BIOPHYSICAL CHARACTERISATION OF MEASLES VIRUS RECEPTORS

Andrew Peter Herbert

A thesis submitted for the degree of Doctor of Philosophy

The University of Edinburgh

September 2002

Unless otherwise stated, the work described in this thesis is my own work and has not been sumitted in whole or in part for a degree or other qualifiaction at this or any other university.

Acknowledgements

I would like to thank my supervisor, Dr. Paul Barlow for allowing me the freedom to explore my own ideas, and for providing help when needed. The members of Labs 120 and 38 for providing an enjoyable working environment. The Barlow group, for providing answers when I needed them. Stanislas Blein, four our endless discussions.

In particular I wish to thank Karen for having the patience to help me through the difficult times, and Suzie and Geoff, for providing endless inspiration, and encouragement.

Abstract

Measles virus (MV) is an important human pathogen that still claims almost one million lives annually, primarily in the developing world. To date two human proteins have been identified that can serve as the receptor for measles virus, namely membrane cofactor protein (MCP) and signalling lymphocyte activation molecule (SLAM). It is thought that a more through understanding of the sructure and behaviour of the interacting regions of these receptors, will facilitate a more detailed understanding of the pathogenesis of MV. This thesis describes efforts to clone, express and characterise the relevant fragments of these receptors.

The natural function of MCP is to prevent activation of complement on self cells. MCP contains four extracellular complement control protein (CCP) modules. The region of MCP that interact with MV has previously been identified to be the first two N-terminal CCP modules. MCP carries a functionally critical N-glycan on its second module (MCP2) and the crystal structure of MCP12 (expressed in mammalian cells) implies a structural role for this carbohydrate moiety. MCP12 and MCP2 were successfully cloned and expressed as secreted proteins in *Pichia pastoris* at levels sufficient for biophysical characterisation. Proton and ^{15}N Nuclear magnetic resonance spectroscopy (NMR) studies, differential scanning calorimetry and mass spectrometry were employed to investigate the recombinant proteins. Both the natural sequences of MCP12 and MCP2 fragments expressed in *P. pastoris* were hyperglycosylated, and therefore these recombinant fragments were investigated in both glycosylated and deglycolylated forms in order to study the effect of the glycans on module 2 of MCP, and with a view to detailed structural and dynamic studies using NMR. The data obtained with the *P. pastoris*-expressed fragments are not inconsistant with the hypothesis that the N-glycan on module 2 of the wild-type MCP is required for this module to be properly folded. After extensive investigations, it was concluded that the inability of *P. pastoris* to attach the natural glycan is a barrier to more detailed structural studies. In an effort to produce MCP2 with a better defined glycosylation profile, efforts were made to express, this fragment in insect cells which are known to produce a 'more natural' glycosylation profile on proteins. Moreover, the MCP2 fragment was also expressed in *E. coli*, in order to produce a form of MCP2 that had never been glycosylated.

Following the discovery that SLAM may be the more physiologically relevant receptor for MV, it had been established that the first V-type immunoglobulin-type domain (SLAM1) is necessary and sufficient for interaction with the virus. SLAM1 was successfully cloned from a human cDNA library using touchdown PCR, and efforts were made towards its expression in *P. pastoris*. Expression levels of 20 mg/l of induction medium were achieved, however the recombinant SLAM1 was also hyperglycosylated by *P. pastoris* which has hampered its purification.

Contents

Table of Figures

Abbreviations

AIEX	Anion exchange chromatography
BMG	Buffered minimal glycerol
BMM	Buffered minimal methanol
CCP module	Complement control protein module
CR1	Complement receptor type 1
CR2	Complement receptor type 2
DAF	Decay accelerating factor
DSC	Differential scanning calorimetry
EndoHf	Endoglycosidase Hf
ESMS	Electospray mass spectrometry
fB	Factor B
fD	Factor D
fH	Factor H
GSH	Reduced glutathione
GSSG	Oxidised glutathione
HIC	Hydrophobic interaction chromatography
HPLC	High performance liquid chromatography
HSQC	Heteronuclear single quantum correlation spectroscopy
MAC	Membrane attack complex
MASP	Mannose binding protein associated serine protease
MBP	Mannose binding protein
MCP	Membrane cofactor protein
NMR	Nuclear magnetic resonance
NOE	Nuclear Overhauser effect
NOESY	Nuclear Overhauser effect spectroscopy
PBS	Phosphate buffered saline
βME	[3-mercaptoethanol
RCA	Regulators of complement activation
RP HPLC	Reverse phase high performance liquid chromatography
SDS PAGE	Sodium dodecylsulphate polyacrylamide gel electrophoresis
SEC	Size exclusion chromatography
TFA	Trifluoroacetic acid
TOCSY	Total correlation spectroscopy
VCP	Vaccinia virus complement control protein

1. Introduction

1.1. Measles Virus

1.1.1. Preliminary remarks

Measles virus has been anecdotally described as the biggest killer of children in history. Whether or not this is true it has undoubtedly been responsible for countless millions of deaths. Measles virus is historically significant and is thought to have been recognised as a disease distinct from smallpox for over 1000 years (Griffin et al., 1994). It is also said that measles is the disease that first introduced the idea of lifelong immunity to medicine, after Panum's observations of a measles epidemic in the Faroe Islands in 1846 (republished as Panum, 1939). He found that islanders who had been infected with measles virus in an earlier measles epidemic remained healthy during the current outbreak. Given the historical significance of measles virus, it is hardly surprising that it has been the subject of such a significant level of research effort.

1.1.2. Measles vaccination:

Since the introduction of an effective attenuated live vaccine based on the Edmonston strain, the incidence of measles has dramatically declined, especially in the developed world (Stalkup, 2002), and the virus is now on the WHO target list for eradication (Wild, 1999). Despite these efforts, however, measles still affects around 50 million people, and is responsible for around one million deaths annually, predominantly in the developing world (Omer, 1999).

One of the major failings of the current vaccine is that its period of protection is necessarily limited. A neonate's initial measles protection comes from transplacentally transferred maternal antibodies, which persist in the blood stream for several months. During this period, vaccination is not a viable option since the vaccine would be rapidly cleared from the blood by the same maternal antibodies that provide the initial immunity (Osterhaus et al., 1994). It is not until this transient immunity has waned that vaccination can proceed, therefore there is inevitably a period of almost a year during which there can be no measles vaccination. There will probably always be clinical cases of measles under this immunisation regime, especially in developing countries, where the incidence of vaccination is lower (El Kasmi et al., 2001). The relative success of this vaccination strategy in the developed world appears to be due to a combination of factors, not least of which is the high rate of vaccination uptake, which over time causes a gradual depletion of the viral reservoir, and provides a herd immunity (Schlenker et al., 1992). This combined with better isolation and treatment of patients with the clinical disease reduces the chances of an individual coming into contact with the virus in the first place.

There also appears to be another problem with the current measles vaccination strategy in the UK, namely that of public concern over the safety of the measles vaccine (Owens, 2002). In the majority of developed countries, the measles vaccine is given in conjunction with those for mumps and rubella as a single dose vaccine known as MMR. However since the publication of some studies that have indicated (albeit fairly inconclusively) that the MMR vaccine may be linked to autism and

bowel disease (Morris et al., 2002; Uhlmann et al., 2002), public confidence in the vaccine has declined significantly. This has resulted in a small but significant drop in the uptake of the vaccine in the UK, which may result in rising rates of clinical measles cases with all the associated problems. The reduction in vaccine uptake also appears to be clustered, and it may be that some regions have a rate of vaccination below that which is necessary to maintain the herd immunity. These problems could be addressed by the use of public information campaigns aiming to reassure parents that the vaccine is safe. This could well be the quickest (and probably the cheapest) solution to this problem. It may also be wise, however, to try developing a new vaccine that has a more favourable distribution of possible times of administration, and would not be tainted with public suspicion that it is unsafe. Such a strategy may serve both to improve the overall levels of measles immunity, and to improve public confidence in the vaccine, and thus help uptake rates to recover to their previous levels. Another reason why it is important to maintain public confidence in the safety of the MMR vaccine is to prevent the emergence of a widespread suspicion of vaccines in general, with all of the associated public health implications.

1.1.3. Disease Progression

The clinical course of an uncomplicated measles virus infection generally follows a standard pattern (for a review see Katz, 1995): the incubation period, the prodromal period, the exanthem period, and the healing of the rash. The incubation period of a typical measles virus infection lasts for 10 to 12 days post-infection, during which time there are virtually no outward symptoms. During this period the virus infects

the endothelial cells of the upper respiratory tract, spreads to the lymph nodes and multiplies. It spreads throughout the bloodstream resulting in the primary viraemia, followed subsequently by a more intense secondary viraemia several days later.

Following the incubation period the patient enters the prodromal phase, which is when the symptoms first appear. The initial symptoms somewhat resemble those of the common cold. Over the next two days, however, they intensify and additional symptoms, including nausea, a rasping cough, and conjunctivitis, may often start to develop. The initial symptoms are followed by the appearance of Koplik's spots shown in Figure 1.1, which are small raised white spots found on the mucous membranes of the mouth and throat that resemble grains of salt. The appearance of Koplik's spots is a characteristic of all clinical measles cases.

Figure 1.1 Koplik's spots in the mouth of a measles patient. Koplik's spots appear like grains of salt on the oral mucosa, and are one of the first noticeable manifestations of the disease. Taken from www.dentistry.leeds.ac.uk/oralpath/viruses/viral%20infections/MMR.htm

The next phase of the disease is the exanthem period, during which the characteristic measles rash starts to develop. The rash is a consequence of the immune response to measles rather that a result of the virus itself, as demonstrated by the lack of presentation of the rash in severely immunocompromised patients (Kaplan et al., 1992; Markowitz et al., 1988). The rash usually appears between 14 and 18 days post-infection and develops initially on the face or behind the ears, spreading over the next two to three days to include the torso and limbs. During this period the body temperature peaks and the Koplik's spots begin to subside. The rash generally lasts less than a week and as it fades the skin tends to peel or flake. The cold symptoms and conjunctivitis usually disappear shortly after the appearance of the rash, although the cough will often persist for several days longer.

1.1.4. The Genetics of a Measles Virus Infection

Measles virus is an enveloped negative (-) sense RNA virus of the order mononegavirales. It is a member of the paramyxoviridae family, subfamily paramyxovirinae and of the genus morbillivirus, and is closely related to several other important viruses including canine distemper and rinderpest viruses. The virus has a 15.9 Kb minus (-) sense RNA genome (Baczko et al., 1983; Dunlap et al., 1983), which consists of six genes in the following order:

N	**60 kDa**
P/C/V	**70 kDa**
M	**37 kDa**
F	**60 kDa**
H	**80 kDa**
L	**~250 kDa**

The N gene encodes for the nucleoprotein, which associates tightly with the RNA genome to form the nucleocapsid (Lund et al., 1984). The P gene encodes for three proteins, (P C and V) (Bellini et al., 1985; Gombart et al., 1992) that when combined with the L gene-product provide the viral RNA-dependent RNA polymerase activity (Horikami et al., 1994). The V and C proteins have also been demonstrated to act as virulence factors (Patterson et al., 2000). The M gene encodes the matrix protein (Bellini et al., 1986), and the F (Buckland et al., 1987; Curran et al., 1988) and H (Gerald et al., 1986) genes encode for the two viral envelope glycoproteins, *i.e.* the fusion protein and haemagglutinin respectively. The gene order of the measles virus genome is: 3' leader - N - P/C/V - M - F - H - L - trailer 5'–. It was originally determined by northern analysis of the mRNAs isolated from measles-infected virus cells (Dowling et al., 1986) and this was subsequently confirmed by sequencing (Takeda et al., 1999). It also appears that the gene order plays a significant role in controlling the levels of expression of each gene product. The RNA-dependant RNA polymerase generally binds to the 3' end of the genome and transcribes each gene in sequence, halting transcription between each gene. A proportion of polymerase seems to 'fall off' the template after the transcription of each gene is finished, and then cannot be replaced. This results in decreased levels of transcription for genes at the 5' end of the genome, compared to the levels for genes at the 3' end (Cattaneo et al., 1987). This phenomenon may be reflected in differential expression levels for the various viral proteins.

1.1.5. Structure of a Measles Virus Virion

Measles virions vary in size from approximately 100 nm to 250 nm in diameter, and are roughly spherical in shape, although they do appear to be somewhat pleomorphic (Horikami et al., 1995). Figure 1.2 shows an electron micrograph of a measles virion. The schematic representation of a measles virion shown in Figure 1.3 shows the virus's key features. It consists of a host cell-derived outer lipid bilayer, which contains the two envelope glycoproteins haemagglutinin (H) and fusion (F) The H protein has a dual role: it both mediates the initial attachment of the virion to the cell and it plays an auxiliary role in the fusion process (Nussbaum et al., 1995). Located just on the inner side of the membrane is the matrix (M) protein, which provides much of the virion's structural stability. The M protein is thought to interact directly with the cytoplasmic domains of the envelope glycoproteins. An idea that is demonstrated by the ability of M protein to specify polarised sorting of these proteins in epithelial cells (Naim et al., 2000). Within the region enclosed by the M protein lies the nucleocapsid, which contains the viral genome. Also present inside the virion are the proteins that provide the viral RNA polymerase activity - the large (L) protein and the phosphoprotein (P).

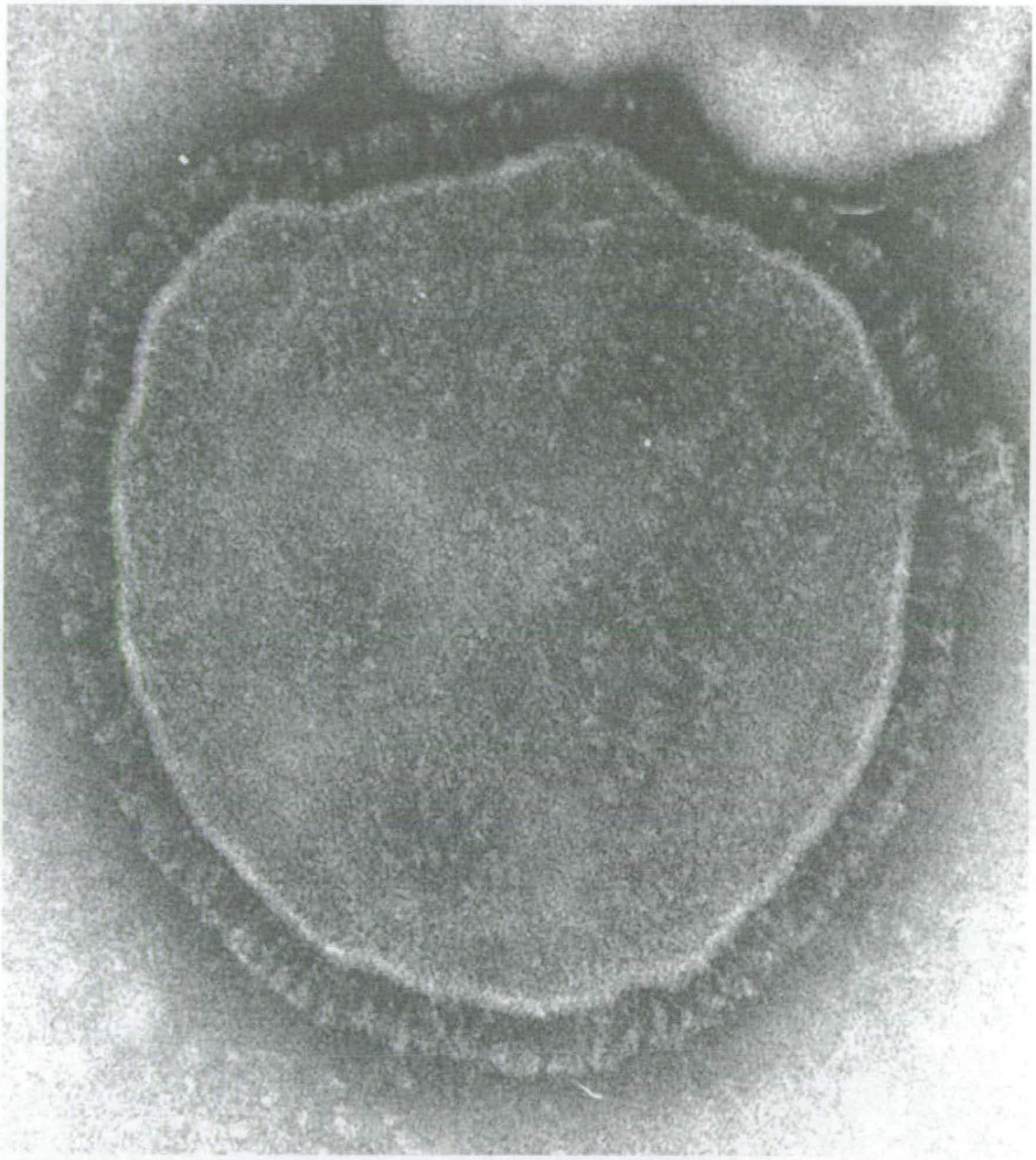


Figure 1.2 Electron micrograph of a measles virion

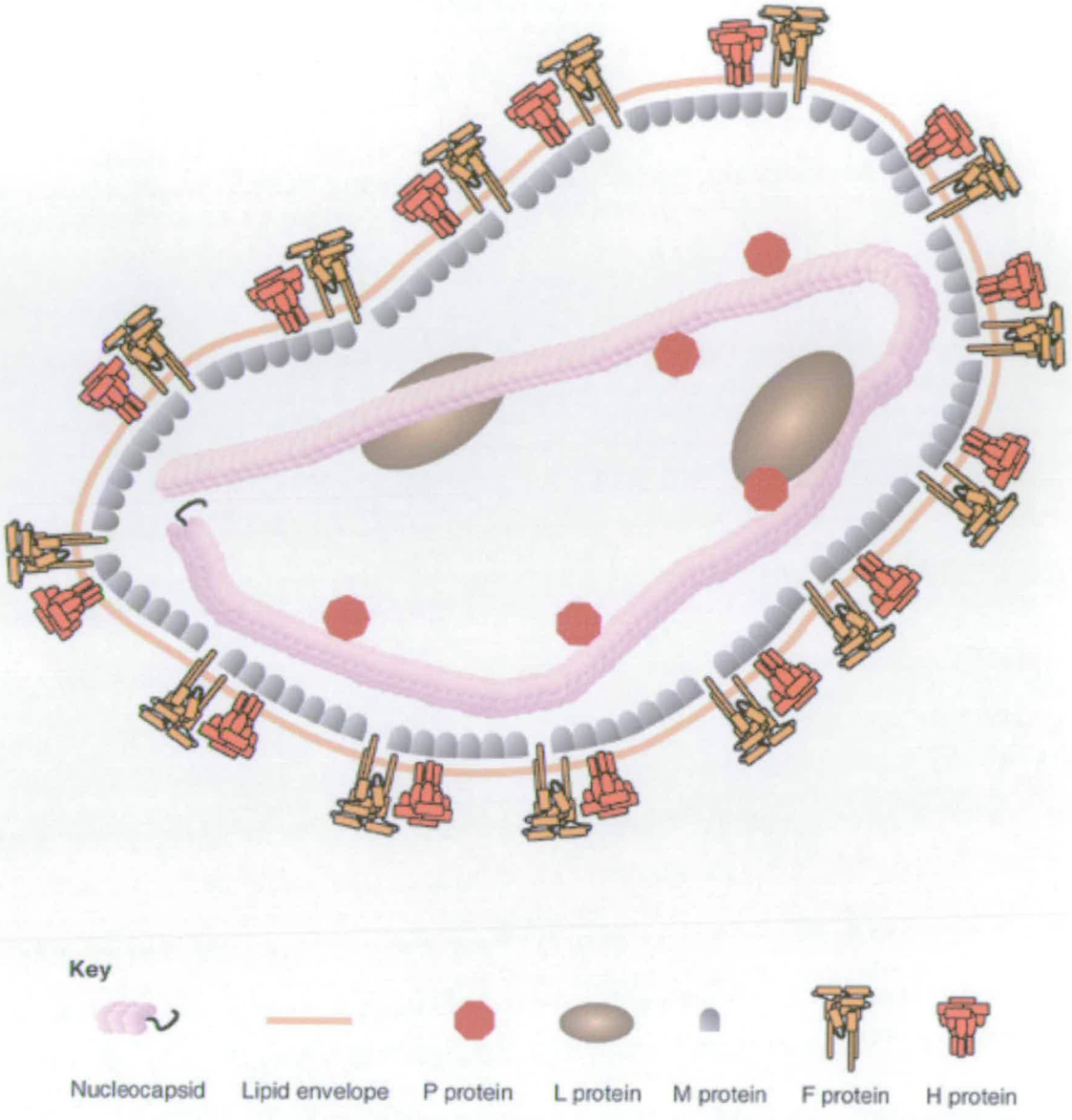


Figure 1.3 Schematic representation of a measles virion structure showing the virion's proteins and major features (taken from Schneider-Schaulies et al., 2002).

1.1.6. The Infectious Cycle

The initial interaction between measles virus and a susceptible cell is the adsorption of the virus to the cell surface. Following the initial binding, the viral envelope fuses with the plasma membrane in a process that appears to require the formation of a

ternary complex between the F and H proteins and the cellular receptor. The fusion protein is synthesised as a precursor protein termed F_0 and is subsequently cleaved and disulphide bonded yielding the active heterodimeric form of the protein (F_1-F_2) (Bolt et al., 1998). Following binding of the H protein to the cellular receptor, a conformational change is induced in F_1-F_2 that facilitates insertion of the hydrophobic fusion domain of F_1-F_2 into the host cell. The fusion process has been shown to require the action of both the H and F proteins. Subsequent interactions between α-helices of the fusion protein result in membrane fusion (Baker et al., 1999; Samuel et al., 2001). Whilst the action of the fusion protein seems to predominate, other host proteins, possibly including moesin, may also be involved in viral fusion. Moesin became a candidate since it was demonstrated that it is present at the sites of MV infection and can form a complex with MCP; and anti-moesin antibodies can block measles infection under certain circumstances; (Schneider-Schaulies et al., 1995b). The schematic representation of the fusion process illustrated in Figure 1.4 shows MV using the protein SLAM as the cellular receptor (see below). This general scheme would hold true also for membrane cofactor protein (MCP) receptor usage (see below). The overall fusion process is, however, still far from fully understood.

Following membrane fusion the nucleocapsid is released into the cytoplasm where the rest of the viral life cycle takes place. The virus initiates the synthesis of its own proteins, using the RNA-dependant RNA polymerase packaged within the virion. This activity provided by a P-L complex appears to only transcribe encapsidated RNAs. The resulting transcripts are processed at the 5' and the 3' ends by the host

cell in the usual way. The cellular machinery is also responsible for the translation of the processed viral transcripts.

