Novel Fusion Proteins Based on Recombinant Lectins for Delivery of a Cytotoxic Peptide Specifically to Cancer Cells

Soran Maruf Mohammed

School of Environment & Life Sciences

Biomedical Research Centre

University of Salford, UK

Submitted in partial fulfilment of the requirements of the degree of

Doctor of Philosophy

2016

Abstract

One of the most common post-translational modifications is covalent attachment of carbohydrates to proteins. At the cell surface, molecular recognition events involved in cancer metastasis are mediated by sugar moieties of glycoproteins. Evidence from both patient (histochemical analysis) and experimental work (tumour models) show that the metastatic tumour phenotype is associated with altered sialylation of tumour cell surfaces. Structural domains that recognize and bind specific carbohydrates without altering the recognized sugars are usually called lectins. Mistletoe lectin isoform I (MLI), from *Viscum album*, is an example which has strong binding affinity for sialic acids especially α2,6 sialic acid.

In the present study, mistletoe lectin isoform I chain B (MLB) was targeted through conducting polymerase chain reaction (PCR) for cDNA which was reverse transcribed from the extracted *V. album* RNA. The targeted sequence was successfully amplified, cloned and sequenced. The sequence data showed a clear indication that the insert is our desired gene product allowing us to go on to design a fusion protein. Both *N*- and *C*-terminus fusion strategies were used to assemble constructs of the MLB chain fused to GFP in the pGFPuv expression vector which was transformed into BL21 *E. coli* and the expression of the target gene/fusion was successfully carried out. The same mechanism was used to generate positive and negative control fusion proteins from elderberry *S. nigra* lectin and snowdrop *G. nivalis* lectin respectively. The expressed fusions proteins were purified through affinity chromatography using CNBR-activated Sepharose 4B column coupled with degalactosylated fetuin so as to confirm the sialic acid binding activity of MLB. The binding of MLB+GFP fusion proteins to the membrane glycans of human cancer cells was also confirmed through the treatment of human metastatic melanoma cells with the fusion protein and the fusions were detected to actively attach and cross the cell

membrane. Following the confirmation of MLB binding to sialic acid residues on fetuin and cancer cell membrane in which a novel attempt, the MLB sequence was further analysed by bioinformatics to identify the glycan binding domain and two main putative sugar binding sites were identified. Then a range of Lectin-Toxin fusions were designed from the whole MLB and the two glycan binding sites linked with wasp venom (Mastoparan) and honey bee venom (Melittin) through a cathepsin-B biodegradable spacer.

Finally, the Lectin-Toxin fusions were screened against a range of cancer cell lines specifically WM-115 and WM 266-4 melanoma lines and a healthy melanocyte cell line (HEMn) in order to confirm the fusions binding and crossing the cell membrane of the tissues and also confirm their apoptotic effect. The fusion proteins were confirmed to have almost no cytotoxic effect on the healthy human melanocytes (HEMn). However, strong cytotoxic effect was observed during the treatment of human primary (WM-115) and metastatic (WM 266-4) melanoma cells with the fusion proteins.

List of Content

List of Figures

List of Tables

Acknowledgement

I would like to express my sincere gratitude to my supervisor Dr Natalie Ferry for her patience and guidance throughout this project. It was a priceless experience working with her. Her open door, continual support and advice over the last four years were instrumental to the quality of this thesis. I am also grateful to my co-supervisor Dr Rhoderick Elder for his support, encouragement and for giving me his Man City season ticket while he was away.

I would like to acknowledge Professor Angharad MR Gatehouse and Dr Martin Edwards for their contributions to this work by providing a DNA sample of GNA lectin and assisting with the fast protein liquid chromatography (FPLC) in Newcastle University.

I am thankful to Kurdistan Regional Government (KRG) and the Innovation Fellowship Programme of Salford University for their financial support for this project.

I would also like to thank all the technicians in Cockcroft building especially Mark Parlby for being a generous man, Dr Tony Bodell for being a good friend (I promise I will send you the private jet when I become the Prime Minister of Kurdistan).

My special and sincere thanks go to my awesome friends Dr Ryan Joynson for the nice moments we shared together in lab 312 and for being my salsa partner, you know you are one of my best friends even though your moustache is very sparse and it never grows, Andrew McGown for being the best flatmate I have ever met, a special friend and a brother, I always count on you even though you are terrified from spiders, Dr Patrick Killoran for being my Irish twin and the source of my magical energy, I am grateful to have you as my friend even though your name means wild man in Kurdish language, Mustafa Albelazi (Mcha) for being a friend with the same dream of mine and for being the

best host during Ramadan, I really enjoyed going to cinema with you but I never understood the concept of the movies because of your destructive snoring, Peter Martin for being a very nice lab&flat mate and for the nice days we were going out for food stamping, I know you will beat Neil3 one day but unfortunately you will be bald, Arvind Swamy for being a good friend and the Indian flavour of our lab, it was pleasure to know you but seriously mate you are more complicated than Bermuda Triangle, Firozeh Ashtiani for being my brown friend and for the nice Iranian food she was sharing with me, Natalie Barnes for being my Mancunian friend and for teaching me how to pronounce citehhh (City) properly, Emyr Bakker for all the chocolates and doughnuts, I know you will regret that you never listened to me about the sexy haircut. You guys are all friends for life.

I would like to express a special word of thanks to my several other friends and colleagues who I shared many wonderful times with including Kurdo Fuad, Dr Shwan Hama, Dr Othman Qadir, Dr Sarchl Hamaqader, Dr Kamila Schmidt, Paz Aranega, Ana Isabel, Chrow Khurshid, Dr Sanat Kumar, Sangkab Sudsaward, Clifford Enyita, Pam, Kuburat, Nyam Chuwang, Hassan Ahmed, Kamran Chaudhry, Alice Guazzelli.

Finally, a few words to the people that I love, although no words can possibly describe how grateful I am to them. To my mother for the prayers, unconditional love, support and encouragement throughout the good and bad times and for so much more. I can just say I love you so much. To my brothers and sisters for always believing in me and giving me the strength to pursue my dreams.

And, last but not least I would like to dedicate this thesis to the memory of my late father. I miss you every day.

Declaration

I declare that no part of this work has been submitted by me for any degree in this or any other university. All of the work presented was conducted by me, except where otherwise stated in the text.

Soran Mohammed

List of Abbreviations

Ala	Alanine
bp	Base pairs
BSA	Bovine serum albumin
cDNA	Complementary DNA
C-terminal/-COOH	Carboxyl terminus
Da	Dalton
DMSO	Dimethyl sulfoxide
dH2O	Deionised water
DMSO	Dimethyl sulphoxide
DNA	Deoxyribonucleic acid
dNTPs	Deoxyribonucleotide triphosphates
DTT	Dithiothreitol
E. coli	Escherichia coli
EDTA	Ethylenediaminetetraacetic acid
ER	Endoplasmic reticulum
FITC	Fluorescein isothiocyanate
GFP	Green fluorescent protein
Gln	Glutamine
Glu	Glutamate
Gly	Glycine
HEMn	Normal Human Epidermal Melanocytes-Neonatal
His	Histidine
HRP	Horseradish peroxidase
IPTG	Isopropyl-β-D-thiogalactopyranoside

Ile	isoleucine
kDA	Kilo Dalton
L	Litre
LB	Luria-Bertani
Leu	Leucine
μM	Micromole
MP	Mastoparan
Mel	Melittin
Met	Methionine
MOPS	4-morpholinepropanesulfonic acid
mRNA	Messenger ribonucleic acid
nM	Nanomole
N-terminal/NH2	Amino terminus
NCBI	National centre for biotechnology information
OD	Optical density
PAGE	Polyacrylamide gel electrophoresis
PBS	Phosphate buffered saline
PCR	Polymerase chain reaction
Pm	Picomole
Pro	Proline
RPMI	Roswell Park Memorial Institute
RT	Reverse transcriptase
RT-PCR	Reverse transcriptase polymerase chain reaction
RNA	Ribonucleic acid
RNase	Ribonuclease

rpm	Revolutions per minute
Ser	Serine
SDS	Sodium dodecyl sulphate
TE	Tris-EDTA
TEMED	N,N,N',N'–retraethylethylenediamine
Tris-CI	Tris(hydroxymethyl)aminoethane
UV	Ultra violet light
Val	Valine
WM-115	Human melanoma cells from primary epitheloid tumour
WM 266-4	Human metastatic melanoma cells

Chapter 1. Introduction

1.1 Glycobiology

Glycosylation is one of the major forms of post-translational modifications in the cells of eukaryotes where it has been found to influence biochemically important protein properties, including protein dynamics, folding, immune responses, ligand binding, solubility and stability (Tsai, Wang & Huang, 2014). All of the chain structures of sugars in a cell or organism are known as the glycome, and the glycosylation machinery is said to be coded by 1 to 2 percent of the genome (Elkund & Freeze, 2005).

Biological information is thought of as a flow from DNA > RNA > Protein, however cells require two other major classes of molecules beyond this template, carbohydrates and lipids. Carbohydrates and lipids generate energy, serve as intermediates in biosynthetic pathways or as structural components or as signalling molecules. Carbohydrate structural roles are important in cell-cell interactions and interactions with the surrounding matrix. Cells and macromolecules are found with covalently attached oligosaccharides or glycans (sugar chains), sometimes these glycans can be found as free standing entities.

Glycans mediate or modulate a wide range of cell to cell and cell to matrix interactions which are crucial in the functioning and development of complex multicellular organisms, they also function in the interactions between different organisms, for example between parasite and host (Rodrigues, Acosta-Serrano et al. 2015). Highly dynamic glycans are plentiful in the cytoplasm and in the nucleus and bound to proteins where they serve as regulatory switches (Varki & Sharon, 2009).

From glycans, glycoconjugates are formed, this is when polysaccharides, oligosaccharides or monosaccharides (known as the glycones) covalently link to proteins or lipids (the aglycones), resulting in the formation of glycolipids, glycoproteins or proteoglycans.

The study of glycan activities related to cellular life is known as glycobiology (Kresge, Simoni et al. 2010). A unique range of biological activities may come from the members of this protein and lipid -glycoform pool. Within one species of one organism and under defined conditions, a distinct array of glycans is expressed by each cell type and the expression patterns tend to be conserved. However, stereotypic species to species changes during developmental stages may change these expression patterns, for example T cells switch to peanut-agglutinin positive from negative during their thymic development (Lowe and Marth 2003).

1.2 Structure of Glycans

Glycans are the most abundant organic biomolecules on the earth and have essential roles in the life of all living organisms as building blocks and well known compounds for energy supply and storage. The structure of glycan compounds is very varied; they can be seen as free mono, oligo and polysaccharides or attached to other biomolecules like proteins or lipids to form glycoconjugates such as glycoproteins, glycolipids and peptidoglycan which are all extensively observed in the human body. For example, about 50% of all human proteins are glycosylated. Genome sequencing projects have shown that sugar-processing enzymes in living organisms from eubacteria to archaea and eukaryotes occupy around 1% of highly conserved regions of each genome (Apweiler, Hermjakob et al. 1999, Wong 2005).

Glycans are molecules made up of a combination of sugar units, mainly the six-carbon hexose sugars, which include mannose, galactose and glucose. These sugars are only presented in the D-enantiomer configuration; due to the asymmetric chirality of the groups (Varki & Sharon, 2009). Proteins and nucleotides being linear polymers, have only one type of basic linkage, however monosaccharides have the ability to theoretically generate

two configurations of the glycosidic linkage α or β (depending on whether the OH group of the anomeric carbon extends up or down from the carbon), furthermore the many hydroxy groups of the sugar also permit several possible regioisomers, plus a monosaccharide can be involved in more than two glycosidic linkages, thus serving as a branchpoint which will link in one of several positions on nearby chains of monosaccharides or other types of molecules, resulting in the variable structure and function of the glycans.

Glycan structures are thus enormously diverse and the main reason of this diversity is the fact that their structures are defined by their functions but not *vice versa*. In consequence, they are very flexible and frequently adjust to physiological needs (Cantor and Schimmel 1980, Pocanschi, Kozlov et al. 2011). The structural complexity is further complicated when it comes to membrane glycans because the fluidity of the outer layer of the lipid bilayer makes it almost impossible to study the structure of a specific surface glycolipid. Despite exceptions in some types of glycoproteins that are tethered to cytoplasm contents, studying cell surface proteins is equally as complex as studying cell surface glycolipids for similar reasons (Roseman 2001).

1.3 Biosynthesis

Different enzymes or transferase complexes are needed for each pathway to initiate glycosylation. The monosaccharide (being the first sugar) which is linked to the lipid or protein defines its pathway, analogous with nucleotides for proteins, then either a single sugar or a sugar chain may be added (Elkund, Freeze & Patterson, 2012). For biosynthesis of glycans activated monosaccharides (nucleotide sugars) are required, these are delivered to specific locations in the Golgi apparatus or the endoplasmic reticulum.

There are two main glycoconjugate structures, *N*-glycans and *O*-glycans.

1.4 *N*-glycans

N-glycans are covalently attached to protein at asparagine (Asn) residues by an *N*-glycosidic bond. Five different N-glycan linkages have been reported, of which *N*-acetylglucosamine to asparagine (GlcNAcβ1-Asn) is the most common. The *N*-glycans are primarily processed in the endoplasmic reticulum (ER) (Mitra et al, 2006). In the first instance there needs to be the assembly of an oligosaccharide precursor structure which is linked to the lipid dolichylphosphate (Dol-P), this is done by linking, via pyrophosphates, to two N-acetylglucosamine residues (GlcNac) in the cytosolic end of the ER. Studies using glucosidase and protease degradation indicate that to the GlcNAc five mannose residues are added sequentially, using GPD-Man as a donor, flipping the glycan across the ER membrane and orienting it in the ER lumen (Herbert, Follemer & Helenius, 1995). To this, using the Dol-P-Man donor, another four mannose residues are added. Three additional glucose residues are added to the terminal ends of the glycan by the donor Dol-P-Glc, this completes assembly of the dolichol oligosaccharide precursor (Hanahima et al, 2014). An ER membrane enzyme, oligosaccharyltransferase (OST) then transfers the precursor to the asparagine residue on a peptide chain (that is being synthesized and translocated through the ER membrane) from the dolichol (the amide group of asparagine is attached to the reduced terminal of *N*-acetylglucosamine to form an aspartylglucosamine linkage). During the formation of this linkage, C-1 of the *N*-acetylglucosamine residue replaces its original attachment with the hydroxyl group to attachment with the amino group. Following attachment, a series of processing reactions, the initial steps of which appear to be conserved in eukaryotes, occur and these have a major role in lysosomal trafficking and protein folding. Firstly, glucosidases I and II and mannosidase I act on the

oligosaccharide precursor to remove the 3 glucoses and 4 mannoses (Hall et al 2014). Post-trimming in the ER, the protein can now be transferred to the Golgi apparatus, here the N-glycan (*N*-acetylglucosamine transferase I) substitutes the mannose residues with GlcNAc moieties. The alterations that have occurred aid in the formation of the substrate for the enzyme α-Mannosidase II, which acts on the *N*-glycan in the medial Golgi and removes two further mannose residues (the α1-3 linked and α1-6 linked mannose). This leaves just 5 residues in the original glycan which is now named the trimannosyl core (Varki & Sharon, 2009). To this resulting *N*-glycan product (named GlcNAc1Man3GlcNAc2-Asn) the GlcNAcT-II enzyme adds another GlcNAc residue to the remaining mannose branch forming an antenna. There are many other enzymes in eukaryotes that can form various complex and diverse glycan structures by adding other sugar residues, such as fructose, galactose and GlcNAc, in fact deficiencies in these enzymes can cause disease, for example, Ehlers-Danlos syndrome (where patients are deficient in collagen processing) and congenital disorder of glycosylation (CDG) (Eklund and Freeze 2006).

All *N*-glycans share a common core sugar sequence, Manα1–6(Manα1–3)Manβ1–4GlcNAcβ1–4GlcNAcβ1-Asn-X-Ser/Thr, and are classified into three types: (1) oligomannose, in which only mannose residues are attached to the core; (2) complex, in which "antennae" initiated by N-acetylglucosaminyltransferases (GlcNAcTs) are attached to the core; and (3) hybrid, in which only mannose residues are attached to the Manα1–6 arm of the core and one or two antennae are on the Manα1–3 arm (Figure 1). Two of the three groups (high mannose and complex oligosaccharides) are the most abundant types of glycosylation found on the majority of glycoproteins (Stanley, Schachter et al. 2009).

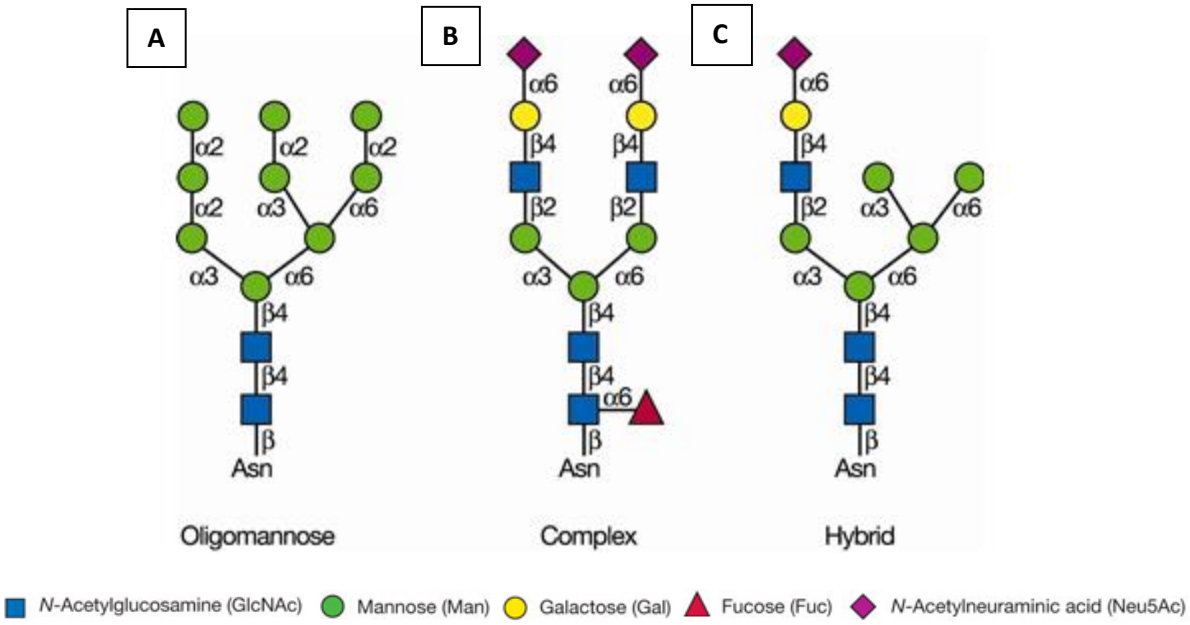


Figure 1. Types of N-glycans. N-glycans added to protein at Asn-X-Ser/Thr sequons are of three general types in a mature glycoprotein: oligomannose, complex, and hybrid. Each N-glycan contains the common core Man3GlcNAc2Asn. A: Oligomannose N-glycan which composed of N-acetyleglucosamine (Blue squares) and mannose (Green circles) residue; B: Complex N-glycan which composed of N-acetyleglucosamine (Blue squares), mannose (Green circles), galactose (Yellow circles), Fucose (Red Triangle) and sialic acid (Burgundy squares) residues; C: Hybrid N-glycan which composed of N-acetyleglucosamine (Blue squares), mannose (Green circles), galactose (Yellow circles) and sialic acid (Burgundy squares) residues (Stanley, Schachter et al. 2009).

The pattern of glycosylation and the modification of the glycans attached to glycoproteins are partially dependent on the structure of the proteins as seen for example in the glycosylation of blood group substances and mannose 6-phosphate (Fredriksson, Podbielska et al. 2010, Sleat, Sun et al. 2013).

1.4.1 Diversity of *N*-glycans

The above description is a general profile of *N*-glycans. However, a range of differences are seen between them especially in those of complex type. In order to comprehend this diversity, it is better to look individually at the outer chain and the core portion of *N*-glycans. The core portion is composed of two *N*-acetylglucosamine residues and three mannoses. The complex types of oligosaccharides also have side chains elongating from α-mannose residues and the first residue of this side chain is *N*-acetylglucosamine. Diversity arises from the variation in the number of *N*-acetylglucosamine residues linked to α-

mannoses and the positions to which *N*-acetylglucosamine are linked (Re, Miyashita et al. 2011) (Figure 2).

Another reason for *N*-glycan diversity is the variation in the side chains that contain non-reducing termini on which majority of the carbohydrate functionalities rely on. The glycan units formed by subultimate and terminal sugars at "outer" positions of a glycan often determine the function(s) or recognition properties of a glycoconjugate.

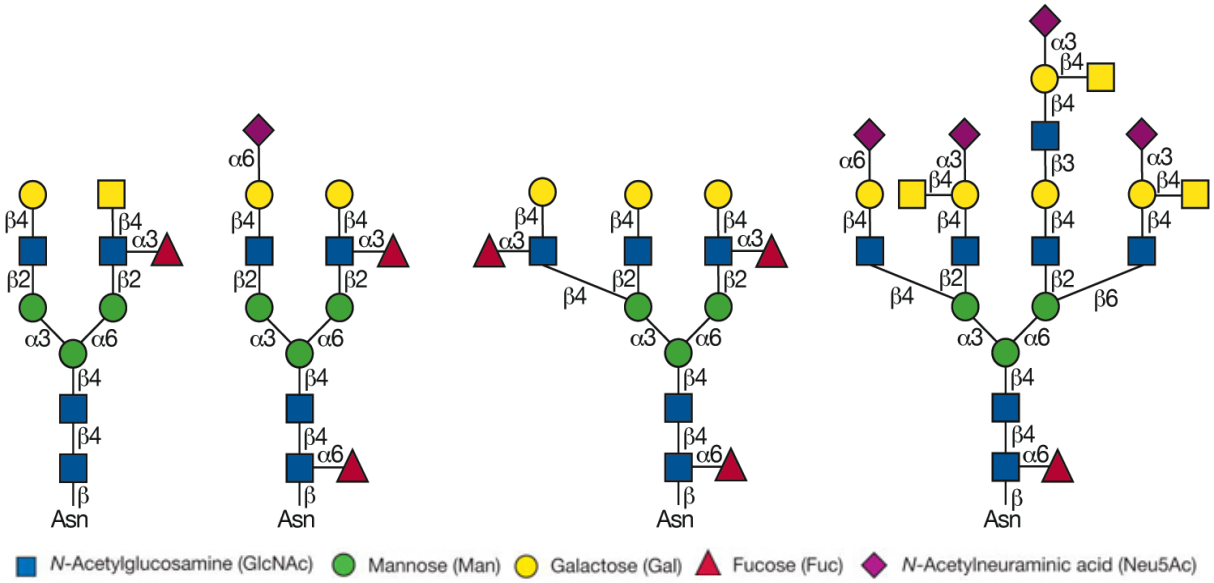


Figure 2. Typical complex *N*-glycan structures found on mature glycoproteins (Stanley, Schachter et al. 2009).

N-Glycans occur on many secreted and membrane-bound glycoproteins at Asn-X-Ser/Thr sequons. Analyses of protein sequence databases have revealed that about two thirds of the entries contain this consensus sequence (Jensen, Karlsson et al. 2012).

1.4.2 Functions of *N*-glycans

Inhibitors of *N*-linked glycosylation have been used to determine the function of *N*-glycans for example tunicamycin; also inhibitors of *N*-glycan processing such as castanospermine, deoxynojirimycin, and swainsonine; have been used as a mutants generated in a gene coding for a glycosylation activity. This has been done in model organisms such as yeast,

cultured mammalian cells, *Drosophila*, *C. elegans*, zebrafish, and mouse. Mouse mutants in particular have provided enormous insights into the functions of *N*-glycans. Deletion of the *Mgat1* gene that encodes GlcNAcT-I prevents the synthesis of complex and hybrid *N*-glycans, and $Man_5GlcNAc_2$ is found at all complex and hybrid *N*-glycan sites. Elimination of GlcNAcT-I in the mouse results in death during embryonic development. The complex *N*-glycans that fail to be synthesized in mice lacking GlcNAcT-I, GlcNAcT-V, GlcNAcT-IVb, and FucT-VIII are important in retaining growth factor and cytokine receptors at the cell surface, probably through interactions with glycan-binding proteins. Furthermore, deletion of genes encoding sialyltransferases, fucosyltransferases, or branching *N*-acetylglucosaminyltranferases (other than GlcNAcT-I) have produced viable mice in the main with defects in immunity or neuronal cell migration or emphysema or inflammation. *N*-Glycans are known to become more branched when cells become cancerous, and this change facilitates cancer progression. Tumors formed in mice lacking GlcNAcT-V or GlcNAcT-III may be retarded in their progression. Thus, glycosyltransferases have been targets for the design of cancer therapeutics (Hibberd, Trevillian et al. 2012, Stanley, Batista et al. 2013, Stanley 2014, Xiao, Smeekens et al. 2016).

1.5 *O*-glycans

O-glycosylation is a common covalent modification of serine and threonine residues of mammalian glycoproteins. *O*-glycans and their structures are comprised of four common core subtypes, based on differentiating linkage reaction of a monosaccharide to unsubstituted GalNAc (GalNacαSer-Thr), so the reduced terminal of *N*-acetylgalactosamine of glycans is attached to the hydroxyl groups of serine and threonine residues of polypeptides. They represent the second most abundant form of carbohydrate-protein

binding and are also known as mucin-type glycans since the initial description of this conjugate was in a mucus substance called mucin (Corfield 2004).

Based upon the binding pattern between glycans and amino acids, *O*-glycans can be divided to different subgroups as follows: (a) mucin type carbohydrates, *N*-acetylgalactosamine attached to serine or threonine; (b) *O*-GlcNAc type, *N*-acetylglucosamine attached to serine or threonine; and (c) xylose attached to serine or threonine in proteoglycans. Animal cells contain these three groups abundantly. Less known groups are (d) galactose attached to hydroxylsine in collagen, (e) xylosyl glucose or glucose attached to serine or threonine in clotting factors, and (f) fucose attached to plasma glycoproteins. In addition, in plants, there are (g) arabinose attached to hydroxyproline and galactose attached to serine, and in fungi, mannose attached to serine or threonine (Sadler 1984, Harris, Leonard et al. 1991, Brockhausen, Schachter et al. 2009, Joshi, Steentoft et al. 2015, Stanley 2015).

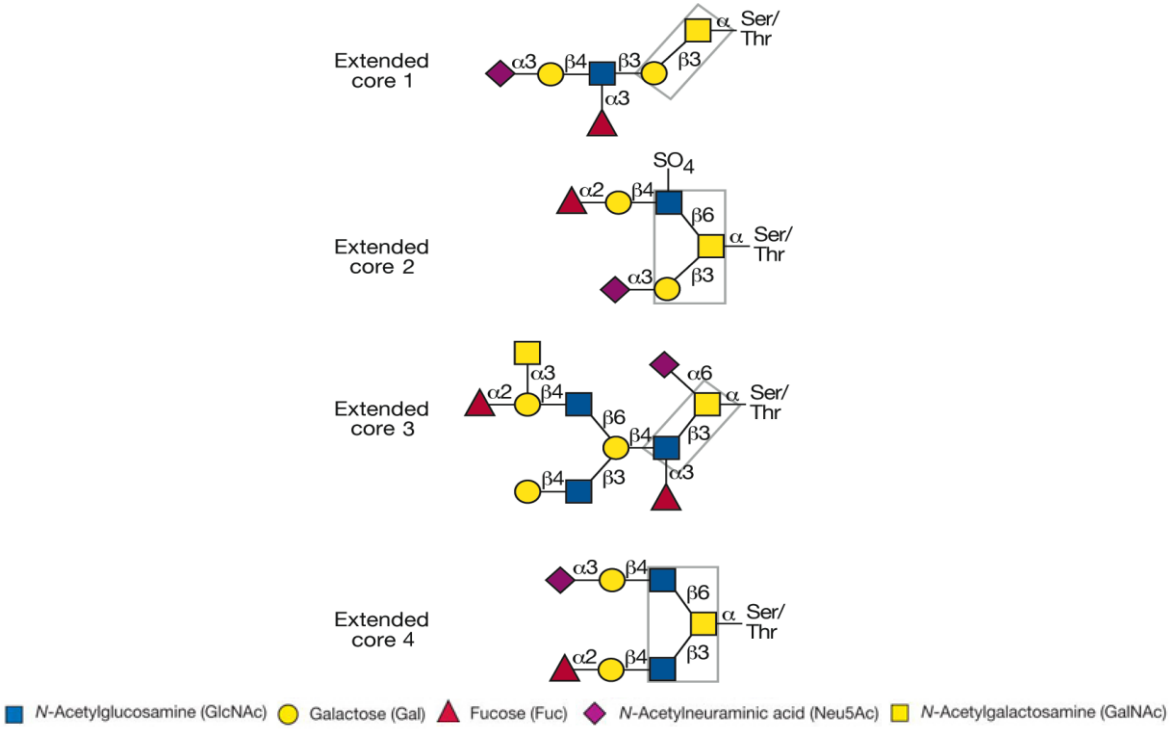


Figure 3. Complex *O*-GalNAc glycans with different core structures. Representative examples of complex *O*-GalNAc glycans with extended core 1, 2, or 4 structures from human respiratory mucins and an *O*-GalNAc glycan with an extended core 3 structure from human colonic mucins. All four core structures (in *boxes*) can be extended, branched, and terminated by fucose in various linkages, sialic acid in α2-3 linkage, or blood group antigenic determinants. Core structures 1 and 3 may also carry sialic acid α2-6-linked to the core *N*-acetylgalactosamine (Brockhausen, Schachter et al. 2009).

1.5.1 Biosynthesis

The transfer of the first sugar from UDP-GalNAc directly to serine or threonine in a protein initiates the biosynthesis of all *O*-GalNAc glycans. Subsequently, with the addition of the next sugar, different mucin *O*-glycan core structures are synthesized. In contrast to *N*-glycosylation and *O*-mannosylation, no lipid-linked intermediates are involved; no glycosidases appear to be involved in the processing of *O*-GalNAc glycans within the Golgi.

The *O*-glycans of mucins produced in a specific tissue predict the enzyme activities that are present in that tissue. All four core structures (figure 3) can be extended, branched, and terminated by fucose in various linkages, sialic acid in α2-3 linkage, or blood group antigenic determinants. All of the core structures can be sialylated and biosynthesis of *O*-glycans can be modified and then terminated through the addition of a sialic acid residue in the early stages of synthesis. These additions give rise to a series of different structures of *O*-glycans that cause restriction in the biosynthetic pathways leading to the four core types. These have been designated tumour associated antigens (T-antigens) and found in association with cancer cells (Xia et al, 2012).

1.5.2 Functions of *O*-glycans

All mammalian cells studies to date show *O*-GalNAc glycosylation. In the secreted mucins of the gastrointestinal, genitourinary and respiratory tracts, the *O*-GalNAc glycans of mucous glycoproteins are essential for hydration and protection of the epithelium. Furthermore, mucins trap bacteria via specific receptor sites within the *O*-glycans. Some sugar residues or their modifications can mask underlying antigens or receptors.

O-GalNAc glycans, especially in the mucins, have a significant effect on the conformation of the attached protein and dependent on the size and hence bulkiness of the *O*-glycans.

The underlying peptide epitopes are variably recognized by antibodies. In addition, *O*-glycosylation of mucins provides almost complete protection from protease degradation.

1.6 Sialic Acids

Sialic acid is a nine-carbon monosaccharide. Sialic acids (Sias) are typically found to be terminating branches of *N*-glycans, *O*-glycans, and glycosphingolipids (gangliosides) (and occasionally capping side chains of GPI anchors) (Yabu, Korekane et al. 2013, Bate, Nolan et al. 2016). They generate biologically significant diversity due to (1) the different α linkages that may be formed between the C-2 of Sias and underlying sugars mediated by specific sialyltransferases (CMP-Sias is used as high-energy donors). Linkages to the C-3 or C-6 positions of galactose residues are most common or to the C-6 position of *N*-acetylgalactosamine residues. Sialic acids may also occupy internal positions within glycans, the most common being one Sia residue attached to another, often at C-8 position and (2) a variety of natural modifications, the C-5 position may have an *N*-acetyl group (giving Neu5Ac) or a hydroxyl group (as in Kdn). The 5-*N*-acetyl group can also be hydroxylated, giving *N*-glycolylneuraminic acid (Neu5Gc). Less common, the 5-amino group is not *N*-acylated, giving neuraminic acid (Neu). These give the four "core" Sia molecules (Neu5Ac, Neu5Gc, Kdn, and Neu) to which one or more additional substitutions may occur at the hydroxyl groups on C-4, C-7, C-8, and C-9 (*O*-acetyl, *O*-methyl, *O*-sulfate, *O*-lactyl, or phosphate groups). The carboxylate group at the C-1 is ionized at physiological pH, but can be condensed into a lactone with hydroxyl groups of adjacent saccharides or into a lactam with a free amino group at C-5. Combinations of different glycosidic linkages with the multitude of possible modifications generate hundreds of ways in which Sias can present themselves.

1.6.1 Physiological and Pathophysiological Roles of Sialic Acids

Sialic acid is considered as one of the most important biomolecules of life since its derivatives and terminal orientation on cell membranes and macromolecules play a major role in many biological and pathological processes (Schauer 2000). They are also negatively charged molecules by which they attract/repulse molecules and cells then they specifically bind to positively charged molecules and transport them. Moreover, as glycoprotein components they protect endothelia through their contribution in giving high viscosity to the lining of mucins such as in the intestine and they contribute in shaping cell membranes supramolecular structures such as gangliosides and consequently influence their functions (Siebert, von der Lieth et al. 1996, Schauer, Ernst et al. 2000).

Sialic acid is the most abundant type of terminal family on vertebrate carbohydrates and their taxonomic distribution is surprising because they are found in both the deuterostome lineage animals and some of their pathogens (Moxon, Rainey et al. 1994).

The nine carbon glycans on erythrocytes are binding sites for parasites and viruses which bind differentially to different type of sialic acids and/or their linkages of different species. These binding sites can be in the form of either decoy binding sites or pathogen receptors (Wybenga, Epand et al. 1996).

The two main and very common types of sialic acids are *N*-glycolylneuraminic acid (Nau5Gc) and *N*-acetylneuraminic acid (Nau5Ac) (Figure 4). The former comes from Nau5Ac after hydroxylation of its acetyl group by CAMP-Nau5Ac hydroxylase (CMAH). While, Nau5Gc exists in the glycosylation pattern of all mammals, *CMAH* gene is deactivated in humans and thus Nau5Gc is absent in their normal glycosylation system. However, this form of sialic acid has been detected in human cancer cells and the reason is believed to be the interaction between glycosylation pattern of cancer tissues and

exogenous sources of Nau5Gc such as red meat and other milk products (Taylor, Gregg et al. 2010). Thus targeting this form of sialic acid may aid in drug delivery.

Despite the general thought that the biosynthesis of a certain oligosaccharide is mainly controlled by a specific cognate glycosyltransferase, the biosynthesis process has remained complex and partially understood. A poor correlation between glycosyltransferase enzymes and their products suggest the existence other factors with relevant role (Kudo, Ikehara et al. 1998, Mehr and Withers 2015). One of the most common glycan structures in the antennae of *N*-linked chains is α2,6-sialylation of lactosamine (Galb1,4GlcNAc) which is catalysed only by b-Galactoside α2,6-sialyltransferase (ST6Gal.I) enzyme. Attempts to compare glycosylation patterns of cancer cells with their surrounding healthy cells have found strong correlations between the elevations of ST6Gal.I activity and glycosylation changes in different types of cancer cells specifically colon cancer. They have also found that this elevation increases further in metastatic colon tissues (Gessner, Riedl et al. 1993, Lise, Belluco et al. 2000). The increase in α2,6-sialylation in membrane glycoconjugates of colon cancer cells was confirmed via using sialic acid binding lectin from *Sambucus nigra* (SNA) and *Tricosanthes japonica* (Dall'Olio and Trere 1992, Yamashita, Fukushima et al. 1995).

Thus lectin mediated binding to sialic acid presents a possible drug targeting mechanisms.

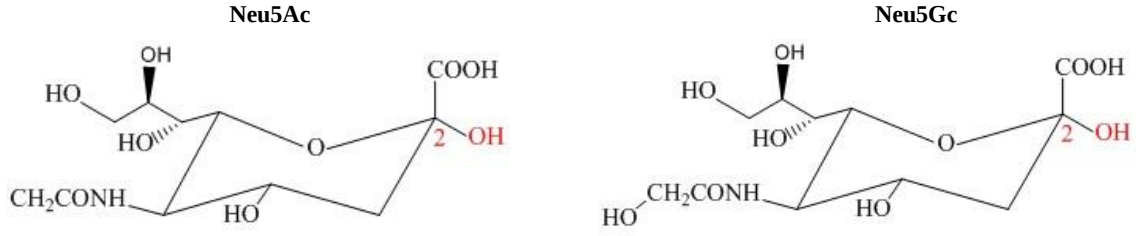


Figure 4. *N*-acetylneuraminic acid (Neu5Ac) and *N*-glycolylneuraminic acid (Neu5Gc).

1.7 Altered Glycosylation in Cancer Cells

It is assumed that cell to cell interactions are probably affected by cell-type specific carbohydrates. It is thus possible that specific signals are provided by the diversity of carbohydrate structures in different cell types and these signals are recognized by counter-receptor molecules. It is then reasonable to assume that different cell recognition molecules are provided to cells according to the alteration of their surface carbohydrates and these molecules are essential when cells interact with each other and differentiate. The tumour cell interaction with other cells, in order to achieve tumour cell colonization and metastasis, is similarly facilitated by alteration of cell surface carbohydrates (Rabinovich and Croci 2012, Tsuboi 2015).

These cell surface carbohydrates have long been a target for drug design (reviewed in Toporkiewicz, Meissner et al. (2015)).

1.8 Glycan Binding Proteins

A wide variety of physical properties of glycans such as shape, mass and charge mediate their biological roles and it appears that nature has taken advantage of this diversity through interacting proteins known as glycan binding proteins (GBPs) which are binding counterparts for glycans. These proteins have the ability to bind different structures of glycans and their recognition property has become the key factor in their exploitation (Simpson, Sacher et al. 2015). The carbohydrate-protein conjugates can be classified into proteoglycans and glycoproteins. In proteoglycans, the carbohydrates are relatively large (>3000dalton) and they are formed by replications of disaccharide units. In glycoproteins, structure repetition is usually absent except for poly-N-acetyllactosamines (Fukuda 1994).

The GBPs are found in all organisms and they are mainly divided into two different classes, glycosamingoglycan-binding proteins and lectins (Simpson, Sacher et al. 2015).

1.8.1 Glycosamingoglycan (GAG)-binding Proteins

Glycosaminoglycans are highly negatively charged linear polysaccharides found on the surface and inside of cells. They are composed of repetitive disaccharide units attached to the core proteins through covalent bonds to build proteoglycans. The constituent of the units is a combination of amino type of sugars such as *N*-acetylglucosamine and *N*-acetylgalactosamine with variably sulphated (except hyaluronic acid) galactose or uronic acid such as glucuronic and iduronic acid (Esko, Kimata et al. 2009). The interaction with many proteins like cytokines and growth factors and regulating their functions are the key roles in the biological functions of GAGs (Handel, Johnson et al. 2005).

The most studied type of GAGs is heparin which is a repetitive unit of uronic acid (10% glucuronic acid and 90% iduronic acid) attached to glucosamine residues in the form of linear copolymer of 1 → 4 and the existence of carboxylate and sulphate residues in their main structure gives them a highly negative charge. While the trisulphated disaccharide is the most common unit in heparins, the average amount of sulphate groups contained in their disaccharides is 2.7 groups (Figure 5A) (Gallagher, Turnbull et al. 1992, Linhardt, Ampofo et al. 1992).

Heparan sulphate is a derivate of heparin with unique structure composed of the same repeated 1→4 linear copolymers but it is predominantly between glucuronic acid and glucosamine (Figure 5B). In heparan sulphates the average of sulphate per disaccharide is less than 1. While, all the disaccharide sequences found in heparin are in heparan sulphate as well, a few extra minor sequences in heparan sulphate have increased its structure and sequence complexity. Further complexity in heparan sulphate comes from observing

substantial levels of iduronic acid in its structure since the structure is mainly composed of glucuronic acid. Also, the average molecular weight of heparan sulphate (29,000Da) is higher than heparin (8000Da) because of longer chains (Griffin, Linhardt et al. 1995).

There are other types of GAGs with the ability to bind proteins such as hyaluronan, chondroitin, dermatan and keratan sulphate. Their main structures are composed of similar components such as *N*-acetylgalactosamine and *N*-acetylglucosamine linked to glucuronic, iduronic and glucose in alternating $1\rightarrow 4$ or $1\rightarrow 3$ linear copolymers (Templeton 1992).

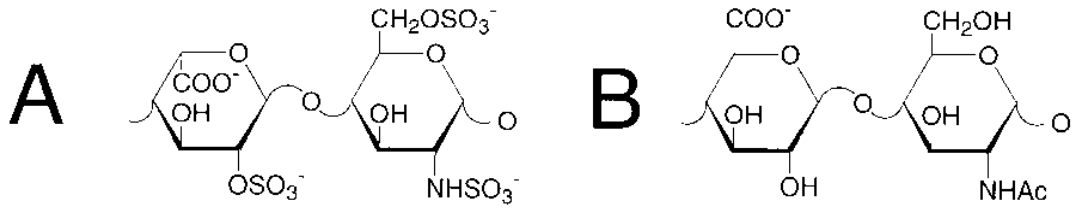


Figure 5. Structure of heparin and heparan sulphate. **A**: Major repeating disaccharide unit in heparin (iduronicacid-2-sulphate = glucosamine-2,6-disulphate; 75–90% of disaccharide sequences). **B**: Major disaccharide repeating unit of heparan sulphate (glucuronic acid =N-acetylglucosamine; 10–50% of disaccharide sequences) (Hileman, Fromm et al. 1998).

1.8.2 Lectins

At the end of 19th century a novel family of proteins were discovered that have ability to agglutinate erythrocytes and they were named as phytoagglutinins because the first discovery was from a plant (The seed of ricin (*Ricinus communis*)) or hemagglutinins because they can agglutinate red blood cells (Sharon and Lis 2004).

Lectin terminology which is a Latin word meaning “to select” and its use in terminology came about with the development of knowledge about this family and its functions and the term soon changed from the individual name to a generic term. At the beginning of their discovery there were relatively few studies about lectins and almost all were from plants. However, lectin research increased very steadily from 1960s on-ward and lectins became a very popular topic for two reasons. Firstly, a wide range of applications of lectin became

apparent such as detection and isolation of glycoconjugates, cell and tissue histochemistry, and observation and characterization of physiological and pathological changes in the cell surface glycans from cell differentiation to cancer. Secondly, simple and quick isolation and purification of a variety of lectins through immobilized carbohydrates affinity chromatography became possible (Sharon and Lis 2003).

Lectins are generally defined as non-immune origin proteins with the ability to specifically attach to carbohydrates. The purpose behind using the non-immune origin term is to differentiate them with carbohydrate specific enzymes such as glycosidase and kinase which cannot agglutinate cells. The majority of lectins have the ability to recognize just a few of the hundreds of monosaccharide found in nature such as mannose, fucose, glucose, galactose, *N*-acetylneuraminic acid, *N*-acetylglucosamine, and *N*-acetylgalactosamine, however, majority of oligosaccharides built from these monosaccharides may also be bound by lectins (Sharon and Lis 2003).

Lectins can also agglutinate cells and attach to polysaccharides and glycoproteins because of their di- or oligovalent structures which gives each lectin molecule at least two glycan binding sites that help cross linking between cells through their cell surface carbohydrates or between other molecules that contain carbohydrates. However, some types of lectins do not possess glycan binding property because of their monovalent structure. It is also important to notice that non-carbohydrate ligands may attach to lectins through one or more binding sites, although the main ligands for lectins are carbohydrates (Goldstein, Hughes et al. 1980, Barondes 1988, Sharon and Lis 2003).

1.8.2.1 Classification of Lectins

Lectins have been classified according to different principles such as origin, carbohydrate specificity, and localization in cells and organisms. The outcome of these classifications overlap each other as a result of common properties shared between different lectins as shown in table 1 (Lakhtin, Afanas' ev et al. 2008, Lakhtin, Lakhtin et al. 2011).

Table 1. Some known classifications of lectins (Lakhtin, Lakhtin et al. 2011).

Microbial, plant, animal lectins
C-type, L-type, R-type, CBM
Phytolectins
C-type (Con A*, PHA)
R-type (RCA-I, II)
Gevein family (WGA, STA)
β-prism fold type
β-trefoil fold type
Others
Animal lectins
S-Lac-type (galectins)
C-type (Ca-зависимые)
Collectins, selectins, lecticans, endocytic receptors
Tachylectins
I-type (Ig-like domains) Siglecs, others
Ficolins (L,M)
Pentraxins (SAP, CRP)
Retrocyclins
Defensins, antimicrobial peptides
Cytokines
F-type
Glycosaminoglycan-binding
Heparin-binding, hyaladherins, pentraxins
Proteoglycan lectins (STA, lecticans)
Others
Bacterial lectins
C-type, F-type, L-type, R-type, CBM hemagglutinins, adhesions, toxins
Surface lectins
Fimbrial lectins, pilins
Non-fimbrial lectins
Symbiotic lectins
Probiotic lectins
Lactobacillus, Bifidobacterium lectins
Extracellular lectins
Intracellular lectins
Membrane-bound lectins
Intracellular eukaryotic lectins
P-type (lysosomal)
L-type (ER, ER-Golgi, Golgi
RNA-sensing lectins (galectins 1, 3)
Nucleoplasmic plant lectins
F-box lectins
Viral lectins
(glycoproteins, assemblies)
Siaα-recognizing
Gal-recognizing
GalNAc-recognizing
Others

Comments. *Lectins: CBM = carbohydrate-binding modules, Con A = concanavalin A (*Canavalia ensiformis*, jack bean), CRP = C-reactive protein, PHA = phytohemagglutinin (*Phaseolus vulgaris*, kidney bean), RCA-I, II = *Ricinus communis* agglutinin 120 or 60 kD, respectively (castor bean), SAP = serum amyloid P substance, STA = *Solanum tuberosum* agglutinin (potato tubers), WGA = wheat germ agglutinin (*Triticum vulgare*). Other abbreviations: ER = endoplasmic reticulum; Ig = immunoglobulin; Gal = galactose; GalNAc = N-acetylgalactosamine; Sia = Sialic acid (mono/diacetylated; varied in link type, and antennary cluster type).

Because of the size of lectin classifications, only their taxonomical origin classification is explained in this review.

1.8.2.2 Plant Lectins

The plant kingdom contains many types of carbohydrate-binding domains spread all over the kingdom and often the occurrence of a specific domain is not restricted to a specific plant family. For instance, hundreds of putative lectin genes were found recently in the full genome analysis of *Arabidopsis*, rice and soybean which belong to all main lectin super families (Jiang, Ma et al. 2010).

While many of the known plant lectins have affinity to specific monosaccharides, results from glycan array studies conducted on different plant lectins showed that one glycan structure cannot characterize a particular glycan-binding domain (CBD) but a carbohydrate-binding profile is required instead (Lannoo, Peumans et al. 2006, Fouquaert, Smith et al. 2009, Taylor and Drickamer 2009). The affinity of a typical CBD is also lower for simple glycans than for complex glycans such as *N*-glycans which have a bulky structure with a particular glycan complementary to a shallow depression on the surface of a particular CBD (Garcia-Pino, Buts et al. 2007, Schouppe, Rougé et al. 2010). Any increase of amino acid numbers in the primary binding site of particular CBD extend the binding with the sugar residues of complex carbohydrates because of increasing the number of hydrogen bonds (H-bonds). Therefore, the affinity of CBDs for simple glycan structures are much lower than for complex structures (*N*-glycans) which are believed to be natural binding targets for the majority of plant lectins (Garcia-Pino, Buts et al. 2007, Van Damme, Lannoo et al. 2008).

According to the expression pattern, the production of lectin in plants can be divided into two main types. First, lectins with regulated expression patterns during plant development and synthesized by the ER associated ribosomes without any effects of external environmental conditions. The cellular positions of this type are inter/extra cellular spaces or vacuolar compartment and they mainly accumulate in seeds and different storage tissues such as bulbs, tubers and rhizomes. For instance, different isoforms of common bean lectin *Phaseolus vulgaris* agglutinin (PHA) and snowdrop lectin *Galanthus nivalis* agglutinin (GNA) are produced during the development of the plants and accumulate in their seeds and bulbs respectively (Van Damme, Peumans et al. 1998a). Second, lectins may be synthesized in cytoplasm under certain external abiotic or biotic stress conditions such as drought, wounding and insect herbivory. This group of lectins is called inducible plant lectins and they are mainly stored in non-storage tissues like root, leaves and flowers. Since their cellular positions are nuclear or cytoplasmic compartments, this type is mainly characterized by cellular localization and relatively low expression levels (Van Damme, Lannoo et al. 2008).

1.8.2.3 Animal Lectins

Animal lectins are a group of glycan-binding proteins with highly diverse amino acid sequences and primary structures. It is believed that the discovery of animal lectin was long before the discovery of plant lectins; however, in the first discoveries many of them were not known as carbohydrate binding proteins (Kilpatrick 2002).

The lectins of different animals have different structures and activities which are both strongly associated with each other and the latter is entirely dependent on the former. Moreover, the primary structure of animal and plant lectins is heterologous even though they share similar preferential carbohydrate binding property. Also, animal lectins are

different from enzymes and antibodies in terms of origin because animal lectins are neither catalyst and nor of immune origin. However, each animal lectin possesses a specific carbohydrate recognition domain (CRD) which helps the lectin to recognise and attach to complex carbohydrate structures (Lis and Sharon 1998, Ghazarian, Idoni et al. 2011).

According to their structures and activities, animal lectins are grouped into eight main families with similar sequence and structural properties as follow: Calcium-requiring lectins (C-type), mannose-6-phosphate binding lectins (P-type), galectins lectins (S-type), immunoglobulin-like lectins (I-type), M-type lectins, L-type lectins, chitinase-like lectins and F- type lectins (Anderson, Evers et al. 2008).

Animal lectins have many important *in vivo* biological functions such as cell trafficking, prevention of autoimmunity, and immune regulation. However, their core function is the recognition of pathogen carbohydrates and inaccessible carbohydrates of the host cells. Thus, recognition function is a direct defence line against pathogens such as ingesting and killing infectious foreign organisms that have mannose on their surface glycans via macrophages mannose specific receptors (Kilpatrick 2002, Singh and Sarathi 2012).

The largest and most diverse family among animal lectins is C-type family. The independently folding carbohydrate binding domain (CRD) of this family is calcium ($Ca^{2+)}$ dependent because during binding CRDs usually bind to Ca^{2+}. The CRD of C-type lectins contains receptors often found in the immune system and their functions change according to the form of glycan recognition. For example, if they recognise endogenous mammalian glycans, they influence glycoprotein turnover and cell adhesion but their influence changes to pathogen recognition in the case of binding to glycans on microorganisms (Kolatkar, Leung et al. 1998).

1.8.2.4 Viral Lectins

Viruses have their own lectins like other organisms and their lectins are generally known as hemagglutinins (HA). By far the most studied viral lectin is influenza virus hemagglutinin which has affinity for terminal sialic acids. The avian and human influenza virus infections primarily happen through binding of the virus hemagglutinin to terminal α2-3- linked sialic acids of intestinal epithelial cell receptors and terminal α2-6- linked sialic acids of respiratory tract cell receptors respectively. In order to adapt to human host and infect human population, avian influenza viruses had to cross species barriers and conduct a switch from α2-3 specific to α2-6 specific (Figure 6) (Stevens, Blixt et al. 2006).

The influenza virus hemagglutinin has ability to agglutinate human red blood cells through binding to cell surface sialic acid-terminated oligosaccharides and the de-sialylation of the cells is shown to inhibit this agglutination. However, sialic acid is not the only ligand for cell agglutination because in some cases the agglutination is shown to happen because of other ligands (Rogers and Paulson 1983). While the type of terminated sialic acid on oligosaccharides mainly affects the binding of hemagglutinin to red blood cells, the binding is also affected by the inner configuration of the oligosaccharides. For example, changing GalNAc to GlcNAc in Siaa2– 3Galß1–3GalNAc oligosaccharide on red blood cells would inhibit agglutination property of two out of four viruses (Gambaryan, Yamnikova et al. 2005).

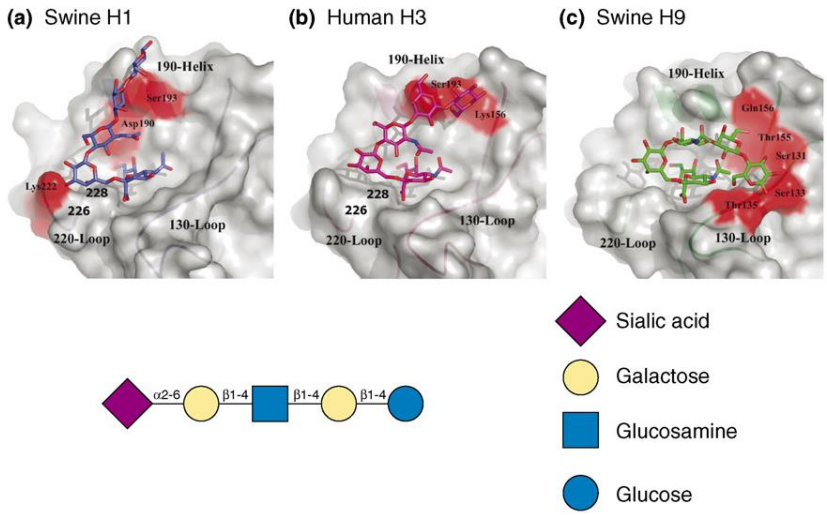


Figure 6. Interaction between influenza HA and a human receptor analog. Modeled interaction surfaces between the binding site of the influenza virus HA (grey) and a 'human' Siaa2–6 receptor analogue lactoseries tetrasaccharide c (LSTc) shown in blue, pink and green. HA of three different influenza viruses from (a) swine H1, (b) human H3 and (c) swine H9 and LSTc, which is terminated by a Sia in an a2–6 configuration. The residues involved in potential hydrogen bonds of the LSTc and the HA are shown in red and demonstrate the different amino acid residues of the HA that are involved in binding and how different sites are involved for different strains of virus. The H9 conformation is different, being folded back on itself. Reproduced, with permission (Lamblin, Degroote et al. 2001).

1.8.2.5 Bacterial Lectins

Bacterial lectins are known as adhesins because of their main role in adhesion process carried out by bacteria during the colonialization of cell surfaces of the host to withstand the cells defence mechanism. The main cause of bacterial adhesion is the interaction between hosts cell surface carbohydrates and surface filamentous proteins of bacteria known as fimbriae which are working as adhesive organelles through their lectin subunits. The interaction between cell surface carbohydrates and bacterial adhesion is a serious threat on any living cells which are in contact with the external environment because it gives access to the colonized bacteria to use the carbohydrate array decorating cell surfaces known as glycocalyx (Sharon 1987).

Many studies have been conducted to investigate the binding affinity of bacterial adhesins with glycans. Type 1 fimbriae (FimH) is one the most-studied adhesins that has affinity for high-mannose containing glycoproteins with multiple terminal α-D-mannosyl units such as

glycoproteins on the surface or urothelial cells which are easy targets for uropathogenic *E. coli* (UPEC) and cause urinary tract infections. About 90% of all known strains of UPEC contain Type 1 fimbriae which helps the bacterial infection through binding to oligomannoside residues on glycoprotein uroplakin Ia. In order to prevent this infection, efforts have been invested to progress an approach called antiadhesion therapy to potentially inhibit the binding ability of this adhesion (Martinez, Mulvey et al. 2000, Ofek, Hasty et al. 2003, Pieters 2007).

1.9 Structure of Plant and Animal Lectins

Plant lectins are generally known as soluble multivalent proteins with multiple glycan-binding sites which allows them to agglutinate cells. On the other hand, animal lectins have different structures and physical natures and are generally grouped into three main categories which are intrinsic membrane-bound glycoproteins, monovalent proteins and glycoproteins, and multivalent proteins and glycoproteins (Lis and Sharon 1986, Gabius and Gabius 1997).

The leguminous plants family is one of the main sources of plant lectins as it contains different type of lectins known as the leguminous lectins. These lectins have different biological activities but similar primary structures and the most studied examples among them are *Phaseolus vulgaris agglutinin* (PHA) from the red kidney bean, *Ricinus communis* agglutinin (RCA) from the castor bean and Concanavalin A (*Canavalia ensiformis*) from jack bean. The non-leguminous plants also contain many lectins with similar structures and carbohydrate binding specificities but their primary structure is completely different from leguminous lectins. The well-studied examples of non-leguminous lectins are mistletoe lectin (ML) from (*Viscum album* L.), tomato lectin from (*Lycoperiscon esculentum*) and jimsonweed agglutinin from (*Datura stramonium*) (Loris,

Hamelryck et al. 1998, Villacampa, Almolda et al. 2013, Nishimoto, Tanaka et al. 2014, Hagens, Klein et al. 2015).

Regarding sequence homology, animal lectins can also be classified into different groups. For example C-type family has a sequence motif that needs Ca^{+2} for their binding activity in contrast to the sequence motif of another family known as galectins that does not need Ca^{+2} for binding activity (Powell and Varki 1995, Gabius and Gabius 1997).

The majority of lectins used in research of animal cell glycoconjugates are plant lectins because they are commercially available and their preparation is cheaper than a high cost preparation of animal lectins (Drickamer and Taylor 1993, Unitt and Hornigold 2011).

1.10 Carbohydrate-Binding Specificities of Lectins

Plant lectins were historically classified according to their monosaccharide specificity as they were first shown to bind glycans through the inhibition of lectin-induced agglutination of cells by specific monosaccharides or their derivatives. For example, lectin from concanavalin A (Con A) and ricin (RCA) were classified as mannose and galactose-binding lectins respectively. Animal lectins can accordingly be classified because their binding can also be inhibited by monosaccharides. Thus, lectins like bovine heart galaectin-1 and asialoglycoprotein receptor from the hepatocyte were classified as galactose and N-acetylgalactosamine/galactose binding lectins respectively (Zhou and Cummings 1992).

Lectins bind to carbohydrates through different types of non-covalent bonds such as hydrophobic interactions, hydrogen bonds and metal coordination. The existence of a large number of hydroxyl groups on the surface of glycans provides lectins appropriate partners especially in the complex networks of hydrogen bonds where the hydroxyl groups function as both donor and acceptor (Bourne and Cambillau 1993, Weis and Drickamer 1996).

There are many factors that affect carbohydrate recognition by lectins; for example, cations such as Ca^{2+} and Mn^{2+} play a major role in some of these conjugations either directly via attaching to glycans or indirectly through building binding spots such as in the C-type lectin or legume lectin conjugations respectively. Despite the general hydrophilic nature of carbohydrates which can clearly be seen in the interaction of Gal-specific lectins with aromatic residues, hydrophobic interaction also strongly affects carbohydrates-lectins binding (Shaanan, Lis et al. 1991, Hamelryck, Loris et al. 1999, Cummings and McEver 2009).

While the multivalent nature of lectins probably gives them ultimate selectivity, primary monosaccharide specificity is also a major factor in the carbohydrate recognition by lectins and it is usually an indicator for a particular branched carbohydrate to be recognized by a specific lectin. Thus, complex carbohydrates bearing Glc or Man are bound by the Glc/Man-specific lectins (e.g. *Lathyrus ockrus* and Con A) but Gal-specific lectins (e.g. *Erythrina corallodendron*, EcorL) need the presence of Gal (or its amino-acetylated derivative) to perform binding (Drickamer 1992).

The cell-cell interactions critically need carbohydrates to be specifically recognized by lectins such as those required throughout lymphocyte-homing, embryogenesis, host-pathogen interactions, fertilization, and tissue development. Investigating lectin-carbohydrate interactions at the molecular level is thus important to broadly understand their intervention in biology and medicine (Sastry and Ezekowitz 1993, Beuth and Pulverer 1994, Bookbinder, Cheng et al. 1995).

1.11 Mistletoe Lectin (ML)

European mistletoe (*Viscum album* L.) grows as a semiparasite plant on deciduous and coniferous trees. Mistletoe extract has been focused upon by many researchers for a long time as a traditional remedy to treat patients with diabetes, cancer, hypertension, arthrosis and/or arthritis, and atherosclerosis (Hajto, Hostanska et al. 1989, Büssing, Suzart et al. 1996).

The immunomodulatory activities like induction of cytokine secretion and increasing the activity of natural killer (NK) cells alongside with cytotoxic effects of mistletoe extracts have also been suggested in the previous studies. Interestingly, the studies showed that these activities mainly come from the active lectins (MLs) in mistletoe extract. The lectins of mistletoe are classified and grouped into three isoforms: MLI (viscumin), MLII and MLIII (Eifler, Pfüller et al. 1994, Gamerith, Amann et al. 2014).

The main and major lectin isoform is ML-I which gives immunomodulatory potency to mistletoe lectin and it is a type II ribosome-inactivating proteins (RIPII). It also consists of two different subunits A (MLA) and B (MLB) which are bound together by a disulphide bond and both of them are *N*-glycosylated (Figure 7). The A-subunit (MLA) is a toxic part of MLI which contains a highly specific *N*-glycosidase which gives the subunit the ability to effectively halt protein synthesis in cells via modifying 28S rRNA of eukaryotic ribosome 60S subunit. On the other hand, the B-subunit of the lectin is responsible for the binding function and also increases the cytotoxicity of the toxic subunit through causing cell agglutination *in vitro* and facilitates cell internalization (Eck, Langer et al. 1999, Niwa, Tonevitsky et al. 2003). Critically, Hajto, Hostanska et al. (1989) showed that using A-subunit alone at the same concentration is completely inactive because it cannot bind to the cellular receptor glycoproteins and perform cell internalization by itself.

Despite similarity between the primary sequences of these three isoforms of mistletoe lectin, they have different carbohydrate binding specificities. The first and main isoform (MLI) was originally shown to have main affinity to D-galactose and 120 folds stronger than its affinity to *N*-acetylgalactosamine (GalNAc) (Wu, Chin et al. 1992). However, Müthing, Meisen et al. (2004) state that ML-I should be considered as sialic acid specific (not galactose-specific) because ML-I has higher affinity to sialylated residues than to galactosylated residues in glycoproteins. The second isoform (MLII) has equivalent affinity to both D-galactose and GalNAc whereas the third (MLIII) isoforms shows affinity only to GalNAc. The molecular weights of these isoforms are also different from each other as follow: both toxic and binding chains of MLI (A: 29kDA, B: 34kDA) are greater than the chains of MLII and MLIII (A: 27kDA, B: 32kDA) (A: 25kDA, B: 30kDA) respectively (Lee, Gabius et al. 1992, Wu, Chin et al. 1992, Tonevitsky, Agapov et al. 2004).

In order to evaluate the antineoplastic activity of mistletoe lectin, human lung carcinoma cells were treated with chemotherapeutic drugs such as doxorubicin and cisplatin alone and in combination with *Viscum album* agglutinin-1 (VAA-1). Stronger synergistic effects were observed in all drugs tested with mistletoe agglutinin (VAA-1) and it is believed that the inhibition of tumour protein synthesis through ribosomes inactivation by MLA is the source of antineoplastic activity of mistletoe lectin (Li 2002)

It has also been suggested to precisely study the anticancer activity of both MLI and MLIII isoforms especially after the observation of strong apoptotic activity from both isoforms in the treatment of leukemic B- and T- cells. Despite the fact that MLIII was less effective than MLI in the treatment of B-cells, both isoforms shown potential apoptotic effects in the treatment of T-cells (Büssing, Wagner et al. 1999).

In order to separately examine the therapeutic effects of these isoforms, designing recombinant protein from each isoform has been suggested. Eck, Langer et al. (1999) reported that the biochemical and biological activity of recombinant rMLIA proteins designed from the toxic MLA chain of MLI remained the same in comparison to the activity of the wild chains. Pevzner, Agapov et al. (2004) also tested the antigenicity of recombinant MLB chains designed from MLI and MLIII and they found that there is an antigenic epitope in the B-chain of MLI which is absent in the B-chain of MLIII. However, the same antigenic epitope was detected even in the recombinant B-chain of MLIII. The proper folding of carbohydrate-binding sites was also observed in recombinant MLIIIB via ELISA tests (Pevzner, Agapov et al. 2004).

Thus the binding chain of mistletoe lectin is an interesting candidate as a therapeutic targeting peptide.

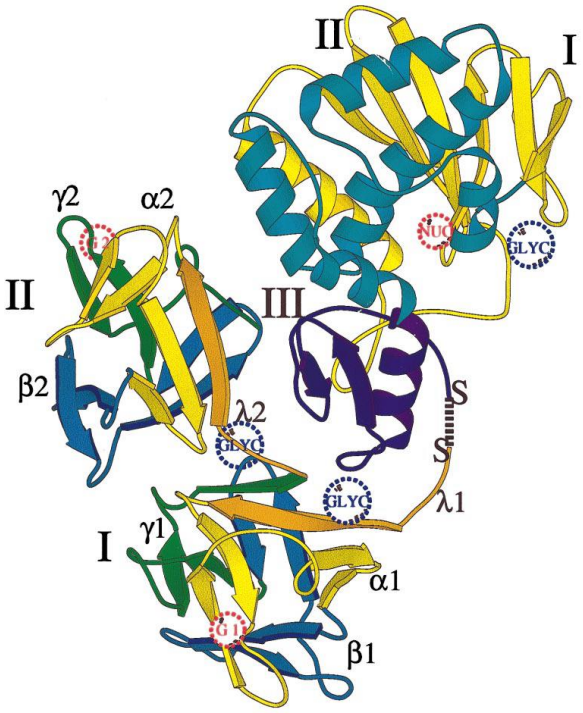


Figure 7. Cartoon plot showing the three domains for the A chain labeled I, II, and III; colored yellow, turquoise and violet. For the B chain domain I and II are colored according to their subdomains: The linker regions l1 and l2 are shown in orange, the homologous subdomains a, b, g are colored yellow, blue, and green. The disulfide bond connecting the two chains is shown as a blue bold dashed line and labeled. Dashed circles indicate the nucleotide binding site ("NUC") for the A chain, the low ("G1") and high affinity ("G2") galactose binding sites in chain B and the three glycosylation sites clearly identified in the electron density (Krauspenhaar, Eschenburg et al. 1999).

1.12 Elderberry Lectin (SNA)

Elderberry lectin (SNA) is a member of carbohydrate-binding proteins which has been extensively studied in the last few decades. The SNA lectin like other type of lectins contains different isoforms with different activities but the main and well-studied isoform is isoform I (SNA-I). The discovery of SNA-I and its carbohydrate affinity has made the lectin to become an important tool in the molecular biology, biochemical and physiological studies. SNA-I belongs to type 2 ribosome-inactivating proteins (RIPII) with strong cytotoxic activity similar to ricin RIPII (Girbes, Ferreras et al. 2004, Stirpe 2004). It is also sold commercially and may be used in cancer detection (Sigma Aldrich, UK).

The SNA-I lectin is composed of two main parts which are a toxic RIPII chain (A chain) and a binding chain (B chain). The binding chain possesses a strong and specific affinity toward NeuAc(α2-6)Gal/GalNAc glycoconjugates. Therefore, it has been widely used in the study of glycoconjugate especially conjugates contain sialic acid residues (Shibuya, Goldstein et al. 1987).

While many studies have been conducted about elderberry lectins, their biological function has remained unclear except in the plant which is believed to be defence against insects and viruses. However, a high level of RIPII in SNA-I lectin was shown to have a harmful toxic effect on animal physiology through the activation of cellular signal of gut mucosa which was detected in the animals fed with a high amount of the elderberry fruit (Tejero, Jiménez et al. 2015).

On the other hand, the cytotoxicity and binding properties of elderberry lectin (SNA-I) have been used as a powerful tool in cancer research. The lectin has been suggested to be used as an agent to detect of Neu5Ac(α-2,6) terminal residues on cancer cells. For example, SNA staining intensity on colorectal cells can be used as an indicator of cancer

recurrence because previous studies have proven a relationship between colon cancer and the induction of particular sialyltranferase on their cell surfaces (Fernández-Rodríguez, Feijoo-Carnero et al. 2000, Dall'Olio, Malagolini et al. 2014).

1.13 Snowdrop Lectin (GNA)

The bulb of snow drops (*Galanthus nivalis*) contains a lectin called *Galanthus nivalis* agglutinin (GNA) which has strong binding affinity for mannose. The structural analysis has shown that the lectin is formed of similar 12.5kDa subunits that initially come from post translationally modified bigger subunits 15kDa synthesized on the rough endoplasmic reticulum (ER) (Van Damme, Allen et al. 1988).

Functionally, GNA is considered as tetrameric protein with strong binding affinity to α-D-mannose carbohydrates. It is also known as a resistant protein to proteolytic activity in which has made the lectin a focus for genetic modification of crops plants with the GNA gene for insect protection and as an appropriate carrier to transport other toxic peptides to the insect gut (Fitches, Audsley et al. 2002).

While the GNA mechanism of action to exploit its cytotoxicity against insects is still unknown, it is believed that the binding of GNA to the glycoproteins of gut epithelium is the key point in GNA activity (Fitches, Woodhouse et al. 2001). In the earlier studies different fusion proteins were produced from GNA *C*-terminally fused to different peptides such as GFP and *Manduca sexta* allatostatin (Manse-AS) which were orally fed to the haemolymph of *L. oleracea* larvae and confirmed that GNA can efficiently deliver biological peptides to the gut of larvae (Raemaekers 2000, Fitches, Audsley et al. 2002). Other studies have also been carried out to exploit the activities of GNA lectin in crop protection through genetic manipulation since the lectin is not harmful to mammals

(Pusztai, Grant et al. 1993, Rao, Rathore et al. 1998, Stoger, Williams et al. 1999, Nakasu, Edwards et al. 2014).

These studies in insect protection demonstrate that a lectin can effectively deliver a toxic peptide across biological membranes. Thus could a similar strategy be used to target and kill cancer cells.

1.14 Glycosylation and Cancer

The focus on the cancer relationship with proteins, DNA and RNA has been driven to a high level and it has undoubtedly revealed crucial information about this disease. However, this lead to glycans (and lipids) being margined for sometimes because it was believed that they are not as related as the aforementioned molecules to cancer. However, recent modifications in proteomic methodologies and modern technologies help advances in glycomic and lipidomic researches and allow extensive studies about the role of these two molecules in cancer progression. To date, all the studies carried out about glycosylation changes in cancer confirm association between altered glycosylation and cancer development and progression (Igl, Polašek et al. 2011).

Changes in cancer cell glycosylation simply means changes in the structure of glycans attached to the cell surface glycoproteins causes by altered expression and function of glycosidases and glycosyltransferases which may finally change the type of proteins that interact with the cell surface glycoproteins and modulate cell signalling and adhesion (Zou 2005).

Abnormal glycosylation is in fact a universal characteristic of cancer cells because it frequently happens in almost all the types of cancer cells and has vital roles in the cell-cell interaction, cancer progression and metastasis (Chen, Jan et al. 2013, Li, Xu et al. 2014).

This aberration was first observed when abnormal glycan structures were detected by antibodies used against cancer cells (Chen, LaRoche et al. 2007). While the mechanism of producing these abnormal glycans is still not clear and the direct relationship between cancer initiation and altered glycosylation has not yet been discovered, the contribution of these alterations in cancer development from the early stage to metastasis has been well established (Fuster and Esko 2005).

The aberrant glycosylation of glycoproteins on cancer cells has been specifically focused on by researchers especially the mucins which are highly glycosylated by GalNAc on serine or threonine (*O*-glycosylation) and contains majority of oncogenic related glycan structures. Despite the variety of changes in glycan structures, two changes are well documented through biomarker studies which are the formation of truncated version of oligosaccharides and the formation of abnormal terminal structures such as sialylation of normal counterparts.It appeared that most of these changes come from glycosyltransferases up or down regulation; therefore, a study was carried out to see the consequences of the absence of a functional enzyme such as Cosmc chaperone protein through performing mutation in its sequence. The mutation was shown to cause the accumulation of cancer-associated precursors (Ju, Lanneau et al. 2008). Thus, it is sensible to consider that non-clonal signature of cell surface glycans is unstable because of a wide range of genetic origin clonal alteration associated with cancer and their expression mosaicism in tissue sections reflect differentiation and variation during cancer development (Nakasaki, Mitomi et al. 1989). Although, cancer cell glycosylation goes through a wide range of variations during cancer progression because of their non-clonal nature, it finally stabilizes through achieving positive properties of the invaded cell populations. Therefore, these glycan alterations can be considered as cancer biomarkers and to be detected via using lectins and antibodies as they are visible on the cell surfaces (Hollingsworth and Swanson 2004).

1.15 Altered Glycosylation in Melanoma Cancer

One of the most fatal malignancies which has broadly increased in the last few decades is melanoma cancer. Melanoma is well known for being unnoticeable during the early stages of progression and becomes metastatic in a very short period of time. To progress and reach metastasis, melanoma is associated with the alterations in the enzymatic activity of some specific glycosyltransferases responsible for the glycosylation of cell surface glycoproteins. These alterations may include changes in the structure of *N*-glycans of the glycoproteins and consequently changes the functionality of the proteins especially those responsible for metastasis (Hakomori 2002). Sialic acid is one of major terminal residues in the cell surface glycans with its properties related to the invasion and malignancy of melanoma tumours, by helping the malignant melanocytes escape from the attachment to the matrix components (Due to changed properties) and thus helps them to easily migrate and interact with platelets and leukocytes through association with selectins and finally facilitate metastasis (Rambaruth and Dwek 2011). Despite the increased amount of sialic acid linkages in melanoma which is often with *N*-acetyllactosamine, many modifications were also detected in their linkages. Increased sialylation of 9-O-acetylation is also detected in melanoma and has become a biomarker for human melanoma cells (Cheresh, Reisfeld et al. 1984, Kohla, Stockfleth et al. 2002).

To screen the changes in glycosylation profile of melanoma cells during the stages of cancer development, *N*-glycans of L1CAM oligosaccharides were precisely studied at different stages of progression. The oligosaccharides were found highly sialilated and alter their glycosylation profile during progression stages form vertical growth phase (VGP) to metastasis. The L1CAM oligosaccharides of melanoma cells at primary stage were found to have both α2-3- and α2-6-linked sialic acids whereas the oligosaccharides of metastatic cells loss α2-3-linked sialic acids and only keep α2-6-linked sialic acids. Therefore, it is

suggested to target specific *N*-glycans of melanoma cells and recognise them as new biomarkers to specifically target the cancer cells (Hoja-Łukowicz, Link-Lenczowski et al. 2013).

1.16 Cancer Therapy

The major drawback of conventional anticancer therapeutics (Those used in conventional chemotherapy) is a lack of specificity that leads to a poor therapeutic index and causes substantial toxicity not just for cancerous tissues but also for the healthy tissues. On the other hand, a high level of specificity may be achieved by using monoclonal antibodies an interest in which has considerably increased recently as well as their applications in cancer medicine. Although, the therapeutic potential of monoclonal antibody may be limited by production time and cost, the antibody-drug conjugation is considered as a potential approach for cancer therapy as the systemic cytotoxicity of the drug is reduced and cancer patients can achieve higher therapeutic benefits from targeted anticancer therapies (Alley, Okeley et al. 2010, Casi and Neri 2012). Table 2 shows different strategies which have been employed for drug targeting.

Table 2. A variety of strategies, previously employed for drug targeting.

Ligand	Target molecule (cell)	MR kDa size (nM)	Immunogenicity	Clinical status	Costs of production
Antibody fragments					
Fab	eg, NCA-90 (granulocyte), CEA (apical surface of gastrointestinal epithelium, lung tissues, breast, and colorectal cancer), VEGF (breast, colon, lung, gastric, renal, and oropharyngeal cancers), HER2 (breast, ovarian, stomach cancer)	50 (5)	Lower than mAb	Approved by FDA: certolizumab pegol, CEA-scan	$615/40 mg
scFv	eg, CEA, HER2	30 (3)	Lower than	Preclinical trial	Low

Monoclonal antibodies (mAb)			Fab		
Rituximab	CD20 (pre-B and B-cell)	145 (15)	*	In clinical use	High cost of final product, about $2,000–$20,000/g
Trastuzumab	HER2	145 (15)	**		
Bevacizumab	VEGF	145 (15)	***		
Alemtuzumab	CD52 (lymphocytes, especially T-cells, monocytes, macrophages, monocyte-derived dendritic cells [moDCs], and the epithelial cells of the distal epididymis)	145 (15)	****		
Panitumumab	EGFR (normal cells and non-small-cell lung cancer [NSCLC], breast, head and neck [squamous cell carcinoma of head and neck], gastric, colorectal, prostate, bladder, pancreatic, ovarian, and renal cancers)	145 (15)	®	In clinical use	High cost: $2,000–$20,000/g
Proteins and peptides					
RGD	Integrins $\alpha_V\beta_3$ and $\alpha_V\beta_5$ (overexpressed on tumor endothelium)	1–5 (1–2)	High	Clinical trials	Low
DARPins	eg, CD4 (T helper cells, monocytes, macrophages, and dendritic cells), HER2.	14–20 (5–10)	High	Phase I/II clinical trials	Low
DNA or RNA oligonucleotides					
Aptamers	proteins, surface receptors	8–13 (3–5)	Low	FDA approved Macugen	Low
Other targeting molecules					
Folates	folate receptors: RFC (all cells), FR (ovarian, brain, head and neck, renal, and breast cancers)	0.44 (0.3)	Low	Yes	Low
Lectins	Lectins receptor: DC-SIGN (dendritic cell), CLR (cancer)	10–200 (2–20)	⊘	Not in clinical use	Low
Transferrin	CD71 (present on all cells, but overexpressed on cancer cells)	80 (5–10)	Low	Not in clinical use	Low

Notes: Size–length in the longitudinal section;

[*]low: 11% positive in HACA (human anti-chimeric antibody) test;
[**]very low: 0.11% positive in HAHA (human anti-human antibody) test (Genentech, Inc., South San Francisco, CA, USA);
[***]very low: 0.63% tested patients positive for treatment-emergent anti-bevacizumab antibodies;
[****]anti-alemtuzumab antibodies were detected in 80.2% of alemtuzumab-treated patients. Titers generally increased during first 3 months of each course, declined by month 12. At month 12, 29.3% of patients remained positive for anti-alemtuzumab antibodies; [ⅇ]low: 2% patients developed binding and neutralizing antibodies;[®]some are potent toxins. Costs are in US$.

Abbreviations: Av, average; CEA, carcino-embryonic antigen; CLR, C-type lectin receptor; DC-SIGN, dendritic cell-specific intercellular adhesion molecule-3-grabbing non-integrin; EGFR, endothelial growth factor receptor; FR, folate receptor; Fv, variable fragments; Fab, antigen-binding fragments; KD, equilibrium dissociation constant; MR, mannose receptor; RFC, reduced folic carrier; mAbs, monoclonal antibodies; scFv, single chain variable fragments; VEGF, vascular endothelial growth factor.

In theory antibody-drug conjugates (ADCs) are simple but practically it may be a considerable challenge. The development of an effective ADC needs many properties to be achieved and optimized such as retention of immunoreactivity, stability (the drug should be protected and attached to the conjugate until it reaches the targeted site *in vivo*), pharmacokinetics (preventing the accumulation of the drug in other non-targeted organs) and finally the drug should be successfully released on the targeted site. However, these parameters were not considered in developing early ADCs which were mainly relying on the existent chemotherapy agents with known cytotoxic profiles and the main target was just increasing tumour specificity of the agents. An example of the very first anti-cancer ADCs is BR96-doxorubicinis which is a combination of BR96 Lewis-Y specific (chimeric) monoclonal antibody attached to doxorubicin chemotherapy through a hydrazone linker (Trail, Willner et al. 1993).

In the development of a new generation of ADCs other requirements have also been taken in to consideration such as designing conjugates from tumour specific monoclonal antibodies (mAbs) with appropriate formats that help tuning the conjugate uptake by tumours and also using appropriate linker to properly manage the drug protection and

release. Thus, the cross-reactivity of the ADCs and their cytotoxicity to healthy tissues and organs have been minimized (Polson, Ho et al. 2011).

While some successful ADCs have been developed and used in cancer therapy, the development of next generation ADCs needs to find more effective drugs to be bound with mAbs. The current ADCs have limited capacity because of the limited amount of drugs can be realistically administrated to patients. The reasons behind this limitation are the high cost of the therapy and the limited amount of drug that can be loaded on the mAbs because loading extra amount of drug often results in loss of potency. Moreover, the amount of drugs that reach malignancies is just a small amount of the injected does (maximum 0.08%/g of tumour) because of the environment of tumours which has a high level of interstitial fluid pressure (IFP) plays as a barrier for diffusion and consequently results in a poor localization of mAbs (Hamblett, Senter et al. 2004, Heldin, Rubin et al. 2004). Therefore, it is important to discover more potent bioactive molecules and use them as a carrier vehicle like mAbs to develop better tumour targeting conjugates. As seen in table 2 lectins are an explored but uncommercialised possibility.

1.17 Drug Targeting

Delivery of a specific drug to a specific location needs the production of targeted drug delivery systems (DDS) and one of the efficient and selective types of DDSs is known as prodrug. A prodrug is a drug conjugate in inactive form until it gets to the target site and becomes active via a specific condition within the site. It typically works in two different manners: first, drug conjugate breaks down within the target site to one or two therapeutic agents; second, the conjugate reacts with two or more compounds within the target site so as to produce therapeutic agents in a special intracellular condition (Minko 2004, Rek, Krenn et al. 2009).

The typical selective drug delivery systems (DDSs) contains three main components which are targeting moiety, drug and carrier. All the components are bound together by the carrier which also increases the solubility of the system. As well as, two crucial features should be found in any DDS system: first, the drug should be protected from degradation by the delivery moiety until it reaches the target site. Second, separation of the drug from the conjugate should be assured within the target site. To achieve these features especially the second one, the therapeutic agent should be attached to the delivery conjugate through specific type of linkages such as a tetrapeptide Gly-Phe-Leu-Gly biodegradable linker which can be cleaved by cathepsin B enzyme and liberate the agent at a specific site (Kopeček, Kopečková et al. 2001, Lu, Shiah et al. 2002). It is also possible to make the whole DDSs moiety biodegradable and degrades within the target site such as the conjugation between indole-3-acetic acid and horseradish peroxidase that changes to a cytotoxic agent through the degradation of the conjugate because of the radical cations results from the oxidization of indole-3-acetic acid by horseradish peroxidase (Minko 2004).

1.18 Lectin-mediated Drug Delivery

The reduced efficacy of drugs delivered through traditional non-targeted therapeutics and their numerous side effects on healthy tissues (Which hence reduce drug dosage and delay treatments) provide the innovation to develop an alternative way to deliver drugs. Therefore, selective drug delivery systems (DDS) became a strong research focus with the promise of potential to specifically deliver drugs to target sites and at the same time reduces the side effects of the drug on the normal tissues. It also increases the efficacy of therapy through increasing exposure between the drug and the target site. In fact lectins

were originally one of the candidates suggested by Woodley and Naisbett (1988) to be used as a targeting moiety in designing DDS (Bies, Lehr et al. 2004).

During designing functional DDSs, identification of specific moieties on target site to be utilized by the drug is crucial. However, there are extra parameters that should be taken into considerations in lectin based drug delivery systems such as the environment of transporting path and the target site itself (Minko 2004, Rek, Krenn et al. 2009). The acidic environment of the gastrointestinal tract and pancreatic enzymes are good examples that cause degradation and early activation of drugs. On the other hand, catabolic assault from bacterial origin enzymes such as β-D-galactosidase and β-D-glucosidase can also damage drugs delivered through colon. However, advantages can be taken from these bacterial enzymes through developing DDSs with ability to utilize the enzymes such as drug-carbohydrate conjugates (Prodrug) (Guarner and Malagelada 2003, Rek, Krenn et al. 2009).

1.19 Lectin-based Drug Targeting

The logic behind using lectins in drug delivery is the simple fact that more than half of cell membrane proteins and lipids are heavily glycosylated with glycans which are targeting sites for lectins. The expression pattern of glycans change from one cell type to another and a range of different glycan array are expressed by different cells. For instance, glycans expressed by carcinogenic cells are different than those expressed by their healthy counterparts. Thus, lectins with affinity to these glycan arrays can be used to deliver drugs to specific cells and tissues (Haltner, Easson et al. 1997).

There are two main mechanisms to accomplish lectin-based drug targeting: direct and reverse lectin targeting. In direct lectin targeting, endogenous cell surface lectins recognize carbohydrate moieties of DDS but in reverse targeting endogenous carbohydrate moieties

on glycoproteins and glycolipids are recognized by the exogenous lectins of DDS (Bies, Lehr et al. 2004, Plattner, Ratzinger et al. 2009). Furthermore, cytotoxicity and apoptosis induction are two other features of lectin based DDSs which increase the potential of the system for cancer treatments (Thies, Nuge et al. 2005, Plattner, Ratzinger GEngleder et al. 2009). In order to exploit their cytotoxicity, lectins can be used in two different approaches. First, a non-toxic lectin can be attached to a therapy that will activate and become toxic within the target site. Second, a toxic lectin can be used to act as both toxic agent and homing moiety through apoptosis induction (Minko 2004, Heinrich, Welty et al. 2005). However, difficulty in finding lectins with selective cytotoxicity towards specific sites is a limitation of the second approach (Ghazarian, Idoni et al. 2011).

Mistletoe lectin (MLI) is one of the potent anticancer candidates with strong cytotoxicity that causes protein synthesis inhibition in cells (Mody, antaram Joshi et al. 1995). The total extract of mistletoe containing all three lectins (MLI, MLII, MLIII) has been widely used in cancer research and many positive results have been achieved especially in breast cancer treatments. However, in the treatment of MV3 melanoma cancer cells with MLI, MLII and MIII separately, isoform I (MLI) in particular showed more cytotoxicity than other isoforms (MLII and MLIII). Therefore, MLI could be a good candidate for the second approach of lectin based drug targeting (Thies, Nugel et al. 2005).

1.20 Cathepsin B Protease and Melanoma Cancer

Cathepsin is a large protease family found in a well-defined organelles of cells and possess unique structure and substrate specificity (Guicciardi, Leist et al. 2004). The family has at least twelve members of different structures of serine-, cysteine- and aspartyl- which are in an inactive proenzyme state inside cells. However, they transform to active state once they come out of cells and enhance the degradation of the components of the basement

membrane and extracellular matrix (Szpaderska and Frankfater 2001, Vasiljeva, Reinheckel et al. 2007, Vasiljeva and Turk 2008). Thus, the metastatic potential of cancer cells can be determined via assessing the proteolytic activity of cathepsin family members (Podgorski and Sloane 2003).

The main activities of aspartyl- or cysteinyl- proteases are the degradation of extracellular matrix and removing physical barriers that limit cell movement and facilitate cell spreading and invasiveness (Jedeszko and Sloane 2004). Different level of cellular activities are possessed by different types of cathepsins but cathepsin B, D, K and L are shown to be the most active ones and play crucial enzymatic roles (Guzińska-Ustymowicz, Zalewski et al. 2004, Lankelma, Voorend et al. 2010).

In the studies carried out to investigate the relationship between cathepsins expression and the progress of metastatic melanoma, direct association was found between the overexpression of cathepsins and rapid melanoma distribution (Roshy, Sloane et al. 2003, Murata, Yatsuda et al. 2010). This association was later observed at the highest rate between cathepsin B and the invasion potential of melanoma cells in a study conducted to modulate metastasis of melanoma cells through using inhibitors of cathepsin B, D and L. The flow cytometry analysis in the study showed that the level of cathepsin B expressed on the cell surfaces was much higher in metastatic melanoma cells than primary melanoma cells, while, comparable amount of cathepsin D and variable amount of cathepsin L were detected on both cells. Real time quantitative RT-PCR was also performed to confirm these results and similar results were achieved as the rate of procathepsin B transcripts was found significantly higher in metastatic melanoma cell lines than primary melanoma cell lines (Matarrese, Ascione et al. 2010).

Thus we can utilise a cathepsin B cleavable linker sequence between a lectin carrier and a cytotoxic peptide, releasing the toxin from the carrier specifically at the cancer cell.

1.21 Cytotoxic Peptides

A wide range of cytotoxic peptides are known and thus this study will focus on only two (Melittin and Mastoparan) due to previous studies conducted with these peptides in anti-cancer research.

1.21.1 Melittin

The cytotoxicity of bee (*Apis mellifera*) venom comes from only 12% of its whole volume which consist of 100ng dry venom and 88% is water. Melittin is the most abundant haemolytic and cardio-toxic peptide in bee venom occupies 50% of the active part (Shaposhnikova, Egorova et al. 1997, Hussein, Nabil et al. 2001).

Melittin is a 26-amino acid (GIGAVLKVLTTGLPALISWIKRKRQQ) cationic peptide with amphipathic features which gives the peptide an ability to interact and damage biological membranes (Habermann 1972, Dempsey 1990). The disruption of cell membrane phospholipid bilayer is believed to be the main cytotoxic effect of melittin on cells that leads to the formation of transmembrane pores (Raghuraman and Chattopadhyay 2007).

Promising cytotoxic activities against cancerous cells were observed from bee venom in general and melittin in particular. In order to assess these activities both *in vitro* and *in vivo* experiments were conducted. Their *in vivo* applications are limited due to the side effects and hemolysis (Liu, Yu et al. 2008, Hoshino, Urakami et al. 2009). The *in vitro* studies showed that bee venom significantly inhibits cancer cell proliferation and clonogenicity.

For instance, treating mouse melanoma cells (K1735M2) with bee venom was seen to positively supress cell growth and migration (Liu, Chen et al. 2002). Significant cell growth suppression, increased apoptosis, and reduced metastasis were also observed through biochemical and morphological analysis of lung cancer cells (NCl-H1299) treated with bee venom (Jang, Shin et al. 2003, Oršolić, Šver et al. 2003). Similarly, the inhibition of tumour growth by the venom was seen in the *in vivo* tests carried out by Liu, Chen et al. (2002) on solid melanoma tumours of mouse .

The haemolytic nature of melittin is the restriction of its systematic administration as anticancer agent. Therefore, recent studies have focused on attaching nanoparticles to the peptide in order to protect it from hemolysis and be able to systematically deliver it to cancer cells (Zetterberg, Reijmar et al. 2011). Yet, the only successful nano carrier is perfluorocarbon nanoemulsion which has been designed by Soman, Baldwin et al. (2009) from lipid monolayer to safely deliver melittin without causing hemolysis. Through this carrier and for the first time melittin was applied via intravenous administration and was seen to inhibit tumour growth. However, the big size of the melittin-nanoparticle molecule (~270 nm) is an obstacle and make it hard to be used with solid tumours because it cannot penetrate them (Minchinton and Tannock 2006, Cabral, Matsumoto et al. 2011).

1.21.2 Mastoparan

The term mastoparan (MP) describes a 14-amino acid (INLKALAALAKKIL) cationic and amphipathic peptide derived from wasp (*Vespula lewisii*) venom (Hirai, Yasuhara et al. 1981). Mastoparan as a major component of wasp venom which has been extensively studied in the last three decades and important cytotoxic activities have been demonstrated from their natural and synthetic analogous (Higashijima, Burnier et al. 1990, Soomets, Haellbrink et al. 1997). Apart from antimicrobial activity, MP was found to induce

phospholipase A2 and C activation and perform cytotoxic activity in variety of cell types such as platelets, mast cells, and pancreatic islet β-cells (Hirai, Yasuhara et al. 1981, Argiolas and Pisano 1983, Yokokawa, Komatsu et al. 1989, Ozaki, Matsumoto et al. 1990).

Both natural and synthetic analogues of MP share similar activity which is binding to heterotrimeric G proteins especially Gi and Go and regulate them. In order to, perform this binding with G proteins or any other intracellular proteins, MP needs to cross cell membrane assisted in doing so by its amphipathic nature. The presence of lipid and their interaction with Gi/Go helps MP to adopt α-helical conformation, while it has disordered structure in aqueous solutions. The amino terminus of MP sequence if the hydrophilic face of MP and the three positively charged lysyl residues in the MP sequence is thought to interact with hetero- trimeric G proteins especially Gi and Go and stimulate the activity of GTPase which in turn catalyses the exchange of GDP/GTP (Weingarten, Ransnäs et al. 1990, Shpakov and Pertseva 2006).

More recently the anti-tumour activity of mastoparan and its analogues has also been demonstrated using different cancer cell lines such as human erythroleukemia (K562) and the results showed that MP cytotoxicity was based on membrane disruption (Yamada, Shinohara et al. 2005). Moreover, a significant difference was observed between treating normal and melanoma cells with mastoparan as the cytotoxicity of MP was attenuated in normal cells in comparison to the activity in cancer cells. The further analysis for this difference revealed that MP induces ROS (Reactive oxygen species) dependent oxidative stress which affects cancer cells more than normal cells because cancer cells are more dependent on their antioxidative systems which can be inhibited by MP and leads to severe ROS accumulation and cell death. However, in normal cells the basal output of ROS is low

and less oxidative stress occurs when they are treated with mastoparan (Irani, Xia et al. 1997, Sattler, Verma et al. 2000, Trachootham, Zhou et al. 2006).

Other studies showed that mastoparan induces the apoptosis of melanoma cells through the mitochondrial intrinsic pathway since the altered morphological properties of mitochondria confirmed that MP as a cell penetrating peptide can cross mitochondria membrane and leads to cell death. Therefore, mastoparan is suggested to be used as anticancer agent (Park, Lee et al. 2000, Armstrong 2006, Rocha, de Barros et al. 2010).

Thus the objective of this study is to:

1- Produce recombinant fusion proteins from the binding chain (MLB) of mistletoe lectin isoform I, the binding chain of elderberry lectin (SNA-B) and snowdrop lectin (GNA) fused to green fluorescent protein (GFP) as a fluorescent marker and express them in *E. coli* in order to demonstrate the glycan binding specificity of the lectins.
2- Generate a range of Carrier:Cytotoxic peptide fusion constructs and express in *E. coli* in order to exploit the glycan binding specificity of the binding chain of mistletoe lectin (MLB) to deliver different cytotoxic peptides specifically to cancer cells.
3- Screen the produced recombinant fusion proteins against different cancer cell lines, including primary and metastatic melanoma,so as to confirm their binding to cell membranes, internalisation and cytotoxic activity.

Chapter 2. Amplification and Cloning of MLA and MLB Chains of Mistletoe Lectin Isoform One (MLI)

2.1 Introduction

Mistletoe lectin (ML) is as a toxic glycoprotein (Lectin) derived from mistletoe (*Viscum album*) and a well-known representative of type II ribosome inactivation proteins (RIPII) with strong cytotoxic activity similar to ricin and abrin toxic glycoproteins (Barbieri, Battelli et al. 1993, Tonevitsky, Agapov et al. 1996).

There are three isoforms among mistletoe lectin (MLI, MLII, MLIII) which are highly similar in their molecular weights but they are different in their amino acid sequences and glycan binding affinities (Urech, Schaller et al. 2006).

The main and most abundant isoform among the three isoforms of mistletoe lectin is isoform one (MLI) which has been structurally and functionally studied in comparison with ricin lectin. The isoform is composed of two main chains with completely different functions. The first chain is a toxic chain (MLA) which is a 29kDa potent *N*-glycosidase and belongs to RIPII with the ability to disrupt protein synthesis through the modification of 28s RNA and it is the source of cytotoxicity of mistletoe lectin. The second chain is a binding chain (MLB) which is a 34kDa peptide folded to two homologous potential galactose binding domains and contain several glycan binding sites and is responsible for the binding property of mistletoe lectin (Krauspenhaar, Eschenburg et al. 1999).

The homology between MLA chain of MLI and the toxic chain of ricin is 41% which is less than the significant 64% homology between the MLB chain of MLI and the binding chain of ricin. The two chains of MLI are bound to each other through a disulphide bond

between Cys 260A and Cys 5B and to activate the enzymatic activity of MLA chain this disulphide bond needs to be reduced so as to let the chain to be released from MLB chain (Tonevitsky, Agapov et al. 1996, Niwa, Tonevitsky et al. 2003).

The glycan binding property of mistletoe lectin gives the lectin an ability to potentially bind to the living cell through cell surface glycans. While MLB chain of MLI binds to cell surface glycans, it also binds to the cell surface receptors which facilitates the internalization of the toxic MLA chain through triggering the endocytotic uptake of cells (Vervecken, Kleff et al. 2000). Recent work (Müthing, Meisen et al. 2004) has suggested that MLB is not actually galactose binding, but is in fact specific for sialic acid.

As cancer cells are known to highly express sialic acid residues at the cell surface (Hoja-Łukowicz, Link-Lenczowski et al. 2013, Kolasińska, Przybyło et al. 2016), the B-chain of mistletoe lectin may represent a potential cancer specific binding peptide. In an attempt to produce recombinant fusion proteins from the cytotoxic A-chain and the carbohydrate B-chain of mistletoe lectin isoform I (MLI), a heterodimeric protein known as rViscumin was produced. The binding activity of rViscumin was later assessed through a solid phase binding assays and confirmed that the protein has a preferential binding to terminally α2-6-sialylated neolacto-series gangliosides IV6Neu5Ac-nLc4Cer, VI6Neu5Ac-nLc6Cer, and VIII6Neu5Ac-nLc8Cer isolated from human granulocytes. The cytotoxicity of rViscumin was also assessed with human bladder carcinoma 5637 cells and was shown to have IC50 value (half maximum cytotoxicity) of 2.1pM (Müthing, Burg et al. 2002).

