The Investigation of Cry toxins from *Bacillus thuringiensis* and their Potential use as Novel Carbohydrate Binding Proteins

Thesis submitted for the degree of
Doctor of Philosophy

by

Norah Cassidy, B. Sc.

Supervised by
Michael O'Connell, B. A (Mod), PhD
and
Brendan O'Connor, B.Sc., PhD.

School of Biotechnology
Dublin City University
Ireland

July 2013

Declaration

I hereby certify that this material, which I now submit for assessment on the programme of study leading to the award of Degree of Doctor of Philosophy is entirely my own work, that I have exercised reasonable care to ensure that the work is original, and does not to the best of my knowledge breach any law of copyright, and has not been taken from the work of others, save and to the extent that such work has been cited and acknowledged within the text of my work.

Signed: _____ ID No: 55067065____

Date: _____

Dedicated to the memory of John Cassidy

The world has more integrity, strength, courage, and love because you lived by those qualities.

You were loved by all and a hero to us. You are always in our hearts

Acknowledgements

I would firstly like to thank my supervisors Mick and Brendan for their, guidance, support and patience throughout the last four years.

I want to thank everyone in the lab for their continuing help and advice especially Colm, Damien, Ruth, Mary, Paul, Roisin and Amy.

Thank you to all my friends, especially Hazel, Carol, Aine, Roya and Eva. Also to my friend and housemate Ross, for putting up with me and for all the fun we had over the last 3 years.

Thank you to my boyfriend Ronan, for believing in me and giving me encouragement when I needed it. Angela, Kieran and Carol thanks for making me feel like part of the family and for your continuing support.

To my sisters, Katie, Sarah and Anna, thank you for always being there for me in good times and bad, ye always cheer me up! A very special thank you to my wonderful Mum and Dad, without them, I would not be where I am today. They supported me through many years of studying, even when it seemed never-ending. Thank you Dad for looking down on me, your courage gave me the strength to keep going.

Abbreviations

2-D	2-dimensional
AAL	*Aleuria aurantia* lectin
Ala	alanine
ALP	alkaline phosphatase
APN	aminopeptidase-N
AmpR	ampicillin resistance
APS	ammonium persulphate
Asn	asparagine
Asp	aspartic acid
ATCC	American type culture collection
Bt	*Bacillus thuringiensis*
bp	base pair
BBMV	brush border membrane vesicles
Bin	binary-like
BSA	bovine serum albumin
CADR	cadherin receptor
CBD	carbohydrate binding domain
CBD VI	cellulose binding domain family VI
CBM	carbohydrate-binding module
CBM6	carbohydrate binding module family 6
CBP	carbohydrate-binding protein
Con A	concanavalin A
CP	capsular polysaccharide
Cry	crystal
CryD3	Cry1Ac domain III
CRP	catabolite gene activator
Cyt	cytotoxic
DBA	*Dolichos biflorus* agglutinin
dH2O	distilled H2O
DNA	deoxyribonucleic acid
dNTP	deoxynucleotide triphosphate
Dol-P	dolichol phosphate

ECL	*Erythrina cristagalli* lectin
EDTA	ethylenediaminetetraacetic acid
ELISA	enzyme linked immunosorbent assay
ELLA	enzyme linked lectin assay
Endo Hf	endoglycosidase Hf
ER	endoplasmic reticulum
FACS	Fluorescence-activated cell sorting
FPLC	fast protein liquid chromatography
Fuc	fucose
Gal	galactose
GalNAc	*N*-acetylgalactosamine
GalNAcT	*N*-acetylgalactosyl transferase
GCR	glyco-conjugate receptor
GFP	green fluorescent protein
GH	glycoside hydrolase
Glc	glucose
GlcA	glucuronic acid
GlcNAc	*N*-acetylglucosamine
Glu	glutamic acid
GNA	*Galanthus nivalis* agglutinin
GPI	glycosylphosphatidylinositol
GT	glycosyltransferase
GU	glucose units
HvALP	*Heliothis virescens* membrane-bound alkaline phosphatise
HPLC	high performance liquid chromatography
HRP	horseradish peroxidise
IMAC	immobilised metal affinity chromatography
IXE	ion Exchange Chromatography
IPTG	isopropyl-_-D-1-thiogalactopyranoside
kb	(kilo)base
kDa	(kilo)Dalton
Kdo	3-deoxy-D-*manno*-octulosonic acid
LAC	lectin affinity chromatography
LB	Luria Bertani

LOS	lipooligosaccharide
LPS	lipopolysaccharide
MAL I (I)	*Maackia amurensis* lectin I(I)
MALDI	matrix assisted laser desorption ionization
MALDI-TOF	matrix-assisted laser desorption/ionization time-of-flight
Man	mannose
MCS	multiple cloning site
MOPS	3-(N-morpholino)propanesulfonic acid
mRNA	messenger ribonucleic acid
MS	mass spectrometry
Mtx	mosquitocidal Cry toxins
NeuNAc	*N*-acetylneuraminic acid (sialic acid)
Ni-NTA	nickel-nitrilotriacetic acid
NCBI	national Centre for Biotechnology Information
NMR	nuclear magnetic resonance
NP-HPLC	normal Phase HPLC
NPL	*narcissus pseudonarcissus* Lectin
O.D	optical density
ORF	open reading frame
OT	oligosaccharyltransferase
PAGE	polyacrylamide gel electrophoresis
PBS	phosphate buffered saline
PCR	polymerase chain reaction
pgl	protein glycosylation
PIPES	piperazine-N,N_-bis(2-ethanesulfonic acid)
PNA	peanut agglutinin
PNGaseF	*N*-glycosidase F
PTM	post translational modification
PFTs	pore forming toxins
ppGalNAcT	polypeptide-N-acetylgalactosaminetransferases
RCA	*Ricinus communis* agglutinin
RNase	ribonuclease
rpm	revolutions per minute
SBA	soybean agglutinin

SDS	sodium dodecyl sulphate
Ser	serine
S-layer	surface-layer
SLP	surface layer protein
SNA	*Sambucus nigra* agglutinin
sp.	species
spp.	species (plural)
subsp.	subspecies
TBS(T)	tris buffered saline (with triton)
tCry1Ac	truncated Cry1Ac
TEMED	N,N,N_,N_-tetramethyl ethylenediamine
Thr	threonine
Tm	melting temperature
TMB	3,3_,5,5_-tetramethylbenzidine
TOF	time-of-flight
Tris	tris (hyroxymethyl) amino methane
UDP	undecaprenyl pyrophosphate
UV	ultraviolet
v/v	volume per volume
Val	valine
w/v	weight per volume
WGA	wheat germ agglutinin
WT	wild type
Xyl	xylose

Table of contents

List of Figures

Chapter 1

Chapter 2

Chapter 3

Chapter 5

Chapter 6

List of Tables

Chapter 1

Chapter 2

Chapter 4

Abstract

It is widely recognised that most proteins are subject to some form of post translational modifications such as glycosylation. Glycosylation has been shown to impact correct protein folding, protein stability, solubility, to aid in cell recognition and to help regulate cell processes. A deeper understanding into the effect of altered glycosylation patterns on cellular processes such as, cell recognition and protein function is necessary. Therefore, the development of new sensitive and accurate technologies that may be used to detect, separate and quantify glycans and glycoproteins is of increasing importance.

The objective of this work was to develop prokaryotic Cry proteins as novel carbohydrate-binding molecules that could be potentially used as ligands for the affinity based detection and separation of glycans and glycoproteins. Recombinant proteins tCry1Ac and CryD3 were designed based on the Cry1Ac toxin from *Bacillus thuringiensis*. tCry1Ac and CryD3 were cloned into pQE60 which contained a 6x histidine tag. The expression of tCry1Ac and CryD3 was carried out in *E. coli* JM109. tCry1Ac and CryD3 were purified by immobilised metal affinity chromatography (IMAC) with additional GalNAc/galactose washes to remove persistent contaminating proteins from *E. coli*. tCry1Ac and CryD3 carbohydrate binding abilities were examined by Enzyme linked lectin assay (ELLA). tCry1Ac was found to have specificity to GalNAc and the Tn antigen while CryD3 had specificity to α1-3, galactose. Site directed mutagenesis was carried out on the predicted binding sites of tCry1Ac and CryD3. Residues N480, Q483, Y487 and N521 for tCry1Ac and N45, Q48, Y52 and, N86 for CryD3 were mutated to alanine, expressed, purified and their carbohydrate specificity examined. The N86A mutant was the only protein that showed significant activity, binding to GalNAc and galactose on an ELLA. This suggests that the binding site for GalNAc and galactose in tCry1Ac and CryD3 is the same and may be altered with relative ease by mutagenesis. This result also confirmed that residue N86 plays an important role in the structure of the binding site and its recognition of the specific sugar as when mutated to alanine it changed the specificity from galactose to GalNAc.

Chapter 1

Introduction

1.1 Protein – Carbohydrate interactions

Protein – carbohydrate interactions are important in biological processes, in nature and in bio-analytical techniques. Many pathogens have developed carbohydrate binding proteins which allow them to recognise and target their host. These specific interactions are particularly important, and need to be studied to give us a deeper understanding of protein – carbohydrate interactions, which will enable the development of novel bio-analytical techniques.

1.2 Microbial Carbohydrate Binding proteins; lectins and toxins

Microorganisms have long been known to exploit host cell-surface glycans as recognition sites and receptors for cell attachment and tissue colonization. A large number of pathogenic species depend on these interactions for infection (Bektas and Rubenstein 2011; Williams and Davies, 2001; Rostand et al. 1997). Viruses, bacteria, fungi and protozoa express a variety of carbohydrate binding proteins including toxins and lectins (Varki et al. 2009). Many of these microbial lectins were originally detected based on their ability to aggregate red blood cells. The first microbial hemagglutinin identified was in the influenza virus, and it was shown by Alfred Gottschalk in the early 1950s to bind to erythrocytes and other cells though sialic acid residues of cell-surface glycoconjugates (Varki et al. 2009). In 1981, the viral hemagglutinin was crystallized to determine its structure. These studies accelerated the research into other types of microbial lectins (Varki et al. 2009).

Bacterial surface lectins were first described in the 1970s. These lectins also exhibited hemagglutinating activity, but their primary function is to facilitate attachment or adherence of bacteria to host cells, a prerequisite for bacterial colonization and infection (Varki et al. 2009). Certain bacteria also produce carbohydrate-binding protein toxins, whose actions often depend on glycan-binding subunits that allow the toxin to combine with membrane glyco-conjugates to deliver the functionally active toxic subunit across the membrane (Varki et al. 2009; Kitov et al. 2000). A number of secreted bacterial toxins also bind to glycans and are summarized in table 1.1. This project is interested in the Cry toxins from *Bacillus thuringiensis,* as they have been identified as having ‘lectin-like’ binding domains,

multiple receptors and they do not harm animals or humans. Therefore, these protein toxins could potentially be used as novel carbohydrate binding molecules.

Table 1.1: Examples of microbial toxins and their targeted glycosylated site. Cer indicates ceramide. Gangliosides are defined using the Svennerholm nomenclature. N-acetylgalactosamine (GalNAc), Galactose (Gal), Glucose (Glc), Mannose (Man), N-Acetylglucosamine (GlcNAc)

Microorganism	**Toxin**	**Proposed receptor sequence**	**Target tissue**
Bacillus thuringiensis	crystal toxins	Galβ1–3/6Galα/β1–3(±Glcβ1–6)GalNAcβ GlcNAcβ1–3Manβ1–4GlcβCer	intestinal epithelia of insects
Clostridium botulinum	botulinum toxins (A–E)	gangliosides GT1b, GQ1b	nerve membrane
Clostridium difficile	toxin A	GalNAcβ1–3Galβ1–4GlcNAcβ1–3Galβ1–4GlcβCer	large intestine
Clostridium tetani	tetanus toxin	Ganglioside GT1b	nerve membrane
Escherichia coli	heat-labile toxin	GM1	intestine
Shigella dysenteriae	Shiga toxin	Galα1–4GalβCer Galα1–4Galβ1–4GlcβCer	large intestine
Vibrio cholerae	cholera toxin	GM1	small intestine

1.3 Introduction and Background of the *Bacillus thuringiensis* toxins

Bacillus thuringiensis (Bt) is a gram positive insect pathogen. It has the ability to produce spores with parasporal crystals. These crystals contain toxins that become activated when the correct pH and environment is reached in the insect gut (Schnepf et al. 1998). These toxins, known as Cry (crystal) or Cyt (cytotoxic) have proven to be very important in controlling crop pests and mosquitoes (Bravo et al. 2011). Bt was first discovered in 1901 in Japan as the cause of wilt disease in silk worm (*Bombyx mori*) (Bravo et al. 2011). Since then many Bt strains have been identified all over the world (Crickmore 2013). Bt strains have been found to be active against Lepidoptera, Diptera, Coleoptera, Hymenoptera, Homoptera, Orthoptera and Mallophaga insect orders (Schnepf et al. 1998). Some Cry toxins have shown activity against nematodes, mites and protozoa (Wei et al. 2003; de Maagd et al. 2001; Crickmore et al. 1998; Schnepf et al. 1998).

Research started by Berliner in 1938, led to the first Bt formulation for use against insect pests on crops (Bravo et al. 2011). This work led to the first study of the toxins that were being expressed by Bt. The *cry* gene toxins were discovered on large plasmids flanked by transposable elements (Pigott and Ellar 2007). In the years following the focus was to clone these toxins and identify their toxicity to insects as well as examining their structure and mode of action (Adang 1991; Sekar et al. 1987; Adang et al. 1985). The success of Bt as a bioinsecticide was in the development of Bt crops that express a *cry* gene, which resulted in insect resistant crops (Adang et al. 1993; Dean and Adang 1992; Adang et al. 1986). The advantage of using Bt toxins is that they are specific to a limited number of insect species and there is no toxicity against humans or other organisms (Bravo et al. 2011). James, (2010) reported that there was more than 58 million hectares of Bt-maize or Bt-cotton grown worldwide.
The main disadvantage of Bt crops is the appearance of insect resistance to the Cry toxins. The emergence of resistance has been well documented for many species (Gassmann et al. 2011; Bagla 2010; Storer et al. 2010; Tabashnik et al. 2008; van Rensburg 2007; Jurat-Fuentes, Gould and Adang 2000; Ramachandran et al. 1998; Moar et al. 1995). From the emergence of resistance many research groups began examining the Cry toxins in greater detail, investigating their mode of action and in particular the binding of the protein to the insect intestinal cell membrane (Hua et al. 2001; Jurat-Fuentes and Adang, 2001; Garczynski and Adang 2000; Luo, Banks and

Adang 1999). Some of the receptors for Cry toxins on the insect cell membranes were isolated through binding studies on the brush border membrane vesicles of insects (BBMV's) (Garczynski and Adang 2000; Hua et al. 2001). To examine toxicity and resistance, domain swapping, mutagenesis and binding studies were carried out on several Cry toxins. In particular, significant work was carried out on Cry1Ac from the strain Kursatki HD-73 as it was found to be one of the most widely used and lethal Cry toxins to insects (Jurat-Fuentes and Adang 2007; Krishnamoorthy et al. 2007; Jurat-Fuentes et al. 2003; McNall and Adang 2003; Banks et al. 2001; Lee et al. 2001; deMaagd et al. 1999; Jenkins et al. 1999; Singsit et al. 1997; Lu et al. 1996; Singsit et al. 1996; Stewart et al. 1996; Masson et al. 1995; Garczynski and Adang 1995; Sangadala et al. 1994).

With this research came a greater understanding of the mode of action of Cry toxins, the identification of receptors on the insect cell, enhanced toxicity against specific insects and the ability to recover toxicity in the case of appearance of resistance (Pardo-Lopez et al. 2009). One finding of particular interest, was the specific binding of Cry1Ac to an N-Acetylgalactosamine (GalNAc) sugar residue on the aminopeptidase N of the insect gut (Rodrigo-Simon, Caccia, and Ferre 2008; Knowles, Knight, and Ellar 1991). It is evident that some Cry toxins have taken advantage of the recognition of the O-glycosylation on the insect gut and have specifically targeted this area (Burton et al. 1999). In recent years the focus of Cry toxin research has been to genetically engineer the Cry toxins to enhance toxicity, discover new *cry* genes through qPCR, identify the receptors through proteomics and characterise these insect receptors (Hua et al. 2009; Perera et al. 2009; Abdullah et al. 2009; Bayyareddy et al. 2009; Jurat-Fuentes and Adang 2007).

1.4 Toxin nomenclature

The *Bacillus thuringiensis* delta-endotoxin nomenclature committee was set up in 1993 in order to update the nomenclature originally devised in 1989 by Hofte and Whiteley. This system provided a useful framework for classifying the ever-expanding set of known genes. However, inconsistencies existed in the scheme due to attempts to accommodate genes that were highly homologous to known genes but did not encode a toxin with a similar insecticidal spectrum. Crickmore et al (1998), proposed a revised nomenclature for the cry and cyt genes. To organise the vast amount of data produced by genomic sequencing efforts, a new naming system emerged, exemplified by the cytochrome P-450 superfamily nomenclature system (Crickmore et al. 1998). The system that they put forward conformed closely to this model both in conceptual basis and in nomenclature format. The underlying basis of this type of system is to assign names to members of gene superfamilies according to their degree of evolutionary divergence as estimated by phylogenetic tree algorithms (Crickmore et al. 1998). Roman numerals were exchanged for Arabic numerals in the primary rank (e.g., Cry1Aa, Cry1Ab) to accommodate the large number of expected new proteins. The definition of a Cry protein put forward by Crickmore et al. (1998), is a parasporal inclusion (crystal) protein from *B. thuringiensis* that exhibits some experimentally verifiable toxic effect to a target organism, or any protein that has obvious sequence similarity to a known Cry protein (Crickmore et al. 1998).

1.5 Cry toxin diversity

The huge variety of known Cry proteins is as a result of continuing efforts to isolate and characterise new strains of Bt. It is hoped that this would result in novel toxins that could be used for the control of pests. Thousands of strains have been screened and there are currently over 150 unique Cry toxins according to the Bt toxin nomenclature webpage. http://www.lifesci.sussex.ac.uk/home/Neil_Crickmore/Bt/

This huge diversity of cry toxins is believed to be due to a high degree of genetic plasticity while many *cry* genes are flanked by transposable elements that may facilitate gene amplification, and in turn developing new toxins. The *cry* genes are found on plasmids, thus horizontal transfer by conjugation may result in the creation of new strains with novel *cry* genes (Pigott and Ellar, 2007).

Due to the large number of known Cry toxins it has permitted comparative sequence analyses and has helped to reveal elements important for both basic toxin function and insect specificity. Hofte and Whiteley carried out the first detailed analysis of Cry protein sequences in 1989 when 13 Cry proteins were known and assigned to one of four groups based on their insect specificity. The sequence alignment of these proteins showed a high degree of diversity among the Cry toxins. In most of the sequences, however, five blocks of conserved amino acids were discovered. Therefore the discovery of new Cry proteins encouraged further analysis by Bravo (1997) and de Maagd et al (2001). Most of the toxins examined had some or all of the five conserved blocks, which suggested that these regions may be important for toxin stability or function.

Natural resources Canada set up a *Bacillus thuringiensis* toxin specificity database as part of their Microbial control agents project. The ongoing discovery of new *Bt* toxin genes and rapid accumulation of information on their insecticidal activities prompted them to construct a relational database on *Bt.* toxin specificity to make the information accessible in a searchable format http://www.glfc.cfs.nrcan.gc.ca/bacillus (Frankenhuyzen et al. 2002).

1.6 Evolution of the Cry toxins

To date, over 700 *cry* gene sequences have been identified due to extensive screening and sequencing of Bt strains (Crickmore et al. 2011). These sequences have been classified according to their amino acid sequence identity in around 70 different *cry* gene groups and the toxins in these groups share less than 40% amino acid identity with toxins in other groups (Crickmore et al. 1998, 2011). The Cry toxins are flanked by transposable elements which are located on large plasmids in *Bt*. It is thought that this could be one reason for the large range of toxins and their similarities in structure. Phylogenetic analysis of Cry toxin sequences showed that the family of Cry proteins belong to four non-phylogenetically related protein families (i) the family of three domain Cry toxins (3d), (ii) the family of mosquitocidal Cry toxins (Mtx), (iii) the family of the binary-like (Bin) and (iv) the cytotoxic (Cyt) family of toxins (Bravo, Gill, and Soberón 2005). Among these toxins, the lineage of 3d-Cry toxins represents the largest group with more than 53 different subgroups of Cry toxins (Crickmore et al. 2011). One particular feature of the 3d-Cry group is the presence of protoxins with two different molecular sizes, namely, 65 and 130 kDa. Cry1Aa, Cry2Aa, Cry3Aa, Cry3Ba, Cry4Aa, Cry4Ba and Cry8Ea all have their crystal structure resolved and this shows a large amount of structural similarity in all toxins (Pigott and Ellar 2007). This group of toxins are known by their production of bacterial protoxins during sporulation which can be either large 130kDa or short protoxins 65-70kDa. The large protoxins are processed by insect gut proteases where much of the C terminal and a short fragment of the N terminal end are cleaved, thus yielding a protease resistant core that is biologically active (de Maagd et al. 2001).

Evolutionary analysis of domain I and II reveal that they co-evolved as different toxins (Crickmore 2000; Bravo 1997). Analysis of domain III however showed domain swapping among several toxins. Therefore domain III swapping among Cry toxins could be an active evolutionary development for determining insect specificity (de Maagd et al. 2001; Bravo 1997). Wu and co-workers (2007) investigated adaptive evolution in Cry toxins and identified several residues that are under positive selection. Positive selection favours the retention of mutations that are beneficial to a protein. Most of the 24 identified residues were located in domains II loop regions or domain III. This may provide some indication that these areas are likely to be involved in receptor binding. It was proposed that high divergence in

these regions could promote rapid evolution to their target receptors (Wu et al. 2007). The diversity of the Cry toxins is speculated to be a result of the independent evolution of the three domains with domains II and III under positive selection and domain III involved in domain swapping. This generated Cry toxins which have similar structure and mode of action but different specificities.

1.7 Structure of the 3d-Cry toxins

1.7.1 Cry Toxin Domain I

Domain I is located on the N terminal of the Cry toxin and is a collection of seven alpha helices where the central helix is hydrophobic and circled by amphipathic helices. (See figure 1.2) This is the helical domain responsible for membrane insertion and pore formation. Domain I shares similarities with other pore forming toxins (PFTs) such as colicin (see figure 1.1) which further supports that it plays a role in pore formation (Bravo, Gill, and Soberón 2008).

Figure 1.1: Crystal structure of colicin N (PDB code, 1A87). The helical bundle with structural similarity to Cry toxin domain I is shown in red (Pigott and Ellar 2007).

1.7.2 Cry Toxin Domain II

Domain II consists of three antiparallel beta sheets, where the loop regions are exposed (See figure 1.2). These areas are thought to be involved in receptor binding. This is supported by the fact that structural similarities have been shown between domain II and several carbohydrate binding proteins like vitelline, lectin jacalin and the lectin Mpa (De Maagd et al. 2003).

1.7.3 Cry Toxin Domain III

Domain III is a beta sandwich and has been reported to have a role in both receptor binding and structure (Pigott and Ellar 2007). Domain III shares structural similarities with carbohydrate binding proteins which include the cellulose binding domain of 1, 4 beta glucanase C, galactose oxidase, sialidase, beta- glucoronaidase, the Carbohydrate Binding Domain (CBD) of of xylanase U and beta-galactosidase (Pigott and Ellar 2007). These enzymes generally consist of a catalytic domain linked to one or more Carbohydrate binding Molecules (CBMs). The function of the CBM is to target the catalytic domain to its polysaccharide substrate (Pigott and Ellar 2007). This enhances the enzyme's catalytic efficiency by increasing its effective concentration at the substrate surface. This could imply that carbohydrate moieties have an important role in receptor binding and toxicity for these toxins.

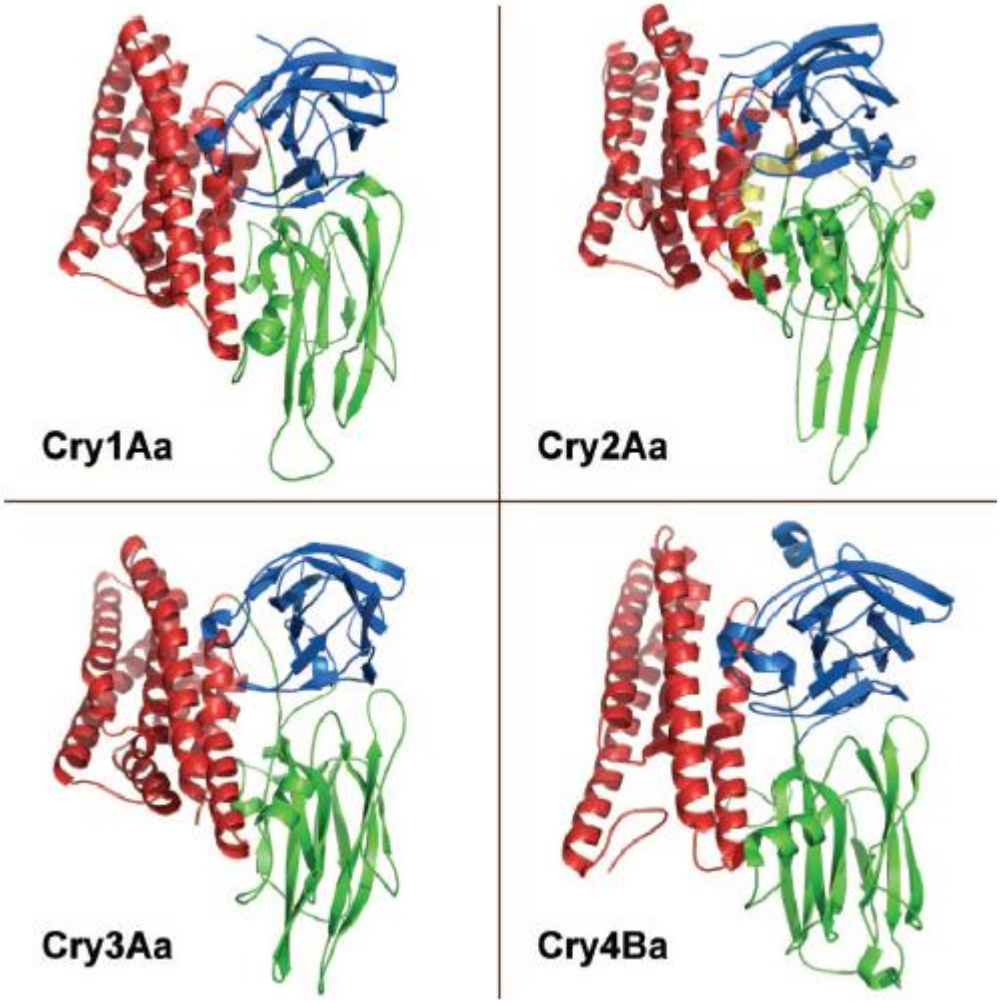


Figure 1.2: Crystal structure of Cry1Aa, Cry2Aa, Cry3Aa, and Cry4Ba. Domain I, domain II, and domain III are shown in red, green, and blue, respectively. The N-terminal protoxin domain of Cry2Aa is shown in yellow (Pigott and Ellar 2007).

1.8 Mode of Action of the 3d-Cry toxins

3d-Cry toxins are recognised as pore forming toxins that lyse insect midgut cells by osmotic shock. For this to happen the parasporal crystals have to be ingested by the larvae, solubilised by the pH of the insect gut and activated by proteases to yield an activated toxin. Activated Cry1A subsequently binds to the insect gut through a mechanism with Glycosylphosphatidylinositol (GPI) anchored proteins such as alkaline phosphatase (ALP) or aminopeptidase-N (APN) and cadherin like protein which result in the formation of a pre-pore structure (Bravo et al. 2011; Pacheco et al. 2009; Pigott & Ellar 2007; Bravo et al. 2004). 3d-Cry toxins may also bind to glycans, glycolipids and certain intracellular proteins (Bayyareddy et al. 2009; Krishnamoorthy et al. 2007; Griffits et al. 2005; McNall and Adang 2003). Figure 1.3 shows the morphological changes in midgut cells of Manduca sexta larvae after ingestion of the toxin. Bravo, Jansens and Peferoen (1992) reported that the toxins were immunolocalized in the apical microvilli of the midgut cells. Figure 1.3 shows the mechanism of action of the 3d-Cry toxins in *Lepidoptera* at the molecular level. Another model for the mode of action of Cry toxins, proposed by Zhang et al (2006) illustrates that binding to cadherin is sufficient to trigger an intracellular signal transduction pathway leading to cell death. This intracellular pathway results in activation of a G protein with subsequent activation of adenylyl cyclase, raising cAMP levels and activating a protein kinase A that in turn leads to cell death without involvement of oligomer formation and toxin pore formation. However nontoxic Cry toxin mutants were constructed that affected toxin oligomerisation and pore formation but this did not affect their binding to cadherin. This illustrates that binding to cadherin alone is not sufficient for toxicity (Rodrıguez-Almazan et al. 2009; Girard et al. 2008; Jimenez-Juarez et al. 2007; Vachon et al. 2002).

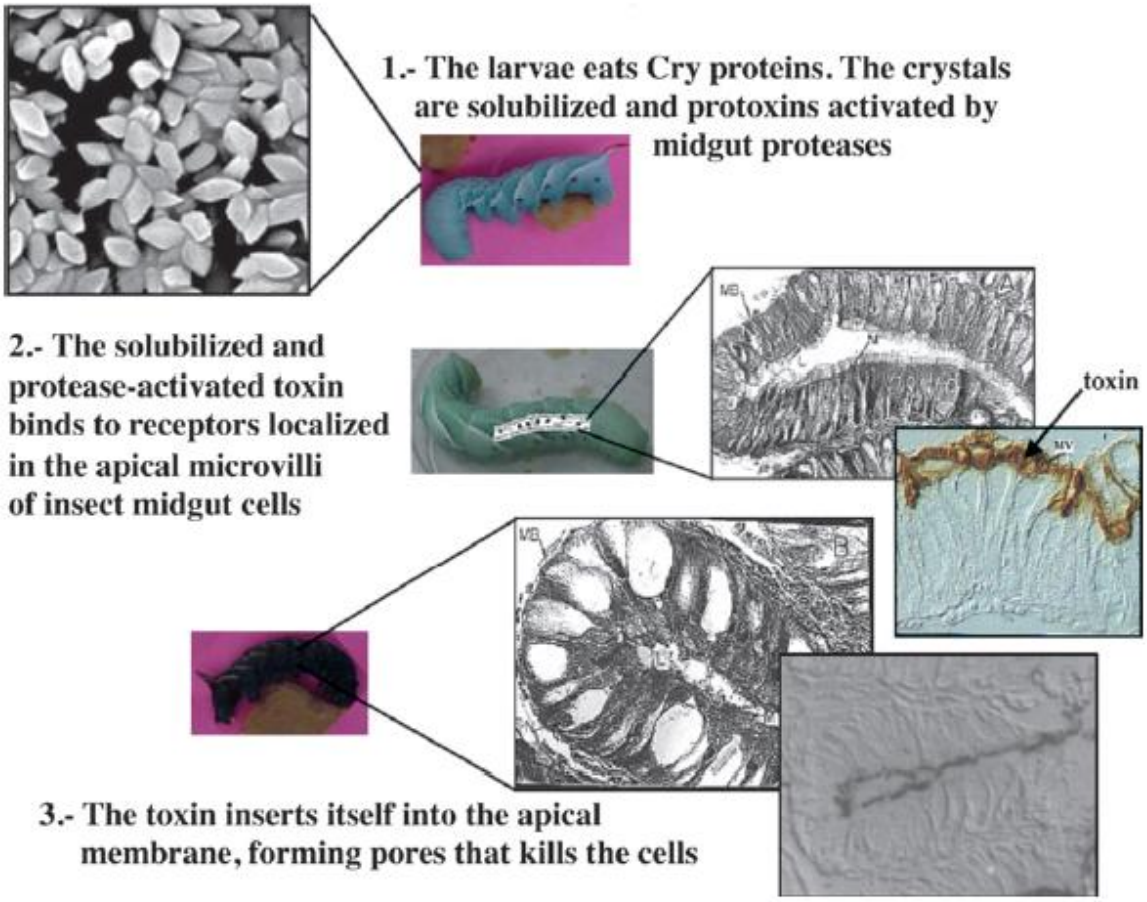


Figure 1.3: The mechanism of action of Cry (3D) toxins in Lepidoptera at the cellular level. Showing the immunolocalization of Cry toxin during toxication (Bravo, Jansens, and Peferoen 1992).

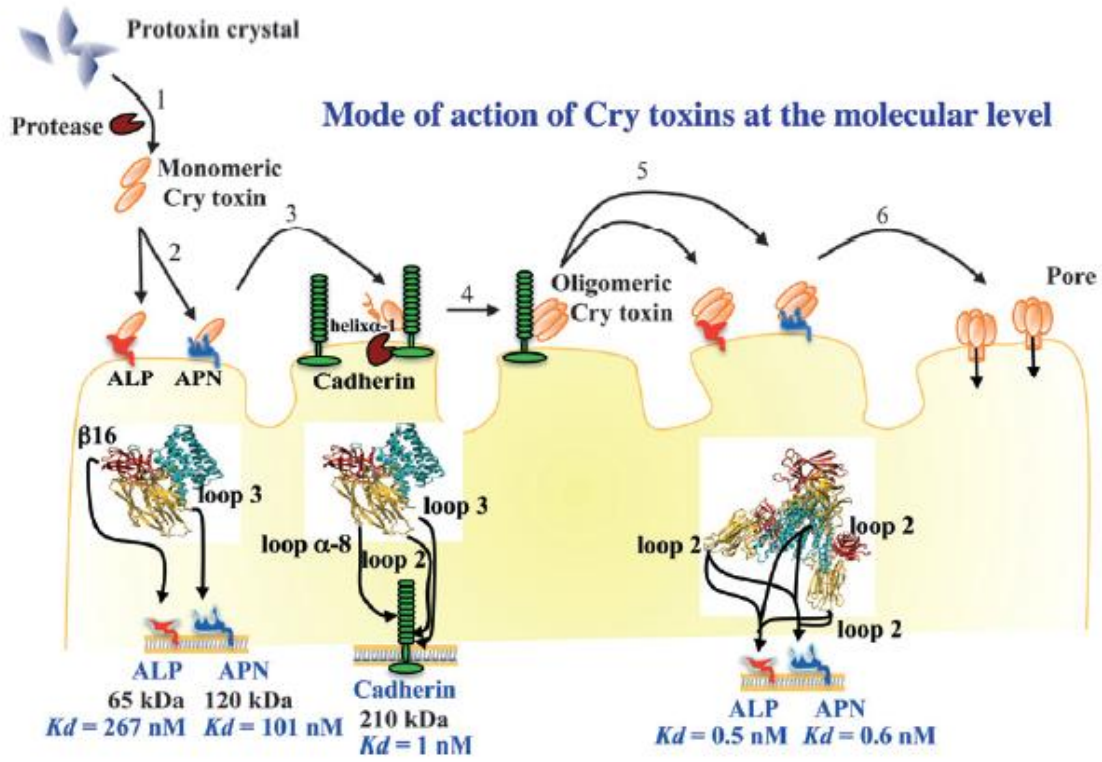


Figure 1.4: The mechanism of action of 3d-Cry toxins in Lepidoptera at the molecular level:

1. The larvae ingest the Cry protoxin, which is solubilised in the midgut due to high pH and activated by gut proteases. 2. The monomeric Cry toxin binds ALP and APN receptors; in a low-affinity interaction, the toxin is then located close to the membrane. 3. The Cry toxin binds the CAD receptor in a high-affinity interaction and this interaction induces proteolytic cleavage of the N-terminal end of the toxin. 4. The cleaved 3d-Cry toxin oligomerises in a toxin prepore oligomer. 5. The oligomeric Cry structure binds to ALP and APN receptors with high affinity. 6. The toxin inserts into the membrane causing pore formation (Pardo-Lopez et al. 2012).

1.9 Investigating the Receptors of the 3d-Cry toxins

Much of the research that has been done to date on receptor binding of Cry toxins is largely focused on how receptor binding affects toxicity and how the toxicity may be increased. Some of the receptors that have been identified are; cadherin receptor (CADR), aminopeptidase-N (APN), alkaline phosphatise (ALP), GalNAc on the APN or ALP and, 270 kDa glyco-conjugate receptor (GCR) (See figure 1.5). Various methods used to analyse receptor binding and pore formation of the toxin have been examined including light scattering of liposomes or brush border membrane vesicles (BBMV) (Carroll and Ellar 1993; Haider and Ellar 1989), leakage of 86Rb+–K+ and leakage of calcein from BBMVs (Rausell et al. 2004; English, Readdy and Bastian 1991) , analysis of single-channel currents in black lipid bilayers (Lorence et al. 1995; Schwartz et al. 1993), analysis of ion amino acid co-transport (Hendrickx et al. 1990), analysis of changes in membrane potential (Munoz-Garay et al. 2006; Lorence et al. 1995) and studies of short-circuit currents in midgut tissues (Liebig, Stetson and Dean 1995).

As outlined in section 1.7, Domains II and III have been identified as the receptor binding domains for the 3d-Cry toxins. A considerable amount of work has been done on domain swapping and mutagenesis to analyse the nature of the binding to receptors on the insect midgut. Mutagenesis has been carried out on domain II for many Cry toxins (Griffitts et al. 2005; Beron, Curatti, and Salerno 2004; De Maagd et al. 1999; Schnepf et al. 1998). Data from these experiments reveal that the mutations may have a negative or positive effect on binding and toxicity thus demonstrating that the binding of receptors and domain II is very complex. It has been observed that all loops of domain II may be involved in receptor binding but not for every receptor and not all at the same time (Xiang et al. 2009). The first example of domain II loop mutants with increased insecticidal activity was Cry1Ab toxin where mutations in loop 2 resulted in higher insecticidal activity against Gypsy moth (*Limantria dispar*) (Rajamohan et al. 1996). The increased insecticidal activity correlated with increased binding affinities to BBMV isolated from Gypsy moth (Rajamohan et al. 1996).

In the case of domain III, there are just a few examples of mutations in two different exposed loop regions with some mutants showing a moderate to non-significant increase in toxicity against different insect species (Lv et al. 2011; Shan et al. 2011; Xiang et al. 2009; Shu et al. 2007). There have only been a small number of studies that have mapped the domain III binding regions with ALP and APN (Arenas et al. 2010; Gómez et al. 2006; Atsumi et al. 2005). However, there has been substantial work done on domain III swapping which yielded in some cases hybrid toxins with increased toxicity (Bosch et al. 1994). One example the hybrid toxin that contains domain I and II from Cry1Ab and domain III from Cry1C; It was found to have more than a sixfold higher toxicity against beet armyworm (*Spodoptera exigua*) compared with Cry1C (de Maagd et al. 2000).

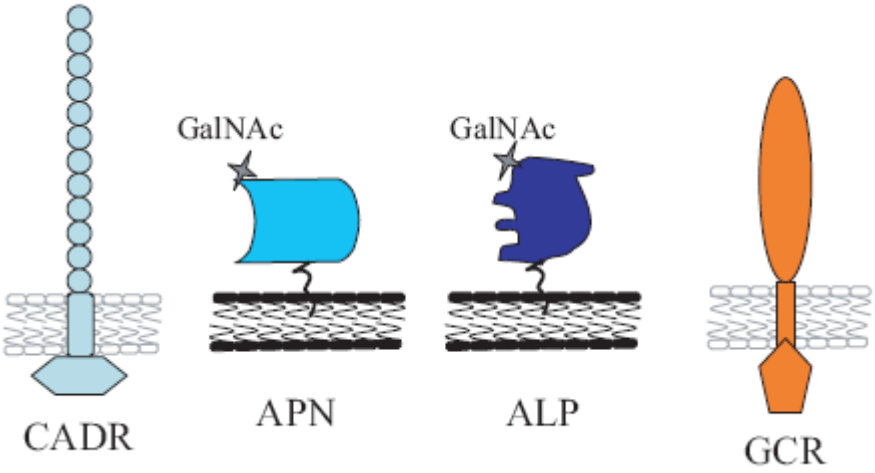


Figure 1.5: Examples of receptor molecules of Cry1A proteins. CADR, cadherin receptor; APN, aminopeptidase-N, ALP, alkaline phosphatase, GalNAc on the APN or ALP and GCR, 270 kDa glyco-conjugate receptor (Bravo, Gill, and Soberón 2007).

1.10 Cry1Ac 3d-Cry toxin

Cry1Ac is one of the 3d- toxins produced by *Bacillus thuringiensis* subsp. *Kurstaki* HD-73 (see Figure 1.7). Adang et al. (1985) characterised the Cry toxins in *Kurstaki* HD-73 and examined their toxicity to *Manduca sexta*. In 1986 the *cry1ac* gene was expressed in tobacco plants. This ultimately led to the commercialisation of insecticidal products and transgenic crops (Adang et al. 1986). Cry1Ac was chosen as the toxin in many transgenic crops because it is extremely lethal to many insect species (Adang et al. 1986). It is now known that binding of Cry1Ac toxin to specific receptors in the midgut brush border membrane is essential for toxicity. Resistance of Cry toxins correlates to changes in the receptors on the brush border membrane of insects (Ferre and Van Rie 2002). In *Heliothis virescens* (tobacco budworm), alterations in the cadherin protein HevCaLP (Jurat-Fuentes et al. 2004; Gahan, Gould and Heckel 2001) and the membrane-bound alkaline phosphatise HvALP (Jurat-Fuentes and Adang 2004) have been reported to be involved in resistance to Cry1Ac toxicity.

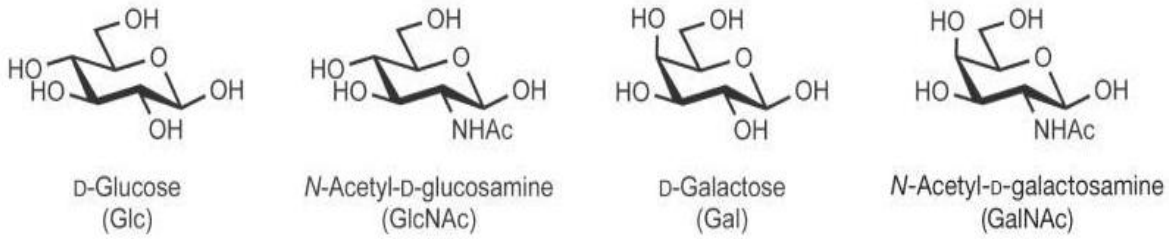


Figure 1.6: Structure of Monosaccharides. Monosaccharides Glucose and Galactose with their derivatives GlcNAc and GalNAc (Varki et al. 2009).

As a result of the research into the receptors on insect cells and mutagenesis of Cry1Ac it was found that this toxin has some unique binding patterns in comparison to other Cry toxins. In an in-direct toxin-binding assay using BBMV's, the modified sugar GalNAc completely abolished toxin binding while GlcNac had no effect (Knowles, Knight, and Ellar 1991) (see Figure 1.6). The presence of GalNAc on the

toxin-binding protein was confirmed by experiments in which both the toxin and the GalNAc-binding lectin soybean agglutinin bound to a 120 kDa protein present in M. sexta BBMV's (Knowles, Knight, and Ellar 1991). The results presented by Burton et al. (1999) point out residues N506, Q509 and Y513 to be involved in the binding of Cry1Ac both to the carbohydrate GalNAc and to aminopeptidase N. The binding of Cry1Ac throughout the larval midgut was examined and it was found that there were different binding patterns in regions of the larva i.e. there were regions where GalNac did not affect binding (Rodrigo et al. 2008). This provides some evidence that the Cry proteins do not recognise one specific receptor and thus would have the ability to affect different species of insects. Xiang et al. (2009), revealed that the residue N546 in β18– β19 loop of the Cry1Ac domain III played an essential role in its insecticidal activity and binding to insect BBMV. Proteomic approaches have also been employed to identify novel Cry1Ac toxin binding proteins (McNall and Adang 2003). Jurat-Fuentes and Adang (2007) used a proteomic approach to study Cry1Ac binding proteins and the alterations on these proteins in resistant *Heliothis virescens* larvae. They demonstrated, through 2D gel electrophoresis, lectin blots and 2D differential in gel electrophoresis (2D-DIGE), that *Heliothis virescens* membrane-bound alkaline phosphatase (HvALP) is a Cry1Ac binding protein and putative receptor. Previous proteomic analysis of Cry1Ac binding proteins in the Brush Border membrane (BBM) proteome allowed for identification of novel toxin binding proteins, including actin and the A subunit of the vacuolar ATPase (Krishnamoorthy et al. 2007).

Detection of Cry1Ac binding to both ATPase and actin suggests the possibility of potentially relevant interactions between part of the Cry1Ac toxin and intracellular proteins. Kitami et al. (2011) reported that Cry1Ac domain III bound to several proteins unrelated to insects in dot blot experiments. Thus even though the main receptor for Cry1Ac would seem to be GalNAc these studies have shown that it may also have specificity for other proteins or glycoproteins. Kitami et al. (2011), postulate that the binding of the unrelated proteins was far away from the binding site of GalNAc. These and other studies that were carried out on Cry1Ac were mainly focused on its impact on toxicity, insect resistance or effect on other organisms (Yang et al. 2008; Jurat-Fuentes and Adang, 2004). The involvement of glycoproteins as Cry toxin receptors has been suggested for several toxin membrane interactions (Burton et al. 1999).

Figure 1.7: Predicted structure of active Cry1Ac. This structure is based on previously solved Cry1A structures and submitted to I-TASSER. Domain I is in blue (beta barrels), domain II is shown in green while domain III is shown in red.

1.11 Carbohydrate Binding Proteins

The understanding of the significance of glycosylation is expanding rapidly. However, progress in understanding cell glycosylation, cell-glycan interactions, host-microbe interactions and aberrant glycosylation, is relatively restricted by the analytical techniques accessible for glycan profiling. Lectins are a general term for proteins that may recognise and bind specific glycan structures (Varki et al. 2009). They have the ability to bind and separate closely related glycoforms of glycoproteins and are increasingly being used in bio-analytical techniques, for glycan/glycoprotein enrichment and separation (reviewed in more detail in section 1.17). The disadvantage of these molecules is the cost and availability as most commercial lectins have been isolated whole from plant seeds, fruit, stems and bark. Lectins, from these sources, are typically multi-subunit proteins in which the subunits themselves may vary or be glycosylated. Thus, it is difficult to produce large quantities of plant lectins recombinantly which can give rise to batch to batch variation depending on the method of purification. In the past number of years prokaryotes have emerged as a good source of novel carbohydrate-binding molecules, for example, *Escherichia coli,* P fimbriae targets Galα1–4Galβ- on the urinary tract and *Helicobacter pylori,* BabA targets Le^b on the stomach (Imberty et al. 2004, Loris et al. 2003). Unlike plant lectins, bacterial lectins are more suited to overproduction in bacterial expression systems. The ability to up-scale production of these molecules and simplified downstream purification methodologies could enable the use of these lectins for wide reaching applications at low costs. DNA methods can be used to further enhance carbohydrate-binding proteins and facilitate their implementation in glycol-analytical platforms. An affinity tag, such as 6x histdine may be engineered onto the protein to enable downstream purification of the recombinant binding protein, to simplify their incorporation or immobilisation onto chromatographic or array based platforms. Mutagenesis may also be used on recombinant carbohydrate-binding proteins to enhance or alter their binding affinities and/or specificities. Potential useful directions for the development and expansion of novel carbohydrate-binding protein libraries include those which target common glycan terminal structures such as GalNAc and GluNAc. The binding domains of the Cry toxins have the potential to be novel carbohydrate binding proteins. These toxins have various receptors in insects, some of which are glycosylated and domain

III in the toxins has shown strong structural similarities to other lectins. Cry1Ac has shown glycan binding properties, in particular to GalNAc. Investigation into the binding of Cry1Ac to GalNac containing glycoproteins for use not as a toxin but as a lectin has not yet been explored. It is proposed that Cry1Ac could be used as a novel carbohydrate-binding bioligand which could have a number of applications, for use in glycan array techniques or in detection and separation of glycans or glycoproteins.

1.12 Glycobiology and Glycosylation

1.12.1 Glycobiology

The term glycobiology first came about in the late 1980s to recognize the disciplines of carbohydrate chemistry and biochemistry coming together with a modern understanding of the cell and molecular biology of glycans and, in particular, their conjugates with proteins and lipids. Glycobiology is the study of the structure, biosynthesis, biology, and evolution of sugar chains or glycans that are widely distributed in nature (Varki et al. 2009). Glycobiology has been one of the fastest growing areas in natural sciences in the last number of years. This is due to the increase in knowledge about the attachment of carbohydrates onto many proteins especially those involved in secretion or insertion into cell membranes.

1.12.2 Glycosylation

Glycosylation is the process of enzymatically attaching carbohydrate molecules to protein or lipid substrates (Rakus et al. 2011). Glycosylation is a class of post-translational modification, which occurs on both proteins and lipids and it is more prevalent than phosphorylation, methylation, or acetylation. It is thought that the majority of proteins in eukaryotic cells are glycosylated. Abberant glycosylation has been observed in diseases such as congenital disorders of glycosylation, and also in a number of pathological states, including inflammation and cancer (Varki et al. 2009). It is important to emphasize that unlike protein sequences, which are primary gene products, glycan chain structures are not encoded directly in the genome and are secondary gene products. Small changes in the environment of the cell can cause dramatic changes in glycans. It is this variable and dynamic nature of glycosylation that makes it a powerful way to generate biological diversity and complexity. Of course, it also makes glycans more difficult to study than nucleic acids and proteins (Varki et al. 2009).

The complexity of carbohydrate structures in an organism is a result of several carbohydrate characteristics. The first is the ability of different types and numbers of sugar residues to form glycosidic bonds with one another. The second is the structural characteristics of these molecules, the type of linkage, the position and the absence or presence of branching (Ghazarian, Idoni, and Oppenheimer 2011). Glycans have the ability to encode a considerable amount of biological information even though they are not encoded by the genome (Varki et al. 2009). However the genome does encode for the enzymes that act on glycans such as glycosyltransferases and glycosidases. These enzymes come together in the endoplasmic reticulum (ER), the Golgi apparatus, and on the cell surface to determine the glycosylation patterns of glycoproteins and glycolipids (Sharon, 1980). There are two main types of glycosylation N linked and O linked glycosylation.

1.13 N – Linked and O – Linked Glycosylation in Eukaryotes

Glycans can form glycosidic bonds with proteins by two types of linkage, the first (N-linked) is the binding of N-acetylglucosamine to the amide side chain of asparagines. The second (O-linked) involves the addition of a C- 1 of N-acetylgalactosamine to the hydroxyl of serine or threonine (Ghazarian, Idoni, and Oppenheimer 2011).

1.13.1 N - Linked Glycan Synthesis

The N – linked glycosylation pathway is the best understood route to protein glycosylation. N-linked glycosylation can be used to demonstrate general principles of glycoconjugate structure and biosynthesis (Taylor and Drickamer 2006). Even though N-linked glycans are diverse they all have a common core structure. These attach to glycoproteins via the amide nitrogens of asparagine side chains. The glycan usually linked to an asparagine residue in animal cells is GlcNAc and this linkage is always in the β configuration (GlcNAcβ1-Asn) (Varki et al. 2009; Taylor and Drickamer 2006) (See figure 1.9). N-linked glycan synthesis can be divided into three stages. 1) The formation of a lipid-linked precursor oligosaccharide which takes place in the endoplasmic reticulum (ER). 2) *en bloc* transfer of the oligosaccharide to the polypeptide and 3) Processing of the oligosaccharide which takes place in the Golgi apparatus. These processing steps may include trimming some of the sugar residues and the addition of new sugars at the non-reducing terminal of the glycan (Taylor and Drickamer 2006).

The donor that initiates N-linked glycan synthesis is a $Glc_3Man_9GlcNac_2$ structure attached to the lipid dolichol through a pyrophosphate linkage. The assembly of a glycan on the dolichol head group takes place in two phases. The first phase occurs on the cytoplasmic side of the ER and the second phase takes place in the lumen (Varki et al. 2009; Taylor and Drickamer 2006) (See figure 1.8). The enzymes that catalyse the attachment of the two GlcNAc residues and the first five mannose

residues use the nucleotide sugar donors UDP-GlcNAc and GDP-Man. The lipid – linked glycan is then translocated across the membrane and becomes inaccessible to cytoplasmic enzymes. Further sugars are added by dolichol – linked sugars once the glycan chain is exposed on the luminal side of the ER membrane. Transfer of the completed glycan to emerging polypeptides also occurs on the luminal side of the ER. The dolichol-linked precursor oligosaccharide can be now transferred to asparagine residues of polypeptides. Glycosylated asparagine residues are almost invariably found in the sequences Asn-Xaa-Ser or Asn-Xaa-Thr, where Xaa can be any amino acid except proline. N – linked glycans are found at the surfaces of proteins while N-glycosylation is only initiated in the lumen of the ER, therefore target asparagine residues are found in secretory proteins and in some parts of transmembrane proteins. Oligosaccharyltransferase can then catalyse the transfer of the completed dolicol-bound precursor glycan to a polypeptide.

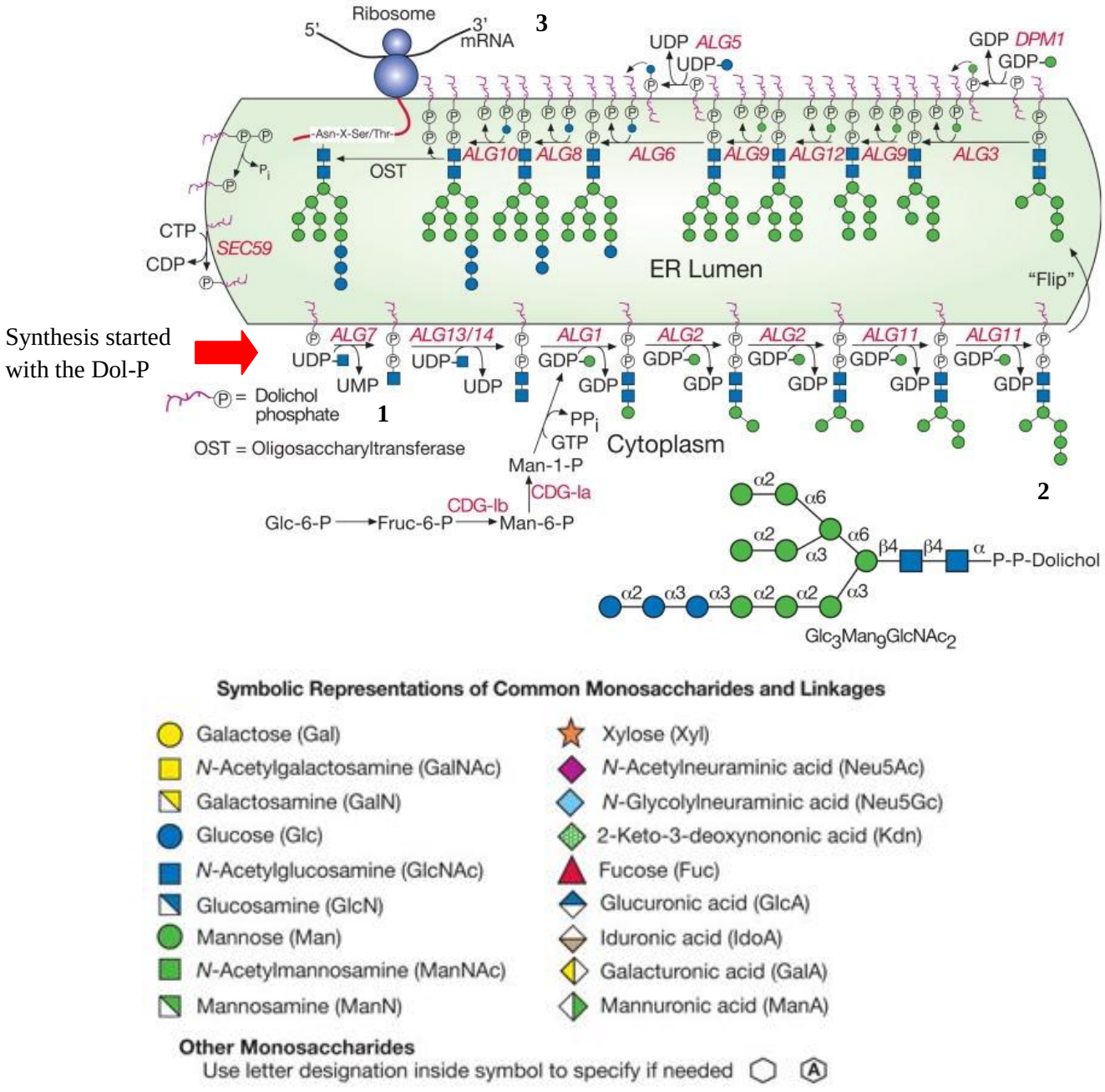


Figure 1.8: Overview of the synthesis of the common core in N-linked glycan in eukaryotes. 1. Dolichol phosphate (Dol-P) on the cytoplasmic face of the ER membrane receives GlcNAc-1-P from UDP-GlcNAc in the cytoplasm to generate Dol-P-P-GlcNAc. **2.** Dol-P-P-GlcNAc is extended to Dol-P-P-$GlcNAc_2Man_5$ and then is flipped across the ER membrane to the lumenal side. **3.** On the lumenal face of the ER membrane, four mannose residues are added from Dol-P-Man and three glucose residues from Dol-P-Glc (Varki et al. 2009).

Following the transfer to the polypeptide, the N-linked glycan is processed, first by the removal of some of the sugar residues by glycosidases. This happens while the polypeptide is in the ER. Glucosidase I removes terminal α1-2 linked glucose and glucosidase II removes the inner α1-3 linked glucose. The removal of glucose signals the glycoprotein to move from the ER to the Golgi apparatus. The glycoprotein is now subject to the action of a series of mannosidases that remove some or all of the four mannose residues in α1-2 linkages. Some of these proteins remain in this state and are high – mannose oligosaccharides, others become complex oligosaccharides and also hybrid oligosaccharides (Varki et al. 2009).

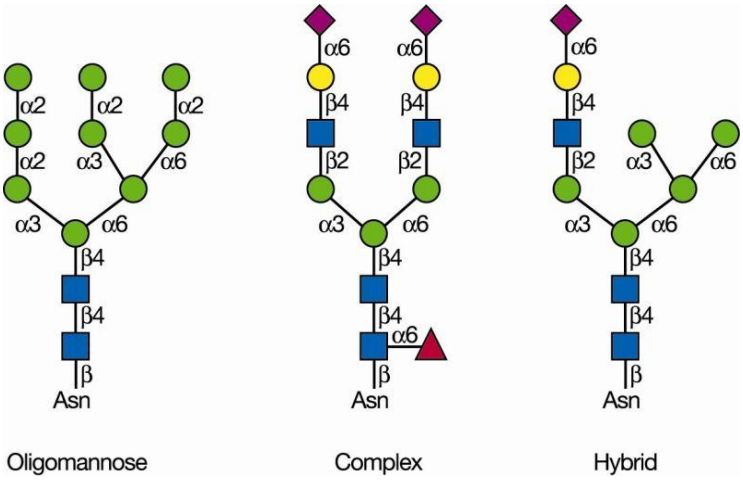
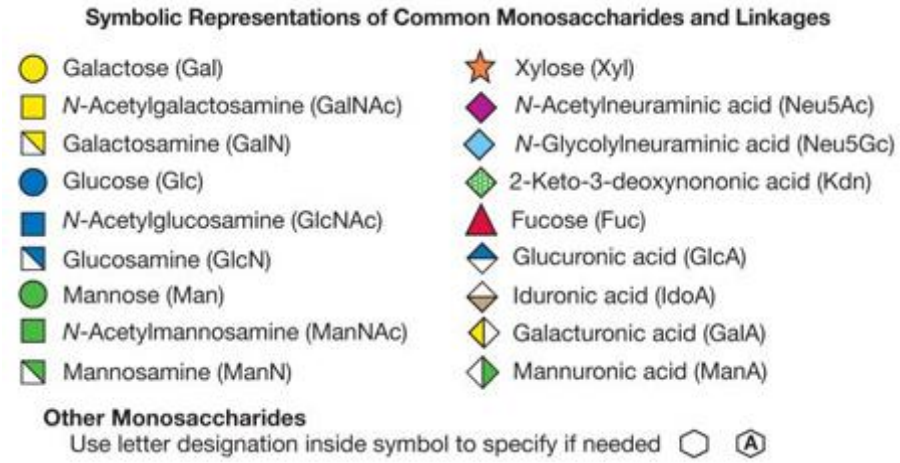


Figure 1.9: Types of N-glycans. Three general types of N-glycans can be attached to a mature glycoprotein: oligomannose, complex, and hybrid. Each N-glycan contains the common core $Man_3GlcNAc_2Asn$ (Varki et al. 2009).



Complex glycans are therefore built on a core that consists of just three mannose residues and two GlcNAc residues which are attached to the glycoprotein. GlcNAc-transferase I initiates re-elongation by the addition of a GlcNAc residue to the 1-3 arm of the core. Following this mannosidases in the golgi remove two additional mannose residues from the 1-6 arm of the core. GlcNAc additions are then added by GlcNAc-transferase and give the glycan its formation by defining a branch structure that is built onto the core. The branched structures may be extended by the addition of a single galactose and sialic acid residue. Galactosyltransferase and sialytransferases are responsible for these steps and are located further along in the secretory pathway. The galactose residue is usually β1-4 linked while the sialic acid is either α2-3 or α2-6 linked. If mutations are made in the dolichol pathway in knockout mice to prevent synthesis of the core oligosaccharide, it generates cells that lack N-linked glycosylation (Varki et al. 2009). Even though cells can be made which have no N-linked glycosylation these cells cannot form an organism. For example, embryos lacking N-linked glyans die shortly after implantation. This shows the importance of N-linked glycosylation and how understanding glycans and their regulation remains a key objective for the future.

1.13.2 O-linked glycosylation

O-glycosylation is a relatively common covalent modification of serine and threonine residues of mammalian glycoproteins. The properties of two groups of glycoproteins, mucins and proteoglycans, are dominated by the large number of O-linked sugars that are on their surface (Varki et al. 2009). The purpose of mucins is to retain water at the cell surface to which they are exposed and therefore mucins must be present at the surfaces of the digestive, genital tracts and the respiratory system. The addition of sialylated glycans to serine and theronine residues results in regions of strong negative charge which allows the mucins bind large amounts of water **(**Tabak 1995). Mucin glycosylation is based on a set of core structures in which GalNAc is linked to the side chain of serine or threonine and has a single β1,3-linked Gal residue to form a core 1 structure (Varki et al. 2009). To date there are about 20 genes so far that have been cloned which encode mucins (Varki et al. 2009). As well as water retention mucins give lubrication and also help protect from invasion by microorganisms (Bektas and Rubenstein 2011). The mucin polypeptides are very long and can contain as many as 1000 amino acids, they can be membrane bound or secreted (Varki et al. 2009). They contain tandem repeats of amino acid sequences rich in serine and threonine. The mucin genes, their transcripts, the resulting mucin proteins, and the attached GalNAc glycans all display large variability. The mucin genes are often polymorphic as different numbers of repeats are observed in different individuals (Varki et al. 2009).

Table 1.2: Structures of O-glycan cores found in mucins (Varki et al. 2009).