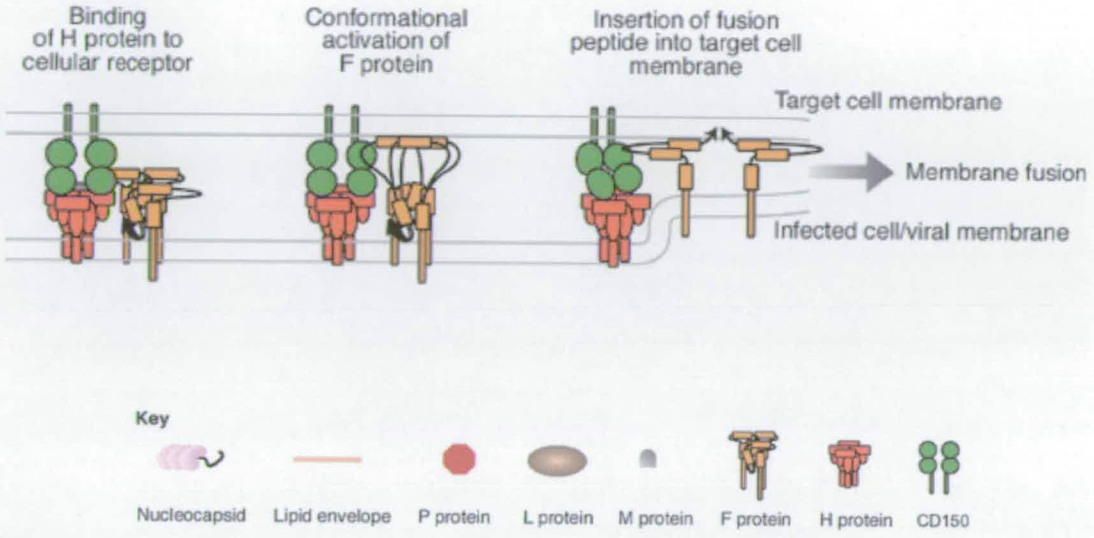


Figure 1.4 Schematic representation of the events involved in the fusion of MV with the cell membrane (taken from Schneider-Schaulies et al., 2002).

It would appear that initially the RNA-dependent RNA polymerase produces only mRNA transcripts. In order for replication to occur, however, the virus also needs to produce a plus (+) sense copy of the complete RNA genome to serve as a template for the production of further copies of the minus (-) sense RNA genome. Based on data obtained on other organisms including Vesicular Stomatitis Virus (VSV), it is thought that the same RNA-dependant RNA polymerase is responsible for the synthesis of the mRNA transcripts, and both the (+) and (-) strand genome length RNAs (Banerjee, 1987). This seems to be achieved through a switching mechanism, whereby the concentration of the viral nucleoprotein upon reaching a critical level triggers the synthesis of whole copies of (+) sense genome. Although the nature of this switching mechanism is not well understood, it is plausible because genome

synthesis appears to be coupled to the encapsidation process. The secondary genomes produced by this mechanism also appear able to serve as the templates for further rounds of transcription, thus amplifying viral mRNAs as well as facilitating viral replication.

1.1.7. Receptors for Measles Virus

As has already been suggested, the initial event in any measles infection (attachment of the virus to the cell) is achieved by an interaction between the H protein and its cellular receptor. Considerable effort was therefore put into identifying the receptor because blocking this interaction could potentially provide a very effective treatment. The breakthrough finally came when (Naniche et al., 1993) discovered that recombinant murine B cells expressing MCP were found to bind measles virus (Edmonston Strain), and that anti-MCP antibodies were able to prevent this binding. This report was further supported when it was reported almost simultaneously that MCP^+ recombinant chinese hamster ovary (CHO) cells could both bind measles virus and support infection, whereas mock transfected cells could not (Dorig et al., 1993). It was also demonstrated that polyclonal anti-MCP antisera could block this infection(Dorig et al., 1993). As further evidence also showed that MCP could act as the cellular receptor for measles virus (Dorig et al., 1994; Gerlier et al., 1994), the idea became generally accepted. The binding site for MCP was mapped on the measles virus H protein, and amino acid 481 was shown to be particularly important for this binding (Bartz et al., 1996; Lecouturier et al., 1996).

Questions regarding this view began to arise in 1996 (Buckland et al., 1997), however, when an accumulation of data indicated that MCP could only act as a receptor for certain laboratory-adapted strains of measles virus (typically the vaccine strains Hallé and Edmonston), and that so called wild-type measles strains appeared to be using a different protein as their receptor (Bartz et al., 1998; Lecouturier et al., 1999).

There is considerable and wide ranging evidence that only certain strains of measles virus use MCP as a receptor. Human red blood cells don't express MCP although African green monkey red blood cells (AGM_RBC) express a simian homologue of MCP. Vaccine strains of measles virus have the ability to haemagglutinate AGM_RBC, whereas other isolates could not, implying that some measles isolates cannot use MCP as a receptor (Saito et al., 1992). Certain MCP^+ transgenic rodents and recombinant cell lines are not permissive for measles infection/attachment (Blixenkrone-Moller et al., 1998; Horvat et al., 1996; Niewiesk et al., 1997). Whilst this could indicate the need for a cofactor, it could also suggest that MCP may not be the 'natural receptor' for measles. One of the consequences of a measles infection with a vaccine strain is the down-regulation of MCP expression, and the amino acids on the H protein responsible for this phenomenon were subsequently identified (Bartz et al., 1996). This characteristic of infection is found only for the vaccine strains and certain laboratory-adapted strains of measles (Lecouturier et al., 1996). This phenomenon is also mirrored in the case of SLAM, since it has been demonstrated that MV-H also down regulates SLAM (Tanaka et al., 2002). The latter case, however, appears to hold true even for measles strains that are normally

associated with MCP receptor usage. Chick embryo fibroblast cultures are thought not to have a homologue of MCP, although certain measles isolates have been successfully passaged through such cultures. It is interesting to note that these data seem to be consistent with one group of isolates using MCP as receptor while another group of isolates use another receptor. This other receptor was identified as CDw150 (SLAM) during the course of this study (Tatsuo et al., 2000), and this result was subsequently supported by similar findings elsewhere (Erlenhoefer et al., 2001; Hsu et al., 2001; Ono et al., 2001a; Yanagi, 2001). In each of these cases the strain of measles used had not been passaged through Vero cells, and was said to be wild-type. The use of SLAM as a receptor for measles virus was further validated when the measles binding site was mapped to the N-terminal Ig domain (Ono et al., 2001b). There are, however, discrepancies also in the data that suggest SLAM is the 'natural' measles virus receptor. Whilst all measles strains tested so far have the ability to use SLAM as a receptor, it appears not to possess the expression profile that one would expect from observing clinical measles infections. SLAM is expressed on activated B and T cells, immature thymocytes and memory cells, but is not found on unstimulated monocytes (Minagawa et al., 2001), and is either not present on immature dendritic cells, or is expressed at very low levels (Ohgimoto et al., 2001). It therefore remains an open question as to how measles viral infections can manifest themselves the way they do, given this restricted expression profile. It has recently been demonstrated that SLAM expression can be induced on monocytes during an inflammatory response (Minagawa et al., 2001), and it may subsequently emerge that this induction is indeed associated with the infection of monocytes with measles virus. Given that these data appear to be counter-intuitive, alternative

hypotheses may be needed, such as invoking the possibility of measles virus using co-receptors.

The problems of receptor identification seem to have arisen from the way in which the virus is isolated. The propagation of measles virus is typically achieved by passaging through monkey kidney cell lines (Vero). This procedure requires several rounds of so called blind passaging when no virus can be propagated. It is only after several weeks that cytopathic effects become observable, and virus can be harvested. It was later found that the isolation procedure could be considerably (10,000-fold) more sensitive if a marmoset lymphoblastic cell line (B95-8, or the adherent relative B95-a) was used in place of Vero cells (Kobune et al., 1990). Furthermore it was discovered that: (i) there were minor antigenic differences between the two types of measles isolate (Kobune et al., 1990); (ii) the B95-8 isolated virus produced disease symptoms in experimentally infected moneys which more closely resembled those of the human form of the disease; and (iii) that when the B95-8 isolated virus was adapted to Vero cells it behaved like the traditional Vero passaged virus. These data imply that some fundamental difference exists between the two types of measles virus, and that the B95-8 passaged measles isolates may more closely resemble the wild type virus than do Vero passaged isolates.

These differences in isolation also seem to translate into the phenotypic differences noted above. Vero-passaged isolates appear to be able to infect MCP^+ recombinant cells, have this infection/binding blocked by anti-MCP antibodies, down-regulate MCP expression, and haemagglutinate AGM_RBC. Whereas, virus passaged on B

cell-derived cell lines on the other hand, appear to display none of those traits (Schneider-Schaulies et al., 1995a). The evidence becomes more compelling when one considers reports that sequence analysis of the H proteins from certain Vero cell-adapted and 'Wild Type' virus differ from each other by two amino acids. Moreover it was found that the phenotype could be switched between 'wild-type' and 'Vero' by mutating E451 and N481 of the H protein in a wild-type Ma93F strain into the valine and tyrosine residues respectively found in the vaccine like Hallé strain (Lecouturier et al., 1996). A later study showed that the Y451 of Edmonston H protein was required for fusion to COS cells, but was unnecessary for binding to the marmoset B95-8 cell line (Xie et al., 1999). MCP down-regulation remained unaffected in this case. There have also been reports that other H-protein amino acids are involved both in MCP down regulation and receptor usage. If G546 is changed to a serine then haemadsorption and MCP receptor usage are conferred upon the virus (Woelk et al., 2001). Similar results have also been obtained elsewhere (Li et al., 2002).

Statistical analysis of the sequences of the H protein from a large number of measles isolates using the maximum likelihood method has generated some interesting findings (Woelk et al., 2001). It was shown that a number of differences in the H protein sequences from different isolates were non-random. Of particular interest was the finding that there was a positive selection for a tyrosine at position 481 in non B-cell passaged virus, and this is apparent in isolates passaged through several different cell lines (Woelk et al., 2001). The authors suggest that this is the result of a natural selection for the MCP receptor, rather than an artefact of Vero passaging. The same study also showed that there was a strong selection for a serine residue at

546 only in virus passaged through Vero cell lines, and as already indicated this change can confer MCP receptor usage. These data show that there are several non-random amino acid changes that are dependant upon the cell line used for propagation, and more work needs to be carried out to determine the most realistic strain of measles. The physiological relevance of MCP as the measles receptor was indeed recently demonstrated (Manchester et al., 2000). There are however precedents for this type of host range switching, for example certain picornaviruses, which have a positive (+) sense RNA genome, appear to be able to switch host range. This occurs not only when passaging the virus through cell lines from different species, but also after passaging through different cell lines from the natural host.

It therefore seems possible, that the ability of different isolates of measles to utilise different receptors could just be an artefact of virus propagation in unnatural cells. On the other hand, it also seems feasible that measles virus could use both receptors at different times during an infection, since the available evidence supporting each receptor acting alone is less than compelling.

1.1.8. The immunosuppressive effect during a measles virus infection

One of the most important consequences of a measles infection is the profound suppressive effect it has on the immune system. This was first noticed by von Pirquet, when he noticed the disappearance of the delayed-type hypersensitivity skin test to tuberculin, and an exacerbation of tuberculosis in measles patients (von Pirquet, 1908). The suppression of the immune system allows the opportunistic

infections, which cause most of the mortality associated with measles, to take hold (Beckford et al., 1985). This immune suppression occurs through a number of different routes. Measles virus causes a decrease in the effectiveness with which antigens are recalled, a suppression of lymphoproliferative responses (Ward et al., 1991), and abnormal lymphokine production (Atabani et al., 2001). It is interesting to note that these immunosuppressive effects occur at the time the immune response to the virus is at its strongest.

Monocytes, which are infected by measles virus, are responsible for the production of accessory factors (e.g. IL-1 and TNF-α). The level of IL-1 produced by infected monocytes is usually slightly raised, whereas the levels of TNF-α produced are reduced. TNF-α is a proinflamatory cytokine that is known to play an important role in the inflammatory response, and the lymphoproliferative effect. This decrease in TNF-α production caused by monocyte infection may be partially responsible for the decrease in lymphoproliferation that is associated with measles virus infections. It is thought that this reduction in TNF-α may reduce the efficacy of the immune response to new infections.

1.2. Membrane Cofactor Protein

1.2.1. Preliminary remarks

Membrane cofactor protein is a member of the regulators of complement activation (RCA) gene family and is expressed on the surface of all human cells examined so

far, with the exception of red blood cells (Johnstone et al., 1993a). It inhibits the complement system by acting as a cofactor for factor I-mediated cleavage of C3b and C4b, and is also involved in some cell signalling pathways. Since MCP is a complement inhibitor that protects host cells from complement-mediated lysis, it has recently become of interest in the field of xenotransplantation (Begum et al., 2000; Diamond et al., 2001; Loveland et al., 1993; Shinkel et al., 1998; Thorley et al., 1997). In addition to its natural role as a complement regulator, MCP has been found to be the cellular receptor for several important human pathogens in addition to measles virus.

1.2.2. The domain structure of MCP

Membrane cofactor protein is a single-chain membrane-bound glycoprotein. It contains an extracellular region, a single pass membrane-spanning segment, and a short cytoplasmic region (although a soluble form is also present in low concentrations in blood plasma, tears and semen) (Hara et al., 1992). The extracellular portion of the protein consists of four complement control protein (CCP) modules (also known as short consensus repeats, or sushi domains), a region rich in serine, threonine and proline residues (the STP region), and a region of unknown function. Following the extracellular region is a single, hydrophobic, transmembrane helix, and a short cytoplasmic sequence that is rich in arginine, lysine, and threonine residues (Liszewski et al., 1991). A schematic representation of MCP is shown in Figure 1.5

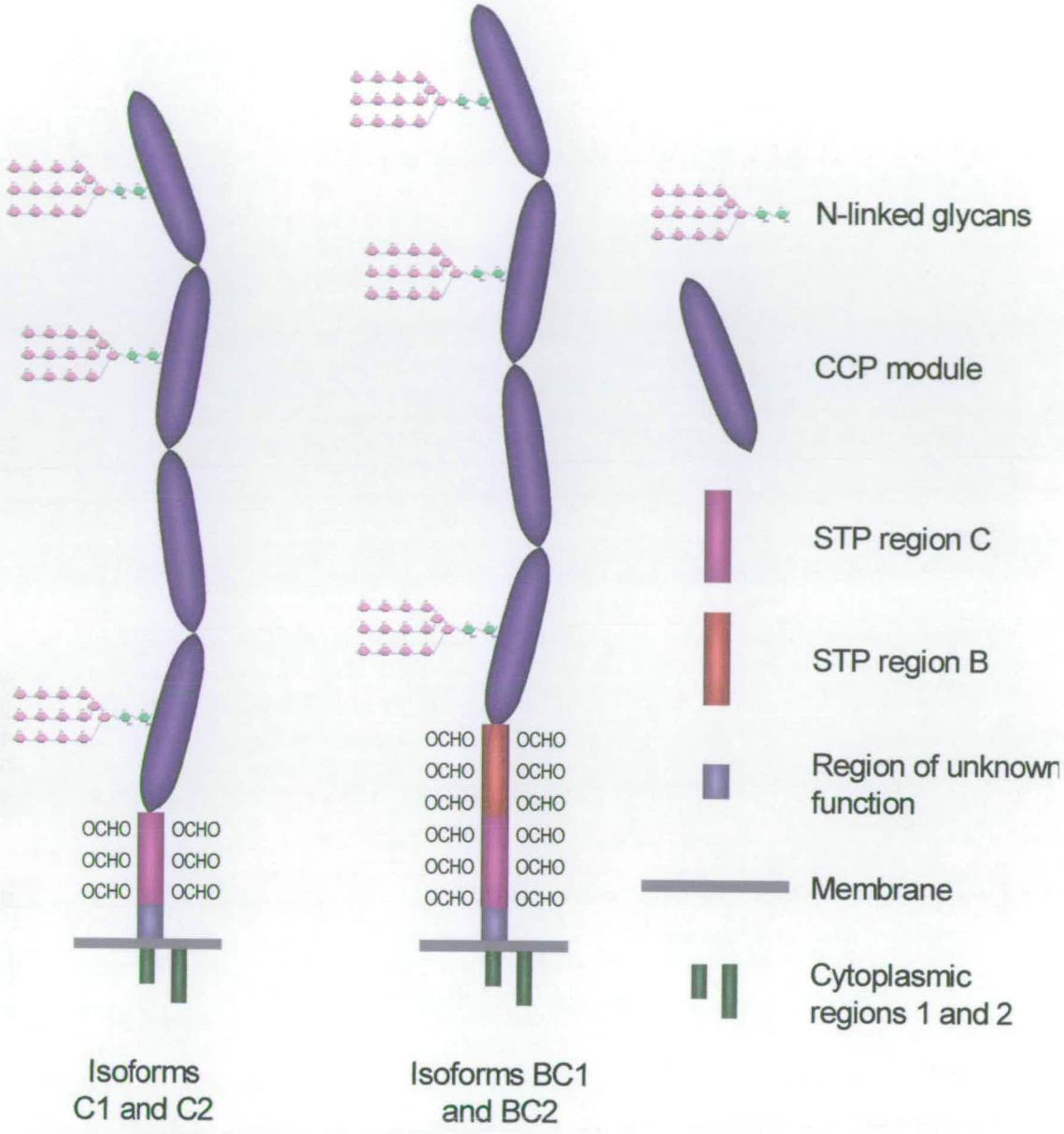


Figure 1.5 Schematic representation the structure of MCP, showing the four common isoforms (adapted from Liszewski et al., 1996).

1.2.3. Structure of MCP12

The publication of the crystal structure of MCP12 shows a number of interesting features (Casasnovas et al., 1999). A hexamer was found in the unit cell and so six

independent structures were solved, and a representative monomer is shown in Figure 1.6. The structures of these six monomers differed in their intermodular angles by ~15 °, implying that the modules may be joined by a flexible linker. The presence of a Ca^{2+} ion at the intermodular interface and crystal packing forces may further act to reduce the flexibility of this junction. Furthermore, the hexameric structure and the presence of a Ca^{2+} ion at the interface between modules, are not supported by physiological evidence, implying that these features may be crystal artefacts. Given the potential for physiologicaly relevent flexibility in the intermodular junction, elucidation of the solution structure and dynamis of this pair of modules should yield further information regarding its fuction.

The crystal structure provides interesting information concerning the structural basis for the functionally important glycans on module 2. The glycans of module 2 are located on the opposite side of MCP12 to the viral binding site. The regions involved in measles virus binding, therefore appear to form an extended glycan free binding site. Moreover, hydrophobic contacts were observed between the glycans and residues Y87 I104 of the polypeptide chain in each of the six independent copies of the protein. Together these data provide further evidence that measles virus does not merely bind to the glycans, and that instead interactions between the glycans and the polypeptide chain are stabilising the conformation of the virus binding surface.

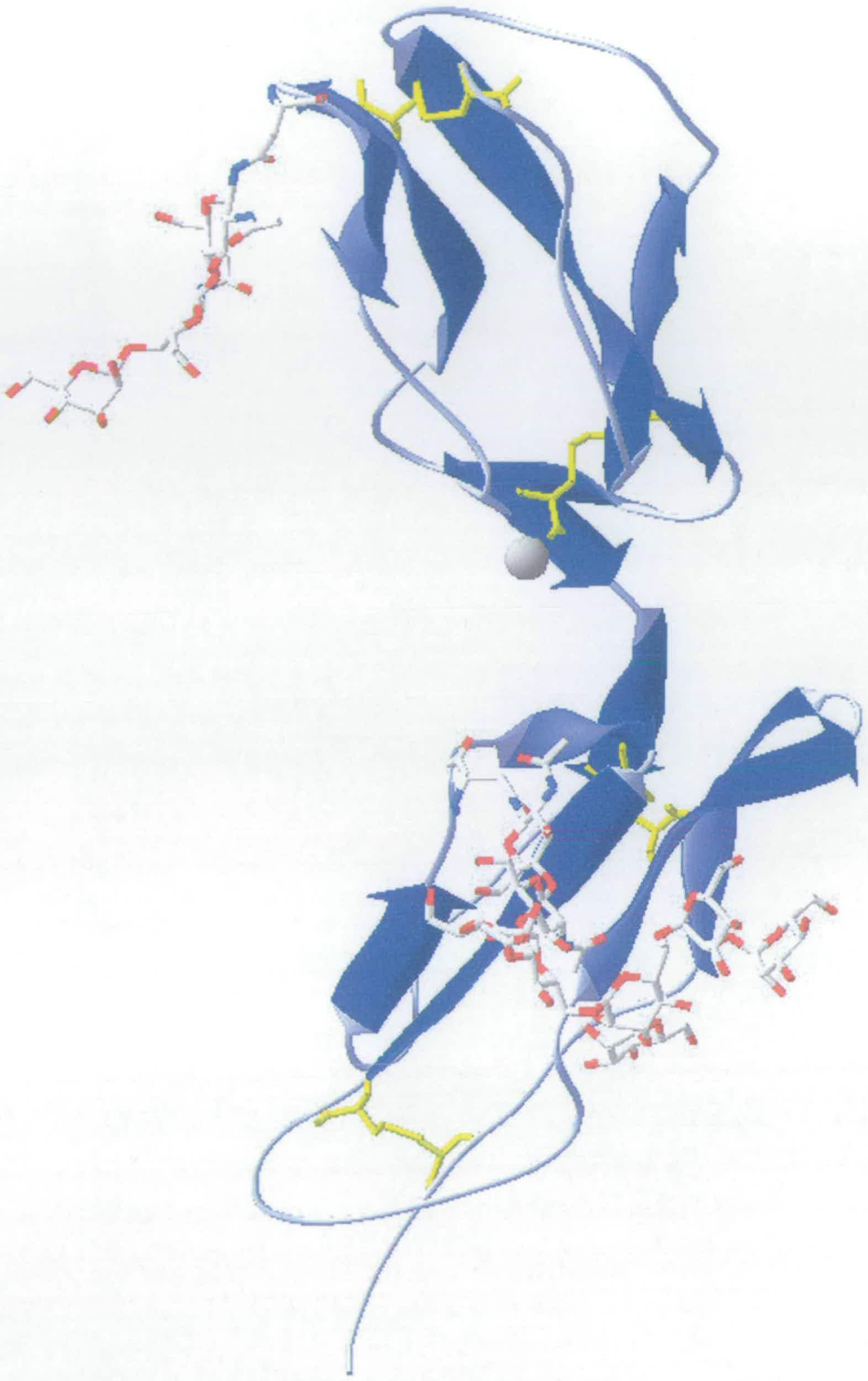


Figure 1.6 Crystal structure of a representative MCP12 monomer, with the N-terminus uppermost. The figure shows the N-linked glycans and Ca^{2+} ion. Cysteines involved in disulphide bonds are shown in yellow.

1.2.4. Post-transcriptional modifications of MCP

Alternative splicing of MCP mRNA generates four common isoforms, although some rarer isoforms also exist. There are two common STP regions (B and C) and two distinct common cytoplasmic regions (1 and 2); the four CCP modules are present in all of the isoforms identified to date. The C version of the STP region is present in all of the major isoforms, and so the presence or absence of the B form of the STP region is responsible for some of the variation; the cytoplasmic sequence is either 1 or 2 in the common isoforms. The prominent isoforms are therefore C1, C2, BC1 and BC2. The exact reasons for the occurrence of different isoforms are not yet fully understood. Most cells appear to express broadly similar levels of each isoform while notable exceptions include sperm, brain, kidney, and the salivary gland (Johnstone et al., 1993b).