Trefzer, Gutzmer et al. (2014) also produced a recombinant protein from mistletoe lectin known as Aviscumine which was shown to enhance T-cell responses and cytokine release through inducing ribotoxic stress at the 28S ribosomal RNA subunit. Aviscumine has been

tested at phase II trial in patients with systemically pre-treated metastatic melanoma stage IV and significant results have been obtained.

Thus the aim of this study is the amplification, cloning and sequencing the toxic chain (MLA) and the binding chain (MLB) of mistletoe lectin isoform I (MLI) from source and analysis of their coding sequences.

A bioinformatics analysis of lectin sequences was also performed in order to design a PCR based screen in other tissues. However, homology between lectin sequences was poor.

2.2 Materials and Methods

2.2.1 Bioinformatics Analysis

Bioinformatic analysis was conducted to identify novel C-type lectins from plant and animal sources (C-lectins are identified in literature as binding molecules to cancer cells) (Yau, Dan et al. 2015).

NCBI (http://www.ncbi.nlm.nih.gov/) was accessed and searched for C-lectins natural resources and the grape *Vitis vinifer* C-lectin was then used as the query sequence to align 14 sequences using Clustal W2 (http://www.ebi.ac.uk/Tools/msa/clustalw2/) (Figure 8). The complete alignment result is shown in appendix 1.

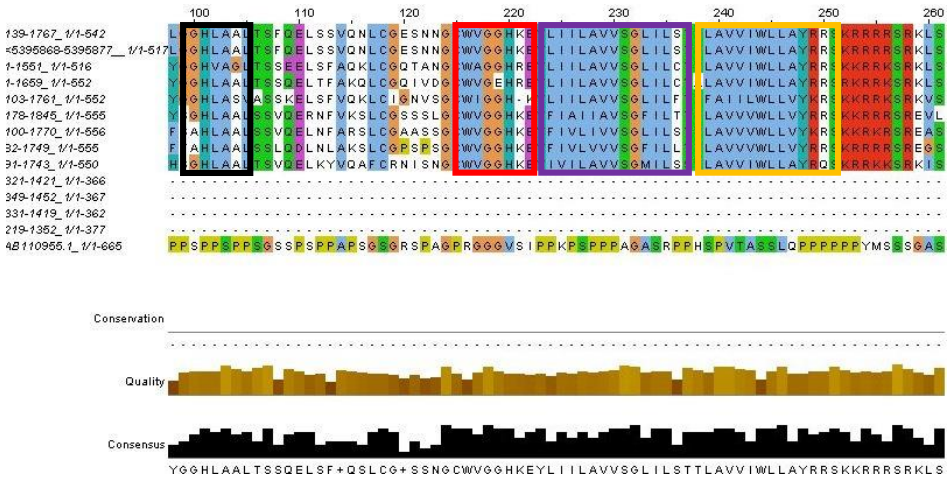


Figure 8. Snapshot of conserved regions in the multiple sequence alignment of C-lectin from different sources.

Degenerate primers were designed corresponding to ((LGGHLA (black box), GCWVGGH (red box), EYLIILAVVSGLIL (purple box) and TTLAVVIWLLAYR (orange box))) and then used to amplify the conserved regions from cDNA generated from a range of plant and invertebrate tissues.

2.2.2 Sample Preparations

Nucleic acids were extracted from fresh leaf tissue of the Mistletoe plant *Viscum album*. Ten leaves were frozen in liquid nitrogen and grounded to a fine powder with mortar and pestle. They were then divided to 100mg samples per eppendorf tube and stored at -20^{o}C.

2.2.3 RNA Extractions

RNA was extracted using ISOLATE II RNA Plant Kit (Bioline, UK). Approximately 100mg of tissue was lysed using 350µl Lysis Buffer RLY (guanidium thiocyanate and guanidium HCl) and 3.5µl ß-mercaptoethanol. The rest of the protocol was performed according to the manufacturer's instructions. The eluted RNA was then purified further using the Bioline silica column, DNase digestion as part of the RNA extraction kit was carried out according to the manufacturer's instructions. RNA was eluted in RNase free water and stored at -80°C until further use.

2.2.4 Agarose gel Electrophoresis Containing Formaldehyde

Sambrook and Russell's protocol (Sambrook and Russel 2006) was followed to perform agarose gel electrophoresis containing formaldehyde to separate RNA according to size. The RNA was denatured in a sterile RNAse-free microcentrifuge tube by mixing the following: 2µl of RNA (up to 20µg), 2µl of 10X MOPS electrophoresis buffer, 4µl of formaldehyde and 10µl of formamide. The RNA solution was incubated at 55C^{o} for 60min and chilled immediately followed by centrifugation for 5 seconds to deposit all of the liquid to the bottom of the microfuge tube. Finally, 2µl of formaldehyde gel-loading buffer (50% (v/v) glycerol, 0.25% (w/v) bromophenol blue, 0.25% (w/v) xylene cyanol FF and 10 mM EDTA pH 8.0) was added to the sample and returned to the ice bucket.

Agarose gel (1.5%) containing 2.2M formaldehyde was used to resolve RNAs. 1.5g of agarose was dissolved in 72ml H_2O and the solution was cooled to 55^oC and then 10ml of 10X MOPS electrophoresis buffer and 18ml of deionized formamide were added. The gel was allowed to set for at least 1 hour at room temperature then covered with saran wrap until the sample was ready to be loaded.

The agarose/formaldehyde gel was installed in a horizontal electrophoresis box. 1X MOPS electrophoresis buffer was added to cover the gel to a depth of approximately 1mm. The gel was run for 5 minutes at 5 volt/cm and then the RNA sample was loaded into the wells of the gel. The two outermost lanes on each side of the gel were left empty and the RNA size standard (Bioline, UK) was loaded in the outside lanes of the gel. The gel was run at 4-5V/cm until the bromophenol blue migrates approximately 8cm. The RNA bands on the gel was visualised and photographed by G-box transilluminator (Syngene, UK).

2.2.5 Two-Step Reverse Transcription Polymerase Chain Reaction

cDNA Synthesis: Synthesis of cDNA was executed using SuperScript^{TM}-II Reverse Transcriptase kit (Invitrogen, UK) following the manufacturer's instructions. Synthesis was performed in a total volume of 20µl and master mix preparation is shown below. Two mixtures were prepared in two different tubes. The first mixture was prepared from the components shown in table 3.

Table 3. Master Mix I for cDNA Synthesis.

Reagent	1 Sample
Oligo (dT) (500 µg/ml)	1.0 µl
RNA template	(1ng to 5µg)
dNTPs 1.0	1.0 µl
Sterile Water	to 12 µl

The mixture was incubated on a heating block at 65ºC for 5 minutes and immediately chilled on ice for 2 minutes.

The second mixture was prepared from the components shown in table 4

Table 4. Master Mix II for cDNA Synthesis.

Reagent	1 Sample
5X First-Strand Buffer	4.0 µl
0.1 M DTT	2.0 µl
Sterile Water	1.0 µl
RNaseOUT™ (40 units/µl)	1.0 µl

The two mixtures were then combined and incubated at 42ºC for 2 min followed by adding 1µl (200 unit) of SuperScript™-II. The reaction was vortexed gently and incubated at 42ºC for 50min then followed by heating at 70ºC for 15 min to inactivate the reaction.

2.2.6 Polymerase Chain Reactions

Amplification of specific DNA products from reverse transcribed cDNA template was carried out using Phusion High-Fidelity PCR Master Mix (NEB, UK). The reaction was prepared from the following components; 10µl of 1x Phusion HF buffer mix (1.5 mM $MgCl_2$), 0.02U μL^{-1} Phusion DNA Polymerase, 200µM dNTPs, < 250 ng DNA template and 0.5 µM primers (Primers used were identical to Soler, Stoeva et al. (1996) and Soler, Stoeva et al. (1998) Table 6). Parameters for the Phusion PCR reaction are shown in table 5. The annealing temperatures were adjusted according to the primers used purchased from Eurofins Company (Eurofins-Germany) and shown in table 6. A no template control was used to check both RNA and DNA contamination in the reagents and consumables.

Table 5. Cycling parameters for PCR using the Phusion Taq system. Annealing temperatures (X) were modified depending on the primers used.

Stage		Conditions
Heated Lid	All cycles	110 °C, constant
Pre-denaturation	1 cycle	98 °C, 30 seconds
Amplification	35 cycles	98 °C, 10 seconds (Denaturation) X °C, 30 seconds (Annealing) 72 °C, 30 seconds (Elongation)
Final extension	1 cycle	72 °C, for 10 minutes
Storage	1 cycle	4 °C, Hold

Table 6. Primers used for Polymerase chain reaction and Sequencing.

Primers	Sequences	TM (°C)
MLA-fwd	AGAGGAAGATGGCCGCTCTC	60
MLA-rev	TACGAGAGGCTAAGACTCAGAG	61
MLB-fwd	GATGATGTTACCTGCAGTGCTTC	60
MLB-rev	TGGCACGGGAAGCCACATTTG	61
M13 Forward	GTAAAACGACGGCCAG	55
M13 Reverse	CAGGAAACAGCTATGAC	55

2.2.7 Agarose Gel Electrophoresis

Agarose gel electrophoresis was performed using a Tris/Borate/EDTA (TBE) buffering system. 0.5X TBE was made from a 5X stock solution contained 45 mM Tris-borate and 1 mM EDTA. Gels were made using TBE with Agarose (Lonza, Tewkesbury, UK). The amount of agarose used was dependent on the matrix density required for specific DNA molecules. DNA was visualised using Gel Red® (Invitrogen, UK) at a final concentration of 0.075µl per ml of gel. DNA HyperLadder I (Bioline, London, UK) was used to quantify the size and amount of DNA molecules of the bands. The gel was visualised and photographed by UV G-box transilluminator (Syngene, UK).

2.2.8 Gel Extraction of DNA

Gel purified PCR products were obtained from agarose gel using Promega Gel extraction kit (Promega, UK) following manufacturer's instructions. Briefly, DNA bands were excised from the gels using sterile scalpel blades and put into separate 1.5ml pre-weight eppendorf tubes. 10µl of Membrane Binding Solution was added to 10mg of the gel slices. The mixture was gently vortexed and incubated at 55°C for 10 minutes or until the gel slices were completely dissolved. Samples were transferred to a SV spin column placed in provided 2ml tubes and incubated for 3min and then centrifuged for 1min at 16000g. Flow through was discarded while placing the SV spin column back in the same tube. The DNA was washed twice by 700µl and 500µl of Membrane Wash Solution (wash buffer, 10mM Tris-HCl, pH 7.5, 80% ethanol) respectively, and centrifuged for 1min at 16,000g both times. The flow through was discarded and the column assembly was centrifuged for 1min with the micro-centrifuge lid open to allow evaporation of any residual ethanol.

The DNA was eluted by adding nuclease-free water to the centre of the SV membrane. The column was incubated for 1min at room temperature and centrifuged for 1min at 16,000g.

2.2.9 General Methods for Cloning of cDNA Fragments

2.2.9.1 Ligation of Blunt Ends of Fragments into Vectors

Purified PCR products (calculated according to the formula shown below) were ligated using Zero Blunt® PCR Cloning Kit (Invitrogen, UK) following the manufacturer's instructions. The PCR Zero Blunt® system is based on lethal selection with untransformed cells (those lacking an insert) being unable to grow. The components shown in table 7 were pipetted into an eppendorf tube for each ligation reaction and the 10:1 molar ratio of insert:vector was used in the ligation reaction using the formula shown below.

$$x\ ng\ insert = \frac{(10)(y\ bp\ PCR\ product)(25\ ng\ linearized\ pCR® - \text{Blunt})}{(3500\ bp\ pCR® - \text{Blunt})}$$

Table 7. Cloning of PCR products.

Reagent	Quantity (µl)
pCR®-Blunt (25 ng) vector	1
Blunt PCR product	1-5
5X ExpressLink ™ T4 DNA Ligase Buffer	2
ExpressLink TM T4 DNA Ligase (5 U/µL)	1
Water	X
Total Volume	10

The reaction was gently mixed by pipetting and incubated overnight at 16°C.

2.2.9.2 Transformation of E. coli with Plasmid DNA

All *E. coli* culturing was performed in LB broth medium (Sigma-Aldrich-USA) containing kanamycin (50ng/ul) antibiotic for the selection of transformants. The LB broth agar was prepared from 15.5 g L^{-1} of LB and 15 g L^{-1} agar (Fluka-Spain) and autoclaved for 15 minutes at 121°C.

TOP10 Chemically Competent *E. coli* were transformed with vector-cDNA constructs using the heat shock method. The TOP10 cells and the vector-cDNA constructs were mixed gently and left on ice for 30 minutes followed by heat shock for 30 seconds at 42°C. The cells were then incubated at 37 °C for 1 hour with 225rpm shaking to allow the expression of the kanamycin resistance gene in the vector. Transformed cells were selected by growth on LB broth agar containing 50µg/mL kanamycin (Invitrogen, UK) for 16 hours. Colonies containing successfully transformed cells were grown overnight in 5mL of liquid LB broth medium with 50µg/mL kanamycin at 37 °C with shaking at 225 rpm.

2.2.10 Plasmid DNA Purification

Plasmids were isolated from bacterial cells using ISOLATE Plasmid Mini kit (Bioline, UK) according to manufacturer's instructions. The extraction of the recombinant plasmid DNA was performed to the overnight cultures of *E. coli* grown in LB broth medium (Sigma-Aldrich-UK) unless otherwise stated 0.5-5 ml of overnight cultures were centrifuged at 11,000 g for 1 min at room temperature. The bacterial cells were re-suspended in 250 µl Resuspension Buffer and transferred into a new 1.5 ml microfuge tube. The cells were lysed by adding 250 µl of Lysis Buffer P and mixed by gently inverting 6-8 times. The plasmid DNA was precipitated using 350 µl Neutralization Buffer and mixed by vigorously inverting the tube 6-8 times and centrifuged at room temperature for 5 minutes at 11,000 g. The supernatant was recovered and applied to spin column P by pipetting. The samples were centrifuged for 1min at 11,000 g and the flow-through discarded. The spin column P washed by adding 500 µl of Wash Buffer AP and centrifuged at 11,000 g for 1 minute. The flow-through was discarded and the spin column P washed again by adding 700 µl of Buffer PE and centrifuged for 1minute at 11,000 g. The flow-through was discarded and column was centrifuged at 11,000 g for additional 2min to remove all the traces of ethanol. Spin column P was placed in a clean 1.5ml microfuge tubes and 50 µl elution buffer was added to the centre of the column, allowed to stand for 3 minutes at room temperature and centrifuged for 1min at 11,000g. The eluted DNA was stored at -20°C for further analysis.

2.2.11 Restriction Digestion Analyses

The cloned PCR products were digested from the pCR®-Blunt plasmid with restriction endonuclease enzymes as instructed by the manufacturer (NEB, UK). Briefly, 1µg of vector was added to a sterile tube containing 10 units of each restriction enzyme (EcoRI)

and 2µl NE Buffer. The mixture was made up to 20µl with sterile water and incubated at 37°C for 1hr. After the incubation time, samples were run on a 1.0% agarose gel at voltage 100V with variable current. The gel was visualised and photographed by G-box transilluminator (Syngene, UK).

2.2.12 Sequencing of Plasmid DNA

Sequencing of the cloned MLA and MLB was carried out by The University of Manchester DNA Sequencing Facility. The sequencing facility utilised a Prism 3100 Genetic Analyser (Applied Biosystems, Life Technologies) with BigDye Terminator cycle sequencing chemistries (Applied Biosystems, Life Technologies) and standard forward and reverse M13 primers. The analysis of the sequenced genes was carried out using Chroma Lite and CLC Sequence Viewer software (Qiagen, UK).

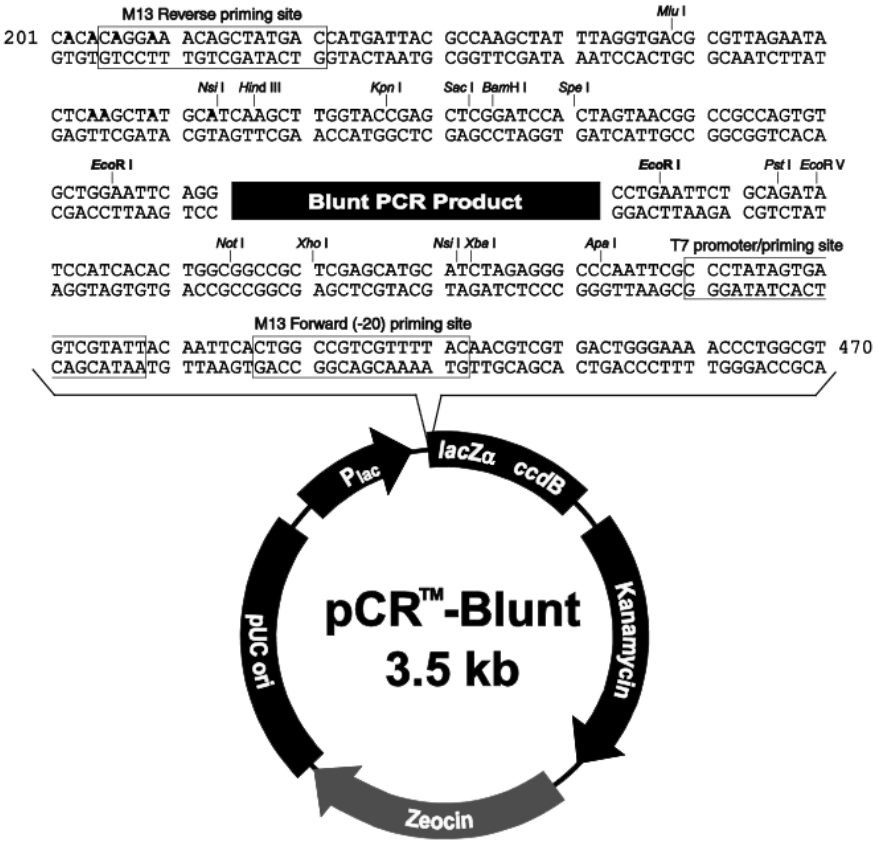


Figure 9. The pCR®-Blunt vector map that summarizes its features and shows restriction sites found in the polylinker.

2.3 Results

The full length DNA sequence of both the toxic A-chain (MLA) and glycan binding B-chain (MLB) of mistletoe lectin I (*Viscum album*) were prepared through PCR amplification reactions utilizing specific primers in order to separately study their biological activity in the following experiments.

2.3.1 C-type Lectin Amplification

Degenerate primers were designed (Table 8) corresponding to the conserved regions [LGGHLA (black box), GCWVGGH (red box), EYLIILAVVSGLIL (purple box) and TTLAVVIWLLAYR (orange box)] shown in figure 8 and identified from a ClustalW2 multiple sequence alignment of previously known C-type lectins from different natural sources.

Table 8. Degenerate primers designed based on the conserved regions of C-type lectin sequences

Conserved Regions	Primers
1- LGGHLA (NUY GGN GGN CAY YUN GCN) 2- GCWVGG (GGN UGY UGC GUN GGN GGN) 3- EYLIILAVVSGLIL (CAR UAY YUN AUH AUH YUN GCN GUN GUN WSN GGN YUN AUH YUN) 4- TTLAVVIWLLAYR (ACN CAN YUN GCU GUN GUN AUH UGG YUN YUN GCN UAY MGV)	A- Fwd $^{5'}$ ANG GNG GNC AYY ANG C $^{3'}$ B- Fwd $^{5'}$ GNA GYA GCG ANG GNG G $^{3'}$ C- Rev $^{5'}$ ACN ACN GCN ARD ATD ATN AR $^{3'}$ D- Rev $^{5'}$ CNA CNG CNA RDA TDA TNA RR $^{3'}$ E- Rev $^{5'}$ CKR TAN GCN ARN ARC C $^{3'}$ F- Rev $^{5'}$ ACN ACA GCN ARN GTN GT $^{3'}$

Based on the multiple sequence alignment we predicted PCR products sizes of 330bp and 390bp (Figure 10).

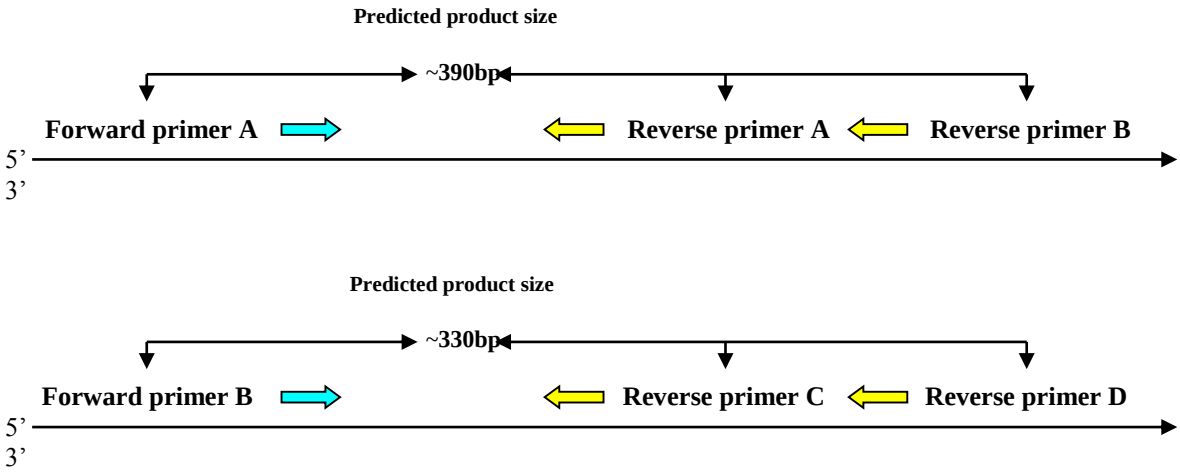


Figure 10. The illustration diagram of degenerate primers design.

PCR was unsuccessful despite repeat attempts and it was decided not to pursue this approach. Based on a survey of the literature (Büssing, Wagner et al. 1999, Niwa, Tonevitsky et al. 2003, Müthing, Meisen et al. 2004) we decided to focus on mistletoe (*Viscum album*) lectin (ML) as a glycan binding molecule.

2.3.2 Mistletoe RNA Extractions

Total RNA was extracted from the leaf tissue of mistletoe (*Viscum album*) using the ISOLATE II RNA plant kit (Bioline, UK). Sufficient quantities of pure RNA were extracted as shown in table 9 and the A260/280 ratio was as expected for pure RNA (Fleige and Pfaffl 2006).

Table 9. Spectrophotometrical determination of RNA concentrations and purity by NanoDrop 2000c (Thermo scientific, USA).

	Mistletoe RNA sample 1	Mistletoe RNA sample 2	Mistletoe RNA sample 3
Yield	297.9 ng/µl	90.9 ng/µl	290.3 ng/µl
λ260	7.446	2.271	5.805
λ280	3.735	1.064	2.851
260/280 Ratio	1.99	2.13	2.04

Formaldehyde gel electrophoresis was also used to check the purity and size of the extracted RNA, results shown in figure 11. As expected the major ribosomal bands were visualized and green tissue frequently yield chloroplast ribosomal bands. Minor degradation products are seen.

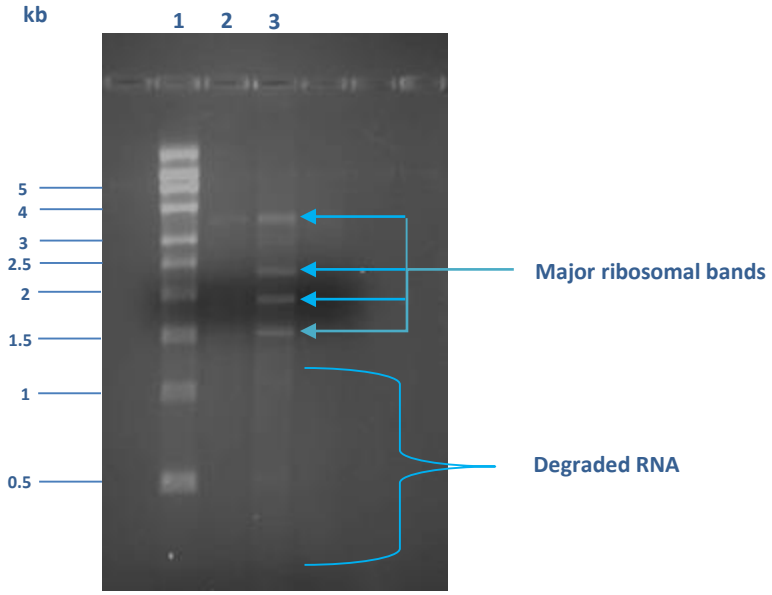


Figure 11. Formaldehyde electrophoreses gel (1%). Lane 1: RNA size marker (Invitrogen, UK); Lane 2 and 3: two different total RNA samples from mistletoe (*Viscum album*).

2.3.3 PCR Amplification of MLA and MLB

PCR amplification was conducted for cDNA derived from extracted RNA. The targeted chains were the toxic chain (MLA) and binding chain (MLB) of mistletoe lectin which are 762bp and 789bp respectively (Figure 12). Both chains were successfully amplified with sizes correspondent to the predicted sizes as shown in figure 13.

GAAGCAAGGAACAATGAATGCGCATTTGGCTTCAAGAAGGGCATGGGTTTGGTATTTT
CTAATGCTGTGCCTAGTTTTTGGTGCGACGGTCAAAGCTGAAACAAAATTCAGCTACG
AGAGGCTAAGACTCAGAGTTACGCATCAAACCACGGGCGAGGAATACTTCCGGTTCA
TCACGCTTCTCCGAGATTATGTCTCAAGCGGAAGCTTTTCCAATGAGATACCACTCTTG
CGTCAGTCTACGATCCCCGTCTCCGATGCGCAAAGATTTGTCTTGGTGGAGCTCACCA
ACCAGGGGGGAGACTCGATCACGGCCGCCATCGACGTTACCAATCTGTACGTGGTGGC
TTACCAAGCAGGCGACCAATCCTACTTTTTGCGCGACGCACCACGCGGCGCGGAAACG
CATCTCTTCACCGGCACCACCCGATCCTCTCTCCCATTCAACGGAAGCTACCCTGATCT
GGAGCGATACGCCGGACATAGGGACCAGATCCCTCTCGGTATAGACCAACTCATTCAA
TCCGTTACGGCGCTTCGTTTTCCGGGCGGCAGCACGCGTACCCAAGCTCGTTCCATTAT
AATCCTCGTTCAGATGATCTCCGAGGCCGCCAGATTCAATCCCATCTTATGGAGGGCT
CGCCAATACATTAACAGTGGGGCGTCATTTCTGCCAGACATGTACATGCTGGAGCTGG
AGACGAGTTGGGGCCAACAATCCACGCAAGTCCAGCAGTCAACCGATGGCGTTTTTAA
TAACCCAATTCGGTTGGCTATACCCCCCGGTAACTTCGTGACGTTGACCAATGTTCGCG
ACGTGATCGCCAGCTTGGCGATCATGTTGTTTGTATGCGGAGAGCGGCCATCTTCCTCT
GACGTGCGCTATTGGCCGCTGGTCATACGACCCGTGATAGCCGATGATGTTACCTGCA
GTGCTTCGGAACCTACGGTGCGGATTGTGGGTCGAAATGGCATGTGCGTGGACGTCCG
AGATGACGATTTCCACGATGGAAATCAGATACAGTTGTGGCCCTCCAAGTCCAACAAT
GATCCGAATCAGTTGTGGACGATCAAAAGGGATGGAACCATTCGATCCAATGGCAGC
TGCTTGACCACGTATGGCTATACTGCTGGCGTCTATGTGATGATCTTCGACTGTAATAC
TGCTGTGCGGGAGGCCACTATTTGGCAGATATGGGGCAATGGGACCATCATCAATCCA
AGATCCAATCTGGTTTTGGCAGCATCATCTGGAATCAAAGGCACTACGCTTACGGTGC
AAACACTGGATTACACGTTGGGACAGGGCTGGCTTGCCGGTAATGATACCGCCCCACG
CGAGGTGACCATATATGGTTTCAGGGACCTTTGCATGGAATCAAATGGAGGGAGTGTG
TGGGTGGAGACGTGCGTGAGTAGCCAACAGAACCAAAGATGGGCTTTGTACGGGGAT
GGTTCTATACGCCCCAAACAAAACCAAGACCAATGCCTCACCTGTGGGAGAGACTCCG
TTTCAACAGTAATCAATATAGTTAGCTGCAGCGCTGGATCGTCTGGGCAGCGATGGGT
GTTTACCAATGAAGGGGCCATTTTGAATTTAAAGAATGGGTTGGCCATGGATGTGGCG
CAAGCAAATCCAAAGCTCCGCCGAATAATTATCTATCCTGCCACAGGAAAACCAAATC
AAATGTGGCTTCCCGTGCCATGATTTAGGTTCATGGCTCGAAGATTGCTTGCATGCGA
CCATCCTTTCTATTTTCTCTTTTCTACCTTTTGAAATAATGTCTGTGGATAATGTGGCAC
GTTGAGGCCCGCCGGAAGAAGCCTTAGCCACCTTGTGTTTGAGAATAAATGAGTTAAT
GCAAGCAATCAACTTCATCCT

Figure 12. The full sequence (AY377890.1) of mistletoe lectin, MLA and MLB sequences and are highlighted in blue and green colours respectively; The linker sequence between MLA and MLB is highlighted in yellow; Start and stop codons are underscored and red.

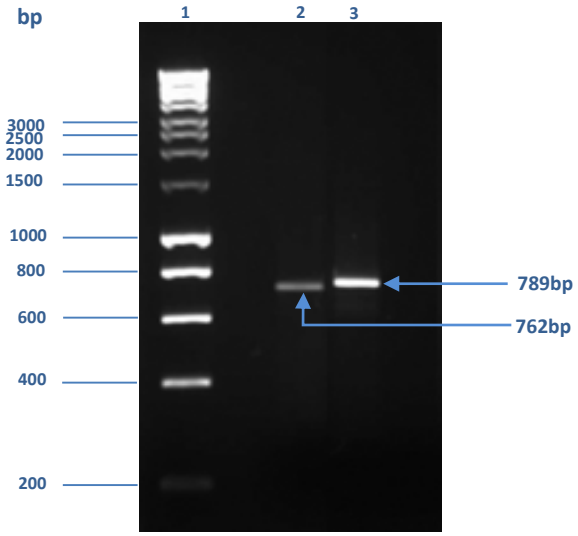


Figure 13. Agarose gel (1%) of PCR showing amplified chain A and chain B of mistletoe lectin. Lane 1: HyperLadder I; Lane 2: amplified chain A (MLA); Lane 3: amplified chain B (MLB).

The PCR products were successfully excised and extracted from the gel and run once again on a separate 1% agarose gel to confirm extraction success. The sizes were again 762bp and 789bp indicating the PCR products were not degraded during the extraction process (Figure 14).

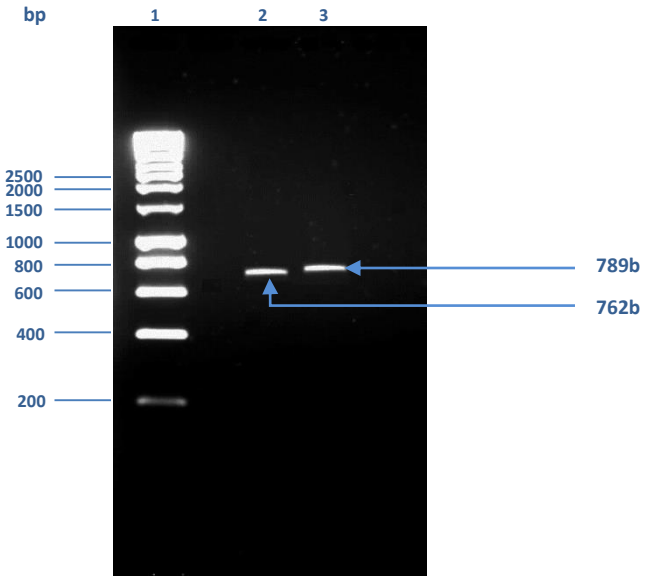


Figure 14. Agarose gel (1%) confirmation of gel purified (Wizard® SV, Peomega) MLA and MLB PCR products. Lane 1: HyperLadder I (Bioline, UK); Lane 2: purified MLA; Lane 3: purified MLB.

The purified PCR products went on to be cloned using the Zero Blunt® cloning kit (Invitrogen) and an overnight culture was prepared from colonies containing a successfully transformed cells grown overnight in 5mL of liquid LB broth medium. Then, a diagnostic restriction digest of extracted plasmids was employed to determine the presence of targeted sequences (Figure 15), confirming successful cloning of the designed product.

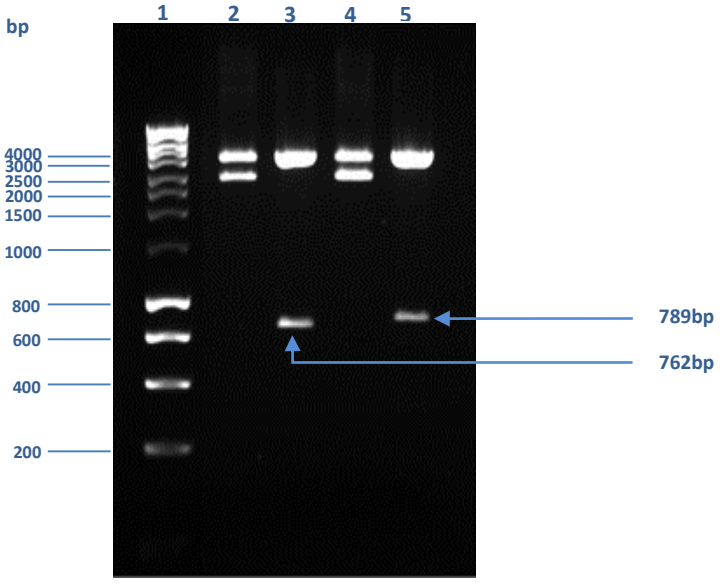


Figure 15. Restriction digest of plasmid pCR–Blunt II with EcoRI to confirm the presence of MLA and MLB PCR product inserts. Lane 1: HyperLadder I (Bioline, UK); Lane 2 and 4: undigested negative controls; Lane 3 and 5: the restriction digests of MLA and MLB respectively.

Several sequencing reactions were performed to confirm the identity of the targeted PCR products, partial sequencing results are shown in figure 16 and figure 17, in order to illustrate the quality of the sequencing reactions.

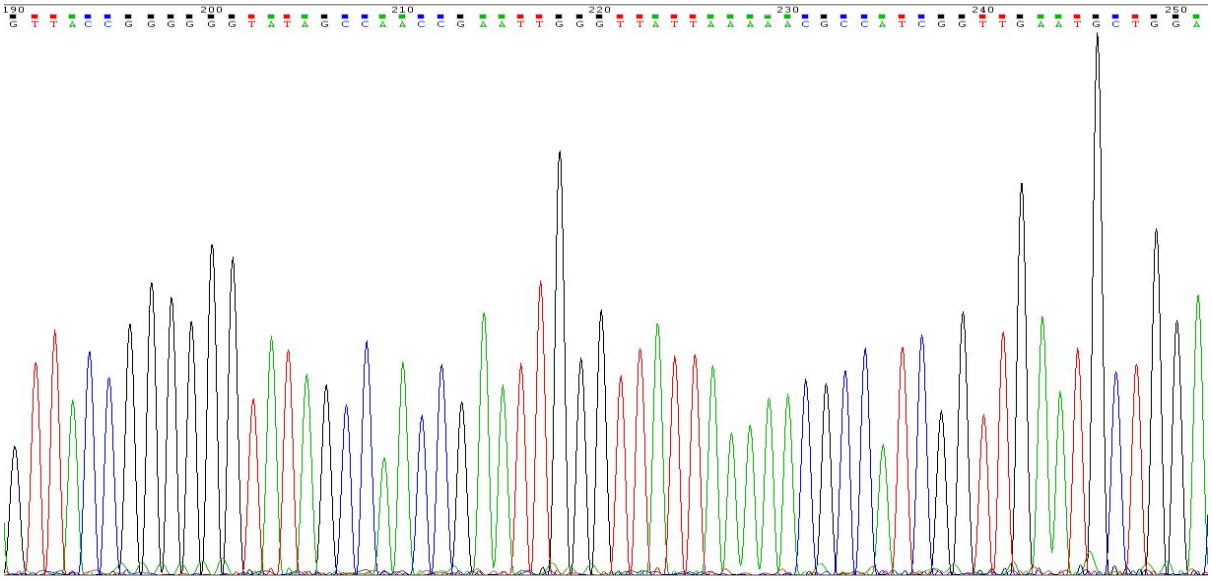


Figure 16. Chromatogram showing the sequence quality of mistletoe lectin type A (MLA) (partial sequence data shown). The selected region is representative of overall sequencing quality and full sequence is shown in appendix 3. The Adenine (A), Guanine (G), Cytosine (C) and Thymine (T) codons are in green, black, blue and red colours respectively.

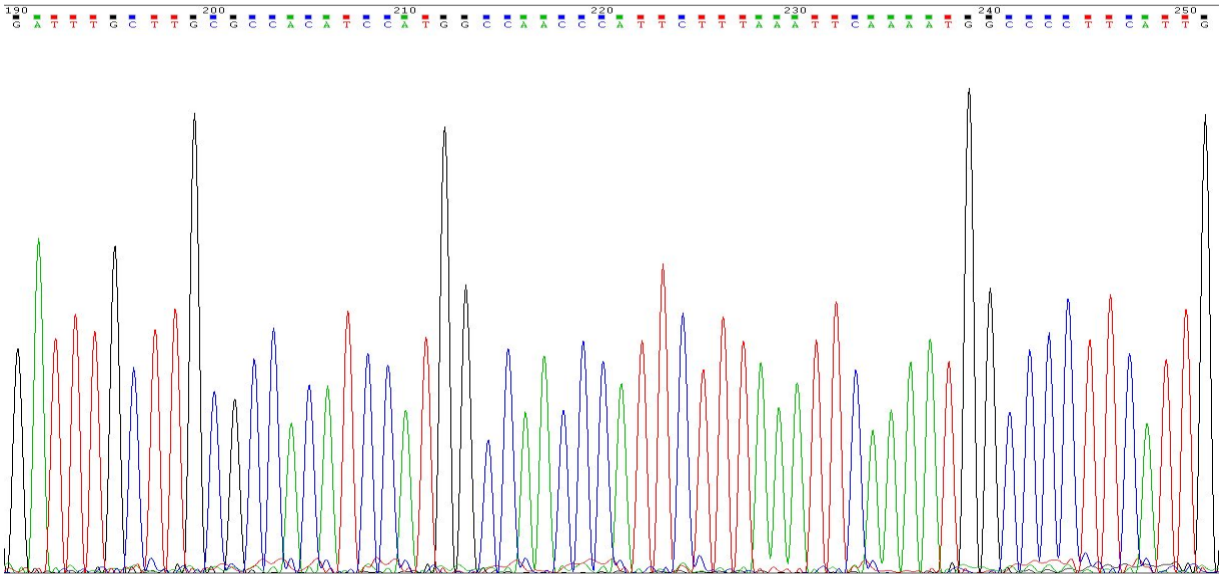


Figure 17. Chromatogram showing the sequence quality of mistletoe lectin type B (MLB) (partial sequence data shown). The selected region is representative of overall sequencing quality and full sequence is shown in appendix 3. The Adenine (A), Guanine (G), Cytosine (C) and Thymine (T) codons are in green, black, blue and red colours respectively.

The sequence data in figure 16 and figure 17 show clear peaks with very low levels of background ‘noise’. Each peak is well defined and fully resolved, giving a clear indication that the data are reliable.

2.3.4 Blast Results

The cloned sequences (MLA and MLB) were queried against a nucleotide database (https://blast.ncbi.nlm.nih.gov) using the BLASTN (Altschul, Gish et al. 1990) search tool to align sequences. The results are shown in table 10 and 11 and confirm the identity of the cloned sequences.

2.3.4.1 Mistletoe Lectin A chain (MLA) and Highly Similar Sequences

Ten out of the top ten hits were from *Viscum album* with the top hit being *Viscum album* lectin A-chain isoform I precursor with a maximum identity of 99%. All hits registered a 0.0 E value (Table 10). Thus confirming the identity of our MLA clone.

Table 10. The NCBI highly similar sequence BLASTN results for the MLA DNA sequence of mistletoe. The result shows that the cloned MLA sequence covers the whole query sequence and they are 99% identical with 0.0 expect value (E).

Accession	Description	Max score	Total score	Query cover	E value	Ident
AY081149.1	*Viscum album* lectin chain A isoform 1 precursor, mRNA, partial cds	1352	1352	100%	0.0	99%
AY377890.1	*Viscum album* clone ml1p lectin I precursor, gene, complete cds	1369	1369	100%	0.0	99%
AY377891.1	*Viscum album* clone cml1p lectin I precursor, mRNA, complete cds	1308	1308	100%	0.0	98%

Figure 18 shows the alignment of the mistletoe lectin A-chain (MLA) sequence from NCBI GeneBank (AY377890.1) and three results of sequencing reactions carried out for the cloned MLA chain. The sequencing reactions were repeated several times to confirm that the mismatch between few base pairs of the sequencing results and the query sequence are not experimental errors. It can clearly be seen that the cloned sequence is almost identical to the reference sequence in NCBI, with 7 variations in bases. The first at 231 is a C substitution for a G in the NCBI accession, the second at position 417 is a C substitution for a T in the NCBI accession, the third at position 465 is a G substitution for a C in the NCBI accession, the forth at position 469 is a T substitution for an A in the NCBI

accession, the fifth at position 478 is an A substitution for a G in the NCBI accession, the sixth at position 568 is a G substitution for a A in the NCBI accession, the seventh at position 624 is a T substitution for a G in the NCBI accession. These base mutations are confirmed by sequencing in triplicate.

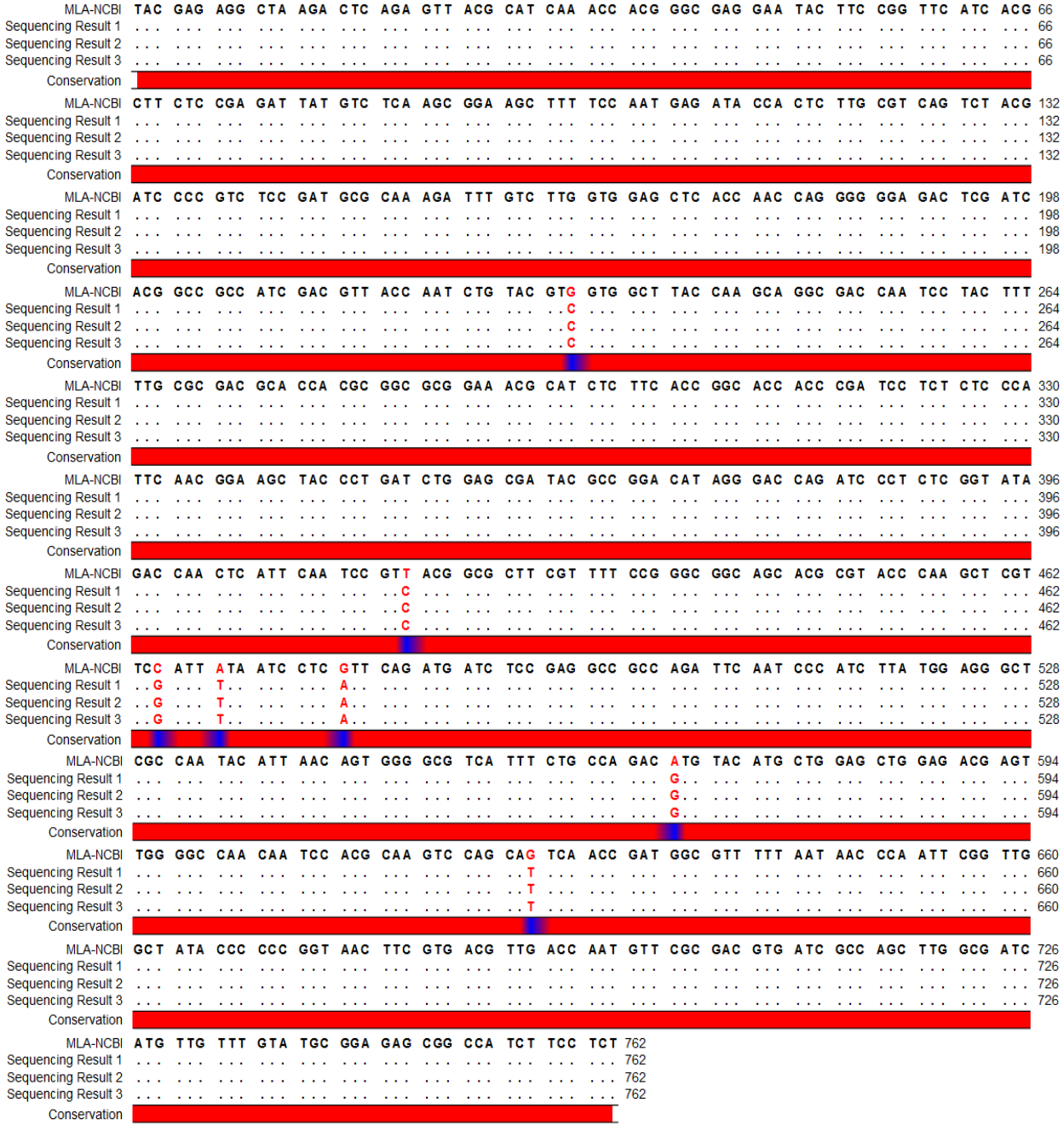


Figure 18. Alignment of reference MLA chain sequence (NCBI: AY377890.1) with three sequencing results of the same DNA sample of cloned MLA chain. The matching residues are shown as dot and the mismatching residues as RED codon letters. The red bar plot represents the similarity between the sequences and indicates the mismatches between the sequences in blue spots.

2.3.4.2 Mistletoe Lectin B-chain (MLB) and Highly Similar Sequences

Ten of the top ten BLASTN search hits were from the *Viscum album* with the top hit being *Viscum album* lectin I precursor with a maximum identity of 99%. All hits registered a 0.0 E value (Table 11).

Table 11. The NCBI highly similar sequence blastn results for the MLB DNA sequence of mistletoe. The result shows that the cloned MLB sequence covers the whole query sequence and they are 99% identical with 0.0 expect value (E).

Accession	Description	Max score	Total score	Query cover	E value	Ident
AY377890.1	*Viscum album* clone ml1p lectin I precursor, gene, complete cds	1441	1441	100%	0.0	99%
AY377891.1	*Viscum album* clone cml1p lectin I precursor, mRNA, complete cds	1402	1402	100%	0.0	99%
AY081149.1	*Viscum album* lectin chain A isoform 1 precursor, mRNA, partial cds	1375	1375	100%	0.0	98%

Figure 19 shows the alignment of the Mistletoe lectin B-chain (MLB) query sequence from NCBI GeneBank (AY377890.1) and three results of sequencing reactions carried out for the cloned MLB chain. The sequencing reactions were repeated several times so as to confirm that the mismatch between few base pairs of the sequencing results and the query sequence are not experimental errors. It can clearly be seen that the query sequence is almost identical to the sequencing date results, except for the following three base differences. The first at 26 is a G substitution for an A in the NCBI accession, the second at position 295 is a G substitution for an A in the NCBI accession, the third at position 479 is a C substitution for a T in the NCBI accession. These base mutations are confirmed by sequencing in triplicate.

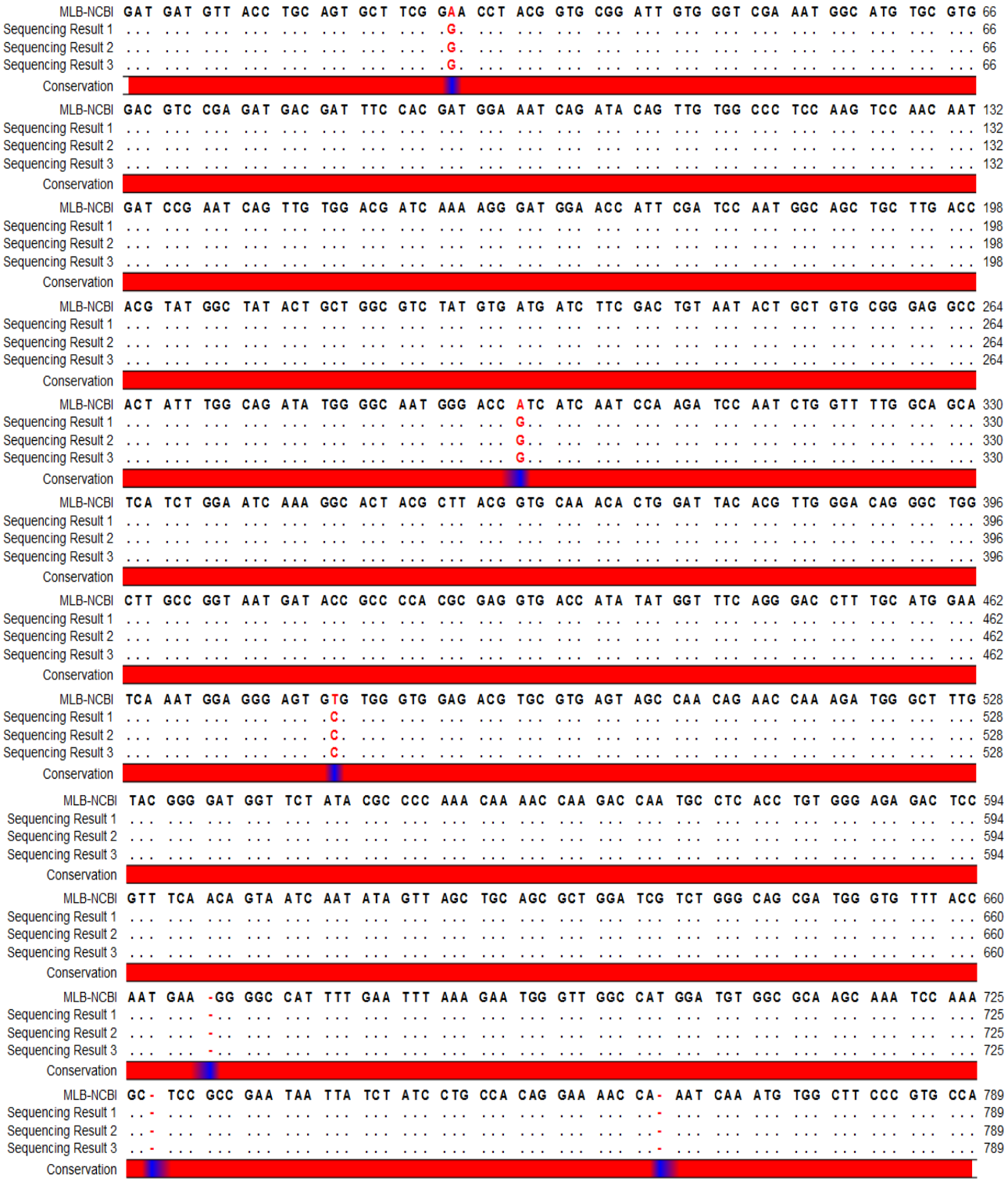


Figure 19. Alignment of reference MLB chain sequence (NCBI: AY377890.1) with three sequencing results of the same DNA sample of cloned MLB chain. The matching residues are shown as dots, the mismatching residues as RED codon letters and the gaps as red dashes. The red bar plot represents the similarity between the sequences and indicates the mismatches between the sequences in blue.

Thus we confirmed the successful cloning of the MLA and MLB chains from source (*Viscum album*).

2.4 Discussion

In this study mistletoe (*Viscum album*) lectin isoform one (MLI) was targeted to be amplified and cloned for further experiments.

The immunomodulatory improvement and many other positive effects of mistletoe extract in the treatment of patients with different type of cancers have increased biomedical focus on this extract (Ding, Cartwright et al. 2014, Galun, Tröger et al. 2015, Lentzen and Witthohn 2016). However, the crude mistletoe extract is composed of a diversity of compounds with different activities, the lectin is believed to be the key compound in the activity of this extract (Maletzki, Linnebacher et al. 2013). Mistletoe has three types of lectins (MLI, MLII, MLIII) which are slightly similar in their primary sequences but they are different in some core amino acids which give them different glycan binding affinity. Among these three isoforms, isoform one (MLI) is the most abundant one which is believed to play a crucial role in the functionality of mistletoe lectin. The MLI is composed of two different chains known as toxic chain (MLA) and binding chain (MLB) which are very similar to ricin A and B chain respectively (Niwa, Tonevitsky et al. 2003).

The MLA chain of MLI is a 29kD ribosome inactivating protein type II (RIPII) consists of 252 amino acids (756bp) and MLB chain is a 34kD galactose/sialic acid binding protein consists of 263 amino acids (789bp). MLA and MLB chains are linked together through a disulfide bond which keeps them attached to each other until B chain binds to the cell surface glycoproteins consisting of galactose/sialic acid residues and subsequently helps the endocytoses of MLA chain through modulating intracellular process such as protein phosphorylation and Ca^{+2} mobilization (Gabius, Walzel et al. 1991).

In order to obtain the coding sequences of MLA and MLB chains separately, a set of primers correspondent to the start and end of each chain was designed so as to be able to

amplify the chains through polymerase chain reactions (PCR). The total RNA was first extracted from fresh leaf of European mistletoe as a start material and went under quality tests such as measuring the purity via using spectrophotometry and running the RNA samples on analytical agarose gel using formaldehyde as denaturing reagent. A reliable value of ~2.0 measured at 260/280nm was achieved from nanodrop readings for the extracted RNA samples (Table 9). Also, the expected plant RNA bands such as major ribosomal bands (28s, 18s) were clearly detected on the analytical gel (Figure 11) (Ghawana, Paul et al. 2011).

The amplification of both MLA and MLB coding sequences from reverse transcribed complementary DNA (cDNA) of mistletoe was successfully performed through polymerase chain reactions (PCR) and using primers correlating to the start and end of MLA and MLB sequences. In order to avoid any possible base mutation during PCR reactions, Q5® High-Fidelity DNA polymerase (NEB, UK) was used which is a thermostable DNA polymerase with an error rate > 100-fold lower than Taq DNA polymerase. The predicted DNA bands of 756bp and 789bp of MLA and MLB respectively were successfully detected on 1% agarose gel (Figure 13) and they were both identical with the size of MLA and MLB bands detected by Eck, Langer et al. (1999) and Eck, Langer et al. (1999).

To perform further analysis of MLA and MLB coding sequences and identify their nucleotide composition, both sequences were cloned into Zero Blunt® TOPO® cloning vector and transformed into Top10 *E. coli* chemical competent cells. The cloning and transformation reactions were successful and the replicated plasmid DNA containing MLA and MLB sequences were successfully isolated from the lysed *E. coli* cells and purified.

In order to confirm the size of the cloned MLA and MLB sequences, restriction digest analysis was performed with EcoRI restriction enzymes which cuts the two EcoRI restriction sites at the start and end of cloned sequences. The digested plasmids were run on agarose gel and the same sizes of 756bp and 789bp were detected for cloned MLA and MLB chains respectively as shown in figure 15. While agarose gel can only show the size of DNA bands and the nucleotide composition of the bands must be detected by sequencing reaction, several Sanger sequencing reactions were performed to analyse the sequences of the cloned MLA and MLB. The reactions were performed in triplicate so as to detect any particular mutation comes from experimental errors and very clear peaks in the sequencing chromatograms were achieved (Figure 16,17).

The sequencing results were first assessed through conducting BLASTN using NCBI GeneBank server and the top 10 hits for both MLA and MLB were from mistletoe and the top hit of both of them scored 99% similarity with the cloned sequences (Table 10,11).

The coding sequences of MLA and MLB achieved from the sequencing reactions were then aligned with the reference sequences derived from NCBI GeneBank (AY377890.1). The alignment of MLA nucleotide sequence with the reference sequence shows very high similarity between the two sequences apart from seven codon mismatches at the positions 231, 417, 465, 469, 478, 568 and 624 shown in figure 18. However, the first three nucleotide mismatches are not counted as mutation because in the first and second positions GTC and GTT are replaced with GTG and GTC respectively and they are all representative codons of valine (Val) amino acid. Similarly, in the third position GTT is replaced with GTC and they are both representative codons of serine (Ser) amino acid. On the other hand, the other four mismatch are counted as mutation because the codons ATA (Isoleucine), GTT (Valine), ATG (Methionine), and CAG (Glutamine) are replaced with

TTA (Leucine), ATT (Isoleucine), GTG (Valine) and CAT (Histidine) respectively (Figure 18).

In the alignment of MLB nucleotide sequence with the reference sequence, only three codon mismatches were detected at the positions 26, 295 and 479 and they are all counted as mutation because the codons GAA (Glutamine acid), ATC (Isoleucine) and GTG (Valine) are replaced with GGA (Glycine), GTC (Valine) and GCG (Alanine) respectively (Figure 19).

In conclusion, high quality sequences of MLA and MLB chains were amplified and cloned into a cloning vector so as to be used in the next experiments.

Chapter 3. Design of Lectin-GFP Fusion Proteins from Mistletoe, Elderberry and Snowdrop Lectins

3.1 Introduction

The Introduction of green florescent protein from jellyfish (*Aequorea Victoria*) into cell biology has increased the feasibility of tracking the subcellular distribution of proteins and monitoring their function on a cellular level. Producing fusion proteins with GFP is considered as a sufficient method of protein labelling beside fixation and immunofluorescence labelling (Reid and Flynn 1997).

In molecular biology and biochemistry, GFP has been widely used as a biomarker for gene and protein expression. The expression level of the fused protein to GFP can thus be detected through the expression level of GFP and the localization of the target protein fused to GFP can also be tracked through visualization of tracking GFP excited under standard UV light (Kaether and Gerdes 1995, Ogawa, Inouye et al. 1995). The method is not without limitation and the main concerns about fusion proteins with GFP attached are the potential cleavage of GFP from the target protein which may cause non-identical distribution between them and also the incapability of the wild type of GFP to produce detectable fluorophore at physiological temperature. This was overcome by the use of mutant forms of GFP which has alleviated the problem through producing temperature-insensitive GFP forms (Siemering, Golbik et al. 1996, Patterson, Knobel et al. 1997).

The screening of lectin binding to glycans and detection of their localization is enhanced by reporter genes like fluorescent proteins fused with the lectin sequence. The availability of labelled lectins helps monitoring of tissue reactivity toward lectins and also extends the

scope of lectin applications beyond that which traditionally involved plant lectins. Previous studies have thus used GFP in a wide range of applications to produce Lectin-GFP fusion proteins. For example, to investigate the glycan binding affinity of *V. vulnificus* hemolysin (VVH) lectin from *Vibrio vulnificus* and perform a glycan array, the VVH sequence was fused to GFPuv which also helped to improve the solubility of the expressed lectin (Kaus, Lary et al. 2014).

Furthermore, in order to overcome the problems which face protein purification directly from natural resources such as cross contamination between the activity of a target protein and the activity of other similar proteins, it is also suggested to produce recombinant protein from the sequence of a target peptide alone or fused to a reporter peptide sequence such as GFP (Sørensen and Mortensen 2005). Medina-Bolivar, Wright et al. (2003) believed that the adjuvant activity of plant lectins purified form natural sources and used in vaccination studies was probably due to the contamination with the toxic subunits. Therefore, they produced a recombinant protein from the binding chain of ricin lectin (ricinB) fused to GFP to assess the potential of ricinB chain as an antigen delivery system without any side activities of other peptide especially the toxic chain of ricin lectin (ricinA). Thus a similar strategy will be employed with MLB lectin.

We aim to produce a range of fusion proteins from the binding chain of mistletoe lectin (MLB, with the binding chain of elderberry lectin (SNA-B) and snowdrop lectin (GNA) also fused to GFPuv and expressed in *E. coli* (as positive and negative control) so as to assess their glycan binding activity in the further studies.

3.2 Material and Methods

Green fluorescent protein (GFP) is a commonly used tag to assess the localization of proteins. Although GFP tagging has numerous advantages, it might alter the folding, function and the localization of the protein of interest because of protein-protein interactions. The location of the tag is believed to be the main reason of this alteration (Magliery, Wilson et al. 2005). Therefore, both *N*- and *C*- terminus cloning strategy were performed to avoid the effect of GFP on the properties and functions of the binding chain (MLB) of mistletoe lectin.

3.2.1 Gene synthesis of Elderberry (*Sambuca nigra*) Lectin Chain B (SNA-B)

The sequence of SNA lectin B-chain (SNA-B) from NCBI (U27122.1) was used as a template sequence for commercial synthesis. Two restriction sites (SacI and EcoRI) were also introduced to the start and end of the SNA-B respectively using SnapGene software (SnapGene-USA). The gene was synthesized by Eurofins Genomics Company (Eurofins, Germany) and sent back as a plasmid DNA in pEX-A2 vector (Table 12).

Table 12. The full sequence of elderberry lectin chain B (SNA-B). The restriction sites are in red and the added nucleotides are in small letters.

gatGAGCTCgctgcagctgcaGGGGGCGAGTACGAAAAAGTATGTTCGGTGGTAGAGGTAACAAGGCGCAT
CAGTGGTTGGGATGGATTGTGTGTGGACGTGAGGTATGGGCACTACATCGATGGGAATCCCGTCCAGC
TGCGGCCGTGTGGAAATGAATGTAACCAACTATGGACGTTCCGCACTGATGGAACAATCCGGTGGTTG
GGTAAATGCCTGACTGCCTCAAGCTCTGTCATGATATACGATTGTAATACTGTTCCTCCAGAGGCCACTA
AGTGGGTAGTATCTATTGACGGCACCATCACCAATCCTCACTCAGGACTCGTCCTTACAGCTCCTCAAGC
TGCAGAGGGAACCGCCCTGTCTCTGGAGAACAATATCCATGCCGCTAGGCAAGGTTGGACTGTAGGAG
ATGTAGAGCCCCTCGTTACTTTTATTGTGGGATATAAACAAATGTGCTTGAGGGAAAACGGTGAAAACA
ATTTTGTATGGTTGGAGGACTGCGTTCTCAACAGGGTGCAGCAAGAGTGGGCACTCTATGGCGACGGC
ACCATTCGAGTAAACAGTAATCGTAGCCTATGTGTGACCTCCGAAGACCACGAGCCCAGTGATCTTATC
GTCATTCTCAAGTGCGAAGGGTCGGGCAACCAGCGCTGGGTATTCAACACCAACGGTACCATCTCAAAC
CCAAACGCTAAACTACTTATGGACGTTGCACAACGCGATGTCTCTCTTCGAAAAATCATTCTCTATCGGC
CCACTGGGAATCCTAACCAGCAATGGATAACTACCACCCATCCAGCTgctgcataatGAATTCcaa

3.2.2 Introducing Restriction Sites

In order to be able to clone the genes of interest into pGFPuv expression vector, correlating restriction sites compatible to the targeted sites for the cloning in the expression vector were introduced to mistletoe lectin B-chain (MLB), elderberry lectin B-chain (SNA-B) and snow drop (*Galanthus nivalis*) lectin (Table 13) (Kindly provided by Professor Gatehouse - Newcastle University) using Q5 high-fidelity DNA polymerase (NEB, UK). The primers were designed using SnapGene software (SnapGene, USA) and are shown in table 14. The PCR reaction was performed in a total volume of 50µL, containing a total of <250ng of template DNA, 10µL of 5x Q5 reaction buffer (2mM MgCl2), 0.5µM primers, 200µM dNTPs, and 0.02U µL-1 Q5 High-Fidelity DNA Polymerase. All reactions followed the thermal cycling parameters: 30sec at 98°C followed by 35 cycles of 10sec at 98°C, 30sec at primer specific annealing temperatures, 30sec at 72, then final denaturation at 72°C for 2min.

Table 13. The full sequence of snow drop (*Galanthus nivalis*) lectin.

ATGGCTAAGGCAAGTCTCCTCATTTTGGCCGCCATCTTCCTTGGTGTCATCACACCATCTTGCCTGAGTG
ACAATATTTTGTACTCCGGTGAGACTCTCTCTACAGGGGAATTTCTCAACTACGGAAGTTTCGTTTTTAT
CATGCAAGAGGACTGCAATCTGGTCTTGTACGACGTGGACAAGCCAATCTGGGCAACAAACACAGGTG
GTCTCTCCCGTAGCTGCTTCCTCAGCATGCAGACTGATGGGAACCTCGTGGTGTACAACCCATCGAACA
AACCGATTTGGGCAAGCAACACTGGAGGCCAAAATGGGAATTACGTGTGCATCCTACAGAAGGATAG
GAATGTTGTGATCTACGGAACTGATCGTTGGGCTACTGGAACTCACACCGGACTTGTTGGAATTCCCGC
ATCGCCACCCTCAGAGAAATATCCTACTGCTGGAAAGATAAAGCTTGTGACGGCAAAGTAATGA

Table 14. Primers used in polymerase chain reaction, introduced restriction sites and added nucleotides shown in red, actual site underscored, main lectin sequence shown in black.

Primers	Sequences	TM(°C)
MLB + SacI (fwd)	GGA GCT CGC TGC AGC TGC AGA TGA TGT TAC CTG	73.2
MLB + EcoRI (rev)	CGA ATT CAT TAT GCA GCT GGC ACG GGA AGC CAC	72
MLB + XbaI (fwd)	GCG TCT AGA AAA AGA TGA TGT TAC CTG CAG TGC TTC G	71.7
MLB + KpnI (rev)	GCA TGA GGT ACC TTT GGC ACG GGA AGC CAC	72.2
SNA-B + HindIII (fwd)	TGA AAG CTT G GGG GGC GAG TAC GAA AAA	66.6
SNA-B + XbaI (rev)	TGA TCT AGA GT AGC TGG ATG GGT GGT AGT	66.7
GNA + XbaI (fwd)	ACA GCG TCT AGA A GCT AAG GCA AGT CTC CTC ATT TTG	70.6
GNA + KpnI (rev)	GCA TGA GGT ACC TT CTT TGC CGT CAC AAG C	69.5
GNA + SacI (fwd)	G GAG CTC GCT GCA GCT AAG GCA AGT CTC	71
GNA + SpeI -rev	C ACT AGT AG TTA TGC AGC CTT TGC CGT CAC AAG C	70.7

3.2.3 Ligation and Transformation of Blunt Ends SNA-B and GNA Fragments into Zero Blunt Cloning Vector

Gel Purified SNA-B and GNA PCR products were ligated into a cloning vector using Zero Blunt® PCR Cloning Kit (Invitrogen, UK) following the manufacturer's instructions shown in section 2.2.9.1. The reaction was gently mixed by pipetting and incubated overnight at 16°C. The One Shot® TOP10 Chemically Competent *E. coli* strain was transformed with the cloning vectors contained SNA-B and GNA constructs using heat shock method described in section 2.2.9.2.

3.2.4 Sequencing of Cloning Plasmid DNA

SNA-B and GNA recombinant plasmid DNA were isolated from bacterial cells using ISOLATE Plasmid Mini kit (Bioline, UK) following manufacturer's instructions described in section 2.2.10. After confirming the correct size band of both SNA-B and GNA sequences by diagnostic restriction digest, sequencing of the cloned SNA-B and GNA to the cloning vector was carried out by The University of Manchester DNA Sequencing Facility. The analysis of the sequenced genes was carried out using Chroma Lite and CLC Sequence Viewer software (Qiagen, UK).

3.2.5 Cloning of Lectin Fragments to the *N*- and *C*- terminus of Green Fluorescent Protein (GFP) Sequence in the pGFPuv Expression Vector

3.2.5.1 Ligation of MLB, SNA-B and GNA Fragments *N*-terminally to GFP Sequence

The *N*-terminus of the purified lectin sequence products with the corresponded introduced restriction sites MLB (SacI+MLB+EcoRI), SNA-B (SacI+SNA-B+EcoRI) and GNA (EcoRI+GNA+ SpeI) from the restriction digest reactions were ligated in frame to the *C*-terminus of GFP sequence between the correspondent restriction sites in the pGFPuv expression vector (Clontech, USA) using T4 ligase enzyme (Invitrogen, UK) following the manufacturer's instructions (Figure 20, 21 and 22). The components of the ligation reaction (Table 15) were pipetted into a PCR tube and incubated overnight at 16°C.

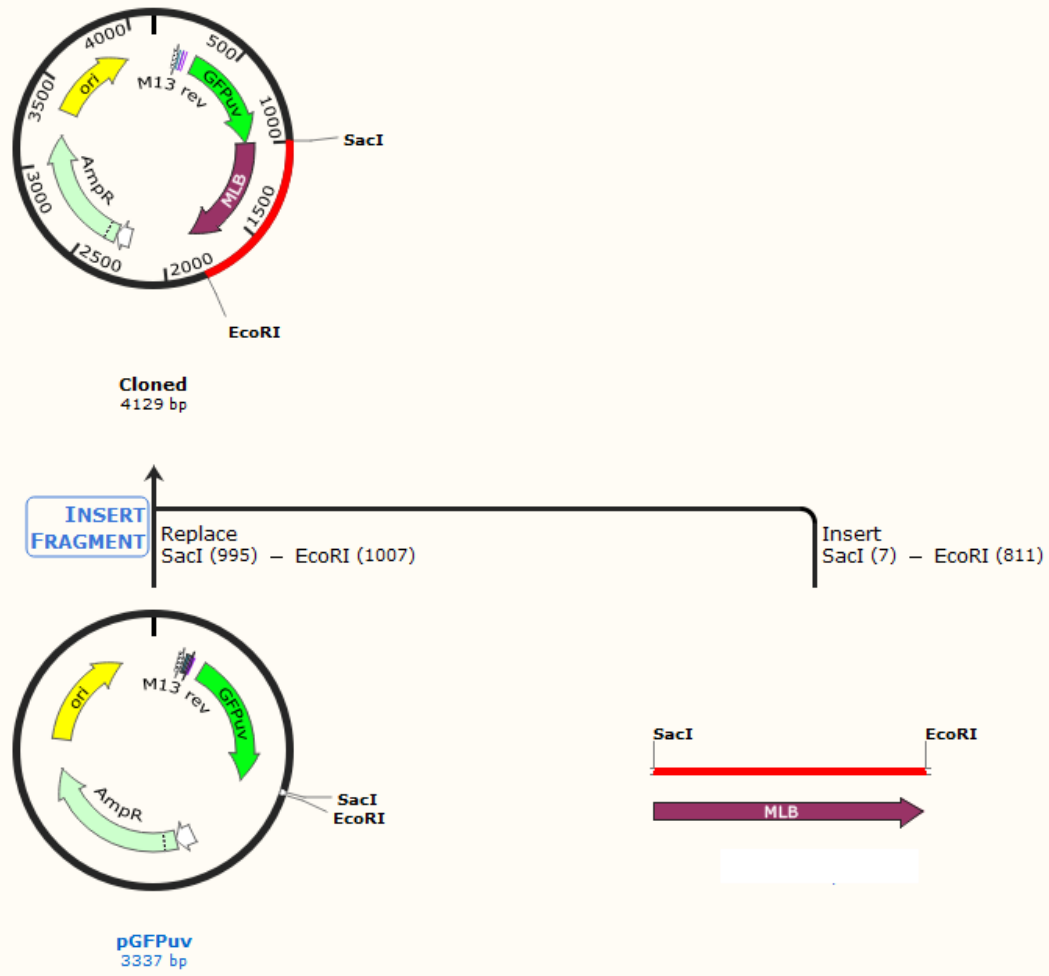


Figure 20. The MLB sequence *N*-terminally cloned to the *N*-terminus of GFP sequence between SacI and EcoRI restriction sites in the pGFPuv expression vector.

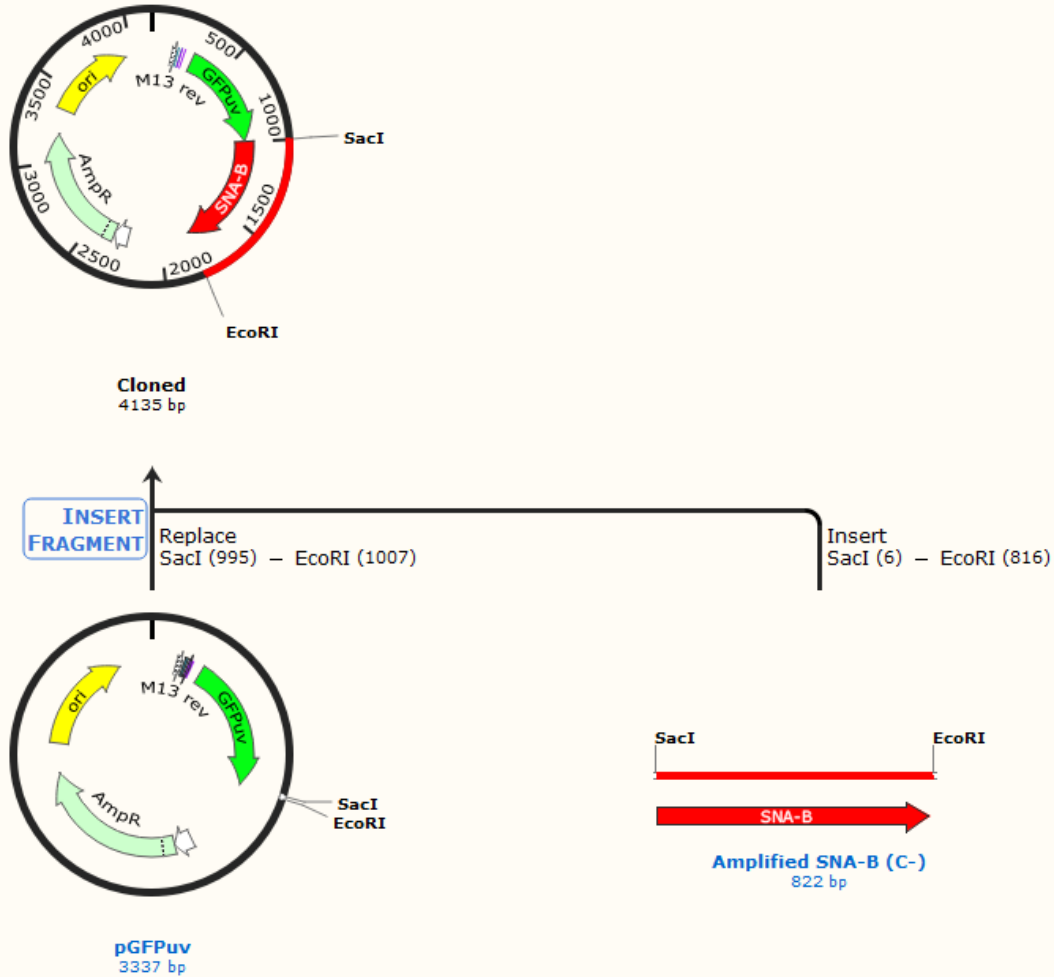


Figure 21. The SNA-B sequence *N*-terminally cloned to the *N*-terminus of GFP sequence between SacI and EcoRI restriction sites in the pGFPuv expression vector.

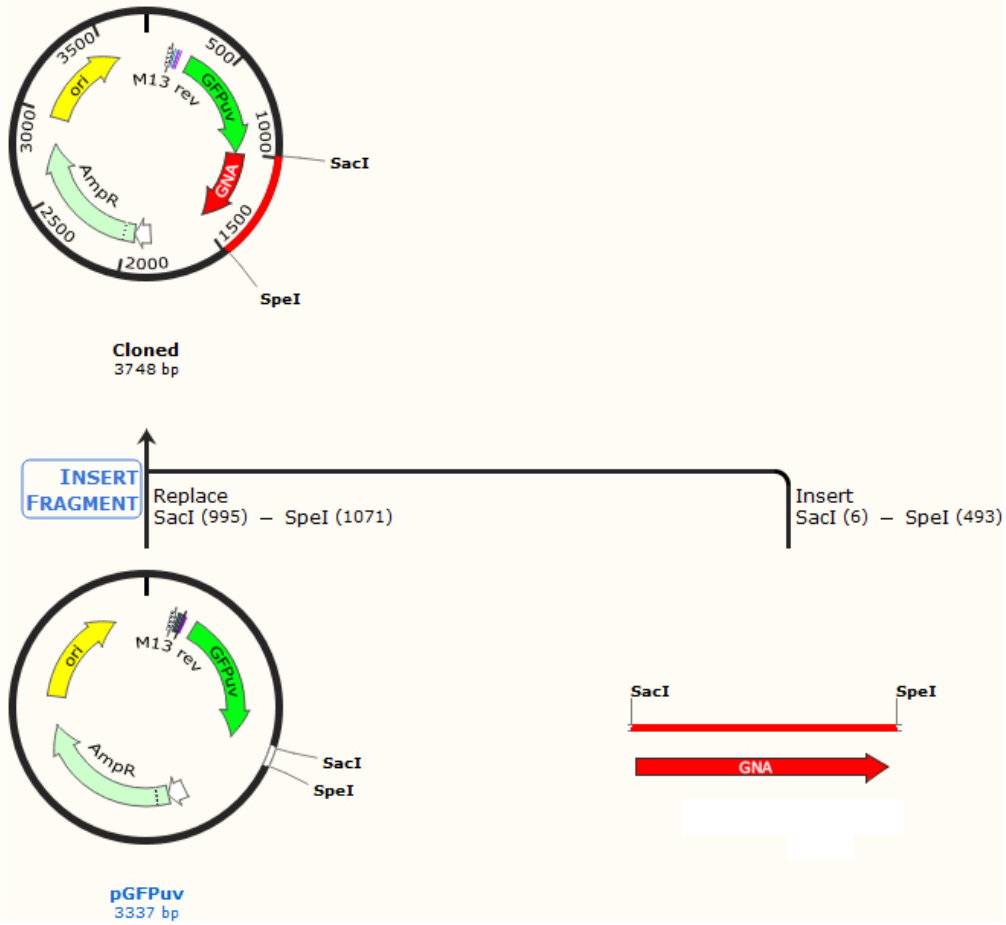


Figure 22. The GNA sequence *N*-terminally cloned to the *N*-terminus of GFP sequence between SacI and SpeI restriction sites in the pGFPuv expression vector.

3.2.5.2 Ligation of MLB, SNA-B and GNA Fragments *C*-terminally to GFP Sequence

The *C*-terminus of the purified lectin sequence products with the corresponded introduced restriction sites MLB (XbaI+MLB+KpnI), SNA-B (HindIII+SNA-B+XbaI) and GNA (XbaI+GNA+KpnI) from the restriction digest reactions were ligated in frame to the *N*-terminus of GFP sequence between the correspondent restriction sites in the pGFPuv expression vector (Clontech, USA) using T4 ligase enzyme (Invitrogen, UK) following the manufacturer's instructions (Figure 23, 24 and 25). The components of the ligation reaction (Table 15) were pipetted into a PCR tube and incubated overnight at 16^{o}C.

Table 15. Cloning of MLB, SNA and GNA products *N*- and *C*- terminally to GFP sequence in pGFPuv expression vector.

Reagent	Quantity (µl)
pGFPuv (50 ng) vector	1
MLB product (37.5 ng)	1-5
T4 DNA Ligase Buffer	2
T4 DNA Ligase (10 U/µL)	1
Total Volume	10

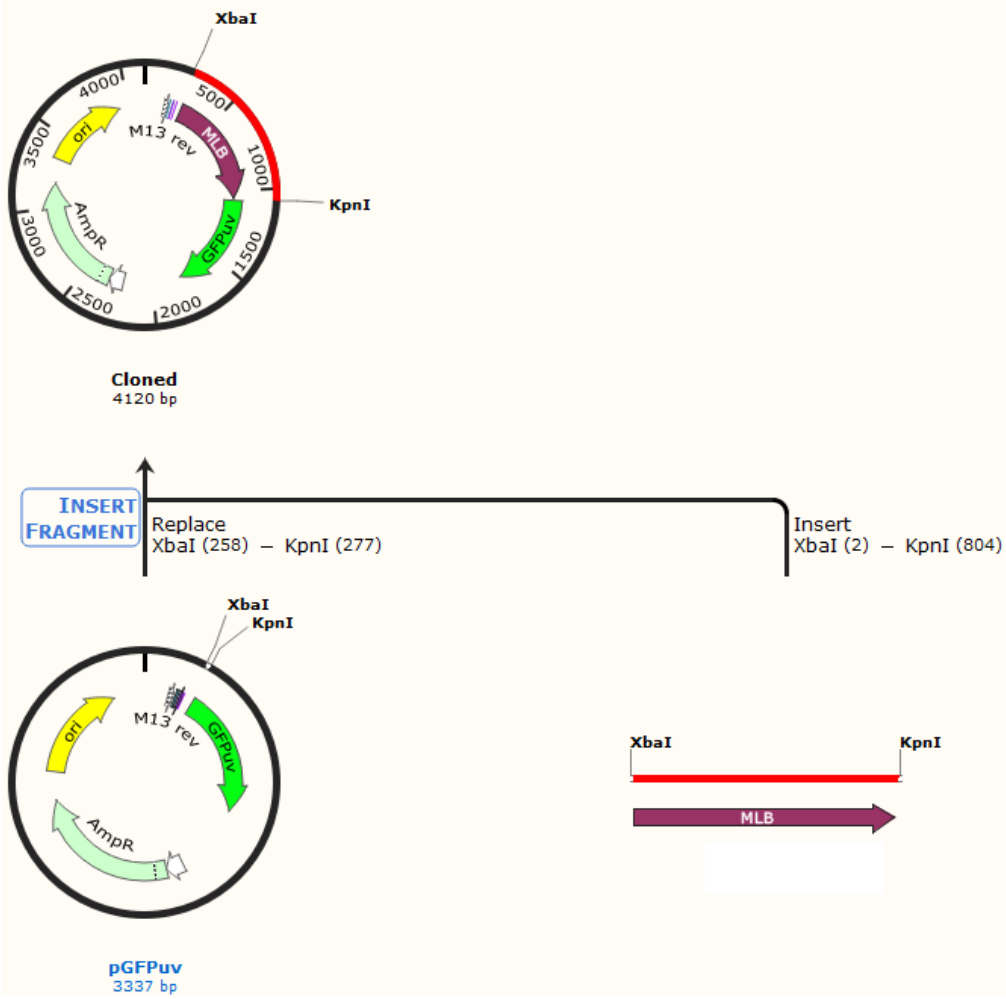


Figure 23. The MLB sequence *C*-terminally cloned to the *C*-terminus of GFP sequence between XbaI and KpnI restriction sites in the pGFPuv expression vector.

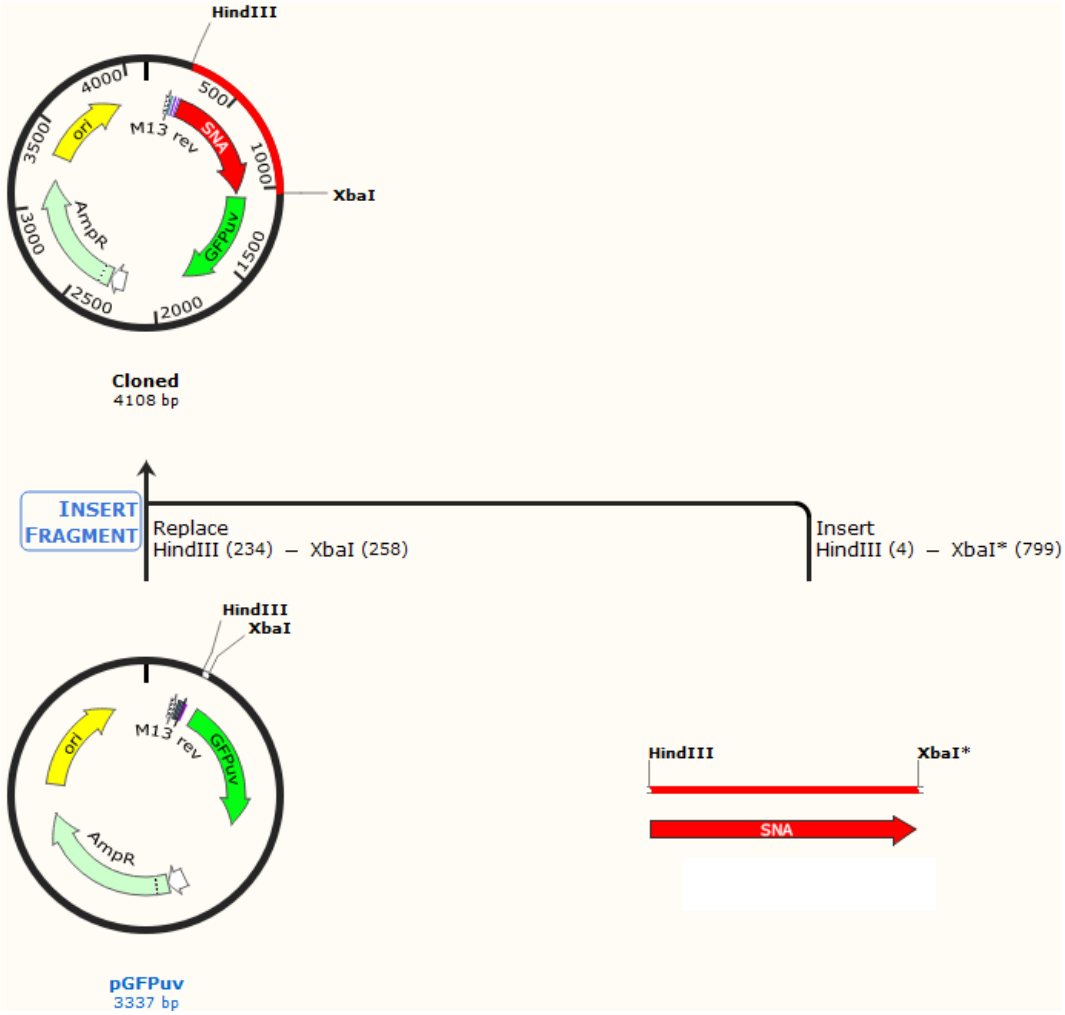


Figure 24. The SNA-B sequence *C*-terminally cloned to the *C*-terminus of GFP sequence between HindIII and XbaI restriction sites in the pGFPuv expression vector.

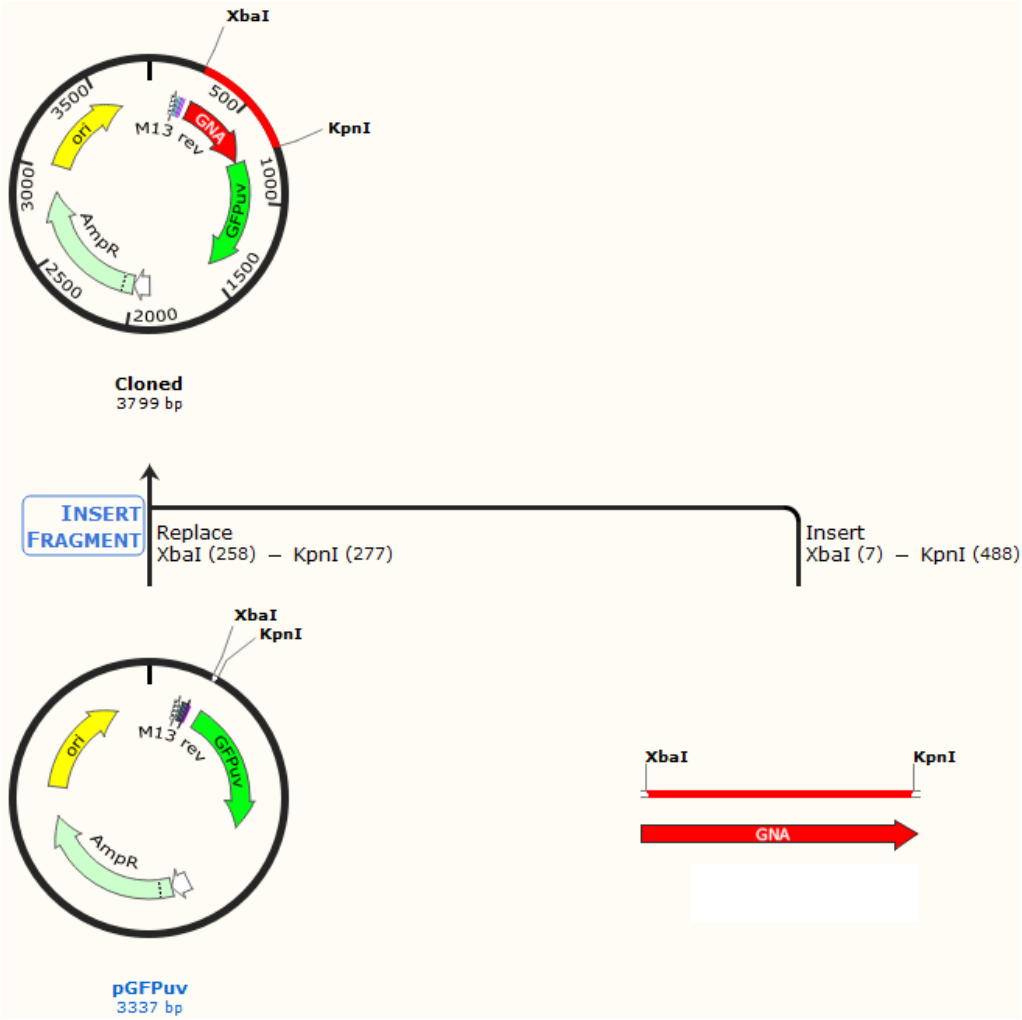


Figure 25. The GNA sequence cloned to the *C*-terminus of GFP sequence between XbaI and KpnI restriction sites in the pGFPuv expression vector.

3.2.6 Transformation of BL21(DE3) E. coli Strain with the pGFPuv Expression Vector Containing Lectin-GFP Constructs

All *E. coli* culturing was performed in LB broth (Luria low salt) medium (Sigma-Aldrich-USA) containing ampicillin (100ng/ul) antibiotic for the selection of transformants. The LB broth agar was prepared from 15.5 g L^{-1} of LB and 15 g L^{-1} agar (Fluka-Spain) and autoclaved for 15 minutes at 121 ^{o}C.

The BL21(DE3) *E. coli* strain was transformed with lectin-GFP constructs using the heat shock method. The chemically competent cells and the lectin-GFP products from the ligation reactions were mixed and left on ice for 30 minutes followed by heat shock for 10

seconds at 42°C. The cells were then incubated at 37 °C for 1 hour with 250rpm shaking to allow the expression of the ampicillin resistance gene in the vector. Transformed cells were selected by growth on LB broth agar plates containing 100µg/mL ampicillin (Fisher, UK) for 16 hours. Plates were screened under standard UV light (395nm-509nm) to visualise GFP expression in the colonies containing MLB, SNA-B and GNA inserts then the successfully transformed cells were grown overnight in 5mL of liquid LB broth medium with 100µg/mL ampicillin at 37 °C with shaking at 250 rpm.

3.2.7 Restriction Digestion Analysis

Plasmids were isolated from bacterial cells using ISOLATE Plasmid Mini kit (BIOLINE, UK) according to the manufacturers instruction shown in section 2.2.10.

The lectin-GFP constructs were digested by restriction endonuclease enzymes as instructed by the manufacturer (NEB, UK). For the *N*-terminus digestion, 1µg of vectors was added to a sterile tube containing 10 units of each restriction enzymes (HindIII, XbaI and KpnI), 2µl Cut Smart buffer and made up to 20µl with sterile water. For the *C*-terminus digestion, 1µg of vector was added to a sterile tube containing 10 units of each restriction enzymes (SacI, EcoRI and SpeI), 2µl Cut Smart buffer and made up to 20µl with sterile water. All tubes were incubated at 37°C for 1hr. After incubation, samples were run on a 1.0% agarose gel at voltage 100V with variable current. The gel was visualised and photographed by UV G-box transilluminator (Syngene, UK).

3.2.8 Sequencing of plasmid DNA

Sequencing of the lectin-GFP constructs (*N*- and *C*-terminus) was carried out by The University of Manchester DNA Sequencing Facility. Standard M13 Reverse primer was used for the sequencing reactions. The analysis of the sequenced genes was carried out using Chroma Lite and CLC Sequence Viewer software (Qiagen, UK).

3.2.9 Expression of Recombinant MLB, SNI, GNA

The recombinant expression vector pGFPuv (Clonthech, USA) was used to express the targeted mistletoe lectin B-chain, elderberry lectin B-chain and GNA lectin genes fuse to GFP. The vector also encodes *Aequorea victoria* green fluorescent protein (GFP) that has been optimized for brighter fluorescent when excited by standard UV light (360–400 nm). The vector is also under the control of strong T7 bacteriophage transcription signals when introduced in *E. coli* host strain BL21(DE3), means the expression of the system is repressed unless induced by IPTG.

The cloned MLB+GFP, SNA-B+GFP and GNA+GFP in the pGFPuv expression vector were transformed into BL21(DE3) cell line then plated on antibiotic selection plates (Ampicillin 100μg/ml) and incubated for 72hrs at 30°C. Plates were screened under a standard UV light (360–400 nm) to visualize glowing colonies containing GFP with MLB, SNI and GNA inserts. The successfully transformed cells were grown overnight in 5mL of liquid LB broth medium conataining 100μg/mL ampicillin at 37°C with shaking at 250 rpm. From the overnight incubated culture 1ml was added to 100 ml liquid culture with antibiotic and incubated at 30°C until OD600 reaches 0.4–0.8 followed by induction through addition of 500 μM IPTG for 18 hours at 30°C. The expression of the targeted genes in the total cell extract was confirmed by InstantBlue Coomassie (Expedeon, UK) stained SDS Page electrophoresis.

3.2.10 Protein Extraction using BugBuster®

BugBuster Protein Extraction Reagent® was used to extract the total proteins from *E. coli* cells as per the manufacturer's instructions. BugBuster is formulated for the gentle disruption of the *E. coli* cell wall, resulting in the liberation of soluble protein. Cells were harvested from liquid culture by centrifugation at 10,000g for 10 min using a pre-weighed centrifuge tubes. The wet weight of the pellet was calculated after removing as much supernatant as possible. The cell pellet was completely resuspended in BugBuster reagent through gentle vortexing at room temperature. The volume of BugBuster used for the extraction was 5 ml per g of cell pellets from the induced bacterial culture. The cell suspension was incubated on a shaking platform at a slow setting for 20 min at room temperature. Cell debris was then removed by centrifugation at 16,000g for 20 min at 4°C. Finally, the supernatant was transferred to a fresh universal tube and stored at –80°C until further use.

3.3 Results

The sialic acid binding property of mistletoe lectin isoform I has been proven previously in addition its well-known affinity for galactose (Müthing, Meisen et al. 2004). However, the activity of the binding chain of mistletoe lectin (MLB) alone without the toxic A chain (MLA) has not been separately tested and has remained as a hypothesis. Therefore, fusion constructs were designed from the MLB chain alone fused to green florescent protein (GFP) to generate constructs and express fusion proteins in this study in order to subsequently test this hypothesis in the next study. The binding chain (SNA-B) of elderberry lectin (*Sambucus nigra*) as a sialic acid binding lectin was used to produce positive control fusion proteins and the mannose binding lectin (GNA) of snowdrop (*Galanthus nivalis*) was used produce negative control fusion proteins.

3.3.1 MLB-GFP Construct Design

Two fusion strategies were used to assemble constructs of the MLB chain fused to GFP in pGFPuv expression vector. These were MLB fused N-terminally and C-terminally to GFP. The cloning map and primers were designed using Snapgene program (Snapgene, USA). The MLB sequence was checked for the presence of internal restriction sites and to confirm that the restriction sites used were not present in the sequence using CLC Sequence Viewer program (Qiagen, UK).

3.3.1.1 *N*-terminus cloning

The MLB chain was successfully cloned *N*-terminally to GFP through fusing the *N*-terminus of MLB to the *C*-terminus of GFP between SacI and EcoRI restrictions sites (Figure 26). The full sequence of pGFPuv expression vector is shown in appendix 2. Primers used to introduce the restriction sites to the start and end of MLB sequence are shown in table 14.