O-Glycan	Structure
Core	
Tn antigen	GalNAcαSer/Thr
Sialyl-Tn antigen	Siaα2-6GalNAcαSer/Thr
Core 1 or T antigen	Galβ1-3GalNAcαSer/Thr
Core 2	GlcNAcβ1-6(Galβ1-3)GalNAcαSer/Thr
Core 3	GlcNAcβ1-3GalNAcαSer/Thr
Core 4	GlcNAcβ1-6(GlcNAcβ1-3)GalNAcαSer/Thr
Core 5	GalNAcα1-3GalNAcαSer/Thr
Core 6	GlcNAcβ1-6GalNAcαSer/Thr
Core 7	GalNAcα1-6GalNAcαSer/Thr
Core 8	Galα1-3GalNAcαSer/Thr

1.13.3 O-glycan structures

The Tn antigen is one of the simplest mucin O-glycan structures. It consists of a single N-Acetylgalactosamine linked to a serine or threonine. This glycan is often antigenic and does not normally occur in O-glycosylated proteins (Brockhausen 2006; Hollingsworth et al. 2004). The most common O-glycan is Galβ1-3GalNAc-, this is the core 1 glycan (T antigen) as it forms the core of many other longer glycans (see figure 1.10) and like the Tn antigen is antigenic (Brockhausen 2006). The Tn and T anitgen may be modified by sialic acid to form sialylated-Tn or -T antigens. Core 2 is an *N*-acetylglucosamine attached to the core 1 (Varki et al. 2009). Linear core 3 and branched core 4 have O-GalNAc glycans and have only been found in mucin secreting tissues such as bronchi, colon, and salivary glands (Perez-Villar and Hill, 1999). The core structures 5-8 do not occur very often. Mucins with core 5 occur in human meconium and intestinal adenocarcinoma tissue, whereas core 6 structures have been found in human intestinal mucin and ovarian cyst mucin (Varki et al. 2009, Brockhausen 1999). All of these structures can be sialylated, however only cores 1–4 and core 6 have been shown to occur as extended, complex O-glycans that carry antigens such as the ABO and Lewis blood group determinants (Varki et al. 2009).

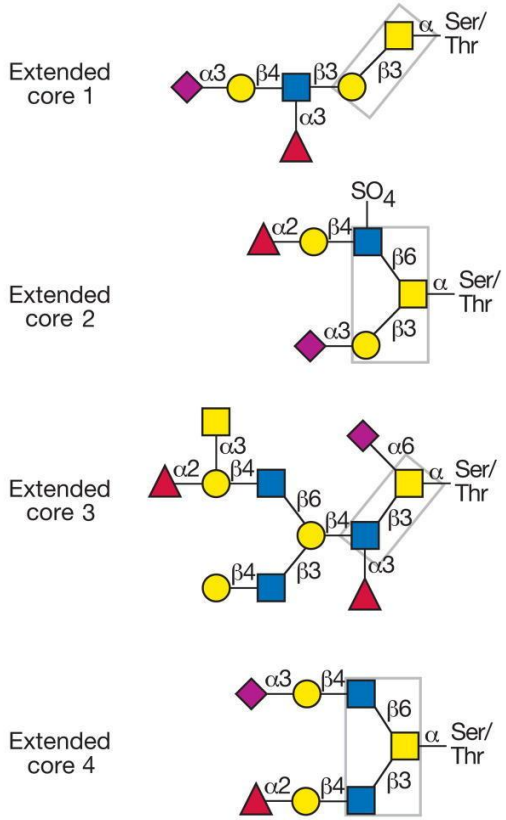
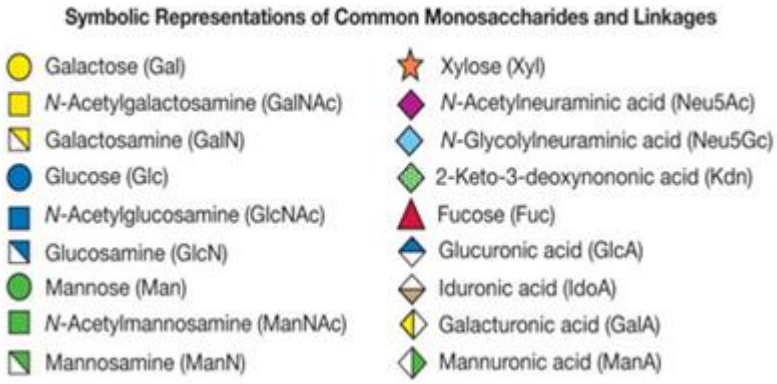


Figure 1.10: Examples of complex O-GalNAc glycans with extended core 1, 2, 3 and 4 (Varki et al. 2009).



1.13.4 O – Glycosylation synthesis

O-glycosylation uses glycosyltransferases that are analogous to those in the N-linked biosynthesis pathway. However, the organisation of these enzymes is very different. Firstly sugars are added one at a time in a stepwise series of reactions, which start with GalNAc attached to a serine or threonine residue (Varki et al. 2009). There is no *en bloc* transfer and there is no target sequence as in N-glycosylation. This is because there are numerous transferases that attach GalNAc to serine and threonine residues. There are at least 21 polypeptide-N-acetylgalactosaminetransferases (ppGalNAcT-1-21). N-acetyl-galactosaminyltransferase (ppGalNAcT) catalyzes the first step of mucin O-glycosylation and transfers GalNAc from UDP-GalNAc to serine or threonine residues **(**Tabak 1995). These transfers take place post-translationally in the Golgi. Subsequently, with the addition of the next sugar, different mucin O-glycan core structures are synthesized (Varki et al. 2009). During the extension of the core GalNAc, no lipid-linked intermediates are involved, and there are no glycosidases involved in the processing of O-GalNAc glycans within the Golgi (Perez-Villar and Hill 1999). The first sugar GalNAc added to the protein creates the Tn antigen (GalNAc-Ser/Thr), which is uncommon in normal mucins, but is often found in mucins derived from tumors (Brockhausen 1999). Another common cancer-associated structure found in mucins is the sialyl-Tn antigen, which contains a sialic acid residue linked to C-6 of *N*-acetylgalactosamine (Brockhausen 2006). O-Acetylation of the sialic acid residue stops anti-sialyl-Tn antibodies recognising the sialyl-Tn. To date, no other sugars are known to be added to the sialyl-Tn antigen. There are eight O-Glycan core structures, and most of these can be substituted with other sugars, such as sialic acid, galactose.

1.14 Prokaryotic glycosylation

It is now known that protein glycosylation is a phenomenon shared by all domains of life. A lot of progress has been made in understanding prokaryotic glycosylation since the general N-glycosylation system was discovered in *Campylobacter jejuni* (Wacker, Linton, and Hitchen 2002). Recently novel general O-glycosylation systems have been observed in pathogenic and symbiotic bacteria (Dell et al. 2010).

Bacterial glycans include peptidoglycan, periplasmic glycans, lipopolysaccharide, glycans of surface layer proteins (SLPs), and extracellular polysaccharides which make up capsules and biofilms (Varki et al. 2009). Prokaryotes produce a variety of glycoconjugates and polysaccharides which can vary in their structural diversity and complexity. These glycans include many sugars not found in eukaryotes, such as Kdo (3-deoxy-D-*manno*-octulosonic acid), heptoses, and modified hexoses (Schäffer, Graninger, and Messner 2001). The resulting glycans have important roles in the biology, and the pathogenicity, of bacterial cells.

Bacteria are divided into Gram-positive and Gram-negative organisms depending on the nature of their cell wall. In Gram-negative bacteria such as *Escherichia coli*, the cell wall consists of inner and outer membranes separated by the periplasm. Peptidoglycan is the major structural component of the periplasm, and it is made up of a polysaccharide covalently cross linked by short peptides (van Heijenoort 2001). In Gram negative bacteria the periplasmic space may also contain β-glucans, while the outer layer mostly contains lipopolysaccharide (Varki et al. 2009) (see figure 1.11 and 1.12). The cells of mucoid strains are surrounded by a polysaccharide capsule, which is thought to play a role in virulence. Gram-positive bacteria do not have the outer membrane, instead they have a much thicker peptidoglycan layer modified with the addition of teichoic acids (Schaffer et al. 2005).

Although much more remains to be discovered regarding the biosynthetic pathways of protein glycosylation in prokaryotes, genetic information so far shows that these pathways vary in complexity (Zarschler et al. 2010). In *Campylobacter jejuni*, a cluster of 12 genes - the *pgl* locus, is responsible for the synthesis of a heptasaccharide GalNAcα1–4GalNAcα1–4(Glcβ1–3)GalNAcα1–4GalNAcα1–4GalNAcα1–3Bac. The transfer of this chain *en bloc* to asparagine residues in a protein involves the same consensus sequence found in eukaryotic glycoproteins (Wacker, Linton, and Hitchen 2002). Oligosaccharyltransferase-mediated O-glycosylation is frequent in gram-negative bacteria, but has not yet been proven to occur in gram-positive bacteria. While cytoplasmic O-glycosylation seems to be more common in Gram-positive bacteria (Weerapana et al. 2006).

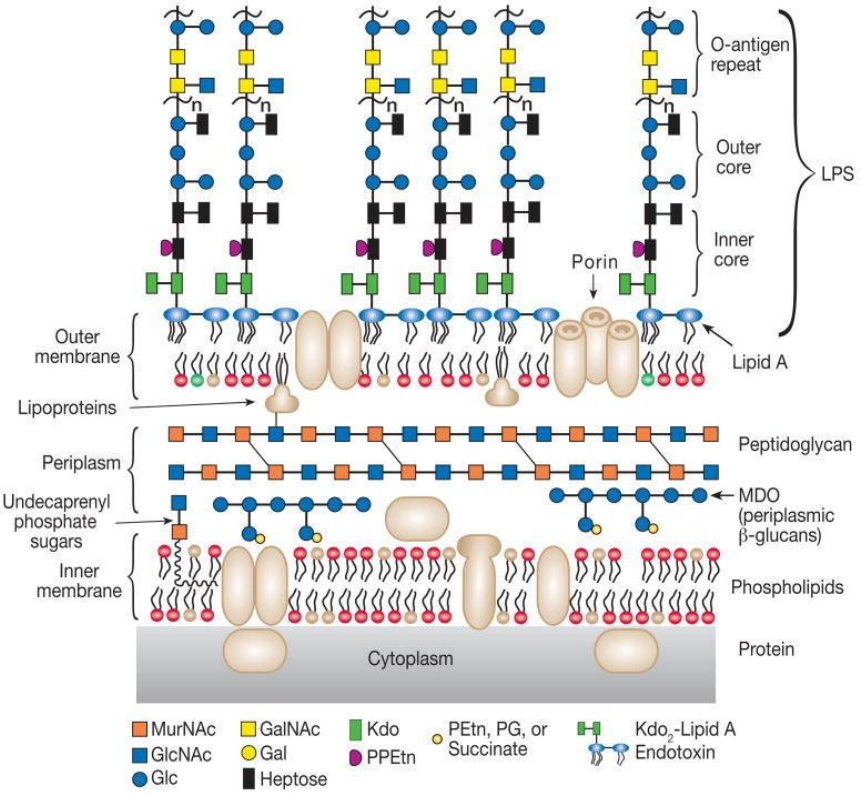


Figure 1.11: Cell wall from Gram-negative bacteria. Diagram of the cell wall of Gram-negative bacteria showing layers of polysaccharides and glycoconjugates (Varki et al. 2009).

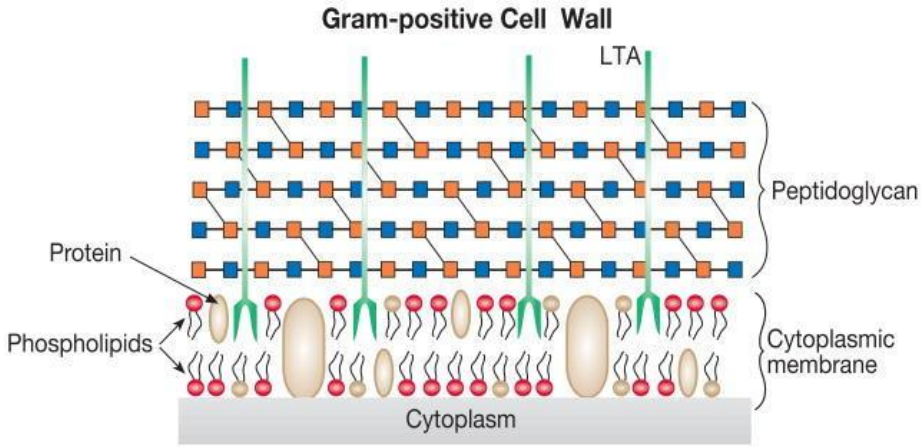


Figure 1.12: Cell wall from Gram-positive bacteria. Gram-positive bacteria lack the outer membrane and lipopolysaccharide (LPS) that is present in Gram-negative organisms. However the peptidoglycan layer is thicker and contains teichoic acids (Varki et al. 2009).

1.15 Significance of Glycosylation

Glycocosylation has proven to be of fundamental importance to many biological processes (Varki et al. 2009). Correct glycosylation plays a role in immune defence, cell growth, cell-cell adhesion, inflammation and fertilisation. Attachment of glycan moities to proteins/glycoproteins functions to stabilise proteins against denaturation and proteolysis, confer structural rigidity, enhance solubility, increase blood retention time and reduce immunogenicity (Varki et al. 2009; Taylor and Drickamer 2006). However, our understanding of the significance of glycans to biological processes arises from the study and genetic engineering of abnormal glycosylation and disease states. The general affect of glycosylation on physiochemical and biological functions are summarised in table 1.3.

Table 1.3: Affect of glycosylation on physiochemical and biological functions.

Type	Function
Physiochemical	• Modify solubility, electrical charge, mass, size and viscosity in solution • Control protein folding • Stabilise protein conformation • Confer thermal stability and protection against proteolysis
Biological	• Regulate intracellular traffic and localisation of glycoproteins • Determine lifetime of glycoproteins in circulation • Determine time of glycoproteins in circulation • Modify immunological properties • Modulate activity of enzymes and hormones • Act as cell surface receptors for lectins, antibodies, toxins • Participate in cell-cell interactions

1.16 Lectins

Lectins are proteins that have the ability to bind to glycans and glycoproteins. Glycans, as discussed previously, can mediate a wide variety of biological roles by virtue of their mass, shape, charge, or other physical properties. However, many of their more specific biological roles are mediated via recognition by Carbohydrate Binding Proteins (GBPs). There is considerable diversity in glycans expressed in organisms because of evolving proteins to recognise glycans that mediate specific physiological or pathological processes (Varki et al. 2009). Lectins were first discovered in plants more than 100 years ago and they are now known to be present in every organism **(**Kwan and Bun 2011) see Table 1.3. Lectins are found throughout the microbial world where they are described as toxins, hemagglutinins and adhesins (Varki et al. 2009; Taylor and Drickamer 2006). Plant Concanavalin A (Con A), was the first pure lectin to be isolated from the Jack Bean (*Canavalia ensiformis)* and to have its sugar specificity demonstrated by James and Stacey in 1936. They suggested that the hemaglutination induced by Con A might be a consequence of a reaction from the plant protein with carbohydrates on the surface of the red cells. Later it was discovered that Con A binds to α1-3 linked mannose residues (see Figure 1.13) Plant lectins are the most commercially available and applied lectins for glycoprotein and glycan research **(**Kwan and Bun 2011).

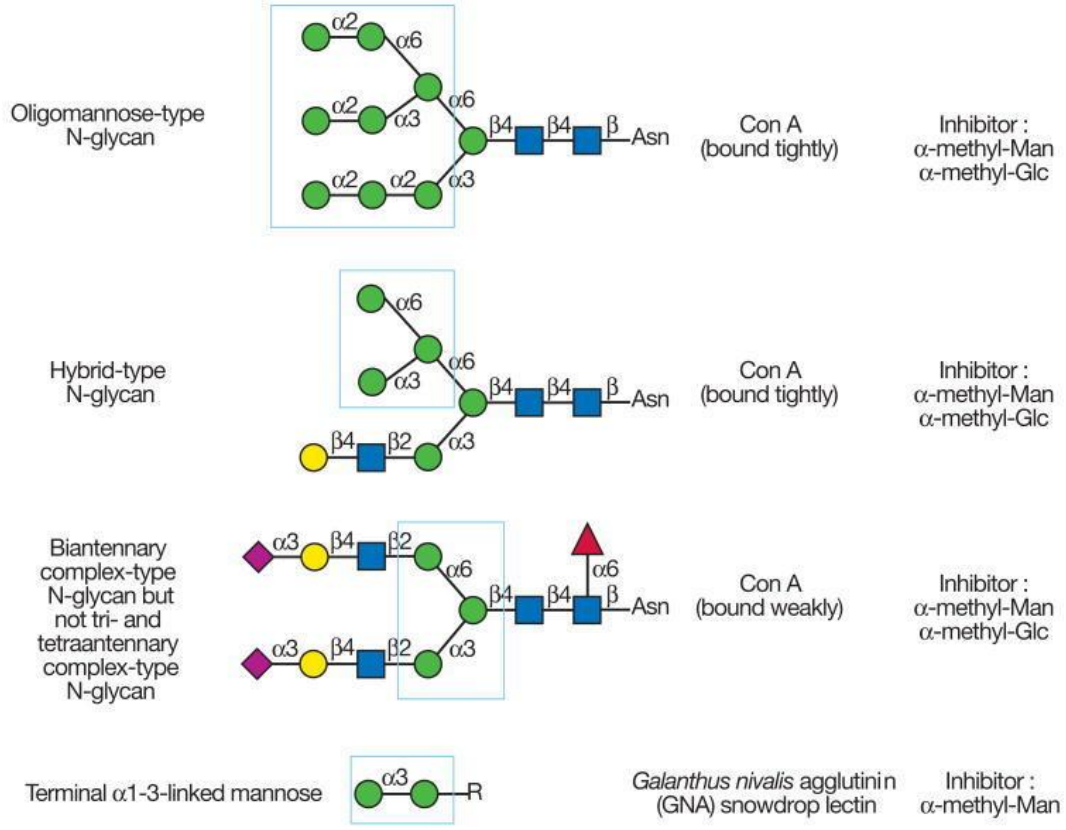


Figure 1.13: Examples of N-glycans recognized by ConA and GNA. Concanavalin A from *Canavalia ensiformis* and *Galanthus nivalis* agglutinin (GNA) (Varki et al. 2009).

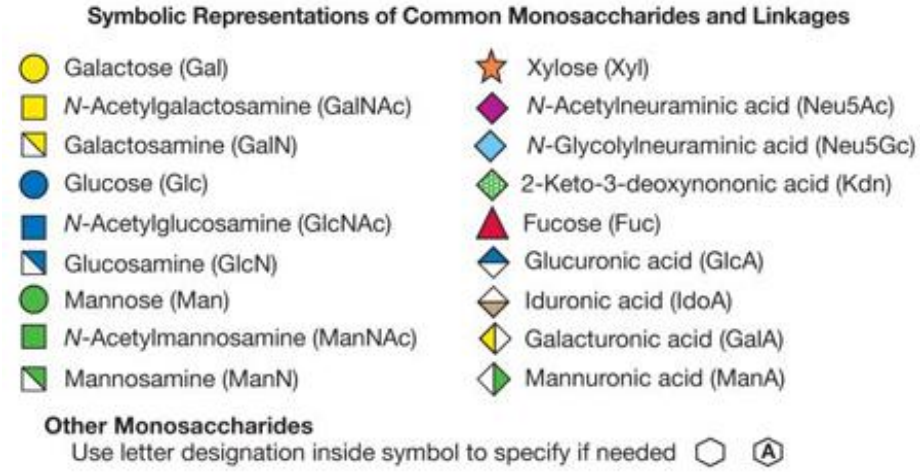


Table 1.4: Examples of types of lectins from different organisms (Varki et al. 2009).

Lectin Family	Lectin	Abbreviation	Sugar specificity
Plant Lectins			
Legume lectins	Concanavalin A	ConA	Man/α1-Me
	lectin 1	GSLI	α-gal
Griffonia simplicifolia			
Griffonia simplicifolia	lectin II	GSLII	β1-4 GlcNAc
Cereal lectins	Wheat germ agglutinin	WGA	NeuNAc, GlcNAc
Plant toxins	Ricin	RCA	β-gal/ lactosamine
Bulb lectins	Snowdrop lectin	GNL	Man
Animal lectins			
Galectins	Galectin 1	Gal-1	Lactose
Galectin 2	Gal-2		Fucose
Galectin 3	Gal-3		Fucose
P-type lectin	Mannose-6 phosphate receptor	M6P	Man-6 phosphate
Bacterial lectins			
P. aeruginosa	Lectin 1	PA-IL	β-gal
P. aeruginosa	Lectin 1	PA-IIL	Fucose/Man
Bacillus thuringiensis	Cry toxins	Cry1Ac	GalNAc
Viral lectins			
Influenza	Hemagglutinin	HA	NeuNAc

1.17 Methods used for the detection and separation of glycans and glycoproteins

1.17.1 Commercial lectins

To date, hundreds of plant and animal lectins have been identified and characterised (Varki et al. 2009). Lectins from many plants have been studied because of their abundance in seeds and tissues (Kwan and Bun 2011). These lectins have been analysed in detail and used for practical research on glycans (Varki et al. 2009). Thus, although monoclonal antibodies might be more specific for glycan epitopes, plant and animal lectins have useful specificities and affinities, are usually less expensive, are better characterised with respect to binding specificity, and are more stable. The availability of the plant lectins and their specificity for complex glycans has helped to expand the field of glycobiology (Varki et al. 2009). Figure 1.14 shows some important examples of the uses of lectins. Lectins with well defined specificity are excellent tools for several analytical and preparative purposes. Lectins may be used to identify and characterise the structure of glycans (Kwan and Bun 2011). They can be used when there is not enough biological material for techniques, such as mass spectrometry or nuclear magnetic resonance and they can also be used as a quantitative method as they can distinguish between subtle differences in oligosaccharide structure (Wu et al. 2009). Disadvantages of using plant lectins include; batch to batch variation depending on the purification method and high costs due to the low yield of lectin (Varki et al. 2009).

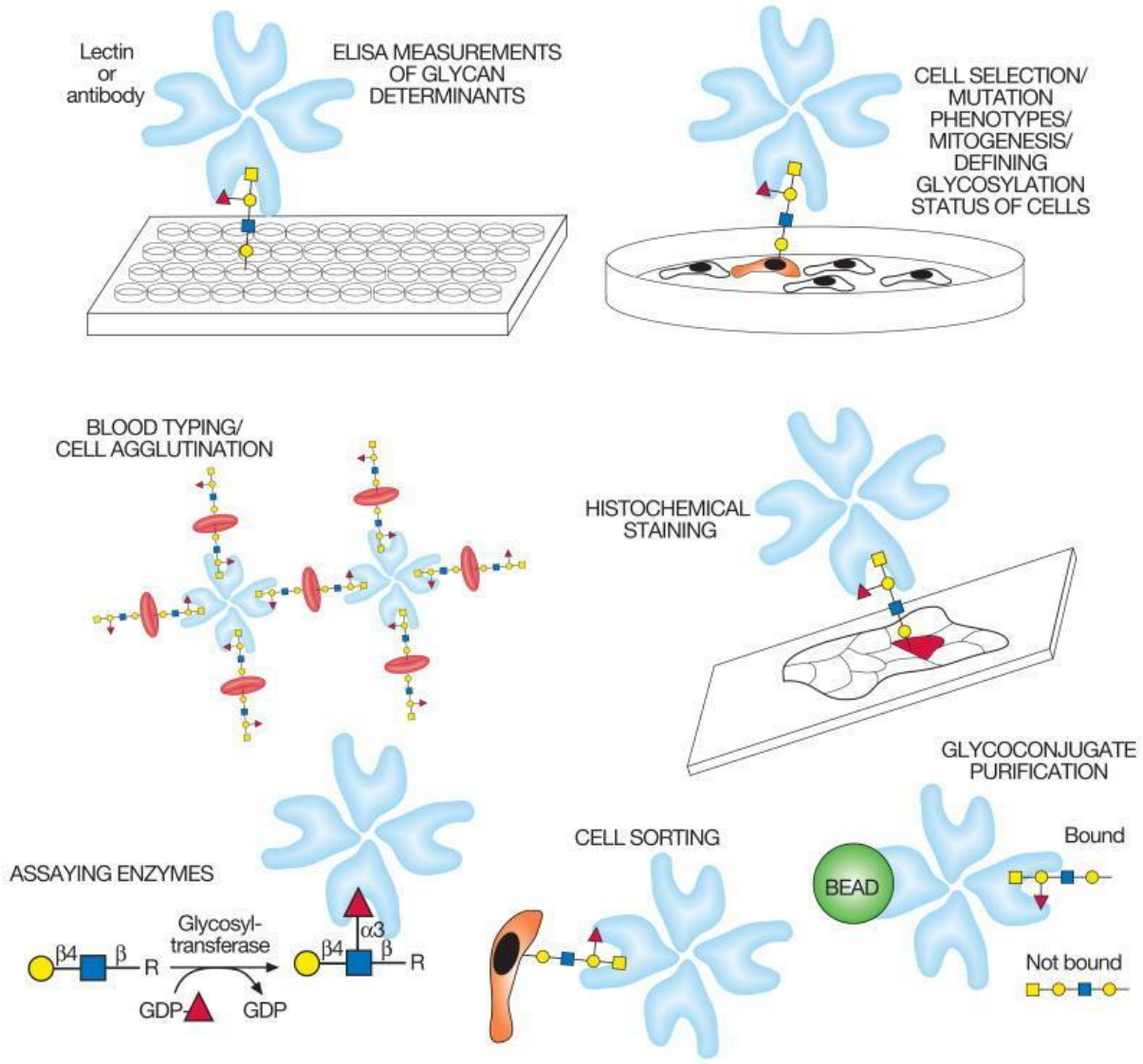


Figure 1.14: Examples of the various uses of plant and animal lectins and antibodies in glycobiology. They can be used to detect glycan structures in all of the formats shown (Varki et al. 2009).

1.17.2 Enzyme Linked Lectin Assay

The Enzyme Linked Lectin Assay (ELLA) is useful technique employed for the generation of glycan profiles. The ELLA which is similar to an ELISA has proven to be important for glycan and lectin profiling (Matsumoto, Shinzaki, and Narisada 2010; Lambre et al. 1991; Andrade et al. 1988; McCoy, Varani, and Goldstein 1983). In an ELLA a protein-glycan conjugate is immobilised onto the surface of a 96 well plate. After this the plate is blocked with appropriate blocking solution and the lectin is added. The binding of the carbohydrate and lectin can be detected using colorimetric techniques (Figure 1.15).

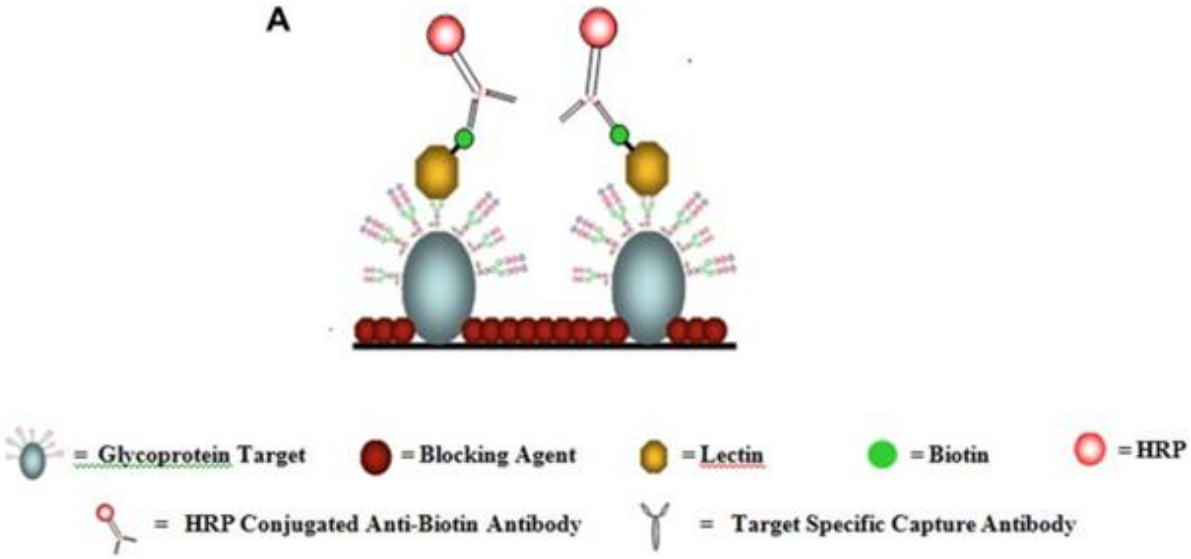


Figure 1.15: Enzyme linked lectin assay (ELLA). A protein carbohydrate conjugate is immobilised on the bottom of a 96 well plate. The plate is blocked with appropriate blocking solution. The glycan complex is probed with labelled lectin(s). The lectins are detected using specific, peroxidase labelled antibodies. The antibody-peroxidase conjugate can be detected using colormetric techniques (Thompson et al. 2011).

In the ELLA the proteins do not need to be denatured or go through harsh chemical steps. Exposure of underlying residues and specific linkages can be achieved by the glycosidase-mediated trimming of terminal glycans and the use of sugar and linkage specific lectins and neo-glycoproteins. As the lectin-glycan interactions are very specific, high concentrations of glycan are not needed. In addition, lectins can also be applied to in vitro and in vivo studies. Commercial lectins can be used to localise

specific glycans in cells by conjugating them to probes or fluorescent tags e.g. Green fluorescent protein (GFP) (Wanchoo, Lewis, and Keyhani 2009; Krist et al. 2004; Fitches et al. 2001), and radiolabelling (Smart et al. 2002). However the biggest disadvantage in the use of lectins is the cost and limited availability of fully characterised lectins **(**Kwan and Bun 2011).

1.17.3 Lectin affinity chromatography

Lectin affinity chromatography is very useful when a heterogeneous glycan mixture is passed over a lectin column. The glycans may be tightly bound to the column, requiring haptenic sugars for elution depending on the relative affinity of the lectin for sugars, therefore separating them from the unbound material (Geyer and Geyer 2006). This is a great technique that can be used in conjunction with MS and NMR where a sample needs to be separated or enriched. This technique can be expanded using multiple columns with several specificities represented.

1.17.4 Two dimensional (2D) gel electrophoresis

Two Dimensional (2D) gel electrophoresis is one of the most efficient protein separation and analytical techniques and is often the first step in glycan analysis. This technique can be used to separate proteins, reflecting both size and/or isoelectric points (Geyer and Geyer 2006). Once separated the glycosylation of the proteins blotted on a membrane can be characterised by probing with carbohydrate-specific stains, lectins or antibodies. A major drawback to this technique is the frequent under-representation of membrane proteins, due to the high degree of insolubility in the solubilisation detergents (Geyer and Geyer 2006). Another disadvantage is that this technique gives an indication of whether glycans are present or not, it does not give a quantitative analysis of the glycans or any information on their oligosaccharide structure.

1.17.5 High Pressure Liquid Chromatography (HPLC)

Due to their structural heterogeneity, glycoproteins represent often complex mixtures of closely related isomeric compounds. HPLC has proven to be good at separating and profiling glycans due to the wide range of adsorbents and solvent systems available, in addition to the speed and reproducibility of the separation (Geyer and Geyer 2006) (See Figure 1.16). Anion exchange chromatography may be used to separate glycans predominantly on the basis of the number of charged groups present. Similar systems can be used to separate glycans on the basis of hydrophilicity (Neutral phase NP-HPLC) and hydrophobicity (Reverse phase RP-HPLC) (Geyer and Geyer 2006). One of the advantages of this technique is the ability to predict possible glycan structures through the comparison of elution times with standard glucose oligomers using only pico moles of sample. A HPLC database for glycans/sugar maps has been constructed based on a large number of well-defined standard oligosaccharides, http://www.gak.co.jp/ECD/Hpg_eng/hpg_eng.htm (Geyer and Geyer 2006).

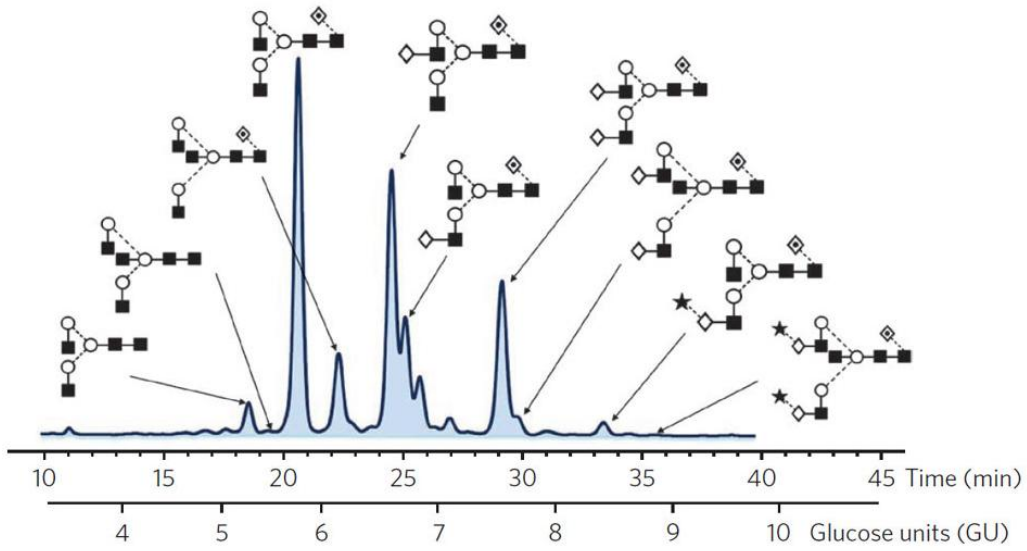


Figure 1.16: An example of a HPLC-HILIC profile of N-glycans released from heavy chain human serum IgG. The scale of glucose units (GU) is based on the elution of a 2-AB–labeled glucose ladder. Only the major species have been shown in this particular figure, but the whole nested family of 36 different structures, including arm-specific isomers, can be identified using this technique (Marino et al. 2010).

1.17.6 Mass Spectrometry (MS)

Mass spectrometry is a technique that detects molecules that can be converted to gas-phase ions. The resulting ions are accelerated out of the ionisation source into a mass analyser, where they are separated according to their mass to charge ratio and detected to produce a mass spectrum (Hitchen and Dell 2006) (See figure 1.17). Mass spectrometric identification of glycoproteins has proven to be an important tool due to its high degree of sensitivity. It can be used to resolve the components of a glycan mixture and derive information about the structures of individual glycans. (Taylor and Drickamer 2006).

Unlike the HPLC method, a substantially larger amount (~10-20 times) of glycan is normally required for a single MS spectrum. Matrix-assisted laser desorption/ionization time-of-flight (MALDI-TOF) mass spectrometry is the most widely used MS technique (Tsarbopoulos, Karas, and Strupat 1994). The information gained by using MALDI MS is mass weight, which can be used to assign putative monosaccharide structures present in a pure oligosaccharide since the mass of a monosaccharide is measured with a high degree of accuracy (Tsarbopoulos, Karas, and Strupat 1994).

For mass determination, glycans must first be separated from the protein backbone. Separation is usually carried out using enzymatic or chemical techniques. Enzymes such as trypsin and PNGaseF, glycosidase that cleaves between the innermost GlcNAc and asparagine residues of N-linked glycans, are most often used to separate glycans prior to analysis. There are disadvantages with both techniques in that the enzymatic release of glycans is dependent on the protein-glycan linkage and that not all glycans may be released, while chemical release may lead to the destruction of the non-carbohydrate parts of the protein backbone. When there are multiple glycosylation sites in a single protein the information as to the specific site and the exact protein-glycan linkage will be lost (Dalpathado and Desaire 2008, Budnik, Lee and Steen 2006).

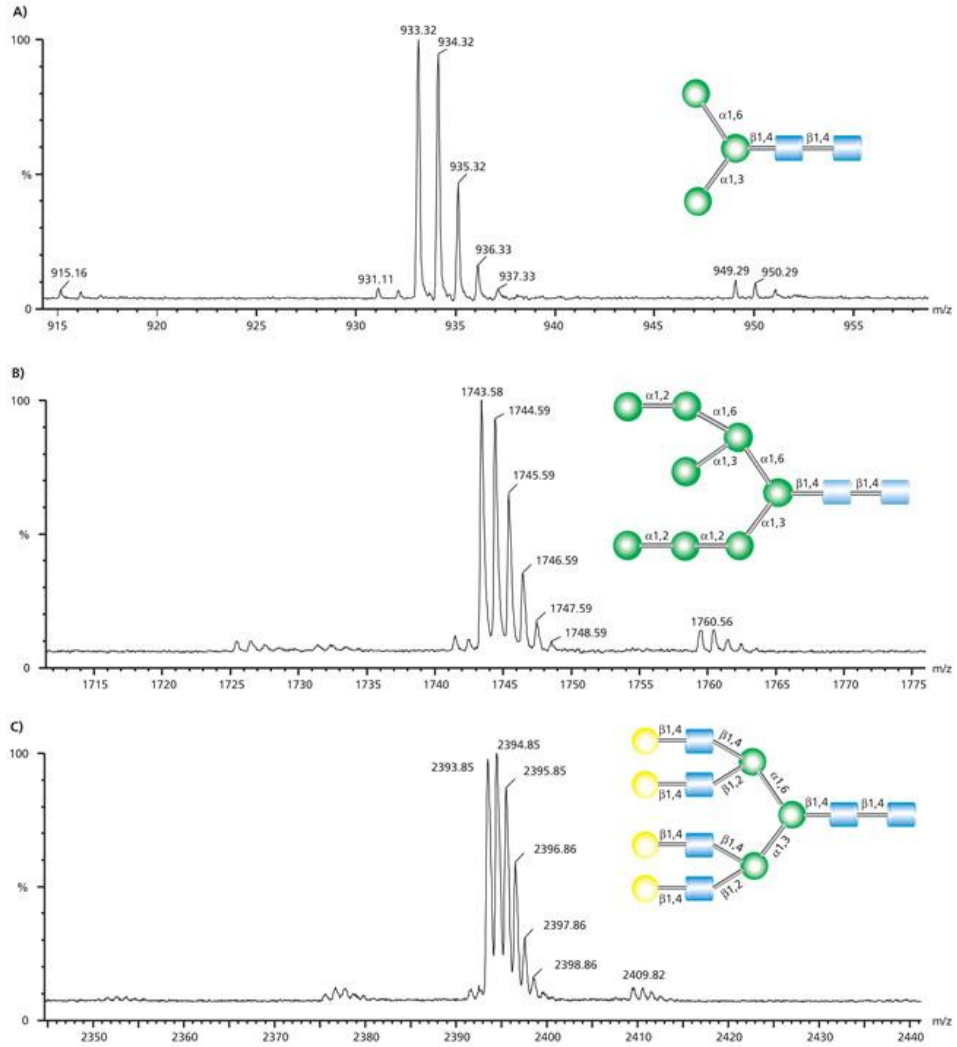


Figure 1.17: An example of MALDI-TOF mass spectra. (A) Man-3 glycan, (B) Man-8 glycan, and (C) NA4 glycan**.** (www.sigmaaldrich.com)

1.17.7 Nuclear Magnetic Resonance (NMR)

NMR analysis of glycoproteins is based on the fact that a glycan distorts in a magnetic field (Budnik, Lee and Steen 2006). It is a non-destructive technique and can provide full structural information. Therefore, NMR is a great source to get definitive information on glycan structures. However there are some limitations associated with this technique, which include the high cost of equipment and trained personnel and the amount of glycan required, concentrations of 3-4 orders of magnitude (10-100 ng) of the purified glycan compared to other techniques (Budnik, Lee and Steen 2006). Although NMR can provide information that cannot be obtained from other approaches and is well suited to the analysis of small sugars, it is not suited to the analysis of mixtures of glycans and the interpretation of the spectra can be made difficult by the similar environments of many of the protons (Taylor and Drickamer 2006).

1.18 Project Aims and Objectives

As explained previously there is a need for novel carbohydrate binding proteins for the detection and analysis of glycans and glycoproteins. In particular there is a need for the detection of cell surface glycans in disease states such as cancer. O-glycosylation has been shown to be affected in many cancers and it would be beneficial to be able to detect these changes. It was thought that the Cry toxins from *Bacillus thuringiensis* may be a good source of novel lectins. The aims of this project were to;

1. Clone a truncated Cry1Ac protein that could be expressed as a soluble active protein, which would have specificity to GalNAc
2. Clone the third domain of tCry1Ac to examine if the third domain on its own would have carbohydrate specificity
3. Produce a large yield of both proteins, by expression in *E. coli* and purification by Immobilised Metal Affinity Chromatography (IMAC) for downstream applications
4. Examine both proteins binding specificities for glycoproteins and carbohydrates by ELLA
5. Determine whether tCry1Ac or CryD3 have carbohydrate binding specificities and whether those specificities may be altered by mutagenesis
6. Determine whether Cry toxins from Bacillus thuringiensis are a good source of novel carbohydrate binding proteins.

Cry1Ac from *Bacillus thuringiensis*, kursatki HD-73 was selected for this study. Bioinformatic and structural analysis of the protein were carried out to examine the best cloning strategy for Cry1Ac (tCry1Ac) and domain III of Cry1Ac (CryD3). Recombinant proteins were produced in *Escherichia coli* based expression systems, with particular focus on the selection of an appropriate strain of *E. coli* and vector system for each, capable of producing soluble, active protein. Following successful expression of these molecules, protocols were developed to yield relatively large quantities of highly purified protein, for glycan binding characterisation studies. As stability is of particular important if these proteins are to be used for bioanalytical platforms, a comprehensive study on optimising their stability was carried out. tCry1Ac and CryD3 were also mutated within the proposed sugar-binding site to

confirm if the sugar specificity of these proteins could be manipulated and ultimately altered.

Chapter 2

Materials and Methods

2.1 Bacterial strains, primers sequences and plasmids

Table 2.1: Bacterial strains

Strain	Genotype	Features	Source
Escherichia coli			
JM109	FtraD36, proAB+ lacIq, lacZ M15 endA1 recA1 hsdR17(rk-, mk+) mcrA supE44- gyrA96 relA, (lacproAB) thi-1	All purpose cloning strain. Produces stable plasmid DNA	Sigma
KRX	[F', traD36, ΔompP, proA+B+, lacIq, Δ(lacZ)M15] ΔompT, endA1, recA1, gyrA96, (Nalr), thi-1, hsdR17 (rk-, mk+), relA1, supE44, Δ(lac-proAB), Δ(rhaBAD)::T7 RNA polymerase	Expression strain	Promega
BL21 (DE3)	FΔ dcm ompT hsdSB(rB-,mB-) gal	Expression strain	Novagen
Bacillus thuringiensis	Serovar kursaki HD-73	Cry1Ac	USDA, ARS Culture Collection in Peoria,I, USA

Table 2.2: Plasmids

Plasmid	Description	Source
Vector		
pQE60	Amp^R Expression vector for C terminus Histidine tag proteins, T5 promoter/lac operon	Qiagen
Constructs		
ptCry1Ac	pQE60 with truncated *cry1Ac*	This project
$ptCry1Ac_{N521A}$	pQE60 containing tCry1Ac with N521A mutation	This project
$ptCry1Ac_{QY487A}$	pQE60 containing tCry1Ac with QY487A mutation	This project
$ptCry1Ac_{N480A}$	pQE60 containing tCry1Ac with N480A mutation	This project
pCryD3	pQE60 with lectin domain of cry1Ac	This project
$CryD3_{PAV16A}$	pQE60 containing CryD3 with PAV16AAA mutation	This project
$CryD3_{N45A}$	pQE60 containing CryD3 with N45A mutation	This project
$CryD3_{QY52A}$	pQE60 containing CryD3 with QY52A mutation	This project
$pCryD3_{N86A}$	pQE60 containing CryD3 with N86A mutation	This project

Table 2.3: Primer Sequences

(Synthesised by Sigma-Aldrich, UK and integrated DNA technologies)

Name	Primer Sequence (5'-3')	T_M (°C)
Cry1Ac P60Df	ATCGATCATGATGGATAACAATCCGAACATCAATG	52.3
Cry1Ac P60Dr	ATCGAAGATCTAACTGGAATAAATTCAAATCT GTCTA	50.1
D3cry_ f	ATCGATCCATGGGTAATAATATAATTGCATCGGATAGTAT	60
D3cry_r	ATCGAAGATCTAATAAATTCAAATCTGTCTATTATCAC	60
mutPAVf	/5Phos/CTC AAA TCG CAG CAG CAA AGG G	60
mutPAVr	/5Phos/TAA TAC TAT CCG ATG CAA TTA TAT TA	48

Name	Primer Sequence (5'-3')	T_M (°C)
Mut480f	/5Phos/TTA AAT AGT AGT GGA GCT AAC ATT	50
Mut480r	/5Phos/TCT AAC TAA GTC CCC ACC AGT AA	55
mutQYf	/5Phos/AAG TTC CAA TTC ACT TCC CAT CGA	57
mutQYr	/5Phos/CAA TAG CCC CTC TAT TCG CAA TG	56
Mut521f	/5Phos/TTA ATT GGG GTG CTT CAT CCA TT	55
Mut521r	/5Phos/CGT TGA GGT GAA TCG GGG TTA	57
D3cry_ f	ATCGATCCATGGGTAATAATATAATTGCATCGGATAGTAT	60
D3cry_r	ATCGAAGATCTAATAAATTCAAATCTGTCTATTATCAC	60
Sequencing Primers		
QE-F	CCCGAAAAGTGCCACCTG	61.6
QE-R	GTTCTGAGGTCATTACTGG	53.5

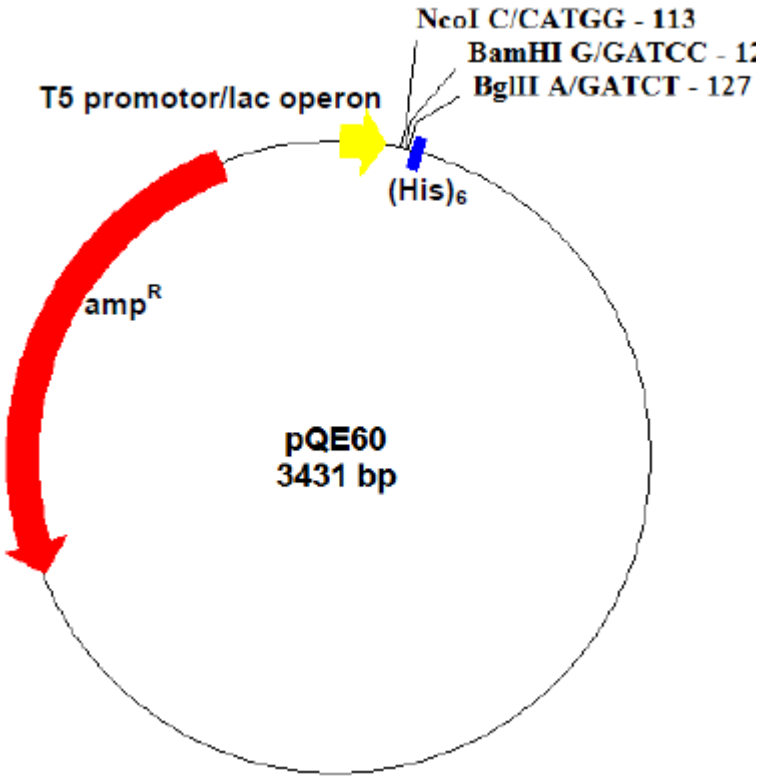


Figure 2.1: The pQE60 vector. The 3431bp pQE60 vector from Qiagen contains the following features. The multiple cloning site is located before the 6x histidine amino acid sequence (blue) which allows for the expression of C-terminally 6x histidine tagged proteins. This is under control of the T5 promotor/lac operon (yellow). The *bla* gene encodes beta-lactamase which confers ampicillin resistance to the bacteria (red). This image was generated using pDRAW32.

2.2 Microbiological Media

Bacteriological agar, tryptone and yeast extract were supplied by Lab M Ltd. All other chemicals were supplied by Sigma-Aldrich Co. All media was sterilised by autoclaving at $121^{O}C$ and $15lb/in^2$ for 20 minutes.

Luria Bertani (LB) Broth

Used for culturing *Escherichia coli* and *Bacillus thuringiensis* strains

Tryptone	10g/L
NaCl	10g/L
Yeast Extract	5g/L
pH	7.0

For LB agar, 15g/L bacteriological agar was included.

SOB medium

Tryptone	20g/L
Yeast Extract	5g/L
NaCl	0.5g/L
KCL	2.5mM
pH	7.0

The solution was sterilised and allowed to cool to $55^{O}C$. Filter sterilised solutions of 1m $MgCl_2$ and 1M $MgSO_4$ were added to final concentrations of 10mM each.

SOC medium

After making SOB medium as described above, filter sterilised 50% glucose was added to a final concentration of 20mM.

TB Medium

Tryptone 12g/L

Yeast extract 24g/L

Glycerol 4mls

Make up to 900mls and sterilise. Allow to cool and add 100mls of Potassium phosphate buffer

2.3 Solutions and Buffers

All Chemicals and solutions were obtained from Sigma Aldrich unless otherwise stated.

RF1

RbCl	100mM
$CaCl_2$	10mM
Potassium Acetate	30mM
Glycerol	15%
pH	5.8

When the pH had been adjusted with HCl $MnCl_2$ was added to a final concentration of 50mM. The solution was filter sterilised through a 0.22µm filter membrane and stored at 4°C.

7% Agarose

Agrose

TAE Buffer (1X)

TE Buffer

Tris-HCL	10mM
Na_2- EDTA	10mM
pH	8.0

TAE Buffer (50X)

Tris	242g/L
Glacial Acetic Acid	5.71% (v/v)
EDTA	50mM (from 0.5M Stock
pH	8.0

The solution was diluted to 1X with dH_2O before use

Solutions for 1, 2, 3 Method of Plasmid Preparation (Birnboim and Doly 1979)

Solution 1

Glucose	50mM
Na_2-EDTA	10mM (from 0.5M stock)
Tris-HCl	25 mM (from 1M stock)
pH	8.0

Solution 2

NaOH	200mM (from 1M stock)
SDS	1% (w/v)

Solution 3

Potassium Acetate	3M
pH	4.8

To 60ml of 5M potassium acetate, 11.5ml of glacial acetic acid and 28.5ml of dH_2O was added. The resulting solution was 3M with respect to potassium and 5M with respect to acetate.

Western Blot Transfer Buffer

Tris	25 mM
Glycine	150 mM
Methanol	10% (v/v)

TBS Buffer

Tris	20 mM
NaCl	150 mM
CaCl2	1 mM
MgCl2	1 mM
pH	7.6

When the pH had been adjusted with HCl MnCl2 was added to a final concentration of 1mM. For TBST, the detergent Triton X-100 was added to a final concentration of 0.1% (v/v)

PBS (10x)

Na^2HPO4	10.9 g/L
NaH^2PO4	3.2 g/L
NaCl	90 g/L

A 1X has a pH of 7.2. For PBST, the detergent Triton X-100 was added to a final concentration of 0.1% (v/v)

Agarose Gel Loading Dye (6X)

Bromophenol Blue	0.25% (w/v)
Xylene Cyanol	0.25% (w/v)
Ficoll (Type 400)	15% (w/v)

Made in dH_2O and sterilised by autoclaving

Ethidium Bromide Stain

A 10mg/ml solution of ethidium bromide was stored at 4^OC in the dark. For the staining of agarose gels, 100ul of this stock was mixed with 1L of dH_2O. The solution was kept in a plastic tray and covered to protect against the light. Fresh stain was made every 1-2 weeks. Used ethidium bromide stain was collected and ethidium bromide was removed by mixing with a de-staining bag (GeneChoice) overnight. The clear liquid was disposed of routinely, while the ethidium waste was incinerated.

SDS-PAGE Sample (Laemmli) Buffer (5X) (Laemmli 1970)

Glycerol	50% (v/v)
2-mercaptoethanol	5% (v/v)
SDS	2% (w/v)
Bromophenol Blue	0.1% (w/v)
Tris-HCl, pH 6.8	62.5mM

SDS-PAGE Running Buffer (5X)

Tris	125mM
Glycine	960mM
SDS	0.5% (w/v)
pH	8.3

Coomassie Blue Stain for SDS-PAGE Gels

Methanol	45% (v/v)
dH_2O	45% (v/v)
Glacial Acetic Acid	10% (v/v)
Coomassie Blue	0.25% (w/v)

To make destaining solution for coomassie blue stained gels, the coomassie blue was omitted.

Buffers for Ni-NTA purification of histidine tagged proteins

Lysis Buffer

NaH2PO4	50 mM
NaCl	300 mM
Imidazole	40 mM
pH	8.0

Wash buffer

NaH2PO4	50 mM
NaCl	300 mM
Imidazole	100 mM
pH	8.0

Elution buffer

NaH2PO4	50 mM
NaCl	300 mM
Imidazole	300 mM
pH	8.0

Column Stripping Buffer

NaH2PO4	50 mM
NaCl	300 mM
EDTA	50 mM
pH	8.0

Table 2.4 Antibiotics

Antibiotic	**Stock Conc**	**Prepared in**	**Working conc. (Solid media)**	**Working conc. (Liquid media)**
Ampicillin	100mg/ml	dH_2O	100ug/ml	100ug/ml

2.4 Storing and Culturing of Bacteria

Glycerol stocks were prepared in duplicate for each strain by picking a colony from a streak plate and inoculating 5ml of LB broth with a specific antibiotic if required overnight at 37^{O}C. A 0.5ml aliquot of this exponentially growing culture was added to 1.0ml of sterile 80% (v/v) glycerol in a cryogenic tube. Stocks were stored at -20^{O}C and the duplicate stored at -80^{O}C. Working *E. coli* and *B. thuringiensis* stocks were stored on agar plates at 4^{O}C.

2.5 Isolation and purification of DNA

2.5.1 Preparation of Total Genomic DNA using the Wizard Genomic DNA Kit

The kit was used according to the manufacturer's instructions (Promega) with some modifications. An overnight culture of the appropriate organism (*Bacillus thuringiensis*) was prepared in LB and a 1ml aliquot of this culture was pelleted by centrifugation at 13,000rpm for 2 minutes. As *Bacillus thuringiensis* is Gram positive some extra steps were required for lysis. The supernatant was removed and the pellet was resuspended in 450µl of 50mM EDTA and 120µl of lysozyme at 10mg/ml. The suspension was then incubated at 37^{O}C for 60 minutes to break down the bacterial cell wall. This was then centrifuged for 2 minutes at 13,000rpm and the supernatant was removed. The pellet was resuspended in 600µl of nuclei lysis solution. The suspension was then incubated at 80^{O}C for 5 minutes to lyse the cells and then cooled to room temperature. 3ul of RNase solution was added to the cell lysate. The tube was inverted several times to mix and incubated at 37^{O}C for 60 minutes. To remove the protein in the sample, 200µl of protein precipitation solution was added and vortexed for 20 seconds. The sample was then incubated on ice for 5 minutes and centrifuged at 13,000rpm for 3 minutes. A phenol choloroform extraction was then carried out.

600µl of the sample supernatant was taken and 400µl of phenol was added. This solution was inverted and centrifuged at 13,000g for 5 minutes. This step was repeated. The supernatant was decanted and transferred to a clean 1.5ml tube. 450µl chloroform was added to this sample and the solution was inverted this was then

centrifuged at 13,000g for 5 minutes. 600µl of isopropanol was added to the solution and inverted, this was then centrifuged at 13,000g for 5 minutes. The pellet was then kept from the last step and washed with 100µl 70% ethanol. This was centrifuged for 3 minutes at 13,000g. The pellet was then air dried and 50µl of DNA rehydration solution was added and incubated overnight at 4 degrees.

2.5.2 Isolation of plasmid DNA

Two procedures for the isolation of plasmid DNA were used. The 1-2-3 method was used for convenient plasmid isolation from large numbers of samples in screening. The Hi Yield plasmid mini Kit from ISIS was used to prepare consistently pure DNA.

2.5.2.1 1-2-3- Method

Bacterial growth was taken off an agar plate with a sterile loop and re-suspended in 200 µL of Solution 1. Alternatively, 1.5 mL of a bacterial culture in a microfuge tube was centrifuged at 13,000rpm for 2 minutes to collect the cells. The supernatant was discarded and the cell pellet re-suspended in 200 µL of Solution 1. The re-suspension was left for 5 minutes at room temperature. 200 µL of Solution 2 was added and the tube was mixed by inversion and left for 5 minutes at room temperature. Finally 200 µL of Solution 3 was added, the tube was mixed by inversion and left at room temperature for 10 minutes. Cell debris was collected by centrifugation at 13,000rpm for 10 minutes. The supernatant was transferred to a fresh microfuge tube and a phenol chloroform extraction was carried out. 500 µL of phenol chloroform isoamylalcohol (25:24:1) was added to the tube and mixed briefly by vortexing. This was then centrifuged at 13,000rpm for 5 minutes. The top aqueous layer was removed and an equal volume of isopropanol was added and mixed by inversion. The tube was centrifuged at 13,000rpm for 5 minutes to pellet the plasmid DNA. The pellet was washed with 70% ethanol and then air dried or dried briefly in a Speed Vac (Savant) vacuum centrifuge. The plasmid DNA was resuspended in 30 µL of TE

buffer and 1 µL of Ribonuclease A was added to digest RNA. The plasmid DNA was stored at -20° C.

2.5.2.2 High Yield Plasmid Mini Prep Kit (BBC Bioscience)

The kit was used according to the manufacturer's instructions with some alterations. 1.5 mL of bacterial culture was added to a microfuge tube and this was centrifuged at 13,000rpm for 1 minute to pellet the cells. Also bacteria can be collected from an agar plate with a sterile loop and resuspended in 200 µL of PD1 buffer. The respuspended cells were then lysed by adding 200 µL pf PD2 buffer and mixed gently by inverting the tube 10 times. The mixture was left to stand for 2 minutes at room temperature until the lysate clears. 300 µL of PD3 buffer was added to neutralise the PD2 and the sample was inverted 10 times. This was then centrifuged for 2 minutes at 13,000rpm. The supernatant was then transferred to a PD column and centrifuged at 13,000rpm for 30 seconds. The flow through was discarded and the PD column returned to the 2mL collection tube. To wash the DNA, 400 µL of W1 buffer was added to the PD column and this was centrifuged for 30 seconds at 13,000rpm. The flow through was again discarded and 600 µL of wash buffer (ethanol) was added to the PD column. The column was centrifuged at 13,000rpm for 30 seconds and the flow through discarded. The column was then dried by centrifugation again at 13,000rpm for 10 minutes. The dried column was then transferred to a fresh microfuge tube. 50 µL of elution buffer or ddH_2O was directly added onto the centre of the membrane. The walls of the column were avoided to stop residual buffer adhering to them. This was then allowed to stand for 2 minutes until the liquid is absorbed. The column was then centrifuged at 13,000rpm for 2 minutes to elute the purified DNA.

2.5.3 Agrose Gel Electrophoresis

DNA was analysed by electrophoresis in agarose gels in a BioRad horizontal gel apparatus. Samples were run on a 0.7-1.0% agrose gel at 120V for 25-30 minutes. Agarose was prepared by adding the required amount to TAE buffer and dissolved by boiling. Agarose was stored at 60°C until needed. TAE was used as the running buffer. Agarose gel loading dye was mixed with the DNA samples to help loading of the DNA into the gel and to give an indication of migration distance during electrophoresis. After running the gels for 25-30 minutes the gels were then put into a bath of ethidium bromide for approximately 15-20 minutes so that an image of the gel could be obtained using a UV transilluminator coupled with an image analyser. On every gel 0.5 µg of 1 Kb Plus DNA ladder (Invitrogen) was run as a molecular size marker.

2.5.4 Isolation of DNA from agarose gels

The Hi Yeild^{TM} Gel/PCR DNA extraction kit from RBC Bioscience was used to isolate DNA from agarose gels. The kit was used according to the manufacturer's instructions: The DNA band to be isolated from the agarose gel was excised using a scalpel. Up to 300mg of a gel slice was transferred to a microfuge tube. 500 µL of DF buffer was added to the sample and mixed by vortexing. The sample was incubated at 55-65°C for 10-15 minutes, inverting every 2-3 minutes until the gel slice was completely dissolved. A DF column was placed into a collection tube and up to 800 µL of the dissolved gel sample was transferred to the DF column. The DF column was centrifuged at 13,000rpm for 30 seconds. The flow through was discarded and 500 µL of wash solution was added to the DF column which was centrifuged again at 13,000rpm for 30 seconds. The flow through was again discarded and the column was dried by centrifugation at 13,000rpm for 10 minutes. 30 µL of elution buffer or ddH_2O was added to the DF membrane. The column was left to stand for 2 minutes and then centrifuged at 13,000rpm for 2 minutes to elute the isolated DNA. The DNA was stored at -20°C.

2.6 Preparation of high efficiency competent cells by the RF method

5mL of LB broth containing the relevant antibiotics was inoculated with a single colony of *E. coli* from a plate and cultured overnight at 37°C. A 1L flask with 200mL of SOB broth was inoculated with 2mL of the overnight culture and incubated at 37°C shaking at 180-220rpm. When the culture had reached an OD_{600} of 0.5 the flask was cooled in ice water. All subsequent steps took place at 4°C. The culture was transferred to a sterile centrifuge bottle. The cells were collected by centrifugation at 3,000rpm for 5 minutes (using a Beckman JA-14 rotor). The supernatant was decanted and the cells gently re-suspended in 60mL of chilled RF1 buffer. The suspension was left on ice for 90 minutes. The cells were again collected by centrifugation at 3,000rpm for 5 minutes. The supernatant was decanted and the cells gently re-suspended in 8mL of chilled RF2 buffer. Aliquots of 400 µL were prepared in sterile 1.5mL microfuge tubes and flash frozen using -80°C metal blocks. The competent cells were stored at -80°C.

2.6.1 Transformation of competent cells

An aliquot of competent cells was thawed on ice. 2 µL of plasmid DNA was mixed with 200 µL of the cell suspension in a sterile 1.5mL microfuge tube. The mixture was incubated on ice for 30 minutes. The cells were heat shocked at 42°C for 45 seconds and placed back into ice for 2 minutes. After this 800 µL of LB broth was added to the cells and they were incubated for 1 hour at 37°C. 200 µL of the sample was spread on LB agar and incubated at 37°C overnight.

2.6.2 Determination of cell efficiency

Competent cell efficiency depends on the number of colonies formed per µg of transformed plasmid DNA. A 10 ng/ µL stock of pUC18 plasmid DNA was diluted to 1ng/ µL, 100pg/ µL and 10 pg/ µL. 1 µL of each dilution was transformed as described above. The cell efficiency was calculated from the number of colonies obtained, taking into account the dilution factor and the fraction of culture transferred to the spread plate.

2.7 Enzymatic reactions

All enzymes and relevant buffers were obtained from Invitrogen Life technologies, New England Biolabs, Sigma Corporation or Bioline and were used according to the manufacturer's instructions.

2.7.1 Polymerase chain reaction

PCR reactions (Mullis and Faloona 1987) were carried out using a Veriti Thermocycler (Applied Biosystems)

2.7.1.1 Standard PCR reaction mixture 50 µL

Buffer (5X) 10 µL
Primers (10µM) 1 µL of each
dNTP's 1 µL
Template DNA 1 µL
dH_2O 35 µL
Taq polymerase 1 µL

2.7.1.2 Standard PCR programme cycle for Velocity taq polymerase reactions

Stage 1: Step 1: 95°C for 10 minutes
Stage 2: Step 1: 95°C for 30 seconds
Step 2: Annealing temperature for 30 seconds
Step 3: 72°C for 30 seconds per kb to be synthesised
Stage 3: Step 1: 72°C for 10 minutes
Stage 2 was carried out for 30 cycles. A 1:3 dilution of velocity Taq polymerase was carried out before use in PCR.