1.2.5. Post-translational modifications of MCP

There are three N-glycosylation sites in MCP, one in each of CCP modules 1, 2 and 4. In addition to its N-glycosylation, MCP is also heavily O-glycosylated in the STP region. The O-glycosylation may help to protect the protein from proteolysis and there are precedents for this in other membrane proteins (Reddy et al., 1989). Both the O-glycosylation, and the N-glycan, on CCP modules 2 and 4 are required for MCP's complement regulatory activity (Liszewski et al., 1998) . Of particular interest though is the glycan on CCP module 2, since it is necessary for both the natural complement regulation activity, and for function as the measles virus receptor (Maisner et al., 1996; Maisner et al., 1994), implying that it could be performing

some structural role. Interestingly it was demonstrated that the exact nature of the N-glycans on module 2 of MCP are unimportant for measles receptor function (Maisner et al., 1995). There is also growing suspicion that the cytoplasmic region can be phosphorylated, and it has been implicated in certain cell signalling pathways (Wang et al., 2000; Wong et al., 1997).

1.2.6. The CCP module

The RCA protein structure is based upon the complement control protein structural motif, which, as its name implies, is present in many of the complement control proteins. The RCAs contain at least three CCP modules connected by short linker sequences (Barlow et al., 1991). CCP modules usually occur in elongated head-to-tail arrays with a functional site that typically involves at least two modules and their associated junctions. It is for this reason that the intermodular orientation is considered to be of particular importance. The genes encoding the complement control proteins are very closely related. CCP modules, as are identified by the presence of a consensus sequence containing a pattern of cysteine residues, and a tryptophan residue, spanning some 60 residues. This consensus sequence occurs numerous times in a wide range of proteins. Although the CCP module was first identified as a structural motif in complement control proteins, it has been found in an increasing number of apparently unrelated proteins, from clotting proteins to neurotransmitter receptors. CCP modules are also found in a wide variety of different organisms, from *C. elegans* and various viruses through to man.

1.2.7. The CCP module structure

A CCP module consists of approximately sixty amino acids including four cysteines which form disulphide bonds in a 1-3, 2-4 pattern as illustrated by the first structure of the 16th CCP module of complement factor H (Norman et al., 1991). Give the structural significance of the disulphide bonds it is perhaps unsurprising that the disulphide bonds were shown to be essential for MCP to serve as a measles virus receptor (Maisner et al., 1994). A sequence alignment of several CCP modules along with the consensus sequence is shown in Figure 1.7. The majority of the structure is composed of small anti-parallel β-sheets, which form a tightly packed barrel-like structure around a compact hydrophobic core that contains a consensus tryptophan side chain (very rarely replaced by a phenylalanine). The modules generally occur in a head-to-tail or 'beads on a string' arrangement; and both well-defined (Wiles et al., 1997) and flexible (Smith et al., 2002) intermodular junctions have been described. Whilst sequence and length variations between CCP modules are considerable, modelling by homology the structures of individual CCP modules is at least feasible. Reliable methods for predicting the relative arrangement of neighbouring modules are, however, as yet a long way from being available. Until such methods are developed, the intermodular arrangement will necessarily require elucidation by empirical methods. It is suspected that in certain circumstances a flexible linker may be critical to biological function. Given the likelihood that crystal packing forces could artefactually rigidify the junction, the use of high field NMR spectroscopy to provide solution structures of CCP module is particularly appropriate.

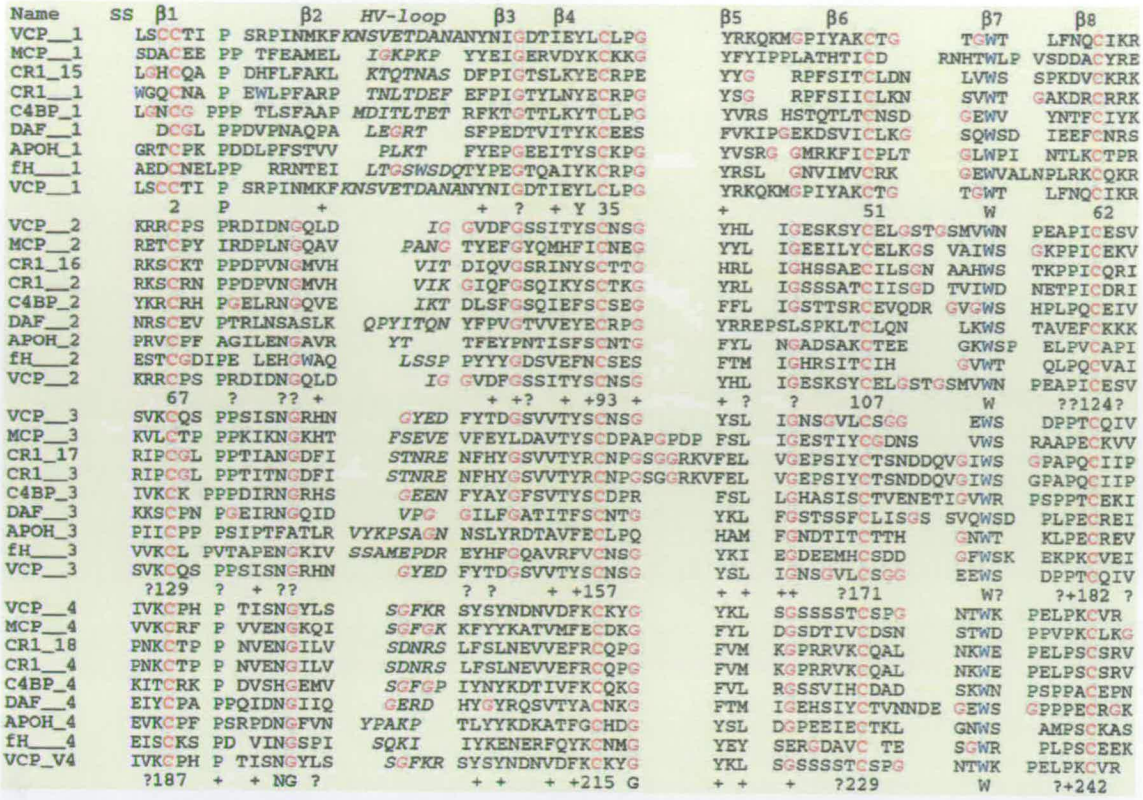


Figure 1.7 Alignment of several CCP modules, showing the consensus sequence. Sequence numbers refer to the sequence of DAF; Prolines (green), cysteines (red), glycines (purple) and tryptophan (blue) are colour-coded; Beneath each block of sequence alignments: the use of the one-letter code indicates a specific residue that is conserved across all sequences shown; a + sign indicates a residue that is conserved or conservatively replaced in all the sequences shown; whereas a ? indicates a residue-type that is not completely conserved (taken from Kirkitadze et al., 2001).

1.2.8. RCA proteins can act as receptors for pathogens

Like MCP, many of the other RCA proteins serve as the receptors for important human pathogens, and every RCA protein has been implicated in interactions with at least one important pathogen. Erythrocyte-borne complement receptor type 1 (CR1, CD35) is important in the rosetting process associated with severe cases of malaria, and has been shown to bind to the *P. falciparum* VAR-1 adhesin proteins on the surfaces of infected erythrocytes (Krych-Goldberg et al., 2002). CCP modules 1-2 of

complement receptor 2 (CR2, CD21) bind the Epstein-Barr virus (EBV) gp350/220 protein conferring on CR2 the ability to act as the receptor for EBV (Molina et al., 1991). Decay accelerating factor (DAF, CD55) acts as the receptor for many of the picornaviruses, for example echoviruses and coxsackieviruses. In addition DAF has been shown to bind the Dr antigen of certain strains of *E. coli* (Lindahl et al., 2000). Complement factor H has been shown to bind to *Streptococcus pyogenes* (through binding the M protein), *Borellia burgdorferi* (through the surface protein OspE) (Hellwage et al., 2001), *Onchocerca volvulus* microfilariae (Meri et al., 2002) and *Neisseria gonorrhoeae* (Ram et al., 1998). The discovery that factor H retains its complement regulatory activity when bound to the *S. pyogenes* M protein may illustrate one reason why the complement control proteins are such a 'popular choice' for a pathogenic receptor. C4b binding protein (C4bp) is also recognised by several important human pathogens including *S. Pyogenes* (Blom et al., 2000), *Bordetella pertussis* (Berggard et al., 2001) and *N. gonorrhoeae* (Ram et al., 2001).

MCP is, therefore, not unique in its ability to act as a receptor for pathogens. Pathogens that use MCP as a receptor include: measles virus (MV); human herpes virus-6 (HHV-6) (Santoro et al., 1999); pathogenic *Neisseria* (Kallstrom et al., 1997) and *S. Pyogenes* (Giannakis et al., 2002; Okada et al., 1995). The attachment sites for *N. gonorrhoeae* have recently been identified as CCP module-3 and the STP rich region (Kallstrom et al., 2001). HHV-6 is implicated in the progression of HIV to AIDS. Like the measles virus, HHV-6 has an immunosuppressive effect. This could be connected in some way to their both utilising MCP as their receptor.

1.2.9. The importance of MCP as a receptor for measles virus

As has already been discussed, MCP can serve as the receptor for at least some strains of measles. A considerable body of data exists for the structural requirements of the interaction between MCP and the measles H protein. The H protein's binding site on MCP was localised to the first two CCP modules (Manchester et al., 1997) using a combination of deletions and chimeric proteins between MCP and the related protein, decay accelerating factor (DAF) that cannot serve as a measles receptor. The binding site was shown to be distinct from the C3b/C4b/factor I interaction sites, which have been located to CCP modules 2/3/4 (Adams et al., 1991; Iwata et al., 1995; Liszewski et al., 2000). The binding of haemagglutinin to MCP appears to require the protein to have structural integrity, since the interaction requires the disulphide bonds to be intact and N-glycan on module two to be present (Maisner et al., 1996; Maisner et al., 1994). It is interesting to note that the exact nature of the glycan on module two seem to be relatively unimportant (Maisner et al., 1995), implying that the glycan serves some structural role rather than acting as an additional attachment site for haemagglutinin. This idea was further supported with the publication of the crystal structure of the first two CCP modules of MCP (Casasnovas et al., 1999). There have been several studies to identify the MCP residues that are required for haemagglutinin binding. These include peptide inhibition studies, site directed mutagenesis changing selected residues to serine (Hsu et al., 1999), and alanine scanning mutagenesis. These results are summarised in Figure 1.8. The residues identified from the above studies were shown to be surface-exposed in the crystal structure, and appear to form an extended binding site that is free from carbohydrates.

DACEEPPTFEAMELIGKPKPYYEIGERVDYKCKKGYFYIPPLATHTICDRNHTWLPVSDDACYRET

CPYIRDPLNGQAVPANGTYEFGYQMHFICNEGYYLIGEEILYCELKGSVAIWSGKPPICEK

Figure 1.8 Summary of the residues of MCP12 that have been implicated in measles virus binding. The consensus cysteines and tryptophans are shown in yellow, and bold respectively. The glycosylated asparagine residues are shown in bold italics. Residues implicated as important on the basis of mutagenesis to alanine and serine are shown as underlined, or red respectively. Residues implicated by peptide inhibition are shown in blue

1.2.10. The function of MCP in cell signalling

The biological role of the intracellular domain of MCP has remained largely a mystery. It is not required for the natural complement function of MCP, as a version of MCP with a GPI anchor rather than the normal transmembrane and cytoplasmic regions still serves to protect cells from complement attack (Shinkel et al., 1998).

Purified pili from pathogenic *Neisseria* cause a calcium flux that can be blocked by anti-MCP antibodies (Kallstrom et al., 1998); partial deletion of the cytoplasmic domain of MCP reduced the effectiveness of a *N. gonorrhoeae* infection (Kallstrom et al., 2001); binding to haemagglutinin or anti-MCP antibodies causes down regulation of IL-12 (Karp et al., 1996); and GST fusions with the intracellular regions of MCP associate with kinases in macrophages (Wong et al., 1997). All of these results suggest a possible role for the intracellular region of MCP in cell signalling pathways. Further evidence of a cell signalling role for MCP came when it was discovered that a tyrosine in the cyt-2 cytoplasmic domain of MCP is phosphorylated by the Lck kinase from the src kinase system (Wang et al., 2000).

Precisely why MCP should posses a cell signalling role still remains moot, although it is not impossible that it may be to provide a link between the cellular immune system and the complement system

1.2.11. The possible importance for MCP in xenotransplantation

It is generally accepted that the scarcity of donor organs is one of the chief limiting factors governing the number of transplant operations that can be performed. Moreover, when human organs are used for transplantation, there exists a small but significant risk that the implanted organs will carry an infection. For these reasons researchers have started to examine the possibility of using alternatives to donated human organs. To date two options are generally considered to have reasonable potential. One option, based on organ engineering and embryonic stem cell research, would involve growing new organs in the laboratory using cloned human embryonic stem cells with the potential recipient acting as the donor of genetic material for their own organs. This approach, the favoured option from a therapeutic standpoint, still requires numerous serious technical problems to be overcome, although encouraging results have been obtained with skin and trachea. In addition to the technical difficulties are the associated moral and ethical problems that still plague any work related to the field of human cloning. The other technique generally referred to as xenotransplantation, involves the use of non-human organs for transplantation. In this case the donor animals of choice appear to be pigs since their organs are of similar size to human organs. In addition there is less public opposition to the experimental use of pigs than there is to the use of primates. Pigs are also assumed

to suffer from fewer diseases that could potentially be transmitted to humans. There are though major problems that stand in the way of xenotransplantation: of uppermost concern is that of hyperaccute rejection (HAR). HAR is a process that involves antibody initiated complement activation (Kadner et al., 2001), and results in rapid degradation of the target. It is thought that using pigs transgenic for human complement regulators, MCP and DAF in particular, could be a potentially important tool in the prevention of HAR, and this approach has indeed delivered some success for *ex vivo* perfusion experiments using transgenic mice (Mulder et al., 1995; Mulder et al., 1996), and in xenotransplantation experiments in baboons (Kadner et al., 2001). In the case of MCP, however, modification of the protein has been undertaken, aiming to produce MCP that retains the ability to regulate complement, whilst losing the ability to act as a receptor for measles virus (Begum et al., 2000). Other approaches that have been proposed, include pigs that are transgenic for several complement regulators simultaneously e.g. DAF and MCP since these two complement regulators appear to act synergistically (Brodbeck et al., 2000). One method that has provided some *ex vivo* success, is the direct addition of a soluble MCP-DAF chimera (named complement activation blocker-2 CAB-2), to human blood that was perfusing a porcine heart (Kroshus et al., 2000). The inclusion of CAB-2 prolonged the survival of the heart. It still remains unclear, however, as to how this approach could be migrated to a human therapy.

1.2.12. The importance of MCP in reproduction

Pregnancy and reproduction in general present the human immune system with an intriguing problem. The foetus is just another piece of foreign tissue as far as the immune system is concerned. It is clear therefore that there must be ways to protect the foetus whilst still retaining the ability to fight infection. It appears that this is achieved, at least in part, through the action of the complement regulatory proteins MCP and DAF. It was shown that mice that were deficient in the murine complement regulator Crry (an analogous protein to MCP and DAF) were unviable, and the foetuses abort (Xu et al., 2000). This was demonstrated to be due to their inability to prevent the activation of complement at the feto-maternal interface.

In addition to protecting the foetus MCP appear to play other roles in reproduction. A soluble form of MCP, lacking the transmembrane domain and cytoplasmic tail, was found in semen at 10-fold greater concentrations than those found in plasma (Seya et al., 1993), and this has been linked to the prevention of sperm rejection. MCP may also play another, seemingly unrelated, role in reproduction since it has been associated with the sperm egg interaction, and certain types of sterility have been associated with abnormal seminal MCP expression (Kitamura et al., 1997).

It therefore remains an important goal of current research to understand all the complement regulators as fully as possible. It is hoped that a fuller knowledge of these proteins could potentially provide highly significant therapeutic benefits in such diverse fields as transplantation, autoimmune diseases, and in the treatment of a diverse number of important human pathogens.

1.3. The complement system

1.3.1. Preliminary remarks

As already indicated MCP is a complement control protein whose natural biological function is to help prevent the complement system from attacking host cells. The complement system is one of the major components of the human immune system, and serves to clear foreign matter from the host. It was originally identified as a heat labile component of blood serum that could 'complement' the ability of antibodies to cause bacterial lysis. It is now known to consist of over thirty proteins, both soluble and membrane bound (Law et al., 1995). Activation of the complement system can occur by one of three mechanisms - the classical, lectin and alternative pathways) - all of which are of necessity tightly regulated. The essence of the complement system is a cascade of proteolytic activations of zymogens, which results in the various effector functions associated with the complement system. Despite, or perhaps because of, the central role of the complement system in the immune system, certain pathogens have evolved methods to hijack the complement system and use it to facilitate an infection. The complement system, when poorly regulated, has also been associated with a number of autoimmune diseases.

1.3.2. Nomenclature of complement proteins

The nomenclature used for the complement system is complex, and reflects by-and-large the chronological order of protein discovery rather than their order of participation in a complement cascade. The proteins of the classical pathway are generally termed 'components' and are given a C followed by a number (e.g. C2).

Many of these proteins are zymogens so require proteolytic activation, and therefore the active enzyme is illustrated by the addition of a bar above the name (e.g. $\overline{C1r}$). Following the proteolytic activation, the cleavage products are differentiated from each other and the parent protein by the addition of a lowercase letter (e.g. C2, C2a & C2b). Usually the small fragment is termed 'a' and the larger active fragment 'b', although an exception is 'C2', where 'C2a' is the larger active fragment and 'C2b' is the smaller one. Other complement proteins are generally termed 'factors', and are named with a capital letter (e.g. factor B), this can be abbreviated to FB or even dropped altogether leaving just B. The complement regulatory proteins are usually given an abbreviation (e.g. MCP), or are referred to by their cluster of differentiation (CD) number (e.g. CD46).

1.3.3. Activation of the complement system

As has already been stated, activation of the complement system occurs either via the classical pathway, the lectin pathway, or the alternative pathway. The classical pathway is activated by antibodies from the adaptive immune system, and results in complement activation which is part of the innate immune system. The classical pathway therefore acts as a critical link between the adaptive and innate immune systems. Very closely related to the classical pathway is the lectin pathway, which is activated by the binding of certain lectins to foreign sugar groups (in particular the terminal mannose residues often found in yeast glycoproteins and bacteria). The lectin pathway differs from the classical pathway only in the way that complement component C4 is activated The final method of activation of the complement system

is known as the alternative pathway that provides a 'tick-over' activation, which can rapidly amplify when needed. A schematic representation of the whole complement system is given in Figure 1.9.

1.3.3.1. Complement activation via the classical pathway

Central to the classical pathway is proteolysis of C4 eventually resulting in the formation of the classical pathway C3 and C5 convertases – $\overline{C4b2a}$ and $\overline{C4b2a3b}$ respectively. The initial step in the classical activation of complement is the activation of C1. C1 is a calcium dependant, soluble serum protein complex consisting of a molecule of C1q, two C1r molecules and two C1s molecules. C1q has no intrinsic enzymatic activity, but is involved in interactions with immune complexes (Reid, 1983), whereas C1r and C1s are zymogens that provide the catalytic activity. C1q has an unusual "bunch of tulips-like" structure, consisting of six copies of a subunit that has a collagen-like region with a globular head. C1q interacts with the F_c region of antibodies with several simultaneous interactions being required for stable binding. The result is that C1q binds strongly to the F_c domains of aggregated IgG molecules (as found in immune complexes), but not to the uncomplexed IgG present in normal serum. The binding of immune complexes to C1 causes a conformational change in C1q resulting in the auto-activation of C1r. Activated $\overline{C1r}$ then activates C1s and further C1r resulting in the active form of C1 ($\overline{C1q(r_2s_2)}$).Activated $\overline{C1q(r_2s_2)}$ then acts on the classical pathway complement component C4 to yield the two components C4a (a weak anaphylatoxin) and activated C4b that contains an exposed thioester. Activated $\overline{C1q(r_2s_2)}$ then acts on

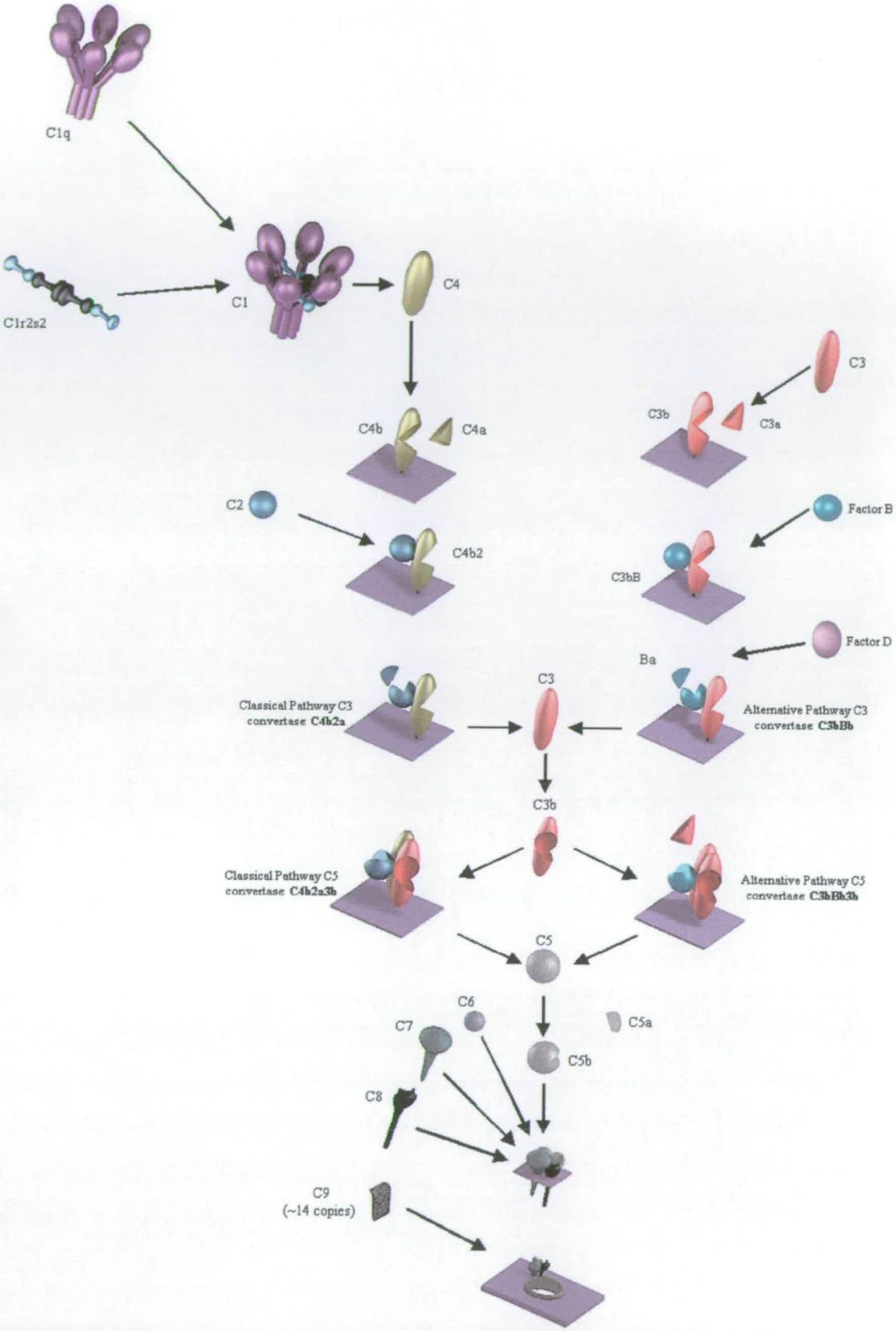


Figure 1.9 Schematic of the complement system. For an explanation, see the text.

the classical pathway complement component C4 to yield the two components C4a (a weak anaphylatoxin) and activated C4b that contains an exposed thioester. Activated C4b is highly unstable and is rapidly attacked by any surrounding nucleophilic groups (mostly from water molecules) to yield inactive C4b (iC4b). A small fraction of the activated C4b molecules bind covalently to hydroxyl groups on proximal cell surfaces, resulting in the presence of surface-bound C4b. The surface-bound C4b generated in this manner can then bind to complement component C2, yielding C4b2. The C2 component of the newly formed C4b2 is now able to act as a substrate for activated C1s, yielding C2b and the classical pathway C3 convertase $\overline{\text{C4b2a}}$. The classical pathway C3 convertase acts on C3 to form C3b and the anaphylatoxin C3a. The creation of C3b exposes a thioester, and like C4b reacts rapidly with free hydroxyl groups on the cell surface, in a process known as opsonisation. In addition C3b can bind to the classical C3 convertase ($\overline{\text{C4b2a}}$) to form the classical pathway C5 convertase (C4b3b2a). This C5 convertase cleaves C5 to generate anaphylotoxic C5a, and C5b. The newly exposed thioester in C5b promotes binding to the surface of the cell under attack by antibody, and the surface-bound C5b then initiates the formation of the membrane attack complex (MAC), which ultimately results in cell lysis.