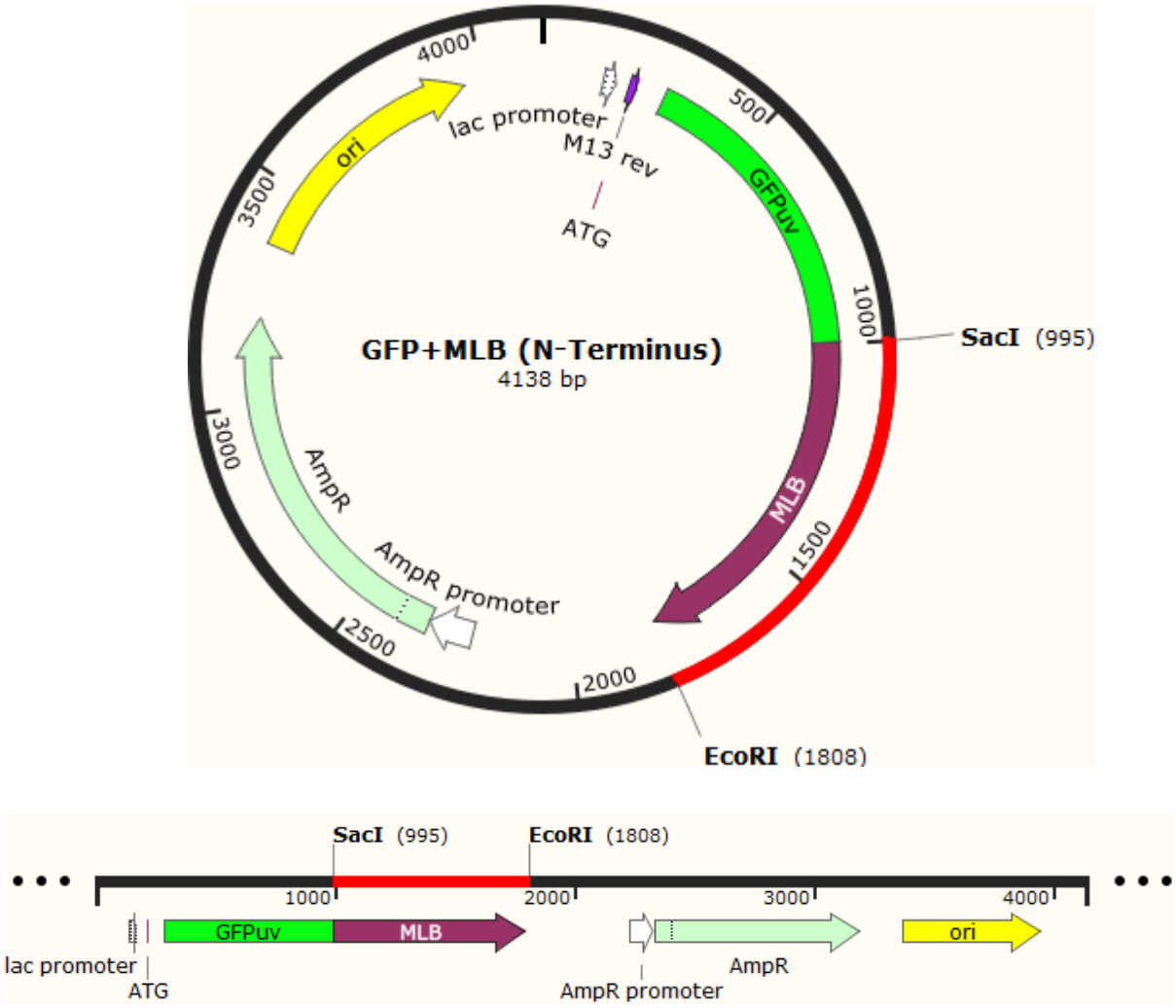
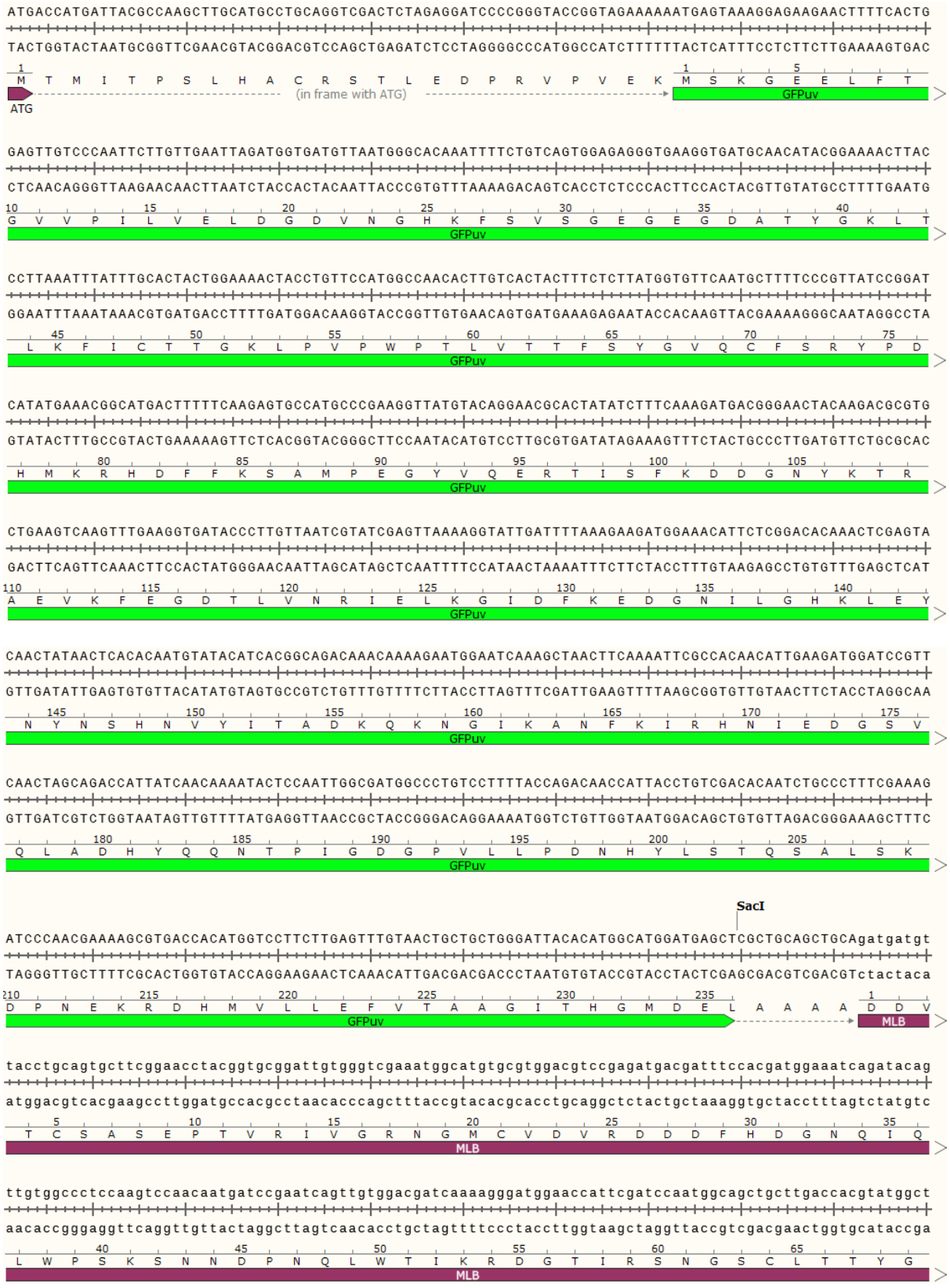
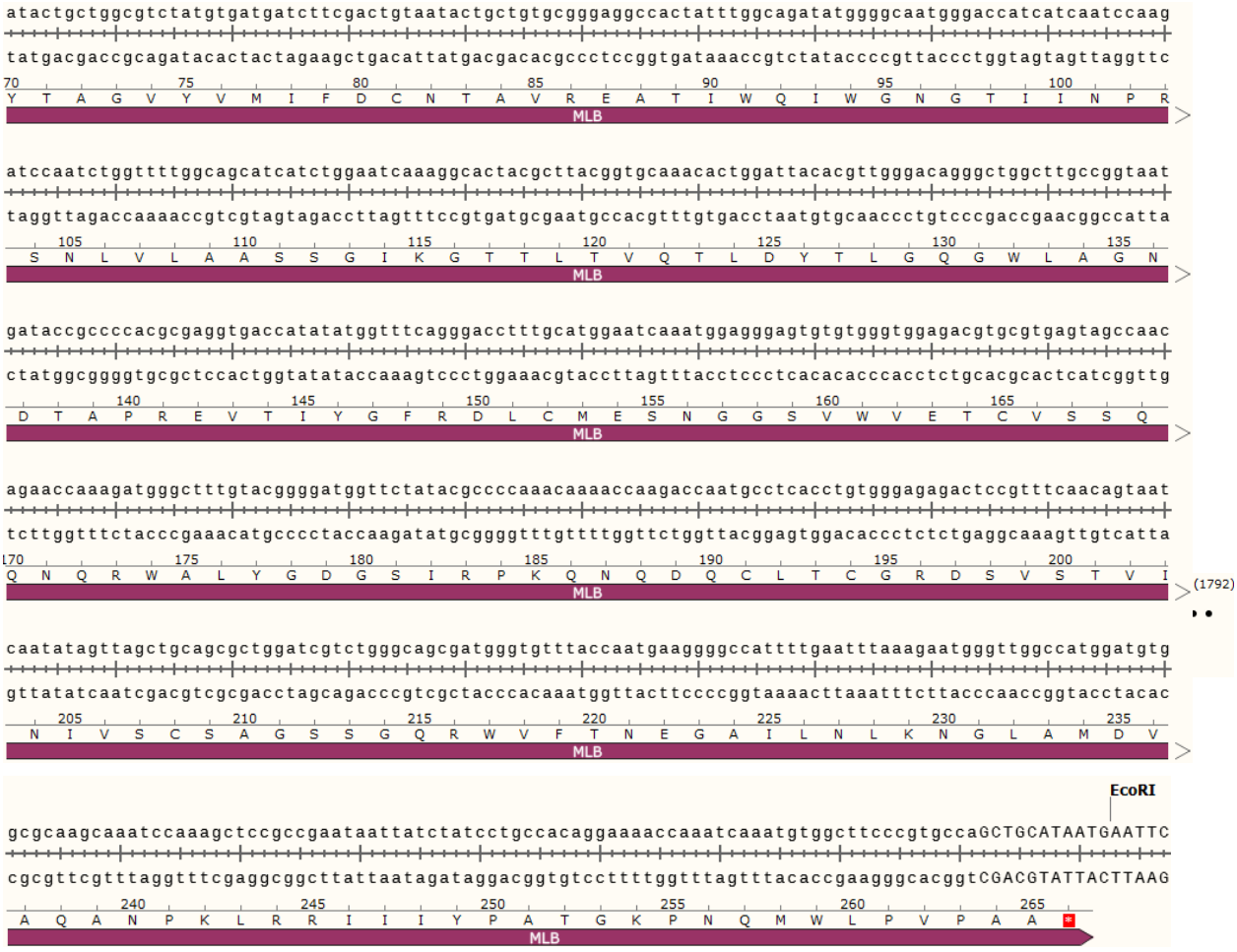


Figure 26. Cloned MLB chain (RED BAR) *N*-terminally fused to GFP protein sequence in the pGFPuv expression vector. The GREEN and MAGENTA parts of the arrow represent GFP and MLB sequences respectively; The light green and yellow arrows represent the ampicillin resistant (AmpR) and signal of origin of replication (ori) sequences of pGFPuv vector.

Figure 27 below shows a detailed view of the *N*-terminus cloning strategy, note that the MLB sequence is in frame with the GFP sequence and the stop codon (TAA).

Figure 27. The full nucleotide and deduced amino acid sequence of MLB chain *N*-terminally fused to GFP between SacI and EcoRI restriction sites in pGFPuv expression vector. The MLB sequence is indicated by a MAGENTA plot bar; The GFP sequence is indicated by a GREEN plot bar.




atactgctggcgtctatgtgatgatcttcgactgtaatactgctgtgcgggaggccactatttggcagatatggggcaatgggaccatcatcaatccaag
tatgacgaccgcagatacactactagaagctgacattatgacgacacgccctccggtgataaaccgtctataccccgttaccctggtagtagttaggttc
70 75 80 85 90 95 100
Y T A G V Y V M I F D C N T A V R E A T I W Q I W G N G T I I N P R
MLB
atccaatctggttttggcagcatcatctggaatcaaaggcactacgcttacggtgcaaacactggattacacgttgggacagggctggcttgccggtaat
taggttagaccaaaaccgtcgtagtagaccttagtttccgtgatgcgaatgccacgtttgtgacctaatgtgcaaccctgtcccgaccgaacggccatta
105 110 115 120 125 130 135
S N L V L A A S S G I K G T T L T V Q T L D Y T L G Q G W L A G N
MLB
gataccgccccacgcgaggtgaccatatatggtttcagggacctttgcatggaatcaaatggagggagtgtgtgggtggagacgtgcgtgagtagccaac
ctatggcggggtgcgctccactggtatataccaaagtccctggaaacgtaccttagtttacctccctcacacacccacctctgcacgcactcatcggttg
140 145 150 155 160 165
D T A P R E V T I Y G F R D L C M E S N G G S V W V E T C V S S Q
MLB
agaaccaaagatgggctttgtacggggatggttctatacgccccaaacaaaaccaagaccaatgcctcacctgtgggagagactccgtttcaacagtaat
tcttggtttctacccgaaacatgcccctaccaagatatgcggggtttgttttggttctggttacggagtggacaccctctctgaggcaaagttgtcatta
170 175 180 185 190 195 200
Q N Q R W A L Y G D G S I R P K Q N Q D Q C L T C G R D S V S T V I
MLB
(1792)
caatatagttagctgcagcgctggatcgtctgggcagcgatgggtgtttaccaatgaaggggccattttgaatttaaagaatgggttggccatggatgtg
gttatatcaatcgacgtcgcgacctagcagacccgtcgctacccacaaatggttacttccccggtaaaacttaaatttcttacccaaccggtacctacac
205 210 215 220 225 230 235
N I V S C S A G S S G Q R W V F T N E G A I L N L K N G L A M D V
MLB
EcoRI
gcgcaagcaaatccaaagctccgccgaataattatctatcctgccacaggaaaaccaaatcaaatgtggcttcccgtgccaGCTGCATAATGAATTC
cgcgttcgtttaggtttcgaggcggcttattaatagataggacggtgtccttttggtttagtttacaccgaagggcacggtCGACGTATTACTTAAG
240 245 250 255 260 265
A Q A N P K L R R I I I Y P A T G K P N Q M W L P V P A A *
MLB

3.3.1.2 *C*-terminus cloning

The MLB chain was cloned *C*-terminally to GFP through fusing the *C*-terminus of MLB to the *N*-terminus of GFP between XbaI and KpnI restriction sites (Figure 28). Primers used to introduce the restriction sites to the start and end of MLB sequence are shown in table 14.

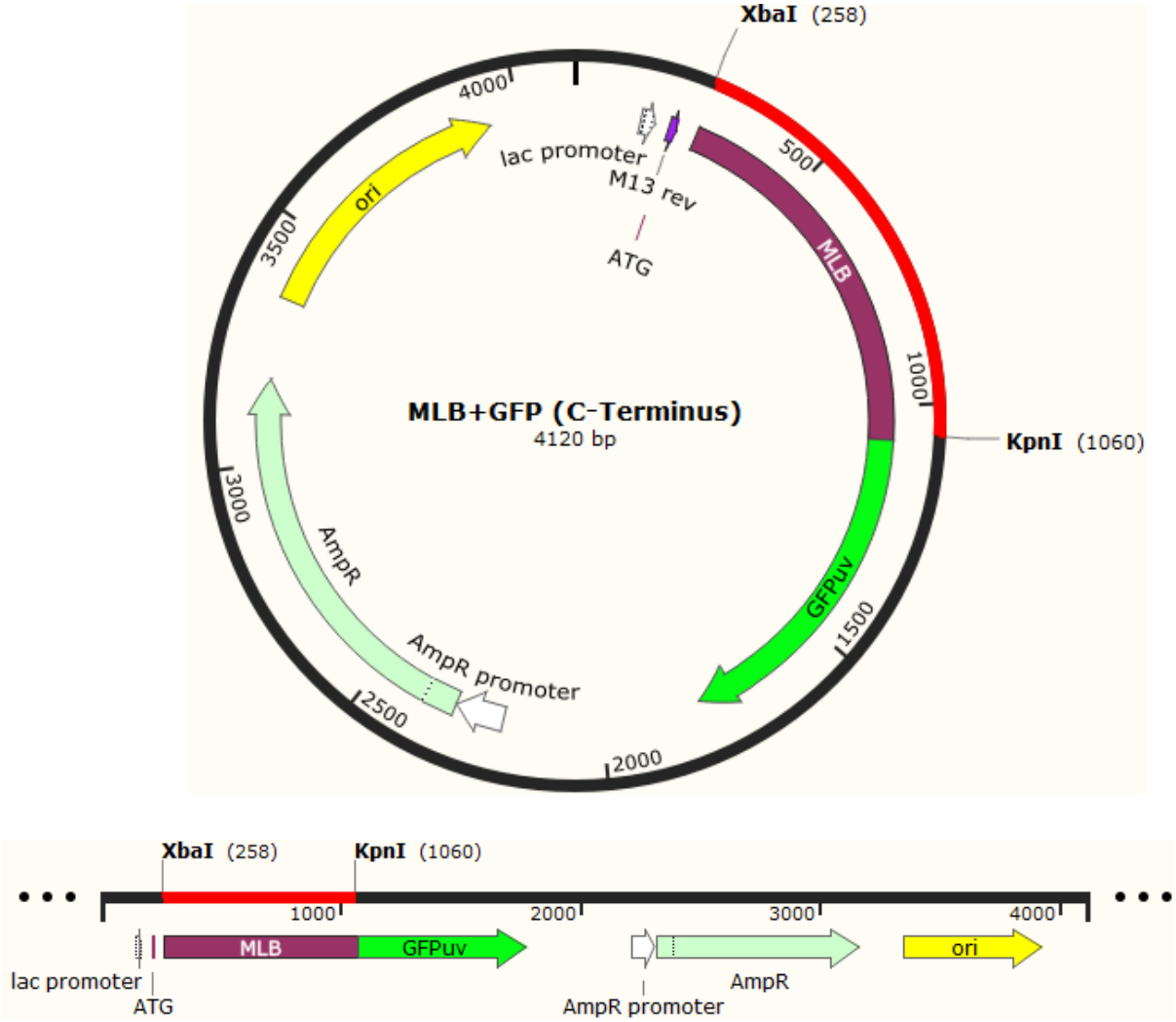
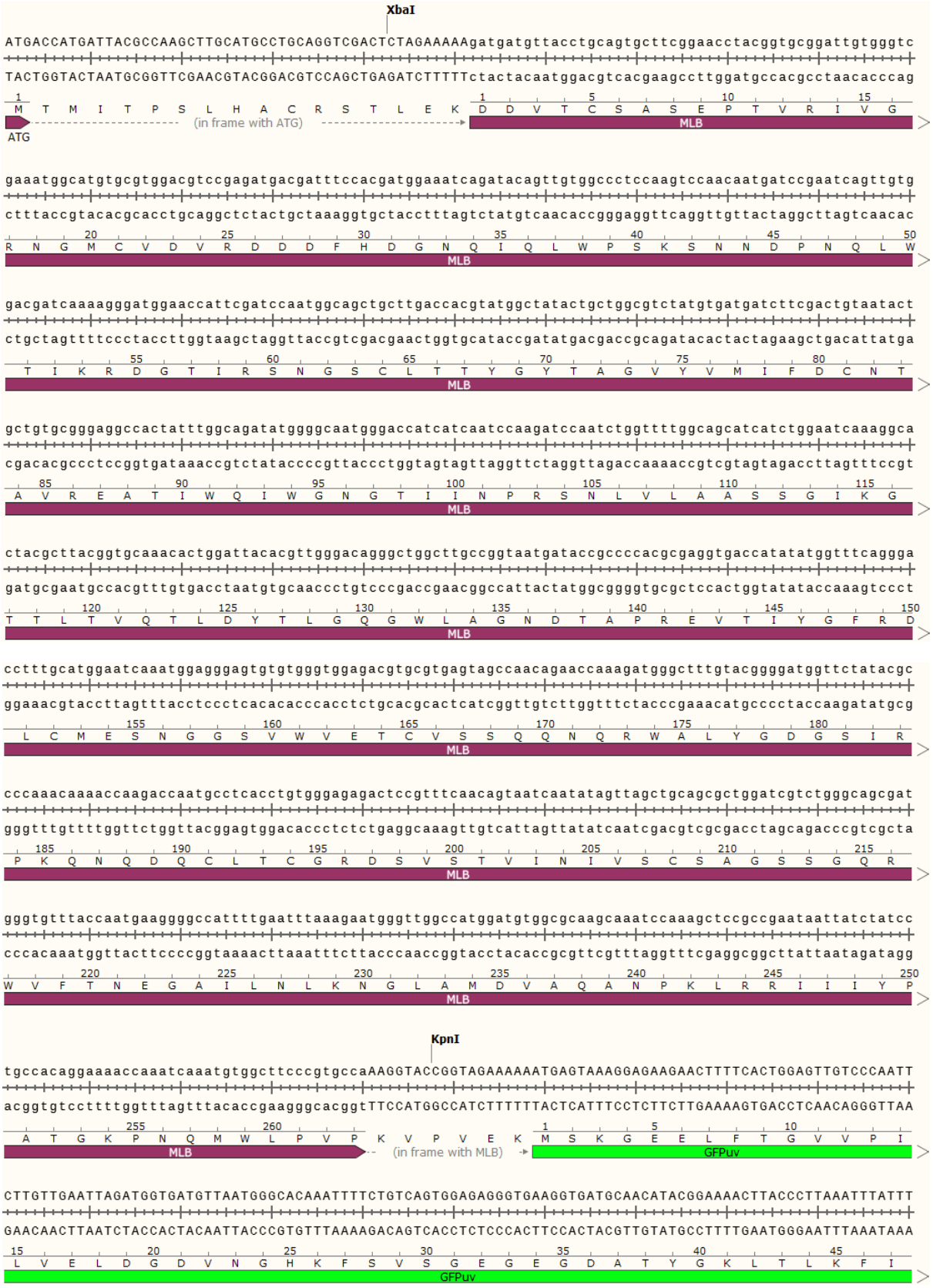
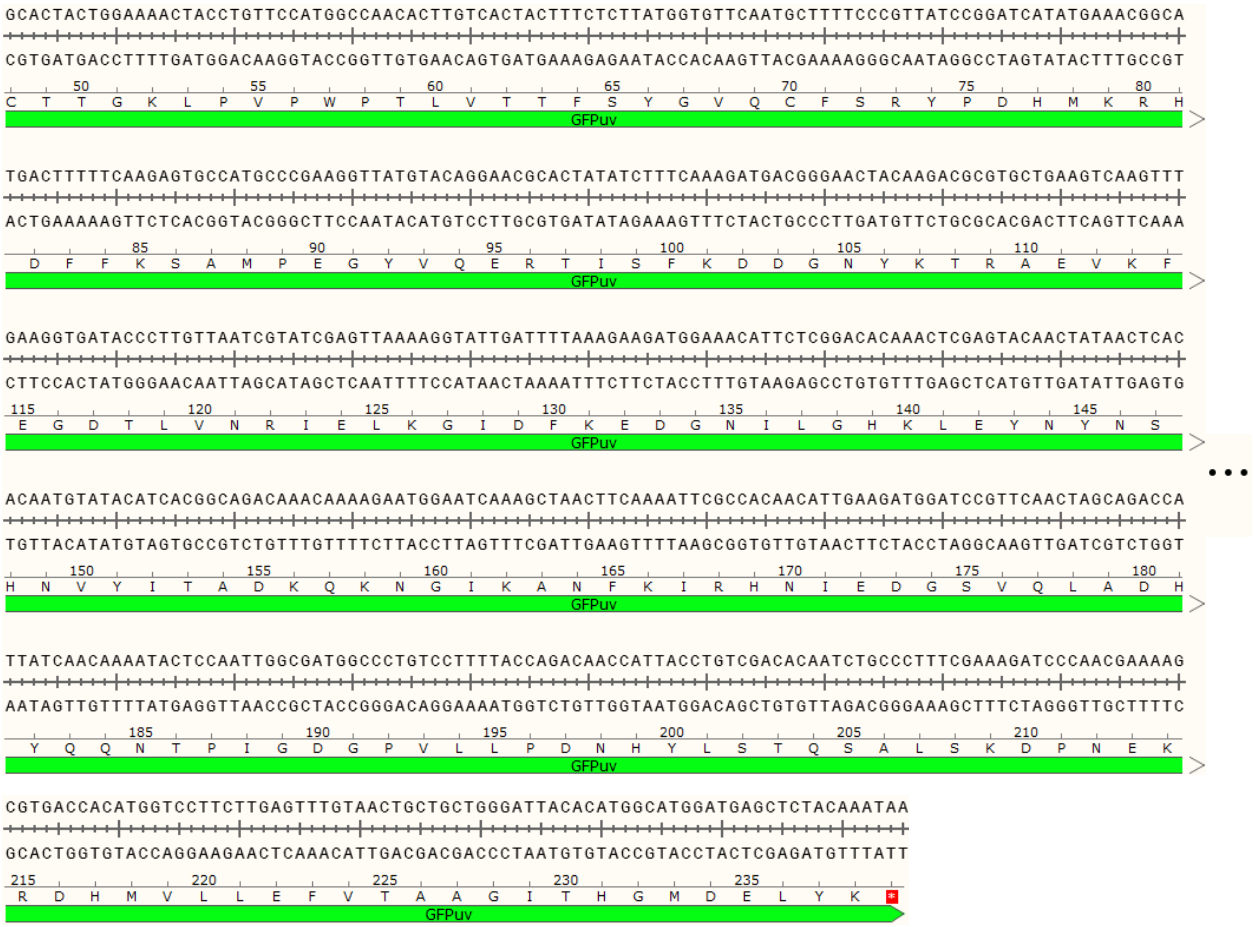


Figure 28. Cloned MLB chain (RED BAR) C-terminally fused to GFP protein sequence in the pGFPuv expression vector. The burgundy and green parts of the arrow represent MLB and GFP sequences respectively; The light green and yellow arrows represent the ampicillin resistant (AmpR) and signal of origin of replication (ori) sequences of pGFPuv vector.

Figure 29 below shows a detailed view of the *C*-terminus cloning strategy, note that the MLB sequence is in frame with the start codon (ATG) of the expression vector and with the GFP sequence.

Figure 29. The full nucleotide and deduced amino acid sequence of MLB chain *C*-terminally fused to GFP between XbaI and KpnI restriction sites in pGFPuv expression vector. The MLB sequence is indicated by a MAGENTA plot bar; The GFP sequence is indicated by a GREEN plot bar.





Following successful transformation of pGFPuv vector containing *N*- and *C*- terminally fused MLB sequence to GFP, several colonies were inoculated separately in 5ml LB medium and the plasmid DNA was extracted from the overnight culture. The diagnostic restriction digest of plasmids extracted was performed to determine the presence of targeted sequences. For *N*- and *C*- terminus fusion proteins a predicted DNA size of 801bp and 805bp was expected respectively and observed in figure 30.

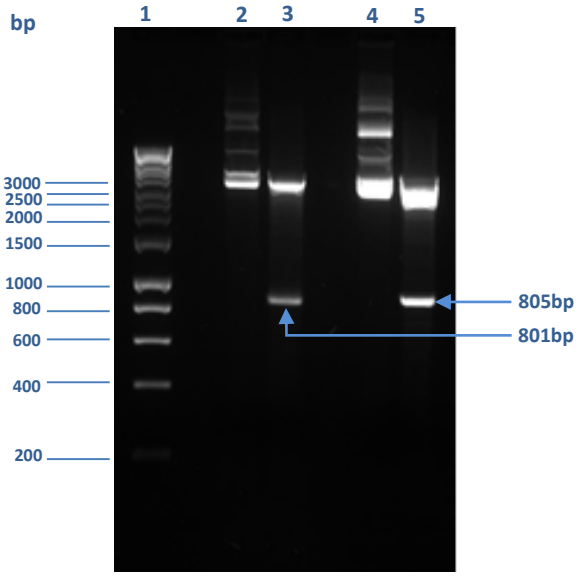


Figure 30. Restriction digest of pGFPuv plasmid with XbaI: KpnI and SacI:EcoRI to confirm the presence of MLB PCR product inserts. Lane 1: HyperLadder I (Bioline, UK); Lanes 3 and 5: the restriction digests of MLB *N*- and *C*-Terminally fused to GFP respectively; Lanes 2 and 4: the undigested negative controls.

Several sequencing reactions were also performed to determine the success of the fusions of the targeted MLB chain. Partial results are shown in figure 31 and figure 32 which show the sequencing quality and the full sequences can be seen in appendix 3.

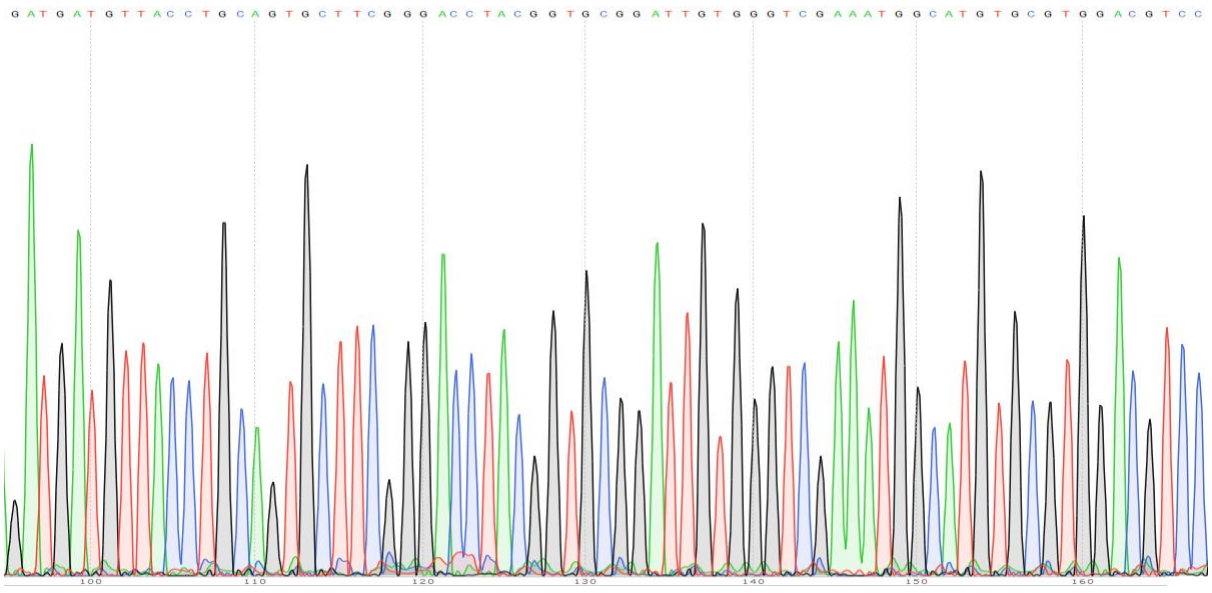


Figure 31. Chromatogram showing the sequence quality of mistletoe lectin B chain (MLB) *N*-Terminally fused to GFP in pGFPuv vector (Partial sequence data shown). The selected region is representative of overall sequencing quality and full sequence is shown in appendix 3. The Adenine (A), Guanine (G), Cytosine (C) and Thymine (T) codons are in green, black, blue and red colours respectively.

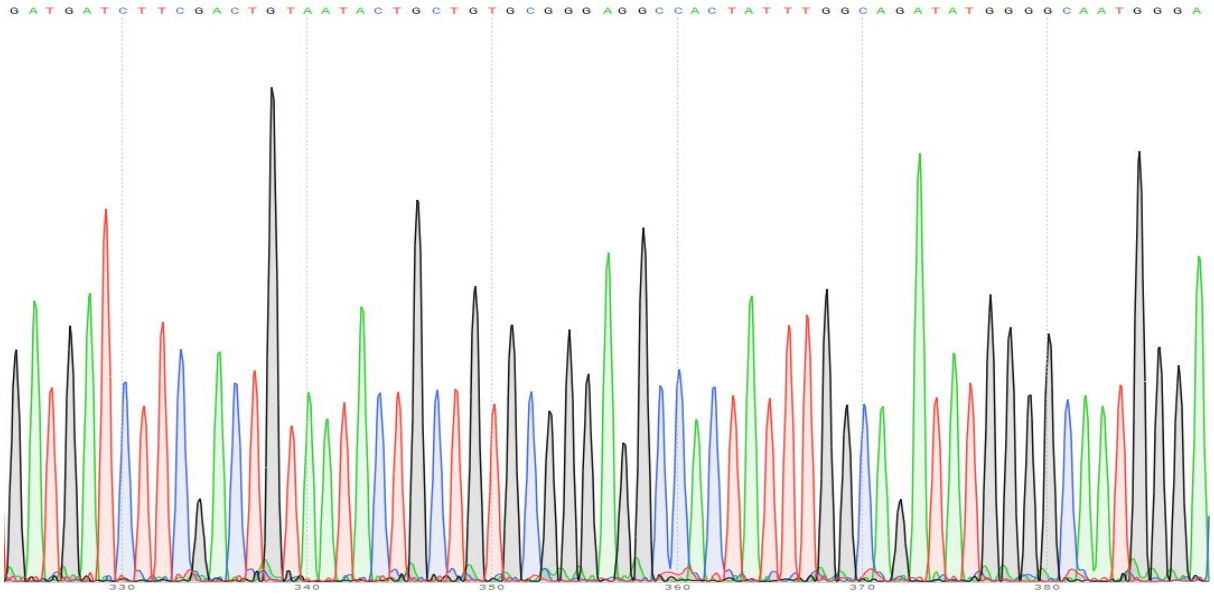


Figure 32. Chromatogram showing the sequence quality of mistletoe lectin B chain (MLB) *C*-Terminally fused to GFP GFP in pGFPuv vector (Partial sequence data shown). The selected region is representative of overall sequencing quality and full sequence is shown in appendix 3. The Adenine (A), Guanine (G), Cytosine (C) and Thymine (T) codons are in green, black, blue and red colours respectively.

The sequence data in figure 31 and figure 32 show clear peaks with very low levels of background 'noise'. Each peak is well defined and fully resolved, giving a clear indication that the data are reliable enough to go on for further experiments.

3.3.2 SNA-GFP and GNA-GFP Construct Design

Both *N*- and *C*- terminus fusion strategies were used to assemble constructs of the SNA B-chain (SNA-B) as positive control and GNA sequences as negative control fused to GFP sequence in pGFPuv expression vector. The sequences were both checked for the presence of internal restriction sites and to confirm that the restriction sites used were not present in the sequence using CLC Sequence Viewer program (Qiagen, UK). The cloning map and primers were designed using SnapGene software (SnapGene- USA).

3.3.2.1 *N*-terminus cloning

The *N*-terminus of SNA-B and GNA sequences were fused to the *C*-terminus of GFP sequence between SacI:EcoRI and SacI:SpeI restrictions sites respectively as shown in figure 33 and figure 34. The partial sequences of pGFPuv expression vector containing SNA-B and GNA are also shown in figure 35 and figure 36. Primers used to introduce the restriction sites to SNA-B and GNA are shown in Table 14.

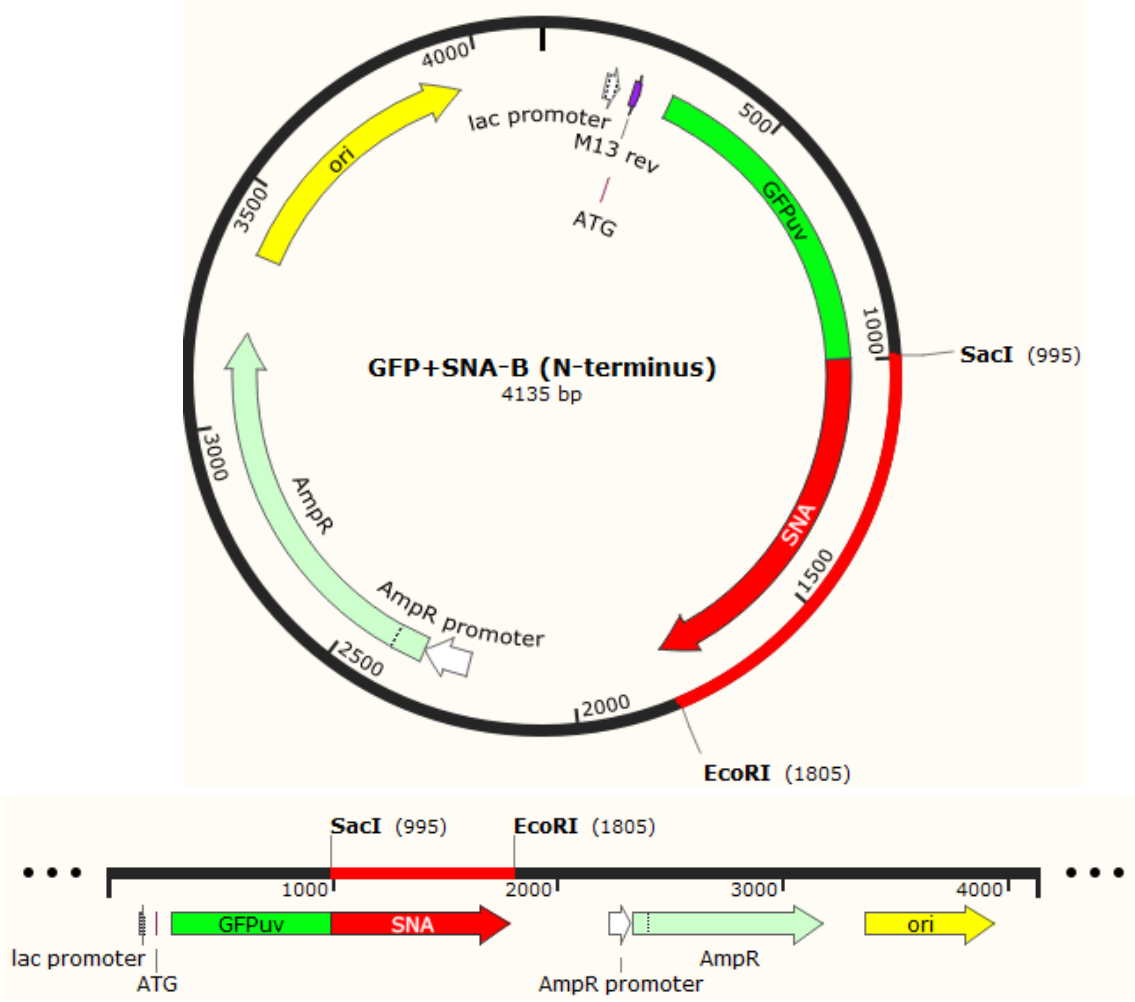


Figure 33. Cloned SNA B-chain *N*-terminally fused to GFP protein sequence in the pGFPuv expression vector. The GREEN and RED parts of the arrow represent GFP and SNA-B sequences respectively; The light green and yellow arrows represent the ampicillin resistant (AmpR) and signal of origin of replication (ori) sequences of pGFPuv vector.

.

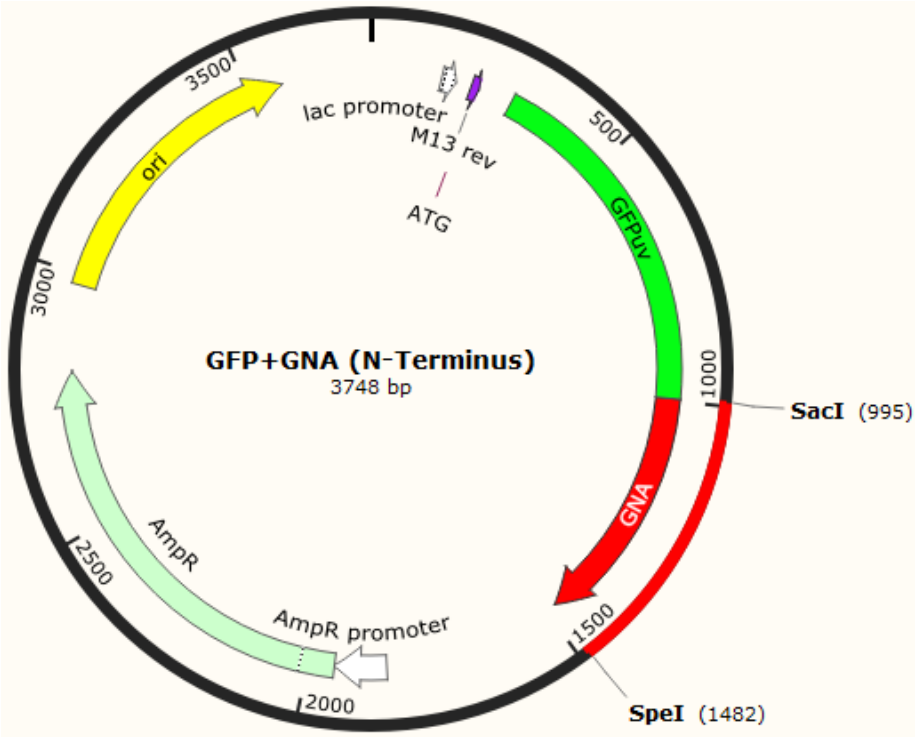
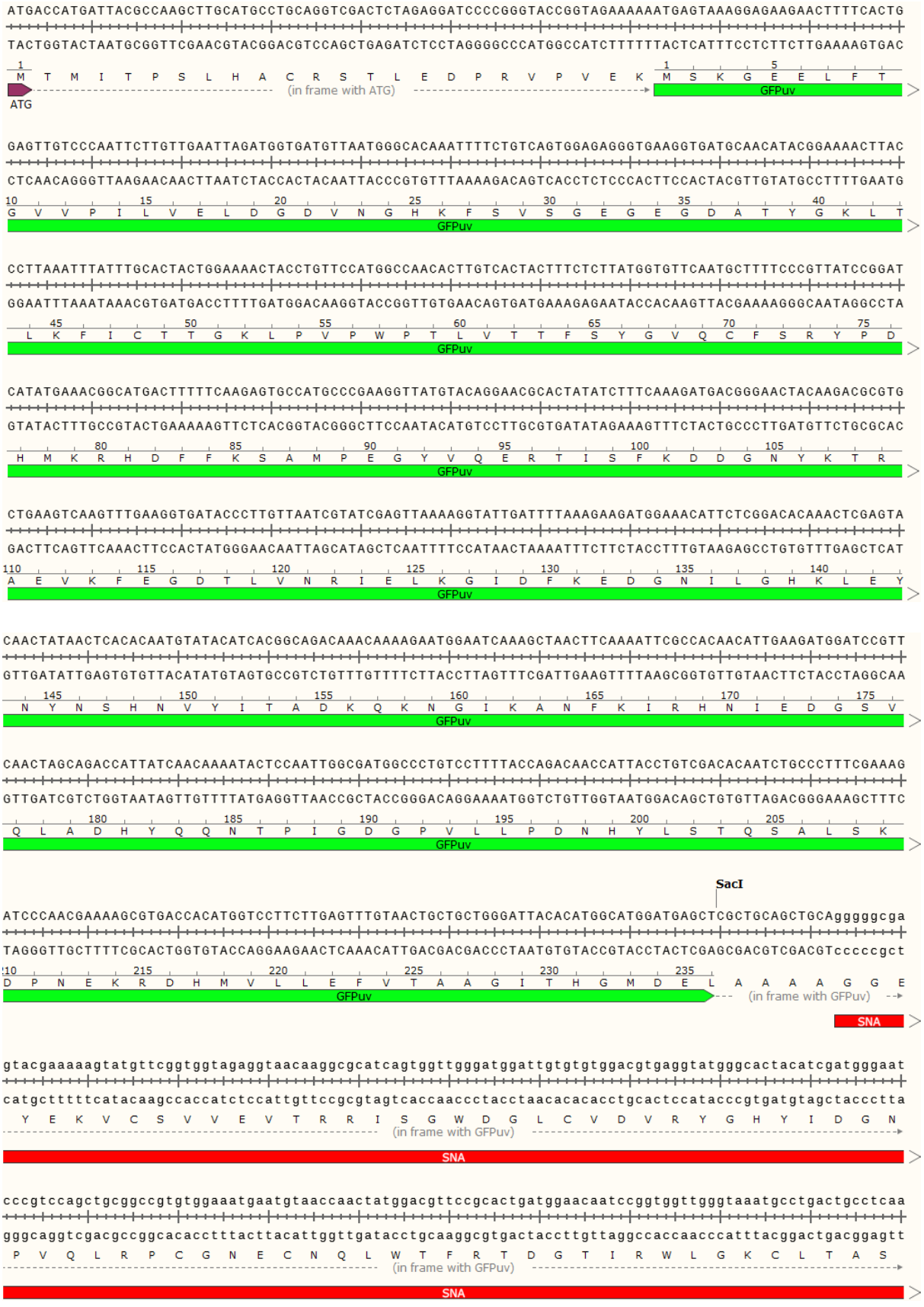


Figure 34. Cloned GNA *N*-terminally fused to GFP protein sequence in the pGFPuv expression vector. The GREEN and RED parts of the arrow represent GFP and GNA sequences respectively; The light green and yellow arrows represent the ampicillin resistant (AmpR) and signal of origin of replication (ori) sequences of pGFPuv vector.

Figure 35 and 36 show a detailed view of the *N*-terminus cloning strategy, note that the SNA-B and GNA sequences are in frame with the GFP sequence and the stop codon (TAA).

Figure 35. The full nucleotide and deduced amino acid sequence of SNA-B chain *N*-terminally fused to GFP between SacI and EcoRI restriction sites in pGFPuv expression vector. The GFP sequence is indicated by a GREEN plot bar; The SNA-B sequence is indicated by a RED plot bar.



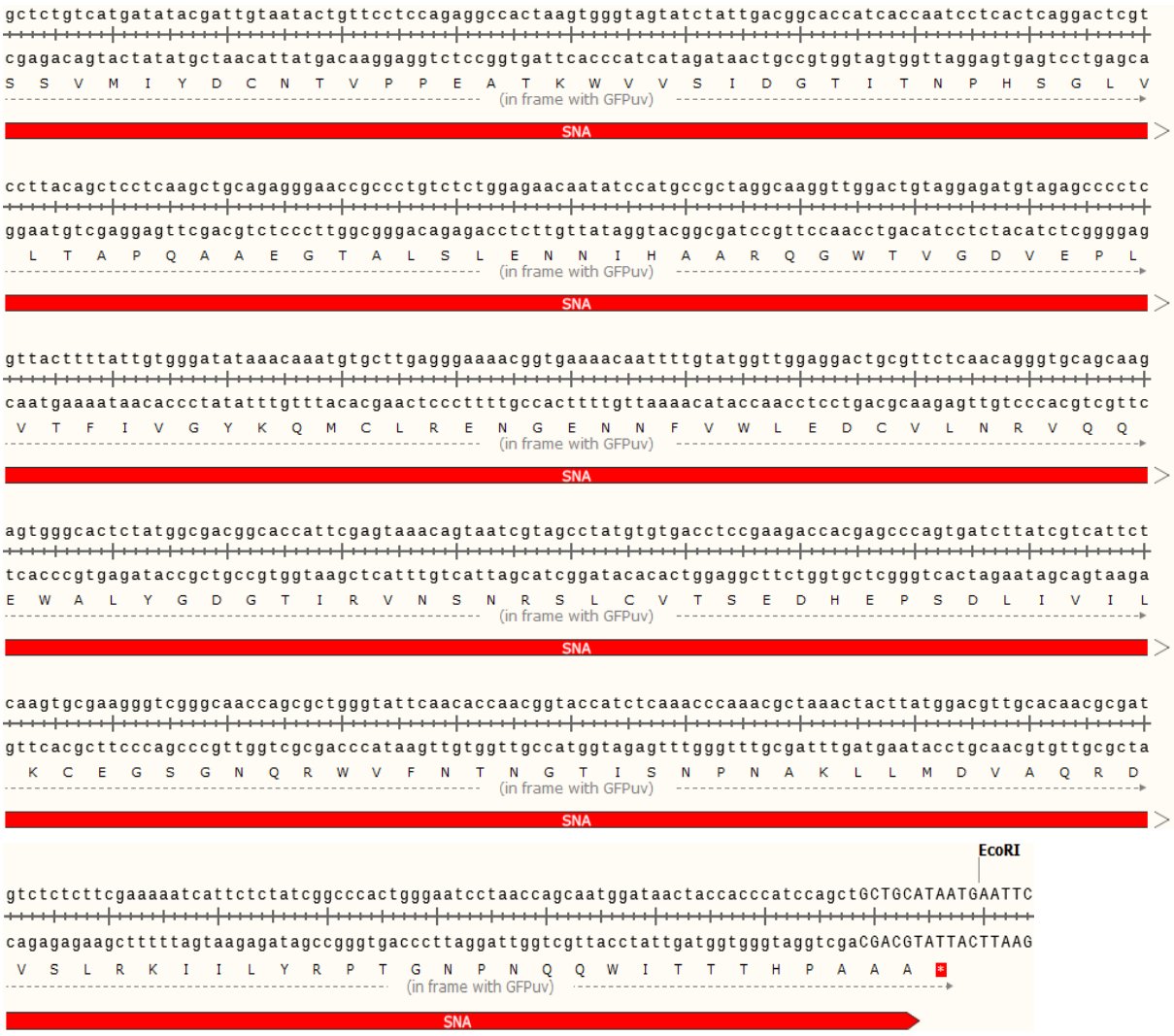
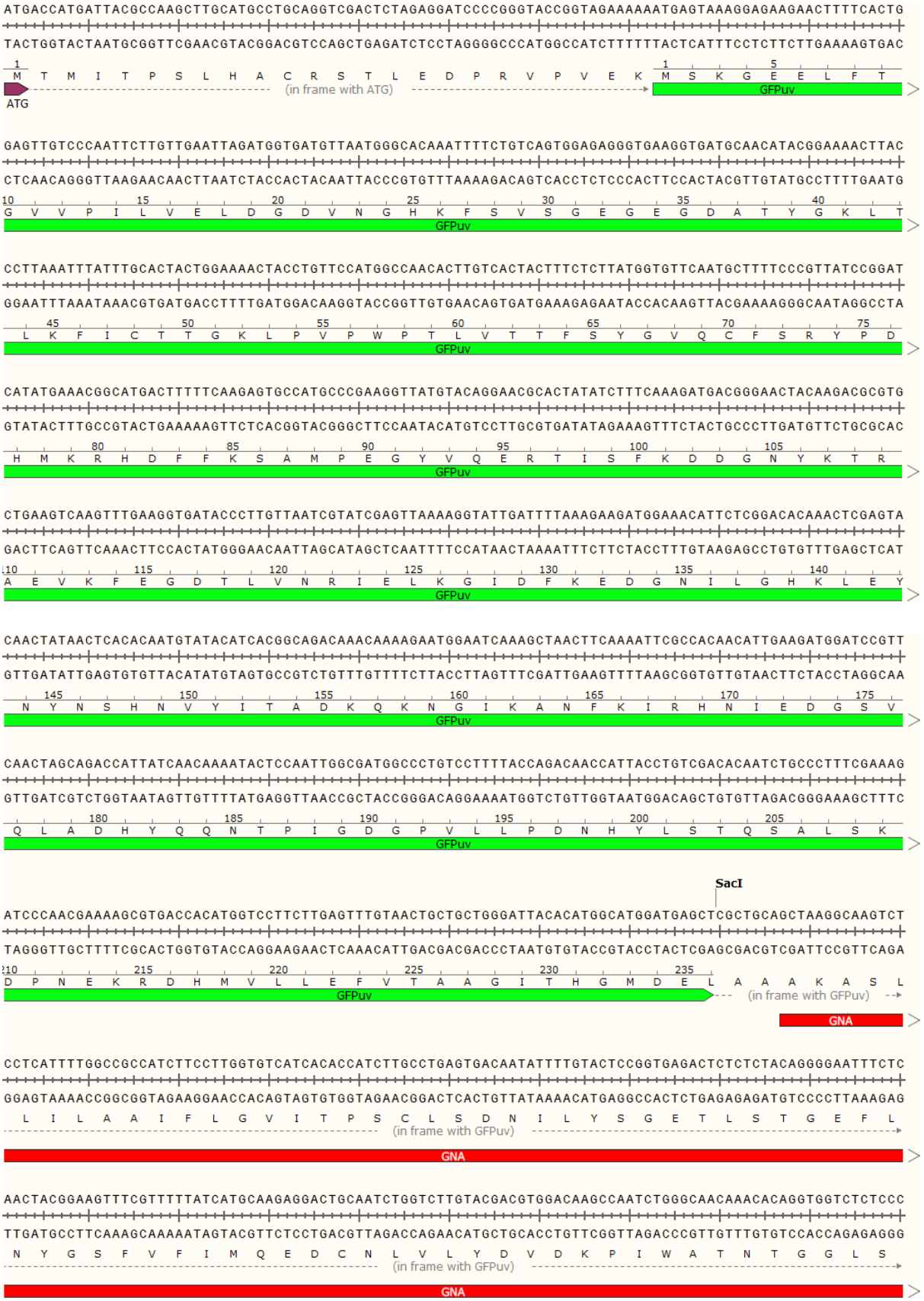

gctctgtcatgatatacgattgtaatactgttcctccagaggccactaagtgggtagtatctattgacggcaccatcaccaatcctcactcaggactcgt
cgagacagtactatatgctaacattatgacaaggaggtctccggtgattcacccatcatagataactgccgtggtagtggttaggagtgagtcctgagca
S S V M I Y D C N T V P P E A T K W V V S I D G T I T N P H S G L V
(in frame with GFPuv)
SNA
ccttacagctcctcaagctgcagagggaaccgccctgtctctggagaacaatatccatgccgctaggcaaggttggactgtaggagatgtagagcccctc
ggaatgtcgaggagttcgacgtctccccttggcgggacagagacctcttgttataggtacggcgatccgttccaacctgacatcctctacatctcggggag
L T A P Q A A E G T A L S L E N N I H A A R Q G W T V G D V E P L
(in frame with GFPuv)
SNA
gttacttttattgtgggatataaacaaatgtgcttgagggaaaacggtgaaaacaattttgtatggttggaggactgcgttctcaacagggtgcagcaag
caatgaaaataacaccctatatttgtttacacgaactccctttgccacttttgttaaaacataccaacctcctgacgcaagagttgtcccacgtcgttc
V T F I V G Y K Q M C L R E N G E N N F V W L E D C V L N R V Q Q
(in frame with GFPuv)
SNA
agtgggcactctatggcgacggcaccattcgagtaaacagtaatcgtagcctatgtgtgacctccgaagaccacgagcccagtgatcttatcgtcattct
tcacccgtgagataccgctgccgtggtaagctcatttgtcattagcatcggatacacactggaggcttctggtgctcgggtcactagaatagcagtaaga
E W A L Y G D G T I R V N S N R S L C V T S E D H E P S D L I V I L
(in frame with GFPuv)
SNA
caagtgcgaagggtcgggcaaccagcgctgggtattcaacaccaacggtaccatctcaaacccaaacgctaaactacttatggacgttgcacaacgcgat
gttcacgcttcccagcccgttggtcgcgacccataagttgtggttgccatggtagagtttgggtttgcgatttgatgaatacctgcaacgtgttgcgcta
K C E G S G N Q R W V F N T N G T I S N P N A K L L M D V A Q R D
(in frame with GFPuv)
SNA
EcoRI
gtctctcttcgaaaaatcattctctatcggcccactgggaatcctaaccagcaatggataactaccacccatccagctGCTGCATAATGAATTC
cagagagaagctttttagtaagagatagccgggtgacccttaggattggtcgttacctattgatggtgggtaggtcgaCGACGTATTACTTAAG
V S L R K I I L Y R P T G N P N Q Q W I T T T H P A A A A *
(in frame with GFPuv)
SNA

Figure 36. The full nucleotide and deduced amino acid sequence of GNA *N*-terminally fused to GFP between SacI and SpeI restriction sites in pGFPuv expression vector. The GFP sequence is indicated by a GREEN plot bar; The GNA sequence is indicated by a RED plot bar.




GTAGCTGCTTCCTCAGCATGCAGACTGATGGGAACCTCGTGGTGTACAACCCATCGAACAAACCGATTTGGGCAAGCAACACTGGAGGCCAAAATGGGAA
CATCGACGAAGGAGTCGTACGTCTGACTACCCTTGGAGCACCACATGTTGGGTAGCTTGTTTGGCTAAACCCGTTCGTTGTGACCTCCGGTTTTACCCTT
R S C F L S M Q T D G N L V V Y N P S N K P I W A S N T G G Q N G N
(in frame with GFPuv)
GNA
TTACGTGTGCATCCTACAGAAGGATAGGAATGTTGTGATCTACGGAACTGATCGTTGGGCTACTGGAACTCACACCGGACTTGTTGGAATTCCCGCATCG
AATGCACACGTAGGATGTCTTCCTATCCTTACAACACTAGATGCCTTGACTAGCAACCCGATGACCTTGAGTGTGGCCTGAACAACCTTAAGGGCGTAGC
Y V C I L Q K D R N V V I Y G T D R W A T G T H T G L V G I P A S
(in frame with GFPuv)
GNA
SpeI
CCACCCTCAGAGAAATATCCTACTGCTGGAAAGATAAAGCTTGTGACGGCAAAGGCTGCATAACTACTAGT
GGTGGGAGTCTCTTTATAGGATGACGACCTTTCTATTTCGAACACTGCCGTTTCCGACGTATTGATGATCA
P P S E K Y P T A G K I K L V T A K A A *
(in frame with GFPuv)
GNA

3.3.2.2 *C*-terminus cloning

The *C*-terminus of SNA-B and GNA sequences were fused to the *N*-terminus of GFP sequence between HindIII:XbaI and XbaI:KpnI restrictions sites respectively as shown in figure 37 and figure 38. The partial sequences of pGFPuv expression vector containing SNA-B and GNA genes are also shown in figure 39 and figure 40. Primers used to introduce the restriction sites to SNA-B and GNA are shown in Table 14.

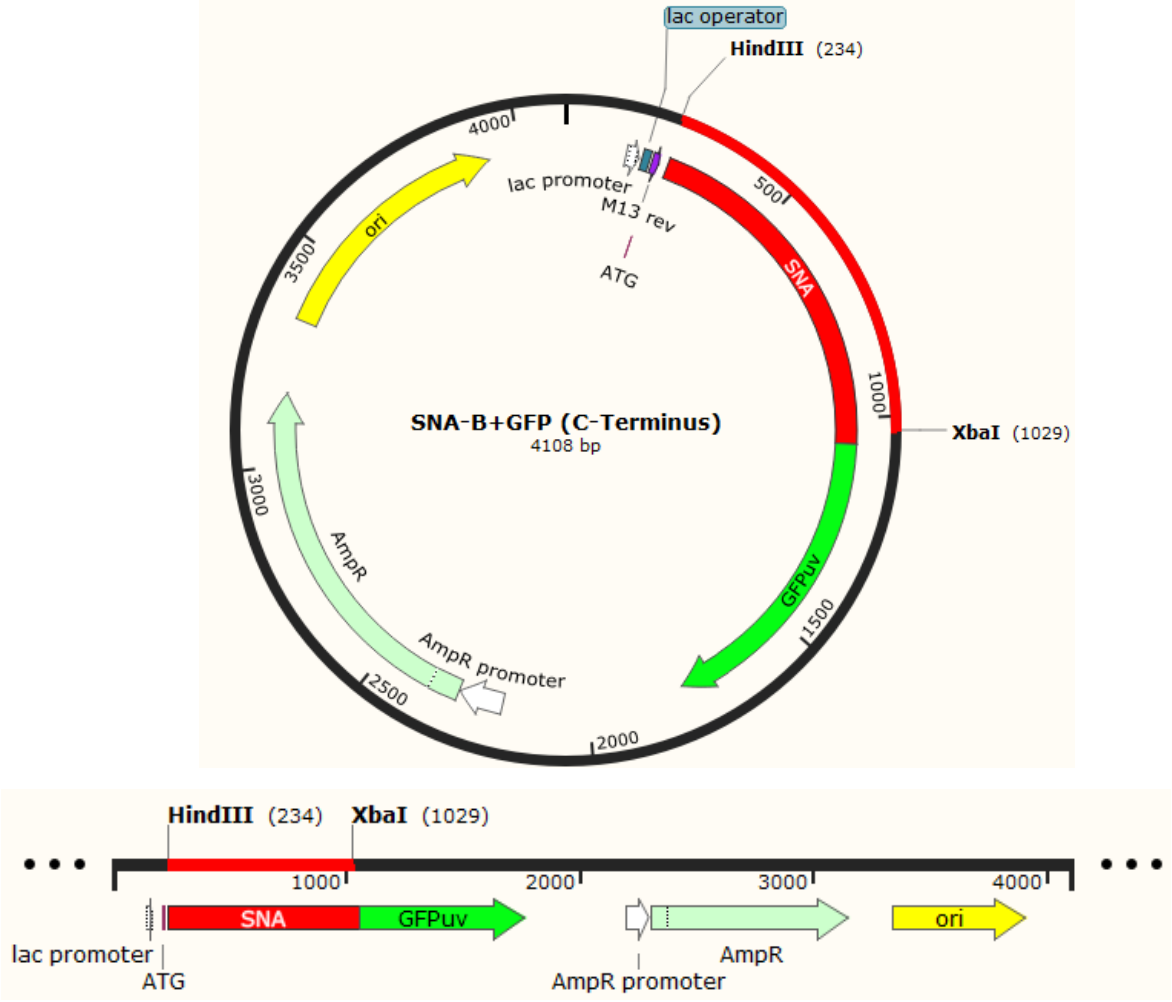


Figure 37. Cloned SNA-B *C*-terminally fused to GFP protein sequence in the pGFPuv expression vector. The RED and GREEN parts of the arrow represent SNA-B and GFP sequences respectively; The light green and yellow arrows represent the ampicillin resistant (AmpR) and signal of origin of replication (ori) sequences of pGFPuv vector.

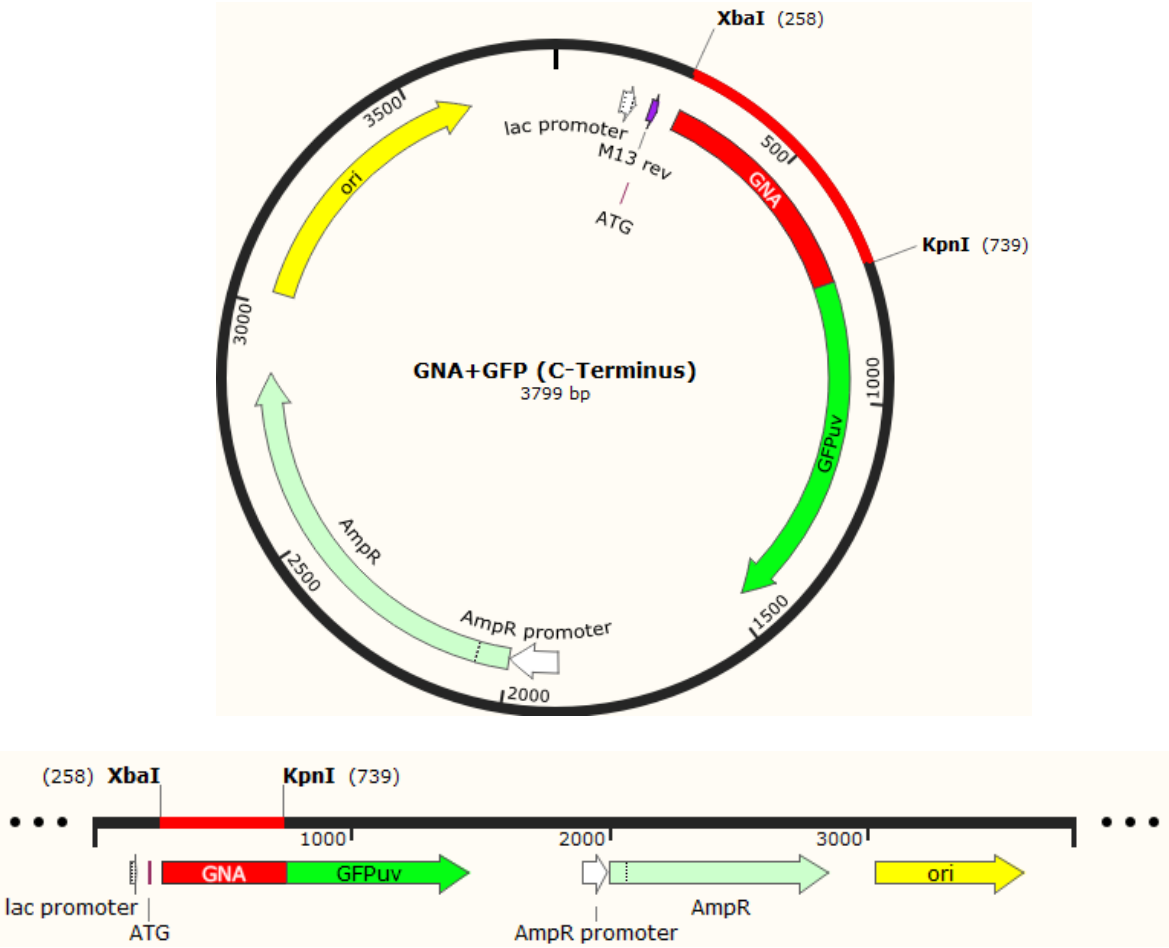


Figure 38. Cloned GNA *C*-terminally fused to GFP protein sequence in the pGFPuv expression vector. The RED and GREEN parts of the arrow represent GNA and GFP sequences respectively; The light green and yellow arrows represent the ampicillin resistant (AmpR) and signal of origin of replication (ori) sequences of pGFPuv vector.

Figure 39, 40 below show a detailed view of the *C*-terminus cloning strategy, note that the SNA-B and GNA sequences are in frame with the start codon (ATG) of the expression vector and with the GFP sequence.

Figure 39. The full nucleotide and deduced amino acid sequence of SNA-B chain *C*-terminally fused to GFP between HindIII and XbaI restriction sites in pGFPuv expression vector. The SNA-B sequence is indicated by a RED plot bar; The GFP sequence is indicated by a GREEN plot bar.

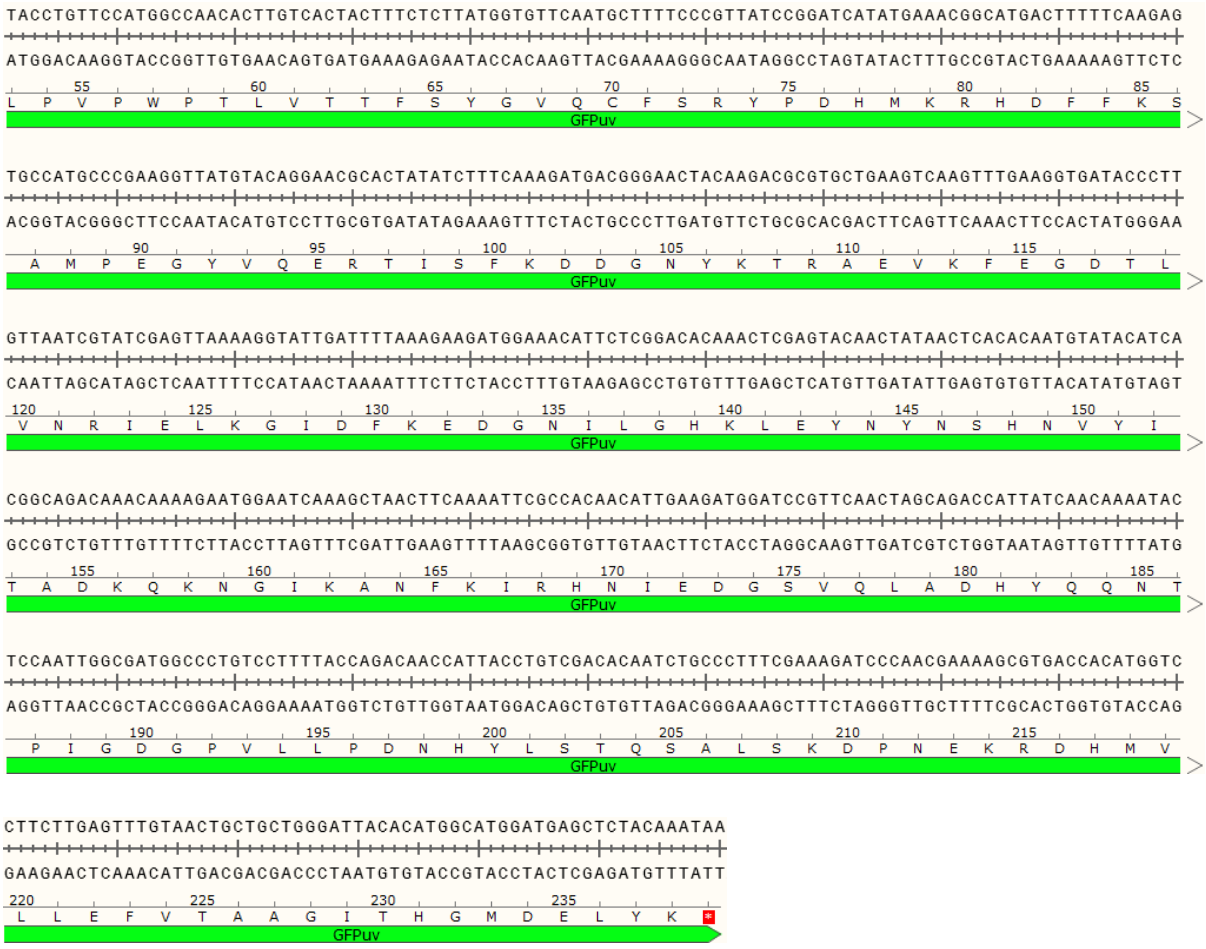
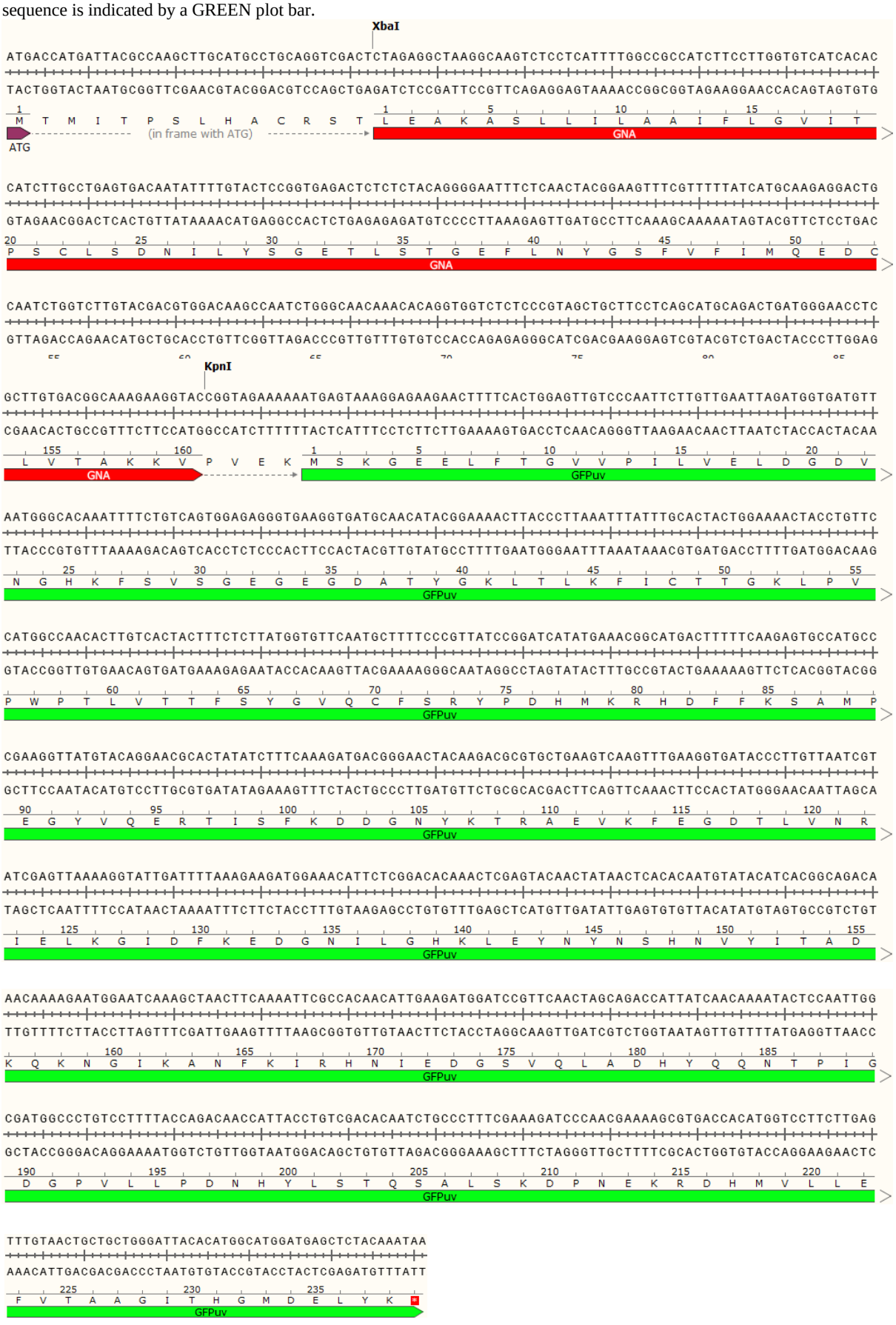

TACCTGTTCCATGGCCAACACTTGTCACTACTTTCTCTTATGGTGTTCAATGCTTTTCCCGTTATCCGGATCATATGAAACGGCATGACTTTTTCAAGAG
ATGGACAAGGTACCGGTTGTGAACAGTGATGAAAGAGAATACCACAAGTTACGAAAAGGGCAATAGGCCTAGTATACTTTGCCGTACTGAAAAAGTTCTC
55 60 65 70 75 80 85
L P V P W P T L V T T F S Y G V Q C F S R Y P D H M K R H D F F K S
GFPuv
TGCCATGCCCGAAGGTTATGTACAGGAACGCACTATATCTTTCAAAGATGACGGGAACTACAAGACGCGTGCTGAAGTCAAGTTTGAAGGTGATACCCTT
ACGGTACGGGCTTCCAATACATGTCCTTGCGTGATATAGAAAGTTTCTACTGCCCTTGATGTTCTGCGCACGACTTCAGTTCAAACTTCCACTATGGGAA
90 95 100 105 110 115
A M P E G Y V Q E R T I S F K D D G N Y K T R A E V K F E G D T L
GFPuv
GTTAATCGTATCGAGTTAAAAGGTATTGATTTTAAAGAAGATGGAAACATTCTCGGACACAAACTCGAGTACAACTATAACTCACACAATGTATACATCA
CAATTAGCATAGCTCAATTTTCCATAACTAAAATTTCTTCTACCTTTGTAAGAGCCTGTGTTTGAGCTCATGTTGATATTGAGTGTGTTACATATGTAGT
120 125 130 135 140 145 150
V N R I E L K G I D F K E D G N I L G H K L E Y N Y N S H N V Y I
GFPuv
CGGCAGACAAACAAAAGAATGGAATCAAAGCTAACTTCAAAATTCGCCACAACATTGAAGATGGATCCGTTCAACTAGCAGACCATTATCAACAAAATAC
GCCGTCTGTTTGTTTTCTTACCTTAGTTTCGATTGAAGTTTTAAGCGGTGTTGTAACTTCTACCTAGGCAAGTTGATCGTCTGGTAATAGTTGTTTTATG
155 160 165 170 175 180 185
T A D K Q K N G I K A N F K I R H N I E D G S V Q L A D H Y Q Q N T
GFPuv
TCCAATTGGCGATGGCCCTGTCCTTTTACCAGACAACCATTACCTGTCGACACAATCTGCCCTTTCGAAAGATCCCAACGAAAAGCGTGACCACATGGTC
AGGTTAACCGCTACCGGGACAGGAAAATGGTCTGTTGGTAATGGACAGCTGTGTTAGACGGGAAAGCTTTCTAGGGTTGCTTTTCGCACTGGTGTACCAG
190 195 200 205 210 215
P I G D G P V L L P D N H Y L S T Q S A L S K D P N E K R D H M V
GFPuv
CTTCTTGAGTTTGTAACTGCTGCTGGGATTACACATGGCATGGATGAGCTCTACAAATAA
GAAGAACTCAAACATTGACGACGACCCTAATGTGTACCGTACCTACTCGAGATGTTTATT
220 225 230 235
L L E F V T A A G I T H G M D E L Y K *
GFPuv

Figure 40. The full nucleotide and deduced amino acid sequence of GNA *C*-terminally fused to GFP between XbaI and KpnI restriction sites in pGFPuv expression vector. The GNA sequence is indicated by a RED plot bar; The GFP sequence is indicated by a GREEN plot bar.



Following successful transformation SNA-B and GNA plasmid DNA was extracted and the diagnostic restriction digest of the plasmid DNA was performed to confirm the presence of targeted sequences (Figure 41)

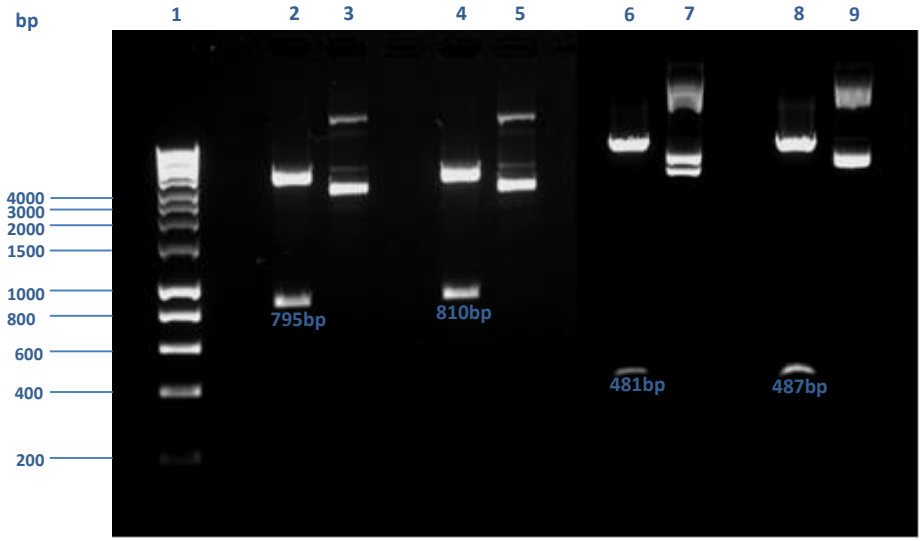


Figure 41. Restriction digest of pGFPuv plasmid to confirm the presence of SNA-B and GNA product inserts. Lane 1: HyperLadder I (Bioline, UK); Lanes 2 and 6: the restriction digests of SNA-B and GNA samples *C*-Terminally fused to GFP respectively; lanes 4 and 8: the restriction digests of SNA-B and GNA samples *N*-Terminally fused to GFP respectively; Lanes 3, 5, 7 and 9: the undigested negative controls.

Figure 41 confirmed the predicted band size of *N*- and *C*- terminus fusion proteins which are 810bp and 795bp for SNA-B fusions and 487bp and 481bp for GNA fusions respectively.

Several sequencing reactions were also performed to determine the quality of the cloned sequences in the recombinant constructs. Partial results are as shown in figure 42, 43, 44, 45 and the full sequences can be seen in appendix 3.

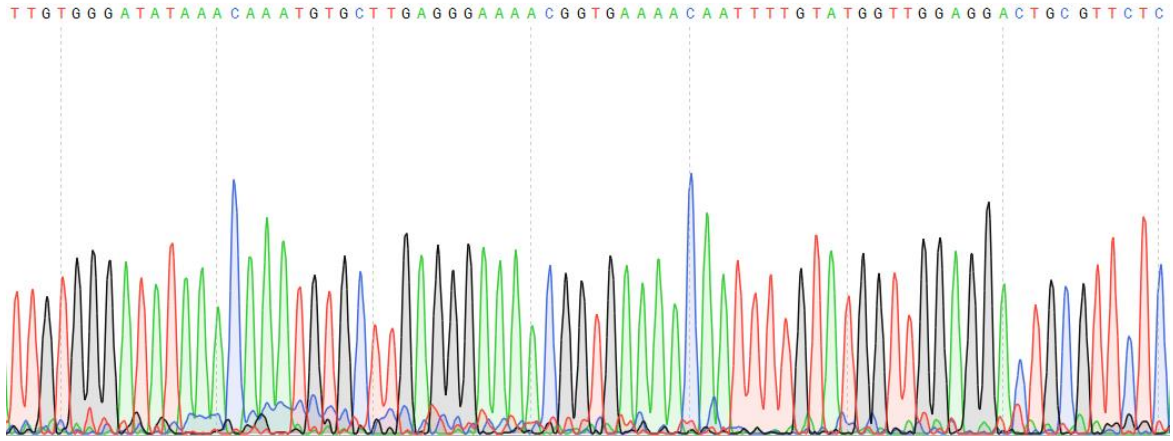


Figure 42. Chromatogram showing the sequence quality of SNA B-chain (SNA-B) *N*-Terminally fused to GFP in pGFPuv vector (Partial sequence data shown). The selected region is representative of overall sequencing quality and full sequence is shown in appendix 3. The Adenine (A), Guanine (G), Cytosine (C) and Thymine (T) codons are in green, black, blue and red colours respectively.

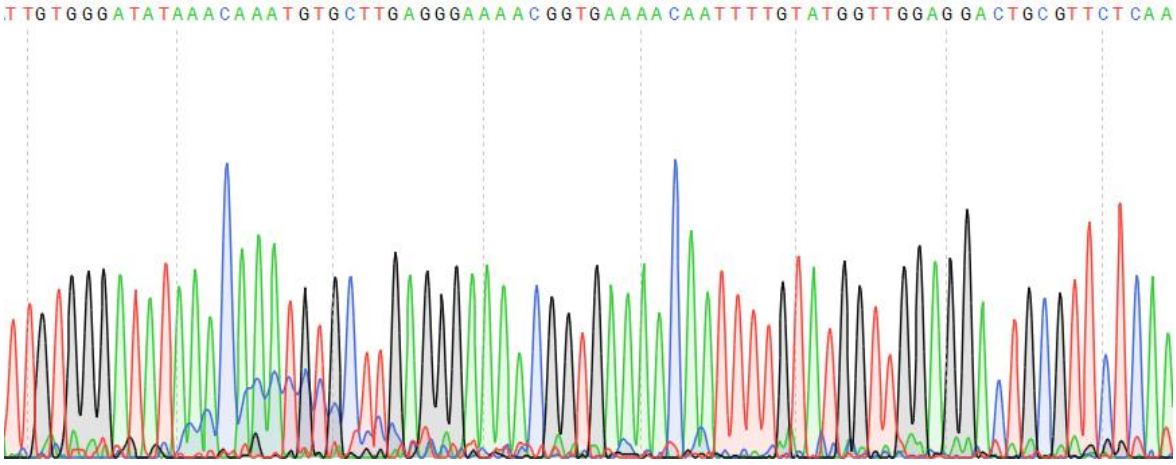


Figure 43. Chromatogram showing the sequence quality of SNA B-chain (SNA-B) *C*-Terminally fused to GFP in pGFPuv vector (Partial sequence data shown). The selected region is representative of overall sequencing quality and full sequence is shown in appendix 3. The Adenine (A), Guanine (G), Cytosine (C) and Thymine (T) codons are in green, black, blue and red colours respectively.

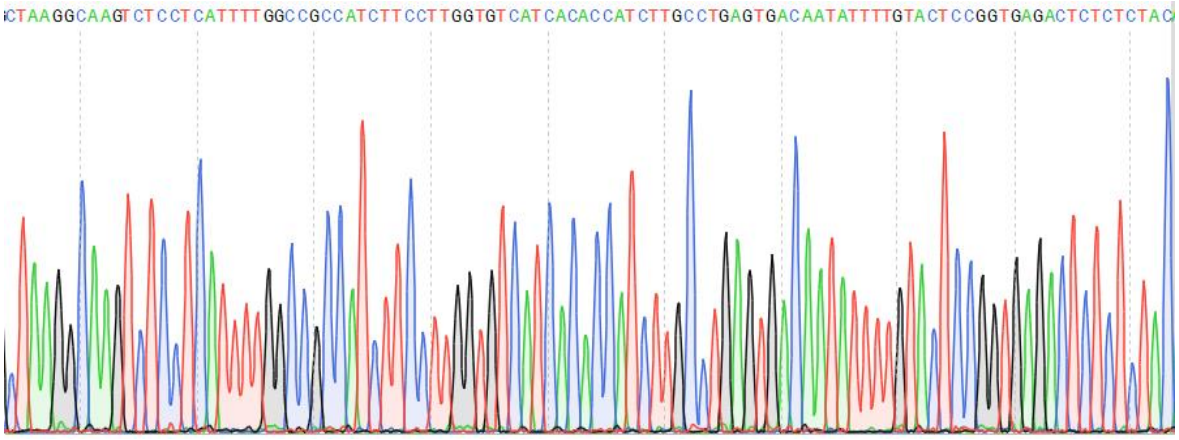


Figure 44. Chromatogram showing the sequence quality of GNA *N*-Terminally fused to GFP in pGFPuv vector (Partial sequence data shown). The selected region is representative of overall sequencing quality and full sequence is shown in appendix 3. The Adenine (A), Guanine (G), Cytosine (C) and Thymine (T) codons are in green, black, blue and red colours respectively.

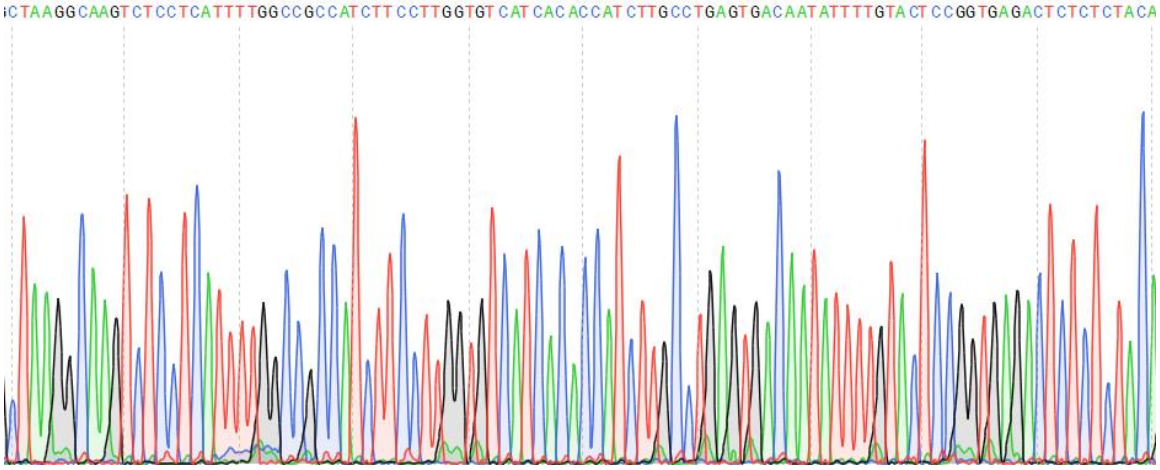


Figure 45. Chromatogram showing the sequence quality of GNA *C*-Terminally fused to GFP in pGFPuv vector (Partial sequence data shown). The selected region is representative of overall sequencing quality and full sequence is shown in appendix 3. The Adenine (A), Guanine (G), Cytosine (C) and Thymine (T) codons are in green, black, blue and red colours respectively.

3.3.3 Transformation of BL21(DE3) *E. coli* Cell Line with pGFPuv Vector Containing MLB, SNA-B and GNA N- and C- Terminally Fused to GFP

The BL21(DE3) *E. coli* strain was transformed with both *N*- and *C*-Terminally fused MLB, SNA-B and GNA to GFP constructs using Chemically Competent BL21(DE3) and the heat shock method. The successfully grown colonies were screened under standard UV light (395nm-509nm) to visualise the emission of GFP in the colonies and successful transformation shown by GFP activity (Figure 46 and Figure 47). The successfully transformed cells were grown overnight in 5mL of liquid LB broth medium with 100µg/mL Ampicillin at 37°C with shaking at 250 rpm.

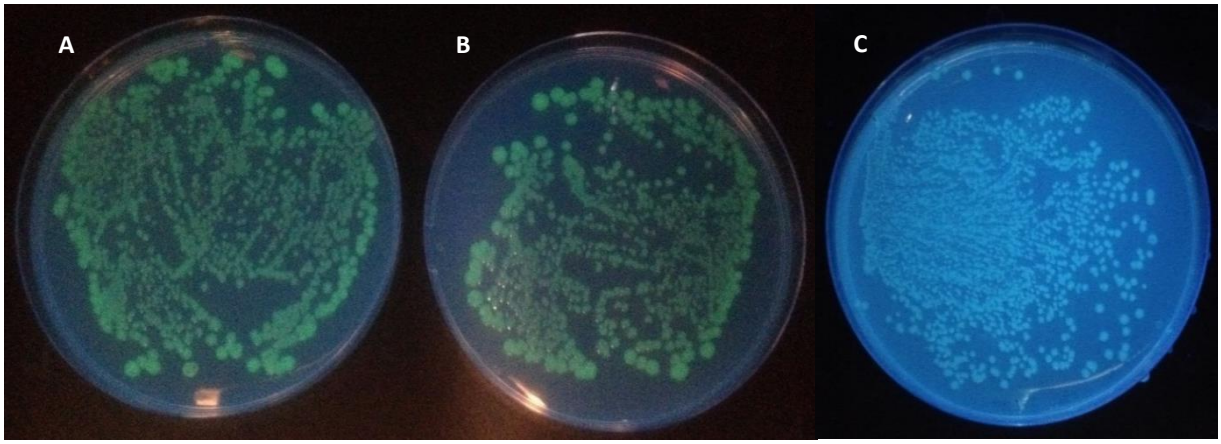


Figure 46. Cloned *N*-terminally fused MLB, SNA-B and GNA to GFP conducted on an Agar and LB plates. Plate A: MLB; Plate B: SNA-B; Plate C: GNA. The emission of fused GFP to MLB and SNA-B means that the GFP is active and gives an indication that the fusion proteins are active. The fused GFP to GNA however showed no emission.

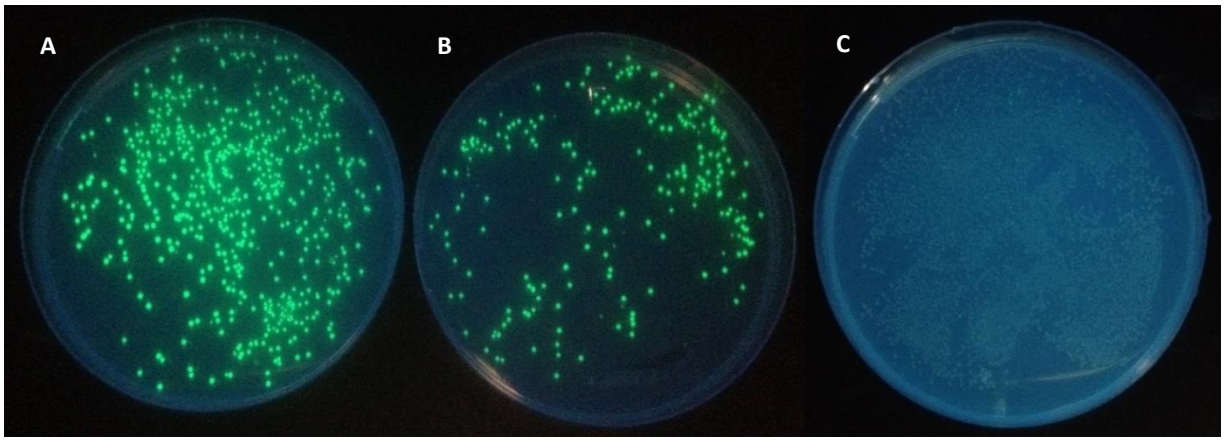


Figure 47. Cloned *C*-terminally fused MLB, SNA-B and GNA to GFP conducted on an Agar and LB plates. Plate A: MLB; Plate B: SNA-B; Plate C: GNA. The emission of fused GFP to MLB and SNA-B means that the GFP is active and gives an indication that the fusion proteins are active. The fused GFP to GNA however showed no emission.

Figure 46 and 47 shows that the fusion proteins made from *C*-terminally fused MLB and SNA-B to GFP has higher GFP activity (based on visualization of GFP) than the fusions made from *N*-Terminally fused MLB and SNA-B to GFP. The emission of GFP is higher in the former which is an indication that GFP was properly folded, similarly for MLB and SNA-B. However, the activity of GFP was almost absent in both *N*- and *C*-terminally fused GNA to GFP which indicates that the fusion protein was either not expressed or not correctly processed in *E. coli*. Therefore, protein expression was performed just for MLB and SNA-B fusion proteins.

3.3.4 Expression of Recombinant MLB+GFP and SNA-B+GFP

The strain of *E. coli* BL21(DE3) harbouring the plasmid pGFPuv containing the MLB and SNA-B genes *N*- and *C*- terminally fused to GFP to be expressed were grown in LB Broth medium (Luria Low Salt) (Sigma-Aldrich, UK) for 18 hours at 30^{o}C after which the total protein was harvested. The emission of GFP in the pellets made from harvested *E. coli* cells show that the fused GFP protein in the fusion proteins is active which gives an indication that the whole fusion proteins are active (Figure 48).

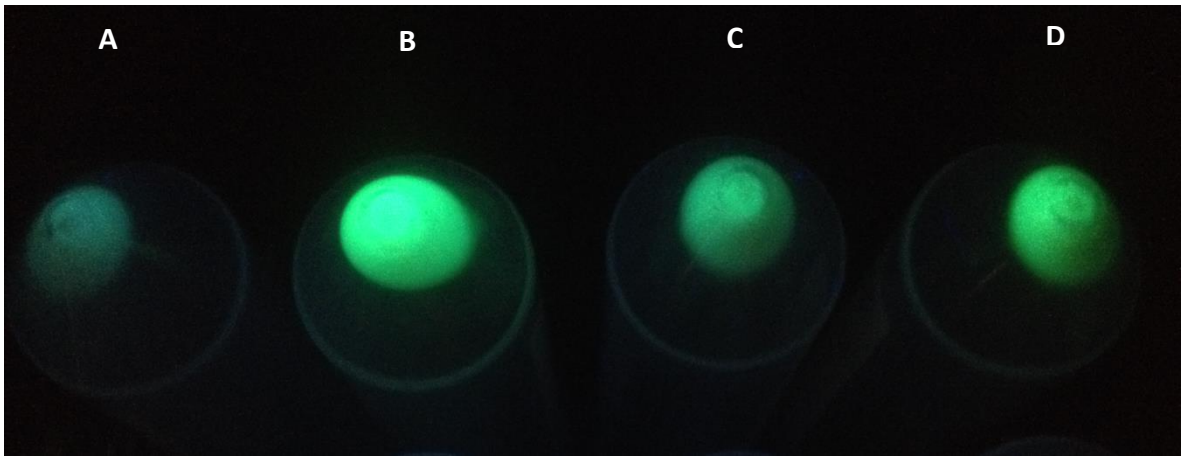


Figure 48. Pelleted *E.coli* cells contain expressed Lectin+GFP fusions. Tube A: *N*-terminus MLB+GFP fusion; Tube B: *C*-terminus MLB+GFP fusion; Tube C: *N*-terminus SNA-B+GFP fusion; Tube D: *C*-terminus SNA-B+GFP fusion.

The harvested proteins from each set of *N*- and *C*- terminus MLB and SNA-B cultures was analysed on 12% SDS-PAGE gel and the predicted band sizes of 58.41 kDa and 58.14 kDa for *N*- and *C*- terminus MLB fusions and 58.77 kDa and 58.08kDa for *N*- and *C*- terminus MLB fusions were detected respectively on the gel (Figure 49).

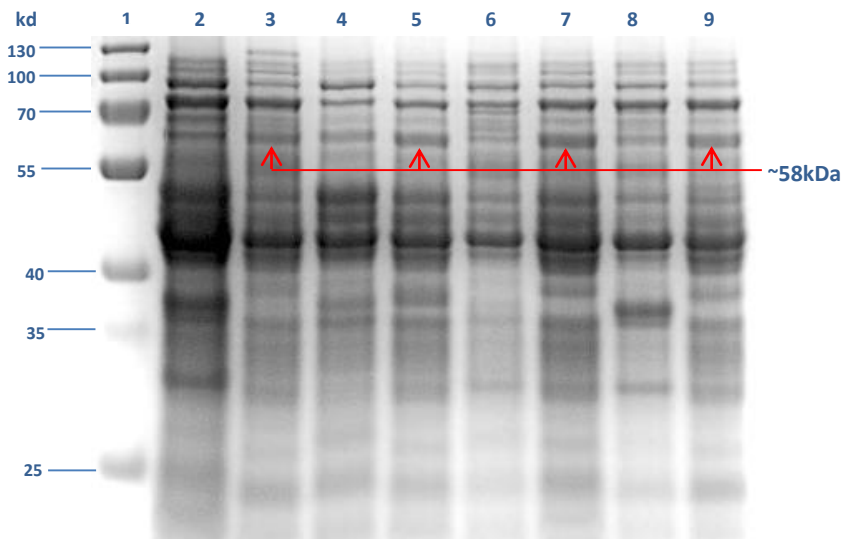


Figure 49. SDS PAGEof expressed *N*- and *C*-terminally fused MLB and SNA-B to GFP and their negative controls, the gel stained with instant blue coomassie stain. Lane 1: PageRuler™ Plus prestained protein ladder (Fisher, UK); Lane 2: Non-induced MLB+GFP fusion (*N*-Term); Lane 3: Induced MLB+GFP fusion (*N*-Term) (58.41kDa); Lane 4: Non-induced MLB+GFP fusion (*C*-Term); Lane 5: Induced MLB+GFP fusion (*C*-Term) (58.14kDa); Lane 6: Non-induced SNA-B+GFP fusion (*N*-Term); Lane 7: Induced SNA-B+GFP fusion (*N*-Term) (58.77kDa); Lane 8: Non-induced SNA-B+GFP fusion (*C*-Term); Lane 7: Induced SNA-B+GFP fusion (*C*-Term) (58.08kDa).

