FUNCTIONAL AND STRUCTURAL STUDIES OF ALGINATE LYASE FROM Persicobacter sp. CCB-QB2

SIM PEI FANG

UNIVERSITI SAINS MALAYSIA

2017

FUNCTIONAL AND STRUCTURAL STUDIES OF ALGINATE LYASE FROM Persicobacter sp. CCB-QB2

by

SIM PEI FANG

Thesis submitted in fulfillment of the requirements
for the degree of
Master of Science

November 2017

ACKNOWLEDGEMENT

I wish to convey my sincere thanks to my supervisor, Dr. Teh Aik Hong, for his detailed and constructive comments. I am very grateful for his guidance, advice and directed me in protein purification and protein crystallization studies for this work. I owe many thanks to Dr. Go Furusawa who gave me guidance during my first steps into molecular work studies and always giving great opinions during the studies. Many thanks to my lab members for the stimulating discussions and companions, and no less way did I learn a lot from them.

I would like to deliver my appreciation to MyMaster program for the financial support. Finally, my special gratitude to my family, without their encouragement and understanding it would have been impossible for me to complete my studies.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATION

A_{280}	absorbance at 280nm
bp	base pair
CBM	carbohydrate-binding module
CD	circular dichroism
C-terminal	carboxy-terminal
DP	degree of polymerisation
IPTG	isopropyl β-D-1-thiogalactopyranoside
kb	kilobase
kDA	kilodalton
L	litre
LB	Luria-Bertani
M	molar
mg	milligram
min	minute
mm	millimetre
mM	millimolar
MW	molecular weight
ng	nanogram
nm	nanometer
N-terminal	amino-terminal
OD	optical density
rpm	revolution per minute
PDB	protein data bank
PCR	polymerase chain reaction

pH	potential hydrogen
pI	isoelectric point
poly-G	polyguluronate
poly-M	polymannuronate
s	second
SDS	sodium dodecyl sulfate
SEC	size exclusion chromatography
sp.	species
V	volt
w/v	weight per volume

KAJIAN TERHADAP FUNGSI DAN STRUKTUR 'ALGINATE LYASE' DARIPADA *Persicobacter* sp. CCB-QB2

ABSTRAK

Gen AlyQ daripada *Persicobacter* sp. CCB-QB2 mengekodkan 'alginate lyase' yang mengandungi dua domain modul pengikat karbohidrat, domain A dan B, di terminal-N domain 'alginate lyase', domain C. 'Alginate lyase' AlyQ tergolong dalam keluarga polisakarida lyase 7 (PL7) di mana domain pertama modul pengikat karbohidrat mempunyai persamaan dengan module pengikat karbohidrat 16 (CBM16) manakala domain kedua pula mempunyai persamaan dengan modul pengikat karbohidrat 32 (CBM32). Kajian-kajian sebelum ini lebih tertumpu kepada pencirian 'alginate lyase' atau modul pengikat karbohydrat secara individu, tetapi jarang pada kajian terhadap pencirian selepas pengabungan 'alginate lyase' dengan modul pengikat karbohydrat, bahkan lebih kurang pada kajian struktur selepas pengabungan kedua-dua domain tersebut. Oleh it, kajian ini memberi tumpuan kepada perbezaan aktiviti enzim 'alginate lyase' semasa kehadiran dan ketiadaan modul pengikat karbohidrat, dan menjelaskan hubungan struktur antara modul pengikat karbohidrat dan 'alginate lyase'. Metodologi projek ini bermula dengan pencarian BLASTp asas diikuti dengan pengklonan dan penulenan protein. Protein terhasil akan dikaji mengunakan kristalografi sinar-X dan pencirian enzim. Dengan kehadiran modul pengikat karbohidrat pertama- CBM16, aktiviti enzim AlyQ adalah lebih cekap dengan nilai k_{cat}/K_m berbanding dengan enzim $AlyQ_{BC}$ dan $AlyQ_C$. Kajian pencirian enzim menunjukkan bahawa $AlyQ_B$ mampu mengikat kepada oligomer 'alginate' terurai di bahagian hujung. AlyQ, $AlyQ_{BC}$ dan $AlyQ_C$ menunjukkan produk degradasi dan aktiviti optimum yang sama pada pH 7.0, namun suhu optimum AlyQ adalah

50°C, manakala $AlyQ_{BC}$ dan $AlyQ_{C}$ adalah pada 40°C. Struktur kristal $AlyQ_{B}$ menggambarkan pertalian antara asid '4-deoxy-alpha-L-erythro-hex-4-enopyranuronic' yang merupakan hasil penguraian oleh 'alginate lyase'. Struktur $AlyQ_{BC}$ menunjukkan bahawa kedua-dua domain B dan C tidak berinteraksi antara satu sama lain. Dengan perbandingan mekanisme 'alginate lyase' lain yang diketahui, menunjukan terdapat tiga asid amino terpulihara di dalam enzim 'alginate lyase' keluarga PL7 iaitu (Gln436, His 438 dan Tyr541), tetapi terdapat dua asid amino lain (Asp446 dan Asp447) yang bercanggah dengan tapak pemangkin enzim. Oleh itu kaedah pemangkinan AlyQ diramal berbeza dengan 'alginate lyase' lain yang telah dikaji setakat ini. Sebagai kesimpulan, domain CBM16 meningkatkan aktiviti enzimatik 'alginate lyase' dan meningkatkan suhu optimumnya. CBM32 boleh mengenal pasti substrat 'alginate' dan 'alginate lyase' keluarga PL-7 mengandungi 2 asid amino tambahan yang mungkin bertanggungjawab dalam aktiviti enzimatik.

FUNCTIONAL AND STRUCTURAL STUDIES OF ALGINATE LYASE FROM *Persicobacter* sp. CCB-QB2

ABSTRACT

AlyQ gene from *Persicobacter* sp. CCB-QB2 encodes alginate lyase which is comprised of two carbohydrate-binding domains, domains A and B, at the N-terminus of the alginate lyase domain, domain C. Alginate lyase domain from AlyQ belongs to the polysaccharide lyase 7 (PL7) family, while the first domain of the carbohydrate-binding module resembles carbohydrate-binding module 16 (CBM 16), and the second domain a CBM 32. Previous studies had mainly focused on activity characterization of alginate lyase or carbohydrate-binding modules individually but rarely on studies of the enzyme characteristics after combining these two domains and even less on the structural studies between alginate lyase and CBM domains. Therefore, this study focused on the alginate lyase enzymatic activity differences with or without the inclusion of the two carbohydrate-binding domains, and to elucidate the structural relationship between carbohydrate-binding domains and alginate lyase. The methodology of this project started with basic BLASTp studies followed by cloning of the targeted domains, and after protein purification, the proteins were subjected to X-ray crystallography studies and enzyme characterization. Comprising the CBM16, AlyQ enzyme had led to a higher k_{cat}/K_m value compared to the truncated versions, $AlyQ_{BC}$ and $AlyQ_C$ enzymes. Substrate binding assay shows that $AlyQ_B$ protein was able to bind the non-reducing end of degraded alginate oligomers. The AlyQ, $AlyQ_{BC}$ and $AlyQ_C$ showed similar degradation product and similar optimum activity at pH 7.0, however the optimum