1.3.3.2. Complement activation via the lectin pathway

As already stated the difference between activation of complement through the classical pathway and lectin pathway lies only in the mechanism of formation of C4b. Activation of the lectin pathway involves mannan-binding protein (MBP) - a

calcium-dependent lectin in the collectin family instead of the C1 complex. Activation occurs when MBP binds sugar residues on foreign oligosaccharides (typically the terminal mannose residues), allowing it to interact with a MBP associated serine protease (MASP) (Holmskov et al., 1994). There are two types of MASP; MASP1 and MASP2 which are homologous in structure to C1r and C1s. MASP activation occurs in a similar fashion to C1r and C1s. The activated MBP-MASP complex can then act on C4 in the same way as $\overline{C1q(r_2s_2)}$ to initiate the formation of the classical pathway C3 convertase $\overline{C4b2a}$.

1.3.3.3. Complement activation via the alternative pathway

Activation of complement via the alternative pathway occurs without any specific initiating event. Instead there is a continuous 'tick-over' mechanism of activation. In host tissue this is kept in check, but it is rapidly amplified in the absence of the necessary control mechanisms. There is an internal thioester bond in C3 that can be hydrolysed by water (Law, 1983). This yields a form of C3 (C3i) that can bind to factor B giving C3iB. Factor D is then able to cleave the factor B potion of C3iB resulting in C3iBb and Ba. The C3iBb produced in this way acts as a fluid phase alternative pathway C3 convertase that cleaves C3 into C3b (and the anaphylatoxin C3a) in a similar way to the classical pathway C3 convertase. The unstable C3b that is generated is quickly degraded unless it attaches to foreign material. When C3b is deposited on a foreign surface, it can also interact with factor B which is subsequently cleaved by factor D. When activated in this manner, the result is the active surface-bound alternative pathway C3 convertase - $\overline{C3bBb}$, and a positive

feedback loop is initiated. It is clear therefore that both the classical pathway and the alternative pathway can cause activation of the alternative pathway's positive feedback loop. The $\overline{C3bBb}$ generated by either pathway is then able to bind to further C3b to form the alternative pathway C5 convertase - $\overline{C3bBb3b}$.

1.3.4. Formation of the membrane attack complex

The final direct effect of the complement cascade is the formation of the membrane attack complex (MAC). This stage is initiated by the proteolytic cleavage of C5 by either the classical or the alternative C5 convertase. For C5 to act as a substrate for the C5 convertases it needs to be bound to C3b and to a surface. Following cleavage of C5 into C5a (an anaphylatoxin that dissociates) and C5b, the formation of the MAC occurs through the sequential binding of the complement components C6-C9 in a non-enzymatic process to the C5b on the foreign surface. After the addition of C7, the nascent MAC becomes hydrophobic and preferentially inserts into the lipid bilayer of the cell surface. Component C8 then binds followed by up to 14 monomers of C9, to form a pore that can lead to cell lysis. Formation of the MAC in the fluid phase is largely prevented by proteins such as vitronectin, which serve to inactivate C5b67 before it can attach to host surfaces. Interestingly a small amount of MAC formation is important for the activation of certain cells of the immune system. The membrane containing a portion of MAC is endocytosed, and can stimulate the cell to produce and metabolise arachidonic acid, and other factors which can serve to stimulate the inflammatory response.

1.3.5. Regulation of the complement system

It is clear that a system such as complement that has positive feedback loops and a "tick over" mechanism of activation requires rigorous regulation. Complement has been called a triggered amplification cascade because a small initial cause leads to an effect that has been greatly amplified. Improper or disproportionate action of complement results in inflammation and, potentially, severe tissue damage. Indeed, inappropriate complement activation is responsible for the debilitating symptoms associated with numerous diseases, although minute amounts of complement activation on host cells actually appears to be beneficial. The extreme destructive power of complement requires that at every stage there are checks and balances to prevent its activation on host cells. This is particularly true for the stages that lead to the generation of the critical C3 and C5 convertases that amplify the system so effectively.

1.3.5.1. Regulation of classical pathway activation

Regulation of the classical pathway occurs primarily at two distinct points in the cascade: the activation of C4 by C1, and the formation and persistence of the C3/C5 convertases.

The first step of regulation in the classical pathway occurs by controlling the activation of C1. Activation of C1 is controlled by the C1 inhibitor (C1-Inh); a serine protease inhibitor (SERPIN) that binds to and inactivates activated $\overline{C1r}$ and $\overline{C1s}$. The (C1-Inh)-C1r-C1s-(C1-Inh) formed in this way dissociates from C1q, thus

preventing further activation of classical pathway. The free C1q thus generated is then able to interact with the C1q receptor and perform several effector functions including inhibition of IL-1 production.

The second and possibly more important points of regulation are at the formation of the classical pathway C3 convertase ($\overline{C4b2a}$) and C5 convertase ($\overline{C4b2a3b}$). These stages are efficiently regulated through two important and distinct mechanisms: by increasing the rate of dissociation of C2a from C4b in $\overline{C4b2a}$; and by promoting the factor I mediated degradation of C4b. Dissociation of the separate components $\overline{C4b2a}$ takes place spontaneously relatively rapidly (~1-5 min) even in the absence of other proteins and this acts as a natural break. Given the activity of $\overline{C4b2a}$, however, efficient regulation requires this rate of dissociation to be further increased, should the cascade need to be halted altogether, as in the case with host cells. The process of increasing the rate of dissociation of C2a from C4b in $\overline{C4b2a}$ is known as decay acceleration. Decay acceleration activity is mediated by the interaction of $\overline{C4b2a}$ with some of the host complement control proteins that can promote the dissociation. The decay accelerating proteins on the cell surface are decay acceleration factor (DAF) and complement receptor type 1 (CR1), and in the fluid phase this role is carried out by C4 binding protein (C4-bp).

Regulation of the C3/C5 convertases through the degradation of C4b is achieved through the action of factor I, and is also mediated through the action of some of the host complement control proteins. Factor I is a highly specific serine protease that

cleaves C4b at two points in its alpha chain, yielding the inactive fragments C4c and C4d whose further functions (if any) are unknown. The activity of factor I alone is relatively low, however it becomes greatly increased when the relevant complement control protein act as cofactors. This process is achieved on the cell surface by MCP and CR1, and in the fluid phase by C4-bp, acting as co-factors for factor I. The actions of C4-bp CR1 and MCP therefore serve to increase the rate of factor I mediated degradation of C4b, and so inhibit the action of the classical pathway C3 convertase.

To summarise the mechanism of classical pathway C3 convertase regulation: on a self surface MCP and CR1 can act as cofactors for factor I, whilst DAF and CR1 promote the dissociation of C2a from C4b in $\overline{C4b2a}$. These two mechanisms work in concert and are complemented by the equivalent action of C4bp in the fluid phase. On a non-self surface however, the absence of these regulators fails to inactivate complement, and lysis of the organism ensues. It can be seen that both C4bp and CR1 possess both decay acceleration and cofactor activities. A summary of the action of all the complement control proteins of both the classical and alternative pathways is shown in Table 1.1

Table 1.1 Summary of the mechanism of action of the complement control proteins (adapted from Weisman et al. 1990)

Protein	Fluid phase or membrane bound	Decay Acceleration		Factor I cofactor Activity	
		Classical pathway	Alternative pathway	Classical pathway cleavage of C4b	Alternative pathway cleavage of C3b
C4bp	Fluid	Yes	No	Yes	No
Factor H	Fluid	No	Yes	No	Yes
DAF	Membrane	Yes	Yes	No	No
MCP	Membrane	No	No	Yes	Yes
CR1	Membrane	Yes	Yes	Yes	Yes

1.3.5.2. Regulation of alternative pathway activation

The alternative pathway is regulated, like the classical pathway, through the action of the complement control proteins and their capacity to accelerate decay and act as co-factors.

Decay acceleration in the alternative pathway is achieved in the fluid phase through the action of the complement control protein factor H promoting dissociation of Bb from activated C3 (C3i and C3b). This is analogous to the decay acceleration activity of C4-bp in the classical pathway. DAF and CR1 also act on the cell surface in the alternative pathway to promote the dissociation of Bb from C3 in an analogous fashion to their activities in the classical pathway.

Factor I is also able to function as part of the alternative pathway to catabolise C3b into C3c and C3dg. The destruction of C3b is promoted in the fluid phase by the cofactor activity of factor H, and on the cell surface by the cofactor activities of MCP and CR1. Factor H is known to bind heparin and other polyanions such as the sialic acid groups often found on many mammalian cell surface glycoproteins. Factor H appears to be able to distinguish between self and non-self cells through its interaction with these polyanions, and its affinity for C3b increases by approximately 10-fold in the presence of polyanions (Meri et al., 1994; Pangburn, 2000; Pangburn, 2002). This increased affinity in the presence of polyanions has been shown to correlate with the protection of host cells, presumably through both cofactor and decay acceleration activities.

To summarise the mechanism of alternative pathway C3 convertase regulation: on a self surface MCP and CR1 can act as cofactors for factor I, whilst DAF and CR1 promote the dissociation of Bb from C3b in $\overline{C3bBb}$. On a non-self surface however, the absence of these regulators fails to inactivate complement, and lysis of the organism ensues. These two mechanisms also work in concert on the cell surface, and are complemented by the powerful selective activity of factor H. Factor H is able to distinguish self from non-self in many cases through the presence or absence of polyanions such as sialic acid on the cell surface. A summary of the action of all the complement control proteins of both the classical and alternative pathways is shown in Table 1.1

1.3.5.3. Regulation of the membrane attack complex

As with the other areas of the complement system, the MAC is also tightly regulated to prevent autologous cell lysis. One way in which MAC formation is regulated is through the action of the S-protein (vitronectin). It appears that when MAC starts to form in the fluid phase, vitronectin binds irreversibly to the C5b67 complex abrogating the membrane binding properties of the MAC. Another such MAC regulatory protein is CD59 (also known a protectin, HRF20, MACIF, MIRL, or MEM-43-Ag), which seems to function to prevent the polymerisation of C9 on the C5b-8 complex in the final stages of pore formation (Lachmann, 1991). The presence of these proteins on a self-surface prevents excessive MAC deposition on autologous tissue, protecting it from complement attack.

1.3.6. Effects of complement activation

The principal biological effects of the complement system are: the enhancement of killing of microbes; the processing and clearing of immune complexes; and the enhancement of the antibody response.

1.3.6.1. The action of complement to enhance the killing of microbes

Complement can enhance the killing of microbes in three ways, simultaneously. The first method is by the direct lysis of a cell by MAC formation. Second, activation of the complement system generates large amounts of the anaphylatoxins C3a and C5a, which act to enhance the inflammatory response. C5a acts as a potent chemotactic agent, activates neutrophils and monocytes, degranulates mast cells, and increases the expression of cell adhesion molecules. This results in: increases blood flow to the area, a respiratory burst (increases the production of oxygen free radicals), and increased production of IL-1, IL6 and histamine. As would be expected the half-life of C5a is very short since carboxypeptidase N rapidly removes its terminal arginine, resulting in C5a-des-Arg, which is considerably less potent. C3a acts in a similar fashion to C5a, but is less potent and has little or no chemotactic activity.

The third method by which the complement system acts to enhance microbial killing is through opsonisation. This process is the result of the coating of the cell surface with the opsonins C3b and C4b. These act as ligands for cell surface receptors on phagocytes, and stimulate phagocytosis. In addition the binding of the opsonised complex to macrophages activates theses cells and results in increased production of proteases, and oxygen free radicals. Finally the opsonised material is recognised by

erythrocyte-borne complement receptor type 1 (the immune adherence receptor) and cleared from the blood stream (see below).

1.3.6.2. The role of complement in clearing immune complexes

It was noted relatively early on that classical pathway complement activation inhibits the formation of precipitating immune complexes, and that in a similar manner alternative pathway complement activation can solubilise precipitated immune complexes. Precipitation of immune complexes is due to the formation of a 'lattice' effect between antibody and antigen, where the antibodies bind multiple different antigen molecules. Since it was found that C3-deficient serum was unable to perform these tasks it was inferred that it was the C3 component of complement that is important. The covalent binding of activated C3 (C3b or C3i) has important effects on immune complexes. Binding of C3 reduces the overall size of an immune complex, possibly by disrupting the binding of antibody to antigen. The presence of C3 allows the immune complex to bind to erythrocytes, by attaching to the complement receptor type 1 (CR1 or CD35), allowing the complex to be transported through the blood to the liver and spleen, and eventually cleared through the action of macrophages.

1.3.6.3. The role of complement in enhancing the antibody response

Whilst the complement system is not solely responsible for the antibody response it does play an important assisting role. Complement enhances the antibody response by assisting in the delivery of immune complexes to antigen presenting cells, and

also in the promotion of B-cell activation (Carroll, 2000). It has also recently been suggested that the complement system plays a role in preserving long-term immunological memory. Complement receptors have been found to be important in localising antigen and C3 to follicular dendritic cells (FDC), which is important in the maintenance of long-term immunity (Barrington et al., 2001; Carroll, 2000).

1.3.7. Human disorders associated with complement abnormalities

Given the central role of complement in the immune system, it is unsurprising that numerous human disorders are associated with complement abnormalities. There are two major categories of complement abnormalities: those associated with deficiencies, and those arising from the complement mediated damage to self resulting from an overactive or poorly regulated complement system.

1.3.7.1. Human disorders associated with complement deficiencies

Just about every important component of the complement system has a disorder associated with its deficiency, and a detailed discussion of all the known complement associated disorders is beyond the scope of this thesis. A summary of some of the more interesting ones will follow.

Disorders associated with classical pathway components include systemic lupus erythematosus (SLE), glomerulonephritis and immune complex diseases. Those associated with alternative pathway deficiencies include severe meningitis, and recurrent upper respiratory infection. Abnormalities in components involved with

MAC formation include recurrent *Neisserial* infections. Abnormalities in complement control proteins include haemolytic uraemic syndrome (HUS), and angioedema.

1.3.7.2. Haemolytic Uremic Syndrome

Familial haemolytic uremic syndrome is an inherited disorder that is associated with haemolytic anaemia, acute renal failure, and thrombocytopenia. This disorder has been associated with a number of mutations in complement factor H (fH). The majority of these mutations appear to be clustered at the C-terminus, and most reside in domain 20 (Perkins et al., 2002). The majority of these mutations therefore, occur within the region of fH that is able to bind C3d, and to one of the regions of fH that is able to bind heparin (modules 19-20). This would suggest that HUS is either caused or exacerbated by the inability of fH to bind to either C3d or polyanions. It is known that HUS is associated with endothelial cell activation, in a complement associated manner, and so it is possible (albeit not proven) that the HUS associated with fH mutations is due to the inability of fH to regulate the alternative pathway due to disrupted polyanion binding. It remains to be seen, however, whether this is indeed the case.

1.3.7.3. Human disorders associated with over active complement systems

There are several ways in which an over active complement system can lead to health problems and one such class is autoimmune disorders. In this case the immune system acts on host tissues and organs causing severe damage and possible

loss of organ function. Within this area are such important human disorders as arthritis and diabetes. It is not yet clear exactly how these disorders are initially triggered, however, what is known is that the patient suffers severe complement mediated tissue damage.

1.3.8. Microbial strategies to evade or exploit the complement system

It seems reasonable to suspect that there has been considerable interplay between the evolution of pathogens and the immune systems they are trying to evade. It should come as no surprise therefore that there are several known methods used by pathogens to evade complement attack. It is known that certain Gram-negative bacteria possess long O-linked polysaccharide chains, which are very efficient at activating the complement system. The problem from the host point of view is that this results in MAC formation at points distal from the cell membrane, and therefore the bacteria escape the eventual cell lysis. Other bacteria use a capsule that is rich in sialic acid, which appears to confer resistance to alternative pathway activation. It appears that possession of a capsule rich in sialic acid promotes the preferential binding of factor H rather than factor B to the tick-over C3b formed, resulting in the inhibition of alternative pathway activation (Jarvis, 1995). One of the most interesting methods of evading the complement system is found in organisms such as trypanosomes and certain viruses. These organisms have gained the ability to express proteins similar to the host's complement control proteins. For example, trypanosomes express variant surface glycoprotein (VSG) (Davitz et al., 1987), and Vaccinia virus expresses VCP which is a highly efficient soluble complement

inhibitor (Isaacs et al., 1992), the crystal structure of which has recently been solved (Murthy et al., 2001). This structure, shows several discrepancies when compared to the NMR structures of VCP34 (Wiles et al., 1997) and VCP 23 (Henderson et al., 2001). It is seems likely from the NMR structures that the intermodular arrangement between modules 2 and 3 is flexible, whereas in the crystal structure, this junction appears rigid. Further investigation is therefore needed into the nature of this junction to determine unequivocally, whether it is indeed flexible as suggested by the NMR structure.

As already stated, certain pathogens use components of the complement system as their receptors, however this is not the only method that pathogens use to exploit the complement system. Certain flavoviruses appear to actually use complement activation to facilitate their infections. An example of this is Dengue virus, which appears to utilise antibody and C3b attached to the virus to enter macrophages via their F_c receptors (Casali et al., 1983).

1.4. Protein glycosylation

1.4.1. Preliminary remarks

One of the most common protein modifications in eukaryotes is glycosylation, yet this near ubiquitous phenomenon is only poorly understood. The majority of extracellular eukaryotic proteins are glycosylated, often with the sugars making up the largest portion of the protein mass. The glycans on glycoproteins appear not to

be limited to a single function, instead there are examples where glycans or glycan recognition have been associated with protein folding (Ronnett et al., 1984; Slieker et al., 1986), sorting (Hauri et al., 2000), quality control (Hammond et al., 1994), structural integrity (Casasnovas et al., 1999; Wyss et al., 1995), the immune system (Rudd et al., 1998), signal transduction (Wells et al., 2001), prion pathogenesis (Rudd et al., 1999; Rudd et al., 2002) and, potentially, information storage (Lehrman, 2001). One of the reasons that sugars perform such a large number of tasks is possibly due to the extreme chemical variability that can occur in oligosaccharides, which vastly out weighs that of both proteins and nucleic acids. Unlike proteins and nucleic acids the sources of variation in the structures of oligosaccharides arises from their ability to branch, and to vary the linkage type and positioning combination with the large number of different monosaccharide units. It is precisely this variability and structural heterogeneity that often results in some of the major difficulties that are associated with studying glycoproteins. It appears that, even on the same protein expressed in the same cell there may be several distinct glycan moieties attached to a particular glycosylation site, referred to as the glycoforms. The function of these individual glycoforms is not understood, and our ignorance will not be improved until purifying and studying individual glycoforms is made routine. However, given the number of important functions that glycans possibly serve, surely it remains one of structural biology's most important goals to fully understand the structural effects of protein glycosylation.

Glycans may form associations with proteins in a number of different forms, such as proteoglycans, and GPI anchors, which are beyond the scope of this thesis. Instead I

shall provide a summary of the most important aspects of protein N-linked glycosylation, with only a cursory description of the many other kinds.

1.4.2. N-linked and other types of protein glycosylation

There are two major types of protein glycosylation; O-linked glycosylation and N-linked glycosylation, although there is increasing evidence for a third type, so called C-glycosylation (Hofsteenge et al., 1994). C-glycosylation occurs through the direct covalent attachment of a glycan to a carbon atom on a tryptophan residue. O-linked glycosylation involves the addition of sugar residues to the hydroxyl groups of certain serine and threonine residues through a condensation reaction. The reason why some serine/threonine residues become O-glycosylated is very poorly understood, although it can be predicted with dificulty and only limited success using comparisons with known O-glycosylated proteins. N-linked glycosylation on the other hand is the attachment of glycans to the side chain nitrogen of asparagine residues in the consensus N-X-S/T where X is any amino acid with the exception of proline. Whilst the presence of this sequon, does not guarantee that N-linked glycosylation will occur at this site, it can only occur at this sequon.

1.4.3. N-Linked glycosylation

The vast majority of secreted proteins receive N-linked glycans during their synthesis, whereas the majority of cytosolic proteins remain non-glycosylated. This phenomenon is largely because the core glycans are synthesised such that they face the lumen of the ER. In itself, this is hardly surprising given that N-linked glycans

are involved in a large number of extracellular recognition events, and enable the host to quickly alter the recognition properties of a protein. As already stated, protein N-linked glycosylation only occurs at the specific sequon N-X-S/T. The initial step in N-linked glycosylation is the synthesis of a specific core saccharide unit that is common to the N-linked glycans found in all organisms, with the sole exception of trypanosomatids. Next comes the attachment of these core sugars to the asparagine residue in the endoplasmic reticulum, in a process that occurs co-translationally (Chen et al., 1999). The factor(s) that govern which of the sequons that are glycosylated are not yet fully understood, however it appears that translation-speed (Ujvari et al., 2001), distance from the C-terminus (Nilsson et al., 2000) and steric effects may all be important. The next phase in N-glycosylation is the initial trimming of the core sugars in the ER with the concomitant association of the nascent protein with chaperones and helper proteins to facilitate disulphide bond formation and correct folding. The final stage of N-linked glycosylation is the further processing of the glycan moiety in the Golgi apparatus until the maturation of the glycoprotein is complete, and the protein is translocated to its final destination. Given the central nature of the glycosylation machinery to protein folding and recognition, it is not surprising that there are a number of diseases and disorders that are associated with aberrant glycosylation including muscular dystrophy (Muntoni et al., 2002). Each of the steps involved in protein N-glycosylation will be described in greater detail in the following sections.