Thus we confirm that both *N*- and *C*- terminus fusion proteins of MLB and SNA-B were successfully expressed in *E. coli* and the activity of GFP was detected which confirms the activity of other candidate proteins in the fusion constructs.

3.4 Discussion

In this study a mutant green florescent gene known as GFPuv was used to produce Lectin-GFP recombinant protein by fusing GFP to the binding chain of mistletoe lectin (MLB), the binding chain of elderberry lectin (SNA-B) or snowdrop lectin (GNA) so as to assess the expression level of these three peptides and to be able to screen their localization in cellular experiments. The coding sequences of the recombinant proteins were cloned into pGFPuv expression vector which has pUC *E. coli* replication origin, in order to express them in *E. coli* bacterial expression system.

Since the main concern in the production of recombinant protein is protein-protein interaction between the fused genes especially between the core amino acids or domains which are at different locations in different genes, it is thus necessary to identify the location of these core portions and fuse the genes in an orientation that causes minimum interaction between the domains responsible for the main activity of the fused proteins such as hydrophobic domains in some membrane proteins important for protein folding and targeting (Zhang and Crandall 2007). Lectins as carbohydrate binding proteins have specific domains responsible for glycan binding activity. Therefore, these domains should be protected from any interaction with other proteins to preserve their binding function.

Despite the fact that the structure of MLB, SNA-B and GNA lectins have been extensively studied and their well-documented 3D structures are available, the impact of fusing these peptides to other peptides on their functionality especially the glycan binding function needs to be assessed experimentally.

In order to identify the optimum orientation of fusing MLB, SNA-B and GNA peptide sequences to GFP, both *N*- and *C*- terminus fusing strategy were applied. In the *N*-terminus constructs, the amino terminus of the lectin sequences was fused to the carboxyl terminus

of GFP sequence. On the other hand, the lectins were *C*-terminally fused to GFP through fusing their carboxyl terminus to the amino terminus of GFP.

The introduction of restriction sites to the start and end of the lectin sequences corresponding the targeted sites on pGFPuv expression vector was successfully performed and few extra amino acids was also placed between lectins and GFP sequences as a linker so as to minimize protein-protein interaction as much as possible. In the *N*-terminus fusing strategy, the amino terminus of lectin sequence was successfully fused in frame with the GFP sequence and the carboxyl terminus was fused in frame with a stop codon as shown in figure 27, 35, 39. In contrast, in the *C*-terminus fusing strategy the lectin amino terminus was successfully fused in frame with a start codon and the carboxyl terminus with the coding sequence of GFP shown in figure 29, 36, 40.

The analytical restriction digest of the recombinant constructs confirmed the correct size of the lectin sequences in the constructs (Figure 30, 41) and the triplicate sanger sequencing reactions also confirmed no mutation or gaps in the lectin coding sequences (Figure 31, 32, 42, 43, 44, 45).

In order to efficiently express the Lectin-GFP fusion constructs and prevent leaky expression, the BL21(DE3) *E. coli* strain was chosen because it has *lacUV5* promoter which is less sensitive than wt *lac* promoter to catabolite repression. Thus, uninduced expression of the target protein can be prevented (Pan and Malcolm 2000).

The expression time, temperature and IPTG concentration of the expression reactions were decided on the base of repeating small scale expressions many times at different conditions. The optimum expression condition was then chosen which was 18 hours, 30C^{o} and 500μM IPTG.

In order to evaluate the expression level of Lectin-GFP fusion proteins, the successfully grown colonies containing recombinant proteins were first screened under standard UV light to excite the expressed GFP in the colonies. Different expression level was detected between different colonies of different constructs and also between different colonies of different orientations of the same construct (Figure 46, 47). The qualitative expression level of *C*-terminally fused MLB and SNA-B to GFP (Plate A&B in figure 46&47) was noticed to be higher than *N*-terminally fused MLB and SNA-B to GFP because a brighter emission of GFP was observed in the former constructs and it is possibly due to less protein-protein interaction in the *C*-terminal constructs and vice versa. However, in the colonies containing GNA-GFP fusion proteins, very low emission of GFP was detected in both *N*- and *C*- terminal cloning strategy. Therefore, the expression level of the fusion constructs was considered too low and the main reason is believed to be the toxicity of GNA lectin which affects *E. coli* cells proliferation and consequently affect the expression of the recombinant constructs and produce inclusion bodies since many very small and ununiformed colonies were observed on the plates containing the GNA-GFP constructs (Plate C in figure 46&47). This result agree with Luo, Zhangsun et al. (2005) who could only produce insoluble inclusion bodies from recombinant GNA expressed in *E.coli*.

Thus, we decided to not proceed the expression of GNA+GFP fusion protein because the activity of GFP in the fusion protein was absent and the protein was considered as misfolded or unfolded. The reason behind a difficulty to express GNA fusion protein is believed to be the toxic effect of GNA which affect the expression process in *E. coli* because previous attempts to produce GNA recombinant fusion proteins for a "systemic" RNAi effect in insects faced the same difficulty and majority of their expressed fusion proteins in *E. coli* was in the insoluble fractions (Min 2016).

However, the expression of MLB+GFP and SNA-B+GFP was successfully carried out and the soluble fractions of MLB and SNA-B recombinant proteins present in both periplasm and cytoplasm of *E. coli* cells were successfully extracted by applying soluble fraction protocol of BugBuster™ protein extraction reagent (EMD Millipore, UK). Then, the fractions of both uninduced and induced Lectin-GFP fusion proteins were analysed on SDS PAGE and the predicted size bands of all the constructs were successfully detected (Figure 49).

Although majority of host proteins-GFP fusions are simple tandem fusions in which the amino terminus (*N*-terminus) of one protein is genetically linked to the carboxyl terminus (*C*-terminus) of the other, the production of successful fusion proteins needs a general rule in order to predict when the host protein function will be preserved and when the fluorescence will be intact. It is common that sometimes one or neither concatenation (*N*- & *C*- terminus) produces functional chimeras. Tsien (1998) believed that the reason could be the splicing of GFP into the middle of the host protein and to prevent this he has suggested to move the *N*- and *C*-terminii of GFP as close to each other as possible through circular permutation or addition of spacers.

In conclusion, recombinant fusion proteins from MLB and SNA-B fused to GFP were successfully designed and expressed in *E. coli.* However, attempts to express GNA-GFP fusion protein failed

Chapter 4. Purification of Expressed Lectin-GFP Fusion Proteins. Confirmation of Sialic Acid Affinity and Demonstration of Biological Activity.

4.1 Introduction

Affinity chromatography is one of the effective systems to purify biological compounds which has been widely used in protein purification because some proteins have the ability to non-covalently attach to a specific molecule in affinity columns known as the ligand. According to the type of the protein to be purified, different type of ligands can be used in affinity chromatography. Lectin as a carbohydrate binding protein needs a ligand that has carbohydrate attachment sites for example a type of ligands known as glycan adsorbent is considered as the most appropriate ligand for lectin purification (Voet and Voet 1995).

The ligands for lectin purification are divided to three main groups as follow: (1) carbohydrates (Monosaccharide, Disaccharide and polysaccharide); (2) glycoproteins and (3) polysaccharide matrices. In the first two groups, the ligand is coupled to a pre-activated matrix but the third group does not need ligands because the matrix itself can be bound by lectins (Pohleven, Štrukelj et al. 2012).

Over the last few decades many types of glycoprotein ligands have been identified and effectively used in protein purification such as glycoprotein ligands attached to cyanogen bromide activated Sepharose. Fetuin is an example of glycoprotein ligands which is a heavily glycosylated protein exists in a variety of glycoforms like bi-, tri-, and tetra-antennary oligosaccharides with variable sialyation. It has been widely used in the purification of glycan binding proteins such as lectins (Nakagawa, Sakamoto et al. 2012).

Protein purification via affinity chromatography can be performed either through gravity flow column or using fast protein liquid chromatography system (FPLC). According to the stability and specificity of both the protein to be purified and the ligand, different eluent and elution techniques can also be used to elute the target protein in affinity purification. For example, elution can be performed through any of the following methods: (a) alternating the pH or ionic strength of the eluent; (b) using urea as eluent; (c) using a specific carbohydrate to perform competitive elution. However, the recommended method to elute lectins is competitive elusion with a carbohydrate which can be effectively bound by the lectin to be purified (Guzmán-Partida, Robles-Burgueno et al. 2004, Pohleven, Štrukelj et al. 2012).

The majority of the previous studies on lectin-carbohydrate binding suggests that lectins primarily bind monosaccharides but lectin binding to complex determinants with higher affinity has also been detected (Cummings 1999). The lectin affinity to monosaccharides is often seen an order of magnitude lower than that for complex carbohydrates and for binding to carbohydrates, lectins require a specific anomeric configuration and certain sugar residues. The carbohydrate-metal binding is another point that needs to be considered in the lectin affinity purification because some lectin-carbohydrate binding often happen in the presence of a specific metal such as the binding of mannose-specific lectins to mannose residues which happens only in the presence of calcium ions (Ca^{+2}). On the other hand, there are other lectins that do not need metal ion to perform binding to their saccharide counterparts such as plant lectins like the mistletoe lectin (Weis, Drickamer et al. 1992, Zheng, Ornstein et al. 1997).

Earlier efforts to study the characteristics of mistletoe lectin Isoform-I (MLI) mainly relied on the extraction of the lectin from the plant (Büssing, Wagner et al. 1999, Hajtò, Krisztina et al. 2007). Also, any attempt to produce recombinant protein from the coding sequence

of mistletoe lectin were to study and confirm the galactose binding ability of MLI and not to study the sialic acid binding ability because MLI is classified as a galactose binding lectin (Wu, Chin et al. 1992, Eck, Langer et al. 1999). However, Müthing, Meisen et al. (2004) carried out comparative solid-phase binding assays along with electrospray ionization tandem mass spectrometry with MLI from the extract of mistletoe and suggested that MLI should be considered as a sialic acid binding lectin instead of galactose binding lectin because the lectin was seen to preferentially bind to terminally α2-6-sialylated neolacto series gangliosides from human granulocytes and only marginal binding of ML-I to terminal galactose residues of neutral glycosphingolipids was detected.

Elderberry (*Sambucus nigra*) lectin known as sambucus nigra agglutinin I (SNA-I) is also a well-known sialic acid binding lectin (Shibuya, Goldstein et al. 1987). The coding sequence of SNA-I has been extensively studied and showed that it is composed of a toxic chain (A chain) and a carbohydrate binding chain (B chain) linked by a disulphide bridge (Tejero, Jiménez et al. 2015). The A chain is a type 2 potent ribosome inactivating protein (RIP2) and the B chain has a specific affinity to terminal α-2,6 sialic acid (α-2,6Neu5Ac) (Damme, Barre et al. 1996).

Altered glycosylation of the majority of the cancer cell surface adhesion proteins has been confirmed to significantly change the biological function of the cells such as cell spreading and signal transduction (Przybylo, Pochec et al. 2008). The most frequently observed altered glycosylation during tumorigenesis is the increased amount of sialylation of the cell surface glycoproteins results from the terminal capping of sialic acid on the additional antennas provided by acetyllactosamine residues (Ranjan and Kalraiya 2013). Human melanoma cancer has been reported as one of the cancers that have aberrant amount of α-2,3 and α-2,6 sialic acid linkages on their cell surface glycoproteins which can be targeted

as a biomarker for the cancer prognosis and treatment (Hoja-Łukowicz, Link-Lenczowski et al. 2013, Kolasińska, Przybyło et al. 2016).

Thus in this study we aim to perform affinity purification using activated-Sepharose bead coupled with fetuin to purify the produced recombinant fusion proteins from the binding chain of both mistletoe lectin (MLB) and positive control elderberry lectin (SNA-B) fused to GFPuv and thus confirm the sialic acid binding ability of the recombinant proteins through assessing their binding to the sialic acid residues of the fetuin based purification system. Finally, we aim to screen the ability of the purified recombinant proteins to specifically bind to melanoma cells and cross their membranes through binding to their cell surface sialic acid residues.

4.2 Material and Methods

4.2.1 Recombinant Protein Purification

Affinity columns were prepared in order to purify sialic acid binding MLB and SNA-B lectin via their binding to fetuin ligand.

4.2.1.1 Column Preparation

The affinity column bead was prepared from the pre-activated CNBr Sepharose 4B which is a good medium for immobilization of ligands containing primary amines (GE Healthcare, UK). The medium was also coupled with degalactosilated fetuin which is a heavily glycosylated protein existing in a variety of glycoforms containing bi-, tri-, and tetra-antennary oligosaccharides with variable sialyation (Sigma-Aldrich, UK). Briefly, the required amount of lyophilized CNBr powder was weighed and suspended in 1 mM HCL. The medium swelled immediately and was washed on a sintered glass filter (porosity G3) with 1mM HCl for 15 minutes. The required amount of fetuin (5–10 mg protein per ml medium) was also dissolved in dissolving buffer (0.1 M CH_3COONa pH 8.3 containing 0.5 M NaCl) and incubated at 37 °C overnight with the required amount of β-Galactosidase to remove galactose molecules from fetuin (Sigma-Aldrich, UK). The β-Galactosidase treated fetuin was then extensively dialyzed against coupling buffer (0.1 M $NaHCO_3$ pH 8.3 containing 0.5 M NaCl) and added to the suspended CNBr medium (5 ml coupling solution/g lyophilized powder) in a stoppered vessel. The mixture was rotated end over end overnight at 4 °C and washed with 5 medium (gel) volumes of coupling buffer to wash away the excess ligands. The mixture was incubated in blocking buffer (0.1 M Tris-HCl,) pH 8.0 for 2 hours to block remaining active groups on Sepharose 4B and washed in three cycles of alternating pH with 5 medium volumes of 0.1 M acetic acid/sodium acetate, pH 4.0 containing 0.5 M NaCl and 0.1 M Tris-HCl, pH 8 containing

0.5 M NaCl. Finally, the coupled CNBr-fetuin was packed in HR 16/10 column (GE Healthcare, UK).

4.2.1.2 Affinity Purification

Following the expression of MLB+GFP and SNA-B+GFP (*N*- and *C*- terminus), the culture supernatant was clarified by centrifugation and the *E.coli* cell walls were gently disrupted by BugBuster reagent, resulting in the liberation of soluble protein. Affinity column chromatography was then conducted to purify targeted fusion proteins from the total extracted proteins from the *E. coli* cells using prepared affinity column described in section 4.2.1.1 and AKTA prime FPLC platform (GE Healthcare, UK). After purification the relevant fractions were analysed by SDS-PAGE and western blotting.

4.2.2 Protein Biochemistry

4.2.2.1 Protein Concentration Assay

The BCA protein quantification assay kit (Merck Millipore, UK) was used to determine the concentration of the purified recombinant proteins in microassay size. The BCA working reagent was prepared according to the manufacturer's instructions and a 7-point standard curve was prepared in triplicate from a serial dilution of stock BSA (2000 µg ml-1). Samples were assayed using 96-well plates, 25µl from each standard and samples was added in replicated into individual wells of a 96-well plate followed by 200 µl BCA working reagent. The plate was covered and incubated at 37^{o}C for 30 minutes. After the incubation time, the plate was cooled to room temperature and the absorbance was measured at 562 nm on Multiskan™ GO Microplate Spectrophotometer (Fisher, UK). The absorbance and known mass of the BSA were then plotted to generate the standard curve which was used to interpolate the measured absorbance of the samples by liner regression (Figure 50).

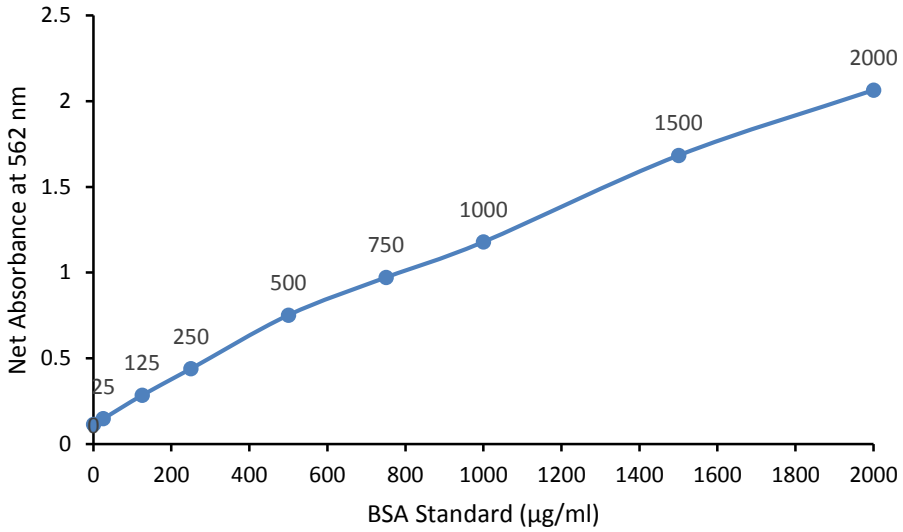


Figure 50. The standard curve of net absorbance at 562nm versus the concentration of a serial dilution of bovine serum albumin (BSA) ranging from 0-2000µm) after incubation with BCA reagent for 30minutes.

The fusion protein sequences were also computationally analysed using ExPASy server (Gasteiger, Gattiker et al. 2003). The computed molecular weight and extinction coefficient of the fusion proteins are shown in table 16.

Table 16. Computed parameters of MLB+GFP, SNA+GFP and GNA+GFP.

Fusion Proteins	Molecular Weight (g/mol)	Extinction Coefficient M-1 cm-1
MLB+GFP (*N*-Terminal)	58023.4	82570
MLB+GFP (*C*-Terminal)	58398.7	81080
SNA+GFP (*N*-Terminal)	58762.3	81080
SNA+GFP (*C*-Terminal)	58072.5	82570
GNA+GFP (*N*-Terminal)	46251.4	47705
GNA+GFP (*c*-Terminal)	46018.2	49195

4.2.2.2 Sodium Dodecyl Sulphate Polyacrylamide Gel Electrophoresis

SDS-PAGE of protein samples was carried out using tris-glycine buffer system (25 mM Tris-HCl PH 6.8, 250 mM glycine, 1% w/v SDS) under denaturing conditions. Compositions of the polyacrylamide gels are shown in table 17. Samples were mixed with 1X reducing sample buffer (50 mM Tris-cl PH 6.8, 100mM 2-mercaptoethanol, 2% w/v

SDS, 0.1% bromophenol blue, 10% v/v glycerol) and boiled at 100°C in a water bath for 3 minutes. Samples were loaded onto the gel along with molecular weight marker PageRuler™ Prestained Protein Ladder (Fisher, UK). Gels were run with the appropriate running buffer with a voltage of 8 V/cm until the dye front moved into the resolving gel then the voltage was increased to 15 V/cm and run until the bromophenol blue reached the bottom of the resolving gel. The gels were stained with InstantBlue (Expedeon-Uk) for 15 minutes and rinsed with water. Finally, the gel was photographed by G-box transilluminator (Syngene, UK).

Table 17. Solutions for preparing 12% resolving gels and 5% stacking gels for Tris-glycin SDS-polyacrylamide gel electrophoresis.

Components	12% Resolving Gel (ml)	5% Stacking Gel (ml)
1.5M Tris-HCL (pH8.8)	2.5	-
1.0M Tris-HCL (pH6.8)	-	0.38
30% Acrylamide mix	4.0	0.5
10% SDS	0.1	0.03
10% Ammonium Persulafate	0.1	0.03
TEMED	0.004	0.003
H_2O	3.3	2.1
Total	10 ml	3 ml

4.2.2.3 Immuno Assay by Western Blotting

Following running the denatured protein samples on SDS PAGE as described in section 4.2.2.2, the proteins were transferred to nitrocellulose membrane (Merck Millipore, UK) using the semi dry blotting technique (Kyhse-Andersen 1984). Briefly, a piece of nitrocellulose membrane the same size as the gel was activated on ice-cold methanol for 10 minutes and the gel was washed in water to remove the traces of SDS. They were then placed on BioRad Mini Transblot blotting apparatus in the following order: ANODE; one sheet of sponge; two sheets of 3MM filter paper soaked in transfer buffer (192 mM Glycine, 25 mM Tris-HCl and 20 % (v/v) Methanol at pH 8.3); 1 sheet of nitrocellulose membrane soaked in PBS; the gel; two sheets of 3MM filter paper; one sheet of sponge

soaked in transfer buffer (192 mM Glycine, 25 mM Tris-HCl and 20 % (v/v) Methanol at pH 8.3); CATHODE.

The transfer was carried out for 120 minutes at a constant current of 400 mA in transfer buffer (192 mM Glycine, 25 mM Tris-HCl and 20 % (v/v) Methanol at pH 8.3). The middle reservoir of the tank was filled with ice to prevent overheating during running. Following the transfer, the membrane was washed in PBS and incubated on a rocker overnight at 4ºC in blocking buffer (5% nonfat dry milk, 0.2% Tween-20 and PBS) to block the free sites on the membranes. The membrane was treated with primary antibody in 10 ml of blocking buffer containing Living Colors® A.v. Monoclonal Ani-GFPuv Antibody 1:5,000 (Clontech, USA) and incubated for 2 hours at room temperature with shaking followed by washing in washing buffer (0.2% Tween-20 and PBS) two times for 5 minutes each wash. The membrane was then incubated for 1 hour in blocking buffer containing Goat IgG Anti-Mouse IgG-HRP conjugated secondary antibody 1:4000, (Sigma-Aldrich-UK) followed by washing in washing buffer (0.2% Tween-20 and PBS) 4 times for 10 minutes each wash. Antibody binding was visualised using Pierce Enhanced chemiluminescence (ECL) substrate solution following the manufacturers instruction (Thermofisher Scientific). The membrane was visualised and photographed by G-box transilluminator (Syngene, UK).

4.2.3 Cell Culture

Human metastatic melanoma cells (WM 266-4) (ECACC, UK) (Table 20) were cultured in RPMI-1640 culture media (Sigma-Aldrich, UK) supplemented with 10% (v/v) fetal bovine serum (FBS) and 1% (v/v) glutamine at 37 ºC in a humidified atmosphere containing 5% carbon dioxide.

The cells were grown until 80% confluency achieved in about 72 to 96 hours followed by subculturing. The culture flask was decanted from the culture media and 0.5 ml trypsin/EDTA (Sigma-Aldrich, UK) was added and incubated for about 5 minutes at 37 ºC to detach the cell monolayer from the flask. The flask content was transferred into a universal tube and centrifuged at 500 x g for 5 minutes. The supernatant was discarded and the pellet was re-suspended in 5 ml of culture media. The split was carried out at recommended 1:3 ratio and transferred into 25 or 75 cm^2 culture flasks according to the amount of cells needed for the further tests and the total volume of the media made up to 5 or 15 ml respectively.

4.2.4 Cell Counting

The harvested cell pellet was suspended in 5-10ml of culture media and 40µl of the culture was transferred into a 1.5ml eppendorf tube to perform 1:2 dilutions with trypan blue dye (Sigma-Aldrich, UK). The diluted culture was then used in the cell counting using a haemocytometer (Labtech, UK) and both viable and dead cells were counted as the dye only stains dead cells blue. The cell number and cell viability were calculated according to the formula A and B shown below.

A-Cells counting equation

$$Totall\ cells/ml = Total\ cells\ counted * \frac{dilution\ factor}{\#\ of\ squares} * 10{,}000\ cells/ml$$

B- Cell viability equation

$$Viability\ (\%) = \frac{live\ cell\ count}{live + dead\ cell\ count} * 100$$

4.2.5 MTS Cell Proliferation Assay

A colorimetric method CellTiter 96®Aqueous One Solution Cell Proliferation Assay (Promega, UK) was used to determine the number of viable cells in the proliferation and cytotoxicity assays. The MTS solution was prepared according to the manufacturers instruction.

A range of cell density (0.5-2 x 10^4 cells/cm^2) was suggested by the cell supplier, therefore the cells were seeded at incremental levels of 1 x 10^3 cells in a range between 2-5 x 10^3 cells/well (surface area, 0.35 cm^2) in 100µl of appropriate culture media on 96-well plate and incubated at 37 ºC in humidified atmosphere containing 5% $C0_2$ for 24, 48 and 72 hours. At the end of each period of time cells were treated with 20µl of the MTS solution and incubated at 37ºC for 3 hours. The colour of the mixture was changed during the incubation time because the MTS was reduced to formazan crystals by the living cells in which dissolved in the culture media and gave a purple colour. The amount of produced formazan crystals was thus directly proportional to cell viability as MTS could be reduced only by the respiring cells. The colour intensity was determined through measuring the absorbance at 490 nm and the blank absorbance at 650 nm using micro plate reader (Dynatech MR5000, USA).

4.2.6 Assessing the Binding of Lectin-GFP Fusion Proteins to Human Melanoma Cells

Human metastatic melanoma cancer cells (WM 266-4) were cultured on 96-well plate with the optimum seeding density 5000 cells/well (~1.4 x 10^4 cells/cm^2) as determined in section 4.2.5 After 24hours incubation at 37 ºC in humidified atmosphere containing 5% $C0_2$, the media was changed to the same RPMI-1640 but with lower amount of FBS (2%). The cells were then treated with 0.3µM of MLB-GFP and SNA-B-GFP fusion proteins and incubated another 24 hours. The binding of the fusion proteins to the cell membrane was

visualised by luminescence imaging microscopy (Nikon, UK) using blue light (488nm) to excite the GFP tagged to the lectins.

4.3 Results

To assess the binding of the lectin-GFP fusion proteins to sialic acid, affinity purification using degalactosilated fetuin was chosen to affinity purify the expressed proteins. Human melanoma cancer cell lines known to overexpress sialic acid on their cell membrane were also used to assess the activity and binding of the designed fusion proteins. Epifluorescence microscopy was employed to visualize binding and demonstrate cellular internalization of the fusion proteins.

4.3.1 Protein Purification

Following the expression of the fusions MLB+GFP and SNA+GFP in both *N*- and *C*-terminus direction and extraction of the soluble protein fractions from *E. coli* section 3.2.10, recombinant proteins were partially purified from culture supernatant by affinity chromatography. A CNBr-activated Sepharose 4B medium was employed for immobilization of ligands containing primary amines (GE Healthcare, UK) and was coupled with fetuin which is a heavily glycosylated protein (Sigma-Aldrich, UK) following manufacturers instruction. Both *N*- and *C*- terminus fusion proteins showed binding activity and successfully bound to the sialic acid of the fetuin ligand on the column bead. However, as higher activity of GFP in the *C*-terminus fusion proteins was shown in the previous study, the similar result was also observed during the purification of the fusion proteins since the binding activity was seen to be higher with *C*-terminus fusion proteins in comparison to the *N*-terminus fusion proteins and in turn resulted in a higher amount of purified products.

Similar purification profiles were shown by all the constructs during their purification of on the sepharose column. Following the addition of the eluent which was PBS containing 0.5M sialic acid, a broad peak of eluted proteins was observed and fractioned into small

fractions so as to be assessed later by SDS page. A typical elution profiles from the sepharose column chromatography purification is shown in figure 51, 52, 53, and 54.

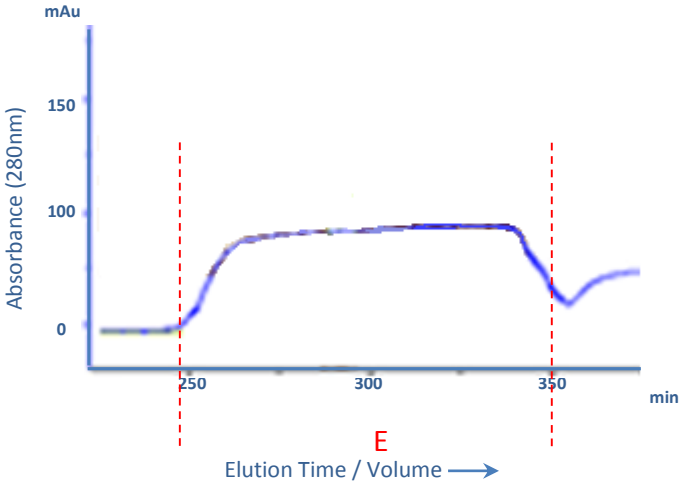


Figure 51. Typical purification profile of CNBr-Activated Sepharose affinity column purification. One elution peak (E) was collected from purified MLB+GFP (*N*-terminus) (~100ml).

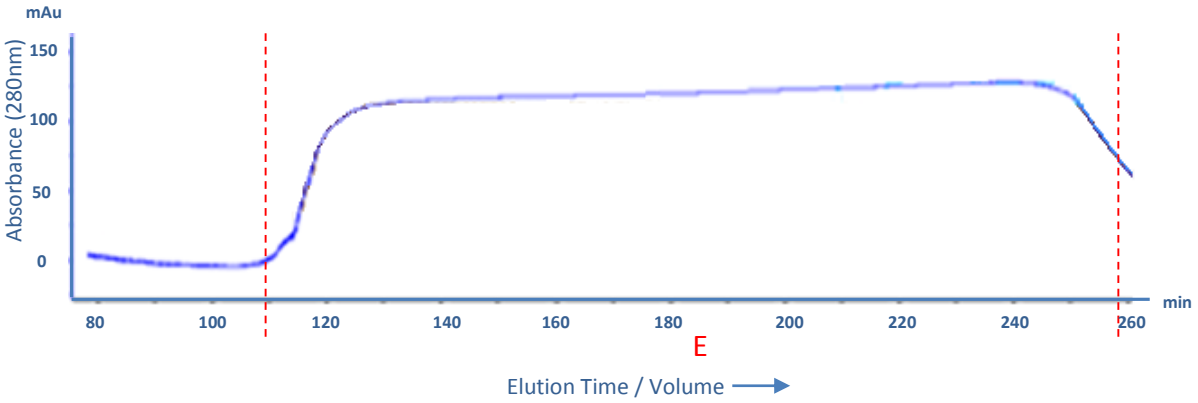


Figure 52. Typical purification profile of CNBr-Activated Sepharose affinity column purification. One elution peak (E) was collected from purified MLB+GFP (*C*-terminus) (~140ml).

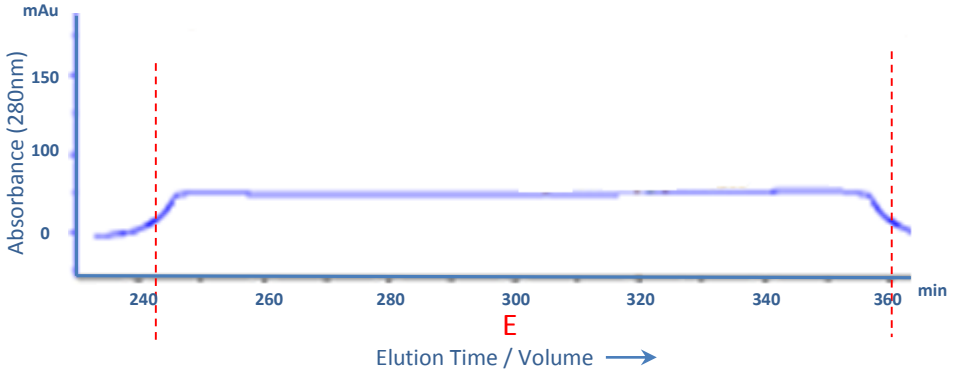


Figure 53. Typical purification profile of CNBr-Activated Sepharose affinity column purification. One elution peak (E) was collected from purified SNA-B+GFP (*N*-terminus) (~110ml).

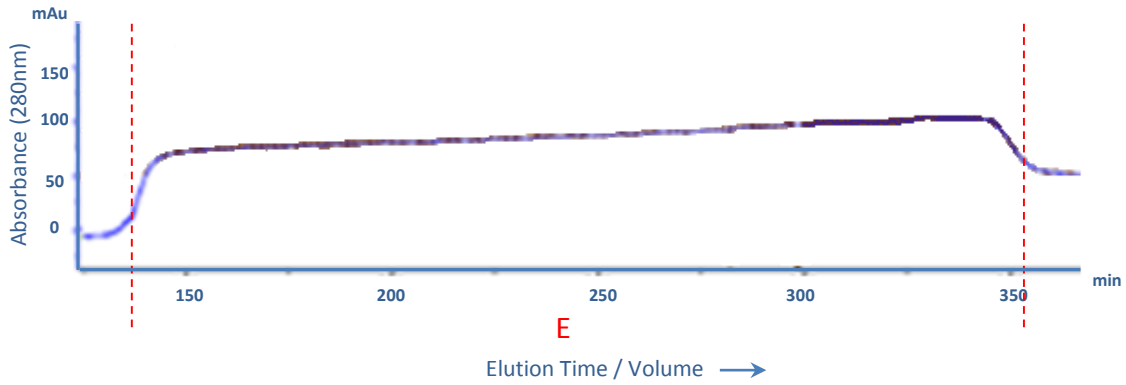


Figure 54. Typical purification profile of CNBr-Activated Sepharose affinity column purification. One elution peak (E) was collected from purified SNA-B+GFP (*C*-terminus) (~225ml).

4.3.2 Sodium Dodecyl Sulphate Polyacrylamide Gel Electrophoresis and Western Blotting

The predicted size of *N*-terminus MLB+GFP and SNA-B+GFP deduced from the amino acid sequences are 58.41kDa and 58.77kDa respectively and of *C*-terminus MLB+GFP and SNA-B+GFP are 58.14kDa and 58.08kDa respectively. The fractions of column eluates were analysed by SDS-PAGE and the predicted sizes of all constructs were observed (Figure 55a) The presence and activity of both *N*- and *C*- terminus MLB+GFP and SNA+GFP were also confirmed by western blotting of culture supernatant using mouse monoclonal anti-GFP antibodies and goat anti-mouse IgG secondary antibody (Figure 55b)

The western blotting experiment also confirmed that all the fusion proteins were expressed and folded properly in *E. coli* since the anti-GFP primary antibody successfully bound to *N*- and *C*-terminus fusion proteins (Figure 55b).

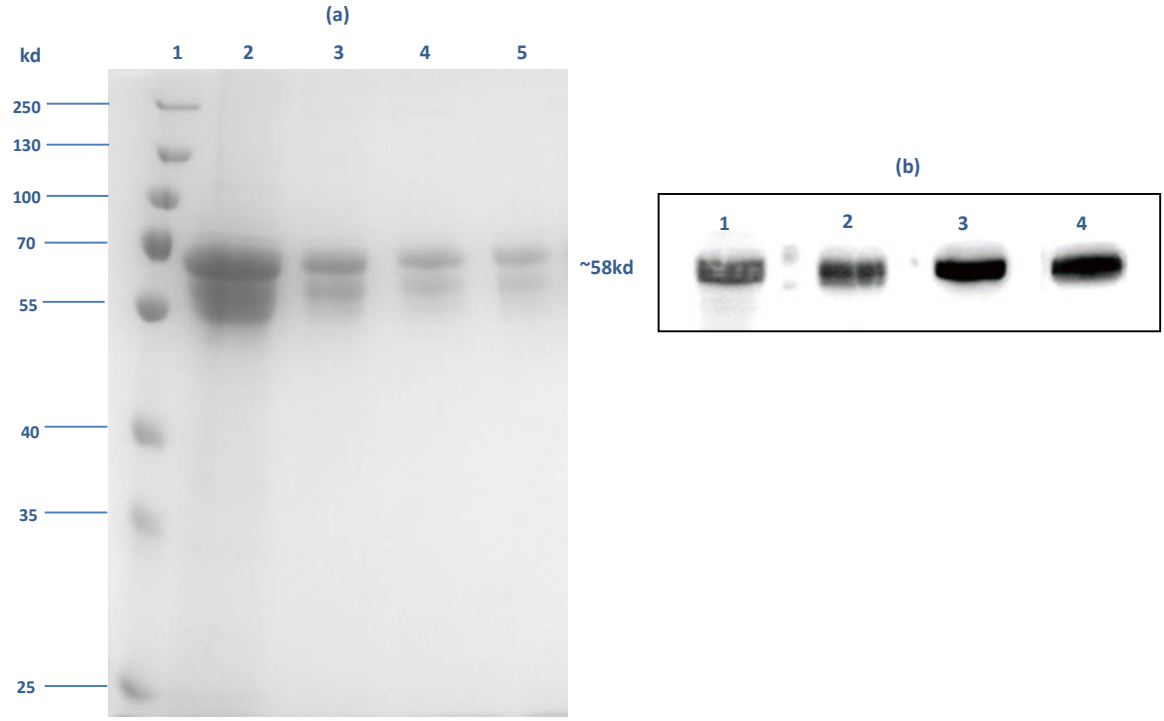


Figure 55. (a) SDS PAGE of purifird MLB+GFP and SNA-B+GFP (N- and C- terminus) fusions stained with InstantBlue (Expedeon,UK) and (b) western blot of purified MLB+GFP and SNA-B+GFP (*N*- and *C*- terminus) fusions using Living Colors® GFPuv anti-GFP primary antibody (Clontech, USA) and goat anti-mouse IgG secondary antibody (Sigma, UK). Lane a1: PageRuler™ Plus prestained protein ladder (Fisher, UK); Lane a2&b1: purified MLB+GFP fusion (*C*-Term) (58.41kDa); Lane a3&b2: purified MLB+GFP fusion (*N*-Term) (58.14kDa); Lane a4&b3: purified SNA-B+MLB fusion (*C*-Term) (58.77kDa); Lane a5&b4: purified SNA-B+GFP fusion (*N*-Term) (58.08kDa).

4.3.3 Assessing the Binding of Lectin-GFP Fusion Proteins to Human Melanoma Cells

In order to assess the binding activity of the lectin-GFP fusion proteins to the sialic acid residues present on the surface of cancer cells, a melanoma cell line (WM 266-4) was used. Previous studies on this cell line have shown overexpression of sialic acid on their cell membranes and it can be used to test the activity of sialic acid binding fusion proteins. The WM 266-4 cells are metastatic melanoma cells with overexpressed α2-6 sialic acid which can be preferentially bound by mistletoe and elderberry lectins (Hoja-Łukowicz and Link-Lenczowsk, 2012).

The cells were prepared on 96-well plates as described in section 4.2.5 and treated with (0.3µM) of each MLB-GFP and SNA-B-GFP (*N*-terminus) fusion proteins. Following the incubation time and to visualise the binding and localization of the fusion proteins, the treated cells were screened and imaged under blue light (488nm) to excite the GFP fused to the lectins using epifluorescence microscope.

Both fusion proteins were successfully attached to the surface of the cancer cells and also crossed their membranes since the emission of GFP was detected not only on the cell surfaces but also inside the cells.

Figure 56 and 57 show the WM 266-4 metastatic melanoma cells treated with MLB+GFP and SNA-B+GFP fusion proteins respectively. It can clearly be seen that the fusion proteins successfully attached to the surface of the cell membranes and caused the agglutination of the cells which is an indication of the activity of the lectins in the fusion proteins. Also, the fusion proteins successfully crossed the cell membranes and reached the cell interiors as the emission of GFP can be seen inside the cells.

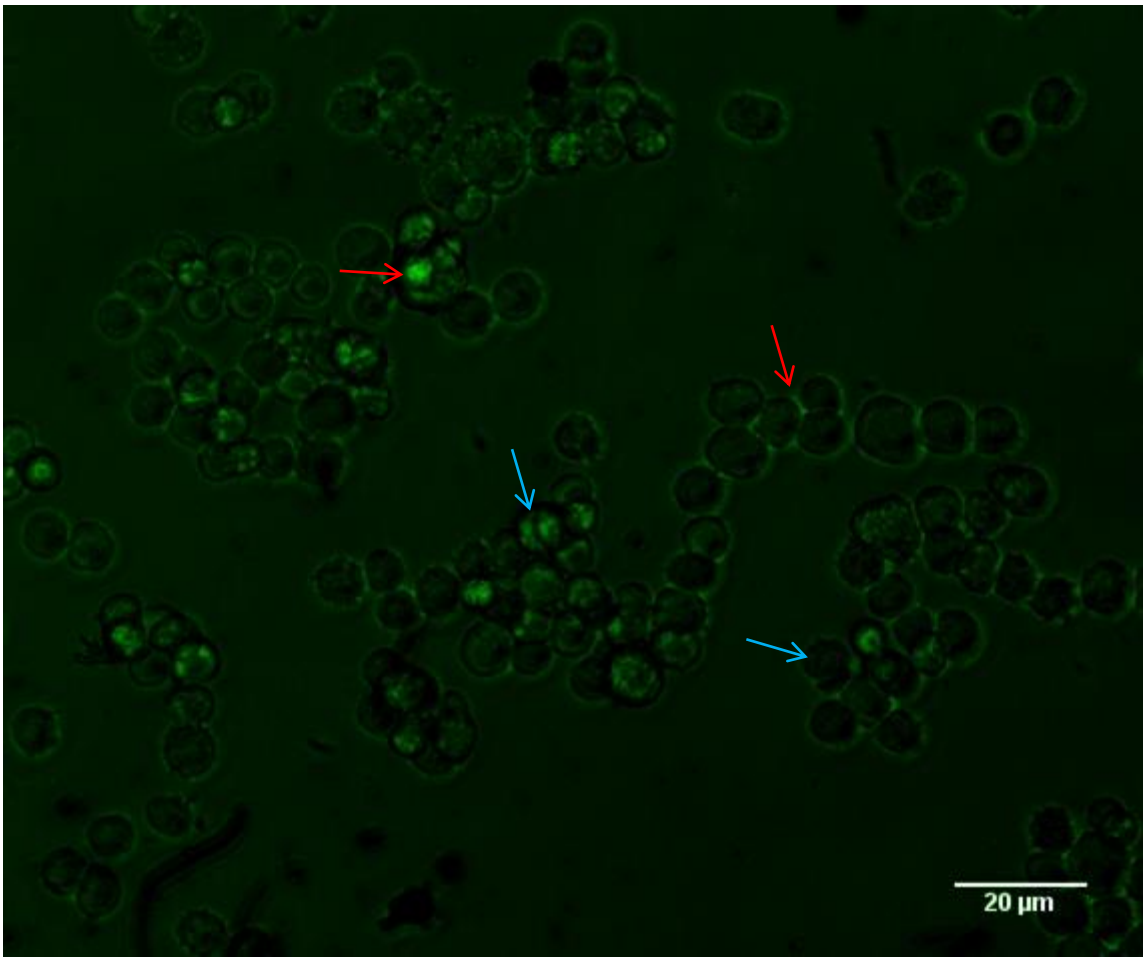


Figure 56. Metastatic melanoma cells (WM 266-4) incubated with 0.3µM of MLB+GFP fusion protein overnight and screened under luminescence imaging microscopy. The RED arrows show the attachment of the fusion protein to the cell membranes and cell agglutination. The BLUE arrows show the cell internalization of the fusion protein.

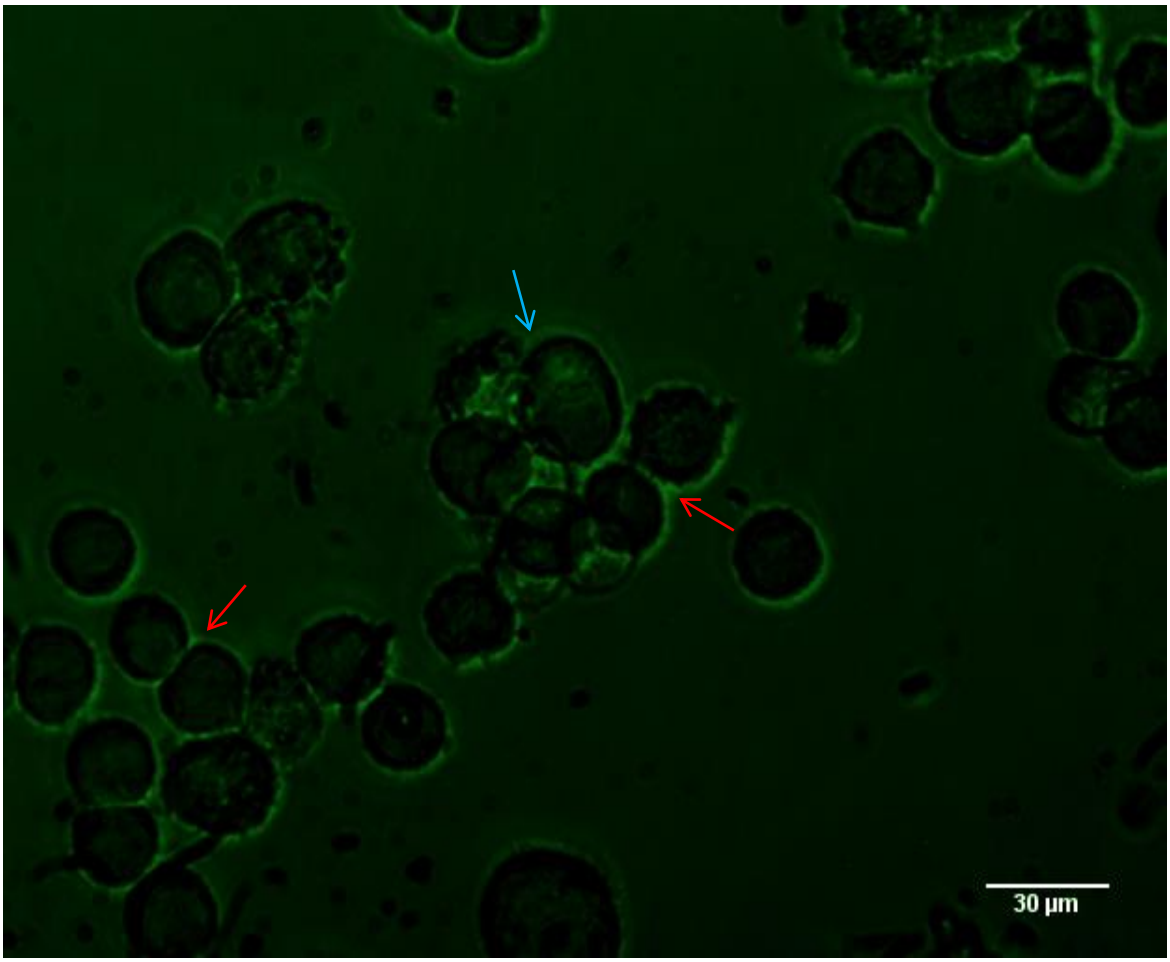


Figure 57. Metastatic melanoma cells (WM 266-4) incubated with 0.3µM of SNA-B+GFP fusion protein overnight and screened under luminescence imaging microscopy. The RED arrows show the attachment of the fusion protein to the cell membranes and cell agglutination. The BLUE arrow show the cell internalization of the fusion protein.

Thus biological activity of MLB-GFP fusion was demonstrated in terms of (a) sialic acid binding (b) activity against cancer cells.

4.4 Discussion

In this study affinity column chromatography was applied to purify Lectin-GFP recombinant fusion proteins produced in the previous study from fusing the binding chain of mistletoe lectin (MLB) and elderberry lectin (SNA-B) to GFP and expressed in *E. coli*.

Protein purification as a process, is a separation of a particular protein from a bulk of proteins or other molecules. There is a diversity of protein purification systems which are different in terms of functionality and application. The selection of an appropriate protein purification system mainly depends on the type of protein that needs to be purified. Proteins can be purified through targeting a specific feature which differentiate the protein of interest from surrounding molecules. This differentiation feature can be naturally possessed by the protein such as glycan binding property or can be added to the target protein such as the addition of six histidine tag (6-His). Therefore, to specifically target a particular property of a specific protein, a purification system should be selected which has ability to recognize this particular property.

Affinity chromatography is one of the potent protein purification systems that relies on specific ligands to which protein of interest can bind. Lectins are well known for their glycan binding affinity and majority of lectins have affinity to one or more specific monosaccharide which can be used as a ligand to specifically isolate lectins from other molecules. Mistletoe and elderberry lectins are both known as sialic acid binding lectins (Müthing, Meisen et al. 2004, Shang and Van Damme 2014). Therefore, ligands containing sialic acid was used to purify both lectins.

In order to minimize the interaction between the column matrix and the target Lectin-GFP fusion proteins during the purification process, pre-activated sepharose matrix was used which is proven to have less interaction with proteins during purification (You, Cao et al.

2015). However, sepharose alone cannot be bound by MLB and SNA-B because it does not have any attachment sites for the lectins. Thus, fetuin as a heavily glycosylated glycoprotein was coupled to the sepharose matrix so as to be bound by the lectins through binding to the sialic acid branches possessed on its glycosylation profile (Wu, Huang et al. 2015).

All the expressed recombinant fusion constructs were observed to successfully bind to the column matrix through binding to available sialic acid sites and they were later eluted through competitive elusion using PBS containing a high concentration of sialic acid as the elution peaks are shown in figure 51, 52, 53, 54. The purification chromatogram was found similar for all the recombinant constructs which was a broad peak. However, a broader peak was observed in the chromatogram of *C*-terminally fused MLB and SNA-B to GFP (Figure 52, 54) in comparison to their *N*-terminally fused counterparts (Figure 51, 53).

The purified recombinant constructs were later analysed by SDS PAGE to assess the size of the purified proteins and to detect any possible degradation happened during the purification process. A correct size band was successfully detected for each construct as shown in figure 55a. Although, the expression of the recombinant constructs was previously confirmed via assessing the emission of GFP under UV light, western blotting was also performed and the anti-GFP primary antibody was observed to successfully bind to all the constructs (Figure 55b).

While the main purpose of conducting purification of the expressed recombinant Lectin-GFP fusions was to achieve pure fusion protein products for further experiments, the affinity purification also confirmed that the fused binding chains of both lectins MLB and SNA-B preserved their glycan binding affinity since they successfully bound to the sialic acid residues of fetuin. This result can thus be counted as a first attempt to experimentally

confirm the sialic acid binding affinity of MLB chain alone without MLA chain since the previous studies hypothetically suggested this affinity or essentially tested the sialic acid binding affinity of the whole mistletoe lectin (MLA+MLB) such as the study conducted Müthing, Meisen et al. (2004).

Aberrant glycosylation during oncogenic transformation is observed in almost all the type of cancers and becomes a universal feature of cancer cells by which different cancer cells achieve different glycosylation profiles. For example, melanoma cancer cells have altered glycosylation profile compared to their healthy counterparts and a series of changes happens in their glycosylation profile during different progression stages. However, this alteration in glycosylation especially in cell surface glycosylation becomes stable at some progression stages and can be used as a biomarker to specifically recognize and target particular cancer cells (Hoja-Łukowicz, Link-Lenczowski et al. 2013). For example, the overexpression of a certain glycan on the surface of a type of cancer cells can targeted by a glycan binding agents such as lectin with affinity to this particular glycan (Rek, Krenn et al. 2009).

In order to exploit the sialic acid binding property of mistletoe lectin in targeting cancer cells, a type of cancer can be targeted which has abundant sialic acid on their cell surface carbohydrates. According to a study conducted by Hoja-Łukowicz, Link-Lenczowski et al. (2013) on *N*-linked glycans of L1 Cell Adhesion Molecule (L1CAM) of melanoma cancer cells, L1CAM of melanoma cells contains abundant amount of sialic acid branches especially on the cell surfaces. Therefore, a metastatic melanoma cell line (WM 266-4) was selected to be targeted by both MLB+GFP and SNA-B+GFP fusion proteins. Following the treatment of melanoma cells with an 0.3 μM of the fusion proteins, the binding of both fusion proteins to the cell surface glycans and their cell internalization was

successfully observed through performing florescence microscopy imaging using blue light (488nm) to excite the GFP fused to MLB and SNA-B (Figure 56, 57).

The toxic side effects of conventional chemotherapeutic drugs on healthy tissues and organs has urged the need to discover alternative therapies for cancer treatment. To date, the immunococonjugates like monoclonal antibodies (mAbs) coupled with toxic agents like toxic drugs or radioisotopes have led to the development of approved therapies in many therapeutic areas especially anti-cancer treatment (Alley, Okeley et al. 2010). Many efforts are underway to increase the therapeutic benefit of cytotoxic agents and reduce their systematic toxicity which is the main limitation of their applications (Chari 2007). However, improvement is still needed to increase the curative potential of these conjugates because the potential of antibody–drug conjugates (ADCs) is still limited and it is not cost effective (Wu and Senter 2005).

To increase the specificity of cytotoxic agents and also reduce their systematic side effects, recombinant fusion proteins can be produced by fusing cytotoxic peptides with known potential anti-cancer property to other peptides with the ability to be specifically directed to a specific site because of their affinity or recognition potential such as lectin affinity to glycans. For that, a less advanced technology and financial support needed because the production of fusion proteins is simple to engineer. Therefore, effective anti-cancer therapy with lower cost can be produced from fusion proteins since they can be produced via genetic engineering and using prokaryotic or eukaryotic expression systems such as the anti-ovarian cancer fusion protein produced by Qin, Zhou et al. (2016) from milk proteins and anti-solid tumours fusion protein produced by Niesen, Hehmann-Titt et al. (2016) from human serine protease granzyme B.

The high level of selectivity of our fusion proteins makes the fusions to be more effective than the previous attempts in targeting cancer cells. Also, because the targeted moiety by

our fusion proteins is sialic acid and it is highly expressed on the surface of different type of cancer cells, the fusion proteins are not limited to target just one type of cancer cells but they can be used to target a range of cancer cells (Büll, Boltje et al. 2013, Büll, Stoel et al. 2014).

In conclusion, affinity purification was successfully carried out to purify MLB+GFP and SNA-B+GFP and sufficient amount of purified fusion proteins was achieved to conduct cellular experiments. The fusion proteins were confirmed to have ability to successfully bind to the surface of human melanoma cells and cross their membranes.

Chapter 5. Construction of a delivery system from MLB and its two putative sugar binding sites to specifically deliver cytotoxic peptides to human melanoma cells

5.1 Introduction

Lectin is a glycan binding moiety which has the ability to specifically attach to a certain glycan and can be used as a carrier vehicle to deliver other peptides to cancer cells through targeting cell surface glycans. Thus the binding chain of mistletoe lectin (MLB) was used to deliver a potent cytotoxic peptide to the surface of human melanoma cells.

The therapeutic potential of two toxic peptides known as melittin and mastoparan from honey bee (*Apis mellifera*) and wasp venom (*Vespula lewisii*) respectively have been characterized for a long time but their application in biomedicine as a drug yet to be explored (Moreno and Giralt 2015). Many studies about whole bee and wasp venom have confirmed their anti-tumour activity and the same activity is expected from melittin and mastoparan as main components of these venoms (Gajski and Garaj-Vrhovac 2013, De Azevedo, Figueiredo et al. 2015).

To date, the studies conducted on melittin as an anti-cancer agent have revealed that melittin has potential to inhibit tumour growth and induce apoptosis. The peptide was seen to mainly work on the disruption of phospholipid bilayers and disorder their function through forming tetramer aggregates. It also aggregates membrane proteins and induces hormone secretion (Ladokhin and White 1999). The disruption and disfunctioning of phospholipid bilayer by melittin consequently alters the cell enzymatic systems either directly or indirectly such as the alteration of protein kinase C and G-proteins (Fukushima, Kohno et al. 1998, Katoh 2002). Moreover, during *in vivo* tests and the treatments of

variety of cancer cells *in vitro*, antiproliferative and anti-angiogeneses activities of melittin were also demonstrated (Song, Lu et al. 2007, Yang, Zhu et al. 2014).

Mastoparan also performs similar functions to melittin such as destabilization of membrane lipid bilayer which consequently causes membrane lysis and perturbing transmembrane signalling through interaction with G-proteins. Also, mastoparan is well-known to interact with the mitochondria membrane and induce its permeability by which mastoparan mediates its cytotoxicity in tumour cells (Grześk, Malinowski et al. 2014, Yamamoto, Ito et al. 2014).

However, the side effects of intravenous administration of both melittin and mastoparan such as hemolysis and liver injury have limited their direct application as a drug. In order to prevent or reduce these side effects, it is suggested to either shield the toxic peptides or attach them to a carrier molecule such as antibodies, targeting peptides, and synthesized carriers so as to keep the peptides inactive until they reach the target site and mediate their cytotoxicity (Dunn, Weston et al. 1996, Jin, Li et al. 2013). For example, Yamada, Shinohara et al. (2005) showed an effective method to encapsulate mastoparan in transferrin (Tf)-modified liposomes decorated with endosomolytic GALA peptides so as to reduce the side effect of mastoparan during targeting Tf receptor of tumour cells.

Regarding attaching the toxic peptides to carrier molecules, many studies have been conducted to produce fusion proteins from fusing melittin and mastoparan sequences to the sequence of other peptides. For example, a recombinant protein was produced from fusing the melittin oligonucleotide with a fragment of murine monoclonal K121 antibody (Dunn, Weston et al. 1996). As well as, the significant results have been achieved from pro-cytotoxic systems produced from fusing mastoparan to a targeting peptide through biodegradable linkers such as matrix metallopeptidase 2 (MMP-2) and cathepsin B (CatB)

which are sensitive to the overexpressed tumour enzymes and allow spatiotemporal control to release the toxic peptides from the polymer on a target site (Moreno, Zurita et al. 2014).

Thus we aim to produce a range of recombinant fusion proteins from fusing the coding sequence of melittin (MLT) and mastoparan (MP) to the coding sequence of the binding chain of mistletoe lectin (MLB) and its two putative sugar binding sites so as to be able to specifically deliver the cytotoxic peptides to the human melanoma cells through the sialic acid binding ability of MLB and exploit their anti-tumour property.

5.2 Material and Methods

The Lectin-GFP fusion constructs showed binding and internalization activity. Therefore, two cytotoxic peptides were selected from literature to generate Lectin-Toxin fusion proteins (Huang, Jin et al. 2013, De Azevedo, Figueiredo et al. 2015).

5.2.1 Conserved Domain Analysis

The binding chain of mistletoe lectin (MLB) was analysed on NCBI conserved domain database and two separate glycan binding domains (1st domain and 2nd domain) were found as shown in figure 68. Therefore, fusion constructs were designed not only for the whole MLB chain fused to the toxin peptides but also each domain separately fused to the toxin peptides because one of these domains might be specifically responsible for the sialic acid binding property of MLB.

5.2.2 Codon Optimization

In order to change the sequence codons of the targeted genes to optimal codons for bacterial expression and efficiently express them in *E. coli*, codon optimization was performed to mistletoe lectin B-chain (MLB), melittin (MLT) and mastoparan (MAS) sequences using GeneOptimizer® facility (Invitrogen, UK) (Figure 72, 73, 74).

5.2.3 Designing and Synthesis of Lectin-Toxin Fusions

Different constructs were designed from mistletoe lectin B-chain (MLB) and its two separate glycan binding domains (1st and 2nd domains) using Snapgene software (Snapgene, USA). They were all *C*-terminally fused through a cathepsin-B (CatB) biodegradable linker peptide to two different toxin peptides namely melittin (MLT), a

toxin peptide consisting of 26 amino acids GIGAVLKVLTTGLPALISWIKRKRQQ, from honey bee (*Apis mellifera*) venom and mastoparan (MAS), a toxin peptide consisting of 14 amino acids INLKALAALAKKIL, from wasp (*Vespula vulgaris*) venom (Huang, Jin et al. 2013, De Azevedo, Figueiredo et al. 2015). A proline rich linker consist of nine proline (Pro) residues was also placed between the lectin sequence and the CatB linker so as to prevent the biodegradable linker to be covered by lectin during folding and to be easily cleaved by cell enzymatic activity. Finally, a $(His)_6$ tag added to the end of *C*-terminus of each construct sequence for purification purpose. Two different restriction sites correspondent to the sites on the pGFPuv expression vector were also introduced to the `5 and 3` primes of each construct as follow: HindIII and EcoRI sites were added respectively to the `5 and 3` primes of the constructs designed from 1st domain and 2nd domain of MLB fused to melittin; XbaI and EcoRI sites were added respectively to the `5 and 3` primes of the constructs designed from MLB fused to melittin and mastoparan, 1st domain, and 2nd domain fused to mastoparan. The designed constructs were synthesized by Genescript Company (Genescript, USA) and they were sent back as plasmid DNA in pUC18 vector as shown in figure 58, 59.

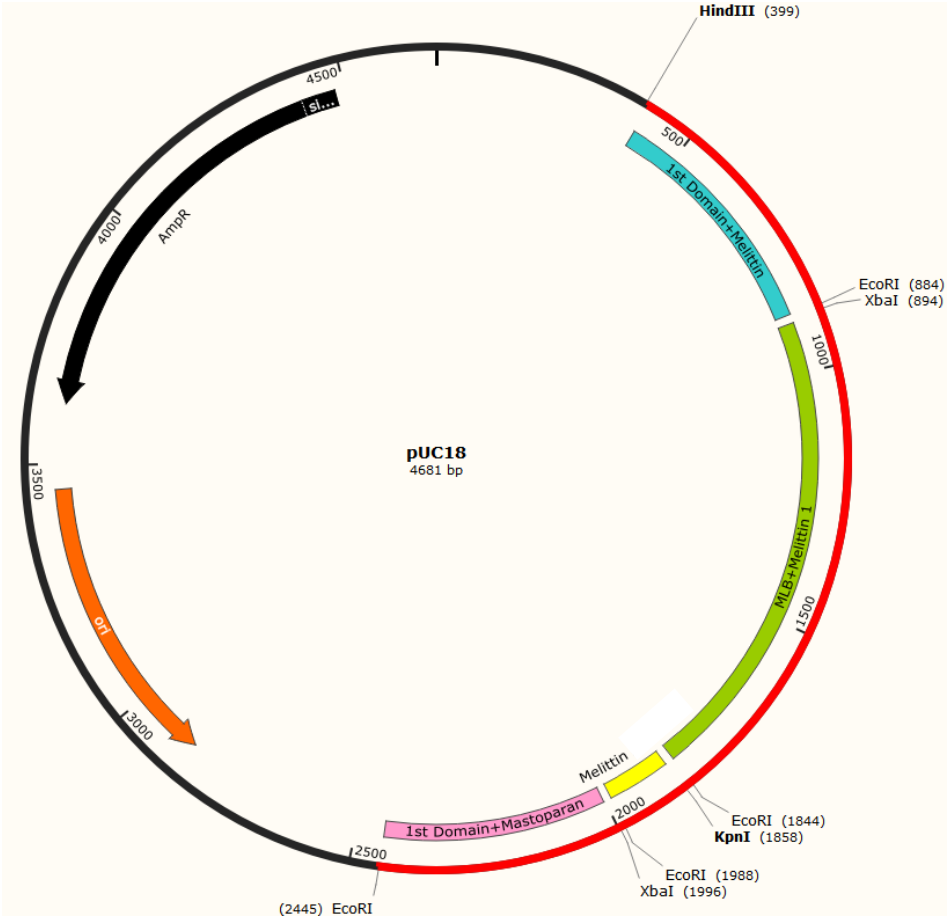


Figure 58. Cloned Lectin-Toxin fusion constructs (RED BAR) pUC18 cloning vector. The BLUE bar represents 1st Domain-Melittin fusion construct; The GREEN bar represents MLB-Melittin 1 fusion construct. The YELLOW bar represents Melittin sequence alone; The PINK bar represents 1st Domain-Mastoparan fusion construct; The BLACK and ORANGE arrows represent the ampicillin resistant (AmpR) and signal of origin of replication (ori) sequences of pUC18 vector.

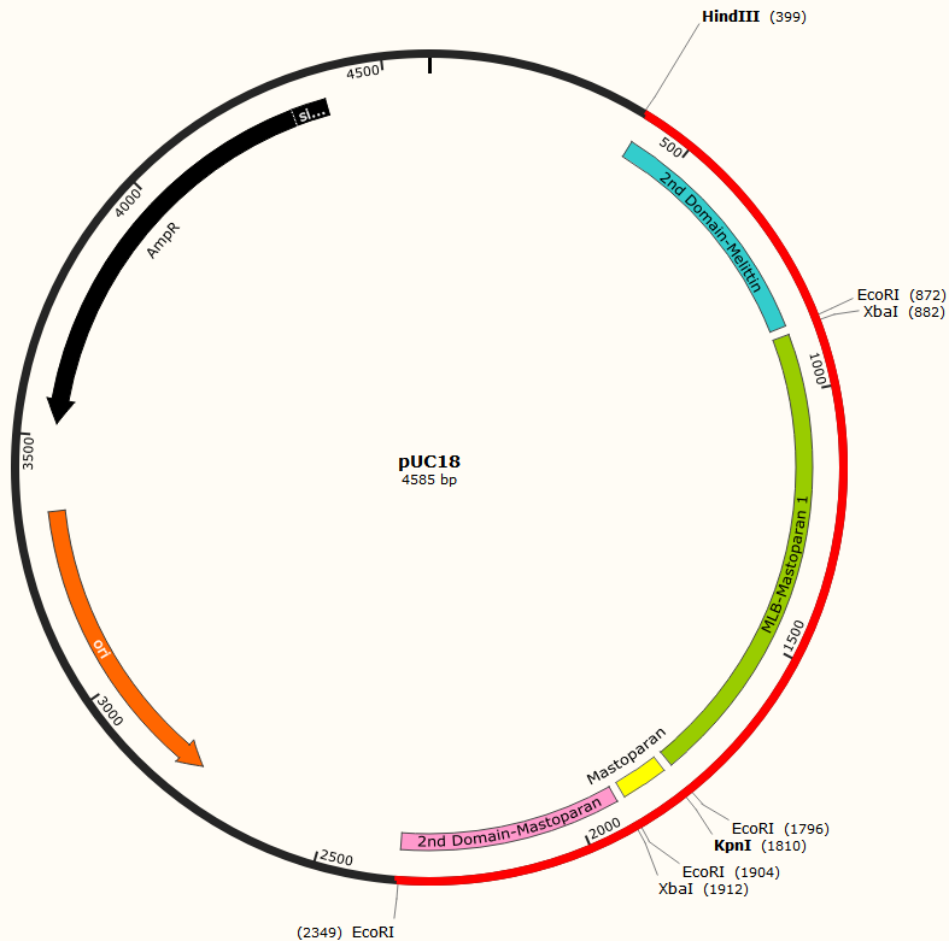


Figure 59. Cloned Lectin-Toxin fusion constructs (RED BAR) pUC18 cloning vector. The BLUE bar represents 2nd Domain-Melittin fusion construct; The GREEN bar represents MLB-Mastoparan 1 fusion construct. The YELLOW bar represents Mastoparan sequence alone; The PINK bar represents 2nd Domain-Mastoparan fusion construct; The BLACK and ORANGE arrows represent the ampicillin resistant (AmpR) and signal of origin of replication (ori) sequences of pUC18 vector.

5.2.4 Restriction Digestion Analyses

The pUC18 containing Lectin-Toxin constructs and the expression vector (pGFPuv) were both digested with appropriate restriction endonuclease enzymes (NEB, UK) so as to sub-clone the fusion constructs from the cloning vector (pUC18) to the expression vector (pGFPuv). Briefly, 1µg of vector was added to different sterile tube containing 10 units of different restriction enzymes (HindIII, XbaI and EcoRI), 2µl Cut Smart buffer and made up to 20µl with sterile water. All tubes were incubated at 37°C for 1hr. After incubation, samples were run on a 1.5% agarose gel at voltage 100V with variable current. The gel was visualised and photographed by UV G-box transilluminator (Syngene, UK).

5.2.5 Ligation of Lectin-Toxin Constructs into pGFPuv Vector

Gel purified Lectin-Toxin fusions (Gel extraction kit, Promega, UK) from the restriction digest reactions were ligated in frame between HindIII, XbaI and EcoRI in the expression vector pGFPuv (Clontech, USA) using T4 ligase enzyme (Invitrogen, UK) following the manufacturers instruction (Figure 60, 61, 62, 63, 64, 65, 66, 67). The components of the ligation reaction table 15 were pipetted into a PCR tube and incubated at 16^{o}C for overnight.

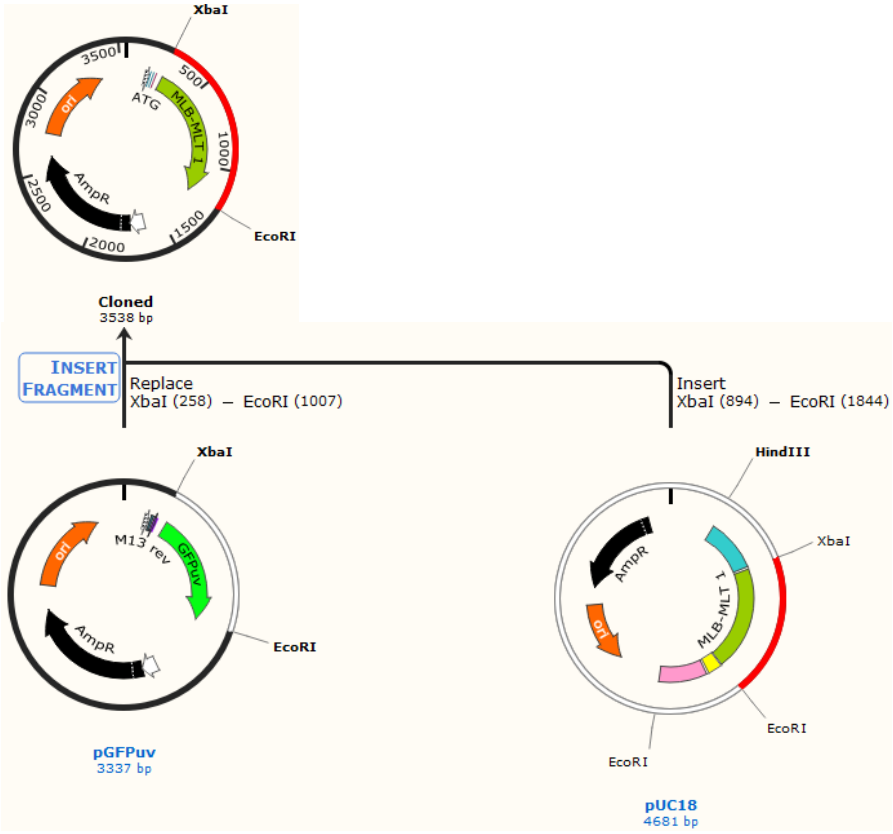


Figure 60. The sub-cloning map of MLB-Melittin 1 fusion construct from pUC18 cloning vector to pGFPuv expression vector. The DARK GREEN arrow represents the MLB-Melittin 1 construct.

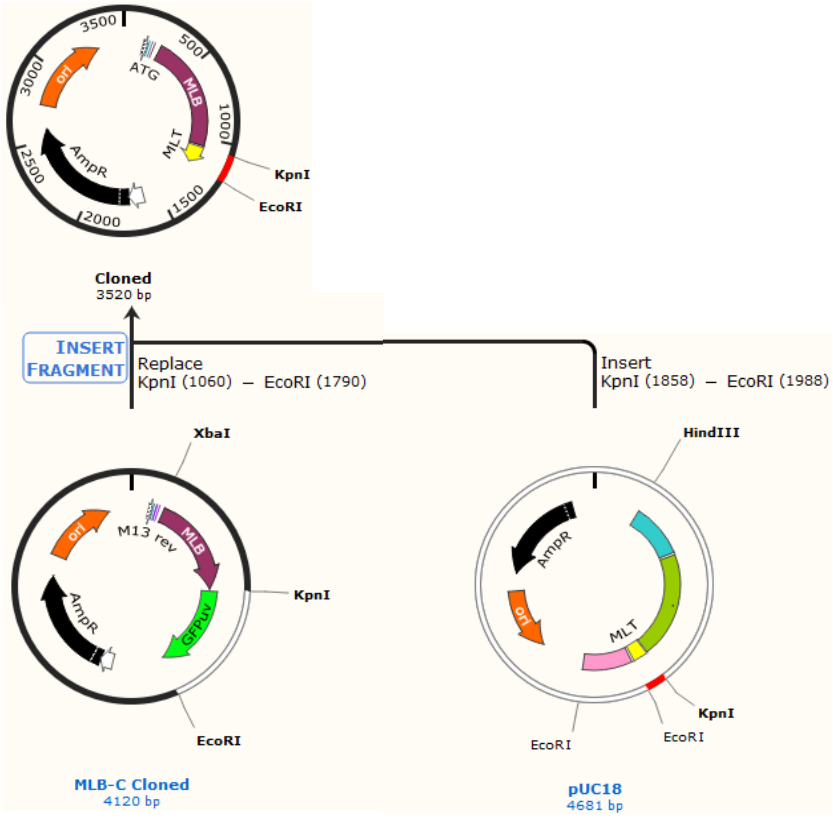


Figure 61. The sub-cloning map of MLB-Melittin 2 fusion construct from pUC18 cloning vector to pGFPuv expression vector. The BURGUNDY arrow represents MLB sequence and the YELLOW arrow represents melittin sequence.

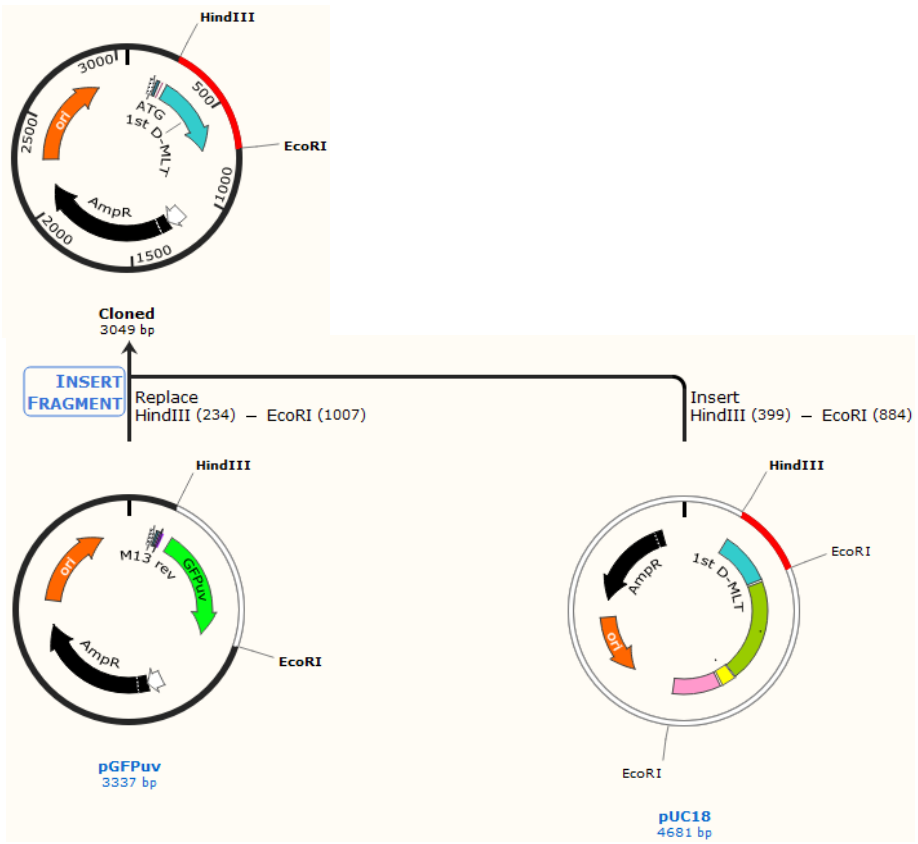


Figure 62. The sub-cloning map of 1st Domain-Melittin fusion construct from pUC18 cloning vector to pGFPuv expression vector. The BLUE arrow represents the 1st Domain-Melittin construct.

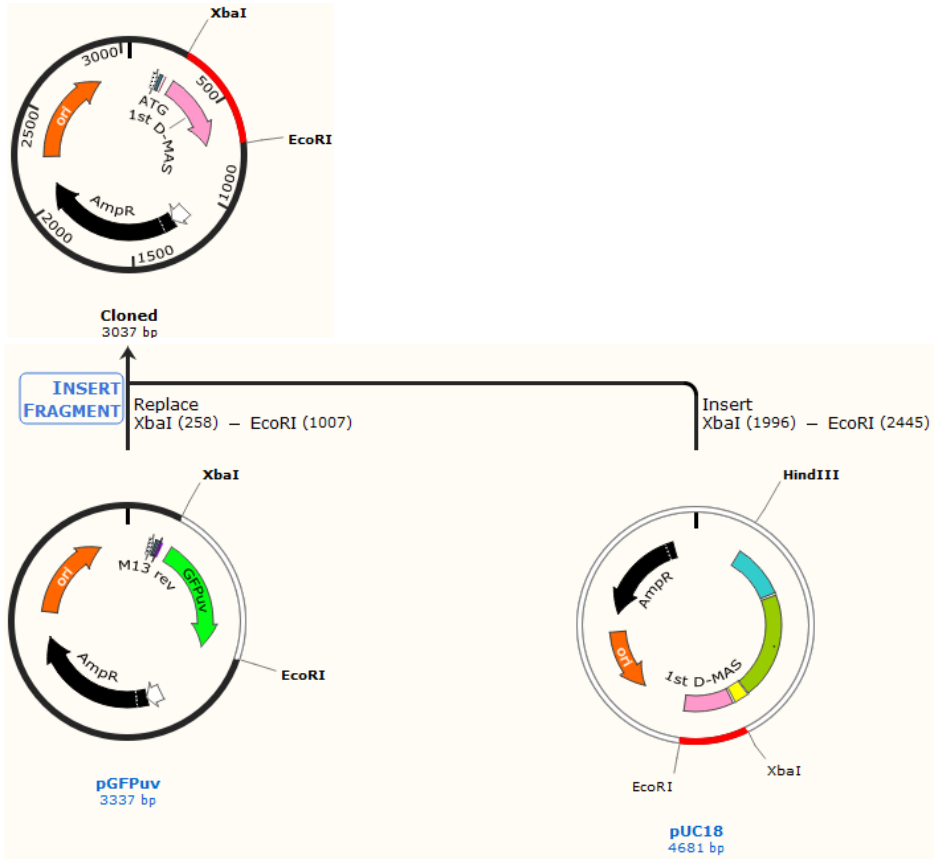


Figure 63. The sub-cloning map of 1st Domain-Mastoparan fusion construct from pUC18 cloning vector to pGFPuv expression vector. The PINK arrow represents the 1st Domain-Mastoparan construct.

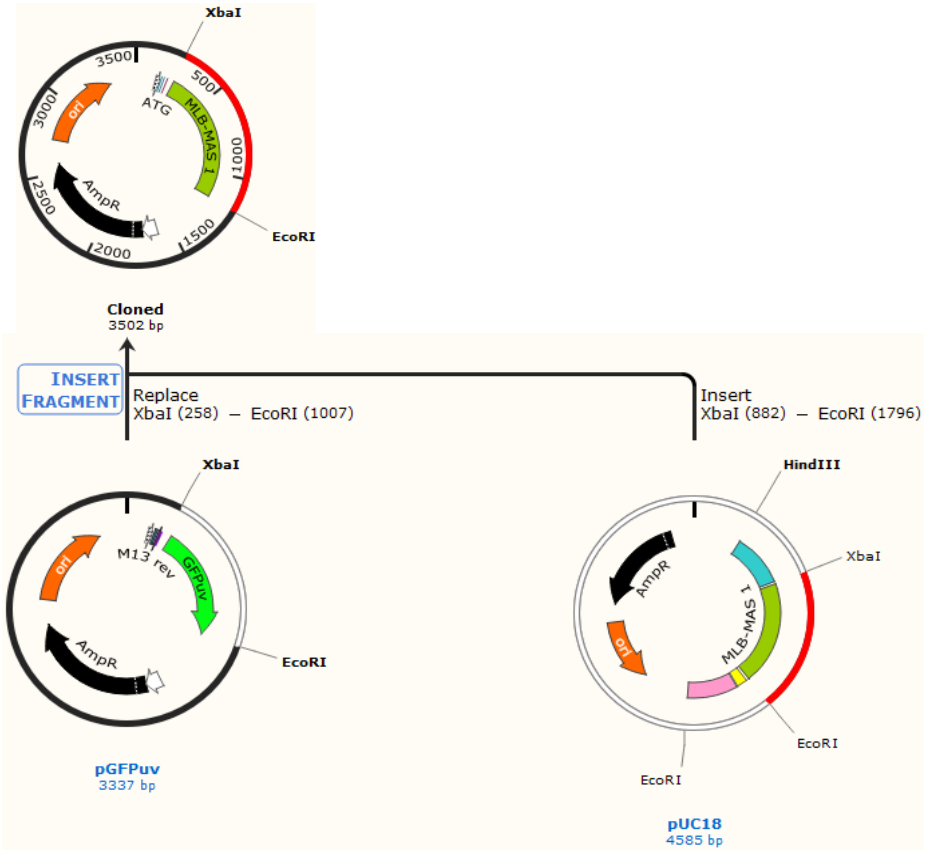


Figure 64. The sub-cloning map of MLB-Mastoparan 1 fusion construct from pUC18 cloning vector to pGFPuv expression vector. The DARK GREEN arrow represents the MLB-Mastoparan 1 construct.

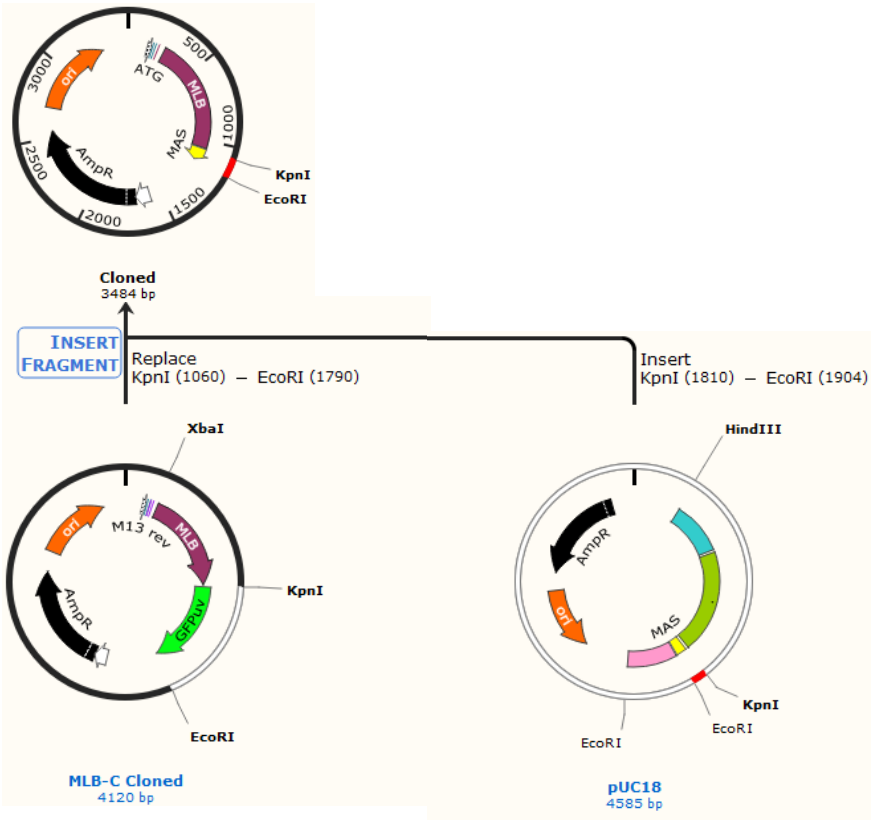


Figure 65. The sub-cloning map of MLB-Mastoparan 2 fusion construct from pUC18 cloning vector to pGFPuv expression vector. The BURGUNDY arrow represents MLB sequence and the YELLOW arrow represents mastoparan sequence.

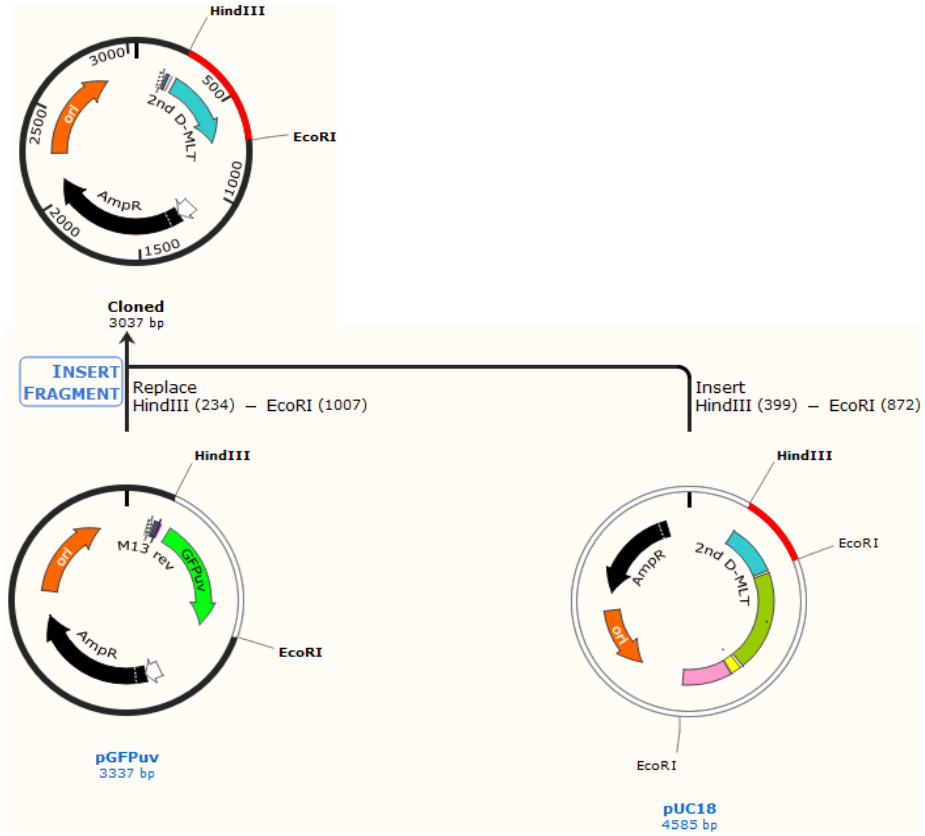


Figure 66. The sub-cloning map of 2nd Domain-Melittin fusion construct from pUC18 cloning vector to pGFPuv expression vector. The BLUE arrow represents the 2nd Domain-Melittin construct.

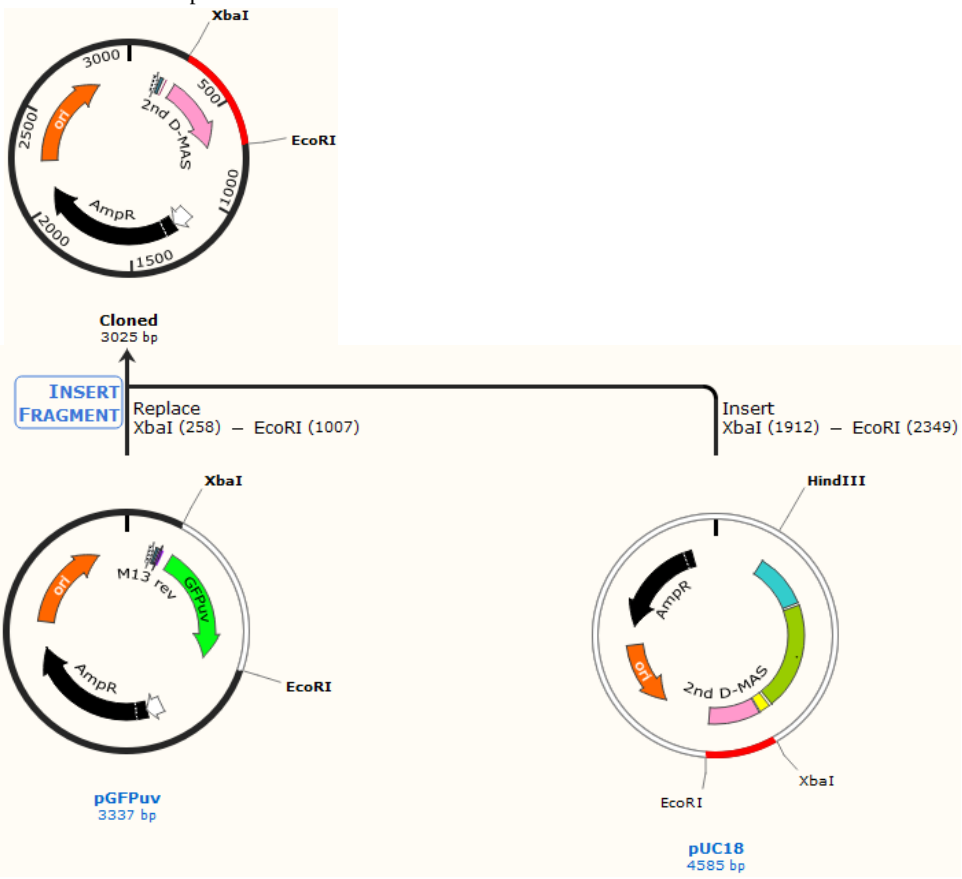


Figure 67. The sub-cloning map of 2nd Domain-Mastoparan fusion construct from pUC18 cloning vector to pGFPuv expression vector. The PINK arrow represents the 2nd Domain-Mastoparan construct.

5.2.6 Transformation of *E. coli* Lemo21 (DE3) with pGFPuv Vector Containing Lectin-Toxin Fusions

To overcome the difficulties in the expression of genes that encode membrane and toxic proteins, a new strain of BL21(DE3) *Escherichia coli* cell line has been constructed known as Lemo21(DE3) which has improved the control over gene expression from T7 promoter. All growth of *E. coli* was performed in LB broth (Luria low salt) medium (Sigma-Aldrich-USA). Ampicillin (100 µg/µl) and chloramphenicol (30 µg/µl) antibiotics were used for the selection of transformants and to maintain pLemo gene encodes T7 lysozyme which is T7 RNA polymerase (RNAP). The LB broth agar was prepared from 15.5 g L^{-1} of LB and 15 g L^{-1} agar (Fluka-Spain) and autoclaved for 15 minutes at 121ºC.

The *E. coli* strain was transformed with pGFPuv expression vector containing Lectin-Toxin constructs using the heat shock method. The Lemo21(DE3) chemically competent cells and the ligation reaction products from section 5.2.5 were mixed and left on ice for 30 minutes followed by heat shock for 10 seconds at 42°C. The cells were then incubated at 37 °C for 1 hour with 250rpm shaking to allow the expression of the ampicillin and chloramphenicol resistance genes in the vector. Transformed cells were selected by growth for 16 hours on LB broth agar containing 100 µg/mL ampicillin and 30 µg/µl chloramphenicol (Fisher, UK). The successfully transformed and grown cells were grown overnight in 5mL of liquid LB broth medium with 100 µg/mL ampicillin and 30 µg/µl chloramphenicol at 37 °C with shaking at 250 rpm.

5.2.7 Plasmid DNA Purification

Plasmids were isolated from bacterial cells using ISOLATE Plasmid Mini kit (BIOLINE, UK), according to manufacturer's instructions shown in section 2.2.10.

5.2.8 Sequencing and Analytical Restriction Digestion

After diagnostic restriction digest and confirming the right size insert in the purified plasmid DNA of all the Lectin-Toxin constructs, sequencing of the constructs was carried out by the Sanger sequencing facility of Source Bioscience Company (Source Bioscience-UK). The standard M13 Reverse primer and PGEX Reverse were used for the sequencing reactions. The sequencing results were analysed through the alignment of the DNA sequences and there translated amino acids with the reference sequences using Snapgene (Snapgene, USA) and CLC Sequence Viewer softwares (Qiagen, UK).

5.2.9 Expression of Recombinant Lectin-Toxins

The recombinant expression vector pGFPuv (Clonthech, USA) was used to express the targeted Lectin-Toxin fusion constructs. The overnight culture was prepared for all the constructs as described in section 3.2.6 and 1ml from the culture was added to 100 ml liquid culture with ampicillin (100 µg/µl) and chloramphenicol (30 µg/µl) antibiotics and incubated at 30°C until OD600 reaches 0.4–0.8 followed by the induction through adding 400 µM IPTG and incubation for 18 hours at 30°C. The expression level of the targeted genes in the total cell extracts was checked by InstantBlue Coomassie (Expedeon, UK) stained SDS PAGE.

5.2.10 Recombinant Protein Purification

Following the expression of Lectin-Toxin fusion proteins, culture supernatant was clarified by centrifugation and the *E. coli* cell pellets were gently resuspended and shaken in BugBuster reagent to disrupt the cell walls resulting in the liberation of soluble protein. To separate the protein of interest from the bulk of total other bacterial protein, a string of nucleotides encode six histidine amino acids can be added to the sequence of the protein to

be purified so as to conduct His-Tag metal ion purification procedure. The affinity of histidine to ions like Ni^{2+} and Co^{2+} makes the expressed gene construct to easily bind to matrix containing these ions and to be easily detached by using solution containing competing ions such as imidazole and EDTA (Hengen 1995).

Immobilized metal ion affinity chromatography (IMAC) was conducted to purify targeted lectin-toxin fusions from the total extracted proteins using the prepacked HiTrap™ column and AKTA Start FPLC platform (GE Healthcare, UK). The relevant fractions were finally analysed by SDS-PAGE and western blotting.

5.2.11 Protein Biochemistry

5.2.11.1 Protein Concentration Assay

The BCA protein assay kit was used to determine the concentration of the purified recombinant fusion proteins in microassay size as described in section 4.2.2.1.

The fusion proteins sequences were also computationally analysed using ExPASy server (Gasteiger, Gattiker et al. 2003). The computed molecular weight and extinction coefficient of the fusion proteins are shown in table 18.

Table 18. Computed parameters of Lectin-Toxin fusion proteins.

Fusion Proteins	Molecular Weight (g/mol)	Extinction Coefficient M-1 cm-1
MLB-Melittin 1	36028.9	66055
MLB-Melittin 2	35329.1	66055
MLB-Mastoparan 1	34661.4	60555
MLB-Mastoparan 2	33609.1	60555
1st Domain-Melittin	18309.9	33585
1st Domain-Mastoparan	17840.3	28085
2nd Domain-Melittin	17595.2	25230
2nd Domain-Mastoparan	17125.6	19730

5.2.11.2 Sodium Dodecyl Sulphate Polyacrylamide Gel Electrophoresis

SDS-PAGE of the protein samples was carried out using Tris-glycin buffer system under denaturing conditions as described in section 4.2.2.2.

5.2.11.3 Western Blot

Following separation of the denatured protein samples on SDS PAGE as described in section 4.2.2.2, the proteins were transferred to nitrocellulose membrane using semi dry blotting technique as described in section 4.2.2.3. The monoclonal Ani-*His* Antibody 1:2,000 (Invitrogen, UK) that recognizes the *His*-tag fused to the *C*- terminus of the targeted proteins was used as a primary antibody and for the secondary antibody Goat IgG Anti-Mouse IgG-HRP conjugated secondary antibody 1:1000, (Sigma-Aldrich-UK) was used. Antibody binding was visualised using Pierce ECL substrate solution as described by the manufacturer (Thermofisher Scientific). The membrane was visualised and photographed by G-box transilluminator (Syngene, UK).

5.3 Results

Following the confirmation of cellular binding and visualization of cellular internalization of the mistletoe lectin-GFP fusion proteins, a reductionist approach was taken to determine the minimal number of glycan binding domains in the MLB chain needed to retain biological activity. Furthermore, GFP was replaced in the fusion protein with a known cytotoxic peptide and fusion proteins expressed to achieve a toxic moiety to cancer cells.

5.3.1 Conserved Domain Analysis

Conserved domain analysis was carried out for the MLB sequence using NCBI Conserved Domain Database. The analysis revealed that the MLB sequence consists of two separate ricin type beta-trefoil glycan binding sites shown in figure 68 and the biological activity of both domains were also studied separately. In order to easily differentiate and recognize the domains in this study, the names 1st Domain and 2nd Domain have been given to the first and second glycan binding sites respectively.

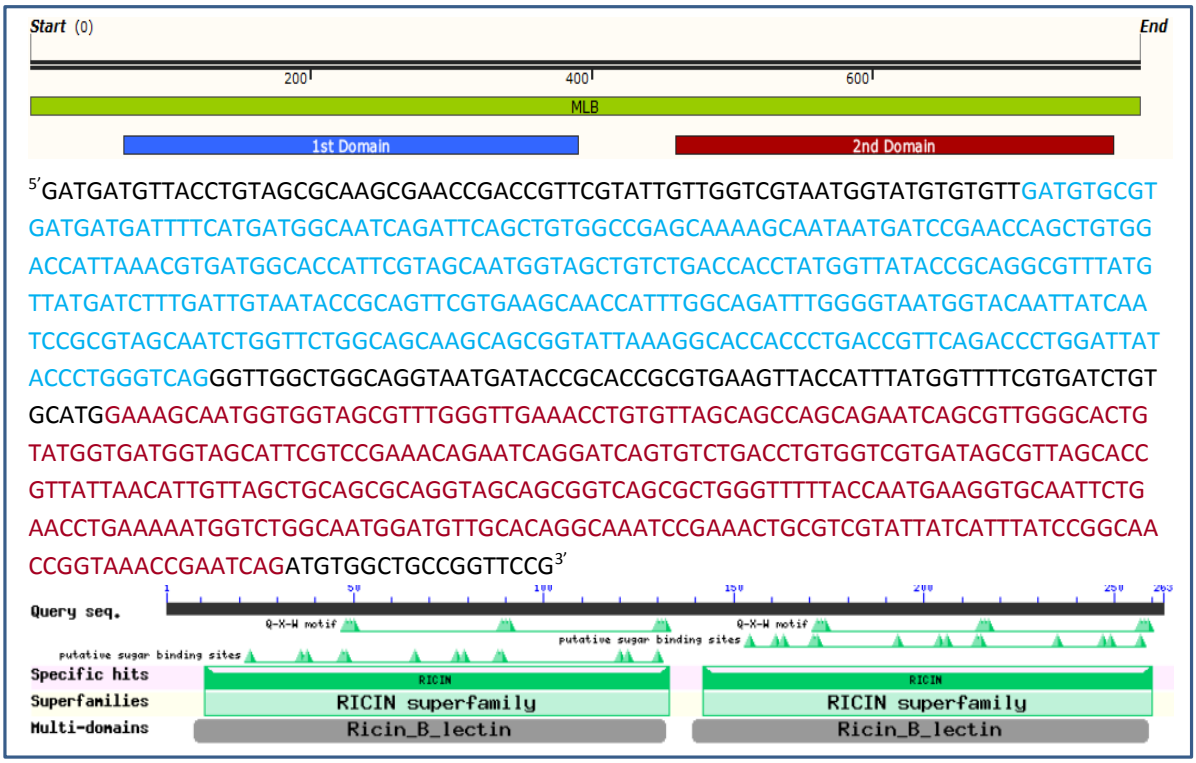


Figure 68. The conserved domain analysis of the full sequence of the glycan binding chain (MLB) of mistletoe lectin. The 1st Domain sequence is in blue and the 2nd Domain sequence is in brown.