2.7.2 Restriction Digests

DNA/PCR 8 µL
Buffer 1 µL
Enzyme 1 µL
Reactions were incubated for 3 hours or overnight at 37°C.

2.7.3 Ligation reaction

PCR product	6 µL
Vector	10 µL
T4 DNA ligase	1 µL
Ligase buffer(10X)	2 µL
dH_2O	1 µL

Reactions were incubated for 3 hours at room temperature of 16 hours at 4°C.

2.7.4 Site specific mutagenesis

Mutations were introduced into open reading frames on plasmid constructs by PCR amplification using complementary phosphorylated primers carrying the desired mutation. A standard PCR reaction mixture as show in Section 2.7.1.1 was set up. The PCR was carried out using the high fidelity Velocity *Taq* polymerase. The standard program cycle for Velocity *Taq* polymerase (Section 2.7.1.2) was used. Typically an extension time of 2 min was used to amplify plasmids of 3.5-4 kb. The template DNA was eliminated by digestion with *Dpn*I restriction endonuclease. *DpN*I is biased towards a methylated recognition sequence. It selectively digests the template DNA from *dam+ E. coli* strains over the newly synthesized DNA. The amplified plasmid, with the incorporated mutation, is circularized by ligation which was facilitated by the use of phosphorylated primers.

2.8 DNA sequencing

Recombinant clones were verified by DNA sequencing. Commercial sequencing services were provided by MWG Biotech/Eurofins and Source BioScience gene service. Suitable sequencing primers for standard vectors were sent with samples to be analysed. Samples were sent as plasmid preparations

2.9 In silico analysis of DNA sequences

DNA sequences for the strains used in this study were obtained from the following sequencing studies;
DNA and protein sequences were aligned using the ClustalW, available at http://www.ebi.ac.uk/Tools/clustalw/ and Gendoc programme, available to download at http://www.nrbsc.org/. DNA sequences were analysed for restriction sites using Webcutter 2.0 programme.

2.10 Protein modelling using I-Tasser and Pymol

Protein sequences were sent to the I-Tasser server in fasta format (http://zhanglab.ccmb.med.umich.edu/I-TASSER/). An e-mail with the predicted protein structure models was sent by the I-Tasser server to the user. Predicted protein models and their scores were given along with any enzyme and lectin binding domain similarity. The protein data files were downloaded and were used as the basis for PyMol modelling. PyMol was downloaded from the PyMol website (http://www.pymol.org/). PyMol interface was opened and the protein data file of the protein of interest was uploaded by the molecular graphics system file > open > data file(.pdb). The protein structure was observed on the PyMol viewer in stick molecule formation. The protein was then modelled using the PyMol viewer command system and the PyMol molecular graphics system. The protein could be rotated and areas selected using the mouse on the PyMol viewer.

- Display a white background for pictures (Display > background > white).
- Obtain the colours for publication (Display > colour space > CMYK for publications).

- Change the protein structure from stick molecules to cartoon ribbons. (Hide all > show cartoon ribbons)
- Show the surface of the protein (select all > show surface)
- Colour the helical and beta barrel structures. (colour > by ss > Helix, sheet, loop)
- Colour selected regions (select region > colour > select colour)
- Mutate regions or amino acids, the sequence was added to the PyMol viewer so residues could be selected with ease (display > sequence). For mutagenesis (wizard > mutagenesis > pick a residue > no mutation, select residue change > apply > done.
- Obtain a high resolution picture (Ray)
- Save the session (file > save session as > file name)
- Save the model as a image (file > save image as > PNG > file name)

2.11 Data analysis of ELLA results

An ELLA was read on the spectrometer at 450nm. The data was transferred into a excel file for analysis. Each sample was carried out in triplicate and the average of the three samples was taken. The averages of the negative controls, PBS and TBST were taken away from the average of the samples. Standard deviation was obtained using the custom standards from the averages of each sample. A column graph was used to graph the ELLA results while concentration ELLA results were graphed using xy scatter plot. The standard deviation was added to the graphs as custom error bars. All ELLAs were carried out an average of three times to confirm conststant results.

2.12 Standard expression culture

An LB plate with the appropriate antibiotic was streaked from a glycerol stock of the strain containing the expression plasmid. A colony was used to inoculate 10 mL of LB broth containing the appropriate antibiotic, and grown over night at 37°C while shaking at 200rpm. A 1 L baffled clonical flask containing 500 mL of LB broth was inoculated with 5 mL of the overnight culture and the appropriate antibiotic was added. The culture was incubated at 37°C, shaking at 200rpm, until an optical absorbance (a_{600}) of 0.4-0.6 was reached. IPTG was added to a final concentration of 50µM to induce expression. The culture was incubated at 20°C -30°C overnight. The culture was then centrifuged at 4,000rpm for 10 minutes (using a Beckman JA-14 rotor) to collect the cells. The supernatant was autoclaved and discarded, and the pellets were stored at 20°C.

2.13 Preparation of time course expression samples

2.13.1 Culture growth and sampling

5mL of an overnight culture was used to inoculate 500mL of LB broth to which the appropriate antibiotic had been added. The culture was incubated at 37°C until the OD_{600} has reached 0.5-0.6. An initial 6mL aliquot (time zero) was taken and placed on ice for 30 minutes prior to harvesting of the cells by centrifugation. IPTG was added to the culture to a final concentration of 50µM and 6mL aliquots were taken every hour for an additional 5 hours (samples time 1 to time 4). All of the samples were placed on ice for at least 30 minutes prior to harvesting by centrifugation. The culture was allowed to grow overnight and a final sample was withdrawn. Having been incubated on ice for 30 minutes all samples were harvested by centrifugation at 4000 rpm for 10 minutes and the cell pellets re-suspended in 6mL of PBS.

2.13.2 Clarified media samples for SDS PAGE – monitoring extracellular protein

A sample of the clarified media was retained for SDS PAGE analysis. Therefore a 500ml aliquot was spun at 13,000rpm for 10 minutes to ensure removal of all cells. The clarified media was transferred to a fresh tube and frozen for analysis on SDS PAGE.

2.13.3 Growth curve preparation

Cell pellets were re-suspended in 6mL of buffer and readings taken at 600_{nm} to enable the construction of growth curves. The readings were also used to calculate the volume of culture required for preparation of samples for SDS-PAGE analysis.

2.13.4 Sample preparation for total protein analysis

The volume of each time course sample needed to prepare samples for total protein analysis was calculated using the following equation.
$(0.7/OD_{600nm})$ x 300 = Volume of culture to be harvested (µL)
The cells were harvested by centrifugation in microfuge tubes at 13,000rpm and were re-suspended in 50µL of 1x SDS-PAGE loading buffer. 20 µL of each sample was then run on SDS-PAGE.

2.13.5 Sample preparation for soluble and insoluble protein analysis

The volume of the time course samples that were harvested to prepare samples for soluble and insoluble protein analysis was calculated using the following equation
$(0.7/OD_{600nm})$ x 2400 = volume of culture to be harvested (µL)
The harvested pellet was then re-suspended in 40 µL of 1x cell lytic (prepared from a 10x stock). The re-suspended cells were left to lyse overnight at 4°C or at room temperature for at least 1 hour. Once the cells were lysed the insoluble cell debris

was pelleted by centrifugation at 13000rpm for 20 minutes and the soluble fraction removed to a fresh tube.

2.13.5.1 Soluble protein analysis

For analysis of soluble protein the soluble fraction was diluted at least 1:5 prior to running on SDS-PAGE, as the components of the cell lytic have a negative impact on the protein separation. Samples were prepared by mixing 5 µL of the soluble fraction with 15 µL of the appropriate buffer and then 5 µL of 5x SDS-PAGE loading dye. The entire sample was loaded onto the SDS-PAGE gel.

2.13.5.2 Insoluble protein analysis

For analysis of insoluble protein the pelleted insoluble material was re-suspended in 50 µL of 1x SDS PAGE loading dye and heated for 10 minutes at 95°C. The samples were centrifuged at 13,000rpm for a few seconds and 15-20 µL was run on the SDS-PAGE gel.

2.14 Preparation of cleared lysate for protein purification

Cell lysis was carried out using a cell disruptor.

2.14.1 Cell lysis by cell disruption (constant cell disruption ltd)

A cell pellet from a 500mL expression culture was re-suspended in 10% of the cell pellet weight in lysis buffer containing 20-40mM imidazole, pH 8.0. 1% Antifoam was added to the cell mixture to prevent foaming in the cell disruptor. 200mL lysis buffer was put over the cell disruptor at 15kpsi to equilibrate the machine. The cell mixture was passed through the machine at 15kpsi to lyse the cells and this was followed by the same amount of lysis buffer to wash any reside from the system. The lysed mixture was collected and spun at 13,000rpm for 50 minutes at 4°C to separate the insoluble and soluble fractions. The cleared lysate was transferred to a fresh container and then passed through a 0.2 µM filter to remove any remaining cell

debris and stored at 4°C. The cell disruptor was cleaned by passing through 500mL of Ethanol followed by 500mL of distilled water at 15-20kpsi.

2.15 Protein Purification

2.15.1 Protein purification by standard IMAC procedure

Protein purification of 6x histidine tagged proteins was carried out by IMAC. A 1mL Ni- Sepharose Fast flow column was washed with 5-10 column volumes of dH_2O. The column was equilibrated with 5-10 column volumes of lysis buffer containing 20-40mM imidazole. The cleared lysate was gently mixed with the nickel resin in a column. The column was washed with 10 column volumes of lysis buffer containing 20-40mM imidazole. A gradient of imidazole in lysis buffer was used to elute contaminating proteins from the column. The target protein was eluted from the column using a high concentration of imidazole in lysis buffer (120-500mM imidazole). The column was washed with 5-10 column volumes of water followed by 3-5 column volumes of 20% ethanol. Samples were taken at each step and analysed on SDS-PAGE electrophoresis.

2.15.2 Cleaning and recharging of IMAC resin

The column was routinely recharged and cleaned prior to re-use of the Ni-NTA resin. The used resin was poured into a column and washed with 5 columns of dH_2O. The resin was then stripped by washing with 5 column volumes of stripping buffer. Ionically bound proteins were removed by washing with 5 column volumes of 1.5M NaCl. Precipitated proteins, hydrophobically bound proteins, lipoproteins and lipids were removed by washing with 5 column volumes of 1M NaOH, contact time 1-2 hours, followed by 10 column volumes of dH_2O. Remaining impurities were removed by washing with 10 column volumes of 30% isopropanol for 30 minutes, followed by 10 column volumes of dH_2O. The resin was charged by adding 2 column volumes of 100mM $NiSO_4$. The resin was washed again with 2 column volumes of dH_2O before storage in 20% ethanol.

2.15.3 Protein purification by Size exclusion Chromatography

The separation of protein contaminants from recombinant proteins under native conditions was done by size exclusion chromatography. Size exclusion chromatography was carried out using FPLC. Before attachment of the gel filtration column to the AKTA purifier the maximum back pressure was set according to the manufacturer's instructions. The column storage valve was disconnected from the gel filtration column, and the top was attached to pump outlet no 1. The bottom of the column was attached directly to the FPLC. To remove the column storage buffer (20% ethanol) two column volumes of water were pumped through the column at a flow rate of 1 mL/min (unless otherwise stated). The column was then equilibrated with 5 column volumes of sample buffer, at the same flow rate, before sample application.

2.15.3.1 Toyopearl HW-55S

With pump A attached to a reservoir of degassed sample buffer 10ml of the protein sample was applied through the sample injection port. The run commenced, with a flow rate of 1 mL/min of buffer A and a maximum back pressure of 0.5 MPa. Following sample injection 2.5 column volumes of buffer was passed through the column. The proteins retention time was measured using the online Monitor U-900 (GE Healthcare), which read the eluent absorbance at 280 nm.

2.15.4 Protein purification by Ion Exchange Chromatography

Ion exchange chromatography was carried out to separate proteins by charge. At a pH lower than a proteins PI the protein will be positively charged and so it binds to a cation exchanger. Toypearl CM sepharose was used as a weak cation exchanger. PBS at a particular pH, generally between the PI of both proteins, was used to equilibrate 0.5ml of CM sepharose in triplicate. The protein sample was buffer

exchanged into PBS at the appropriate pH in triplicate. Both samples were mixed gently overnight at 4°C. The CM sepharose was spun at 13,000rpm for 5 minutes and the PBS supernatant was taken off. The protein sample was added to the CM sepharose and gently mixed for 1 hour at 4°C. The sample was then spun again at 13,000rpm for 5 minutes and the supernatant extracted. This contained the negatively charged proteins that did not bind to the resin. The protein that bound to the CM sepharose was eluted by adding PBS that has a higher pH than the protein's PI and mixing gently for one hour. The sample is spun at 13,000rpm for 5 minutes and the supernatant extracted. Analysis of the protein samples was carried out on SDS PAGE.

2.15.5 Purification of proteins using detergent: Triton X100

During the cell disruption and IMAC process a concentration of Triton X100 ranging from 1 to 10% was added to the sodium phosphate buffer. IMAC was carried out as in section 2.21. The protein was then eluted in 300mM imidazole with added Triton X100. The eluted protein was then analysed on SDS PAGE and buffer exchanged into PBS to be analysed on ELLAs.

2.15.6 Purification of proteins by competing sugars

The IMAC procedure was carried out as normal and during the washes 100mM of a sugar eg. Galactose was added to the Sodium phosphate buffer with imidazole. The column was again washed with sodium phosphate buffer and imidazole and then the protein was eluted in 300mM imidazole. The eluted protein was analysed on SDS PAGE and buffer exchanged to remove any residual imidazole and sugar.

2.16 Examining the stability of a protein

The protein stability was analysed through these steps and then examined by protein quantification and on ELLA's. The pH was tested in steps of one pH unit in between pH 2 and 10. The salt stability was tested with 0.2M NaCl and 0-2M $(NH_4)_2SO_4$ in steps of 0.5M. Acetonitrile and methanol was tested in 10% steps between 0 and 50%. The temperature stability was tested in +10°C steps from 4°C to 40°C. The stability and occurrence of proteolytic activity was tested by leaving an aliquot of the sample at room temperature overnight. The aliquot was then centrifuged and the activity and UV absorbance at 280nm was measured in the supernatant.

2.17 Protein quantification by 280nm readings

The quantification of protein concentration by 280nm reading is based on the proteins ability to absorb ultraviolet light in solution with the absorbance of 280nm. Prior to reading all samples were spun down at 13,000rpm for 10 minutes to minimize any debris present in solution that may interfere in absorbance readings. The UV lamp was set to 280nm and the absorbance was calculated with buffer solution. The absorbance of the protein solution was read in triplicate. The concentration of the protein was then calculated using Beer-Lambert law;

$A=\text{Ɛ}cl$

Equation: Beer – Lambert equation for the quantitative determination of protein concentration from absorbance readings at 280nm. A=absorbance at 280nm, Ɛ=molecular extinction coefficient, c=concentration (in units corresponding to Ɛ) and l=path length.

2.18 Sodium Dodecyl Sulfate Polyacrylamide Gel Electrophoresis (SDS-PAGE)

2.18.1 Preparation of SDS gels

Protein samples were analysed by Sodium Dodecyl Sulfate Polyacrylamide Gel Electrophoresis (SDS-PAGE), based on the method outlined by Laemmli (Laemmli 1970).
10%, 12.5% and 15% gels were prepared as outlined in the table

Table 2.5: Preparation of SDS gels

Resolving Gel	**10%**	**15%**	**4%**
30% Acrylamide (mL)	2.501	3.750	0.325
H_2O (mL)	3.007	1.758	1.540
1.5M Tris-HCl pH8.8 (mL)	1.875	1.875	-
0.5M Tris-HCl pH 8.8 (µL)	-	-	625
10% APS (µL)	37.5	37.5	12.5
10% SDS (µL)	75	75	25
Temed (µL)	3.75	3.75	2.5

2.18.2 Sample Preparation

20 µL of the protein sample was added to 5µL of 5x SDS-PAGE loading buffer. Samples were boiled for 5 minutes and applied to wells that had been flushed of un-polymerised acrylamide.

2.18.3 Sample application

10-15 µL of the prepared sample was loaded to each SDS-PAGE well. One lane on each gel was kept for the molecular weight protein marker (figure 2.1). A 20 µL aliquot of the marker was loaded to the gels for coomassie staining. For a western blot, the NEB pre-stained and colour plus pre-stained protein markers were used.

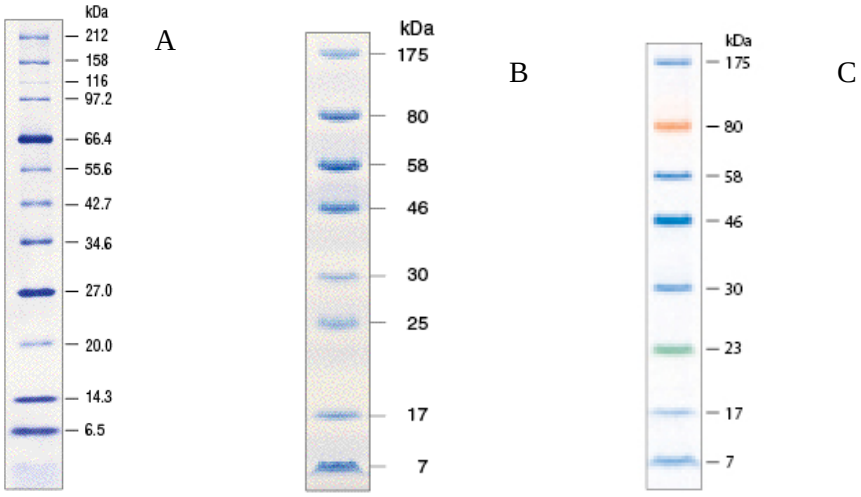


Figure 2.2: A; NEB broad range protein marker, B; Pre-stained protein marker and C; Colour plus pre-stained protein marker. Representative images of the NEB broad range protein marker, pre-stained protein marker and colour plus pre-stained protein marker used in this study. Images obtained from www.NEB.com.

Table 2.6: NEB Protein marker, representative bands.

Protein	**Source**	**Molecular Weight (daltons)**
Myosin	rabbit muscle	212,000
MBP-β-galactosidase	*E. coli*	158,194
β-Galactosidase	*E. coli*	116,351
Phosphorylase b	rabbit muscle	97,184
Serum albumin	**bovine**	**66,409**
Glutamic dehydrogenase	bovine liver	55,561
MBP2*	*E. coli*	42,710
Thioredoxin reductase	*E. coli*	34,622
Triosephosphate isomerase	***E. coli***	**26,972**
Trypsin inhibitor	soybean	20,100
Lysozyme	chicken egg white	14,313
Aprotinin	bovine lung	6,517
Insulin A	bovine pancreas	3,400
B chain	bovine pancreas	2,340

2.18.4 Gel analysis

Polyacrylamide gels were removed from the electrophoresis box and washed with dH_2O. Gels were stained for one hour with coomassie blue stain solution. An overnight de-stain followed using coomassie blue de-stain solution. The gels were then put into dH_2O to enhance the protein bands further.

2.19 Western blot

A SDS-PAGE gel was run using NEB pre-stained molecular weight markers. Four pieces of blotting paper and a piece of nitrocellulose membrane were cut to the same dimensions as the SDS-PAGE gel. The blotting paper, nitrocellulose and the gel were soaked in transfer buffer, and arranged on the semi-dry electroblotter. Transfer then occurred at a constant 20V for 15 minutes. For the detection of recombinant 6x histidine tagged proteins, the membrane was blocked with 3% BSA-TBST for one hour. The membrane was washed four times for 5 minutes with TBST, and then incubated with 1:10,000 murine anti-His antibody- HRP in 3% BSA-TBST. The membrane was washed four times for 5 minutes with TBST and incubated in 15mL dH_2O containing sigmafast 3,3-diaminobenzidine tablets. dH_2O was used to stop the development.

To detect glycoproteins the membrane was blocked with Carbo-Free blocking solution (Vector Laboratories), diluted to the correct concentration in PBS. The membrane was washed four times for 5 minutes in TBST then incubated for 1 hour with a 1 μg/ml solution of a biotin labelled lectin of choice in TBST. The membrane was washed again then incubated with 1:10,000 horseradish peroxidase (HRP) labelled murine anti-biotin antibody in TBS. After a final washing the membrane was developed using 15 ml of dH2O containing a SIGMA*FAST* Ô 3,3_-diaminobenzidine tetrahydrochloride tablet. The developed blot was washed with dH_2O.

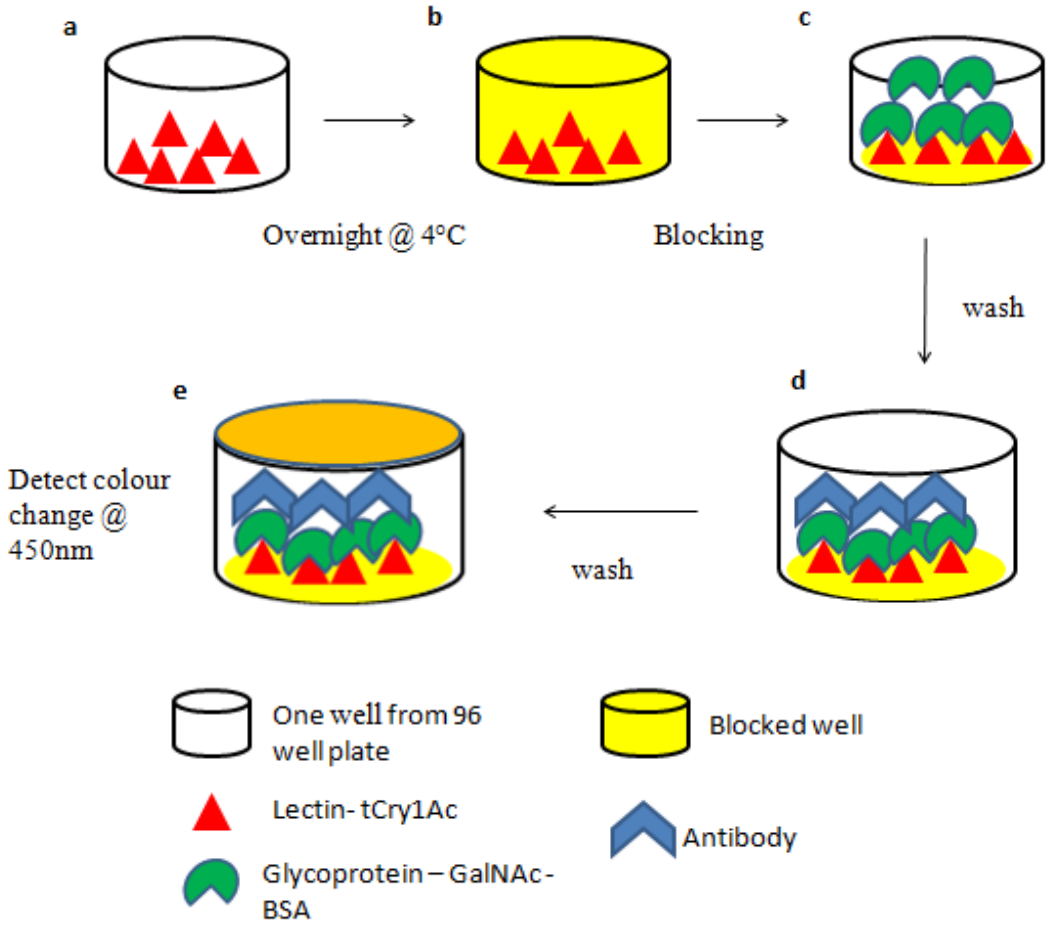
2.20 Protein identification by MS peptide mapping and sequencing analysis

The protein samples were reduced and alkylated with iodoacetamide, i.e. carbamidomethylated, and subsequently digested with trypsin that cleaves after lysine and arginine residues. The resulting peptides were spotted onto an anchorchip target for analysis on a Bruker Autoflex Speed MALDI TOF/TOF instrument. The peptide mixture was analyzed in positive reflector mode for accurate peptide mass determination. MALDI MS/MS was performed on 15 peptides for peptide fragmentation analysis. The MS and MS/MS spectra were combined and used for database searching using the Mascot software. The data are searched against in-house protein databases downloaded from NCBI, including the NRDB database containing more than 20 million known non-redundant protein sequences. The data can also be searched against a custom database containing specific protein sequences. The Mascot software finds matching proteins in the database by their peptide masses and peptide fragment masses. The protein identification is based on a probability-scoring algorithm (www.matrixscience.com).

2.21 Enzyme linked lectin assay

50µL of a glycoprotein was immobilised in a well of a NUNC MaxiSorp ELISA plate at 4°C overnight. Each sample was done in triplicate, at a concentration of 5µg/mL. The plate was inverted to remove any unbound glycoprotein and the wells were blocked with 200 µL 3% BSA in PBS for two hours at 25°C. The plate was inverted to remove the solution and the plate was washed with TBST four times. A 50 µL aliquot of lectin in TBST was added at a concentration of 5µg/mL and incubated at 25°C for one hour. This was removed by inverting the plate and washed four times with TBST. 50 µL of 1:10,000 murine anti-histidine-HRP or anti-biotin antibody-HRP was then added to the wells. Antibody was diluted fresh each time with TBST and the plate was incubated for an hour at 25°C. Unbound antibody was removed by inversion and the plate was washed four times with TBST and once with

PBS. 100 µL of TMB substrate was added to each well and the reaction was left for 10 minutes. The reaction was stopped by the addition of 50 µL of 10% H_2OS_4. The absorbance could then be read on a plate reader at 450nm.



2.3 The ELLA assay methodology. (a) Lay down lectin or carbohydrate binding protein (e.g. tCry1Ac), (b) block with BSA, (c) incubate with neo-glycoprotein (GalNAc-BSA), (d) incubate with antibody, (e) detect antibody and colour change at Abs 450nm.

Chapter 3

Bioinformatic analysis and structural modelling of Cry1A Toxins; Cloning of the Cry1Ac derivatives tCry1Ac and CryD3

3.1 Overview

This chapter describes the bioinformatic analysis and structural modelling carried out on the Cry1Ac toxin from *Bacillus thuringiensis* kursatki HD-73, for the design and subsequent cloning of two recombinant genes, encoding tCry1Ac, a truncated Cry1Ac and CryD3, the third domain of Cry1Ac. It was hypothesised that soluble active proteins would be expressed in *E. coli* for both tCry1Ac and CryD3. The gene encoding tCry1Ac was designed from the gene encoding the activated Cry1Ac toxin that binds to the insect cell membrane (Bravo et al. 2007). The expressed protein if active should bind to N-acetylgalactosamine (GalNAc). Bioinformatic and structural analysis of Cry1Ac domain III, revealed that this domain alone may yield an active protein that would potentially bind to galactose. The gene encoding CryD3 would result in a protein that encodes only the third domain of Cry1Ac. A recombinant expression system developed in *E. coli*, using the plasmid pQE60 which encodes a C- terminal 6 x histidine tag, was used for protein purification.

3.2 Analysis of the Cry1Ac protein sequence

A blast search was carried out on the protein sequence of Cry1Ac to identify homologues of the three domains and in particular to compare domain III of Cry1Ac with carbohydrate binding proteins (Figure 3.1). The sequence identity was highly conserved between Cry1Ac and related Cry toxins. The blast search revealed the approximate location of the three domains of Cry1Ac, domain I (aa 10-250), domain II (aa 250-460) and domain III (aa 460-640). The similarity of domain III to a lectin superfamily Carbohydrate binding module family 6 (CBM6) was also observed (Figure 3.1)

CBM6, is also known as cellulose binding domain family VI (CBD VI), and has related CBMs, CBM35 and CBM36. These carbohydrate binding families are non-catalytic carbohydrate binding domains found in a range of enzymes that display activities against a diverse range of carbohydrate targets, including mannan, xylan, beta-glucans, cellulose, agarose, and arabinans (http://blast.ncbi.nlm.nih.gov). The function of the domain is to facilitate the strong binding of the catalytic modules to their dedicated, insoluble substrates. Many of these CBMs are associated with glycoside hydrolase (GH) domains. CBM6 is an unusual CBM as it represents a chimera of two distinct binding sites with different modes of binding: binding site I within the loop regions and binding site II on the concave face of the beta-sandwich fold (Varki et al. 2009). CBM36 comprises a group of calcium-dependent xylan binding domains. Members of CBM35 display conserved specificity through extensive sequence similarity, but divergent function through their appended catalytic modules. The alignment model in figure 3.2 also contains the C-terminal domains of bacterial insecticidal toxins, as they may be involved in determining insect specificity through carbohydrate binding functionality (NCBI blast search).

A diagram of the full Cry1Ac, the activated Cry1Ac (tCry1Ac) and domain III of the activated toxin (CryD3) is shown in Figure 3.3. It highlights the domains and the regions that are cleaved prior to activation. Domain III represents the binding region of Cry1Ac.

A multiple sequence alignment was performed on Cry1Aa, Cry1Ab and Cry1Ac using ClustalW and Genedoc. Figure 3.4 highlights the few areas where there was

sequence divergence. Overall the sequences were highly conserved in the N and C termini, non conserved blocks were mostly confined to domains II and III (Figure 3.4). From amino acids 470 to approx 620 there is a lot of variation between the sequence of Cry1Ac and Cry1Aa. However the main differences are found in the lengths, orientations, and sequences of the loops (Pigott and Ellar 2007). The importance of these differences is evident with Cry1Aa and Cry1Ac, where a loop extension in Cry1Ac creates a unique GalNAc binding pocket involved in receptor binding (Pigott and Ellar 2007).

This sequence divergence in domain III accounts for the variety in receptors for the Cry toxins. A maximum likelihood method was employed to detect evidence of adaptive evolution in Cry proteins by Wu et al. (2007). They identified positively selected residues, which are all located in Domain II or III, revealing that the Cry toxins are adapting and evolving in these regions (Figure 3.5). This is possibly the reason for multiple receptors, variation in targeted species of insects among toxins and the recent increase in Cry toxin susceptibility among nematodes and mosquitoes. Therefore the sites under positive selection could prove interesting targets for structural and functional analysis in relation to receptor binding, in particular carbohydrate recognition moieties. Multiple sequence alignments were preformed on Cry1Ac and other related toxins which revealed similar results. The rooted phylogenetic tree, of the sequence alignment reveals that Cry1Ac diverged earlier from Cry1Aa, Cry1Ab and also Cry1Ad (Figure 3.6).

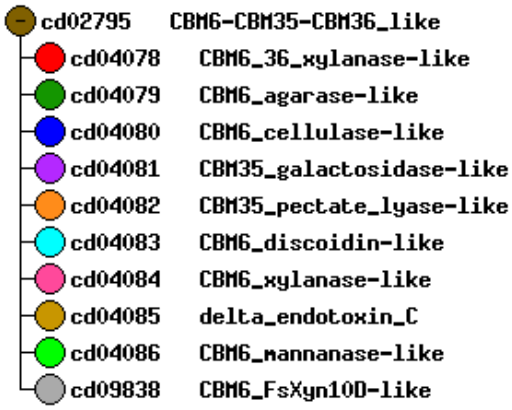
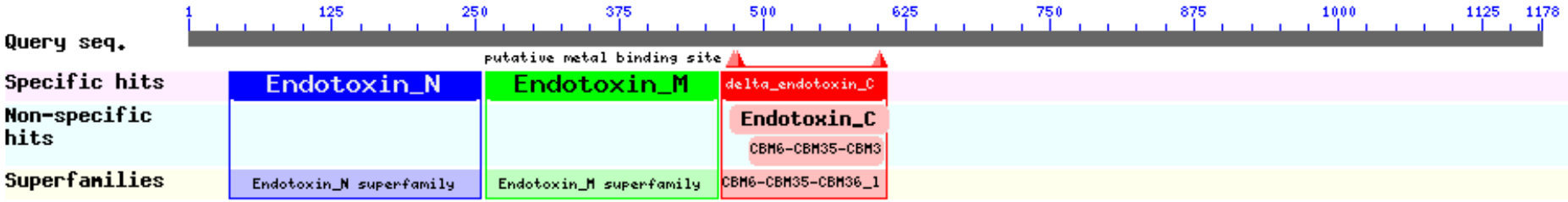


Figure 3.1: Hierarchy of the functions of related proteins to domain III from the blast search result. The hierarchy shows all the families related to the CBM6 family. This is a tree that starts with the highest related families down to the lowest similarities. Domain III shows similarity to the CBM6-CBM35-CBM36 superfamilies, the members of which have related structures and functions to xylanase, agarose, cellulose galactosidase, and delta endotoxin proteins (Marchler-Bauer et al. 2013, NCBI blast search).

A



B

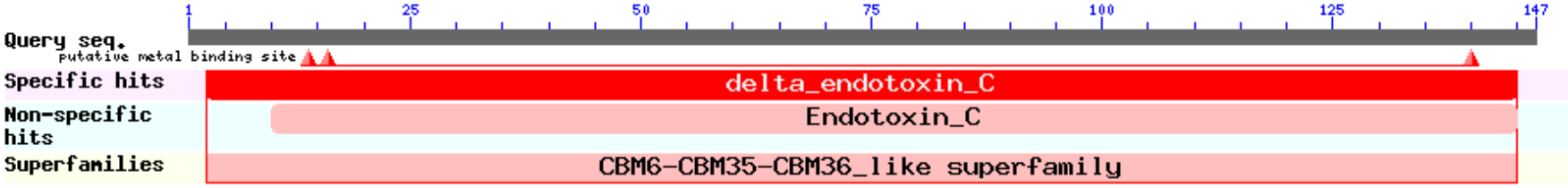


Figure 3.2: Blast search results of Cry1Ac. (a) The blue box from ~10-250, represents domain I, the green ~250-460, domain II, and the red, ~460-620, domain III. (b) Domain III has a specific hit with endotoxin C, while also showing similarities to a lectin superfamily CBM6-CBM35-CBM6.

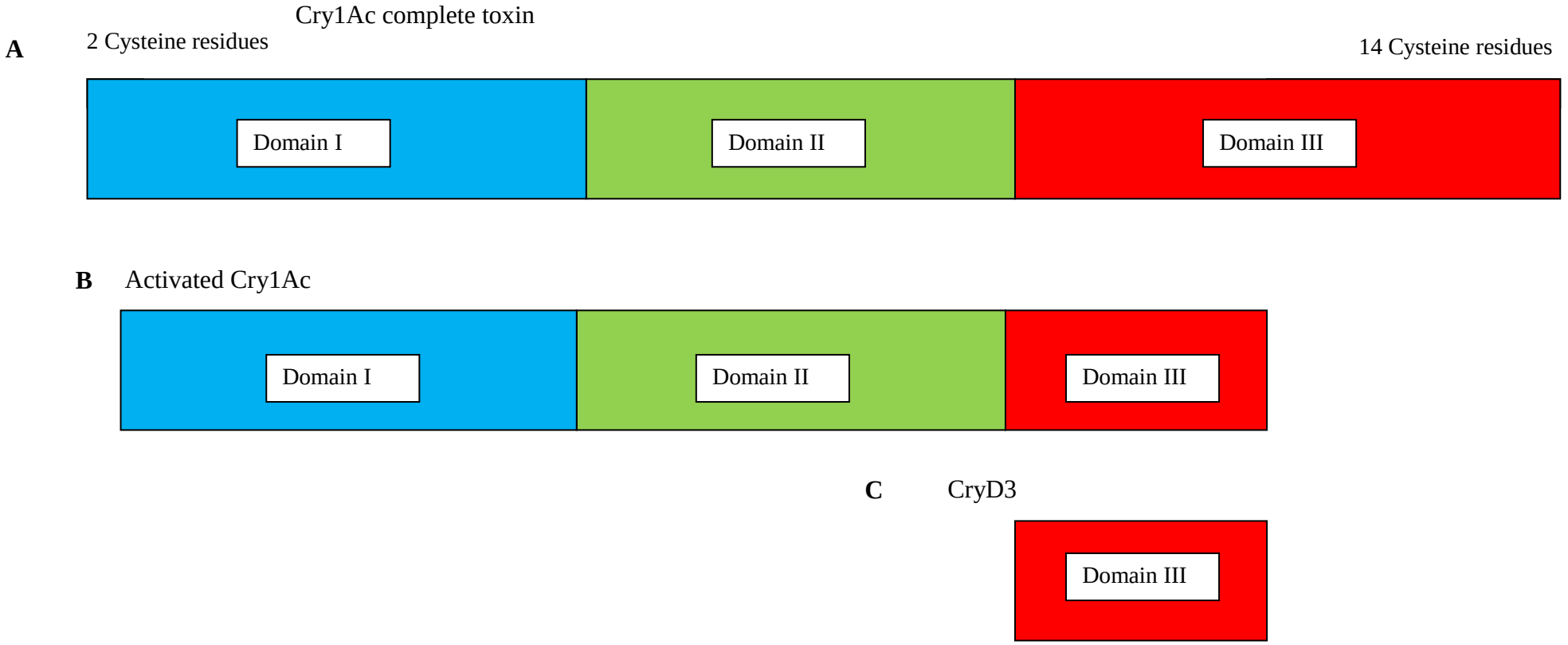


Figure 3.3: Diagram of (A) Cry1Ac complete toxin, (B) Activated Cry1Ac toxin (tCry1Ac) and (C) Domain III (CryD3). (A) Cry1Ac complete toxin contains domain I, domain II and domain III. This protein is not activated and is protected by the cysteine residues in domain III and some in domain I, these parts of the protein are highlighted yellow. (B) The activated toxin (tCry1Ac) which has been cleaved to yield, domain I, domain II and partial domain III. There are no cysteine residues in the protein and it can now induce binding and toxicity in insects. (C) Only the activated part of domain III makes up CryD3. It contains the binding site but cannot induce toxicity in insects.

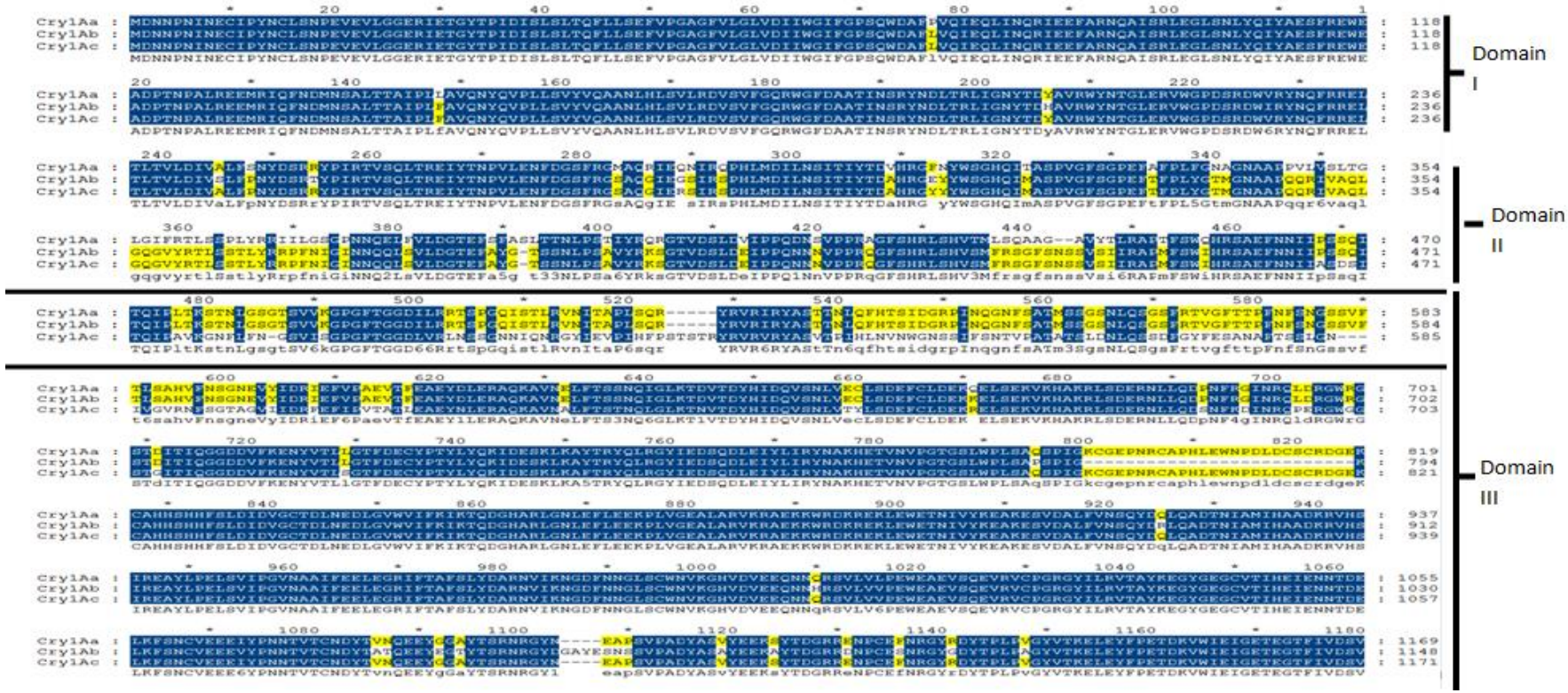


Figure 3.4: Multiple sequence alignment of Cry homologous proteins. The alignment shows the large sequence homology of Cry1Aa, Cry1Ab and Cry1Ac. The sequence is almost identical for each domain I which spans from approx 1-250 amino acids. Domain II spans from approx 250-460 amino acids and shows a considerable sequence similarity between Cry1Ab and Cry1Ac. The active part of Domain III spans 460-620 and shows the dissimilarity of this region in Cry1Ac compared to Cry1Aa and Cry1Ab. Alignment created using ClustalW and Genedoc 2.7.

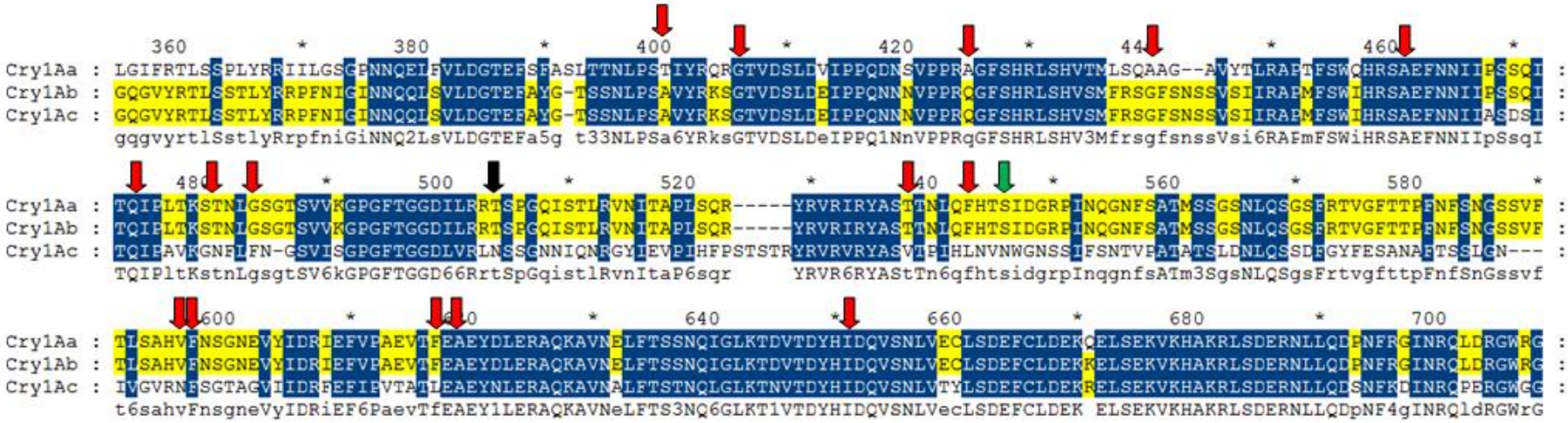


Figure 3.5: Multiple sequence alignment of homologous Cry proteins, focusing on specific residues in domain III. Domain III in each toxin has considerable sequence similarity; however there are some differences especially between the Cry1Ac sequence and the other two sequences. The black arrow represents the amino acid known to be involved in GalNAc binding (Xiang et al. 2009). The green arrow points out the amino acid that is involved in GalNAc binding and is under positive selection for adaptive evolution (Wu et al. 2007). The red arrows highlight the residues that are under positive selection for adaptive evolution, as observed by Wu et al. (2007). Alignment created using ClustalW and Genedoc 2.7.

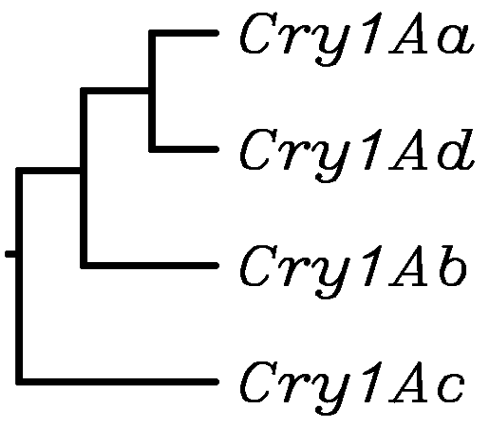


Figure 3.6: Rooted phylogenetic tree from the results of the sequence alignment of Cry1Aa, Cry1Ab, Cry1Ac and Cry1Ad. Cry1Ac diverged early from Cry1Aa, Cry1Ab and Cry1Ad, even though these toxins all form a group distinguishable from other Cry toxins. Alignment and phylogenetic tree created using ClustalW.

3.3 Designing the recombinant proteins tCry1Ac and CryD3

The design of tCry1Ac was based on the activated form of Cry1Ac, as described in Figure 3.3, which binds to the insect gut intestine (Figure 3.8). This design eliminated all the cysteine residues in Cry1Ac shown in Figure 3.7, which cause an insoluble protein. It is predicted that tCry1Ac would bind to GalNAc. CryD3 was designed around domain III of tCry1Ac. (Figure 3.9) It was proposed that if CryD3 would be produced as a soluble protein with a similar structure to that of domain III in Cry1Ac, it may also have carbohydrate binding abilities. CryD3 therefore would not have the toxin domain but it still may bind carbohydrates. Previously, Cry proteins would have been harvested as inclusion bodies and extracted using methods such as cell-free extract for excreted toxins and autolysis, sonication, French press, freeze/thaw for *Bacillus* expressed proteins. The inclusion bodies were then purified by density gradient or recrystallization and the crystals were solubilised using pH and activated by proteolytic enzymes. The toxins were further purified using methods such as precipitation and dialysis. Making the proteins soluble and more amenable to expression in *E. coli* allows them to be readily produced for use as carbohydrate binding lectins for various applications.

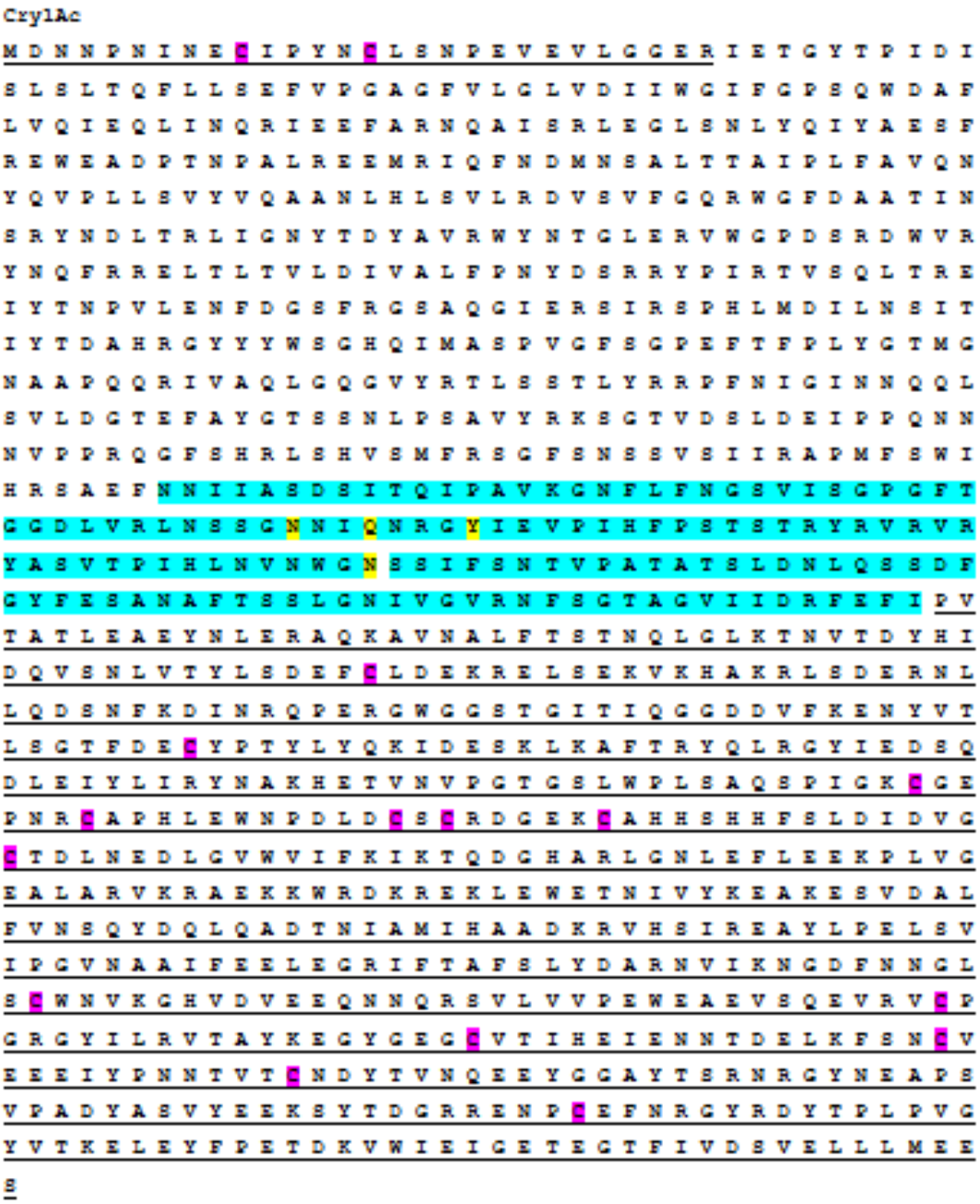


Figure 3.7: Complete protein sequence of Cry1Ac. The amino acids underlined were not cloned. tCry1Ac, the region that is not underlined, represents the active toxin. Cystine residues are highlighted in pink, domain III is highlighted in blue, and the residues thought to play a role in GalNAc binding are highlighted in yellow.

tCry1Ac

MIIETGYTPIDISLSLTKFLLSEFVPGAGFVLGLVDIIWGIFGPSQWDAFLVQIEQLINQ
RIEEFARNQAISRLEGLSNLYQIYAESFREWEADPTNPALREEMRIQFNDMNSALTTAIP
LFAVQNYQVPLLSVYVQAANLHLSVLRDVSVCGQRWGFDAATINSRYNDLTRLIGNYTDY
AVRWYNTGLERVWGPDSRDWVRYNQFRRELTLTVLDIVALFPNYDSRRYPIRTVSQLTRE
IYTNPVLENFDGSFRGSAQGIERSIRSPHLMDILNSITIYTDAHRGYYYWSGHQIMASPV
GFSGPEFTFPLYGTMGNAAPQQRIVAQLGQGVYRTLSSTLYRRPFNIGINNQQLSVLDGT
EFAYGTSSNLPSAVYRKGETVDSLDEIRPQNNNVPPRQGFSHRLSHVSMFRSGFSNSSVS
IIRAPMFSWIHRSAEFNNIIASDSITQIPAVKGNFLFNGSVISGPGFTGGDLVRLNSSGN
NIQNRGYIEVPIHFPSTSTRYRVRVRYASVTPIHLNVNWGNSSIFSNTVPATATSLDNLQ
SSDFGYFESANAFTSSLGNIVGVRNFSGTAGVIIDRFEFIPVRSHHHHHH

Figure 3.8: Protein sequence of tCry1Ac, a truncated form of Cry1Ac. This protein is predicted to express as a soluble active protein, similar to that which is activated in the intestine of insects. A 6x histidine tag was added to the C terminus of the recombinant protein.

CryD3

MNNIIASDSITQIPAVKGNFLFNGSVISGPGFTGGDLVRLNSSGNNI
QNRGYIEVPIHFPSTSTRYRVRVRYASVTPIHLNVNWGNSSIFSNTV
PATATSLDNLQSSDFGYFESANAFTSSLGNIVGVRNFSGTAGVIIDR
FEFIRSHHHHHH

Figure 3.9: Protein sequence of CryD3, the third domain of tCry1Ac. CryD3 is 153 amino acids long, and encoding a 6x histidine tag at the C terminus of the recombinant protein. This protein would not have the toxin capabilities of Cry1Ac but it may have kept the ability to bind GalNAc.

3.4 Structure predictions of tCry1Ac and CryD3 using I-TASSER

The protein sequences of tCry1Ac and CryD3 were submitted to the online service I-Tasser. The I-Tasser server is an internet service for protein structure and function predictions. It automatically generates high-quality predictions of 3D structure and biological functions of protein molecules from their amino acid sequences (Roy, Alper, and Yang 2010; Roy, Jianyi, and Yang 2012; Zang, 2008). Figure 3.10 shows how the server generates structure and function predictions by first retrieving templates of proteins that have similar secondary structure from the Protein Data Bank (PDB) library by a threading approach. In the second step, the continuous fragments excised from the PDB templates are reassembled into full-length models by replica-exchange Monte Carlo simulations with the threading of unaligned regions built by *ab initio* modeling. In the third step, SPICKER (a clustering algorithm to identify the near-native models from a pool of protein structure decoys) assembles the fragments again to refine the clusters and remove steric clash. The decoys generated in the second simulations are then clustered and the lowest energy structures are selected. The final full-atomic models are obtained by REMO which builds the atomic details from the selected I-TASSER decoys through the optimization of the hydrogen-bonding network (Figure 3.10) (Roy, Alper, and Yang 2010; Roy, Jianyi, and Yang 2012; Zang, 2008). I-Tasser then send the predicted protein structures, models and protein data files to the user.

Figure 3.11 shows three different views of the predicted structure of tCry1Ac by I-Tasser. These models show a very similar structure to other Cry toxins, (De maagd, et al. 2001) which would predict that tCry1Ac would fold into an active protein. The predicted structure of CryD3 is shown in Figure 3.12, while also highlighting the predicted binding site in the protein. This predicted binding site was in a different position than the expected binding site in domain III of Cry1Ac. I-Tasser predicted residues P14, A15 and V16 to be involved in galactose binding. It also showed how CryD3 would fit into the Cry toxin protein model. This structure for CryD3 is very similar to that of domain III in tCry1Ac however it does have some small differences and this could affect the binding site. Therefore further analysis on these protein models and predicted binding sites was done using PyMol.

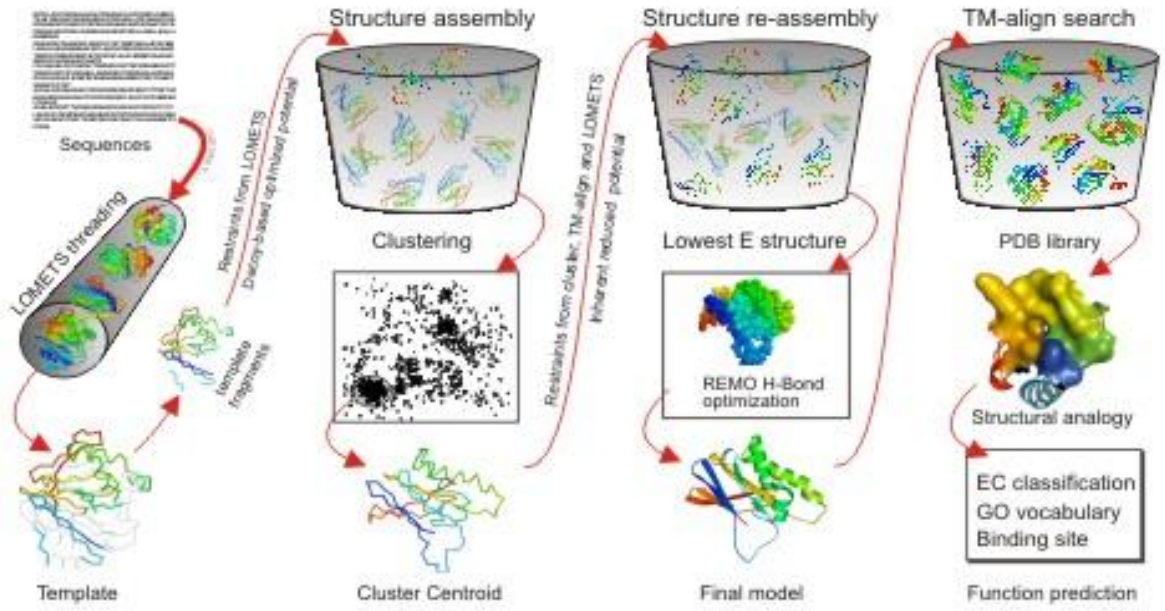


Figure 3.10: I-Tasser protocol for protein structure and function predictions (Roy, Alper, and Yang 2010).

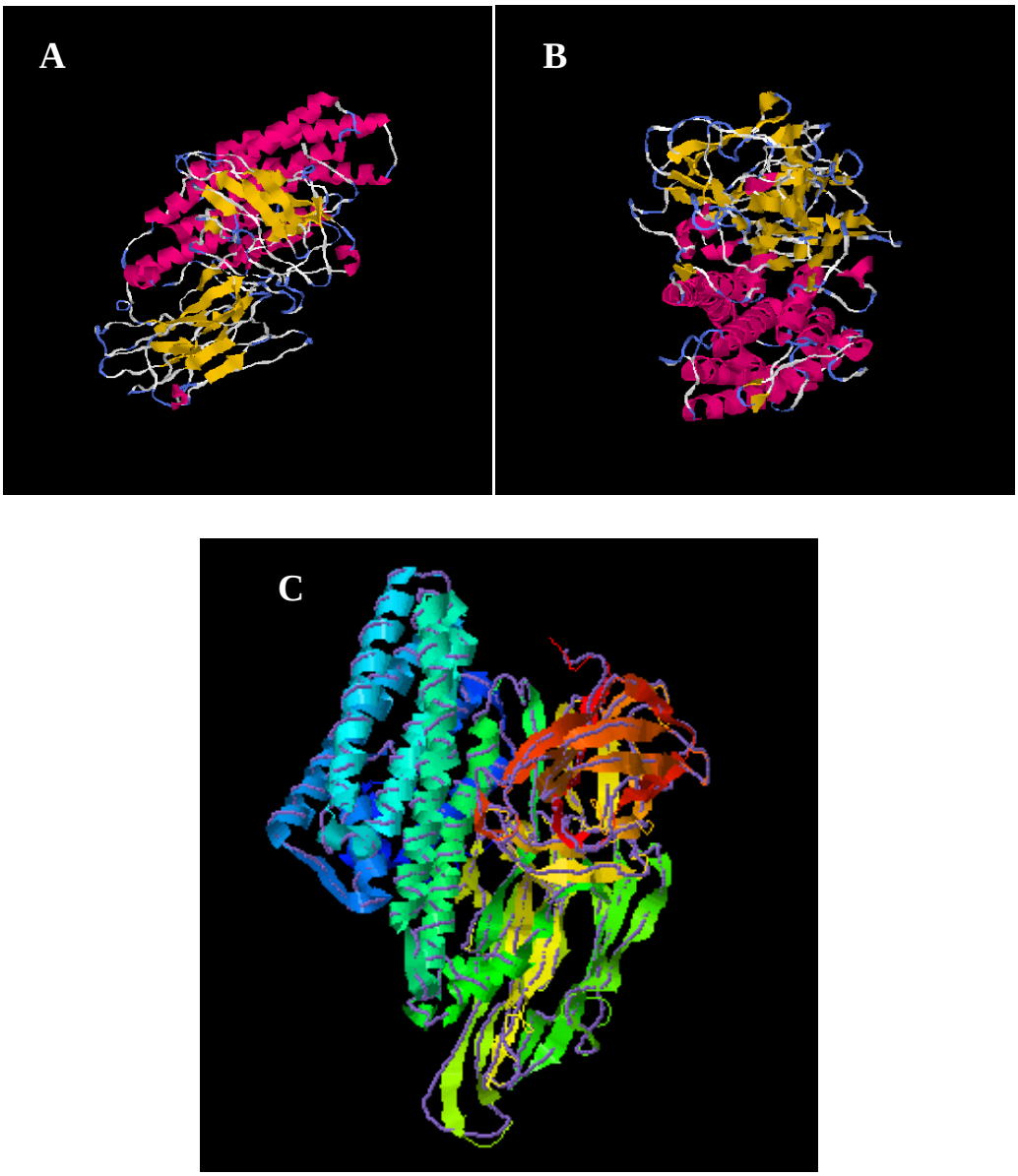


Figure 3.11: Predicted structure of tCry1Ac from I-Tasser. A and B show the helical domain I in pink and domains II and III in yellow. C) Domain I is in blue (beta barrels), domain II is shown in green while domain III is shown in red. Models from I-Tasser.

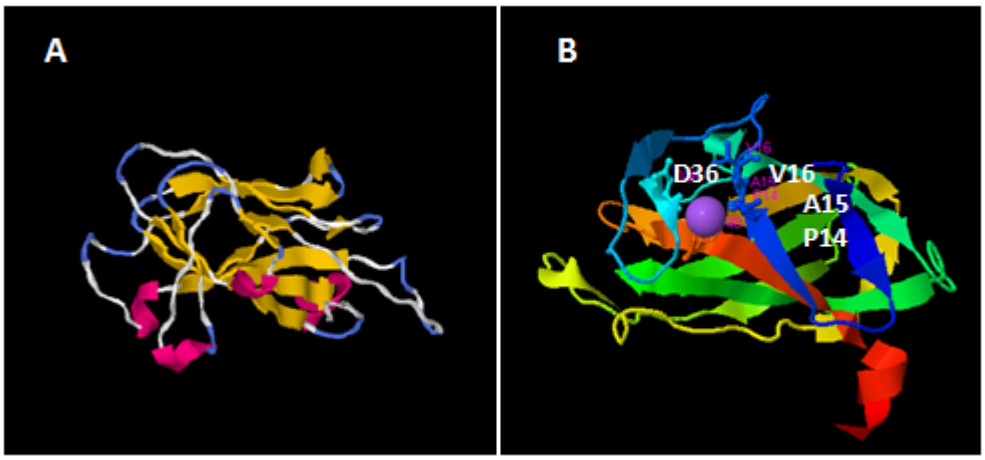


Figure 3.12: Predicted structure models of CryD3 by I-Tasser. A) Predicted structure of CryD3, helical formation in pink, beta sheets in yellow. B) Predicted structure of CryD3 with highlighted residues for the galactose binding site, 6x histidine tag and C terminus shown in red, N terminus shown in blue. Models from I-Tasser.

Figure 3.13: Predicted structure of CryD3 fitting into the Cry toxin model. This predicts that CryD3 will have a similar structure and, possibly, function to that of the third domain in Cry1Ac.