1.4.4. Synthesis of the core sugars

The 'core sugars' are a 14-saccharide unit that is synthesised as a membrane bound dolicholpyrophosphate precursor ($Glc_3Man_9GlcNAc_2$-*PP*-dol) partly in the cytosol and partly in the lumen of the ER. This synthesis is a multistep process and requires the activity of enzymes that are located on either side of the ER membrane.

The underlying principals behind the synthesis of the core sugars and the early events in N-glycosylation have been known for some time, and will be summarised here (for a review see Burda et al., 1999). The initial stage in the synthesis of the core sugars is the synthesis of Man_5-$GlcNAc_2$-PP-dol, which occurs in the cytoplasmic side of the ER membrane. This stage occurs in two major steps (steps 1 & 2 in the schematic shown in Figure 1.10), the first of which is the synthesis of $GlcNAc_2$-PP-dol, and the second of which is the synthesis of $Man_5GlcNAc_2$-*PP*-dol. During the first step two moieties of UDP-GlcNAc are sequentially transferred to P-dol to give $GlcNAc_2$-*PP*-dol. Following this five moieties of GDP-Man are sequentially transferred to the $GlcNAc_2$-*PP*-dol produced in the previous phase. This yields a $Man_5GlcNAc_2$-*PP*-dol, which is also located on the cytoplasmic side of the ER membrane. This species is subsequently acted on by flippase (step 3 in the schematic shown in Figure 1.10), an enzyme with no synthetic activity, but which simply transfers the species to the lumenal side of the ER. All of the subsequent steps involved with N-linked glycosylation occur within the secretory pathway.

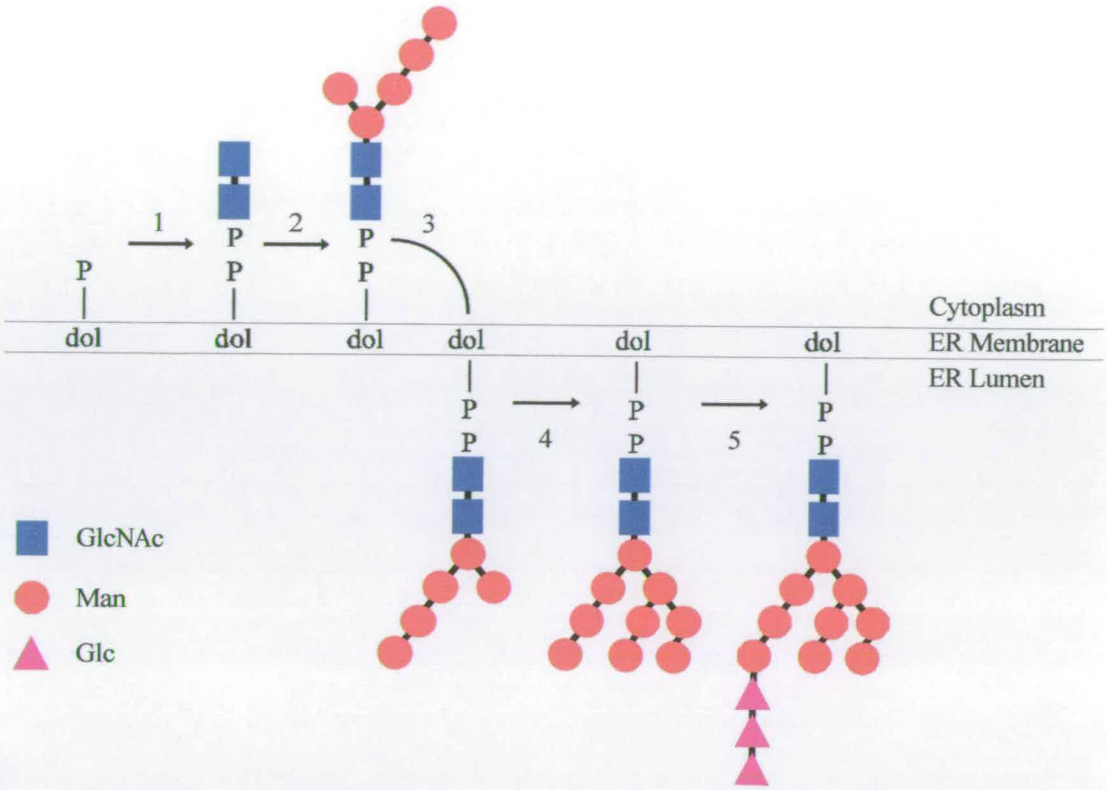


Figure 1.10 Schematic illustration of the steps involved in the synthesis of the core glycans of N-linked glycosylation. Step 1: transfer of two GlcNAc residues from UDP-GlcNAc to Dol-*P* yielding $GlcNAc_2$-*PP*-dol. Step 2: transfer of 5 mannose residues from GDP-Man to $GlcNAc_2$-*PP*-dol yielding $Man_5GlcNAc_2$-*PP*-dol. Step 3: flipping of the glycans from the cytoplasmic to lumenal side of the ER membrane. Step 4: transfer of 4 mannose residues from dol-*P*-Man to $Man_5GlcNAc_2$-*PP*-dol yielding $Man_9GlcNAc_2$-*PP*-dol. Step 5: transfer of 3 glucose residues from dol-*P*-Glc to $Man_9GlcNAc_2$-*PP*-dol yielding the fully formed $Glc_3Man_9GlcNAc_2$-*PP*-dol.

The next stage in the synthesis of the core sugars, involves the transfer of four moieties of dol-*P*-Man onto the $Man_5GlcNAc_2$-*PP*-dol to yield $Man_9GlcNAc_2$-*PP*-dol (step 4 in Figure 1.10). Following this, three moieties of dol-*P*-Glc are subsequently sequentially transferred to $Man_9GlcNAc_2$-*PP*-dol yielding the final core species of $Glc_3Man_9GlcNAc_2$-*PP*-dol. Interestingly these dol-*P*-Man and dol-*P*-Glc moieties are also synthesised on the cytoplasmic side of the ER membrane, and subsequently 'flipped' onto the lumenal side. The steps involved in the synthesis of the core sugars is summarised in the schematic illustration shown in Figure 1.10. It

should be noted that this scheme of core glycan synthesis is highly conserved, and is the case for wild type fungi, plants, and mammalian cells, however, exceptions are known to exist. In the case of trypanosomatid cells, the core sugars are synthesised unglucosylated as $Man_9GlcNAc_2$-*PP*-dol, $Man_7GlcNAc_2$-*PP*-dol or $Man_6GlcNAc_2$-*PP*-dol depending on the species. This case will not be discussed further, as none of the organisms connected to this study exhibit this type of glycosylation

1.4.5. Attachment of the core sugars to the nascent polypeptide chain

Attachment of the core sugars to the polypeptide chain takes place cotranslationally with synthesis of the nascent polypeptide chain (for a review see Parodi, 2000). The transfer occurs through the action of the enzyme oligosaccharyltransferase (OST). The reaction releases dol-*PP* with the concomitant covalent attachment of the remainder of the precursor onto the consensus Asn residue of the N-X-S/T sequon through a condensation reaction. OST, a highly complex enzyme, consists of eight subunits and is located in the ER membrane. Moreover, it has been demonstrated that in yeast cells OST can transfer the standard triglucosylated core sugars to the protein species 20 fold faster than it can transfer unglucosylated species (Ballou et al., 1986). It is thought that this difference in transfer efficiency is important for ensuring that the only fully glucosylated core sugars are transferred to the protein. This is thought to be an important point as the subsequent trimming of the glucose residues from the glycans on a nascent glycoprotein are directly involved in the folding process (see section 1.4.6 for a fuller description of the role of glycan processing during glycoprotein folding).

1.4.6. Initial processing of the core sugars and the role in protein folding

The initial steps in N-glycan processing appear to be reasonably well conserved between many diverse species. Indeed it has been demonstrated that there are few differences between these processes between mammalian cells and *S. cerevisiae* This observation is probably largely due to the effects that these processes exhibit during the protein folding process. It is thought that the initial processing of the N-glycans follows the basic scheme: partial deglucosylation, full deglucosylation followed by demannosylation. The deglycosylation steps are highly important for recruitment of appropriate chaperones, whereas the initial demannosylation step seems to acts like an indicator that folding is complete. The initial step of partial deglucosylation is thought to be due to the action of glucosidase I, removing the outermost glucose residue, followed by the action of glucosidase II removing the next glucose residue (see Figure 1.11 for clarification) from the N-glycans of the nascent glycoprotein. This may not always be the case, however, and it may be that the glycans are fully deglucosylated (either by glucosidases I and II, or by glucosidase II alone), followed by partial reglucosylation yielding the monoglucosylated glycan. The exact nature of this deglucosylation process is largely unimportant, however, as the important factor is that the glycans achieve a monoglucosylated state. This monoglucosylated form of the glycan facilitates the recognition of the nascent glycoprotein by the ER resident lectins calreticulin (CRT) and calnexin (CNX), which have been shown to play an important role in the folding

of many proteins (Peterson et al., 1995). Moreover, it appear that it is not simply the monoglucosylated glycan that is recognised by CRT and CNX, but instead it appears that recognition by these lectins requires interactions between the lectin and both glycan and protein moieties.

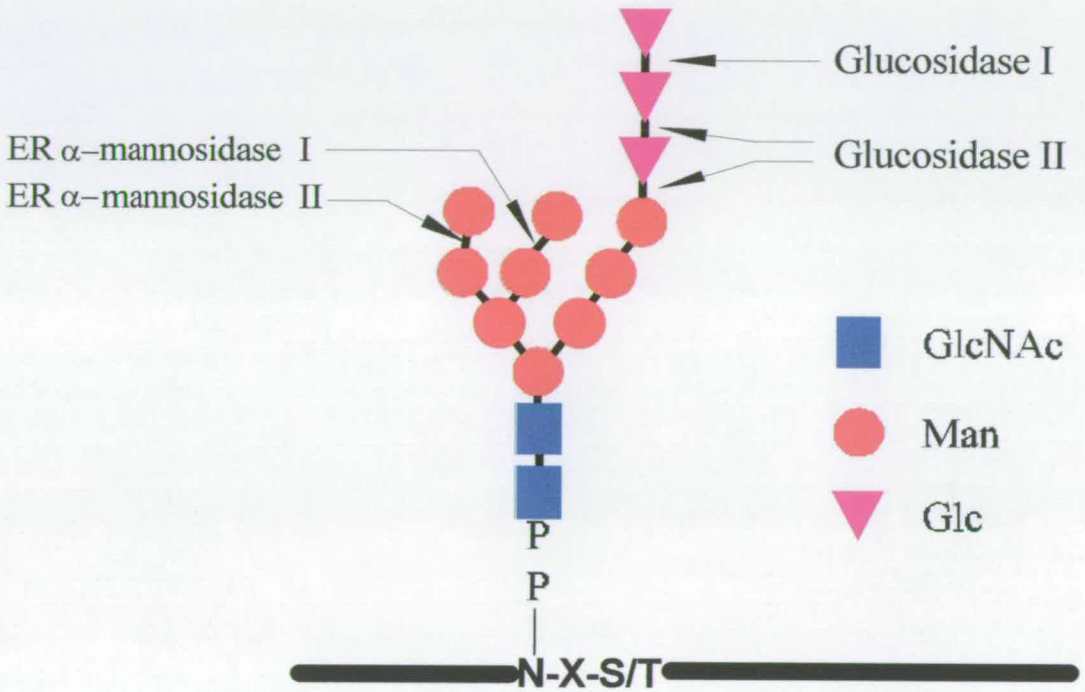


Figure 1.11 Schematic representation of the core glycans attached to the consensus Asn residue of a protein. The diagram shows the sites involved in the initial trimming of the glycans during protein folding.

The binding of either CNX or CRT to the monoglucosylated nascent glycoprotein facilitates the binding of Erp57. Erp57 is a cellular folding factor that possesses thiol oxidoreductase activity, and is able to bind both either CNX or CRT. The association of the nascent monoglucosylated protein with CNX or CRT and Erp57 facilitates the folding and the formation of any necessary disulphide bonds through disulphide bond shuffling. The glycans of the nascent glycoprotein are subsequently

deglucosylated and the nascent glycoprotein dissociates from Erp57 and CNX or CRT. Following this primary folding step the fate of the protein and the glycans are significantly linked. The glycans may be reglucosylated, facilitating the reassociation with CNX or CRT and Erp57, with concomitant refolding and retention in the ER, or this reglucosylation may not occur resulting in the glycoprotein being 'released' and passing through the remainder of the secretory pathway The enzymes involved in the attachment and removal of the terminal glucose residue to/from the glycans are glucosidase II, and UDP-glucose: glycoprotein glucosyl transferase (GT). It important to note that GT appears to be able to differentiate between properly folded and improperly folded glycoproteins, and only appears to reglucosylate the glycans attached to improperly folded glycoproteins. In this way GT appears to act as a folding sensor allowing correctly folded glycoproteins to pass through the secretory pathway, whilst improperly folded proteins are retained in the ER for further folding attempts until such time as they become correctly folded, or are subsequently targeted for proteosomal degradation. It can be seen therefore that the early events in N-linked glycosylation are intimately linked with disulphide bond formation and proper folding of nascent glycoproteins.

1.4.7. Further processing of the glycans in the Golgi apparatus

Following the initial processing of the N-linked glycans in the ER, and the concomitant folding of the nascent protein, these glycans may be further processed in the Golgi apparatus to yield the mature glycoform of the protein. At this point the glycosylation profiles observed tend to differ between higher and lower eukaryotes.

There are three major classes of N-glycans: high mannose, hybrid, and complex, and a schematic representation of some of the more common types of N-glycans in these classes is illustrated in Figure 1.12.

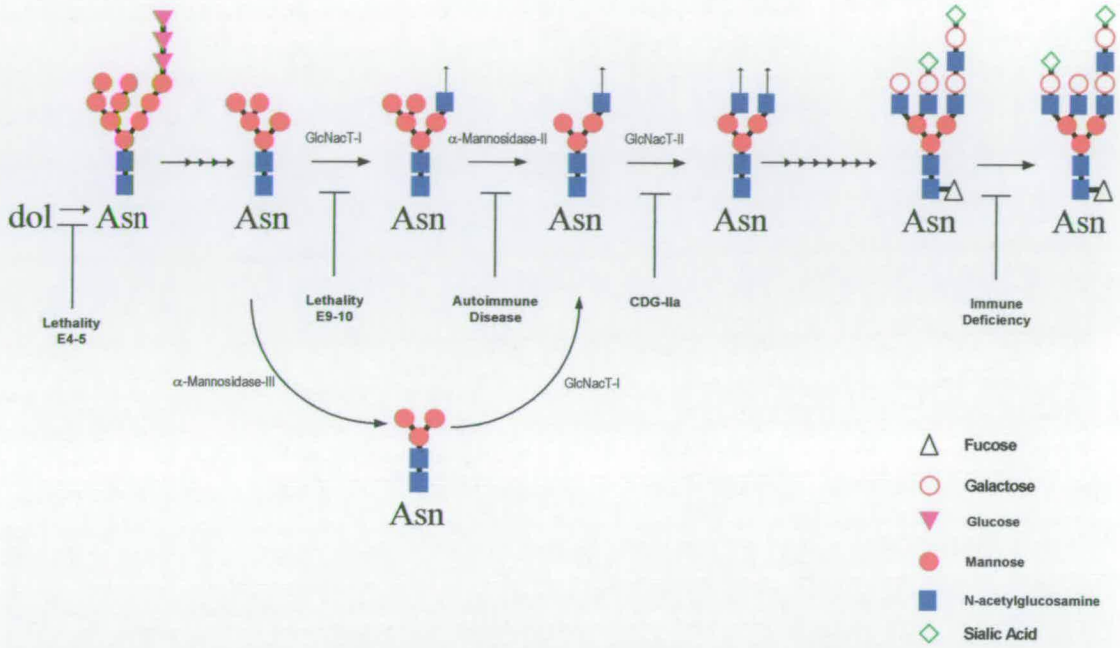


Figure 1.12 Schematic illustration of different types of N-glycan structure, the enzymes involved in their processing, and associated diseases and disorders (adapted from Chui et al 2001).

High mannose-type glycans are in effect little more than the core sugars. They consist solely of GlcNac and Man residues and exist commonly as $Man_{8-9}GlcNac_2$ structures, although they may become trimmed down. This type of glycosylation pattern is the typical of the types produced by many types of lower eukaryotes, particularly yeasts, and is the one that is most commonly produced by *P. pastoris*. In the case of some heterologous proteins produced by *P. pastoris*, however, the phenomenon of hyperglycosylation may be observed. During this process the arms of the glycan are significantly extended by the addition of further mannose residues (Cereghino et al., 2000). For hyperglycosylation in *P. pastoris* it is known that these

glycans are comprised entirely of GlcNac and mannose residues, although it has been demonstrated that in some cases at least, these glycans are only partially sensitive to α-mannosidase. It was subsequently discovered that there are phosphate groups between some of these mannose residues that were not sensitive to alkaline phosphatases (Martinet et al., 1997). Hyperglycosylation of proteins may reduce the activity of some enzymes, and may result in the protein being highly antigenic (Cereghino et al., 2000) The hyperglycosylation of recombinant proteins therefore becomes an interesting problem, that will need to be addressed before their full therapeutic use can be realised

Complex glycans can exhibit great diversity that is dependant on a number of factors, including: organism, cell type, position in the developmental cycle etc. There are however a number of generalisations that differentiate complex glycans from high mannose type glycans. Complex glycans are fucosylated on the GlcNac residue attached to the asparagine residue. In addition to the fucosylation, the Man residues get trimmed until three remain. These are then elongated through the addition of other types of sugars: initially GlcNac, then Gal, and finally the terminal branches are often ended with sialic acid. Hybrid glycans, as the name suggests, lie between the extremes of high mannose and complex glycans, and have no sialic acid or fucose residues. Hybrid glycans usually have only one or two elongations to the central mannose sugars.

The terminal sialic acids found in complex glycans, can have important consequences in the context of the immune system, since there is evidence that these

allow self from non-self discrimination in the complement system (Meri et al., 1990; Meri et al., 1994; Pangburn, 2000). It is also interesting to note that mutation in the α-mannosidase II gene induces a systemic autoimmune disease similar to systemic lupus erythematosus (Chui et al., 2001). α-mannosidase II is one of the enzymes involved in the hybrid to complex transition, and therefore one would expect a good number of cell surface glycoproteins to have fewer than the normal number of sialic acid residues. If this were the case, it is not inconceivable that the ensuing autoimmune disease is due, at least in part to a failure of the complement system alternative pathway to discriminate self from non-self.

1.4.8. Effects of the N-linked glycans on glycoproteins

The precise function of the N-linked glycans in many glycoproteins is often poorly understood. Although, as previously described, much of the importance associated with N-linked glycosylation lies with the links between the initial steps in N-glycosylation and protein folding, there are, cases where the glycans themselves play a functional role.

The role of glycans are many, and protein glycan interactions may be involved in several important ways, including recognition, immunological processes, and structural phenomena. Glycans can mask important regions in a protein that would otherwise be exposed, including the sites of protease sensitivity, or antigenic sites. Examples include HIV, which, uses glycosylation to mask potential antigenic sites (Reitter et al., 1998). The interaction between a protein and a N-glycan can take a

number of forms, and cases exist where the protein affects the structure of the glycan, and where the glycan affects the structure of the protein. Cases where the protein affects the glycan include that of CD2, where contacts between the protein and the first two GlcNac residues reduce the mobility of the core sugar residues resulting in the glycan 'lying' parallel to the peptide (Wyss et al., 1995). An example of a case where the glycans are affecting the peptide is that of the α-subunit of human chorionic gonadotropin (hCG), where the glycans have a significant effect on the mobility of the protein (De Beer et al., 1996; van Zuylen et al., 1997), without having a drastic effect on the overall conformation of the protein. It is further interesting to note that contacts between the glycan and protein are only observed in the free α-subunit rather than in the intact heterodimer.

1.5. Signalling Lymphocyte Activation Molecule

1.5.1. Preliminary remarks

Signalling lymphocyte activation molecule, also referred to as SLAM, or IPO-3, is a co-stimulatory protein on B lymphocytes, and dendritic cells. SLAM was initially identified using a monoclonal antibody that bound to an unidentified antigen, and activated T-cells (Cocks et al., 1995). SLAM facilitates the proliferation of activated B cells, upon binding to its ligands. This proliferation of activated B cells is associated with increased immunoglobulin production. SLAM, therefore is a receptor for co-stimulatory molecules, rather than a co-stimulatory molecule in its own right. Signalling through SLAM is also associated with increased levels of

interferon-gamma production. SLAM is expressed on thymocytes, B lymphocytes, dendritic cells, endothelial cells and certain subpopulations of T lymphocytes. Following activation , the expression of SLAM is up regulated.

1.5.2. Domain structure of SLAM

SLAM is a member of the immunoglobulin superfamily, and is within the CD2 subset. SLAM is a type I membrane protein, and exist as one of three isoforms - the result of alternative splicing. The full length protein has a Mr of 70 kDa. Of the remaining isoforms, one is a secreted form lacking the transmembrane domain, and the other has C-terminal truncation, yielding a cytoplasmic domain that is 36 amino acids shorter than that of the full length version.

The transmembrane region of SLAM is predicted to be comprised of between 21 and 27 amino acids and is followed by an intracellular domain comprising the C-terminal 77 amino acids. This region contains four tyrosine phosphorylation sites, three of which are predicted to form SH2 binding domains and which are thought to be involved in subsequent cell signal transduction pathways.

Figure 1.13 Model of the extracellular Ig domains of SLAM. The model is based on the structure of the CD2-CD58 heterodimeric complex, and has the glycans shown in random orientations (taken from Mavaddat et al., 2000).

The extracellular region of SLAM is comprised of two immunoglobulin domains. The N-terminal domain is of the Ig V type, and the membrane proximal domain is of the Ig C2 type. The V type domain contains five potential N-linked glycosylation sites, and two Cys residues which are predicted to be disulphide bonded. The C2 type domain, contains four potential N-linked glycosylation sites, and four Cys residues, which are predicted to form disulphide bonds in a 1-4, 2-3 pattern. From the sequence data it appears that SLAM would be expected to be highly glycosylated, although information concerning which of the glycosylation sites are recognised is as yet unavailable.

A model of the extracellular immunoglobulin domains of SLAM based on its homology to CD2 and CD 58 is shown in Figure 1.13, and is shown with the potential N-linked glycans in a random arrangement (Mavaddat et al., 2000). The model shows SLAM as a homodimer, because it is predicted to be homophilic, and have the ability to act as a self ligand (see section 1.5.4).

1.5.3. Receptors for SLAM

As SLAM is involved in cell signalling events, it is clear that SLAM must form interactions on both sides of the cell membrane. As yet the only extracellular receptor for SLAM that has been identified is SLAM itself, however, this does not rule out the possibility that as yet undiscovered binding partners exist. There are two important intracellular receptors for SLAM that have been discovered to date. One -

EAT-2 is expressed in B-cells, and is responsible for SLAM's B-cell signalling; the other - SLAM associated protein - SAP (also referred to as SH2D1A) is expressed in T-cells and is responsible for SLAM's T-cell signalling activity.