5.3.2 Codon Optimization

Protein production through heterologous gene expression often faces difficulties because of inefficient level of transcription and translation of a particular coding sequence placed in a non-native context. The main reason is that the tRNA pools are different between organisms and to overcome this problem the codons of a heterologous gene can be optimized for tRNA pools in a specific host. The sequence of MLB, melittin and mastoparan were optimized to *E. coli* tRNA pools so as to increase their expression and differences between the original sequences and the optimized sequences are shown in figure 69, 70, 71.

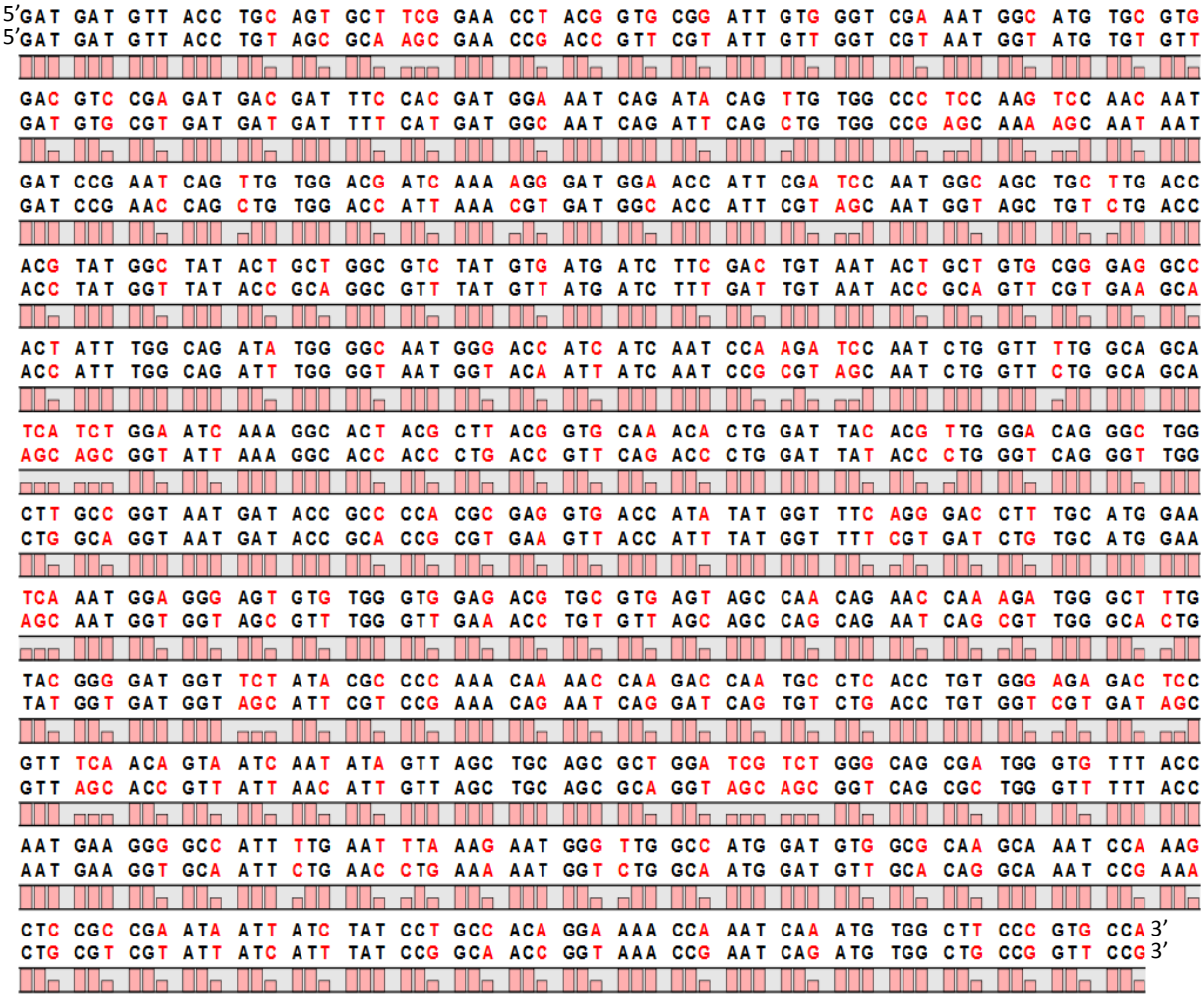


Figure 69. Alignment of reference MLB chain sequence (NCBI: AY377890.1) with the optimized MLB chain. The matched nucleotides are in BLACK; the mismatched (Optimized) nucleotides are in RED and also indicated as gap in the plot bar.

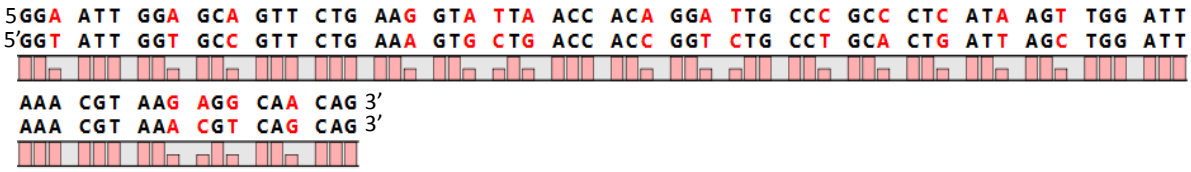


Figure 70. Alignment of reference Melittin sequence (NCBI: AY745248.1) with the optimized Melittin sequence. The matched nucleotides are in BLACK; the mismatched (Optimized) nucleotides are in RED and also indicated as gap in the plot bar.

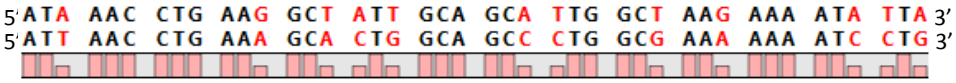


Figure 71. Alignment of reference Mastoparan sequence (NCBI: DQ865260.1) with the optimized Mastoparan sequence. The matched nucleotides are in BLACK; the mismatched (Optimized) nucleotides are in RED and also indicated as gap in the plot bar.

It can clearly be seen in the figures that many of the nucleic acids in the sequences have been changed so as to give them better consistency which consequently improve the expression of the genes in *E. coli*. Also, codon optimization reduces the number of codon class and makes the sequence codons to fall into a certain quality class. Figure 72, 73, 74 show the codon quality distribution of MLB, melittin and mastoparan respectively before and after codon optimization. In the unoptimized sequences, the codons are distributed to several class (Figure 72, 73, 74 A). However, in the unoptimized sequences majority of the codons belongs to one class (Figure 72, 73, 74 B) which helps successful expression in *E. coli* (Sharp and Li 1987).

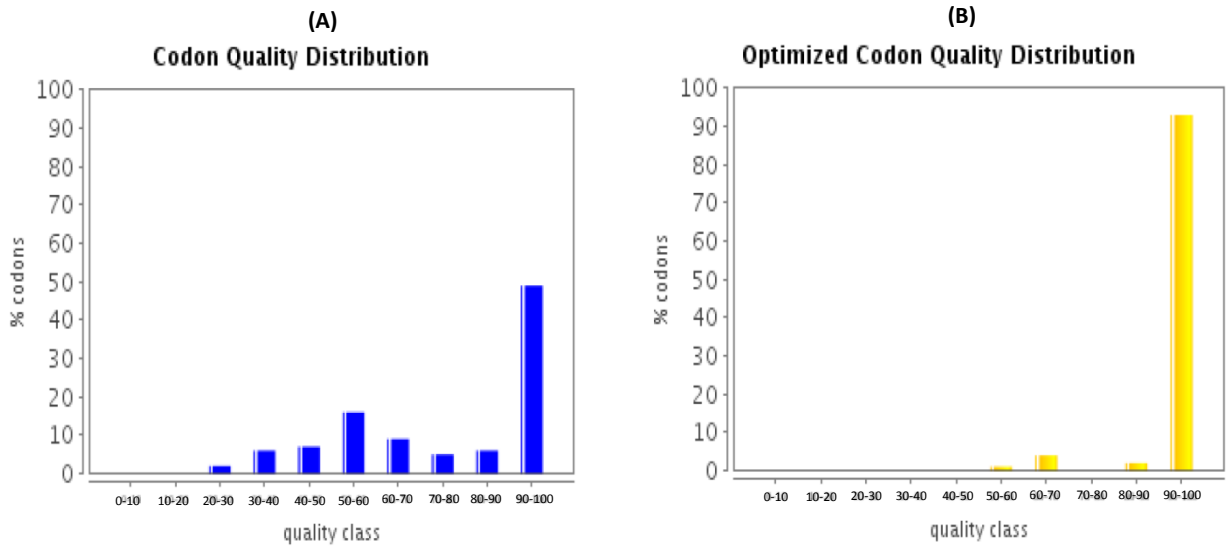


Figure72. Histograms show the percentage of unoptimized and optimized sequence codons of MLB which fall into a certain quality class. The quality value of the most frequently used codon for a given amino acid in the desired expression system is set to 100, the remaining codons are scaled accordingly.

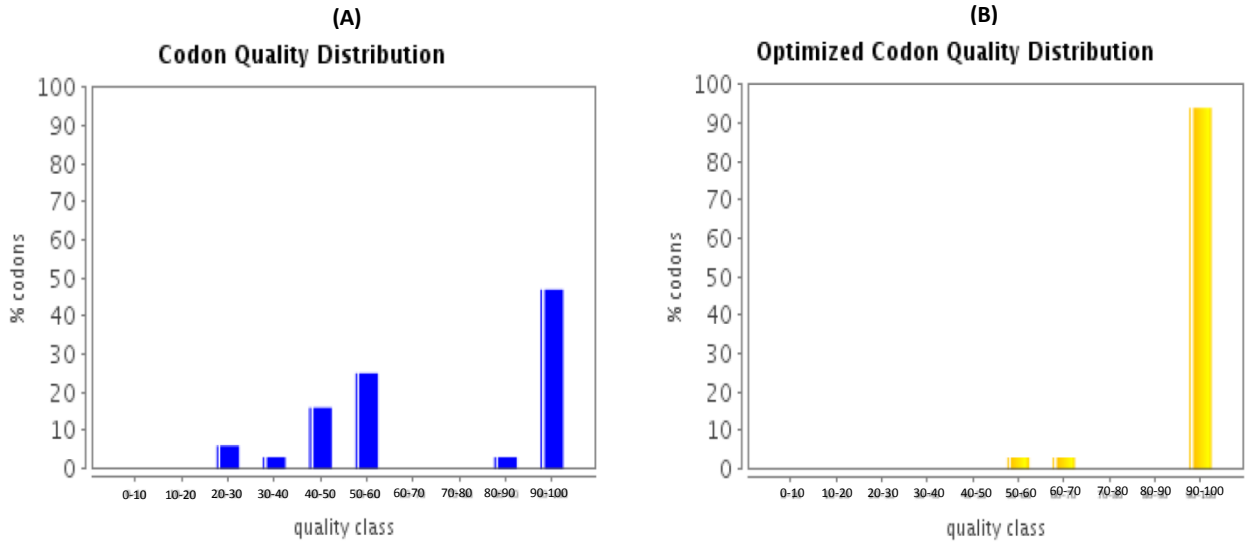


Figure73. Histograms show the percentage of unoptimized and optimized sequence codons of melittin (ML) which fall into a certain quality class. The quality value of the most frequently used codon for a given amino acid in the desired expression system is set to 100, the remaining codons are scaled accordingly.

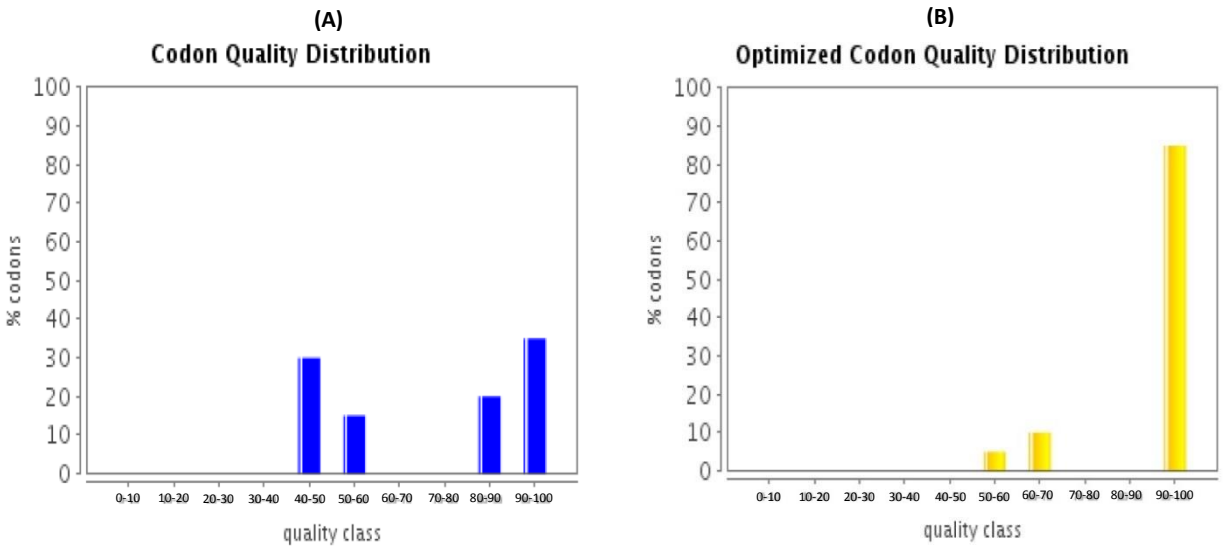


Figure74. Histograms show the percentage of unoptimized and optimized sequence codons of mastoparan (Mas) which fall into a certain quality class. The quality value of the most frequently used codon for a given amino acid in the desired expression system is set to 100, the remaining codons are scaled accordingly.

5.3.3 Lectin-Toxin Fusion Proteins

To assess the suitability of mistletoe lectin B chain (MLB) to be used as a 'carrier' to deliver cytotoxic agents specifically to cancer cells, recombinant fusion proteins were designed from MLB and its two glycan binding domains *C*- terminally fused to two different toxin peptides derived from *Apis mellifera* (Honey bee) and *Vespula vulgaris* (Wasp).

Due to the large number of Lectin-Toxin fusion proteins produced and large amount of data generated, Table 19 summarises the name and composition of the fusion proteins.

Table 19. Names and summary of the fusion proteins produced from MLB and the toxin peptides.

Fusion Proteins	Linker	Detail
MLB-Melittin 1	No Linker	The whole MLB chain fused to Melittin
MLB-Melittin 2	CatB	The whole MLB chain fused to Melittin
MLB-Mastoparan 1	No Linker	The whole MLB chain fused to Mastoparan
MLB-Mastoparan 2	CatB	The whole MLB chain fused to Mastoparan
1st Domain-Melittin	CatB	The 1st glycan binding domain of MLB chain fused to Melittin
1st Domain-Mastoparan	CatB	The 1st glycan binding domain of MLB chain fused to Mastoparan
2nd Domain-Melittin	CatB	The 2nd glycan binding domain of MLB chain fused to Melittin
2nd Domain-Mastoparan	CatB	The 2nd glycan binding domain of MLB chain fused to Mastoparan

5.3.3.1 Production of the MLB-Melittin and MLB-Mastoparan Fusion Proteins Expression Constructs

Expression constructs for four MLB-Toxin fusion proteins (designated MLB-Melittin 1, MLB-Melittin 2, MLB-Mastoparan 1 and MLB-Mastoparan 2) were designed and produced by the cloning strategy described in section 5.2.5

The MLB-Melittin 1 construct contains MLB *C*-terminally fused to Melittin toxin peptide (GIGAVLKVLTTGLPALISWIKRKRQQ) via a cathepsin-B (CatB) (GIPVSLRSK) biodegradable linker. The MLB-Melittin 2 construct contains MLB *C*-terminally fused to Melittin toxin peptide without cathepsin-B (CatB) biodegradable linker. The MLB-Mastoparan 1 construct contains MLB *C*-terminally fused to Mastoprant toxin peptide (INLKALAALAKKIL) via a cathepsin-B (CatB) (GIPVSLRSK) biodegradable linker. The MLB-Mastoparan 2 construct contains MLB *C*-terminally fused to Mastoprant toxin peptide without cathepsin-B (CatB) biodegradable linker. All the expression constructs also had a *C*-terminal extension encoding a $(His)_6$ tag. The complete expression constructs of the four fusion proteins were cloned in-frame with the start codon (ATG) within the

pGFPuv expression vector. The nucleotide, translated amino acid sequences and schematic diagrams of the expression constructs are shown in figures 75, 76, 77, 78.

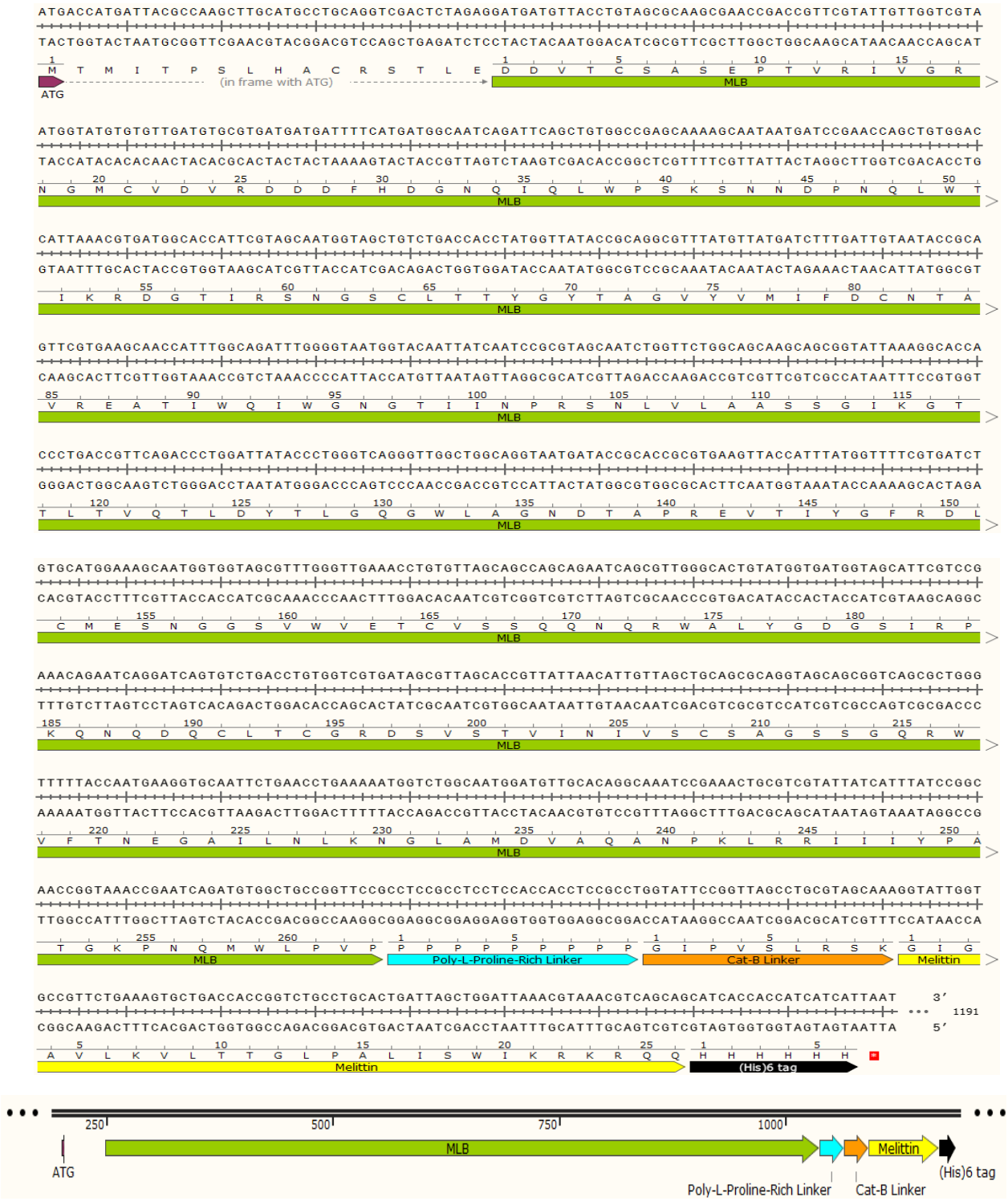
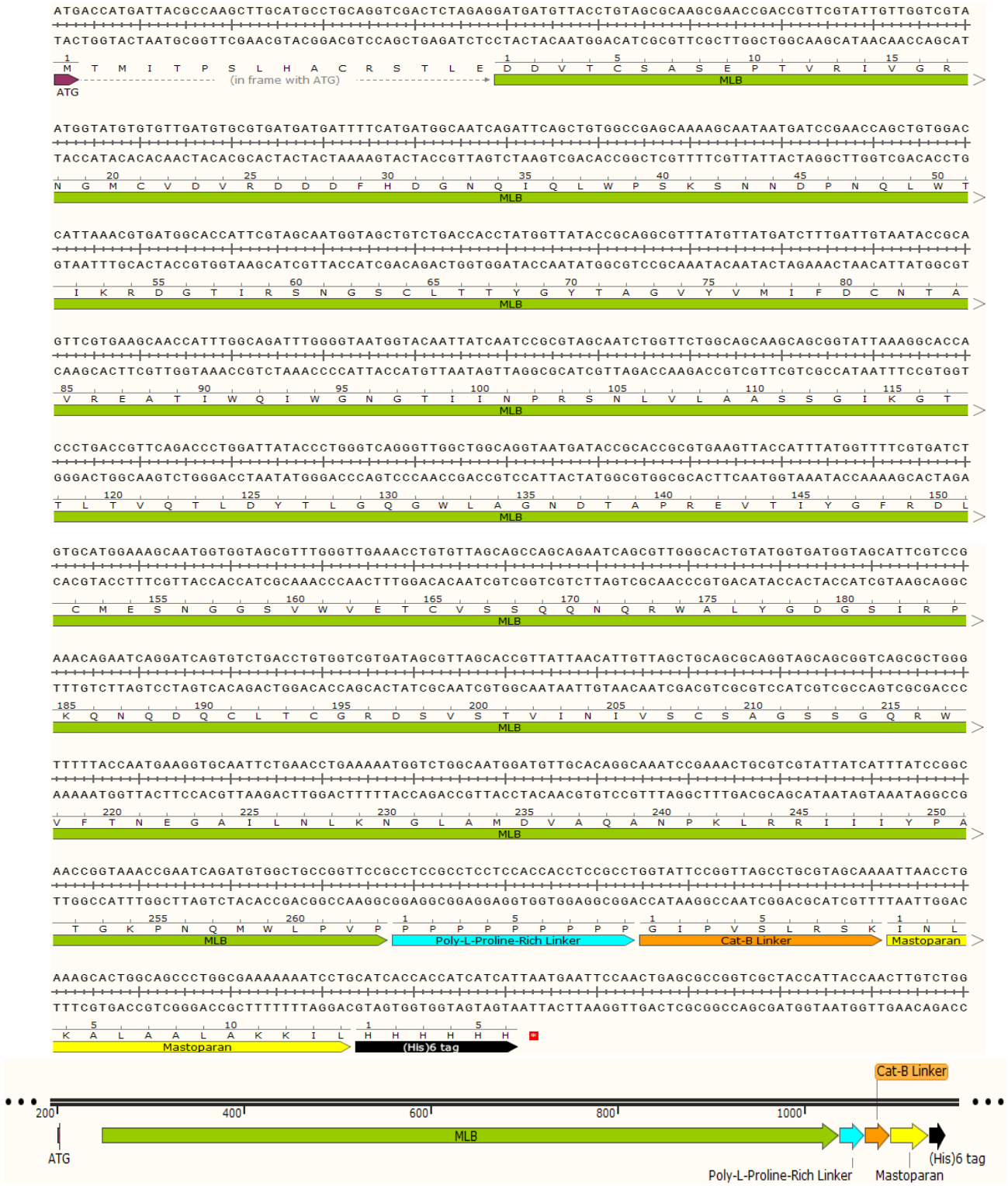


Figure75. Nucleotide sequence, deduced amino acid sequence and schematic representation of the MLB-Melittin 1 fusion protein construct. The gene encoding MLB is highlighted in GREEN (bp 265-1053). The poly-L-proline-rich linker is highlighted in BLUE (bp 1054-1080). The cathepsin-B (CatB) biodegradable linker is highlighted in ORANGE (bp 1081-1107). The gene encoding Melittin is highlighted in YELLOW (bp 1108-1185). The (His)6 tag is highlighted in BLACK (bp 1186-2204). The stop codon is highlighted in a red star.

[Figure76](). Nucleotide sequence, deduced amino acid sequence and schematic representation of the MLB-Melittin 2 fusion protein construct. The gene encoding MLB is highlighted in Green (bp 265-1053). The poly-L-proline-rich linker is highlighted in BLUE (bp 1054-1080). The gene encoding Melittin is highlighted in YELLOW (bp 1081-1158). The (His)6 tag is highlighted in BLACK (bp 1159-1176). The stop codon is highlighted in a red star.



[Figure77](). Nucleotide sequence, deduced amino acid sequence and schematic representation of the MLB-Mastoparan 1 fusion protein construct. The gene encoding MLB is highlighted in GREEN (bp 265-1053). The poly-L-proline-rich linker is highlighted in BLUE (bp 1054-1080). The cathepsin-B (CatB) biodegradable linker is highlighted in ORANGE (bp 1081-1107). The gene encoding Mastoparan is highlighted in YELLOW (bp 1108-1149). The (His)6 tag is highlighted in BLACK (bp 1150-1167). The stop codon is highlighted in a red star.

Figure78. Nucleotide sequence, deduced amino acid sequence and schematic representation of the MLB-Mastoparan 2 fusion protein construct. The gene encoding MLB is highlighted in GREEN (bp 265-1053). The poly-L-proline-rich linker is highlighted in BLUE (bp 1054-1080). The gene encoding Mastoparan is highlighted in YELLOW (bp 1081-1122). The (His)6 tag is highlighted in BLACK (bp 1123-1140). The stop codon is highlighted in a red star.

5.3.3.2 Glycan Binding Domains -Toxin Fusion Proteins

In order to explore which domain is responsible for the sialic acid binding ability of MLB chain, fusion proteins were also designed for the two glycan binding domains (1st & 2nd Domains) separately.

5.3.4.2.1 Production of the 1st Domain-Melittin and 1st Domain-Mastoparan Fusion Proteins Expression Constructs

Expression constructs for two 1st Domain-Toxin fusion proteins (designated 1st Domain-Melittin and 1st Domain-Mastoparan) were designed and produced by the cloning strategy described in section 5.2.5.

The 1st Domain-Melittin construct contains 1st Domain *C*-terminally fused to Melittin toxin (GIGAVLKVLTTGLPALISWIKRKRQQ) via a cathepsin-B (CatB) (GIPVSLRSK) biodegradable linker. The 1st Domain-Mastoparan construct contains 1st Domain *C*-terminally fused to Mastoparan toxin (INLKALAALAKKIL) via cathepsin-B (CatB) (GIPVSLRSK) biodegradable linker. Both expression constructs also had a *C*-terminal extension encoding a $(His)_6$ tag. The complete expression constructs of the four fusion proteins were cloned in-frame with the start codon (ATG) within the pGFPuv expression vector. The nucleotide, translated amino acid sequences and schematic diagrams of the expression constructs are shown in figures 79, 80.

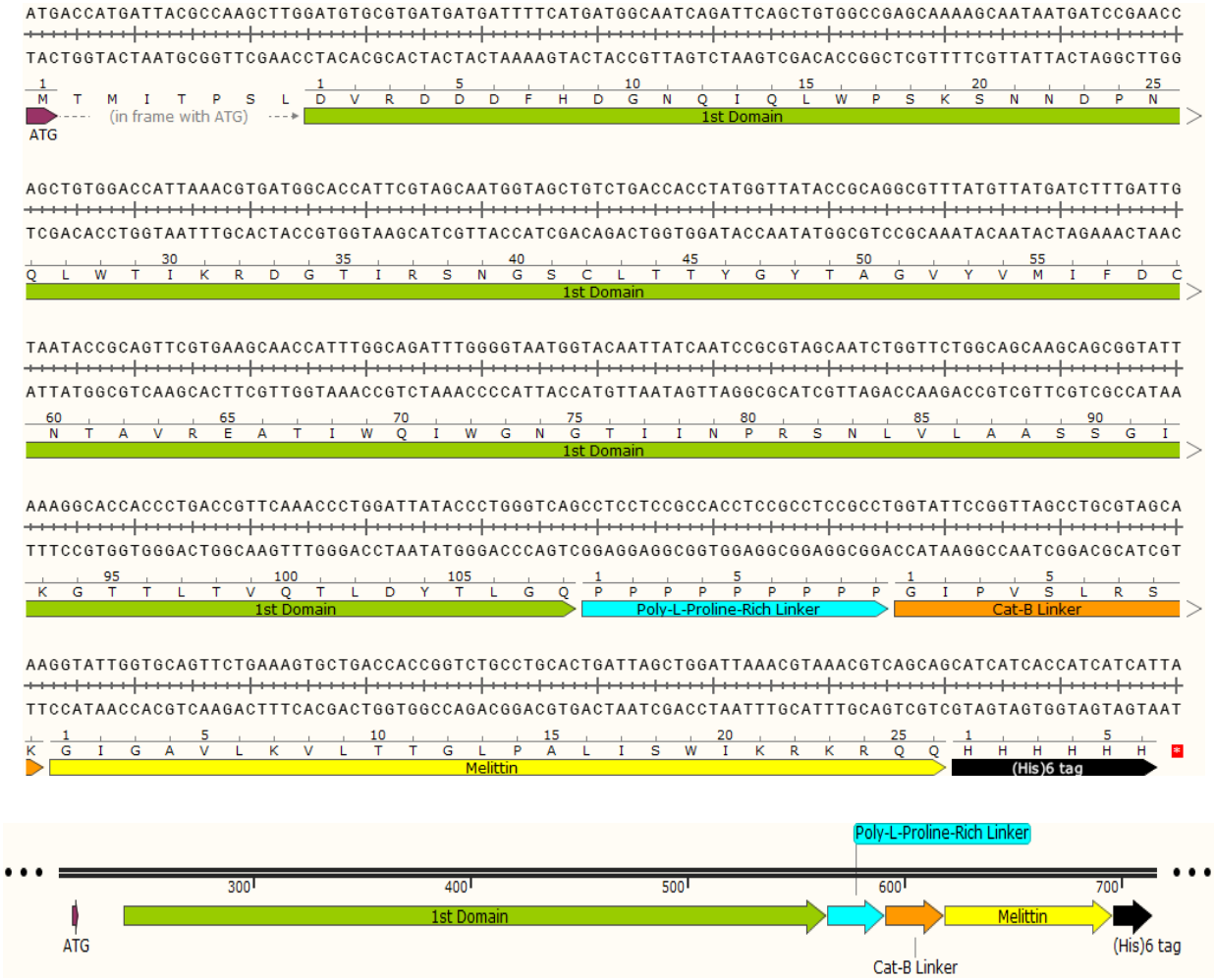


Figure79. Nucleotide sequence, deduced amino acid sequence and schematic representation of the 1st Domain-Melittin fusion protein construct. The gene encoding 1st Domain of MLB is highlighted in GREEN (bp 241-564). The poly-L-proline-rich linker is highlighted in BLUE (bp 565-591). The cathepsin-B (CatB) biodegradable linker is highlighted in ORANGE (bp 592-618). The gene encoding Melittin is highlighted in YELLOW (bp 619-696). The (His)6 tag is highlighted in BLACK (bp 697-714). The stop codon is highlighted in a red star.

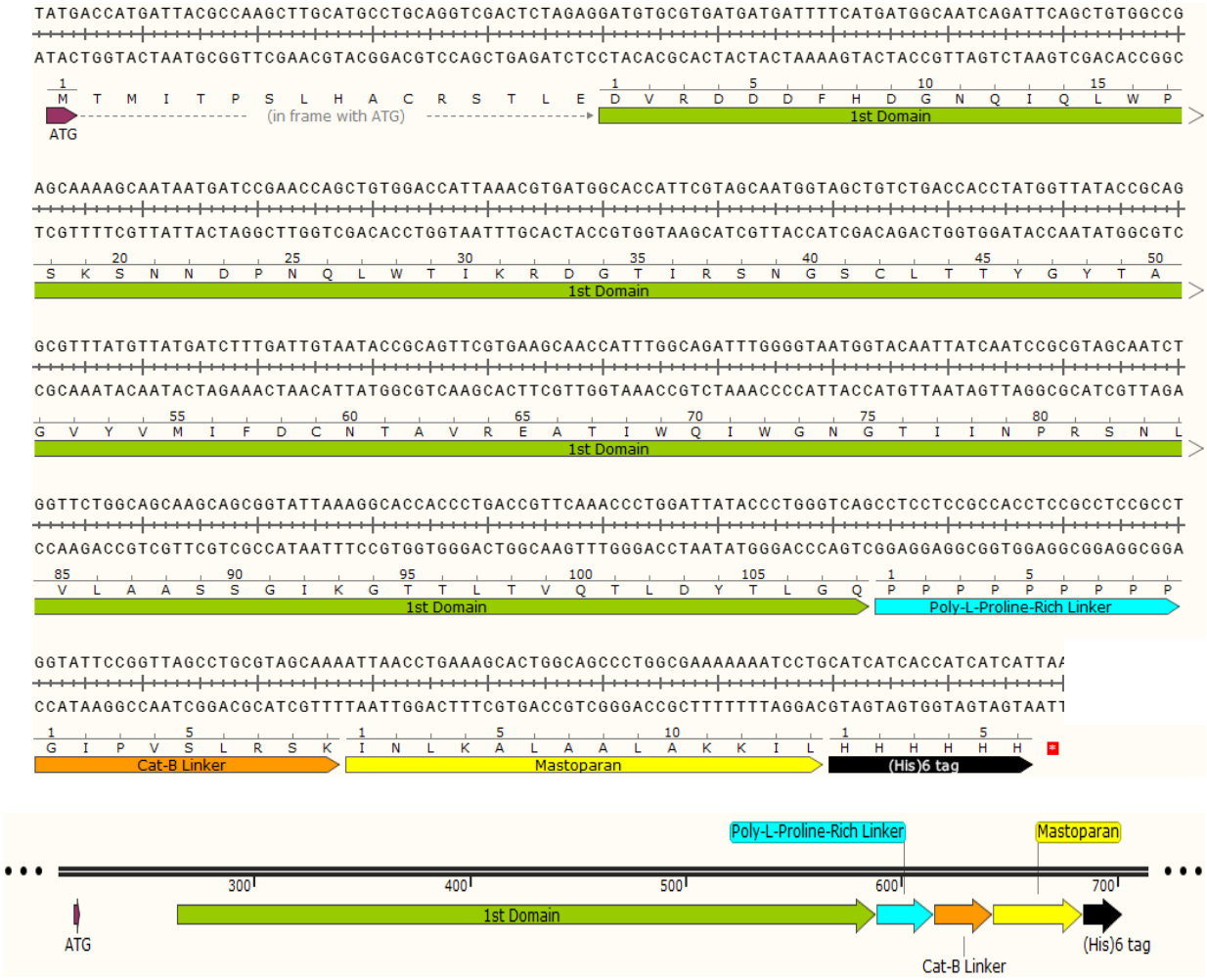


Figure80. Nucleotide sequence, deduced amino acid sequence and schematic representation of the 1st Domain-Mastoparan fusion protein construct. The gene encoding 1st Domain of MLB is highlighted in GREEN (bp 265-588). The poly-L-proline-rich linker is highlighted in BLUE (bp 589-615). The cathepsin-B (CatB) biodegradable linker is highlighted in ORANGE (bp 616-642). The gene encoding Mastoparan is highlighted in YELLOW (bp 643-684). The (His)6 tag is highlighted in BLACK (bp 685-702). The stop codon is highlighted in a red star.

5.3.4.2.2 Production of the 2nd Domain-Melittin and 2nd Domain-Mastoparan Fusion Proteins Expression Constructs

Expression constructs for two 2nd Domain fusion proteins (designated 2nd Domain-Melittin and 2nd Domain-Mastoparan) were designed and produced by the cloning strategy described in section 5.2.5.

The 2nd Domain-Melittin construct contains 2nd Domain *C*-terminally fused to Melittin toxin (GIGAVLKVLTTGLPALISWIKRKRQQ) via a cathepsin-B (CatB) (GIPVSLRSK) biodegradable linker. The 2nd Domain-Mastoparan construct contains 1st Domain *C*-terminally fused to Mastoparan toxin (INLKALAALAKKIL) via cathepsin-B (CatB) (GIPVSLRSK) biodegradable linker. Both expression constructs also had a *C*-terminal extension encoding a $(His)_6$ tag. The complete expression constructs of the four fusion proteins were cloned in-frame with the start codon (ATG) within the pGFPuv expression vector. The nucleotide, translated amino acid sequences and schematic diagrams of the expression constructs are shown in figures 81, 82.

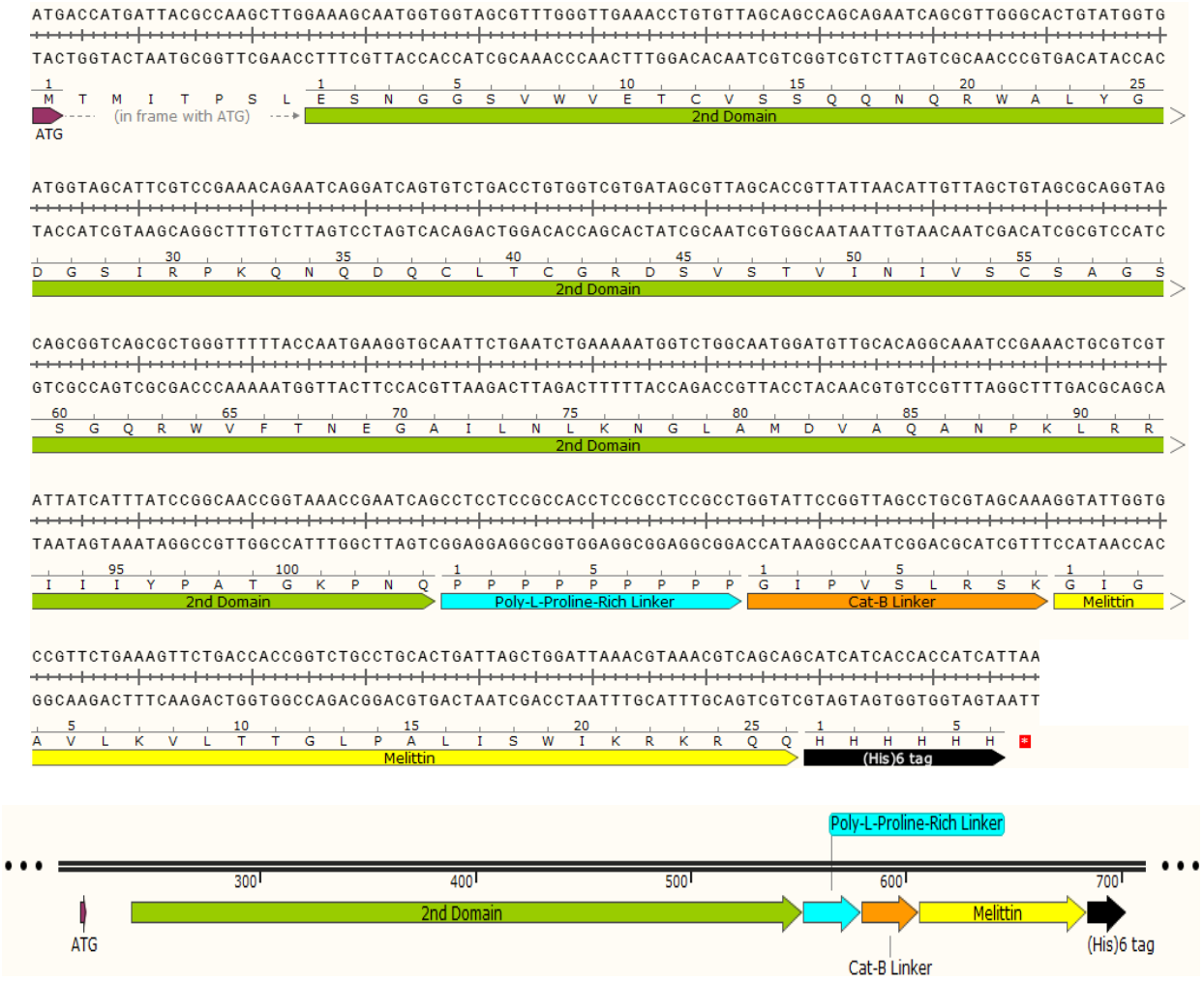


Figure81. Nucleotide sequence, deduced amino acid sequence and schematic representation of the 2nd Domain-Melittin fusion protein construct. The gene encoding 2nd Domain of MLB is highlighted in GREEN (bp 241-552). The poly-L-proline-rich linker is highlighted in BLUE (bp 553-579). The cathepsin-B (CatB) biodegradable linker is highlighted in ORANGE (bp 580-606). The gene encoding Melittin is highlighted in YELLOW (bp 607-684). The (His)6 tag is highlighted in BLACK (bp 685-702). The stop codon is highlighted in a red star.

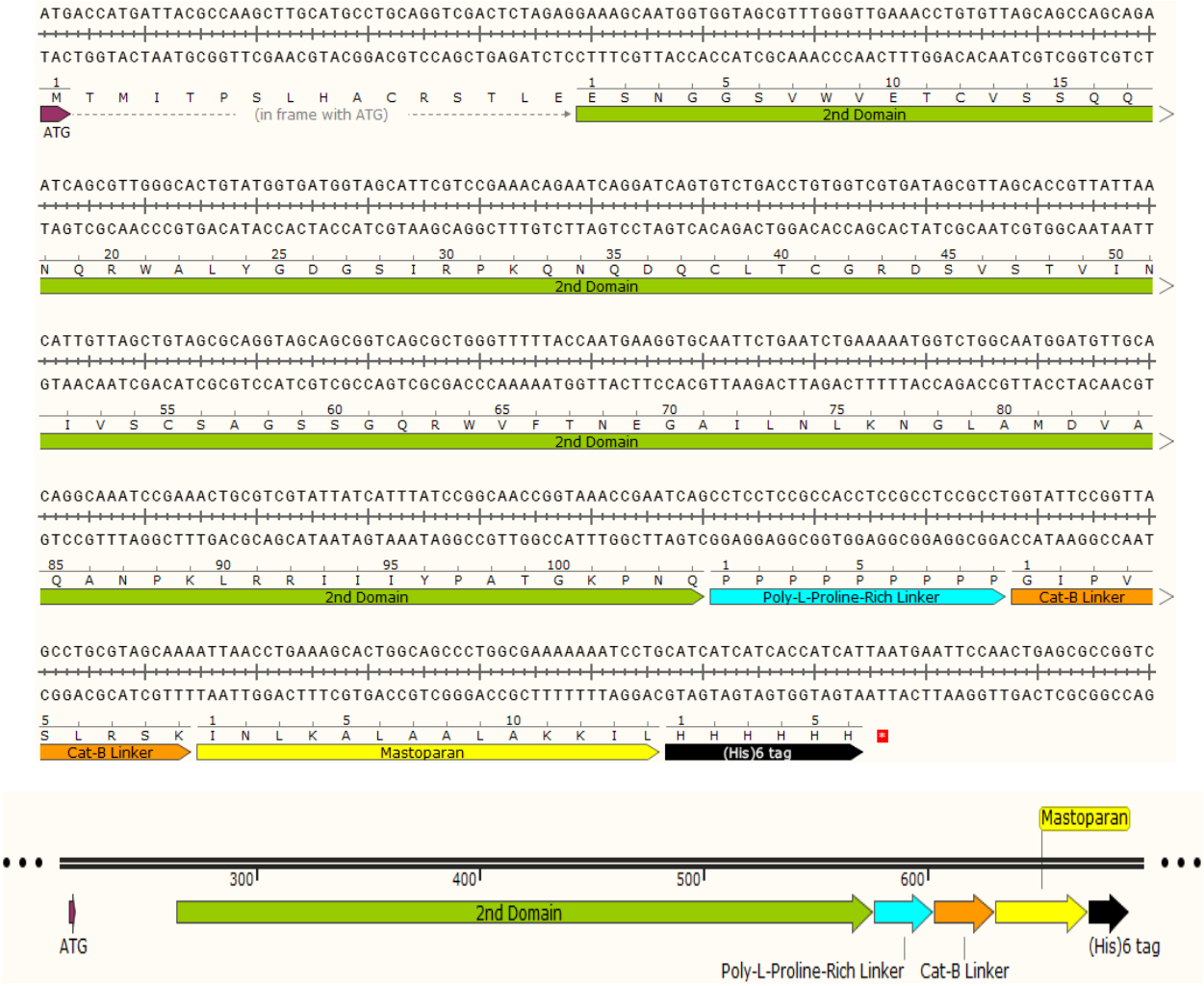


Figure82. Nucleotide sequence, deduced amino acid sequence and schematic representation of the 2nd Domain-Mastoparan fusion protein construct. The gene encoding 2nd Domain of MLB is highlighted in GREEN (bp 265-576). The poly-L-proline-rich linker is highlighted in BLUE (bp 577-603). The cathepsin-B (CatB) biodegradable linker is highlighted in ORANGE (bp 604-630). The gene encoding Mastoparan is highlighted in YELLOW (bp 631-672). The (His)6 tag is highlighted in BLACK (bp 673-690). The stop codon is highlighted in a red star.

After successful transformation of Lemo21(DE3) *E. coli* cells with the pGFPuv expression vector containing Lectin-Toxin fusions, diagnostic restriction digest was performed to the plasmid DNA extracted from overnight grown cultures with appropriate restriction enzymes to confirm the existence of the targeted fusion constructs in the expression vector (Figure 83, 84).

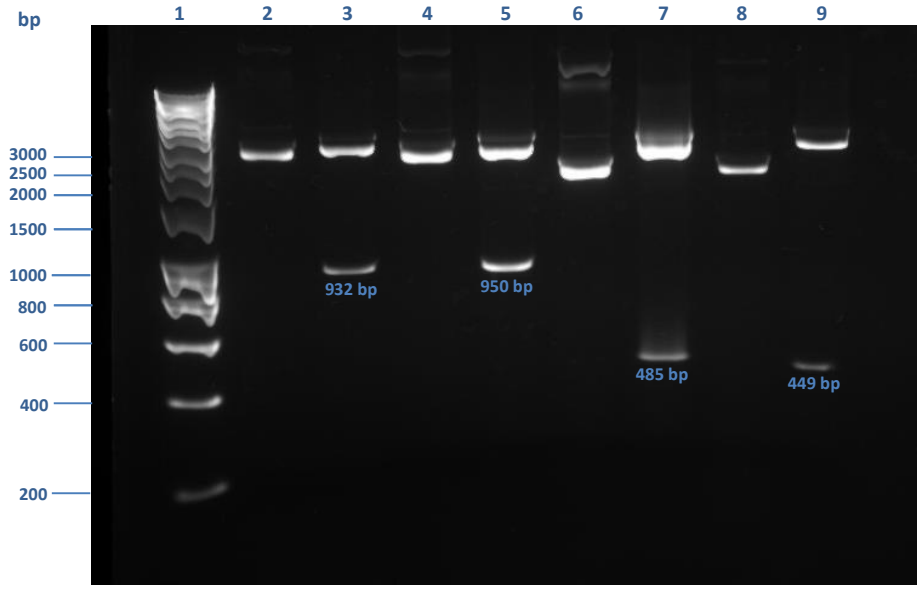


Figure 83. Restriction digest of pGFPuv plasmid to confirm the presence of Lectin-Toxin inserts. Lane 1: HyperLadder I (Bioline, UK); Lanes 3,5,7 and 9: the restriction digests of MLB-Melittin 2, MLB-Melittin 1, 1st Domain-Melittin and 1st Domain-Mastoparan fusion constructs respectively; Lanes 2, 4, 6 and 8: the undigested negative controls.

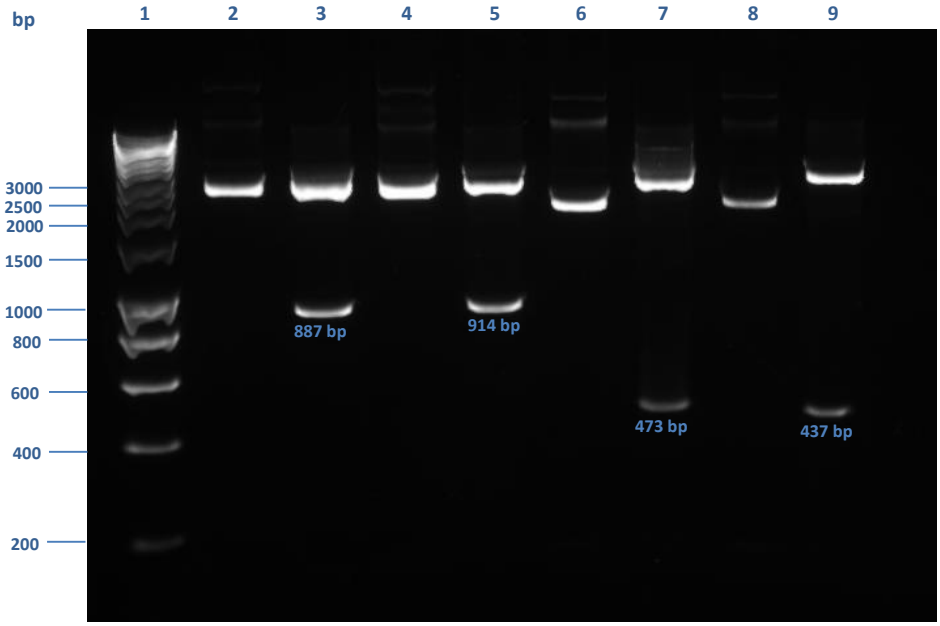


Figure 84. Restriction digest of pGFPuv plasmid to confirm the presence of Lectin-Toxin inserts. Lane 1: HyperLadder I (Bioline, UK); Lanes 3,5,7 and 9: the restriction digests of MLB-Mastoparan 2, MLB-Mastoparan 1, 2nd Domain-Melittin and 2nd Domain-Mastoparan fusion constructs respectively; Lanes 2, 4, 6 and 8: the undigested negative controls.

The sequence of the Lectin-Toxin constructs in the pGFPuv expression vector was confirmed through sequencing reactions. The partial results are shown in figure 85-92 and the full results are shown in appendix 4.

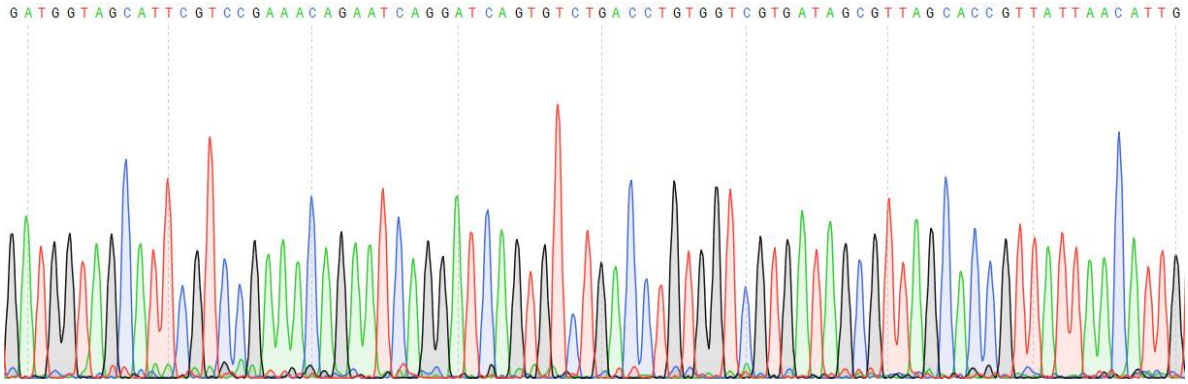


Figure 85. Chromatogram showing the sequence quality of MLB-Melittin 1 fusion construct (Partial sequence data shown). The selected region is representative of overall sequencing quality and the full sequencing result is shown in appendix 4. The Adenine (A), Guanine (G), Cytosine (C) and Thymine (T) codons are in green, black, blue and red colours respectively.

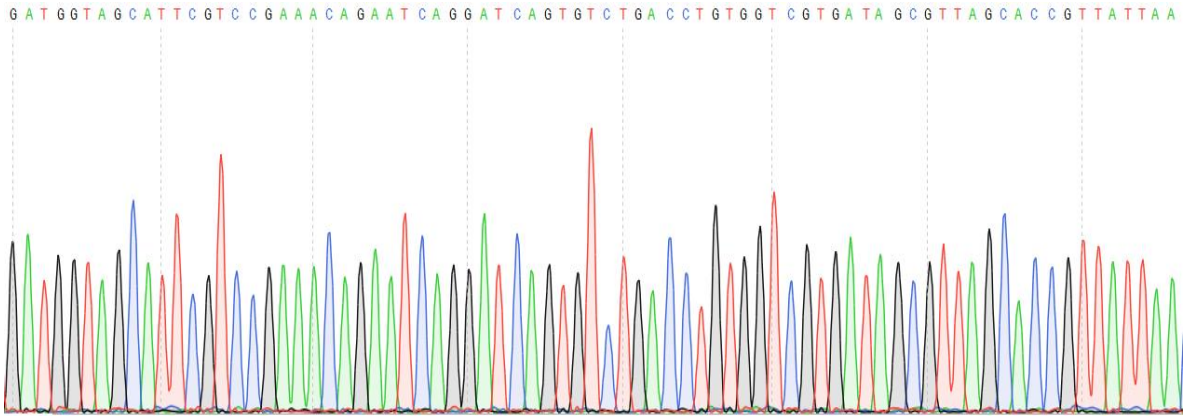


Figure 86. Chromatogram showing the sequence quality of MLB-Melittin 2 fusion construct (Partial sequence data shown). The selected region is representative of overall sequencing quality and the full sequencing result is shown in appendix 4. The Adenine (A), Guanine (G), Cytosine (C) and Thymine (T) codons are in green, black, blue and red colours respectively.

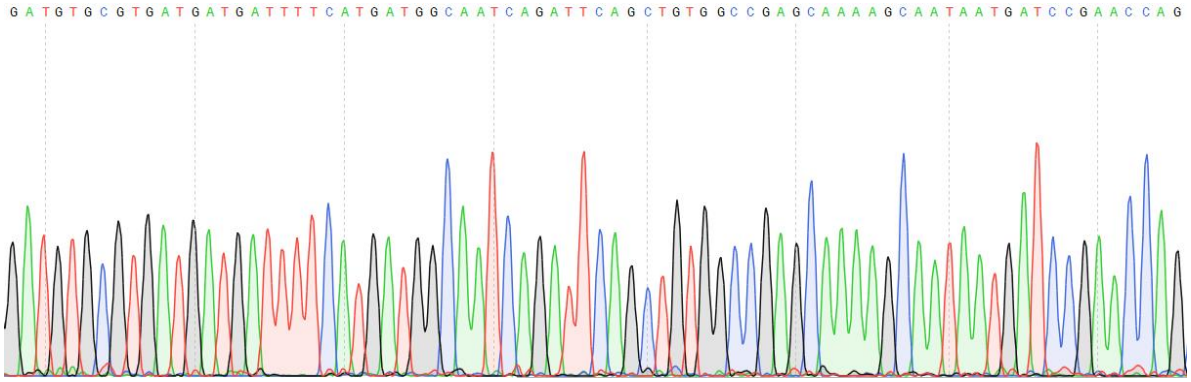


Figure 87. Chromatogram showing the sequence quality of 1st Domain-Melittin fusion construct (Partial sequence data shown). The selected region is representative of overall sequencing quality and the full sequencing result is shown in appendix 4. The Adenine (A), Guanine (G), Cytosine (C) and Thymine (T) codons are in green, black, blue and red colours respectively.

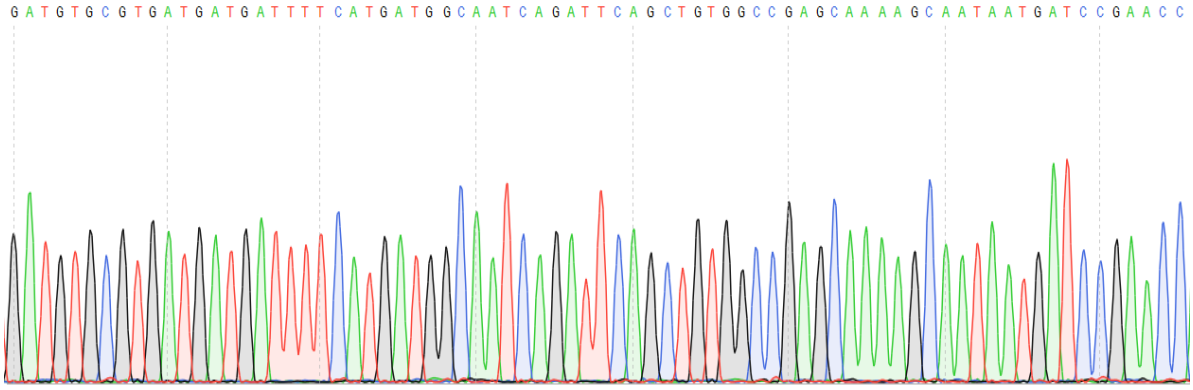


Figure 88. Chromatogram showing the sequence quality of 1st Domain-Mastoparan fusion construct (Partial sequence data shown). The selected region is representative of overall sequencing quality and the full sequencing result is shown in appendix 4. The Adenine (A), Guanine (G), Cytosine (C) and Thymine (T) codons are in green, black, blue and red colours respectively.

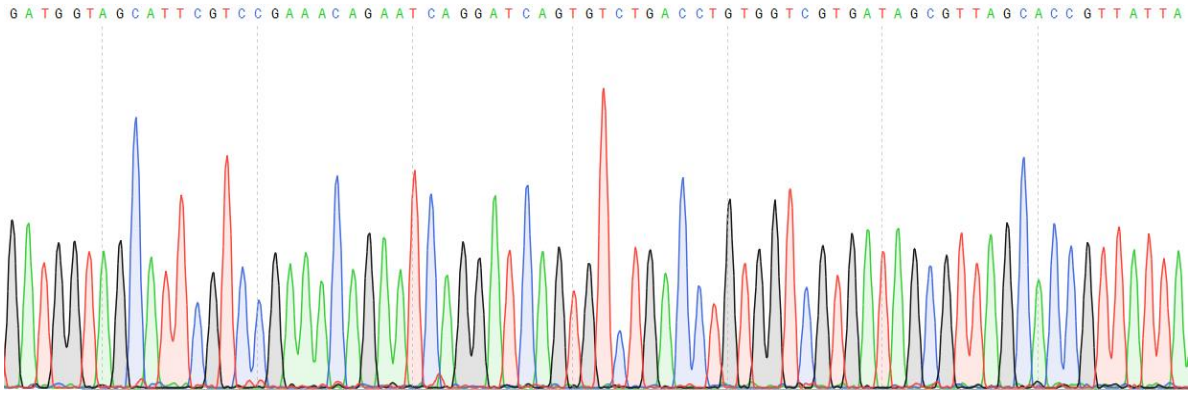


Figure 89. Chromatogram showing the sequence quality of MLB-Mastoparan 1 fusion construct (Partial sequence data shown). The selected region is representative of overall sequencing quality and the full sequencing result is shown in appendix 4. The Adenine (A), Guanine (G), Cytosine (C) and Thymine (T) codons are in green, black, blue and red colours respectively.

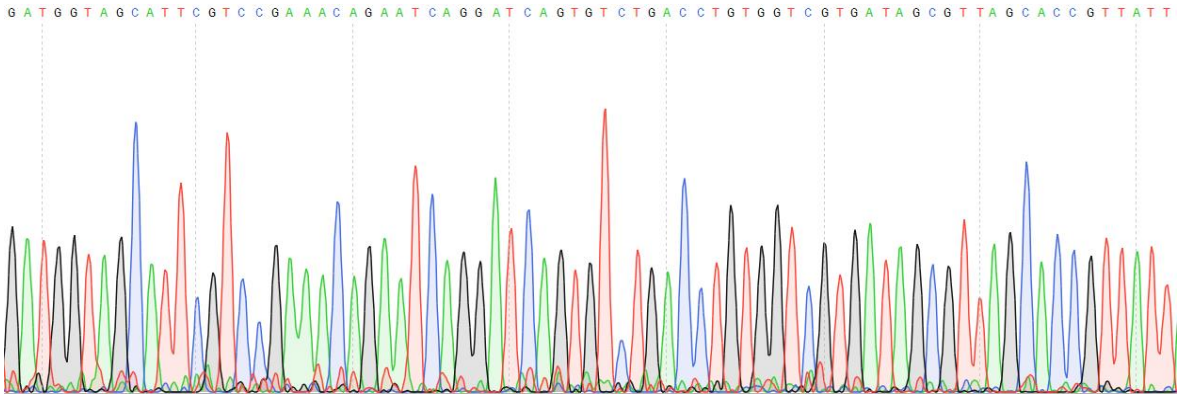


Figure 90. Chromatogram showing the sequence quality of MLB-Mastoparan 2 fusion construct (Partial sequence data shown). The selected region is representative of overall sequencing quality and the full sequencing result is shown in appendix 4. The Adenine (A), Guanine (G), Cytosine (C) and Thymine (T) codons are in green, black, blue and red colours respectively.

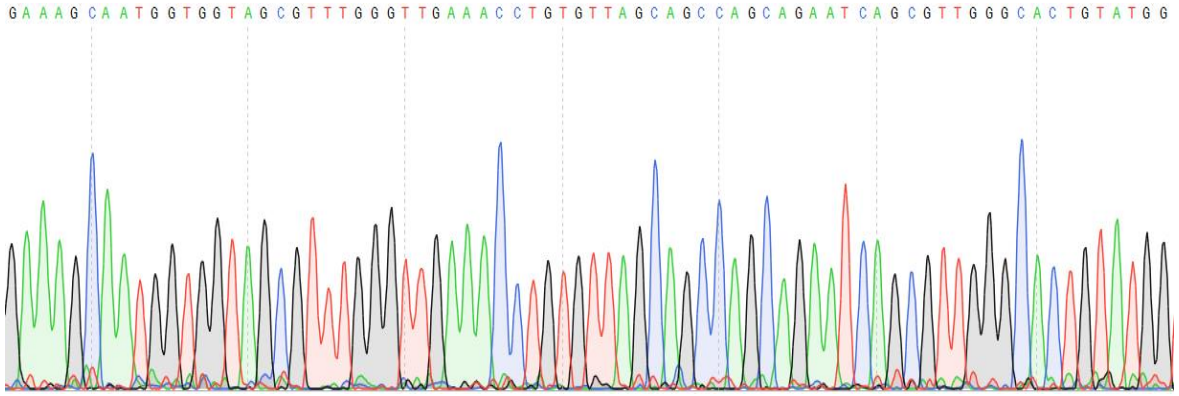


Figure 91. Chromatogram showing the sequence quality of 2nd Domain-Melittin fusion construct (Partial sequence data shown). The selected region is representative of overall sequencing quality and the full sequencing result is shown in appendix 4. The Adenine (A), Guanine (G), Cytosine (C) and Thymine (T) codons are in green, black, blue and red colours respectively.

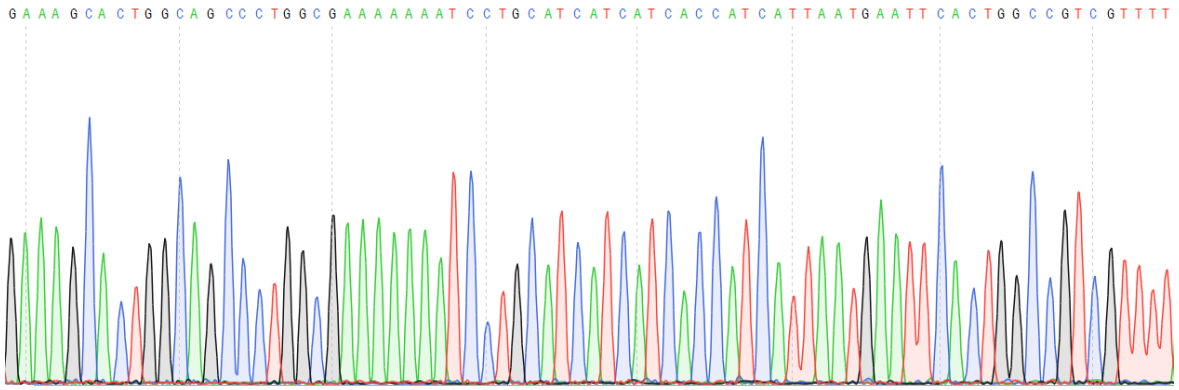


Figure 92. Chromatogram showing the sequence quality of 2nd Domain-Mastoparan fusion construct (Partial sequence data shown). The selected region is representative of overall sequencing quality and the full sequencing result is shown in appendix 4. The Adenine (A), Guanine (G), Cytosine (C) and Thymine (T) codons are in green, black, blue and red colours respectively.

5.3.5 Recombinant Expression of the Fusion Proteins

The confirmation of the correct size and the sequence confirmation of the Lectin-Toxin constructs gave an indication that the constructs were reliable enough to go on for protein expression. To alleviate growth inhibitory effects from the toxic genes in the fusion constructs and formation of inclusion bodies during the expression, Lemo21 (DE3) *E.coli* cells were used in which has tunable T7 polymerase that allows an efficient expression of difficult clones especially toxic and membrane proteins.

All the fusion constructs were expressed successfully and the total harvested proteins from each set of culture was analysed on 12% SDS-PAGE gel and the expected bands of the constructs were detected on the gels. Western blotting analysis was also used to identify clones expressing the fusion proteins using anti-His primary antibodies. The antibody successfully bound to all the fusion protein bands transferred to the nitrocellulose membrane and strong signals were detected from chemiluminescent substrate (Figure 93).

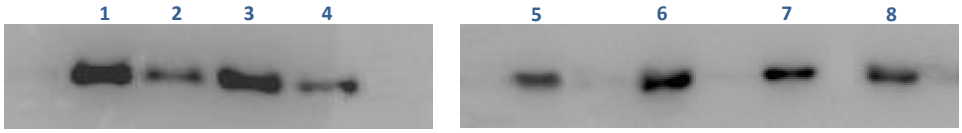


Figure 93. Western blot of Lectin-Toxin fusions proteins using Anti-His(C-term) primary antibody (ThermoFisher, UK) and goat anti-mouse IgG secondary antibody (Sigma, UK). Lane 1: MLB+Melittin 1 fusion; Lane 2: MLB+Melittin 2 fusion; Lane 3: MLB+Mastoparan 1 fusion; Lane 4: MLB+Mastoparan 2 fusion; Lane 5: 1st Domain-Melittin fusion; Lane 6: 1st Domain-Mastoparan fusion; Lane 7: 2nd Domain-Melittin fusion; Lane 8: 2nd Domain-Mastoparan fusion.

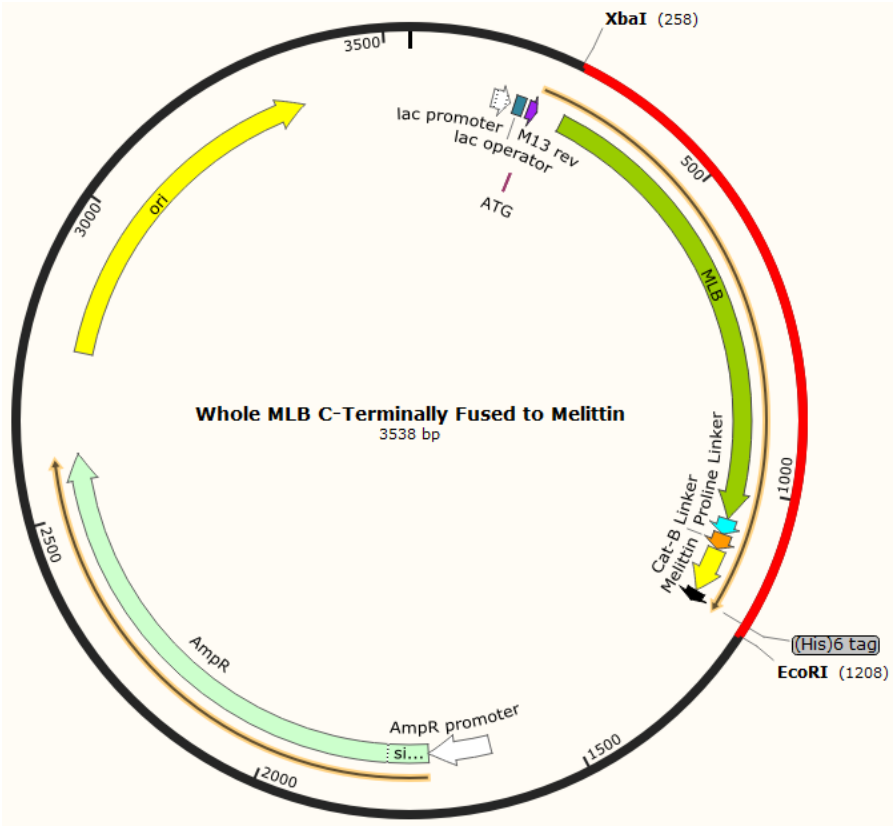


Figure 94. Cloned MLB chain *C*-terminally fused to melittin toxin peptide through cathepsin B biodegradable linker (CatB) in the pGFPuv expression vector. The green, orange and yellow parts of the first arrow represent MLB, CatB and melittin sequences respectively; The light green and yellow arrows represent the ampicillin resistant (AmpR) and signal of origin of replication (ori) sequences of pGFPuv vector.

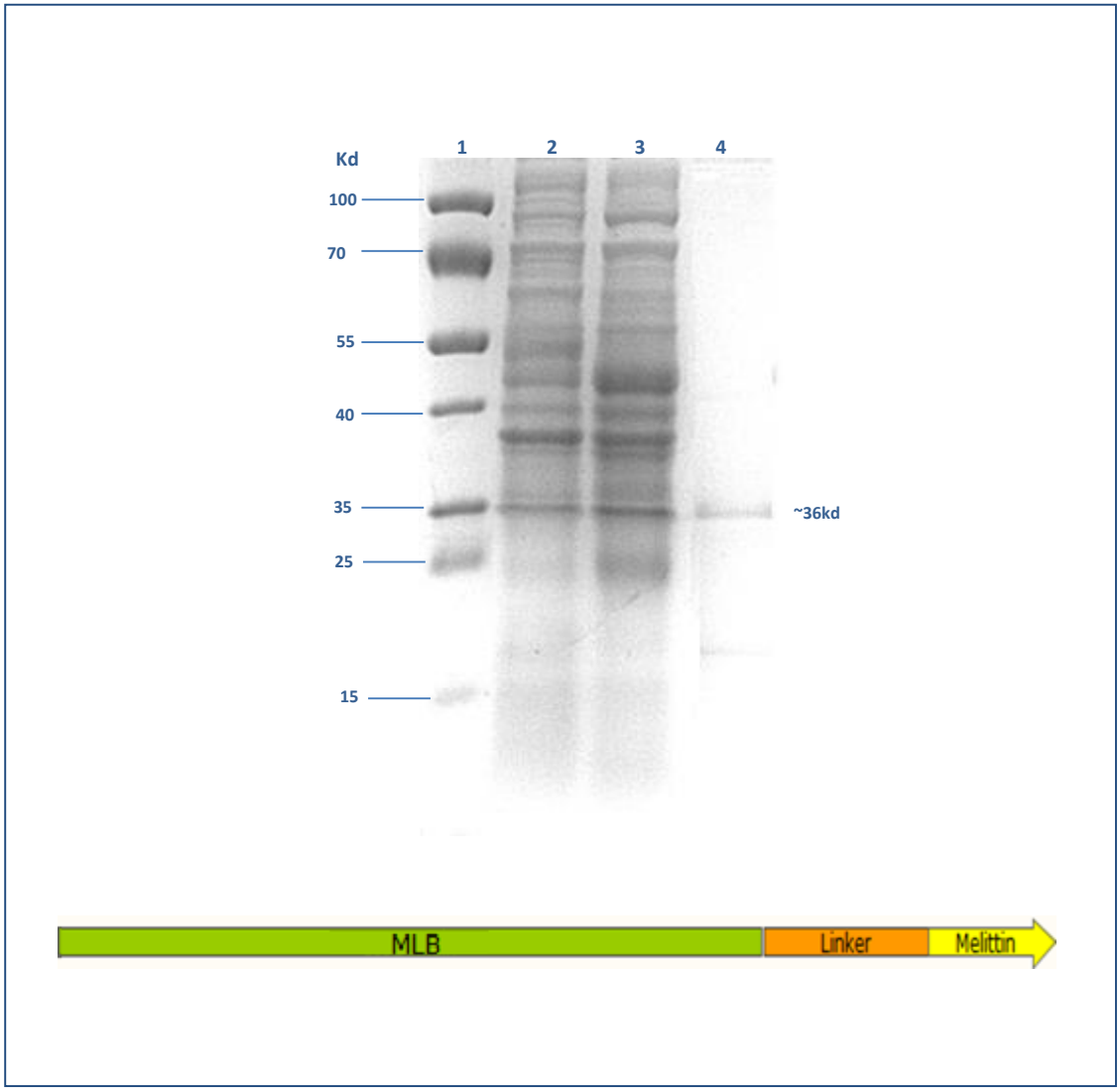


Figure 95. SDS PAGE stained with InstantBlue (Expedeon,UK) of expressed MLB fused to melittin through cathepsin B linker. Lane 1: Protein marker; Lane 2: Uninduced MLB-Melittin 1 fusion; Lane 3: Induced MLB-Melittin 1 fusion; Lane 4: Purified MLB-Melittin 1 fusion (~36kDa).

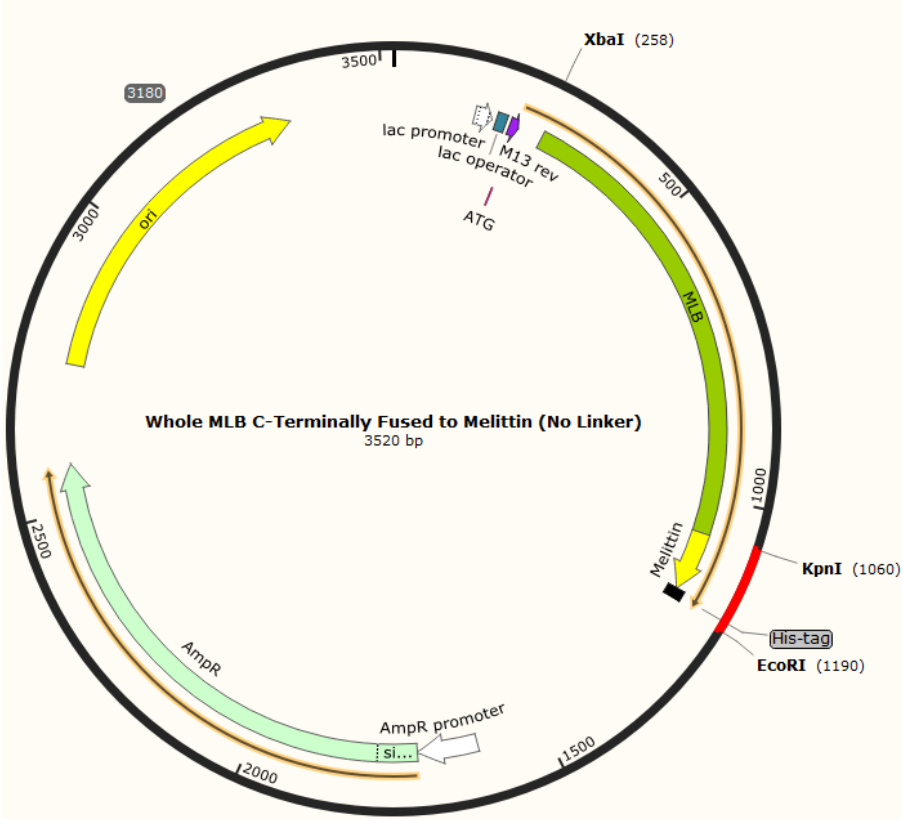


Figure 96. Cloned MLB chain *C*-terminally fused to melittin toxin peptide without cathepsin B biodegradable linker (CatB) in the pGFPuv expression vector. The green and yellow parts of the first arrow represent MLB and melittin sequences respectively; The light green and yellow arrows represent the ampicillin resistant (AmpR) and signal of origin of replication (ori) sequences of pGFPuv vector.

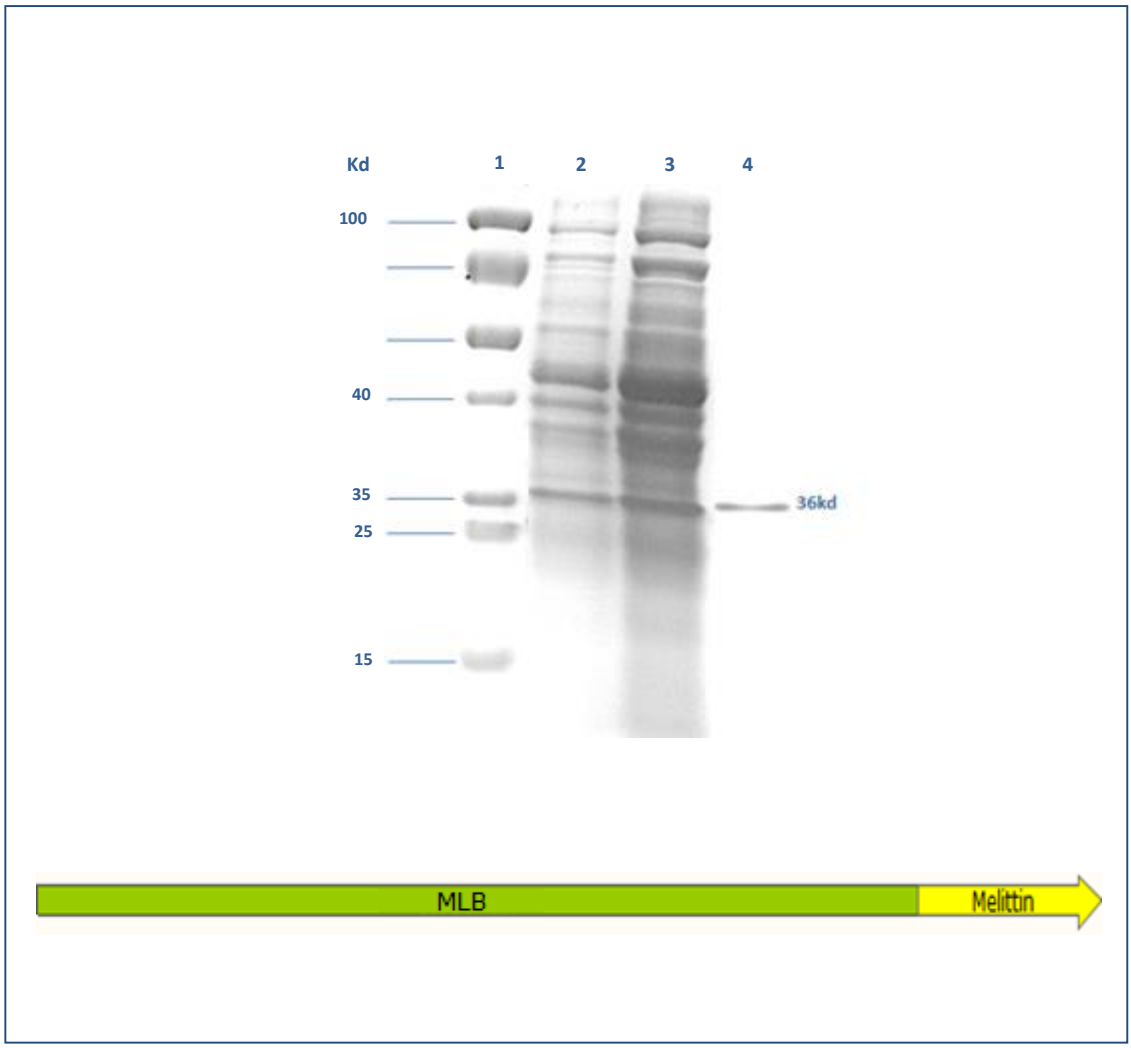
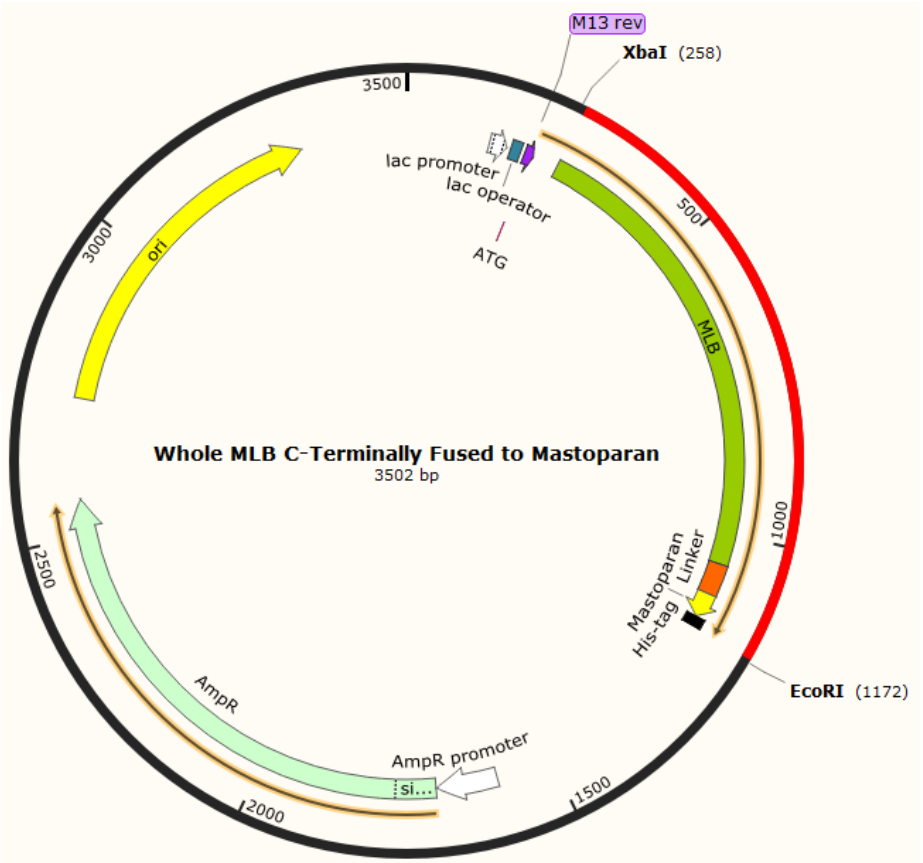


Figure 97. SDS PAGE stained with InstantBlue (Expedeon,UK) of expressed MLB fused to melittin without cathepsin B linker. Lane 1: Protein marker; Lane 2: Uninduced MLB-Melittin 2 fusion; Lane 3: Induced MLB-Melittin 2 fusion; Lane 4: Purified MLB-Melittin 2 fusion (~36kDa).



[Figure 98](). Cloned MLB chain *C*-terminally fused to mastoparan toxin peptide through cathepsin B biodegradable linker (CatB) in the pGFPuv expression vector. The green, orange and yellow parts of the first arrow represent MLB, CatB and mastoparan sequences respectively; The light green and yellow arrows represent the ampicillin resistant (AmpR) and signal of origin of replication (ori) sequences of pGFPuv vector.

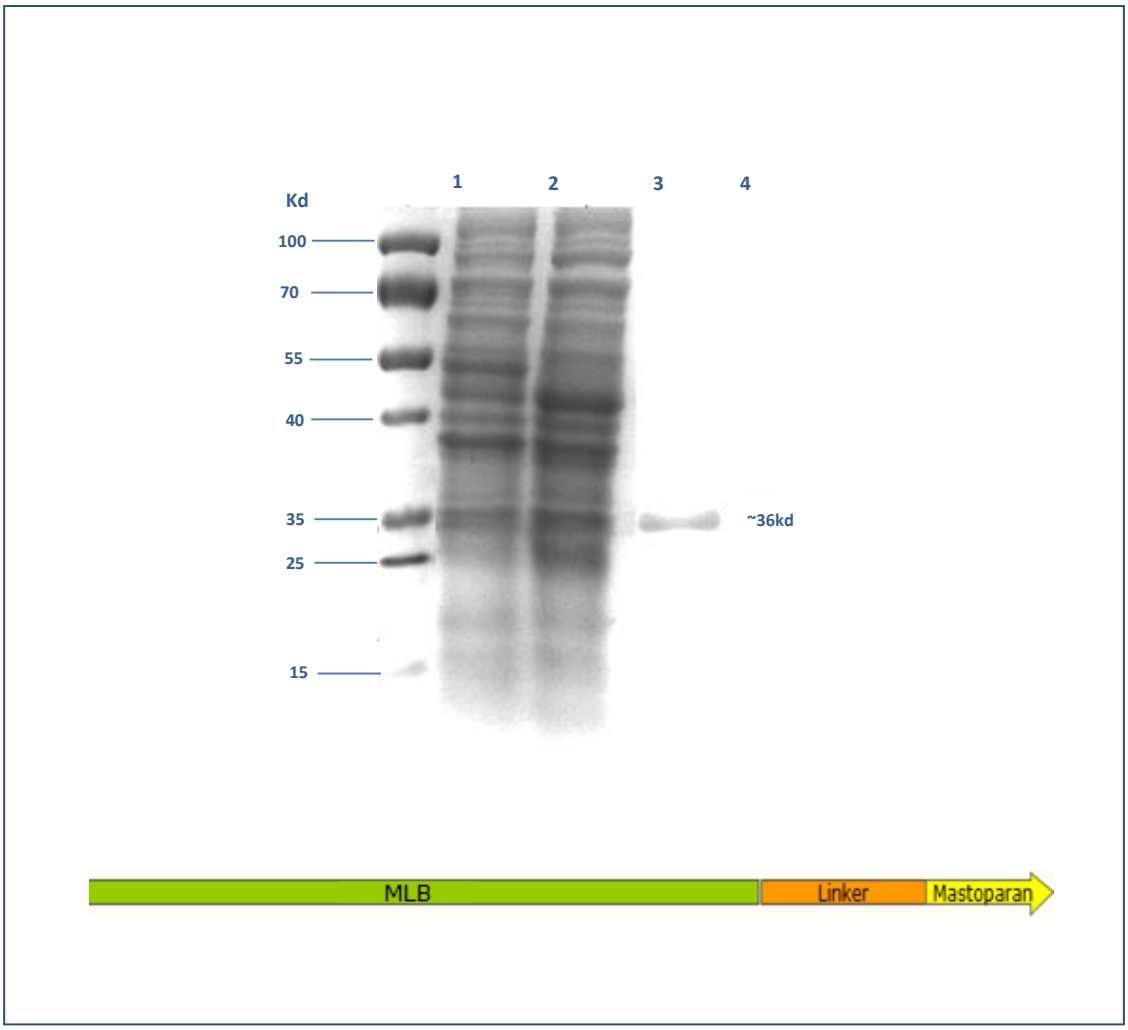


Figure 99. SDS PAGE stained with InstantBlue (Expedeon,UK) of expressed MLB fused to mastoparan through cathepsin B linker. Lane 1: Protein marker; Lane 2: Uninduced MLB-Mastoparan 1 fusion; Lane 3: Induced MLB-Mastoparan 1 fusion; Lane 4: Purified MLB-Mastoparan 1 fusion (~35kDa).

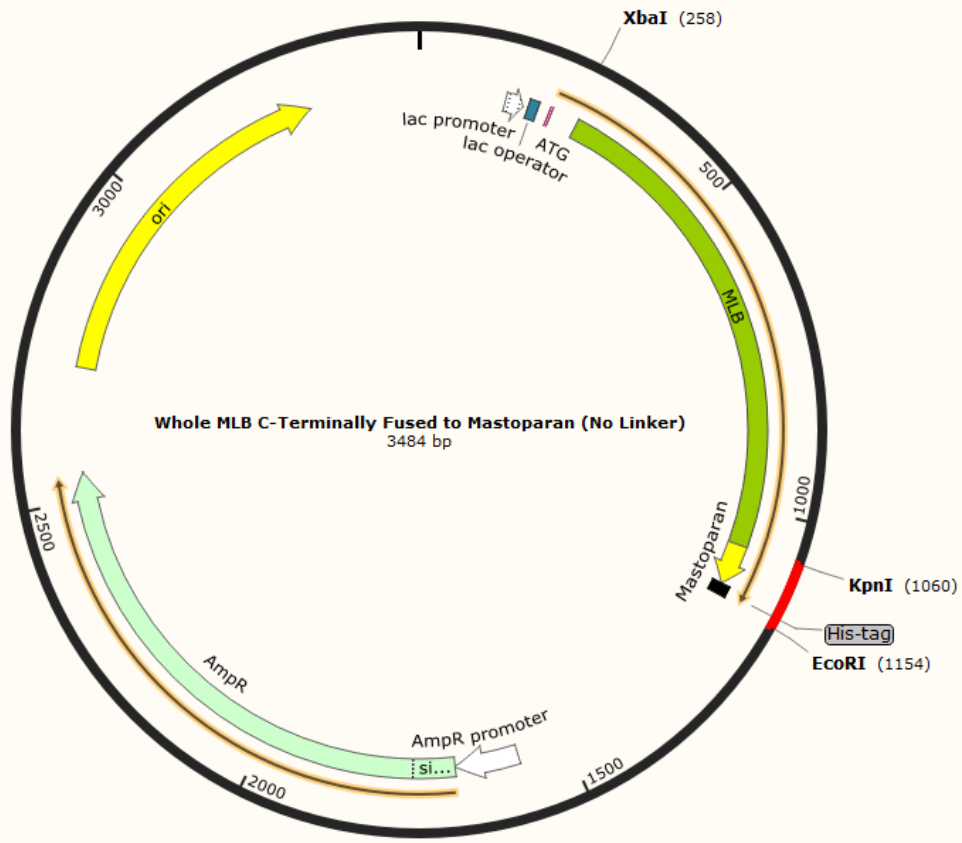


Figure 100. Cloned MLB chain *C*-terminally fused to mastoparan toxin peptide without cathepsin B biodegradable linker (CatB) in the pGFPuv expression vector. The green and yellow parts of the first arrow represent MLB and mastoparan sequences respectively; The light green and yellow arrows represent the ampicillin resistant (AmpR) and signal of origin of replication (ori) sequences of pGFPuv vector.

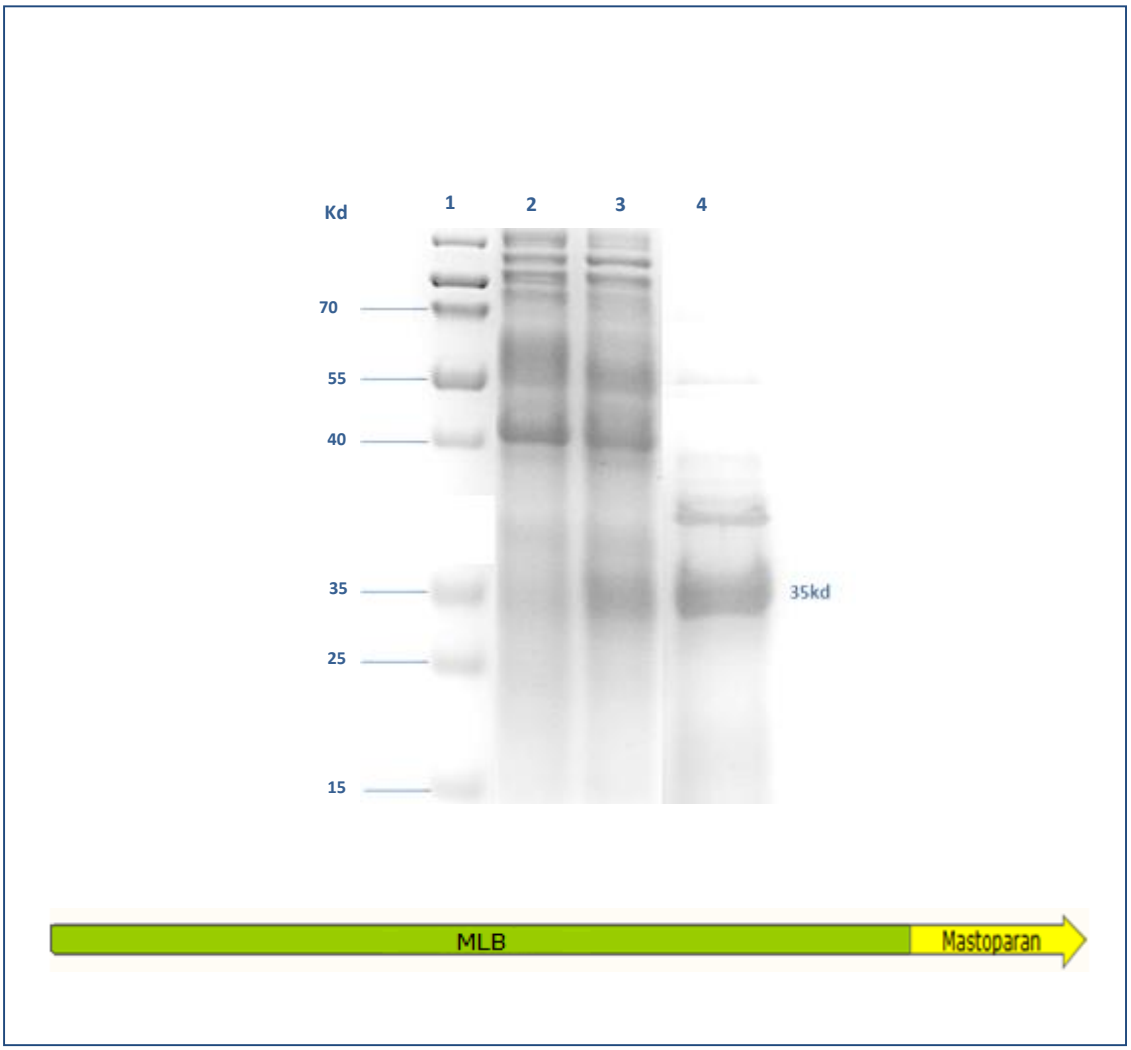


Figure 101. SDS PAGE stained with InstantBlue (Expedeon,UK) of expressed MLB fused to mastoparan without cathepsin B linker. Lane 1: Protein marker; Lane 2: Uninduced MLB-Mastoparan 2 fusion; Lane 3: Induced MLB-Mastoparan 2 fusion; Lane 4: Purified MLB-Mastoparan 2 fusion (~35kDa).