temperature of AlyQ was 50°C, while $AlyQ_{BC}$ and $AlyQ_{C}$ 40 °C. The crystal structure of $AlyQ_{B}$ illustrates its affinity towards 4-deoxy-alpha-L-erythro-hex-4-enopyranuronic acid, the degraded unsaturated alginate substrate. The domains B and C from the $AlyQ_{BC}$ structure showed that both are independent and do not interact. From the comparison with other known alginate-binding mechanisms, the active sites of alginate lyase domain consist of three conserved catalytic amino acids (Gln436, His 438 and Tyr541), however there are two additional amino acids (Asp446 and Asp447) found to be in proximity with the predicted substrate active site. Hence, AlyQ's mode of binding is expected to be different from that of presently known alginate lyases. In conclusion, having a CBM16 domain enhances the enzymatic activity of alginate lyase and increases its optimal temperature. The second CBM belongs to CBM32 family that binds to degraded alginate substrate, and the alginate lyase domain belongs to PL-7 family and contains two additional amino acids that might be responsible for its enzymatic activity.

1.0 INTRODUCTION

1.1 Background of Research

Alginates are a major constituent in the cell walls of brown algae and exopolysaccharides in marine bacteria (Jagtap *et al.*, 2014; Kim *et al.*, 2011 and Uchimura *et al.*, 2010). Alginates are either made up of homopolymeric blocks of poly α-1,4-L-guluronate(poly-G) and poly β-1,4-D-mannuronate(poly-M), or heteropolymeric blocks (poly-MG) with random arrangements of both monomers. Alginates are depolymerized to oligomeric alginate via enzymatic hydrolysis of alginate lyases. These oligomeric alginates have been shown to possess various biological activities such as enhancing the germination in plants and stimulating production of cytokines (Hu *et al.*, 2004 & Iwamoto *et al.*, 2005).

Alginate lyase enzymes can be found in a wide range of organisms, including algae, marine invertebrates, and marine and terrestrial microorganisms and most of the organisms that produce alginate lyase use depolymerized alginates as carbon and energy sources. (Wong *et al.*, 2000). Alginate lyases exhibit poly-M and poly-G specific activity which can either cleave the alginate polysaccharide internally (endo-type) or cleave at the end of polysaccharide to form monosaccharide (exotype) by a β-elimination mechanism which targets the glycosidic bond (1→4 O linkage) and produce an unsaturated oligosaccharide with a double bond at the non-reducing end (Kim *et al.*, 2011; Wong *et al.*, 2000, Zhu & Yin, 2015).

Carbohydrate-binding modules (CBM) are non-catalytic carbohydrate-binding accessory modules from larger modular enzymes such as xylanase, alginate lyases, and chitinase that are enzymes that breakdown polysaccharides (Abbott *et al.*, 2007).

CBMs have been found in polysaccharide degrading enzymes, and the enzymatic complexes are found to work more efficiently with CBMs, and the enzymatic activity decreases dramatically when the CBMs are removed (Dutta & Wu 2014; Mingardon *et al.*, 2011). However, not many research have been conducted to analyze the role of CBM towards the enzymatic activity of alginate lyase, with only a recently published research by Li *et al.*, 2015 that demonstrates CBM13 indeed aids in alginate lyase enzymatic activity by enhancing thermostability and altering the substrate preferences. Besides that, currently the CBM domain and alginate lyase crystal structures were studied separately, and no crystal structure of CBM and alginate lyase has ever been studied together, hence posing a research gap to elucidate the structural relationship of CBM with alginate lyase.

The AlyQ gene in this experiment containing three domains was obtained from *Persicobacter* sp. CCB-QB2, an agarolytic bacterium isolated from seaweed (genus Ulva) collected from a coastal area of Malaysia (Furusawa *et al.*, 2016). The genome sequence of *Persicobacter* sp. CCB-QB2 could be obtained in DDBJ/EMBL/GeneBank database with the accession number LBGV00000000 (Furusawa *et al.*, 2016). This report aims to express and purify the targeted proteins – full-length protein containing domain A, B and C (AlyQ), domain B and C ($AlyQ_{BC}$), domain A ($AlyQ_A$), domain B ($AlyQ_B$) and domain C ($AlyQ_C$) to elucidate the relationship between carbohydrate-binding domains and alginate lyase by studying the enzymatic activity and the structure of the protein.

1.2 Objectives

1. To clone, express and purify stable targeted proteins - AlyQ, $AlyQ_{BC}$, $AlyQ_{A}$, $AlyQ_{B}$ and $AlyQ_{C.}$
2. To investigate the role of the carbohydrate-binding modules (CBM) in the catalytic function of AlyQ.
3. To study the structure of AlyQ by X-ray crystallography.
4. To elucidate the binding mechanism of carbohydrate-binding modules (CBM) and the alginate lyase domain.

2.0 LITERATURE REVIEW

2.1 *Persicobacter* sp.

Persicobacter sp. is a rod shaped, Gram-negative, exhibiting gliding motility and salmon-pink pigmented bacterium, and also requires seawater or NaCl for growth (Muramatsu *et al.*, 2010, Furusawa *et al.*, 2016). According to Muramatsu *et al.*, (2010) the *Persicobacter* genus belongs to the family '*Flammeovirgaceae*' of the phylum *Bacteroidetes*. *Persicobacter* sp. CCB-QB2 grows aerobically at 30 ℃, and were isolated from seaweed (genus *Ulva*) from the Queens Bay of Penang Island, Malaysia (Furusawa *et al.*, 2016). *Persicobacter* sp. CCB-QB2 DNA sequence analysis showed that it was closely related to *Persicobacter diffluens* and *Persicobacter psychrovividus* which are all known to degrade agar (Muramatsu *et al.*, 2010, Furusawa *et al.*, 2016).

2.2 Alginate

Alginate or alginic acids are linear polysaccharides of β-D mannosyluronic acid and C5 epimer α-L-gulosyluronic acid that are covalently (1-4)-linked (Figure 2.1). Alginate can be arranged in three forms, that is homopolymeric G block, homopolymeric M block and heteropolymeric G/M blocks (Wong *et al.*, 2000). The content of guluronic acid monomer plays an important role in rigidity and flexibility of the alginate (Kim *et al.*, 2011). The ratio of the M and G residues differs depending on the sources of the alginates, with a higher content of guluronic acid in alginates increases the rigidity and gel strength (Swift *et al.*, 2014). Due to alginates ability as gelling and viscosifying agent, it applications are wide across many industries including printing, medical, texture and food industry (Kim *et al.*, 2011; Zhu & Yin 2015).

