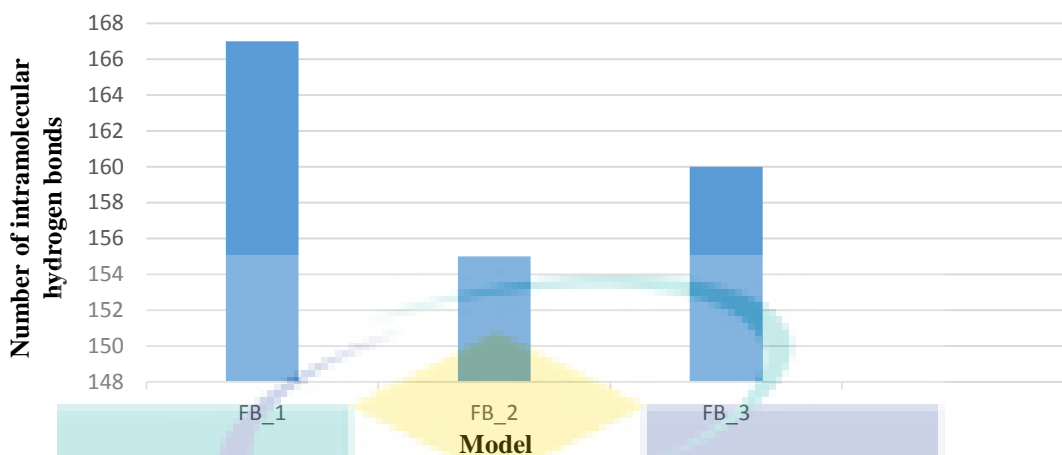
Figure 4.33   Number of intramolecular hydrogen bonds in the starting structure of FB_1, FB_2 and FB_3.

The number of intramolecular hydrogen bonds of fruit bromelain decreased during the simulation regardless of temperature. At 313 K, FB_1 retained an average number of 151.78 intramolecular hydrogen bonds at 313 K (Figure 4.34 (a)). This number further decreased to 140.54 at 333 K and 144.55 at 353 K respectively. FB_2 has an average number of 144.49 intramolecular hydrogen bonds at 313 K, 147.18 at 333 K and 139.09 at 353 K respectively (Figure 4.34 (b)). In addition, FB_3 was able to keep an average of 143.22, 144.52 and 141.45 intramolecular hydrogen bonds at 313 K, 333 K and 353 K respectively (Figure 4.34 (c)). This result indicates that the hydrogen bond network within the fruit bromelain structures was disrupted, which weakens the fruit bromelain stability as the temperature rises (Mallamace, Fazio, Mallamace, & Corsaro, 2018).

(a)   (b)
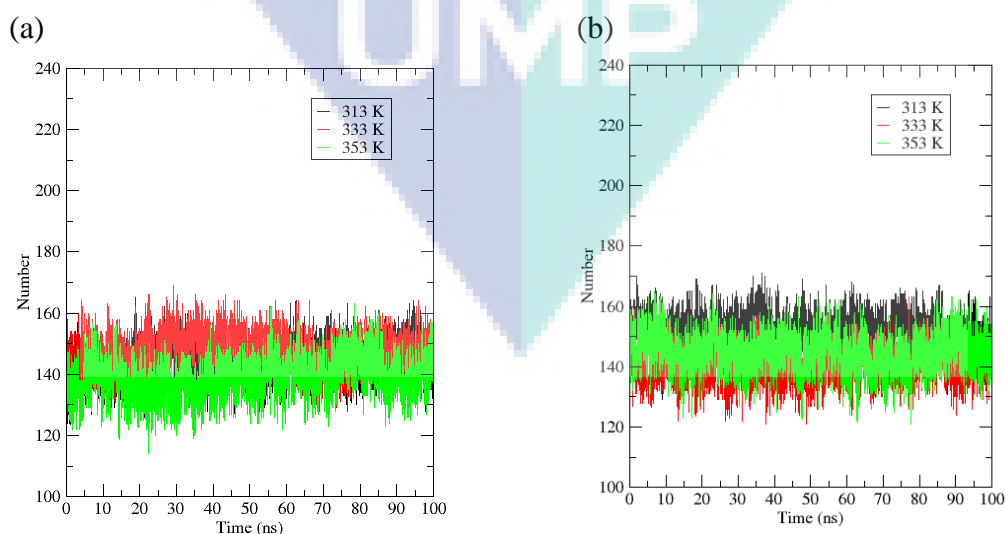
Figure 4.34   Number of intramolecular hydrogen bonds of (a) FB_1 (b) FB_2 and (c) FB_3 as a function of time at 313 K, 333 K and 353 K.

(c)
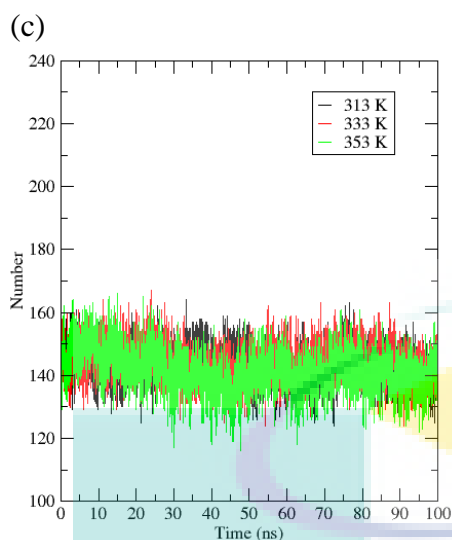
Figure 4.34   Continued.

At the same time, fruit bromelain models also formed intermolecular hydrogen bonds (protein-water hydrogen bonds) with water in the simulation box. FB_1 has an average number of 420 protein-water hydrogen bonds at 313 K, 428 at 333 K and 401 at 353 K (Figure 4.35 (a)). Meanwhile, FB_2 formed an average number of 434 intermolecular hydrogen bonds at 313 K which decreased to 412 at 333 K and 414 at 353 K (Figure 4.35 (b)). In addition, FB_3 demonstrated an average number of 454 protein-water hydrogen bonds at 313 K followed by 435 and 418 at 333 K 353 K respectively (Figure 4.35 (c)). Based on the result, the number of protein-water hydrogen bond decreased at the elevated temperature and this change is in line with the changes in SASA. This is because the raise of temperature increased the total surface area of hydrophobic residues exposed to the solvent, which decreased the interaction efficiency between fruit bromelain and water, and eventually leads to the deduction of the number of protein-water hydrogen bonds (Paul et al., 2014). On the contrary, less number of protein-water hydrogen bonds was formed in FB_3 at 353 K which is speculated to be due to aggregation, which buries the polar residues inside the hydrophobic core and hence decreases its overall contact with water (Pace et al., 2014).
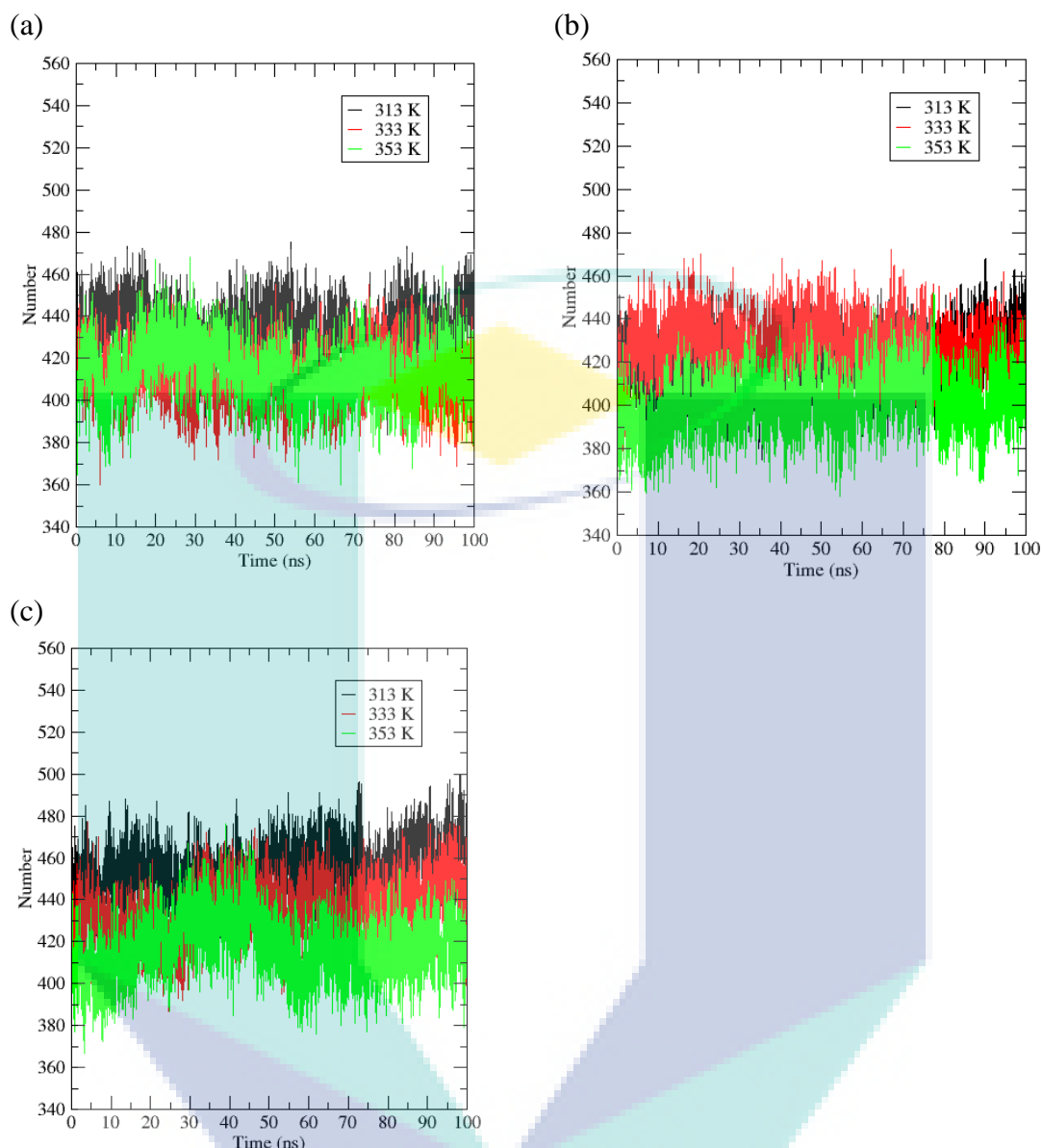
Figure 4.35   Number of protein-water hydrogen bonds of (a) FB_1 (b) FB_2 and (c) FB_3 as a function of time at 313 K, 333 K and 353 K.

Furthermore, the stability of fruit bromelain was also examined at residue level. RMSF depicted the flexibility and mobility of protein residues at different temperatures. In FB_1, fluctuation was observed at residue position 115-119, 137-143, 155-161, 186-198, 210-211, 255-256 and 265-269 (Figure 4.36 (a)). Meanwhile, FB_2 showed flexibility at residue position 107-111, 129-135, 147-153, 178-190, 202-203 and 257-261 (Figure 4.36 (b)). In addition, FB_3 displayed residue fluctuation at 116-120, 138-145, 156-162, 187-199, 211-212, 267-271, 281, 286 and 293-297 (Figure 4.36 (c)). From the RMSF analysis, it can be seen that fluctuations are more apparent at the first 100 protein residues in the mature part of fruit bromelain while the flexibility for the remaining part

is relatively smaller. Higher RMSF values in these regions imply that they have higher conformational flexibility and may become unstable to the increase of temperature (Du et al., 2017; Ning et al., 2018).
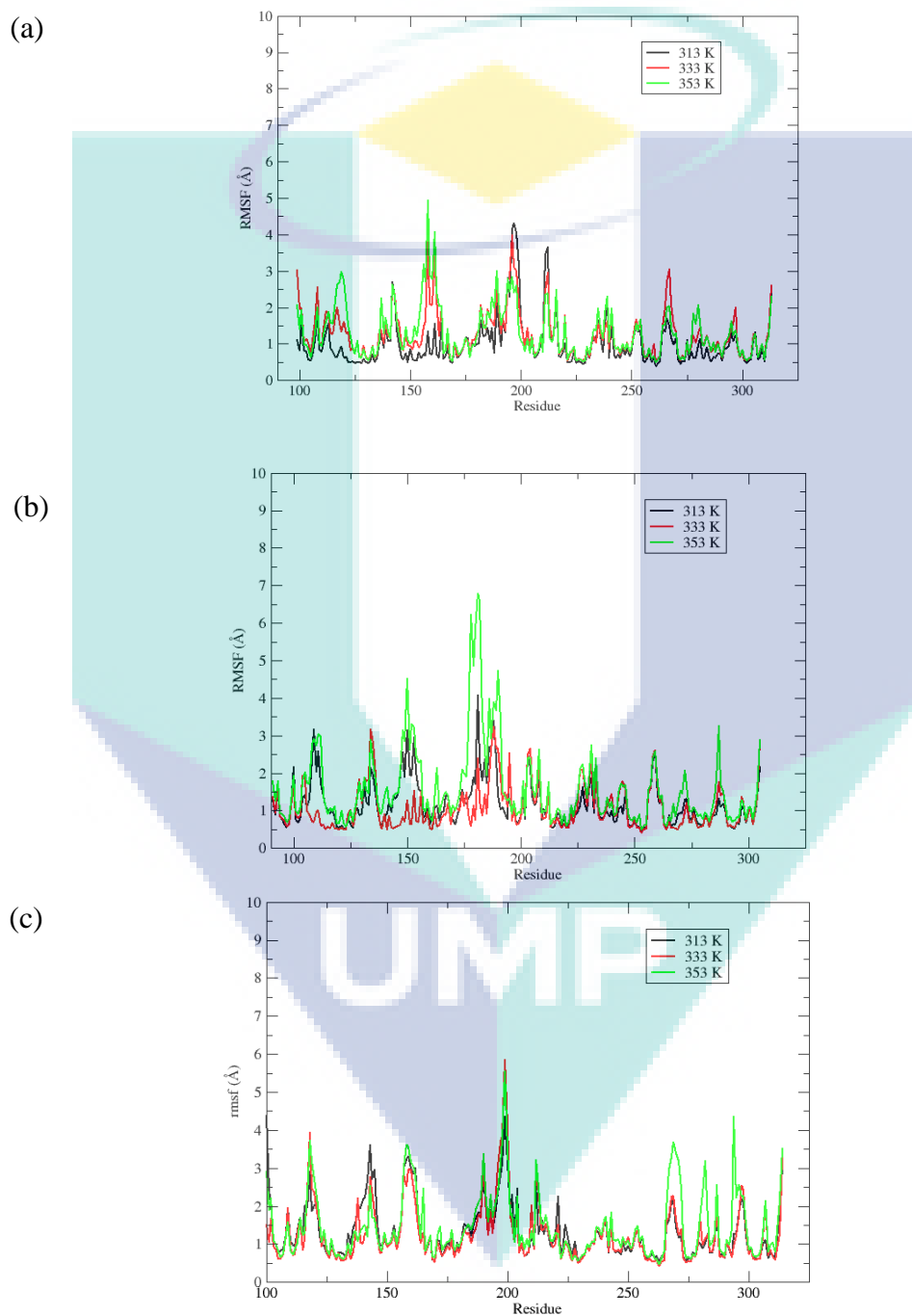
Figure 4.36   RMSF of (a) FB_1, (b) FB_2 and (c) FB_3 according to residue number at 313 K, 333 K and 353 K.

Superimposition of the fruit bromelain structures showed that the fluctuations occurred at a structurally equivalent position which is at the loop regions of the L-subdomain (Figure 4.37 & Figure 4.38). The α-helical content rich L-subdomain is more susceptible to heat denaturation because its overall structure is less compact (Burgos, Ochoa, & Perillo, 2019). Besides that, α-helix tends to be flexible due to its less rigid configuration (Emberly, Mukhopadhyay, Wingreen, & Tang, 2003). In contrast, the higher stability in the R-subdomain is ascribed to its anti-parallel β-barrel configuration which is difficult to unfold (Chaturvedi & Mahalakshmi, 2017; Perczel, Gaspari, & Csizmadia, 2005). The extensive inter-chain hydrogen bonding networks make these β-pleated-sheets as the most stable protein secondary structure (Cebe et al., 2013; Gessmann et al., 2011). Consequently, the L-subdomain is more readily destabilised by heat compared to the R-subdomain.
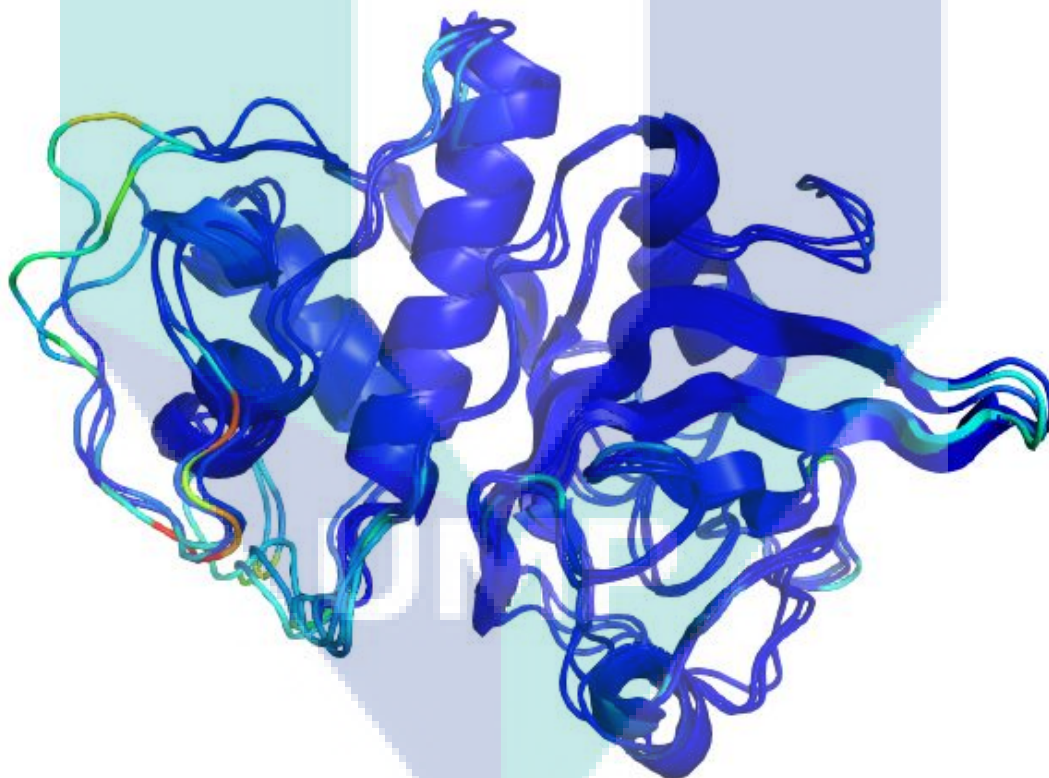
Figure 4.37   Superimposition of mature part of FB_1, FB_2 and FB_3. The fruit bromelain models are coloured according to the b-factors calculated from the simulation, blue representing the most stable region and red denoted as the region with highest fluctuation.

(a)

(b)
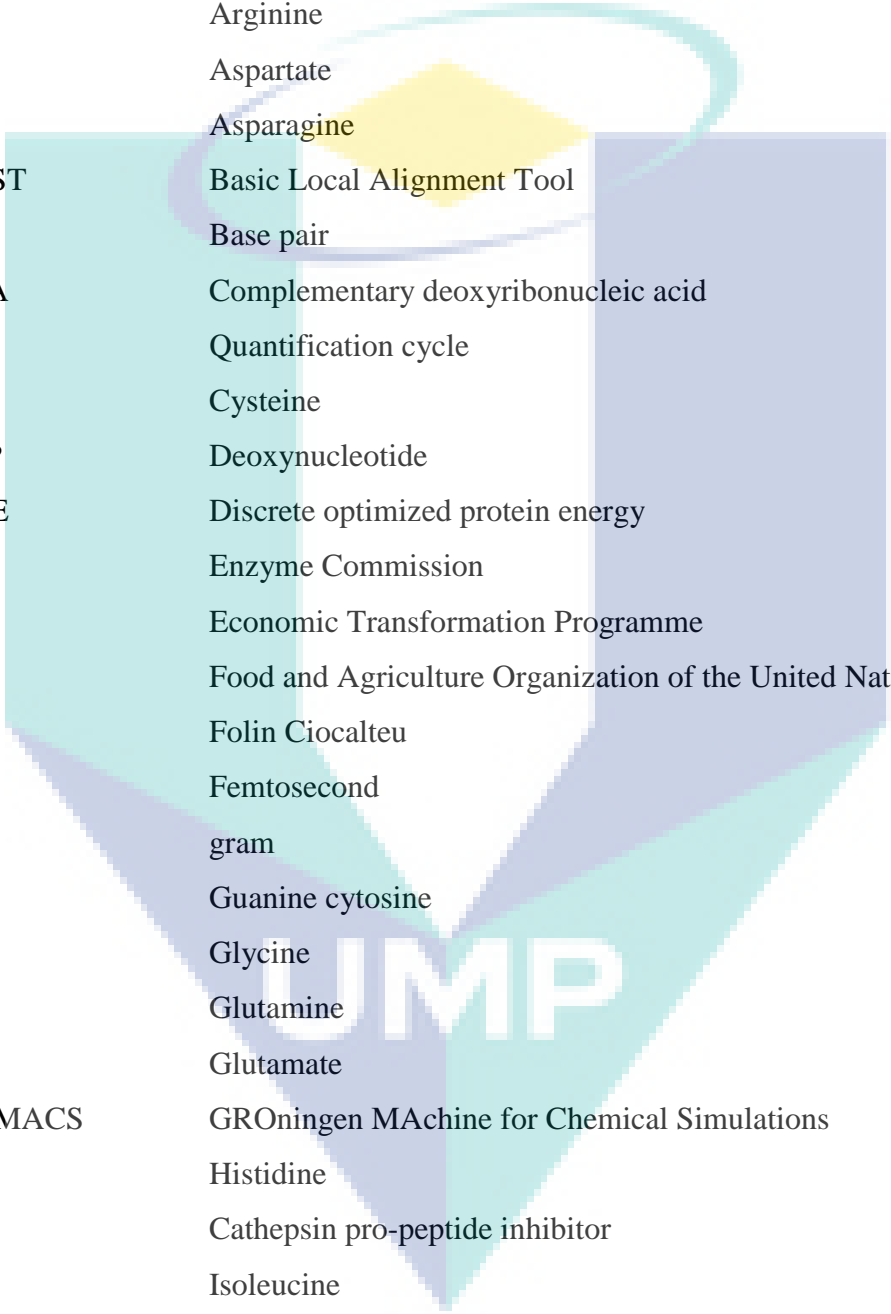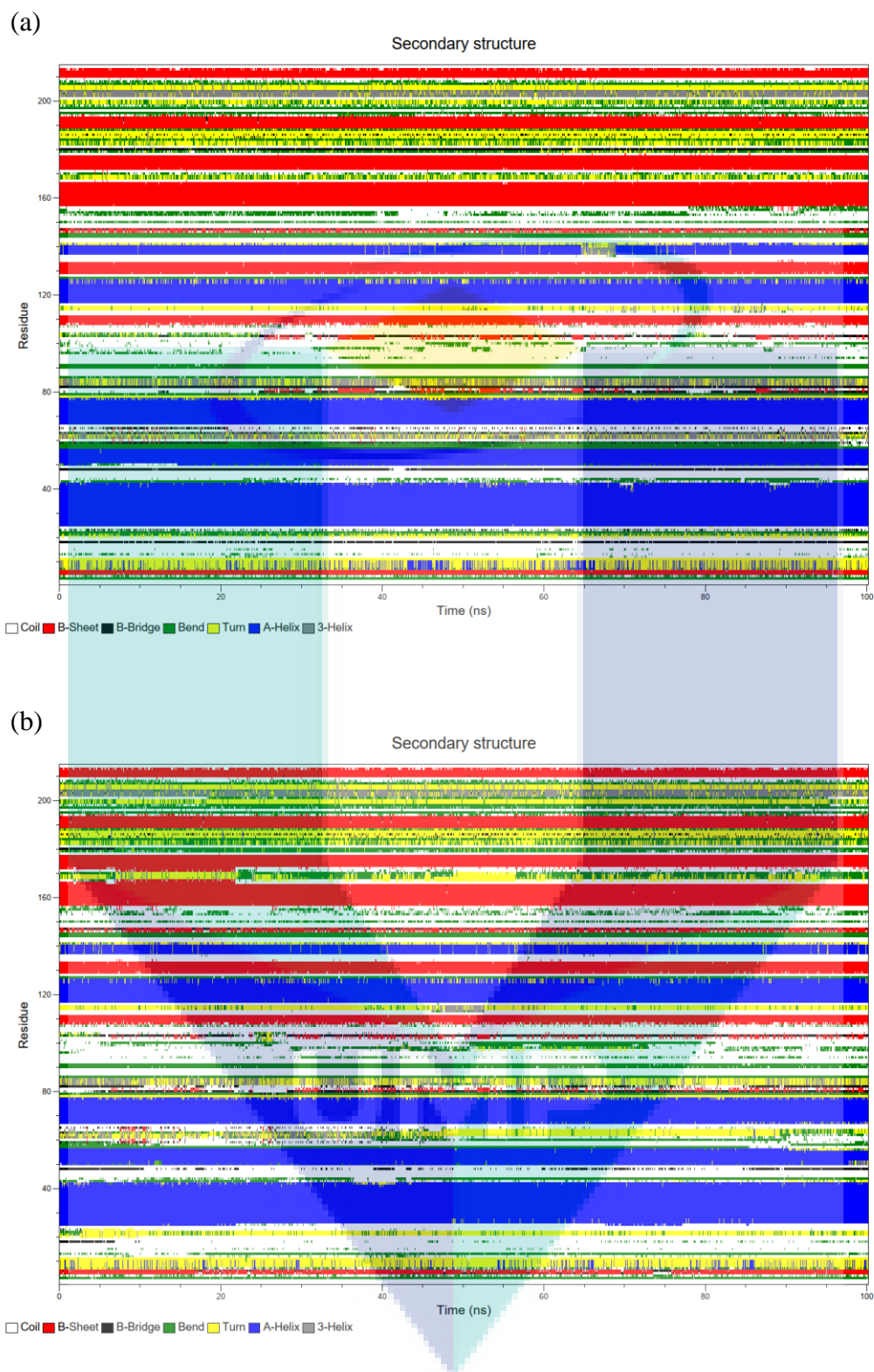
Figure 4.38   DSSP showing the evolution of secondary structures of (a) FB_1 at 313 K, (b) FB_1 at 333 K (c) FB_1 at 353 K, (d) FB_2 at 313 K, (e) FB_2 at 333 K, (f) FB_2 at 353 K, (g) FB_3 at 313 K, (h) FB_3 at 333 K and (i) FB_3 at 353 K over the course of simulation.