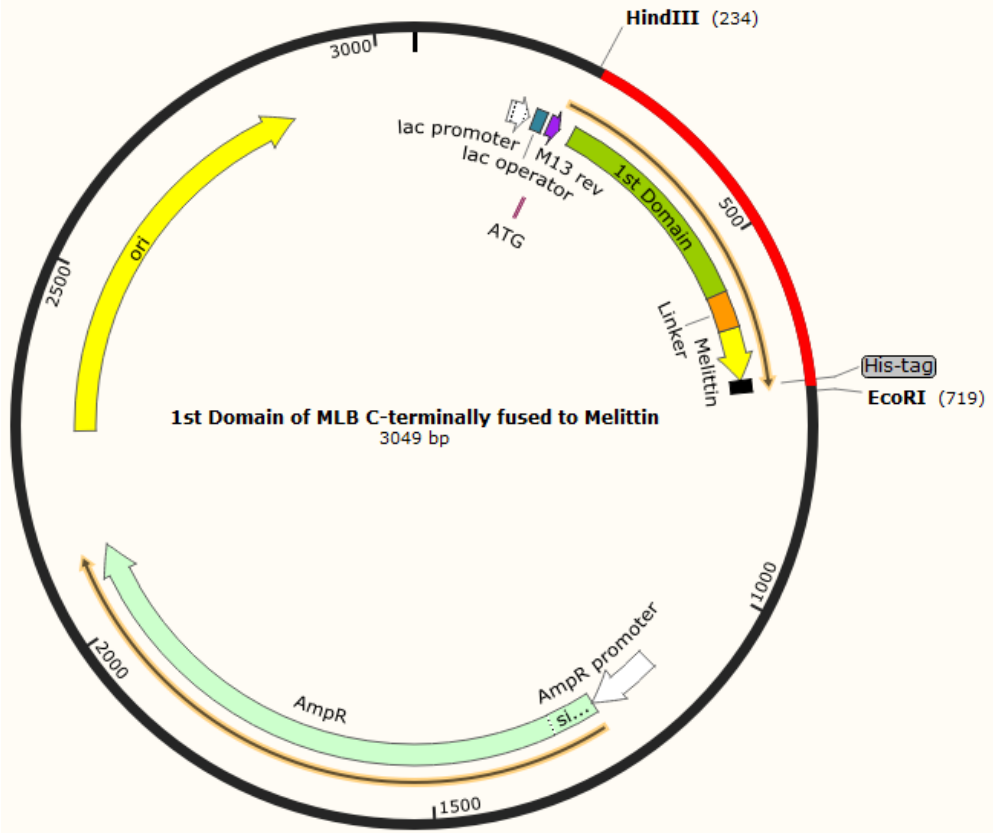


Figure 102. Cloned 1st domain of MLB chain *C*-terminally fused to melittin toxin peptide through cathepsin B biodegradable linker (CatB) in the pGFPuv expression vector. The green, orange and yellow parts of the first arrow represent MLB, CatB and melittin sequences respectively; The light green and yellow arrows represent the ampicillin resistant (AmpR) and signal of origin of replication (ori) sequences of pGFPuv vector.

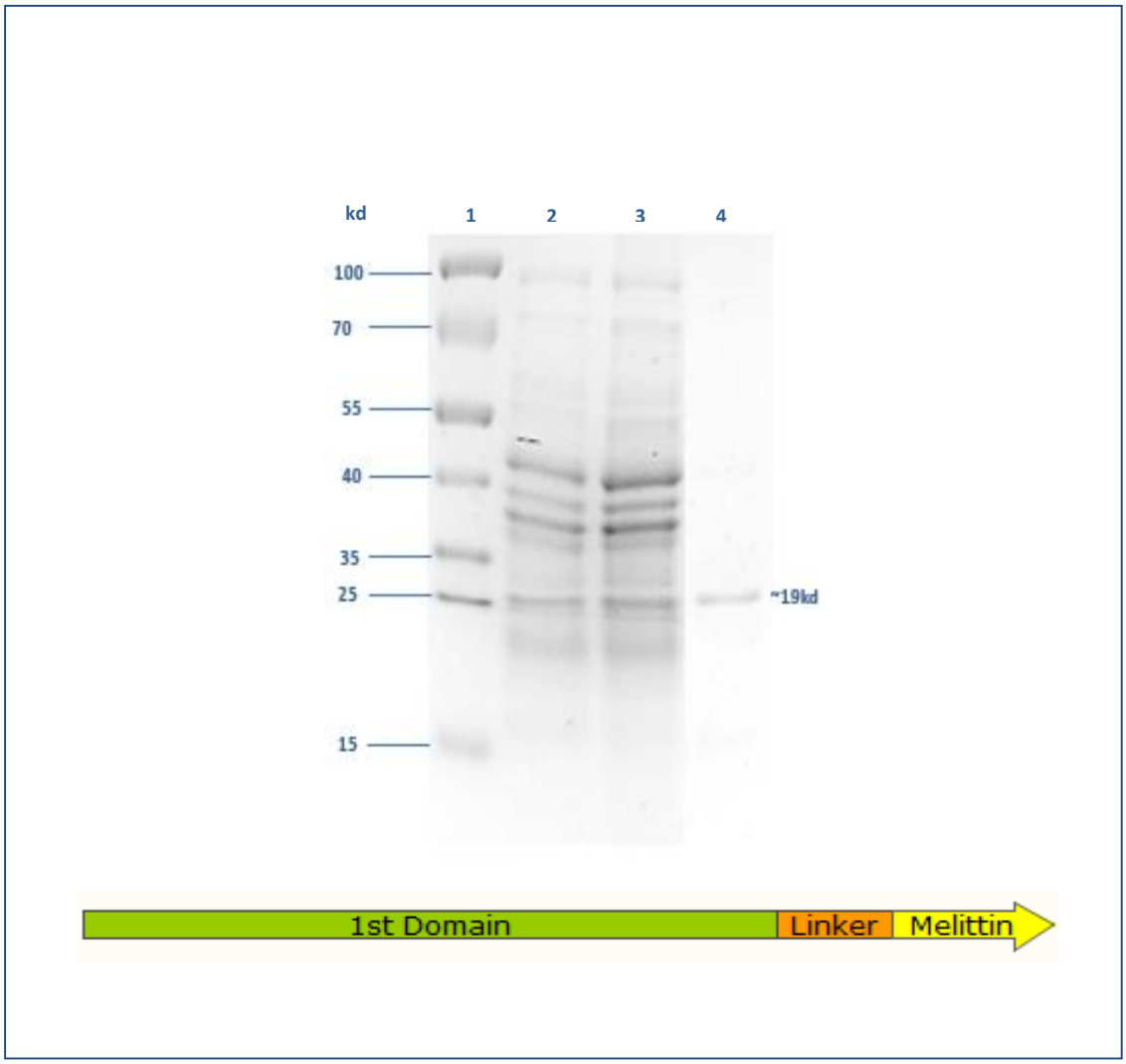


Figure 103. SDS PAGE stained with InstantBlue (Expedeon,UK) of expressed 1st domain of MLB fused to melittin through cathepsin B linker. Lane 1: Protein marker; Lane 2: Uninduced 1st Domain-Melittin fusion; Lane 3: Induced 1st Domain-Melittin fusion; Lane 4: Purified 1st Domain-Melittin fusion (~19kDa).

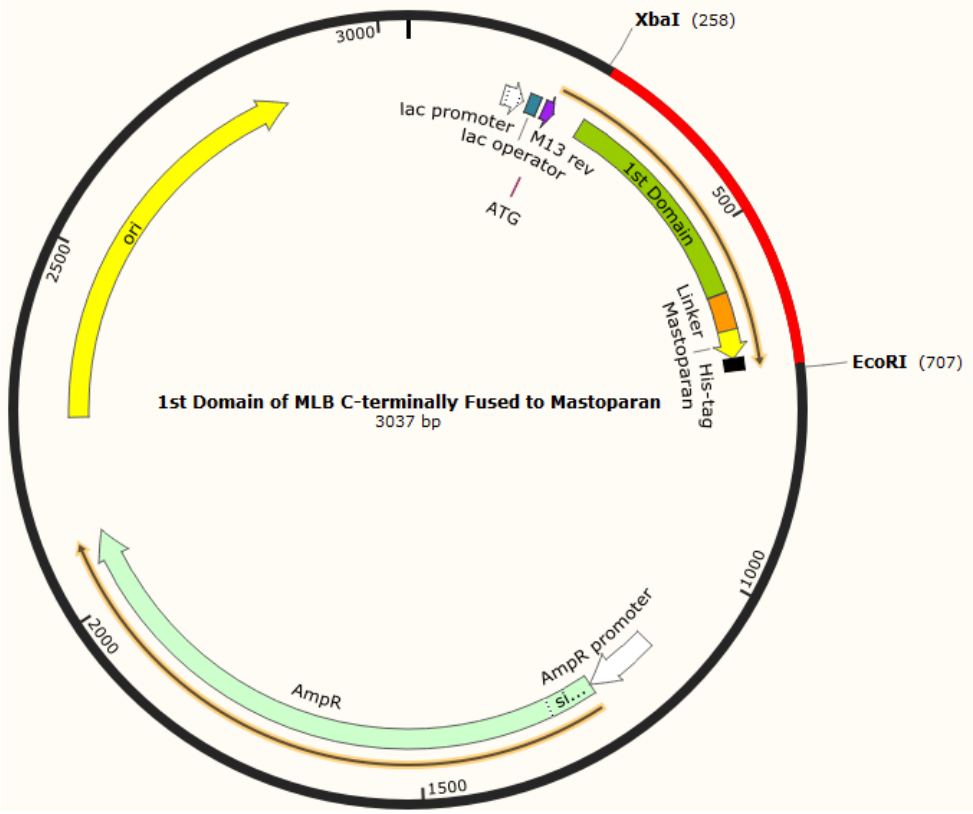


Figure 104. Cloned 1st domain of MLB chain *C*-terminally fused to mastoparan toxin peptide without cathepsin B biodegradable linker (CatB) in the pGFPuv expression vector. The green and yellow parts of the first arrow represent MLB and mastoparan sequences respectively; The light green and yellow arrows represent the ampicillin resistant (AmpR) and signal of origin of replication (ori) sequences of pGFPuv vector.

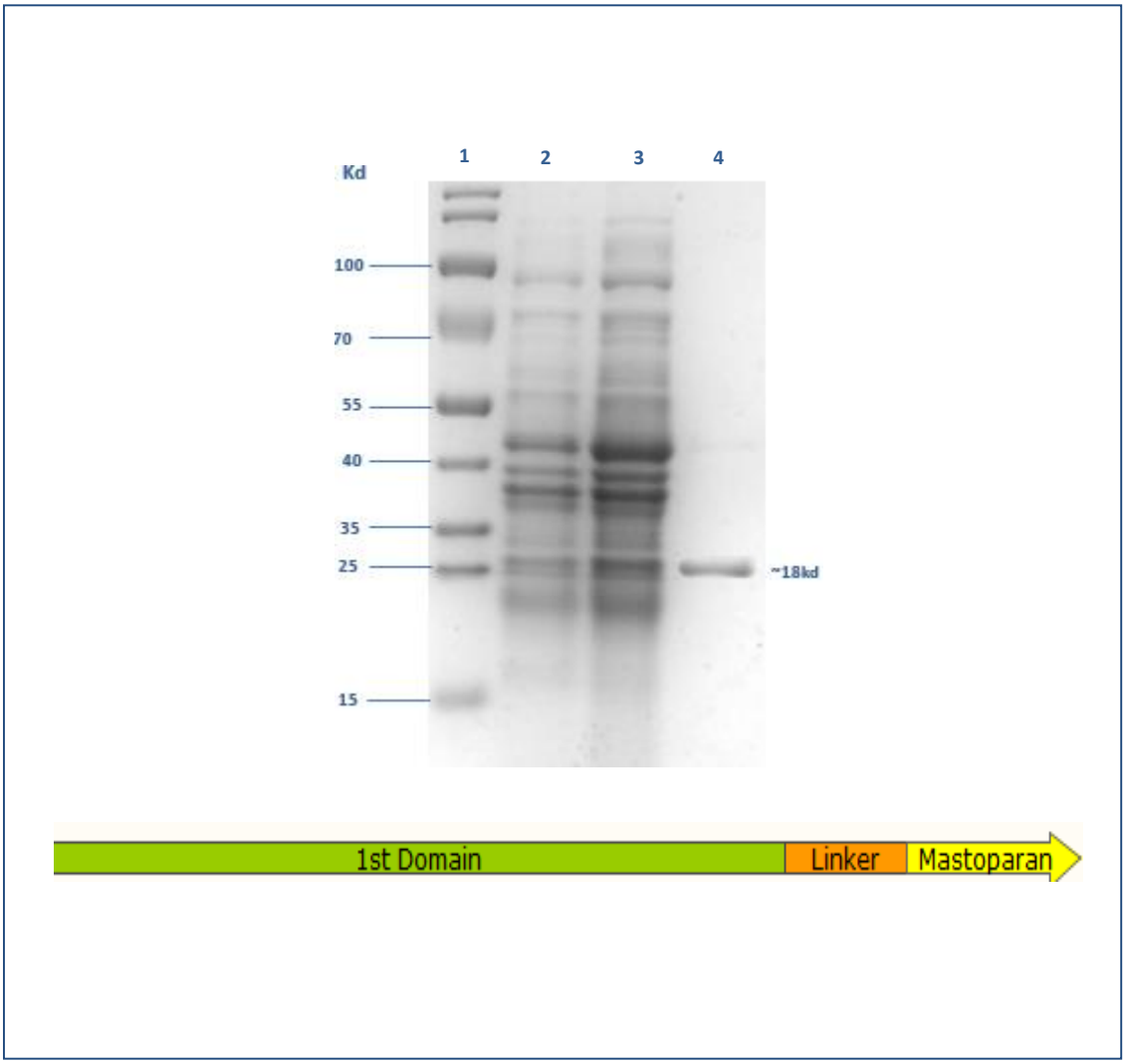


Figure 105. SDS PAGE stained with InstantBlue (Expedeon,UK) of expressed 1st domain of MLB fused to mastoparan through cathepsin B linker. Lane 1: Protein marker; Lane 2: Uninduced 1st Domain-Mastoparan fusion; Lane 3: Induced 1st Domain-Mastoparan fusion; Lane 4: Purified 1st Domain-Mastoparan fusion (~18kDa).

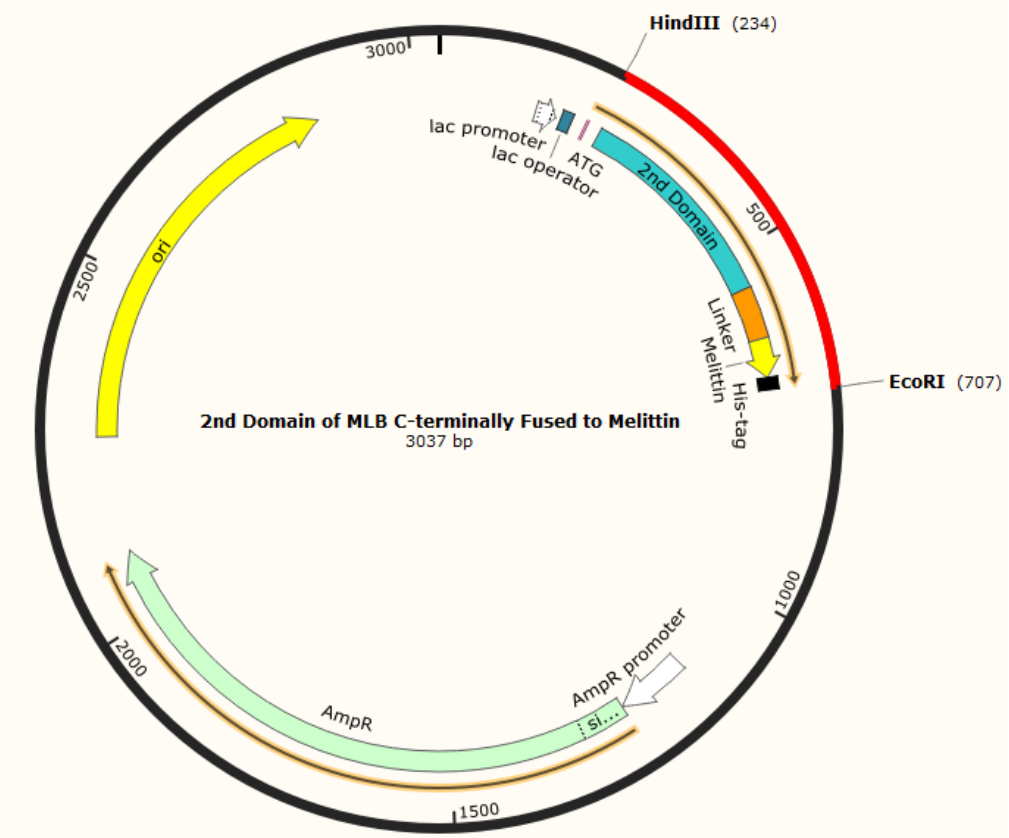


Figure 106. Cloned 2nd domain of MLB chain *C*-terminally fused to melittin toxin peptide through cathepsin B biodegradable linker (CatB) in the pGFPuv expression vector. The green, orange and yellow parts of the first arrow represent MLB, CatB and melittin sequences respectively; The light green and yellow arrows represent the ampicillin resistant (AmpR) and signal of origin of replication (ori) sequences of pGFPuv vector.

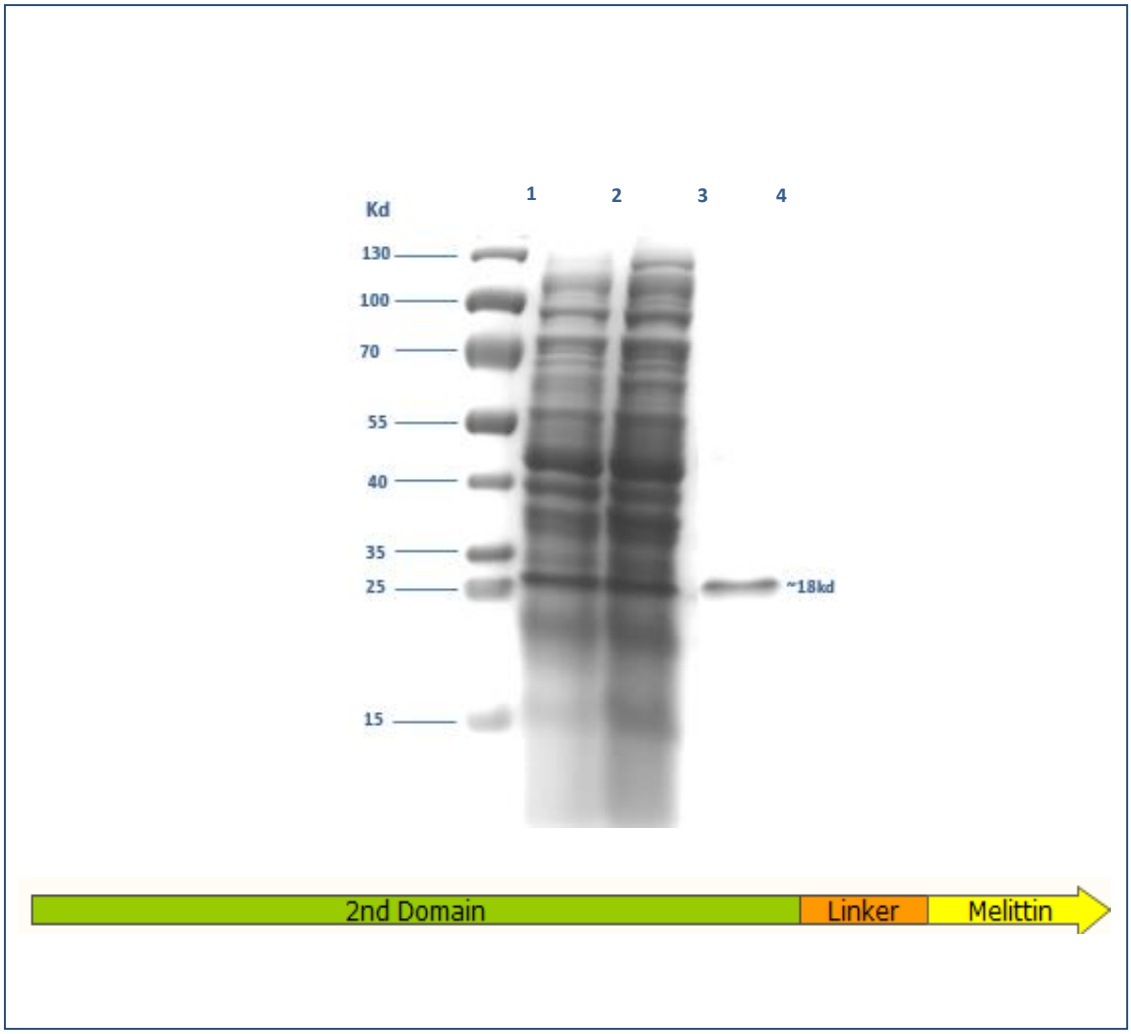


Figure 107. SDS PAGE stained with InstantBlue (Expedeon,UK) of expressed 2nd domain of MLB fused to melittin through cathepsin B linker. Lane 1: Protein marker; Lane 2: Uninduced 2nd Domain-Melittin fusion; Lane 3: Induced 2nd Domain-Melittin fusion; Lane 4: Purified 2nd Domain-Melittin fusion (~18kDa).

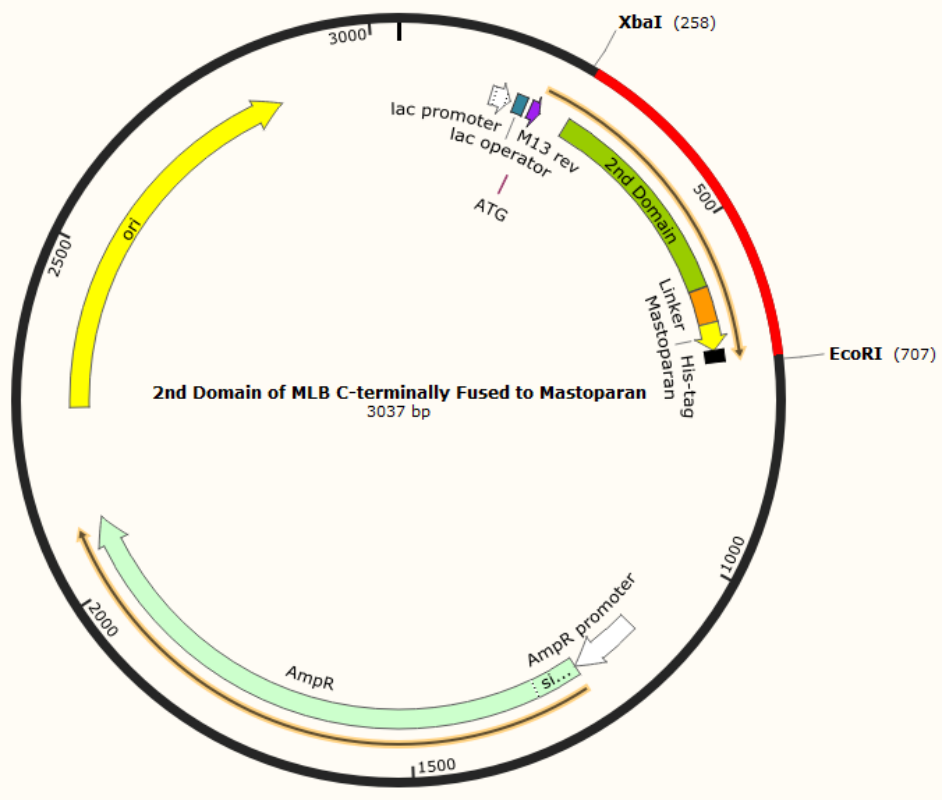


Figure 108. Cloned 2nd domain of MLB chain *C*-terminally fused to mastoparan toxin peptide without cathepsin B biodegradable linker (CatB) in the pGFPuv expression vector. The green and yellow parts of the first arrow represent MLB and mastoparan sequences respectively; The light green and yellow arrows represent the ampicillin resistant (AmpR) and signal of origin of replication (ori) sequences of pGFPuv vector.

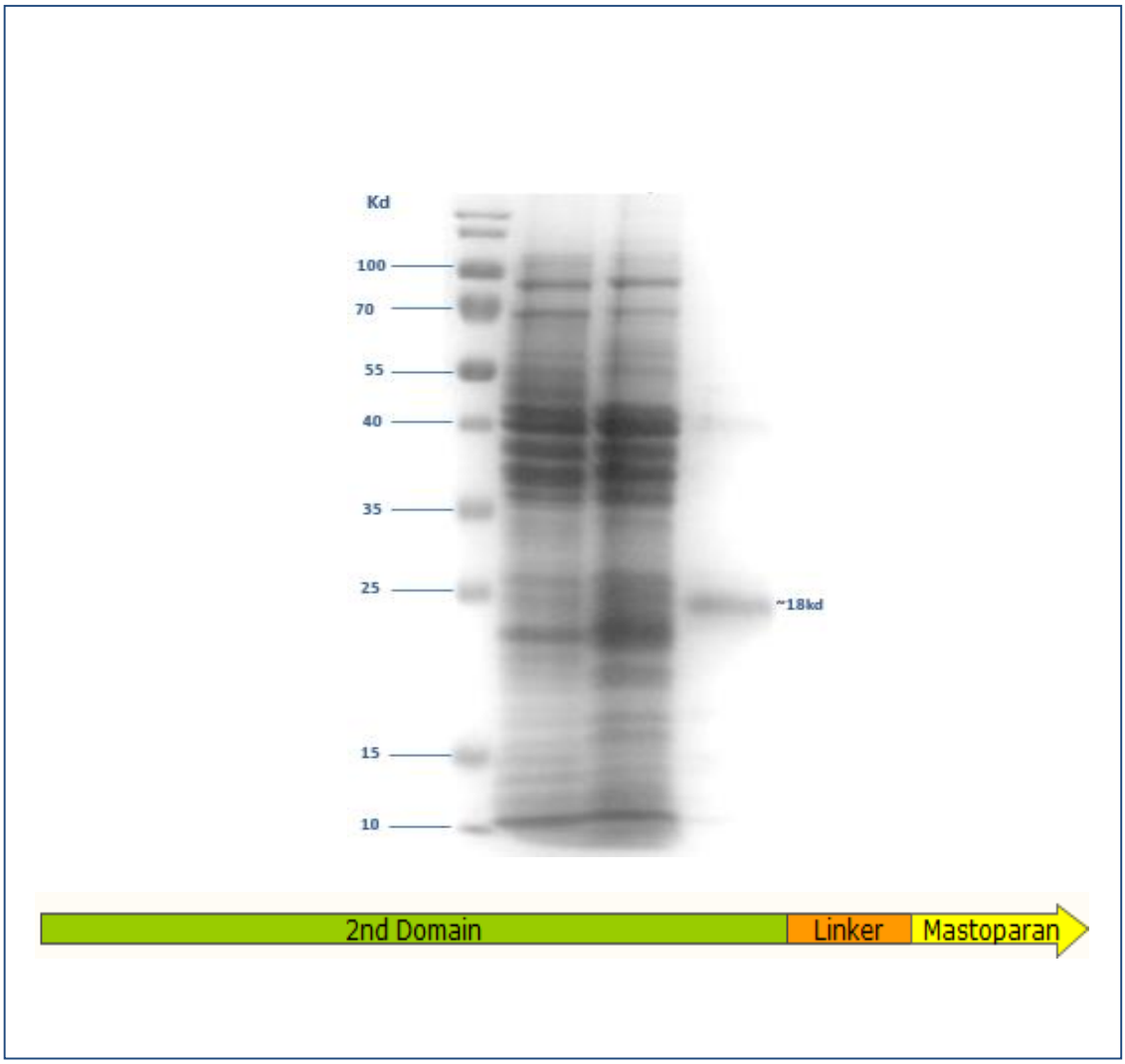


Figure 109. SDS PAGE stained with InstantBlue (Expedeon,UK) of expressed 2nd domain of MLB fused to mastoparan through cathepsin B linker. Lane 1: Protein marker; Lane 2: Uninduced 2nd Domain-Mastoparan fusion; Lane 3: Induced 2nd Domain-Mastoparan fusion; Lane 4: Purified 2nd Domain-Mastoparan fusion (~18kDa).

All the recombinant fusion proteins were purified from culture supernatant by immobilized nickel ion affinity chromatography utilising the added $(His)_6$ tag. A typical elution profile of purifications by nickel affinity chromatography is shown in figures 110.

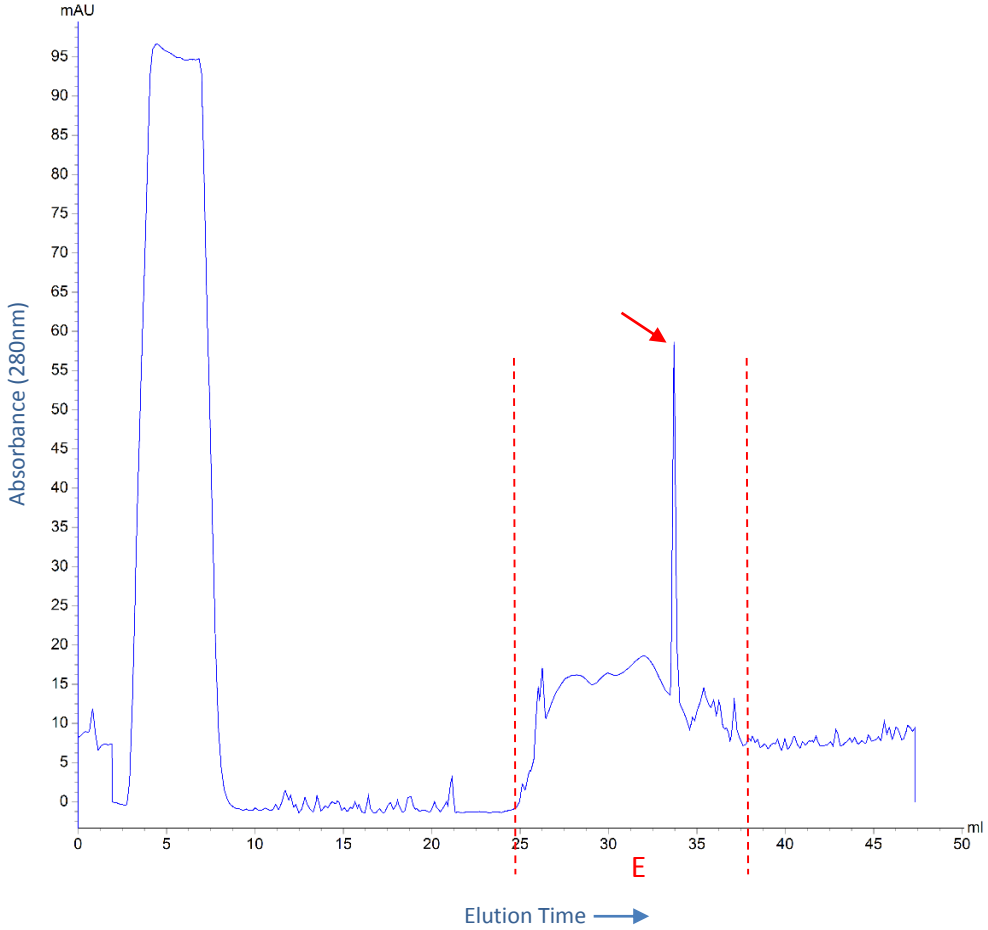


Figure 110. Typical purification profile of nickel affinity column purification of the fusion proteins tagged with (His)6 tag. The gradient elution fractions are (E) (~13ml) and the purified MLB-Melittin 1 fraction is pointed with a red arrow.

From the amino acid sequences, the predicted sizes of MLB-Melittin 1, MLB-Melittin 2, MLB-Mastoparan 1, MLB-Mastoparan 2, 1st Domain-Melittin, 1st Domain-Mastoparan, 2nd Domain-Melittin and 2nd Domain-Mastoparan are 36.0kDa, 35.3kDa, 34.7kDa, 34.0kDa, 18.3kDa, 17.8kDa, 17.6kDa and 17.3kDa respectively. The protein expression was successfully performed and the presence of all the fusion constructs was confirmed by sodium dodecyl sulphate gel electrophoresis (SDS-PAGE) and western blotting using anti-His primary antibodies

Thus, we demonstrate recombinant expression of cytotoxic peptides fused to full length MLB or 1st and 2nd domain truncations of MLB.

5.4 Discussion

This study presents the experiments carried out with the aim of producing fusion conjugates using the binding chain of mistletoe lectin (MLB) and the toxic peptides melittin and mastoparan.

Following confirming the ability of MLB+GFP fusion protein to successfully bind to human melanoma cancer cells, further experiments were carried out to exploit this ability to deliver different cytotoxic peptides specifically to cancer cells. Since MLB has a long coding sequence of 789bp which is not a desired feature in peptides to be used in drug design, the sequence was further analysed through bioinformatics to identify any possible truncations that could be performed (Torchilin 2008). The analysis revealed that the sequence comprised of two very similar putative sugar binding domains with Q-X-W motif and they are both very similar to the domains of the binding site of ricin lectin (Figure 69). Therefore, a range of Lectin-Toxin conjugates were designed not only from the whole MLB but also from the putative sugar binding domains separately because it was believed that one of the domains might be responsible for the sialic acid binding activity and could be as active as the whole MLB in binding to cancer cells.

In order to efficiently express the fusion proteins in *E. coli* and avoid the difficulties of heterologous gene expression, the sequences of MLB, melittin and mastoparan were optimized to *E. coli* tRNA pools through optimizing majority of their nucleotides and better nucleic acid consistency was achieved (Figure 72, 73, 74) which potentially improved the expression of the recombinant constructs and was previously suggested by Nørholm, Toddo et al. (2013) that high-level of expression can be achieved for membrane proteins expressed in *E. coli* through codon optimization. The codon optimization thus helped to successfully express soluble fractions of the fusion proteins, despite a high level

of toxicity of the cytotoxic peptides in the fusion constructs which was previously shown to cause insolubility in the production of fusion proteins (Peng, Hong et al. 2012, Buhrman, Rayahin et al. 2013).

Despite the demonstrated role of mistletoe lectin in the Lectin-Toxin conjugates as a recognition molecule to specifically recognize cancer cells and avoid healthy cells, further specificity of the conjugates toward cancer cells was attempted through placing a biodegradable linker between the binding chains and the cytotoxic peptides. Cathepsine-B (Cat-B) linker was thus chosen to link lectin peptides with the toxic peptides because this linker can be cleaved by cathepsin-B enzyme which is overexpressed in human melanoma cancer cells especially metastatic melanoma according to Matarrese, Ascione et al. (2010). In addition of its main function as a biodegradable linker, Cat-B also keeps melittin and mastoparan attached to MLB in inactive form until they reach the target site by which reduces the side effects of the toxic peptides since the active form of melittin and mastoparan was shown by Huang, Jin et al. (2013) and Kim, Son et al. (2016) respectively to cause strong haemolysis in the body and to avoid this side effect as suggested by Moreno and Giralt (2015) to utilise toxic peptides in recombinant fusion proteins.

Therefore, two sets of recombinant fusion proteins were produced. In the first set, MLB-Melittin 1 and MLB-Mastoparan 1 fusion proteins (Figure 75, 77) were produced based on the whole sequence of MLB fused to melittin and mastoparan respectively through a Cat-B linker. Also, two more fusion proteins (MLB-Melittin 2 and MLB-Mastoparan 2) similar to the previous ones were designed but without Cat-B linker (Figure 76, 78) so as to evaluate the role of Cat-B linker in the conjugates.

The second set of recombinant fusion proteins was produced based on the two putative sugar binding domains of MLB (1st Domain and 2nd Domain). Two conjugates (1st

Domain-Melittin and 1st Domain-Mastoparan) were designed from fusing the 1st Domain of MLB to melittin and mastoparan respectively through a Cat-B linker (Figure 79, 80). Another two conjugates (2nd Domain-Melittin and 2nd Domain-Mastoparan) were designed from fusing the 2nd Domain of MLB to melittin and mastoparan respectively through a Cat-B linker (Figure 81, 82). While Cat-B linker keeps the toxic peptides inactive until the conjugate binds to melanoma cells, the overexpressed Cat-B enzyme on the cancer cells was supposed to cleave the linker and releases the toxic peptides from the conjugates and trigger their cytotoxicity as a similar experiment was previously demonstrated by Moreno and Giralt (2015).

All the recombinant conjugates were successfully cloned into pGFPuv expression vector with a $(His)_6$ tag and the correct size of the cloned constructs was confirmed by restriction digest reactions with appropriate enzymes (Figure 83, 84). Sequencing reaction results also confirmed the complete sequences of all the conjugates without any mutation and gaps (Figure 85-92).

To express toxic peptides like melittin and mastoparan in *E. coli* cells, a specific cell line has to be selected which has potential to express toxic genes because the cytotoxicity of the expressed peptides might inhibit the proliferation of the bacterial cells and halt the expression process. Therefore, Lemo21(DE3) was selected which is a derivative of BL21(DE3) and highly suggested for the expression of toxic and membrane proteins (Schlegel, Löfblom et al. 2012).

The lectin-toxin conjugates were successfully expressed and purified through immobilized metal affinity chromatography (IMAC). The induced, uninduced and purified products of the conjugates were then analysed by SDS PAGE. The predicted size bands of the conjugates made from the whole MLB sequence fused to melittin and mastoparan which

are MLB-Melittin 1 (36.04kD), MLB-Melittin 2 (35.34), MLB-Mastoparan 1 (34.67kD), MLB-Mastoparan 2 (33.97kD) were successfully detected on SDS gels as shown in figure 95, 97, 99, 101.

Also, the correct size bands of the conjugates made from 1st Domain of MLB fused to melittin and mastoparan named as 1st Domain-Melittin (18.31kD) and 1st Domain-Mastoparan (17.84kD) and the conjugates made from 2nd Domain of MLB fused to melittin and mastoparan named as 2nd Domain-Melittin (17.60kD) and 2nd Domain-Mastoparan (17.30kD) were successfully detected (Figure 103, 105, 107, 109). However, all the constructs made from 1st and 2nd domains of MLB showed slightly (~3kD) bigger size bands on the SDS gels which may be due to "Gel shifting". The main reason of gel shifting observations is believed to be the tertiary structure of the expressed proteins which affects the level of detergent-loading on protein and consequently affects the migration rates of polypeptide on SDS-PAGE and this result agree with Rath, Glibowicka et al. (2009).

Western blotting reactions of Lectin-Toxin conjugates confirmed the identity of the conjugates since correct bands were detected for all the fusion proteins transferred to nitrocellulose membranes and treated with Anti-His primary antibody (Figure 93).

In conclusion, two sets of fusion proteins were successfully designed from the whole MLB chain and its two putative sugar binding sites (1st Domain and 2nd Domain) fused to melittin and mastoparan toxic peptides. The fusion proteins were successfully expressed and purified for further cell experiments.

Chapter 6. Cellular cytotoxicity of Lectin-Toxin Fusion Proteins

6.1 Introduction

The specific biological activities of some peptides have given them potential to be used as therapeutic agents especially in cancer biomedicine. Research to produce effective drugs from these peptides are at different stages of progression for examples the lanreotide peptide, a synthetic analogue of Somatostatin which is a naturally occurring inhibitory hormone which blocks the release of several other hormones, is at clinical stage for the treatment of gastrointestinal tumours but there are many other peptides under investigation such as depsipeptides found in microorganisms and have recently been characterized as potent anticancer agents (Pokuri, Fong et al. 2016, Serrill, Wan et al. 2016).

To improve the efficiency of chemotherapies, the main recent focus has been on the identification of appropriate carrier vehicles to incorporate with the existence chemotherapies so as to alter their pharmacokinetics and/or attach the chemotherapies to targeting moieties to shift their pharmacodynamics (Kurrikoff and Suhorut 2012). The effective way to perform these improvements is the attachment of the drugs to a tumour specific moiety such as receptor ligand, monoclonal antibodies and targeting peptide. However, a large-scale screening also needs be conducted to identify an appropriate moiety. For example, extensive screening of peptide libraries is needed to characterize an efficient targeting peptide with selectivity only to cancer cells (Pasqualini, Koivunen et al. 1997).

While the combination of drugs and targeting vehicles often faces difficulties in penetrating tumours to reach their interior, cell internalization capacity also needs to be added to the drug targeting moiety through adding other molecules with the ability to cross

cells membranes such as cell penetrating peptides (CPPs) (Essler and Ruoslahti 2002, Myrberg, Zhang et al. 2007).

The binding chain of mistletoe lectin (MLB) has ability to bind to specific cell surface oligosaccharides containing sialic acid and has been shown to be internalised by the cells. Therefore, MLB can be used as a carrier vehicle to deliver therapeutic agents like cytotoxic peptides specifically to cancer cells which have sialic acid on their cells surface oligosaccharides (Lyu, Kwon et al. 2004).

The aim of this study is the *in vitro* assessment of the ability of the recombinant fusion proteins produced from melittin and mastoparan cytotoxic peptides fused to the binding chain of mistletoe lectin (MLB), and to truncated peptides containing its two putative sugar binding site, to specifically recognize and bind to the human melanoma cells at the primary and metastatic progression stages and also to assess their selectivity to avoid binding the healthy melanocyte cells and eliminate them from their cytotoxic effects.

6.2 Material and Methods

6.2.1 Cell Culture

Two human melanoma cell lines at two different stages of cancer development (WM-115 and WM 266-4) (ECACC, UK) were cultured in RPMI-1640 culture media (Sigma-Aldrich, UK) supplemented with 10% (v/v) fetal bovine serum (FBS) and 1% (v/v) glutamine at 37 ºC in a humidified atmosphere containing 5% carbon dioxide.

Healthy human melanocyte cells (HEMn) (ECACC, UK) (Table 20) were cultured in melanocyte growth media (Sigma-Aldrich, UK) supplemented with bovine pituitary extract (0.004 ml/ml), basic fibroblast growth factor (1 ng/ml), insulin (5 µg/m), hydrocortisone (0.5 µg/ml) and phorbol myristate acetate (10 ng/ml) at 37ºC in a humidified atmosphere containing 5% carbon dioxide.

All the cell lines were grown until 80% confluency achieved in about 72 to 96 hours followed by subculturing. The culture flask was decanted from the culture media and 0.5 ml trypsin/EDTA (Sigma-Aldrich, UK) was added and incubated for about 5 minutes at 37 ºC to detach the cell monolayer from the flask. The flask content was transferred into a universal tube and centrifuged at 500 x g for 5 minutes. The supernatant was discarded and the pellet was re-suspended in 5 ml of culture media. The split was carried out at recommended 1:3 ratio and transferred into 25 or 75 cm^2 culture flasks according to the amount of cells needed for the further tests and the total volume of the media made up to 5 or 15 ml respectively.

Table 20. Human cell lines used in cell culture experiments.

Cell line	Source	Tumor Type	Passage No.
HEMn	ECACC	Primary human epidermal melanocytes isolated from neonatal foreskin	-
WM-115	ECACC	Derived from the skin of 58 years old homo sapiens female with primary melanoma.	85
WM 266-4	ECACC	Derived from the skin of the same 58 years old homo sapiens female at the metastatic level of melanoma.	+3

6.2.2 The apoptotic Effect of Lectin-Toxin Fusion Proteins on Cells

Human melanoma cells (WM-115, WM 266-4) and the healthy melanocyte cells (HEMn) were cultured on 96-well plate with the optimum seeding density 5000 cells/well (~1.4 x 10^4 cells/cm2) detected in section 4.2.5. After 24hours incubation at 37 ºC in humidified atmosphere containing 5% $C0_2$, the media was changed to the same RPMI-1640 but with lower amount of FBS (2%). The cells were then treated with 0.5µM from each of the fusion proteins and incubated another 24 and 48hours. The cell viability was measured through adding 20µl of the MTS solution and incubated at 37ºC for 3 hours. The absorbance of the mixture was then measured at 490 nm and the blank absorbance at 650 nm using micro plate reader (Dynatech MR5000, USA).

6.2.3 Cell Morphology Assessment

The effect of Lectin-Toxin fusion proteins on cell morphology was assessed through treating human metastatic melanoma cells (WM 266-4) with different concentration of the fusions. The cancer cells were cultured in 25 cm^2 flasks as described in section 4.2.5. When the cells reached 80% confluency the culture media was removed and the cells were

washed two times with 1X PBS. The cells were then detached from the flask using trypsin/EDTA and transferred into a universal tube followed by centrifugation. The cells were re-suspended in 3ml of appropriate culture media and the counted. The seeding density 5000 cells/well was used to plate the cells on the 96-well plate and the plate was incubated for 24 hours at 37^{o}C in humidified atmosphere containing 5% $C0_{2.}$ After the incubation time, the cells were treated with 0.1, 0.3 and 0.5µM fusion proteins overnight. Finally, the cells were screened and photographed under the microscope.

6.2.4 Cell membrane Binding and Internalization of Lectin-Toxin Fusions

The ability of the Lectin-Toxin fusion proteins to bind to the cell glycocalyx and cross the cell membrane was determined through Fluorescein isothiocyanate (FITC) staining. The subsequent morphological apoptotic signs caused by the incubation of the cells with the labelled proteins were also observed.

6.2.4.1 FITC labelling of the Lectin-Toxin Fusions

The FITC stain (Sigma-Aldrich, UK) was dissolved in 1X PBS with the concentration of 0.5µg/µl. A fresh 1 M carbonate buffer pH 9.5 was prepared from 0.32 M sodium carbonate (Na_2CO_3) and 0.68 M sodium bicarbonate (NaHCO3). The conjugation reaction was carried out by mixing 200 µl of fusion protein (~1µg/µl), 25 µl carbonate buffer, 25 µl 1X PBS and 50 µl FITC solution in light-protected eppendorf tube and incubated at room temperature for 2 hours with gentle agitation. The reaction mixture was then extensively dialyzed overnight at 4^{o}C in dark to remove unbound FITC. The FITC labelled fusion proteins were stored at 4^{o}C until further use.

6.2.4.2 Cell Preparation

Human melanoma cells (WM-155 and WM 266-4) and the healthy human melanocyte cells (HEMn) were cultured in 25cm^2 flasks as described in section 4.2.3. The cells were then plated on the 96-well plate with the seeding density of 5000 cells/well as described in section 4.2.5 and incubated for 24 hours at 37^{o}C in humidified atmosphere containing 5% $C0_2$.

6.2.4.3 Epifluorescence Imaging Microscopy

Live confocal microscopy was conducted on two melanoma cell lines (WM-115 and WM 266-4) and one healthy melanocyte cell line (HEMn) to investigate the binding and cell penetrating capacity of the Lectin-Toxin fusions and also to see the intracellular distribution of the internalized fusions. To plate the cells, a glass coverslip was placed in 6 well plates and the cells were grown in the wells overnight. The cells were then treated with 0.5µM of FITC labelled fusion proteins for four hours followed by washing with ice-cold PBS. The binding of the fusion proteins to the cell glycocalyx and their colocalization were screened under epifluorescence imaging microscopy equipped with 10X and 20X objectives (Nikon, UK). The FITC fluorophore was excited with blue light (488nm).

6.3 Results

The binding of the recombinant MLB-Toxins, 1st Domain-Toxins and 2nd Domain-Toxins fusion proteins specifically to melanoma cancer cells and their utility in delivering cytotoxic peptides to the cells were assessed by treating two human melanoma cell lines (WM-115 and WM 266-4) and one healthy human melanocytes (HEMn) with the recombinant fusion proteins.

6.3.1 Detection of Cell Variability by MTS Assay

A colorimetric assay (MTS) was used to determine induced cell death in human melanoma and healthy melanocyte cells by the recombinant fusion proteins.

Significant increases were observed in the apoptosis of both melanoma cells (WM-115, WM 266-4) specifically metastatic cells (MW 266-4) when they were treated with (0.5µM) of the recombinant fusion proteins for 24 hours. However, all the fusion proteins showed negligible effects on the healthy human melanocyte cells (HEMn).

Figure 111, 112, 113 show the effects of MLB-Melittin 1, MLB-Melittin 2, 1st Domain-Melitin and 1st Domain-Mastoparan fusion proteins on the healthy human melanocytes (HEMn), human primary melanoma cells (WM-115) and human metastatic melanoma cells (WM 266-4) respectively.

The treatment of the healthy melanocyte cells with the fusion proteins had almost no effects on the cells (Figure 111) whereas MLB-Melittin 1 and 1st Domain-Mastoparan showed a significant apoptotic effects on both primary and metastatic melanoma cells. However, MLB-Melittin 2 and 1st Domain-Melittin were seen to have no significant effects on the cancer cells as well (Figure 112, 113).

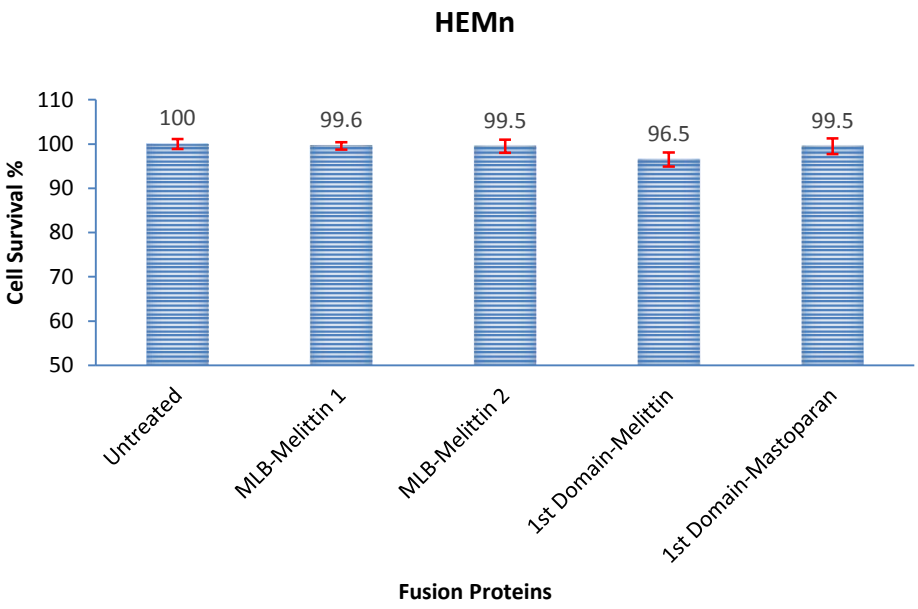


Figure 111. The MTS assay conducted on human healthy melanocytes cells (HEMn) after 24hours incubation time. First column: Untreated cells and control; Second column: Cells treated with 0.5µM of MLB-Melittin 1 fusion protein; Third column: Cells treated with 0.5µM of MLB-Melittin 2 fusion protein; Forth column: Cells treated with 0.5µM of 1st Domain-Melittin fusion protein; Fifth column: Cells treated with 0.5µM of 1st Domain-Mastoparan fusion protein.

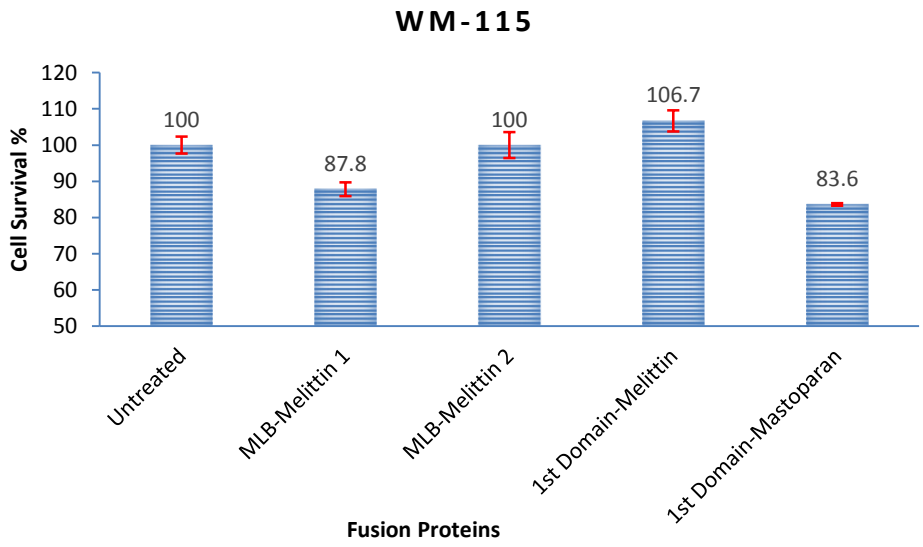


Figure 112. The MTS assay conducted on human primary melanoma cells (WM-115) after 24hours incubation time. First column: Untreated cells and control; Second column: Cells treated with 0.5µM of MLB-Melittin 1 fusion protein; Third column: Cells treated with 0.5µM of MLB-Melittin 2 fusion protein; Forth column: Cells treated with 0.5µM of 1st Domain-Melittin fusion protein; Fifth column: Cells treated with 0.5µM of 1st Domain-Mastoparan fusion protein.

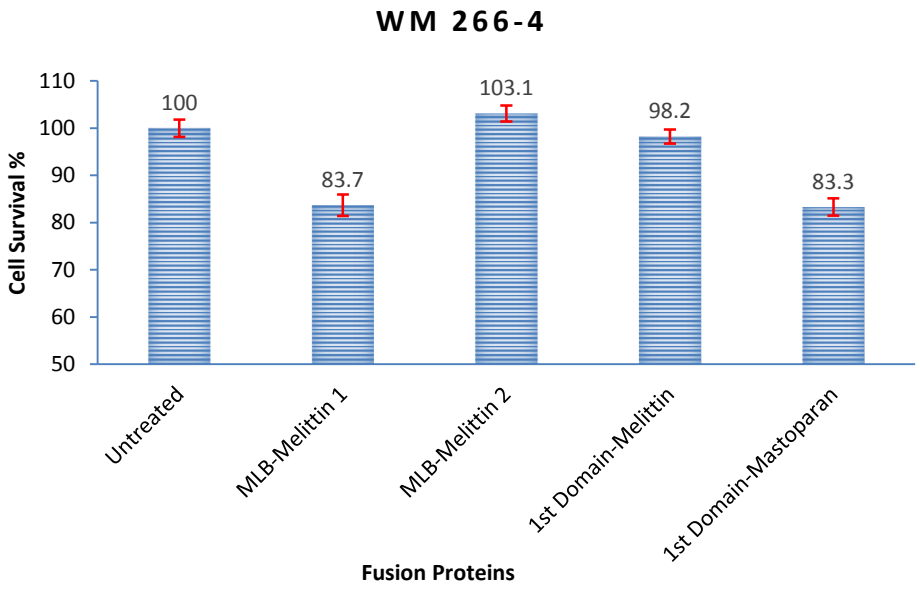


Figure 113. The MTS assay conducted on human metastatic melanoma cells (WM 266-4) after 24hours incubation time. First column: Untreated cells and control; Second column: Cells treated with 0.5µM of MLB-Melittin 1 fusion protein; Third column: Cells treated with 0.5µM of MLB-Melittin 2 fusion protein; Forth column: Cells treated with 0.5µM of 1st Domain-Melittin fusion protein; Fifth column: Cells treated with 0.5µM of 1st Domain-Mastoparan fusion protein.

In figure 114, 115 and 116 the effect of MLB-Mastoparan 1, MLB-Mastoparan 2, 2nd Domain-Melitin and 2nd Domain-Mastoparan fusion proteins on the healthy human melanocytes (HEMn), human primary melanoma cells (WM-115) and human metastatic melanoma cells (WM 266-4) are shown respectively.

The treatment of the healthy melanocyte cells with the fusion proteins showed almost no effects of the fusions on the cells (Figure 114) whereas MLB-Mastoparan 1 and 2ndDomain-Mastoparan showed a significant apoptotic effects on metastatic melanoma cells but less effects on the primary cells. However, the effects of MLB-Mastoparan 2 and 2nd Domain-Melittin were again seen insignificant on the cancer cells as well but higher than MLB-Melittin 2 and 1stDomain-Melittin (Figure 115, 116).

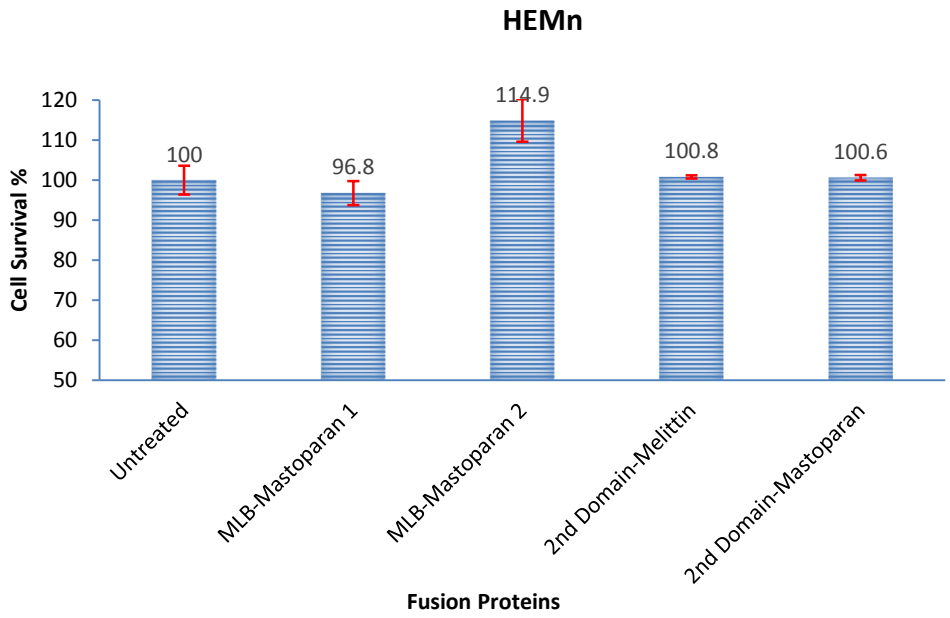


Figure 114. The MTS assay conducted on human healthy melanocytes cells (HEMn) after 24hours incubation time. First column: Untreated cells and control; Second column: Cells treated with 0.5µM of MLB-Mastoparan 1 fusion protein; Third column: Cells treated with 0.5µM of MLB-Mastoparan 2 fusion protein; Forth column: Cells treated with 0.5µM of 2nd Domain-Melittin fusion protein; Fifth column: Cells treated with 0.5µM of 2nd Domain-Mastoparan fusion protein.

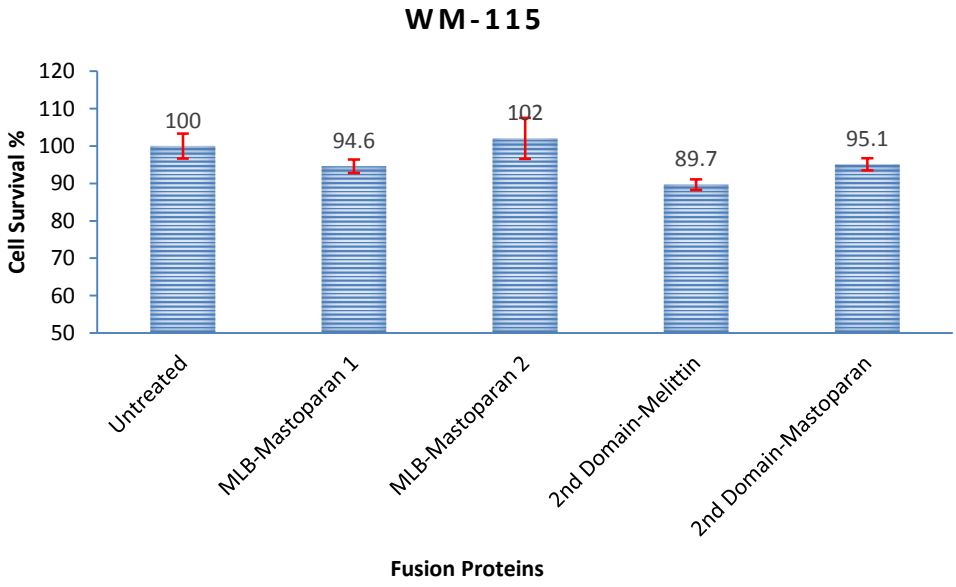


Figure 115. The MTS assay conducted on human primary melanoma cells (WM-115) after 24hours incubation time. First column: Untreated cells and control; Second column: Cells treated with 0.5µM of MLB-Mastoparan 1 fusion protein; Third column: Cells treated with 0.5µM of MLB-Mastoparan 2 fusion protein; Forth column: Cells treated with 0.5µM of 2nd Domain-Melittin fusion protein; Fifth column: Cells treated with 0.5µM of 2nd Domain-Mastoparan fusion protein.

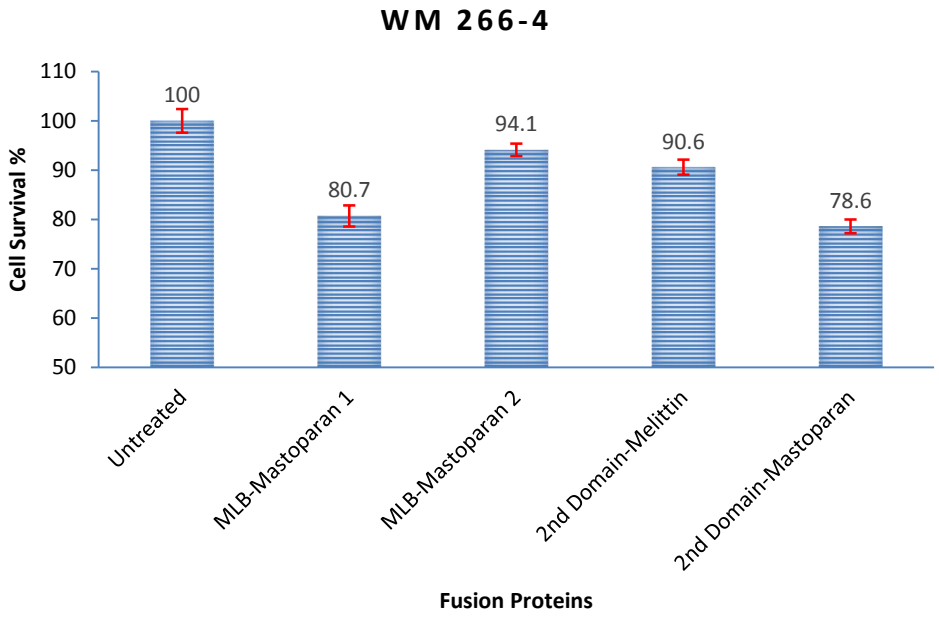


Figure 116. The MTS assay conducted on human metastatic melanoma cells (WM 266-4) after 24hours incubation time. First column: Untreated cells and control; Second column: Cells treated with 0.5µM of MLB-Mastoparan 1 fusion protein; Third column: Cells treated with 0.5µM of MLB-Mastoparan 2 fusion protein; Forth column: Cells treated with 0.5µM of 2nd Domain-Melittin fusion protein; Fifth column: Cells treated with 0.5µM of 2nd Domain-Mastoparan fusion protein.

Following the confirmation of the cytotoxic effect of the fusion protein on the melanoma cells especially on the metastatic cells (WM 266-4), a longer incubation time of the fusion proteins with the metastatic cells was tried in order to assess their cytotoxicity. For that, the cells were treated with the same concentration (0.5µM) of the fusion proteins but for 48hours.

In comparison with the 24hours incubation, a dramatic change was observed in the cytotoxic effect of the fusion proteins after 48hours incubation. The fusion proteins were seen more cytotoxic to the cells and significantly reduced their viability.

Figure 117 shows the cytotoxic effect of MLB-Melittin 1, MLB-Melittin 2, 1st Domain-Melitin and 1st Domain-Mastoparan fusion proteins on the metastatic melanoma cells (WM 266-4) after 48hours incubation. It can clearly be seen that all the fusion proteins had more cytotoxic effect on the cells especially MLB-Melittin 2 and 1st Domain-Melittin which showed much higher activity than the 24hours incubation time.

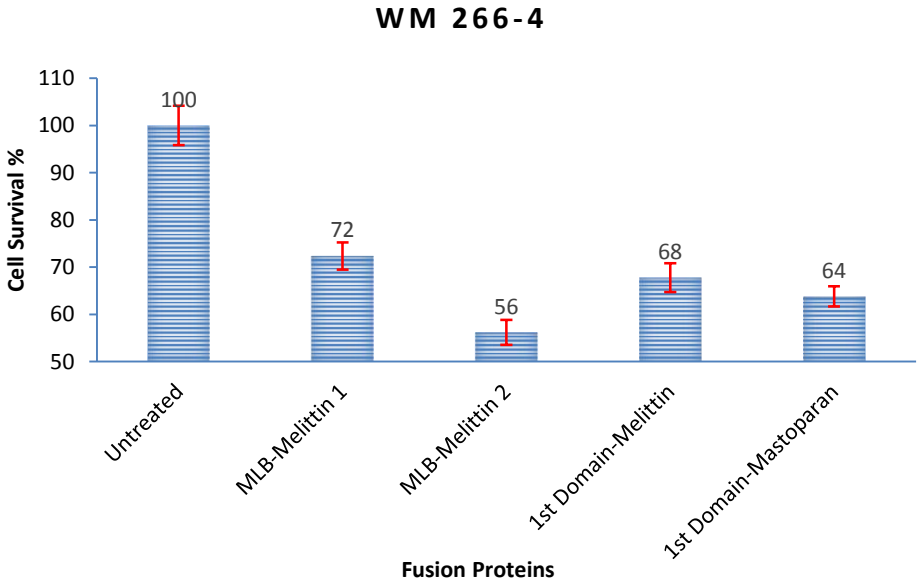


Figure 117. The MTS assay conducted on human metastatic melanoma cells (WM 266-4) after 48hours incubation time. First column: Untreated cells and control; Second column: Cells treated with 0.5µM of MLB-Melittin 1 fusion protein; Third column: Cells treated with 0.5µM of MLB-Melittin 2 fusion protein; Forth column: Cells treated with 0.5µM of 1st Domain-Melittin fusion protein; Fifth column: Cells treated with 0.5µM of 1st Domain-Mastoparan fusion protein.

Figure 118 shows the cytotoxic effect of MLB-Mastoparan 1, MLB-Mastoparan 2, 2nd Domain-Melittin and 2nd Domain-Mastoparan fusion proteins on the metastatic melanoma cells (WM 266-4) after 48hours incubation. The longer incubation time was seen to dramatically increase the cytotoxic effect of the fusion proteins since longer exposure time of the cells to MLB-Mastoparan 1 and 2nd Domain-Mastoparan reduced their viability to almost half.

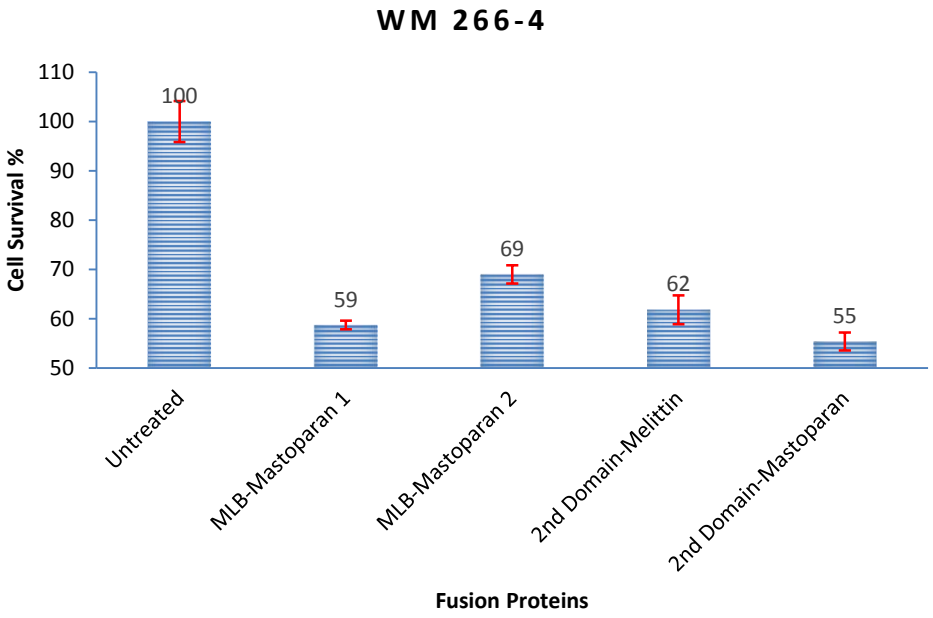


Figure 118. The MTS assay conducted on human metastatic melanoma cells (WM 266-4) after 48hours incubation time. First column: Untreated cells and control; Second column: Cells treated with 0.5µM of MLB-Mastoparan 1 fusion protein; Third column: Cells treated with 0.5µM of MLB-Mastoparan 2 fusion protein; Forth column: Cells treated with 0.5µM of 2nd Domain-Melittin fusion protein; Fifth column: Cells treated with 0.5µM of 2nd Domain-Mastoparan fusion protein.

6.3.2 Cell Morphology Assessment

The interaction of melittin and mastoparan with the cell membrane and their ability to break it was assessed through screening the morphology of metastatic melanoma cells (WM266-4) under the microscope following treating them with various concentration of the Lectin-Toxin fusion proteins. The untreated cells were used as a control to distinguish the morphological changes happened during the treatment. The effect of the fusion proteins on the shape of the cells was clearly seen but different level of activity was shown by the proteins at different concentrations. The fusion proteins were seen to effectively interact with the cell membranes and lyse them. Figure 119-124 shows WM266-4 cells treated with three different concentrations (0.1, 0.3 and 0.5µM) of MLB-Melittin 1, 1st Domain-Melittin, 1st Domain-Mastoparan, MLB-Mastoparan 1, 2nd Domain-Melittin and 2nd Domain-Mastoparan fusion protein respectively. The untreated cells in picture A was

used as control in which have clear shape and smooth cell membranes. In picture B, few number of deformed cells and chromatin shrinking can be seen and their number increased with the increase of the fusion protein concentration shown in picture C. In picture D, a large number of cells are deformed and a large amount of apoptotic bodies can be seen around the cells as a results of a higher ratio of cell membranes lysis caused by the higher concentration of the fusion proteins. Despite that all the fusion proteins showed effect on the morphology of the cells, MLB-Mastoparan 1 and 1st Domain-Mastoparan fusion proteins were seen to have the highest impact on the cells since the highest amount of apoptotic bodies was observed around the cells treated with these two fusion protein which is an indication that the fusion protein possesses a potent membrane lytic activity (Picture D of figure 121,122).

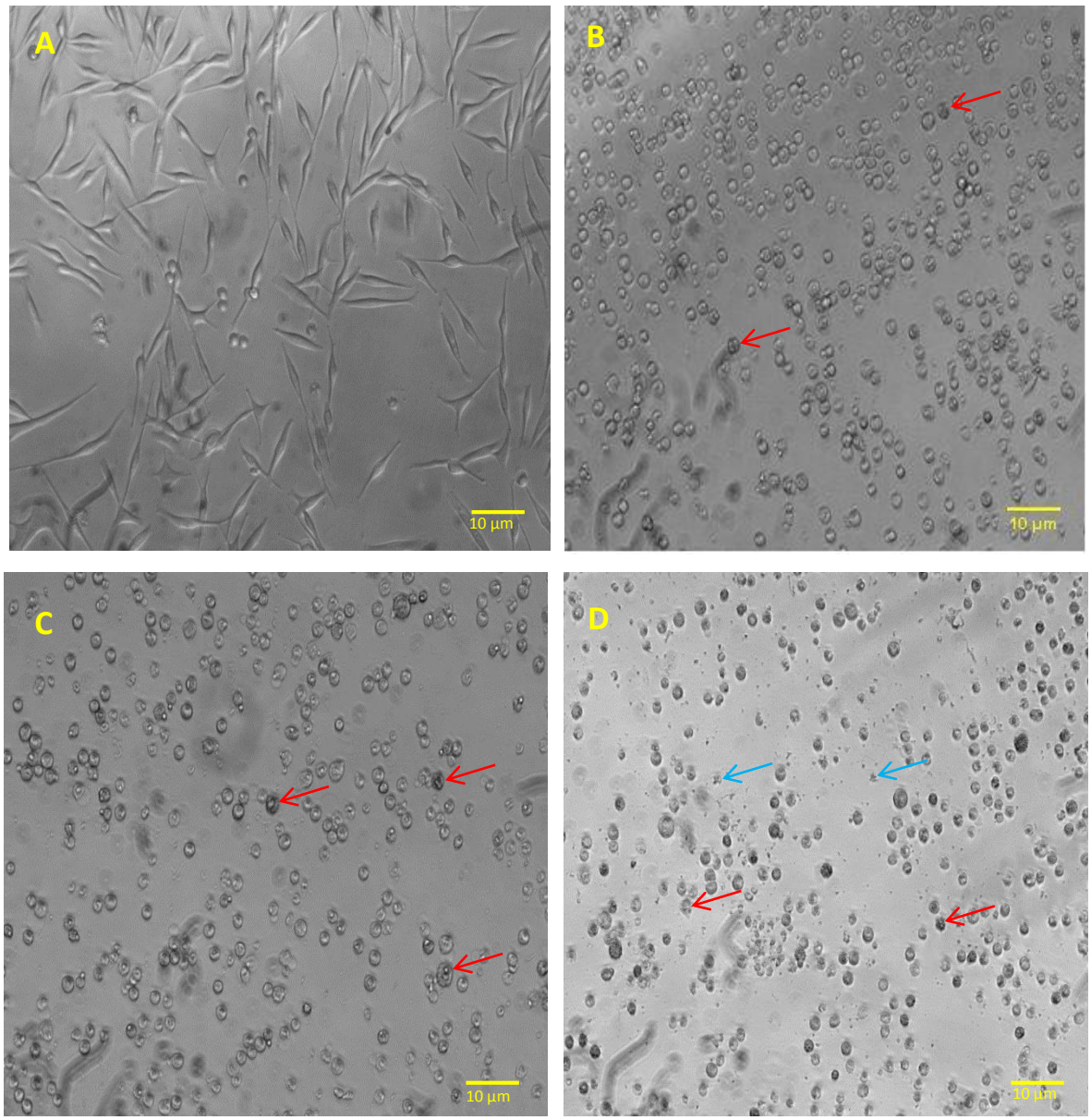


Figure 119. Human metastatic melanoma cells (WM 266-4) treated with MLB-Melittin 1 fusion protein. Picture A: Untreated cells (control); Picture B, C, D: Cells treated with 0.1, 0.3, 0.5µM fusion protein respectively. The red arrows show the deformed cells and the blue arrows show the apoptotic bodies.

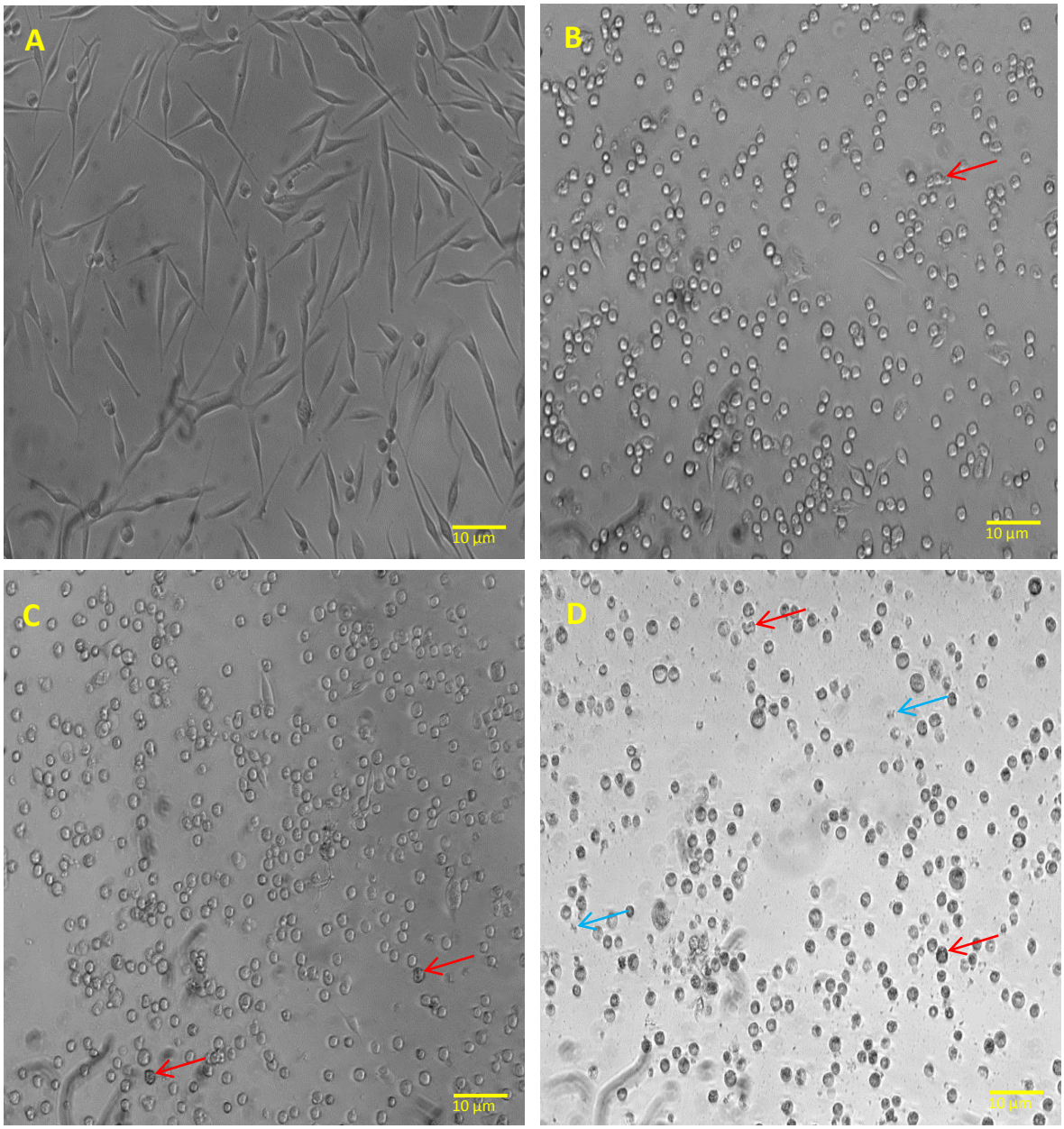


Figure 120. Human metastatic melanoma cells (WM 266-4) treated with 1st Domain-Melittin fusion protein. Picture A: Untreated cells (control); Picture B, C, D: Cells treated with 0.1, 0.3, 0.5µM fusion protein respectively. The red arrows show the deformed cells and the blue arrows show the apoptotic bodies.

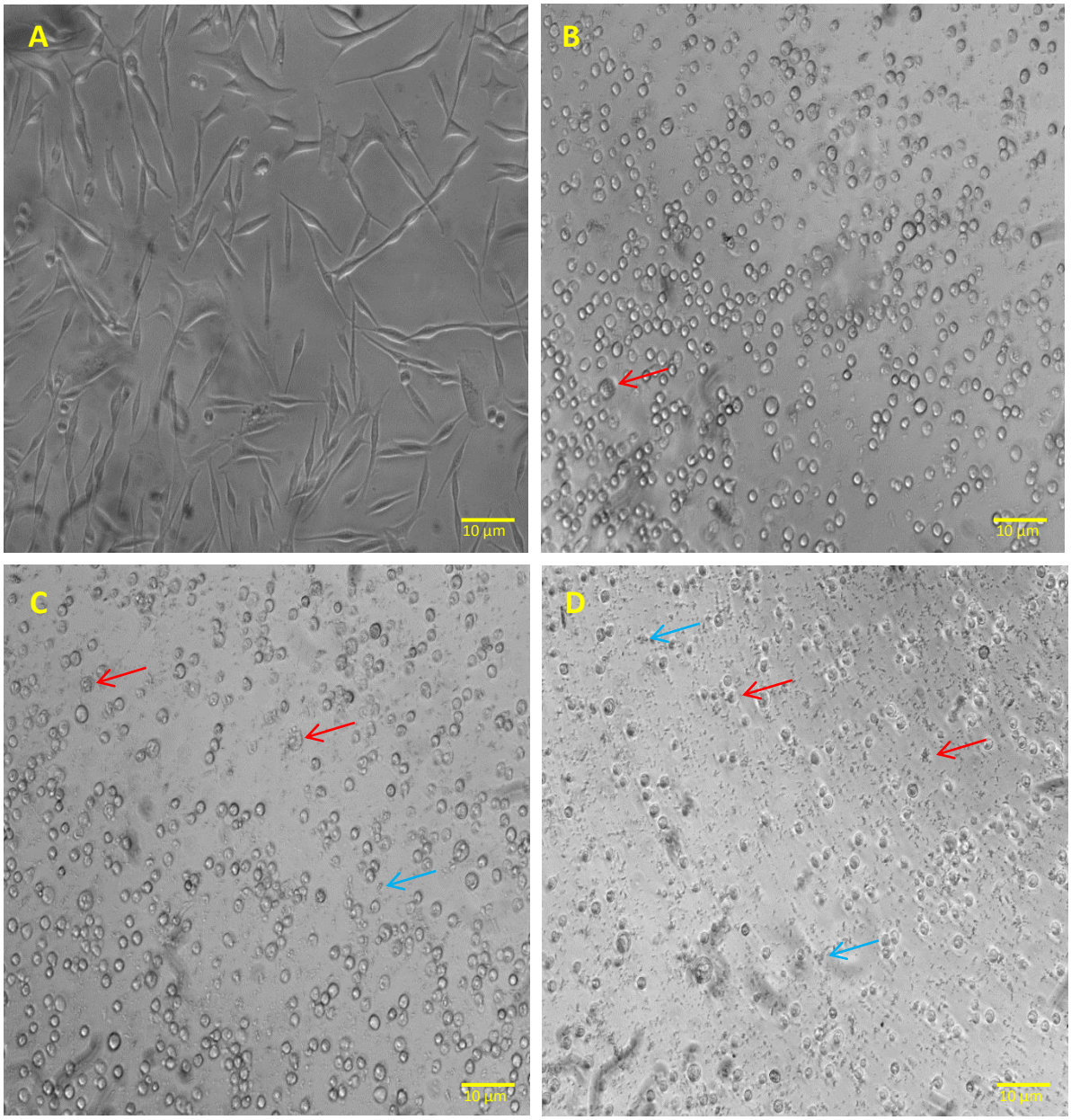


Figure 121. Human metastatic melanoma cells (WM 266-4) treated with 1st Domain-Mastoparan fusion protein. Picture A: Untreated cells (control); Picture B, C, D: Cells treated with 0.1, 0.3, 0.5µM fusion protein respectively. The red arrows show the deformed cells and the blue arrows show the apoptotic bodies.

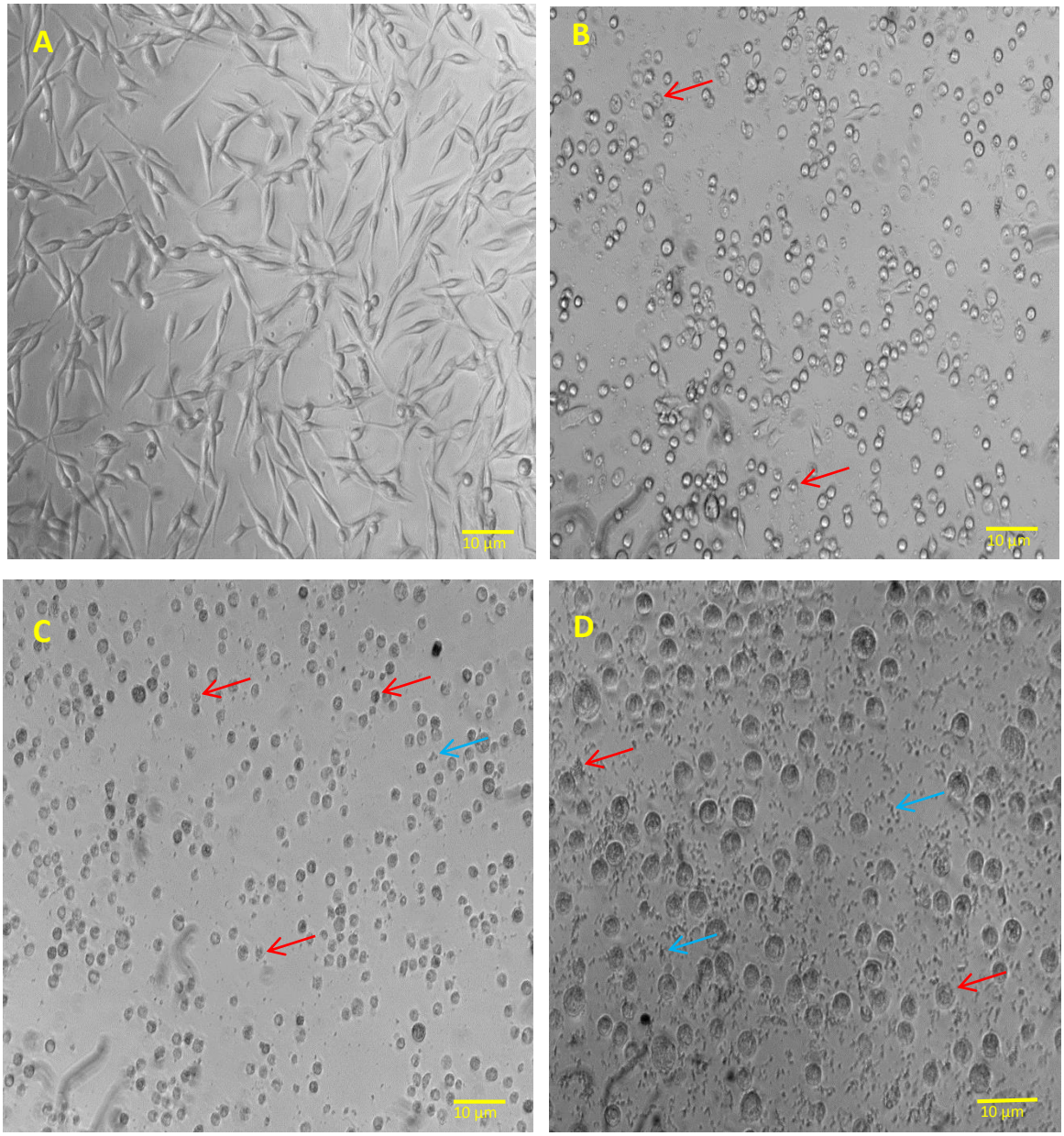


Figure 122. Human metastatic melanoma cells (WM 266-4) treated with MLB-Mastoparan 1 fusion protein. Picture A: Untreated cells (control); Picture B, C, D: Cells treated with 0.1, 0.3, 0.5µM fusion protein respectively. The red arrows show the deformed cells and the blue arrows show the apoptotic bodies.

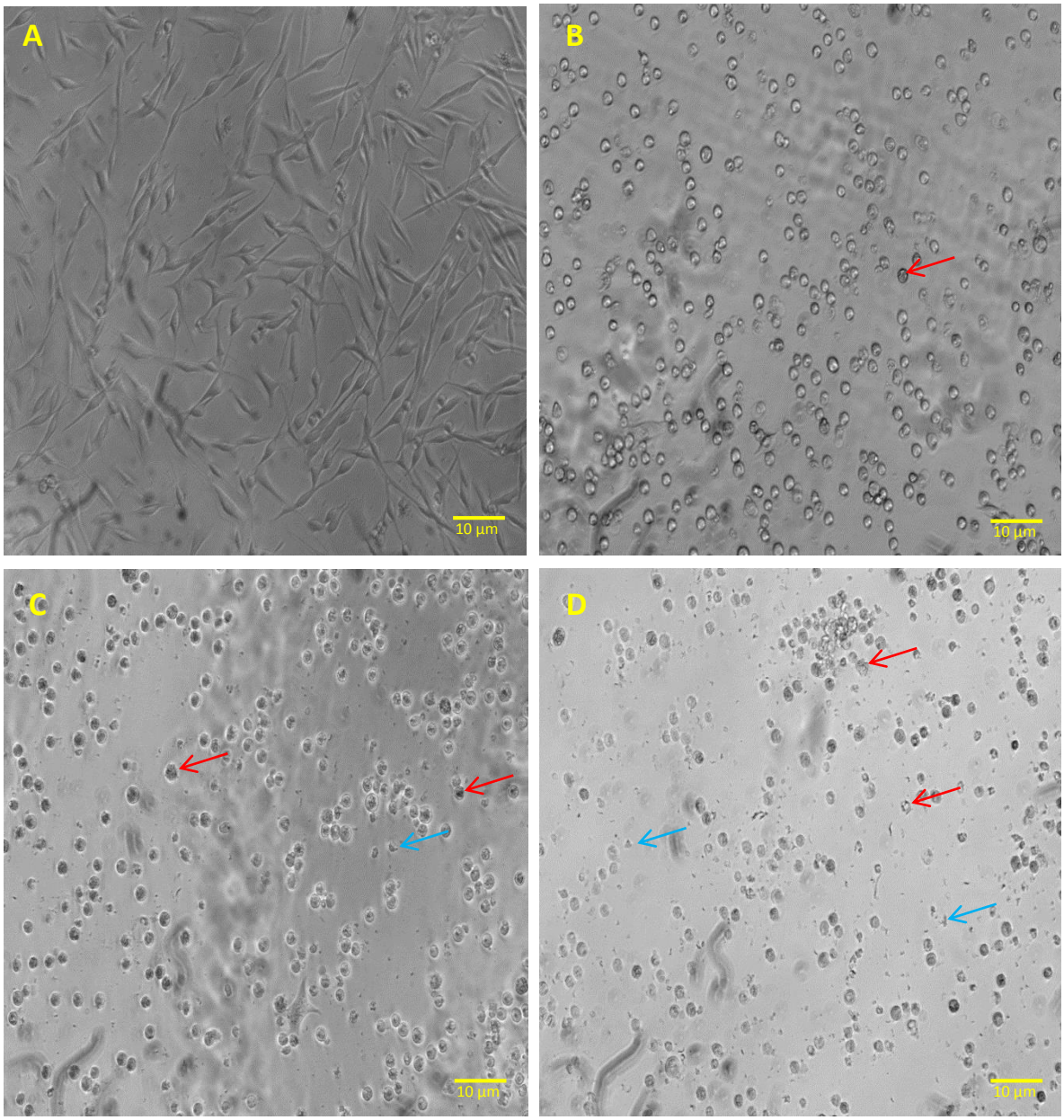


Figure 123. Human metastatic melanoma cells (WM 266-4) treated with 2nd Domain-Melittin fusion protein. Picture A: Untreated cells (control); Picture B, C, D: Cells treated with 0.1, 0.3, 0.5µM fusion protein respectively. The red arrows show the deformed cells and the blue arrows show the apoptotic bodies.

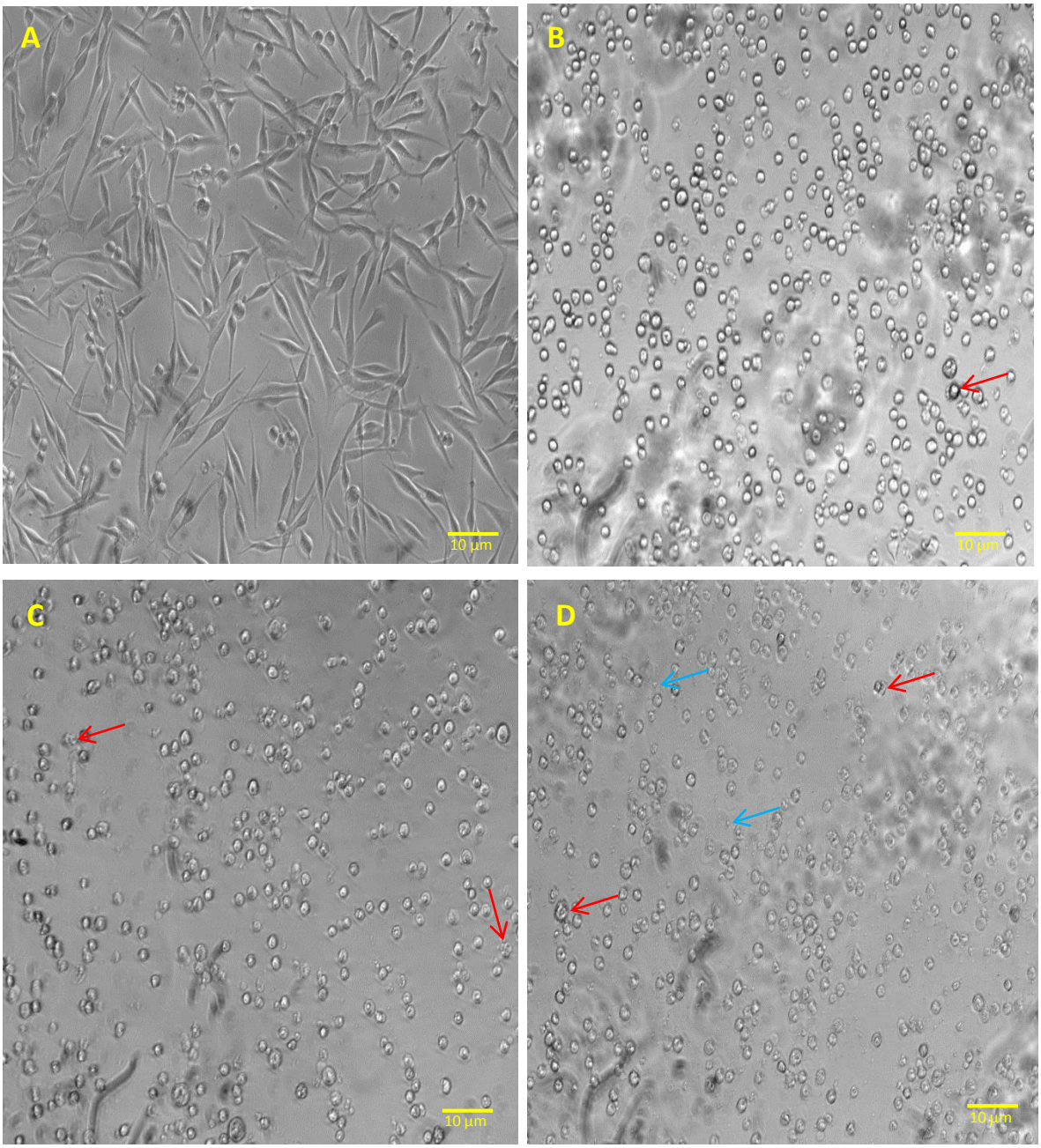


Figure 124. Human metastatic melanoma cells (WM 266-4) treated with 2nd Domain-Mastoparan fusion protein. Picture A: Untreated cells (control); Picture B, C, D: Cells treated with 0.1, 0.3, 0.5µM fusion protein respectively. The red arrows show the deformed cells and the blue arrows show the apoptotic bodies.

6.3.3 Epifluorescence Microscopy

To visualise the binding and cell internalization of the recombinant fusion proteins, they were labelled with fluorescein isothiocyanate (FITC). Subsequent morphological assessment of cells was also used to determine the effect of the recombinant fusion proteins on the cells. One day prior to infecting the cells with the fusion proteins, the melanoma cells (MW-115 and MW 266-4) and healthy melanocyte cells (HEMn) were plated on 96-well plated with the cell density of 5000 cell/well as explained in section 4.2.5. The cells were then treated with 0.5µM of each recombinant fusion protein and incubated for 4 hours. Cells were later screened under epifluorescence microscopy using blue light (488nm) to excite the FITC fluorophore (Figure 126-133). The typical shape of HEMn, WM-115 and WM 266-4 cells without any treatment is also shown in figure 125.

From the microscopic images the attachment of the fusion proteins to the cell surfaces and crossing the cell membranes can clearly be seen. Furthermore, in comparison to the untreated cells a significantly higher proportion of cell death was observed in cells treated with fusion proteins. Such results were not observed in healthy cells.

The fluorescent microscope imaging results also confirmed the results to the colorimetric assay (MTS) which was the observation of minor binding of the fusion proteins to the healthy melanocyte cells (HEMn) and minor cell agglutination. Also, the fusions proteins did not show cytotoxic effects to the healthy cells since the adhesion and growth morphology of the cells was not affected by their incubation with the fusion proteins as shown in the picture A of figure 126-133. On the other hand, picture B and C of figure 126-133 show that all the fusion proteins effectively bound to the melanoma cells (WM-115 and WM 266-45) and caused cell agglutination followed by cell internalization. However, different level of activity was observed between the fusion proteins and 1st Domain-Mastoparan fusion protein was again seen to have the strongest cell binding and

internalization to both WM-115 and WM 266-4 cells since the highest intensity of the fluorescence was detected around and inside the cells treated with this fusion protein as shown in picture B and C of figure 129.