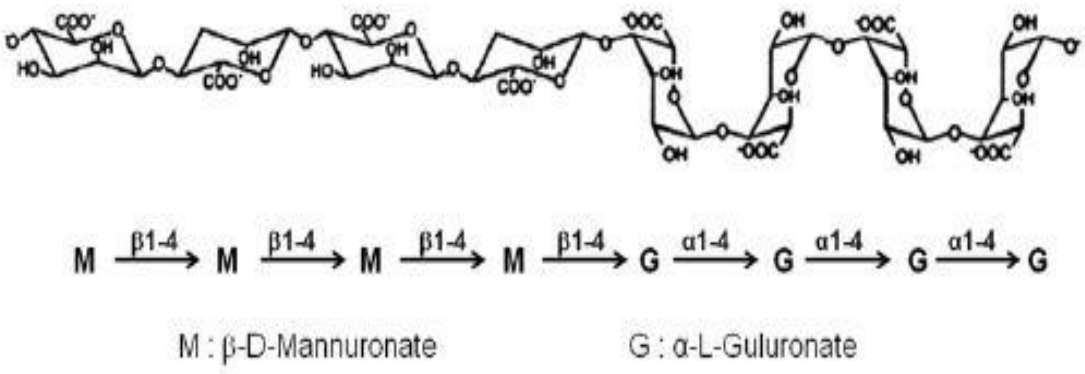

Figure 2.1: Chemical structure of alginates. Alginate is a linear polysaccharide in which β-D-mannuronate (M) and its C5 epimer, α-L-guluronate (G), are covalently (1-4)-linked in different sequences (Kim *et al.*, 2011).

Commercial alginates are mostly derived from brown seaweed with the major seaweed species being from *Macrocystis*, *Laminaria*, and *Ascophyllum* (Kim *et al.*, 2011; Swift *et al.*, 2014). In the food industry, alginates extracted from seaweeds are mainly used as food additives as a thickening or gelling agent, as well as a stabilizer. While in the medical industry, alginate oligosaccharides are shown to stimulate proliferation of human endothelial cells. For example, alginate hydrogels are used in some bandages for treating burns, besides that alginate hydrogel capsules can be used as a drug delivery system through oral route due to alginate's ability to interact with the mucus gastrointestinal tract, hence enhancing the performance of drug by prolonging the time of drug in the gastrointestinal tract. (Sosnik 2014; Lee *et al.*, 2012).

In addition, a research done by Takeda *et al.*, (2011) showed that a genetically modified microorganism that contained both alginate lyases gene and ethanologenic bacteria, did successfully produce ethanol solely from unsaturated uronic acid that

was broken down from alginate carbon source. Alginate can bind divalent cations such as Ca^{2+} and Mg^{2+} to develop gel formation. Alginate gel forming ability is temperature independent and biocompatible hence it is used for many biotechnology applications such as immobilization of cells. However, alginate gel formation is greatly affected by pH, because the pH could change the ionic form of the uronic acid residue (Kim *et al.*, 2011).

2.3 Alginate lyase

Alginate lyases come from a subgroup of polysaccharide lyase family including PL5, PL6, PL7, PL14, PL15, PL17, and PL18 families (Table 2.1). The alginate lyase are divided base on their enzyme specificity towards alginates, mode of degradation and the sources or origins of the enzymes. There are two types of alginate lyase enzyme, one is poly(α-L-guluronate) lyase (EC 4.2.2.11), and the other one is poly(β-D-mannuronate) lyase (EC 4.2.2.3), while most reported alginate lyase preferred poly-M, and fewer poly-G preferred alginate lyase is found (Wong *et al.*, 2000; Zhu & Yin, 2015). Although the alginate lyase is divided poly-M or poly-G specific lyases, they still possess low to moderate ability to degrade their opposite homopolymer, (Zhu & Yin, 2015).

Alginate lyases, also known as alginases or alginate depolymerase, have been isolated from a wide range of organisms, including algae, marine mollusks, and marine and terrestrial microorganisms. Alginate lyases have been extracted from several species of microorganism such as *Pseudoalteromonas sp.* (Chen *et al.*, 2016; Li *et al.*, 2011), *Vibrio sp.* (Badur *et al.*, 2015; Kim *et al.*, 2013; Li *et al.*, 2016), and *Microbulbifer sp.* (Swift *et al.*, 2014). Most of the organisms that produce alginate lyase are able to incorporate M and G units from depolymerized alginates as carbon

and energy sources. Interestingly, some of the alginate lyase producing bacteria also synthesize alginate. The majority of alginate lyases is endolytic that cleave the polymer with an endo-mode of action to form end product of 2-4 degree of polymerization, while only a few alginase possesses an exo-mode of action to cleave unsaturated and saturated monomers from the nonreducing end to form end products with 1 degree of polymerization, most endolytic lyases were group into PL-5 and PL-7 while exolyases were group into PL-15 and PL-17 (Wong *et al.*, 2000). Microorganisms that feed on alginate, utilizes extracellular alginate lyase to degrade alginate then transport into cytoplasm of the cell (Kim *et al.*, 2011).

Besides, alginate lyase may differ in molecular mass, and are divided into three groups: 20-35kDa, ~40kDa and ~60kDa. PL-7 alginate lyase family is the most studied group and the sequence analysis has showed that it has three highly conserve sequences that are (R/E) (S/T/N) EL, Q (I/V) H, YFKAG (V/I) YNQ. The YFKAG (V/I) YNQ is observed in the C-termini of both poly-M lyases and poly-G lyases, these different substrates recognizing lyases propose that this conserve region is an important region that is responsible in maintaining the stable three-dimensional conformation of the lyases (Wong *et al.*, 2000). Alginate lyases with the conserved QVH regions are reportedly specific to poly-M alginates while having the conserve QIH regions are responsible for alginate lyases to degrade poly-G or poly-MG alginates (Zhu & Yin 2015).

Table 2.1 Alginate lyase types based on polysaccharide lyase family (Zhu & Yin 2015)

PL family	Fold family	Common Characteristics
PL-5	$(\alpha/\alpha)_6$ barrel family	Endotype degradation mechanism, Poly-M specific
PL-6	Parallel β-helix family	Endotype degradation mechanism
PL-7	β-jelly roll family	Endotype degradation mechanism, Ability to degrade poly-MG, Poly-G and Poly-M,
PL-14	β-jelly roll family	Endotype degradation mechanism
PL-15	$(\alpha/\alpha)_6$ barrel family	Exotype degradation mechanism
PL-17	open α-barrel and β-sheets (Park *et al.*, 2014)	Exotype degradation mechanism
PL-18	β-jelly roll family	Endolytic degradation mechanism, Ability to degrade poly-MG, Poly-G and Poly-M

2.3.1 Alginate lyase degradation mechanism

Alginate lyase catalyses the degradation of alginate polymers by β-elimination mechanism was first proposed by Gacesa (1987). Firstly, by targeting the glycosidic 1-4 O-linkage between monomers and resulting in a double bond between C4 and C5 carbons of the sugar ring, hence forming 4-deoxy-L-erythro-hex-4-enopyranosyluronic acid as the non- reducing terminal moiety (Alekseeva *et al.*, 2004; Lambard *et al.*, 2010). With several alginate lyase 3-dimentional structures having been elucidated (Dong *et al.*, 2014; Thomas *et al.*, 2013), the catalytics mechanism to depolymerize alginate are shown in three steps (Figure 2.2). Firstly, the carboxyl group is neutralized by the charge of a salt bridge (Histidine or Lysine), next is via general base-catalyzed abstraction of the proton on C5, one residue (Aspartic acid or Cystine) act as the proton abstractor and another residue as the proton donor or one residue act as both proton donor and acceptor. Finally, transfer of an electron from the carboxyl group forms a double bond between C4 and C5, causing the elimination of the 4-O-glycosidic bond.