3.5 Modelling of tCry1Ac and CryD3 structures using PyMOL

After tCry1Ac and CryD3 were submitted to I-Tasser a protein data file was generated for the two proteins. This protein data file was then inputted to PyMOL. PyMOL is a **user-sponsored** molecular visualization system on an **open-source** foundation, which allows the user to download the software to visualise, render and animate 3D molecular structures. Protein models for both tCry1Ac and CryD3 were made (Figures 3.14 and 3.15). The binding site of GalNAc in domain III was then investigated. Previously, Burton et al. (1999) and Xiang et al. (2009) have shown residues N506, Q509, Y513 and N546 to be involved in GalNAc binding. PyMOL allowed regions of the protein to be highlighted and examined in more detail for both tCry1Ac and CryD3 (Figure 3.16). Examination of the region in CryD3 that I-Tasser predicted as the carbohydrate binding site was also carried out (Figure 3.15).

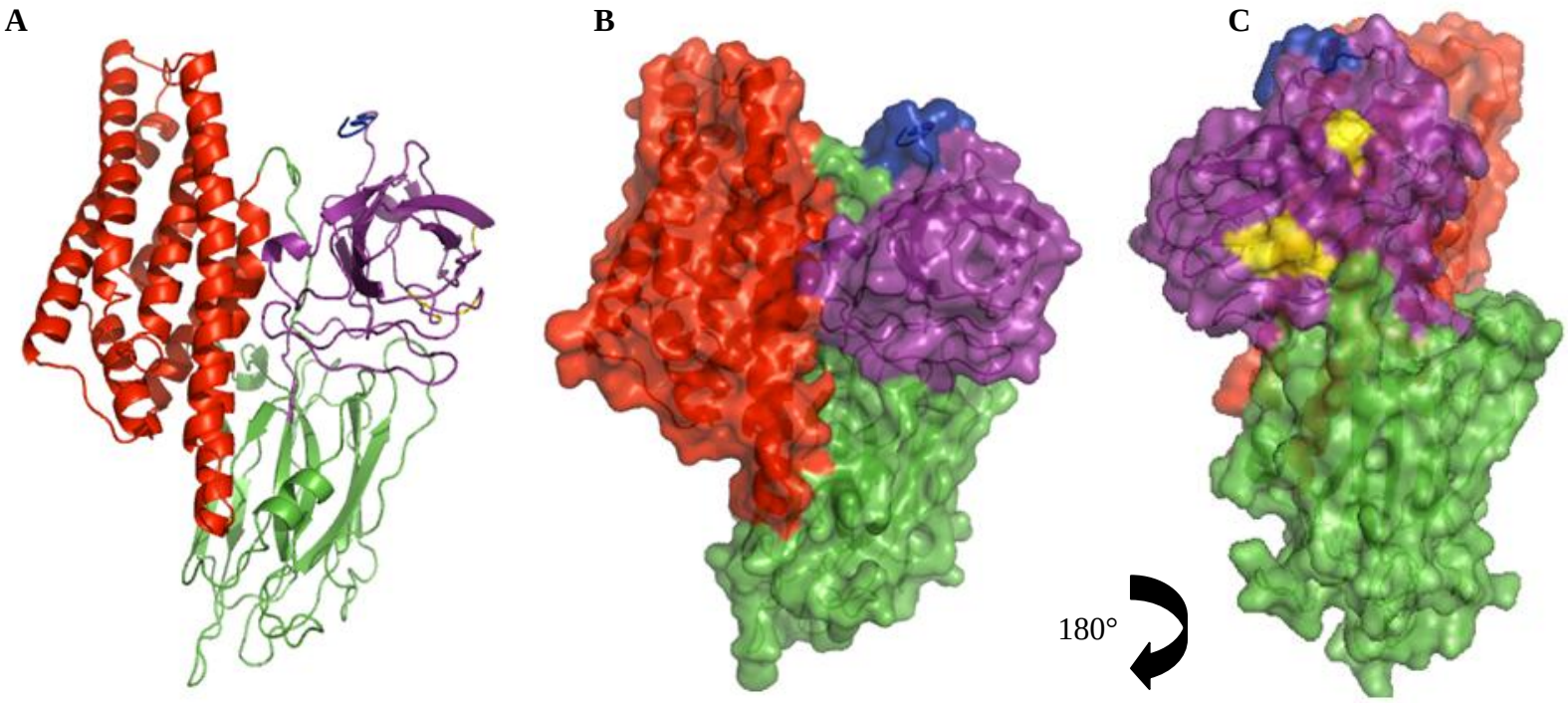


Figure 3.14: Predicted protein structure of tCry1Ac, edited using PyMOL. For all structures domain I is coloured red, domain II, green and domain III, purple. The 6x histidine tag is blue and the predicted binding site is yellow. **A**) A ribbon cartoon of tCry1Ac showing the three domains, the α-helical structure of domain I is shown in red, while the β-sheets of domains II and III are shown in green and purple respectively. This structure is very similar to other Cry1A toxin structures. **B**) and **C**) show the cartoon surface of tCry1Ac as it is easier to identify potential binding sites. C) 180° turn to the left revealing the residues N480, Q583, Y587 and N521 in yellow. These residues are examined in more detail in Figure 3.15.

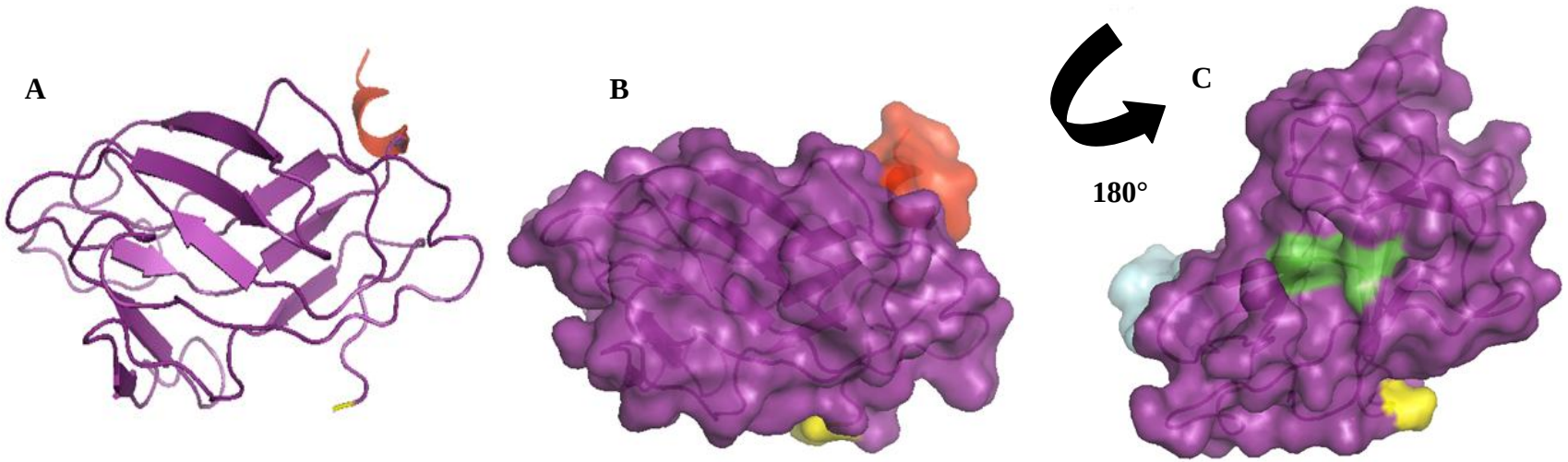


Figure 3.15: Predicted protein structure of CryD3, edited using PyMOL. A) Cartoon ribbon structure of CryD3, N terminus highlighted yellow and the 6x histidine in red. **B)** Cartoon ribbon of CryD3 showing the surface, N terminus highlighted yellow and the 6x histidine in red. **C**) CryD3 turned 180° and angled towards the predicted carbohydrate binding site from I-Tasser, which is coloured green. The 6x histidine tag is blue and the N terminus is yellow.

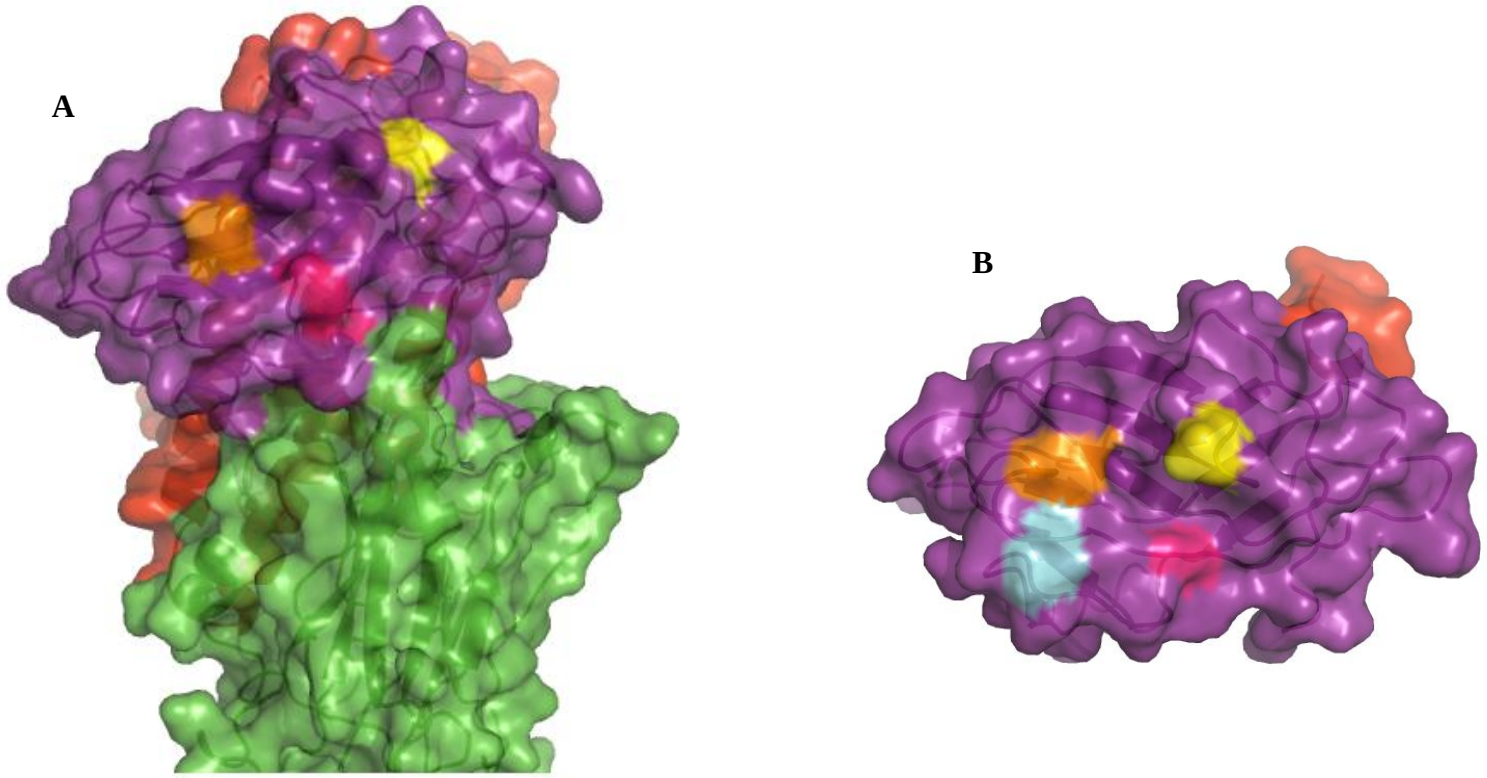


Figure 3.16: Models of tCry1Ac and CryD3 showing expected binding site of GalNAc. A) tCry1Ac, domain I in red, domain II in green and domain III in purple with highlighted residues believed to be involved in GalNAc binding, N480- orange, Q483-cyan (buried in this structure), Y487 – pink, N521- yellow. **B**) CryD3 in purple with highlighted residues that relate to tCry1Ac, N45 – orange, Q48 – cyan, Y52 – pink, N86 – yellow. The 6x histidine tag in both proteins is red. These protein structures are from the same sequence however some structural differences are predicted. The residues that are highlighted are located in different positions in the different proteins. In tCry1Ac, the domains I and II may be supporting domain III and cause its structure to change slightly.

3.6 Cloning of *cry1Ac* and *cryd3* from *Bacillus thuringiensis*

The most effective systems in terms of simplicity and reproducibility for protein expression are generally considered to be prokaryotic systems. Over expression of specific proteins in the natural prokaryotic host cell however can be difficult to achieve. *Escherichia coli* a gram negative bacterium is the most widely used host for expression of recombinant proteins. This is mainly due to its simplicity, safety, and rapid high density growth rate on inexpensive substances, allowing for large scale protein expression.

There are of course a number of different *E. coli* strains and also a broad range of cloning and expression vectors, which can all help to increase the expression or amount of soluble protein. Depending on the expression system, fusion tags can be incorporated onto the protein, to facilitate downstream purification and analysis.

An *E.coli* expression system will be used for the expression of tCry1Ac and CryD3. The Qiagen vector pQE60 (Figure 3.17) was used as the main expression vector. This vector has been designed to produce high amounts of recombinant protein. This is due to an optimised promoter, the phage T5 transcriptional promoter, a lac operator sequence, extensive multiple cloning sites, and a tag site designed to introduce a 6x histidine at the C terminus of a protein expressed from an insertion in the multiple cloning sites. The pQE60 vector confers on the host ampicillin resistance as it expresses the β-lactamase gene.

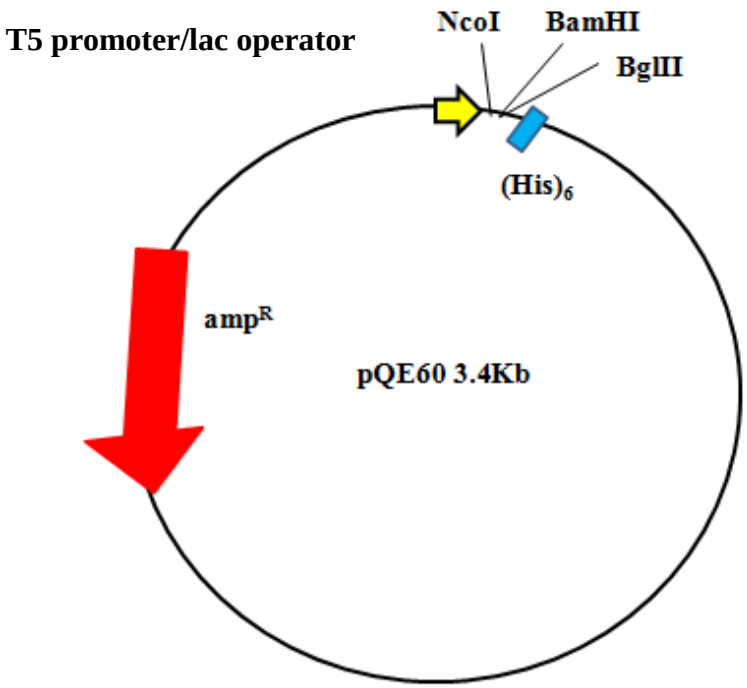


Figure 3.17: pQE60 vector from Qiagen. pQE60 contains a PT5 promoter for increased expression of the gene of interest, a *lac* operator to allow control of expression with IPTG, a multiple cloning site with restriction sites and codons for a 6xHis tag that will be added to the C terminus of the expressed protein.

The *B. thuringiensis* stain kurstaki HD73 was obtained from USDA, ARS Culture Collection in Peoria, IL. The strain kurstaki HD73 itself is not sequenced, however the sequence for the related organism kurstaki str. T03a001 (Accession no NZ_ACND00000000.1) was used for designing the primers for the DNA region encoding tCry1Ac. Primers Cry1Ac P60Df and Cry1Ac P60Dr were designed to amplify a truncated region of *cry1Ac* which would be cloned into the pQE60 vector. The primers used for the amplification of the truncated region of the gene contained BspHI and BglII restriction sites (see section 2.1). Genomic DNA was extracted from *B. thuringiensis* strain kurstaki HD73 (see section 2.5.1) and PCR was carried out using the primers Cry1Ac P60Df and Cry1Ac P60Dr. The PCR product was analysed by agarose gel electrophoresis and the DNA band corresponding to the expected size was present (Figure 3.20). The PCR products were then restricted

using BspHI and BglII, gel extracted and ligated to the NcoI-BglII digested pQE60 vector (Figure 3.18). Following transformation of *E .coli* JM109 cells, plasmid DNA was isolated and screened for inserts using gel electrophoresis and restriction analysis. The presence of a correct insert was confirmed by DNA sequencing.

This plasmid was named ptCry1Ac and was then used to amplify the DNA region encoding the CryD3 domain. The plasmid ptCry1Ac DNA was isolated using a High Yield Plasmid Mini Prep Kit (see section 2.5.2.2). The DNA was then used for PCR along with the primers Cryd3f and Cryd3r (see section 2.1), which incorporated restriction sites for NcoI and BglII respectively. The PCR product was analysed and a DpnI restriction digest was carried out to remove any double stranded plasmid DNA (see section 2.7.2). The PCR product was then restricted using NcoI and BglII and ligated to pQE60 (Figure 3.19). *E. coli* JM109 was transformed with ligated DNA (see section 2.6.1). Plasmid DNA was isolated and screened for inserts using gel electrophoresis and restriction analysis. The presence of a correct insert was confirmed by DNA sequencing.

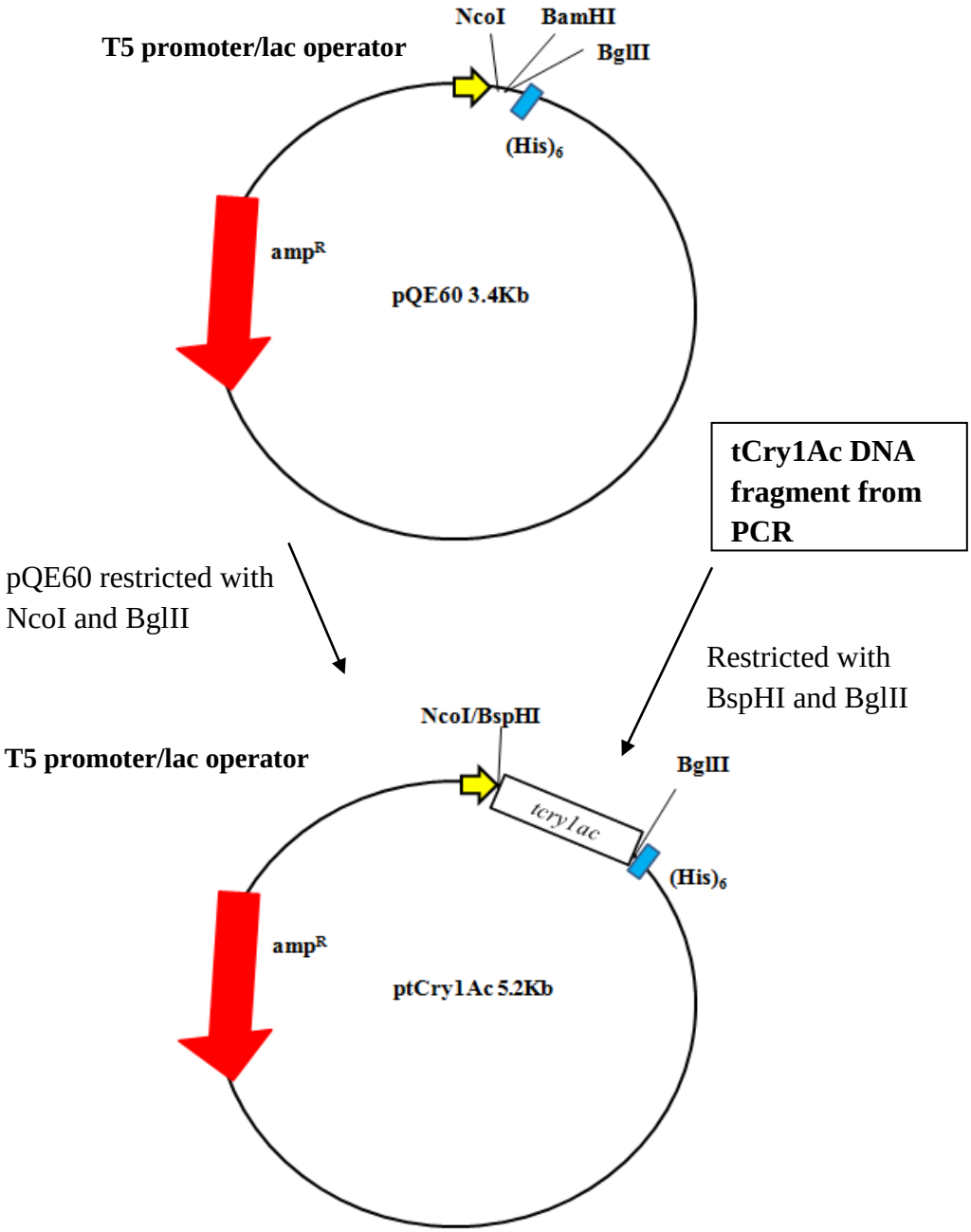


Figure 3.18: Cloning strategy for the region encoding tCry1Ac in the expression vector pQE60. The PCR product was restricted using the enzymes BspHI and BglII. The NcoI site in the PCR product prohibited the use of this enzyme to cut the product for cloning. Restriction with BspHI results in compatible ends for ligation with NcoI. The pQE60 vector was restricted using NcoI and BglII. The products were then ligated to form ptCry1Ac.

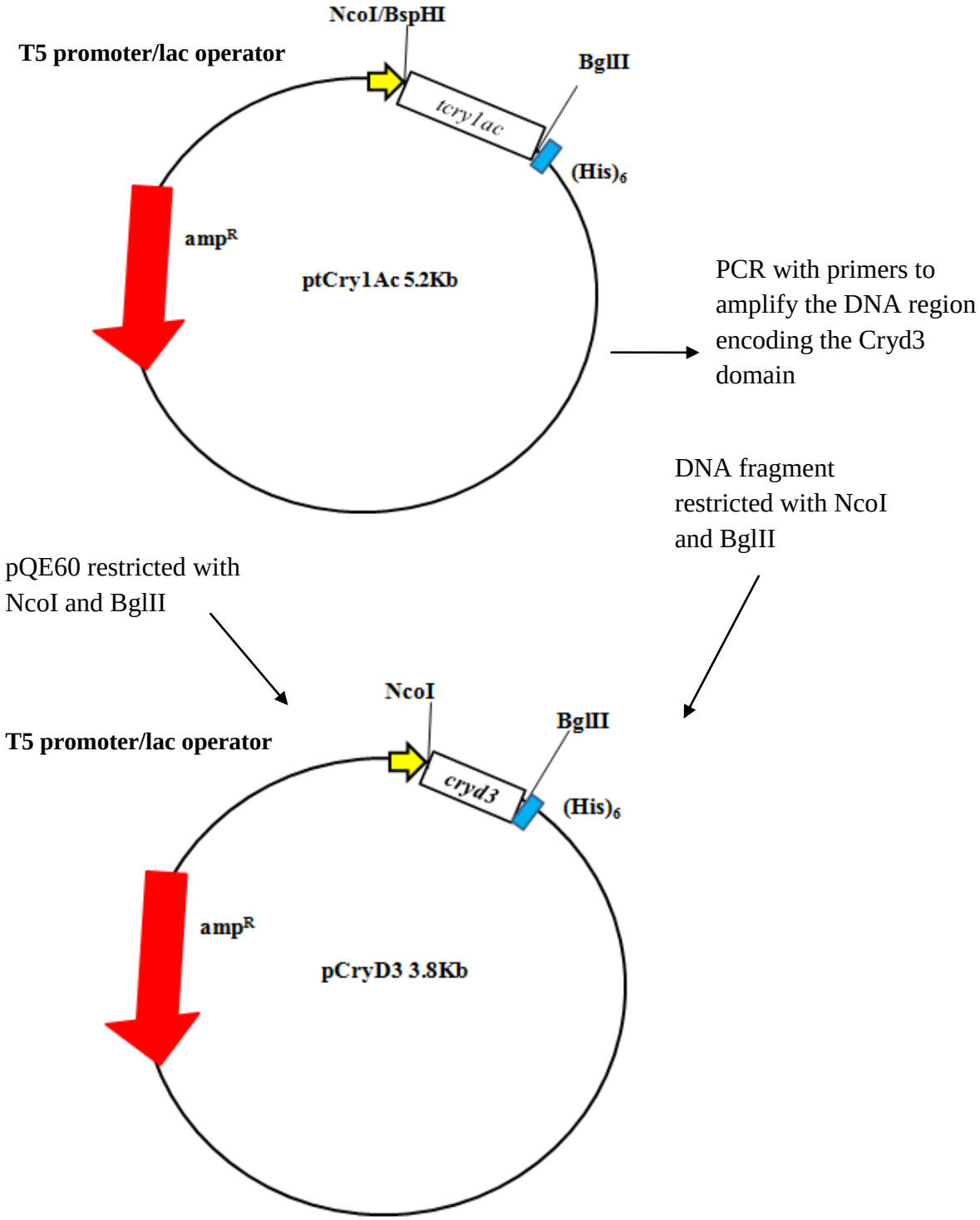


Figure 3.19: Cloning strategy for the DNA region encoding the CryD3 domain in the pQE60 vector. PCR on ptCry1Ac DNA using primers for the region encoding the target domain was carried out. The product was cleaned and digested using DpnI to eliminate any double stranded DNA in the reaction. The PCR fragment and pQE60 were both digested using NcoI and BglII. They were then ligated to form pCryD3.

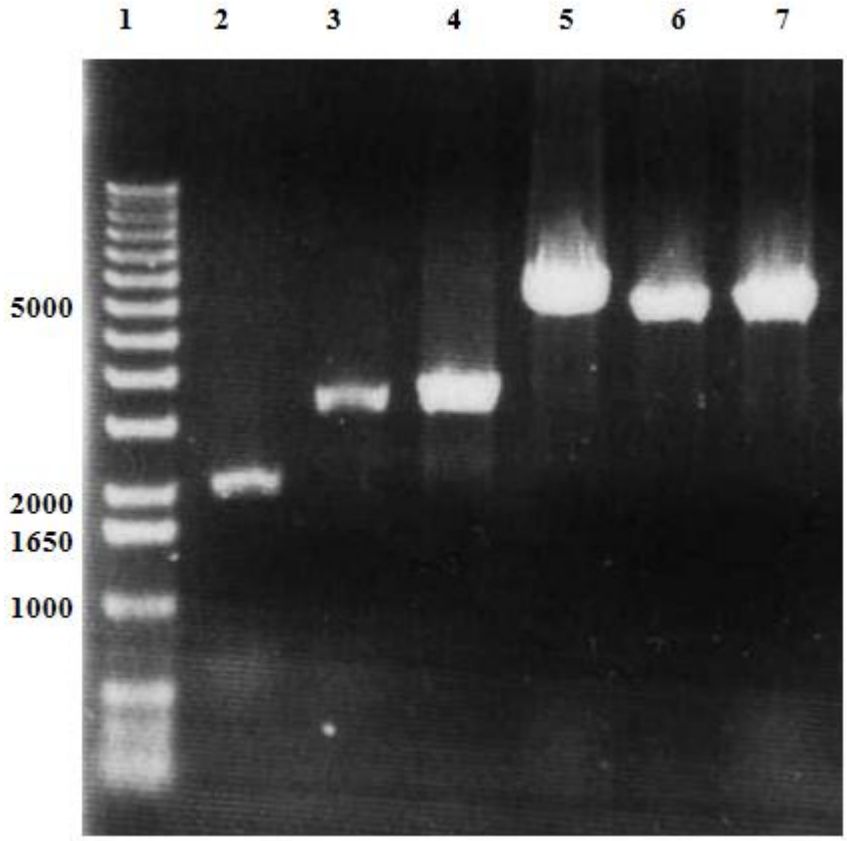


Figure 3.20: **1% DNA agrose gel electrophoresis on plasmids pQE60, pCryD3 and ptCry1Ac.** Lane **1;** 1Kb DNA Plus ladder, **2;** pQE60, **3;** and **4;** pCryD3, **5-7;** ptCry1Ac.

The linear DNA fragments in the ladder are used for confirmation of staining and for comparative analysis of migration. They are not used for direct size calculation of the closed circular plasmid molecules.

3.7 Discussion

This chapter has described the bioinformatic analysis of the Cry1Ac protein sequence, the design of two recombinant proteins tCry1Ac and CryD3, the extensive structural modelling of these proteins by I-Tasser and PyMOL and the subsequent cloning of these genes into the pQE60 vector.

The blast search revealed the high sequence homology between Cry1Ac and closely related toxins such as Cry1Aa and Cry1Ab. It also showed where the three domains were located on Cry1Ac and this was important information for the design of tCry1Ac and CryD3. This blast search revealed Domain III of Cry1Ac has similarity to the CBM6-CBM35-CBM36 superfamilies, which is an important indicator that the third domain of Cry1Ac may have carbohydrate binding abilities.

A multiple sequence alignment was carried out on Cry1Ac, Cry1Aa and CryAb to examine the sequences for any differences particularly in the third domain. This sequence alignment showed considerable homology between the sequences especially in the N and C terminus. However, there was a great deal of variation particularly in Cry1Ac domain III, from amino acid 480 to 620. The altered sequence may give Cry1Ac its receptor binding capacity for GalNAc. From the results of the sequence alignment between Cry1Aa, Cry1Ab, and Cry1Ac a rooted phylogenetic tree showed that Cry1Ac diverged early from Cry1Aa, Cry1Ab and Cry1Ad giving an indication as to why Cry1Ac has carbohydrate specificities but Cry1Aa, Cry1Ab and Cry1Ad do not.

The recombinant proteins tCry1Ac and CryD3 were then designed based on the Cry1Ac activated toxin when it has been processed following ingestion by insects (Bravo et al. 2007) and from the highlighted domains on the blast search of Cry1Ac. These proteins were designed so that they would be expressed as soluble active proteins with 6x histidine tags on their C termini. Therefore this would limit the downstream experiments which would be needed for solubilisation or activation of the proteins. This would be the first time any Cry toxins would be cloned in this way and if they can be expressed and purified in E. coli as active proteins it could be used as a model going further for analysing Cry toxins.

Further analyses were carried out on tCry1Ac and CryD3 by using I-TASSER to get structure predictions of both proteins. These structures were very similar to Cry1Ac

and the third domain of Cry1Ac. There were a few slight differences which could be due to the structure predictions of the proteins in I-TASSER or the slight alteration in protein folding. PyMol was used to model the structures of tCry1Ac and CryD3 and examine the predicted binding sites further. By using PyMol the expected binding site of GalNAc was mapped on tCry1Ac and the corresponding sites were highlighted in CryD3. This showed some differences in the binding site and so CryD3 may have a different binding specificity than tCry1Ac, or it could have a higher or lower affinity for GalNAc.

Cloning of tCry1Ac and CryD3 and the construction of ptCry1Ac and pCryD3 was carried out using the pQE60 vector. Positive clones were identified and the plasmid inserts sequenced. These plasmids were then available for the expression and characterisation of the recombinant proteins tCry1Ac and CryD3, described in Chapter 4.

This work has shown conclusively that Cry1Ac domain III has protein sequence homology to lectins in the superfamilies, CBM6-CBM35-CBM36 and backs up previous research which showed the structure of domain III has similarities to lectins (Bravo et al. 2012). The residues of tCry1Ac previously shown to have an effect in GalNAc binding were mapped on tCry1Ac and CryD3 and they show the potential binding site for GalNAc on the protein. This may allow for further mutagenesis studies on this protein to elucidate the binding site for GalNAc on Cry1Ac and whether tCry1Ac and CryD3 have carbohydrate binding specificities also.

Chapter 4

Optimisation of tCry1Ac and CryD3 protein expression and purification

4.1 Overview

This chapter describes the optimisation of the expression and purification of tCry1Ac and CryD3 and the separation of these proteins from a persistent contaminating protein from *E. coli.* tCry1Ac was expressed in low amounts in small scale expression and purification. CryD3 had relatively good expression yields but there were contaminating proteins in the purified samples. To assess tCry1Ac and CryD3 binding specificities for different glycoproteins they were examined by Enzyme Linked Lectin Assay (ELLA). This revealed that tCry1Ac bound to GalNAc while CryD3 had specificity towards galactose. These results correlated with the structural modelling and predictions described in chapter 3. However the relative affinity of CryD3 for galactose was lower than the affinity tCry1Ac had for GalNAc. The expression of tCry1Ac was optimised, by varying the host cell choice (i.e. the strain of *E. coli*), the type of medium and the temperature used during expression. Optimisation of the purification of tCry1Ac and CryD3 was carried out by increasing imidazole concentrations during the elution phase of Immobilized Metal Affinity Chromatography (IMAC). Instability of the purified protein and relative sensitivity to proteases were thought to be a problem and so optimisation of the conditions during purification were also varied and tested. Contaminating proteins continued to be a problem throughout the purification process with the main contaminant being identified by MS peptide mapping and sequencing analysis. This contaminating protein was found to be a cAMP regulatory protein from *E. coli*. Various purification systems were examined to remove the persistent contaminating proteins. These included ion exchange chromatography, gel filtration, IMAC purification with added detergents and the addition of galactose washes. These modifications ultimately succeeded in greatly improving the purification and stability of the proteins. Functional analysis of the pure tCry1Ac and CryD3 proteins were then carried out by use of the ELLA.

4.2 Small scale expression of tCry1Ac and CryD3 proteins

The proteins tCry1Ac and CryD3 were expressed in *E. coli* JM109 from ptCry1Ac and pCryD3 respectively as described in section 2.10. The cell lysates from 100ml cultures were examined by SDS-PAGE. From the protein sequence the predicted protein size of tCry1Ac and CryD3 was determined by ExPASY. tCry1Ac was predicted at 69kDa and CryD3 was less than 20kDa. A protein at approx 69kDa was observed on the SDS PAGE gel for tCry1Ac (see Figure 4.1 b). However, over expression of the protein was not obvious. Therefore, a western blot was carried out to probe for the 6x histidine tag on the C terminus of tCry1Ac (Figure 4.1 a). A protein was observed in low amounts on the western blot in the soluble fraction. Thus optimising the expression of tCry1Ac was needed to obtain a higher yield of recombinant protein for functional analysis. CryD3 expression was also examined on SDS-PAGE gels, as the protein is approx 20kDa it is difficult to view on 15% gels. An increased expression of a 20kDa protein was observed in the overnight soluble samples (see Figure 4.2). The overnight cultures of tCry1Ac and CryD3 were then lysed by cell disruption and the soluble fraction taken for preliminary IMAC purification.

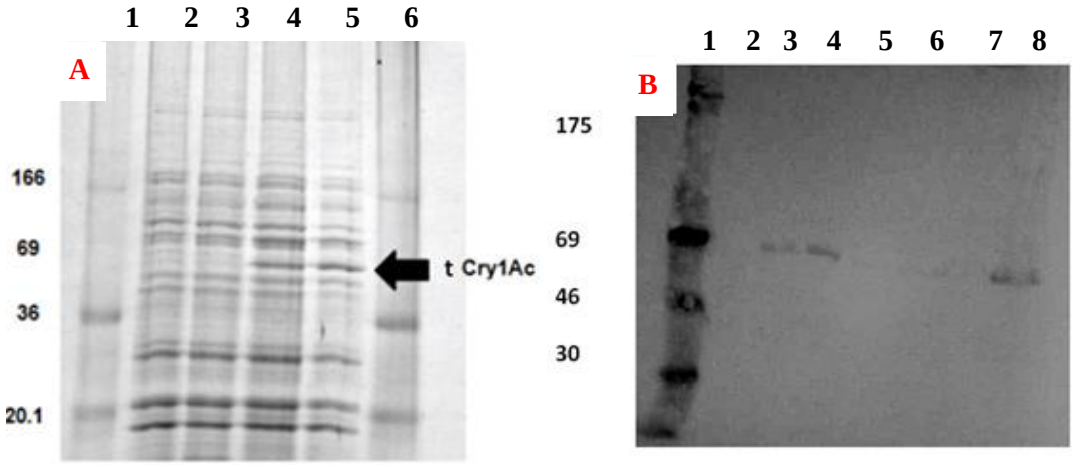


Figure 4.1: SDS PAGE and western blot analysis of small scale expression of tCry1Ac in *E. coli* JM109. (A) 1 and 6; protein ladder, 2; 2 hours post induction, 3; 3 hours post induction, 4; 4 hours post induction, 5; overnight post induction. (B) Western blot, 1; protein marker, 2; pQE60, 3; not induced ptCry1Ac total protein, 4; induced ptCry1Ac total protein, 5; not inducted ptCry1Ac soluble protein, 6; induced ptCry1Ac soluble protein, 7; not induced ptCry1Ac insoluble, 8; induced ptCry1Ac insoluble.

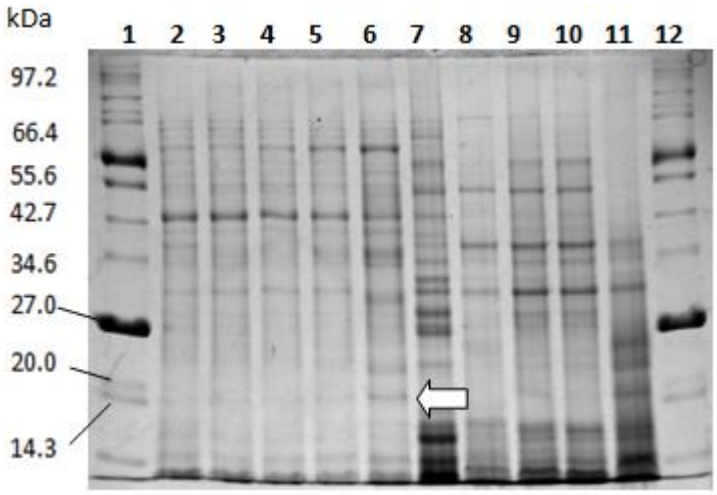


Figure 4.2: SDS PAGE analysis, soluble and insoluble expression of CryD3 in *E. coli* JM109.1 and 12; Broad range protein marker, 2-6; soluble post induction at time 1hr, 2hr, 3hr, 4hr and overnight respectively, 7-11; insoluble post induction at time 1hr, 2hr, 3hr, 4hr and overnight respectively. White arrow represents CryD3 expression.

4.3 Preliminary Purification of tCry1Ac and CryD3 by Immobilised Metal Affinity Chromatography (IMAC)

Affinity chromatography separates proteins on the basis of a reversible interaction between a protein and a specific ligand coupled to a chromatography matrix. The technique is ideal for a capture or intermediate step in a purification protocol and may be used whenever a suitable ligand is available for the protein of interest. This method potentially gives high selectivity, high resolution, and high capacity for the protein of interest. The target protein may then be collected in a purified, concentrated form.

The 6x histidine tag is one of the most common affinity tags used to facilitate the purification and detection of recombinant proteins. Polyhistidine tags such as 4x histidine or 10x histidine are also used. They may provide useful alternatives to 6x histidine for improving purification results. For example, since 10x histidine binds more strongly to the affinity medium, a higher concentration of eluent (imidazole) can be used during the washing step before elution. This may facilitate the removal of contaminants which may otherwise be co-purified with a 6x histidine fusion protein.

Chelating Sepharose, when charged with Ni2+ ions, selectively binds proteins if complex forming amino acid residues, in particular histidine, are exposed on the protein surface. 6x histidine fusion proteins can be easily bound and then eluted with buffers containing imidazole. tCry1Ac and CryD3, each with the histidine tag fusion were purified from the soluble fraction by IMAC on two separate columns. As described in section 2.13.1 the soluble fraction was added to the resin and washed with 40, 60 and 70mM imidazole to elute any contaminating proteins. Both proteins were eluted using 250mM imidazole. Samples were taken from each fraction to be analysed on SDS PAGE. Figure 4.3 presents the results obtained from the preliminary analysis of IMAC purification, revealing tCry1Ac at 69kDa in the elution fractions. From the SDS PAGE gel it is clear that tCry1Ac does not express very well and further optimisation of expression is needed. Figure 4.4 shows CryD3 as a large band at approximately 20kDa. There are also some contaminating bands on the gel; therefore optimisation of purification was needed for CryD3.

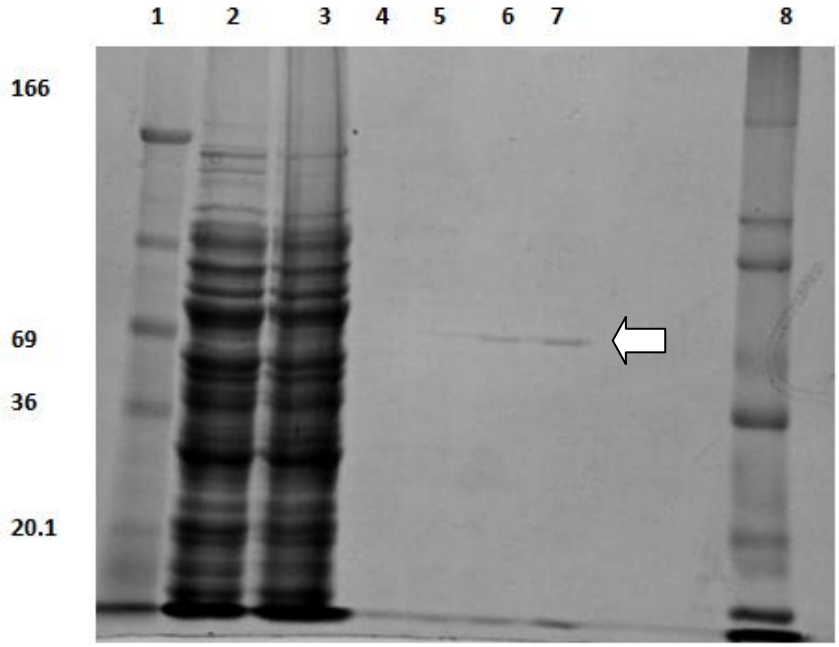


Figure 4.3: SDS PAGE analysis of the preliminary purification of tCry1Ac by IMAC. 1;SigmaMarker Wide Range , 2; lysate, 3; flow through, 4 and 5; 60mM and 70mM imidazole washes, 6 and 7; 250mM imidazole elutions, 8; protein marker. The white arrow indicates tCry1Ac.

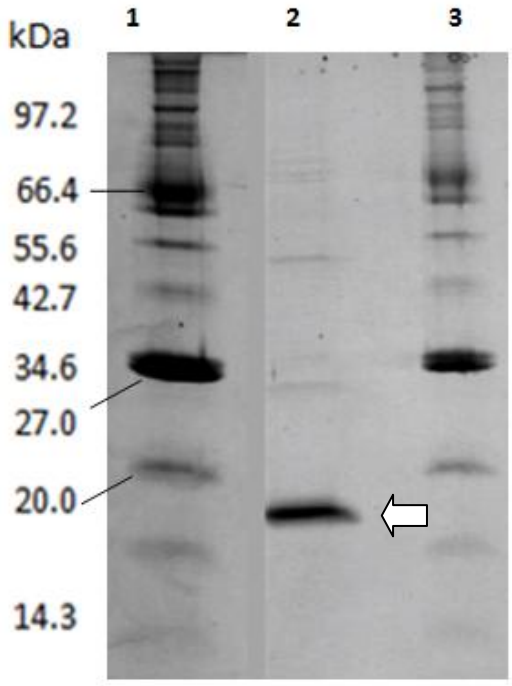


Figure 4.4: SDS PAGE analysis of the preliminary purification of CryD3 by IMAC. 1; broad range protein marker NEB, 2; elution of CryD3 in 250mM imidazole, 3; broad range protein marker. The white arrow represents CryD3.

4.4 Buffer exchange of tCry1Ac and CryD3

Buffer exchange and desalting are important steps following IMAC purification. This is because downstream processes, such as protein quantification and subsequent assays can be affected by the presence of imidazole. Both desalting and buffer exchange may be carried out simultaneously using viva spin columns. This also helps to concentrate low concentrations of tCry1Ac protein which facilitates its analysis by ELLA. tCry1Ac and CryD3 were concentrated and buffer exchanged into PBS. These protein samples were then be used for binding analysis using the ELLA.

4.5 Examining Lectin binding of tCry1Ac and CryD3 using the ELLA

The Enzyme Linked Lectin Assay (ELLA) method (which is outlined in section 2.21 and figure 4.5) is similar to the ELISA. The primary antibody in an ELISA is replaced by a carbohydrate-binding protein and the target is the glycans of the glycoprotein. The 6x histidine affinity tag which was engineered into tCry1Ac and CryD3 allow for the detection of the interaction between tCry1Ac and the glycoprotein by use of an anti-His antibody. The relative affinities and specificities of recombinant carbohydrate-binding proteins may be studied through the comparison with biotinylated commercial plant lectins. The dependency of tCry1Ac and CryD3 on metal ions is not known, but many plant lectins require calcium, magnesium and manganese for the function of the binding pocket.

Typically N-glycosylated glycoproteins are used in the ELLA's to study binding specificity. However as tCry1Ac binds to GalNAc, a glycoprotein with GalNAc on its surface was needed for the assay. Therefore a neoglycoprotein was obtained as this consists of numerous GalNAc linked onto BSA. The binding of tCry1Ac to the neoglycoprotein with a GalNAc surface and also to an array of glycoproteins was investigated. The commercial lectins used in this assay were ECL, SBA, RCA, PNA and DBA (see figure 4.6). The results presented in Figure 4.6, seem to demonstrate that the purified tCry1Ac displayed a strong specificity for GalNAc in comparison to the plant commercial lectins. A dilution series of the neoglycoprotein and tCry1Ac were carried out in order to examine the relative affinities of tCry1Ac to GalNAc. (Figures 4.7 and 4.8) This further analysis showed a linear affinity of tCry1Ac for GalNAc and binding at low concentrations of tCry1Ac.

CryD3 was predicted to have Galactose specificity. Various glycoproteins with galactose on the terminus were used in the ELLA such as asialofetuin and a Galα1-3, Gal neoglycoprotein. The GalNAc neoglycoprotein was also used to assess if CryD3 had any GalNAc specificity. The commercial lectin ECL was used as a positive control for Galactose, while tCry1Ac was used as a positive control for GalNAc.

Figure 4.9 shows the dilutions of CryD3 to asialofetuin, the affinity is not very strong as it correlates to the drop after each dilution.

4.5.1. ELLA Assay Methodology

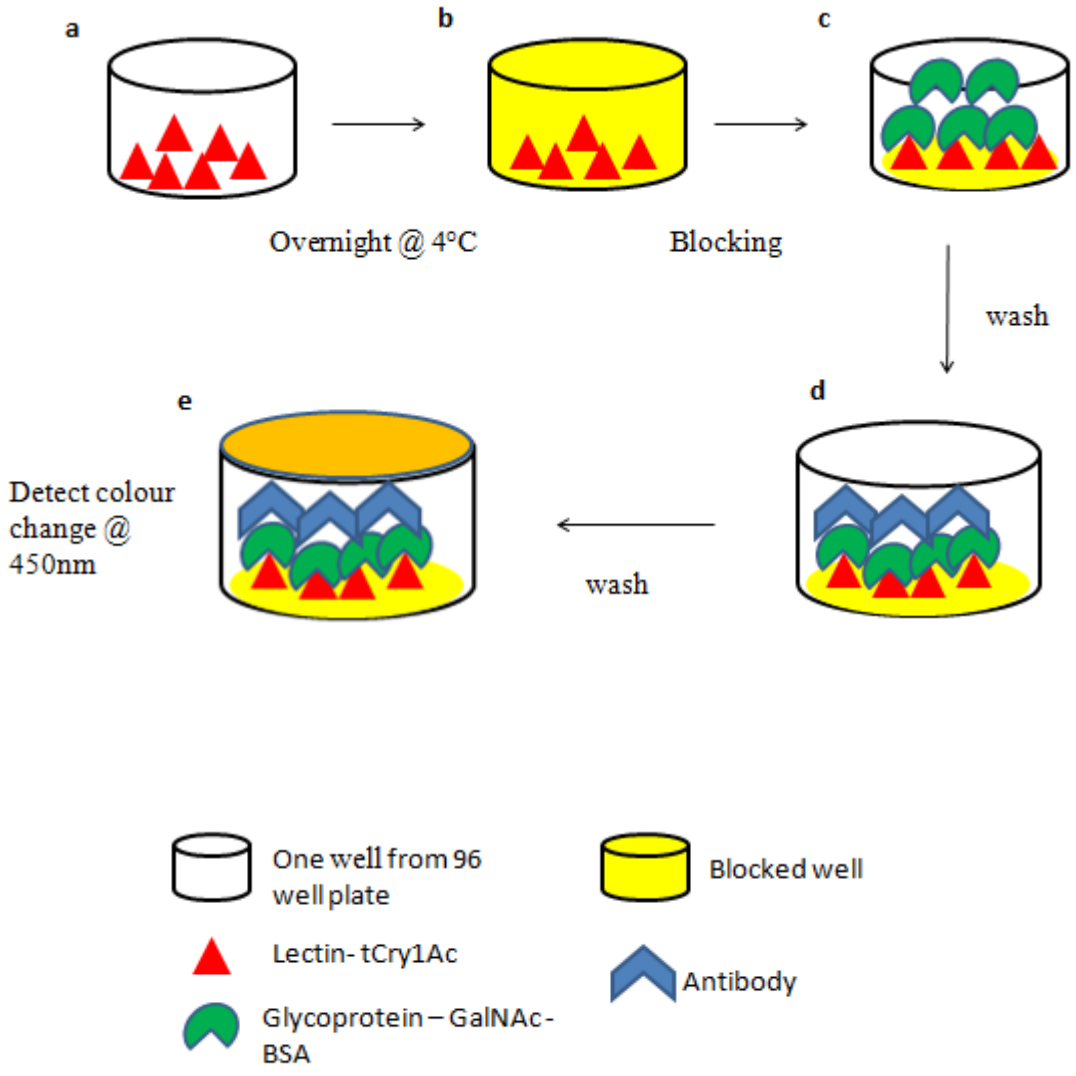


Figure 4.5: The ELLA assay methodology. (a) Lay down lectin or carbohydrate binding protein (e.g. tCry1Ac), (b) block with BSA, (c) incubate with neo-glycoprotein (GalNAc-BSA), (d) incubate with antibody, (e) detect antibody and colour change at Abs 450nm.

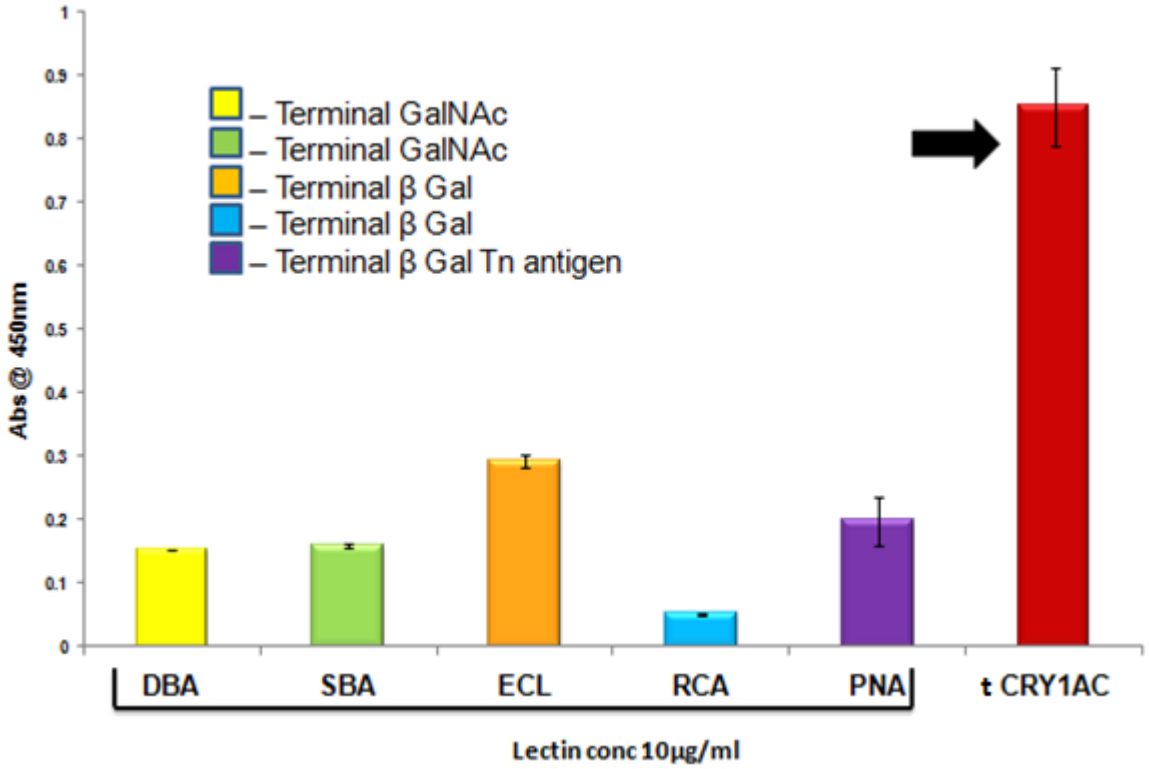


Figure 4.6: Comparison of tCry1Ac and commercial plant lectin binding profiles by ELLA analysis. ELLA analysis of tCry1Ac 10 µg/ml and commercial lectins binding to GalNAc linked BSA, ELLA samples averaged in triplicate and ELLA experiment carried out three times. The legend shows the specificity of each commercial lectin. Interestingly tCry1Ac showed strong binding to GalNAc-BSA compared to the commercial plant lectins. The black arrow shows the binding of tCry1Ac to GalNAc-BSA.

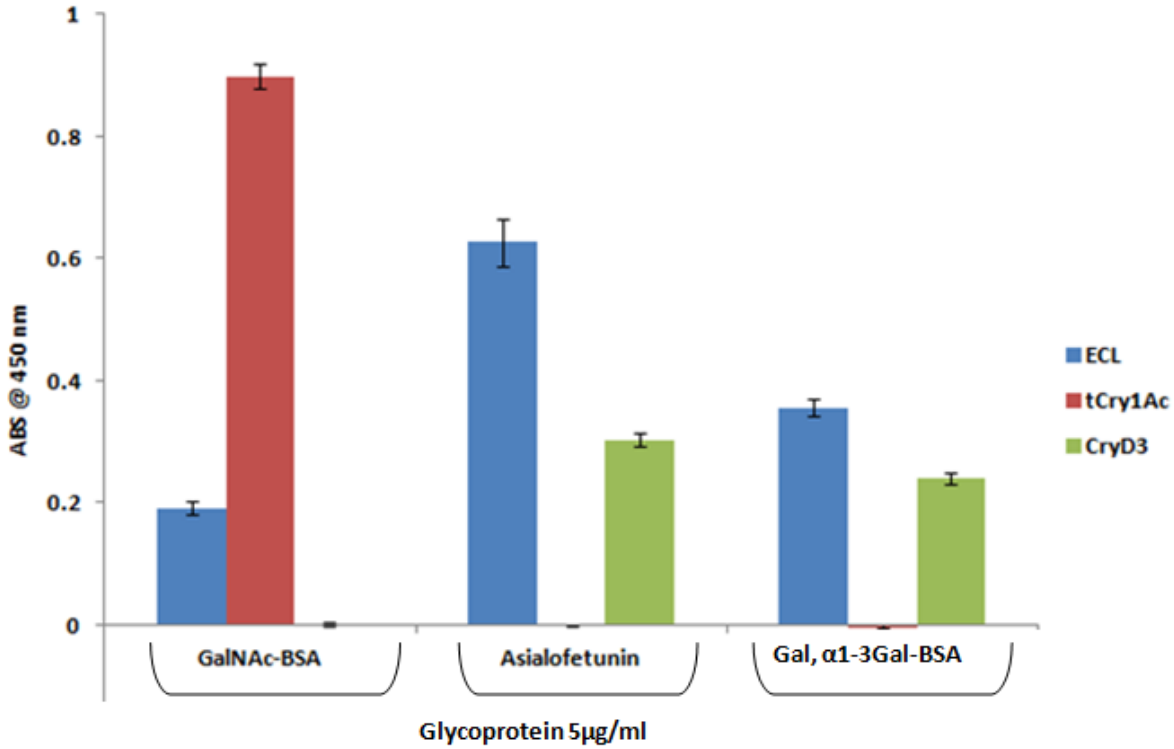


Figure 4.7: ELLA analysis showing binding of tCry1Ac, CryD3 compared to the commercial lectin ECL. ELLA of tCry1Ac10µg/mL, CryD310µg/mL, and the commercial lectin ECL10µg/mL to neo-glycoprotein GalNAc-BSA, the glycoprotein asialofetuin and neo-glycoprotein Gal,α1-3Gal-BSA. ELLA samples averaged in triplicate and ELLA experiment carried out three times. tCry1Ac shows binding to GalNAc-BSA. Significantly CryD3 has specificity for asialofetuin and galactose α1-3.

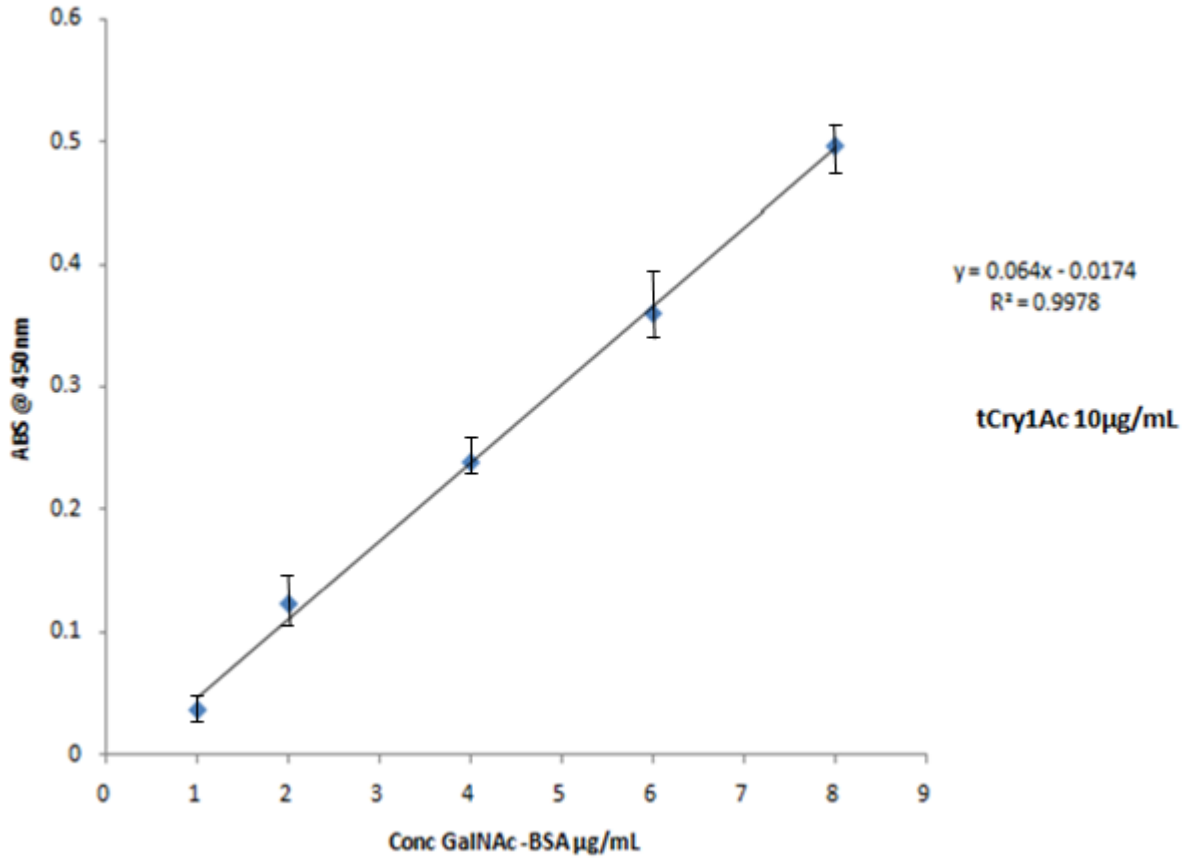


Figure 4.8: Effect of increasing concentrations of GalNAc-BSA on tCry1Ac binding activity by ELLA analysis. GalNAc-BSA concentration from 1µg/mL – 8µg/mL, concentration of tCry1Ac 10µg/mL. ELLA samples averaged in triplicate and ELLA experiment carried out three times. The graph is a linear correlation of the affinity between GalNAc-BSA and tCry1Ac.

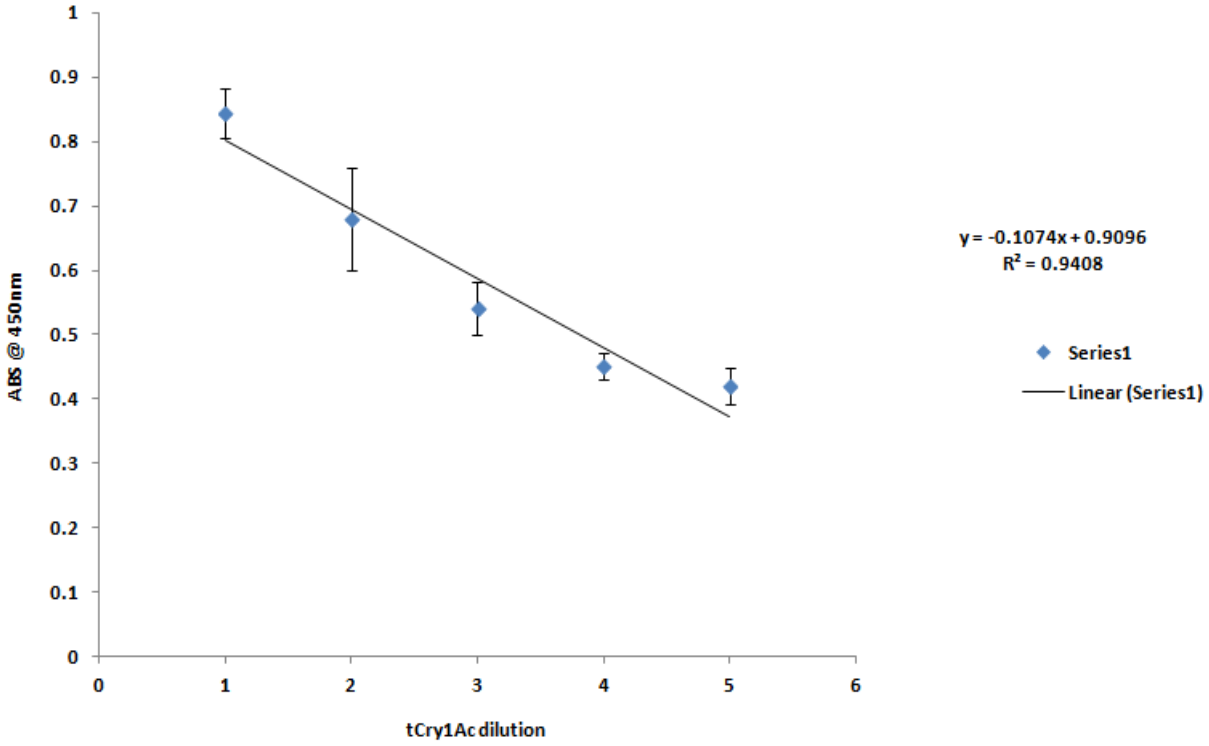


Figure 4.9: Effect of decreasing concentrations of tCry1Ac protein on binding to 5µg/mL GalNAc-BSA by ELLA analysis. The working concentration of tCry1Ac was 100µg/mL, 1 is 20 µg/mL, 2 is 15 µg/mL, 3 is 10 µg/mL, 4 is 5 µg/mL and 5 is 2.5 µg/mL. ELLA samples averaged in triplicate and ELLA experiment carried out three times. The affinity here is slightly better showing a slower drop in signal with each dilution.