(c)

Secondary structure

(d)

Secondary structure

Figure 4.38   Continued.

(e)

(f)

Figure 4.38   Continued.

(g)

(h)

Figure 4.38   Continued.

(i)
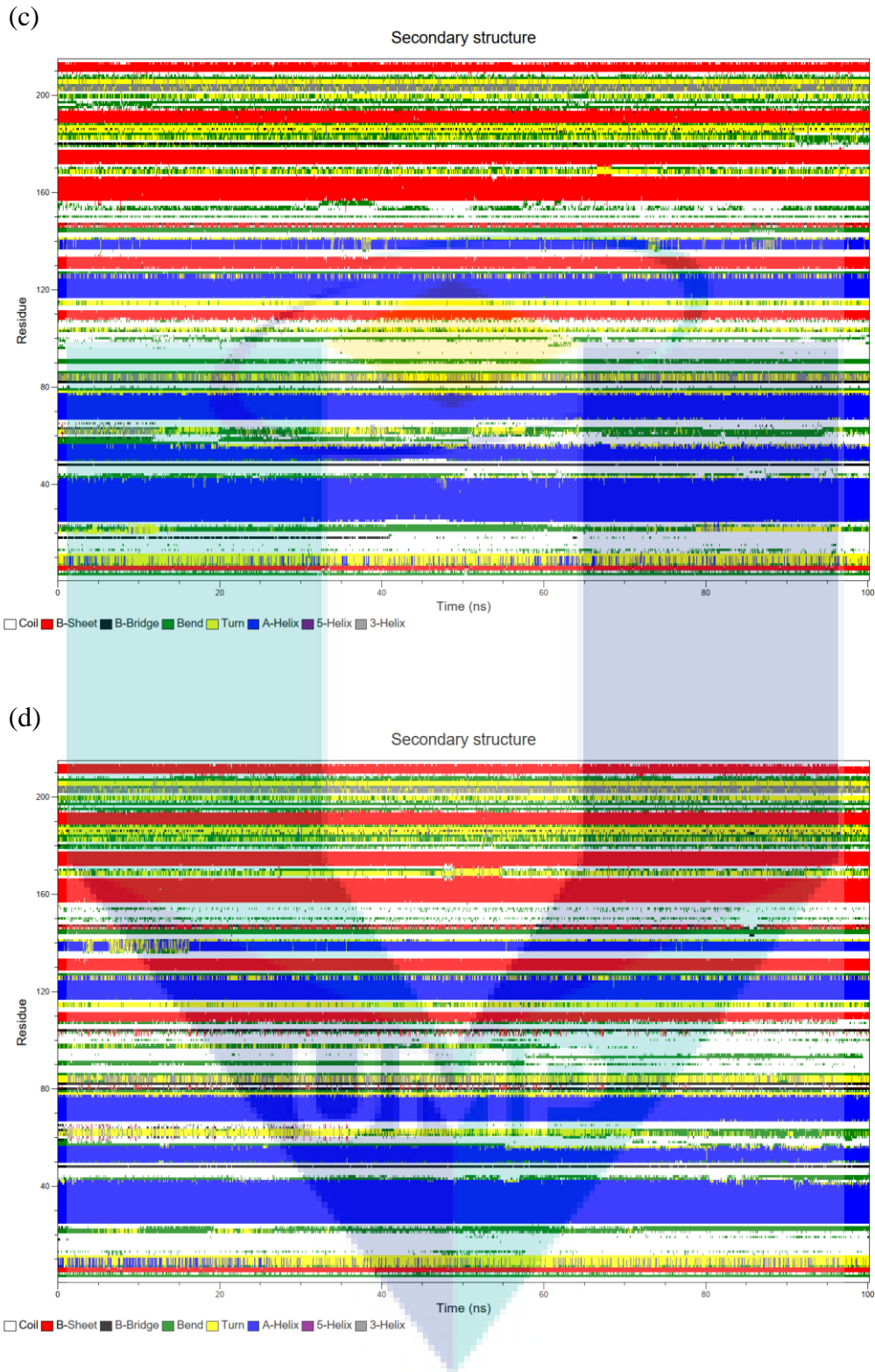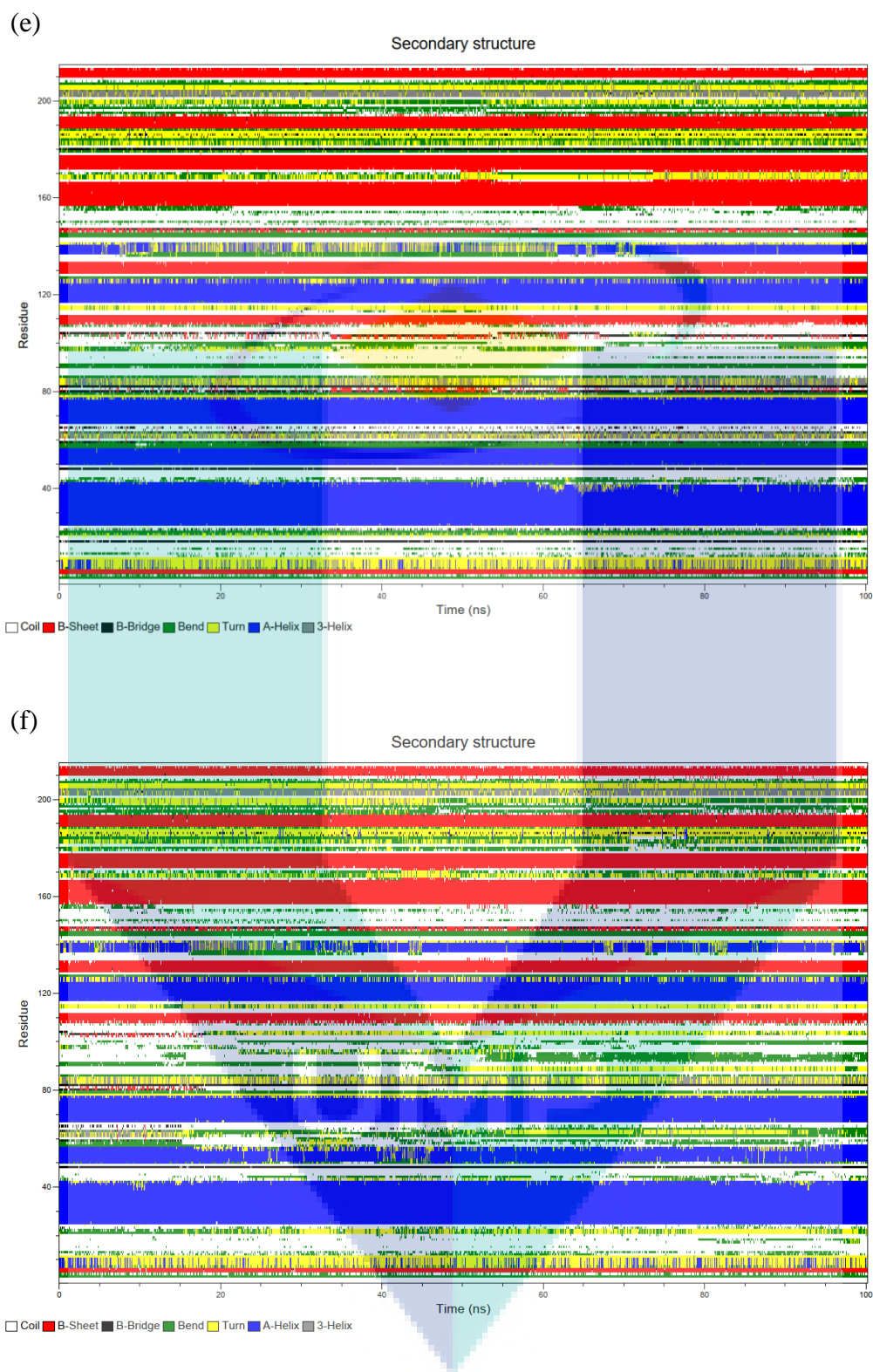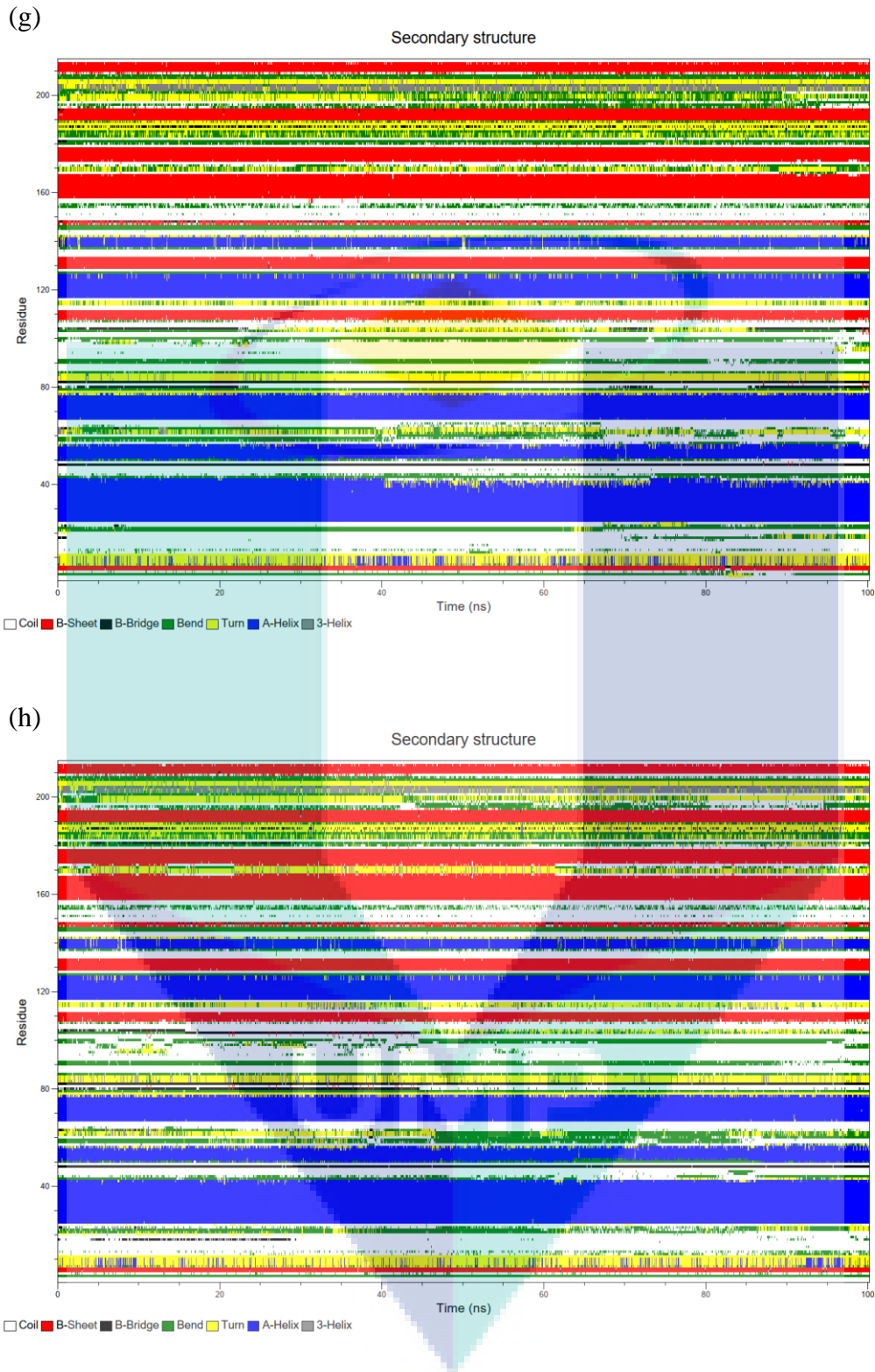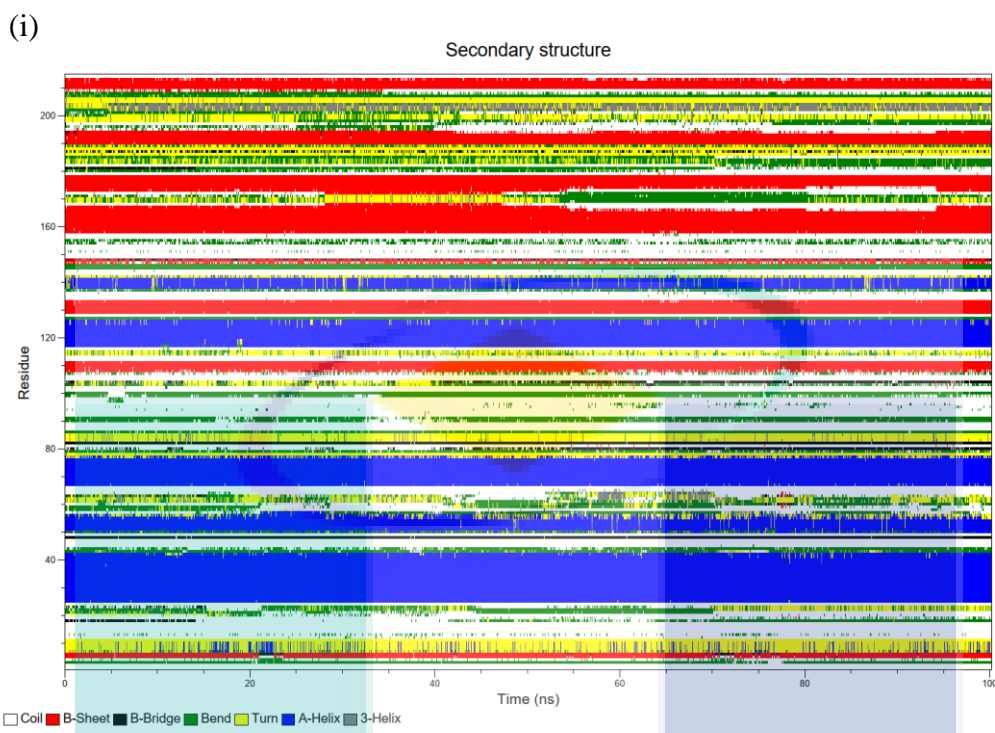
Figure 4.38   Continued.

Moreover, protein loops are irregular and less ordered regions due to their limited non-covalent interactions which make them have larger temperature dependency (Ahmad, Kumar, Ramanand, & Rao, 2012; Zeiske et al., 2016). This explains why the loop regions undergo significant fluctuations and are prone to unfolding as the temperature increases. Besides that, flexible protein loops are usually located on the protein surface exposed to the solvent (Shehu & Kavraki, 2012; Yedavalli & Rao, 2013; Yu, Yan, Zhang, & Dalby, 2017). Hence, the protein loop fluctuation is related to the increment of SASA and intermolecular hydrogen bonds form between fruit bromelain and water as the rapid displacement of these loops causes the exposure of hydrophobic region to water (Wintrode, Zhang, Vaidehi, Arnold, & Goddard, 2003). In essence, protein loops play a pivotal role in protein function such as binding and catalysis albeit their less disorganised structure in nature (Chang et al., 2014; Wong, Liu, & Kou, 2017). In this context, fluctuation on the loop regions cause the loss of proteolytic activity when exposing the fruit bromelain at high temperature in a prolonged period as discussed in section 4.3.5.1.

The physicochemical properties of amino acid residues contribute to the flexibility and stability of proteins (Alvarez-Ponce et al., 2018; Sosa-Pagán, Iversen, & Grandl, 2017; Yu et al., 2017). To investigate the effect of the composition of amino acid on fruit

bromelain flexibility, sequence information was extracted from the thermal sensitive regions. The identified temperature sensitive sequences are KNQNP, KDQNP, KIKTGYL, KYVTGYL, AVSYGCK, AVSNGCD, YKAYQGTCNANSF, YQAYQGTCNANSF, YQAYQDCAANSW, and DSSGT (Table 4.8). Intriguingly, the most common amino acid residues in these sequences are asparagine, aspartic acid, cysteine, lysine, glutamine, glycine, serine, threonine and tyrosine which are believed to be responsible for the fluctuations observed. Among the identified amino acid residues, asparagine, cysteine, glutamine, glycine, serine, threonine and tyrosine are non-charged amino acids which cannot not provide sufficient electrostatic interactions such as hydrogen bonds to stabilise fruit bromelain structures (Ramli et al., 2012; Szilágyi & Závodszky, 2000).

Table 4.8   Sequence information on the fluctuated regions

Model	Fluctuated region	Sequence
FB _1	115-119	KNQNP
	137-143	KIKTGYL
	155-161	AVSYGCK
	186-198	YKAYQGTCNANSF
	265-269	DSSGT
FB_2	107-111	KNQNP
	129-135	KIKTGYL
	147-153	AVSYGCK
	178-190	YQAYQGTCNANSF
	257-261	DSSGT
FB_3	116-120	KDQNP
	138-145	KYVTGYL
	156-162	AVSNGCD
	187-199	YQAYQDCAANSW
	267-271	DSSGT

As discussed above, protein folding is driven by the hydrophobic effect which sequester hydrophobic surface from water and maintain the overall protein structure in a lower energy state. The present hydrophilic amino acid residues include aspartic acid, asparagine, glutamine, lysine, serine and threonine which decreased the hydrophobicity leading to the increase of fruit bromelain's mobility at an elevated temperature (Brewer et al., 2012; Kazlauskas, 2018; Sinha & Khare, 2013). Furthermore, asparagine, cysteine and glutamine are recognised as thermolabile amino acid residues because they tend to undergo deamination and oxidation at an elevated temperature (Russell, Ferguson, Hough, Danson, & Taylor, 1997). In addition, glycine is a small aliphatic amino acid which

induces the flexibility on protein loops by allowing chain rotations and dihedral angles to become unavailable to other residues (Kumar, Sharma, & Bhalla, 2014; Ramli et al., 2013; Yennamalli, Rader, Wolt, & Sen, 2011).

In summary, FB_1, FB_2 and FB_3 showed noticeable differences in their relative conformational flexibility and stability with increasing of temperature. FB_1 is the most thermostable fruit bromelain model demonstrated by its least deviated dynamics in the course of simulation. Meanwhile, RMSD suggests that FB_2 is the least stable fruit bromelain model as the temperature raises. FB_2 displayed a denature behaviour in Rg, SASA and changes of intramolecular and intermolecular hydrogen bonds. In contrast, FB_3 is suspected to aggregate at elevated temperature because it exhibited a reduction in overall dimension, SASA and protein-water hydrogen bonds when the simulation temperature increased to 353 K. Based on the RMSF analysis, the major fluctuated regions with increase of temperature are protein loops located at the L-subdomain because this region is less rigid as well as lacking in electrostatic interactions and hydrophobicity.

CHAPTER 5

CONCLUSION

5.1   Conclusion

In conclusion, gene expression and the proteolytic activity level of fruit bromelain from ripe and unripe *A. comosus* cultivar pineapples were characterised. Based on the analysis, unripe pineapples fruits have higher fruit bromelain gene expression level and catalytic activity compared to the ripe fruits. This observation is most likely related to the defensive role of fruit bromelain during pineapple fruit growth cycle. Besides that, three fruit bromelain transcripts were selected and isolated from the *A. comosus* cultivar MD 2. The identity and properties of the isolated fruit bromelain sequences were carefully examined using different bioinformatic tools. It was found that the isolated fruit bromelain sequences exhibited alterations comparing to their respective reference sequences arising from sequencing error. Physicochemical analysis revealed the isolated fruit bromelain sequences are acidic proteins with molecular weight around 24 kDa. Moreover, fruit bromelain 3D models were generated based on the isolated sequences. The structural comparison analysis revealed similarities and differences between these models as well as with papain. This provides an insight on how pro-peptide functions as a scaffold in fruit bromelain folding and contributes to the inactivation of mature protein. Furthermore, binding residues in fruit bromelain His160, S2 Trp183, Glu19 and Asn159 were also identified. Lastly, the thermal stability of fruit bromelain was studied. Molecular dynamics simulation disclosed how fruit bromelain responded to the elevation of temperature. Instability was observed at loop regions and was influenced by the nature of amino acids. The FB_1 model exhibited highest thermostability while FB_2 is the least thermal stable fruit bromelain structure.

5.2   Recommendations for future research

Up to now, the majority of the bromelain studies focused on using different extraction and purification techniques to obtain bromelain for industrial and therapeutic applications. A breakthrough to obtain bromelain with higher activity is required. The result in this study demonstrated that the extraction of fruit bromelain from unripe pineapple is promising. Besides that, a comparison of gene expression level and proteolytic activity of fruit bromelain from different unripe cultivars is worth an in-depth investigation. Moreover, fruit bromelain is speculated to enhance the pest resistance of the pineapple fruits. Overexpression of fruit bromelain in important pest susceptible agricultural crops can eliminate the use of pesticides to protect farmers and the environment from its hazardous effects. The abnormality found between the fruit bromelain sequences is also worth for further investigation. In addition, by utilising the structural information obtained, it is expected that a thermal stable fruit bromelain will be produced through protein engineering in the future.

REFERENCES

Abraham, M. J., Murtola, T., Schulz, R., Páll, S., Smith, J. C., Hess, B., & Lindah, E. (2015). Gromacs: high performance molecular simulations through multi-level parallelism from laptops to supercomputers. SoftwareX, 1–2, 19–25. https://doi.org/10.1016/j.softx.2015.06.001

Afshar-Mohammadian, M., Rahimi-Koldeh, J., & Sajedi, R. H. (2010). The comparison of protease activity and total protein in three cultivars of kiwifruit of Northern Iran during fruit development. Acta Physiologiae Plantarum, 33(2), 343–348. https://doi.org/10.1007/s11738-010-0553-3

Ahmad, S., Kumar, V., Ramanand, K. B., & Rao, N. M. (2012). Probing protein stability and proteolytic resistance by loop scanning: a comprehensive mutational analysis. Protein Science, 21(3), 433–446. https://doi.org/10.1002/pro.2029

Al-Sa'ady, A., Al-Hadban, W., & Al-Zubaidy, M. (2016). Optimal conditions for bromelain extraction from pineapple fruit (Ananas comosus). Engineering and Technology Journal, 34(5), 675–682. Retrieved from https://www.researchgate.net/publication/318967578_Optimal_Conditions_for_Br omelain_Extraction_from_Pineapple_Fruit_Ananas_comosus

Ali, S., & Muhammad, Y. G. (2017). Industrial application of microbial proteases. European Journal of Pharmaceutical and Medical Research, 4(6), 623–629. Retrieved from http://www.ejpmr.com/admin/assets/article_issue/1496474101.pdf

Altschup, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. Journal of Molecular Biology, 215(3), 403–410. https://doi.org/10.1016/S0022-2836(05)80360-2

Alvarez-Ponce, D., Ruiz-González, M. X., Vera-Sirera, F., Feyertag, F., Perez-Amador, M. A., & Fares, M. A. (2018). Arabidopsis heat stress-induced proteins are enriched in electrostatically charged amino acids and intrinsically disordered regions. International Journal of Molecular Sciences, 19(8), 1–15. https://doi.org/10.3390/ijms19082276

Alves, L. C., Melo, R. L., Sanderson, S. J., Mottram, J. C., Coombs, G. H., Caliendo, G., … Juliano, M. A. (2001). S1 subsite specificity of a recombinant cysteine proteinase, CPB, of Leishmania mexicana compared with cruzain, human cathepsin L and papain using substrates containing non-natural basic amino acids. European Journal of Biochemistry, 268(5), 1206–1212. https://doi.org/10.1046/j.1432-1327.2001.01973.x

Amid, A., Ismail, N. A., Yusof, F., & Salleh, H. M. (2011). Expression, purification, and characterization of a recombinant stem bromelain from Ananas comosus. Process Biochemistry, 46(12), 2232–2239. https://doi.org/10.1016/j.procbio.2011.08.018

Amri, E., & Mamboya, F. (2012). Papain, a plant enzyme of biological importance: a review. American Journal of Biochemistry and Biotechnology, 8(2), 99–104. https://doi.org/10.3844/ajbbsp.2012.99.104

Armenteros, J. J. A., Tsirigos, K. D., Sønderby, C. K., Petersen, T. N., Winther, O., Brunak, S., … Nielsen, H. (2019). SignalP 5.0 improves signal peptide predictions using deep neural networks. Nature Biotechnology, 37(4), 420–423. https://doi.org/10.1038/s41587-019-0036-z

Asim, M., Abdan, K., Jawaid, M., Nasir, M., Dashtizadeh, Z., Ishak, M. R., & Hoque, M. E. (2015). A review on pineapple leaves fibre and its composites. International Journal of Polymer Science, 2015, 1–16. https://doi.org/10.1155/2015/950567

Banerjee, S., Kumar, J., Apte-Deshpande, A., & Padmanabhan, S. (2010). A novel prokaryotic vector for identification and selection of recombinants: direct use of the vector for expression studies in E. coli. Microbial Cell Factories, 9(30), 1–8. https://doi.org/10.1186/1475-2859-9-30

Beck, T. F., Mullikiin, J. C., & Biesecker, L. G. (2016). Systematic evaluation of Sanger validation of NextGen sequencing variants. Clinical Chemistry, 62(4), 647–654. https://doi.org/10.1038/nrg3575.Systems

Bellucci, M., De Marchis, F., & Pompa, A. (2018). The endoplasmic reticulum is a hub to sort proteins toward unconventional traffic pathways and endosymbiotic organelles. Journal of Experimental Botany, 69(1), 7–20. https://doi.org/10.1093/jxb/erx286