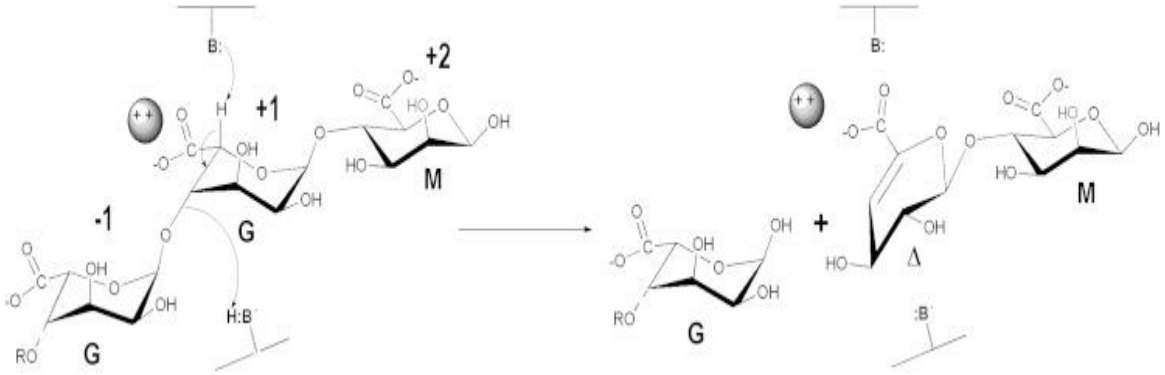

Figure 2.2: Schematic representation of catalytic mechanism of G-specific lyases (Thomas *et al.*, 2013)

2.3.2 Applications of alginate lyase

The ability of alginate lyase to degrade alginate substrate into oligosaccharides that possessed various biological activities has made alginate lyase a crucial enzyme across various disciplines (Kim *et al.*, 2011). In the plantation industry, alginate oligomers degraded by alginate lyase acted as root growth promoters, germination promoters and aided in shoot proliferation for plantation such as lettuce, red amaranth, rice, and peanut seed (Hien *et al.*, 2000; Iwasaki & Matsubara, 2000; Qualhato *et al.*,2013)

In the medical field, alginate as a microcapsule could act as an effective drug delivery system (Cañibano-Hernández *et al.*, 2017). The alginate oligosaccharides also could stimulate secretion of cytokines in human macrophage cells and stimulate the bioactivity of VEGF in human endothelial cells (Iwamoto *et al.*, 2005). Moreover, studies by Cotton, Graham and Lee (2009) showed that alginate lyases had been effective on cystic fibriosis patients due to these patients have mucus building up in the respiratory tract produced by *Pseudomonas aeruginosa*, and alginate lyase's ability to degrade extracellular alginate that was produced by *Pseudomonas aeruginosa*.

In the biofuel industry, alginates consist of up to 40 % dry weight in algae cells could act as feedstocks for bioconversion into biofuels. Alginate lyases could contribute to bioethanol and biofuel because it could convert alginate into unsaturated monosaccharide. The monosaccharide would then spontaneously rearrange into 4-deoxy-L-erythro-5-hexoseulose uronic acid (DEH) then be converted into 2-keto-3-deoxygluconate (KDG) via DEH reductase and further converted to pyruvate and glyceraldehydes-3-phosphate via Entner-Doudoroff (ED) pathway, and then the pyruvate could be converted into biofuels (Wargacki *et al.*, 2012).

2.4 Carbohydrate-binding module (CBM)

Carbohydrate-binding modules (CBM) were first identified as cellulose binding domain (CBD) in cellobiohydrolase from *Trichoderma reesei* (Tomme *et al.*, 1988) and cellulases from *Cellulomonas fimi* (Gilkes *et al.*, 1988). The truncation of CBD domains causes the ability of the catalytic enzyme to hydrolyse cellulose substrate decreases significantly, hence it was proposed that this domain is responsible to target substrate and enhances the enzyme activity. It quickly became apparent that CBDs are not only attached to cellulases but also other types of enzymes that degrade carbohydrates such as xylan, plant cell wall, starch, chitin and mammalian glycans. Hence, carbohydrate-binding modules (CBM) were assigned to represent these proteins that bind to carbohydrate in a non-catalytic way (Boraston *et al.*, 1999).

According to the carbohydrate active enzyme (CAZy) database (Cantarel *et al.*, 2009), to date, carbohydrate-binding modules have been divided into 81 families, and they are a non-catalytic polysaccharide-recognizing module that promotes the

association of enzyme with the substrate. CBMs may play an important role in enzyme targeting by prolonging the association of the catalytic domain to the substrate and are sometimes involved in substrate presentation for catalysis. Currently, due to the expanding diversity of the ligand specificity, the CBM families are now divided by their amino acid sequence similarity, structure, and variation of ligand specificity (Duan *et al.*, 2016). CBMs could range from 30-200 amino acids and exist as a single domain, double domains, or triple domains in one protein. Besides, alginate lyase enzymes has been found to work more efficiently together with presence of CBM 13 and the enzymatic activity decreases dramatically when CBM domain is removed (Li *et al.*, 2015).

Studies have also shown that CBMs could alter the surface properties of fibers by disrupting the tightly packed polysaccharides and reduction of surface polarity of the fiber causing the substrate to go loose and become more exposed to the catalytic enzyme for more efficient degradation (Shoseyov *et al.*, 2006). Besides that, Duan *et al.*, (2016) showed that CBMs have also large ligand specificity because they contain three aromatic amino acid that are responsible for ligand recognition.

CBMs are grouped into three types based on their function and structure (Table 2.2) (Boraston *et al.*, 2004, Gilbert *et al.*, 2013). Type A is a surface binding CBM, its platform like-binding site is described to be complementary to the flat surface of cellulose or chitin crystals. Type A CBM shows non-to little affinity towards soluble carbohydrate. Type B is a glycan chain binding CBM that contains grooves or clefts and several subsites to bind the sugar units of polymeric ligands. Type B CBM shows increase affinities towards substrates with higher degree of polymerization (DP) up to hexasaccharides, while showing decreased affinity towards oligomers with three DP or less. Type C is a small-sugar binding CBM, which shows optimum affinity in

binding mono-, di- and tri-saccharides (Gilbert *et al.*, 2013).

Table 2.2 CBM types (Boraston *et al.*, 2004, Gilbert *et al.*, 2013)

Type	Fold Family	CBM families
A	β-Sandwich, Cysteine knot, Unique, oligonucleotide/oligosaccharide binding folds	1, 2a, 3, 5, 10, 16
B	β-Sandwich	2b, 4, 6, 15, 17, 20, 22, 27, 28, 29, 34, 36
C	β-Sandwich, β-Trefoil, hevein-like fold	9, 13, 14, 18, 32

2.4.1 Applications of Carbohydrate-binding module

Carbohydrate-binding modules have many practical applications in biotechnology due to three main basic properties that is, an independent folding unit, multiple attachment sites, and multiple binding specificities, hence CBM can be easily adapted to address the needs of certain application (Shoseyov *et al.*, 2006). In the bioprocess industry, CBMs have been reported to fuse to targeted protein for usage as affinity purification tags (Kiyohara *et al.*, 2009). This purification technique took advantage of CBM affinity towards cellulose hence enable effective purification of targeted protein at a lower cost (Rodriguez *et al.*, 2004). Furthermore, Haynes *et al.*, (2000) presented a novel protein purification technique that took advantage of CBM4 ability to bind with water-soluble cellulose, separating the targeted protein with two different phases of solution, hence the invention of two-phase liquid separation system. CBM also plays an important role in bioethanol production from lignocellulose, which is needed to increase the catalytic activity of the enzymes to breakdown the cellulosic material (Doi *et al.*, 2004).