There are interesting differences between SLAM's two main intracellular partners, concerning their ability to interact with SLAM in different phosphorylation states. EAT-2 is able to interact with SLAM only when both Tyr307 and Tyr327 of SLAM are phosphorylated, whereas SAP is able to interact with SLAM in both its phosphorylated, and dephosphorylated states (Morra et al., 2001). It seems likely that this could explain the differences between the effects of SLAM activation in B-cells as compared to T-cells.

1.5.4. SLAM may be able to perform bidirectional signalling

As indicated the only known extracellular receptor for SLAM is itself, and that SLAM – SLAM interactions allow bidirectional signalling between B-cells and T-cells, *i.e.* SLAM has the ability to act as a self-ligand. It had been proposed that the self affinity for SLAM is very high, lying somewhere in the low nM range (Punnonen et al., 1997). Subsequent experiments using SPR technology, however, imply that the self-affinity is rather weak, with an affinity in the region of 200 μM (Mavaddat et al., 2000). This would be a lower affinity that that usually found with interactions on the leukocyte cell surface, which usually have dissociation constants in the range of 0.3 – 20 μM (Davis et al., 1998). The difference in the affinity data between the two reports is quite hard to reconcile, and it was suggested by

(Mavaddat et al., 2000) that the apparent extremely high affinity could be due to protein aggregation occurring during the course of the experiments, although this appears to be speculation. It would seem illogical however for a protein involved in signalling events, to act in what would be essentially an irreversible manner. On the other hand, the low affinity suggested may rule out the possibility that SLAM self-association has any physiological relevance. It is clear, however, that this question needs to be resolved before any firm conclusions can be drawn.

1.5.5. SLAM is associated with X-linked lymphoproliferative disease

In addition to the afore mentioned action of SLAM as a receptor for measles virus, SLAM is associated with another clinical condition, namely X-linked proliferative disease (XLP). XLP was identified as a distinct clinical disorder in 1969, when Purtilo recorded the deaths of six out of 18 male members of the Duncan family, (reported as Purtilo et al., 1975). The patients were all between the ages of 2 and 19 years of age, and all died of a lymphoproliferative disease, which was initially termed Duncan's disease. The disease is characterised by dysgammaglobulinaemia, atypical lymphocytosis and lymphadenopathy, and a significant number of the affected boys presented with infectious mononucleosis, towards the latter stages of the disease. Initially it was thought that EBV infection was required for a patient to develop XLP, however subsequent studies have indicated that this is not necessarily the case. The gene associated with XLP was subsequently identified as SAP – SLAM's intracellular T-cell partner (Sayos et al., 1998). It is hoped that a more

thorough understanding of SLAM signalling and its interaction with SAP could lead to potential therapeutic agents to treat XLP.

1.6. Summary and project aims

Measles virus is able to use two different proteins for its cellular receptor – MCP and SLAM. MCP possesses an N-linked glycan on its second module that is essential for functionality in a manner that implies it is structuraly significant. This thesis describes the investigations performed on the measles binding regions of its receptors MCP and SLAM.

The expression and characterisation of both the MCP12 module pair and MCP2 single module are described. Particular attention was payed to understanding the underlying mechanisms by which the functionally essential glycans on the second module of MCP exert their effects through the use of biophysical techniques. These investigations form part of an ongoing effort aimed towards a complete elucidation of the mechanisms behind both the effects the module 2 glycans have on MCP and the interaction between MCP and the measles virus.

The expression of the measles interacting region of the second measles receptor are also described, with the eventual goal of understanding the reasons why measles virus is able to utilise two different receptors from different protein families.

2. Materials and Methods

All aqueous buffers were made up in Milli-Q grade water. The buffer components were of the highest available grade and were typically purchased from Fischer or Sigma-Aldrich. Centrifugation of volumes greater than 1.5 ml was carried out either in a Sorvall RC 5C plus, a Sorvall 26 plus or a Sorvall legend RT centrifuge. Centrifugation of volumes less than 1.5 ml was carried out in a Hereus Biofuge Pico. Concentration of large volumes of protein (>300 ml) was achieved using hollow fibre prep scale 'TFT' cartridges (Millipore), 5 kDa molecular weight cut off. Concentration of protein volumes 10-300 ml was performed using a high-pressure stirred cell with ultrafiltration membranes (Amicon) with a molecular weight cut off as described. Concentration of protein volumes of less than 10 ml was performed using centrifugal concentrators (Viva-science) with molecular weight cut off membranes as described. UV/Vis spectroscopy was carried out in an 8543 UV/Vis spectrophotometer (Hewlet Packard) using either quartz ($\lambda < 290$ nm) or disposable cuvettes ($\lambda > 290$ nm).

2.1. General DNA Methods

2.1.1. Polymerase Chain Reaction (PCR)

DNA fragments were amplified using the PCR in a Techne Cyclogene™, a Stratagene Robocycler™ or Eppendorf 'Master Cycler™ Personal' thermal cycler. Reactions carried out in the Techne Cyclogene thermal cycler were overlaid with 50 µl of mineral oil. This step was unnecessary for reactions carried out in the other two thermal cyclers. The high fidelity DNA polymerases 'Deep Vent' (New England Biolabs, NEB) 'Herculase' or 'Herculase Hot Start' (Stratagene) were used for gene amplification. The reactions were assembled on ice in sterile 0.5 ml (Cyclogene thermal cycler) or 0.2 ml (other thermal cyclers) thin walled PCR tubes. The thermal cycler was preheated to 96 °C before the tubes were inserted. Reactions were carried out on a 50 µl scale, and were typically set up as follows:

Component	Quantity	Final concentration
10 X reaction buffer	5 µl	1 X concentration
dNTP mix (10 mM each)	1 µl	200 µM
Template DNA (10 ngμl^{-1})	1 µl	200 pg.μl^{-1}
Primers (100 ngμl^{-1})	1 µl each	2 ng.μl^{-1}
DMSO	2.5 µl	5 %
Polymerase (5 Uμl^{-1})	0.5 µl	0.05 U.μl^{-1}
d.H_2O	40 µl	
Total reaction volume	38 µl	

2.1.2. Agarose gel electrophoresis

Agarose gels were made using molecular biology grade agarose (Bio-Rad) in TAE buffer (see Appendix A) using the gel percentages given containing 0.2 µg.ml^{-1}

ethidium bromide, and were run in a minielectrophoreisis gel tank (Anachem) at a constant current of 80 mA. The DNA bands were visualised using long wavelength UV radiation.

2.1.3. DNA extraction from Agarose gels

Following electrophoresis the relevant DNA bands were excised form the agarose and purified using the Qiagen gel extraction kit, which is based on (Vogelstein et al., 1979), according to the manufacturer's instructions. The DNA was eluted from the spin columns in autoclaved deionised H_2O, and the concentration estimated by A_{260} measurement.

2.1.4. Ethanol precipitation of DNA

DNA was precipitated using ethanol. Sodium acetate, pH 5.2, was added to the DNA to a final concentration of 300 mM, followed by the addition of two volumes of 95 % (v/v) ethanol (-20 °C). The mixture was then incubated at 0 °C for 30 minutes, centrifuged at 13,000 g for 30 minutes, and the supernatant poured off. The pellet was then washed twice with 1 ml of 70 % ethanol and centrifuged at 13,000 g for 5 minutes between each wash. The pellet was dried at 37 °C, and redissolved in autoclaved milli-Q water for 30 minutes at room temperature before storage at –20 °C until required.

2.1.5. Restriction Digests

Restriction enzymes were typically purchased from NEB. Restriction digests were carried out on a 20 µl scale unless otherwise stated. Digests were carried out for 2 hours at 37 °C unless otherwise stated. As excessive concentrations of glycerol can inhibit the action of some restriction enzymes, the total amount of enzyme added to any reaction was less than or equal to 5 %, to keep the final glycerol concentration less than 2.5 %.

2.1.6. DNA Ligation

DNA ligation was conducted according to (Sambrook et al., 1989) at 16 °C for 16 hours using T4 DNA ligase (NEB). Alternatively ligation was performed using the Roche 'rapid ligation' kit for 5 minutes at room temperature according to the manufacturer's instructions. The ligation reactions were carried out using the vector: insert ratios described; and were typically performed on a 10 µl scale with 1 Weiss unit of T4 DNA ligase per reaction unless otherwise stated.

2.1.7. *E. coli* Transformation

E. coli (strains as described) was transformed with circular plasmid DNA using the calcium chloride method (Sambrook et al., 1989), or by using 'One Shot' chemically competent cells (Invitrogen), according to the manufacturer's instructions. The cells were transformed with 5 µl of a ligation reaction unless otherwise stated. Following the transformation procedure the cells were plated onto LB (see Appendix A)

containing the appropriate antibiotic – 25 $\mu g.ml^{-1}$ Zeocin for pPICZα, or 100 $\mu g.ml^{-1}$ Ampicillin for pPIC9, pGEM-T, Pet15b, and Pet32. In the case of LB Zeocin plates the LB contained only 5 $g.l^{-1}$ of NaCl instead of the usual 10 $g.l^{-1}$, to prevent inactivation of the zeocin. Following transformation the plates were incubated at 37 °C for 16 hours after which time colonies could be picked for "miniprep". In the case of pGEM-T, the plates were then refrigerated for a further 8 hours to allow the blue colour to develop in negative transformants, without the formation of satellite colonies.

2.1.8. Plasmid minipreps, midipreps and maxipreps

Plasmids were recovered from *E. coli* using the Qiagen "Qiaprep Spin Miniprep" kit (~12 μg); the Promega "Wizard Plus Midiprep", or Sigma "GeneElute Plasmid Maxiprep" kits (250 μg), according to the manufacturer's instructions. Each kit uses a modified version of the alkaline lysis method followed by high salt-mediated adsorption of DNA onto a silica matrix. In all cases cultures were inoculated with a single colony from a freshly incubated selective plate. Cultures for miniprep were grown in 5 ml LB (plus appropriate antibiotic) and typically incubated for 5-6 hours at 37 °C with shaking at 250 rpm. Cultures for midiprep and maxiprep were grown in 100 ml, and were incubated overnight at 37 °C with shaking at 250 rpm. The DNA produced by each kit was finally eluted with autoclaved deionised water. The DNA concentration was determined using A_{260} measurements after the purity of the plasmid DNA had been established by confirming that the A_{260}/A_{280} ratio was 1.7-1.8 (Sambrook et al., 2000).

2.1.9. DNA sequencing

Fluorescent automated DNA sequencing was performed using a modification of the dideoxy method (Sanger et al., 1977), using either 'Big-Dye' or dRhodamine dye terminators, according to the manufacturer's instructions (Applied Biosystems). The sequencing gels were run by the Automated DNA Sequencing Service (Institute of Cell and Molecular Biology - University of Edinburgh) and were run on an ABI 377 XL sequencer.

2.1.10. Transformation of *P. pastoris* by electroporation

Expression of the proteins in *P. pastoris* was done in one of three strains: KM71, KM71-His, or X33. The yeast was transformed by electroporation in a 0.2 cm electroporation cuvette in a Bio-Rad *Gene Pulser* (1.5 kV, 200 Ω, 25 μF). All the steps in the electroporation process were carried out at 0 °C except where stated. A single colony of *P. pastoris* was taken from a freshly streaked plate and used to inoculate 5 ml of YPD (see Appendix A). The culture was incubated overnight (30 °C, 200 rpm). An aliquot containing 1.0 ml of this culture was used to inoculate a further 100 ml of YPD, and was incubated (30 °C, 250 rpm) until an A_{600} of 1.3-1.5 was achieved. The cells were harvested by centrifugation under sterile conditions (1500 g, 5 minutes at 4 °C) and resuspended in 100 ml of autoclaved Milli-Q water. The cells were then harvested again under the same conditions and resuspended in 50 ml of water followed by harvesting and resuspension in 4 ml of 1 M sorbitol. The cells were then harvested for a final time and resuspended in 200 μl of 1 M sorbitol. The cells (80 μl) were then incubated on ice for 5 minutes with 10 μg of linearised

plasmid (1 $\mu g.\mu l^{-1}$), and pulsed (1500 V, 25 μF and 200 Ω). Immediately after pulsing 1 ml of ice-cold 1 M sorbitol was added to the cuvette, and the contents transferred to a 15 ml sterile centrifuge tube. The mixture was then incubated at 30 °C with no shaking for 2 hours. Transformants were selected on YPDS (yeast extract peptone dextrose sorbitol – see Appendix A) + 100 $\mu g.ml^{-1}$ Zeocin (pPICZα,), or MD plates (pPIC9).

2.2. General protein methods

2.2.1. Trichloroacetic acid (TCA) precipitation of proteins

A half-sample volume of TCA (30 % w/v) was added to each sample. The samples were then incubated on ice for 30 minutes, and the resulting precipitate was collected by centrifugation at 13,000 g for 20 minutes at room temperature. The supernatant was then poured off, and the pellet washed twice through the addition of 1 ml aliquots of a mixture of ethanol and diethyl ether (1:1 v/v) followed by centrifugation at 13,000 × g for 2 minutes at room temperature. The pellet was then dried by incubation at 37 °C for 10 minutes prior to sodium dodecyl sulphate (SDS)-polyacrylamide gel electrophoresis (PAGE).

2.2.2. SDS-PAGE

SDS-PAGE was carried out according to the method of (Laemmli, 1970) either on a Bio-Rad Mini-protean-3 mini SDS-PAGE system using 10-20 % acrylamide gradient

Tris-HCl 'Ready-Gels', or on a Bio-Rad Criterion SDS-PAGE system using 10-20 % acrylamide gradient Tris-HCl Criterion gels. SDS-PAGE gels were run at a constant voltage of 200 V for 50 minutes (Ready-Gels) or 80 minutes (Criterion gels).

2.2.3. Immuno-blotting

The protein was subjected to SDS-PAGE (as above) and transferred to nitrocellulose membrane in 'Towbin Buffer' (see Appendix A) at a constant current of 150 mA for 2 hour in a Bio-Rad western blotting system. Following transfer the membrane was blocked by incubation for 2 hours at room temperature in PBS (see Appendix A) supplemented with 5 % non-fat dried milk (blocking buffer) on a bottle-rolling platform. The membrane was then incubated overnight with the primary antibody diluted in blocking buffer (dilutions and identity as given). Following extensive washing in PBS supplemented with 0.05 % (w/v) TWEEN20. The membrane was then incubated for 2 hours with the secondary antibody diluted in blocking buffer (dilutions and identity as given) at room temperature on a rolling bottle platform. Following extensive washing in PBS (peroxidase conjugated secondary) or TBS (phosphatase conjugated secondary) the membrane was stained with 3,3'-Diaminobenzidine (DAB) (peroxidase conjugated secondary) or 5-Bromo-4-chloro-3-indolyl phosphate/Nitro blue tetrazolium (BCIP/NBT) (phosphatase conjugated secondary).

2.2.4. Test inductions of recombinant *P. pastoris* clones

Single colonies were purified by streaking on the appropriate solid media, and a single colony used to inoculate 10 ml of buffered minimal glycerol medium (BMG - see Appendix A) in a 50 ml sterile disposable centrifuge tube. The culture was incubated for 24 hours at 30 °C with agitation at 225 rpm, and the cells harvested by centrifugation at 2000 g for 5 minutes; the supernatant decanted; and the pellet resuspended in 5 ml of buffered minimal methanol medium (BMM - see Appendix A). The culture was incubated as before for 4 days, and was fed daily to 0.5 % (v/v) with methanol. Aliquots of 1 ml were taken and tested for expression using SDS-PAGE.

2.2.5. Large scale growth and induction of recombinant *P. pastoris* KM71/KM71H clones

A single colony was picked from a freshly streaked plate of the clone of interest, and used to inoculate 10 ml of BMG. YPD + 100 $\mu g.ml^{-1}$ Zeocin were used for clones containing pPICZα, and minimal dextrose medium (MD - see Appendix A) plates used for clones containing pPIC9. The culture was incubated overnight at 30 °C with shaking at 250 rpm. This culture (5 ml) was used to inoculate 100 ml of fresh BMG, and the resulting culture was incubated overnight at 30 °C with shaking at 250 rpm. Four two litre conical flasks each containing 500 ml of BM were inoculated with 20 ml of the previous culture and incubated overnight at 30 °C with shaking at 250 rpm. The cells were then harvested by centrifugation at 1000 g for 5 minutes at 4 °C, in sterile centrifuge pots, and resuspended in 1 l of BMM. The resuspended

culture was then split equally between four two litre conical flasks and incubated at 29 °C with shaking at 250 rpm for the optimal induction time. The cultures were fed daily to a final concentration of 1 % (v/v) with methanol to replace methanol that had been used by the yeast or had been lost by evaporation. Following induction the culture was harvested by centrifugation at 3000 g for 10 minutes at 4 °C, and the supernatant retained.

2.2.6. Chromatography

All chromatographic steps with the exception of ConA affinity chromatography were carried out on a Waters 650E FPLC system fitted with a 996 photodiode array detector, or on a BioCad 700E protein purification system (Applied Biosystems).

2.2.7. Protein deglycosylation

Protein samples were deglycosylated using the endoglycosidase $EndoH_f$ (NEB). $EndoH_f$ is a fusion protein between the EndoH gene from *Streptomyces plicatus* (Robbins et al., 1984) and the maltose binding protein affinity tag. Protein samples were deglycosylated with 1,000 U of $EndoH_f$ for each glycosylation site per milligram of recombinant protein (1 NEB unit is equivalent to 1 IUB milliunit) in 50 mM sodium phosphate, pH 5.5, with 1 mM Phenylmethylsulfonyl fluoride (PMSF) and 5 mM EDTA for 4 hours at 37 °C typically in a total volume of 5 ml.

2.2.8. Protein N-terminal sequencing

N-terminal sequencing was performed by Edman degradation on an ABI 477A automated sequencer at the Welmet protein characterisation facility, University of Edinburgh. Purified protein samples were run on SDS-PAGE and transferred to 'ProBlot' polyvinylidene difluoride (PVDF) membrane, in Towbin buffer (see Appendix A) at a constant current of 150 mA for 2 hours. Protein bands were stained using 0.1 % Coomassie Blue G-250 in methanol, and destained with 10 % (v/v) acetic acid, 50 % (v/v) methanol. Following de-stain, the membrane was dried and the band excised.

2.2.9. Mass spectrometry (MS)

Mass spectrometry was performed using positive ionisation atmospheric pressure electrospray and was carried out on a Platform II single quadrupole mass spectrometer (Micromass). Purified samples were desalted using reverse phase chromatography on a C4 column and the fractions containing protein were freeze-dried and redissolved in acetonitrile/H_20 (1:1, v/v) containing 0.5 % formic acid. Cone voltages were ramped to give the optimum signal-to-noise ratio over a wide m/z range.

2.2.10. Carboxymethylation of MCP2 to determine the number of disulphide bonds

MCP2 was deglycosylated with PNGaseF (NEB) by incubating MCP2 with 1000 U of PNGaseF per mg of MCP2, in 50 mM potassium phosphate, pH 7.5, at 37 °C for 4

hours. Three samples of the PNGaseF-deglycosylated MCP2 in 100 mM Tris, pH 8.5, were prepared (100 µl of 100 $\mu g.\mu l^{-1}$). One sample was subjected to liquid chromatography-mass spectrometry (LC-MS) untreated; another was incubated in the presence of 40 mM neutralised iodoacetate at room temperature for 30 minutes, and was then subjected to LC-MS. The third was denatured and reduced by incubating for 1 hour at 37 °C in the presence of 10 mM DTT, 8 M Urea, and 100 mM Tris, pH 8.5. Neutralised iodoacetate was subsequently added to a final concentration of 40 mM, and the reaction incubated at room temperature for 30 minutes. This sample was then also subjected to LC-MS.

2.2.11. Removal of glycoproteins and cleaved sugars

The oligosaccharides cleaved during deglycosylation, and any protein that remained glycosylated following deglycosylation, were removed from the sample using "ConA Sepharose" (Amersham-Pharmacia Biotech). ConA is a lectin isolated from the jack bean *Canabalia ensiformis*, and binds to the terminal α-D-mannose and α-D-glucose residues of certain oligosaccharides, along with sterically related residues with available C-3, C-4 or C-5 hydroxyl groups (Reeke et al., 1975). Following enzymatic deglycosylation of proteins, the sample pH was increased to 7.5 using 1 M Trizma base (Sigma) and 4 M sodium chloride was added to a final concentration of 0.5 M. Removal of the sugars was typically achieved using 5 ml of ConA sepharose per litre of BMM *P. pastoris* induction medium. The sample was applied to a gravity fed, 5 ml, ConA sepharose column that had previously been equilibrated with binding buffer (20 mM Tris, pH 7.5, 0.5 M NaCl), and the flow-through collected in

2 ml fractions. A_{280} was recorded for each fraction and when the A_{280} reading had returned to near baseline, elution buffer (20 mM Tris pH 7.5, 0.5 M NaCl, 1 M methyl-α-D-glucopyranoside) was added. Again 2 ml fractions were collected and A_{280} readings taken.

2.3. Expression, purification and characterisation of MCP12

2.3.1. Cloning of MCP12 into pPIC9

The fragment corresponding to the N-terminal two modules of MCP (MCP12) was amplified by PCR using MCP cDNA as a template (BC2isoform cloned into pSG5, kindly supplied by Dr. J. Atkinson from the Washington University School of Medicine, St. Louis). The PCR reaction was set up as described in section 2.1.1, and was cycled according to the following parameters:

Temperature	Time	No. Cycles
95 °C	2 min	1
95 °C	30 Sec	35
55 °C	45 Sec	
74 °C	1 min	
74 °C	10 min	1
4 °C	Hold	

The sequences of the primers were:

MCP1 5'

5' GGGGAATTCTCCGATGCCTGTGAGGAGCCACCAAC 3'

MCP2 3'

5' GCGGCGGCCGCCTATCACAAAACCTTTTCACATATTGG 3'

The restriction sites *Eco*R1 and *Not*1 were incorporated into the 5' and 3' primers respectively to facilitate directional cloning. Stop codons were incorporated into the 3' primer to allow expression of the protein without the extra residues that would otherwise be incorporated from the vector.

The fragment obtained from PCR was subjected to agarose gel electrophoresis (1.4% gel), and was subsequently purified from the gel using the Qiagen gel extraction kit according to the manufacturer's instructions. The fragment was then precipitated with ethanol according to section 2.1.4, and was redissolved in 20 µl of autoclaved milli-Q water. The fragment was then digested with *Eco*R1 and *Not*1, for 2 hours at 37 °C in the reaction described below:

Ethanol-precipitated MCP12 PCR product	17 µl
10 X reaction buffer	2 µl
*Eco*R1 (NEB) 20 $U\mu l^{-1}$	0.5 µl
*Not*1 (NEB) 20 $U\mu l^{-1}$	0.5 µl

Following the restriction digest, the reaction was heat-inactivated by incubating for 20 minutes at 65 °C. The digested DNA was then precipitated according to section 2.1.4, and redissolved in 10 µl of autoclaved milli-Q water. The purified digested fragment was then ligated into pPIC9 previously cut with *Eco*R1 and *Not*1 using a 1:1 molar ratio of insert to vector. The resulting construct was used to transform chemically competent *E. coli* Top10 (Invitrogen) according to the calcium chloride method (Sambrook et al., 1989). Positive transformants were selected on LB agar plates containing 100 $\mu g.ml^{-1}$ ampicillin. Ten colonies were picked, and each used to

inoculate 5 ml of LB containing 100 $\mu g.ml^{-1}$ ampicillin. The cultures were incubated overnight at 37 °C with shaking at 250 rpm. They were then "miniprepped", and 0.5 μg of each plasmid was digested with *Eco*R1 and *Not*1; and run on an agarose gel (1.4 %) to confirm the presence of the insert. Plasmids containing the insert were sequenced (dRhodamine) to confirm the identity and accuracy of the insert.