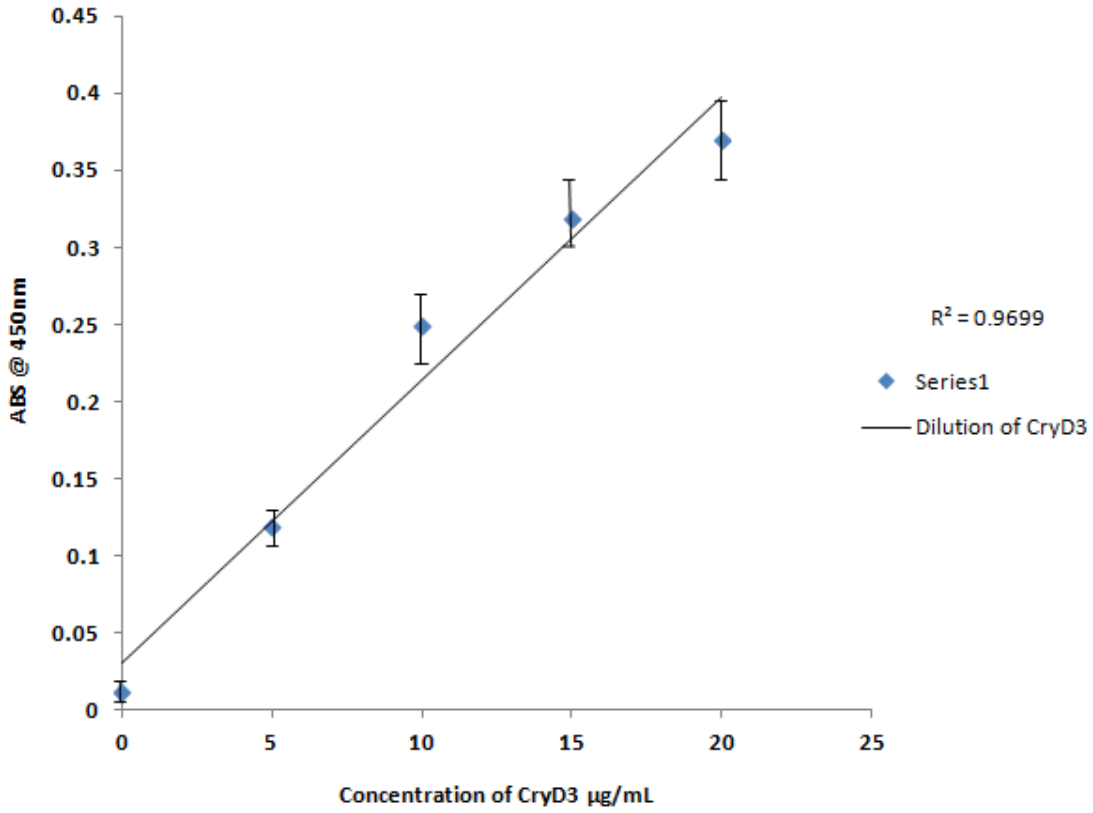


Figure 4.10: Effect of increasing concentrations of CryD3 on the binding to asialofetuin by ELLA analysis. CryD3 in concentrations of 20 µg/mL, 15 µg/mL, 10 µg/mL, and 5µg/mL was tested against 5µg/mL of asialofetuin in order to assess the affinity of their binding. ELLA samples averaged in triplicate and ELLA experiment carried out three times.

4.6 Optimisation of expression of recombinant tCry1Ac

As the expression of tCry1Ac was low in the soluble fraction, optimisation of the expression was carried out to obtain a higher yield of protein. Expression studies on various factors that influence expression of a recombinant protein, such as strain of *E. coli*, temperature, IPTG concentration and media were carried out. Time-course expression was carried out for each condition and the resulting soluble and insoluble fractions were analysed on SDS PAGE gels.

4.6.1 Choice of host strain for tCry1Ac

Two strains of *E. coli* were examined JM109 and KRX for tCry1Ac protein expression. JM109 and KRX are both strains designed for protein expression type strains. KRX has been modified to protect the expressed protein from degradation. Although the KRX cells grew faster than JM109 after induction (Figure 4.11) there was some increase in expression of tCry1Ac in the overnight soluble fraction of JM109 when compared to KRX. Therefore it was decided to use JM109 for expression of tCry1Ac (Figure 4.12).

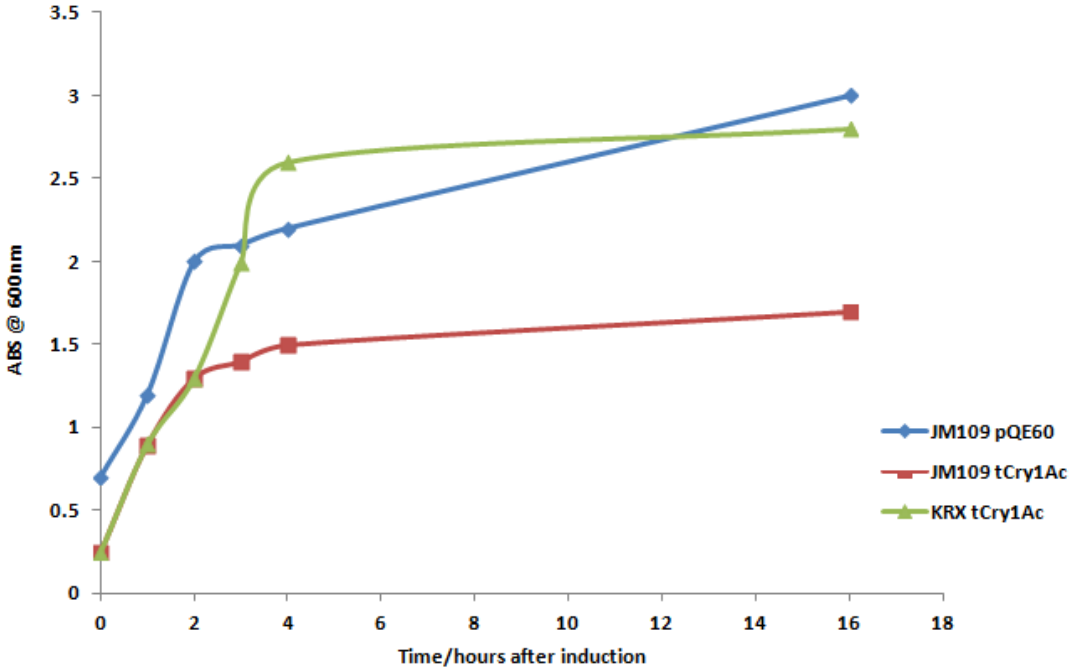


Figure 4.11: Comparasion of *E. coli* strains under expression conditions. The growth curve of *E. coli* strains JM109 and KRX were compared, while expressing tCry1Ac. The graph shows pQE60 in JM109, tCry1Ac in JM109 and tCry1Ac in KRX. Timepoints were taken post induction with IPTG.

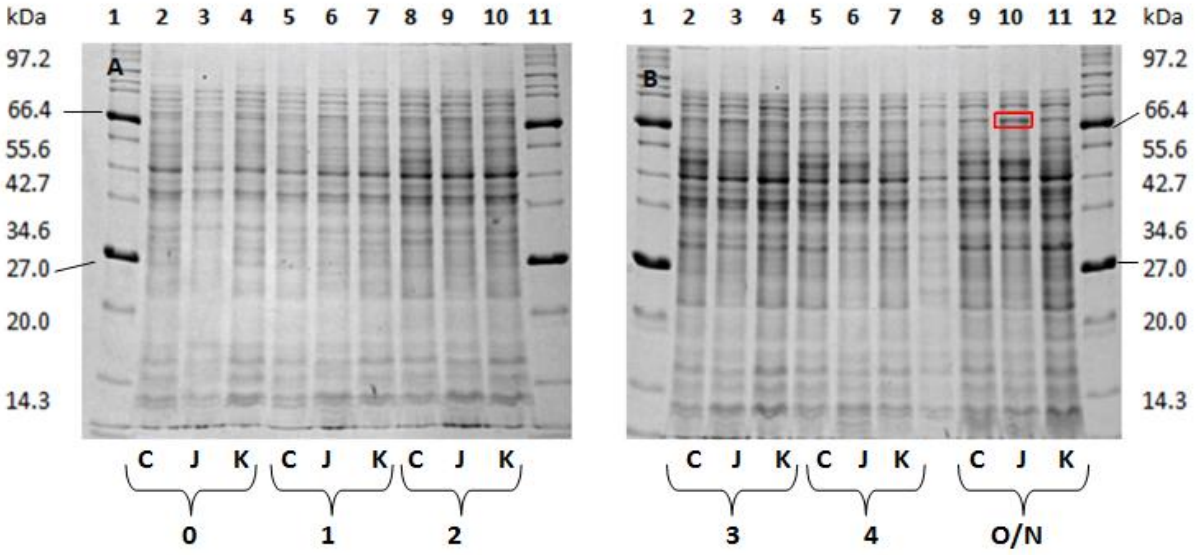


Figure 4.12: SDS PAGE of soluble expression of tCry1Ac in different expression *E. coli* strains, JM109 and KRX. **C** represents JM109 pQE60 as the control, **J** represents JM109 tCry1Ac, **and K** represents KRX tCry1Ac.

Panel A: 1 and 11; broad range protein marker, **2**; Control, time of induction, **3**; tCry1Ac JM109, time of induction, **4**; tCry1Ac KRX, time of induction, **5**; Control, 1 hour post induction, **6**; tCry1Ac JM109, 1 hour post induction, **7**; tCry1Ac KRX, 1 hour post induction, **8**; Control, 2 hours post induction, **9**; tCry1Ac JM109, 2 hour post induction, **10**; tCry1Ac KRX, 2 hour post induction.

Panel B: 1 and 12; broad range protein marker, **2**; Control, 3 hours post induction, **3**; tCry1Ac JM109, 3 hours post induction, **4**; tCry1Ac KRX, 3 hours post induction, **5**; Control, 4 hour post induction, **6**; tCry1Ac JM109, 4 hour post induction, **7**; tCry1Ac KRX, 4 hour post induction, **8**; Repeated, tCry1Ac KRX, 4 hour post induction **9**; Control, overnight, **10**; tCry1Ac JM109, overnight, **11**; tCry1Ac KRX, overnight. The red box indicates tCry1Ac JM109, overnight.

4.6.2 IPTG concentration optimisation during expression of tCry1Ac

IPTG controls the *lac* promoter by inducing expression, therefore a higher concentration of IPTG will give a higher level of expression and in turn a higher protein yield. However too much IPTG could be inhibitory to the cells and may limit protein production. Therefore different IPTG concentrations were used to induce the cultures. The gel (Figure 4.14) shows that there was no significant increase in expression over 50µM. The graph (Figure 4.13) shows that 50µM was the optimal amount of IPTG as it produced no quantifiable decrease in growth rate or cell number when compared to 5µM, an amount that does not induce gene expression. Therefore 50µM was chosen as the best concentration to induce tCry1Ac protein expression.

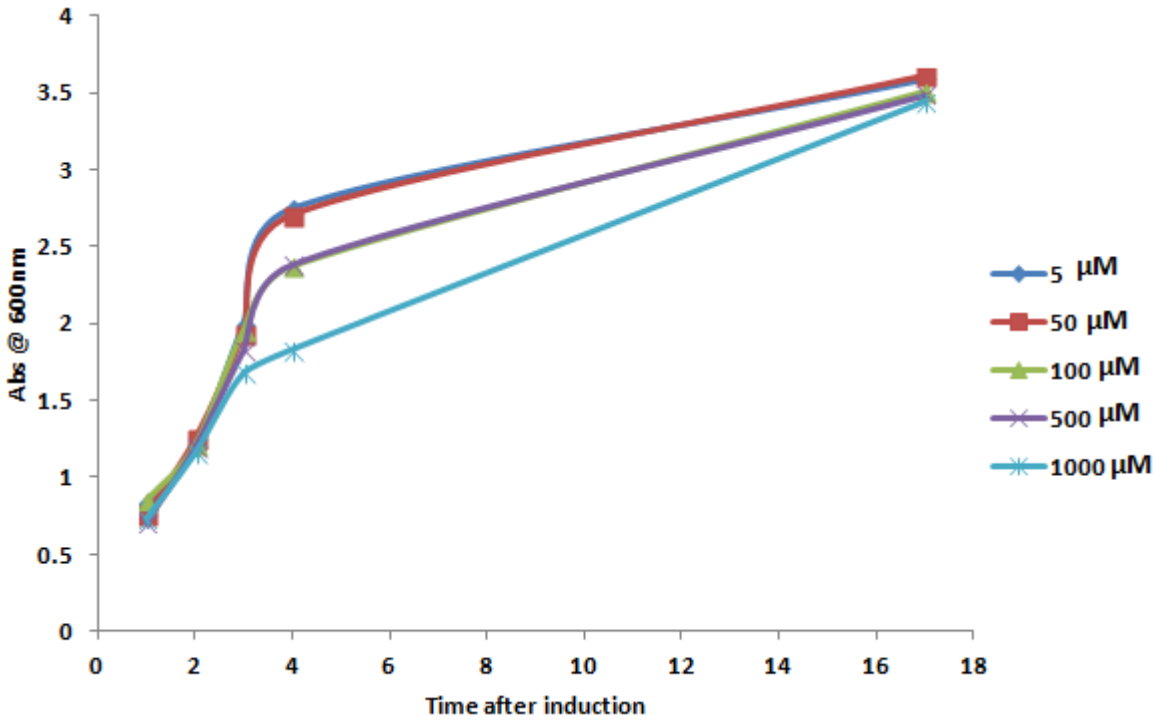


Figure 4.13: Expression growth curve of E. coli JM109 with tCry1Ac in varying concentrations of IPTG. At higher concentratation than 50µM IPTG the cells do not grow as well and have a slower growth rate.

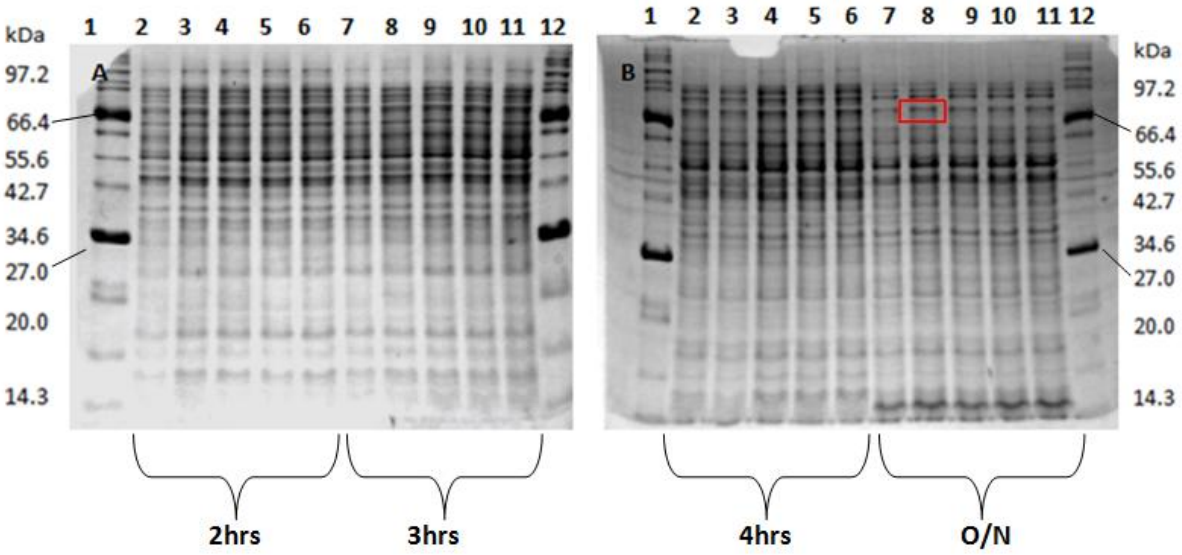


Figure 4.14: Time course study of the expression of tCry1Ac using varying concentrations of IPTG by SDS PAGE Analysis.

Panel A: 1, 12; broad range protein marker, **2**; 5µM, 2 hours post induction, **3**; 50µM, 2 hours post induction, **4**; 100µM, 2 hours post induction, **5**; 500µM, 2 hours post induction, **6**; 1000µM, 2 hours post induction, **7**; 5µM, 3 hours post induction, **8**; 50µM, 3 hours post induction, **9**; 100µM, 3 hours post induction, **10**; 500µM, 3 hours post induction, **11**; 1000µM, 3 hours post induction.

Panel B: **1, 12**; broad range protein marker, **2**; 5µM, 4 hours post induction, **3**; 50µM, 4 hours post induction, **4**; 100µM, 4 hours post induction, **5**; 500µM, 4 hours post induction, **6**; 1000µM, 4 hours post induction, **7**; 5µM, overnight, **8**; 50µM, overnight, **9**; 100µM, overnight, **10**; 500µM, overnight, **11**; 1000µM, overnight. Red box indicates tCry1Ac at 50µM overnight.

4.6.3 Temperature optimisation of expression of tCry1Ac

Temperature is an important factor for expression. The growth of *E. coli* is optimal at 37°C. However, the stability of a protein may be increased by lowering the temperature to 30°C when induced with IPTG. Expressing the Cry toxins in *E. coli* at a lower temperature of 20-25°C may yield a higher level of stable protein in the soluble fraction. Therefore, expression of tCry1Ac was carried out and compared at 30°C and 20°C. Figure 4.15 shows that there is a slightly higher expression of soluble tCry1Ac at 20°C. Therefore, protein expression was carried out at 20°C post induction. As tCry1Ac is a large protein at about 69kDa. *E. coli* may have some difficulty over expressing the protein. Slowing the growth rate down by lowering the expression temperature to 20°C gives the cells more time to produce the protein.

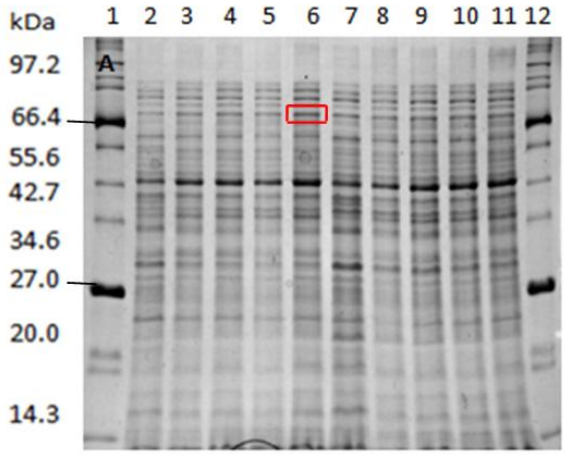


Figure 4.15: Time course study of the expression of tCry1Ac using varying temperature by SDS PAGE analysis.

(A) 1, 12; broad range protein marker, **2**; 20°C, one hour post induction, **3**; 20°C two hour post induction, **4**; 20°C three hour post induction, **5**; 20°C four hour post induction, **6**; 20°C overnight, **7**; 30°C one hour post induction, **8**; 30°C two hour post induction, **9**; 30°C three hour post induction, **10**; 30°C four hour post induction, **11**; 30°C overnight.

4.6.4 Comparison of TB and LB for the expression of tCry1Ac

The two types of bacterial growth media used here were LB and TB. LB is a medium used routinely for growing *E. coli*. TB is a highly nutritious medium that is used for protein expression. *E. coli* grows very well in this media and continues to grow and produce protein to a higher OD when compared to LB (Figure 4.16). The results presented in Figure 4.16 and in Figure 4.17 clearly show that there is higher expression of tCry1Ac in the overnight TB than in the overnight LB. TB was thus chosen as the media for the expression of tCry1Ac and CryD3 at 20°C.

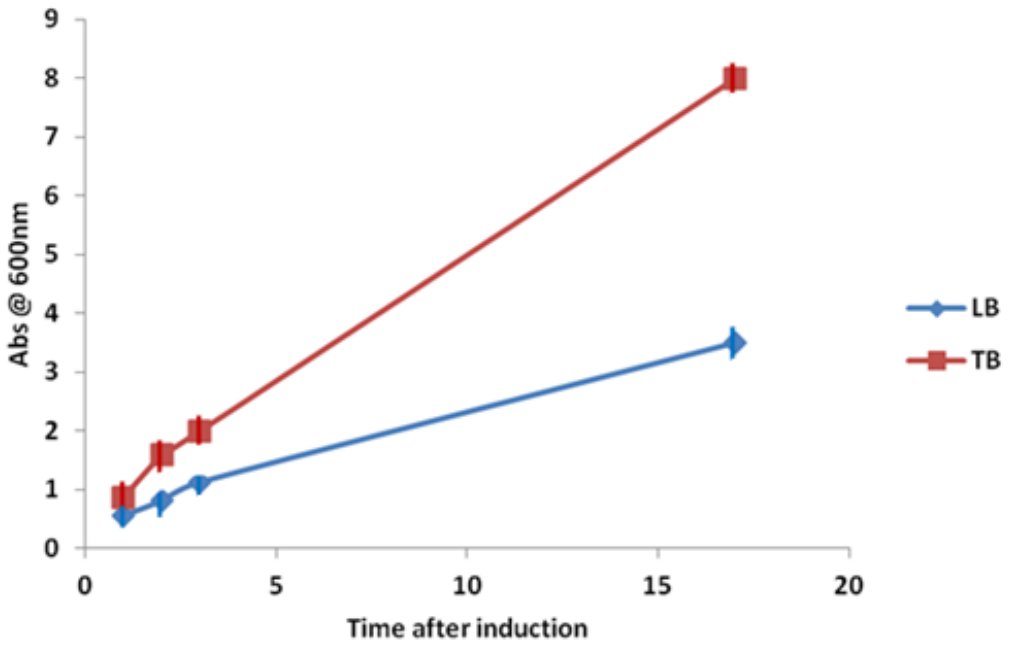


Figure 4.16: Growth curve comparasion, TB and LB media expressing tCry1Ac. Comparasion of growth of tCry1Ac JM109 in TB and LB broth.

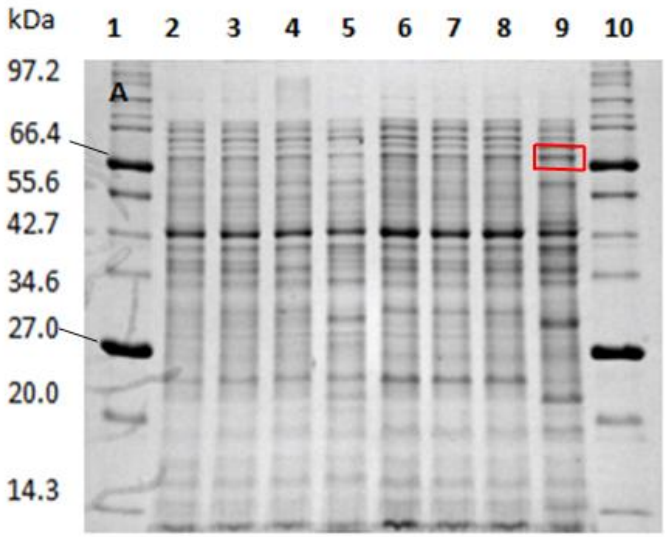


Figure 4.17: SDS PAGE analysis of the time-course expression of tCry1Ac in TB and LB growth medium.

1 and 10; Broad range protein marker, **2**; LB 2 hours post induction, **3** LB 3 hours post induction, **4**; LB 4 hours post induction, **5**; LB overnight, **6**; TB 2 hours post induction, **7**; TB 3 hours post induction, **8**; TB 4 hours post induction, **9**; TB overnight.

4.7 Further optimisation of the purification of the tCry1Ac and CryD3 proteins

The conditions for the expression of tCry1Ac were now optimised. The expression strain was *E. coli* JM109, the IPTG concentration used was 50µM, the temperature of expression was 20°C and the media used was TB. tCry1Ac was then purified by IMAC after expression with the new conditions. Figure 4.18 shows the purification of tCry1Ac by IMAC and reveals numerous protein bands in the eluted fractions. Even though stringent imidazole washes were carried out during tCry1Ac there was still a significant amount of contaminating proteins. At higher elution fractions there are many lower bands observed. Therefore a western blot was carried out to examine these elution fractions. As can be seen from figure 4.19, numerous bands are detected. These bands may represent cleaved fragments of tCry1Ac that have a 6x histidine tag.

Previously, in the small scale expression of CryD3 contaminating proteins were evident in the elution fractions also. Figure 4.20 shows the results obtained from the optimised purification of tCry1Ac and CryD3. It shows a similar protein band pattern for both proteins. Remarkably, there is a higher protein band in the CryD3 preparation which correlates with a band observed in the tCry1Ac preparation.

These protein samples were analysed on an ELLA and neither of the protein preparations showed activity. This could be due to protein inactivity because of protease activity or interference by the contaminating proteins. CryD3 also fell out of solution and aggregated after one day of storage at 4°C. As both proteins were active and binding to GalNAc and galactose in section 4.5, it is clear that both proteins are being affected by the contaminating proteins or instability. Therefore the stability of both proteins and protease activity during purification were examined further.

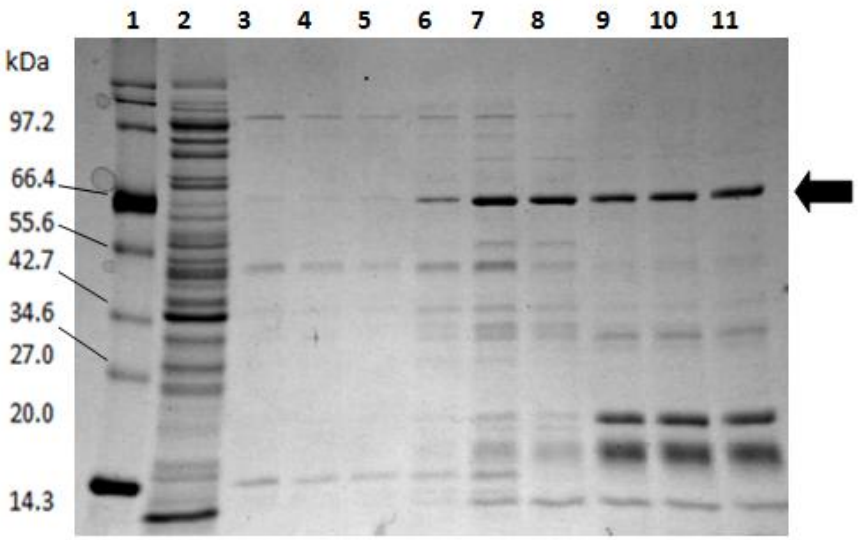


Figure 4.18: SDS PAGE analysis of the tCry1Ac protein purified by IMAC with increasing imidazole washes. 1; Broad range protein marker, **2**; flow through, **3**; 40mM imidazole wash, **4**; 60mM imidazole wash, **5**; 80mM imidazole wash, **6**; 100mM imidazole wash, **7-11**; 300mM imidazole elution. The black arrow indicates tCry1Ac.

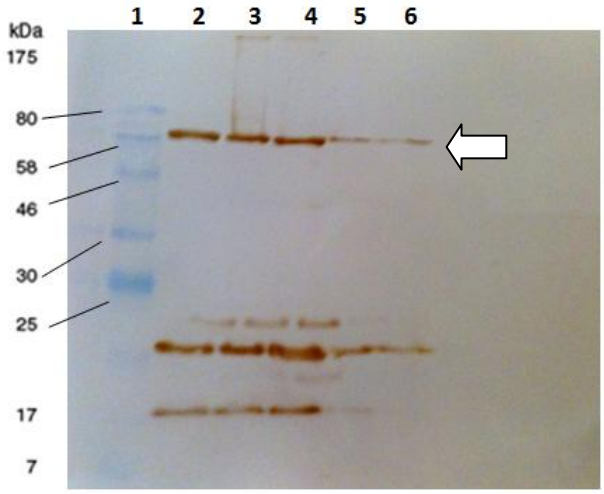


Figure 4.19: Western blot analysis of the tCry1Ac protein following IMAC purification. To detect if the contaminating bands were cleaved parts of tCry1Ac a western blot was done to detect the 6x histidine tag on the C terminus of the protein. **1**; Prestained plus protein ladder, **2-6**; Elutions of tCry1Ac from IMAC. The white arrow indicates tCry1Ac.

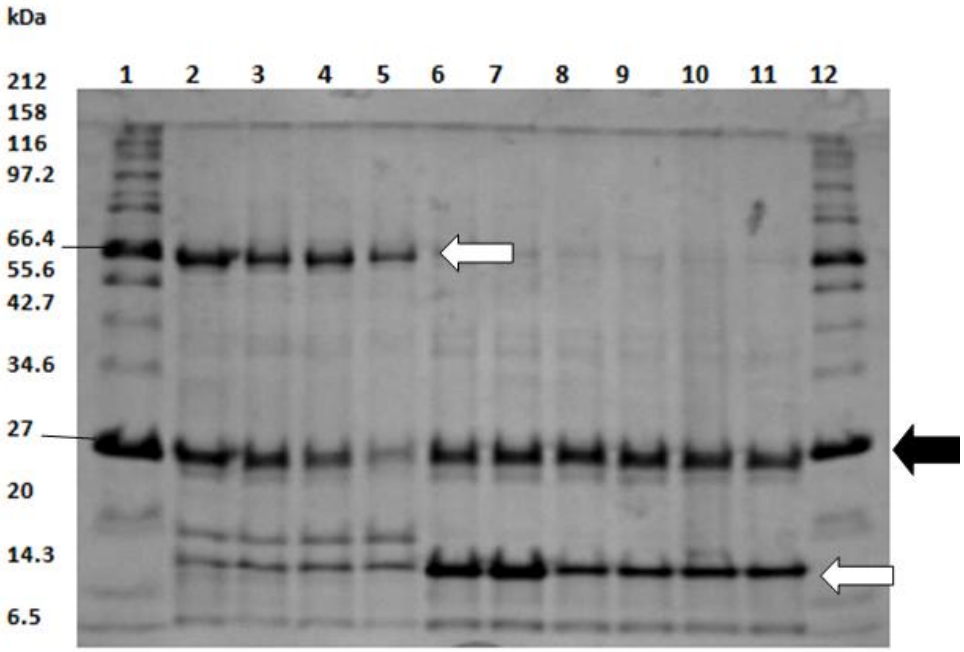


Figure 4.20: SDS PAGE analysis of the optimised purification by IMAC of tCry1Ac and CryD3. These purifications had washes of 100mM imidazole, the 300mM imidazole elutions of tCry1Ac and CryD3 are shown on the gel. 1 and 12; Broad range protein marker, 2-5; tCry1Ac 300mM imidazole elutions, 6-11; CryD3 300mM imidazole elutions. There is a one band that is common to both purifications at approx 30kDa. White arrows indicate tCry1Ac and CryD3. There is a band that is common to both purifications at approx 30kDa (black arrow).

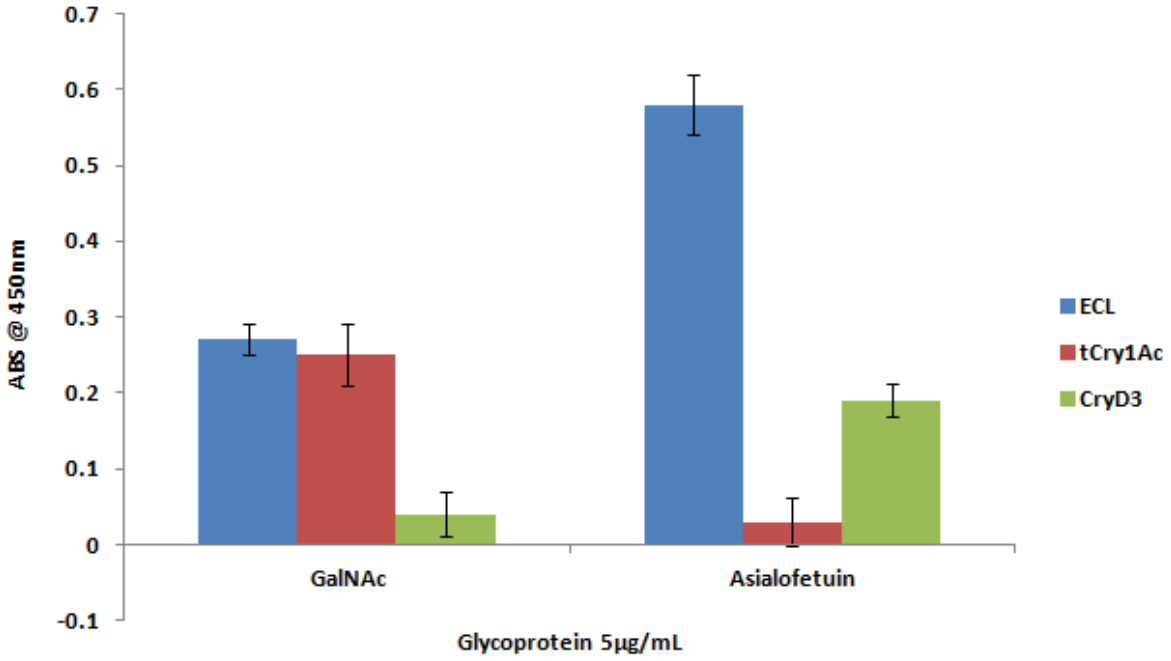


Figure 4.21: ELLA analysis of tCry1Ac and CryD3 proteins compared to the commercial lectin ECL. The purified samples from the IMAC optimisation of tCry1Ac 10µg/ml (red) and CryD3 10µg/ml (green) were analysed on an ELLA. There was a very low signal for tCry1Ac (red) and CryD3 (green).

4.7.1 Examining the stability and degradation of tCry1Ac and CryD3 during IMAC purification

SDS PAGE analysis of the tCry1Ac and CryD3 proteins both show similar patterns of contaminating protein bands following purification by the optimisation of the IMAC procedure. As these bands also showed up on the western blots (using anti-His tag antibodies) they were thought to be histidine tagged C terminal cleaved protein fragments in the case of tCry1Ac and/or possibly aggregations of CryD3. Therefore, experiments were carried out to examine the stability of both proteins and how their degradation and/or aggregation could be reduced. Firstly protease inhibitors were introduced and the plasmid containing tCry1Ac was introduced into the protease resistant *E. coli* KRX strain as described in section 2.6.1. The protein was expressed and purified and compared to *E. coli* JM109 with no protease inhibitors. The results of the purifications may be seen in figures 4.22 and 4.23. It is apparent that the protease inhibitors and the KRX strain had no effect on protein degradation, and surprisingly, there were even more bands found in this purification over the original purification. ptCry1Ac and pCryD3 were used to transform *E. coli* strain BL21 to examine if the there was any difference in protein expression or purification. Figure 4.24 shows the purified samples of tCry1Ac and CryD3 expressed in BL21 compared to JM109. There is no significant difference in protein expression, and the same protein band pattern is observed for both purifications.

As stated earlier, temperature was found to have a significant effect on the protein expression as a change from 30°C to 20°C increased soluble protein expression. Figure 4.25 shows aliquots of CryD3 at various temperatures. The sample becomes cloudier at 30°C and at 37°C the protein was found to fall out of solution. Therefore, after the expression of the proteins, all experiments were carried out at 4°C. This improved greatly the quality of the protein, (see Figure 4.26.a). However, there were still some contaminating bands and consistently there was a band at about 30kDa in both protein preparations. Therefore, another western blot was carried out on these samples (see Figure 4.26. b). This western blot showed tCry1Ac at 69kDa and CryD3 at 20kDa. However, there was no protein at the 30kDa band to correspond to the SDS PAGE gel. Therefore, it was concluded that this protein is either a degraded part of the Cry protein that has bound to the purified proteins and so is co-purified, or is a contaminating protein from the host strain of *E. coli*. From this SDS PAGE gel

the two corresponding bands at 30kDa were excised and sent to Alphalyse to be identified by MS peptide mapping and sequencing analysis.

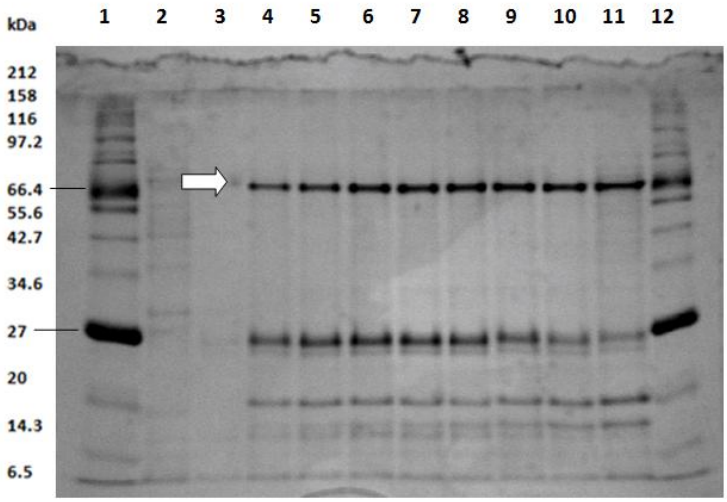


Figure 4.22: **SDS PAGE analysis of the purified of tCry1Ac, expressed in *E. coli* JM109.** 1 and 12; Broad range protein marker, 2; flow through, 3; 100mM imidazole wash, 4-11; 300mM imidazole elution. The white arrow represents tCry1Ac.

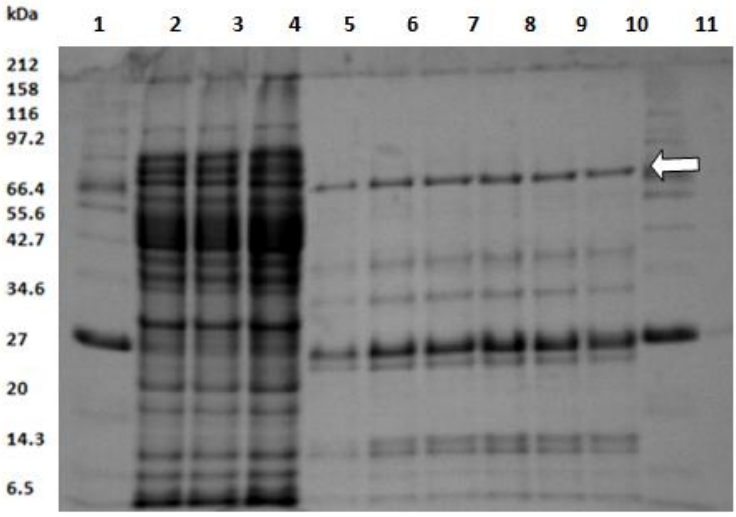


Figure 4.23: SDS PAGE analysis of the purified of tCry1Ac, expressed in *E. coli* KRX.1 and 11; Broad range protein marker, 2; Lysate, 3; Flow through, 4; 100mM imidazole wash, 5-10; 300mM imidazole elutions. The white arrow represents tCry1Ac.

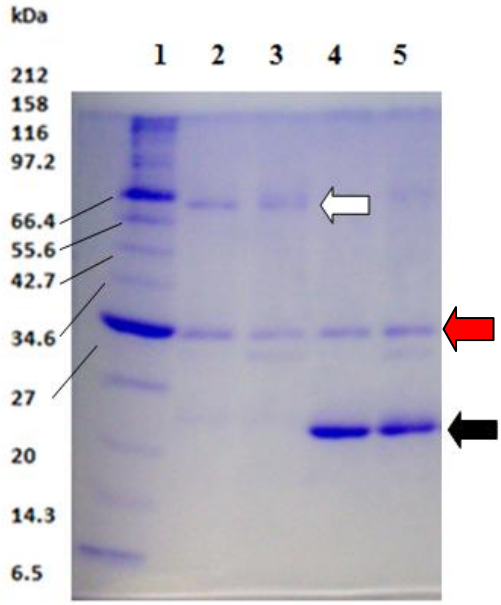


Figure 4.24: Investigation of *E. coli* strain BL21 for the expression of tCry1Ac and CryD3. ptCry1Ac and pCryD3 were used to transform *E. coli* strain BL21 to investigate if there was any increased expression or differences in contaminating protein. Lane 1; broad range protein marker, 2; tCry1Ac from JM109, 3; tCry1Ac from BL21, 4; CryD3 from JM109, 5; CryD3 from BL21. It is clear that there was no difference in expression or any change in contaminating protein. The white arrow indicates tCry1Ac, the black CryD3 and the red the contaminating protein.

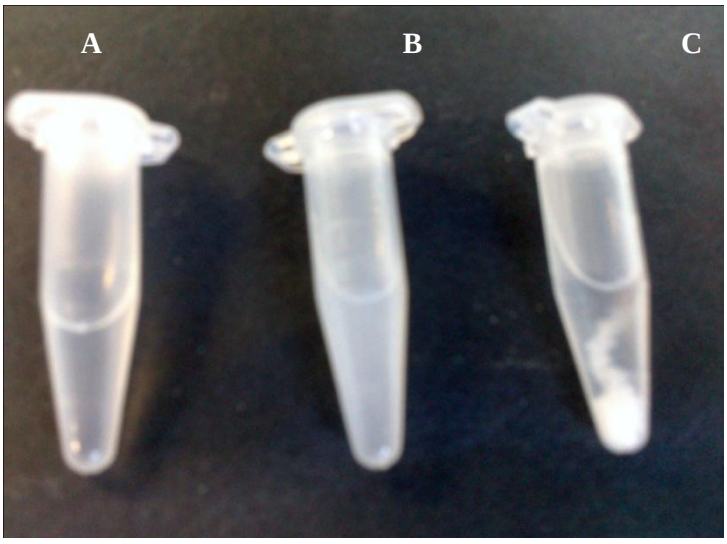


Figure 4.25: CryD3 solubility tested at various temperatures. CryD3 at A: 4°C, B: 30°C, C: 37°C. CryD3 starts to falls out of solution as the temperature rises. At 30°C the solution is cloudy and at 37°C there is a visible white precipitation in the solution.

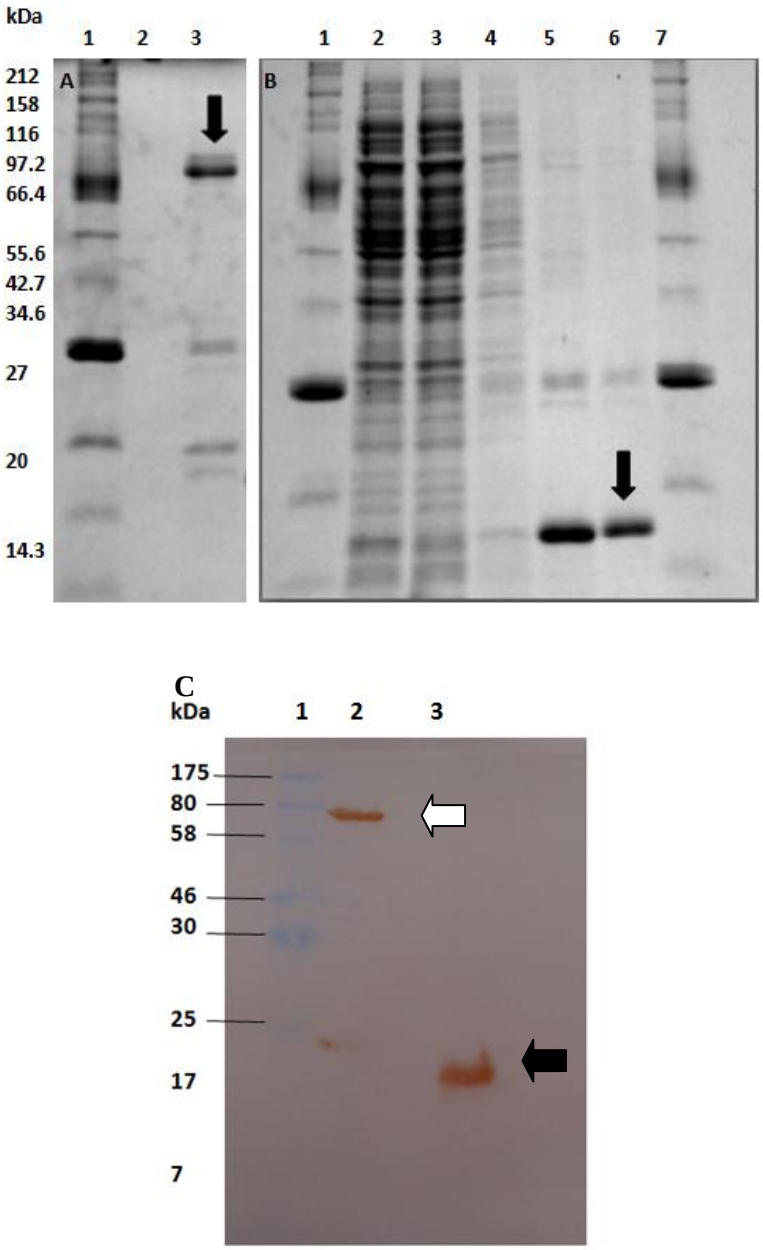


Figure 4.26: SDS PAGE gel and western blot of tCry1Ac and CryD3 purifications from JM109 at 4°C (A) SDS PAGE analysis of the purification of tCry1Ac at 4°C, 1; broad range protein marker, 3; tCry1Ac 300mM imidazole eluted sample. The black arrow represents tCry1Ac. **(B)** SDS PAGE analysis of the CryD3 purification at 4°C, 1 and 7; broad range protein marker, 2; Lysate, 3; Flow through, 4; 100mM imidazole wash, 5 and 6; 300mM imidazole elutions. The black arrow represents CryD3, **(C)** Western blot of SDS PAGE of purification of tCry1Ac and CryD3, 1; pre-stained protein marker, 2; tCry1Ac, 3; CryD3. The white arrow represents tCry1Ac while the black arrow indicated CryD3.

4.8 Identification of the contaminating protein bands by MS peptide mapping and sequencing analysis

The samples were extracted from the SDS PAGE gel in figure 4.26 and were sent to Alphalyse for protein analysis. The samples were identified by MS peptide mapping and sequence analysis as outlined in section 2.18. The results can be seen in table 4.1, revealing the major contaminating protein to be a cAMP-regulatory protein also known as the catabolite gene activator (CRP) from *E. coli.* The standards used in the experiment were transferrin and BSA. The sequence of the protein can be seen in figure 4.27.

The CRP protein complexes with cyclic AMP and binds to specific DNA near the promoter to regulate the transcription of several catabolite-sensitive operons (Aiba, 1985). The protein induces a severe bend in the DNA and it acts as a negative regulator of its own synthesis as well as for adenylate cyclase (cyaA), which generates cAMP. Constitutively expressed, levels decrease in stationary phase. The structure of CRP with highlighted histidine residues was made using PyMol and can be seen in figure 4.28. It was not known how the CRP protein was co-purified with both tCry1Ac and CryD3. Investigation into various purification and separation techniques was then done to remove CRP from the protein samples.

Table 4.1: Results from the MS peptide mapping and protein identification. The unknown protein was a cAMP-regulatory protein from *E. coli*.

Sample name	Protein found in database	GI-number	MW	Score	Seq. cov.
Cry unknown	cAMP-regulatory protein [*Escherichia coli* E24377A]	gi|157159326	23812	346	68%

Quality control standards included in the analysis Standard	Protein found in database	GI-number	MW	Score	Seq. cov.
1 pmol transferrin	transferrin [Homo sapiens]	gi|115394517	79190	1020	57%
62 fmol BSA	Chain A, Crystal Structure Of Bovine Serum Albumin	gi|367460260	68416	268	39%

```
  1 MVLGKPQTDP TLEWFLSHCH IHKYPSKSTL IHQGEKAETL YYIVKGSVAV
 51 LIKDEEGKEM ILSYLNQGDF IGELGLFEEG QERSAWVRAK TACEVAEISY
101 KKFRQLIQVN PDILMRLSAQ MSRRLQVTSE KVGNLAFLDV TGRIAQTLLN
151 LAKQPDAMTH PDGMQIKITR QEIGQIVGCS RETVGRILKM LEDQNLISAH
201 GKTIVVYGTR
```

Figure 4.27: Protein sequence of cAMP- regulatory protein. Matched peptides shown in bold underline with 68% predictability match.

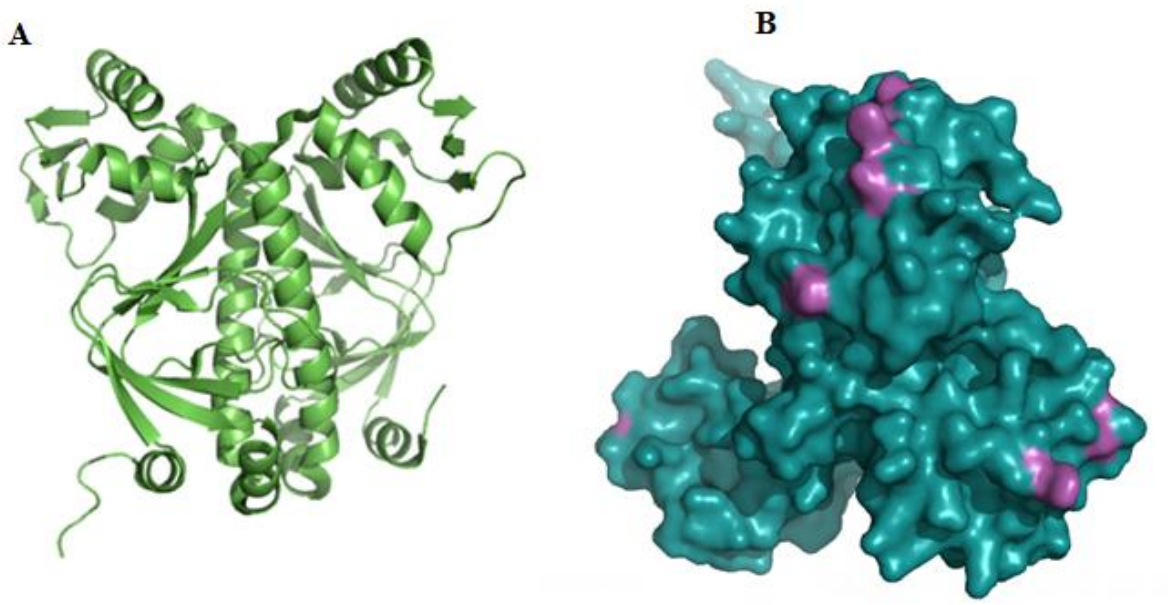


Figure 4.28: Protein structure of the catabolite gene activator. (A) Cartoon ribbon structure of CRP showing a dimer formation. (B) Cartoon model of the CRP surface with all histidine residues coloured in pink. Most of the histidines are exposed on the surface of the protein. Structure made using PyMol.

4.9 Purification of tCry1Ac and CryD3 from the catabolite gene activator

As the CRP was identified as the consistent contaminating protein in tCry1Ac and CryD3 purifications, experiments were designed to separate and purify these proteins using various chromatography techniques. The three techniques further discussed are IMAC using both detergent and competing sugars, Gel filtration and Ion exchange chromatography (IEX). These techniques were chosen to remove the contaminating protein that was consistently found in the purified samples of tCry1Ac and CryD3 and was postulated to be responsible for the loss of carbohydrate binding activities.

4.9.1 Purification of tCry1Ac and CryD3 by Ion exchange chromatography

Ion Exchange Chromatography (IXE) for the separation of biomolecules was introduced in the 1960s and continues to play a major role in the separation and purification of proteins, peptides, nucleic acids and other charged biomolecules (Yigzaw et al. 2009). It offers high resolution and group separations with high loading capacity. The technique is capable of separating molecular species that have only minor differences in their charge properties, for example two proteins differing by one charged amino acid. These features make IEX well suited for capture, intermediate purification or polishing steps in a purification protocol. IEX separates molecules on the basis of differences in their net surface charge. Molecules vary considerably in their charge properties and will exhibit different degrees of interaction with charged chromatography media according to differences in their overall charge, charge density and surface charge distribution. The charged groups within a molecule that contribute to the net surface charge possess different *pKa* values depending on their structure and chemical microenvironment. Since all molecules with ionizable groups can be titrated, their net surface charge is highly pH dependent. In the case of proteins, which are built up of many different amino acids containing weak acidic and basic groups, their net surface charge will change gradually as the pH of the environment change. IEX chromatography takes advantage of the fact that the relationship between net surface charge and pH is unique for a specific protein. In an IEX separation reversible interactions between charged molecules and oppositely charged IEX media are controlled in order to

favour binding or elution of specific molecules and achieve separation. A protein that has no net charge at a pH equivalent to its isoelectric point (pI) will not interact with a charged medium. However, at a pH above its isoelectric point, a protein will bind to a positively charged medium or anion exchanger and, at a pH below its pI, a protein will behind to a negatively charged medium or cation exchanger.

Here cation exchange chromatography was employed to separate tCry1Ac while anion exchange chromatography was used to separate CryD3 (see section 2.13.5). The pI for CRP is 8.38, CryD3 is 9.39 and tCry1Ac is 6.96. For the separation of tCry1Ac from CRP a pH of 7.7 was utilised to give CRP an overall positive charge while tCry1Ac should have an overall negative charge. Therefore at this pH the CRP should bind to CM sepharose with the tCry1Ac should be eluted. To separate CryD3 from CRP a pH of 8.8 was employed to give CRP a negative overall charge and CryD3 a positive overall charge. At this pH CryD3 should bind to the CM sepharose with the CRP remaining unbound. The CryD3 could then be eluted with a pH of 9.4. After both protein samples were purified and eluted they were analysed by SDS PAG. Figure 4.29 reveals that there was no significant improvement in tCry1Ac or CryD3 purification by the use of cation exchange chromatography. There was some retention of CryD3 on the CM sepharose but most of the protein did not bind to the column. CM sepharose is a weak exchanger therefore the experiment was repeated changing the pH to pH 7 and pH 8.5, for tCry1Ac and CryD3 respectively. The results show in Figure 4.30 show some improvement in protein purification but there are still some protein contaminants in the samples. An ELLA analysis of the eluted protein samples was carried out and the results displayed in Figure 4.31. This shows no binding activity for either tCry1Ac or CryD3. Overall, ion exchange chromatography did not effectively separate tCry1Ac and CryD3 from CRP. Therefore other methods of purification were examined.

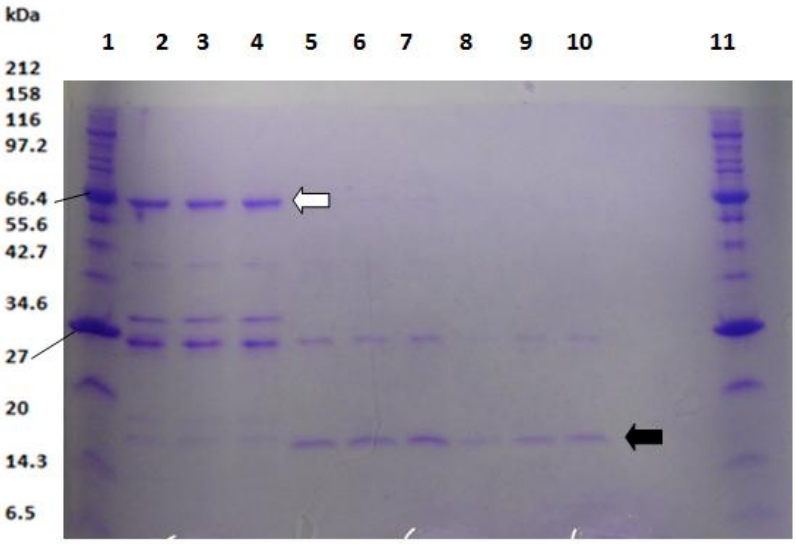


Figure 4.29: SDS PAGE analysis of purified tCry1Ac and CryD3 by cation exchange chromatography - using pH 7.8 and 8.8 respectively. Lanes 1 and 11; Broad range protein marker, 2-4; tCry1Ac in triplicate from the Ion exchange, 5-7; CryD3 washes in triplicate, 8-10; CryD3 elution in triplicate. The white arrow represents tCry1Ac, the black arrow CryD3.

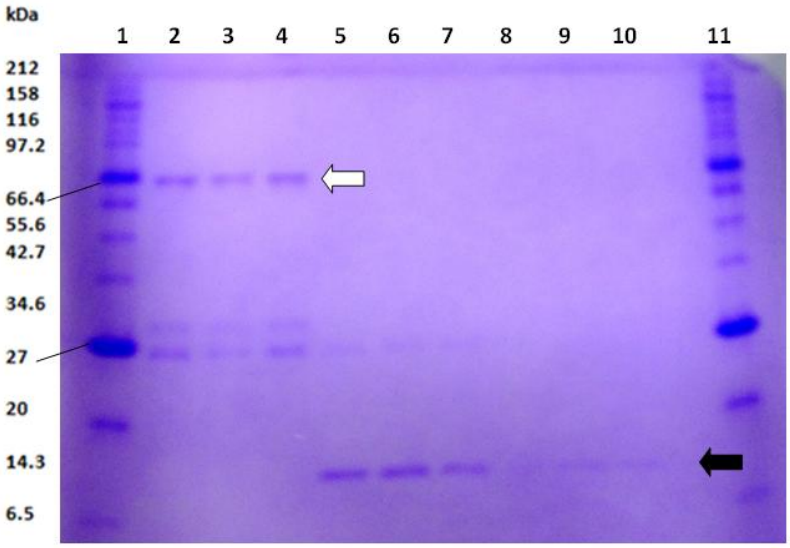


Figure 4.30: SDS PAGE analysis of purified tCry1Ac and CryD3 by cation exchange chromatography - pH 7 and pH 8.5 respectively. Optimised pH to pH 7 for tCry1Ac and pH 8.5 for CryD3 for stronger binding to the CM sepharose. Lane 1 and 11; Broad range protein marker, 2-4; tCry1Ac elutions, 5-7; CryD3 washes, 8-10; elutions of CryD3. The white arrow represents tCry1Ac, the black arrow CryD3.

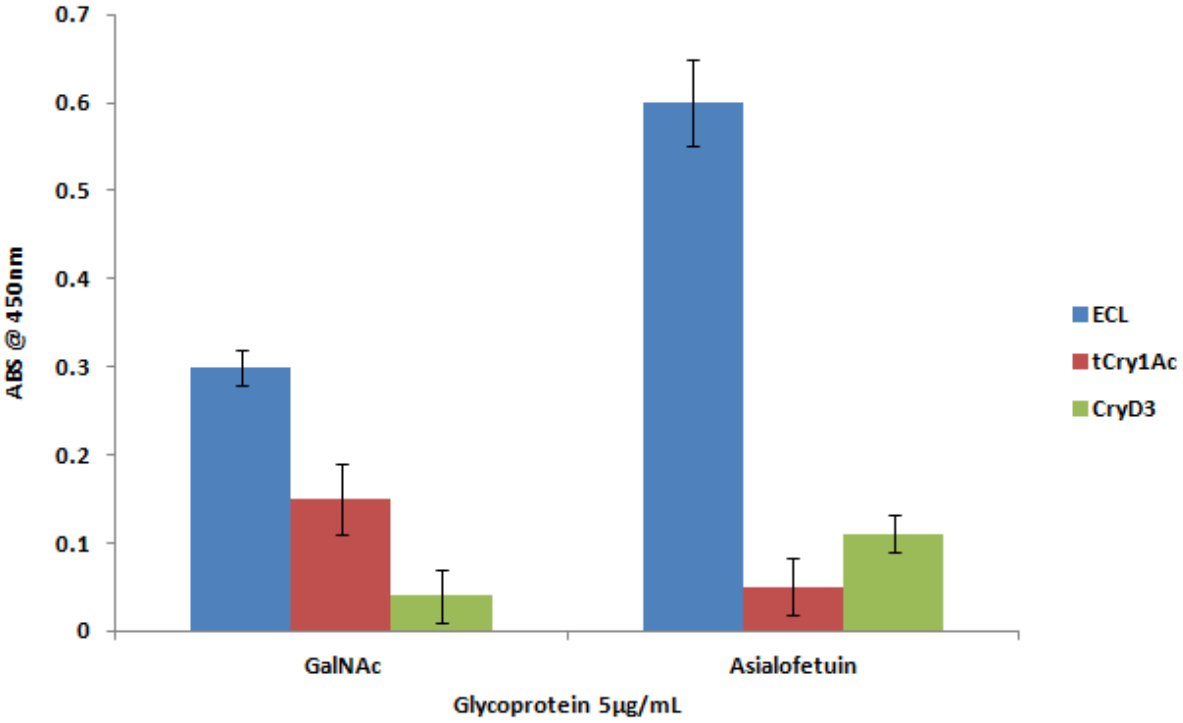


Figure 4.31: ELLA analysis of tCry1Ac and CryD3 purified by cation exchange chromatography. The samples from the cation exchange chromatography at pH 7 for tCry1Ac and pH 8.5 for CryD3. tCry1Ac (red) and CryD3 (green) were analysed on an ELLA and compared to the commercial lectin ECL. There is no significant binding for tCry1Ac (red) or CryD3 (green).

4.9.2 tCry1Ac purification using Gel filtration (size exclusion) chromatography

Gel filtration or size exclusion chromatography separates proteins based on differences in size as they pass through a packed column. The larger proteins will pass through the matrix faster than smaller proteins, due to the retardation of lower relative molecular mass proteins within the porous beads of the gel filtration medium. This technique can be used to characterise proteins, separate them from contaminating proteins and also examine how the proteins interact in a non denatured environment. The separation of tCry1Ac (69kDa) from CRP (30kDa) was attempted using gel filtration as there is a sufficient difference in relative molecular masses. As CryD3 (20kDa) is only approximately 10kDa smaller than CRP their separation would be difficult to achieve on this type of column.

4.9.2.1 Investigation of the purification of tCry1Ac by Toyopearl HW-55S

The Toyopearl SuperButyl-550 (HW-55S base bead) (Tosoh Biosciences, Germany) column matrix is composed of methyl-methacrylate. The compounds structure is displayed in Figure 4.32 and has no carbohydrate component. As a result, this matrix would be more suitable for the examination of carbohydrate binding molecules that might otherwise bind to a Sepharose matrix.

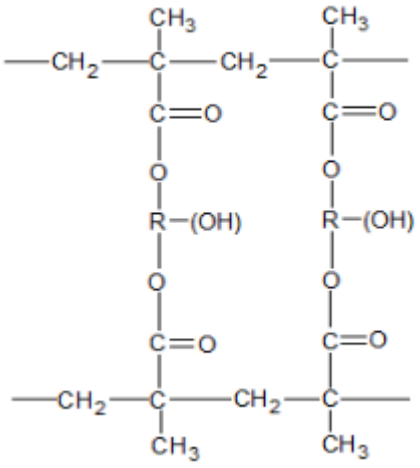


Figure 4.32: Polymeric structure of methyl methracylate. The hydroxylated polymeric structure that forms the filtration matrix in the SEC column HW-55S. R is a hydroxylated aliphatic group. Image produced using Chemsketch.

tCry1Ac was applied to the Toyopearl HW-55S column as outlined in section 2.13.3 and was run through the column. The results of that run are shown in figures 4.33 and 4.34. Figure 4.33 presents the full elution profile. The elution profile shows 4 protein peaks. The profile shows that there was some good separation of peak one and four in solution, however peak two and three did not separate very well as they similar in size. Figure 4.34 is magnified to show peaks 2, 3 and 4 on the elution profile. Further analysis of these proteins was carried out by SDS PAGE and ELLA. SDS PAGE (see Figure 4.35) indicates that peak one contains tCry1Ac with some contaminant proteins. Therefore this implies that these contaminating proteins are being co-purified with tCry1Ac as they cannot be separated by size exclusion chromatography. The results do show however that the two smaller bands in the tCry1Ac protein sample are not in peak one and so could be the proteins for peaks two and three on the elution profile. The results of the ELLA analysis, shown in Figure 4.36, show that peak one has a very low activity towards GalNAc but it is not significant. It is difficult to say how the CRP is binding to both tCry1Ac and CryD3. It could be due to hydrophobic bonds or even through the lectin domain. Therefore, further attempts to purify and separate tCry1Ac and CryD3 from the CRP were then carried out by IMAC in the presence of (a) detergent and (b) competing Galactose.