In addition, carbohydrate-binding modules are also involved in the bioremediation process that involves bioengineering of CBMs and organophosphorus hydrolase into bacteria cell, CBMs function as immobilization for bacteria to bind to

cellulose, and the active enzymes are used to hydrolyse nerve gas (Wang *et al.*, 2002). Moreover, CBMs had been used as a diagnostic tool to detect pathogenic microbes in food samples, the CBM appended to the bacterium specific monoclonal antibody was able to capture the bacteria cell, hence enabled the rapid detection (Shoseyov *et al.*, 1999). McCartney *et al.*, (2004) introduced CBM as a probe to detect various polysaccharide on plant cell wall by using different types of recombinant CBM.

Besides that, a recent study by Nardi *et al.*, (2015) showed that overexpression of CBM in plant cell wall could alter the structure and composition of the plant cell wall by decreasing the catabolism of pectin, causing a higher amount of pectin in the cell wall. In turn, this higher cell wall integrity by pectin reduces the plant susceptibility to fungal growth known as *Botrytis cinerea*.

2.5 Crystal Structures of Carbohydrate Binding Module and Alginate Lyase

There are up to 81 families in carbohydrate binding modules according to CAZy database, however to date only a handful of CBMs are found to be associated with alginate lyase which are CBM13 (Li *et al.*, 2015), CBM16 (Li *et al.*, 2011) and CBM32 (Badur *et al*. 2015; Thomas *et al.*,2013). According to Fujimoto (2013), CBM13 usually contains 150 amino acid and possesses a β-trefoil fold meaning it contains three repeated homologous sequences, where each of the sequence is composed of four β-strands that fold into a Y-shaped β-hairpin (Figure 2.3A), examples of CBM13 structures include, 1KNL (Notenboom *et al.*, 2002), 1ISX (Fujimoto *et al.*, 2002) and 4OWJ (Kaus *et al.*, 2014).

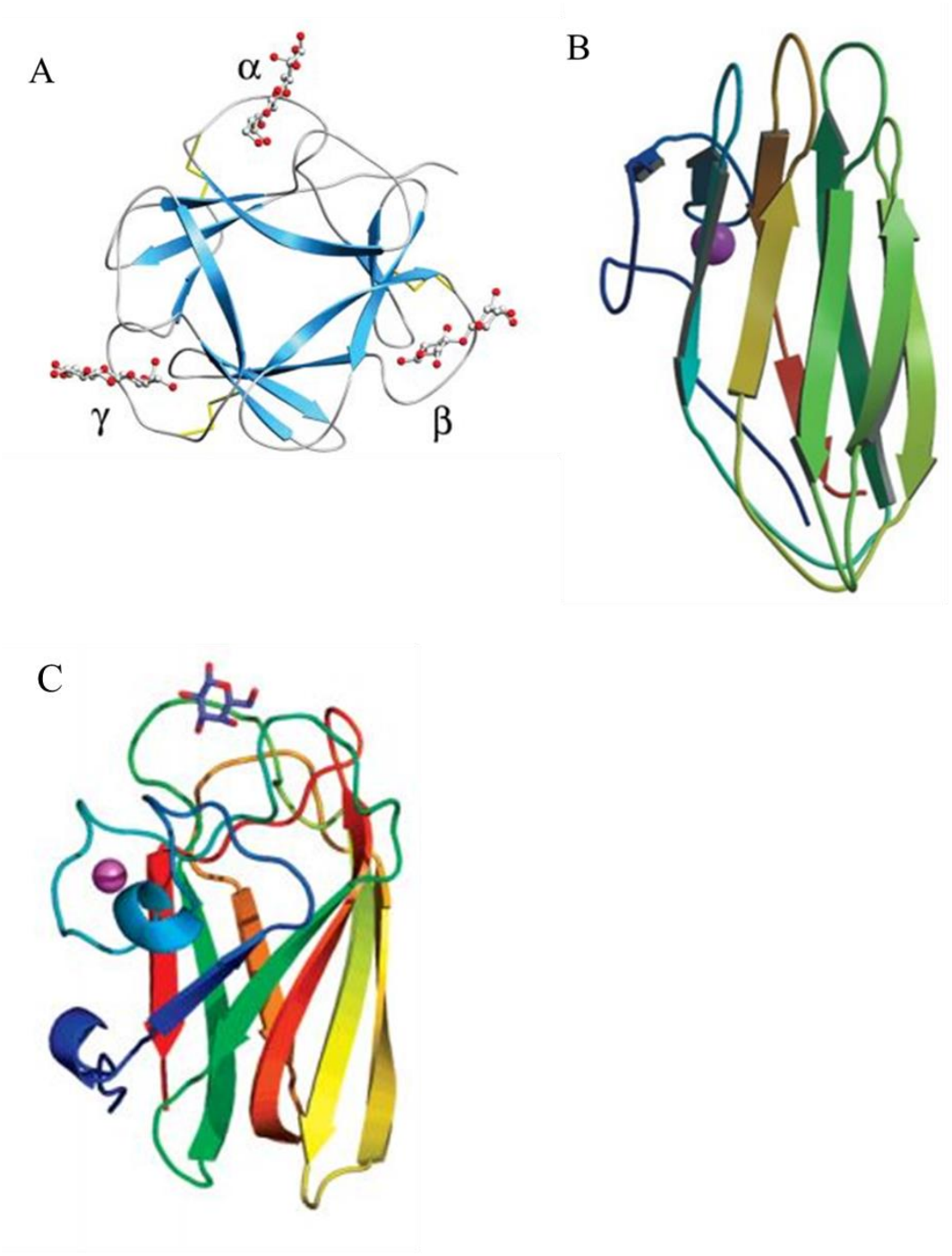

Figure 2.3: Ribbon diagram of CBM structures. A. Ribbon diagram of CBM13 structure with lactose moieties bound in all three subsites of the trefoil fold (PDB code: 1KNL) (Notenboom *et al*., 2002). B. Ribbon diagram of CBM16-1 structure (PDB code: 2ZEW), the purple sphere represents the calcium ion. (Bae *et al.*, 2008). C. Ribbon diagram of CBM32 structure (PDB code: 2V72) with the violet sphere representing calcium ion and stick figure as a bound galactose. (Boraston *et al.*, 2007).