Berger, A., & Schechter, I. (1970). Mapping the active site of papain with the aid of peptide substrates and inhibitors. Philosophical Transactions of the Royal Society B: Biological Sciences, 257(813), 249–264. https://doi.org/10.1098/rstb.1970.0024

Betts, M. J., & Russel, R. B. (2003). Amino- acid properties and consequences of substitutions. In M. R. Barnes & I. C. Gray (Eds.), Bioinformatics for geneticists (pp. 311–342). https://doi.org/10.1002/0470867302.ch14

Beveridge, A. J. (1996). A theoretical study of the active sites of papain and S195C rat trypsin: implications for the low reactivity of mutant serine proteinases. Protein Science, 5(7), 1355–1365. https://doi.org/10.1002/pro.5560050714

Bhattacharyya, B. K. (2008). Bromelain: an overview. Natural Product Radiance, 7(4), 359–363. Retrieved from http://nopr.niscair.res.in/handle/123456789/5694

Bianco, V., Iskrov, S., & Franzese, G. (2012). Understanding the role of hydrogen bonds in water dynamics and protein stability. Journal of Biological Physics, 38(1), 27–48. https://doi.org/10.1007/s10867-011-9235-7

Botella, J. R., & Smith, M. (2008). Genomics of pineapple, crowning the king of tropical fruits. In P. H. Moore & R. Ming (Eds.), Genomics of tropical crop plants (pp. 441–452). https://doi.org/https://doi.org/10.1007/978-0-387-71219-2

Bowie, J. U., Ltcy, R., & Eisenberg, D. (1991). A method to identify protein sequences that fold into a known three-dimensional structure. Science, 253(58), 164–170. https://doi.org/10.1126/science.1853201

Brasil, I. M., Lourdes, M. De, Otoch, O., & Costa, J. H. (2008). Isolation of total RNA from ripe and unripe soursop (Annona muricata L.) fruit. African Journal of Plant Science, 2(9), 94–98. Retrieved from https://academicjournals.org/journal/AJPS/article-stat/6853A579627

Bresolin, I. R. A. P., Bresolin, I. T. L., Silveira, E., Tambourgi, E. B., & Mazzola, P. G. (2013). Isolation and purification of bromelain from waste peel of pineapple for therapeutic application. Brazilian Archives of Biology and Technology, 56(6), 971–979. https://doi.org/10.1590/S1516-89132013000600012

Brewer, S. H., Tang, Y., Vu, D. M., Gnanakaran, S., Raleigh, D. P., & Dyer, R. B. (2012). Temperature Dependence of Water Interactions with the Amide Carbonyls of α-Helices. Biochemistry, 51(26), 5293–5299. https://doi.org/10.1021/bi3006434

Burgos, M. I., Ochoa, A., & Perillo, M. A. (2019). β-sheet to α-helix conversion and thermal stability of β-Galactosidase encapsulated in a nanoporous silica gel. Biochemical and Biophysical Research Communications, Vol. 508, pp. 270–274. https://doi.org/10.1016/j.bbrc.2018.11.077

Bustin, S. A., Beaulieu, J. F., Huggett, J., Jaggi, R., Kibenge, F. S. B., Olsvik, P. A., … Toegel, S. (2010). MIQE précis: practical implementation of minimum standard guidelines for fluorescence-based quantitative real-time PCR experiments. BMC Molecular Biology, 11(74), 1–5. https://doi.org/10.1186/1471-2199-11-74

Bustin, S. A., Benes, V., Garson, J. A., Hellemans, J., Huggett, J., Kubista, M., … Wittwer, C. T. (2009). The MIQE guidelines: minimum information for publication of quantitative real-time PCR experiments. Clinical Chemistry, 55(4), 611–622. https://doi.org/10.1373/clinchem.2008.112797

Butts, C. T., Zhang, X., Kelly, J. E., Roskamp, K. W., Unhelkar, M. H., Freites, J. A., … Martin, R. W. (2016). Sequence comparison, molecular modeling, and network analysis predict structural diversity in cysteine proteases from the Cape sundew, Drosera capensis. Computational and Structural Biotechnology Journal, 14, 271–282. https://doi.org/10.1016/j.csbj.2016.05.003

Camilloni, C., Bonetti, D., Morrone, A., Giri, R., Dobson, C. M., Brunori, M., … Vendruscolo, M. (2016). Towards a structural biology of the hydrophobic effect in protein folding. Scientific Reports, 6, 1–9. https://doi.org/10.1038/srep28285

Cebe, P., Hu, X., Kaplan, D. L., Zhuravlev, E., Wurm, A., Arbeiter, D., & Schick, C. (2013). Beating the heat-fast scanning melts silk beta sheet crystals. Scientific Reports, 3, 1–7. https://doi.org/10.1038/srep01130

Chakravarty, S., Godbole, S., Zhang, B., Berger, S., & Sanchez, R. (2008). Accuracy of comparative models of protein structure. BMC Structural Biology, 8(31), 1–13. https://doi.org/10.1186/1472-6807-8-31

Chan, Y. K., d'Eeckenbrugge, G. C., & Sanewski, G. M. (2003). Breeding and variety improvement. In D. P. Bartholomew, R. E. Paull, & K. G. Rohrbach (Eds.), The pineapple botany, production and uses (pp. 33–56). Hawaii: CABI Publishing.

Chang, H. J., Jian, J. W., Hsu, H. J., Lee, Y. C., Chen, H. Sen, You, J. J., … Yang, A. S. (2014). Loop-sequence features and stability determinants in antibody variable domains by high-throughput experiments. Structure, 22(1), 9–21. https://doi.org/10.1016/j.str.2013.10.005

Chaturvedi, D., & Mahalakshmi, R. (2017). Transmembrane β-barrels Evolution, folding and energetics. Biochimica et Biophysica Acta (BBA) - Biomembranes, 1859(12), 2467–2482. https://doi.org/10.1016/j.bbamem.2017.09.020

Chen, G., Huang, K., Miao, M., Feng, B., & Campanella, O. H. (2019). Molecular dynamics simulation for mechanism elucidation of food processing and safety: state of the art. Comprehensive Reviews in Food Science and Food Safety, 18(1), 243–263. https://doi.org/10.1111/1541-4337.12406

Chew, L. Y., Toh, G. T., & Ismail, A. (2018). Application of proteases for the production of bioactive peptides. In M. Kuddus (Ed.), Enzymes in food biotechnology (pp. 247–261). https://doi.org/10.1016/b978-0-12-813280-7.00015-3

Childers, M. C., & Daggett, V. (2017). Insights from molecular dynamics simulations for computational protein design. Molecular Systems Design and Engineering, 2(1), 9–33. https://doi.org/10.1039/c6me00083e

Cloete, R., Kapp, E., Joubert, J., Christoffels, A., & Malan, S. F. (2018). Molecular modelling and simulation studies of the Mycobacterium tuberculosis multidrug efflux pump protein Rv1258c (. PLoS ONE, 13(12), 1–18. https://doi.org/10.1371/journal.pone.0209717

Colovos, C., & Yeates, T. O. (1993). Verification of protein structures: patterns of nonbonded atomic interactions. Protein Science, 2(9), 1511–1519. https://doi.org/10.1002/pro.5560020916

Cordara, G., Van Eerde, A., Grahn, E. M., Winter, H. C., Goldstein, I. J., & Krengel, U. (2016). An unusual member of the papain superfamily: mapping the catalytic cleft of the Marasmius oreades agglutinin (MOA) with a caspase inhibitor. PLoS ONE, 11(2), 1–19. https://doi.org/10.1371/journal.pone.0149407

Corzo, C. A., Waliszewski, K. N., & Welti-Chanes, J. (2012). Pineapple fruit bromelain affinity to different protein substrates. Food Chemistry, 133(3), 631–635. https://doi.org/10.1016/j.foodchem.2011.05.119

Coulombe, R., Grochulski, P., Sivaraman, J., Ménard, R., Mort, J. S., & Cygler, M. (1996). Structure of human procathepsin L reveals the molecular basis of inhibition by the prosegment. The EMBO Journal, 15(20), 5492–5503. https://doi.org/10.1002/j.1460-2075.1996.tb00934.x

Cupp-enyard, C., & Sigma-Aldrich. (2008). Sigma ' s non-specific protease activity assay - casein as a substrate. Journal of Visualised Experiments, 19, 1–2. https://doi.org/10.3791/899

da Silva López, R. . (2017). Debridement applications of bromelain: a complex of cysteine proteases from pineapple. Advances in Biotechnology & Microbiology, 3(5), 6109–6111. https://doi.org/10.19080/AIBM.2017.03.555624

Daniel, R. M., Dines, M., & Petach, H. H. (1996). The denaturation and degradation of stable enzymes at high temperatures. Biochemical Journal, 317(1), 1–11. https://doi.org/10.1042/bj3170001

Das, S., & Bhattacharyya. (2018). Bromelain from pineapple: its stability and therapeutic potentials. In C. S. Bogsan & S. D. Todorov (Eds.), Tropical fruits: from cultivation to consumption and health benefits, pineapple (pp. 43–100). New York: Nova Science Publishers.

Dawson, C. (2016). Pineapple: united nation conference on trade and development. In An INFOCOMM Commodity Profile. Geneva 10.

Demidyuk, I. V., Shubin, A. V., Gasanov, E. V., & Kostrov, S. V. (2010). Propeptides as modulators of functional activity of proteases. BioMolecular Concepts, 1(3–4), 305–322. https://doi.org/10.1515/bmc.2010.025

Dhukani, A. (2013). Ananas comosus (L.) Merr., Bromeliaceae. In C. L. Quave (Ed.), Medicinal plant monographs (pp. 46–57). Atlanta: Emory University.

Díaz-Mendoza, M., Velasco-Arroyo, B., González-Melendi, P., Martínez, M., & Díaz, I. (2014). C1A cysteine protease-cystatin interactions in leaf senescence. Journal of Experimental Botany, 65(14), 3825–3833. https://doi.org/10.1093/jxb/eru043

Dieffenbach, C. W., Lowe, T. M. J., & Dveksler, G. S. (1993). General concepts for PCR primer design. Genome Research, 3(3), 530–537. https://doi.org/10.1101/gr.3.3.S30

Dinner, A. R., Sali, A., Smith, L. J., Dobson, C. M., & Karplus, M. (2000). Understanding protein folding via free-energy surfaces from theory and experiment. Trends in Biochemical Sciences, 25(7), 331–339. https://doi.org/10.1016/S0968-0004(00)01610-8

Dong, R., Pan, S., Peng, Z., Zhang, Y., & Yang, J. (2018). mTM-align : a server for fast protein structure database search and multiple protein structure alignment. Nucleic Acids Research, 46(W1), 380–386. https://doi.org/10.1093/nar/gky430

Dorn, M., E Silva, M. B., Buriol, L. S., & Lamb, L. C. (2014). Three-dimensional protein structure prediction: methods and computational strategies. Computational Biology and Chemistry, 53(Part B), 251–276. https://doi.org/10.1016/j.compbiolchem.2014.10.001

Drenth, J., Jansonius, J. N., Koekoek, R., Swen, H. M., & Wolthers, B. . G. (1968). Structure of papain. Nature, 218, 929–932.

Dror, R. O., Dirks, R. M., Grossman, J. P., Xu, H., & Shaw, D. E. (2012). Biomolecular simulation: a computational microscope for molecular biology. Annual Review of Biophysics, 41, 429–452. https://doi.org/10.1146/annurev-biophys-042910-155245

Du, X., Sang, P., Xia, Y. L., Li, Y., Liang, J., Ai, S. M., … Liu, S. Q. (2017). Comparative thermal unfolding study of psychrophilic and mesophilic subtilisin-like serine proteases by molecular dynamics simulations. Journal of Biomolecular Structure and Dynamics, 35(7), 1500–1517. https://doi.org/10.1080/07391102.2016.1188155

Dubey, V. K., Pande, M., Singh, B. K., & Jagannadham, M. V. (2007). Papain-like proteases : Applications of their inhibitors. African Journal of Biotechnology, 6(9), 1077–1086. https://doi.org/10.4314/ajb.v6i9.57108

Emberly, E. G., Mukhopadhyay, R., Wingreen, N. S., & Tang, C. (2003). Flexibility of α-helices: results of a statistical analysis of database protein structures. Journal of Molecular Biology, 327(1), 229–237. https://doi.org/10.1016/S0022-2836(03)00097-4

FAO. (2019). Pineapples. Retrieved March 1, 2019, from http://www.fao.org

Feig, M. (2016). Local rotein structure refinement via molecular dynamics simulations with locPREFMD. Journal of Chemical Information and Modeling, 56(7), 1304–1312. https://doi.org/10.1021/acs.jcim.6b00222

Feig, M. (2017). Computational protein structure refinement: almost there, yet still so far to go. Wiley Interdisciplinary Reviews: Computational Molecular Science, 7(3). https://doi.org/10.1002/wcms.1307

Feig, M., & Mirjalili, V. (2016). Protein structure refinement via molecular-dynamics simulations: what works and what does not? Proteins, 84(S1), 282–292. https://doi.org/10.1002/prot.24871

Feng, T., Li, M., Zhou, J., Zhuang, H., Chen, F., Ye, R., … Fang, Z. (2015). Application of molecular dynamics simulation in food carbohydrate research - a review. Innovative Food Science and Emerging Technologies, 31, 1–13. https://doi.org/10.1016/j.ifset.2015.06.015

Fields, P. A., Dong, Y., Meng, X., & Somero, G. N. (2015). Adaptations of protein structure and function to temperature: there is more than one way to "skin a cat." Journal of Experimental Biology, 218(12), 1801–1811. https://doi.org/10.1242/jeb.114298

Fiser, A. (2010). Template-based protein structure modeling. Methods in Molecular Biology, 673, 73–94. https://doi.org/10.1007/978-1-60761-842-3

Fox, T., Mason, P., Storer, A. C., & Mort, J. S. (1995). Modification of S1 subsite specificity in the cysteine protease cathepsin b. Protein Engineering, Design and Selection, 8(1), 53–57. https://doi.org/10.1093/protein/8.1.53

Ganugapati, J., & Akash, S. (2017). Multi-template homology based structure prediction and molecular docking studies of protein 'L' of Zaire ebolavirus (EBOV). Informatics in Medicine Unlocked, 9, 68–75. https://doi.org/10.1016/j.imu.2017.06.002

Gao, Y., Mei, Y., & Zhang, J. Z. H. (2015). Treatment of hydrogen bonds in protein simulations. In J. Liu (Ed.), Advanced Materials for Renewable Hydrogen Production, Storage and Utilization (pp. 111–136). https://doi.org/10.5772/61049

Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S., Wilkins, M. R., Appel, R. D., & Bairoch, A. (2005). Protein analysis tools on the ExPASy server. In J. M. Walker (Ed.), The proteomics protocols handbook (pp. 571–607). https://doi.org/10.1385/1592598900

Gelpi, J., Hospital, A., Goñi, R., & Orozco, M. (2015). Molecular dynamics simulations: advances and applications. Advances and Applications in Bioinformatics and Chemistry, 37. https://doi.org/10.2147/AABC.S70333

George, S., Bhasker, S., Madhav, H., Nair, A., & Chinnamma, M. (2014). Functional characterization of recombinant bromelain of Ananas comosus expressed in a prokaryotic system. Molecular Biotechnology, 56(2), 166–174. https://doi.org/10.1007/s12033-013-9692-2

Gessmann, D., Mager, F., Naveed, H., Arnold, T., Weirich, S., Linke, D., … Nussberger, S. (2011). Improving the resistance of a eukaryotic β-barrel protein to thermal and chemical perturbations. Journal of Molecular Biology, Vol. 413, pp. 150–161. https://doi.org/10.1016/j.jmb.2011.07.054

Ginalski, K. (2006). Comparative modeling for protein structure prediction. Current Opinion In Structural Biology, 16(2), 172–177. https://doi.org/10.1016/j.sbi.2006.02.003

Groves, M. R., Taylor, M. A. J., Scott, M., Cummings, N. J., Pickersgill, R. W., & Jenkins, J. A. (1996). The prosequence of procaricain forms an α-helical domain that prevents access to the substrate-binding cleft. Structure, 4(10), 1193–1203. https://doi.org/10.1016/S0969-2126(96)00127-X

Gu, J., Tong, H., Sun, L., & Lin, Z. (2019). Molecular dynamics perspective on the thermal stability of mandelate racemase. Journal of Biomolecular Structure and Dynamics, 37(2), 383–393. https://doi.org/10.1080/07391102.2018.1427631

Guex, N., & Peitsch, M. C. (1997). SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling. Electrophoresis, Vol. 18, pp. 2714–2723. https://doi.org/10.1002/elps.1150181505

Haimov, B., & Srebnik, S. (2016). A closer look into the α-helix basin. Scientific Reports, 6, 1–12. https://doi.org/10.1038/srep38341

Hall, T. A. (1999). BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucleic Acids Symposium Series, 41, 95–98. https://doi.org/citeulike-article-id:691774

Han, J., Cai, Y., Xie, X., Wang, Y., Wang, L., Li, C., … Ni, L. (2018). A simple method for purification of bromelain in a thermosensitive triblock copolymer-based protection system and recycling of phase components. Separation Science and Technology (Philadelphia), 53(4), 636–644. https://doi.org/10.1080/01496395.2017.1398757

Heo, L., & Feig, M. (2018a). Experimental accuracy in protein structure refinement via molecular dynamics simulations. Proceedings of the National Academy of Sciences, 115(52), 13276–13281. https://doi.org/10.1073/pnas.1811364115

Heo, L., & Feig, M. (2018b). PREFMD: a web server for protein structure refinement via molecular dynamics simulations. Bioinformatics, 34(6), 1063–1065. https://doi.org/10.1093/bioinformatics/btx726

Heredia-Sandoval, N. G., Valencia-Tapia, M. Y., de la Barca, A. M. C., & Islas-Rubio, A. R. (2016). Microbial proteases in baked goods: modification of gluten and effects on immunogenicity and product quality. Foods, 5(3), 1–10. https://doi.org/10.3390/foods5030059

Hidayat, T., Chandrika, K., Izana, A. F., Azman, S. A., & Alina, W. (2013). Phylogenetic abnalysis of Malaysian pineapples cultivars based on the DNA sequence of the internal transcribed spacer region. Jurnal Teknologi (Sciences and Engineering), 62(2), 43–46. https://doi.org/10.11113/jt.v62.1878

Hossain, M. . (2016). World pineapple production: an overview. African Journal of Food Agriculture, Nutrition and Development, 16(4), 11443–11456. https://doi.org/10.18697/ajfand.76.15620

Huang, J., Rauscher, S., Nawrocki, G., Ran, T., Feig, M., De Groot, B. L., … MacKerell, A. D. (2016). CHARMM36m: an improved force field for folded and intrinsically disordered proteins. Nature Methods, 14(1), 71–73. https://doi.org/10.1038/nmeth.4067

Hui, K., & Feng, Z.-P. (2013). Efficient experimental design and analysis of real-time PCR assays. Channels, 7(3), 160–170. https://doi.org/10.4161/chan.24024

Ishitani, R., Terada, T., & Shimizu, K. (2008). Refinement of comparative models of protein structure by using multicanonical molecular dynamics simulations. Molecular Simulation, 34(3), 327–336. https://doi.org/10.1080/08927020801930539

Ismail, B., Mohammed, H., & Nair, A. J. (2019). Influence of proteases on functional properties of food. In B. Parameswaran, S. Varjani, & S. Raveendran (Eds.), Green Bio-processes: Enzymes in Industrial Food Processing (Vol. 55, pp. 31–53). https://doi.org/10.1007/978-981-13-2324-9

Janssen, B. J., Thodey, K., Schaffer, R. J., Alba, R., Balakrishnan, L., Bishop, R., … Ward, S. (2008). Global gene expression analysis of apple fruit development from the floral bud to ripe fruit. BMC Plant Biology, 29(16), 1–29. https://doi.org/10.1186/1471-2229-8-16

Jisha, V. N., Smitha, R. B., Pradeep, S., Sreedevi, S., Unni, K. N., Sajith, S., … Benjamin, S. (2013). Versatility of microbial proteases. Advances in Enzyme Research, 1(3), 39–51. https://doi.org/10.4236/aer.2013.13005

Jones, P., Binns, D., Chang, H. Y., Fraser, M., Li, W., McAnulla, C., … Hunter, S. (2014). InterProScan 5: Genome-scale protein function classification. Bioinformatics, 30(9), 1236–1240. https://doi.org/10.1093/bioinformatics/btu031

Joy, P. P., & Anjana, R. (2015). Evolution of pineapple. In KV Peter (Ed.), Evolution oF Horticultural Crops (Vol. 5, pp. 1–39). New Delhi: Astral International Pvt.Ltd.

Jung, Y., Choi, C., Park, J., Kang, H., Choi, J., Nou, I., … Kang, K. (2008). Overexpression of the pineapple fruit bromelain gene (BAA) in transgenic Chinese cabbage (Brassica rapa) results in enhanced resistance to bacterial soft rot. Electronic Journal of Biotechnology, 11(1), 1–8. https://doi.org/10.2225/vol10-issue4-fulltext-5

Jutamongkon, R., & Charoenrein, S. (2010). Effect of temperature on the stability of fruit bromelain from Smooth Cayenne pineapple. Kasetsart Journal - Natural Science, 44, 943–948. Retrieved from https://research.rdi.ku.ac.th/world/showitem.php?itemID=118661&lang=en

Kamphuis, I. G., Kalk, K. H., Swarte, M. B. A., & Drenth, J. (1984). Structure of papain refined at 1.65 Å resolution. Journal of Molecular Biology, 179(2), 233–256. https://doi.org/10.1016/0022-2836(84)90467-4

Kato, K., Nakayoshi, T., Fukuyoshi, S., Kurimoto, E., & Oda, A. (2017). Validation of molecular dynamics simulations for prediction of three-dimensional structures of small proteins. Molecules, 22(10), 1–15. https://doi.org/10.3390/molecules22101716

Kazlauskas, R. (2018). Engineering more stable proteins. Chemical Society Reviews, 47(24), 9026–9045. https://doi.org/10.1039/c8cs00014j

Kelly, G. (1996). Bromelain: a literature review and discussion of its therapeutic applications. Alternative Medicine Review, 11(44), 243–257.

Ketnawa, S., Chaiwut, P., & Rawdkuen, S. (2012). Pineapple wastes: a potential source for bromelain extraction. Food and Bioproducts Processing, 90(3), 385–391. https://doi.org/10.1016/j.fbp.2011.12.006

Khouri, H. E., Vernet, T., Menard, R., Parlati, F., Laflamme, P., Tessier, D. C., … Storer, A. C. (1991). Engineering of papain: selective alteration of substrate specificity by site-directed mutagenesis. Biochemistry, 30(37), 8929–8936. https://doi.org/10.1021/bi00101a003

Koia, J. H., Moyle, R. L., & Botella, J. R. (2012). Microarray analysis of gene expression profiles in ripening pineapple fruits. BMC Plant Biology, 12(240), 1–13. https://doi.org/10.1186/1471-2229-12-240

Kothare, A., Pardhi, V., Chivte, P., Muley, S., & Shanbhag, T. (2017). Purification , partial characterization and activity profiling of bromelain protease. International Journal of Pharmacy and Biological Sciences, 7(3), 1–9. Retrieved from https://www.researchgate.net/publication/330727139_PURIFICATION_PARTIAL _CHARACTERIZATION_AND_ACTIVITY_PROFILING_OF_BROMELAIN_ PROTEASE

Kryshtafovych, A., & Fidelis, K. (2009). Protein structure prediction and model quality assessment. Drug Discovery Today, 14(7–8), 386–393. https://doi.org/10.1016/j.drudis.2008.11.010

Kumar, V., Sharma, N., & Bhalla, T. C. (2014). In silico analysis of β -galactosidases primary and secondary structure in relation to temperature adaptation. Journal of Amino Acids, 2014, 1–9. https://doi.org/10.1155/2014/475839

Kwon, C. W., Park, K. M., Kang, B. C., Kweon, D. H., Kim, M. D., Shin, S. W., … Chang, P. S. (2015). Cysteine protease profiles of the medicinal plant Calotropis procera R. Br. revealed by de Novo transcriptome analysis. PLoS ONE, 10(3), 1–15. https://doi.org/10.1371/journal.pone.0119328

Larocca, M., Rossano, R., Santamaria, M., & Riccio, P. (2010). Analysis of pineapple [Ananas comosus (L.) Merr.] fruit proteinases by 2-D zymography and direct identification of the major zymographic spots by mass spectrometry. Food Chemistry, 123(4), 1334–1342. https://doi.org/10.1016/j.foodchem.2010.06.016

Laskowski, R. A., MacArthur, M. W., Moss, D. S., & Thornton, J. M. (2012). PROCHECK - a program to check the stereochemical quality of protein structures. Journal of Applied Crystallography, 26, 283–291. Retrieved from papers://c33b182f-cf88-47e8-a9c5-ad67b5626483/Paper/p1776

Leal, F., & D'Eeckenbrugge, G. C. (2002). Morphology, anatomy and taxonomy. In D. P. Bartholomew, R. E. Paull, & K. G. Rohrbach (Eds.), The Pineapple Botany, Production and Uses (pp. 13–32). CABI Publishing.

Lee, A. R., Bak, H. J., Kim, N. Y., Kim, M.-S., Go, H.-J., Han, J. W., … Lee, H. H. (2012). Cloning, heterologous expression, and enzymatic characterization of cathepsin L from starfish (Asterina pectinifera). Bioscience, Biotechnology, and Biochemistry, 76(12), 2342–2346. https://doi.org/10.1271/bbb.120568

Lee, J., Freddolino, P. L., & Zhang, Y. (2017). Ab initio protein structure prediction. In J. R. Daniel (Ed.), From protein structure to function with bioinformatics (pp. 3–35). Dordrecht: Springer Netherlands.