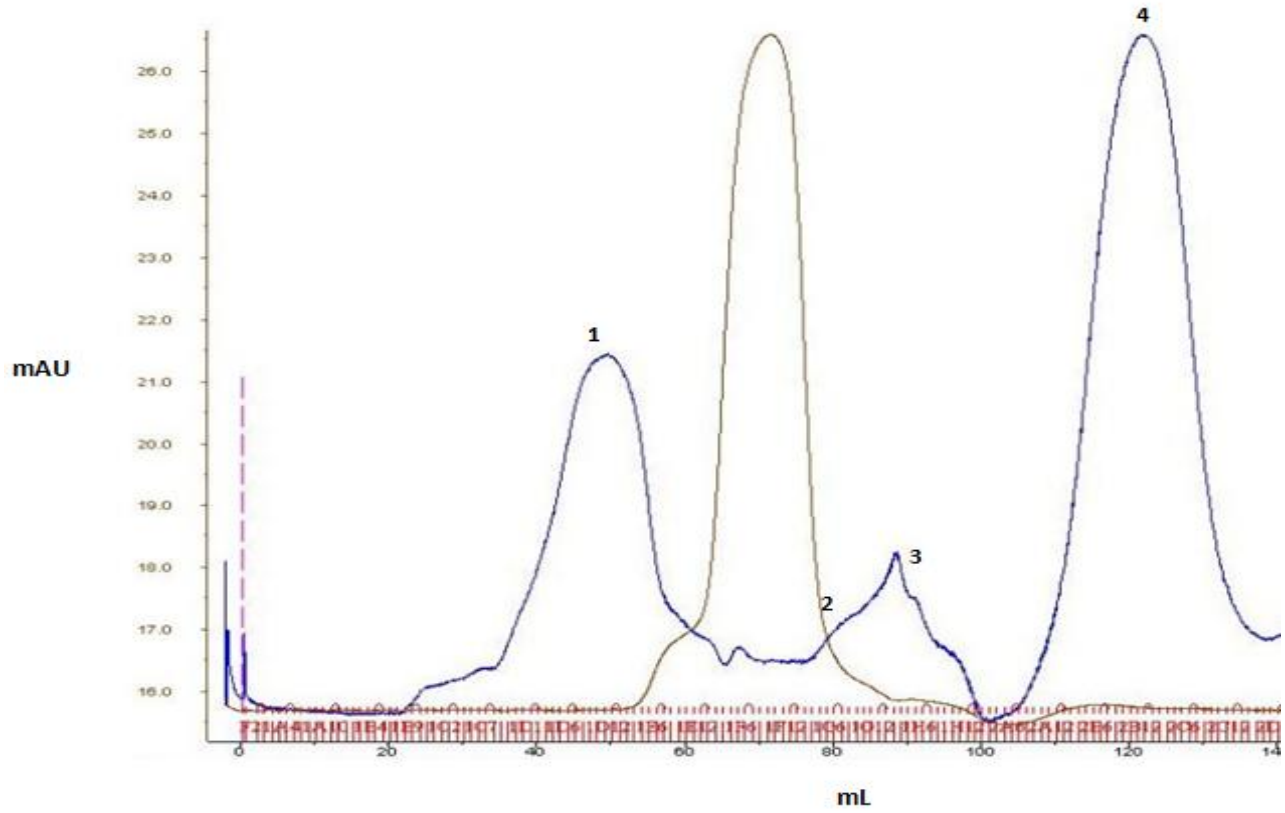


Figure 4.33: Elution profile of tCry1Ac in PBS pH 7.2 from the Toyopearl HW-55S SEC matrix. The elution profile of tCry1Ac using SEC with the Toyopearl HW-55S column. The optical density at OD280nm was plotted against the elution volume. There are 4 main protein peaks in blue.

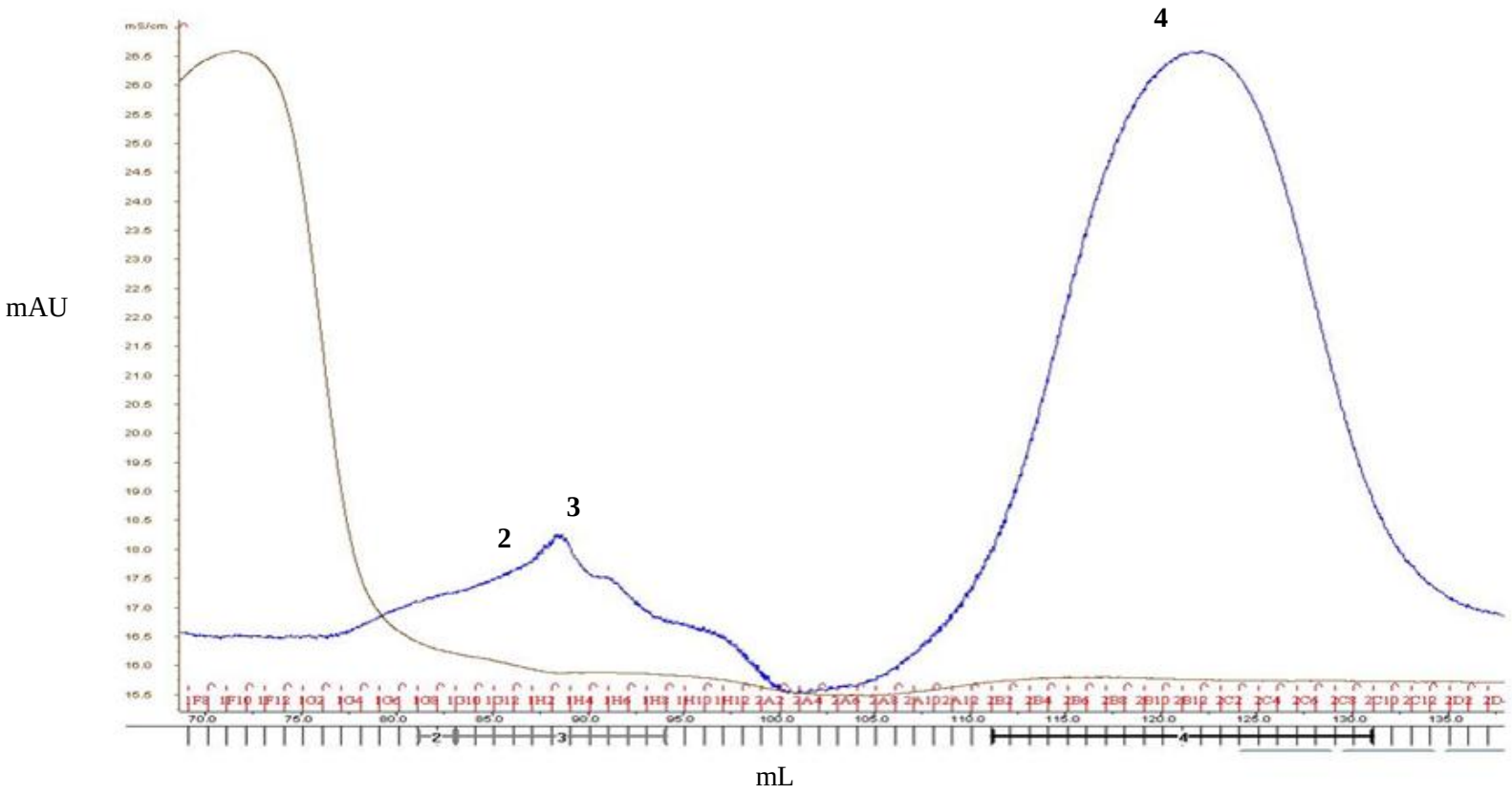


Figure 4.34: Peaks two, three and four from tCry1Ac elution profile. Magnified peaks 2, 3 and 4 from the elution profile of tCry1Ac by size exclusion chromatography.

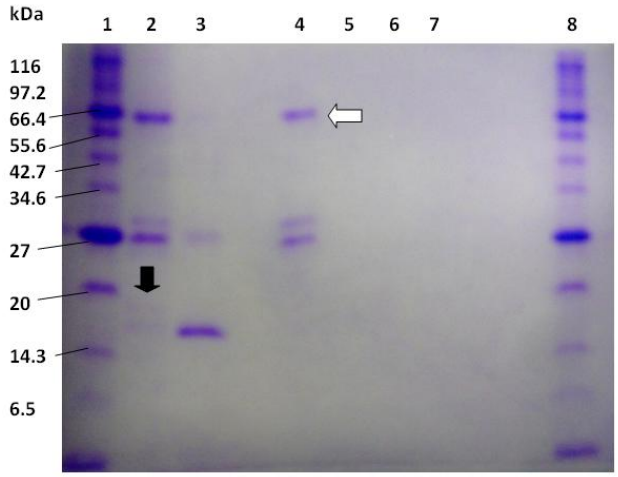


Figure 4.35: SDS PAGE of the peaks from the size exclusion column. 1 and 8; Broad range protein marker, 2; tCry1Ac before being put over the column, 3; CryD3, 4; Peak one, 5; Peak two, 6; Peak three, 7; Peak four. The white arrow represents tCry1Ac. The black arrow post to two small bands that are not in the elution fraction of peak one.

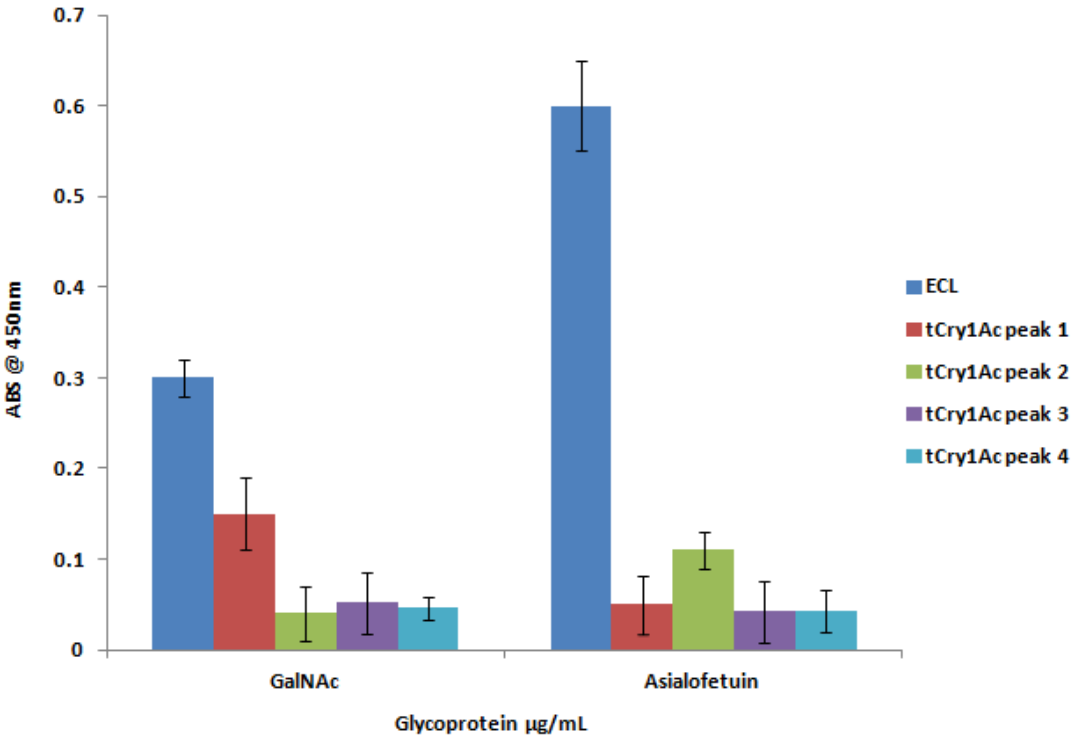


Figure 4.36: ELLA analysis of the four peaks of the tCry1Ac sample produced by gel filtration. tCry1Ac peak 1, 2, 3 and, 4 were analysed on an ELLA in comparison with the ECL. There is a small amount of binding for tCry1Ac peak 1 to GalNAc. Overall there was no significant activity for tCry1Ac to GalNAc.

4.9.3 Separation of tCry1Ac and the Catabolite Gene Activator using IMAC in the presence of a non ionic detergent Triton X-100.

As the other separation and purification protocols did not separate tCry1Ac or CryD3 it was thought that the catabolite gene activator may be binding to tCry1Ac through hydrophobic interactions. Triton X-100 is the most characteristic example of a non-ionic surfactant. Non-ionic detergents contain a headgroup that is uncharged and hydrophilic (see Figure 4.37). They are considered mild surfactants as they break protein-lipid, lipid-lipid but not protein-protein interactions and most of them do not denature proteins. Therefore, proteins are solubilised and isolated in their native and active form retaining their protein activity. This is their main advantage and preferred in isolation of membrane proteins. To test this theory tCry1Ac was expressed and purified using IMAC with the addition of 10% triton X100 in the lysis, wash and elution buffers. A significant reduction in protein yield was observed in the SDS PAGE gel (Figure 4.38) from the tCry1Ac compared to tCry1Ac with added Triton X100. There was a small band barely visible on the gel that corresponds to tCry1Ac. The eluted sample was examined on an ELLA (see Figure 4.39) but it did not show activity for tCry1Ac. Therefore purification by non ionic detergent was found not to yield positive results.

Figure: 4.37: The structure of Triton X-100. Adapted from G- Biosciences.

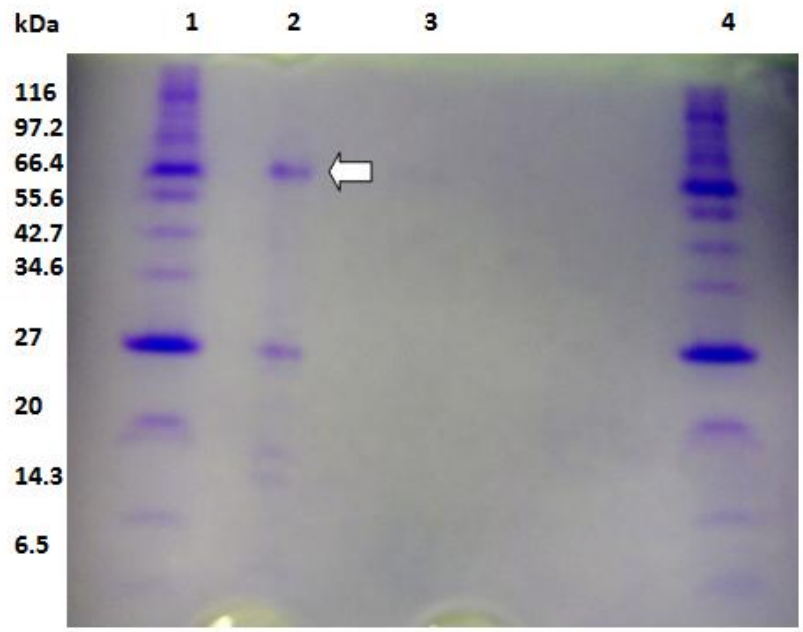


Figure 4.38: SDS PAGE analysis of tCry1Ac purified by IMAC in the presence of 10% triton X 100. Lane 1 and 4; Broad range protein marker, 2; tCry1Ac, 3 tCry1Ac with added Triton X100. The white arrow indicates tCry1Ac.

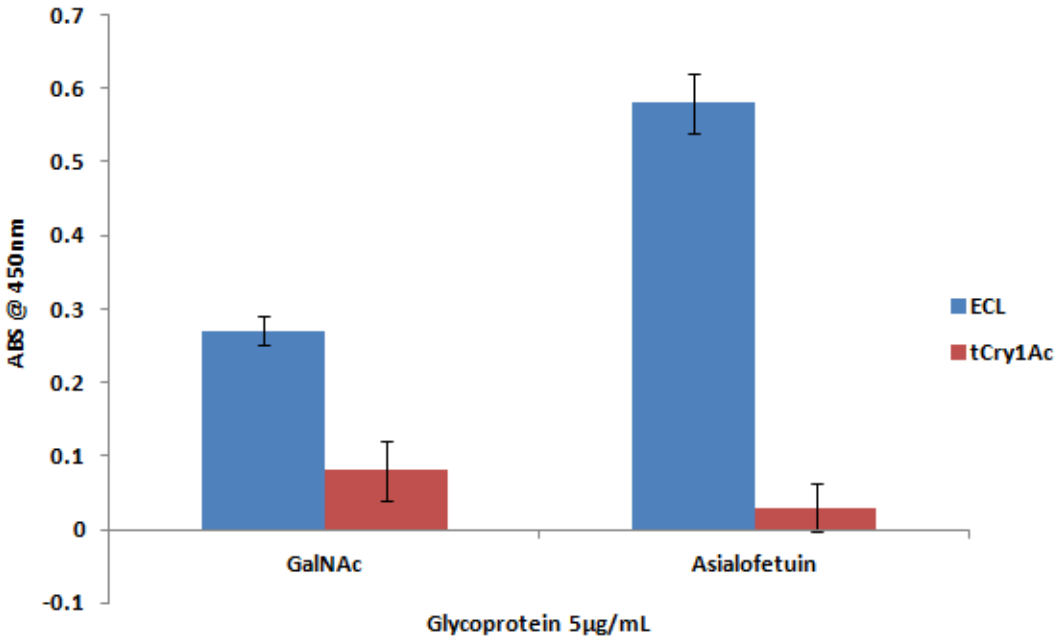


Figure 4.39: ELLA analysis of tCry1Ac, purified by IMAC with 10% Triton X100. Analysis of the binding activity of tCry1Ac purified by IMAC with added triton X100. No increase in binding was detected for tCry1Ac to GalNAc.

4.9.4 Separation of CryD3 and Catabolite Gene Activator by competing Galactose

As CryD3 has been shown to bind the sugar galactose, it was thought that adding free galactose during the IMAC washes may result in competition with CRP for the binding of CryD3. The experiment was carried out as described in section 2.13.7. Figure 4.40 shows a significant improvement in purification as the eluted fraction of CryD3 (after the galactose washes) revealed a single band for CryD3. However, there were some differences in the protein yield which means that the contaminant proteins may be potentially diluted. Therefore, the purification was repeated with and without galactose washes. Figure 4.41 shows the difference in contaminating proteins at approx 30kDa between CryD3 purified with 100mM galactose and CryD3 purified without galactose. Figure 4.42 shows the results from the ELLA analysis of the purified sample of CryD3. It is apparent from this graph that CryD3 is binding to asialofetuin and therefore is active. This would suggest therefore that CRP could be glycosylated. A lectin blot was carried out in order to detect any glycosylated protein in the sample.

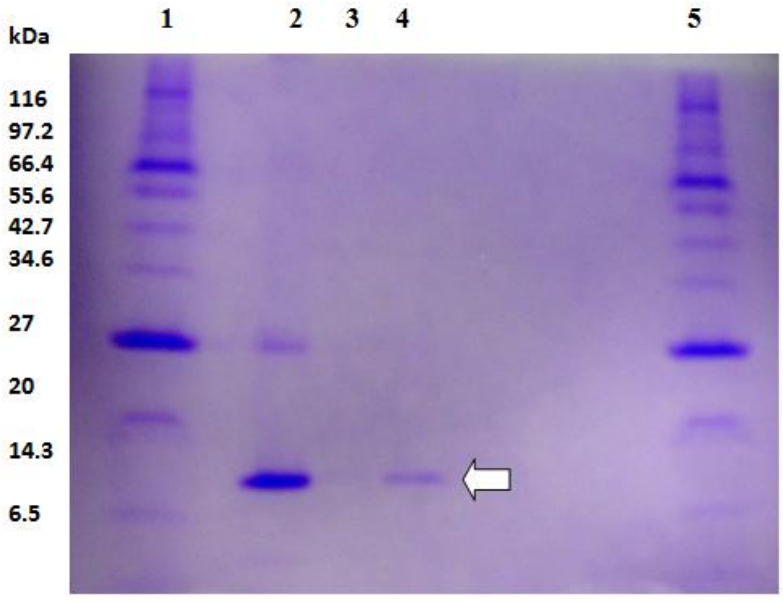


Figure 4.40: SDS PAGE analysis of CryD3 purification by IMAC with competing Galactose. Lane 1 and 5; Broad range protein marker, 2; CryD3 purified by IMAC, 3; wash with imidazole 100mM and galactose 100mM, 4; Elution of CryD3 with imidazole 100mM and galactose 100 mM. The white arrow indicates CryD3.

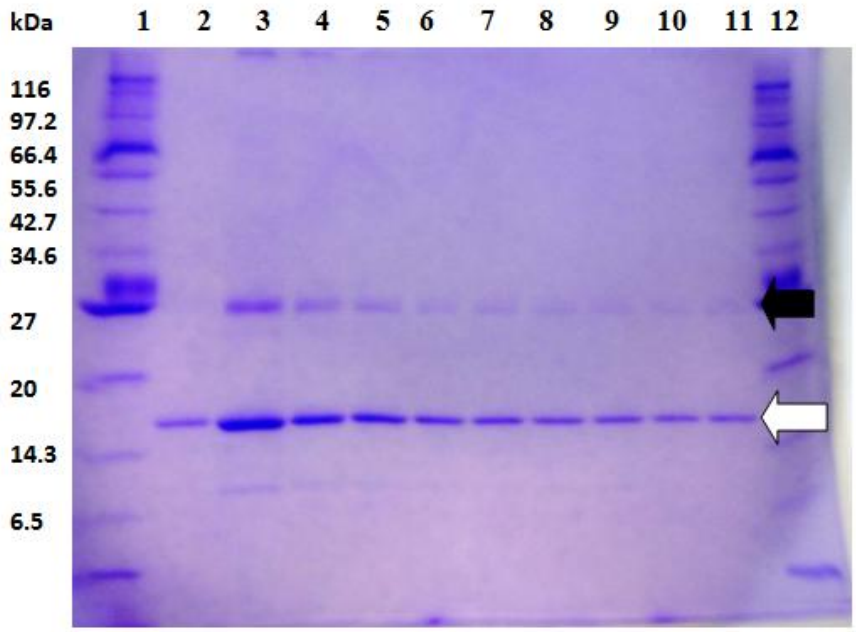


Figure 4.41: SDS PAGE analysis of CryD3 purification by IMAC with competing galactose compared to dilutions of CryD3 purified by IMAC with no Galactose. Lane 1 and 12; Broad range protein marker, 2; CryD3 purified by IMAC with galactose 100mM, 3-11; dilutions from 1/2 to 1/10 respectively of CryD3 purified by IMAC with no galactose. The white arrow indicates CryD3 while the black arrow indicates the contaminating protein.

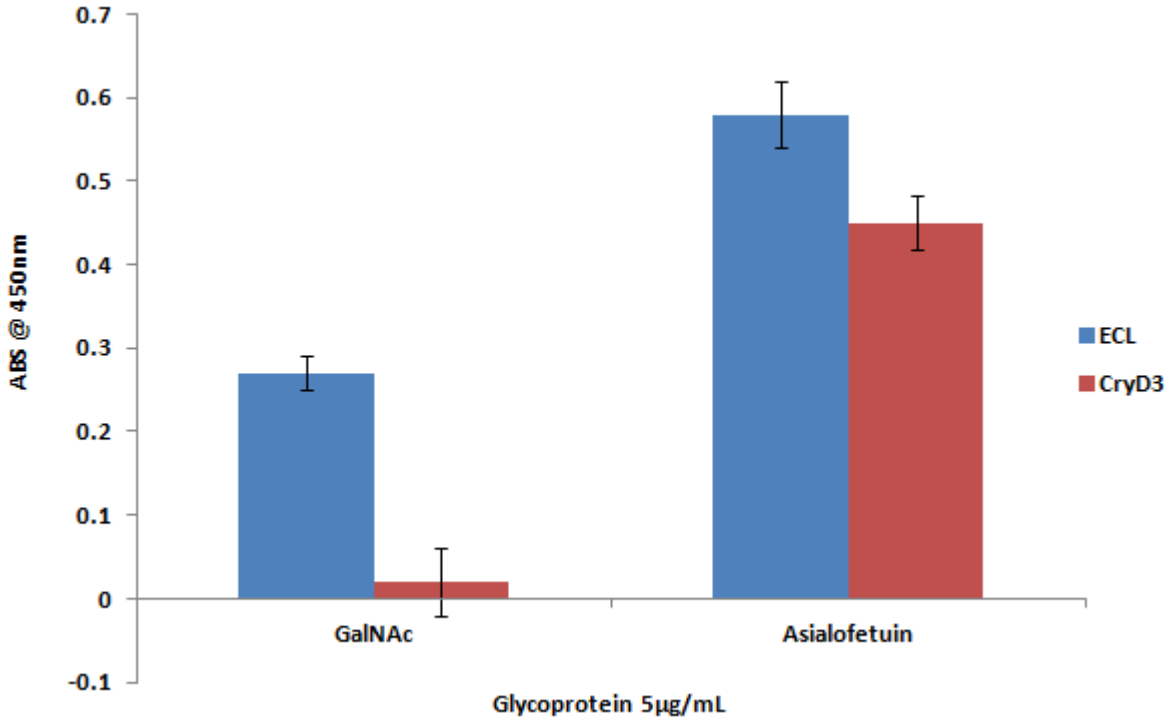


Figure 4.42: ELLA analysis of CryD3 purified by IMAC with added 100mM galactose washes. CryD3 was purified by IMAC and galactose was added to the washes with the resulting eluted protein analysed by ELLA. There was significant binding of CryD3 to asialofetuin. Interestingly, this correlation is related to the loss of the CRP protein.

4.10 Investigation of the possible glycosylation of the catabolic gene activator

As competitve galactose washes (during the IMAC purification of CryD3) were found to prevent the CRP from binding to CryD3 it was proposed that CRP may be glycosylated. Therefore, a lectin blot using ECL to detect galactose residues was carried out on the purified samples of tCry1Ac and CryD3 with CRP (See section 2.17). Figure 4.43 showed that no bands appeared on the lectin blot. While this would suggest that CRP may not be glycosylated, it is still interesting that galactose has the ability to separate CryD3 from CRP. Therefore there may be a type of glycosylation that does showing up on this type of lectin blot.

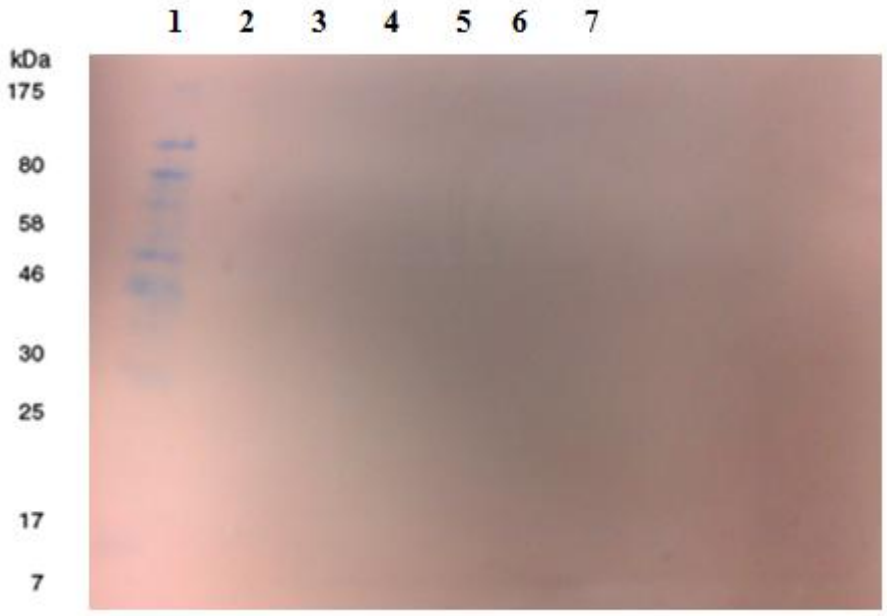


Figure 4.43: Lectin blot using ECL of purified samples of tCry1Ac and CryD3 with contaminating CRP protein. A lectin blot was carried out on the protein eluted from IMAC to possibly identify if CRP is glycosylated. 1; pre-stained protein ladder, 2-4; tCry1Ac eluted peaks, 5-7; CryD3 eluted peaks.

4.11 IMAC: Increasing imidazole washes for the removal of bound CRP to tCry1Ac and CryD3

tCry1Ac and CryD3 both bound very well to the nickel resin of the IMAC via the 6x histidine tag on their C terminus. Even at 300mM and 500mM imidazole elutions it took several fractions to elute the proteins. It was proposed that as tCry1Ac and CryD3 are over expressed in the host expression strain *E. coli* there should be a lot more of these proteins than the CRP protein. Therefore, IMAC purification was carried out in the presence of increasing imidazole washes ranging from 100mM, 120mM, 140mM 160mM, 180mM, 200mM and 300mM. It was postulated that the tCry1Ac-CRP or CryD3- CRP 'complexes' would be significantly larger than the unbound proteins and would therefore elute from the column first, leaving the unbound tCry1Ac and CryD3 to come off the column last (assuming saturation of CRP). Figure 4.44 presents some interesting results as it shows a decrease in the CRP and contaminating bands corresponding with an increasing amount of eluting imidazole.

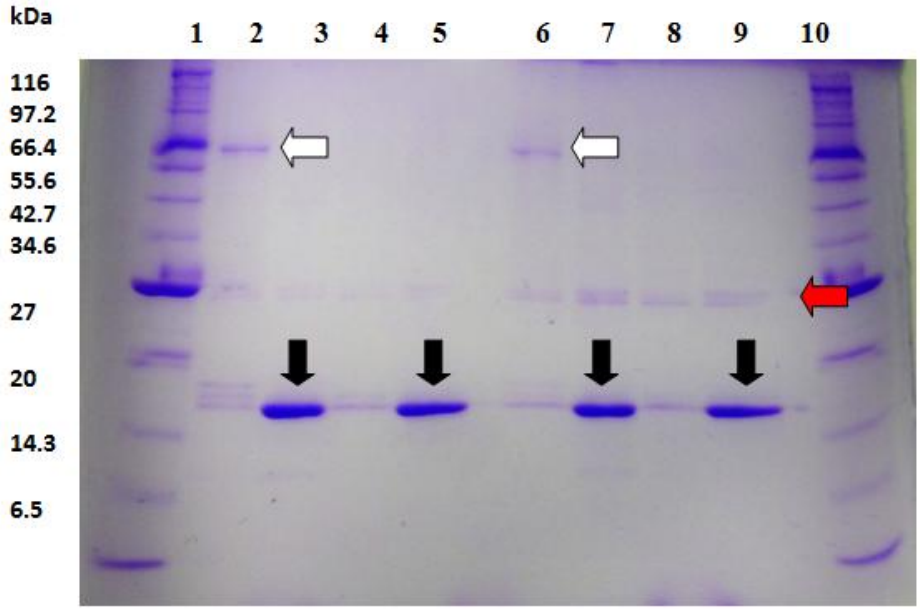


Figure 4.44: SDS PAGE analysis of the optimised IMAC purifications of tCry1Ac and CryD3 using higher imidazole elutions. tCry1Ac and CryD3 were purified and imidazole washes of 200mM were used before the proteins were eluted in 1mL fractions with 300mM imidazole. 1 and 10; Broad range protein marker, 2; tCry1Ac 300mM elution, 3 and 5; CryD3 300mM imidazole elution, 6; tCry1Ac 200mM imidazole wash, 7 and 9; 200mM imidazole wash. The white arrow represents tCry1Ac, the black arrow, CryD3 and the red arrow, the contaminating protein CRP.

4.11.1 Optimising protein purification by IMAC using combined increased imidazole and sugar washes

IMAC purification was carried out on tCry1Ac and CryD3 by combining both very high levels of imidazole washes with (a) a 100mM galactose wash for CryD3 and (b) a 100mM GalNAc wash for tCry1Ac. Figure 4.45 shows a single band for CryD3 and tCry1Ac, revealing that the combined method of increased imidazole washes and a galactose/GalNAc wash results in a highly purified protein with no bound contamination from *E. coli* proteins. The yield of the proteins was 1mg/ml for CryD3 and 500µg/ml for tCry1Ac. This degree of purity would allow the sugar specificities of tCry1Ac and CryD3 to be investigated more exactly.

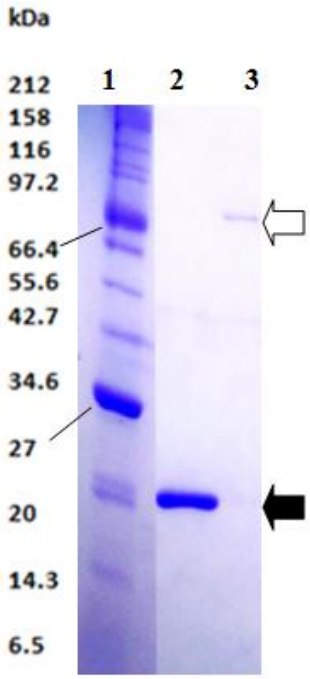


Figure 4.45: SDS PAGE analysis of tCry1Ac and CryD3 purified by IMAC combined with high imidazole and galactose/GalNAc 100mM washes. tCry1Ac, CryD3 elutions after purification by IMAC and galactose/GalNAc washes. **Lane 1**; broad range protein marker, **2**; CryD3 **3**; tCry1Ac. The white arrow indicates tCry1Ac and the black arrow indicates CryD3.

4.12 Investigation of glycan binding specificities of tCry1Ac and CryD3

ELLA analysis was carried out on the purified samples of tCry1Ac and CryD3 to examine their binding and relative affinities for different glycoproteins. Firstly they were tested against GalNAc-BSA, the Tn antigen and asialofetuin. Figure 4.46 shows the results of this ELLA analysis. tCry1Ac showed specificity for GalNAc and the Tn antigen, while CryD3 displayed specificity for galactose. This would seem to indicate that without the interference of CRP both of these proteins have active glycan binding. Dilutions of tCry1Ac and CryD3 binding to GalNAc –BSA and asialofetuin respectively were examined by ELLA analysis. In figure 4.47 tCry1Ac displayed a relatively good affinity for GalNAc. Figure 4.48 reveals that CryD3 has a slightly lower affinity for galactose, as there is a steep drop in signal with each dilution. This demonstrates that even though CryD3 binds to galactose specifically it does not have a high affinity for galactose and therefore does not bind very tightly. In figure 4.49 tCry1Ac and CryD3 were tested against GalNAc, galactose, sialic acid and GlcNAc. tCry1Ac and CryD3 showed no activity towards sialic acid or invertase. This would seem to indicate that tCry1Ac and CryD3 are selective for binding to GalNAc and galactose respectively.

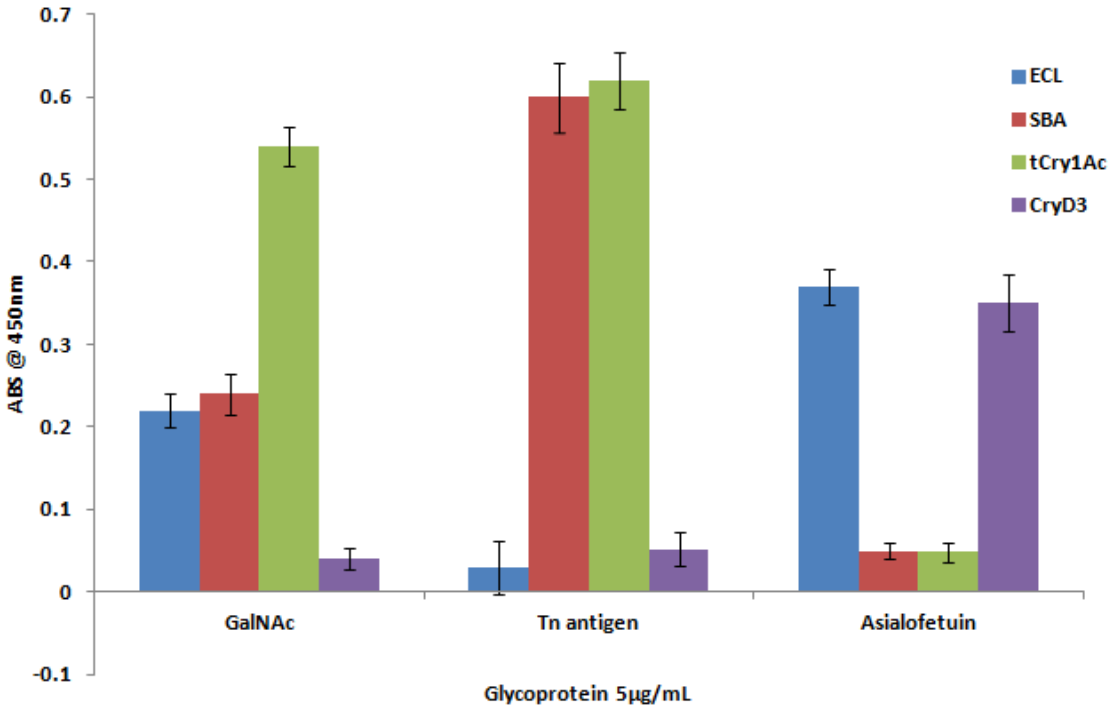


Figure 4.46: ELLA analysis of the purified tCry1Ac 10µg/ml and CryD3 10µg/ml compared to the commercial lectins SBA 10µg/ml and ECL 10µg/ml. Neoglycoprotein GalNAc-BSA and the Tn antigen were used along with the glycoprotein Asialofetuin. ELLA samples averaged in triplicate and ELLA experiment carried out three times. The positive controls, SBA binding to GalNAc and the Tn antigen while ECL binds to Asialofetuin and GalNAc. The graph shows the increased activity of tCry1Ac (green) and CryD3 (purple), which correlates to the purification of the proteins from CRP. Interestingly tCry1Ac (green) shows very strong binding to the Tn antigen and GalNAc, while CryD3 (purple) is specific to Asialofetuin.

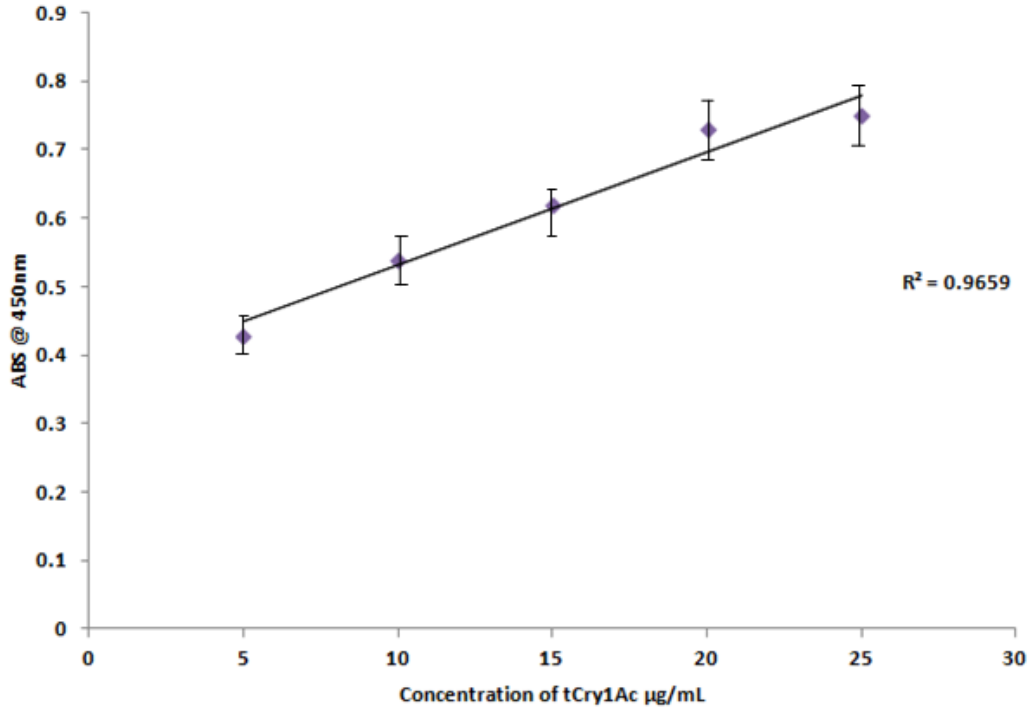


Figure 4.47: ELLA analysis of tCry1Ac affinity for GalNAc. A dilution series was done on tCry1Ac and analysed against GalNAc. ELLA samples averaged in triplicate and ELLA experiment carried out three times. tCry1Ac seems to have a good affinity for GalNAc as the slope of the line is gradual. tCry1Ac still binds with good specificity at 5µg/mL of tCry1Ac.

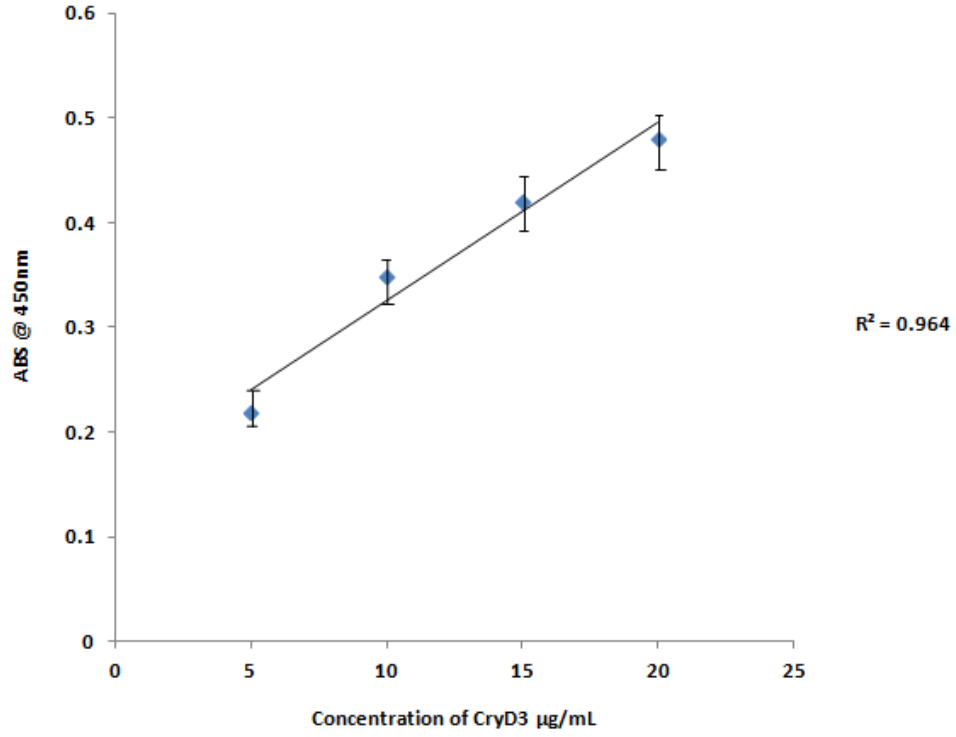


Figure 4.48: ELLA analysis of CryD3 affinity for galactose. Dilutions of CryD3 were assayed against 5µg/mL of asialofetuin to examine the affinity of CryD3 for galactose. ELLA samples averaged in triplicate and ELLA experiment carried out three times. The affinity was an improvement from figure 4.10. Nevertheless the affinity is still slightly weak as the signal drops quickly after each dilution.

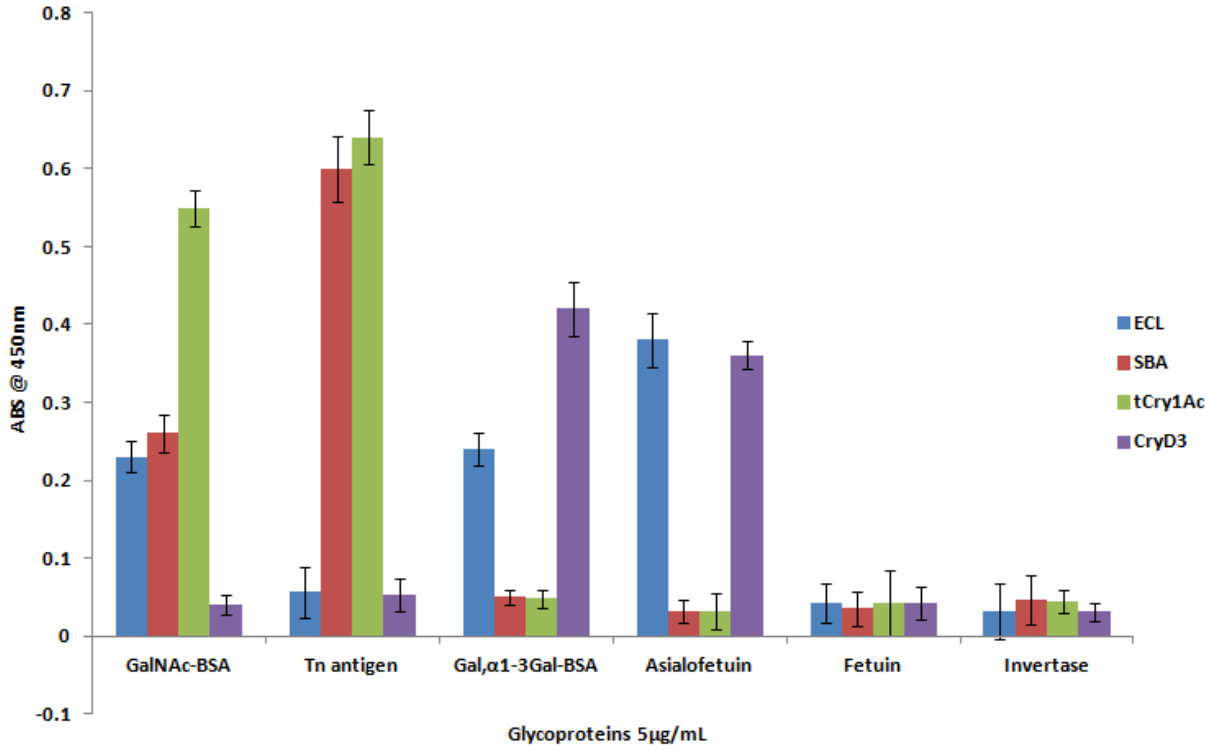


Figure 4.49: ELLA analysis of tCry1Ac and CryD3 specificity for various glycoproteins with different glycan epitopes. This ELLA analysed the binding specificity of tCry1Ac (green) and CryD3 (purple) to various glycoproteins which contained, GalNAc, galactose, sialic acid, and GlcNAc. ELLA samples averaged in triplicate and ELLA experiment carried out three times.

4.13 Discussion

tCry1Ac and CryD3 were successfully expressed, purified and separated from a persistently contaminating protein, namely the cAMP regulatory protein (CRP). The purified proteins where then characterised for their sugar binding properties, in particular their ability to bind to GalNAc and galactose. Extensive optimisation of tCry1Ac purification was essential due to the amounts of purified protein required for characterisation. The ability to produce large amounts of recombinant protein is also a pre-requisite if this molecule was to be incorporated onto an affinity or protein microarray formats. Optimisation of protein purification was also required for tCry1Ac and CryD3 as there was a persistent contaminating protein from *E. coli* in the eluted samples. The contaminating protein also seemed to block or inhibit the binding activity of tCry1Ac and CryD3 proteins.

A preliminary small scale expression and purification was carried on tCry1Ac and CryD3. The small scale expression revealed that tCry1Ac did not express well and therefore optimisation of expression was required. Contaminating proteins were observed in the purified sample of CryD3, and later in the optimised tCry1Ac expressions. This lead to a requirement of the optimisation of IMAC conditions for both proteins. The subsequent proteins were analysed on a preliminary ELLA to determine if tCry1Ac and CryD3 had sugar/glycan binding abilities. tCry1Ac was found to strongly bind to GalNAc while CryD3 was found to bind to galactose. This was a significant result as it showed that the Cry1Ac toxin could be cloned and expressed as a soluble toxin while still retaining its carbohydrate specificity. This finding supports previous research carried out on Cry1Ac, where it was found to have specificity for GalNAc (Xiang et al. 2009, Burton et al. 1999, Knowles, Knight, and Ellar 1991). Surprisingly CryD3 showed specificity to galactose. This finding has not been documented previously and shows that it is possible that domain three of Cry1Ac may be produced as a soluble protein and still retains carbohydrate specificity.

Following the successful expression of soluble tCry1Ac, expression optimisation was carried out as described in section 4.6. Two *E. coli* strains, KRX and JM109 were transformed with ptCry1Ac, to express C-terminally 6x histidine tagged tCry1Ac. The clones from each strain were submitted to expression analysis by SDS-PAGE, with JM109 selected as the most efficient clone for recombinant tCry1Ac expression. IPTG induction levels, incubation temperatures and growth media were all varied to optimise protein production.

As IPTG concentrations and incubation temperature influence cell growth rate, cultures may not be directly comparable, due to different cell densities. To combat this, all cultures were inoculated from the same starter culture and samples were also diluted to the same OD prior to cell harvest and analysis by SDS-PAGE. The effect of induction with varying levels of IPTG was investigated by looking at the amount of soluble and insoluble product produced at certain stages of expression. It was concluded that increasing the IPTG concentration did not result in a significant increase in levels of recombinant protein expressing in the soluble protein fraction and that dramatically increasing the IPTG concentration had an adverse affect on the cells growth rate.

The optimal temperature for expression was also investigated. It was determined that that optimal incubation at 20°C overnight, lead to an increased amount of protein per cell. A morning harvest time would also lend itself to immediate downstream processing, thus reducing any prolonged protein exposure to proteases.

The media chosen for protein expression was TB (Terrific broth) which would enable the cells to continue to grow and produce protein overnight, at the reduced temperature of 20°C.

Having evaluated the conditions for optimal expression of tCry1Ac from *E. coli* a purification optimisation was investigated for both tCry1Ac and CryD3. Due to the presence of the 6x histidine tag on the proteins C-terminus an IMAC purification strategy could be employed. Imidazole elution concentrations were investigated in order to establish an optimised strategy for IMAC purification of 6x histidine tagged tCry1Ac and CryD3. The optimised elution profile consisted of using a concentration of 40mM Imidazole in the binding buffer (wash buffer 1), followed by an intermediate wash step using 100 mM Imidazole (wash buffer 2), with a

concentration of 300 mM Imidazole used for the elution of tCry1Ac and CryD3 from the matrix. However, there were still some contaminants in the eluted samples of both proteins.

A western blot analysis was carried out in order to determine if the contaminating bands were potentially cleaved fragments of the histidine tagged proteins. The histidine results seemed to indicate this to be the case. ELLA analysis on the proteins showed no sugar binding activity. Therefore analysis on the stability of tCry1Ac and CryD3 was carried out with particular attention to protease inactivation/sensitivity and temperature denaturation. The host strain was compared to KRX *E. coli*, which is a protease resistant strain and protease inhibitor tablets were added to the cell lysate prior to cell disruption. This surprisingly resulted in more protein bands on the purification gels. Expression of protein was then carried out at 20°C with the actual purifications being carried out at 4°C as CryD3 aggregated and fell out of solution at temperatures between 30°C and 37°C. Analysis of the resulting purified protein showed less contaminating bands. However, there was still one prominent band that consistently showed up in both tCry1Ac and CryD3 samples. A western blot analysis showed a single band for tCry1Ac and CryD3. This suggested that a protein was being co-purified and was having an effect on the activity and stability of both proteins.

The 30kDa protein band was sent to be identified by MS. The results of this revealed that the protein was a cAMP regulatory protein from *E. coli* also known as the catabolite gene activator (CRP). It was not clear how CRP was being co-purified with tCry1Ac and CryD3. The next steps were to identify how this protein was being co-purified with these proteins, to ultimately separate it from both the tCry1Ac and CryD3 proteins.

Ion exchange and gel filtration chromatography were both found to not have any significant effect on the removal of the contaminant protein. In order to test the possibility that the binding of tCry1Ac to CRP may be hydrophobic in nature 10% TritonX100 was added during the IMAC purification of tCry1Ac. This resulted in a very low yield of protein and it was difficult to tell if the CRP protein had been separated or diluted. The resulting ELLA showed the tCry1Ac protein to have little or no binding activity for GalNAc.

It was hypothesized that as both tCry1Ac and CryD3 lost the ability to bind glycans when CRP is to them bound, then it could be binding to their sugar recognition site. Therefore CryD3 was purified by IMAC with the addition of 100mM galactose washes. The resulting gel showed a band for CryD3, with a lower yield than normal but with no CRP band present. The CRP was found to be present in the purification with no galactose washes but was not present in the purification with galactose washes. This would indicate that the galactose might be competing with CRP binding. ELLA analysis showed that the CryD3 purified by IMAC and galactose was active and bound to asialofetuin.

A lectin blot carried out on a sample containing CRP showed no glycosylation was present. A general feature of lectin binding sites is that they consist of a primary binding site that is capable of binding an individual monosaccharide unit, usually with a low affinity. This weak affinity is most often compensated for by the presence of multiple subunits (Loris 2002). Consequently, any disruption to the lectins quaternary structure may have a greater impact on the lectins ability to interact with its target molecule. Therefore even though CRP may not be glycosylated, how it binds to tCry1Ac and CryD3 has an effect on their sugar binding abilities.

Investigation of IMAC conditions was again analysed and it was found that if the tCry1Ac and CryD3 in particular are eluted in 300mM imidazole fractions, the larger proteins tCry1Ac –CRP complex and the CryD3-CRP complex elute off the column first. Therefore, these can be discarded and the ‘unbound’ tCry1Ac and CryD3 may be eluted. Thus when a combined purification was carried out with increased imidazole washes and GalNAc/galactose washes, it resulted in a purified protein for both tCry1Ac and CryD3.

The pure samples of tCry1Ac and CryD3 were then analysed on an ELLA, showing significant activity towards GalNAc and galactose respectively. Further ELLA analysis was carried out on these purified samples. These results indicate that tCry1Ac is specific for GalNAc and the Tn antigen while CryD3 is specific for galactose. These findings are consistent with those of the structural modelling and binding predictions carried out in chapter 3. These experiments confirm that tCry1Ac binds to GalNAc, while the third domain on its own binds to galactose. This provides new information that domain III of Cry1Ac binds to galactose when produced as a

soluble protein. Previous studies have not characterised the third domain and this study shows that it is possible to produce a soluble active protein, while also showing that the third domain can have altered carbohydrate specificity. (Xiang et al. 2009, Burton et al. 1999, Knowles, Knight, and Ellar 1991).

Chapter 5

Site Directed Mutagenesis of tCry1Ac and CryD3

5.1 Overview

This chapter describes the site directed mutagenesis of tCry1Ac and CryD3 in both the known and predicted binding sites for the substrates GalNAc and galactose respectively. Alanine scanning was used to examine the effect these amino acids have on the protein structure and function. The effects of altering the amino acids of the binding sites on specificity and affinity were then studied. The structural models of the proteins were examined by PyMol to observe the effect mutagenesis had on each protein. Whole vector amplification of tCry1Ac and CryD3 was used to introduce the mutations. The resulting proteins were named, $tCry1Ac_{N480A}$, $tCry1Ac_{QY487A}$, $tCry1Ac_{N521A}$, $CryD3_{N45A}$, $CryD3_{QY52A}$, $CryD3_{N86A}$, $CryD3_{PAV16A}$. Small scale protein expression and IMAC purification was carried out for each protein. Contaminating proteins were still an issue during purification. An ELLA investigating the sugar binding abilities of each protein was carried out. $CryD3_{N86A}$ was the only protein that demonstrated significant biological activity. Therefore $CryD3_{N86A}$ was selected to analyse further. $CryD3_{N85A}$ was purified to a high degree and displayed significant binding to GalNAc with residual binding to galactose. This finding could suggest that the binding site for GalNAc and galactose in tCry1Ac and CryD3 is located in the same region but altered due to changes in protein folding when CryD3 is isolated.

5.2 Site directed mutagenesis of tCry1Ac and CryD3

tCry1Ac and the lectin domain CryD3 both display specificity for oligosaccharides tCry1Ac as expected showed specificity to GalNAc, while CryD3 bound to galactose. This may be because of a different binding site or an alteration in the binding region due to protein folding. Therefore alanine mutations on the known binding regions and expected binding regions were carried out to examine the binding specificities of both proteins. From previous studies on Cry1Ac sites N506, Q509 and Y513 and N546 were shown to have an essential role in binding (Xiang et al. 2009; Burton et al. 1999). These sites correspond to N480, Q483, Y487 and N521 in tCry1Ac and N45, Q48, Y52 and, N86 in CryD3. From I-Tasser the predicted galactose binding site for CryD3 was around the region, P14, A15 and V16.

Alanine scanning is the most commonly used technique to map functional epitopes (Moreira et al. 2007, Cunningham and Wells, 1989, Morrison and Weiss 2001, Wells, 1990). Alanine substitution nullifies the amino acid side chain without altering the main-chain conformation, nor does it impost extreme electrostatic or steric effects (Cunningham and Wells, 1989)

Alanine residues were introduced to tCry1Ac and CryD3 via site directed mutagenesis using Velocity tac polymerase whole vector amplification as described in section 2.7 (see Figure 5.1). The residues that were targeted for mutagenesis were the known binding residues of tCry1Ac N480, Q483, Y487 and N521 and the respective residues in CryD3; N45, Q48, Y52 and, N86. (See section 3.4 and 3.5). Protein modelling showed the expected binding region of CryD3 to involve residues P14, A15 and V16 (Figure 3.15). Therefore a double mutant of PAV was altered to AAA.

The resulting genes were named *tcry1ac*$_{N480A}$, *tcry1ac*$_{QY487A}$, *tcry1ac*$_{N521A}$, *cryc3*$_{N45A}$, *cryd3*$_{QY52A}$, *cryd3*$_{N86A}$, *cryd3*$_{PAV16A}$. The resulting proteins were named, tCry1Ac$_{N480A}$, tCry1Ac$_{QY487A}$, tCry1Ac$_{N521A}$, CryD3$_{N45A}$, CryD3$_{QY52A}$, CryD3$_{N86A}$, CryD3$_{PAV16A}$ respectively. The construction of the mutants is outlined in Figures 5.2-5.5. The predicted impact to the protein structure is shown in Figures 5.6 to 5.10. The model that displayed the most promising results was CryD3$_{N86A}$ in Figure 5.9. Figure 5.9 showed that when CryD3 is mutated from N86A, it opens up a possible binding site in CryD3.

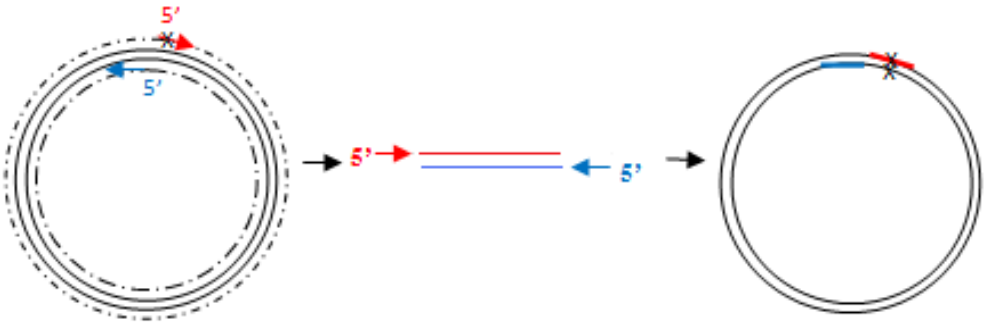


Figure 5.1: Schematic of Velocity whole vector amplification site directed mutagenesis. Two primers are designed to amplify the entire expression plasmid, one of which contains the desired mutation. Following PCR (Section 2.7.1) a DpnI restriction digest can be carried out to restrict any of the methylated template DNA. The phosphorylated ends are then ligated. The ligation is subsequently transformed directly into an *E. coli* expression strain (Section 2.6). For the mutation of tCry1Ac the template plasmid was ptCry1Ac while CryD3 was pCryD3.

```
AATAATATAATTGCATCGGATAGTATTACTCAAATCCCTGCAGTGAAGGGAAACTTTCTTTTTAATGG
                          TACTGGTGGGGACTTAGTTAGATTAAATAGTAGTGGAGCTAACATT
TTCTGTAATTTCAGGACCAGGATTTACTGGTGGGGACTTAGTTAGATTAAATAGTAGTGGAAATAACATT
CAGAATAGAGGGTATATTGAAGTTCCAATTCACTTCCCATCGACATCTACCAGATATCGAGTTCGTGTAC
GGTATGCTTCTGTAACCCCGATTCACCTCAACGTTAATTGGGGTAATTCATCCATTTTTTCCAATACAGT
ACCAGCTACAGCTACGTCATTAGATAATCTACAATCAAGTGATTTTGGTTATTTTGAAAGTGCCAATGCT
TTTACATCTTCATTAGGTAATATAGTAGGTGTTAGAAATTTTAGTGGGACTGCAGGAGTGATAATAGACA
GATTTGAATTTATT
```

Figure 5.2: Outline of construction of tCry1Ac$_{N480A}$ and CryD3$_{N45A}$. Representation of the position of the primers used for the mutation of tCry1Ac$_{N480A}$ and CryD3$_{N45A}$, Mut480f (blue), Mut480r (red) (Table 2.3) within the tCry1Ac and CryD3 sequence which creates the mutation N480A and N45A respectively. The nucleotides to be mutated are highlighted in yellow, with the nucleotides to be introduced on each primer underlined in bold.

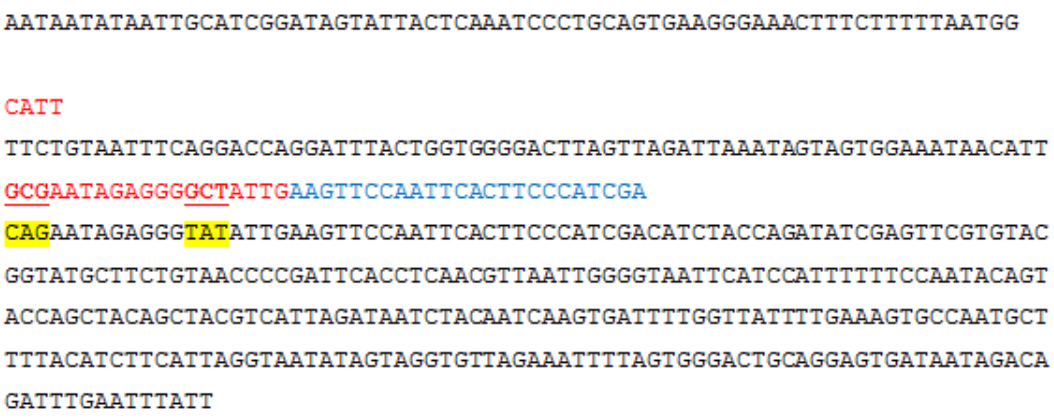


Figure 5.3: Outline of construction of tCry1Ac$_{QY487A}$ and CryD3$_{QY52A}$. Representation of the position of the primers used for the mutations of tCry1Ac$_{QY487A}$ and CryD3$_{QY52A}$, Mut487f (blue), Mut487r (red) (Table 2.3) within the tCry1Ac and CryD3 sequence which creates the mutations Q483A, Y487A and Q48A, Y52A respectively. The nucleotides to be mutated are highlighted in yellow, with the nucleotides to be introduced on each primer underlined in bold.