To date only two CBM16 structure were define, CBM16-1 (PDB code: 2ZEW) and CBM16-2 (PDB code: 2ZEZ) from *Caldanaerobius polysaccharolyticus* by Bae *et al.,* (2008). Su *et al.,* (2010) reported the mutational studies of CBM16-1 derived from Bae *at al.,* (2008) that could bind to cellopentaose PDB code: 3OEA) and mannopentaose (PDB code: 3OEB). Both CBM16-1 and CBM-2 proteins fold into a compact domain of β-sandwich that contains two layers of five antiparallel β-sheets, besides that it also contains a calcium ion that bestow stability towards the structure (Figure 2.3B) (Bae *et al.,* 2008). In the mutational studies by Su *et al.,* (2010) demonstrates that two tryptophan amino acids as a critical player to the ligand binding via hydrophobic stacking interaction.

CBM32 is one of the most diverse and valuable model families of CBMs, due to its variability of binding specificity, which extends to both plant cell wall polysaccharides and eukaryotic glycans (Grondin *et al.,* 2014). The first crystal structure of CBM32 was found in the N-terminus of a galactose oxidase from *F. graminearum* (Ito *et al.,* 1991). CBM32 mainly binds to galactose and its various isomers such as CBM32 from *Yersinia enterolitica* was found to bind to polygalacturonic acid (Abbott *et al.* 2007), and CBM32 from *Clostridium perfringens* was found to bind beta-D-glcNAc-beta(1,3) galNAc (Ficko-Blean & Boraston, 2009). CBM32 structure is a β-sandwich fold and consists of a metal ion usually a calcium ion presumeably as a structural component (Figure 2.3C) (Boraston *et al.,* 2007).

Alginate lyases are divided into seven groups, PL-5, 6, 7, 14, 15, 17 and 18. Where according to Carbohydrate Active Enzymes database (http://www.cazy.org/) to date PL-7 alginate lyase is the most studied alginate lyase comprising of six

crystal structures, while PL-5, 14 and 18 each consists of two crystal structures and PL-6, PL-15 and 17 each having one defined structure. The overall structure of PL-5 is made up of alpha helixes and has a deep tunnel-like cleft in a novel (a/a)-barrel structure as the catalytic active domain (Figure 2.4A). The only crystal structure of PL-6 was recently studied by Xu *et al*., (2017), it formed an asymmetric unit of four monomeric molecules, with only two of the four have interactions. The monomeric PL-6 enzyme was built by parallel β-helixes fold forming a 'twin tower like' structure with a calcium ion bound to it (Figure 2.4B). The PL-7 structure was a β-sandwich fold like structure with a large active cleft covered by two short flexible loops (Figure 2.4C). The PL-14 overall structure was two antiparallel β-sheets with a deep cleft showing a β-jelly roll fold (Figure 2.4D). The PL-15 overall structure is consisted of N-terminal small β-sheet and $(\alpha/\alpha)_6$ barrel and a C-terminal anti-parallel β-sheet fold (Figure 2.4E). The overall structure of PL-17 is a combination of $(\alpha/\alpha)_6$ barrel and β-sandwich fold containing a zinc ion (Figure 2.4F). Both the PL-15 and PL-17 possess similar structures and exolytic degradation mechanism hence correlation between the structure similarity and the degradation mechanism is significant. Lastly, the overall structure of PL-18 displays a β-jelly roll scaffold composed mainly of 2 anti-parallels β-sheets, and a calcium ion chelating in the structure (Figure 2.4G).

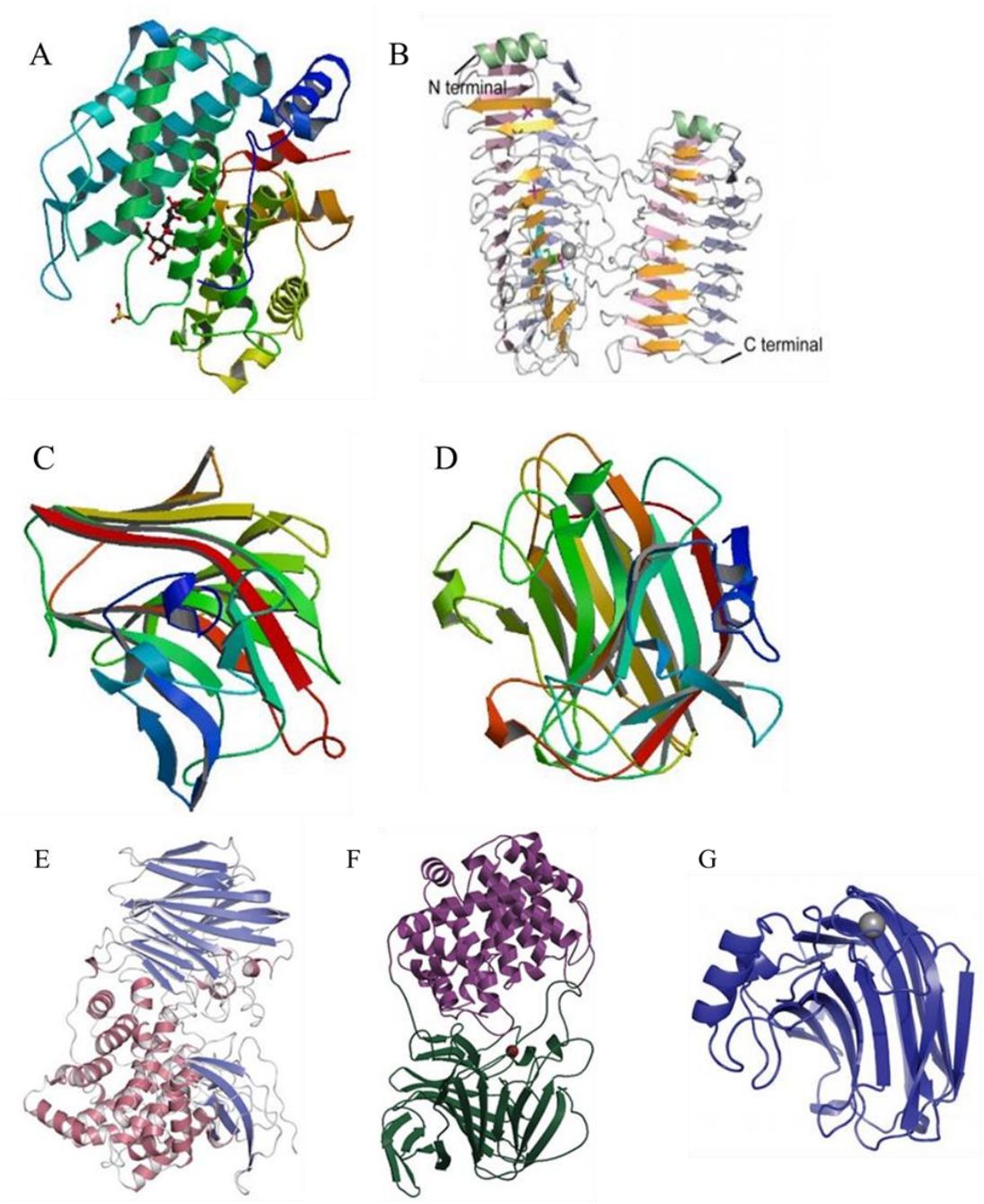

Figure 2.4: Ribbon diagrams of alginate lyases. A. PL-5 structure, black stick figure represents the ligand product and yellow stick figure represents sulfate ion (PDB code: 1HV6) (Yoon *et al.*, 2001) B. PL-6 monomeric structure, grey sphere represents calcium ion (PDB code: 5GKD) (Xu *et al.*, 2017). C. PL-7 structure (PDB code: 1UAI) (Osawa *et al.*, 2005). D. PL-14 structure (PDB code: 5GMT) (Qin *et al.*, 2017). E. PL-15 structure (PDB code: 3A0O) (Ochiai *et al.*, 2010). F. PL-17 structure, the red sphere represents a zinc ion (PDB code: 4NEI) (Park *et al.*, 2014). G. PL-18 structure, with grey sphere representing calcium ion (PDB code: 4Q8K) (Dong *et al.*, 2014)