2.3.1.1. Transformation of *P. pastoris* KM71 with MCP12 pPIC9

The MCP12 pPIC9 clone was subjected to midiprep using the Promega "Wizard Plus Midiprep" kit according to the manufacturer's instructions, and 35 μg of the product was linearised with *Bgl*II (100 NEB units), in a total volume of 300 μl at 37 °C for 3 hours. Complete digestion was confirmed using agarose gel electrophoresis (1.2 %). The enzyme *Bgl*II cuts pPIC9 in two places, so the region containing the insert was purified away from the other fragment using agarose gel electrophoresis followed by gel extraction. The purified, linearised plasmid fragment containing the insert was precipitated using ethanol and redissolved to a final concentration of 1 $\mu g.\mu l^{-1}$ in autoclaved, deionised water. The *P. pastoris* KM71 cells were transformed with 5 μg of this linear plasmid fragment by electroporation, and transformants were selected on MD (minimal dextrose – see Appendix A) plates. The colonies were then screened using membrane immuno-screening to identify which clones were most likely to express the highest levels of recombinant protein. Twelve transformants that appeared to express at reasonable levels were purified by streaking on MD plates and these were then used for test inductions to estimate which one expressed the protein at the highest level. The identity of the expressed protein was

verified by immunoblot. N-terminal sequencing and mass spectrometry confirmed the integrity of the protein sequence.

2.3.1.2. Membrane immunoscreening of MCP12 pPIC9 KM71 clones

The following procedure was carried out in a laminar flow hood to reduce the chances of contamination. Once the colonies from a *P. pastoris* transformation were of sufficient size to be manipulated (approx 2-4 mm in diameter) each clone was tested for protein expression using an immuno-screening procedure. A "Hybond-C Extra" membrane disk (Amersham-Pharmacia Biotech) was overlaid on each plate. A standard plate spreader was then used to apply sufficient pressure to the reverse of the membrane to allow a small amount of each colony to be transferred onto the membrane. The membrane and master plate were then marked with a sterile hypodermic needle, in such a way that unambiguous alignment of the membrane and master plate was possible. The membrane was then removed from the master plate and applied reverse side down to a MG (minimal glycerol – see Appendix A) plate, ensuring that the membrane was in contact with the plate across its entire area. The plate and membrane were then incubated at 30 °C for 24 hours after which time protein expression was induced by transfering the membrane to a MM (minimal methanol - see Appendix A) again ensuring that the membrane's entire surface was in contact with the plate. The plate was incubated at 30 °C for 48 hours, and was fed daily by adding 0.5 ml of methanol to the lid of the plate.

2.3.1.3. Time Course optimisation of MCP12 Expression

A single colony was picked from a freshly streaked MD plate of MCP12 pPIC9 KM71 and used to inoculate 10 ml of BMG in a 50 ml sterile cell culture centrifuge tube. The culture was incubated overnight at 30 °C with shaking at 250 rpm. This culture was subsequently used to inoculate a further 100 ml of BMG in a 500 ml baffled conical flask, closed with a foam stopper. This culture was incubated for 2 days at 30 °C with shaking at 250 rpm, and was subsequently harvested by centrifugation at 1500 g, at room temperature using aseptic technique. The pellet was then resuspended in 50 ml of BMM, and incubated at 29 °C with shaking at 250 rpm for a total of 6 days. The culture was 'fed' daily with methanol by the addition of 1 % of the culture volume of HPLC grade methanol. Starting on day 2, 1.5 ml samples of the culture were taken, and the supernatant harvested by centrifugation at 12,000 × g, and subsequently stored at -20 °C until the end of the sixth day of induction. Once all five samples had been collected, the 1 ml aliquots of the supernatants were deglycosylated by the addition of 2000 U of $EndoH_f$ (NEB), and incubating at 37 °C for four hours. The samples were then precipitated using TCA (see section 2.2.1), and subjected to SDS-PAGE

2.3.2. Purification of recombinant MCP12

2.3.2.1. Initial purification of MCP12 on a Resource Iso column

Large scale growths of MCP12 were performed as described in section 2.2.5. The culture was harvested after induction by centrifugation (2500 x g, 10 mins at 4 °C).

Proteases inhibitors were added - PMSF (0.5 mM final concentration) and EDTA (5 mM final concentration), and microbial growth was inhibited with NaN_3 (0.05 % final concentration). The culture supernatant was concentrated using ultrafiltration, down to a final volume of 20 ml per litre of induced culture supernatant.

After it was concentrated the protein was exchanged into 100 mM Tris buffer, pH 8.3, on a gravity-fed PD10 gel filtration column (Amersham-Pharmacia Biotech), and an equal volume of 4 M $(NH_4)_2SO_4$ was added and the pH readjusted to 8.3, giving the final sample conditions of 50 mM Tris, 2 M $(NH_4)_2SO_4$, pH 8.3. This was then applied in 4 ml aliquots to a "Resource Iso" hydrophobic interaction chromatography column (Amersham-Pharmacia Biotech) that had been pre-equilibrated with 2 M $(NH_4)_2SO_4$, 50 mM Tris, pH 8.3, with a flow rate of 2 $ml.min^{-1}$. Unbound protein was washed from the column with five column volumes of starting buffer. The bound MCP12 was then eluted with a linear gradient from 2M $(NH_4)_2SO_4$ to 0 M $(NH_4)_2SO_4$ (both containing 50 mM Tris pH 8.3) over 20 column volumes. Samples of the fractions containing protein were exchanged into 50 mM sodium phosphate, pH 5.5, using 500 µl centrifugal concentrators with a molecular weight cut off of 5 kDa. These samples were then deglycosylated by the addition of 2 µl of $EndoH_f$, TCA-precipitated and subjected to SDS-PAGE. Fractions containing MCP12 (as judged by SDS-PAGE) were pooled, concentrated to a final volume of 5 ml using a 20 ml centrifugal concentrator with a 5 kDa molecular weight cut-off membrane and exchanged into 50 mM sodium phosphate, pH 5.5. PMSF was added to a final concentration of 1 mM and EDTA to a final concentration of 5 mM.

2.3.2.2. Purification of EndoHf deglycosylated MCP12

The protein obtained as described in section 2.3.2.1 was then deglycosylated with $EndoH_f$, as described in section 2.2.7. The deglycosylated MCP12 was subsequently purified using a ConA sepharose column (5 ml) to remove the cleaved sugars and any remaining glycoprotein as described in section 2.2.11. The protein was then exchanged into 0.1 M Tris, pH 8.5, using a PD10 column. The resulting partially purified MCP12 in 0.1 M Tris, pH 8.5, was then applied in 1 ml aliquots to a MonoQ HR5/5 column (Amersham-Pharmacia Biotech) that had been pre-equilibrated with the same buffer using a flow rate of 1 $ml.min^{-1}$, and the unbound protein was washed away with five column volumes of starting buffer. MCP12 was eluted using a linear gradient of 0 M to 1 M NaCl, in 0.1 M Tris pH 8.5 over 15 column volumes.

2.3.2.3. Purification of glycosylated MCP12

The MCP12 purified on the "Resource Iso" column described in section 2.3.2.1 was applied in 0.5 ml aliquots to a C4 reverse phase column, that had previously been equilibrated with 95 % (v/v) acetonitrile; 0.05 % (v/v) trifluoroacetic acid (TFA) in water with a flow rate of 1 $ml.min^{-1}$. Unbound salts and contaminants were washed away with 10 column volumes of the starting buffer. Bound glycosylated MCP12 was eluted with a linear gradient from 95 % acetonitrile, 0.05 % (v/v) TFA to 5 % acetonitrile, 0.05 % v/v TFA. Fractions containing glycosylated MCP12 (as judged by SDS-PAGE) were lyophilised and subsequently redissolved in 20 mM potassium phosphate pH 6.0.

2.4. Recombinant MCP2 from P. pastoris

2.4.1. Cloning MCP2 into pPICZαA

The region corresponding to module 2 of MCP (MCP2) was amplified by PCR using the same template DNA as that for MCP12 (MCP-BC2-pSG5). The PCR reaction was set up as described in section 2.1.1, and was cycled according to the following parameters:

Temperature	Time	No. Cycles
95 °C	1 min	1
95 °C	30 Sec	35
55 °C	30 Sec	
74 °C	30 Sec	
74 °C	10 min	1
4 °C	Hold	

The sequences of the primers were:

MCP2 5' EcoR1
5' CCGGAATTCTATAGAGAAACATGTCC 3'

MCP2 3' Not1
5' GCGGCGGCCGCCTATCACAAAACCTTTTCACATATTGG 3'

Two stop codons were incorporated to allow expression of the protein without the tags present in the vector. The restriction sites *Eco*RI and *Not*I were incorporated into the 5' and 3' primers respectively to facilitate directional cloning. The fragment obtained from PCR was subjected to agarose gel electrophoresis (1.4% gel), and was subsequently purified from the gel using the Qiagen gel extraction kit according to the manufacturer's instructions. The fragment was then precipitated with ethanol according to section 2.1.4, and was redissolved in 20 μl of autoclaved milli-Q water.

The fragment was then digested with *EcoR*1 and *Not*1, for 2 hours at 37 °C in the reaction described below:

Ethanol precipitated MCP2 PCR product	17 µl
10 × reaction buffer	2 µl
*Eco*R1 (NEB) 20 $U.\mu l^{-1}$	0.5 µl
*Not*1 (NEB) 20 $U.\mu l^{-1}$	0.5 µl

Following the restriction digest the reaction was heat-inactivated by incubating for 20 minutes at 65 °C. The digested DNA was then precipitated according to section 2.1.4, and redissolved in 10 µl of autoclaved milli-Q water. The purified digested fragment was then ligated into pPICZαA previously cut with *Eco*R1 and *Not*1 using a 1:1 molar ratio of insert to vector. The result was used to transform chemically competent OneShot *E. coli* Top10 cells according to the manufacturer's instructions. Positive transformants were selected on plates of low salt LB agar + 25 $\mu g.ml^{-1}$ Zeocin. Ten colonies were picked, and each used to inoculate 5 ml of low salt LB containing 25 $\mu g.ml^{-1}$ Zeocin for selection. The cultures were incubated overnight at 37 °C with shaking at 250 rpm. They were then subjected to miniprep, and 0.5 µg of each plasmid was digested with *Eco*R1 and *Not*1; and run on an agarose gel (1.4 %) to confirm the presence of the insert. Plasmids containing an insert of the predicted size for MCP2 were sequenced (Big Dye) to confirm the identity and accuracy of the insert.

2.4.1.1. Transformation of *P. pastoris* KM71His with MCP2 pPICZα

The MCP2 pPICZαA clone was obtained by maxiprep using the Sigma Genelute Maxiprep kit according to the manufacturer's instructions, and 50 μg of the product was linearised with *Sac*1 (200 NEB units), in a total volume of 300 μl at 37 °C for 3 hours. Complete digestion was confirmed using agarose gel electrophoresis (1.2 %). The purified linearised plasmid fragment containing the insert was phenol: chloroform extracted, and then precipitated with ethanol. The resulting pellet was redissolved in 10 μl of autoclaved deionised water. The *P. pastoris* KM71H cells were transformed using 5 μg of this linearised plasmid by electroporation (1.5 kV, 200 Ω, 25 μF) using a Bio-Rad Gene pulser. Transformants were selected on YPDS agar + 100 $\mu g.\mu l^{-1}$ zeocin plates. Five colonies from the initial transformation plate were purified by streaking on YPD agar + 100 $\mu g.\mu l^{-1}$ zeocin plates and these were then used for test inductions to estimate which expressed MCP2 at the highest level. N-terminal sequencing and mass spectrometry were used to confirm the integrity of the sequence.

2.4.1.2. Test inductions of recombinant MCP2 *P. pastoris* clones

Single MCP2 KM71H colonies were purified by streaking on YPD agar plates containing 100 $\mu g.\mu l^{-1}$ zeocin, and a single colony used to inoculate 10 ml of BMG in a 50 ml sterile tissue culture centrifuge tube. The culture was incubated for 24 hours at 30 °C with agitation at 225 rpm, and the cells harvested by centrifugation at 2000 g for 5 minutes; the supernatant decanted and discarded; and the pellet resuspended in 5 ml of BMM. The culture was then incubated as before for 4 days,

and was fed daily by the addition of HPLC grade methanol (0.5 % of culture volume). Aliquots of 1 ml were taken from each clone, deglycosylated with 1000 U of $EndoH_f$ and tested for expression of MCP2 using SDS-PAGE.

2.4.2. Purification of recombinant MCP2

2.4.2.1. Initial purification of MCP2 on a Resource Iso column

Large scale growths of MCP2 were performed as described in section 2.2.5. The culture was harvested after induction by centrifugation (2500 g, 10 mins at 4 °C). Proteases inhibitors were added - PMSF (0.5 mM final concentration) and EDTA (5 mM final concentration), and microbial growth was inhibited with NaN_3 (0.05 % final concentration). The culture supernatant was concentrated using ultrafiltration, down to a final volume of 20 ml per litre of induced culture supernatant.

After it was concentrated the protein was exchanged into 100 mM Tris buffer, pH 8.3, on a gravity-fed PD10 gel filtration column (Amersham-Pharmacia Biotech), and an equal volume of 4 M $(NH_4)_2SO_4$ was added and the pH readjusted to 8.0, giving the final sample conditions of 50 mM Tris, 2 M $(NH_4)_2SO_4$, pH 8.0. This was then applied in 4 ml aliquots to a Resource Iso hydrophobic interaction chromatography column that had been pre-equilibrated with 2 M $(NH_4)_2SO_4$,50 mM Tris, pH 8.0, with a flow rate of 2 $ml.min^{-1}$. Unbound protein was washed from the column with five column volumes of starting buffer. The bound MCP12 was then eluted with a linear gradient from 2 M $(NH_4)_2SO_4$ to 0 M $(NH_4)_2SO_4$ (both containing 50 mM Tris pH 8.0) over 20 column volumes. Samples of the fractions

containing protein were exchanged into 50 mM sodium phosphate, pH 5.5, using 500 µl centrifugal concentrators with a molecular weight cut off of 5 kDa. These samples were then deglycosylated by the addition of 2 µl of $EndoH_f$, TCA-precipitated and subjected to SDS-PAGE. Fractions containing MCP2 (as judged by SDS-PAGE) were pooled, concentrated to a final volume of 5 ml and exchanged into 50 mM sodium phosphate, pH 5.5, using a 20 ml centrifugal concentrator with a 5 kDa molecular weight cut off membrane. PMSF and EDTA were added to final concentrations of 0.5 mM and 5 mM respectively.

2.4.2.2. Purification of $EndoH_f$ deglycosylated MCP2

The protein obtained as described in section 2.4.2.1 was then deglycosylated with $EndoH_f$, as described in section 2.2.7. The deglycosylated MCP2 was subsequently purified using a ConA sepharose column (5 ml) to remove the cleaved sugars and any remaining glycoprotein as described in section 2.2.11. The protein was then exchanged into 0.1 M Tris, pH 8.5, using a PD10 column. The resulting partially purified MCP12 in 0.1 M Tris pH 8.5 was then applied in 1 ml aliquots to a Poros20Q column (Applied Biosystems) that had been pre-equilibrated with the same buffer using a flow rate of 5 $ml.min^{-1}$, and the unbound protein was washed away with five column volumes of starting buffer. MCP12 was eluted using a linear gradient of 0 M to 1 M NaCl in 0.1 M Tris, pH 8.5, over 20 column volumes.

2.4.2.3. Purification of glycosylated MCP2

The MCP2 purified on the Resource Iso column described in section 2.4.2.1 was applied in 0.5 ml aliquots to a C4 reverse phase column that had previously been equilibrated with 95 % acetonitrile, 0.05 % v/v trifluoroacetic acid (TFA) solution in water with a flow rate of 1 $ml.min^{-1}$. Unbound salts and contaminants were washed away with 10 column volumes of the starting buffer. Bound glycosylated MCP12 was eluted with a linear gradient from 95 % acetonitrile, 0.05 % v/v TFA to 5 % acetonitrile, 0.05 % v/v TFA. Fractions containing glycosylated MCP2 (as judged by SDS-PAGE) were lyophilised and subsequently redissolved in 20 mM potassium phosphate pH 6.0.

2.5. Recombinant MCP2 from organisms other than *P. pastoris*

2.5.1. Cloning and expression of MCP2 in E. coli

2.5.1.1. Cloning MCP2 into pET15b

The region of DNA representing module 2 of MCP was amplified by PCR (annealing temp 55 °C; elongation time 30 seconds), using the same template as that for MCP12 (MCP-BC2-pSG5).

The sequences of the primers were:

MCP2 5' Nde1
5' GGGAATTCCATATGTATAGAGAAACATGTCC 3'

MCP2 3' Xho1
5' CCGCTCGAGCGGCTATCACAAAACCTTTTCA 3'

The restriction sites *Nde*1 and *Xho*1 were incorporated into the 5' and 3' primers respectively to facilitate directional cloning. Stop codons were incorporated in the 3' primer to truncate the protein so that the minimum of cloning artefacts were expressed. The band of the predicted size for MCP2 was purified using agarose gel electrophoresis (1.4 %), followed by gel extraction. The band was then digested with *Nde*1 and *Xho*1, and was then repurified by agarose gel electrophoresis, followed by gel extraction. The digested fragment was then ligated into pET-15b using the 'Rapid Ligation Kit' (Roche), and used to transform chemically competent One Shot Top10 *E. coli* cells. The transformants were selected on LB Ampicillin plates. Colonies were taken for miniprep and the plasmids digested with *Nde*1 and *Xho*1 to confirm the presence of an insert of the correct size. Plasmids containing an insert of the predicted size for MCP2 were sequenced (Big Dye) to confirm the identity and sequence of the insert.

2.5.1.2. Cloning MCP2 into pET32

MCP2-pET15b was digested with *Nco*1 and *Xho*1, and the result run on agarose gel electrophoresis (1.2 %). The smaller fragment corresponding to the region of interest was purified form the gel, precipitated (see section 2.1.4) and the pellet redissolved in 10 µl of autoclaved Milli-Q water. This fragment was then ligated into pET32a (previously cut with *Nco*1 and *Xho*1) using the Roche Rapid Ligation kit according to the manufacturer's instructions, and the result used to transform chemically competent OneShot Top10 *E. coli* cells. Transformants were selected on LB

Ampicillin plates. Positive clones were selected and the plasmid retrieved by maxiprep using the Sigma Gene Elute Maxiprep kit according to the manufacturer's instructions. This plasmid DNA was subsequently further purified using phenol: chloroform extraction and ethanol precipitation.

2.5.2. Cloning and Expression of MCP2 in Drosophila S2 Cells

2.5.3. Growth and Maintenance of S2 Cells

2.5.3.1. General maintenance

Drosophila S2 cells were cultured in Invitrogen's Complete Drosophila Expression Medium (DES)™ containing 10 % heat inactivated fœtal bovine serum; or in Ultimate Insect™ serum free medium. The cells were resuspended before seeding a fresh culture. The S2 cells were grown in adherent culture and flasks were seeded to a density of 2×10^6 cells ml^{-1} in 10 ml of fresh medium in new T75 flasks with vent closures. The cells were incubated at 28 °C until a cell density of 1×10^7 cells ml^{-1} was achieved, at which point the cells were split to a density of 2×10^6 cells ml^{-1} with fresh media, and T75 plates. The general health of the culture was followed using microscopic examination, and trypan blue cell viability assay.

2.5.4. Cloning of MCP2 into pMT/BiP/V5-His-A

The MCP2 gene fragment containing a N-terminal His tag and thrombin cleavage sequence previously cut from the MCP2 pET15b construct using *Nco*1 and *Xho*1 (see section 2.5.1.2). The insert was ligated into pMT/BiP/V5-His-A previously cut with *Nco*1 and *Xho*1, using the Roche Rapid Ligation Kit according to the manufacturer's instructions. The result was used to transform chemically competent OneShot Top10 *E. coli* cells according to the manufacturer's instructions and positive transformants selected on LB agar plates containing 100 $\mu g.ml^{-1}$ ampicillin.

2.5.4.1. Transfection of S2 cells with pMT/BiP/V5-His-A-MCP2

Plates (35 mm) were seeded with 3×10^6 S2 cells in 3 ml of complete DES. The cells were incubated for 16 hours at 28 °C, in a humidified plate incubator, by which time the cells had reached a cell density of 3.4×10^6. The cells were transfected using calcium phosphate DNA precipitation with calcium chloride and Hepes-buffered saline. Transfection by this method is thought to be mediated by endocytosis of the DNA/calcium phosphate precipitate. The DNA used for transfection was a 20:1 mixture of pMT/BiP/V5-His-A-MCP2, and the selection vector was pCoHygro. The cells were incubated in the presence of the calcium phosphate – DNA solution for 16 hours, after which time the cells were washed twice with complete DES, and incubated for a further 2 days at 28 °C in the same plate. The cells were then resupended in complete DES containing 300 $\mu g.ml^{-1}$ hygromycin, and the cells incubated for 3 weeks until resistant cells started to grow out.

2.5.4.2. Induction of expression of MCP2 in S2 cells

Once a stable cell line had been established, protein expression was induced by the addition of copper sulphate to a final concentration of 500 μM to the growth medium. The stable cell line was adapted to use serum free medium, by passaging twice through serum free medium, in T25 flasks. Following adaptation to serum free medium, a T25 flask was seeded with 2×10^6 cells. The cells were incubated for 16 hours at 28 °C, followed by the addition of 100 mM copper sulphate to a final concentration of 500 μM. The cells were then incubated for a further 24 hours before assaying the culture medium for the presence of MCP2.

2.6. Cloning and Expression of SLAM

2.6.1. Cloning of SLAM into pGEM-T

The cDNA of SLAM (CDw150) was amplified using a modification of touchdown PCR using 2.5 U of 'Herculase HotStart' DNA polymerase (Stratagene). The Quick-Clone universal human cDNA library (Clontech) was used as a template (8 ng); 100 ng of each primer; a final concentration of 0.2 mM of each dNTP; and 5% (v/v) DMSO.

The sequences of the primers were:

5' primer
5' CTTTTGGGGCAAGCTACGGAACAGGTG 3'

3' primer
5' CACTAGCATAGACTGTGATGGAATTTGTTTC 3'

Neither primer had a restriction sites built in to reduce the possibility of forming unwanted side products.