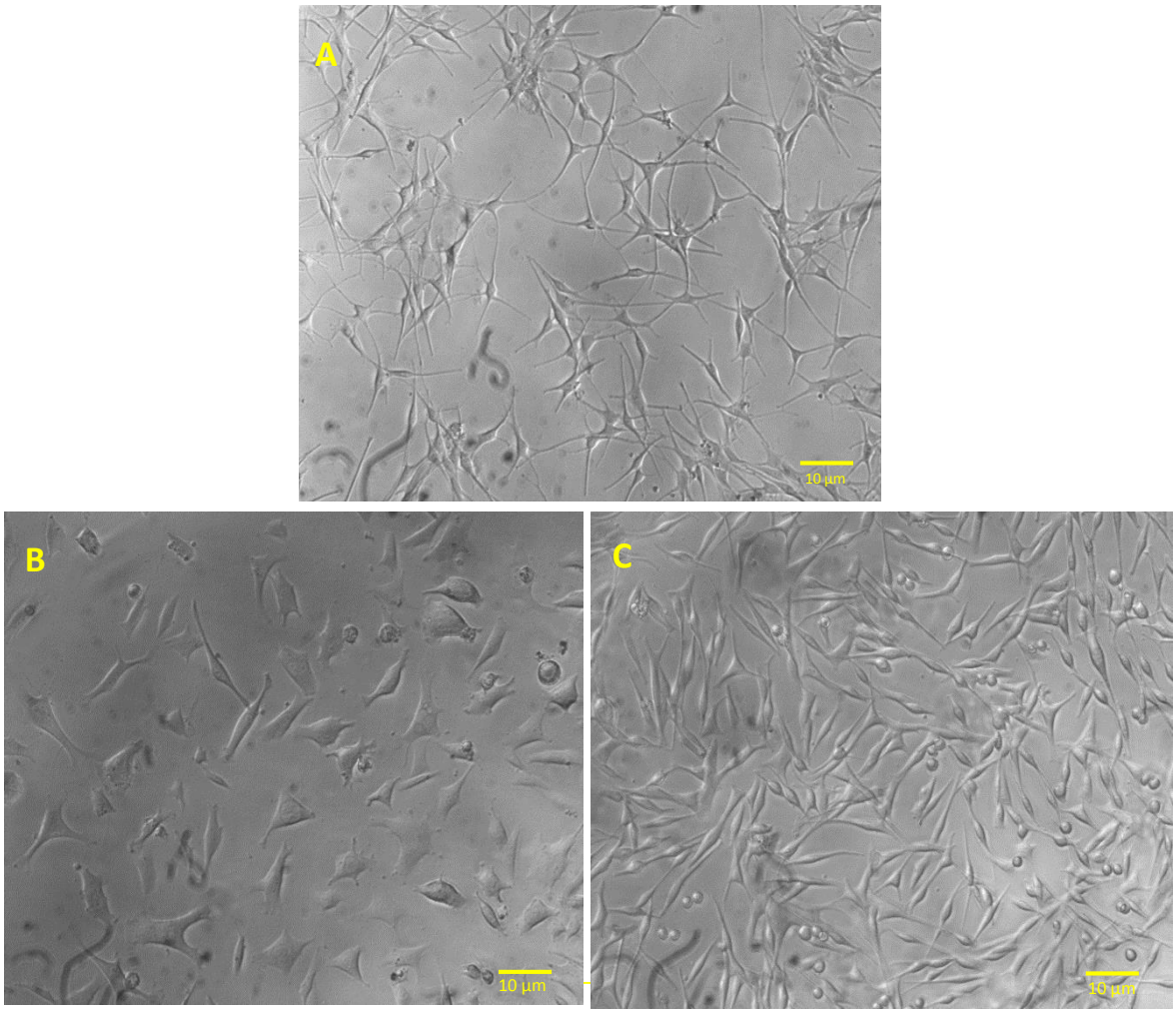


Figure 125. Typical shape of three human skin cells without any treatment. Picture A: Healthy human melanocyte cells (HEMn); Picture B: Primary human melanoma cells (WM-115); Picture C: Metastatic human melanoma cells (WM 266-4).

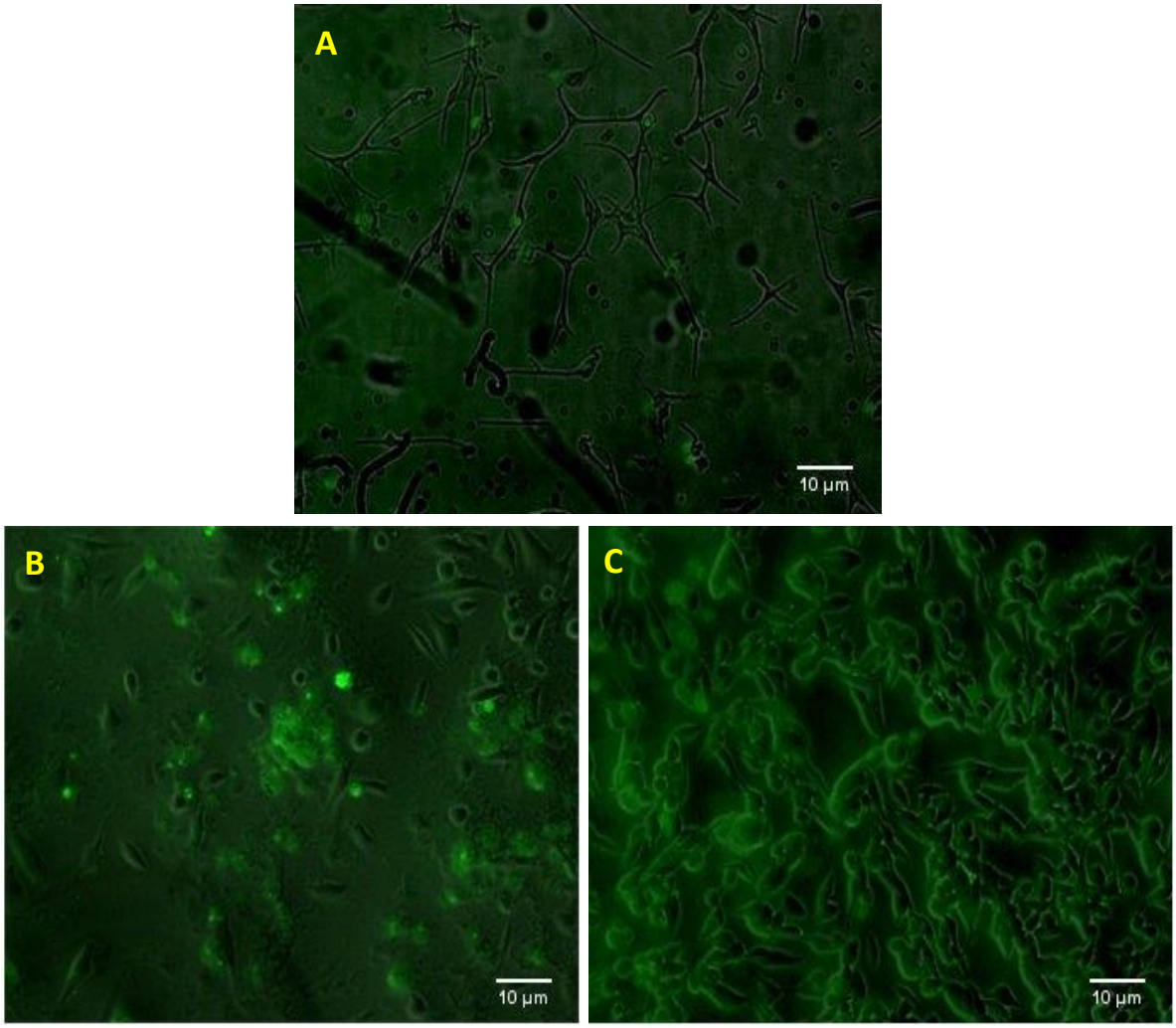


Figure 126. Human healthy melanocyte and melanoma cells incubated with 0.5µM of FITC labelled MLB-Melittin 1 fusion protein overnight and screened under luminescence imaging microscopy. Picture A: Healthy human melanocyte cells (HEMn); Picture B: Primary human melanoma cells (WM-115); Picture C: Metastatic human melanoma cells (WM 266-4).

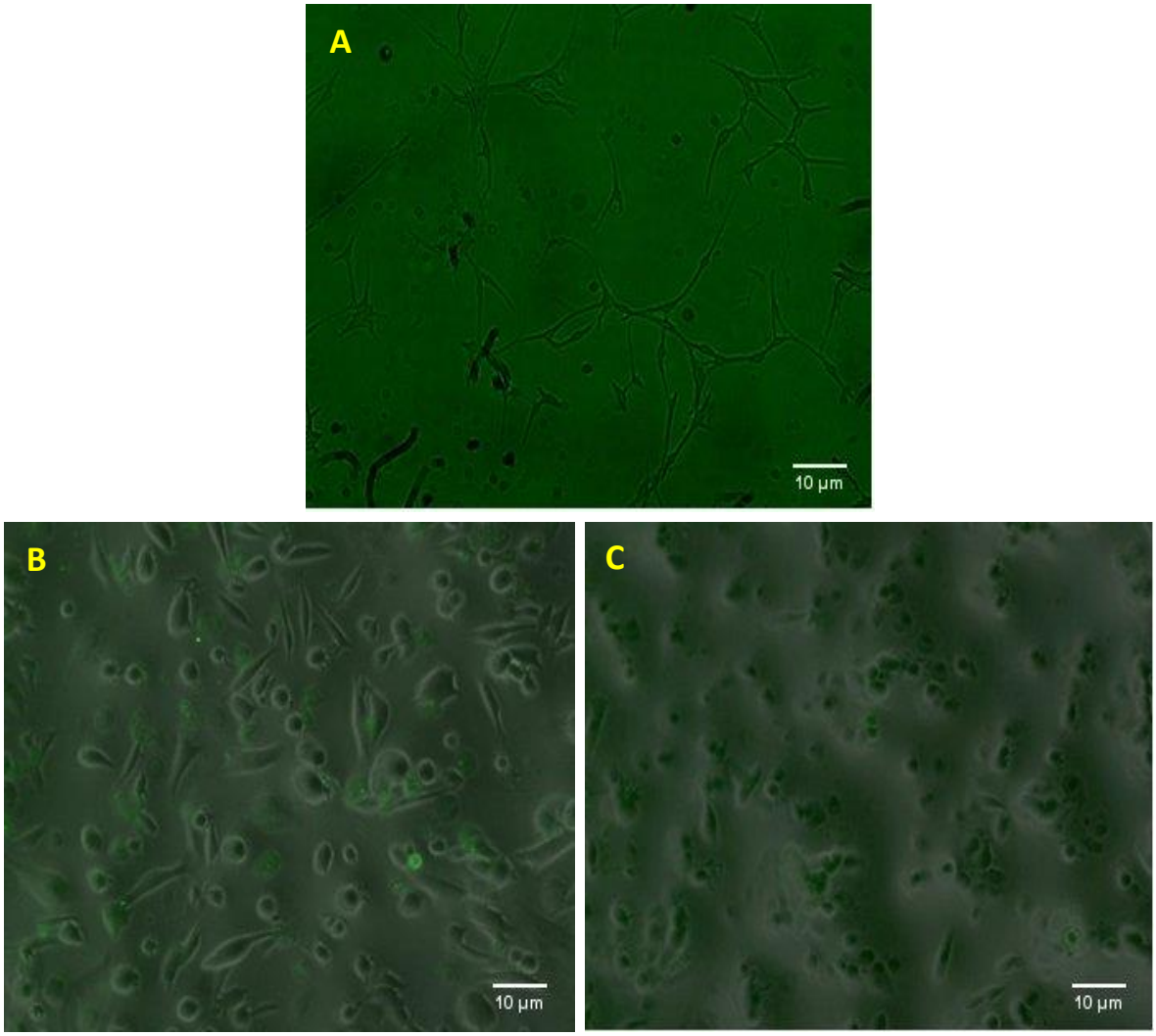


Figure 127. Human healthy melanocyte and melanoma cells incubated with 0.5µM of FITC labelled MLB-Melittin 2 fusion protein overnight and screened under luminescence imaging microscopy. Picture A: Healthy human melanocyte cells (HEMn); Picture B: Primary human melanoma cells (WM-115); Picture C: Metastatic human melanoma cells (WM 266-4).

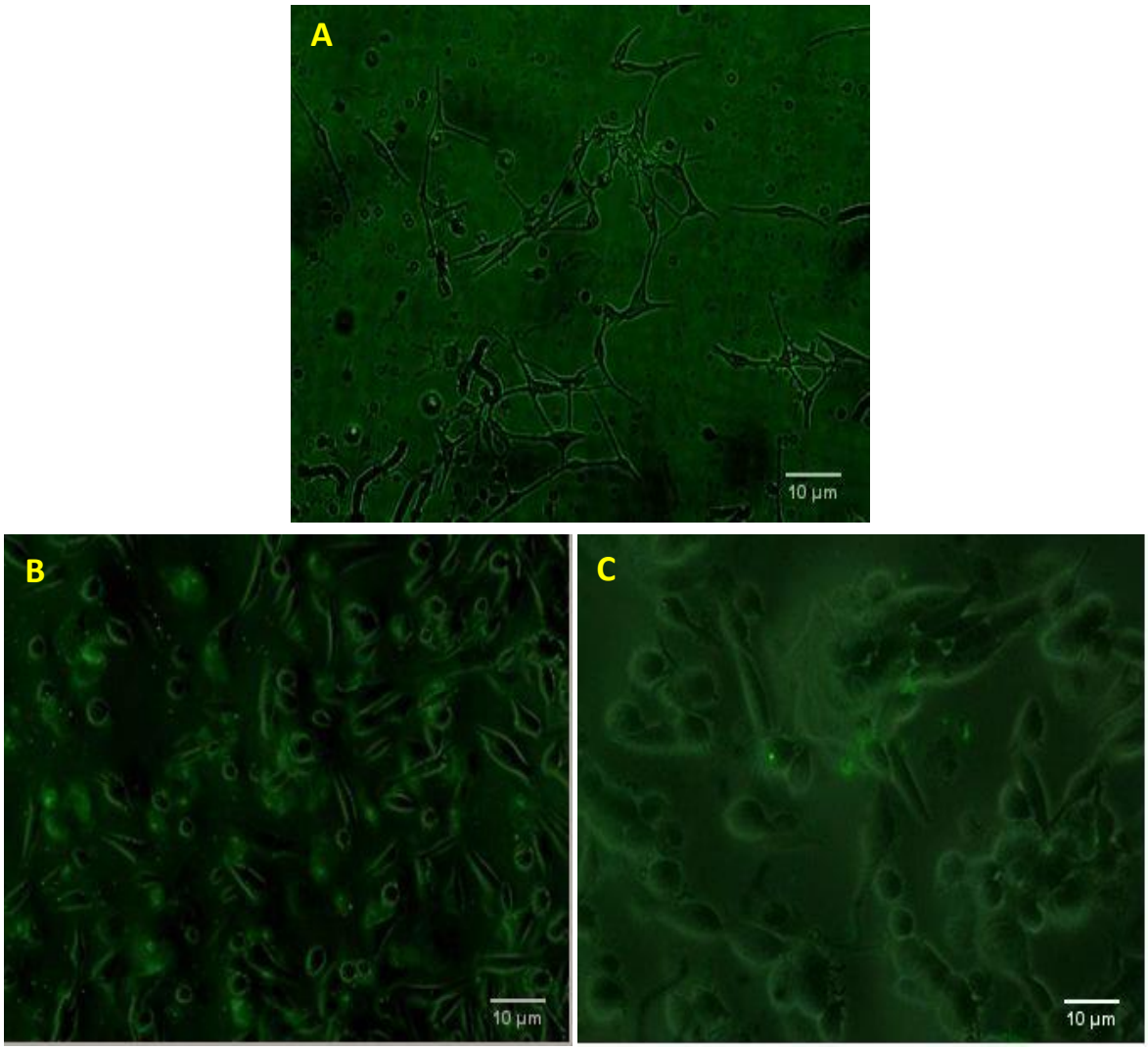


Figure 128. Human healthy melanocyte and melanoma cells incubated with 0.5µM of FITC labelled 1st Domain-Melittin fusion protein overnight and screened under luminescence imaging microscopy. Picture A: Healthy human melanocyte cells (HEMn); Picture B: Primary human melanoma cells (WM-115); Picture C: Metastatic human melanoma cells (WM 266-4).

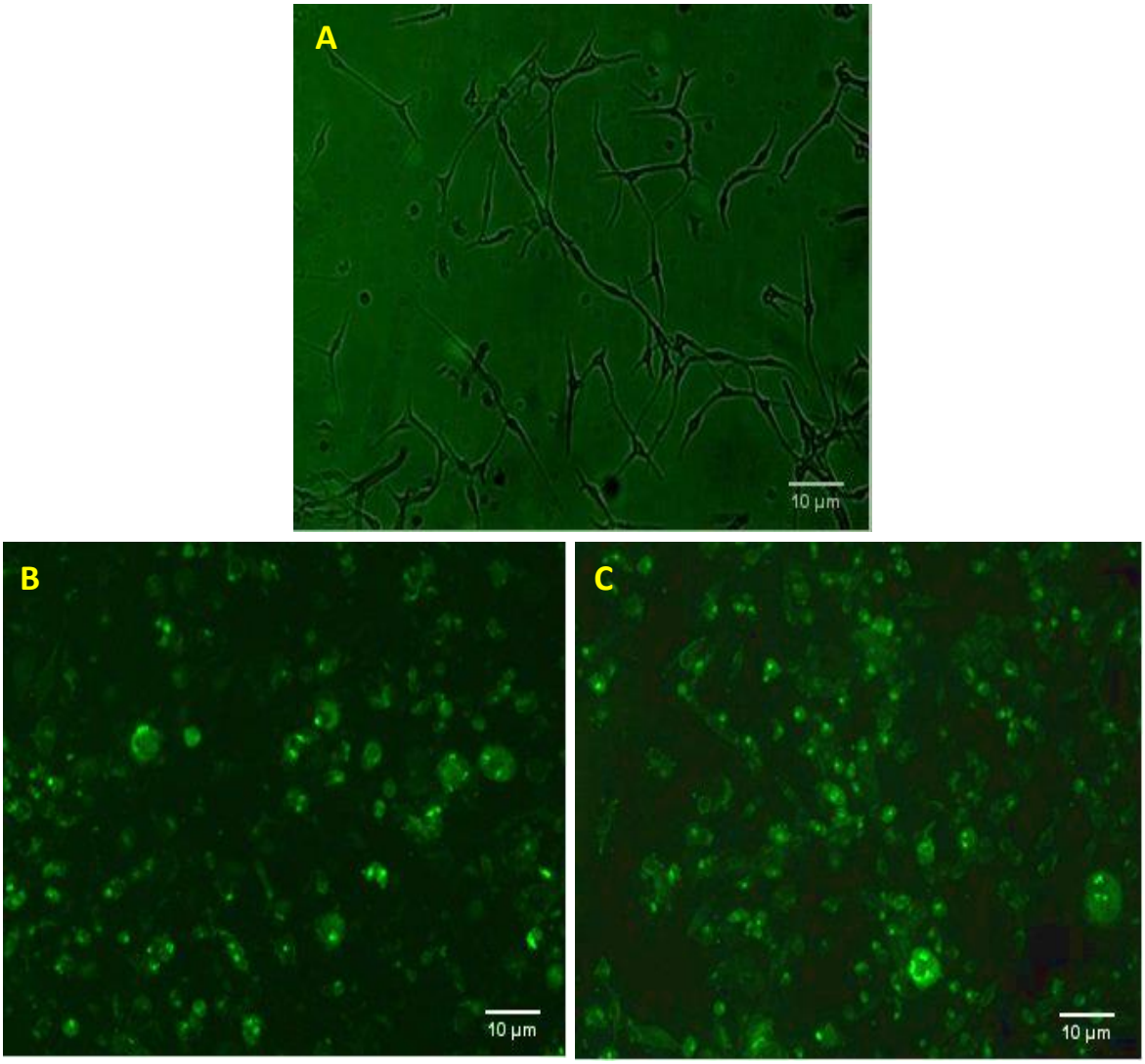


Figure 129. Human healthy melanocyte and melanoma cells incubated with 0.5µM of FITC labelled 1st Domain-Mastoparan fusion protein overnight and screened under luminescence imaging microscopy. Picture A: Healthy human melanocyte cells (HEMn); Picture B: Primary human melanoma cells (WM-115); Picture C: Metastatic human melanoma cells (WM 266-4).

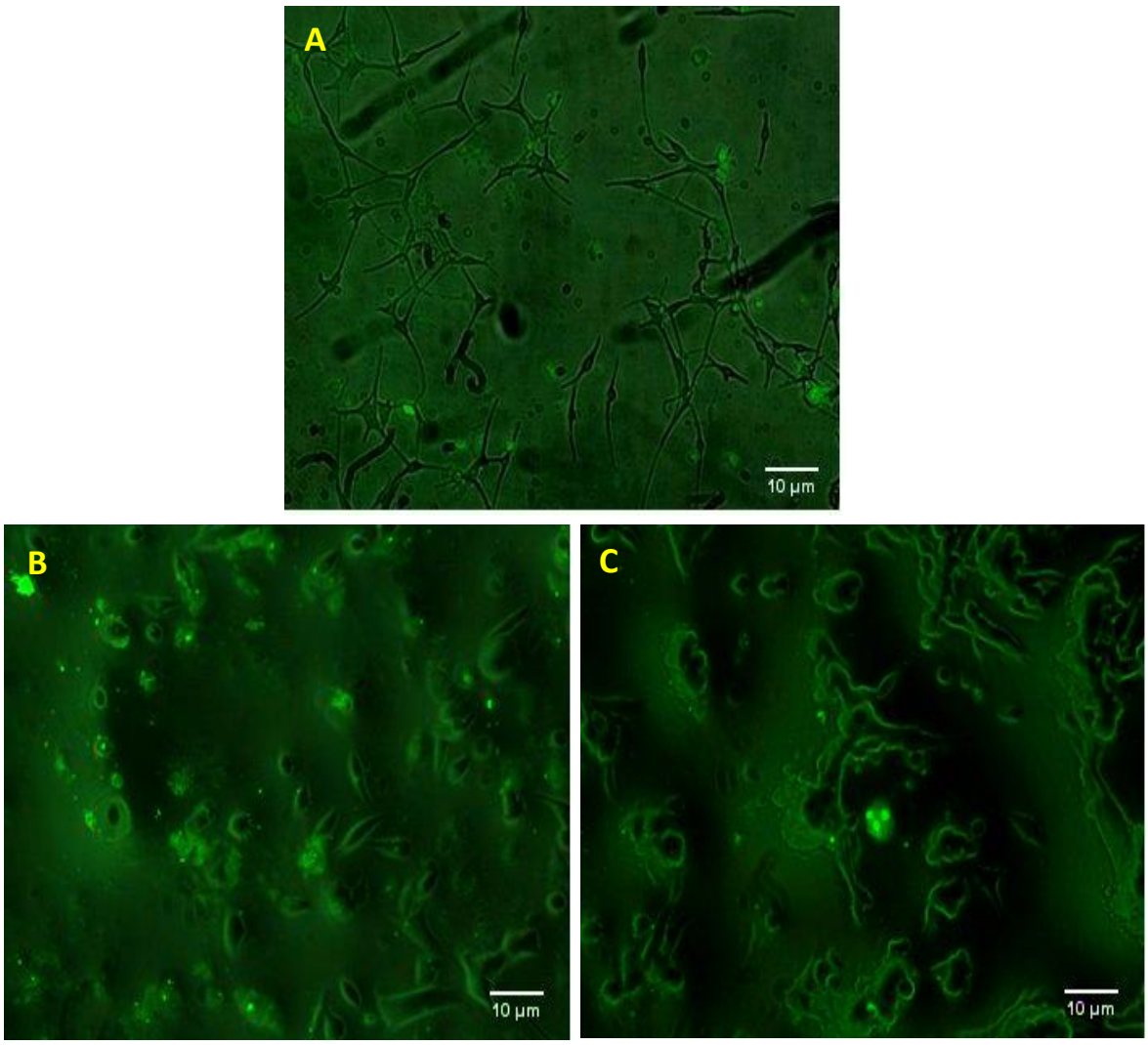


Figure 130. Human healthy melanocyte and melanoma cells incubated with 0.5µM of FITC labelled MLB-Mastoparan 1 fusion protein overnight and screened under luminescence imaging microscopy. Picture A: Healthy human melanocyte cells (HEMn); Picture B: Primary human melanoma cells (WM-115); Picture C: Metastatic human melanoma cells (WM 266-4).

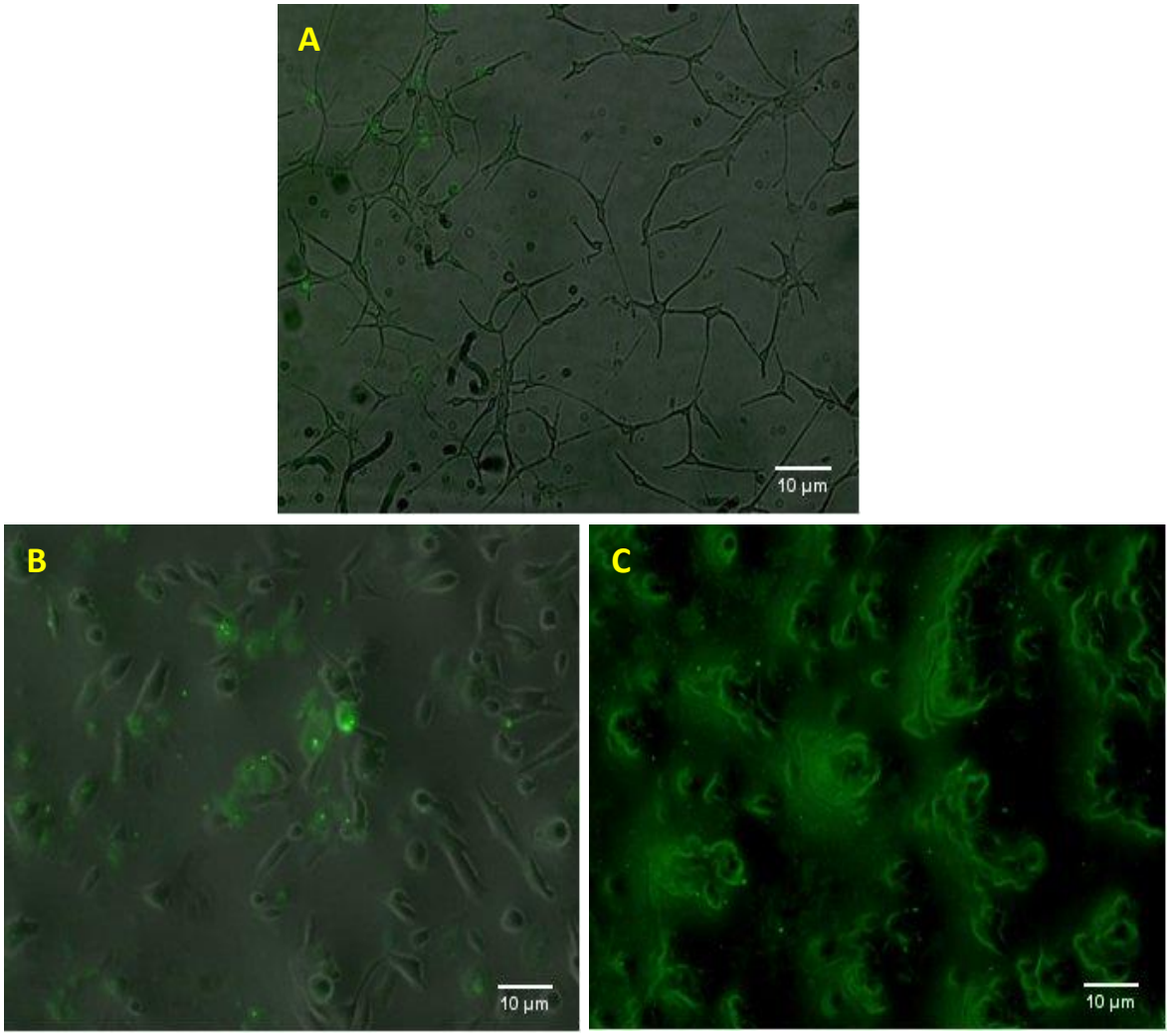


Figure 131. Human healthy melanocyte and melanoma cells incubated with 0.5µM of FITC labelled MLB-mastoparan 2 fusion protein overnight and screened under luminescence imaging microscopy. Picture A: Healthy human melanocyte cells (HEMn); Picture B: Primary human melanoma cells (WM-115); Picture C: Metastatic human melanoma cells (WM 266-4).

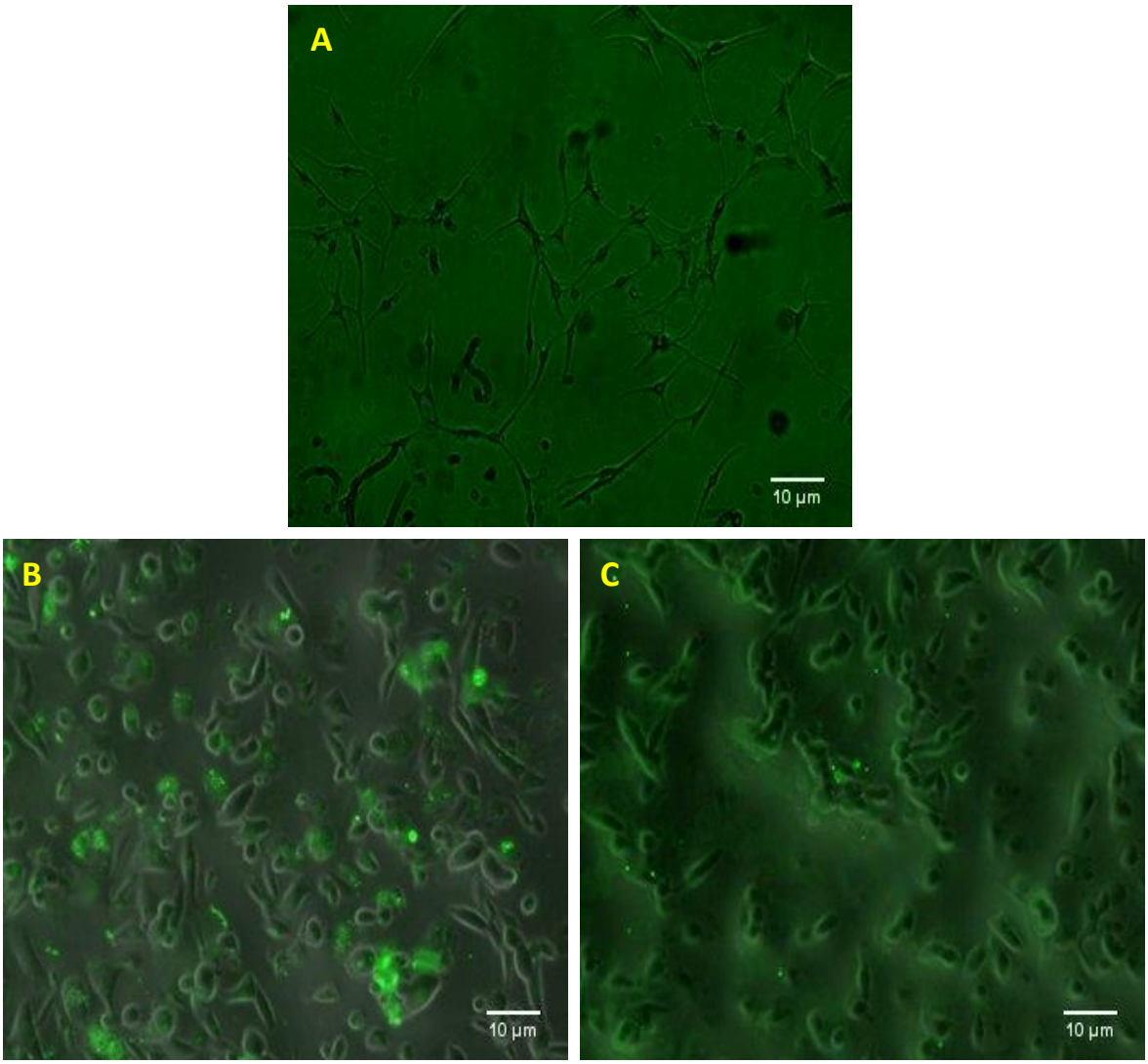


Figure 132. Human healthy melanocyte and melanoma cells incubated with 0.5µM of FITC labelled 2nd Domain-Melittin fusion protein overnight and screened under luminescence imaging microscopy. Picture A: Healthy human melanocyte cells (HEMn); Picture B: Primary human melanoma cells (WM-115); Picture C: Metastatic human melanoma cells (WM 266-4).

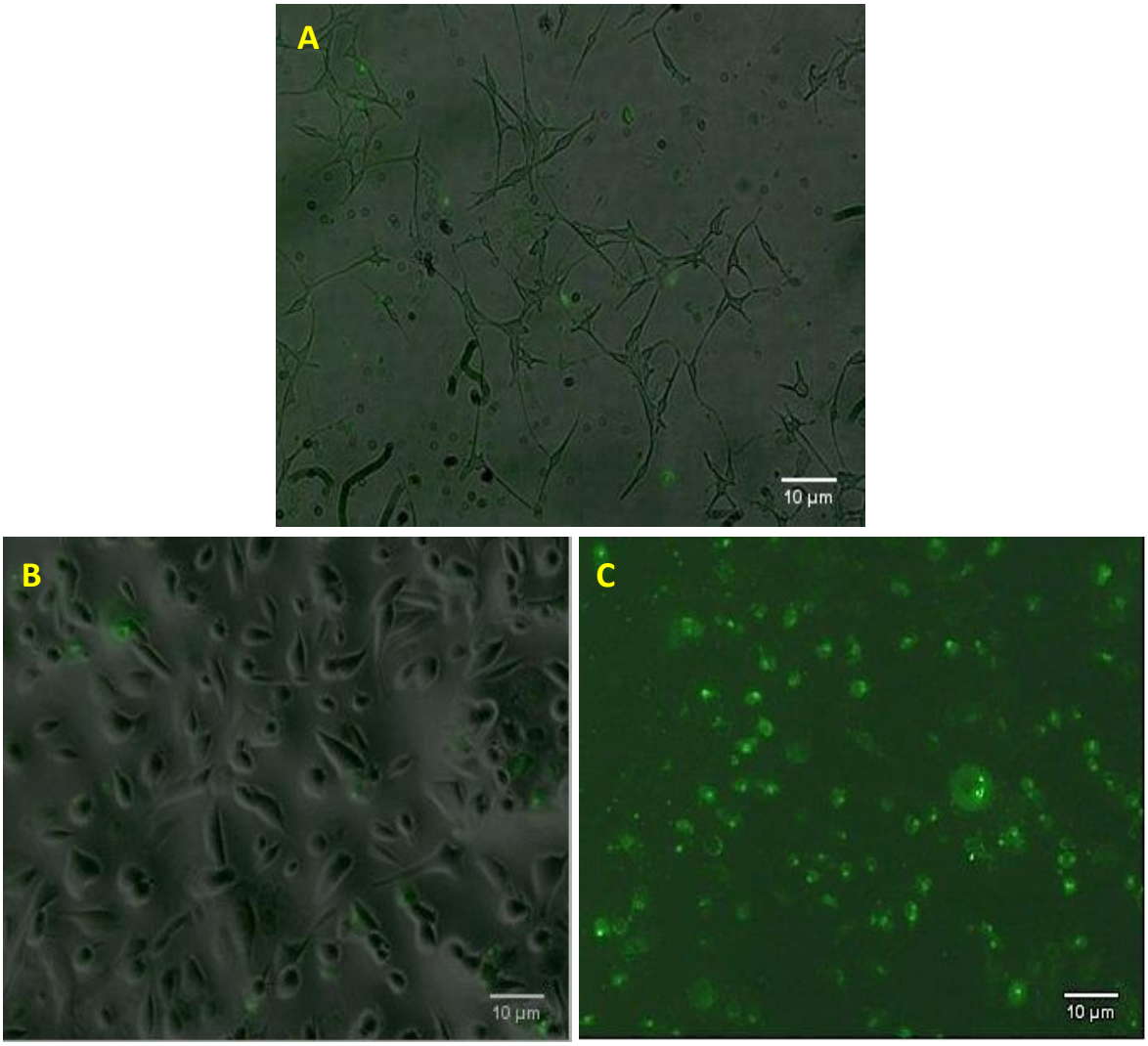


Figure 133. Human healthy melanocyte and melanoma cells incubated with 0.5µM of FITC labelled 2nd Domain-Mastoparan fusion protein overnight and screened under luminescence imaging microscopy. Picture A: Healthy human melanocyte cells (HEMn); Picture B: Primary human melanoma cells (WM-115); Picture C: Metastatic human melanoma cells (WM 266-4).

6.4 Discussion

This study presents the application of the recombinant Lectin-Toxin conjugates in cell culture experiments to assess their binding and cytotoxic activity to different human healthy and cancer cell lines.

In the previous chapters, the recombinant fusion protein produced from MLB fused to GFP were confirmed to have ability to bind to melanoma cells and cross their membrane and accordingly new recombinant fusion proteins were designed to use MLB as carrier vehicle to deliver cytotoxic peptides like melittin and mastoparan to melanoma cells and applied in this study.

The first step of the cell experiments was the selection of cancer cells which was informed by previous knowledge on the availability of sialic acid on the cell surfaces. Hoja-Łukowicz, Link-Lenczowski et al. (2013) carried out a study on two human melanoma cell lines known as WM793 and WM1205Lu which were at primary and metastatic progression stages respectively and showed that both lines have abundant sialic acid on their cell surface oligosaccharides. Furthermore, they observed that the two cell lines have two different sialic acid profiles on their L1CAM oligosaccharide. The primary melanoma cell line (WM793) was shown to have both α2-3 and α2-6 on their L1CAM, while, only α2-6 was observed on the metastatic melanoma cells (WM1205Lu). We believe that these two cell lines can be specifically targeted by the Lectin-Toxin conjugates since the sialic acid binding affinity of MLB fusions was confirmed in the previous chapters. Moreover, we suggested that a stronger binding of the fusion proteins to the metastatic cells (WM 266-4) should be achieved compared to the binding to the primary cells (WM-115) because Müthing, Burg et al. (2002) have shown that the preferential affinity of MLB is to terminally α2-6 sialic acid.

The colorimetric assay (MTS) was carried out to assess the toxicity of the fusion proteins toward human healthy melanocyte (HEMn), human primary melanoma (WM-115) and human metastatic melanoma (WM 266-4) treated with the recombinant fusion proteins. All the fusion proteins showed minor or almost no effect on the healthy cells (HEMn) (Figure 111, 114). This is an indication that in the treatment of HEMn, the designed delivery systems meet the most important criteria of selective delivery systems which is the elimination of healthy cells from their cytotoxic effects. On the other hand, differing levels of cytotoxicity of the fusion proteins were observed with the cancer cells. MLB-Melittin 1 and MLB-Mastoparan 1 fusion constructs showed significant effects on both WM-115 and WM 266-4 melanoma cells, however, the lowest cytotoxic effect observed on both cell lines was obtained with the MLB-Melittin 2 and MLB-Mastoparan 2 fusions, these fusions do not have the Cat-B linker in their structure (Figure 112, 113, 115, 116). This confirms the importance of Cat-B in the activity of the recombinant fusion proteins in which helps the cytotoxic peptide to be quickly released from the lectin peptide and trigger their cytotoxicity because without the linker the toxic peptide remains attached to the lectin peptide in an inactive form and does not have any cytotoxic impact on the cells. The 1st Domain-Melittin fusion also showed limited effects on either melanoma cell lines (106.7%and 98.2%) (Figure 112, 113) although 2nd Domain-Melittin showed slightly higher effects (89.7%and 90.6%) (Figure 115, 116). However significant cytotoxicity was shown by 1st Domain-Mastoparan on both primary and metastatic melanoma cells (83.6% and 83.3%) (Figure 112, 113). Interestingly the effect of the 2nd Domain-Mastoparan fusion was shown to be higher (78.6%) on metastatic melanoma cells (WM 266-4) than on primary melanoma cells (WM-115) (95.1%) (Figure 115, 116), confirming that MLB specifically targets α2-6 sialic acids.

The treatment of WM 266-4 cells with the same concentration (0.5μM) of the fusions but for a longer incubation time (48hours) was seen to potentially increases the cytotoxic effect of all the fusion proteins especially those fusions containing melittin which showed lower cytotoxic effects in the 24hours incubation time (Figure 117, 118). This result agrees with a study conducted by Černe, Erman et al. (2013) and demonstrated that the cytotoxic effect of melittin is time dependent.

Further cell morphological assessment was carried out to assess the effect of the recombinant fusion proteins on WM 266-4 metastatic melanoma cells, and again different levels of activity were shown by the fusion proteins. Deformed cells and cell membrane lysis was detected within the cells treated with all the recombinant fusion proteins (Figure 119-124) but MLB-Mastoparan 1 and 1st Domain-Mastoparan fusion proteins were seen to have the highest impact on the cells as they caused more disruption of the cell membranes and produced more apoptotic bodies around the cells (Figure 121, 122), thus confirming the results with calorimetric assays.

Fluorescein isothiocyanate reagent (FITC) was also used to label the fusion proteins so as to be able to visualise and track their localization during cell binding and internalization. The activity of the FITC labelled fusion proteins were tested with the healthy human melanocyte cells (HEMn) and both primary and metastatic human melanoma cells (WM-115, WM 266-4) and the cells were screened by using florescence microscopy. The fusion proteins were seen to have a minor effect on the healthy cells (HEMn) (Picture A of figure 125-133). However, the binding of the fusion conjugates to the surface of the cancer cells and cell agglutination which is a well-known activity of lectin family were clearly detected within the treated cells. Also we could observe abundant broken cell membranes as well as deformed cells being detected within the treated cell groups. This is believed to be due to their interaction with melittin and mastoparan well known to cause destabilization of the

membranes lipid bilayer and consequently causes membrane lysis as shown in picture B and C of figure 125-133. These results thus agree with Damianoglou, Rodger et al. (2010) and Park, Lee et al. (2000) who previously demonstrated the cellular toxicity of melittin and mastoparan.

Due to the limited amount of highly purified fusion proteins, only one concentration of the proteins was able to be tested with the cancer cells in the time frame of this study. Therefore, to precisely understand the activity of the fusion proteins and their level of cytotoxicity toward cancer cells, different concentrations of the proteins should be tested so as to identify the half maximal inhibitory concentration (IC50) of the fusion proteins and also assess the dose response of the cells (Samant, Blair et al. 2013).

In conclusion, the cytotoxic effect of Lectin-Toxin fusion constructs was tested with human healthy melanocyte and melanoma cell lines. The fusion proteins were confirmed to have almost no cytotoxic effect on the healthy human melanocytes (HEMn). However, some of the fusion proteins showed positive cytotoxicity on human primary (WM-115) and metastatic (WM 266-4) melanoma cells, particularly the fusions containing Mastoparan.

Chapter 7. Summary of Main Findings and Future Research

7.1 Summary of Main Findings

One of the greatest challenges in medicine is producing a drug to work only at the intended targeted site. Despite the fact that modern technology has widely improved the efficiency of the new generation of therapies, limited advances have been achieved in cancer therapies. Inefficient delivery and lack of specificity are two main factors that mainly affect the efficacy of the majority of the existent anti-cancer therapies such as chemotherapies because they limit the effective dose of the drugs reaches the target site and also expose the normal tissues to the large amount of the administrated doses. Therefore, substantial improvement is demanded in anti-cancer therapies especially in drug delivery and selectivity (Estanqueiro, Amaral et al. 2015, Xiao, Morris-Natschke et al. 2016). While promising anti-tumour effects have been detected in some drugs which have poor recognition capacity for cancer cells, recent studies about anti-tumour drugs have mainly focused on the discovery of appropriate carrier vehicles to specifically deliver these drugs to cancer cells. Although a diversity of drug delivery systems has been tried based on different carrier vehicles ranging from macro biomolecules to nanoparticles, the most significant advances have been achieved using anti-bodies to target caner tissues (Rodriguez-Ruiz, Maksimenko et al. 2015, Wei, Chen et al. 2015, Muralidhara, Baid et al. 2016, Shojaei, Tabatabaeian et al. 2016). Despite this, some types of cancers like melanoma have remained without effective therapy (Rek, Krenn et al. 2009, Casi and Neri 2012, Kurrikoff and Suhorut 2012). To date, the most effective drug for the advanced stage of melanoma is a combination of two drugs namely Nivolumab (Humanized IgG4 anti-PD-1 monoclonal antibody) and Ipilimumab (Anti–cytotoxic T-lymphocyte–associated antigen 4 (CTLA-4) antibody) which has recently been approved for clinical use. The drug

acts to up-regulate antitumor immunity but the drug cannot be used with every melanoma patient because of the treatment-related adverse events resulting from the combination of the two drugs (Larkin, Chiarion-Sileni et al. 2015).

Thus, the aim of the research in this thesis was to identify a novel bioactive peptide to deliver effective anti-cancer agents specifically to cancer cells, especially melanoma cells. Initially a range of novel fusion proteins were designed to potentially deliver cytotoxic peptides to cancer cells through targeting the overexpressed sialic acid residues on their surfaces using biochemical, molecular biology and cell culture methodologies. The binding chain (MLB) of Mistletoe lectin (*Viscum album*) isoform one (MLI) was thus chosen to be studied as a carbohydrate binding peptide and to be used as a carrier molecule in the fusion proteins.

The whole isoform one (MLI) of mistletoe lectin which is comprised of two chains known as toxic chain (MLA) and binding chain (MLB) was amplified and cloned so as to be used later in designing different fusion constructs. Despite a few mismatched bases between the cloned sequences and the reference sequences from NCBI GeneBank server, a high quality sequences of both chains were produced so as to be used in the further experiments.

In order to assess the sialic acid binding ability of MLB and identify an optimum orientation to fuse MLB to other peptides without causing major protein-protein interaction that affect the glycan binding ability of MLB, the coding sequence of a muted version of green florescent protein known as GFPuv was fused with MLB. Also, the coding sequence of the binding chain of elderberry (*Sambucus nigra*) lectin (SNA-B) which is a sialic acid binding lectin and snow drop (*Galanthus nivalis*) lectin (GNA) which is a mannose binding lectin were used as a positive and negative control respectively. The lectin sequences were fused to the GFP sequence in both *N*- and *C*- terminal orientations

and their expression level in the fusion constructs were determined according to the expression level of GFP since the correctly expressed GFP can be excited and visualized under standard UV light.

The results presented in chapter 2 showed that the expression of both cloning strategies (*N*- and *C*- terminal) of MLB+GFP and SNA-B+GFP was successfully carried out and the emission of GFP in *E. coli* colonies containing the fusion constructs were clearly visualized under standard UV light. The SDS PAGE analysis also clearly showed the predicted size bands of all the fusion constructs (Figure 49). However, attempts to express GNA+GFP fusion constructs failed potentially due to the toxicity of GNA which caused the formation of inclusion bodies and also slowed down the proliferation of the *E. coli* cells.

The affinity purification conducted to purify the expressed Lectin+GFP fusion proteins shown in chapter 3 demonstrated that the binding chain of both lectins (MLB, SNA-B) preserved their sialic acid binding ability as they successfully bound to the sialic acid residues of fetuin. This was the first experimental attempt to confirm the sialic acid binding affinity of MLB alone without MLA since the previous attempt carried out by Müthing, Meisen et al. (2004) was to assess the sialic acid binding of the whole mistletoe lectin (MLA+MLB) and Wu, Chin et al. (1992) that showed that mistletoe lectin is a galactose binding lectin. The *in vitro* cell culture experiments also confirmed that MLB and SNA-B can potentially bind to the sialic acid on the surface of melanoma cells (WM 266-4) and cross their membranes (Figure 56, 57).

Due to the confirmation of the binding ability of MLB to cancer cell in chapter 3, we decided to use MLB as a carrier vehicle to produce a selective drug delivery system (SDS) to deliver melittin (MLT) and mastoparan (MP) cytotoxic peptides specifically to

melanoma cells. Therefore, the coding sequence of MLB was further analysed (Section 5.2.1) using bioinformatics analysis and consequently two separate putative sugar binding domains were identified in the sequence (Figure 68). The sequence of MLB, melittin and mastoparan were later optimized to *E. coli* tRNA pools so as to optimize their nucleic acid consistency which was previously proven that improve the expression of recombinant fusion constructs in bacteria. Then, a range of fusion constructs was designed from fusing the coding sequence of MLB and its two sugar binding domains to the coding sequence of MLT and MP through Cat-B biodegradable linker which hypothetically helps to keep the cytotoxic peptides in an inactive form and attached to the carrier peptides until they reach the intended site. Finally, the fusion constructs were successfully expressed in *E. coli* and purified. In this study we identified a potential method to effectively express and purify an active and soluble form of MLB fused to melittin and mastoparan peptides which were previously shown by Buhrman, Rayahin et al. (2013) and Peng, Hong et al. (2012) to be insoluble when they are expressed in *E. coli*. our study also yielded eight active fusion proteins to be used in the further cell culture experiments.

Finally, the cell culture experiments were undertaken in chapter 6 to assess the binding and cytotoxic activity of the fusion proteins produced in chapter 5. Two melanoma cell lines (WM-115 and WM 266-4) at primary and metastasis progression stages respectively were selected to be targeted by the fusion proteins because they were previously studied by Hoja-Łukowicz, Link-Lenczowski et al. (2013) who demonstrated that they have sialic acid branches on their surfaces. A healthy melanocyte cell line (HEMn) was also used as a control to assess the ability of the fusion proteins to eliminate the healthy cells from their cytotoxic effects.

The colorimetric assay (MTS) was also carried out to assess the cytotoxic effect of the fusion proteins on the viability of the melanoma cells (WM-115, WM 266-4) and the

healthy melanocyte cells (HEMn) and confirmed the results of the morphological assessments in which showed that majority of the fusion proteins especially MLB-Melittin 1, MLB-Mastoparan 1, 1st Domain-Mastoparan and 2nd Domain-Mastoparan have significant impact on the viability of the melanoma cells, specifically on the metastatic cells which have abundant α2-6 sialic acid on their surface oligosaccharides (Hoja-Łukowicz, Link-Lenczowski et al. 2013) (Figure 112, 113, 115, 116, 117, 118). This impact on the metastatic cells thus confirmed that MLB played a crucial role in delivering the cytotoxic peptides specifically to the melanoma cells since the primary affinity of mistletoe lectin to terminally α2-6 sialic acid was previously confirmed by Müthing, Burg et al. (2002). In contrast, the fusion proteins were seen to have almost no effect on the healthy melanocyte cells (HEMn) which again confirmed the specificity of the fusion proteins toward melanoma cells (Figure 111, 114).

The morphological study revealed that all the fusion proteins containing the Cat-B linker had effects on the morphology of the cancer cells. However, two of the fusion proteins 1st Domain-Mastoparan and MLB-Mastoparan 1 were seen to have the highest impact on the cells morphology because more apoptotic bodies were detected with the cells treated with these two fusion proteins (Figure 121, 122). This is an indication that the binding peptide (MLB) successfully delivered the toxic peptides to the cancer cells and also the toxic peptides preserved their cytotoxic effect in the fusion protein especially mastoparan which caused more cell membranes breakdown.

On the other hand, the labelling of the fusion proteins with fluorescein isothiocyanate reagent (FITC) facilitated visualising and tracking the cell binding and internalization of the fusion proteins and confirmed the lytic ability of the fusion proteins such as those detected in the treatment of the primary (WM-115) and metastatic (WM 266-4) melanoma cells with the Lectin-Toxin fusion proteins which clearly showed that the cell membranes

were lysed by the activity of some of the fusion proteins as shown in picture B and C of figure 125-133.

The ability of MLB to specifically and potentially bind to the sialic acid residues on the surface of cancer cells and its ability to cross their membranes confirmed in the previous chapters supports the idea of using MLB and other potent peptides to specifically deliver anti-cancer agents to cancer cells which subsequently reduce the side effect of the drugs on the healthy tissues and increase the effective amount of drug reaches the target side.

In summary, this thesis has made multiple contributions to knowledge: firstly, the confirmation of sialic acid binding affinity of the binding chain (MLB) of mistletoe lectin alone without the cytotoxic chain (MLA); secondly, confirming the ability of MLB to specifically bind to the sialic acid residues on the surface of cancer cells and also the ability to cross the cell membranes; thirdly, we have confirmed that the overexpressed sialic acid on the surface of melanoma cells (WM-115 and WM 266-4) can be used as a specific biomarker to specifically target the cells by sialic acid binding molecules. Finally, the production of a range of fusion proteins with the ability to specifically target melanoma cells through targeting their cell surface overexpressed sialic acid residues through the binding ability of MLB to sialic acid and inhibit the cells proliferation through the lytic ability of melittin and mastoparan.

7.2 Future Research

7.2.1 Protein Structure

This work, although providing preliminary knowledge about the structure and properties of the binding chain of mistletoe lectin (MLB), should lead to further attempts to analyse the coding sequence of the peptide and its two putative sugar binding chains in order to identify the core amino acids responsible for their sialic acid binding activity. Consequently, shorter coding sequences could be achieved for the sugar binding peptides which facilitates their movement in the biological systems like cell membrane and increases their application in further drug delivery systems (Bechara and Sagan 2013).

7.2.2 Dose Response

Due to time limitations a single concentration of fusion proteins was used in the cell line tests. In order to demonstrate the full therapeutic potential of these fusions a dose-response curve should be generated as per standard drug design protocols (Samant, Blair et al. 2013), this would allow calculation of the minimum amount of therapeutic agent required for a biological effect. Furthermore, a wider range of cancer cell types should be tested as well as moving into tests with solid tissue biopsies (Coulstock, Sosabowski et al. 2013). Flow cytometry should also be conducted on the treated cells in order to accurately quantify the effects of the fusion protein (Sica, Maiuri et al. 2016). Finally, many lectins are known to provoke immune reactions in humans (Hajto, Horváth et al. 2016), the MLB peptide should at a minimum be screened against the database of known human allergens (Gendel and Jenkins 2006).

Ultimately the MLB chain may be linked to a variety of cytotoxic peptides or other cytotoxic conjugates and utilised as a cell specific delivery tool.

7.2.3 Generating Peptide-siRNA Conjugate

In this work, lectin fusion constructs were generated from mistletoe B-chain (glycan binding portion) using only a cathepsin B biodegradable cleavage site to link the targeting 'carrier' peptide (MLB) and a selected cytotoxic peptide.

For future work we propose to synthesise a different linker sequence with the necessary modifications for click chemistry in order to generate a peptide-siRNA conjugate. In order to utilise siRNA as drugs, one potential alternative to the use of liposomal transfection agents is the covalent conjugation of a cell binding and cell-penetrating peptide (CPP), with the intention of imparting on the siRNA an enhanced ability to enter mammalian cells and reach the appropriate RNA target (Turner et al., 2007).

Bibliography

Alley, S. C., N. M. Okeley and P. D. Senter (2010). "Antibody–drug conjugates: targeted drug delivery for cancer." Current Opinion In Chemical Biology **14**(4): 529-537.

Altschul, S. F., W. Gish, W. Miller, E. W. Myers and D. J. Lipman (1990). "Basic local alignment search tool." J Mol Biol **215**(3): 403-410.

Anderson, K., D. Evers and K. G. Rice (2008). Structure and function of mammalian carbohydrate-lectin interactions. Glycoscience, Springer**:** 2445-2482.

Apweiler, R., H. Hermjakob and N. Sharon (1999). "On the frequency of protein glycosylation, as deduced from analysis of the SWISS-PROT database." Biochimica et Biophysica Acta (BBA)-General Subjects **1473**(1): 4-8.

Argiolas, A. and J. J. Pisano (1983). "Facilitation of phospholipase A2 activity by mastoparans, a new class of mast cell degranulating peptides from wasp venom." Journal of Biological Chemistry **258**(22): 13697-13702.

Armstrong, J. S. (2006). "Mitochondria: a target for cancer therapy." British journal of pharmacology **147**(3): 239-248.

Barbieri, L., M. G. Battelli and F. Stirpe (1993). "Ribosome-inactivating proteins from plants." Biochimica et Biophysica Acta (BBA)-Reviews on Biomembranes **1154**(3): 237-282.

Barondes, S. H. (1988). "Bifunctional properties of lectins: lectins redefined." Trends Biochem Sci. **13**: 480-482.

Bate, C., W. Nolan and A. Williams (2016). "Sialic Acid on the Glycosylphosphatidylinositol Anchor Regulates PrP-mediated Cell Signaling and Prion Formation." Journal of Biological Chemistry **291**(1): 160-170.

Bechara, C. and S. Sagan (2013). "Cell-penetrating peptides: 20years later, where do we stand?" FEBS letters **587**(12): 1693-1702.

Beuth, J. and G. Pulverer (1994). Lectin Blocking: New Strategies For The Prevention And Therapy Of Tumor Metastasis And Infectious Diseasesgustav Fischer Verlag. New York, Stuttgart, Jena.

Bies, C., C.-M. Lehr and J. F. Woodley (2004). "Lectin-mediated drug targeting: history and applications." Advanced drug delivery reviews **56**(4): 425-435.

Bies, C., C. M. Lehr and J. F. Woodley (2004). "Lectin-mediated drug targeting: history and applications." Adv. Drug Deliv. Rev. **56**: 425–435.

Bookbinder, L., A. Cheng and J. Bleil (1995). "Tissue-and species-specific expression of sp56, a mouse sperm fertilization protein." Science **269**(5220): 86-89.

Bourne, Y. and C. Cambillau (1993). The role of structural water molecules in protein-saccharide complexes. Water And Biological Macromolecules, Springer**:** 321-337.

Brockhausen, I., H. Schachter and C. R. P. Stanley in Varki A, Esko JD, et al. (2009). O-GalNAc Glycans. Essentials Of Glycobiology. Cold Spring Harbor (NY), Cold Spring Harbor Laboratory Press.

Buhrman, J. S., J. E. Rayahin, L. C. Cook, M. J. Federle and R. A. Gemeinhart (2013). "Active, Soluble Recombinant Melittin Purified by Extracting Insoluble Lysate of Escherichia coli Without Denaturation." Biotechnology Progress **29**(5): 1150-1157.

Büll, C., T. J. Boltje, M. Wassink, A. M. de Graaf, F. L. van Delft, M. H. den Brok and G. J. Adema (2013). "Targeting aberrant sialylation in cancer cells using a fluorinated sialic acid analog impairs adhesion, migration, and in vivo tumor growth." Molecular cancer therapeutics 12(10): 1935-1946.

Büll, C., M. A. Stoel, M. H. den Brok and G. J. Adema (2014). "Sialic acids sweeten a tumor's life." Cancer research 74(12): 3199-3204.

Büssing, A., K. Suzart, J. Bergmann, U. Pfüller, M. Schietzel and K. Schweizer (1996). "Induction of apoptosis in human lymphocytes treated with Viscum album L. is mediated by the mistletoe lectins." Cancer letters **99**(1): 59-72.

Büssing, A., M. Wagner, B. Wagner, G. Stein, M. Schietzel, G. Schaller and U. Pfüller (1999). "Induction of mitochondrial Apo2. 7 molecules and generation of reactive oxygen-intermediates in cultured lymphocytes by the toxic proteins from Viscum album L." Cancer letters **139**(1): 79-88.

Cabral, H., Y. Matsumoto, K. Mizuno, Q. Chen, M. Murakami, M. Kimura, Y. Terada, M. Kano, K. Miyazono and M. Uesaka (2011). "Accumulation of sub-100 nm polymeric micelles in poorly permeable tumours depends on size." Nature nanotechnology **6**(12): 815-823.

Cantor, C. R. and P. R. Schimmel (1980). Biophysical chemistry: Part III: the behavior of biological macromolecules, Macmillan.

Casi, G. and D. Neri (2012). "Antibody–drug conjugates: basic concepts, examples and future perspectives." Journal Of Controlled Release **161**(2): 422-428.

Černe, K., A. Erman and P. Veranič (2013). "Analysis of cytotoxicity of melittin on adherent culture of human endothelial cells reveals advantage of fluorescence microscopy over flow cytometry and haemocytometer assay." Protoplasma **250**(5): 1131-1137.

Chari, R. V. (2007). "Targeted cancer therapy: conferring specificity to cytotoxic drugs." Accounts of chemical research **41**(1): 98-107.

Chen, C.-Y., Y.-H. Jan, Y.-H. Juan, C.-J. Yang, M.-S. Huang, C.-J. Yu, P.-C. Yang, M. Hsiao, T.-L. Hsu and C.-H. Wong (2013). "Fucosyltransferase 8 as a functional regulator of nonsmall cell lung cancer." Proceedings of the National Academy of Sciences **110**(2): 630-635.

Chen, S., T. LaRoche, D. Hamelinck, D. Bergsma, D. Brenner, D. Simeone, R. E. Brand and B. B. Haab (2007). "Multiplexed analysis of glycan variation on native proteins captured by antibody microarrays." Nature methods **4**(5): 437-444.

Cheresh, D. A., R. A. Reisfeld and A. P. Varki (1984). "O-acetylation of disialoganglioside GD3 by human melanoma cells creates a unique antigenic determinant." Science **225**(4664): 844-846.

Corfield, A. P. (2004). Structure/function of O-glycans. Encyclopedia of Genetics, Genomics, Proteomics and Bioinformatics, John Wiley & Sons, Ltd.

Coulstock, E., J. Sosabowski, M. Ovečka, R. Prince, L. Goodall, C. Mudd, A. Sepp, M. Davies, J. Foster and J. Burnet (2013). "Liver-targeting of interferon-alpha with tissue-specific domain antibodies." PLoS one **8**(2): e57263.

Cummings, R. (1999). "Structure and function of the selectin ligand PSGL-1." Brazilian journal of medical and biological research **32**(5): 519-528.

Cummings, R. D. and R. P. McEver (2009). C-type lectins. Essentials of Glycobiology. Cold Spring Harbor (NY), Cold Spring Harbor Laboratory Press.

Dall'Olio, F., N. Malagolini, M. Trinchera and M. Chiricolo (2014). "Sialosignaling: Sialyltransferases as engines of self-fueling loops in cancer progression." Biochimica et Biophysica Acta (BBA)-General Subjects **1840**(9): 2752-2764.

Dall'Olio, F. and D. Trere (1992). "Expression of alpha 2, 6-sialylated sugar chains in normal and neoplastic colon tissues. Detection by digoxigenin-conjugated Sambucus nigra agglutinin." European journal of histochemistry: EJH **37**(3): 257-265.

Damianoglou, A., A. Rodger, C. Pridmore, T. R Dafforn, J. A Mosely, J. M Sanderson and M. R Hicks (2010). "The synergistic action of melittin and phospholipase A2 with lipid membranes: development of linear dichroism for membrane-insertion kinetics." Protein and peptide letters **17**(11): 1351-1362.

Damme, E. J., A. Barre, P. Rougé, F. Leuven and W. J. Peumans (1996). "The NeuAc (α-2, 6)-Gal/GalNAc-Binding Lectin from Elderberry (Sambucus Nigra) Bark, a type-2 Ribosome-Inactivating Protein with an Unusual Specificity and Structure." European journal of biochemistry **235**(1-2): 128-137.

De Azevedo, R. A., C. R. Figueiredo, F. Adilson K, A. L. Matsuo, M. H. Massaoka, N. Girola, A. V. Auada, C. F. Farias, K. F. Pasqualoto and C. P. Rodrigues (2015). "Mastoparan induces apoptosis in B16F10-Nex2 melanoma cells via the intrinsic mitochondrial pathway and displays antitumor activity in vivo." Peptides **68**: 113-119.

Dempsey, C. E. (1990). "The actions of melittin on membranes." Biochimica et Biophysica Acta (BBA)-Reviews on Biomembranes **1031**(2): 143-161.

Ding, X., C. Cartwright, L. Tan, R. Lee and P. Yang (2014). "Mistletoe extract inhibits the proliferation of human hepatocellular carcinoma cells by induction of apoptosis and downregulation of c-MYC." Cancer Research **74**(19 Supplement): 3206-3206.

Drickamer, K. (1992). "Engineering galactose-binding activity into a C-type mannose-binding protein." Nature **12;360(6400)**: 183-186.

Drickamer, K. and M. E. Taylor (1993). "Biology of Animal Lectins." Annu. Rev Cell Biol. **64**: 9-237.

Dunn, R. D., K. M. Weston, T. J. Longhurst, G. G. Lilley, D. E. Rivett, P. J. Hudson and R. L. Raison (1996). "Antigen binding and cytotoxic properties of a recombinant immunotoxin incorporating the lytic peptide, melittin." Immunotechnology **2**(3): 229-240.

Eck, J., M. Langer, B. Möckel, A. Baur, M. Rothe, H. Zinke and H. Lentzen (1999). "Cloning of the mistletoe lectin gene and characterization of the recombinant A-chain." European Journal of Biochemistry **264**(3): 775-784.

Eck, J., M. Langer, B. Möckel, K. Witthohn, H. Zinke and H. Lentzen (1999). "Characterization of recombinant and plant-derived mistletoe lectin and their B-chains." European journal of biochemistry **265**(2): 788-797.

Eifler, R., K. Pfüller, W. Göckeritz and U. Pfüller (1994). Improved Procedures For Isolation Of Mistletoe Lectins And Their Subunits: Lectin Pattern Of The European Mistletoe. Lectins Biology Biochemistry Clinical Biochemistry. New Delhi, India, Wiley Eastern. **9:** 144-151.

Eklund, E. A. and H. H. Freeze (2006). "The congenital disorders of glycosylation: a multifaceted group of syndromes." NeuroRx **3**(2): 254-263.

Esko, J. D., K. Kimata and U. Lindahl (2009). Proteoglycans and sulfated glycosaminoglycans. Essentials of Glycobiology. Cold Spring Harbor (NY), Cold Spring Harbor Laboratory Press.

Essler, M. and E. Ruoslahti (2002). "Molecular specialization of breast vasculature: a breast-homing phage-displayed peptide binds to aminopeptidase P in breast vasculature." Proceedings of the National Academy of Sciences **99**(4): 2252-2257.

Estanqueiro, M., M. H. Amaral, J. ConceiÇÃo and J. M. S. Lobo (2015). "Evolution of liposomal carriers intended to anticancer drug delivery: An overview." International Journal of Current Pharmaceutical Research **7**(4): 26-33.

Fernández-Rodríguez, J., C. Feijoo-Carnero, A. Merino-Trigo, M. Pâez De la Cadena, F. J. Rodríguez-Berrocal, A. de Carlos, M. Butrón and V. S. Martínez-Zorzano (2000). "Immunohistochemical analysis of sialic acid and fucose composition in human colorectal adenocarcinoma." Tumor Biology **21**(3): 153-164.

Ferreira, S., J. Vasconcelos, C. Cavalcanti, M. Rêgo and E. Beltrao (2013). "Sialic acid differential expression in non-melanoma skin cancer biopsies." Medical molecular morphology **46**(4): 198-202.

Fitches, E., N. Audsley, J. A. Gatehouse and J. P. Edwards (2002). "Fusion proteins containing neuropeptides as novel insect contol agents: snowdrop lectin delivers fused allatostatin to insect haemolymph following oral ingestion." Insect Biochemistry and Molecular Biology **32**(12): 1653-1661.

Fitches, E., S. D. Woodhouse, J. P. Edwards and J. A. Gatehouse (2001). "In vitro and in vivo binding of snowdrop (Galanthus nivalis agglutinin; GNA) and jackbean (Canavalia ensiformis; Con A) lectins within tomato moth (Lacanobia oleracea) larvae; mechanisms of insecticidal action." Journal of Insect Physiology **47**(7): 777-787.

Fleige, S. and M. W. Pfaffl (2006). "RNA integrity and the effect on the real-time qRT-PCR performance." Molecular aspects of medicine **27**(2): 126-139.

Fouquaert, E., D. F. Smith, W. J. Peumans, P. Proost, J. Balzarini, S. N. Savvides and E. J. Van Damme (2009). "Related lectins from snowdrop and maize differ in their carbohydrate-binding specificity." Biochemical and biophysical research communications **380**(2): 260-265.

Fredriksson, S.-Å., M. Podbielska, B. Nilsson, B. Krotkiewska, E. Lisowska and H. Krotkiewski (2010). "ABH blood group antigens in N-glycan of human glycophorin A." Archives of biochemistry and biophysics **498**(2): 127-135.

Fukuda, M. (1994). Molecular Glycobiology. New York, Oxford University Press.

Fukushima, N., M. Kohno, T. Kato, S. Kawamoto, K. Okuda, Y. Misu and H. Ueda (1998). "Melittin, a metabostatic peptide inhibiting G s activity." Peptides **19**(5): 811-819.

Furukawa, J.-i., M. Tsuda, K. Okada, T. Kimura, J. Piao, S. Tanaka and Y. Shinohara (2015). "Comprehensive Glycomics of a Multistep Human Brain Tumor Model Reveals Specific Glycosylation Patterns Related to Malignancy." PloS one **10**(7): e0128300.

Fuster, M. M. and J. D. Esko (2005). "The sweet and sour of cancer: glycans as novel therapeutic targets." Nature Reviews Cancer **5**(7): 526-542.

Gabius, H. and S. Gabius (1997). Glyco-Sciences Status and prespectives. Weinheim, Chapman & Hall.

Gabius, H., H. Walzel, S. Joshi, J. Kruip, S. Kojima, V. Gerke, H. Kratzin and S. Gabius (1991). "The immunomodulatory beta-galactoside-specific lectin from mistletoe: partial sequence analysis, cell and tissue binding, and impact on intracellular biosignalling of monocytic leukemia cells." Anticancer research **12**(3): 669-675.

Gajski, G. and V. Garaj-Vrhovac (2013). "Melittin: A lytic peptide with anticancer properties." Environmental Toxicology and Pharmacology **36**(2): 697-705.

Gallagher, J. T., J. E. Turnbull and M. Lyon (1992). Heparan sulphate proteoglycans: molecular organisation of membrane-associated species and an approach to polysaccharide sequence analysis. Heparin and Related Polysaccharides, Springer**:** 49-57.

Galun, D., W. Tröger and M. Reif (2015). "Mistletoe extract therapy versus no antineoplastic therapy: a randomized clinical trial on overall survival and quality of life in pancreatic cancer patients." Phytomedicine(22): S6.

Gambaryan, A., S. Yamnikova, D. Lvov, A. Tuzikov, A. Chinarev, G. Pazynina, R. Webster, M. Matrosovich and N. Bovin (2005). "Receptor specificity of influenza viruses from birds and mammals: new data on involvement of the inner fragments of the carbohydrate chain." Virology **334**(2): 276-283.

Gamerith, G., A. Amann, B. Schenk, T. Auer, J. Huber, K. Cima, H. Lentzen, J. Löffler-Ragg, H. Zwierzina and W. Hilbe (2014). "P24. Aviscumine enhances NK-cytotoxicity against tumor cells." Journal for immunotherapy of cancer **2**(Suppl 2): P15.

Garcia-Pino, A., L. Buts, L. Wyns, A. Imberty and R. Loris (2007). "How a plant lectin recognizes high mannose oligosaccharides." Plant physiology **144**(4): 1733-1741.

Gasteiger, E., A. Gattiker, C. Hoogland, I. Ivanyi, R. D. Appel and A. Bairoch (2003). "ExPASy: the proteomics server for in-depth protein knowledge and analysis." Nucleic acids research **31**(13): 3784-3788.

Gendel, S. M. and J. A. Jenkins (2006). "Allergen sequence databases." Mol Nutr Food Res **50**(7): 633-637.

Gessner, P., S. Riedl, A. Quentmaier and W. Kemmner (1993). "Enhanced activity of CMP-NeuAc: Galβ1-4GlcNAc: α2, 6-sialyltransferase in metastasizing human colorectal tumor tissue and serum of tumor patients." Cancer letters **75**(3): 143-149.

Ghawana, S., A. Paul, H. Kumar, A. Kumar, H. Singh, P. K. Bhardwaj, A. Rani, R. S. Singh, J. Raizada, K. Singh and S. Kumar (2011). "An RNA isolation system for plant tissues rich in secondary metabolites." BMC Research Notes **4**(1): 1-5.

Ghazarian, H., B. Idoni and S. B. Oppenheimer (2011). "A glycobiology review: carbohydrates, lectins and implications in cancer therapeutics." Acta histochemica **113**(3): 236-247.

Ghazarian, H., B. Idoni and S. B. Oppenheimer (2011). "A glycobiology review: carbohydrates, lectins, and implications in cancer therapeutics." Acta Histochem. **113(3)**: 236–247.

Girbes, T., J. M. Ferreras, F. J. Arias and F. Stirpe (2004). "Description, distribution, activity and phylogenetic relationship of ribosome-inactivating proteins in plants, fungi and bacteria." Mini Reviews In Medicinal Chemistry **4**(5): 461-476.

Goldstein, I. J., R. C. Hughes, M. Monsigny, T. Osawa and N. Sharon (1980). "What should be called a lectin?" Nature **285**: 66.

Griffin, C. C., R. J. Linhardt, C. L. Van Gorp, T. Toida, R. E. Hileman, R. L. Schubert and S. E. Brown (1995). "Isolation and characterization of heparan sulfate from crude porcine intestinal mucosal peptidoglycan heparin." Carbohydrate research **276**(1): 183-197.

Grześk, G., B. Malinowski, E. Grześk, M. Wiciński and K. Szadujkis‐Szadurska (2014). "Direct regulation of vascular smooth muscle contraction by mastoparan‐7." Biomedical reports **2**(1): 34-38.

Guarner, F. and J. R. Malagelada (2003). "Gut flora in health and disease." Lancet **361**: 512–519.

Guicciardi, M. E., M. Leist and G. J. Gores (2004). "Lysosomes in cell death." Oncogene **23**(16): 2881-2890.

Guzińska-Ustymowicz, K., B. Zalewski, I. Kasacka, Z. Piotrowski and E. Skrzydlewska (2004). "Activity of cathepsin B and D in colorectal cancer: relationships with tumour budding." Anticancer research **24**(5A): 2847-2852.

Guzmán-Partida, A., M. Robles-Burgueno, M. Ortega-Nieblas and I. Vázquez-Moreno (2004). "Purification and characterization of complex carbohydrate specific isolectins from wild legume seeds: Acacia constricta is (vinorama) highly homologous to Phaseolus vulgaris lectins." Biochimie **86**(4): 335-342.

Habermann, E. (1972). "Bee and wasp venoms." Science **177**(4046): 314-322.

Hagens, C. v., R. Klein, A. Staudt, A. Glenz, B. Reinhard-Hennch, A. Loewe-Mesch, J. J. Kuehn, U. Abel, J. Munzinger and A. Schneeweiss (2015). "Mistletoe lectin-(ML-1) and viscotoxin (VT) antibodies during therapy with Viscum album in patients with breast cancer." Phytomedicine **22**: S27.

Hajto, T., A. Horváth and S. Papp (2016). "Improvement of Quality of Life in Tumor Patients after an Immunomodulatory Treatment with Standardized Mistletoe Lectin and Arabinoxylan Plant Extracts." Int J Neurorehabilitation **3**(205): 2376-0281.1000205.

Hajto, T., K. Hostanska and H.-J. Gabius (1989). "Modulatory potency of the β-galactoside-specific lectin from mistletoe extract (Iscador) on the host defense system in vivo in rabbits and patients." Cancer Research **49**(17): 4803-4808.

Hajtò, T., F. Krisztina, A. Ildikò, P. Zsolt, B. Pèter, N. Pèter and P. Pàl (2007). "Unexpected different binding of mistletoe lectins from plant extracts to immobilized lactose and N-acetylgalactosamine." Analytical chemistry insights **1**: 43.

Hakomori, S. (2002). "Glycosylation defining cancer malignancy: new wine in an old bottle." Proceedings of the National Academy of Sciences **99**(16): 10231-10233.

Haltner, E., J. Easson and C. M. Lehr (1997). "Lectins and bacterial invasion factors for controlling endo- and transcytosis of bioadhesive drug carrier systems." Eur. J. Pharm. Biopharm. **44**: 3– 13.

Hamblett, K. J., P. D. Senter, D. F. Chace, M. M. Sun, J. Lenox, C. G. Cerveny, K. M. Kissler, S. X. Bernhardt, A. K. Kopcha and R. F. Zabinski (2004). "Effects of drug loading on the antitumor activity of a monoclonal antibody drug conjugate." Clinical Cancer Research **10**(20): 7063-7070.

Hamelryck, T. W., R. Loris, J. Bouckaert, M.-H. Dao-Thi, G. Strecker, A. Imberty, E. Fernandez, L. Wyns and M. E. Etzler (1999). "Carbohydrate binding, quaternary structure and a novel hydrophobic binding site in two legume lectin oligomers fromDolichos biflorus." Journal of molecular biology **286**(4): 1161-1177.

Handel, T., Z. Johnson, S. Crown, E. Lau, M. Sweeney and A. Proudfoot (2005). "Regulation of protein function by glycosaminoglycans-as exemplified by chemokines." Annu. Rev. Biochem. **74**: 385-410.

Harris, R. J., C. K. Leonard, A. W. Guzzetta and M. W. Spellman (1991). "Tissue plasminogen activator has an O-linked fucose attached to Threonine-61 in the epidermal growth factor domain." Biochemistry **2311**: 30.

Heinrich, E. L., L. A. Welty, L. R. Banner and S. B. Oppenheimer (2005). "Direct targeting of cancer cells: a multiparameter approach." Acta Histochem. **107**: 335–344.

Heldin, C.-H., K. Rubin, K. Pietras and A. Östman (2004). "High interstitial fluid pressure—an obstacle in cancer therapy." Nature Reviews Cancer **4**(10): 806-813.

Hengen, P. N. (1995). "Purification of His-Tag fusion proteins from Escherichia coli." Trends in biochemical sciences **20**(7): 285-286.

Hibberd, A. D., P. R. Trevillian, D. A. Clark, P. Mcelduff and W. B. Cowden (2012). "The effects of Castanospermine, an oligosaccharide processing inhibitor, on mononuclear/endothelial cell binding and the expression of cell adhesion molecules." Transplant immunology **27**(1): 39-47.

Higashijima, T., J. Burnier and E. Ross (1990). "Regulation of Gi and Go by mastoparan, related amphiphilic peptides, and hydrophobic amines. Mechanism and structural determinants of activity." Journal of Biological Chemistry **265**(24): 14176-14186.

Hileman, R. E., J. R. Fromm, J. M. Weiler and R. J. Linhardt (1998). "Glycosaminoglycan-protein interactions: definition of consensus sites in glycosaminoglycan binding proteins." Bioessays **20**(2): 156-167.

Hirai, Y., T. Yasuhara, H. Yoshida and T. Nakajima (1981). A new mast-cell degranulating peptide, mastoparan-m, in the venom of the hornet vespa-mandarinia, Biomed Res Found Kanda Po Box 182 Chiyodaku. **2:** 447-449.

Hoja-Łukowicz, D., P. Link-Lenczowski, A. Carpentieri, A. Amoresano, E. Pocheć, K. A. Artemenko, J. Bergquist and A. Lityńska (2013). "L1CAM from human melanoma carries a novel type of N-glycan with Galβ1-4Galβ1-motif. Involvement of N-linked glycans in migratory and invasive behaviour of melanoma cells." Glycoconjugate journal **30**(3): 205-225.

Hollingsworth, M. A. and B. J. Swanson (2004). "Mucins in cancer: protection and control of the cell surface." Nature Reviews Cancer **4**(1): 45-60.

Hoshino, Y., T. Urakami, T. Kodama, H. Koide, N. Oku, Y. Okahata and K. J. Shea (2009). "Design of synthetic polymer nanoparticles that capture and neutralize a toxic peptide." Small **5**(13): 1562-1568.

Huang, C., H. Jin, Y. Qian, S. Qi, H. Luo, Q. Luo and Z. Zhang (2013). "Hybrid melittin cytolytic peptide-driven ultrasmall lipid nanoparticles block melanoma growth in vivo." ACS nano **7**(7): 5791-5800.

Hussein, A., Z. Nabil, S. Zalat and M. Rakha (2001). "Comparative study of the venoms from three species of bees: effects on heart activity and blood." Journal of natural toxins **10**(4): 343-357.

Igl, W., O. Polašek, O. Gornik, A. Knežević, M. Pučić, M. Novokmet, J. Huffman, C. Gnewuch, G. Liebisch and P. M. Rudd (2011). "Glycomics meets lipidomics-associations of N-glycans with classical lipids, glycerophospholipids, and sphingolipids in three European populations." Molecular BioSystems **7**(6): 1852-1862.

Irani, K., Y. Xia, J. L. Zweier, S. J. Sollott, C. J. Der, E. R. Fearon, M. Sundaresan, T. Finkel and P. J. Goldschmidt-Clermont (1997). "Mitogenic signaling mediated by oxidants in Ras-transformed fibroblasts." Science **275**(5306): 1649-1652.

Jang, M.-H., M.-C. Shin, S. Lim, S.-M. Han, H.-J. Park, I. Shin, J.-S. Lee, K.-A. Kim, E.-H. Kim and C.-J. Kim (2003). "Bee venom induces apoptosis and inhibits expression of cyclooxygenase-2 mRNA in human lung cancer cell line NCI-H1299." Journal of pharmacological sciences **91**(2): 95-104.

Jedeszko, C. and B. F. Sloane (2004). "Cysteine cathepsins in human cancer." Biological chemistry **385**(11): 1017-1027.

Jensen, P. H., N. G. Karlsson, D. Kolarich and N. H. Packer (2012). "Structural analysis of N-and O-glycans released from glycoproteins." Nature protocols **7**(7): 1299-1310.

Jiang, S.-Y., Z. Ma and S. Ramachandran (2010). "Evolutionary history and stress regulation of the lectin superfamily in higher plants." BMC evolutionary biology **10**(1): 1.

Jin, H., C. Li, D. Li, M. Cai, Z. Li, S. Wang, X. Hong and B. Shi (2013). "Construction and Characterization of a CTLA-4-Targeted scFv–Melittin Fusion Protein as a Potential Immunosuppressive Agent for Organ Transplant." Cell biochemistry and biophysics **67**(3): 1067-1074.

Joshi, H. J., C. Steentoft, T.-B. S. Katrine, S. Y. Vakhrushev, H. H. Wandall and H. Clausen (2015). Protein O-GalNAc Glycosylation: most complex and differentially regulated PTM. Glycoscience: Biology and Medicine, Springer**:** 1049-1064.

Ju, T., G. S. Lanneau, T. Gautam, Y. Wang, B. Xia, S. R. Stowell, M. T. Willard, W. Wang, J. Y. Xia and R. E. Zuna (2008). "Human tumor antigens Tn and sialyl Tn arise from mutations in Cosmc." Cancer research **68**(6): 1636-1646.

Kaether, C. and H.-H. Gerdes (1995). "Visualization of protein transport along the secretory pathway using green fluorescent protein." FEBS letters **369**(2-3): 267-271.

Katoh, N. (2002). "Inhibition by melittin of phosphorylation by protein kinase C of annexin I from cow mammary gland." Journal of Veterinary Medical Science **64**(9): 779-783.

Kaus, K., J. W. Lary, J. L. Cole and R. Olson (2014). "Glycan Specificity of the Vibrio vulnificus Hemolysin Lectin Outlines Evolutionary History of Membrane Targeting by a Toxin Family." Journal of Molecular Biology **426**(15): 2800-2812.

Kilpatrick, D. C. (2002). "Animal lectins: a historical introduction and overview." Biochimica et Biophysica Acta (BBA)-General Subjects **1572**(2): 187-197.

Kim, Y., M. Son, E.-Y. Noh, S. Kim, C. Kim, J.-H. Yeo, C. Park, K. W. Lee and W. Y. Bang (2016). "MP-V1 from the Venom of Social Wasp Vespula vulgaris Is a de Novo Type of Mastoparan that Displays Superior Antimicrobial Activities." Molecules **21**(4): 512.

Kohla, G., E. Stockfleth and R. Schauer (2002). "Gangliosides with O-acetylated sialic acids in tumors of neuroectodermal origin." Neurochemical research **27**(7-8): 583-592.

Kolasińska, E., M. Przybyło, M. Janik and A. Lityńska (2016). "Towards understanding the role of sialylation in melanoma progression." Acta biochimica Polonica **63**(1221).

Kolatkar, A. R., A. K. Leung, R. Isecke, R. Brossmer, K. Drickamer and W. I. Weis (1998). "Mechanism of N-acetylgalactosamine binding to a C-type animal lectin carbohydrate-recognition domain." Journal of Biological Chemistry **273**(31): 19502-19508.

Kopeček, J., P. Kopečková, T. Minko, Z.-R. Lu and C. Peterson (2001). "Water soluble polymers in tumor targeted delivery." Journal of Controlled Release **74**(1): 147-158.

Krauspenhaar, R., S. Eschenburg, M. Perbandt, V. Kornilov, N. Konareva, I. Mikailova, S. Stoeva, R. Wacker, T. Maier and T. Singh (1999). "Crystal Structure of Mistletoe Lectin I fromViscum album." Biochemical and biophysical research communications **257**(2): 418-424.

Kresge, N., R. D. Simoni and R. L. Hil (2010). "Historical perspectives: Glycobiology and carbohydrates." The Journal Of Biological Chemistry **285**(NO. 5): 3524.

Kudo, T., Y. Ikehara, A. Togayachi, K. Morozumi, M. Watanabe, M. Nakamura, S. Nishihara and H. Narimatsu (1998). "Up-regulation of a set of glycosyltransferase genes in human colorectal cancer." Laboratory investigation; a journal of technical methods and pathology **78**(7): 797-811.

Kurrikoff, K. and J. Suhorut (2012). "Cell-Penetrating Peptides in Cancer Targeting." Drug Delivery in Oncology: From Basic Research to Cancer Therapy: 1187-1217.

Kyhse-Andersen, J. (1984). "Electroblotting of multiple gels: a simple apparatus without buffer tank for rapid transfer of proteins from polycrylamide to nitrocellulose." Journal of biochemical and biophysical methods **10**(3): 203-209.

Ladokhin, A. S. and S. H. White (1999). "Folding of amphipathic α-helices on membranes: energetics of helix formation by melittin." Journal of molecular biology **285**(4): 1363-1369.

Lakhtin, V., S. Afanas' ev, V. Aleshkin, I. Nesvizhskiĭ, M. Lakhtin, V. Shubin, I. Cherepanova and V. Pospelova (2008). "Classification of lectins as universal regulator molecules of biological systems." Vestnik Rossiiskoi Akademii Meditsinskikh Nauk/Rossiiskaia Akademiia Meditsinskikh Nauk(3): 36-43.

Lakhtin, V., M. Lakhtin and V. Alyoshkin (2011). "Lectins of living organisms. The overview." Anaerobe **17**(6): 452-455.

Lamblin, G., S. Degroote, J.-M. Perini, P. Delmotte, A. Scharfman, M. Davril, J.-M. Lo-Guidice, N. Houdret, V. Dumur and A. Klein (2001). "Human airway mucin glycosylation: a combinatory of carbohydrate determinants which vary in cystic fibrosis." Glycoconjugate journal **18**(9): 661-684.

Lankelma, J. M., D. M. Voorend, T. Barwari, J. Koetsveld, A. H. Van der Spek, A. P. De Porto, G. Van Rooijen and C. J. Van Noorden (2010). "Cathepsin L, target in cancer treatment?" Life sciences **86**(7): 225-233.

Lannoo, N., W. J. Peumans, E. V. Pamel, R. Alvarez, T.-C. Xiong, G. Hause, C. Mazars and E. J. Van Damme (2006). "Localization and in vitro binding studies suggest that the cytoplasmic/nuclear tobacco lectin can interact in situ with high-mannose and complex N-glycans." FEBS letters **580**(27): 6329-6337.

Larkin, J., V. Chiarion-Sileni, R. Gonzalez, J. J. Grob, C. L. Cowey, C. D. Lao, D. Schadendorf, R. Dummer, M. Smylie and P. Rutkowski (2015). "Combined nivolumab and ipilimumab or monotherapy in untreated melanoma." N Engl J Med **2015**(373): 23-34.

Lee, R. T., H. Gabius and Y. C. Lee (1992). "Ligand binding characteristics of the major mistletoe lectin." Journal of Biological Chemistry **267**(33): 23722-23727.

Lentzen, H. and K. Witthohn (2016). Drug containing recombinant mistletoe lectins for treating malignant melanoma, Google Patents.

Li, C. (2002). "Poly (L-glutamic acid)–anticancer drug conjugates." Advanced drug delivery reviews **54**(5): 695-713.

Li, N., H. Xu, K. Fan, X. Liu, J. Qi, C. Zhao, P. Yin, L. Wang, Z. Li and X. Zha (2014). "Altered β1, 6-GlcNAc branched N-glycans impair TGF-β-mediated Epithelial-to-Mesenchymal Transition through Smad signalling pathway in human lung cancer." Journal of cellular and molecular medicine **18**(10): 1975-1991.

Linhardt, R. J., S. A. Ampofo, J. Fareed, D. Hoppensteadt, J. Folkman and J. B. Mulliken (1992). "Isolation and characterization of human heparin." Biochemistry **31**(49): 12441-12445.

Lis, H. and N. Sharon (1986). "Lectin in molecular and tools." Annu Rev. Biochem **67**: 55:35.

Lis, H. and N. Sharon (1998). "Lectins: carbohydrate-specific proteins that mediate cellular recognition." Chemical reviews **98**(2): 637-674.

Lise, M., C. Belluco, S. P. Perera, R. Patel, P. Thomas and A. Ganguly (2000). "Clinical correlations of α2, 6-sialyltransferase expression in colorectal cancer patients." Hybridoma **19**(4): 281-286.

Liu, S., M. Yu, Y. He, L. Xiao, F. Wang, C. Song, S. Sun, C. Ling and Z. Xu (2008). "Melittin prevents liver cancer cell metastasis through inhibition of the Rac1-dependent pathway." Hepatology **47**(6): 1964-1973.

Liu, X., D. Chen, L. Xie and R. Zhang (2002). "Effect of honey bee venom on proliferation of K1735M2 mouse melanoma cells in-vitro and growth of murine B16 melanomas in-vivo." Journal of pharmacy and pharmacology **54**(8): 1083-1089.

Loris, R., T. Hamelryck, J. Bouckaert and L. Wyns (1998). "Legume lectin structure." Biochimica et Biophysica Acta (BBA)-Protein Structure and Molecular Enzymology **1383**(1): 9-36.

Lowe, J. B. and J. D. Marth (2003). "A genetic approach to mammalian glycan function." Annual Review of Biochemistry **72**(1): 643-691.

Lu, Z.-R., J.-G. Shiah, S. Sakuma, P. Kopečková and J. Kopeček (2002). "Design of novel bioconjugates for targeted drug delivery." Journal of controlled release **78**(1): 165-173.

Luo, S., D. Zhangsun and K. Tang (2005). "Functional GNA expressed in Escherichia coli with high efficiency and its effect on Ceratovacuna lanigera Zehntner." Applied microbiology and biotechnology **69**(2): 184-191.

Lyu, S.-Y., Y.-J. Kwon, H.-J. Joo and W.-B. Park (2004). "Preparation of alginate/chitosan microcapsules and enteric coated granules of mistletoe lectin." Archives of pharmacal research **27**(1): 118-126.

Magliery, T. J., C. G. Wilson, W. Pan, D. Mishler, I. Ghosh, A. D. Hamilton and L. Regan (2005). "Detecting protein-protein interactions with a green fluorescent protein fragment reassembly trap: scope and mechanism." Journal of the American Chemical Society **127**(1): 146-157.

Maletzki, C., M. Linnebacher, R. Savai and U. Hobohm (2013). "Mistletoe lectin has a shiga toxin-like structure and should be combined with other Toll-like receptor ligands in cancer therapy." Cancer Immunology, Immunotherapy **62**(8): 1283-1292.

Martinez, J. J., M. A. Mulvey, J. D. Schilling, J. S. Pinkner and S. J. Hultgren (2000). "Type 1 pilus-mediated bacterial invasion of bladder epithelial cells." The EMBO journal **19**(12): 2803-2812.

Matarrese, P., B. Ascione, L. Ciarlo, R. Vona, C. Leonetti, M. Scarsella, A. M. Mileo, C. Catricalà, M. G. Paggi and W. Malorni (2010). "Cathepsin B inhibition interferes with metastatic potential of human melanoma: an in vitro and in vivo study." Molecular cancer **9**(1): 1.

Medina-Bolivar, F., R. Wright, V. Funk, D. Sentz, L. Barroso, T. D. Wilkins, W. Petri and C. L. Cramer (2003). "A non-toxic lectin for antigen delivery of plant-based mucosal vaccines." Vaccine **21**(9): 997-1005.

Mehr, K. and S. G. Withers (2015). "Mechanisms of the Sialidase and Trans-sialidase Activities of Bacterial Sialyltransferases from Glycosyltransferase Family 80 (GT80)." Glycobiology **26**(4): 353-359.

Min, C. (2016). Limitations of RNA interference as a potential technique for crop protection against insect pests. Ph.D., Durham University.

Minchinton, A. I. and I. F. Tannock (2006). "Drug penetration in solid tumours." Nature Reviews Cancer **6**(8): 583-592.

Minko, T. (2004). "Drug targeting to the colon with lectins and neoglycoconjugates." Adv Drug Deliv Rev. **56**: 491–509.

Mody, R., S. antaram Joshi and W. Chaney (1995). "Use of lectins as diagnostic and therapeutic tools for cancer." Journal of pharmacological and toxicological methods **33**(1): 1-10.

Moreno, M. and E. Giralt (2015). "Three valuable peptides from bee and wasp venoms for therapeutic and biotechnological use: Melittin, apamin and mastoparan." Toxins **7**(4): 1126-1150.

Moreno, M., E. Zurita and E. Giralt (2014). "Delivering wasp venom for cancer therapy." Journal of controlled release **182**: 13-21.

Moxon, E. R., P. B. Rainey, M. A. Nowak and R. E. Lenski (1994). "Adaptive evolution of highly mutable loci in pathogenic bacteria." Current biology **4**(1): 24-33.

Muralidhara, B. K., R. Baid, S. M. Bishop, M. Huang, W. Wang and S. Nema (2016). "Critical considerations for developing nucleic acid macromolecule based drug products." Drug discovery today **21**(3): 430-444.

Murata, R. M., R. Yatsuda, M. H. dos Santos, L. K. Kohn, F. T. Martins, T. J. Nagem, S. M. Alencar, J. E. de Carvalho and P. L. Rosalen (2010). "Antiproliferative effect of benzophenones and their influence on cathepsin activity." Phytotherapy Research **24**(3): 379-383.

Müthing, J., M. Burg, B. Möckel, M. Langer, W. Metelmann-Strupat, A. Werner, U. Neumann, J. Peter-Katalinic and J. Eck (2002). "Preferential binding of the anticancer drug rViscumin (recombinant mistletoe lectin) to terminally α2-6-sialylated neolacto-series gangliosides." Glycobiology **12**(8): 485-497.

Müthing, J., I. Meisen, P. Bulau, M. Langer, K. Witthohn, H. Lentzen, U. Neumann and J. Peter-Katalinic (2004). "Mistletoe lectin I is a sialic acid-specific lectin with strict preference to gangliosides and glycoproteins with terminal Neu5Acα2-6Galβ1-4GlcNAc residues." Biochemistry **43**(11): 2996-3007.

Myrberg, H., L. Zhang, M. Mäe and Ü. Langel (2007). "Design of a tumor-homing cell-penetrating peptide." Bioconjugate chemistry **19**(1): 70-75.

Nakagawa, Y., H. Sakamoto, H. Tateno, J. Hirabayashi and S. Oguri (2012). "Purification, Characterization, and Molecular Cloning of Lectin from Winter Buds of Lysichiton camtschatcensis (L.) Schott." Bioscience, biotechnology, and biochemistry **76**(1): 25-33.

Nakasaki, H., T. Mitomi, T. Noto, K. Ogoshi, H. Hanaue, Y. Tanaka, H. Makuuchi, H. Clausen and S.-i. Hakomori (1989). "Mosaicism in the expression of tumor-associated carbohydrate antigens in human colonic and gastric cancers." Cancer research **49**(13): 3662-3669.

Nakasu, E. Y., M. G. Edwards, E. Fitches, J. A. Gatehouse and A. M. Gatehouse (2014). "Transgenic plants expressing ω-ACTX-Hv1a and snowdrop lectin (GNA) fusion protein show enhanced resistance to aphids." Frontiers in plant science **5**: 673.

Niesen, J., G. Hehmann-Titt, M. Woitok, R. Fendel, S. Barth, R. Fischer and C. Stein (2016). "A novel fully-human cytolytic fusion protein based on granzyme B shows in vitro cytotoxicity and ex vivo binding to solid tumors overexpressing the epidermal growth factor receptor." Cancer letters **374**(2): 229-240.

Nishimoto, K., K. Tanaka, T. Murakami, H. Nakashita, H. Sakamoto and S. Oguri (2014). "Datura stramonium agglutinin: Cloning, molecular characterization and recombinant production in Arabidopsis thaliana." Glycobiology **25**(2): 157-169.

Niwa, H., A. G. Tonevitsky, I. I. Agapov, S. Saward, U. Pfüller and R. A. Palmer (2003). "Crystal structure at 3 Å of mistletoe lectin I, a dimeric type-II ribosome-inactivating protein, complexed with galactose." European Journal of Biochemistry **270**(13): 2739-2749.

Nørholm, M. H., S. Toddo, M. T. Virkki, S. Light, G. von Heijne and D. O. Daley (2013). "Improved production of membrane proteins in Escherichia coli by selective codon substitutions." FEBS letters **587**(15): 2352-2358.

Ofek, I., D. L. Hasty and N. Sharon (2003). "Anti-adhesion therapy of bacterial diseases: prospects and problems." FEMS Immunology & Medical Microbiology **38**(3): 181-191.

Ogawa, H., S. Inouye, F. I. Tsuji, K. Yasuda and K. Umesono (1995). "Localization, trafficking, and temperature-dependence of the Aequorea green fluorescent protein in cultured vertebrate cells." Proceedings of the National Academy of Sciences **92**(25): 11899-11903.

Oršolić, N., L. Šver, S. Verstovšek, S. Terzić and I. Bašić (2003). "Inhibition of mammary carcinoma cell proliferation in vitro and tumor growth in vivo by bee venom." Toxicon **41**(7): 861-870.

Ozaki, Y., Y. Matsumoto, Y. Yatomi, M. Higashihara, T. Kariya and S. Kume (1990). "Mastoparan, a wasp venom, activates platelets via pertussis toxin-sensitive GTP-binding proteins." Biochemical and biophysical research communications **170**(2): 779-785.

Pan, S.-h. and B. A. Malcolm (2000). "Reduced background expression and improved plasmid stability with pET vectors in BL21 (DE3)." Biotechniques **29**(6): 1234-1238.

Park, H. S., S. Y. Lee, Y. H. Kim, J. Y. Kim, S. J. Lee and M.-U. Choi (2000). "Membrane perturbation by mastoparan 7 elicits a broad alteration in lipid composition of L1210 cells." Biochimica et Biophysica Acta (BBA)-Molecular and Cell Biology of Lipids **1484**(2): 151-162.