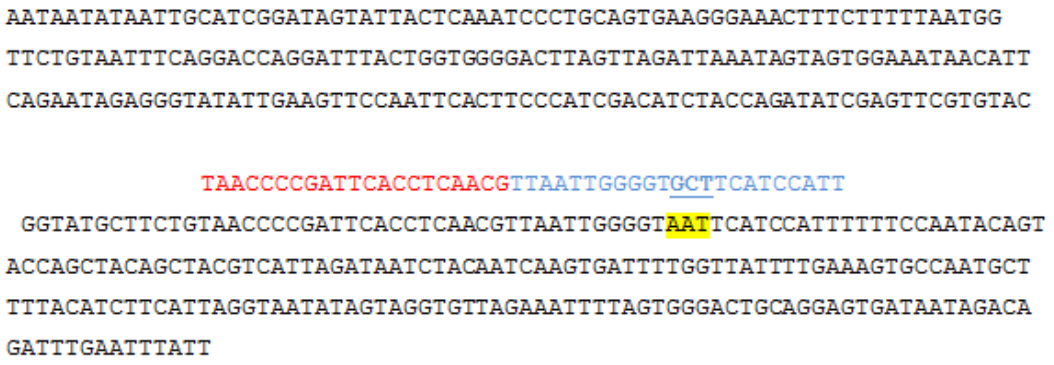


Figure 5.4: Outline of construction of tCry1Ac$_{N521A}$ and CryD3$_{N86A}$. Representation of the position of the primers used for the mutation of tCry1Ac$_{N521A}$ and CryD3$_{N86A}$, Mut521f (blue), Mut521r (red) (Table 2.3) within the tCry1Ac and CryD3 sequence which creates the mutation N521A and N86A respectively. The nucleotides to be mutated are highlighted in yellow, with the nucleotides to be introduced on each primer underlined in bold.

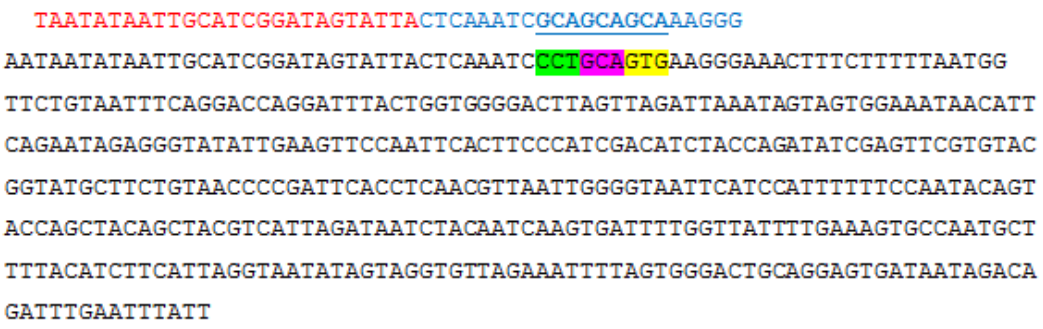


Figure 5.5: Outline of construction of CryD3$_{PAV16A}$. Representation of the position of the primers used for the mutation of CryD3$_{PAV16A}$, MutPAVf (blue), MutPAVr (red) (Table 2.3) within the CryD3 sequence which creates the mutation P14A, A15 and V16A. The nucleotides to be mutated are highlighted in yellow and green, the pink is the highlighted alanine, with the nucleotides to be introduced on each primer underlined in bold.

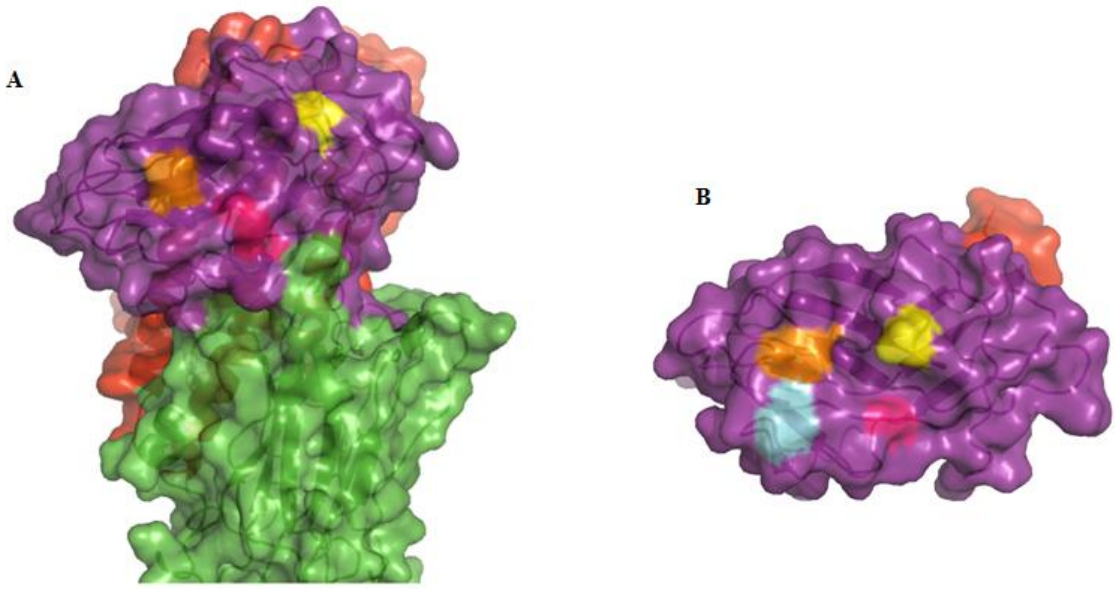


Figure 5.6: PyMol 3D Protein models of tCry1Ac (partial view) and CryD3 showing expected binding site of GalNAc and galactose. **A**) tCry1Ac, domain I in red, domain II in green and domain III in purple with highlighted residues believed to be involved in GalNAc binding, N480- orange, Q483-cyan (buried in this structure), Y487 – pink, N521- yellow. **B**) CryD3 in purple with highlighted residues that relate to tCry1Ac, N45 – orange, Q48 – cyan, Y52 – pink, N86 – yellow. The 6 x histidine tag in both proteins is red. These protein structures are from the same sequence but some structural differences are predicted. The residues that are highlighted are located in different positions in the different proteins. In tCry1Ac, the domains I and II may be supporting domain III and cause its structure to change slightly.

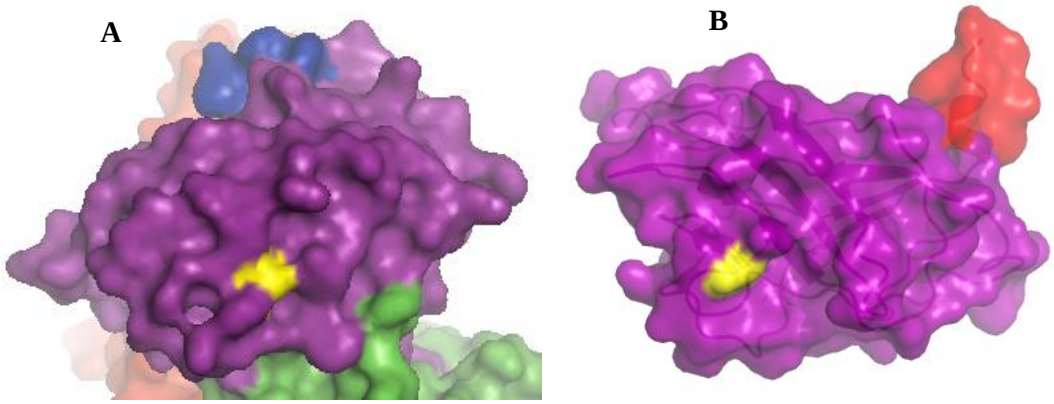


Figure 5.7: PyMol Protein models of tCry1Ac$_{N480A}$ (partial view) and CryD3$_{N45A}$. (A) tCry1Ac$_{N480A}$ with highlighted residue in yellow that was mutated to alanine, domain III; purple, domain II; green, 6x histidine tag; blue. (B) CryD3$_{N45A}$ with highlighted residue in yellow that was mutated to alanine, 6x histidine tag highlighted in red.

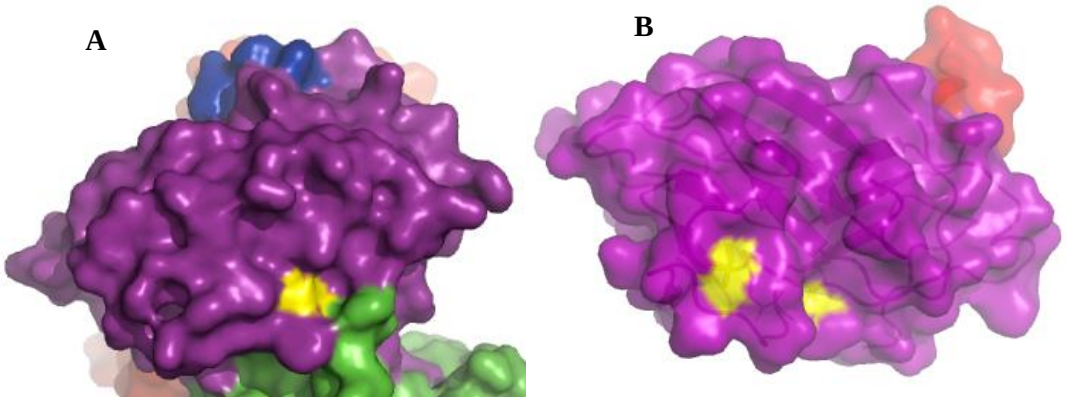


Figure 5.8: PyMol Protein models of tCry1Ac$_{QY487A}$ (partial view) and CryD3$_{QY52A}$. (A) tCry1Ac$_{QY487A}$ with highlighted residues in yellow that were mutated to alanine, domain III; purple, domain II; green, 6x histidine tag; blue. (B) CryD3$_{QY52A}$ with residues that were mutated to alanine in yellow, 6x histidine tag highlighted in red.

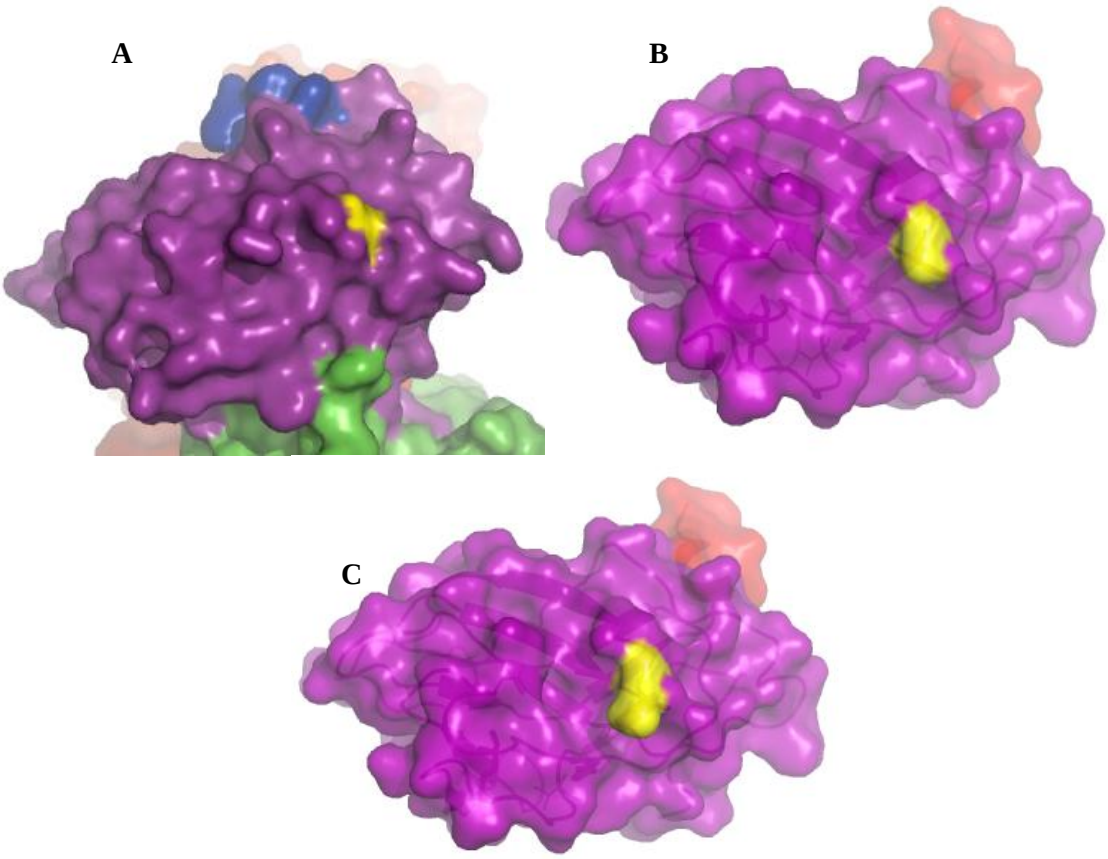


Figure 5.9: PyMol Protein models of tCry1Ac$_{N521A}$ (partial view), CryD3$_{N86A}$. tCry1Ac$_{N521A}$, CryD3$_{N86A}$ and CryD3 with mutations to alanine in yellow. (A) tCry1Ac$_{N521A}$ showing a partial view of tCry1Ac$_{N521A}$ with the third domain in purple, histidine tag in blue and N521A change in yellow. (B) CryD3$_{N86A}$ with histidine tag in red and mutation to alanine in yellow. (C) CryD3 with histidine tag in red and N86 in yellow. This shows that when CryD3 is mutated N86A, it opens up a possible binding site in the CryD3.

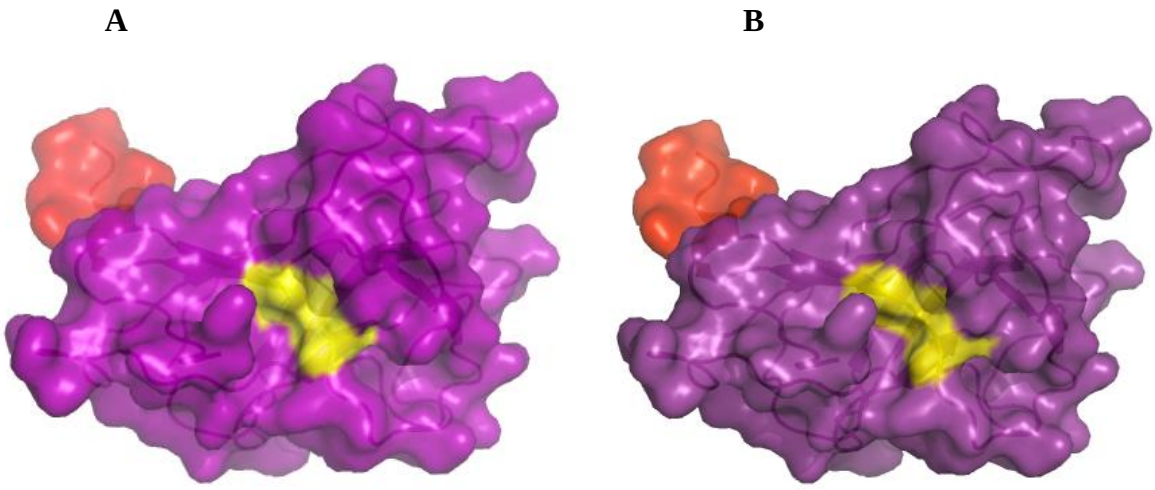


Figure 5.10: PyMol Protein model of CryD3$_{PAV16A}$ and CryD3. (A) CryD3$_{PAV16A}$ with mutations to alanine in yellow, 6x histidine tag highlighted in red. (B) CryD3 with PAV highlighted in yellow and 6x histidine tag highlighted in red.

5.3 Preliminary expression and purification of mutants of tCry1Ac and CryD3

Following successful construction of the tCry1Ac and CryD3 mutants, each plasmid was verified by sequencing and transformed into the *E. coli* JM109 expression strain. The optimised expression and purification by IMAC strategies for tCry1Ac and CryD3 were used for all derivatives. The only difference was that galactose was not added in the preliminary purification to examine CRP binding to the mutants. This was to examine if there was any difference in contaminating bands compared to tCry1Ac and CryD3. The SDS PAGE gels showing the protein expressions are presented in Figure 5.11. They show low expression yields for tCry1Ac mutants, while CryD3 mutants expressed well, especially $CryD3_{N86A}$ and $CryD3_{PAV16A}$**.** Following this the producing cells were lysed by cell distruption and the mutant proteins purified by IMAC. The results obtained from the preliminary purification of the mutant proteins are presented in Figures 5.12, 5.13 and 5.14 respectively. While the gels show the purified proteins, there are several contaminating bands including a large band at 30kDa which correlates to the CRP protein from *E. coli*.

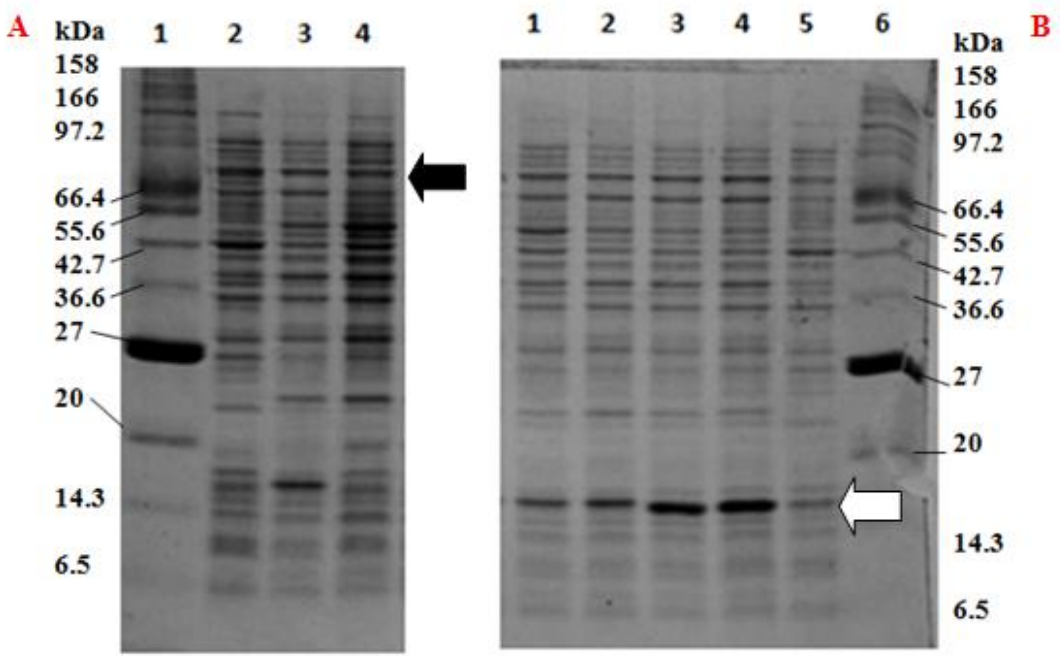


Figure 5.11: SDS PAGE gels of small scale protein expression of tCry1Ac and CryD3 mutants. The mutants; tCry1Ac$_{N480A}$, tCry1Ac$_{QY487A}$, tCry1Ac$_{N521A}$, CryD3$_{N45A}$, CryD3$_{QY52A}$, CryD3$_{N86A}$, CryD3$_{PAV16A}$ were expressed in *E. coli* JM109 and the total protein analysed on a SDS PAGE gel.

(A) Expression of tCry1Ac mutants, **1**; broad range protein marker, **2**; tCry1Ac$_{N480A,}$ **3**; tCry1Ac$_{QY487A}$, **4**; tCry1Ac$_{N521A}$. The black arrow indicates tCry1Ac.

(B) Expression of CryD3 mutants, **1**; CryD3$_{N45A}$, **2**; CryD3$_{QY52A}$, **3**; CryD3$_{N86A}$, **4**; CryD3$_{PAV16A,}$ **5**; non induced CryD3, **6**; Broad range protein marker. The white arrow indicates CryD3.

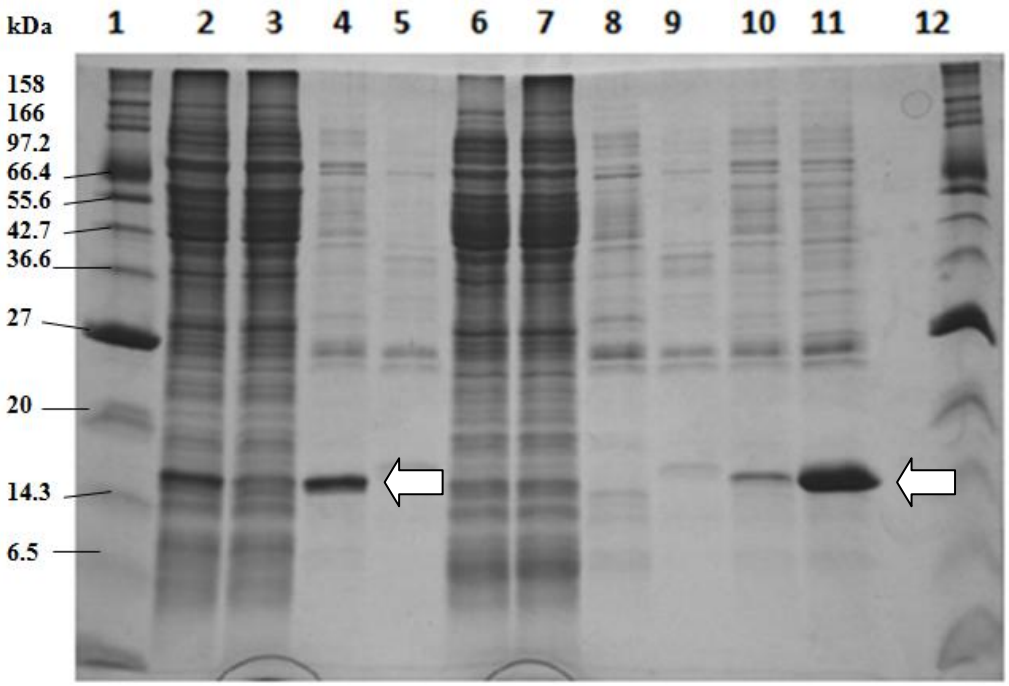


Figure 5.12: SDS PAGE gels of the preliminary purification of CryD3$_{N45A}$ and CryD3$_{QY52A}$. IMAC purifications on CryD3$_{N45A}$ and CryD3$_{QY52A}$ reveal strong expression for the mutants but there are a large number of contaminating bands in the purifications. **1** and **12**; Broad range protein marker, **2**; CryD3$_{N45A}$ lysate, **3**; CryD3$_{N45A}$ flow through, **4** and **5**; CryD3$_{N45A}$ elutions, **6**; CryD3$_{QY52A}$ lysate, **7**; CryD3$_{QY52A}$ flow through, 8-10; imidazole washes, **11**; CryD3$_{QY52A}$ elution. White arrows indicate CryD3$_{N45A}$ and CryD3$_{QY52A}$.

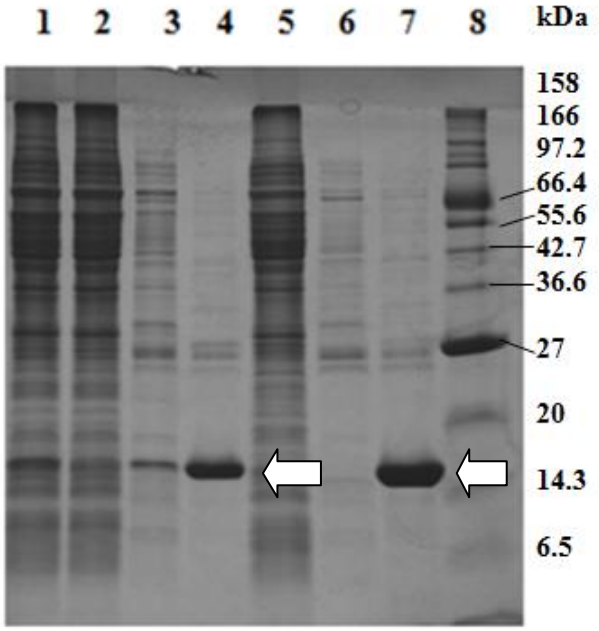


Figure 5.13: SDS PAGE gels of the preliminary purifications of CryD3$_{N86A}$ and CryD3$_{PAV16A}$. IMAC purifications on CryD3$_{N86A}$ and CryD3$_{PAV16A}$ reveal strong expression for the mutants but there are a large number of contaminating bands in the purifications. **1**; CryD3$_{N86A}$ lysate, **2**; CryD3$_{N86A}$ flow through, **3**; CryD3$_{N86A}$ imidazole wash, **4**; CryD3$_{N86A}$ elution, **5**; CryD3$_{PAV16A}$ flow through, **6**; CryD3$_{PAV16A}$ imidazole wash, **7**; CryD3$_{PAV16A}$ elution, **8**; Broad range protein markers. White arrows indicate CryD3$_{N86A}$ and CryD3$_{PAV16A}$.

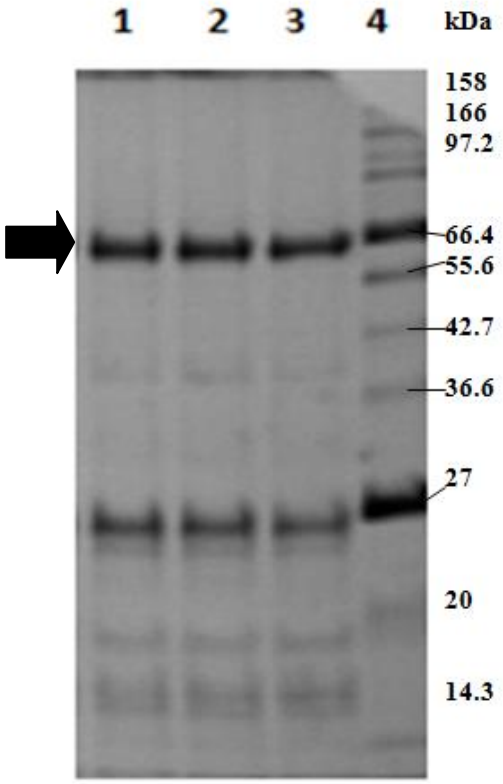


Figure 5.14: SDS PAGE gel of the preliminary purifications of tCry1Ac$_{N480A}$, tCry1Ac$_{QY487A}$, and tCry1Ac$_{N521A}$. IMAC purifications on tCry1Ac$_{N480A}$, tCry1Ac$_{QY487A}$, tCry1Ac$_{N521A,}$ show expression at 69kDa, with contaminating bands again at approx 30kDa and lower. **1**; tCry1Ac$_{N480A}$ elution, **2**; tCry1Ac$_{QY487A}$ elution, **3**; tCry1Ac$_{N521A}$ elution, **4**; broad range protein marker. Black arrow indicates tCry1Ac$_{N480A}$, tCry1Ac$_{QY487A}$, and tCry1Ac$_{N521A}$

5.4 Investigating glycan binding of tCry1Ac and CryD3 mutants

To examine if either tCry1Ac or CryD3 mutants had specficity towards glycans on glycoproteins an ELLA was carried out as described in section 2.19. As can be seen from Figure 5.15, CryD3$_{N86}$ was the only mutant that displayed binding activity to GalNAc. This is significant as CryD3 previously had specificity to galactose. It was therefore decided to repeat the IMAC purification of CryD3$_{N86}$ with the addition of galacotse washes to produce a purer prepation of CryD3$_{N86}$ with no contaminating proteins.

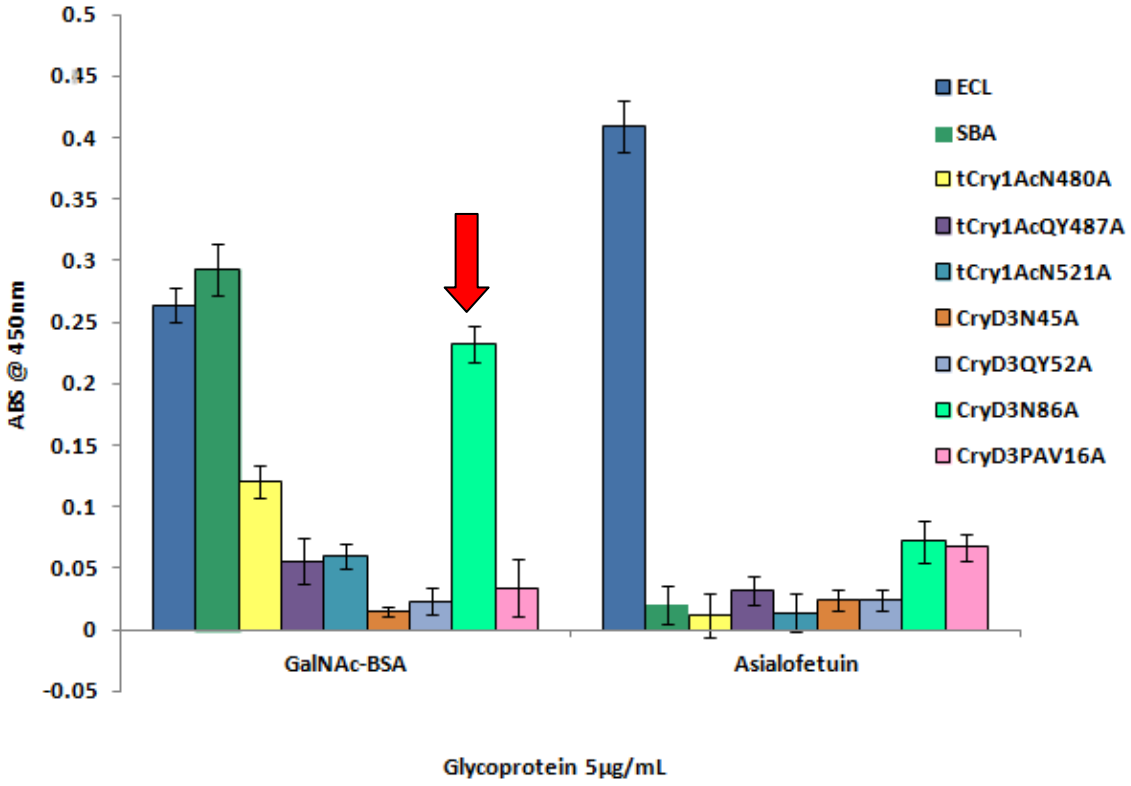


Figure 5.15: Preliminary ELLA on purified mutants from tCry1Ac and CryD3 compared to commercial lectins ECL and SBA. ELLA on purified mutants from tCry1Ac and CryD3 compared to commercial lectins ECL and SBA. CryD3$_{N86A}$ showed specificity to GalNAc, while tCry1Ac$_{N480A}$ showed a small amount of specificity to GalNAc also. The red arrow indicates CryD3$_{N86A}$ binding to GalNAc-BSA.

5.5 Optimised purification of CryD3$_{N86A}$ by IMAC incorporating galactose washes

As described previously the initial purification of CryD3$_{N86A}$ contained contaminating protein bands and its specificity to GalNAc was low. Therefore CryD3$_{N86A}$ was purified by IMAC purification with the addition of galactose washes as described in section 2.13.7. This was expected to help purify CryD3$_{N86A}$, as the galactose would compete for the binding of CryD3$_{N86A}$. This was shown to be sucessful for the purification of tCry1Ac and CryD3 with additional washes of GalNAc and galactose. Figure 5.16 shows a significant improvement of the purification with a single band for both CryD3$_{N86A}$ and CryD3 seen. This result also demonstrated that the purification method, incorporating imidazole and galactose washes, used for CryD3 also works for the protein CryD3$_{N86A.}$ These purified protein samples were both then re-examined on an ELLA.

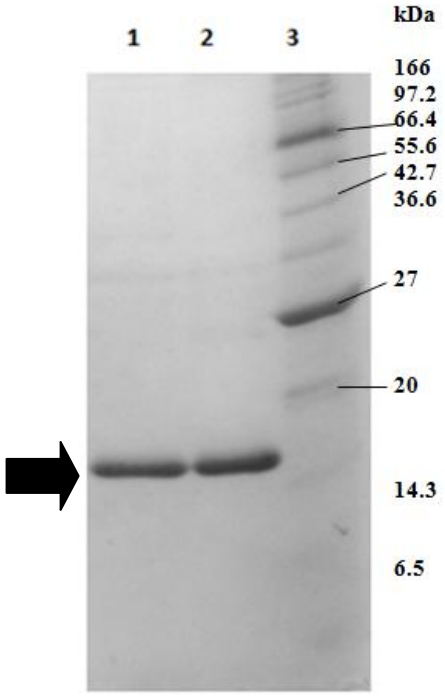


Figure 5.16: SDS PAGE of IMAC and galactose purification of CryD3 and CryD3$_{N86A}$. CryD3 and CryD3$_{N86A}$ were purified by IMAC and galactose washes resulting in a single protein band for CryD3 and CryD3$_{N85A}$. **1**; CryD3 elution, **2**; CryD3$_{N86A}$ elution, **3**; broad range protein marker. Arrow indicates CryD3 and CryD3$_{N86A.}$

5.6 Investigation of CryD3$_{N86A}$ specificity for glycoproteins

CryD3$_{N86A}$ had previously shown binding to GalNAc in the ELLA (see figure 5.15). However, as this protein also contained contaminating proteins it was decided to optimise the purification. This purer sample of CryD3$_{N86A}$ was retested by ELLA. As can be seen in Figure 5.17 it is clear that CryD3$_{N86A}$ bound to GalNAc and asialofetuin (galactose). This contrasts with CryD3 which had specificity for galactose only. It is apparent that one change N86A may have a significant effect on the binding specificity of the protein. These results seem to indicate that the binding site for tCry1Ac and CryD3 could be very similar. However, slight alterations of the protein may have an effect on its specificity towards sugars. CryD3$_{N86A}$ continues to display relatively good affinity towards GalNAc-BSA as it becomes progressively diluted and examined by ELLA (Figure 5.18). Even at low concentrations (5µg/ mL and 2.5µg/ mL) CryD3$_{N86A}$ still displayed specificity to GalNAc.

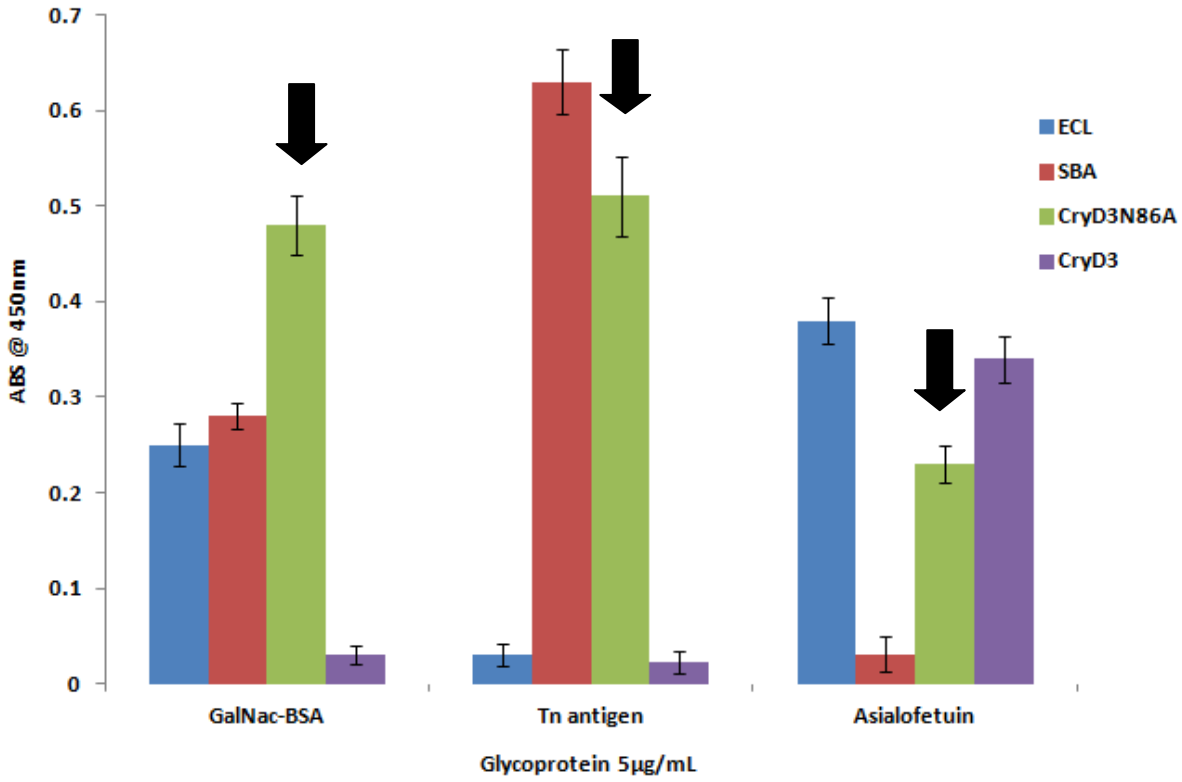


Figure 5.17: ELLA of CryD3 and CryD3$_{N86A}$ compared to commercial lectins ECL and SBA. An ELLA was carried out on the purified samples of CryD3 and CryD3$_{N86A}$ to examine and compare their specificities. CryD3$_{N86A}$ (in green) has altered binding specificity to CryD3. It now binds to GalNAc-BSA, the Tn antigen and has some weaker binding to asialofetuin. The arrows indicate the activity of CryD3$_{N86A}$.

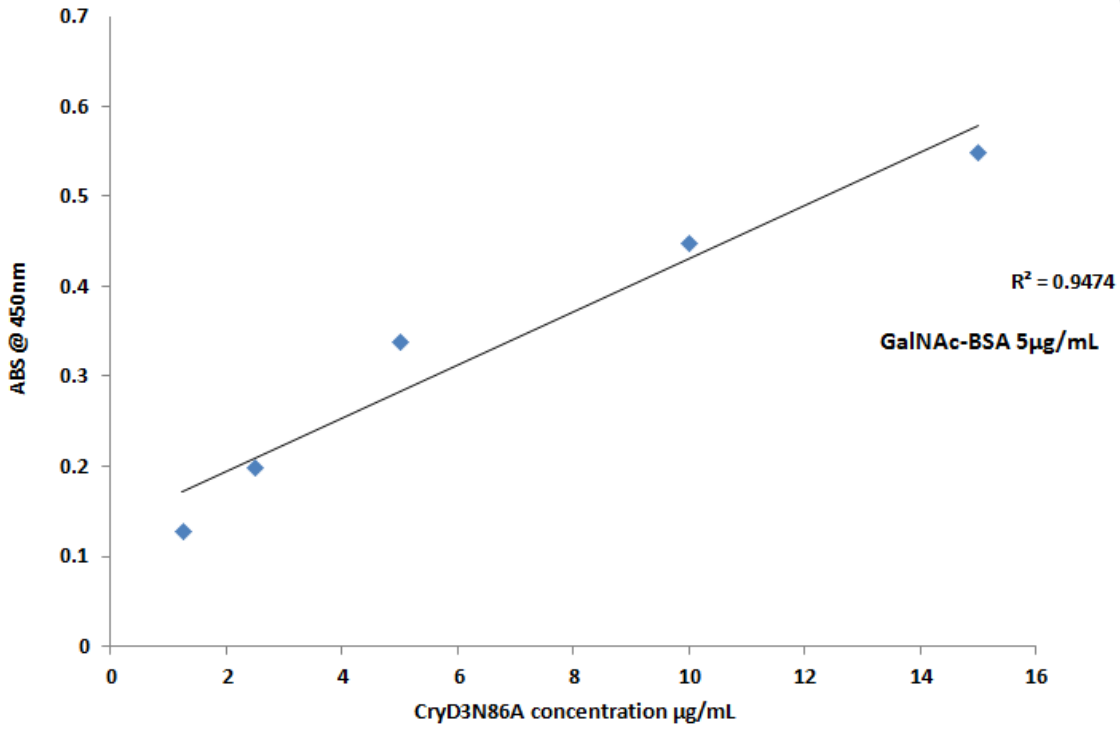


Figure 5.18: The effect of CryD3$_{N86A}$ dilutions on the binding to GalNAc-BSA (5µg/mL) analysed by ELLA. CryD3$_{N86A}$ concentrations 15µg/mL, 10µg/mL, 5µg/mL, 2.5µg/mL, 1.25µg/mL were assayed against GalNAc-BSA 5µg/mL to examine the affinity of CryD3$_{N86A}$ to GalNAc by ELLA as described in section 2.19.

5.7 Discussion

The mutants of tCry1Ac and CryD3; $tCry1Ac_{N480A}$, $tCry1Ac_{QY487A}$, $tCry1Ac_{N521A}$, $CryD3_{N45A}$, $CryD3_{QY52A}$, $CryD3_{N86A}$, $CryD3_{PAV16A}$ were successfully produced , cloned and expressed in *E. coli* JM109. The initial purifications showed each mutant with contaminating proteins in their preparations. This indicates that none of these mutations had an effect on the binding of CRP to domain III. This may suggest that the CRP is not "binding" to domain III near the GalNAc binding site; however as shown in chapter 4 the introduction of either GalNAc or galactose in the IMAC purifications eliminates CRP from the protein sample.

ELLA analysis on the mutant proteins revealed that only $CryD3_{N86A}$ displayed sugar-binding activity. Unfortunately, $tCry1Ac_{N480A}$, $tCry1Ac_{QY487A}$, $tCry1Ac_{N521A}$, $CryD3_{N45A}$, $CryD3_{QY52A}$, and $CryD3_{PAV16A}$ did not show any significant results in relation to sugar binding. Due to time limitations these proteins were not investigated further.

The purification of $CryD3_{N86A}$ was optimised by the incorporation of galactose washes, which produced pure active protein. $CryD3_{N86A}$ displayed specificity to GalNAc and some specificity for asialofetuin. This result would indicate that the change N86A increases GalNAc binding. It was revealed by Xiang *et al*., (2009), that the residue N546 in β18– β19 loop of the Cry1Ac domain III played an essential role in its insecticidal activity and binding to insect BBMV. N546 in Cry1Ac corresponds to N86 in CryD3. This would strongly indicate that the binding site for galactose on CryD3 and for GalNAc on tCry1Ac is in the same region and this mutation N86A can alter the binding specificity from galactose to GalNAc.

Additional ELLA testing is required to characterise $CryD3_{N86A}$, in order to analyse its binding abilities in more detail. It is interesting to note that only one amino acid change gave rise to an altered specificity for this protein. This shows the potential of these Cry toxin domains as carbohydrate binding proteins as they can be manipulated to give various sugar specificities. Further mutations could be carried out to potentially produce a mutant with only GalNAc and Tn antigen binding activity. This would potentially give rise to 3 distinct proteins from the CryD3 domain with

specificities to GalNAc, galactose and the Tn antigen respectively. These proteins would be good candidates for incorporation into novel bioanalytical platforms that would have the potential for detection and separation of closely related glycoproteins or in the detection of cell surface glycans in disease states such as cancer (Ghazarian, Idoni, and Oppenheimer 2011, Brockhausen 2006, Dube and Bertozzi 2005).

Chapter 6

Conclusion and Discussion

Interest in the area of glycobiology and protein glycosylation has increased significantly in the past twenty years. This is due to an increasing understanding of the importance of glycosylation such as the study of changes in protein glycosylation found in disease states such as cancer, discovery of bacterial glycosylation, and the increase in the use of glycoprotein therapeutics by the newly emerging biopharmaceutical industry.

Glycoproteins now represent a large percentage of marketed and clinical development phase therapeutic proteins (Hossler et al. 2009). Thus there is an increased need to understand the nature and function of the carbohydrate moieties on these proteins and the impact on their physiochemical properties, such as stability (half-life), solubility and immunogenicity. This is essential for discovering and developing safe and efficacious glycoprotein biopharmaceuticals. N and O linked glycosylation has a significant impact on the biological activity, pharmacokinetics, clearance, and immunogenicity of the therapeutic proteins (Li and d'Anjou 2009). New techniques are required to understand and study these glycoproteins and to investigate the impact carbohydrates have on the physiochemical properties of a protein.

In many early stage cancers there are significant changes in protein glycosylation and these changes represent an area for the development of biomarkers (Ghazarian, Idoni, and Oppenheimer 2011, Dube and Bertozzi 2005). For example, in cancers such as colon and breast, mucin-type O-glycans such as the T and the Tn antigen are often found on the surface of cells (Brockhausen 2006).

Carbohydrate binding proteins such as lectins, have huge potential here as they may be used to detect, separate and quantify glycans, glycoproteins and even closely related glycoforms (Ghazarian, Idoni, and Oppenheimer 2011, Kwan and Bun 2011). As discussed in the Introduction, it has become apparent that bacteria are an excellent source of novel lectins as many have evolved carbohydrate binding proteins to specifically target their host cells. Bacterial lectins are much more amenable to recombinant production than plant lectins and their specificities may be further altered or enhanced by mutagenesis. In addition to this bacteria have been found to

produce lectins with novel binding abilities with specific binding (Gilboa-Garber et al. 2011).

The focus of this research was to investigate the protein based Cry toxins from *Bacillus thuringiensis* and their potential for use as lectins in the detection, separation and quantification of glycans and glycoproteins. Analysis of the tCry1Ac, CryD3 and, $CryD3_{N86A}$ proteins (Cry1Ac toxin derivatives), revealed that each had specificity for glycans/oligosaccharides/carbohydrates and glycoproteins. tCry1Ac specifically bound to GalNAc, CryD3 bound to α1-3, galactose while $CryD3_{N86A}$ bound to GalNAc and galactose. This information provides us with good basis to conclude that the Cry toxins from *Bacillus thuringiensis* are potentially a good source of novel lectins and that they can have their carbohydrate specificity changed with relative ease by mutagenesis.

The bioinformatics carried out on Cry1Ac gave a good insight into the protein sequence in comparison to other Cry toxins. It highlighted the three domains that enabled the design of tCry1Ac and CryD3. The protein derivatives tCry1Ac and CryD3 were designed from the blast search. tCry1Ac was based on the activated form of Cry1Ac and should bind to GalNAc. CryD3 is designed based on the binding domain (domain III) of tCry1Ac.

The protein structures of tCry1Ac and CryD3 were predicted by I-TASSER and PyMol was used to model these predictions and to investigate any structural differences that could potentially affect binding. It was found that tCry1Ac had a very similar structure to the active Cry1Ac, while CryD3 had a slightly altered structure to domain III of tCry1Ac, which may affect its binding activity. The modelling of tCry1Ac and CryD3 gave a good insight into the protein folding and also the potential binding site for GalNAc. The residues thought to be involved in GalNAc binding were highlighted in both proteins, which would be useful for the design of the site directed mutagenesis in chapter 5.

tCry1Ac and CryD3 were successfully cloned into the expression vector pQE60 containing a C terminal 6x histidine tag, which allowed for downstream applications. Cloning both proteins in this way was significant as they would be expressed in *E. coli* as soluble active proteins. Previously Cry toxins have been expressed as a insoluble protoxins and various downstream processes were needed to activate the

toxins. This method allowed for less downstream processes and a higher yield of active protein produced by *E. coli*.

tCry1Ac and CryD3 were expressed in *E. coli* JM109 and purified by IMAC in laboratory scale. Preliminary ELLA analysis on these proteins showed that tCry1Ac bound to GalNAc and CryD3 bound, with low affinity, to galactose. This was a significant result as the Cry toxins have never been examined on ELLA's and it showed that tCry1Ac bound to GalNAc, compared to other commercial lectins, whole the carbohydrate specificity of the third domain of Cry1Ac had specificity for galactose. Both of these proteins showed good potential as CBP's therefore optimisation of the protein expression and purification was carried out.

Expression optimisation was carried out on tCry1Ac under various conditions which included altering the strains of *E. coli*, the concentration of the inducer IPTG, the temperature of expression and the type of growth medium. This optimisation of expression showed a significant increase in tCry1Ac expression. Optimisation of IMAC purification was then carried out for tCry1Ac and CryD3, in particular, the removal of persistent contaminating proteins. tCry1Ac and CryD3 protein purifications were carried out at 4°C to increase protein stability.

The main contaminant protein was identified as CRP (cAMP regulatory protein) from *E. coli*. Various purification strategies to remove this contaminating protein were then examined. CryD3 was purified by IMAC with added galactose washes of 100mM. tCry1Ac and Cry1Ac were then purified with high concentrations of imidazole 300mM and sugar washes 100mM (galactose and GalNAc). The purifications ultimately produced a pure protein for tCry1Ac and CryD3.

When analysed on an ELLA, both proteins were found to have active binding to their corresponding glycans i.e. tCry1Ac binding to GalNAc while CryD3 bound to galactose. tCry1Ac and CryD3 showed relatively good affinity for GalNAc and galactose respectively. The glycoprotein binding specificities of tCry1Ac and CryD3 were then investigated by ELLA analysis. tCry1Ac only showed specificity for GalNAc and the Tn antigen, while CryD3 had specificity for terminal α1-3, galactose residues, for example, asialofetuin. This result is significant as tCry1Ac binds specifically to GalNAc and the Tn antigen, which is a important cancer associated antigen. The ability to detect and quantify this antigen on cell surfaces would be of

great benefit in detecting and staging certain cancers (Brockhausen 2006). The finding that the third domain has a very specific binding to α1-3, galactose residues, is important as this protein can detect or separate glycoproteins with slight change at terminal galactose residues. The ELLA represents a new technique that can be employed for the analysis of Cry toxins specificity to carbohydrates. This assay is much more specific for carbohydrate binding compared to previous methods such as indirect toxin binding assays to BBMVs and light scattering assays (Carroll and Ellar 1993, Hendrickx et al. 1990).

Site directed mutagenesis was then carried out on tCry1Ac and CryD3 to investigate the important amino acids at their binding site and their binding characteristics. Alanine scanning was used to examine the effect certain amino acids have on the protein structure and function. The resulting proteins were named, $tCry1Ac_{N480A}$, $tCry1Ac_{QY487A}$, $tCry1Ac_{N521A}$, $CryD3_{N45A}$, $CryD3_{QY52A}$, $CryD3_{N86A}$, $CryD3_{PAV16A}$. These proteins were cloned, expressed in *E. coli* JM109 and purified by IMAC. An ELLA carried out on the proteins showed that only $CryD3_{N86A}$ had significant binding activity to GalNAc with some binding to galactose. The IMAC affinity purification of $CryD3_{N86A}$ was optimised with galactose washes at 100mM, which ultimately produced a pure preparation of $CryD3_{N86A}$. When analysed on an ELLA, $CryD3_{N86A}$ was found to recognise and bind both GalNAc and galactose.

Instead of cancelling out the galactose binding specificity on CryD3 N86A, altered the binding specificity to GalNAc. This result was significant as it showed that a single point mutation at N86A had the ability to change the proteins specificity from galactose to GalNAc and galactose. This suggests that the binding site for GalNAc and galactose in tCry1Ac and CryD3 are in the same region. It also indicated that N86 plays an important role in the structure and binding of the protein. This correlates to the study carried out by Xiang et al. (2009), where they revealed that the residue N546 in β18– β19 loop of the Cry1Ac domain III played an essential role in its insecticidal activity and binding to insect BBMV. N546 in Cry1Ac corresponds to N86 on CryD3, showing that this residue has an important role in the carbohydrate specificity and binding of this protein. Figure 6.1 shows the structure of tCry1Ac, CryD3 and $CryD3_{N86A}$ modelled using PyMol and the corresponding residue N86 highlighted on each protein. This shows that N86 is located close to the predicted

binding site on each protein and therefore is predicted to have an effect on the carbohydrate specificity of each protein.

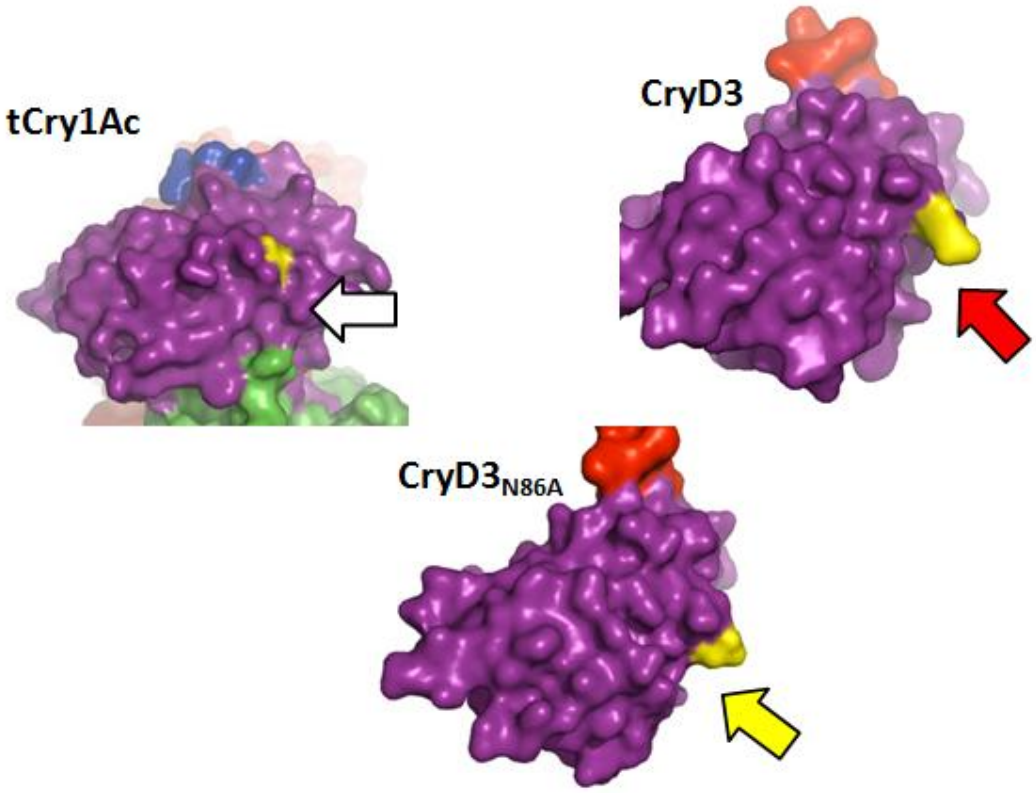


Figure 6.1: Structural modelling of the third domain of tCry1Ac, CryD3 and CryD3$_{N86A}$ highlighting corresponding residue N86. The third domain is shown in purple with the 6x histidine tag in red, and the residue N86 in yellow. The white arrow indicates the predicted binding site of GalNAc on tCry1Ac. The red arrow indicates the predicted binding site of galactose on CryD3. The yellow arrow indicates the predicted binding site of GalNAc and galactose on CryD3$_{N86A.}$

While this study did not show that the CRP protein from *E. coli* was glycosylated, further analysis on the nature of the affinity of CRP to tCry1Ac, CryD3 and $CryD3_{N86A}$ is required. Limitations exist if tCry1Ac, CryD3 and $CryD3_{N86A}$ are able to bind to numerous proteins via protein interactions. This may affect the binding of tCry1Ac, CryD3 and $CryD3_{N86A}$ to glycans and they may give false positives or negatives if they were introduced into an assay with a multitude of glycoproteins. However, the fact that GalNAc and galactose compete with the binding of CRP to tCry1Ac, CryD3 and $CryD3_{N86A}$ suggest that the interaction is a glycan/sugar based one There have been many discoveries of bacterial glycosylation since the *N*-linked general protein glycosylation pathway in *C. jejuni* was published in 2002 (Young et al. 2002). This includes a potential novel glycosylation system in *P.luminescens* (Fox 2011). Future work is needed to examine the possibility of *E. coli* having some type of glycosylation system.

Further work is required to characterise tCry1Ac, CryD3 and $CryD3_{N86A}$ in detail. These proteins have shown that they (or their binding domains) could potentially be used as lectins and that their carbohydrate binding affinities and specificities may be altered by mutagenesis. The applications of such proteins could be very useful. The creation of various lectins from the Cry1Ac toxin displaying a wide variety of specificities and affinities would be highly advantageous in the field of glycobiology, specifically, in the manufacture of proteins and lectin microarrays. Arrays that consist of recombinant molecules from a common protein scaffold would mean that each CBP would have the same requirements for activity, such as temperature, pH and buffer. The ability to produce these proteins through recombinant production has various advantages. It would ensure consistent production (which is not seen with plant lectins) and it would allow incorporation of affinity tags into the protein enabling more direct orientation specific immobilisation strategies. This in turn would lead to a decrease in cost, and greater accessibility of recombinant prokaryotic lectins.

tCry1Ac, CryD3 and $CryD3_{N86A}$ also have the potential to be used in various bioanalytical and separation platforms. CryD3 and $CryD3_{N86A}$ could be attached to cyanogen bromide (CNBr) activated matrices such as sepharose beads which could then be used in lectin affinity chromatography. They are smaller proteins than

tCry1Ac and they also bind their respective sugars with a lower affinity, which should enable an easier elution of the bound glycoproteins from the column with the corresponding free sugar. The coupling of molecules to CNBr-activated matrices is usually via amine groups. Therefore CryD3 and CryD3$_{N86A}$ could be coupled with CNBr sepharose via exposed lysine residues. Our laboratory is currently looking at the immobilisation of carbohydrate binding proteins to gold and silica nano-particles as well as carbon monoliths and PolyHIPE (poly High Internal Phase Emulsions) in order to develop novel bioanalyical platforms for glycoprotein separation and analysis.

tCry1Ac could also prove useful for analysing cell surface glycoproteins in disease states by microscopy. tCry1Ac binds specifically to GalNAc and the Tn antigen, which is found on the surface of many cancer cells such as breast and colon (Brockhausen 2006). tCry1Ac could be coupled with a florescent tag which would facilitate the labelling and detection of the Tn antigen on the surface of cells by microscopy. Fluorescently labelled tCry1Ac, CryD3 and CryD3$_{N86A}$ could also be used in Fluorescence-activated cell sorting (FACS) to identify the glycoproteins on the cell surface. This may be also used to examine any changes in surface glycosylation that occur during apoptosis (programmed cell death) and to distinguish diseased from healthy cells.

Previously Cry toxins were not cloned as soluble toxins. This study has shown that Cry1Ac can be cloned as a soluble protein and retain its activity to GalNAc. It also shows that the third domain of Cry1Ac can also be produced as a soluble protein, which has altered carbohydrate specificity to galactose, but can be returned to a GalNAc binder with the mutation N86A.

The present study confirms previous findings that Cry1Ac binds to GalNAc and contributes additional evidence that suggests the Cry toxins may have their carbohydrate specificities and affinities altered with relative ease by mutagenesis (Xiang et al. 2009, Burton et al. 1999, Knowles, Knight, and Ellar 1991). However the finding that CryD3 binds to galactose has not been observed by any other studies carried out on Cry1Ac. This is most likely due to the design of the protein CryD3 and the folding of this protein without domain I and II for structural support, which seems to alter the carbohydrate specificity of the third domain.

There are hundreds of Cry toxins whose binding characteristics have yet to be identified and there are many areas where Cry toxins could prove useful if their binding was extensively investigated and documented. Cry toxins potentially present a huge reservoir and variety of natural CBP's to study and are amenable to recombinant production and modification. This work presents a model of how domain III of a Cry toxin could be potentially used as a sugar binding lectin. The work carried out on tCry1Ac and CryD3 may potentially be applied to other Cry toxins thereby increasing the library of carbohydrate binding protein ligands.

6.2 Potential publications

- Carbohydrate binding properties of Cry1Ac toxin and its future applications
- Cry1Ac and its potential in glycoanalysis
- Mutation N86A in domain III of Cry1Ac derivative, CryD3, gives altered carbohydrate binding specificity

References

Abdulla, M.F.A. Alzate, O., Mohammad, M., McNall, R.J., Adang, M.J., and Dean, D.H. 2003. Introduction of Culex toxicity into *Bacillus thuringiensis* Cry4Ba by protein engineering. *Appl. Environ. Microbiol*. 69, pp. 5343-5353.

Abdullah, M.A., Moussa, S., Taylor, M.D., Adang, M.J. 2009. *Manduca sexta* Lepidoptera: Sphingidae cadherin fragments function as synergists for Cry1A and Cry1C *Bacillus thuringiensis* toxins against noctuid moths *Helicoverpa zea*, *Agrotis ipsilon* and *Spodoptera exigua*. *Pest Manag Sci*, June 1.

Adams, T.T., Eiteman, M.A., and Adang, M.J. 1999. *Bacillus thuringiensis* subsp. *kurstaki* spore production in batch culture using broiler litter extracts as complex media. *Bioresource Technology*. 67: pp.83-87.

Adang, M.J. 2004. *[M638] Insect aminopeptidase N*. In. A., and F. Woessner, eds. Handbook of Proteolytic Enzymes, 2nd Edition. Elsevier Science. Oxford.

Adang, M.J. 1991. *Bacillus thuringiensis* insecticidal crystal proteins: Structure, action, and utilization. *In*: K. Maramorosch (ed.) , *Biotechnology for Biological Control of Pests and Vectors*. CRC Press, Boca Raton, FL, Chapter 1, pp. 3-24.

Adang, M.J., Brody, M.S., Cardineau, G., Eagan, N., Roush, R.T., Shewmaker, C.K., Jones, A., Oakes, J.V., and McBride, K.E. 1993 . The reconstruction and expression of a *Bacillus thuringiensis cryIIIA* gene in protoplasts and potato plants. *Plant Mol. Biol*, 21, pp. 1131-1145.

Adang, M.J., Firoozabady, E., Klein, J., DeBoer, D., Sekar, V., Kemp, J.D., Murray, E., Rocheleau, T.A., Rashka, K., Staffeld, G., Stock, C., Sutton, D., and Merlo, D.J. 1986 . Expression of a *Bacillus thuringiensis* insecticidal protein gene in tobacco plants. *In*: Arntzen and C. Ryan (eds.) , *Molecular Strategies for Crop Protection. UCLA Symposia on Molecular and Cellular Biology, New Series, Volume 48*. Alan R. Liss, Inc. New York, NY.

Adang, M.J. and Jurat-Fuentes, J.L 2006 . *Bacillus thuringiensis* Cry proteins. *In* J.N. All and M.F. Tracy, (eds.) *Use and Management of Insecticides, Acaricides and Transgenic Crops*. Entomological Society of America. Lanham, MD.

Adang, M.J., Staver, M.J., Rocheleau, T.A., Leighton, J., Barker, R.F., and Thompson, D.V. 1985 . Characterized full-length and truncated plasmid clones of the crystal protein of *Bacillus thuringiensis* subsp. *kurstaki* HD-73 and their toxicity to *Manduca sexta*. *Gene*, 36, pp. 289-300.

Aiba H. 1985 . Transcription of the *Escherichia coli* adenylate cyclase gene is negatively regulated by cAMP-cAMP receptor protein. *J. Biol. Chem.* 260, pp. 3063-3070

Akiba, T., Abe, Y., Kitada, S., Kusaka, Y., Ito, A., Ichimatsu, T., Katayama, H., Akao, T., Higuchi, K., Mizuki, E., Ohba, M., Kanai, R. and Harata, K. 2009 . Crystal Structure of the Parasporin-2 *Bacillus thuringiensis* Toxin That Recognizes Cancer Cells. *Journal of Molecular Biology*, 386 (1) , pp.121-133.

Anderson N. L, Regnier F.E, Gibson B. W, and Fisher S. J. 2010 Sweetening the Pot: Adding Glycosylation to the Biomarker Discovery Equation. *Clinical Chemistry* 56,(2), pp.223–236.

Arenas, I., Bravo, A., Soberon, M., and Gomez, I. 2010. Role of alkaline phosphatase from *Manduca sexta* in the mechanism of action of *Bacillus thuringiensis* Cry1Ab toxin. *J Biol Chem* 285, pp. 12497–12503.

Atsumi, S., Mizuno, E., Hara, H., Nakanishi, K., Kitami, M., Miura, N., *et al*. 2005. Location of the *Bombyx mori* aminopeptidase N type I binding site on *Bacillus thuringiensis* Cry1Aa toxin. *Appl Environ Microbiol* 71, pp. 3966–3977.