3.0 MATERIALS AND METHODS

3.1 Overview of Methodology

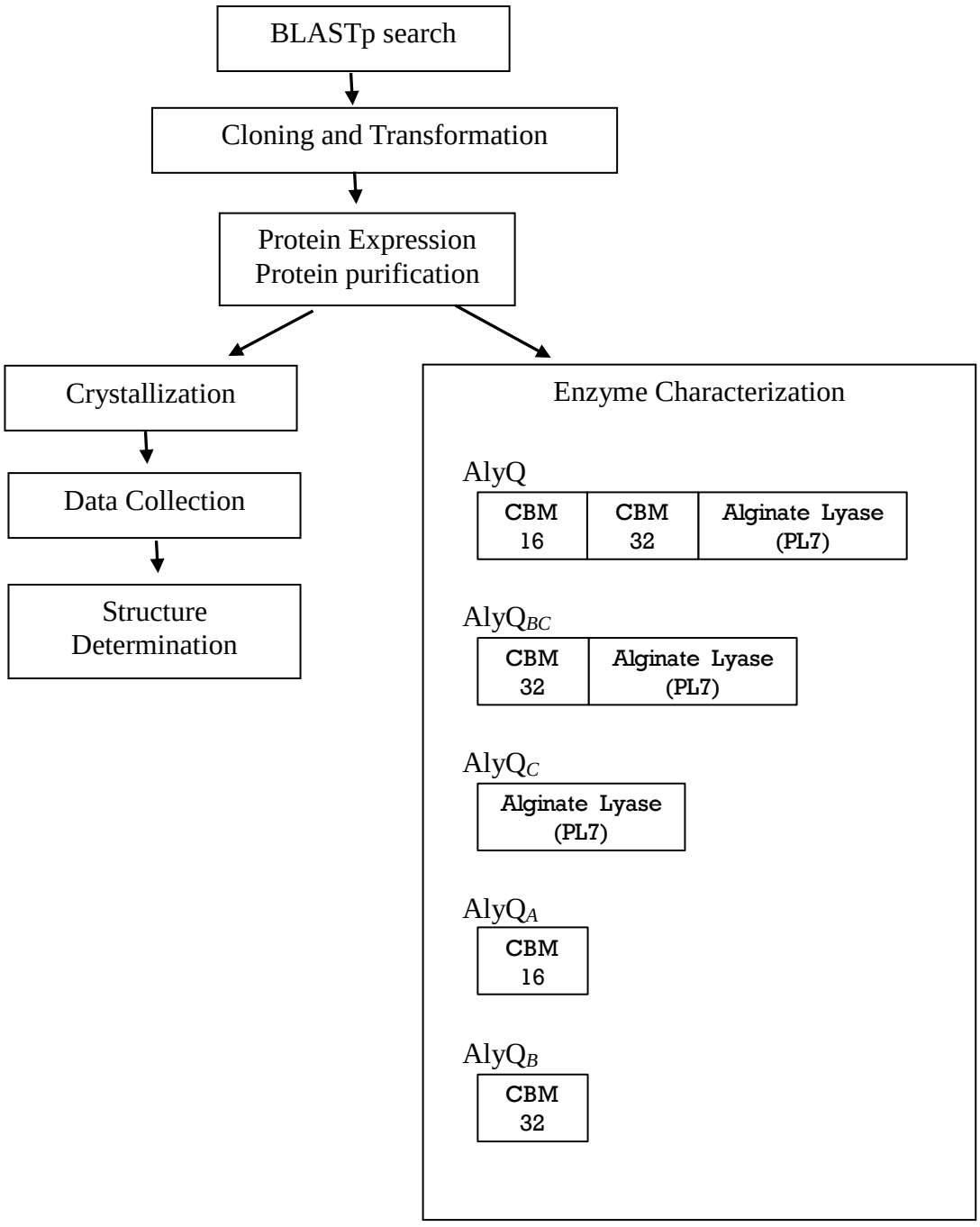

Figure 3.1: Flowchart of the methodology.

3.2 BLASTp Search, Cloning and Transformation

3.2.1 BLASTp Search

The AlyQ gene sequence (Appendix 1) were translated and subjected to NCBI BLASTp search program (https://blast.ncbi.nlm.nih.gov/Blast.cgi). The gene sequence was then divided into three domains and cloned into five different protein by truncating the full-length AlyQ to $AlyQ_{BC}$, $AlyQ_{C}$, $AlyQ_{A}$ and $AlyQ_{B}$. The detail of the gene sequences could be found from Appendix 1 to Appendix 5.

3.2.2 Polymerase Chain Reaction cloning and PCR purification.

The genomic DNA of *Persicobacter* sp. CCB-QB2 was used as a template in PCR for obtaining the alginate lyase gene. High-fidelity PCR was performed using KAPA HiFi HotStart ReadyMix PCR Kit (Kapa Biosystems). The primers, template and KAPA HiFi HotStart Ready Mix were added together then followed by the PCR cycling protocol suggested from technical data sheet from KAPA HiFi HotStart Ready Mix. The AlyQ, $AlyQ_{BC}$, $AlyQ_{A}$, $AlyQ_{B}$ and $AlyQ_{C}$. genes were added with extra six Histidine site at the C-terminal to enable nickel beads purification. The PCR cycling protocols are summarise in (Table 3.1).

Table 3.1 PCR Cycling Protocol

Step	Temperature	Duration	Cycles
Initial Denaturation	95°C	3 min	1
Denaturation	98°C	20 sec	30
Annealing	AlyQ :65°C $AlyQ_{BC}$:67.7°C $AlyQ_{A}$:65.5°C	15 sec	

	$AlyQ_B$:68.9°C $AlyQ_C$:66.7°C		
Extension	72°C	1 min/kb	
Final Extension	72°C	1 min/kb	1

QIAquick PCR Purification Kit (Qiagen, Germany) was used to purify PCR DNA fragments. The PCR product was mixed with five volumes of PB buffer, applied to a QIAquick spin column, and centrifuged for 1 minute at 13,000 rpm. The flow-through was discarded and the bound DNA was washed with 750 µl of PE Buffer by centrifugation. After the flow-through was discarded, the column was centrifuged for an additional 1 minute to remove the residual ethanol. The QIAquick column was placed in a clean micro centrifuge tube. To elute the DNA, 30 µl of EB buffer was added to the centre of the QIAquick membrane, allowed to stand for 1-2 minutes, then centrifuged for 1 minute, this step was repeated at least three times using the same eluted EB buffer to maximize the concentration of DNA product.