Pasqualini, R., E. Koivunen and E. Ruoslahti (1997). "Alpha v integrins as receptors for tumor targeting by circulating ligands." Nature biotechnology **15**(6): 542-546.

Patterson, G. H., S. M. Knobel, W. D. Sharif, S. R. Kain and D. W. Piston (1997). "Use of the green fluorescent protein and its mutants in quantitative fluorescence microscopy." Biophysical journal **73**(5): 2782.

Peng, C.-C., J.-Y. Hong and W.-C. Tu (2012). "Cost-Effective Expression and Purification of Recombinant Venom Peptide Mastoparan B of Vespa Basalis in Escherichia Coli." International Journal of Bioscience, Biochemistry and Bioinformatics **2**(5): 309.

Pevzner, I. B., I. I. Agapov, H. Niwa, N. V. Maluchenko, M. M. Moisenovich, U. Pfüller and A. G. Tonevitsky (2004). "Differences in amino acid sequences of mistletoe lectin I and III B-subunits determining carbohydrate binding specificity." Biochimica et Biophysica Acta (BBA)-General Subjects **1675**(1): 155-164.

Pieters, R. J. (2007). "Intervention with bacterial adhesion by multivalent carbohydrates." Medicinal research reviews **27**(6): 796-816.

Plattner, V., G. Ratzinger, E. Engleder, S. Gallauner, F. Gabor and M. Wirth (2009). "Alteration of the glycosylation pattern of monocytic THP-1 cells upon differentiation and its impact on lectin-mediated drug delivery." European Journal of Pharmaceutics and Biopharmaceutics **73**(3): 324-330.

Plattner, V. E., E. T. Ratzinger GEngleder and S. Gallauner (2009). "Alteration of the glycosylation pattern of monocytic THP-1 cells upon differentiation and its impact on lectin-mediated drug delivery." Eur J Pharm Biopharm **73**: 361–365.

Pocanschi, C. L., G. Kozlov, U. Brockmeier, A. Brockmeier, D. B. Williams and K. Gehring (2011). "Structural and functional relationships between the lectin and arm domains of calreticulin." Journal of Biological Chemistry **286**(31): 27266-27277.

Podgorski, I. and B. F. Sloane (2003). Cathepsin B and its role (s) in cancer progression. Biochemical Society Symposia, Portland Press Limited.

Pohleven, J., B. Štrukelj and J. Kos (2012). "Affinity chromatography of lectins." Affinity Chromatography. InTech: 49-74.

Pokuri, V. K., M. K. Fong and R. Iyer (2016). "Octreotide and Lanreotide in Gastroenteropancreatic Neuroendocrine Tumors." Current oncology reports **18**(1): 1-9.

Polson, A. G., W. Y. Ho and V. Ramakrishnan (2011). "Investigational antibody-drug conjugates for hematological malignancies." Expert Opinion On Investigational Drugs **20**(1): 75-85.

Powell, L. D. and A. Varki (1995). "I-type lectins." J. Biol.Chem. **14243-6**: 270.

Przybylo, M., E. Pochec, P. Link-Lenczowski and A. Litynska (2008). "Beta1-6 branching of cell surface glycoproteins may contribute to uveal melanoma progression by up-regulating cell motility." Mol Vis **14**: 625-636.

Pusztai, A., G. Grant, R. Spencer, T. J. Duguid, D. Brown, S. Ewen, W. Peumans, E. J. Damme and S. Bardocz (1993). "Kidney bean lectin-induced Escherichia coli overgrowth in the small intestine is blocked by GNA, a mannose-specific lectin." Journal of Applied Bacteriology **75**(4): 360-368.

Qin, Y., J. Zhou, W. Zhang, X. Yang, J. Wang, C. Wei, F. Gu and T. Lei (2016). "Construction of An Anticancer Fusion Peptide (ACFP) Derived from Milk Proteins and An Assay of Anti-ovarian Cancer Cells in vitro." Anti-cancer agents in medicinal chemistry.

Rabinovich, G. A. and D. O. Croci (2012). "Regulatory circuits mediated by lectin-glycan interactions in autoimmunity and cancer." Immunity **36**(3): 322-335.

Raemaekers, R. J. (2000). Expression of functional plant lectins in heterologous systems, Durham University.

Raghuraman, H. and A. Chattopadhyay (2007). "Melittin: a membrane-active peptide with diverse functions." Bioscience reports **27**(4-5): 189-223.

Rambaruth, N. D. and M. V. Dwek (2011). "Cell surface glycan–lectin interactions in tumor metastasis." Acta histochemica **113**(6): 591-600.

Ranjan, A. and R. D. Kalraiya (2013). "α2, 6 Sialylation associated with increased β1, 6-branched N-oligosaccharides influences cellular adhesion and invasion." Journal of biosciences **38**(5): 867-876.

Rao, K., K. S. Rathore, T. K. Hodges, X. Fu, E. Stoger, D. Sudhakar, S. Williams, P. Christou, M. Bharathi and D. P. Bown (1998). "Expression of snowdrop lectin (GNA) in transgenic rice plants confers resistance to rice brown planthopper." The Plant Journal 15(4): 469-477.

Rath, A., M. Glibowicka, V. G. Nadeau, G. Chen and C. M. Deber (2009). "Detergent binding explains anomalous SDS-PAGE migration of membrane proteins." Proceedings of the National Academy of Sciences **106**(6): 1760-1765.

Re, S., N. Miyashita, Y. Yamaguchi and Y. Sugita (2011). "Structural diversity and changes in conformational equilibria of biantennary complex-type N-glycans in water revealed by replica-exchange molecular dynamics simulation." Biophysical journal **101**(10): L44-L46.

Reid, B. G. and G. C. Flynn (1997). "Chromophore formation in green fluorescent protein." Biochemistry **36**(22): 6786-6791.

Rek, A., E. Krenn and A. J. Kungl (2009). "Therapeutically targeting protein–glycan interactions." Br J Pharmacol **157**: 686–694.

Rocha, T., L. L. S. de Barros, K. Fontana, B. M. de Souza, M. S. Palma and M. A. da Cruz-Höfling (2010). "Inflammation and apoptosis induced by mastoparan Polybia-MPII on skeletal muscle." Toxicon **55**(7): 1213-1221.

Rodrigues, J. A., A. Acosta-Serrano, M. Aebi, M. A. Ferguson, F. H. Routier, I. Schiller, S. Soares, D. Spencer, A. Titz and I. B. Wilson (2015). "Parasite Glycobiology: A Bittersweet Symphony." PLoS Pathog **11**(11): e1005169.

Rodriguez-Ruiz, V., A. Maksimenko, R. Anand, S. Monti, V. Agostoni, P. Couvreur, M. Lampropoulou, K. Yannakopoulou and R. Gref (2015). "Efficient “green” encapsulation of a highly hydrophilic anticancer drug in metal–organic framework nanoparticles." Journal of drug targeting **23**(7-8): 759-767.

Rogers, G. N. and J. C. Paulson (1983). "Receptor determinants of human and animal influenza virus isolates: differences in receptor specificity of the H3 hemagglutinin based on species of origin." Virology **127**(2): 361-373.

Roseman, S. (2001). "Reflections on glycobiology." Journal of Biological Chemistry **276**(45): 41527-41542.

Roshy, S., B. F. Sloane and K. Moin (2003). "Pericellular cathepsin B and malignant progression." Cancer and Metastasis Reviews **22**(2-3): 271-286.

Sadler, J. E. (1984). "Biosynthesis of glycoproteins: formation of O-linked oligosaccharides." Biology of carbohydrates **2**: 199.

Samant, S., D. Blair, A. Chen, J. Mettetal, W. C. Shyu, M. Hixon, J. Ecsedy, S. Palani and A. Chakravarty (2013). "Abstract B214: Application of an evolutionary model of cancer cell response to dose-response viability curves to assess the potential for pre-existing resistance." American Association for Cancer Research **12**(11 Supplement): B214-B214.

Sambrook, J. and D. W. Russel (2006). The condensed protocols. New York, Cold Spring Harbor.

Sastry, K. and R. A. Ezekowitz (1993). "Collectins: pattern recognition molecules involved in first line host defense." Current opinion in immunology **5**(1): 59-66.

Sattler, M., S. Verma, G. Shrikhande, C. H. Byrne, Y. B. Pride, T. Winkler, E. A. Greenfield, R. Salgia and J. D. Griffin (2000). "The BCR/ABL tyrosine kinase induces production of reactive oxygen species in hematopoietic cells." Journal of Biological Chemistry **275**(32): 24273-24278.

Schauer, R. (2000). "Achievements and challenges of sialic acid research." Glycoconjugate journal **17**(7-9): 485-499.

Schauer, R., B. Ernst, G. W. Hart and P. Sinaý (2000). Biochemistry of Sialic Acid Diversity, Wiley Online Library.

Schlegel, S., J. Löfblom, C. Lee, A. Hjelm, M. Klepsch, M. Strous, D. Drew, D. J. Slotboom and J.-W. de Gier (2012). "Optimizing membrane protein overexpression in the Escherichia coli strain Lemo21 (DE3)." Journal of molecular biology **423**(4): 648-659.

Schouppe, D., P. Rougé, Y. Lasanajak, A. Barre, D. F. Smith, P. Proost and E. J. Van Damme (2010). "Mutational analysis of the carbohydrate binding activity of the tobacco lectin." Glycoconjugate journal **27**(6): 613-623.

Serrill, J. D., X. Wan, A. M. Hau, H. S. Jang, D. J. Coleman, A. K. Indra, A. W. Alani, K. L. McPhail and J. E. Ishmael (2016). "Coibamide A, a natural lariat depsipeptide, inhibits VEGFA/VEGFR2 expression and suppresses tumor growth in glioblastoma xenografts." Investigational new drugs **34**(1): 24-40.

Shaanan, B., H. Lis and N. Sharon (1991). "Structure of a legume lectin with an ordered N-linked carbohydrate in complex with lactose." Science **254**(5033): 862-866.

Shaposhnikova, V., M. Egorova, A. Kudryavtsev, M. K. Levitman and Y. N. Korystov (1997). "The effect of melittin on proliferation and death of thymocytes." FEBS letters **410**(2): 285-288.

Sharon, N. (1987). "Bacterial lectins, cell-cell recognition and infectious disease." FEBS letters **217**(2): 145-157.

Sharon, N. and H. Lis (2003). Lectins. London, Kluwer Academic Publishers.

Sharon, N. and H. Lis (2004). "History of lectins: from hemagglutinins to biological recognition molecules." Glycobiology **14**: 11 pp. 53R–62R.

Sharp, P. M. and W.-H. Li (1987). "The codon adaptation index-a measure of directional synonymous codon usage bias, and its potential applications." Nucleic acids research **15**(3): 1281-1295.

Shibuya, N., I. Goldstein, W. Broekaert, M. Nsimba-Lubaki, B. Peeters and W. Peumans (1987). "The elderberry (Sambucus nigra L.) bark lectin recognizes the Neu5Ac (alpha 2-6) Gal/GalNAc sequence." Journal of Biological Chemistry **262**(4): 1596-1601.

Shojaei, A. F., K. Tabatabaeian, S. Shakeri and F. Karimi (2016). "A novel 5-fluorouracile anticancer drug sensor based on ZnFe 2 O 4 magnetic nanoparticles ionic liquids carbon paste electrode." Sensors and Actuators B: Chemical **230**: 607-614.

Shpakov, A. and M. Pertseva (2006). "Molecular mechanisms for the effect of mastoparan on G proteins in tissues of vertebrates and invertebrates." Bulletin of experimental biology and medicine **141**(3): 302-306.

Sica, V., M. C. Maiuri, G. Kroemer and L. Galluzzi (2016). "Detection of Apoptotic Versus Autophagic Cell Death by Flow Cytometry." Programmed Cell Death: Methods and Protocols: 1-16.

Siebert, H.-C., C.-W. von der Lieth, X. Dong, G. Reuter, R. Schauer, H.-J. Gabius and J. F. Vliegenthart (1996). "Molecular dynamics-derived conformation and intramolecular interaction analysis of the N-acetyl-9-O-acetylneuraminic acid-containing ganglioside GD1a and NMR-based analysis of its binding to a human polyclonal inununoglobulin G fraction with selectivity for O-acetylated sialic acids." Glycobiology **6**(6): 561-571.

Siemering, K. R., R. Golbik, R. Sever and J. Haseloff (1996). "Mutations that suppress the thermosensitivity of green fluorescent protein." Current Biology **6**(12): 1653-1663.

Simpson, D. J., J. C. Sacher and C. M. Szymanski (2015). "Exploring the interactions between bacteriophage-encoded glycan binding proteins and carbohydrates." Current opinion in structural biology **34**: 69-77.

Singh, H. and S. P. Sarathi (2012). "Insight of Lectins-A review." International Journal of Scientific and Engineering Research **3**: 1-9.

Sleat, D. E., P. Sun, J. A. Wiseman, L. Huang, M. El-Banna, H. Zheng, D. F. Moore and P. Lobel (2013). "Extending the mannose 6-phosphate glycoproteome by high resolution/accuracy mass spectrometry analysis of control and acid phosphatase 5-deficient mice." Molecular & Cellular Proteomics **12**(7): 1806-1817.

Soler, M. H., S. Stoeva, C. Schwamborn, S. Wilhelm, T. Stiefel and W. Voelter (1996). "Complete amino acid sequence of the A chain of mistletoe lectin I." FEBS Letters **399**(1-2): 153-157.

Soler, M. H., S. Stoeva and W. Voelter (1998). "Complete amino acid sequence of the B chain of mistletoe lectin I." Biochem Biophys Res Commun **246**(3): 596-601.

Soman, N. R., S. L. Baldwin, G. Hu, J. N. Marsh, G. M. Lanza, J. E. Heuser, J. M. Arbeit, S. A. Wickline and P. H. Schlesinger (2009). "Molecularly targeted nanocarriers deliver the cytolytic peptide melittin specifically to tumor cells in mice, reducing tumor growth." The Journal of clinical investigation **119**(9): 2830-2842.

Song, C., X. Lu, B. Cheng, J. Du, B. Li and C. Ling (2007). "Effects of melittin on growth and angiogenesis of human hepatocellular carcinoma BEL-7402 cell xenografts in nude mice." Ai zheng= Aizheng= Chinese journal of cancer **26**(12): 1315-1322.

Soomets, U., M. Haellbrink, M. Zorko and Ü. Langel (1997). "From Galanin And Mastoparan To Galparan And Transportan." Current Topics In Peptide & Protein Research **Letn. 2 (1997)**: 83-113.

Sørensen, H. P. and K. K. Mortensen (2005). "Advanced genetic strategies for recombinant protein expression in Escherichia coli." Journal of biotechnology **115**(2): 113-128.

Stanley, P. (2014). "Mannosyl (Alpha-1, 3-)-Glycoprotein Beta-1, 2-N-Acetylglucosaminyltransferase (MGAT1)." Handbook of Glycosyltransferases and Related Genes: 183-194.

Stanley, P. (2015). O-Glycans in Mammalian Notch Signaling. Glycoscience: Biology and Medicine, Springer**:** 857-864.

Stanley, P., F. Batista and H.-H. Huang (2013). "Roles for Glycans in Mammalian Development and Spermatogenesis." The FASEB Journal **27**(1_MeetingAbstracts): 211.212.

Stanley, P., H. Schachter and N. Taniguchi (2009). N-glycans. Essentials of Glycobiology. Cold Spring Harbor (NY), Cold Spring Harbor Laboratory Press.

Stevens, J., O. Blixt, L. Glaser, J. K. Taubenberger, P. Palese, J. C. Paulson and I. A. Wilson (2006). "Glycan microarray analysis of the hemagglutinins from modern and pandemic influenza viruses reveals different receptor specificities." Journal of molecular biology **355**(5): 1143-1155.

Stirpe, F. (2004). "Ribosome-inactivating proteins." Toxicon **44**(4): 371-383.

Stoger, E., S. Williams, P. Christou, R. E. Down and J. A. Gatehouse (1999). "Expression of the insecticidal lectin from snowdrop (Galanthus nivalis agglutinin; GNA) in transgenic wheat plants: effects on predation by the grain aphid Sitobion avenae." Molecular Breeding **5**(1): 65-73.

Szpaderska, A. M. and A. Frankfater (2001). "An intracellular form of cathepsin B contributes to invasiveness in cancer." Cancer research **61**(8): 3493-3500.

Taylor, M. E. and K. Drickamer (2009). "Structural insights into what glycan arrays tell us about how glycan-binding proteins interact with their ligands." Glycobiology **19**(11): 1155-1162.

Taylor, R. E., C. J. Gregg, V. Padler-Karavani, D. Ghaderi, H. Yu, S. Huang, R. U. Sorensen, X. Chen, J. Inostroza and V. Nizet (2010). "Novel mechanism for the generation of human xeno-autoantibodies against the nonhuman sialic acid N-glycolylneuraminic acid." The Journal of experimental medicine **207**(8): 1637-1646.

Tejero, J., P. Jiménez, E. J. Quinto, D. Cordoba-Diaz, M. Garrosa, M. Cordoba-Diaz, M. J. Gayoso and T. Girbés (2015). "Elderberries: A source of ribosome-inactivating proteins with lectin activity." Molecules **20**(2): 2364-2387.

Templeton, D. M. (1992). "Proteoglycans in cell regulation." Critical reviews in clinical laboratory sciences **29**(2): 141-184.

Thies, A., D. Nuge, U. Pfuller, I. Moll and U. Schumacher (2005). "Influence of mistletoe lectins and cytokines induced by them on cell proliferation of human melanoma cells in vitro." Toxicology **207**: 105–116.

Thies, A., D. Nugel, U. Pfüller, I. Moll and U. Schumacher (2005). "Influence of mistletoe lectins and cytokines induced by them on cell proliferation of human melanoma cells in vitro." Toxicology **207**(1): 105-116.

Tonevitsky, A., I. Agapov, I. Pevzner, N. Malyuchenko, M. Moisenovich, U. Pfueller and M. Kirpichnikov (2004). "Cloning and expression of catalytic subunit of MLIII, the ribosome-inactivating protein from Viscum album." Biochemistry (Moscow) **69**(6): 642-650.

Tonevitsky, A. G., I. I. Agapov, A. T. Shamshiev, D. E. Temyakov, P. Pohl and M. P. Kirpichnikov (1996). "Immunotoxins containing A-chain of mistletoe lectin I are more active than immunotoxins with ricin A-chain." FEBS letters **392**(2): 166-168.

Toporkiewicz, M., J. Meissner, L. Matusewicz, A. Czogalla and A. F. Sikorski (2015). "Toward a magic or imaginary bullet? Ligands for drug targeting to cancer cells: principles, hopes, and challenges." Int J Nanomedicine **10**: 1399-1414.

Torchilin, V. (2008). "Intracellular delivery of protein and peptide therapeutics." Drug Discov Today Technol **5**(2-3): e95-e103.

Trachootham, D., Y. Zhou, H. Zhang, Y. Demizu, Z. Chen, H. Pelicano, P. J. Chiao, G. Achanta, R. B. Arlinghaus and J. Liu (2006). "Selective killing of oncogenically transformed cells through a ROS-mediated mechanism by β-phenylethyl isothiocyanate." Cancer cell **10**(3): 241-252.

Trail, P., D. Willner, S. Lasch, A. Henderson, S. Hofstead, A. Casazza, R. Firestone, I. Hellstrom and K. Hellstrom (1993). "Cure of xenografted human carcinomas by BR96-doxorubicin immunoconjugates." Science **261**(5118): 212-215.

Trefzer, U., R. Gutzmer, T. Wilhelm, F. Schenck, K. C. Kähler, V. Jacobi, K. Witthohn, H. Lentzen and P. Mohr (2014). "Treatment of unresectable stage IV metastatic melanoma with aviscumine after anti-neoplastic treatment failure: a phase II, multi-centre study." Journal for immunotherapy of cancer **2**(1): 1.

Tsien, R. Y. (1998). "The green fluorescent protein." Annual review of biochemistry 67(1): 509-544.

Tsuboi, S. (2015). Roles of Glycans in Immune Evasion from NK Immunity. Sugar Chains, Springer**:** 177-188.

Unitt, J. and D. Hornigold (2011). "Plant lectins are novel Toll-like receptor agonists." Biochemical pharmacology **81**(11): 1324-1328.

Urech, K., G. Schaller and C. Jaggy (2006). "Viscotoxins, mistletoe lectins and their isoforms in mistletoe (Viscum album L.) extracts Iscador." Arzneimittelforschung **56**(6A): 428-434.

Van Damme, E. J., A. K. Allen and W. J. Peumans (1988). "Related mannose-specific lectins from different species of the family Amaryllidaceae." Physiologia Plantarum **73**(1): 52-57.

Van Damme, E. J., N. Lannoo and W. J. Peumans (2008). "Plant lectins." Advances in botanical research **48**: 107-209.

Van Damme, E. J., W. J. Peumans, A. Barre and P. Rougé (1998a). "Plant lectins: a composite of several distinct families of structurally and evolutionary related proteins with diverse biological roles." Critical Reviews in Plant Sciences **17**(6): 575-692.

Vasiljeva, O., T. Reinheckel, C. Peters, D. Turk, V. Turk and B. Turk (2007). "Emerging roles of cysteine cathepsins in disease and their potential as drug targets." Current pharmaceutical design **13**(4): 387-403.

Vasiljeva, O. and B. Turk (2008). "Dual contrasting roles of cysteine cathepsins in cancer progression: apoptosis versus tumour invasion." Biochimie **90**(2): 380-386.

Vervecken, W., S. Kleff, U. Pfüller and A. Büssing (2000). "Induction of apoptosis by mistletoe lectin I and its subunits. No evidence for cytotoxic effects caused by isolated A-and B-chains." The international journal of biochemistry & cell biology **32**(3): 317-326.

Villacampa, N., B. Almolda, B. González and B. Castellano (2013). "Tomato lectin histochemistry for microglial visualization." Microglia: Methods and Protocols: 261-279.

Voet, D. and J. G. Voet (1995). Biochemistry. New York, John Wiley & Sons.

Wei, T., C. Chen, J. Liu, C. Liu, P. Posocco, X. Liu, Q. Cheng, S. Huo, Z. Liang and M. Fermeglia (2015). "Anticancer drug nanomicelles formed by self-assembling amphiphilic dendrimer to combat cancer drug resistance." Proceedings of the National Academy of Sciences **112**(10): 2978-2983.

Weingarten, R., L. Ransnäs, H. Mueller, L. A. Sklar and G. Bokoch (1990). "Mastoparan interacts with the carboxyl terminus of the alpha subunit of Gi." Journal of Biological Chemistry **265**(19): 11044-11049.

Weis, W. I. and K. Drickamer (1996). "Structural basis of lectin-carbohydrate recognition." Annual review of biochemistry **65**(1): 441-473.

Weis, W. I., K. Drickamer and W. A. Hendrickson (1992). "Structure of a C-type mannose-binding protein complexed with an oligosaccharide." Nature **360**(6400): 127-134.

Wong, C.-H. (2005). "Protein glycosylation: new challenges and opportunities." The Journal of organic chemistry **70**(11): 4219-4225.

Woodley, J. and B. Naisbett (1988). The potential of lectins for delaying the intestinal transit of drugs. Proc. Int. Symp. Control Rel. Bioact. Mater.

Wu, A. M., L.-K. Chin, H. Franz, U. Pfüller and A. Herp (1992). "Carbohydrate specificity of the receptor sites of mistletoe toxic lectin-I." Biochimica et Biophysica Acta (BBA)-General Subjects **1117**(2): 232-234.

Wu, A. M. and P. D. Senter (2005). "Arming antibodies: prospects and challenges for immunoconjugates." Nature biotechnology **23**(9): 1137-1146.

Wu, Z. L., X. Huang, A. J. Burton and K. A. Swift (2015). "Probing sialoglycans on fetal bovine fetuin with azido-sugars using glycosyltransferases." Glycobiology **26**(4): 329-334.

Wybenga, L. E., R. F. Epand, S. Nir, J. W. Chu, F. J. Sharom, T. D. Flanagan and R. M. Epand (1996). "Glycophorin as a receptor for Sendai virus." Biochemistry **35**(29): 9513-9518.

Xiao, H., J. M. Smeekens and R. Wu (2016). "Quantification of tunicamycin-induced protein expression and N-glycosylation changes in yeast." Analyst **141**: 3737-3745.

Xiao, Z., S. L. Morris-Natschke and K. H. Lee (2016). "Strategies for the optimization of natural leads to anticancer drugs or drug candidates." Medicinal research reviews **36**(1): 32-91.

Yabu, M., H. Korekane, H. Takahashi, H. Ohigashi, O. Ishikawa and Y. Miyamoto (2013). "Accumulation of free Neu5Ac-containing complex-type N-glycans in human pancreatic cancers." Glycoconjugate journal **30**(3): 247-256.

Yamada, Y., Y. Shinohara, T. Kakudo, S. Chaki, S. Futaki, H. Kamiya and H. Harashima (2005). "Mitochondrial delivery of mastoparan with transferrin liposomes equipped with a pH-sensitive fusogenic peptide for selective cancer therapy." International journal of pharmaceutics **303**(1): 1-7.

Yamamoto, T., M. Ito, K. Kageyama, K. Kuwahara, K. Yamashita, Y. Takiguchi, S. Kitamura, H. Terada and Y. Shinohara (2014). "Mastoparan peptide causes mitochondrial permeability transition not by interacting with specific membrane proteins but by interacting with the phospholipid phase." FEBS Journal **281**(17): 3933-3944.

Yamashita, K., K. Fukushima, T. Sakiyama, F. Murata, M. Kuroki and Y. Matsuoka (1995). "Expression of Siaα2→ 6Galβ1→ 4GlcNAc residues on sugar chains of glycoproteins including carcinoembryonic antigens in human colon adenocarcinoma: applications of Trichosanthes japonica agglutinin I for early diagnosis." Cancer research **55**(8): 1675-1679.

Yang, X., H. Zhu, Y. Ge, J. Liu, J. Cai, Q. Qin, L. Zhan, C. Zhang, L. Xu and Z. Liu (2014). "Melittin enhances radiosensitivity of hypoxic head and neck squamous cell carcinoma by suppressing HIF-1α." Tumor Biology **35**(10): 10443-10448.

Yau, T., X. Dan, C. C. W. Ng and T. B. Ng (2015). "Lectins with potential for anti-cancer therapy." Molecules **20**(3): 3791-3810.

Yokokawa, N., M. Komatsu, T. Takeda, T. Aizawa and T. Tamada (1989). "Mastoparan, a wasp venom, stimulates insulin release by pancreatic islets through pertussis toxin sensitive GTP-binding protein." Biochemical and biophysical research communications **158**(3): 712-716.

You, Y., Y. Cao, S. Guo, J. Xu, Z. Li, J. Wang and C. Xue (2015). "Purification and identification of α 2–3 linked sialoglycoprotein and α 2–6 linked sialoglycoprotein in edible bird's nest." European Food Research and Technology **240**(2): 389-397.

Zetterberg, M. M., K. Reijmar, M. Pränting, Å. Engström, D. I. Andersson and K. Edwards (2011). "PEG-stabilized lipid disks as carriers for amphiphilic antimicrobial peptides." Journal of controlled release **156**(3): 323-328.

Zhang, J. and I. Crandall (2007). "Expression of both N-and C-terminal GFP tagged huCD36 and their discrepancy in OxLDL and pRBC binding on CHO cells." Lipids in health and disease **6**(1): 1.

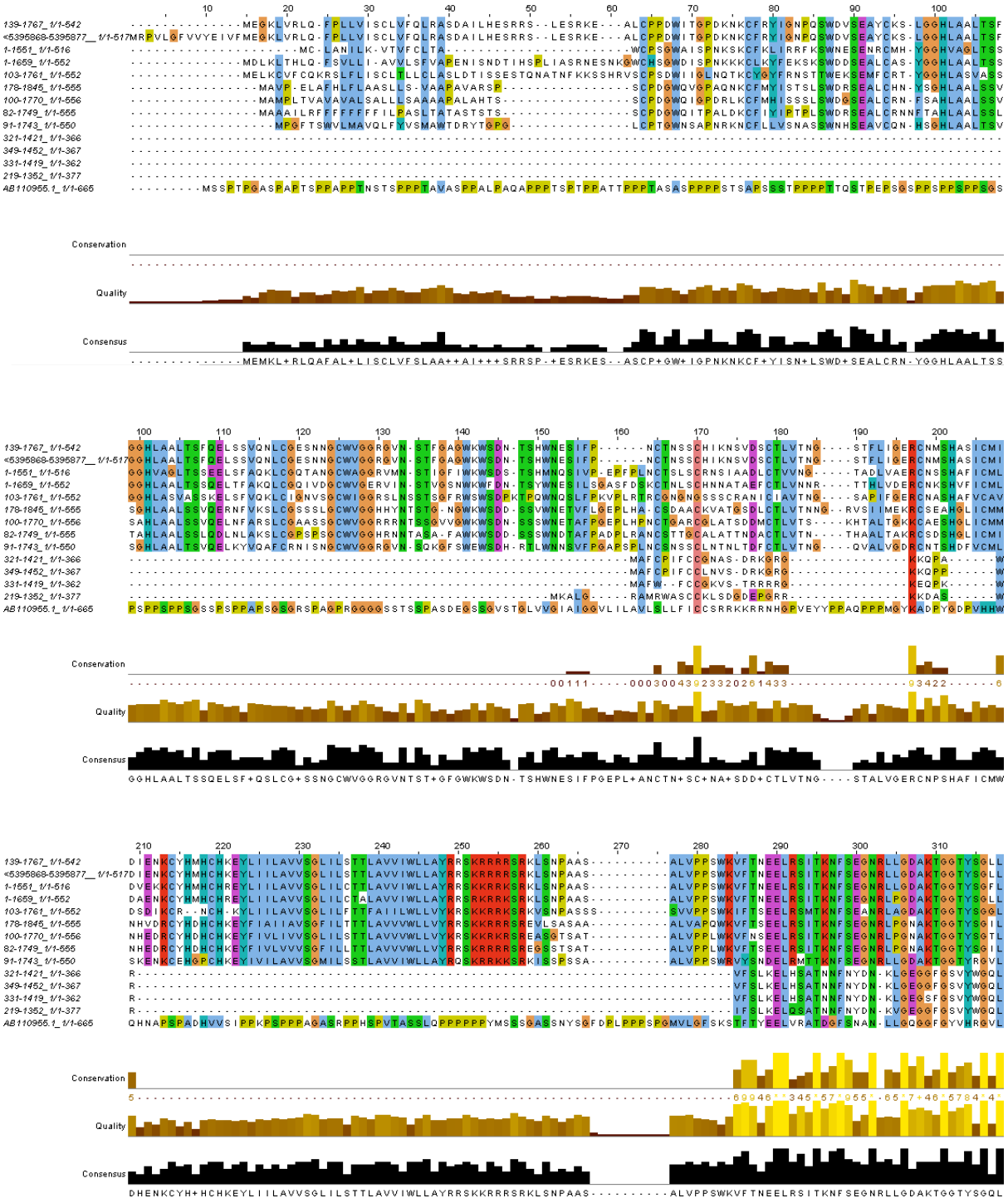
Zheng, Y.-J., R. L. Ornstein and J. A. Leary (1997). "A density functional theory investigation of metal ion binding sites in monosaccharides." Journal of Molecular Structure: THEOCHEM **389**(3): 233-240.

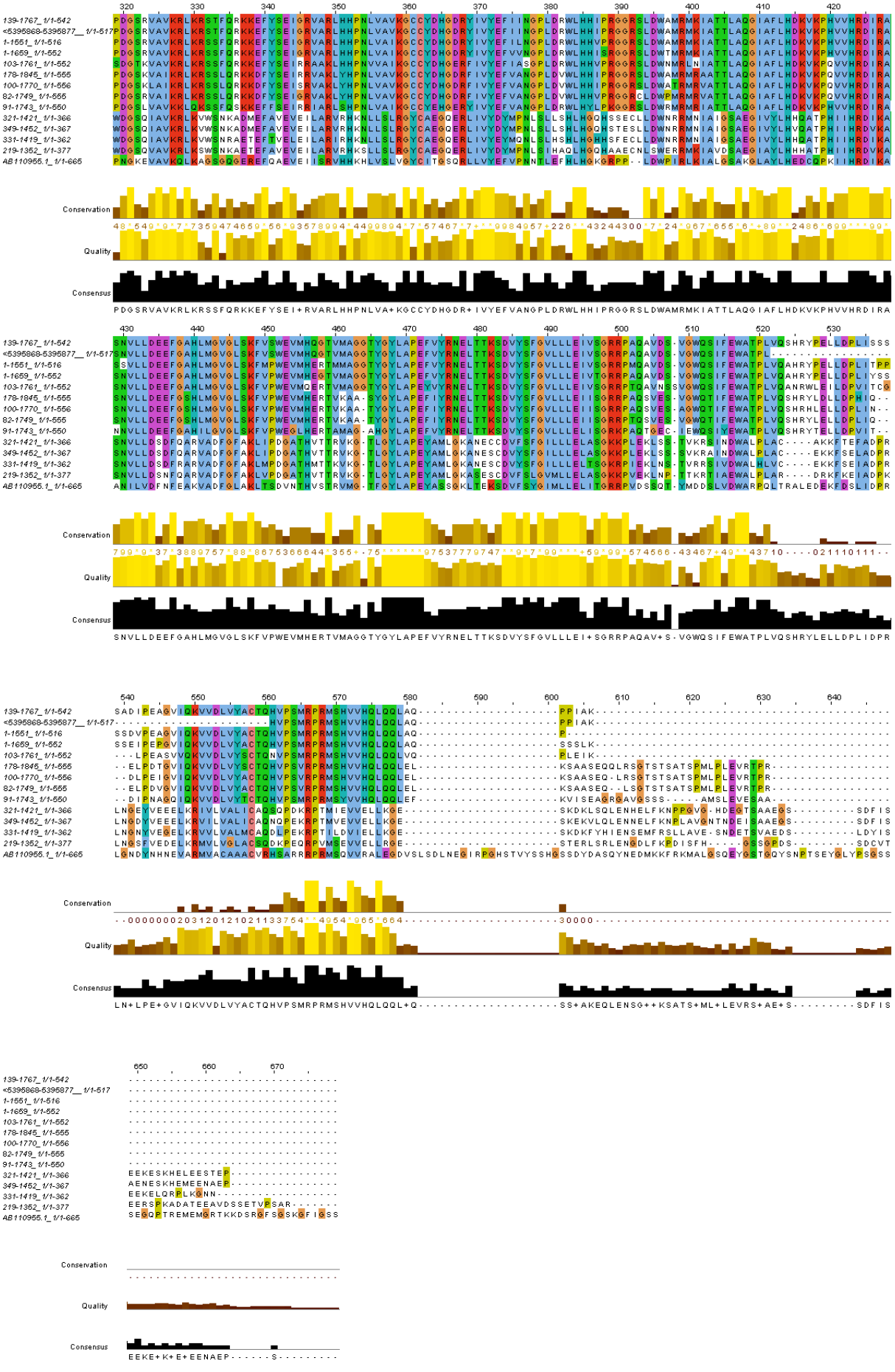
Zhou, Q. and R. Cummings (1992). Animal lectins: a distinct group of carbohydrate-binding proteins involved in cell adhesion, molecular recognition, and development. Cell surface carbohydrates and cell development, CRC Press Boca Raton, FL: 99-125.

Zou, W. (2005). "C-glycosides and aza-C-glycosides as potential glycosidase and glycosyltransferase inhibitors." Current topics in medicinal chemistry **5**(14): 1363-1391.

Appendix 1

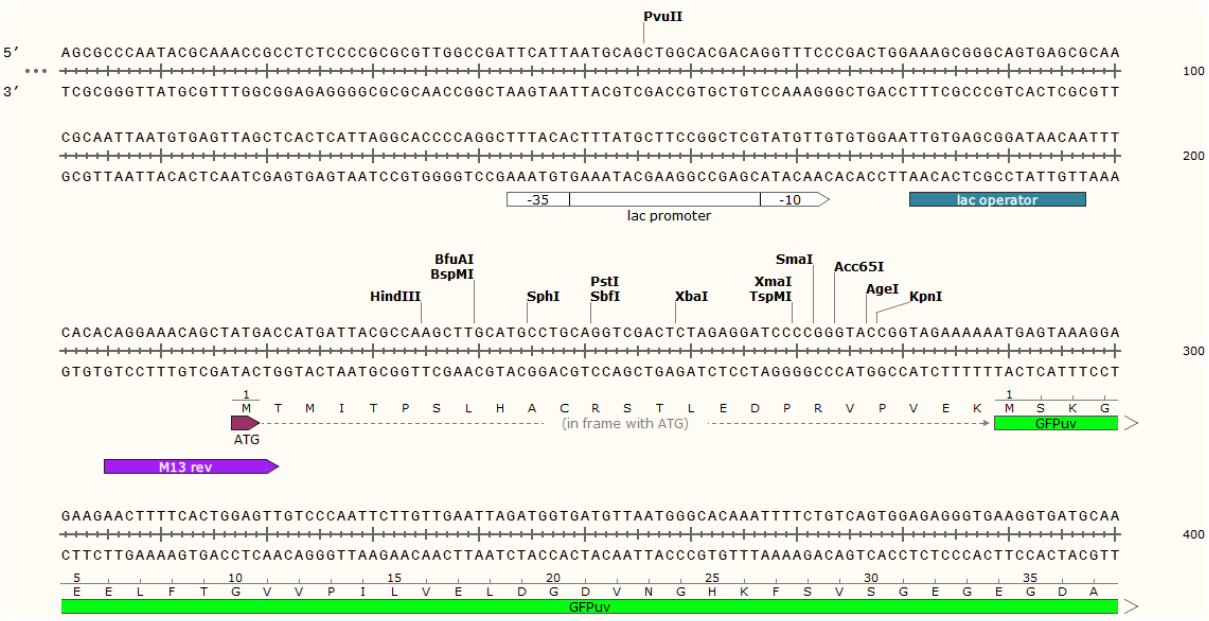
Clustal W2 alignment to identify novel C-type lectins

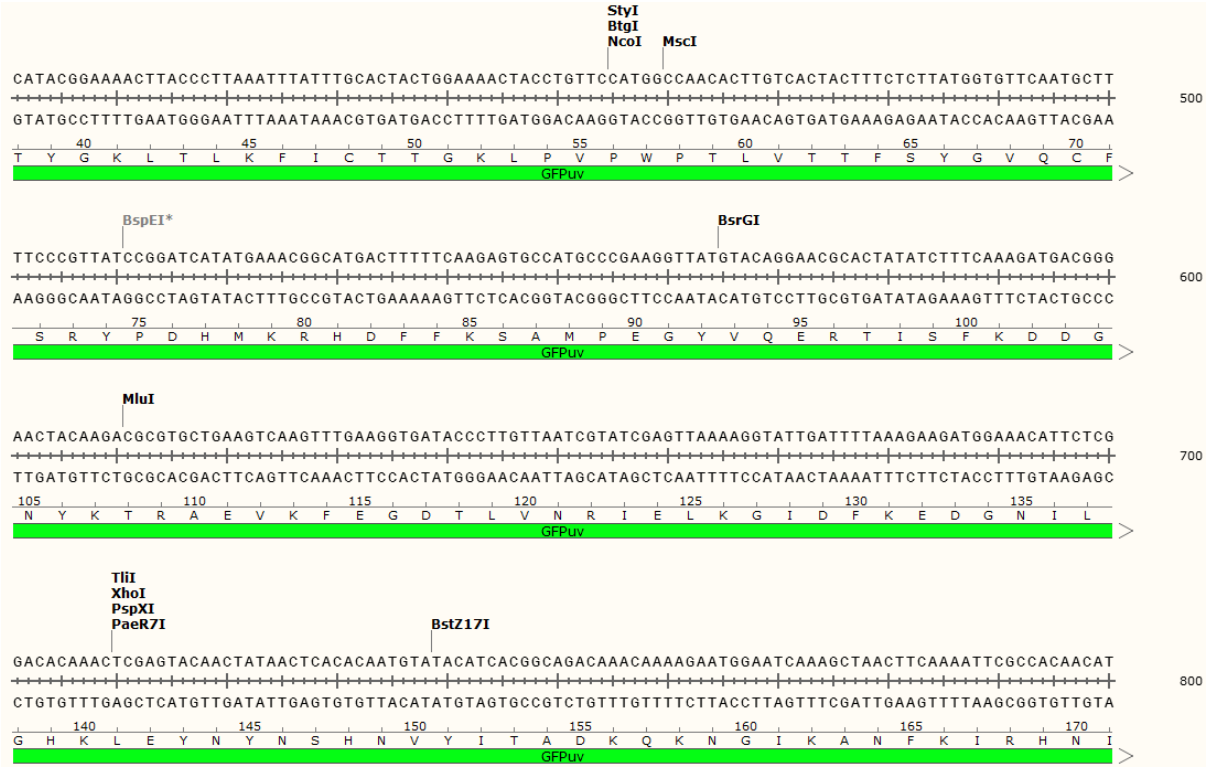

Conservation
Quality
Consensus

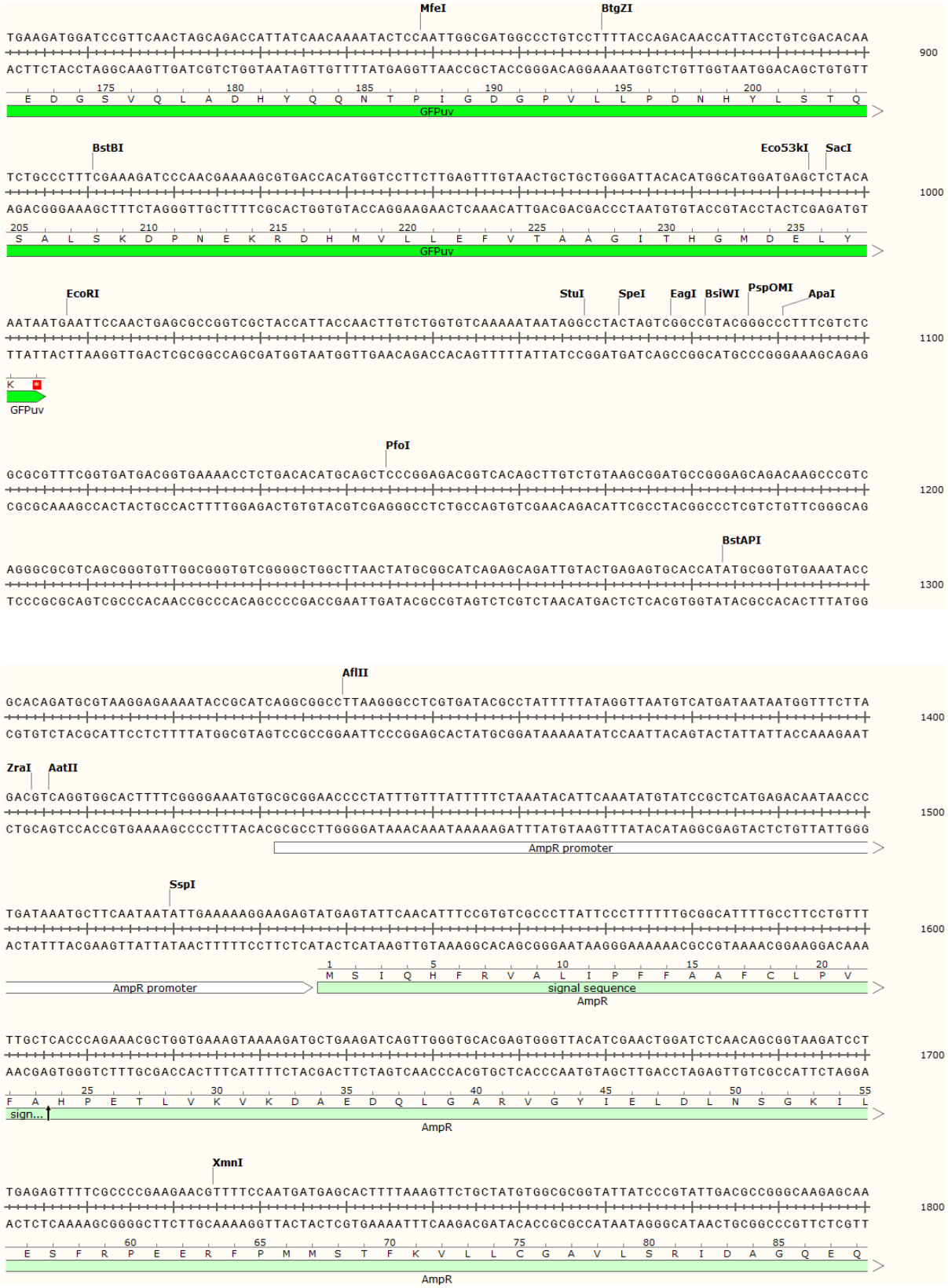

139-1767_1/1-542
<5395868-5395877__1/1-517
1-1551_1/1-516
1-1659_1/1-552
103-1761_1/1-552
178-1845_1/1-555
100-1770_1/1-556
82-1749_1/1-555
91-1743_1/1-550
321-1421_1/1-366
349-1452_1/1-367
331-1419_1/1-362
219-1352_1/1-377
AB110955.1_1/1-665
Conservation
Quality
Consensus

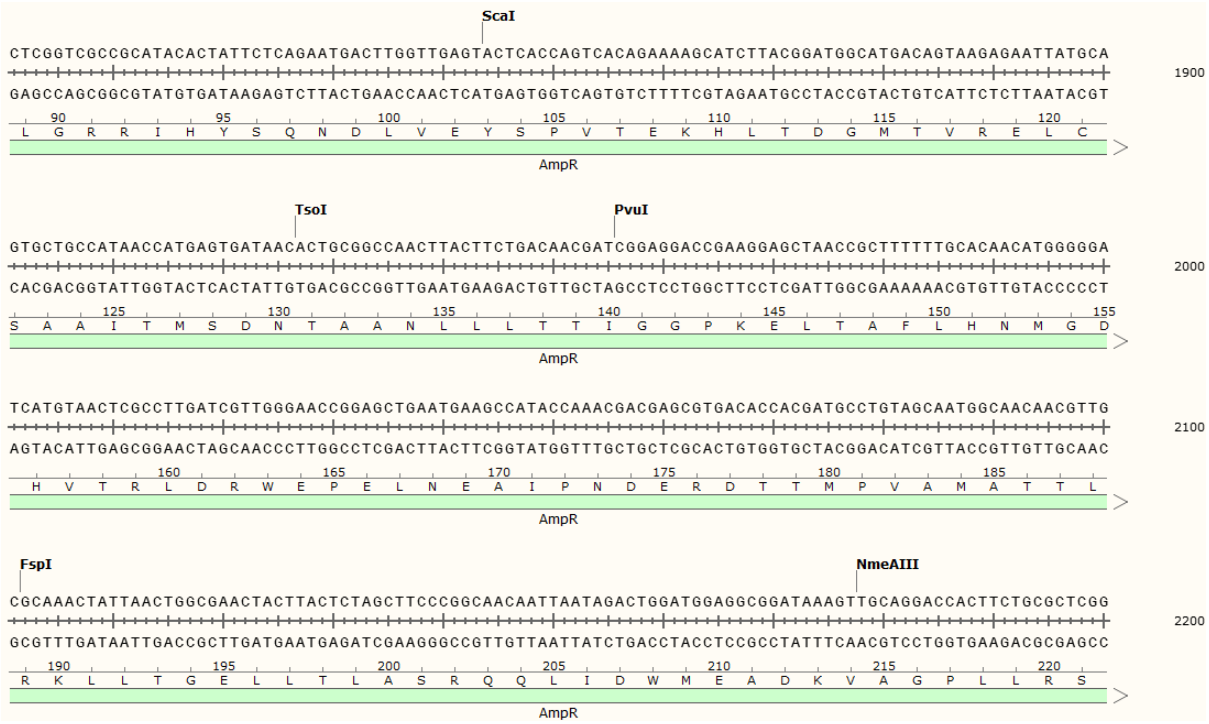
Appendix 2

The full sequence of pGFPuv expression vector


PvuII
5′ AGCGCCCAATACGCAAACCGCCTCTCCCCGCGCGTTGGCCGATTCATTAATGCAGCTGGCACGACAGGTTTCCCGACTGGAAAGCGGGCAGTGAGCGCAA
3′ TCGCGGGTTATGCGTTTGGCGGAGAGGGGCGCGCAACCGGCTAAGTAATTACGTCGACCGTGCTGTCCAAAGGGCTGACCTTTCGCCCGTCACTCGCGTT
100
CGCAATTAATGTGAGTTAGCTCACTCATTAGGCACCCCAGGCTTTACACTTTATGCTTCCGGCTCGTATGTTGTGTGGAATTGTGAGCGGATAACAATTT
GCGTTAATTACACTCAATCGAGTGAGTAATCCGTGGGGTCCGAAATGTGAAATACGAAGGCCGAGCATACAACACACCTTAACACTCGCCTATTGTTAAA
200
-35
-10
lac promoter
lac operator
BfuAI
BspMI
HindIII
SphI
PstI
SbfI
XbaI
SmaI
XmaI
TspMI
Acc65I
AgeI
KpnI
CACACAGGAAACAGCTATGACCATGATTACGCCAAGCTTGCATGCCTGCAGGTCGACTCTAGAGGATCCCCGGGTACCGGTAGAAAAAATGAGTAAAGGA
GTGTGTCCTTTGTCGATACTGGTACTAATGCGGTTCGAACGTACGGACGTCCAGCTGAGATCTCCTAGGGGCCCATGGCCATCTTTTTTACTCATTTCCT
300
1
M T M I T P S L H A C R S T L E D P R V P V E K
1
M S K G
ATG
(in frame with ATG)
GFPuv
M13 rev
GAAGAACTTTTCACTGGAGTTGTCCCAATTCTTGTTGAATTAGATGGTGATGTTAATGGGCACAAATTTTCTGTCAGTGGAGAGGGTGAAGGTGATGCAA
CTTCTTGAAAAGTGACCTCAACAGGGTTAAGAACAACTTAATCTACCACTACAATTACCCGTGTTTAAAAGACAGTCACCTCTCCCACTTCCACTACGTT
400
5 10 15 20 25 30 35
E E L F T G V V P I L V E L D G D V N G H K F S V S G E G E G D A
GFPuv


StyI
BtgI
NcoI
MscI
CATACGGAAAACTTACCCTTAAATTTATTTGCACTACTGGAAAACTACCTGTTCCATGGCCAACACTTGTCACTACTTTCTCTTATGGTGTTCAATGCTT
GTATGCCTTTTGAATGGGAATTTAAATAAACGTGATGACCTTTTGATGGACAAGGTACCGGTTGTGAACAGTGATGAAAGAGAATACCACAAGTTACGAA
500
40 45 50 55 60 65 70
T Y G K L T L K F I C T T G K L P V P W P T L V T T F S Y G V Q C F
GFPuv
BspEI*
BsrGI
TTCCCGTTATCCGGATCATATGAAACGGCATGACTTTTTCAAGAGTGCCATGCCCGAAGGTTATGTACAGGAACGCACTATATCTTTCAAAGATGACGGG
AAGGGCAATAGGCCTAGTATACTTTGCCGTACTGAAAAAGTTCTCACGGTACGGGCTTCCAATACATGTCCTTGCGTGATATAGAAAGTTTCTACTGCCC
600
75 80 85 90 95 100
S R Y P D H M K R H D F F K S A M P E G Y V Q E R T I S F K D D G
GFPuv
MluI
AACTACAAGACGCGTGCTGAAGTCAAGTTTGAAGGTGATACCCTTGTTAATCGTATCGAGTTAAAAGGTATTGATTTTAAAGAAGATGGAAACATTCTCG
TTGATGTTCTGCGCACGACTTCAGTTCAAACTTCCACTATGGGAACAATTAGCATAGCTCAATTTTCCATAACTAAAATTTCTTCTACCTTTGTAAGAGC
700
105 110 115 120 125 130 135
N Y K T R A E V K F E G D T L V N R I E L K G I D F K E D G N I L
GFPuv
TliI
XhoI
PspXI
PaeR7I
BstZ17I
GACACAAACTCGAGTACAACTATAACTCACACAATGTATACATCACGGCAGACAAACAAAAGAATGGAATCAAAGCTAACTTCAAAATTCGCCACAACAT
CTGTGTTTGAGCTCATGTTGATATTGAGTGTGTTACATATGTAGTGCCGTCTGTTTGTTTTCTTACCTTAGTTTCGATTGAAGTTTTAAGCGGTGTTGTA
800
140 145 150 155 160 165 170
G H K L E Y N Y N S H N V Y I T A D K Q K N G I K A N F K I R H N I
GFPuv


MfeI
BtgZI
TGAAGATGGATCCGTTCAACTAGCAGACCATTATCAACAAAATACTCCAATTGGCGATGGCCCTGTCCTTTTACCAGACAACCATTACCTGTCGACACAA
900
ACTTCTACCTAGGCAAGTTGATCGTCTGGTAATAGTTGTTTTATGAGGTTAACCGCTACCGGGACAGGAAAATGGTCTGTTGGTAATGGACAGCTGTGTT
E D G S V Q L A D H Y Q Q N T P I G D G P V L L P D N H Y L S T Q
GFPuv
BstBI
Eco53kI
SacI
TCTGCCCTTTCGAAAGATCCCAACGAAAAGCGTGACCACATGGTCCTTCTTGAGTTTGTAACTGCTGCTGGGATTACACATGGCATGGATGAGCTCTACA
1000
AGACGGGAAAGCTTTCTAGGGTTGCTTTTCGCACTGGTGTACCAGGAAGAACTCAAACATTGACGACGACCCTAATGTGTACCGTACCTACTCGAGATGT
S A L S K D P N E K R D H M V L L E F V T A A G I T H G M D E L Y
GFPuv
EcoRI
StuI
SpeI
EagI
BsiWI
PspOMI
ApaI
AATAATGAATTCCAACTGAGCGCCGGTCGCTACCATTACCAACTTGTCTGGTGTCAAAAATAATAGGCCTACTAGTCGGCCGTACGGGCCCTTTCGTCTC
1100
TTATTACTTAAGGTTGACTCGCGGCCAGCGATGGTAATGGTTGAACAGACCACAGTTTTTATTATCCGGATGATCAGCCGGCATGCCCGGGAAAGCAGAG
K
GFPuv
PfoI
GCGCGTTTCGGTGATGACGGTGAAAACCTCTGACACATGCAGCTCCCGGAGACGGTCACAGCTTGTCTGTAAGCGGATGCCGGGAGCAGACAAGCCCGTC
1200
CGCGCAAAGCCACTACTGCCACTTTTGGAGACTGTGTACGTCGAGGGCCTCTGCCAGTGTCGAACAGACATTCGCCTACGGCCCTCGTCTGTTCGGGCAG
BstAPI
AGGGCGCGTCAGCGGGTGTTGGCGGGTGTCGGGGCTGGCTTAACTATGCGGCATCAGAGCAGATTGTACTGAGAGTGCACCATATGCGGTGTGAAATACC
1300
TCCCGCGCAGTCGCCCACAACCGCCCACAGCCCCGACCGAATTGATACGCCGTAGTCTCGTCTAACATGACTCTCACGTGGTATACGCCACACTTTATGG
AflII
GCACAGATGCGTAAGGAGAAAATACCGCATCAGGCGGCCTTAAGGGCCTCGTGATACGCCTATTTTTATAGGTTAATGTCATGATAATAATGGTTTCTTA
1400
CGTGTCTACGCATTCCTCTTTTATGGCGTAGTCCGCCGGAATTCCCGGAGCACTATGCGGATAAAAATATCCAATTACAGTACTATTATTACCAAAGAAT
ZraI
AatII
GACGTCAGGTGGCACTTTTCGGGGAAATGTGCGCGGAACCCCTATTTGTTTATTTTTCTAAATACATTCAAATATGTATCCGCTCATGAGACAATAACCC
1500
CTGCAGTCCACCGTGAAAAGCCCCTTTACACGCGCCTTGGGGATAAACAAATAAAAAGATTTATGTAAGTTTATACATAGGCGAGTACTCTGTTATTGGG
AmpR promoter
SspI
TGATAAATGCTTCAATAATATTGAAAAAGGAAGAGTATGAGTATTCAACATTTCCGTGTCGCCCTTATTCCCTTTTTTGCGGCATTTTGCCTTCCTGTTT
1600
ACTATTTACGAAGTTATTATAACTTTTTCCTTCTCATACTCATAAGTTGTAAAGGCACAGCGGGAATAAGGGAAAAAACGCCGTAAAACGGAAGGACAAA
M S I Q H F R V A L I P F F A A F C L P V
AmpR promoter
signal sequence
AmpR
TTGCTCACCCAGAAACGCTGGTGAAAGTAAAAGATGCTGAAGATCAGTTGGGTGCACGAGTGGGTTACATCGAACTGGATCTCAACAGCGGTAAGATCCT
1700
AACGAGTGGGTCTTTGCGACCACTTTCATTTTCTACGACTTCTAGTCAACCCACGTGCTCACCCAATGTAGCTTGACCTAGAGTTGTCGCCATTCTAGGA
F A H P E T L V K V K D A E D Q L G A R V G Y I E L D L N S G K I L
sign...
AmpR
XmnI
TGAGAGTTTTCGCCCCGAAGAACGTTTTCCAATGATGAGCACTTTTAAAGTTCTGCTATGTGGCGCGGTATTATCCCGTATTGACGCCGGGCAAGAGCAA
1800
ACTCTCAAAAGCGGGGCTTCTTGCAAAAGGTTACTACTCGTGAAAATTTCAAGACGATACACCGCGCCATAATAGGGCATAACTGCGGCCCGTTCTCGTT
E S F R P E E R F P M M S T F K V L L C G A V L S R I D A G Q E Q
AmpR


ScaI
CTCGGTCGCCGCATACACTATTCTCAGAATGACTTGGTTGAGTACTCACCAGTCACAGAAAAGCATCTTACGGATGGCATGACAGTAAGAGAATTATGCA
GAGCCAGCGGCGTATGTGATAAGAGTCTTACTGAACCAACTCATGAGTGGTCAGTGTCTTTTCGTAGAATGCCTACCGTACTGTCATTCTCTTAATACGT
1900
90 95 100 105 110 115 120
L G R R I H Y S Q N D L V E Y S P V T E K H L T D G M T V R E L C
AmpR
TsoI
PvuI
GTGCTGCCATAACCATGAGTGATAACACTGCGGCCAACTTACTTCTGACAACGATCGGAGGACCGAAGGAGCTAACCGCTTTTTTGCACAACATGGGGGA
CACGACGGTATTGGTACTCACTATTGTGACGCCGGTTGAATGAAGACTGTTGCTAGCCTCCTGGCTTCCTCGATTGGCGAAAAAACGTGTTGTACCCCCT
2000
125 130 135 140 145 150 155
S A A I T M S D N T A A N L L L T T I G G P K E L T A F L H N M G D
AmpR
TCATGTAACTCGCCTTGATCGTTGGGAACCGGAGCTGAATGAAGCCATACCAAACGACGAGCGTGACACCACGATGCCTGTAGCAATGGCAACAACGTTG
AGTACATTGAGCGGAACTAGCAACCCTTGGCCTCGACTTACTTCGGTATGGTTTGCTGCTCGCACTGTGGTGCTACGGACATCGTTACCGTTGTTGCAAC
2100
160 165 170 175 180 185
H V T R L D R W E P E L N E A I P N D E R D T T M P V A M A T T L
AmpR
FspI
NmeAIII
CGCAAACTATTAACTGGCGAACTACTTACTCTAGCTTCCCGGCAACAATTAATAGACTGGATGGAGGCGGATAAAGTTGCAGGACCACTTCTGCGCTCGG
GCGTTTGATAATTGACCGCTTGATGAATGAGATCGAAGGGCCGTTGTTAATTATCTGACCTACCTCCGCCTATTTCAACGTCCTGGTGAAGACGCGAGCC
2200
190 195 200 205 210 215 220
R K L L T G E L L T L A S R Q Q L I D W M E A D K V A G P L L R S
AmpR
BsaI
BmrI
CCCTTCCGGCTGGCTGGTTTATTGCTGATAAATCTGGAGCCGGTGAGCGTGGGTCTCGCGGTATCATTGCAGCACTGGGGCCAGATGGTAAGCCCTCCCG
GGGAAGGCCGACCGACCAAATAACGACTATTTAGACCTCGGCCACTCGCACCCAGAGCGCCATAGTAACGTCGTGACCCCGGTCTACCATTCGGGAGGGC
2300
225 230 235 240 245 250 255
A L P A G W F I A D K S G A G E R G S R G I I A A L G P D G K P S R
AmpR
AhdI
TATCGTAGTTATCTACACGACGGGGAGTCAGGCAACTATGGATGAACGAAATAGACAGATCGCTGAGATAGGTGCCTCACTGATTAAGCATTGGTAACTG
ATAGCATCAATAGATGTGCTGCCCCTCAGTCCGTTGATACCTACTTGCTTTATCTGTCTAGCGACTCTATCCACGGAGTGACTAATTCGTAACCATTGAC
2400
260 265 270 275 280 285
I V V I Y T T G S Q A T M D E R N R Q I A E I G A S L I K H W *
AmpR
TCAGACCAAGTTTACTCATATATACTTTAGATTGATTTAAAACTTCATTTTTAATTTAAAAGGATCTAGGTGAAGATCCTTTTTGATAATCTCATGACCA
AGTCTGGTTCAAATGAGTATATATGAAATCTAACTAAATTTTGAAGTAAAAATTAAATTTTCCTAGATCCACTTCTAGGAAAAACTATTAGAGTACTGGT
2500
AAATCCCTTAACGTGAGTTTTCGTTCCACTGAGCGTCAGACCCCGTAGAAAAGATCAAAGGATCTTCTTGAGATCCTTTTTTTCTGCGCGTAATCTGCTG
TTTAGGGAATTGCACTCAAAAGCAAGGTGACTCGCAGTCTGGGGCATCTTTTCTAGTTTCCTAGAAGAACTCTAGGAAAAAAAGACGCGCATTAGACGAC
2600
ori
CTTGCAAACAAAAAAACCACCGCTACCAGCGGTGGTTTGTTTGCCGGATCAAGAGCTACCAACTCTTTTTCCGAAGGTAACTGGCTTCAGCAGAGCGCAG
GAACGTTTGTTTTTTTGGTGGCGATGGTCGCCACCAAACAAACGGCCTAGTTCTCGATGGTTGAGAAAAAGGCTTCCATTGACCGAAGTCGTCTCGCGTC
2700
ori
ATACCAAATACTGTTCTTCTAGTGTAGCCGTAGTTAGGCCACCACTTCAAGAACTCTGTAGCACCGCCTACATACCTCGCTCTGCTAATCCTGTTACCAG
TATGGTTTATGACAAGAAGATCACATCGGCATCAATCCGGTGGTGAAGTTCTTGAGACATCGTGGCGGATGTATGGAGCGAGACGATTAGGACAATGGTC
2800
ori


AlwNI
TGGCTGCTGCCAGTGGCGATAAGTCGTGTCTTACCGGGTTGGACTCAAGACGATAGTTACCGGATAAGGCGCAGCGGTCGGGCTGAACGGGGGGTTCGTG
ACCGACGACGGTCACCGCTATTCAGCACAGAATGGCCCAACCTGAGTTCTGCTATCAATGGCCTATTCCGCGTCGCCAGCCCGACTTGCCCCCCAAGCAC
2900
ori
CACACAGCCCAGCTTGGAGCGAACGACCTACACCGAACTGAGATACCTACAGCGTGAGCTATGAGAAAGCGCCACGCTTCCCGAAGGGAGAAAGGCGGAC
GTGTGTCGGGTCGAACCTCGCTTGCTGGATGTGGCTTGACTCTATGGATGTCGCACTCGATACTCTTTCGCGGTGCGAAGGGCTTCCCTCTTTCCGCCTG
3000
ori
AGGTATCCGGTAAGCGGCAGGGTCGGAACAGGAGAGCGCACGAGGGAGCTTCCAGGGGGAAACGCCTGGTATCTTTATAGTCCTGTCGGGTTTCGCCACC
TCCATAGGCCATTCGCCGTCCCAGCCTTGTCCTCTCGCGTGCTCCCTCGAAGGTCCCCCTTTGCGGACCATAGAAATATCAGGACAGCCCAAAGCGGTGG
3100
ori
TCTGACTTGAGCGTCGATTTTTGTGATGCTCGTCAGGGGGGCGGAGCCTATGGAAAAACGCCAGCAACGCGGCCTTTTTACGGTTCCTGGCCTTTTGCTG
AGACTGAACTCGCAGCTAAAAACACTACGAGCAGTCCCCCCGCCTCGGATACCTTTTTGCGGTCGTTGCGCCGGAAAAATGCCAAGGACCGGAAAACGAC
3200
ori
PciI
GCCTTTTGCTCACATGTTCTTTCCTGCGTTATCCCCTGATTCTGTGGATAACCGTATTACCGCCTTTGAGTGAGCTGATACCGCTCGCCGCAGCCGAACG
CGGAAAACGAGTGTACAAGAAAGGACGCAATAGGGGACTAAGACACCTATTGGCATAATGGCGGAAACTCACTCGACTATGGCGAGCGGCGTCGGCTTGC
3300
SapI
BspQI
ACCGAGCGCAGCGAGTCAGTGAGCGAGGAAGCGGAAG 3′
3337
TGGCTCGCGTCGCTCAGTCACTCGCTCCTTCGCCTTC 5′

Appendix 3

The sequencing results of *N*- and *C*- terminally fused MLB, SNA, and GNA to GFP in pGFPuv expression vector

MLB *N*-terminally cloned in pGFPuv Expression Vector

NNNNNNNNNNNNNGGGCGATTGGGCCCTCTAGATGCATGCTCGAGCGGCCGCCAGTGTGAT
GGATATCTGCAGAATTCAGGACAGCGGAGCTCGCTGCAGCTGCAGATGATGTTACCTGCAGTGC
TTCGGGACCTACGGTGCGGATTGTGGGTCGAAATGGCATGTGCGTGGACGTCCGAGATGACGAT
TTCCACGATGGAAATCAGATACAGTTGTGGCCCTCCAAGTCCAACAATGATCCGAATCAGTTGTG
GACGATCAAAAGGGATGGAACCATTCGATCCAATGGCAGCTGCTTGACCACGTATGGCTATACT
GCTGGCGTCTATGTGATGATCTTCGACTGTAATACTGCTGTGCGGGAGGCCACTATTTGGCAGAT
ATGGGGCAATGGGACCGTCATCAATCCAAGATCCAATCTGGTTTTGGCAGCATCATCTGGAATCA
AAGGCACTACGCTTACGGTGCAAACACTGGATTACACGTTGGGACAGGGCTGGCTTGCCGGTAA
TGATACCGCCCCACGCGAGGTGACCATATATGGTTTCAGGGACCTTTGCATGGAATCAAATGGAG
GGAGTGCGTGGGTGGAGACGTGCGTGAGTAGCCAACAGAACCAAAGATGGGCTTTGTACGGGG
ATGGTTCTATACGCCCCAAACAAAACCAAGACCAATGCCTCACCTGTGGGAGAGACTCCGTTTCA
ACAGTAATCAATATAGTTAGCTGCAGCGCTGGATCGTCTGGGCAGCGATGGGTGTTTACCAATGA
AGGGGCCATTTTGAATTTAAAGAATGGGTTGGCCATGGATGTGGCGCAAGCAATCCAAAGCTCC
GCCGAATAATTATCTATCCTGCCACAGGAAAACCAATCAAATGTGGCTTCCCGTGCCAGCTGCAT
AAGAATTCTCATGCTCTGAATTCCAGCACACTGGCGGCCGTTACTAGTGGATCCGAGCTCGGTAC
CAAGCTTGATGCATAGCTTGAGTATTCTACGCGTCACCTAATAGCTTGGCGTAATCATGGTCATAG
CTTGTTTCCTGTGTGAAATTGTTATCCGCTCAC

MLB *C*-terminally cloned in pGFPuv Expression Vector

GNNNNNNNNNNATAGGGCGATTGGGCCCTCTAGATGCATGCTCGAGCGGCCGCCAGTGTGAT
GGATATCTGCAGAATTCAGGACAGCGTCTAGAAAAAGATGATGTTACCTGCAGTGCTTCGGGAC
CTACGGTGCGGATTGTGGGTCGAAATGGCATGTGCGTGGACGTCCGAGATGACGATTTCCACGA
TGGAAATCAGATACAGTTGTGGCCCTCCAAGTCCAACAATGATCCGAATCAGTTGTGGACGATCA
AAAGGGATGGAACCATTCGATCCAATGGCAGCTGCTTGACCACGTATGGCTATACTGCTGGCGTC
TATGTGATGATCTTCGACTGTAATACTGCTGTGCGGGAGGCCACTATTTGGCAGATATGGGGCAA
TGGGACCGTCATCAATCCAAGATCCAATCTGGTTTTGGCAGCATCATCTGGAATCAAAGGCACTA
CGCTTACGGTGCAAACACTGGATTACACGTTGGGACAGGGCTGGCTTGCCGGTAATGATACCGC
CCCACGCGAGGTGACCATATATGGTTTCAGGGACCTTTGCATGGAATCAAATGGAGGGAGTGCG
TGGGTGGAGACGTGCGTGAGTAGCCAACAGAACCAAAGATGGGCTTTGTACGGGGATGGTTCT
ATACGCCCCAAACAAAACCAAGACCAATGCCTCACCTGTGGGAGAGACTCCGTTTCAACAGTAAT
CAATATAGTTAGCTGCAGCGCTGGATCGTCTGGGCAGCGATGGGTGTTTACCAATGAAGGGGCC
ATTTTGAATTTAAAGAATGGGTTGGCCATGGATGTGGCGCAAGCAAATCCAAAGCTCCGCCGAAT
AATTATCTATCCTGCCACAGGAAAACCAAATCAAATGTGGCTTCCCGTGCCAGGTACCTCATGCCC
TGAATTCCAGCACACTGGCGGCCGTTACTAGTGGATCCGAGCTCGGTACCAAGCTTGATGCATAG
CTTGAGTATTCTAACGCGTCACCTAAATAGCTTGNGTAATCATGGTCATAGCTGTTTCCTGTGTGA
AATTGTATCCGCTCACATT

SNA *N*-terminally cloned in pGFPuv Expression Vector

GGNNNNNNNCCTNTNGGGNGATTGGGCCCTCTAGATGCATGCTCGAGCGGCCGCCAGTGTGAT
GGATATCTGCAGAATTCAGGACAGCGGAGCTCGCTGCAGCTGCAGGGGGCGAGTACGAAAAAG
TATGTTCGGTGGTAGAGGTAACAAGGCGCATCAGTGGTTGGGATGGATTGTGTGTGGACGTGA
GGTATGGGCACTACATCGATGGGAATCCCGTCCAGCTGCGGCCGTGTGGAAATGAATGTAACCA
ACTATGGACGTTCCGCACTGATGGAACAATCCGGTGGTTGGGTAAATGCCTGACTGCCTCAAGCT
CTGTCATGATATACGATTGTAATACTGTTCCTCCAGAGGCCACTAAGTGGGTAGTATCTATTGACG
GCACCATCACCAATCCTCACTCAGGACTCGTCCTTACAGCTCCTCAAGCTGCAGAGGGAACCGCC
CTGTCTCTGGAGAACAATATCCATGCCGCTAGGCAAGGTTGGACTGTAGGAGATGTAGAGCCCC
TCGTTACTTTTATTGTGGGATATAAACAAATGTGCTTGAGGGAAAACGGTGAAAACAATTTTGTA
TGGTTGGAGGACTGCGTTCTCAACAGGGTGCAGCAAGAGTGGGCACTCTATGGCGACGGCACCA
TTCGAGTAAACAGTAATCGTAGCCTATGTGTGACCTCCGAAGACCACGAGCCCAGTGATCTTATC
GTCATTCTCAAGTGCGAAGGGTCGGGCAACCAGCGCTGGGTATTCAACACCAACGGTACCATCTC
AAACCCAAACGCTAAACTACTTATGGACGTTGCACAACGCGATGTCTCTCTTCGAAAAATCATTCT
CTATCGGCCCACTGGGAATCCTAACCAGCAATGGATAACTACCACCCATCCAGCTGCTGCATAAT
GAATTCTCATGCCCTGAATTCCAGCACACTGGCGGCCGTTACTAGTGGATCCGAGCTCGGTACCA
AGCTTGATGCATAGCTTGAGTATTCTAACGCGTCACCTAAATAGCTTGGNGTAATCATGGGTCAT
AGCTGTTTCCTGTGTGAAATTGTTATCCG

SNA *C*-terminally cloned in pGFPuv Expression Vector

GNNNNNANNCNTNTANGGCGATTGGAAGCTTGGGGGGCGAGTACGAAAAAGTATGTTCGGTG
GTAGAGGTAACAAGGCGCATCAGTGGTTGGGATGGATTGTGTGTGGACGTGAGGTATGGGCAC
TACATCGATGGGAATCCCGTCCAGCTGCGGCCGTGTGGAAATGAATGTAACCAACTATGGACGTT
CCGCACTGATGGAACAATCCGGTGGTTGGGTAAATGCCTGACTGCCTCAAGCTCTGTCATGATAT
ACGATTGTAATACTGTTCCTCCAGAGGCCACTAAGTGGGTAGTATCTATTGACGGCACCATCACC
AATCCTCACTCAGGACTCGTCCTTACAGCTCCTCAAGCTGCAGAGGGAACCGCCCTGTCTCTGGA
GAACAATATCCATGCCGCTAGGCAAGGTTGGACTGTAGGAGATGTAGAGCCCCTCGTTACTTTTA
TTGTGGGATATAAACAAATGTGCTTGAGGGAAAACGGTGAAAACAATTTTGTATGGTTGGAGGA
CTGCGTTCTCAACAGGGTGCAGCAAGAGTGGGCACTCTATGGCGACGGCACCATTCGAGTAAAC
AGTAATCGTAGCCTATGTGTGACCTCCGAAGACCACGAGCCCAGTGATCTTATCGTCATTCTCAA
GTGCGAAGGGTCGGGCAACCAGCGCTGGGTATTCAACACCAACGGTACCATCTCAAACCCAAAC
GCTAAACTACTTATGGACGTTGCACAACGCGATGTCTCTCTTCGAAAAATCATTCTCTATCGGCCC
ACTGGGAATCCTAACCAGCAATGGATAACTACCACCCATCCAGCTACTCTAGACNCCAGGGTTTT
CCCAGTCNNGACGTGTAAAACGACGGCCAGTGAATTGTAATACGACTCACTATAGGGCGAATTG
GGC

GNA *N*-terminally cloned in pGFPuv Expression Vector

ANNNNNNNNNNTNNNNGGCGAATTGGGCCCTCTAGATGCATGCTCGAGCGGCCGCCAGTGTG
ATGGATATCTGCAGAAAACNGGGGAGCTCGCTGCAGCTAAGGCAAGTCTCCTCATTTTGGCCGC
CATCTTCCTTGGTGTCATCACACCATCTTGCCTGAGTGACAATATTTTGTACTCCGGTGAGACTCTC
TCTACAGGGGAATTTCTCAACTACGGAAGTTTCGTTTTTATCATGCAAGAGGACTGCAATCTGGTC
TTGTACGACGTGGACAAGCCAATCTGGGCAACAAACACAGGTGGTCTCTCCCGTAGCTGCTTCCT
CAGCATGCAGACTGATGGGAACCTCGTGGTGTACAACCCATCGAACAAACCGATTTGGGCAAGC
AACACTGGAGGCCAAAATGGGAATTACGTGTGCATCCTACAGAAGGATAGGAATGTTGTGATCT
ACGGAACTGATCGTTGGGCTACTGGAACTCACACCGGACTTGTTGGAATTCCCGCATCGCCACCC
TCAGAGAAATATCCTACTGCTGGAAAGATAAAGCTTGTGACGGCAAAGGCTGCATAACTACTAGT
GCCTGAATTCCAGCACACTGGCGGCCGTTACTAGTGGATCCGAGCTCGGTACCAAGCTTGATGCA
TAGCTTGAGTATTCTAACGCGTCACCTAAATAGCTTGGCGTAATCATGGTCATAGCTGTTTCCTGT
GTGAAATTGTTATCCGCTCACAATTCCACACAACATACGAGCCGGAAGCATAAAGTGTAAAGCCT
GGGGTGCCTAATGAGTGAGCTAACTCACATTAATTGCGTTGCGCTCACTGCCCGCTTTCCAGTCG
GGAAACCTGTCGTGCCAGCTGCATTAATGAATCGGCCAACGCGCGGGGAGAGGCGGTTTGCGTA
TTGGGCGCTCTTCCGCTTCCTCGCTCACTGACTCGCTGCGCTCGGTCGTTCGGCTGCGGCGAGCG
GTATCAGCTCACTCAAAGGCGGTAATACNGTTATCCACANAATCNNGGGATACGCNNNAAGAAC
ATGTGAGCCAAANG

GNA *C*-terminally cloned in pGFPuv Expression Vector

GNNNNNNNNCTNNTNGGGGGCGATTGGGCCCTCTAGATGCATGCTCGAGCGGCCGCCAGTGT
GATGGATATCTGCAGAAAACAGGACAGCGTCTAGAAGCTAAGGCAAGTCTCCTCATTTTGGCCG
CCATCTTCCTTGGTGTCATCACACCATCTTGCCTGAGTGACAATATTTTGTACTCCGGTGAGACTCT
CTCTACAGGGGAATTTCTCAACTACGGAAGTTTCGTTTTTATCATGCAAGAGGACTGCAATCTGGT
CTTGTACGACGTGGACAAGCCAATCTGGGCAACAAACACAGGTGGTCTCTCCCGTAGCTGCTTCC
TCAGCATGCAGACTGATGGGAACCTCGTGGTGTACAACCCATCGAACAAACCGATTTGGGCAAG
CAACACTGGAGGCCAAAATGGGAATTACGTGTGCATCCTACAGAAGGATAGGAATGTTGTGATC
TACGGAACTGATCGTTGGGCTACTGGAACTCACACCGGACTTGTTGGAATTCCCGCATCGCCACC
CTCAGAGAAATATCCTACTGCTGGAAAGATAAAGCTTGTGACGGCAAAGAAGGTACCTCATGCC
CTGAATTCCAGCACACTGGCGGCCGTTACTAGTGGATCCGAGCTCGGTACCAAGCTTGATGCATA
GCTTGAGTATTCTAACGCGTCACCTAAATAGCTTGGCGTAATCATGGTCATAGCTGTTTCCTGTGT
GAAATTGTTATCCGCTCACAATTCCACACAACATACGAGCCGGAAGCATAAAGTGTAAAGCCTGG
GGTGCCTAATGAGTGAGCTAACTCACATTAATTGCGTTGCGCTCACTGCCCGCTTTCCAGTCGGG
AAACCTGTCGTGCCAGCTGCATTAATGAATCGGCCAACGCGCGGGGAGAGGCGGTTTGCGTATT
GGGCGCTCTTCCGCTTCCTCGCTCACTGACTCGCTGCGCTCGGTCGTTCGGCTGCGGCGAGCGGT
ATCAGCTCACTCAAANGCGGTAATACGGTTATCCNCAGAATCNNGGGGATAACGCAGNAAGAA
CATGTGAGCAAAGGCC

Appendix 4

The sequencing results of Lectin-Toxin fusion constructs cloned into pGFPuv expression vector

MLB-Melittin 1

GAGGTCGACTCTAGAGGATGATGTTACCTGTAGCGCAAGCGAACCGACCGTTCGTATTGTTGGTC
GTAATGGTATGTGTGTTGATGTGCGTGATGATGATTTTCATGATGGCAATCAGATTCAGCTGTGG
CCGAGCAAAAGCAATAATGATCCGAACCAGCTGTGGACCATTAAACGTGATGGCACCATTCGTA
GCAATGGTAGCTGTCTGACCACCTATGGTTATACCGCAGGCGTTTATGTTATGATCTTTGATTGTA
ATACCGCAGTTCGTGAAGCAACCATTTGGCAGATTTGGGGTAATGGTACAATTATCAATCCGCGT
AGCAATCTGGTTCTGGCAGCAAGCAGCGGTATTAAAGGCACCACCCTGACCGTTCAGACCCTGG
ATTATACCCTGGGTCAGGGTTGGCTGGCAGGTAATGATACCGCACCGCGTGAAGTTACCATTTAT
GGTTTTCGTGATCTGTGCATGGAAAGCAATGGTGGTAGCGTTTGGGTTGAAACCTGTGTTAGCA
GCCAGCAGAATCAGCGTTGGGCACTGTATGGTGATGGTAGCATTCGTCCGAAACAGAATCAGGA
TCAGTGTCTGACCTGTGGTCGTGATAGCGTTAGCACCGTTATTAACATTGTTAGCTGCAGCGCAG
GTAGCAGCGGTCAGCGCTGGGTTTTTACCAATGAAGGTGCAATTCTGAACCTGAAAAATGGTCTG
GCAATGGATGTTGCACAGGCAAATCCGAAACTGCGTCGTATTATCATTTATCCGGCAACCGGTAA
ACCGAATCAGATGTGGCTGCCGGTTCCGCCTCCGCCTCCTCCACCACCTCCGCCTGGTATTCCGGT
TAGCCTGCGTAGCAAAGGTATTGGTGCCGTTCTGAAAGTGCTGACCACCGGTCTGCCTGCACTGA
TTAGCTGGATTAAACGTAAACGTCAGCAGCATCACCACCATCCATTCATTATTGAATTTCCACCTG
GAGACCGCCGGTTCGCCTATCACAT

MLB-Melittin 2

NNNNNGNTTNNNANNGAAAGCGGGCAAGTGAGCGNANGCATTAANNNTGNANTTAGCTCNC
TCNTTAGGCNCCCCAGGCTTTACACTTTANGCTTCCGGCTCGTATGNTNNNGNGGAANTNNTGA
GCGGATAACATTTCACACAGGAAACAGCTATGACCATGATTACGCCAAGCTTGCATGCCTGCAGG
TCGACTCTAGAGGATGATGTTACCTGCAGTGCTTCGGGACCTACGGTGCGGATTGTGGGTCGAA
ATGGCATGTGCGTGGACGTCCGAGATGACGATTTCCACGATGGAAATCAGATACAGTTGTGGCC
CTCCAAGTCCAACAATGATCCGAATCAGTTGTGGACGATCAAAAGGGATGGAACCATTCGATCCA
ATGGCAGCTGCTTGACCACGTATGGCTATACTGCTGGCGTCTATGTGATGATCTTCGACTGTAAT
ACTGCTGTGCGGGAGGCCACTATTTGGCAGATATGGGGCAATGGGACCGTCATCAATCCAAGAT
CCAATCTGGTTTTGGCAGCATCATCTGGAATCAAAGGCACTACGCTTACGGTGCAAACACTGGAT
TACACGTTGGGACAGGGCTGGCTTGCCGGTAATGATACCGCCCCACGCGAGGTGACCATATATG
GTTTCAGGGACCTTTGCATGGAATCAAATGGAGGGAGTGCGTGGGTGGAGACGTGCGTGAGTA
GCCAACAGAACCAAAGATGGGCTTTGTACGGGGATGGTTCTATACGCCCCAAACAAAACCAAGA
CCAATGCCTCACCTGTGGGAGAGACTCCGTTTCAACAGTAATCAATATAGTTAGCTGCAGCGCTG
GATCGTCTGGGCAGCGATGGGTGTTTACCAATGAAGGGGCCATTTTGAATTTAAAGAATGGGTT
GGCCATGGATGTGGCGCAAGCAAATCCAAAGCTCCGCCGAATAATTATCTATCCTGCCACAGGAA
AACCAAATCAAATGTGGCTTCCCGTGCCACGGGTACCGCCTCCTCCGCCACCTCCGCCTCCGCCTG
GTATTGGTGCAGTTCTGAAAGTTCTGACCACCGGTCTGCCTGCACTGATTAGCTGGATTAAACGT
AAACGTCAGCAGCATCATCACCATCATCATTAATGAATTCCAACTGAGCGCCGGTCGCTACCATTA
CCAACTTGTCTGGTGTCAAAAATAATAGGCCTACTAGTCGGCCGTACGGGCCNNNNNTCTCGCG
NGTTNNNNNNNN

1st Domain-Melittin

ATGACCATGATTACGCCAAGCTTGGATGTGCGTGATGATGATTTTCATGATGGCAATCAGATTCA
GCTGTGGCCGAGCAAAAGCAATAATGATCCGAACCAGCTGTGGACCATTAAACGTGATGGCACC
ATTCGTAGCAATGGTAGCTGTCTGACCACCTATGGTTATACCGCAGGCGTTTATGTTATGATCTTT
GATTGTAATACCGCAGTTCGTGAAGCAACCATTTGGCAGATTTGGGGTAATGGTACAATTATCAA
TCCGCGTAGCAATCTGGTTCTGGCAGCAAGCAGCGGTATTAAAGGCACCACCCTGACCGTTCAAA
CCCTGGATTATACCCTGGGTCAGCCTCCTCCGCCACCTCCGCCTCCGCCTGGTATTCCGGTTAGCC
TGCGTAGCAAAGGTATTGGTGCAGTTCTGAAAGTGCTGACCACCGGTCTGCCTGCACTGATTAGC
TGGATTAAACGTAAACGTCAGCAGCATCATCACCATCATCATTAATGAATTCCAACTGAGCGCCG
GTCGCTACCATTACCAACTTGTCTGGTGTCAAAAATAATAGGCCTACTAGTCGGCCGTACGGGCC
CTTTCGTCTCGCGCGTTTCGGTGATGACGGTGAAAACCTCTGACACATGCAGCTCCCGGAGACGG
TCACAGCTTGTCTGTAAGCGGATGCCGGGAGCAGACAAGCCCGTCAGGGCGCGTCAGCGGGTGT
TGGCGGGTGTCGGGGCTGGCTTAACTATGCGGCATCAGAGCAGATTGTACTGAGAGTGCACCAT
ATGCGGTGTGAAATACCGCACAGATGCGTAAGGAGAAAATACCGCATCAGGCGGCCTTAAGGG
CCTCGTGATACGCCTATTTTTATAGGTTAATGTCATGATAATAATGGTTTCTTAGACGTCAGGTGG
CACTTTTCGGGGAAATGTGCGCGGAACCCCTATTTGG

1st Domain-Mastoparan

GAGGTCGACTCTAGAGGATGTGCGTGATGATGATTTTCATGATGGCAATCAGATTCAGCTGTGGC
CGAGCAAAAGCAATAATGATCCGAACCAGCTGTGGACCATTAAACGTGATGGCACCATTCGTAG
CAATGGTAGCTGTCTGACCACCTATGGTTATACCGCAGGCGTTTATGTTATGATCTTTGATTGTAA
TACCGCAGTTCGTGAAGCAACCATTTGGCAGATTTGGGGTAATGGTACAATTATCAATCCGCGTA
GCAATCTGGTTCTGGCAGCAAGCAGCGGTATTAAAGGCACCACCCTGACCGTTCAAACCCTGGAT
TATACCCTGGGTCAGCCTCCTCCGCCACCTCCGCCTCCGCCTGGTATTCCGGTTAGCCTGCGTAGC
AAAATTAACCTGAAAGCACTGGCAGCCCTGGCGAAAAAAATCCTGCATCATCACCATCATCATTA
ATGAATTCCAACTGAGCGCCGGTCGCTACCATTACCAACTTGTCTGGTGTCAAAAATAATAGGCC
TACTAGTCGGCCGTACGGGCCCTTTCGTCTCGCGCGTTTCGGTGATGACGGTGAAAACCTCTGAC
ACATGCAGCTCCCGGAGACGGTCACAGCTTGTCTGTAAGCGGATGCCGGGAGCAGACAAGCCCG
TCAGGGCGCGTCAGCGGGTGTTGGCGGGTGTCGGGGCTGGCTTAACTATGCGGCATCAGAGCA
GATTGTACTGAGAGTGCACCATATGCGGTGTGAAATACCGCACAGATGCGTAAGGAGAAAATAC
CGCATCAGGCGGCCTTAAGGGCCTCGTGATACGCCTATTTTTATAGGTTAATGTCATGATAATAAT
GGTTTCTTAGACGTCAGGTGGCACTTTTCGGGGAAATGTGCGCGGAACCCCTATTTGTTTATTTTT
CTAAATA

MLB-Mastoparan 1

GGAAAAAAACACCTCTGGTGAGGTCGACTCTAGAGGATGATGTTACCTGTAGCGCAAGCGAACC
GACCGTTCGTATTGTTGGTCGTAATGGTATGTGTGTTGATGTGCGTGATGATGATTTTCATGATG
GCAATCAGATTCAGCTGTGGCCGAGCAAAAGCAATAATGATCCGAACCAGCTGTGGACCATTAA
ACGTGATGGCACCATTCGTAGCAATGGTAGCTGTCTGACCACCTATGGTTATACCGCAGGCGTTT
ATGTTATGATCTTTGATTGTAATACCGCAGTTCGTGAAGCAACCATTTGGCAGATTTGGGGTAAT
GGTACAATTATCAATCCGCGTAGCAATCTGGTTCTGGCAGCAAGCAGCGGTATTAAAGGCACCAC
CCTGACCGTTCAGACCCTGGATTATACCCTGGGTCAGGGTTGGCTGGCAGGTAATGATACCGCAC
CGCGTGAAGTTACCATTTATGGTTTTCGTGATCTGTGCATGGAAAGCAATGGTGGTAGCGTTTGG
GTTGAAACCTGTGTTAGCAGCCAGCAGAATCAGCGTTGGGCACTGTATGGTGATGGTAGCATTC
GTCCGAAACAGAATCAGGATCAGTGTCTGACCTGTGGTCGTGATAGCGTTAGCACCGTTATTAAC
ATTGTTAGCTGCAGCGCAGGTAGCAGCGGTCAGCGCTGGGTTTTTACCAATGAAGGTGCAATTCT
GAACCTGAAAAATGGTCTGGCAATGGATGTTGCACAGGCAAATCCGAAACTGCGTCGTATTATCA
TTTATCCGGCAACCGGTAAACCGAATCAGATGTGGCTGCCGGTTCCGCCTCCGCCTCCTCCACCAC
CTCCGCCTGGTATTCCGGTTAGCCTGCGTAGCAAAATTAACCTGAAAGCACTGGCAGCCCTGGCG
AAAAAAATCCTGCATCACATCATCATATGATCACCTGAACGGCGTTCGCTACATTACCACTTGTCT
GCGTCCAAAAATAATAGTCCTACTAGTCGGCCGTAG

MLB-Mastoparan 2

GATTCNTNAATGCAGNTGGCANGNCNNTTNNCCGACTGGAAAGCGGGCAGTGAGCGCAACGC
AATTTAATGTGAGTTAGCTCACTCNTTAGGCACCCCAGGCTTTACACTTTATGCTTCCGGCTCGTA
TGTTGTGTGGAATTGTGAGCGGATAACAATTTCACACAGGAAACAGCTATGACCATGATTACGCC
AAGCTTGCATGCCTGCAGGTCGACTCTAGAGGATGATGTTACCTGCAGTGCTTCGGGACCTACGG
TGCGGATTGTGGGTCGAAATGGCATGTGCGTGGACGTCCGAGATGACGATTTCCACGATGGAAA
TCAGATACAGTTGTGGCCCTCCAAGTCCAACAATGATCCGAATCAGTTGTGGACGATCAAAAGGG
ATGGAACCATTCGATCCAATGGCAGCTGCTTGACCACGTATGGCTATACTGCTGGCGTCTATGTG
ATGATCTTCGACTGTAATACTGCTGTGCGGGAGGCCACTATTTGGCAGATATGGGGCAATGGGA
CCGTCATCAATCCAAGATCCAATCTGGTTTTGGCAGCATCATCTGGAATCAAAGGCACTACGCTTA
CGGTGCAAACACTGGATTACACGTTGGGACAGGGCTGGCTTGCCGGTAATGATACCGCCCCACG
CGAGGTGACCATATATGGTTTCAGGGACCTTTGCATGGAATCAAATGGAGGGAGTGCGTGGGTG
GAGACGTGCGTGAGTAGCCAACAGAACCAAAGATGGGCTTTGTACGGGGATGGTTCTATACGCC
CCAAACAAAACCAAGACCAATGCCTCACCTGTGGGAGAGACTCCGTTTCAACAGTAATCAATATA
GTTAGCTGCAGCGCTGGATCGTCTGGGCAGCGATGGGTGTTTACCAATGAAGGGGCCATTTTGA
ATTTAAAGAATGGGTTGGCCATGGATGTGGCGCAAGCAAATCCAAAGCTCCGCCGAATAATTATC
TATCCTGCCACAGGAAAACCAAATCAAATGTGGCTTCCCGTGCCACGGGTACCGCCTCCTCCGCC
ACCTCCGCCTCCGCCTATTAACCTGAAAGCACTGGCAGCCCTGGCGAAAAAAATCCTGCATCATC
ATCACCATCATTAATGAATTCCAACTGAGCGCCGGTCGCTACCATTACCAACTTGTCTGGTGTCAA
AAATAATAGGCCTACTAGTCGGCCGTACGGGCCNNTNGTCTCGCGNGNNNNNNNNNNNNN

2nd Domain-Melittin

CTGATACCCGCTCGCCGCAGCCGAACGACCGAGCGCAGCGAGTCAGTGAGCGAGGAAGCGGAA
GAGCNCCCAATACGCAAACCGCCTCTCCCCGCGCGTTGGCCGATTCATTAATGCAGCTGGCACGA
CAGGTTTCCCGACTGGAAAGCGGGCAGTGAGCGCAACGCAATTAATGTGAGTTAGCTCACTCATT
AGGCACCCCAGGCTTTACACTTTATGCTTCCGGCTCGTATGTTGTGTGGAATTGTGAGCGGATAA
CAATTTCACACAGGAAACAGCTATGACCATGATTACGCCAAGCTTGGAAAGCAATGGTGGTAGC
GTTTGGGTTGAAACCTGTGTTAGCAGCCAGCAGAATCAGCGTTGGGCACTGTATGGTGATGGTA
GCATTCGTCCGAAACAGAATCAGGATCAGTGTCTGACCTGTGGTCGTGATAGCGTTAGCACCGTT
ATTAACATTGTTAGCTGTAGCGCAGGTAGCAGCGGTCAGCGCTGGGTTTTTACCAATGAAGGTGC
AATTCTGAATCTGAAAAATGGTCTGGCAATGGATGTTGCACAGGCAAATCCGAAACTGCGTCGTA
TTATCATTTATCCGGCAACCGGTAAACCGAATCAGCCTCCTCCGCCACCTCCGCCTCCGCCTGGTA
TTCCGGTTAGCCTGCGTAGCAAAGGTATTGGTGCCGTTCTGAAAGTTCTGACCACCGGTCTGCCT
GCACTGATTAGCTGGATTAAACGTAAACGTCAGCAGCATCATCACCACCATCATTAATGAATTCCA
ACTGAGCGCCGGTCGCTACCATTACCAANTTGTCTGGTGTCAAAAATAATAGGCCTACTAGTCGG
CCGTACGGGCCCTTTCNTCNCGCGNGTTNNNGNNNNNNN

2nd Domain-Mastoparan

GAGGTCGACTCTAGAGGAAAGCAATGGTGGTAGCGTTTGGGTTGAAACCTGTGTTAGCAGCCAG
CAGAATCAGCGTTGGGCACTGTATGGTGATGGTAGCATTCGTCCGAAACAGAATCAGGATCAGT
GTCTGACCTGTGGTCGTGATAGCGTTAGCACCGTTATTAACATTGTTAGCTGTAGCGCAGGTAGC
AGCGGTCAGCGCTGGGTTTTTACCAATGAAGGTGCAATTCTGAATCTGAAAAATGGTCTGGCAAT
GGATGTTGCACAGGCAAATCCGAAACTGCGTCGTATTATCATTTATCCGGCAACCGGTAAACCGA
ATCAGCCTCCTCCGCCACCTCCGCCTCCGCCTGGTATTCCGGTTAGCCTGCGTAGCAAAATTAACC
TGAAAGCACTGGCAGCCCTGGCGAAAAAAATCCTGCATCATCATCACCATCATTAATGAATTCCA
ACTGAGCGCCGGTCGCTACCATTACCAACTTGTCTGGTGTCAAAAATAATAGGCCTACTAGTCGG
CCGTACGGGCCCTTTCGTCTCGCGCGTTTCGGTGATGACGGTGAAAACCTCTGACACATGCAGCT
CCCGGAGACGGTCACAGCTTGTCTGTAAGCGGATGCCGGGAGCAGACAAGCCCGTCAGGGCGC
GTCAGCGGGTGTTGGCGGGTGTCGGGGCTGGCTTAACTATGCGGCATCAGAGCAGATTGTACTG
AGAGTGCACCATATGCGGTGTGAAATACCGCACAGATGCGTAAGGAGAAAATACCGCATCAGGC
GGCCTTAAGGGCCTCGTGATACGCCTATTTTTATAGGTTAATGTCATGATAATAATGGTTTCTTAG
ACGTCAGGTGGCACTTTTCGGGGAAATGTGCGCGGAACCCCTATTTGTTATTTTTCTAAATACATT
CA