Li, J., & Cheng, J. (2016). A stochastic point cloud sampling method for multi-template protein comparative modeling. Scientific Reports, 6, 1–16. https://doi.org/10.1038/srep25687

Li, Q., Yi, L., Marek, P., & Iverson, B. L. (2013). Commercial proteases: present and future. FEBS Letters, 587(8), 1155–1163. https://doi.org/10.1016/j.febslet.2012.12.019

Liang, H., Li, M., Shi, M., Liao, A., & Wu, R. (2012). Study on the stability of fruit bromelain. Advanced Materials Research, 421, 19–22. https://doi.org/10.4028/www.scientific.net/AMR.421.19

Lin, E., Burns, D. J. W., & Gardner, R. C. (1993). Fruit developmental regulation of the kiwifruit actinidin promoter is conserved in transgenic petunia plants. Plant Molecular Biology, 23(3), 489–499. https://doi.org/10.1007/BF00019297

Liu, C. H., & Liu, Y. (2017). Fruit quality and differentially expressed genes of winter-harvested pineapple in response to elevated temperature over a short postharvest period. Postharvest Biology and Technology, 130, 21–27. https://doi.org/10.1016/j.postharvbio.2017.03.016

Livak, K. J., & Schmittgen, T. D. (2001). Analysis of relative gene expression data using real-time quantitative PCR and the 2-ΔΔCT method. Methods, 25(4), 402–408. https://doi.org/10.1006/meth.2001.1262

Lobanov, M. Y., Bogatyreva, N. S., & Galzitskaya, O. V. (2008). Radius of gyration as an indicator of protein structure compactness. Structural-Functional Analysis of Biopolymers and Their Complexes, 42(4), 623–628. https://doi.org/10.1134/S0026893308040195

M Mukaka, M. (2012). Statistics corner: a guide to appropriate use of correlation coefficient in medical research. Malawi Medical Journal, 24(3), 69–71. https://doi.org/10.1016/j.cmpb.2016.01.020

Mahajan, R. T., & Badgujar, S. B. (2010). Biological aspects of proteolytic enzymes : a review. Journal of Pharmacy Research, 3(9), 2048–2068. Retrieved from https://www.researchgate.net/publication/282212064_Biological_aspects_of_proteolytic_enzymes_A_Review

Malek, K., Norazan, M., Ramaness, P., Othman, Z., Malek, R., Aziz, R., … Enshasy, H. El. (2016). Cysteine proteases from Carica papaya: an important enzyme group of many industrial applications. IOSR Journal of Pharmacy and Biological Sciences, 11(2), 11–16. https://doi.org/10.9790/3008-11211116

Mallamace, D., Fazio, E., Mallamace, F., & Corsaro, C. (2018). The role of hydrogen bonding in the folding/unfolding process of hydrated lysozyme: a review of recent NMR and FTIR results. International Journal of Molecular Sciences, 19(12), 3825. https://doi.org/10.3390/ijms19123825

Manohar, J., Gayathri, R., & Vishnupriya, V. (2016). Tenderisation of meat using bromelain from pineapple extract. International Journal of Pharmaceutical Sciences Review and Research, 39(1), 81–85. Retrieved from http://globalresearchonline.net/journalcontents/v39-1/17.pdf

Manzoor, Z., Nawaz, A., Mukhtar, H., & Haq, I. (2016). Bromelain: methods of extraction, purification and therapeutic applications. Brazilian Archives of Biology and Technology, 59(0), 1–16. https://doi.org/10.1590/1678-4324-2016150010

Marchand, Tanguy, L., De Rosa, M., Salvi, N., Sala, B. M., Andreas, L. B., Barbet-Massin, E., … Ricagno, S. (2018). Conformational dynamics in crystals reveal the molecular bases for D76N beta-2 microglobulin aggregation propensity. Nature Communications, 9(1), 1–11. https://doi.org/10.1038/s41467-018-04078-y

Martinez, M., Cambra, I., Gonzalez-Melendi, P., E. Santamarıa, M., & Diaz, I. (2012). C1A cysteine-proteases and their inhibitors in plants. Physiologia Plantarum, 145(1), 85–94. https://doi.org/10.1111/j.1399-3054.2012.01569.x

Maurer, H. R. (2001). Bromelain: biochemistry, pharmacology and medical use. Cellular and Molecular Life Sciences, 58(9), 1234–1245. https://doi.org/10.1007/PL00000936

Meier, A., & Söding, J. (2015). Automatic prediction of protein 3D structures by probabilistic multi-template homology modeling. PLoS Computational Biology, 11(10), 1–20. https://doi.org/10.1371/journal.pcbi.1004343

Menard, R, Carmona, E., Plouffe, C., Bromme, D., Konishi, Y., Lefebvre, J., & Storer, A. C. (1993). The specificity of the S1' subsite of cysteine proteases. FEBS Letters, 328(1), 107–110. https://doi.org/10.1016/0014-5793(93)80975-z

Menard, Robert, Carriere, J., Laflamme, P., Plouffe, C., Khouri, H. E., Vernet, T., … Storer, A. C. (1991). Contribution of the glutamine 19 side chain to transition-state stabilization in the oxyanion hole of papain. Biochemistry, 30(37), 8924–8928. https://doi.org/10.1021/bi00101a002

Merkel, J. S., Strutevant, J. M., & Regan, L. (1999). Sidechain interactions in parallel β sheets: the energetics of cross-strand pairings. Structure, 7(11), 1333–1343. https://doi.org/10.1016/S0969-2126(00)80023-4

Mihăşan, M. (2010). Basic protein structure prediction for the biologist: a review. Archives of Biological Sciences, 62(4), 857–871. https://doi.org/10.2298/ABS1004857M

Misas-villamil, J. C., Hoorn, R. A. L. Van Der, & Doehlemann, G. (2016). Papain-like cysteine proteases as hubs in plant immunity. New Phytologist, 212(4), 902–907. https://doi.org/10.1111/nph.14117

Mishra, A., Ranganathan, S., Jayaram, B., & Sattar, A. (2018). Role of solvent accessibility for aggregation-prone patches in protein folding. Scientific Reports, 8(1), 1–13. https://doi.org/10.1038/s41598-018-31289-6

Mohamed, S. A. E. H., Elloumi, M., & Thompson, J. D. (2016). Motif discovery in protein sequences. In S. Ramakrishnan (Ed.), Pattern recognition: analysis and applications (pp. 3–18). https://doi.org/10.5772/57353

Moree, B., Connell, K., Mortensen, R. B., Liu, C. T., Benkovic, S. J., & Salafsky, J. (2015). Protein conformational changes are detected and resolved site specifically by second-harmonic generation. Biophysical Journal, 109(4), 806–815. https://doi.org/10.1016/j.bpj.2015.07.016

Moretti-Almeida, G. (2018). Pineapple taxonomy. In Cristina Stewart Bogsan & S. D. Todorov (Eds.), Tropical fruits: from cultivation to consumption and health benefits, pineapple (pp. 1–14). New York: Nova Science Publishers.

Moyle, R., Fairbairn, D. J., Ripi, J., Crowe, M., & Botella, J. R. (2005). Developing pineapple fruit has a small transcriptome dominated by metallothionein. Journal of Experimental Botany, 56(409), 101–112. https://doi.org/10.1093/jxb/eri015

MPIB. (2019). Origin of Pineapple. Retrieved May 7, 2019, from http://www.mpib.gov.my/en/origin-of-pineapple/

Muhammad, Z. A., & Ahmad, T. (2017). Therapeutic uses of pineapple-extracted bromelain in surgical care — a review. Journal of the Pakistan Medical Association, 67(1), 121–125. Retrieved from https://jpma.org.pk/article-details/8055?article_id=8055

Mukherjee, S., Majumdar, S., & Bhattacharyya, D. (2005). Role of hydrogen bonds in protein-DNA recognition: effect of nonplanar amino groups. Journal of Physical Chemistry B, 109(20), 10484–10492. https://doi.org/10.1021/jp0446231

Müller, A., Barat, S., Chen, X., Bui, K. C., Bozko, P., Malek, N. P., & Plentz, R. R. (2016). Comparative study of antitumor effects of bromelain and papain in human cholangiocarcinoma cell lines. International Journal of Oncology, 48(5), 2025–2034. https://doi.org/10.3892/ijo.2016.3411

Nägler, D. K., Tam, W., Storer, A. C., Krupa, J. C., Mort, J. S., & Menard, R. (1999). Interdependency of sequence and positional specificities for cysteine proteases of the papain family cysteine protease with unique carboxypeptidase activity. Biochemistry, 38(15), 4868–4874. https://doi.org/10.1021/bi982632s

Nair, I. C., & Jayachandran, K. (2019). Aspartic Proteases in Food industry. In B. Parameswaran, S. Varjani, & S. Raveendran (Eds.), Green Bio-processes: Enzymes in Industrial Food Processing (pp. 15–30). https://doi.org/10.1007/978-981-13-2324-9

Nieuwenhuizen, N. J., Maddumage, R., Tsang, G. K., Fraser, L. G., Cooney, J. M., De Silva, H. N., … Atkinson, R. G. (2012). Mapping, complementation, and targets of the cysteine protease actinidin in kiwifruit. Plant Physiology, 158(1), 376–388. https://doi.org/10.1104/pp.111.187989

Nieuwenhuizen, Niels J., Beuning, L. L., Sutherland, P. W., Sharma, N. N., Cooney, J. M., Bieleski, L. R. F., … Atkinson, R. G. (2007). Identification and characterisation of acidic and novel basic forms of actinidin, the highly abundant cysteine protease from kiwifruit. Functional Plant Biology, 34(10), 946–961. https://doi.org/10.1071/fp07121

Ning, X., Zhang, Y., Yuan, T., Li, Q., Tian, J., Guan, W., … Zhang, Y. (2018). Enhanced thermostability of glucose oxidase through computer-aided molecular design. International Journal of Molecular Sciences, 19(2). https://doi.org/10.3390/ijms19020425

Nishiyama, I. (2007). Fruits of the Actinidia genus. In Advances in food and nutrition research (Vol. 52, pp. 293–324). https://doi.org/10.1016/S1043-4526(06)52006-6

Ota, S., & Muta, E. (1985). Reinvestigation of fractionation active and some components properties of stem of the proteolytically and fruit bromelains. The Journal of Biochemistry, 98(1), 219–228. https://doi.org/10.1093/oxfordjournals.jbchem.a135261

Pace, C. N, & Scholtz, J. M. (1998). A helix propensity scale based on experimental studies of peptides and proteins. Biophysical Journal, 75(1), 422–427. https://doi.org/10.1016/s0006-3495(98)77529-0

Pace, C. Nick, Fu, H., Fryar, K. L., Landua, J., Trevino, S. R., Schell, D., … Grimsley, G. R. (2014). Contribution of hydrogen bonds to protein stability. Protein Science, 23(5), 652–661. https://doi.org/10.1002/pro.2449

Padmanabhan, S., Banerjee, S., & Mandi, N. (2011). Screening of bacterial recombinants: strategies and preventing false positives. In B. Gregory (Ed.), Molecular cloning - selected applications in medicine and biology (pp. 1–20). https://doi.org/10.5772/22140

Papamichael, E. M., Roustas, M. K., & Bieth, J. G. (2017). Detection of S1-P1 and S3-P3 interactions between papain and four synthetic substrates. Brazilian Archives of Biology and Technology, 42(3), 277–280. https://doi.org/10.1590/s1516-89131999000300002

Park, H., Ovchinnikov, S., Kim, D. E., DiMaio, F., & Baker, D. (2018). Protein homology model refinement by large-scale energy optimization. Proceedings of the National Academy of Sciences, 115(12), 3054–3059. https://doi.org/10.1073/pnas.1719115115

Paul, M., Hazra, M., Barman, A., & Hazra, S. (2014). Comparative molecular dynamics simulation studies for determining factors contributing to the thermostability of chemotaxis protein "CheY." Journal of Biomolecular Structure and Dynamics, 32(6), 928–949. https://doi.org/10.1080/07391102.2013.799438

Pavan, R., Jain, S., Shraddha, & Kumar, A. (2012). Properties and therapeutic application of bromelain: a review. Biotechnology Research International, 2012, 1–6. https://doi.org/10.1155/2012/976203

Pavlopoulou, A., & Michalopoulos, I. (2011). State-of-the-art bioinformatics protein structure prediction tools (review). International Journal of Molecular Medicine, 28(3), 295–310. https://doi.org/10.3892/ijmm.2011.705

Perczel, A., Gaspari, Z., & Csizmadia, I. G. (2005). Structure and stability of the beta-pleated sheets. Journal of Computational Chemistry, 26(11), 1155–1168. https://doi.org/10.1002/jcc.20255

Portaro, F. C. V., Santos, A. B. F., Cezari, M. H. S., Juliano, M. A., Juliano, L., & Carmona, E. (2000). Probing the specificity of cysteine proteinases at subsites remote from the active site: analysis of P4, P3, P2′ and P3′ variations in extended substrates. Biochemical Journal, 347(1), 123–129. https://doi.org/10.1042/bj3470123

Pucci, F., & Rooman, M. (2017). Physical and molecular bases of protein thermal stability and cold adaptation. Current Opinion in Structural Biology, 42, 117–128. https://doi.org/10.1016/j.sbi.2016.12.007

Quail, M. A., Smith, M., Coupland, P., Otto, T. D., Harris, S. R., Connor, T. R., … Gu, Y. (2012). A tale of three next generation sequencing platforms: comparison of Ion torrent, Pacific Biosciences and Illumina MiSeq sequencers. BMC Genomics, 13(1), 1–13. https://doi.org/10.1186/1471-2164-13-341

Raimbault, A. K., Zuily-Fodil, Y., Soler, A., Mora, P., & de Carvalho, M. H. C. (2013). The expression patterns of bromelain and AcCYS1 correlate with blackheart resistance in pineapple fruits submitted to postharvest chilling stress. Journal of Plant Physiology, 170(16), 1442–1446. https://doi.org/10.1016/j.jplph.2013.05.008

Ramalingam, C., Srinath, R., & Islam, N. N. (2012). Isolation and characterization of bromelain from pineapple (Ananas Comosus) and comparing its anti-browning activity on apple juice with commercial anti- browning agents. Elixir Food Science, 45, 7822–7826.

Ramli, A. N. M., Azhar, M. A., Shamsir, M. S., Rabu, A., Murad, A. M. A., Mahadi, N. M., & Md. Illias, R. (2013). Sequence and structural investigation of a novel psychrophilic α-amylase from Glaciozyma antarctica PI12 for cold-adaptation analysis. Journal of Molecular Modeling, 19(8), 3369–3383. https://doi.org/10.1007/s00894-013-1861-5

Ramli, A. N. M., Mahadi, N. M., Shamsir, M. S., Rabu, A., Joyce-Tan, K. H., Murad, A. M. A., & Md. Illias, R. (2012). Structural prediction of a novel chitinase from the psychrophilic Glaciozyma antarctica PI12 and an analysis of its structural properties and function. Journal of Computer-Aided Molecular Design, 26(8), 947–961.

https://doi.org/10.1007/s10822-012-9585-7

Ramli, A. N. M., Manas, N. H. A., Hamid, A. A. A., Hamid, H. A., & Illias, R. M. (2018). Comparative structural analysis of fruit and stem bromelain from Ananas comosus. Food Chemistry, 266, 183–191. https://doi.org/10.1016/j.foodchem.2018.05.125

Ramsaroop, R. E. S., & Saulo, A. A. (2007). Comparative consumer and physicochemical analysis of Del Monte Hawai'I Gold and Smooth Cayenne pineapple cultivars. Journal of Food Quality, 30(2), 135–159. https://doi.org/10.1111/j.1745-4557.2007.00111.x

Raskovic, B., Lazic, J., & Polovic, N. (2016). Characterisation of general proteolytic, milk clotting and antifungal activity of Ficus carica latex during fruit ripening. Journal of the Science of Food and Agriculture, 96(2), 576–582. https://doi.org/10.1002/jsfa.7126

Rathnavelu, V., Alitheen, N., Sohila, S., Kanagesan, S., & Ramesh, R. (2016). Potential role of bromelain in clinical and therapeutic applications (review). Biomedical Reports, 5(3), 283–288. https://doi.org/10.3892/br.2016.720

Raval, A., Piana, S., Eastwood, M. P., Dror, R. O., & Shaw, D. E. (2012). Refinement of protein structure homology models via long, all-atom molecular dynamics simulations. Proteins, 80(8), 2071–2079. https://doi.org/10.1002/prot.24098

Rawlings, N. D., Barrett, A. J., Thomas, P. D., Huang, X., Bateman, A., & Finn, R. D. (2018). The MEROPS database of proteolytic enzymes, their substrates and inhibitors in 2017 and a comparison with peptidases in the PANTHER database. Nucleic Acids Research, 46(D1), D624–D632. https://doi.org/10.1093/nar/gkx1134

Redwan, R. M., Saidin, A., & Kumar, S. V. (2016). The draft genome of MD-2 pineapple using hybrid error correction of long reads. DNA Research, 23(5), 427–439. https://doi.org/10.1093/dnares/dsw026

Rhoads, A., & Au, K. F. (2015). PacBio Sequencing and its applications. Genomics, Proteomics and Bioinformatics, 13(5), 278–289. https://doi.org/10.1016/j.gpb.2015.08.002

Robinson, P. K. (2015). Enzymes: principles and biotechnological applications. Essays In Biochemistry, 59(0), 1–41. https://doi.org/10.1042/bse0590075

Rohrbach, K. G., & Johnson, M. W. (2003). Pests, diseases and weeds. In D. P. Bartholomew, R. E. Paull, & K. G. Rohrbach (Eds.), The pineapple: botany,

production and uses (pp. 203–252). https://doi.org/10.1144/GSL.MEM.2002.024.01.09

Rosa, M., Roberts, C. J., & Rodrigues, M. A. (2017). Connecting high-temperature and low temperature protein stability and aggregation. PLoS ONE, 12(5), 1–12. https://doi.org/10.1371/journal.pone.0176748

Rosnah, S., Daud, W. R. R., Takrif, M. S., & Hassan, O. (2009). Physico-mechanical properties of the Josapine pineapple fruits. Pertanika Journal of Science & Technology, 17(1), 117–123.

Rowan, A. D., Buttle, D. J., & Barrett, A. J. (1990). The cysteine proteinases of the pineapple plant. Biochemical Journal, 266(3), 869–875. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/2327970%0Ahttp://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC1131219

Roy, S., Choudhury, D., Aich, P., Dattagupta, J. K., & Biswas, S. (2012). The structure of a thermostable mutant of pro-papain reveals its activation mechanism. Acta Crystallographica Section D: Biological Crystallography, 68(12), 1591–1603. https://doi.org/10.1107/s0907444912038607

Russell, R. J. M., Ferguson, J. M. C., Hough, D. W., Danson, M. J., & Taylor, G. L. (1997). The crystal structure of citrate synthase from the hyperthermophilic archaeon Pyrococcus furiosus at 1.9 Å resolution. Biochemistry, 36(33), 9983–9994. https://doi.org/10.1021/bi9705321

Sali, A., & Blundell, T. L. (1993). Comparative protein modelling by satisfaction of spatial restraints. Journal of Molecular Biology, 234(3), 779–815. https://doi.org/10.1006/jmbi.1993.1626

Salleh, A. B., Rahim, A. S. M. A., Rahman, R. N. Z. R. A., Leow, T. C., & Basri, M. (2012). The role of Arg157Ser in improving the compactness and stability of ARM lipase. Journal of Computer Science & Systems Biology, 05(02), 39–46. https://doi.org/10.4172/jcsb.1000088

Salmela, L., Walve, R., Rivals, E., Ukkonen, E., & Sahinalp, C. (2017). Accurate self-correction of errors in long reads using de Bruijn graphs. Bioinformatics, 33(6), 799–806. https://doi.org/10.1093/bioinformatics/btw321

Sambrook, J., & Russell, D. W. (2006). Alkaline Agarose Gel Electrophoresis. In Cold Spring Harbor Protocols. https://doi.org/10.1101/pdb.prot4027

Sefidbakht, Y., Ranaei Siadat, O., & Taheri, F. (2017). Homology modeling and molecular dynamics study on Schwanniomyces occidentalis alpha-amylase. Journal of Biomolecular Structure and Dynamics, 35(3), 574–584. https://doi.org/10.1080/07391102.2016.1154892

Shaji., D. (2017). Multi-template homology modeling of human MCT8 Protein. International Journal of Advanced Research, 5(7), 1025–1036. https://doi.org/10.21474/ijar01/4811

Sharma, A., & Chatterjee, B. (2017). Review study on enzymes in fruits. SciFed Journal of Protein Science, 1(1), 1–6. Retrieved from http://scifedpublishers.com/fulltext/review-study-on-enzymes-in-fruits/21893

Shehu, A., & Kavraki, L. E. (2012). Modeling structures and motions of loops in protein molecules. Entropy, 14(2), 252–290. https://doi.org/10.3390/e14020252

Shen, M. Y., & Sali, A. (2006). Statistical potential for assessment and prediction of protein structures. Protein Science, 15(11), 2507–2524. https://doi.org/10.1110/ps.062416606.Instead

Shendure, J., & Ji, H. (2008). Next-generation DNA sequencing. Nature Biotechnology, 26(10), 1135–1145. https://doi.org/10.1038/nbt1486

Shinde, U. P., Liu, J. J., & Inouye, M. (1997). Protein memory through altered folding mediated by intramolecular chaperones. Nature, 389(6650), 520–522. https://doi.org/10.1038/39097

Sinha, R., & Khare, S. K. (2013). Thermostable protease. In T. Satyanarayana, J. Littlechild, & Y. Kawarabayasi (Eds.), Thermophilic microbes in environmental and industrial biotechnology: biotechnology of thermophiles (pp. 859–880). https://doi.org/10.1007/978-94-007-5899-5

Sokalingam, S., Raghunathan, G., Soundrarajan, N., & Lee, S. G. (2012). A study on the effect of surface lysine to arginine mutagenesis on protein stability and structure using green fluorescent protein. PLoS ONE, 7(7), 1–12. https://doi.org/10.1371/journal.pone.0040410

Sosa-Pagán, J. O., Iversen, E. S., & Grandl, J. (2017). TRPV1 temperature activation is specifically sensitive to strong decreases in amino acid hydrophobicity. Scientific Reports, 7(1), 1–10. https://doi.org/10.1038/s41598-017-00636-4

Srujana, N. S. V., & Narayana, M. K. (2017). Extraction , purification and characterization of bromelain from pineapple and in silico annotation of the protein. Helix, 7(4), 1799–1805.

Szilágyi, A., & Závodszky, P. (2000). Structural differences between mesophilic, moderately thermophilic and extremely thermophilic protein subunits: results of a comprehensive survey. Structure, 8, 493–504. Retrieved from papers://47009fda-d2bb-40d3-9a68-dc08a98706c7/Paper/p2854

Tap, F. M., Majid, F. A. A., & Khairudin, N. B. A. (2016). Structure prediction of stem bromelain from pineapples (Ananas comosus) using procaricain enzyme as a modelling template. International Journal of Applied Engineering Research, 11(9), 6109–6111.

Tap, F. M., Majid, F. A. A., & Khairudin, N. B. A. (2017). Tertiary structure prediction of bromelain from Ananas comosus using comparative modelling method. International Journal of Advanced and Applied Sciences, 4(12), 31–35. https://doi.org/10.21833/ijaas.2017.012.007

Taralp, A., Kaplan, H., Sytwu, I. I., Vlattas, I., Bohacek, R., Knap, A. K., … Hasnain, S. (1995). Characterization of the S3 subsite specificity of cathepsin B. Journal of Biological Chemistry, 270(30), 18036–18043. https://doi.org/10.1074/jbc.270.30.18036

Taylor, S., Wakem, M., Dijkman, G., Alsarraj, M., & Nguyen, M. (2010). A practical approach to RT-qPCR-publishing data that conform to the MIQE guidelines. Methods, 50(4), S1–S5. https://doi.org/10.1016/j.ymeth.2010.01.005

Tchoupé, J. R., Moreau, T., Gauthier, F., & Bieth, J. G. (1991). Photometric or fluorometric assay of cathepsin B, L and H and papain using substrates with an aminotrifluoromethylcoumarin leaving group. Biochimica et Biophysica Acta (BBA)/Protein Structure and Molecular, 1076(1), 149–151. https://doi.org/10.1016/0167-4838(91)90232-O

Thalip, A. A., Tong, P. S., & Ng, C. (2015). The MD2 "Super Sweet" pineapple (Ananas comosus). Utar Agriculture Science Journal, 1(4), 14–17. Retrieved from http://eprints.utar.edu.my/1982/1/The_MD2_(Super_Sweet)_pineapple_(Ananas_comosus).pdf

Tsai, J., Bonneau, R., Morozov, A. V., Kuhlman, B., Rohl, C. A., & Baker, D. (2003). An improved protein decoy set for testing energy functions for protein structure prediction. Proteins: Structure, Function and Genetics, 53(1), 76–87. https://doi.org/10.1002/prot.10454

Turk, D., Guncar, G., Podobnik, M., & Turk, B. (1998). Revised definition of substrate binding sites of papain-like cysteine proteases. Biological Chemistry, 379(2), 137–147.