3.2.3 Restriction Enzyme Digestion

The PCR product contain NdeI restriction site at the N-terminal, and HindIII restriction site at the C-terminal. Hence, these PCR products were digested with both restriction enzymes to facilitate the cloning of genes into vector. The restriction enzyme and the enzyme buffer were purchased from Fermentas, USA. The preparation of the digestion reaction was set up and incubated in 37°C incubator for 1 hour for PCR product and 5 minutes for plasmid DNA. The plasmid DNA used was pColdIII and pET21A (Appendix 6 and 7). Next, the digested product was incubated in 80°C for 20 minutes to inactivate the enzyme. AlyQ and $AlyQ_A$ were

cloned in pET21A vector while AlyQ$_{BC}$, AlyQ$_{C}$ and AlyQ$_{B}$ were cloned in pColdIII vector due to protein solubility issues.

3.2.4 Extraction of DNA from Agarose Gels

The QIAquick gel extraction kit (Qiagen, Germany) was used to purify DNA from agarose gel. The digested DNA was run on an agarose gel and the desired bands of DNA were cut out and placed in a 1.5 ml tube. About 600 µl of QG Buffer was added to the sliced agarose gel that weight around 0.1 g in the Eppendorf tube and heated at 50°C, shaken at 500 rpm using thermomixer for 10 minutes. After the gel slice had completely dissolved, 1 gel volume of isopropanol was added to sample. To bind the DNA, the sample was applied to a QIAquick spin column and centrifuged at 13,000 rpm for 1 minute. Next for the washing step, the flow-through was discarded and 750 µl of PE buffer was added to the column and centrifuged at 13,000 rpm for 1 minute. Then to dry out the column, the QIAquick spin column was spin for an additional 1 minute. To elute the DNA, the column was applied to a new microcentrifuge tube and 30µl of TE buffer was added to the membrane filter and allowed to stand for 1 minute, then centrifuged for 1 minute.

3.2.5 Ligation of the fragment of PCR and expression vectors

The DNA purified by QIAquick gel extraction kit (Qiagen, Germany) was ligated overnight at 16°C using ligation reaction setup. The T4 ligase and its buffer were purchased from Promega, USA by using 1:3 molar ratio of vector: insert DNA when cloning.

3.2.6 Transformation of BL21 DE3 and DH5α Competent cells and plasmid isolation

DE3 and DH5α *E. coli* cells were used during bacterial transformation process for routine cloning. DH5α cells are used for transformation of ligation mixture before isolating the successful clones to be transformed to BL21 DE3 cells. This is because DH5α can transform efficiently by reducing homologous recombination and intracellular endonuclease activity, hence increase the stability of plasmid to store in the cell. While BL21 DE3 is used for protein expression. The competent cell was taken incubated on ice for 10 minutes after taken out from -80°C freezer, and then added with 100 µg of ligation mixture or DNA plasmid. The cells were further incubated on ice for 30 minutes before heat shocked at 42°C for 1 minute. Then it the heat shocked cells were incubated on ice for 2 minutes then added with 200µl of LB broth and incubated at 37°C for 1 hour with 200 rpm shaking. Then 100 µl of the sample was spread onto a LB agar plate containing ampicilin. The plates were incubated at 37°C overnight.

Plasmids were isolated using the QIAprep Spin Miniprep Kit (Qiagen). Single colonies were inoculated into 2ml of LB broth with 100 µg/ml of ampicilin and grown overnight at 37°C with 200 rpm shaking. The cells were harvested by centrifugation, cell pellets were resuspended with 250 µl of P1 Buffer, then 250 µl P2 Buffer was added to the mixture and gently inverted 7-8 times to mix. After 2 minutes of room temperature incubation, 350 µl of N3 Buffer added and mixed by inverting the tubes. The tubes were incubated for 10 minutes on ice, and then centrifuged at 13,000 rpm for 10 minutes at 4°C to remove the precipitated cell debris, proteins, and genomic DNA. The supernatant was transfer to a QIAprep spin column (Qiagen, Germany) and centrifuged at 13,000 rpm for 1 minute, the flow

through was discarded. The column was then washed by adding 750 µl of PE Buffer and centrifuged for 1 minute. The column undergoes another set of centrifuging for 1 minute to dry the column. The plasmid DNA was eluted with 30 µl of TE buffer of EB Buffer. The construct was analysed by restriction enzyme digestion and DNA sequencing.

3.3 Storage of Strains

DH5α strains harbouring the plasmids encoding the correct fusion protein were maintained as glycerol stock. A single colony was inoculated into 2 ml LB broth containing ampicillin and incubated with shaking at 37°C until the OD_{600} reached 0.6-0.8 and 1 ml was removed from the culture and transferred to a cryovial. 0.5 ml of 50% glycerol was added, mixed well, and stored at -80°C freezer.

3.4 Protein Expression and Purification

3.4.1 Protein Expression

Solubility of protein and molecular weight estimation were analysed in small-scale expression. 50 ml LB medium with ampicillin were inoculated with 500 µl of an overnight preculture. Cells grew for 2 hours at 37°C with 200 rpm agitation until the OD_{600} reached 0.4-0.6. 1 ml of culture was removed and labelled as uninduced. The remaining culture was induced with IPTG with a final concentration of 0.1 mM-1 mM of IPTG to test the protein solubility. The culture was then incubated at 15°C overnight.

After stabilizing the protein expression and solubility, the clones were ready for large scale expression. 1L of LB broth containing 100 µg/ml of ampicilin and

inoculated with 10 ml of an overnight culture. The culture was grown at 37 °C until the OD_{600} reached 0.5-0.6 then incubated at 15 °C for 30 minutes before adding isopropyl-β-D-thiogalactopyranoside (IPTG). For AlyQ and $AlyQ_{BC}$, the final IPTG concentration is 1 mM, $AlyQ_A$ and $AlyQ_B$ IPTG concentration is 0.2 mM, while for $AlyQ_C$ the final IPTG concentration is 0.1 mM. The induced culture was the incubated in 15 °C with agitation of 200 rpm overnight for AlyQ, $AlyQ_{BC}$, $AlyQ_A$ and $AlyQ_B$ culture. $AlyQ_C$ was cultured in 15 °C with agitation of 100 rpm for 40 hours.

3.4.2 Protein Extraction by Sonication

The induced culture was centrifuged at 8000 rpm for 20 minutes to collect the cell pellets. The cell pellets were then washed with phosphate buffered saline (PBS) to wash the excess LB broth in the cell. Supernatant was discarded after centrifugation. Next the cell pellets were resuspended in 50 ml of buffer A (100 mM NaCl, 50 mM Tris, pH 8.0) and lysed by sonication. The samples were sonicated in ice bucket to prevent denaturation of protein, with five sets of 1 second on 1 second off for 1minute and amplitude of 35 %. In between the sets, 1 minute resting period was provided to prevent extensive frothing of the sample. After sonication, the samples undergo centrifugation at 14,000 rpm for 15 minutes at 4 °C. The supernatant would contain the soluble protein.

3.4.3 Protein Purification

The AKTA explorer (GE healthcare, USA) was used to perform all the purification techniques (Affinity chromatography and Gel filtration). The column (HiTrap IMAC HP 5 ml, GE healthcare, USA) was prepared and charged by washing