Avisar, D., Eilenberg, H., Keller, M., Reznik, N., Segal, M., Sneh, B. and Zilberstein, A. 2009 . The *Bacillus thuringiensis* delta-endotoxin Cry1C as a potential bioinsecticide in plants. *Plant Science*, 176 (3), pp. 315-324.

Bagchi T, 2000. Partial purification of Cry1A toxin-binding proteins from Helicoverpa armigera larval midgut. *World Journal of Microbiology and Biotechnology*. 16, pp. 635-340.

Bagla, P. 2010. Hardy cotton-munching pests are latest blow to GM crops. *Science*, 327 , pp. 1439

Banks, D.J., and Adang, M.J. 2003 . Cloning of a *Heliothis virescens* 110-kDa aminopeptidase N and expression in Drosophila S2 cells. *Insect Biochem. Molec. Biol*. 33, pp. 409-508.

Banks, D. J., Jurat-Fuentes, J.L., Dean, D.H., and Adang, M.J. 2001 . *Bacillus thuringiensis* Cry1Ac and Cry1Fa -endotoxin binding to a novel 110 kDa aminopeptidase in *Heliothis virescens* is not N-acetylgalactosamine mediated. *Insect Biochem. Molec. Biol*, 31, pp. 909-918.

Bayyareddy, K., Andacht, T.M., Abdullah, M.A., Adang, M.J. 2009. Proteomic identification of *Bacillus thuringiensis subsp. israelensis* toxin Cry4Ba binding proteins in midgut membranes from *Aedes Stegomyia aegypti* Linnaeus Diptera, Culicidae larvae. *Insect Biochem Mol Biol*, 39 (4), pp. 279-86.

Bektas, M, and Rubenstein, D.S. 2011. The role of intracellular protein O-glycosylation in cell adhesion and disease, *Journal of biomedical research*, 25, (4), pp. 227-236.

Benz I, Schmidt M. A. 2002. Never say never again: protein glycosylation in pathogenic bacteria. *Molecular Microbiology*, 45, (2) , pp. 267–276.

Beron C M, Curatti L, and Salerno G. L. 2004. New strategy for identification of Novel cry-Type Genes from *Bacillus thuringiensis* Strains. *Applied and Environmental Microbiology*. 71, (2), pp. 761-765.

Bosch, D., Schipper, B., van der Kleij, H., de Maagd, R.A., and Stiekema, J. 1994. Recombinant *Bacillus thuringiensis* insecticidal proteins with new properties for resistance management. *Biotechnology* 12, pp. 915–918.

Brar, S.K., Verma, M., Tyagi, R.D., Surampalli, R.Y., Barnabé, S. and Valéro, J.R. 2007. *Bacillus thuringiensis* proteases: Production and role in growth, sporulation and synergism. *Process Biochemistry,* 42 (5) , pp. 773-790.

Bravo, A. 1997. Phylogenetic relationships of *Bacillus thuringiensis* delta-endotoxin family proteins and their functional domains. *J Bacteriol* 179, pp 2793–27801.

Bravo A, Del Rincon-Castro MC, Ibarra JE & Sobero´n M. 2011. Towards a healthy control of insect pest: potential use of microbial insecticides. *Green Trends in Insect Control*, Lo´ pez O & Fernandez-Bolan˜os JG, (eds), pp. 266–299. Royal Society of Chemistry, London.

Bravo, A., Gill, S.S. and Soberón, M. 2007. Mode of action of *Bacillus thuringiensis* Cry and Cyt toxins and their potential for insect control. *Toxicon,* 49 (4) , pp. 423-435.

Bravo, A., Gill, S.S., and Soberón, M. 2005. *Bacillus thuringiensis* mechanisms and use. *In, Comprehensive Molecular Insect Science*. Gilbert, L.I., Iatrou, K., and Gill, S.S. (eds). Oxford: ELSEVIER, pp. 175–206.

Bravo A, Go´mez I, Conde J, Mun˜oz-Garay C, Sa´nchez J, Miranda R, Zhuang M, Gill SS & Sobero´n M. 2004. Oligomerization triggers binding of a *Bacillus thuringiensis* Cry1Ab pore-forming toxin to aminopeptidase N receptor leading to insertion into membrane microdomains. *Biochim Biophys Acta* 1667, pp. 38–46.

Bravo A, Gómez I, Porta H, García-Gómez B. I, Rodriguez-Almazan C, Pardo L and Soberón M. 2012. Evolution of *Bacillus thuringiensis* Cry toxins insecticidal activity. *Microbial Biotechnology*, 6, pp. 17–26.

Bravo A, Jansens S & Peferoen M. 1992. Immunocytochemical localization of *Bacillus thuringiensis* insecticidal crystal proteins in intoxicated insects. *J Invertebr Pathol* 60, pp. 237–247.

Bravo, A., Likitvivatanavong, S., Gill, S.S., and Soberón, M. 2011. *Bacillus thuringiensis*: a story of a successful bioinsecticide. *Insect Biochem Mol Biol* 41, pp. 423–431.

Bravo A and Soberon M. 2008. How to cope with insect resistance to Bt toxins? *Trends Biotechnol* 26, pp. 573–579.

Brockhausen I. 2006. Mucin-type O-glycans in human colon and breast cancer: glycodynamics and functions, *EMBO reports,* 7, (6), pp.599-604

Brockhausen I. 1999. Pathways of O-glycan biosynthesis in cancer cells. *Biochim. Biophys. Acta*. 1473, pp. 67–95.

Brockhausen, I., Yang, J. M., and Burchell, J. 1995 . Mechanisms Underlying Aberrant Glycosylation of Muc1 Mucin in Breast-Cancer Cells. *European Journal of Biochemistry*. 233 (2) , pp.607-617.

Broderick N A, Raffa K F, and Handelsman J. 2006. Midgut bacteria required for *Bacillus thuringiensis* insecticidal activity. *PNAS*, 103, (41), pp. 15196-15199.

Budnik, B. A., Lee, R. S. and Steen, J. A. J. 2006 . Global methods for protein glycosylation analysis by mass spectrometry. *Biochimica Et Biophysica Acta-Proteins and Proteomics*. 1764 (12) , pp1870-1880.

Burton, S.L., Ellar, D.J., Li, J. and Derbyshire, D.J. 1999. N-acetylgalactosamine on the putative insect receptor aminopeptidase N is recognised by a site on the domain III lectin-like fold of a *Bacillus thuringiensis* insecticidal toxin. *Journal of Molecular Biology,* 287 (5), pp. 1011-1022.

Carroll J & Ellar DJ . 1993. An analysis of *Bacillus thuringiensis* insecticidal d-endotoxins action on insect-midgut-membrane permeability using a light-scattering assay. *Eur J Biochem* 214, pp. 771–778.

Chang, G., Chen, C., Lin, C., Chen, H. and Chen, H. 2003. Improvement of glycosylation in insect cells with mammalian glycosyltransferases. *Journal of Biotechnology,* 102 (1), pp. 61-71.

CHANG, H., LIANG, G., WANG, G., YU, H., GUO, Y. and WU, K. 2008. Expression of Aminopeptidase N1 APN1 , the Main Receptor Protein for *Bacillus thuringiensis* Cry1A Toxin from Helicoverpa armigera Larval Midgut in Trichoplusia ni cells. *Agricultural Sciences in China,* 7 (3) , pp.329-335.

Chang, V.T., Crispin, M., Aricescu, A.R., Harvey, D.J., Nettleship, J.E., Fennelly, J.A., Yu, C., Boles, K.S., Evans, E.J., Stuart, D.I., Dwek, R.A., Jones, E.Y., Owens, R.J. and Davis, S.J. 2007. Glycoprotein Structural Genomics: Solving the Glycosylation Problem. *Structure,* 15(3) , pp. 267-273.

Chen, J., Aimanova, K.G., Pan, S. and Gill, S.S. 2009. Identification and characterization of Aedes aegypti aminopeptidase N as a putative receptor of *Bacillus thuringiensis* Cry11A toxin. *Insect Biochemistry and Molecular Biology,* 39 (10) , pp. 688-696.

Chen, J., Brown, M.R., Hua, G., and Adang, M.J. 2005 . Comparison of the localization of *Bacillus thuringiensis* Cry1A d-endotoxins and their binding proteins in larval midgut of tobacco hornworm, *Manduca sexta*. *Cell Tissue Res*. 321, pp. 123-129.

Chen J, Hua G, Jurat-Fuentes JL, Abdullah MA, Adang MJ . 2007 . Synergism of *Bacillus thuringiensis* toxins by a fragment of a toxin-binding cadherin.*Proc Natl Acad Sci U S A.* 104 (35), pp. 13901-6.

Crickmore, N. 2013. *Bacillus thuringiensis Toxin Nomenclature* [Online]. Available from: http://www.lifesci.sussex.ac.uk/home/Neil_Crickmore/Bt/. Assessed 19 November 2012.

Crickmore, N. 2000. The diversity of *Bacillus thuringiensis* d-endotoxins. *In Enthomopathogenic Bacteria: From Laboratory to Field Application*. Charles, J.-F., Delécluse, A., and Nielsen-LeRoux, C. (eds) . Dordrecht: Kluwer Academic Publishers, pp. 65–79.

Crickmore, N., Zeigler, D.R., Feitelson, J., Schnepf, E., Van Rie, J., Lereclus, D., Baum, J. and Dean, D.H. 1998. Revision of the Nomenclature for the *Bacillus thuringiensis* Pesticidal Crystal Proteins. *Microbiology and Molecular Biology Reviews*, 62 (3), pp. 807-813.

Crickmore, N., Zeigler, D.R., Schnepf, E., Van Rie, J., Lereclus, D., Baum, J., *et al*. 2011. '*Bacillus thuringiensis* toxin nomenclature' [WWW document]. URL http://www.lifesci. sussex.ac.uk/Home/Neil_Crickmore/Bt/.

Cummings R.D. 2009. The repertoire of glycan determinants in the human glycome. *Mol. BioSyst*., 5, pp. 1087-1104.

Cunningham, B.C. and Wells, J.A. 1989 . High-Resolution Epitope Mapping of Hgh-Receptor Interactions by Alanine-Scanning Mutagenesis. *Science,* 244 (4908) , pp.1081-1085.

Dalpathado, D. S. and Desaire, H. 2008 . Glycopeptide analysis by mass spectrometry. *Analyst*. 133 (6) , pp.731-738.

Daniel, A., Dean, D.H., and Adang, M.J. 2001 . Analyses of the pore forming ability of *Bacillus thuringiensis* Cry1A mutant toxins using a light-scattering technique. *Pesticide Biochem. Physiol*. 70, pp. 7-18.

Daniel, A., Sangadala, S., Dean, D.H., and Adang, M.J. 2002 . Denaturation of either *Manduca sexta* aminopeptidase N or *Bacillus thuringiensis* Cry1A toxins exposes binding epitopes hidden under nondenaturing conditions. *Appl. Environ. Microbiol*. 68, pp. 2106-2112.

Dean, D.H. and Adang, M.J. 1992. Protein engineering of *Bacillus thuringiensis* delta-endotoxins and genetic manipulation for plant protection. , *In*: P. R. Shewry and S. Gutteridge (eds.), *Plant protein engineering*. University Press, Cambridge, UK. Chapter 16, pp. 293-311

Dell A, Galadari A, Sastre F, Hitchen P. 2010. Similarities and Differences in the Glycosylation Mechanisms in Prokaryotes and Eukaryotes. *International Journal of Microbiology*, 10, (14).

Dojima, T., Nishina, T., Kato, T., Uno, T., Yagi, H., Kato, K. and Park, E.Y. 2009. Comparison of the N-linked glycosylation of human β1,3-N-acetylglucosaminyltransferase 2 expressed in insect cells and silkworm larvae. *Journal of Biotechnology*, 143(1) , pp. 27-33.

Dorsch, J.A., Candas, M., Griko, N.B., Maaty, W.S.A., Midboe, E.G., Vadlamudi, R.K. and Bulla, L.A. 2002. Cry1A toxins of *Bacillus thuringiensis* bind specifically to a region adjacent to the membrane-proximal extracellular domain of BT-R1 in Manduca sexta:: involvement of a cadherin in the entomopathogenicity of *Bacillus thuringiensis*. *Insect Biochemistry and Molecular Biology*, 32 (9) , pp. 1025-1036.

Drake P.M, Cho W, Li B, Prakobphol A, Johansen E, Lebrilla C. B. and Joo H. 2009. The prospects of glycan biomarkers for the diagnosis of diseases. *The Royal Society of Chemistry Mol. BioSyst*, 5, pp. 17–20.

Dube D H. and Bertozzi C R. 2005. Glycans in Cancer and inflammation: Potential for therapeutics and diagnostics. *Nature reviews, drug discovery*, 4, pp. 477-488.

English L, Readdy TL and Bastian AE . 1991. Delta-endotoxininduced leakage of 86Rb+-K+ and H2O from phospholipids vesicles is catalyzed by reconstituted midgut membrane. *Insect Biochem* 21, pp.177–184.

Fernández, L.E., Gómez, I., Pacheco, S., Arenas, I., Gilla, S.S., Bravo, A. and Soberón, M. 2008. Employing phage display to study the mode of action of *Bacillus thuringiensis* Cry toxins. *Peptides,* 29 (2), pp. 324-329.

Ferre´ J and Van Rie J . 2002. Biochemistry and genetics of insect resistance to *Bacillus thuringiensis*. *Annu Rev Entomol* 47, pp. 501–533.

Fischbach M A, Lin H, Liu D.R, Walsh C. T. 2006. How pathogenic bacteria evade mammalian sabotage in the battle for iron. *Nature Chemistry Biology*, 2, pp. 132-138.

Fitches, E., Woodhouse, S. D., Edwards, J. P. and Gatehouse, J. A. 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) , pp. 777-787.

Frankenhuyzen, K.v. 2009 Insecticidal activity of *Bacillus thuringiensis* crystal proteins. *Journal of Invertebrate Pathology,* 101 (1) ,pp. 1-16.

van Frankenhuyzen, K. and C. Nystrom. 2002 *The Bacillus thuringiensis toxin specificity database*. [Online]. Available from http://cfs.nrcan.gc.ca/projects/119/2 [Assessed 10 September 2012].

Gahan LJ, Gould F and Heckel DG. 2001. Identification of a gene associated with Bt resistance in *Heliothis virescens. Science* 293, pp. 857–860.

Garczynski, S.F., Adang, M.J. 2000. Investigations of *Bacillus thuringiensis* Cry1 Toxin Reeceptor Structure and Function. *In* (eds.) J.-F. Charles, A. Delecluse and C. Nielsen-LeRoux. *Entomopathogenic Bacteria: from laboratory to field application*. Kluwer Academic Publishers, The Netherlands.

Garczynski S.F, Adang M. J. 1995. *Bacillus thuringiensis* CryIA C δ-Endotoxin Binding Aminopeptidase in the Manduca sexta Midgut has Glycosyl-phosphatidylinositol Anchor. *Insect Biochem. Molec. Biol*, 25, (4), pp. 409-415.

Garczynski, S.F., Crim, J.W., and Adang, M.J. 1991 . Identification of putative insect brush border membrane binding molecules specific to *Bacillus thuringiensis* delta-endotoxin by protein blot analysis. *Appl. Environ. Microbiol*. 57, pp. 2816-2820.

Gassmann, A.J., Petzold-Maxwell, J.L., Keweshan, R.S., and Dunbar, M.W. 2011. Field-evolved resistance to Bt maize by western corn rootworm. *PLoS ONE* 6, (7), pp. 22629

Geyer, H. and Geyer, R. 2006. Strategies for analysis of glycoprotein glycosylation. *Biochimica Et Biophysica Acta-Proteins and Proteomics*. 1764 (12) , pp. 1853-1869.

Ghazarian H, Idoni B, Oppenheimer S.B. 2011. A glycobiology review: Carbohydrates ,lectins and implications in cancer therapeutics. *acta histochemica*, 113, pp. 236–247.

Gilboa-Garber N , Zinger-Yosovich K.D. , Sudakevitz D , Lerrer B , Imberty A , Wimmerova M , Wu A.M. , Garber N.C. 2011 . The Five Bacterial Lectins PA-IL, PA-IIL, RSL, RS-IIL, and CV-IIL : Interactions with Diverse Animal Cells and Glycoproteins. The Molecular Immunology of Complex Carbohydrates-3, *Advances in Experimental Medicine and Biology*. Volume 705, pp. 155-211.

Gill, D.J., Clausen, H. and Bard, F. Location, location, location: new insights into O-GalNAc protein glycosylation. *Trends in Cell Biology,* In Press, Corrected Proof

Girard F, Vachon V, Pre´fontaine G, Marceau L, Su Y, Larouche G, Vincent G, Schwartz JL, Masson L & LapradeR. 2008. Cysteine scanning mutagenesis of a-4 a putative pore lining helix of the *Bacillus thuringiensis* insecticidal toxin Cry1Aa. Appl Environ Microbiol 74, pp. 2565–2572.

Griffitts, J.S., Haslam, S.M., Yang, T., Garczynski, S.F., Mulloy, B., Morris, H., Cremer, P.S., Dell, A., Adang, M.J., Aroian, R.V. 2005 . Glycolipids as receptors for *Bacillus thuringiensis* crystal toxin. *Science*, 307, pp. 922-925.

Gómez, I., Arenas, I., Benitez, I., Miranda-Ríos, J., Becerril, B., Grande, G., *et al*. 2006. Specific epitopes of Domains II and III of *Bacillus thuringiensis* Cry1Ab toxin involved in the sequential interaction with cadherin and aminopeptidase-N receptors in *Manduca sexta*. *J Biol Chem,* 281, pp. 34032–34039.

Gómez, I., Dean, D.H., Bravo, A., and Soberón, M. 2003. Molecular basis for *Bacillus thuringiensis* Cry1Ab toxin specificity: two structural determinants in the *Manduca sexta* Bt-R1 receptor interact with loops a-8 and 2 in domain II of Cy1Ab toxin. *Biochem* 42, pp. 10482–10489.

Gómez, I., Pardo-López, L., Muñoz-Garay, C., Fernandez, L.E., Pérez, C., Sánchez, J., Soberón, M. and Bravo, A. 2007. Role of receptor interaction in the mode of action of insecticidal Cry and Cyt toxins produced by *Bacillus thuringiensis*. *Peptides,* 28 (1) , pp. 169-173.

Gooday, G. W. 1990. The Ecology of Chitin Degradation. *Advances in Microbial Ecology*. 11, pp. 387-430.

Griffitts J S, Whitacre J L, Stevens D. E, Aroian R.V. 2001. Bt Toxin Resistance from Loss of a Putative Carbohydrate-Modifying Enzyme. Science. 293, pp. 860-864.

Grochulski P, Masson L, Borisova S, Pusztai-Carey M, Schwartz J-L, Brousseau R, Cygler M. 1995. *Bacillus thuringiensis* CryIA a Insecticidal Toxin: crystal structure and Channel Formation. *J. Mol. Biol*. 254, pp. 447-464.

Haider MZ & Ellar DJ. 1989. Mechanism of action of *Bacillus thuringiensis* insecticidal d-endotoxin: interaction with phospholipids vesicles. *Biochim Biophys Acta* 978, pp. 216–222.

Hakomori S and Cummings R.D. 2012. Glycosylation effects on cancer development, *Glycoconj J*, 29, pp. 565-566.

van Heijenoort J. 2001. Formation of the glycan chains in the synthesis of bacterial peptidoglycan. *Glycobiology*.11, pp. 25R–36R.

Hendrickx K, De Loof A & Van Mellart H. 1990. Effect of *Bacillus thuringiensis* delta-endotoxin on the permeability of brush border membrane vesicles from tobacco hornworm Manduca sexta midgut. *Comp Biochem Physiol* 95C, pp. 241–245.

Hitchen, P. G. and Dell, A. 2006. Bacterial glycoproteomics. *Microbiology-Sgm*. 152, pp.1575-1580.

Hofte H, Whiteley H.R. 1989. Insecticidal Crystal Proteins of *Bacillus thuringiensis*. *Microbiology Reviws*, 53, (2),pp. 242-255.

Holger R A.N, Frankenberger S, Zahringer U, Mack D. 2010. Structure, function and contribution of polysaccharide intercellular adhesin PIA to Staphylococcus epidermidis biofilm formation and pathogenesis of biomaterial-associated infections. *European Journal of Cell Biology*, 89, pp. 103–111.

Hollingsworth MA, Swanson BJ. Mucins in cancer: 2004. Protection and control of the cell surface. *Nat. Rev. Cancer*, 4, pp. 45–60.

Hossler P, Khattak S.F, and Jian L.Z. 2009 . Optimal and consistent protein glycosylation in mammalian cell culture. *Glycobiology*.19 (9) , pp. 936-949.

Hua, G., Jurat-Fuentes, J.L. and Adang M.J. 2004a . BtR1a extracellular cadherin repeat 12 mediates *Bacillus thuringiensis* Cry1Ab binding and cytotoxicity. *J. Biol. Chem.* 279, pp. 28051-28056.

Hua, G., Jurat-Fuentes, J.L., and Adang, M.J. 2004b . Fluorescent-based assays establish *Manduca sexta* Bt-R1a cadherin as a receptor for multiple *Bacillus thuringiensis* Cry1A toxin in Drosophila S2 cells. *Insect Biochem. Molec. Biol*. 34, pp. 193-202.

Hua, G., Masson, L. Jurat-Fuentes, J.L., Schwab, G. and Adang, M.J. 2001 . Binding analyses of *Bacillus thuringiensis* Cry delta-endotoxin using brush border membrane vesicles of *Ostrinia nubilalis*. *Appl. Environ. Microbiol*. 67, pp. 872-879.

Hua G, Zhang R, Abdullah MA, Adang MJ.2008. Anopheles gambiae cadherin AgCad1 binds the Cry4Ba toxin of *Bacillus thuringiensis* israelensis and a fragment of AgCad1 synergizes toxicity. *Biochemistry*. 47 (18), pp. 5101-10

Hua, G., Zhang, R., Bayyareddy, K., Adang, M.J. 2009. Anopheles gambiae alkaline phosphatase is a functional receptor of *Bacillus thuringiensis* jegathesan Cry11Ba toxin. *Biochemistry*. 48 (41), pp. 9785-93

Hu X, Hansen M.B, Eilenberg J, Hendriksen N.B, Smidt L, Yuan Z, Jensen G.B. 2004, conjugative transfer, stability and expression of a plasmid encoding a cry1Ac gene in Bacillus cereus group strains. *FEMS Microbiology Letters*, 231, pp. 45-52.

Iacovache, I., van der Goot, F.G. and Pernot, L. 2008. Pore formation: An ancient yet complex form of attack. *Biochimica Et Biophysica Acta BBA - Biomembranes*, 1778 (7-8) ,pp. 1611-1623.

Ibargutxi, M.A., Muñoz, D., Escudero, I.R.d. and Caballero, P. 2008. Interactions between Cry1Ac, Cry2Ab, and Cry1Fa *Bacillus thuringiensis* toxins in the cotton pests Helicoverpa armigera Hübner and Earias insulana Boisduval . *Biological Control*, 47 (1) , pp. 89-96.

Ideo H, Fukushima K, Gengyo-Ando K, Mitani S, Dejima K, Nomura K, Yamashita K. 2009 A C.Elegans Glycolipid-Binding Galectin functions in host Defence Against Bacterial Infection. ASBMB, Downloaded from www.jbc.org at IReL Dublin City University , on November 16, 2009.

Imberty, A., Wimmerova, M., Mitchell, E. P. and Gilboa-Garber, N. 2004 . Structures of the lectins from Pseudomonas aeruginosa: insights into the molecular basis for host glycan recognition. *Microbes and Infection*. 6 (2) , pp. 221-228.

James C. 2010. Global status of commercialized biotech/GM crops. *ISAAA Briefs* 42: ISAAA, Ithaca, NY.

Jenkins, J.L., Lee, M.K., Sangadala, S., Adang, M.J., Dean, D.H. 1999. Binding of *Bacillus thuringiensis* Cry1Ac toxin to *Manduca sexta* aminopeptidase-N receptor is not directly related to toxicity. *FEBS Letters*, 462, pp. 373-376.

Jime´nez-Jua´rez A, Mun˜oz-Garay C, Go´mez I, Saab-Rincon G, Damian-Alamazo JY, Gill SS, Sobero´n M & Bravo A. 2007. *Bacillus thuringiensis* Cry1Ab mutants affecting oligomer formation are non-toxic to Manduca sexta larvae. *J Biol Chem* 282, pp. 21222–21229.

Jurat-Fuentes, J.L. and Adang, M.J. 2007. A proteomic approach to study Cry1Ac binding proteins and their alterations in resistant Heliothis virescens larvae. *J Invertebr Pathol*. May 25 .

Jurat-Fuentes, J.L. and Adang, M.J 2006 . The Heliothis virescens cadherin protein expressed in Drosophila S2 cells functions as a receptor for *Bacillus thuringiensis* Cry1A but not Cry1Fa toxins.*Biochemistry*. Aug 15;45 (32), pp. 9688-95

Jurat-Fuentes, J.L. and Adang, M.J. 2006 . Cry toxin mode of action in susceptible and resistant *Heliothis virescens* larvae. *J. Invertbr. Pathol.* 92, pp. 166-171.

Jurat-Fuentes, J.L. and Adang, M.J. 2004 . Characterization of a Cry1Ac receptor alkaline phosphatase in susceptible and resistant *Heliothis virescens* larvae. *Eur. J. Biochem*. 271, pp. 3127-3135.

Jurat-Fuentes, J.L. and Adang, M.J. 2001 . Importance of Cry1 d -endotoxin domain II loops for binding specificity in *Heliothis virescens* L. . *Appl. Environ. Microbiol.* 67, pp. 323-329.

Jurat-Fuentes, J.L., Gahan, L.J., Gould, F.L., Heckel, D.G. and Adang, M.J. 2004 . The HevCaLP protein mediates binding specificity of the Cry1A class of *Bacillus thuringiensis* toxins in *Heliothis virescens* *Biochem*. 43, pp. 14299-14305.

Jurat-Fuentes, J.L., Gould, F.L., and Adang, M.J. 2003. Dual resistance to *Bacillus thuringiensis* Cry1Ac and Cry2Aa toxins in *Heliothis virescens* suggests multiple mechanisms of resistance. *Appl. Environ. Microbiol*, 69, pp. 5898-5906.

Jurat-Fuentes, J.L., Gould, F.L., and Adang, M.J. 2002 . Altered glycosylation of 63- and 68-kDa microvillar proteins in *Heliothis virescens* correlates with reduced Cry1 toxin binding, pore formation and increased resistance to

Bacillus thuringiensis Cry1 toxins. *Appl. Environ. Microbiol.* 68, pp. 5711-5717.

Jurat-Fuentes, J.L., Gould, F.L. and Adang, M.J. 2000 . High levels of resistance and cross resistance to *Bacillus thuringiensis* Cry1 toxins in *Heliothis virescens* are due to reduced toxin binding and pore formation. *Resistant Pest Management,* 11, pp. 23-24.

Karlova, R., Weemen-Hendriks, M., Naimov, S., Ceron, J., Dukiandjiev, S. and de Maagd, R.A. 2005. *Bacillus thuringiensis* δ-endotoxin Cry1Ac domain III enhances activity against Heliothis virescens in some, but not all Cry1-Cry1Ac hybrids. *Journal of Invertebrate Pathology,* 88 (2) ,pp. 169-172.

Karumbaiah, L., Oppert, B., Juan L. Jurat-Fuentes, J.-L., and Adang, M.J. 2007 . Analysis of midgut proteinases from *Bacillus thuringiensis*- susceptible and -resistant *Heliothis virescens* Lepidoptera: Noctuidae .*Comp. Biochem. Physiol*. B. Jan;146 (1), pp. 139-46

Kasman, L.A., Lukowiak, A.A., Garczynski, S.F., McNall, R.J., Youngman, P., and Adang, M.J. 1998 . Phage display of a biologically active *Bacillus thuringiensis* toxin. *Appl. Environ. Microbiol.* 64, pp. 2995-3003.

Kato, T., Higuchi, M., Endo, R., Maruyama, T., Haginoya, K., Shitomi, Y., Hayakawa, T., Mitsui, T., Sato, R. and Hori, H. 2006. *Bacillus thuringiensis* Cry1Ab, but not Cry1Aa or Cry1Ac, disrupts liposomes. *Pesticide Biochemistry and Physiology,* 84 (1), pp. 1-9.

Kim, Y., Roh, J.Y., Kang, J.N., Wang, Y., Shim, H.J., Li, M.S., Choi, J.Y. and Je, Y.H. 2008 Mutagenesis of *Bacillus thuringiensis cry1Ac* gene and its insecticidal activity against Plutella xylostella and Ostrinia furnacalis. *Biological Control,* 47 (2), pp. 222-227.

KITAM M, KADOTANI T, NAKANISHI K, ATSUMI S, HIGURASHI S, ISHIZAKA T, WATANABE A, SATO R. 2011. *Bacillus thuringiensis* Cry

Toxins Bound Specifically to Various Proteins *via* Domain III, Which Had a Galactose-Binding Domain-Like Fold. *Bioscience, Biotechnology, and Biochemistry*, Vol. 75, (2), pp. 305-312.

Kitov, P.I., Sadowska, J.M., Mulvey, G., Armstrong, G.D., Ling, H. 2000. Shiga-like toxins are neutralized by tailored multivalent carbohydrate ligands. *Nature*, 403, pp. 669–672.

Kline K A, Falker S, Dahlberg S, Normark S, Normark B H-. 2009. Bacterial Adhesins in Host-Microbe Interactions. *Cell Host and Microbe*, 5.

Knight, P.J.K., Carroll, J. and Ellar, D.J. 2004. Analysis of glycan structures on the 120 kDa aminopeptidase N of Manduca sexta and their interactions with *Bacillus thuringiensis* Cry1Ac toxin. *Insect Biochemistry and Molecular Biology*, 34 (1) , pp.101-112.

Knowles, B.H., Knight, P.J., Ellar, D.J. 1991. N-acetyl Galactosamine is Part of the Receptor in Insect Gut Epithelia that Recognizes an Insecticidal Protein from *Bacillus thuringiensis*. *Biological Sciences*, 245, pp. 1312, 31-35.

Kornfeld R, Kornfeld S. 1985. Assembly of asparagine-linked oligosaccharides. *Annu Rev Biochem*.;54, pp. 631–664.

Krishnamoorthy, M., Jurat-Fuentes, J.L., McNall, R.J., Andacht, T.and Adang, M.J. 2007. Identification of novel Cry1Ac binding proteins in midgut membranes from *Heliothis virescens* using proteomic analyses. *Insect Biochemistry and Molecular Biology*, 37 (3), pp. 189-201.

Krist, P., Vannucci, L., Kuzma, M. 2004. Fluorescent labelled thiourea-bridged glycodendrons. *Chembiochem*. 5 (4) , pp. 445-452.

Kwan L.S and Bun N.T . 2011. Lectins: production and practical applications. *Appl Microbiol Biotechnol*. 89 (1), pp. 45–55.

Lambre, C. R., Terzidis, H., Greffard, A. and Webster, R. G. 1991 . An Enzyme-Linked Lectin Assay for Sialidase. *Clinica Chimica Acta*. 198 (3) , pp183-194.

Laemmli, U. K. 1970. Cleavage of Structural Proteins during the Assembly of the Head of Bacteriophage T4. *Nature* 227 (5259), pp680-685.

Larragy, Ruth. 2011. *The cloning , expression and characterisation of bacterial chitin-binding proteins from pseudomonas aeruginosa , serratia marcescens, photorhabdus luminescens and photorhabdus asymbiotica.* PhD thesis, Dublin City University

Lee, M.K., Jenkins, J.L., You, T.H., Curtiss, A., Son, J.J., Adang, M.J. and Dean, D.H. 2001 . Mutations at the arginine residues in a8 loop of Cry1Ac *Bacillus thuringiensis* d-endotoxin affect toxicity and binding to *Manduca sexta* and *Lymantria dispar* aminopeptidase *N*. *FEBS Letts*, 497, pp. 108-112.

Lee, M.K., Miles, P. and Chen, J. 2006. Brush border membrane binding properties of *Bacillus thuringiensis* Vip3A toxin to Heliothis virescens and Helicoverpa zea midguts. *Biochemical and Biophysical Research Communications,* 339 (4) , pp. 1043-1047.

Li H and d'Anjou M. 2009. Pharmacological significance of glycosylation in therapeutic proteins, *Current opinion in Biotechnology*, 20, pp. 678-684.

Liebig B, Stetson DL and Dean DH. 1995. Quantification of the effect of *Bacillus thuringiensis* toxins on short-circuit current in the midgut of Bombyx mori. *J Insect Physiol* 41, pp. 17–22.

Lis H and Sharon N. 1993. Protein glycosylation: Structural and functional aspects, *Eur. J. Biochem,* 218, pp. 1-27.

Lorence A, Darszon A, Dı´az C, Lie´vano A, Quintero R & Bravo A. 1995. Delta-endotoxins induce cation channels in Spodoptera frugiperda brush border

membrane in suspension and in planar lipid bilayers. *FEBS Lett* 360, pp. 353–356.

Loris, R. 2002. Principles of structures of animal and plant lectins. *Biochimica Et Biophysica Acta-General Subjects*. 1572 (2-3) , pp. 198-208.

Loris, R., Tielker, D., Jaeger, K. E. and Wyns, L. 2003. Structural basis of carbohydrate recognition by the lectin LecB from Pseudomonas aeruginosa. *Journal of Molecular Biology*. 331 (4) , pp .861-870.

Lopez, M., Tetaert, D., Juliant, S., Gazon, M., Cerutti, M., Verbert, A. and Delannoy, P. 1999. O-Glycosylation potential of lepidopteran insect cell lines. *Biochimica Et Biophysica Acta BBA - General Subjects*, 1427 (1) ,pp. 49-61.

Lu, Y.-J. and Adang, M.J. 1996 . Conversion of *Bacillus thuringiensis* CryIAc-binding aminopeptidase to a soluble form by endogenous phosphatidylinositol phospholipase C. *Insect Biochem. Mol. Biol*. 26, pp. 33-40.

Ludwig, J. A. and Weinstein, J. N. 2005 . Biomarkers in cancer staging, prognosis and treatment selection. *Nature Reviews Cancer*. 5 (11) , pp. 845-856.

Luo, K., Banks, D., and Adang, M.J. 1999. Toxicity, binding, and permeability analyses of four *Bacillus thuringiensis* Cry 1 d-endotoxins to brush border membrane vesicles of *Spodoptera exigua* and *Spodoptera frugiperda*. *Appl. Environ. Microbiol*, 65, pp. 457-464

Luo, K., McLachlin, J., Brown, M.R. and Adang, M.J. 1999 . Expression of a glycosylphosphatidylinositol-linked *Manduca sexta* aminopeptidase N in insect cells. *Protein Expr. Purification* 17, pp. 113-122.

Luo, K., Sangadala, S., Masson, L., Mazza, A., Brousseau, R., and Adang, M.J. 1997 . The *Heliothis virescens* 170 kDa aminopeptidase function as

"Receptor A" by mediating specific *Bacillus thuringiensis* Cry1A d -endotoxin binding and pore formation. *Insect Biochem. Molec.* Biol. 27, pp. 735-743.

Luo, K., Tabashnik,, B.E., and Adang, M.J. 1997 . Binding of *Bacillus thuringiensis* Cry1Ac toxin to aminopeptidase in susceptible and resistant *Plutella xylostella*. *Appl. Environ. Microbiol*. 63, pp. 1024-1027.

Lv, Y., Tang, Y., Zhang, Y., Xia, L., Wang, F., Ding, X., 2011 The role of b20-b21 loop structure in insecticidal activity of Cry1Ac toxin from *Bacillus thuringiensis*. *Curr Microbiol* 62, pp. 665–670.

de maagd, R.A., Alejandra, B. and Crickmore, N. 2001. How the *Bacillus thuringiensis* has evolved specific toxins to colonize the insect world. *Trends in Genetics*, 17 (4), pp. 193-199.

de Maagd R A, Bakker P L, Masson L, Adang M J, Sangadala S, Stiekema W, Bosch D. 1999 . Domain III of the *Bacillus thuringiensis* delta-endotoxin Cry1Ac is involved in binding to Manduca sexta brush border membranes and to its purified aminopeptidase N. *Molecular Microbiology*, 31, (2), pp. 463-471.

de maagd R. A, Bravo A, Berry C, Crickmore N, Schnepf H.E. 2003 Structure, Diversity, and Evolution of Protein Toxins from Spore- Forming Entomopathogenic Bacteria. *Annu. Rev. Genet*, 37, pp. 409-33.

de Maagd, R.A., Weemen-Hendriks, M., Stiekema, W., and Bosch, D. 2000 Domain III substitution in *Bacillus thuringiensis* delta-endotoxin Cry1C domain III can function as a specific determinant for *Spodoptera exigua* in different,but not all, Cry1-Cry1C hybrids. *Appl Environ Microbiol* 66, pp. 1559–1563.

Marchler-Bauer A. 2013, *CDD:* conserved domains and protein three-dimensional structure., *Nucleic Acids Res*. 41 (D1), pp. 384-52

Mariño K , Bones J , Kattla J a J and Rudd PM. 2010. A systematic approach to protein glycosylation analysis: a path through the maze. *Nature Chemical Biology* , 6, pp. 713–723.

Martins, É.S., Monnerat, R.G., Queiroz, P.R., Dumas, V.F., Braz, S.V., de Souza Aguiar, R.W., Gomes, A.C.M.M., Sánchez, J., Bravo, A. and Ribeiro, B.M. Midgut GPI-anchored proteins with alkaline phosphatase activity from the cotton boll weevil Anthonomus grandis can be the putative receptors for the Cry1B protein of *Bacillus thuringiensis*. *Insect Biochemistry and Molecular Biology,* In Press, Uncorrected Proof

Masson, L., Lu, Y.-J., Mazza, A., Brousseau, R., and Adang, M.J. 1995. The CryIAc receptor purified from *Manduca sexta* displays multiple specificities. *J. Biol. Chem*, 270, pp. 20309-20315.

Masson, L. Mazza, A., Sangadala, S., Adang, M.J., Brousseau, R. 2002 . Polydispersity of *Bacillus thuringiensis* Cry1 toxins in solution and its effect on receptor binding kinetics. *Biochim. Biophys. Acta*, 1594, pp. 266-275.

Matsumoto, H., Shinzaki, S., Narisada, M. 2010 . Clinical application of a lectin-antibody ELISA to measure fucosylated haptoglobin in sera of patients with pancreatic cancer. *Clinical Chemistry and Laboratory Medicine*. 48 (4) , pp. 505-512.

Mccoy, J. P., Varani, J. and Goldstein, I. J. 1983 . Enzyme-Linked-Lectin Assay Ella - Specific Enzyme Assay for Detection and Quantitation of Terminal Carbohydrate Groups. *Federation proceedings*. 42 (4) , pp. 933-933.

Mestecky J, Tomana M, Moldoveanu Z, Julian B.A, Suzuki H, Matousovic K, Renfrow M.B, Novak L, Wyatt R.J, Novak J. 2008. Role of Aberrant Glycosylation of IgA1 Molecules in the Pathogenesis of IgA Nephropathy. *Kidney Blood Press Res*, 31,pp. 29–37

Michelakakis, H., Moraitou, M., Mavridou, I. and Dimitriou, E. 2009 . Plasma lysosomal enzyme activities in congenital disorders of glycosylation, galactosemia and fructosemia. *Clinica Chimica Acta,* 401 (1-2) , pp. 81-83.

Moar, W.J., Pustzai-Carey, H. Van Fassen, D. Bosch, R. Frutos, C. Rang, K. Luo, and Adang, M.J. 1995. Development of *Bacillus thuringiensis* CryIC resistance by *Spodoptera exigua* Hubner Lepidoptera: Noctuidae . *Appl. Environ. Microbiol*. 6, pp. 2086-2092.

Moreira, I.S., Fernandes, P.A. and Ramos, M.J. 2007 . Hot spots-A review of the protein-protein interface determinant amino-acid residues. *Proteins-Structure Function and Bioinformatics,* 68 (4) , pp.803-812.

Moremen K.W, Tiemeyer M and Narin A.V. 2012. Vertebrate protein glycosylation: diversity, synthesis and function, *Nature reviews*, 13, pp. 448-462.

Morrison, K.L. and Weiss, G.A. 2001 . Combinatorial alanine-scanning. *Current Opinion in Chemical Biology*, 5 (3) , pp.302-307.

Morse R.J, Yamamoto T, and Stroud R.M. 2001. Structure of Cry2Aa Suggests an Unexpected Receptor Binding Epitope. *Structure.* 9, pp. 409-417.

Mun˜oz-Garay C, Sa´nchez J, Darszon A, de Maagd RA, Bakker P, Sobero´n M & Bravo A. 2006. Permeability changes of Manduca sexta midgut brush border membranes induced by oligomeric structures of different Cry toxins. *J Membr Biol,* 212, pp. 61–68.

Murray, E., Stock, C., Eberle, M., Sekar, V., Rocheleau, T., and Adang, M.J. 1991 . Analysis of unstable RNA transcripts of insecticidal protein genes of *Bacillus thuringiensis* in transgenic plants and electroporated protoplasts. *Plant Mol. Biol.* 16, pp. 1035-1050

mcNall, R. and Adang, M.J. 2003. Identification of novel *Bacillus thuringiensis* Cry1Ac binding proteins in *Manduca sexta* midgut through proteomic analysis. *Insect Biochem. Molec. Biol*, 33, pp. 999-1010.

Narita Fumitake Gejyo. 2008. Pathogenetic significance of aberrant glycosylation of IgA1 in IgA nephropathy. *Clin Exp Nephrol*, 12, pp. 332–338

Oltean D.I, Pullikuth A. K, Lee H-K, Gill S S. 1999 Partial Purification and Characterization of *Bacillus thuringiensis* Cry1A Toxin Receptor A from *Heliothis virescens* and Cloning of the Corresponding cDNA. *Applied and Environmental Microbiology*. 65, (11), pp. 4760-4766.

Otto Michael. 2009. *Staphylococcus epidermidis*- the 'accidental' pathogen. *Nature reviews Microbiology*, 7, pp. 555-567.

Pacheco, S., Gómez, I., Gill, S.S., Bravo, A. and Soberón, M. 2009 Enhancement of insecticidal activity of *Bacillus thuringiensis* Cry1A toxins by fragments of a toxin-binding cadherin correlates with oligomer formation. *Peptides*, 30 (3) , pp. 583-588.

Pardo-López, L., Muñoz-Garay, C., Porta, H., Rodríguez-Almazán, C., Soberón, M. and Bravo, A. 2009. Strategies to improve the insecticidal activity of Cry toxins from *Bacillus thuringiensis*. *Peptides*, 30 (3), pp. 589-595.

Pardo-Lopez L, Soberon M & Bravo A. 2012 *Bacillus thuringiensis* insecticidal three-domain Cry toxins: mode of action, insect resistance and consequences for crop protection. *FEMS Microbiol Rev* .pp. 37 –22

Park Y, Abdullah MA, Taylor MD, Rahman K, Adang MJ 2009 . Enhancement of *Bacillus thuringiensis* Cry3Aa and Cry3Bb toxicities to coleopteran larvae by a toxin-binding fragment of an insect cadherin. *Appl Environ Microbiol*.75 (10), pp. 3086-92.

Perera, O.P., Willis, J.D., Adang, M.J. and Jurat-Fuentes, J.L. 2009. Cloning and characterization of the Cry1Ac-binding alkaline phosphatase HvALP from Heliothis virescens. *Insect Biochemistry and Molecular Biology,* 39 (4), pp. 294-302.

Perez-Villar, J and Hill, R.L . 1999. The structure and assembly of secreted mucins. *Journal of Biological Chemistry,* 274, pp. 31751-31754.

Perera OP, Willis JD, Adang MJ, Jurat-Fuentes JL. 2009 . Cloning and characterization of the Cry1Ac-binding alkaline phosphatase HvALP from *Heliothis virescens*. *Insect Biochem Mol Biol,* 39 (4), pp. 294-302.

Pigott, C.R., and Ellar, D.J. 2007. Role of receptors in *Bacillus thuringiensis* crystal toxin activity. *MicrobiolMol Biol Rev* 71, pp. 255–281.

Rajamohan, F., Alzate, O., Cotrill, J.A., Curtiss, A., and Dean, D.H. 1996 Protein engineering of *Bacillus thuringiensis* delta-endotoxin: mutations at domain II of CryIAb enhance receptor affinity and toxicity toward gypsy moth larvae. *Proc Natl Acad Sci USA* 93, pp. 14338–14343.

Rakus J F. and Mahal L. K. 2011. New Technologies for Glycomic Analysis: Toward a Systematic Understanding of the Glycome, *Annu. Rev. Anal. Chem*. 4, pp. 367–92

Ramachandran, S., Buntin, G.D., All, J.N., Tabashnik, B.E., Raymer, P.L., Adang, M.J., Pulliam, D.A., Stewart, Jr, C.N. 1998. Survival, development, and oviposition of resistant diamondback moth Lepidoptera: Plutellidae on transgenic canola producing a *Bacillus thuringiensis* toxin. *J. Econ. Entomol*. 91, pp. 1239-1244.

Rausell C, Garcı´a-Robles I, Sa´nchez J, Mun˜oz-Garay C, Martı´nez-Ramı´rez AC, Real MD & Bravo A. 2004a . Role of toxin activation on binding and pore formation activity of the *Bacillus thuringiensis* Cry3 toxins in membranes of Leptinotarsa decemlineata . *Biochem Biophys Acta* 1660, pp. 99–105.

Rausell C, Mun˜oz-Garay C, Miranda-CassoLuengo R, Go´mez I, Rudin˜o-Pin˜era E, Sobero´n M & Bravo A. 2004b. Tryptophan spectroscopy studies and black lipid bilayer analysis indicate that the oligomeric structure of Cry1Ab toxin from *Bacillus thuringiensis* is the membrane-insertion intermediate. *Biochemistry* 43, pp. 166–174.

Rausell C, Pardo-Lo´ pez L, Sa´nchez J, Mun˜oz-Garay C, Morera C, Sobero´n M & Bravo A. 2004c. Unfolding events in the water-soluble monomeric Cry1Ab toxin during transition to oligomeric pre-pore and membrane inserted pore channel. *J Biol Chem* 279, pp. 55168–55175.

van Rensburg, J.B.J. 2007. First report of field resistance by stem borer Busseola fusca Fuller to Bt-transgenic maize. *S Afr J Plant Soil*, 24 (3), pp. 147-151.

Rodrı´guez-Almaza´n C, Reyes EZ, Zun˜iga-Navarrete F, Mun˜oz- Garay C, Go´mez I, Evans AE, Likitvivatanavong S, Bravo A, Gill SS & Sobero´n M. 2012. Cadherin binding is not a limiting step for *Bacillus thuringiensis* subs. israelensis Cry4Ba toxicity to Aedes aegypti larvae. *Biochem* J, 443, pp 711–717.

Rodrı´guez-Almaza´n CR, Zavala LE, Mun˜oz-Garay C, Jime´nez- Jua´rez N, Pacheco S, Masson L, Sobero´n M & Bravo A. 2009. Dominant negative mutants of *Bacillus thuringiensis* Cry1Ab toxin function as anti-toxins: demonstration of the role of oligomerization in toxicity. *PLoS ONE* 4, pp. 5545.

Rodrigo-Simon, A., Caccia, S., and Ferre, J. 2008. *Bacillus thuringiensis* Cry1Ac Toxin-Binding and Pore-Forming Activity in Brush Border Membrane Vesicles Prepared from Anterior and Posterior Midgut Regions of Lepidopteran Larvae. *Applied and Environmental Microbiology*, 47 (6), pp. 1710-1716.

Rosenfeld, R., Bangio, H., Gerwig, G.J., Rosenberg, R., Aloni, R., Cohen, Y., Amor, Y., Plaschkes, I., Kamerling, J.P. and Maya, R.B. 2007 A lectin array-based methodology for the analysis of protein glycosylation. *Journal of Biochemical and Biophysical Methods,* 70 (3),pp. 415-426.

Rostand KS and Esko JD. 1997 . Microbial adherence to and invasion through proteoglycans. *Infect Immun*. 65, pp. 1–8.

Roy A, Alper K and Yang Z. 2010. I-TASSER: a unified platform for automated protein structure and function prediction. *Nature Protocols*, vol 5, pp. 725-738.

Roy A, Jianyi Y and Yang Z. 2012. COFACTOR: an accurate comparative algorithm for structure-based protein function annotation. *Nucleic Acids Research*, vol 40, pp. W471-W477.

Saeland E, Belo A I, Mongera S, Van Die I Meijer G A, andVan Kooyk Y. 2011. Differential glycosylation of MUC1 and CEACAM5 between normal mucosa and tumour tissue of colon cancer patients. *International Journal of Cancer* accepted to print

Sangadala, S., Azadi, P., Carlson, D.H., and Adang, M.J. 2001 . Carbohydrate analyses of *Manduca sexta* aminopeptidase N, co-purifying neutral lipids and their functional interactions with *Bacillus thuringiensis* Cry1Ac toxin. *Insect Biochem, Molec. Biol.* 32, pp. 97-107.

Sangadala, S., Walters, F., English, L.H., and Adang, M.J. 1994. A mixture of *Manduca sexta* aminopeptidase and phosphatase enhance *Bacillus thuringiensis* insecticidal CryIA c toxin binding and 86Rb+-K+ leakage *in vitro*. *J. Biol. Chem,* 269, pp. 10088-10092.

Schäffer C, Graninger M, and Messner P. 2001 Prokaryotic glycosylation. *Proteomics*, 1, pp. 248–261.

Schaffer C, Messner P. 2005. The structure of secondary cell wall polymers: How Gram-positive bacteria stick their cell walls together. *Microbiology*.151, pp. 643–651.

Schenk B, Fernandez F and Waechter CJ. 2001 The ins ide and outs ide of dolichyl phosphate biosynthesis and recycling in the endoplasmic reticulum. *Glycobiology.*;11, pp. 61R–70R.

Schnepf, E., Crickmore, N., Van Rie, J., lereclus, D., Baum, J., Feitelson, J., Zeigler, D.R and Dean, D.H. 1998. *Bacillus thuringiensis* and Its Pesticidal Crystal Proteins. *Microbiology and Molecular Biology Reviews*, 62 (3), pp. 775-806.

Schofield, L., Hewitt, M. C., Evans, K., Siomos, M. A. & Seeberger, P. H. 2002 .Synthetic GPI as a candidate anti-toxic vaccine in a model of malaria. *Nature* 418, pp. 785.

Schwartz JL, Garneau L, Savaria D, Masson L, Brousseau R & Rousseau E 1993. Lepidopteran-specific crystal toxins from *Bacillus thuringiensis* form cation and anion selective channels in planar lipid bilayers. *J Membr Biol* 132, pp. 53–62.

Schwartz, J.L., Lu, Y.-J., Sohnlein, P., Brouseau, R., Laprade, R., Masson, L., and Adang, M.J. 1997 . Ion channels formed in planar lipid bilayers by *Bacillus thuringiensis* toxins in the presence of *Manduca sexta* midgut receptors. *FEBS Letters* 412, pp. 270-276.

Sekar, V., Thompson, D.V., Maroney, M.J., Bookland, R., and Adang, M.J. 1987 . Molecular cloning and characterization of the insecticidal crystal protein gene of *Bacillus thuringiensis* var. *tenebrionis*. *Proc. Natl. Acad. Sci. USA* 84, pp. 7036-7040.

Shan, S., Zhang, Y., Ding, X., Hu, S., Sun, Y., Yu, Z., *et al*. 2011 A Cry1Ac toxin variant generated by directed evolution has enhanced toxicity against Lepidopteran insects. *Curr Microbiol* 62, pp. 358–365.

Sharon N. 1980 . Carbohydrates. *Sci. Am*. 243,(5), pp. 90–116.

Shriver Z, Raguram. S and Sasisekharan R. 2004. Glycomics: A pathway to a class of new and improved therapeutics, *Nature reviews drug discovery*, 3, pp. 863-873.

Shu, C., Liu, R., Wang, R., Zhang, J., Feng, S., Huang, D., and Song, F. 2007. Improving toxicity of *Bacillus thuringiensis* strain contains the cry8Ca gene specific to *Anomala corpulenta* larvae. *Curr Microbiol* 55, pp. 492–496.

Singsit, C., Adang, M. J., Lynch, R. E., Anderson, W. F., Wang, A., Cardineau, G., and Ozias-Akins, P. 1997. Expression of a *Bacillus thuringiensis cryIA c* gene in transgenic peanuts and its efficacy against lesser cornstalk borer. *Transgenic Research*, 6, pp. 169-176.

Singsit, C., Adang, M.J., Lynch, R.E., Anderson, W.F., Wang, A., Cardineau, G., and Ozias-Akins, P. 1996 . The expression of a *Bacillus thuringiensis cryIA c* gene in somatic embryos and peanut plants against lesser cornstalk borer. *Transgene Res*, 5, pp. 1-8.

Slovin, S. F. 2003. Fully synthetic carbohydrate-based vaccines in biochemically relapsed prostate cancer: Clinical trial results with α-*N*-acetylgalactosamine-Oserine/ threonine conjugate vaccine. *J. Clin. Oncol.* 21, pp. 4292–4298.

Smart, J. D., Nantwi, P. K. K., Rogers, D. J. and Green, K. L. 2002 . A quantitative evaluation of radiolabelled lectin retention on oral mucosa in vitro and in vivo. *European Journal of Pharmaceutics and Biopharmaceutics*. 53 (3) , pp. 289-292.

Stewart Jr., C.N., Adang, M.J., All, J.N., Boerma, H.R., Cardineau, G., Tucker, D., and Parrott, W.A. 1996 . Genetic transformation, recovery and characterization of soybean *Glycine max* L. Merrill transgenic for a synthetic *Bacillus thuringiensis cryIA c* gene. *Plant Physiol,* 112, pp. 121-129.

Stewart Jr., C.N., Adang, M.J., All, J.N., Raymer, P.L., Ramachandran, S., and Parrott, W.A. 1996. Insect control and dosage effect in transgenic canola, *Brassica napus* L. Brassicaceae , containing a synthetic *Bacillus thuringiensis cryIA c* gene. *Plant Physiol,* 112, pp. 115-120.

Storer, N.P., Babcock, J.M., Schlenz, M., Meade, T., Thompson, G.D., Bing, J.W., and Huckaba, R.M. 2010. Discovery and characterization of field resistance to Bt Maize: *Spodoptera frugiperda* Lepidoptera: Noctuidae in Puerto Rico. *J Econ Entomol* 103, pp. 1031–1038.

Tabak, L.A. 2010. The role of mucin-type O-glycans in eukaryotic development. *Seminars in Cell & Developmental Biology,* 21 (6), pp. 616-621.

Tabak LA. 1995. In defense of the oral cavity, Structure, biosynthesis, and function of salivary mucins. *Annu Rev Physiol*.;57, pp. 547–564.

Tabashnik, B.E., Finson, N., Groeters, F.R., Moar, W.J., Johnson, M.W., Luo, K., and Adang, M.J. 1994 . Reversal of resistance to *Bacillus thuringiensis*. Reversal of resistance to *Bacillus thuringiensis*. *Proc. Natl. Acad. Sci. U.S.A.* 91, pp. 4120-4124.

Tabashnik, B.E., Gassman, A.J., Crowdwer, D.W., and Carriere, Y. 2008. Insect resistance to *Bt* crops: evidence versus theory. *Nat Biotechnol,* 26, pp. 199–202.

Tabashnik, B.E., Huang, F., Ghimire, M.N., Leonard, B.R., Siegfried, B.D., Randasamy, M. 2011. Efficacy of genetically modified Bt toxins against

insects with different mechanism of resistance. *Nat Biotechnol* 29, pp. 1128–1131

Taylor M.E, Drickamer K. 2006 . *Introduction to Glycobiology*. Second edition. Oxford. Oxford University Press.

Taylorpapadimitriou, J. & Epenetos, A. A. 1994. Exploiting altered glycosylation patterns in cancer- progress and challenges in diagnosis and therapy. *Trends Biotech.* 12, pp. 227–233.

Theopold, U., Rissler, M., Fabbri, M., Schmidt, O. and Natori, S. 1999 . Insect Glycobiology: A Lectin Multigene Family in Drosophila melanogaster. *Biochemical and Biophysical Research Communications,* 261 (3), pp. 923-927.

Thompson, R, Creavin, A, O'Connell, M, O'Connor, B, and Clarke, Paul. 2011. *Optimisation of the enzyme-linked lectin assay for enhanced glycoprotein and glycoconjugate analysis. Analytical Biochemistry*, 413 (2), pp. 114-122.

Tian E, Ten Hagen K G. 2009 Recent insights into the biological roles of mucin-type O-glycosylation. *Glycoconj J*, 26, pp. 325–334.

Torres, J., Lin, X. and Boonserm, P. 2008. A trimeric building block model for Cry toxins in vitro ion channel formation. *Biochimica Et Biophysica Acta BBA - Biomembranes,* 1778 (2) , pp. 392-397.

Tsarbopoulos, A; Karas, M; Strupat, K. 1994. Comparative Mapping of Recombinant Proteins and Glycoproteins by Plasma Desorption and Matrix-Assisted Laser Desorption/Ionization Mass Spectrometry. *Analytical chemistry,* 66, (S). pp. 2062-2070

Vachon V, Prefontaine G, Coux F, Rang C, Marceau L, Masson L, Broussea R, Frutos R, Schwartz JL & Laprade R. 2002. Role of helix 3 in pore formation

by *Bacillus thuringiensis* insecticidal toxin Cry1Aa. *Biochemistry* 41, pp. 6178–6184.

Vadlamudi R K, Weber E, Ji I, Ji T.H, Bulla L.A Jr. 1994. Cloning and Expresion of a Receptor for an Insecticidal Toxin of *Bacillus thuringiensis*. *The journal of Biological Chemistry*, 270, (10), pp. 5490-5494.

Varki, A. 1999. *Essentials of glycobiology.* Cold Spring Harbor, NY: Cold Spring Harbor Laboratory.

Varki, A. 1993 . Biological Roles of Oligosaccharides - all of the Theories are Correct. *Glycobiology*. 3 (2) , pp. 97-130.

Varki, A., Cummings, R.D., Esko J.D., Freeze H.H., Stanley P, Bertozzi, C.R, Hart G.W. and Etzler, M.E. 2009. *Essentials of Glycobiology*, 2nd edition, Cold Spring Harbor NY.

Wacker, M., Feldman, M. F., Callewaert, N. 2006. Substrate specificity of bacterial oligosaccharyltransferase suggests a common transfer mechanism for the bacterial and eukaryotic systems. *Proc. Natl. Acad. Sci. U.S.A.* 103 (18) , pp. 7088-7093.

Wacker, M., Linton, D., Hitchen, P. G. 2002. *N*-Linked Glycosylation in *Campylobacter jejuni* and its Functional Transfer into *E. coli*. *Science* 298 (5599) , pp. 1790-1793.

Wanchoo, A., Lewis, M. W. and Keyhani, N. O. 2009 . Lectin mapping reveals stage-specific display of surface carbohydrates in in vitro and haemolymph-derived cells of the entomopathogenic fungus Beauveria bassiana. *Microbiology-Sgm*. 155, pp. 3121-3133.

Watson R.J, Millichap P, Joyce S.A, Reynolds S and Clarke D.J. 2010. The role of iron uptake in pathogenicity and symbiosis in *Photorhabdus luminescens* TT01. *BMC Microbiology*, 10, pp. 177.

Weerapana, E. and Imperiali, B. 2006. Asparagine-linked protein glycosylation: from eukaryotic to prokaryotic systems. *Glycobiology* 16 (6) , pp. 91R-101.

Wei, J.Z., Hale, K., Carta, L., Platzer, E., Wong, C., Fang, S-C., Aroian, R.V. 2003. *Bacillus thuringiensis* crystal proteins that target nematodes. *PNAS Microbiology,* 100 (5), pp. 2760-2765.

Wells, J.A. 1990. Additivity of Mutational Effects in Proteins. *Biochemistry,* 29 (37) , pp.8509-8517.

Williams, S.J. and Davies, G.J. 2001. Protein–carbohydrate interactions: Learning lessons from nature. *Trends Biotechnol*, 19, pp. 356–362.

Wilson, I.B.H. 2002. Glycosylation of proteins in plants and invertebrates. *Current Opinion in Structural Biology,* 12 (5) , pp. 569-577.

Wu, S.J., Koller, C.N., Miller, D.L., Bauer, L.S., and Dean, D.H. 2000 Enhanced toxicity of *Bacillus thuringiensis* Cry3A delta-endotoxin in coleopterans by mutagenesis in a receptor binding loop. *FEBS Lett* 473, pp. 227–232.

Wu, J.-Y., Zhao, F.-Q., Bai, J., Deng, G., Qin, S., and Bao, Q.-Y. 2007. Adapatative evolution of cry gene in *Bacillus thuringiensis*: Implications for their specificity determination. *FEBS Lett* 5, pp. 102–110.

Xiang W F, Qiu X L, Zhi D X, Min Z X, Yuan L, Quan Y Z. 2009. N546 in β18-β19 loop is important for binding and toxicity of the *Bacillus thuringiensis* Cry1Ac toxin. *Journal of Invertebrate Pathology*. 101, 119-123.

Yang Z. 2008. I-TASSER server for protein 3D structure prediction. *BMC Bioinformatics*, 9, 40.

Yigzaw Y, Hinckley P, Hewig A, Vedantham G. 2009 Ion exchange chromatography of proteins and clearance of aggregates. *Curr Pharm Biotechnol*.10 (4), pp. 421-6.

Young, N. M., Brisson, J., and Kelly, J. 2002. Structure of the *N*-Linked Glycan Present on Multiple Glycoproteins in the Gram-negative Bacterium, *Campylobacter jejuni. J. Biol. Chem.* 277 (45) , pp. 42530-42539.

Zarschler K, Janesch B, Pabst M, Altmann F, Paul M, Christina S. 2010. Protein tyrosine O-glycosylation—A rather unexplored prokaryotic glycosylation system. *Glycobiology*, 20, (6), pp. 787–798.

Zhang R, Hua G, Andacht TM, and Adang MJ. 2008. A 106-kDa aminopeptidase is a putative receptor for *Bacillus thuringiensis* Cry11Ba toxin in the mosquito Anopheles gambiae. *Biochemistry.* 47 (43), pp. 11263-72.

Zhang, S., Cheng, H., Gao, Y., Wang, G., Liang, G. and Wu, K. 2009. Mutation of an aminopeptidase N gene is associated with Helicoverpa armigera resistance to *Bacillus thuringiensis* Cry1Ac toxin. *Insect Biochemistry and Molecular Biology*, 39 (7), pp. 421-429.

Zhang X, Candas m, Griko NB, Rose-Young L, Bulla LA Jr. 2005. Cytotoxicity of *Bacillus thuringiensis* Cry1Ab toxin depends on specific binding of the toxin to the cadherin receptor BT-R expressed in the insect cells. *Cell Death and Differentiation*, 12, pp. 1407-1416.

Zhang, X., Candas, M., Griko, N.B., Taussig, R., and Bulla, L.A., Jr. 2006. A mechanism of cell death involving an adenylyl cyclase/PKA signaling pathway is induced by the Cry1Ab toxin of *Bacillus thuringiensis*. *Proc Natl Acad Sci USA* 103, pp. 9897–9902.