Turk, V., Stoka, V., Vasiljeva, O., Renko, M., Sun, T., Turk, B., & Turk, D. (2012). Cysteine cathepsins: from structure, function and regulation to new frontiers. Biochimica et Biophysica Acta - Proteins and Proteomics, 1824(1), 68–88. https://doi.org/10.1016/j.bbapap.2011.10.002

Upadhyay, A., Lama, J. P., & Tawata, S. (2010). Utilization of pineapple waste: a review. Journal of Food Science and Technology Nepal, 6, 10–18. https://doi.org/10.3126/jfstn.v6i0.8255

van der Hoorn, R. A. L., Leeuwenburgh, M. A., Bogyo, M., Joosten, M. H. A., & Peck, S. C. (2004). Activity profiling of papain-like cysteine proteases in plants. Plant Physiology, 135, 1170–1178. https://doi.org/10.1104/pp.104.041467.1170

Vel, T., & Stanley, S. (2015). Studies on the effect of the different modes of extraction on enzyme activity of proteolytic enzymes from the peels of three fruit samples Carica papaya (papaya), Ananas comosus (pineapple) and Actinidia deliciosa (kiwi). The SciTech Journal, Vol. 02, pp. 27–32. Retrieved from http://oaji.net/articles/2015/215-1424676162.pdf

Verma, S., Dixit, R., & Pandey, K. C. (2016). Cysteine proteases: modes of activation and future prospects as pharmacological targets. Frontiers in Pharmacology, 7(107), 1–12. https://doi.org/10.3389/fphar.2016.00107

Vogt, G., & Argos, P. (1997). Protein thermal stability: hydrogen bonds or internal packing? Folding and Design, 2(4), 40–46. https://doi.org/10.1016/S1359-0278(97)00062-X

Wali, N. (2019). Pineapple (Ananas comosus). In S. Nabavi & A. Silva (Eds.), Nonvitamin and nonmineral nutritional supplements (Vol. 1, pp. 367–373). https://doi.org/10.1016/B978-0-12-812491-8.00050-3

Wang, Wei, Zhang, L., Guo, N., Zhang, X., Zhang, C., Sun, G., & Xie, J. (2014). Functional properties of a cysteine proteinase from pineapple fruit with improved resistance to fungal pathogens in Arabidopsis thaliana. Molecules, 19(2), 2374–2389. https://doi.org/10.3390/molecules19022374

Wang, Weihao, Cai, J., Wang, P., Tian, S., & Qin, G. (2017). Post-transcriptional regulation of fruit ripening and disease resistance in tomato by the vacuolar protease SlVPE3. Genome Biology, 18(1), 1–23. https://doi.org/10.1186/s13059-017-1178-2

Wang, Z., Eickholt, J., & Cheng, J. (2010). MULTICOM: a multi-level combination approach to protein structure prediction and its assessments in CASP8. Bioinformatics, 26(7), 882–888. https://doi.org/10.1093/bioinformatics/btq058

Wintrode, P. L., Zhang, D., Vaidehi, N., Arnold, F. H., & Goddard, W. A. (2003). Protein dynamics in a family of laboratory evolved thermophilic enzymes. Journal of Molecular Biology, 327(3), 745–757. https://doi.org/10.1016/S0022-2836(03)00147-5

Wong, M. L., & Medrano, J. F. (2005). Real-time PCR for mRNA quantitation. BioTechniques, 39(1), 75–85. https://doi.org/10.2144/05391RV01

Wong, S. W. K., Liu, J. S., & Kou, S. C. (2017). Fast de novo discovery of low-energy protein loop conformations. Proteins: Structure, Function and Bioinformatics, 85(8), 1402–1412. https://doi.org/10.1002/prot.25300

Woodman, M. E., Savage, C. R., Arnold, W. K., & Stevenson, B. (2018). Direct PCR of intact bacteria (colony PCR). Current Protocols in Microbiology, 42(1), A.3D.1-A.3D.7. https://doi.org/10.1002/cpmc.14

Wu, J., Xiao, J., Zhang, Z., Wang, X., Hu, S., & Yu, J. (2014). Ribogenomics: the science and knowledge of RNA. Genomics, Proteomics and Bioinformatics, 12(2), 57–63. https://doi.org/10.1016/j.gpb.2014.04.002

Wu, W. C., Ng, H. S., Sun, I. M., & Lan, J. C. W. (2017). Single step purification of bromelain from Ananas comosus pulp using a polymer/salt aqueous biphasic system. Journal of the Taiwan Institute of Chemical Engineers, 79, 158–162. https://doi.org/10.1016/j.jtice.2017.04.001

Wu, X., Xu, P., Wang, J., Xu, Y., Fu, T., Zhang, D., … Li, G. (2015). Folding mechanisms of trefoil knot proteins studied by molecular dynamics simulations and go-model. In D. Wei, Q. Xu, T. Zhao, & H. Dai (Eds.), Advance in structural bioinformatics (pp. 93–110). https://doi.org/10.1007/978-94-017-9245-5

Yao, S., Hart, D. J., & An, Y. (2016). Recent advances in universal TA cloning methods for use in function studies. Protein Engineering, Design and Selection, 29(11), 551–556. https://doi.org/10.1093/protein/gzw047

Yedavalli, P., & Rao, N. M. (2013). Engineering the loops in a lipase for stability in DMSO. Protein Engineering, Design and Selection, 26(4), 317–324. https://doi.org/10.1093/protein/gzt002

Yennamalli, R. M., Rader, A. J., Wolt, J. D., & Sen, T. Z. (2011). Thermostability in endoglucanases is fold-specific. BMC Structural Biology, 11(10). https://doi.org/10.1186/1472-6807-11-10

Yu, H., Yan, Y., Zhang, C., & Dalby, P. A. (2017). Two strategies to engineer flexible loops for improved enzyme thermostability. Scientific Reports, 7, 1–15. https://doi.org/10.1038/srep41212

Zeiske, T., Stafford, K. A., & Palmer III, A. G. (2016). Thermostability of enzymes from molecular dynamics simulations. Journal of Chemical Theory and Computation, 12(6), 2489–2492. https://doi.org/10.1021/acs.jctc.6b00120

Zhu, X., Wang, J., Peng, B., & Shete, S. (2016). Empirical estimation of sequencing error rates using smoothing splines. BMC Bioinformatics, 17(1), 1–11. https://doi.org/10.1186/s12859-016-1052-3

Zou, Z., Huang, Q., Xie, G., & Yang, L. (2018). Genome-wide comparative analysis of papain-like cysteine protease family genes in castor bean and physic nut. *Scientific Reports*, *8*(1), 1–13. https://doi.org/10.1038/s41598-017-18760-6
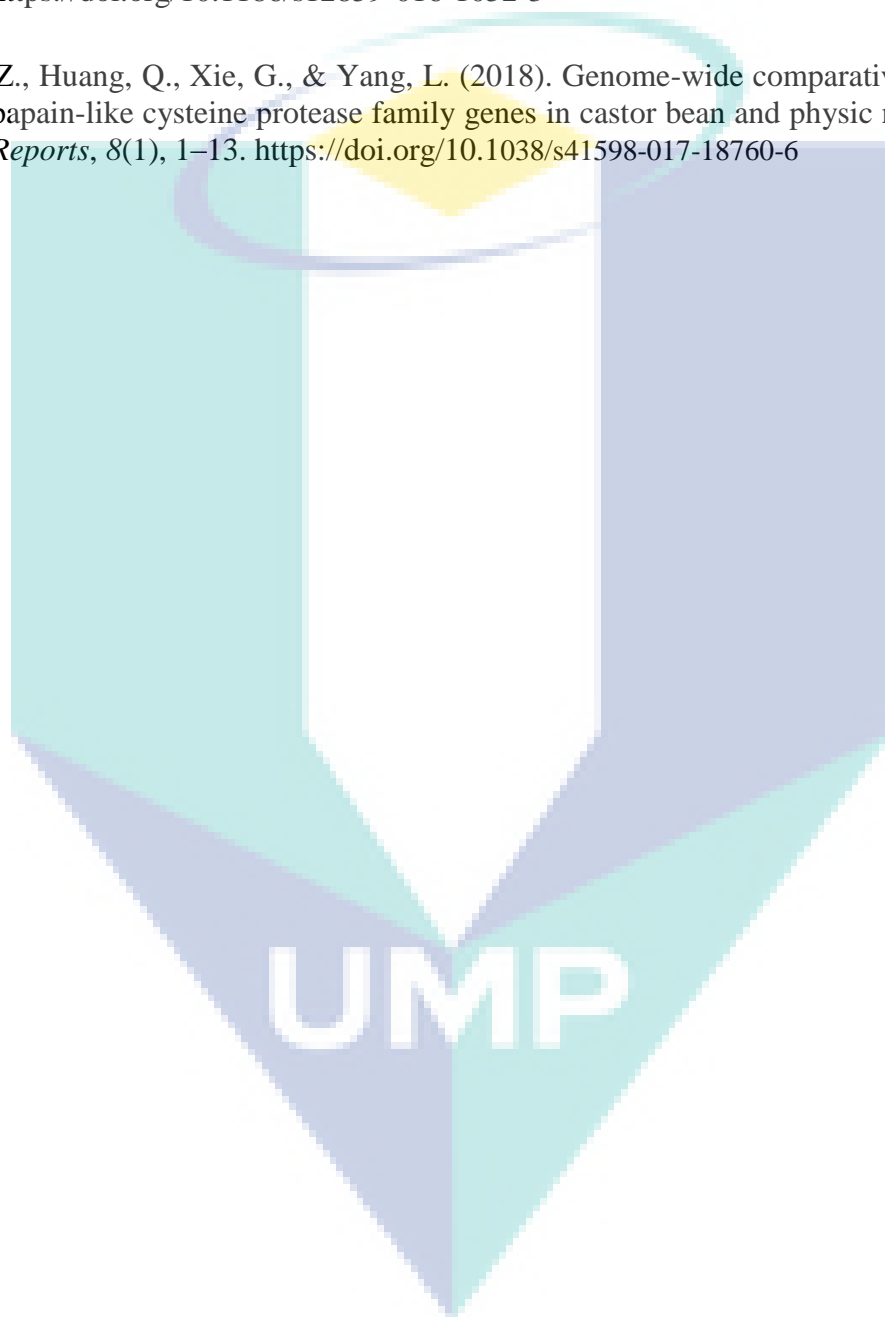
APPENDIX A
QPCR ANALYSIS OF FRUIT BROMELAIN

Table A1 C_q values of fruit bromelain and actin at different annealing temperature

Fruit bromelain primers				
Temperature (°C)	Replicate 1	Replicate 2	Replicate 3	Mean reading ± SD
55	28.02	28.00	28.07	28.03 ± 0.04
57	27.47	26.96	27.50	27.31 ± 0.30
60	27.54	27.92	27.73	27.73 ± 0.19
62	30.68	30.73	30.72	30.71 ± 0.03
65	0.00	0.00	0.00	0.00 ± 0.00
Actin primers				
Temperature (°C)	Replicate 1	Replicate 2	Replicate 3	Mean reading ± SD
55	24.24	23.87	24.61	24.24 ± 0.37
57	24.27	24.17	24.33	24.26 ± 0.09
60	25.34	25.40	24.94	25.23 ± 0.25
62	28.65	28.76	28.74	28.73 ± 0.06
65	0.00	0.00	0.00	0.00 ± 0.00

Table A2 C_q values of different starting quantity in PCR reaction

Fruit bromelain				
Starting quantity (5^)	Replicate 1	Replicate 2	Replicate 3	Mean reading ± SD
-1	22.24	22.20	21.94	22.13 ± 0.16
-2	24.12	24.22	23.98	24.11 ± 0.12
-3	28.11	28.13	27.82	28.02 ± 0.17
-4	30.35	30.59	30.42	30.45 ± 0. 12
-5	34.70	34.61	35.01	34.77 ± 0.21
Actin				
Starting quantity (5^)	Replicate 1	Replicate 2	Replicate 3	Mean reading ± SD
-1	24.87	25.00	25.32	25.06 ± 0.23
-2	26.94	27.18	26.77	26.96 ± 0.21
-3	30.85	31.45	31.44	31.25 ± 0.34
-4	33.39	33.31	33.29	33.33 ± 0.05
-5	37.30	37.95	37.34	37.53 ± 0.36
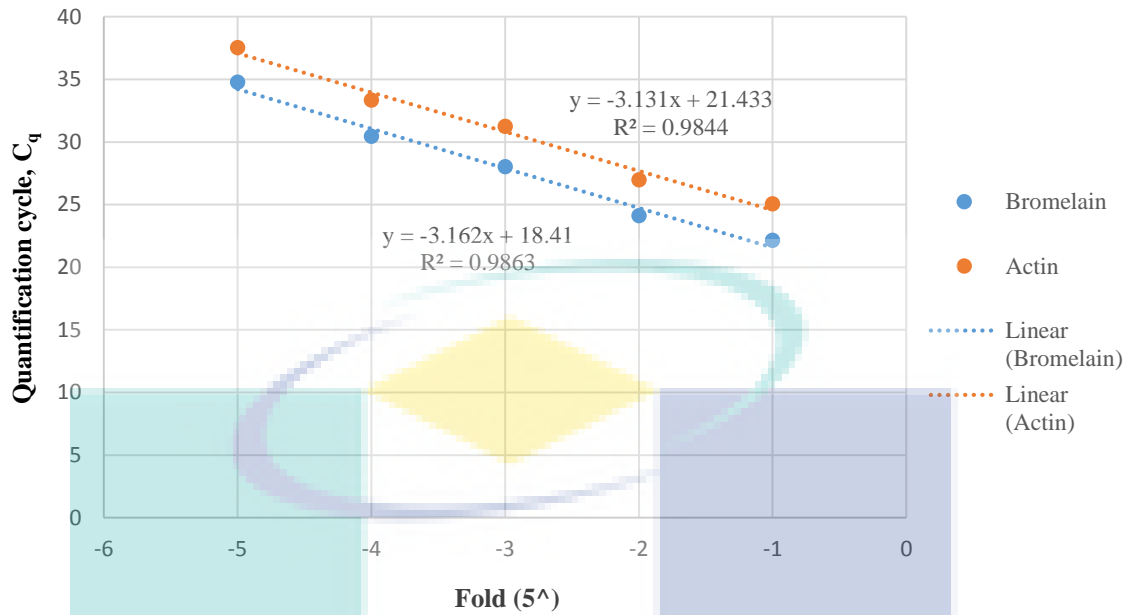
Figure A1   Standard curve of fruit bromelain and actin generated with 5-fold serial dilution.

Table A3 C_q values of fruit bromelain and actin in unripe and ripe *A. comosus* cultivar MD 2 normalised by actin

	Fruit bromelain			
	Replicate 1	**Replicate 2**	**Replicate 3**	**Mean reading ± SE**
Unripe 1	18.94	18.95	18.60	20.06 ± 0.70
Unripe 2	21.25	21.42	21.14	
Unripe 3	20.05	20.05	20.12	
Ripe 1	27.57	27.20	27.19	26.35 ± 0.65
Ripe 2	26.47	26.72	26.67	
Ripe 3	24.97	25.04	25.35	
	Actin			
	Replicate 1	**Replicate 2**	**Replicate 3**	**Mean reading ± SE**
Unripe 1	23.62	23.92	24.36	24.10 ± 0.28
Unripe 2	24.32	24.80	24.77	
Unripe 3	23.50	23.82	23.76	
Ripe 1	23.19	23.19	22.94	27.10 ± 2.08
Ripe 2	28.17	28.00	28.00	
Ripe 3	30.00	30.24	30.15	

ENZYMATIC ANALYSIS OF FRUIT BROMELAIN

Table B1   Absorbance readings of L-tyrosine with different concentration at 660 nm

Conc. (µmol)	Replicate 1	Replicate 2	Replicate 3	Mean reading ± SD
0.055	0.090	0.093	0.095	0.093 ± 0.003
0.111	0.189	0.196	0.190	0.192 ± 0.004
0.221	0.356	0.349	0.360	0.355 ± 0.006
0.442	0.700	0.680	0.675	0.685 ± 0.013
0.553	0.851	0.856	0.850	0.852 ± 0.003
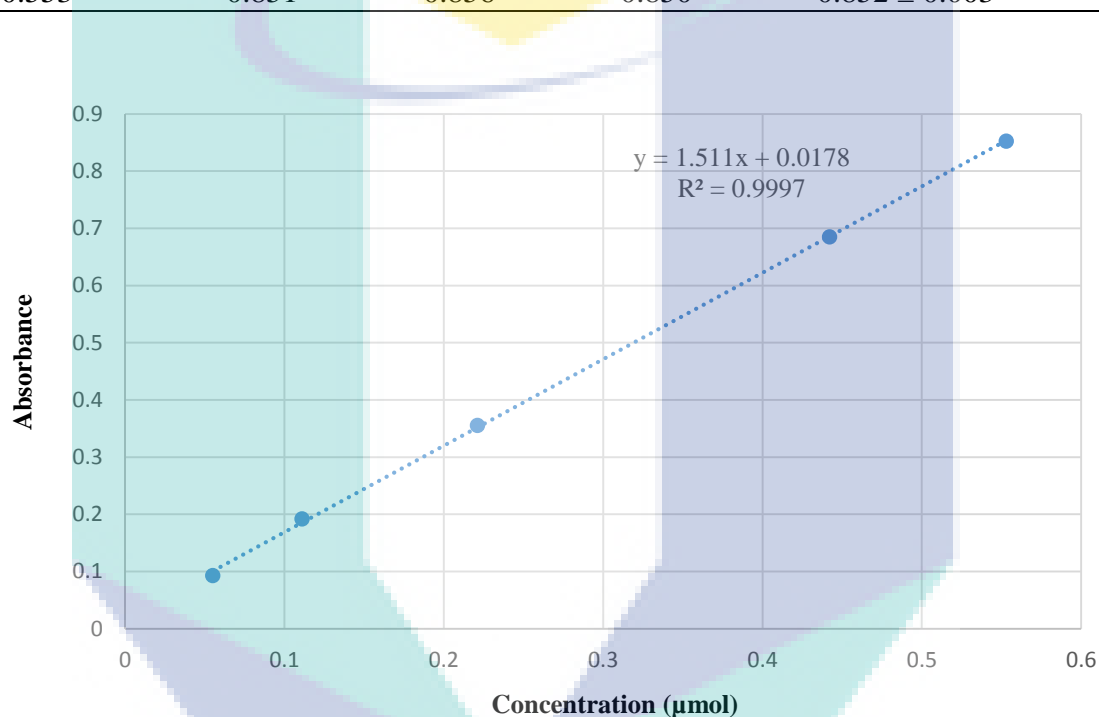
Figure B1   Standard curve of L-Tyrosine of different concentration at 660 nm.

Table B2   Enzymatic activity of crude fruit bromelain from unripe and ripe *A.comosus* cultivar MD 2

Sample	Replicate 1 (U/mL)	Replicate 2 (U/mL)	Replicate 3 (U/mL)	Mean reading ± SE
Unripe 1	2.00	1.82	1.98	1.91 ± 0.08
Unripe 2	1.80	1.78	1.69	
Unripe 3	1.97	2.09	2.07	
Ripe 1	1.25	1.19	1.35	1.13 ± 0.09
Ripe 2	0.97	0.96	0.96	
Ripe 3	1.16	1.14	1.22	

APPENDIX C
SEQUENCE ANALYSIS

Table C1   Accession number of *A. comosus* cultivar MD 2 fruit bromelain sequences, length and location of the I29 and PLCE domains

| Accession number	Length of amino acid	Location of I29 domain	Location of PLCE domain
OAY62650.1	351	37-93	122-334
OAY65848.1	302	1-53	86-299
OAY67114.1	150	-	1-147
OAY68270.1	319	5-61	91-302
OAY68387.1	258	37-93	122-258
OAY68854.1	359	36-92	123-336
OAY68894.1	326	38-94	129-324
OAY71019.1	326	37-93	121-324
OAY76881.1	666	37-93	121-323
		358-414	445-664
OAY80099.1	170	1-54	69-169
OAY80102.1	714	37-93	123-334
		386-442	472-652
OAY83410.1	1530	42-98	129-340
		356-411	441-648
		656-712	742-937
		960-1016	1024-1213
		1229-1285	1316-1479
OAY85826.1	352	37-93	126-335
OAY85828.1	170	1-54	69-170
OAY85856.1	1924	61-117	151-340
		361-417	447-645
		661-717	749-953
		959-1015	1046-1254
		1280-1336	1367-1565
		1602-1658	1689-1901
OAY85857.1	721	37-93	123-334
		398-454	485-698
OAY85858.1	351	37-93	125-334
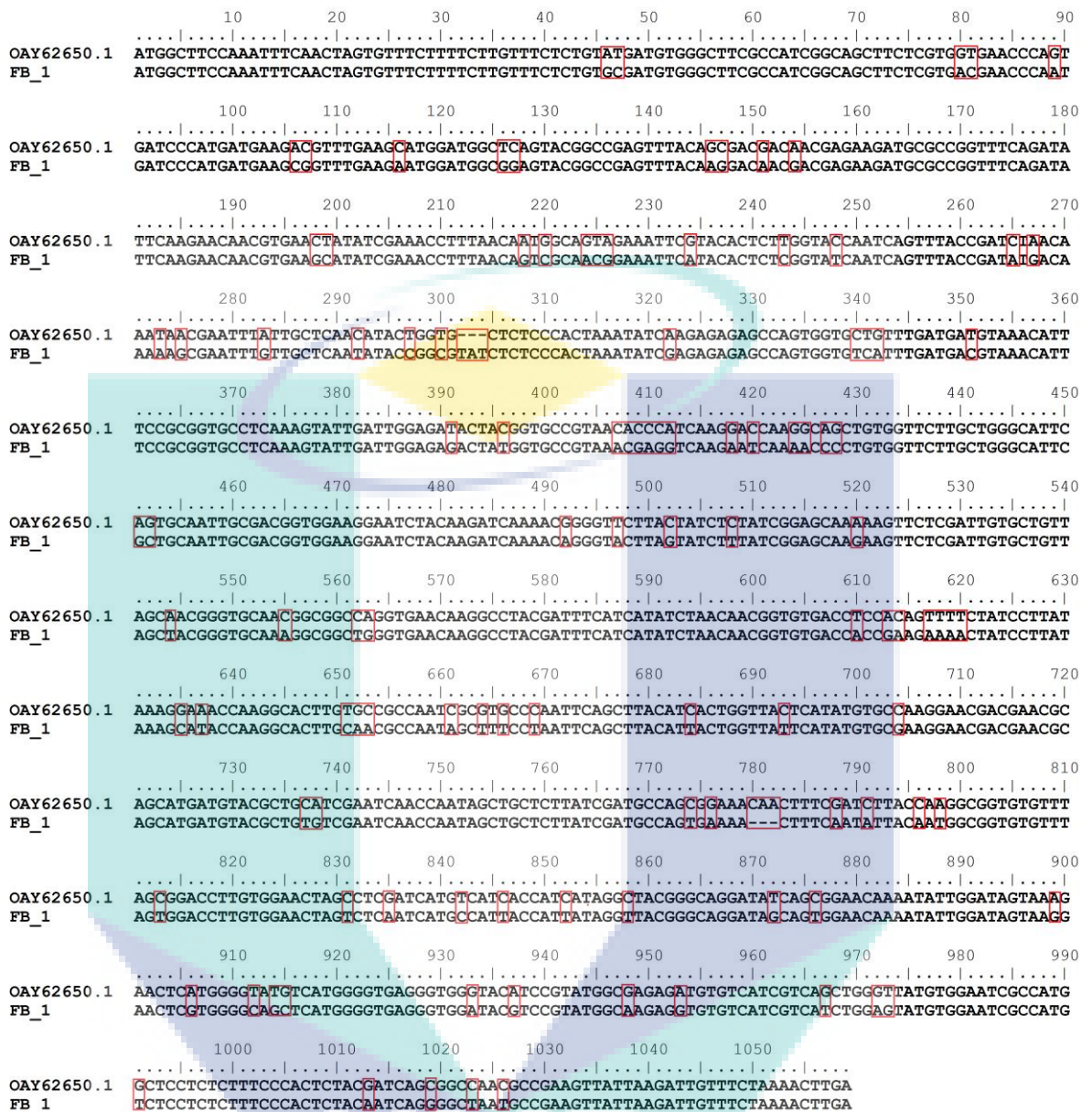
Figure C1   Sequence alignment of the isolated fruit bromelain nucleotide sequences FB_1 compared with original reference sequences OAY62650.1 retrieved from NCBI Genbank.

```
                10        20        30        40        50        60        70        80        90
            ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
OAY68270.1  ATGATGAAGCGGTTTGAAGAATGGATGGCGGAGTACGGCCGAGTTTACAAGGACAACGACGAGAAAATGCGCCGGTTTCAGATATTCAAG
FB_2        ATGATGAAGCGGTTTGAAGAATGGATGGCGGAGTACGGCCGAGTTTACAAGGACAACGACGAGAAAATGCGCCGGTTTCAGATATTCAAG

                100       110       120       130       140       150       160       170       180
            ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
OAY68270.1  AACAACGTGAACCATATCGAAACCTTTAACAGTCGCAACGGAAATTCGTACACTCTCGGTATCAATCAGTTTACCGATATGACAAAAAGC
FB_2        AACAACGTGAACCATATCGAAACCTTTAACAGTCGCAACGGAAATTCGTACACTCTCGGTATCAATCAGTTTACCGATATGACAAAAAGC

                190       200       210       220       230       240       250       260       270
            ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
OAY68270.1  GAATTTGTTGCTCAATATACCGGCGTATCTCTCCCACTAAATATCGAGAGAGAGCCAGTGGTGTCATTTGATGACGTAAACATCTCCGCA
FB_2        GAATTTGTTGCTCAATATACCGGCGTATCTCTCCCACTAAATATCGAGAGAGAGCCAGTGGTGTCATTTGATGACGTAAACATCTCCGCA

                280       290       300       310       320       330       340       350       360
            ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
OAY68270.1  GTGCCTCAAAGTATTGATTGGAGAGACTATGGTGCCGTAAACGAGGTCAAGAATCAAAACCCCTGTGGTTCTTGCTGGTCATTCGCTGCA
FB_2        GTGCCTCAAAGTATTGATTGGAGAGACTATGGTGCCGTAAACGAGGTCAAGAATCAAAACCCCTGTGGTTCTTGCTGGTCATTCGCTGCA

                370       380       390       400       410       420       430       440       450
            ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
OAY68270.1  ATTGCGACGGTGGAGGGAATCTACAAGATCAAAACAGGGTATTTAGTATCTTTATCAGAGCAAGAAGTTCTCGATTGTGCTGTTAGCTAC
FB_2        ATTGCGACGGTGGAGGGAATCTACAAGATCAAAACAGGGTATTTAGTATCTTTATCAGAGCAAGAAGTTCTCGATTGTGCTGTTAGCTAC

                460       470       480       490       500       510       520       530       540
            ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
OAY68270.1  GGGTGCAAAGGCGGCTGGGTGAACAAGGCCTACGATTTCATCATATCTAACAACGGTGTGACCACCGAAGAAAACTATCCTTATCAAGCA
FB_2        GGGTGCAAAGGCGGCTGGGTGAACAAGGCCTACGATTTCATCATATCTAACAACGGTGTGACCACCGAAGAAAACTATCCTTATCAAGCA

                550       560       570       580       590       600       610       620       630
            ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
OAY68270.1  TACCAAGGCACTTGCAACGCCAATAGCTTTCCTAATTCAGCTTACATTACTGGTTATTCATATGTGCGAAGGAACGACGAACGCAGTATG
FB_2        TACCAAGGCACTTGCAACGCCAATAGCTTTCCTAATTCAGCTTACATTACTGGTTATTCATATGTGCGAAGGAACGACGAACGCAGTATG

                640       650       660       670       680       690       700       710       720
            ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
OAY68270.1  ATGTACGCTGTGTCGAATCAACCAATAGCTGCTCTTATCGATGCCAGTGAAAACTTTCAATATTACAATGGCGGTGTGTTTAGCGGACCT
FB_2        ATGTACGCTGTGTCGAATCAACCAATAGCTGCTCTTATCGATGCCAGTGAAAACTTTCAATATTACAATGGCGGTGTGTTTAGCGGACCT

                730       740       750       760       770       780       790       800       810
            ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
OAY68270.1  TGTGGAACTAGTCTCAATCATGCCATTACCATTATAGGTTACGGGCAGGATAGCAGTGGAACAAAATATTGGATAGTAAGGAACTCGTGG
FB_2        TGTGGAACTAGTCTCAATCATGCCATTACCATTATAGGTTACGGGCAGGATAGCAGTGGAACAAAATATTGGATAGTAAGGAACTCGTGG

                820       830       840       850       860       870       880       890       900
            ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
OAY68270.1  GGCAGCTCATGGGGTGAGGGTGGATACGTCCGTATGGCAAGAGGTGTGTCATCGTCATCTGGAGTATGTGGAATCGCCATGTCTCCTCTC
FB_2        GGCAGCTCATGGGGTGAGGGTGGATACGTCCGTATGGCAAGAGGTGTGTCATCGTCATCTGGAGTATGTGGAATCGCCATGCCTCCTCTC

                910       920       930       940       950       960
            ....|....|....|....|....|....|....|....|....|....|....|....|....|
OAY68270.1  TTTCCCACTCTACAATCAGGGGCTAATGCCGAACTTATTAAGATGGTTTCTGAAACTTGA
FB_2        TTTCCCACTCTACAATCAGGGGCTAATGCCGAACTTATTAAGATGGTTTCTGAAACTTGA
```

Figure C2   Sequence alignment of the isolated fruit bromelain nucleotide sequences FB_2 compared with original reference sequences OAY68270.1 retrieved from NCBI Genbank.

```
              10        20        30        40        50        60        70        80        90
         ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
OAY8585.1 ATGGCTTCCAAAGTTCAACTCGTGTTTCTTTTCTTGTTTCTCTGTGTGATGTGGGCTTCGCCATCGGCAGCTTCTCGTGACGAACCCAGT
FB_3      ATGGCTTCCAAAGTTCAACTCGTGTTTCTTTTCTTGTTTCTCTGTGTGATGTGGGCTTCGCCATCGGCAGCTTCTCGTGACGAACCCAGT

                       100       110       120       130       140       150       160       170       180
         ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
OAY8585.1 GATCCCATGATGAAGCGGTTTGAAGAATGGATGGCAGAGTACGGCCGAGTGTACAAGGACAACGACGAGAAGATGCGCCGGTTTCAGATA
FB_3      GATCCCATGATGAAGCGGTTTGAAGAATGGATGGCAGAGTACGGCCGAGTGTACAAGGACAACGACGAGAAGATGCGCCGGTTTCAGATA

                       190       200       210       220       230       240       250       260       270
         ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
OAY8585.1 TTCAAGAACAACGTGAATCATATCGAAACCTTTAACAATCGCAACGGAAATTCATACACTCTCGGTATCAATAAGTTTACCGATATGACA
FB_3      TTCAAGAACAACGTGAATCATATCGAAACCTTTAACAATCGCAACGGAAATTCATACACTCTCGGTATCAATAAGTTTACCGATATGACA

                       280       290       300       310       320       330       340       350       360
         ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
OAY8585.1 AATAACGAATTTGTTGCTCAATATACTGGTG---TATCTCTCCCACTAAATTTCAAGAGAGAGCCAGTGGTGTCATTTGATGACGTAAAC
FB_3      AATAACGAATTTGTTGCTCAATATACCGGTGGTATATCTCGCCCATTAAATATCGAGAAAGAGCCAGTGGTGTCATTTGATGACGTAAAC

                       370       380       390       400       410       420       430       440       450
         ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
OAY8585.1 ATCTCTGCGGTGGGTCAAAGTATTGATTGGAGAGACTATGGTGCCGTAACAGAGGTCAAAGACCAAAACCCCTGTGGTTCTTGCTGGGCA
FB_3      ATCTCTGCGGTGGGTCAAAGTATTGATTGGAGAGACTATGGTGCCGTAACAGAGGTCAAGGACCAAAACCCCTGTGGTTCTTGCTGGGCA

                       460       470       480       490       500       510       520       530       540
         ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
OAY8585.1 TTCAGTGCAATTGCGACGGTGGAAGGAATCTACAAGATCGTAACAGGGTACTTAGTATCTCTATCGGAGCAAGAAGTTCTCGATTGTGCT
FB_3      TTCAGTGCAATTGCGACGGTGGAAGGAATCTACAAGATCGTAACAGGGTACTTAGTATCTCTATCGGAGCAAGAAGTTCTCGATTGTGCT

                       550       560       570       580       590       600       610       620       630
         ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
OAY8585.1 GTTAGCAATGGGTGCGACGGCGGCTTTGTGGACAATGCCTACGATTTCATCATATCTAACAACGGTGTGGCCTCCGAAGCTGACTATCCT
FB_3      GTTAGCAATGGGTGCGACGGCGGCTTTGTGGACAATGCCTACGATTTCATCATATCTAACAACGGTGTGGCCTCCGAAGCTGACTATCCT

                       640       650       660       670       680       690       700       710       720
         ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
OAY8585.1 TATCAAGCATACCAAGGCGATTGCGCCGCCAATAGCTGGCCCAATTCAGCGTACATTACTGGTTATTCATATGTGCGAAGCAACGACGAA
FB_3      TATCAAGCATACCAAGGCGATTGCGCCGCCAACAGCTGGCCCAATTCAGCATACATTACTGGTTATTCATATGTGCGAAGCAACGACGAA

                       730       740       750       760       770       780       790       800       810
         ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
OAY8585.1 AGCAGCATGAAGTACGCTGTGTGGAATCAACCAATAGCTGCTGCTATCGATGCCAGTGGAGACAACTTTCAATATTACAATGGCGGTGTG
FB_3      AGCAGCATGAAGTACGCTGTGTGGAATCAACCAATAGCTGCTGCTATCGATGCCAGTGGAGACAACTTTCAATATTACAATGGCGGTGTG

                       820       830       840       850       860       870       880       890       900
         ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
OAY8585.1 TTTAGTGGACCTTGTGGAACTAGTCTCAATCATGCCATCACCATTATAGGTTACGGGCAGGATAGCAGTGGAACACAATATTGGATTGTA
FB_3      TTTAGTGGACCTTGTGGAACTAGTCTCAATCATGCCATCACCATTATAGGTTACGGGCAGGATAGCAGTGGAACACAATATTGGATTGTA

                       910       920       930       940       950       960       970       980       990
         ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
OAY8585.1 AAGAACTCATGGGGTAGCTCATGGGGTGAACGTGGATACGTCCGTATGGCGAGAGGTGTGTCTTCGTCTGGATTATGTGGAATCGCCATG
FB_3      AAGAACTCATGGGGTAGCTCATGGGGTGAACGTGGATACATCCGTATGGCGAGAGGTGTGTCTTCGTCTGGATTATGTGGAATCGCCATG

                      1000      1010      1020      1030      1040      1050
         ....|....|....|....|....|....|....|....|....|....|....|....|....
OAY8585.1 GATCCTCTCTATCCCACTCTACAATCAGGGGCTAATGTCGCAGTTATTAAGATGGTTTCTGAAACTTGA
FB_3      GATCCTCTCTATCCCACTCTACAATCAGGGGCTAATGTCGCAGTTATTAAGATGGTTTCTGAAACTTGA
```

Figure C3   Sequence alignment of the isolated fruit bromelain nucleotide sequences FB_3 compared with original reference sequences OAY85858.1 retrieved from NCBI Genbank.
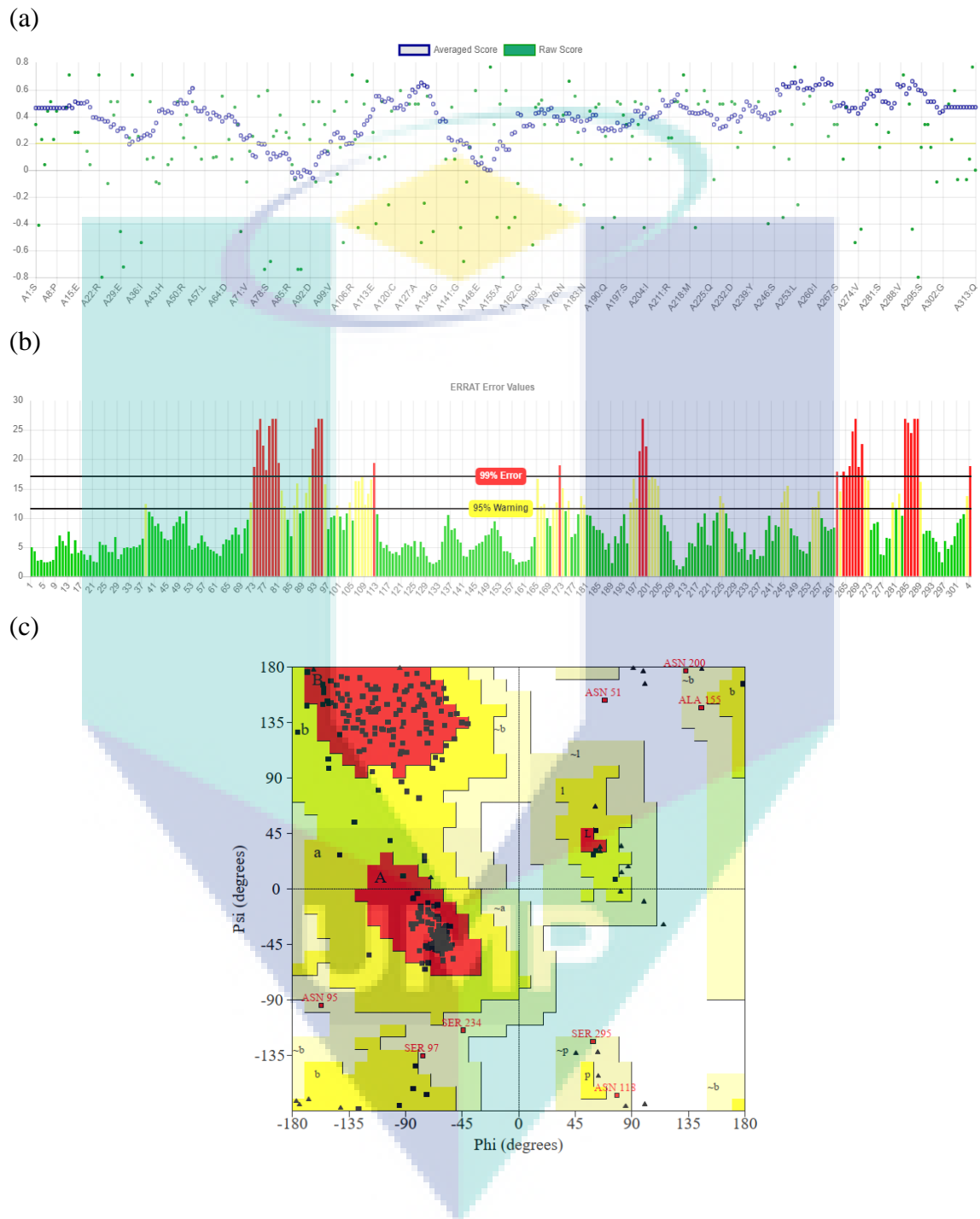
(a)

(b)

(c)

Figure D1   Structure validation of FB_1 using 1PCI as template. (a) Compatibility of atomic model with its own amino acid sequence in Verify 3D profile. (b) Plots the value of the error function in ERRAT. (c) Ramachandran plot in PROCHECK analysis.
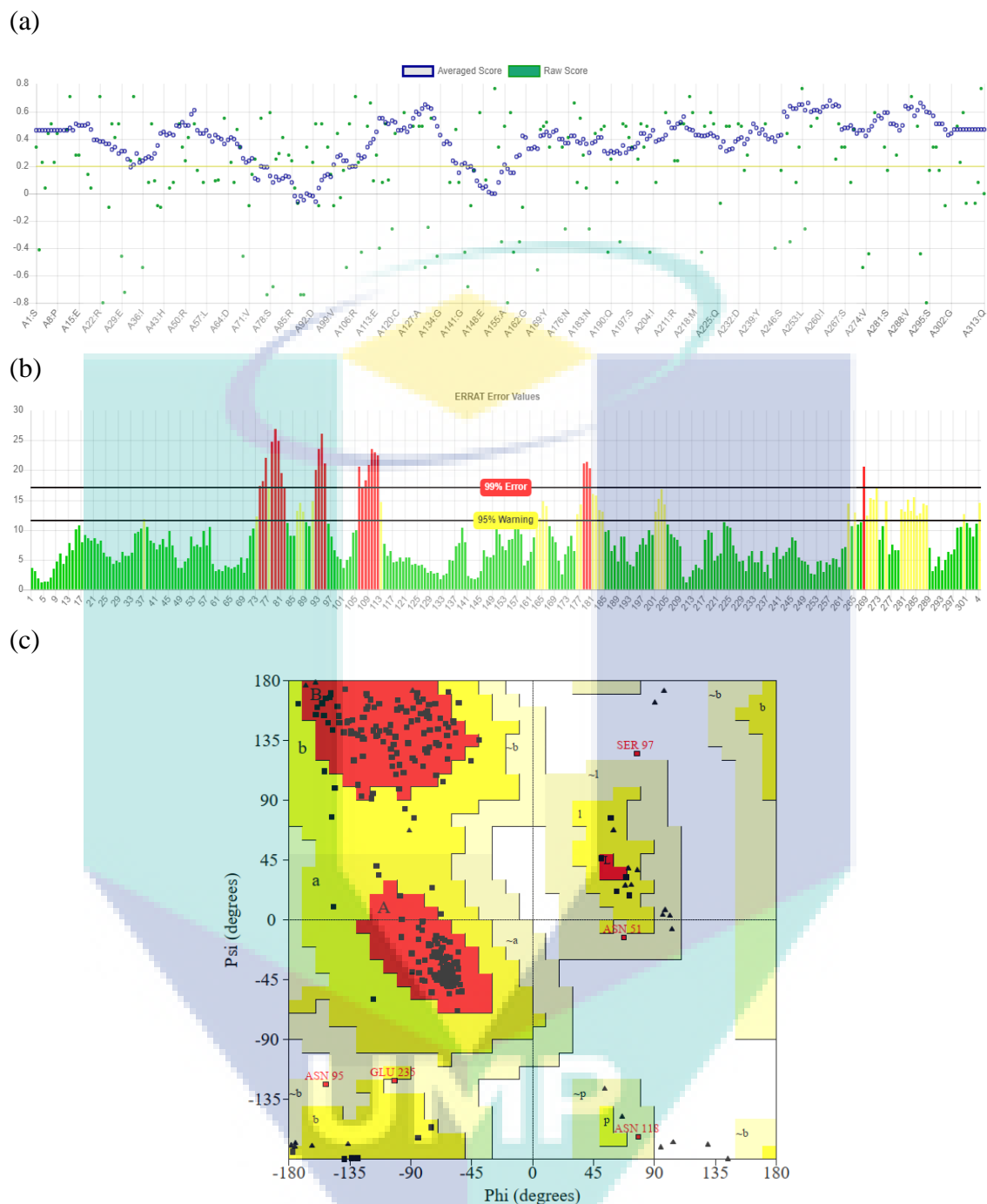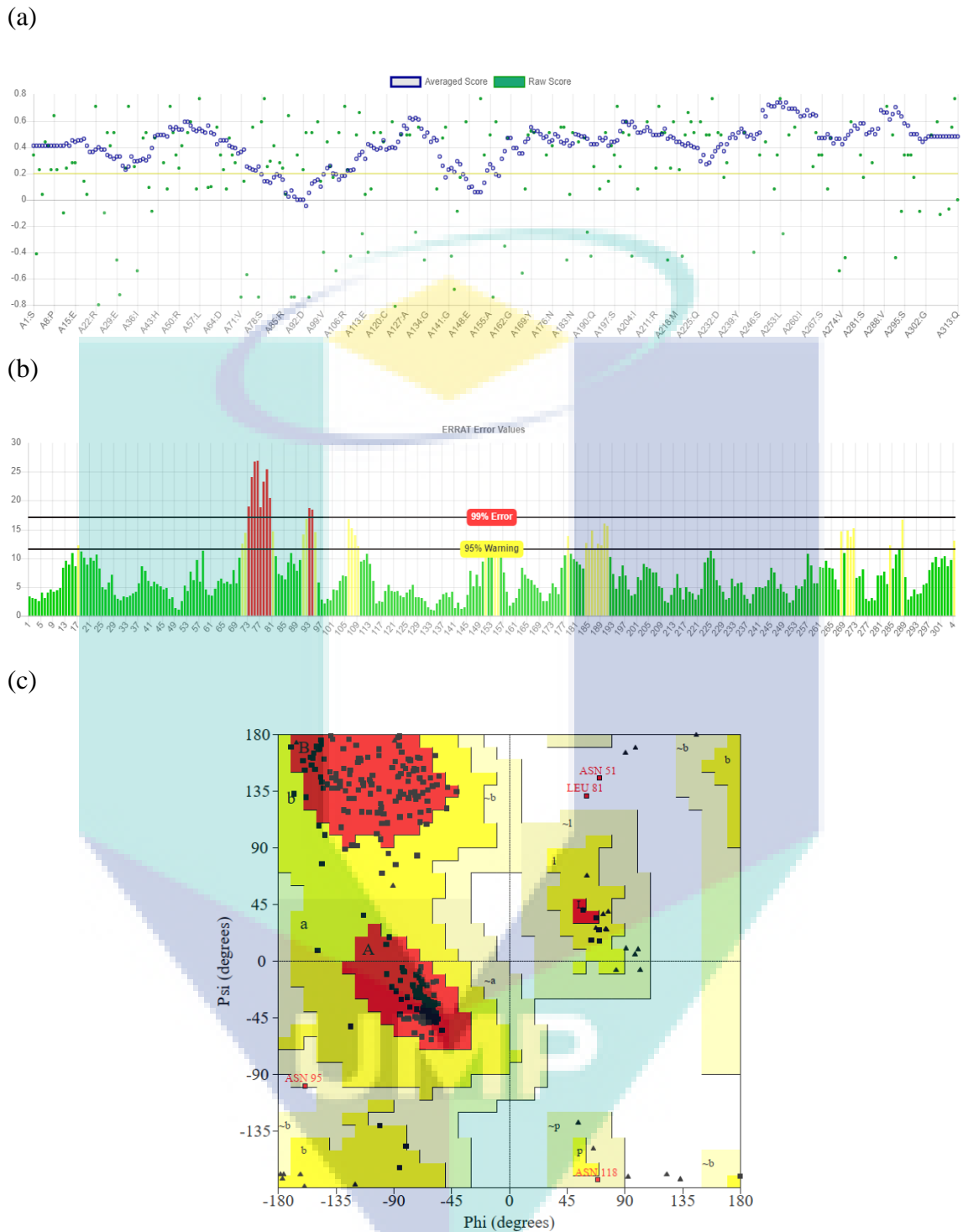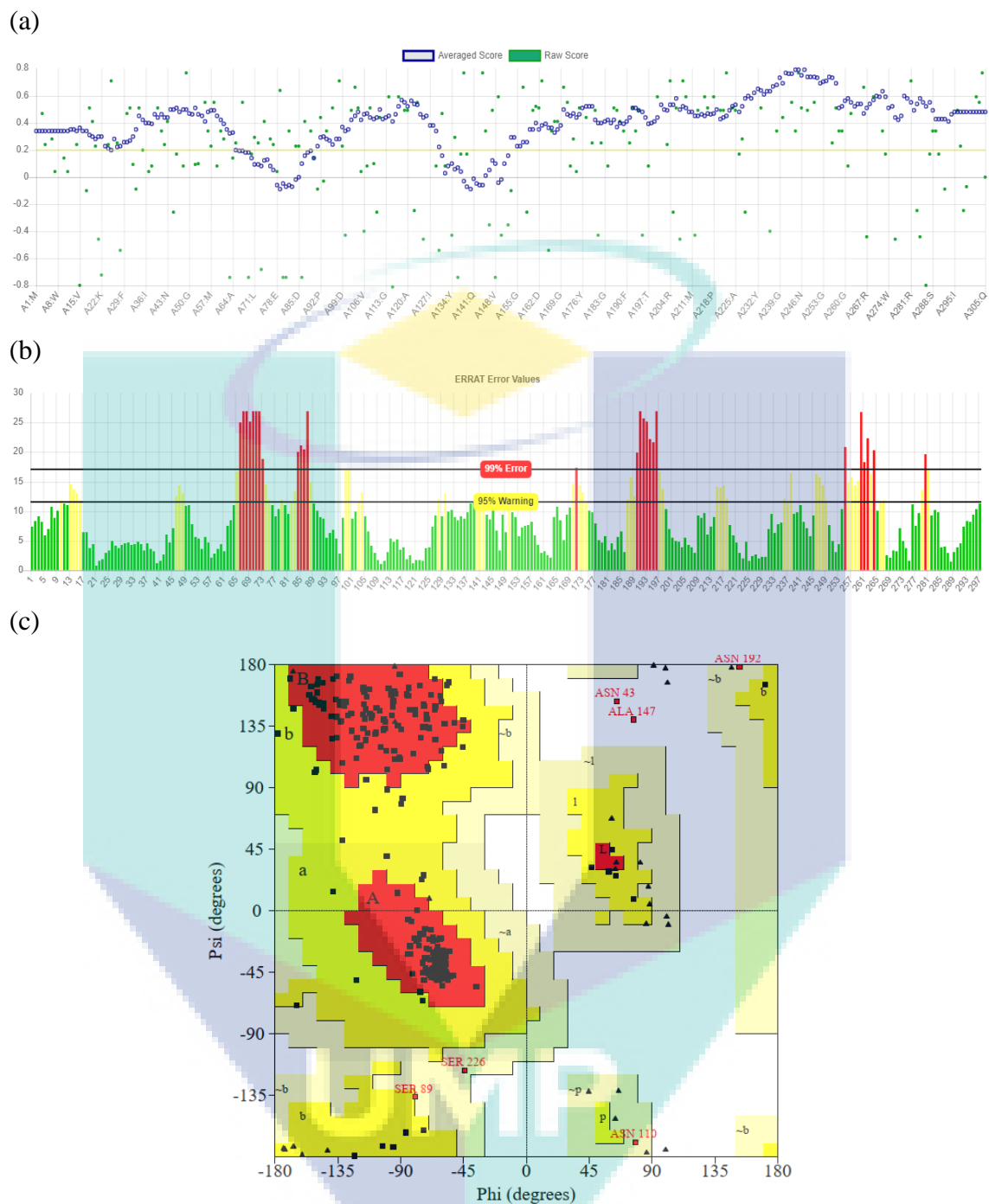
(a)

(b)

(c)

Figure D2   Structure validation of FB_1 using 1PCI-6MIR as template. (a) Compatibility of atomic model with its own amino acid sequence in Verify 3D profile. (b) Plots of values of the error function in ERRAT. (c) Ramachandran plot in PROCHECK analysis.

Figure D3   Structure validation of FB_1 using 1PCI-6MIR-4QRV as template. (a) Compatibility of atomic model with its own amino acid sequence in Verify 3D profile. (b) Plots of values of the error function in ERRAT. (c) Ramachandran plot in PROCHECK analysis.
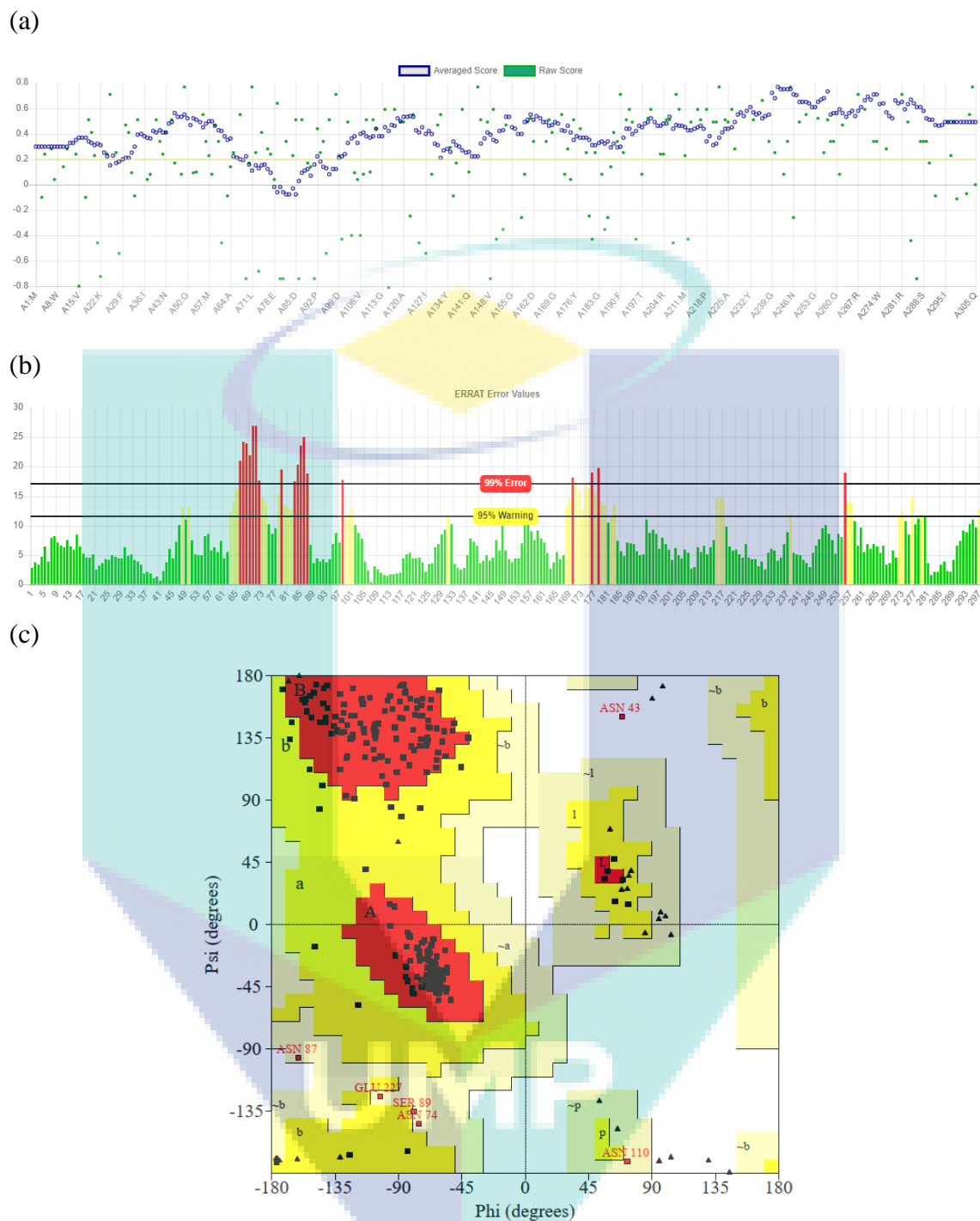
(a)

(b)

(c)

Figure D4   Structure validation of FB_2 using 1PCI as template. (a) Compatibility of atomic model with its own amino acid sequence in Verify 3D profile. (b) Plots of values of the error function in ERRAT. (c) Ramachandran plot in PROCHECK analysis.
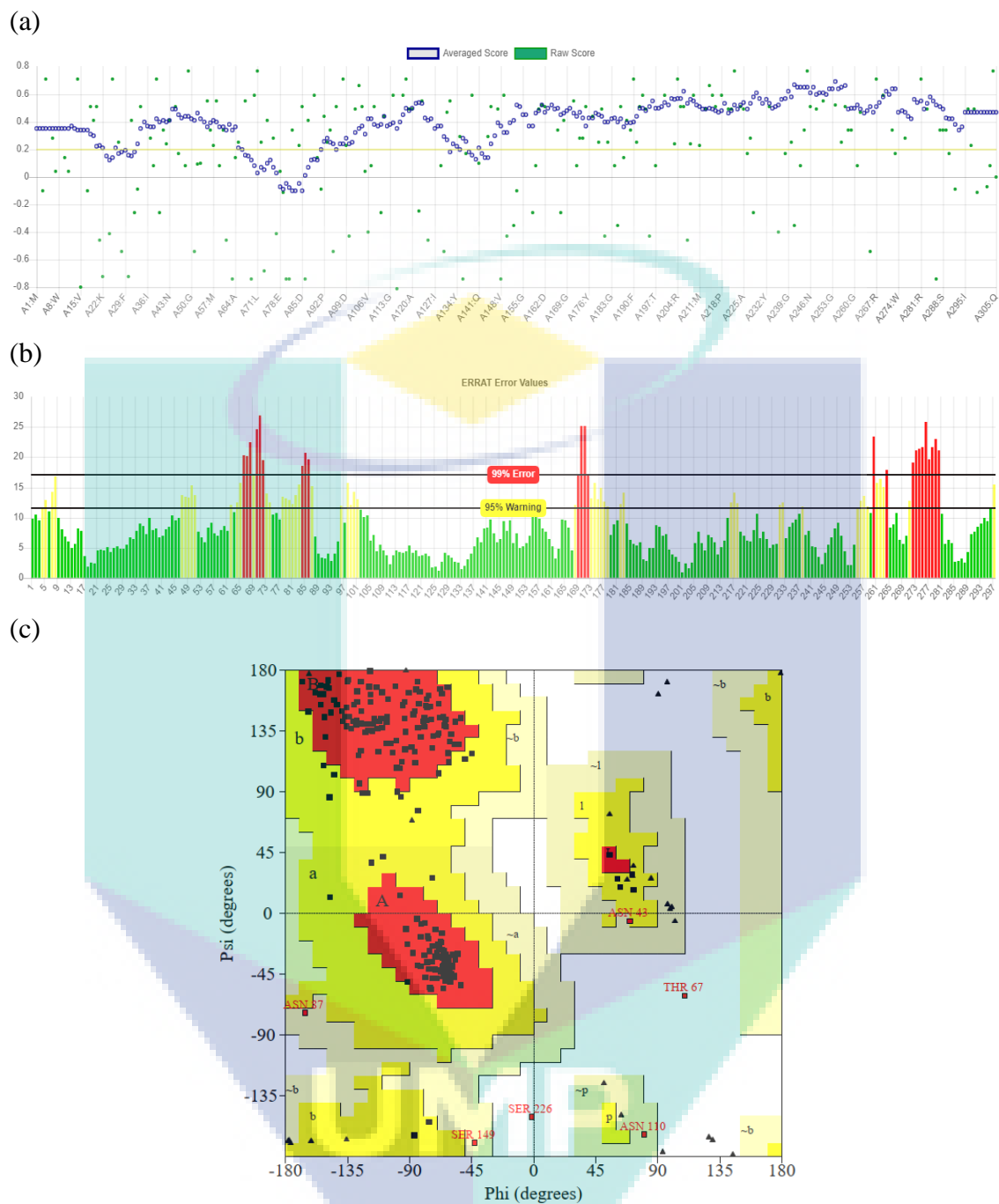
(a)

(b)

(c)

Figure D5   Structure validation of FB_2 using 1PCI-6MIR as template. (a) Compatibility of atomic model with its own amino acid sequence in Verify 3D profile. (b) Plots of values of the error function in ERRAT. (c) Ramachandran plot in PROCHECK analysis.

Figure D6   Structure validation of FB_2 using 1PCI-6MIR-4QRV as template. (a) Compatibility of atomic model with its own amino acid sequence in Verify 3D profile. (b) Plots of values of the error function in ERRAT. (c) Ramachandran plot in PROCHECK analysis.
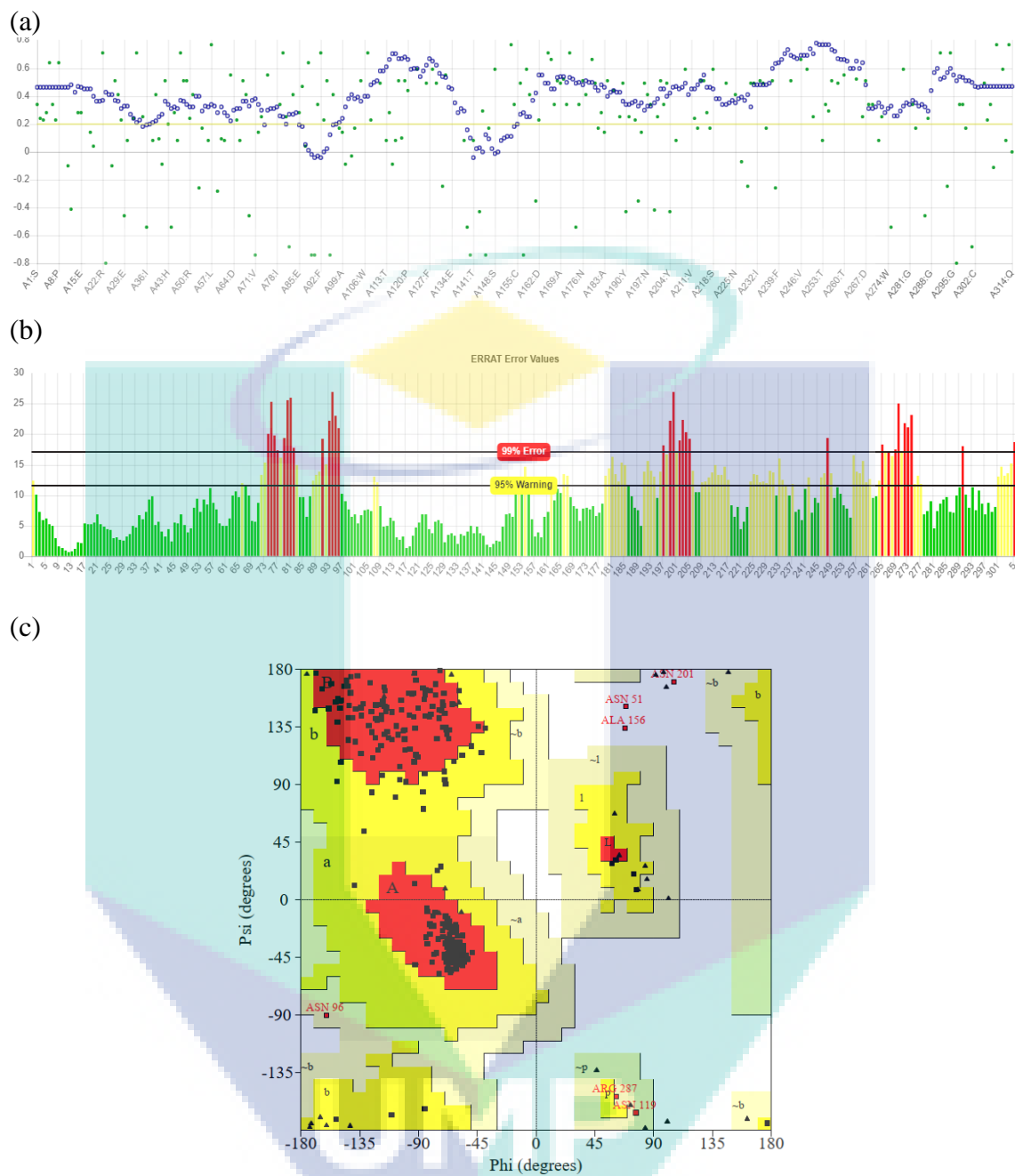
(a)

(b)

(c)

Figure D7   Structure validation of FB_3 using 1PCI as template. (a) Compatibility of atomic model with its own amino acid sequence in Verify 3D profile. (b) Plots of values of the error function in ERRAT. (c) Ramachandran plot in PROCHECK analysis.
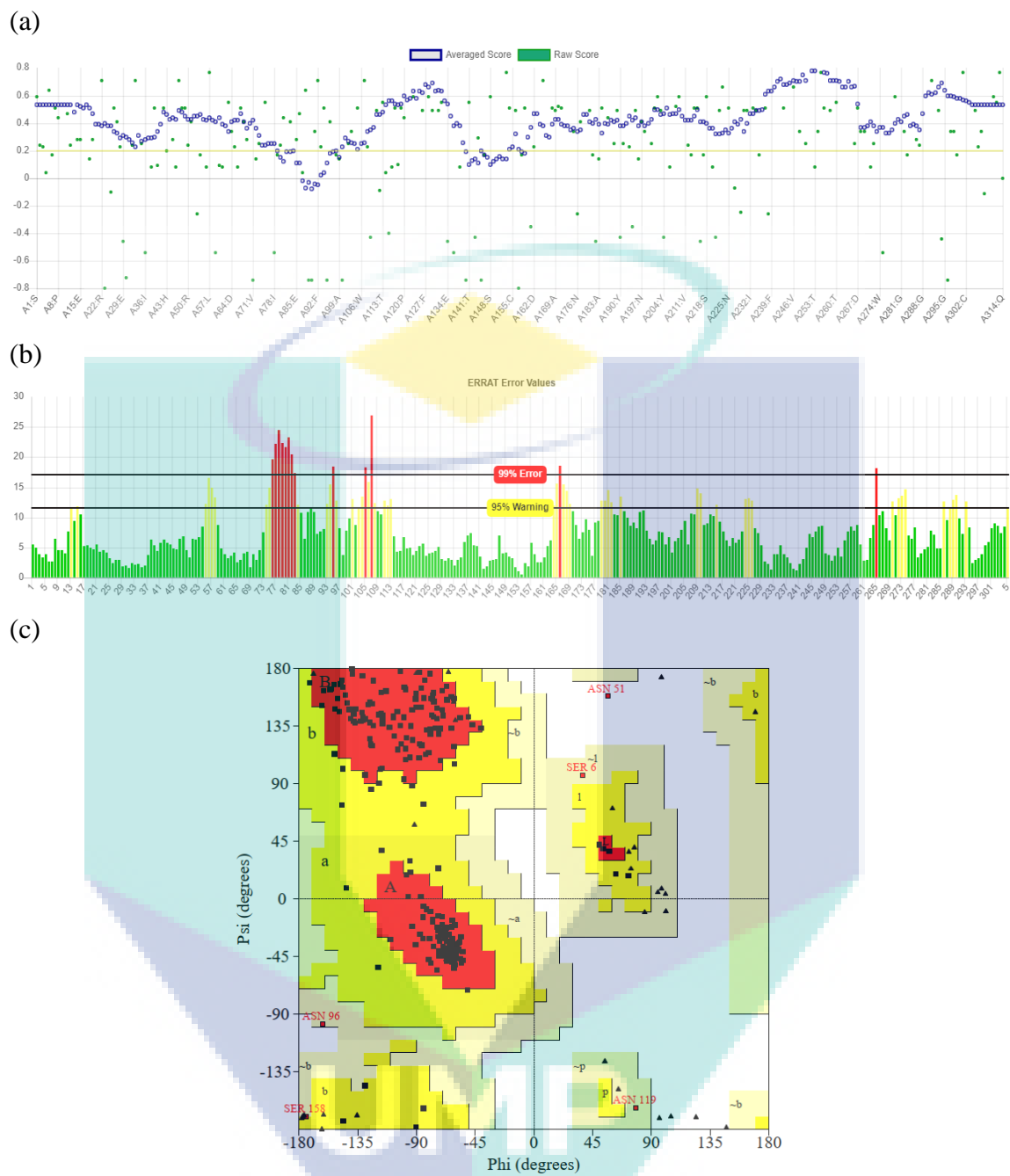
(a)

(b)

(c)

Figure D8   Structure validation of FB_3 using 1PCI-6MIR as template. (a) Compatibility of atomic model with its own amino acid sequence in Verify 3D profile. (b) Plots of values of the error function in ERRAT. (c) Ramachandran plot in PROCHECK analysis.
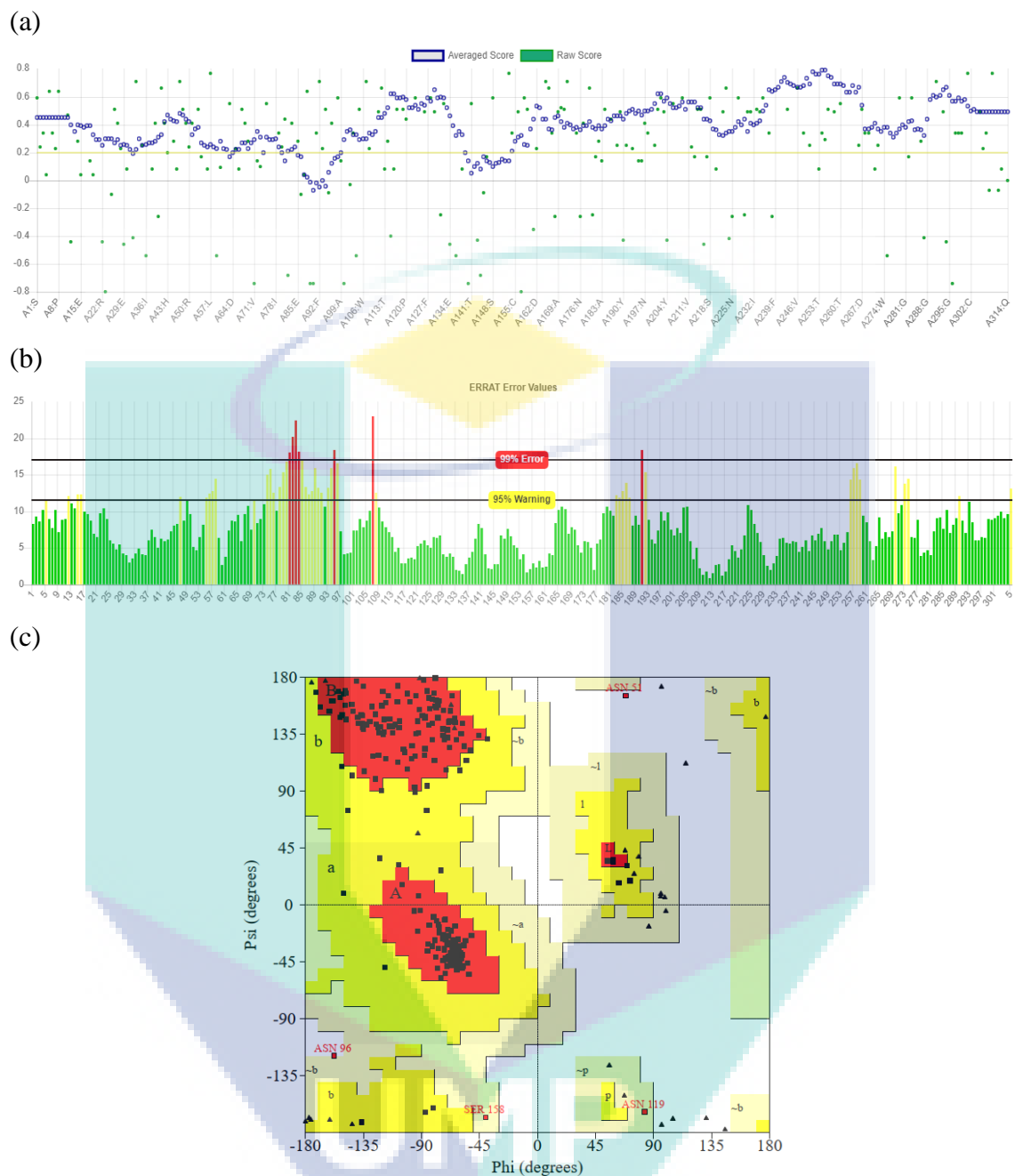
(a)

(b)

(c)

Figure D9   Structure validation of FB_3 using 1PCI-6MIR-3TNX as template. (a) Compatibility of atomic model with its own amino acid sequence in Verify 3D profile. (b) Plots of values of the error function in ERRAT. (c) Ramachandran plot in PROCHECK analysis.

Figure D10   Structure validation of refined FB_1. (a) Compatibility of atomic model with its own amino acid sequence in Verify 3D profile. (b) Plots of values of the error function in ERRAT. (c) Ramachandran plot in PROCHECK analysis.

(a)

(b)

(c)

Figure D11   Structure validation of refined FB_2. (a) Compatibility of atomic model with its own amino acid sequence in Verify 3D profile (b) Plots of values of the error function in ERRAT. (c) Ramachandran plot in PROCHECK analysis.
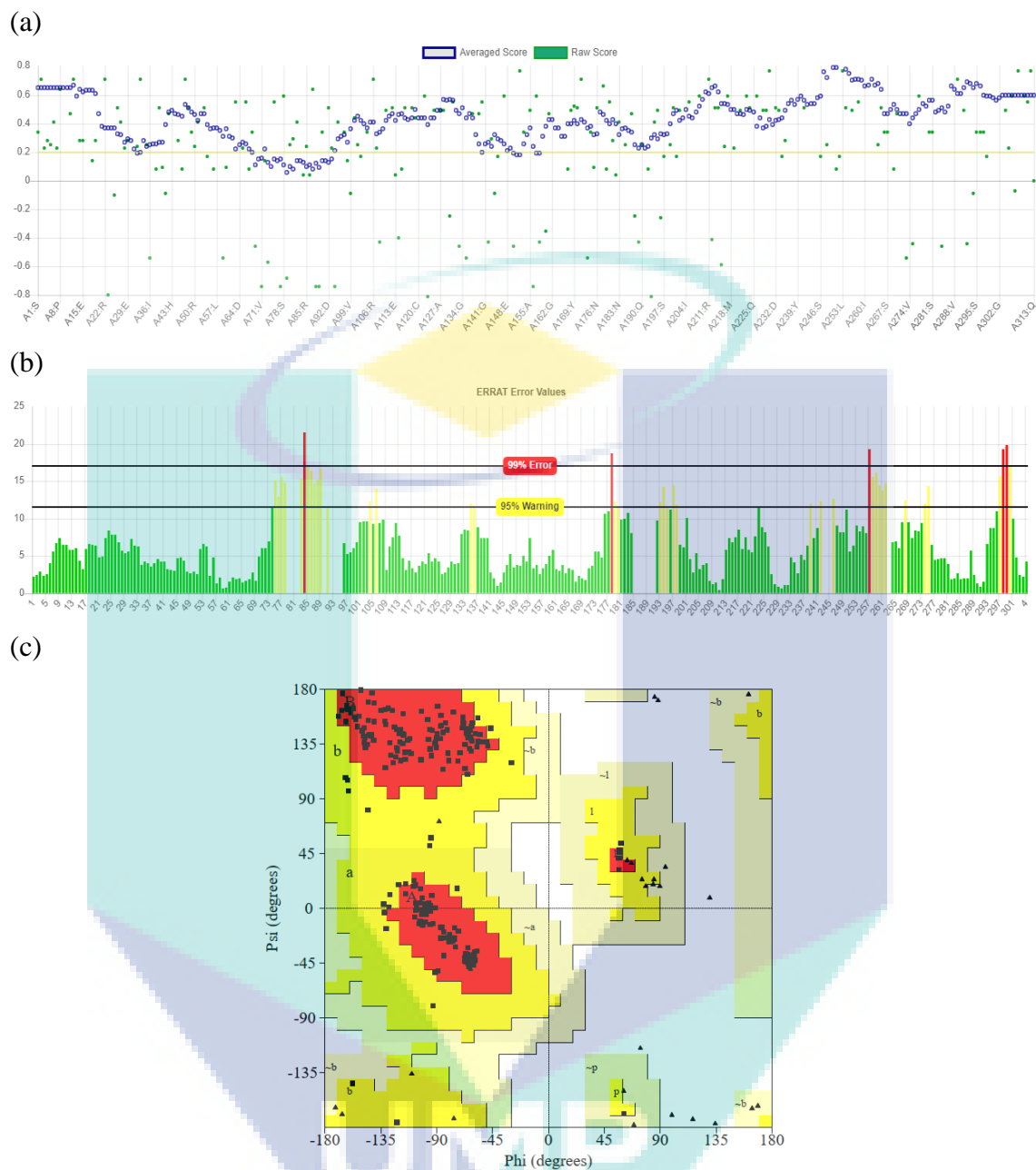
138

Figure D12   Structure validation of refined FB_3. (a) Compatibility of atomic model with its own amino acid sequence in Verify 3D profile. (b) Plots of values of the error function in ERRAT. (c) Ramachandran plot in PROCHECK analysis.
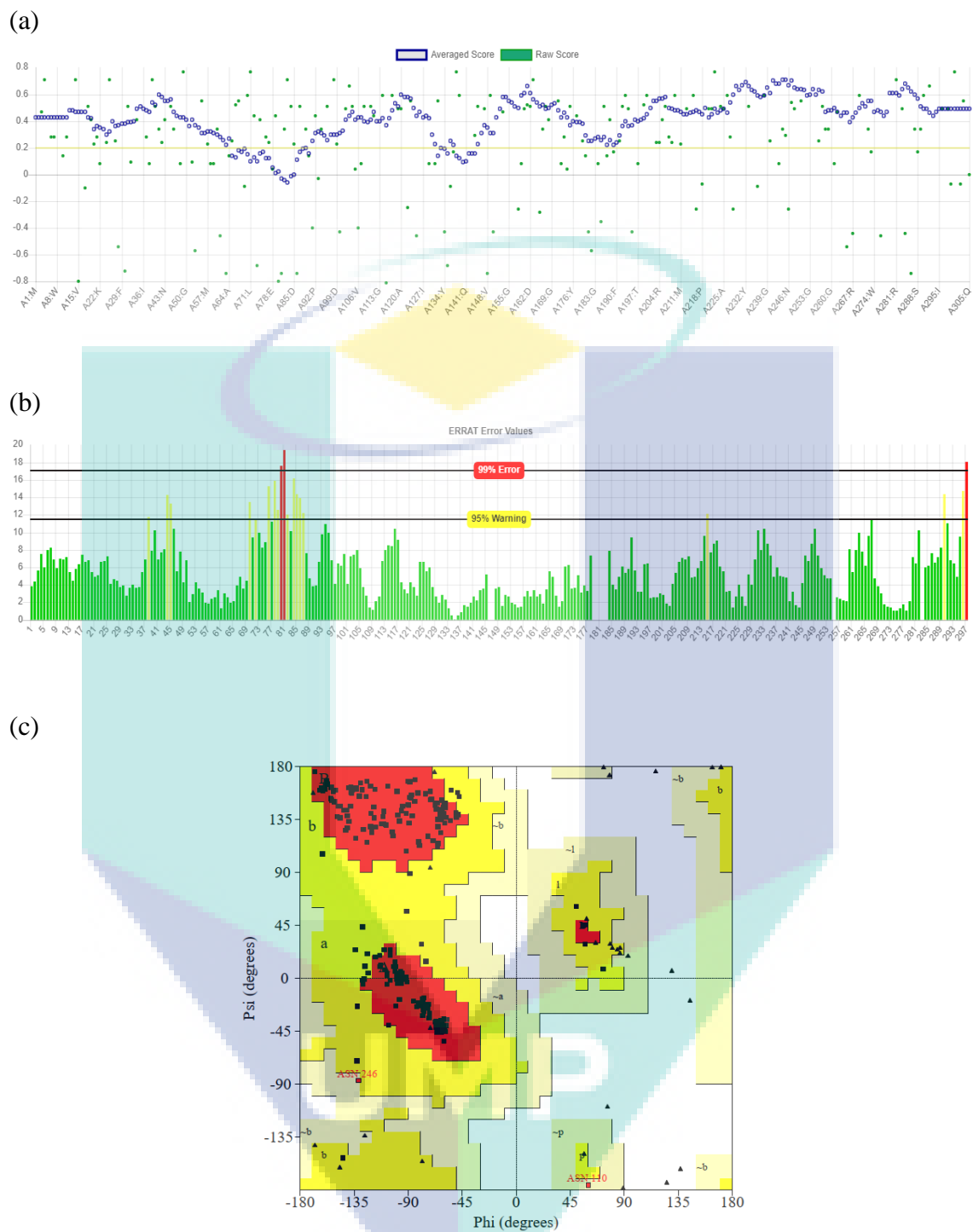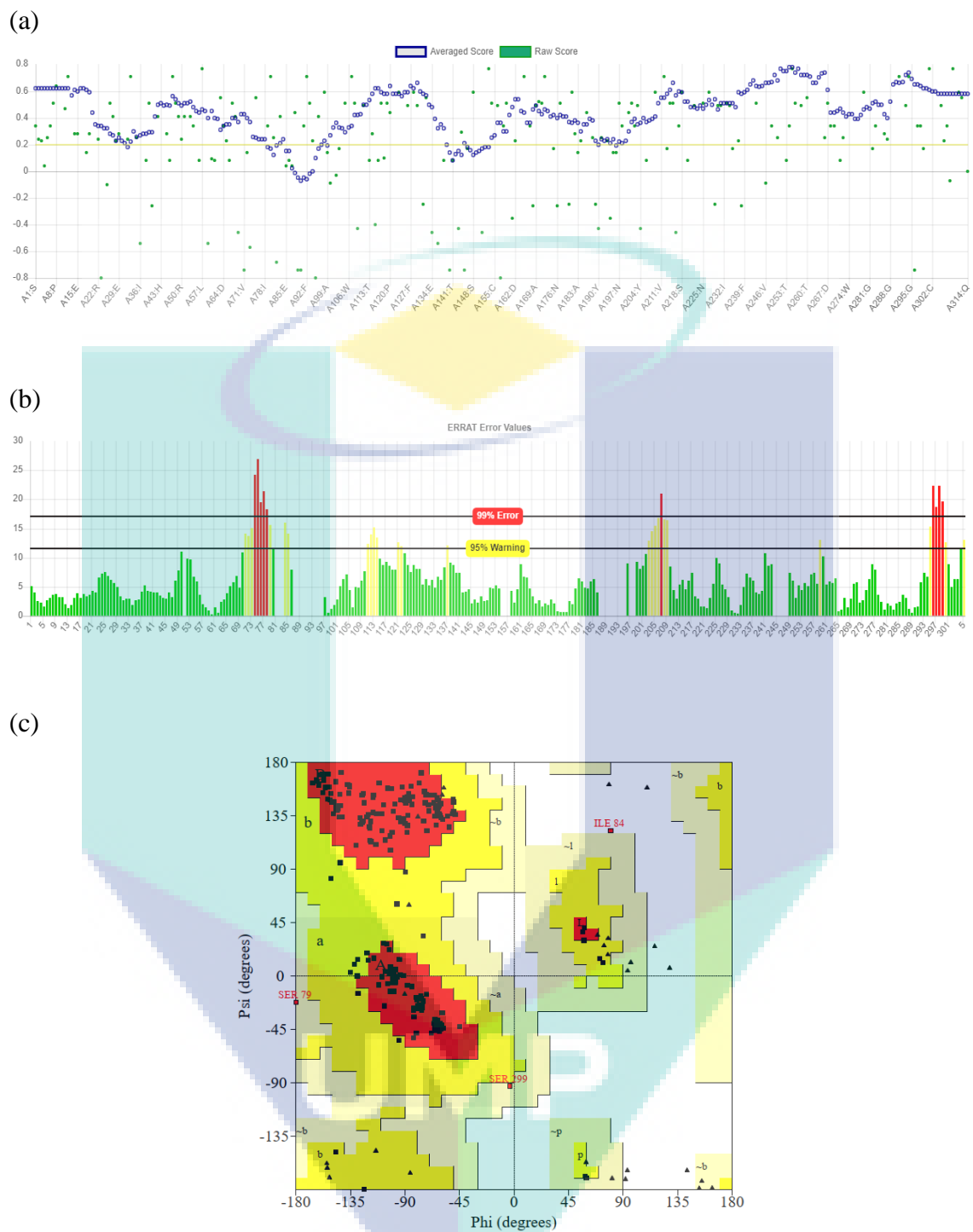
APPENDIX E
OPTIMUM TEMPERATURE AND THERMOSTABILITY ANALYSIS OF FRUIT BROMELAIN

Table E1   Enzymatic activity of the crude fruit bromelain at different temperatures

Temperature (°C)	Enzyme activity (U/mL)
40	2.00 ± 0.06
50	2.11 ± 0.05
60	2.69 ± 0.01
70	2.09 ± 0.10
80	1.67 ± 0.13

Table E2   Enzymatic activity of the crude fruit bromelain after incubated for one hour at different temperatures

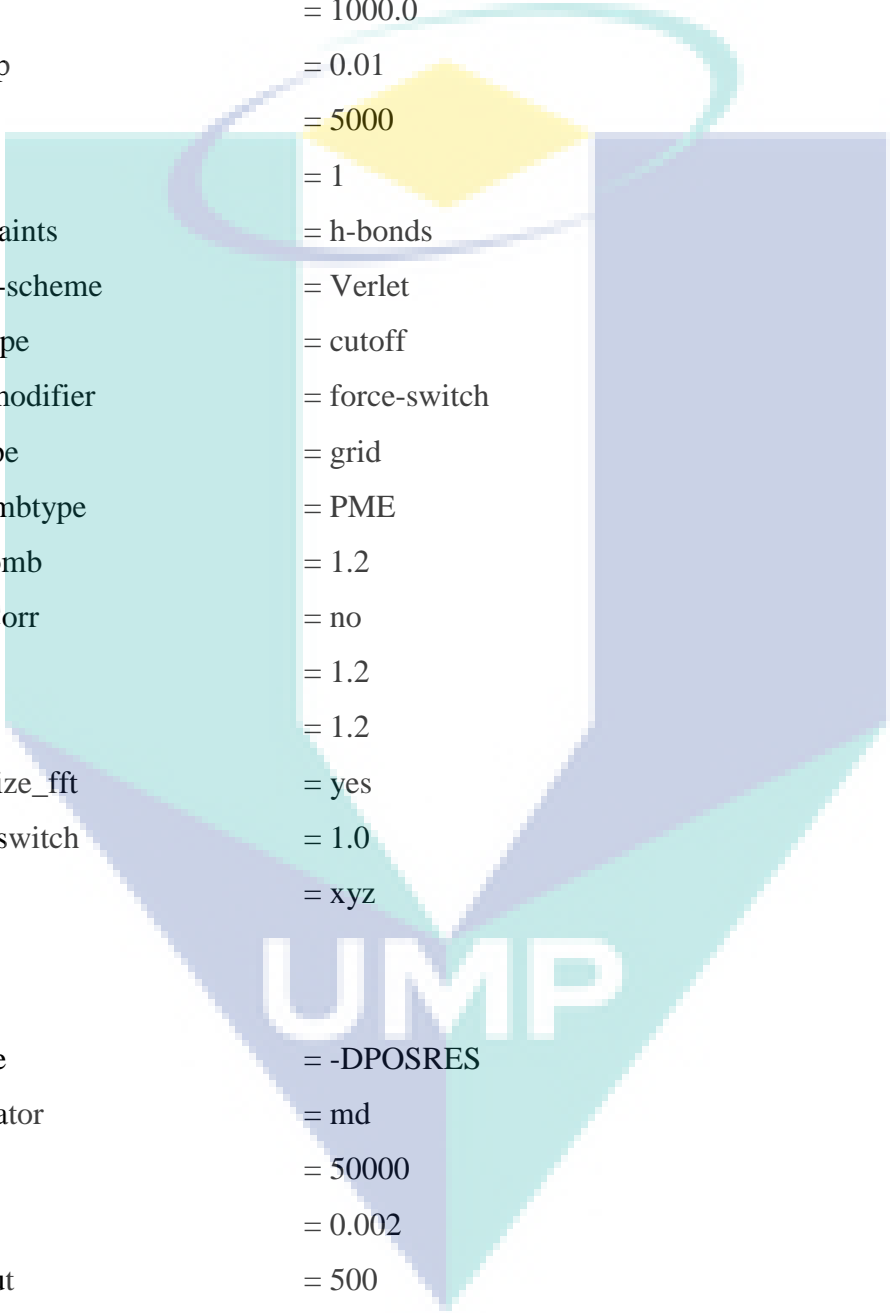| Temperature (°C)	Enzyme activity (U/mL)
40	1.54 ± 0.04
50	1.35 ± 0.10
60	0.54 ± 0.10
70	0.08 ± 0.04
80	0.00 ± 0.00

APPENDIX F
COMMAND LINE USED IN GROMACS

1. gmx pdb2gmx -f FB.pdb -o FB_processed.gro -water spce

2. gmx editconf -f FB_processed.gro -o FB_newbox.gro -c -d 1.0 -bt cubic

3. gmx grompp -f ions.mdp -c FB_solv.gro -p topol.top -o ions.tpr

4. gmx genion -s ions.tpr -o FB_solv_ions.gro -p topol.top -pname NA -nname CL -neutral

5. gmx grompp -f minim.mdp -c FB_solv_ions.gro -p topol.top -o em.tpr

6. gmx mdrun -v -deffnm em

7. gmx grompp -f nvt.mdp -c em.gro -r em.gro -p topol.top -o nvt.tpr

8. gmx mdrun -deffnm nvt

9. gmx grompp -f npt.mdp -c nvt.gro -r nvt.gro -t nvt.cpt -p topol.top -o npt.tpr

10. gmx mdrun -deffnm npt

11. gmx grompp -f md.mdp -c npt.gro -t npt.cpt -p topol.top -o md_0_1.tpr

12. gmx mdrun -deffnm md_0_1

13. gmx trjconv -f md.trr -s md.tpr -pbc mol -ur compact -center -skip 10 -o md_center_skip.xtc

14. gmx rms -f md_center_skip.xtc -s md.tpr -o rmsd.xvg -tu ns

15. gmx rmsf -f md_center_skip.xtc -s md.tpr -o rmsf.xvg -oq rmsf.pdb -res

16. gmx gyrate -f md_center_skip.xtc -s md.tpr -o gyrate.xvg

17. gmx hbond -f md_center_skip.xtc -num hbond.xvg -s md.tpr -tu ns

18. gmx do_dssp -f md_center_skip.xtc -s md.tpr -map ss.map -o ss.xpm -sc ss.xvg -tu ns

19. gmx xpm2ps -f ss.xpm -o ss.eps -bx 0.2 -by 3

20. ps2pdf -sPAPERSIZE=ledger ss.eps ss.pdf

21. gmx sasa -f md_center_skip.xtc -s md.tpr -o totalarea.xvg -or residue.xvg -oa atom.xvg -tv volume.xvg -odg solvation.xvg -q sasa.pdb

22. gmx trjconv -f md_center_skip.xtc -fit rot+trans -s md.tpr -o fix.xtc

23. gmx trjconv -f fix.xtc -s md.tpr -b 40000 -e 50000 -o last10ns-no-skip.xtc

24. gmx cluster -f last10ns-no-skip.xtc -s md.tpr -n index.ndx -o cluster.xpm -minstruct 10 -sz clust-size.xvg -clid -cl -method gromos -dist rms-dist.xvg -cutoff 0.1

APPENDIX G
PARAMETER FILES USED IN GROMACS
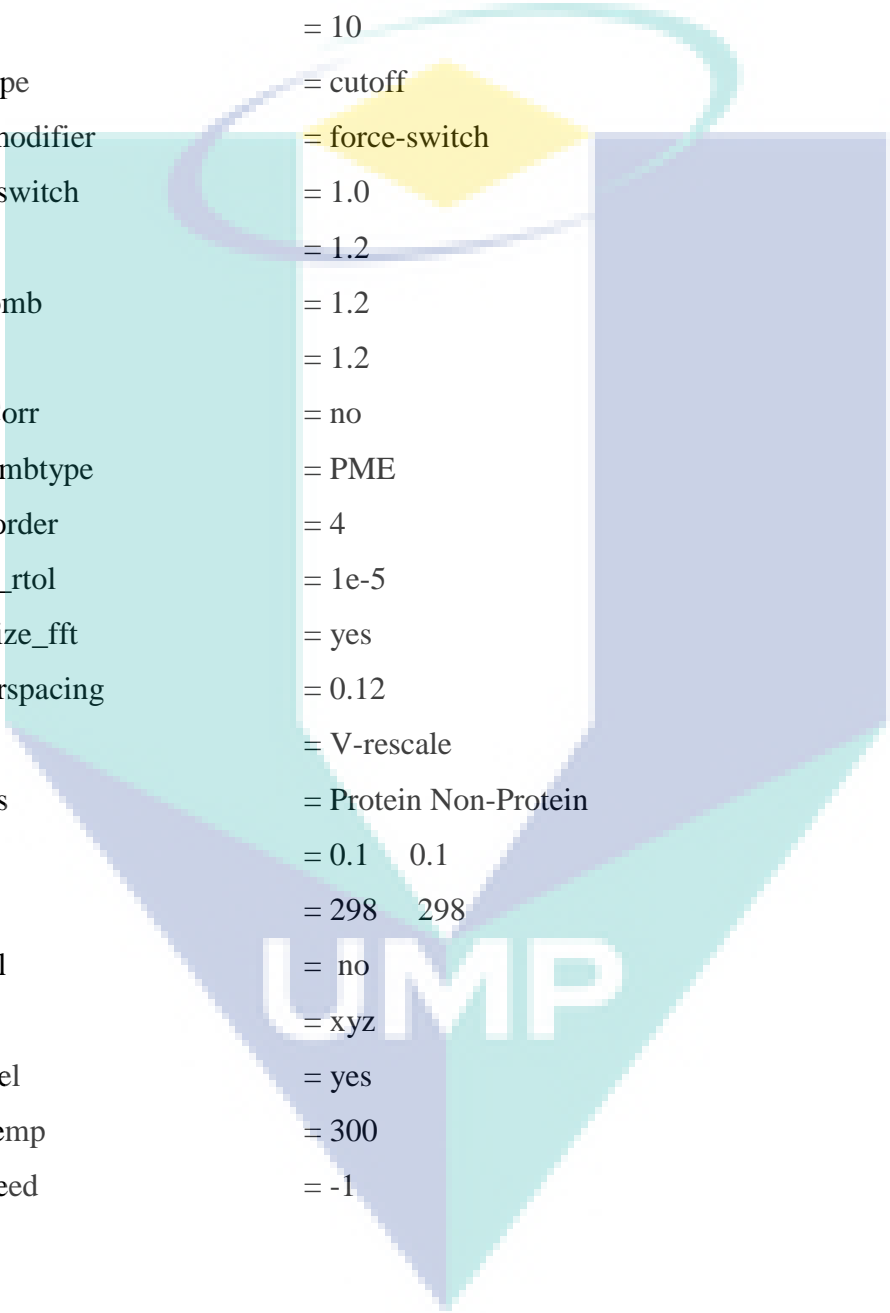
Energy minimisation

integrator	= steep
emtol	= 1000.0
emstep	= 0.01
nsteps	= 5000
nstlist	= 1
constraints	= h-bonds
cutoff-scheme	= Verlet
vdwtype	= cutoff
vdw-modifier	= force-switch
ns_type	= grid
coulombtype	= PME
rcoulomb	= 1.2
DispCorr	= no
rlist	= 1.2
rvdw	= 1.2
optimize_fft	= yes
rvdw-switch	= 1.0
pbc	= xyz

NVT

Define	= -DPOSRES
integrator	= md
nsteps	= 50000
dt	= 0.002
nstxout	= 500
nstvout	= 500
nstenergy	= 500
nstlog	= 500
continuation	= no
constraint_algorithm	= lincs

constraints	= h-bonds
lincs_iter	= 1
lincs_order	= 4
cutoff-scheme	= Verlet
ns_type	= grid
nstlist	= 10
vdwtype	= cutoff
vdw-modifier	= force-switch
rvdw-switch	= 1.0
rlist	= 1.2
rcoulomb	= 1.2
rvdw	= 1.2
DispCorr	= no
Coulombtype	= PME
pme_order	= 4
ewald_rtol	= 1e-5
optimize_fft	= yes
fourierspacing	= 0.12
tcoupl	= V-rescale
tc-grps	= Protein Non-Protein
tau_t	= 0.1 0.1
ref_t	= 298 298
pcoupl	= no
pbc	= xyz
gen_vel	= yes
gen_temp	= 300
gen_seed	= -1

NPT

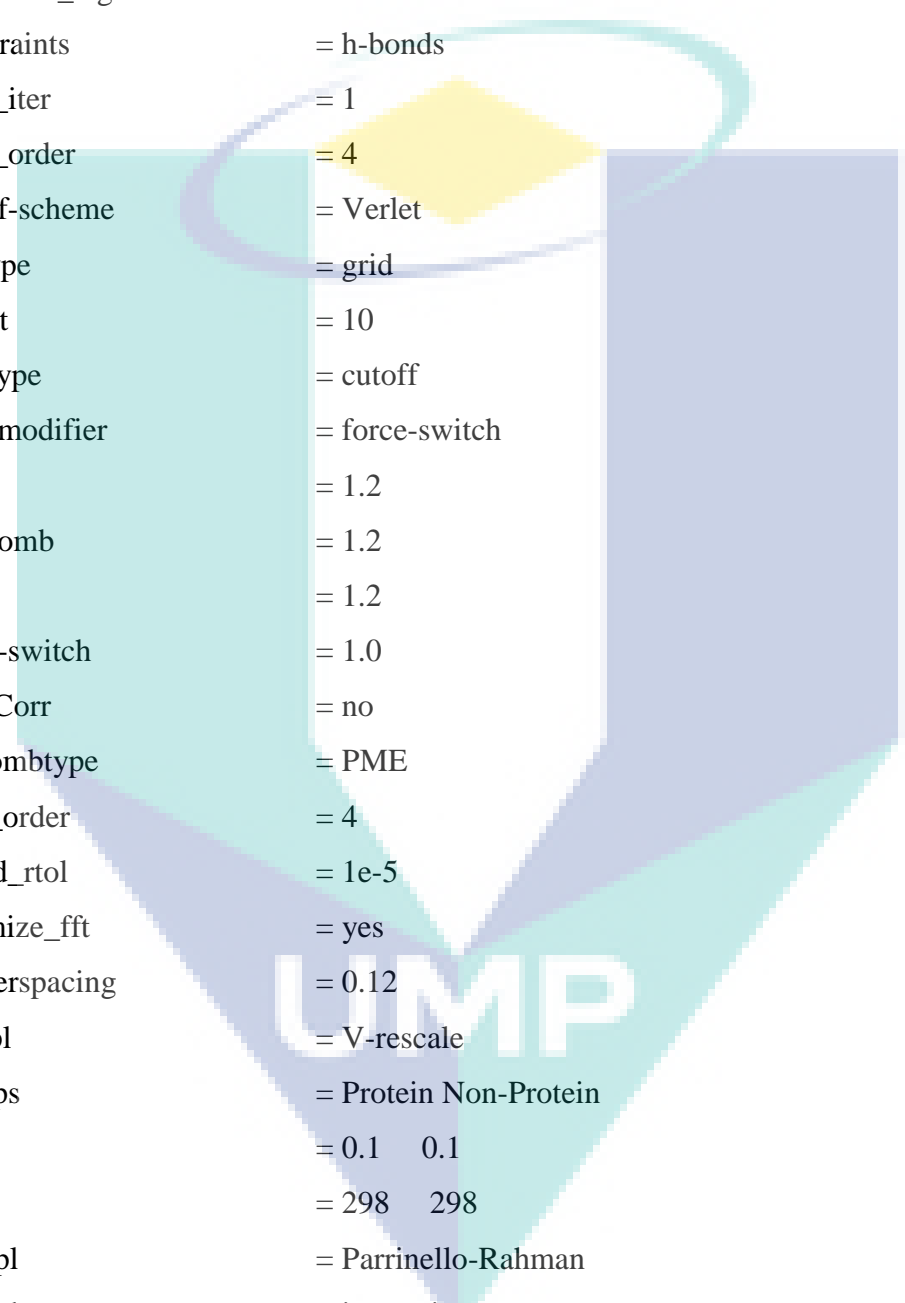|	
define	= -DPOSRES
integrator	= md
nsteps	= 50000
dt	= 0.002
nstxout	= 500

```
nstvout                 = 500
nstenergy               = 500
nstlog                  = 500
continuation            = yes
constraint_algorithm    = lincs
constraints             = h-bonds
lincs_iter              = 1
lincs_order             = 4
cutoff-scheme           = Verlet
ns_type                 = grid
nstlist                 = 10
vdwtype                 = cutoff
vdw-modifier            = force-switch
rlist                   = 1.2
rcoulomb                = 1.2
rvdw                    = 1.2
rvdw-switch             = 1.0
DispCorr                = no
coulombtype             = PME
pme_order               = 4
ewald_rtol              = 1e-5
optimize_fft            = yes
fourierspacing          = 0.12
tcoupl                  = V-rescale
tc-grps                 = Protein Non-Protein
tau_t                   = 0.1    0.1
ref_t                   = 298    298
pcoupl                  = Parrinello-Rahman
pcoupltype              = isotropic
tau_p                   = 2.0
ref_p                   = 1.0
compressibility         = 4.5e-5
refcoord_scaling        = com
pbc                     = xyz
```
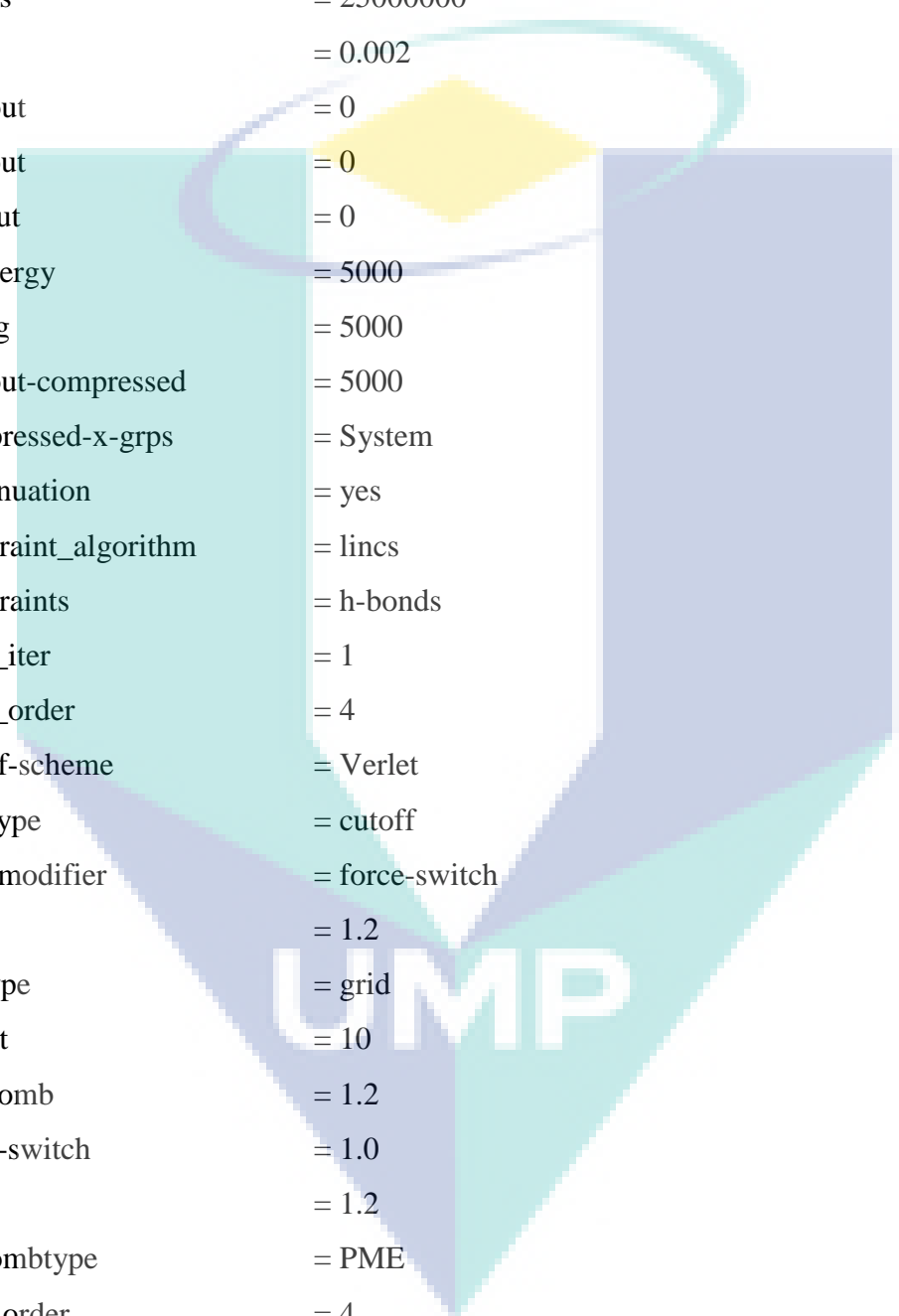
gen_vel   = no

Production run

integrator   = md
nsteps   = 25000000
dt   = 0.002
nstxout   = 0
nstvout   = 0
nstfout   = 0
nstenergy   = 5000
nstlog   = 5000
nstxout-compressed   = 5000
compressed-x-grps   = System
continuation   = yes
constraint_algorithm   = lincs
constraints   = h-bonds
lincs_iter   = 1
lincs_order   = 4
cutoff-scheme   = Verlet
vdwtype   = cutoff
vdw-modifier   = force-switch
rlist   = 1.2
ns_type   = grid
nstlist   = 10
rcoulomb   = 1.2
rvdw-switch   = 1.0
rvdw   = 1.2
coulombtype   = PME
pme_order   = 4
fourierspacing   = 0.12
tcoupl   = V-rescale
tc-grps   = Protein Non-Protein
tau_t   = 0.1   0.1

ref_t	= 298 298
pcoupl	= Parrinello-Rahman
pcoupltype	= isotropic
tau_p	= 2.0
ref_p	= 1.0
compressibility	= 4.5e-5
pbc	= xyz
DispCorr	= no
gen_vel	= no