The reaction was cycled as follows

Temp	Time	No. Cycles
98 °C	3 min	1 Cycle
98 °C	40 s	20 Cycles
60 °C	30 s	
74 °C	1 min	
98 °C	40 S	20 Cycles
55 °C	30 S	
74 °C	1 min 30 s	
98°C	40 s	10 Cycles
52 °C	30 s	
74 °C	1 min 50 s	
74 °C	10 min	1 Cycle
4 °C	hold	

The PCR product was purified using agarose gel electrophoresis (1.4 %) followed by gel extraction. They were A-tailed and ligated into pGEM-T (Promega) according to the manufacturer's instructions. Transformants were selected on LB-amp containing X-gal, and white colonies that potentially contain the SLAM gene were subjected to miniprep, and digested with *Nco*1 and *Nde*1 that are present in the flanking sequences of the vector. Plasmids containing an insert of the predicted length for SLAM were sequenced (Big-Dye) to confirm both the identity and sequence of the insert.

2.6.2. Cloning of SLAM1 into pPICZαA

The N-terminal domain of SLAM was amplified by PCR using SLAM-pGEM-T as a template (annealing temperature 60 °C; and extension time 45 seconds).

The sequences of the primers were:

5' primer:
5' CCGGAATTCGCAAGCTACGGAAC 3'

3' primer:
5' TTTTCCTTTTGCGGCCGCCTACTGGGTCTTG 3'

The restriction sites *Eco*R1 and *Not*1 were incorporated into the 5' and 3' ends of the PCR product respectively, to facilitate directional cloning. A stop codon was incorporated at the 3' end of the gene to allow expression of the N-terminal domain without the purification tags present on the expression vector. Following amplification the PCR product was purified using agarose gel electrophoresis and gel extraction. The purified PCR product (0.5 μg) was digested using *Eco*R1 (10 NEB units) and *Not*1 (5 NEB units), and was then repurified using agarose gel electrophoresis followed by gel extraction. The digested PCR product (50 ng) was ligated using the rapid ligation kit (Roche), into pPICZαA (25 ng) pre-cut with *Eco*R1 and *Not*1. The reaction was used to transform One Shot Top10 *E. coli* and transformants selected on LB-Zeocin plates. Ten colonies were subjected to miniprep and the plasmid digested with *Eco*R1 and *Not*1 to test for the presence of the insert. Plasmids containing an insert, of the predicted size for the N-terminal domain of SLAM, were sequenced (Big Dye) to confirm the identity and sequence of the insert.

2.6.3. Transformation of *P. pastoris* X33 with SLAM1 pPICZαA

The SLAM1 pPICZαA clone was "maxipreped" using the GenElute Maxiprep kit (Sigma) according to the manufacturer's instructions, and 15 μg of this was linearised with *Sac*1 (100 NEB units), in a total volume of 300 μl at 37 °C for 3 hours. Complete linearisation was checked using agarose gel electrophoresis (1.2 %). The linearised plasmid was purified using phenol: chloroform extraction followed by ethanol precipitation. The X33 cells were transformed with 5 μg of linearised plasmid by electroporation and transformants were selected on YPDS Zeocin plates. Twelve transformants were purified by streaking on YPD Zeocin plates and these were then used for test inductions to estimate which one expressed the protein at the highest level.

2.6.4. Optimising the expression conditions for SLAM1

The SLAM1 X33 clone that appeared to express the highest levels of protein was used for expression optimisation. A single colony was picked from a freshly streaked YPDZ plate and used to inoculate 10 ml of BMG. This culture was incubated over night at 30 °C with shaking at 250 RPM and used to inoculate 50 ml of BMG. This culture was incubated over night as before and used to inoculate 500 ml of BMG. This culture was incubated overnight as before until it reached OD_{600} of 21. The cells were harvested by centrifugation at 1500 g for 5 minutes at room temperature and resuspended in 1 litre of BMM (0.5 % (v/v) methanol). The culture was then split into 4 × 250 ml, each had an OD_{600} of 13. The cultures were fed with

methanol daily to 0.5 %; 1 %; 2 % or 4 %. The absorbance at 600 nm was recorded each day at the time of feeding, and a sample was taken for analysis of protein expression. Each sample was deglycosylated with 100 NEB units of $EndoH_f$ prior to SDS-PAGE.

The identity of the expressed protein was confirmed using immunoblotting, employing a goat polyclonal IgG raised against amino acid residues 33-51 of human SLAM (*sc-1334* - Autogen-Bioclear) as the primary (1:500 dilution). A rabbit anti-goat IgG peroxidase conjugate (A-5420 Sigma) was used as a secondary antibody (1:5000 dilution). The membrane was stained for 5 minutes with the precipitating peroxidase substrate DAB with nickel enhancer (Sigma).

The expressed protein was tested for the presence of N-glycans using a band shift type assay. Two 1 ml samples of culture supernatant were taken following induction, and one was deglycosylated with 1000 NEB units of EndoHf at 37 °C for 2 hours, whilst the other was simply incubated at 37 °C for 2 hours. The samples were then precipitated with TCA and subjected to SDS-PAGE, and stained with a colloidal coomassie blue stain (Sigma).

3. Expression, Purification, and characterisation of MCP12

3.1. Preliminary remarks

The importance of the first two modules of MCP (MCP12) as a receptor for measles virus, and potentially other pathogens, has been demonstrated, and to date the whole of this protein has been characterised in reasonable detail. A greater understanding of MCP12 may prove useful for fully determining its interaction with measles virus and other potentially pathogens. We were therefore interested in characterising this region in greater detail, with the intention of understanding the mechanism by which these pathogens use MCP to gain entry into cells.

Since this region of the protein is only a fragment of the whole it cannot be obtained by purification from a natural source, instead some form of artificial method will be required to produce this fragment. Limited proteolysis can be used to produce domain fragments; however, this approach has problems with consistency and purity. Another possible method for producing domain fragments is chemical synthesis, and there has been some success using this method for producing some relatively small proteins (Bersch et al., 1998), but the yield rapidly falls off above 50 - 60 amino acids. This technique is further hampered by the requirement to artificially fold the protein following synthesis. An additional disadvantage is the expense of isotopic labelling. A promising technique based on inteins for splicing together smaller fragments is still not widely proven (Muir et al., 1998; Xu et al., 1999). At the present time, the most successful means of producing the quantities of larger or more complex proteins required for structural and biophysical characterisation is to

employ a heterologous gene expression system. The work described in the following section describes the successful expression and purification of the MCP12 fragment in the methylotrophic yeast *P. pastoris*.

3.2. Cloning and Expression of MCP12

The ease with which the expression of foreign genes in *E. coli* can be achieved has markedly increased over the last decade, and *E. coli* is now the expression host of choice for producing a wide variety of foreign genes from a number of different organisms. Amongst the major advantages of using *E. coli* as an expression host are: the general level of understanding of *E. coli* genetics; the ease of performing genetic manipulations; the high levels expression that are often achieved; and the speed with which initial protein samples can be obtained (Baneyx, 1999). There are, however, a number of serious limitations associated with using *E. coli* as an expression host. Being a prokaryote, *E. coli* has notably different cellular machinery to higher eukaryotes, with the result that *E. coli* is unable to perform many of the processes involved in the folding of multi-modular proteins (Georgiou et al., 1996), although efforts have been made to address this issue by genetically modifying *E. coli* with various chaperones etc (Thomas et al., 1997). In addition to this *E. coli* is unable to perform many of the necessary post-translational modifications that are common in eukaryotic proteins. In the case of MCP the most important of these are disulphide bond formation and N-glycosylation, both of which have been shown to be necessary for MCP to be fully active (Liszewski et al., 1998; Maisner et al., 1994). The use of *E. coli* in this case would therefore probably require the use of a refolding protocol,

and any N-linked glycans would need to be added chemically at a later stage. Instead it would seem a natural choice to use one of the eukaryotic expression systems as a starting point for expressing fragments of MCP.

There are several available eukaryotic expression systems, each with its own advantages and disadvantages. They are generally divided into two main classes, namely higher eukaryotic and lower eukaryotic expression systems. The former typically employ the culture of immortalised cell lines from various higher eukaryotes including mammals, and examples of such cell lines include Chinese hamster ovary (CHO) cells. In mammalian expression systems, the gene to be expressed is transferred to the cell line of interest using such methods as viral infection (e.g. adenovirus or lentivirus), electroporation (Shigekawa et al., 1988) lipid mediated transfection (Felgner et al., 1989) or calcium phosphate transfection (Chen et al., 1987). Since mammalian systems are derived from higher eukaryotes and are the closest available expression host to man, they are able to produce correctly folded and glycosylated proteins. Although even with mammalian expression systems, the glycosylation profile of the recombinant protein may not fully reflect its natural state. These expression systems have, therefore, been particularly useful for the expression of proteins such as membrane proteins that proved difficult to express in other systems.

There are, however, several limitations of the mammalian expression systems. The most important of these are: low yields of protein; the complex media required for growth and maintenance of the cells; and the large period of time required between

cloning the gene and obtaining the initial protein samples. Biophysical and structural studies usually require tens of milligrams of highly purified and homogeneous protein. Therefore, the tens of micrograms per litre yields that are often associated with mammalian expression systems would prove unsuitable. As such these expression systems are of primary use in such cases, only once it has been established that the protein cannot be obtained using another system. The complex media requirements of mammalian cell culture makes the isotopic enrichment required by one of the most powerful biophysical techniques, high field NMR spectroscopy, unattractive from a cost point of view (~ £2,000 per litre for ^{15}N CHO cell growth medium, at the time of writing). Bearing these points in mind it seems logical to initially attempt to use a non-mammalian eukaryotic expression system for the expression of MCP fragments.

Insect cells provide another route to higher eukaryotic expression of proteins. These expression systems employ similar cell culture techniques to the mammalian expression systems already discussed. The cell lines are derived from species such as *Spodoptera frugiperda* (Sf9, and Sf21 cells) *Trichoplusia ni* (High Five™) and *Drosophila melanogaster* (S2 cells). The insect expression systems have similar capabilities to their mammalian counterparts as far as post-translational modifications are concerned. The major difference in protein processing between the two expression systems lies in the glycosylation profiles of the resultant glycoproteins. Insect cells are generally unable to produce the complex glycans that are observed in proteins expressed in mammalian systems (Altmann et al., 1999). Insect expression systems do, however, have a number of advantages over

mammalian expression systems. These advantages include: less stringent media requirements and the ability to easily adapt to growth on serum-free medium; generally higher expression levels - 8-10 $\mu g.ml^{-1}$ for human melantoransferrin (Hegedus et al., 1999); and reduced equipment requirements – generally a CO2 incubator is not required.

Of the other available eukaryotic expression systems in common use, yeast has many advantages when it comes to expressing complex proteins for biophysical studies. There are two yeast expression systems currently in common usage – *Saccharomyces cerevisiae* and *Pichia pastoris*. Both of which are able to form multimodular and disulphide bonded proteins (White et al., 1994), and they can perform rudimentary glycosylation, (high mannose-type glycosylation patterns rather than the complex or hybrid patterns often found in higher eukaryotes). One potential advantage of using *P. pastoris* over *S. cerevisiae* for the expression of MCP12 is due to the phenomenon of hyperglycosylation. Hyperglycosylation is the result of the organism adding excessive numbers of mannose residues to the core sugars during glycan maturation in the Golgi apparatus. This phenomenon is frequently observed in proteins expressed in *S. cerevisiae*; however its occurrence in *P. pastoris* is considerably less common (Lundblad, 1999). Expression in *S. cerevisiae*, of CCP module pairs and individual CCP modules from factor H, for NMR structural studies, has been reported although the yields are typically quite low. In the case of factor H module 16 the yield of purified protein was only around 0.1 $mg.l^{-1}$ (Barlow et al., 1991). The other commonly used yeast expression system - *P. pastoris* is particularly powerful (for a review see Cereghino et al., 2000). In addition to its ability to produce

properly processed complex proteins, the levels of foreign protein that have been achieved are often very high, and levels of over 2.5 $g.l^{-1}$ of secreted protein have been reported (Paifer et al., 1994). These expression levels combined with *P. pastoris*'s ability to grow on minimal medium, permits isotopic labelling of recombinant proteins without the prohibitive costs associated with the higher eukaryotic systems. Moreover, *P. pastoris* previously has been used successfully to express and isotopically label several different complement control protein fragments for biophysical and structural investigations (Guthridge et al., 2001; Henderson et al., 2001; Smith et al., 2002; Wiles et al., 1997). It is logical therefore to employ *P. pastoris* as the initial expression system for the production of the N-terminal fragments of MCP in the current study. The following section describes the successful cloning and expression of the two N-terminal modules of MCP (MCP12) in *P. pastoris*.

3.2.1. Cloning and expression of MCP12 in *P. pastoris*

The *P. pastoris* expression vector pPIC9 was chosen for the initial expression of MCP12, because it is an inducible expression vector that targets the recombinant protein to the secretory pathway, and subsequently into the culture medium. Secretion of recombinant proteins offers several advantages over intracellular expression. Secretion of proteins facilitates both disulphide bond formation and N-linked glycosylation, both of which are known to be important in this instance. In addition, secretion is generally associated with less troublesome purification

protocols. This is especially true in the case of *P. pastoris* which is known to naturally secrete only very low levels of protein (Higgins, 1998). The pPIC9 vector (see Figure 3:1 for pPIC9 plasmid map) is a shuttle vector that uses ampicillin resistance for selection in *E. coli*, and the histidine auxotrophic marker for selection in *P. pastoris*. The gene of interest is cloned behind the powerful inducible promoter from the native *P. pastoris* alcohol oxidase gene AOX1, and induction of protein expression is subsequently achieved by switching the carbon source from the glycerol used during the growth phase to methanol. The protein was targeted for secretion by cloning the gene of interest behind the secretion signal from the α-mating factor from *S. cerevisiae*. The α-factor secretion signal was chosen in over the native *P. pastoris* PH01 secretion signal, since generally more success has been achieved using the α-factor secretion signal (Laroche et al., 1994). This secretion signal contains the KEX2 cleavage, followed by two EA repeats. Cleavage of the secretion signal is a two step process (Achstetter, 1989); the initial cleavage is by the KEX2 gene product; which cleaves on the carboxyl side of R in the secretion signal sequence EKR↓EAEA. The second step is cleavage by the STE13 gene product of the two EA repeats. STE13 is a membrane bound dipeptidase that cleaves on the carboxyl side of A in the sequence NH-XA (Julius et al., 1983).

This type of cloning strategy inevitably leaves a number of extra amino acid residues on the expressed protein as a result of cloning artefacts. In this case efficient cleavage of the secretion signal by both KEX2 and STE13 is predicted to leave extra YVE residues on the protein's N-terminus. This would therefore result in the predicted sequence of the N-terminus of the protein, before the first cysteine of the

consensus CCP module sequence, being YVEFSDA. However is has been reported that STE13 processing of the secretion signal does not always occur efficiently for heterologous proteins in *P. pastoris* (Goda et al., 2000); in which cases the protein is secreted with extra residues on its N-terminus (either EA or EAEA depending on the efficiency of STE13 cleavage).

Comments for pPIC9:
8023 nucleotides

5' AOX1 promoter fragment: bases 1-948
5' AOX1 primer site: bases 855-875
a-Factor secretion signal (S): bases 949-1218
a-Factor primer site: bases 1152-1172
Multiple Cloning Site Region: bases 1192-1241
3' AOX1 primer site: bases 1327-1347
3' AOX1 transcription termination (TT) fragment: bases 1253-1586
HIS4 ORF: bases 4514-1980
3' AOX1 fragment: bases 4870-5626
ColE1 origin: bases 6708-6034
Ampicillin resistance gene: bases 7713-6853

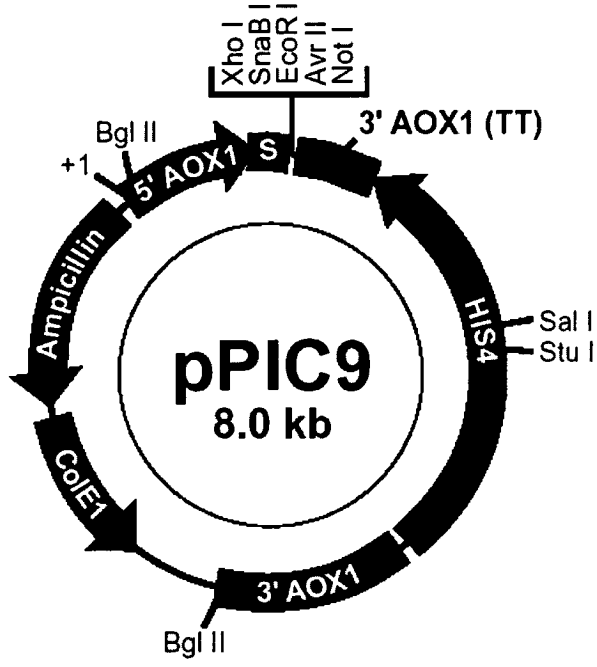
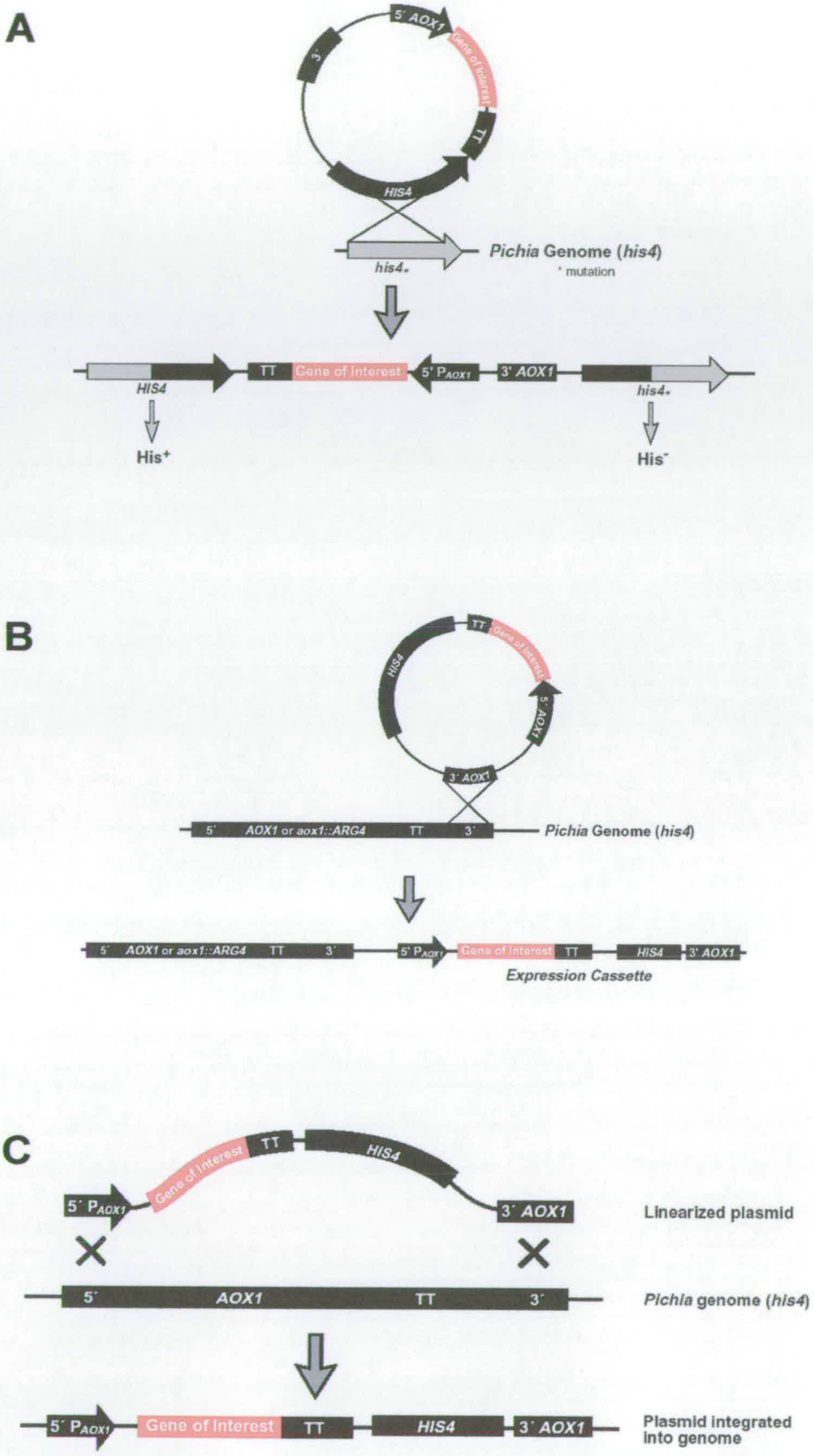


Figure 3:1 A plasmid map of pPIC9 showing the key features of the AOX1 promoter, *α*-factor secretion signal, ampicillin resistance, and HIS4 gene to provide the histidine auxotrophic marker for selection in *P. pastoris* (taken from Invitrogen, 2002a).

The transformation of *P. pastoris* involves the integration of linearised recombinant vector into the host genome by the process of homologous recombination. The vector is linearised in a region that contains homologous sequences to ones found in the host genome (Cregg et al., 1985). The presence of several homologous regions in the vector permits a choice of methods by which the recombinant plasmid is integrated into the host genome. This has important implications for the eventual

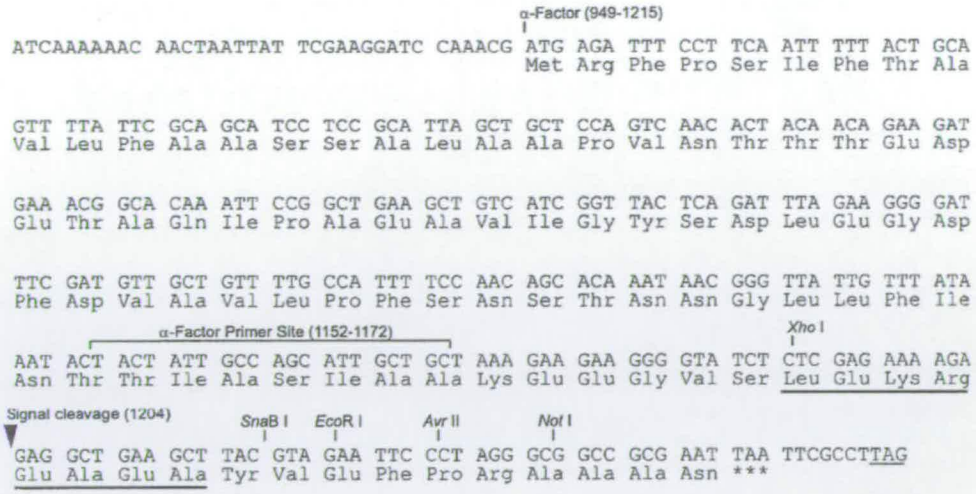
phenotype of the recombinant *P. pastoris*, which is dependant on both the region in which the plasmid is linearised and the genotype of the original host strain. Aside from the His phenotype used to select recombinants, there is another important phenotype to consider when generating recombinant *P. pastoris*. *P. pastoris* has two alcohol oxidase genes, AOX1 and AOX2, and of these the AOX1 gene contributes the majority of the alcohol oxidase activity in the cell (Cregg et al., 1989). This occurs despite the 97% sequence identity between the AOX1 and AOX2 genes. Yeasts that contain wild type AOX1 and AOX2 have wild type methanol utilisation (Mut^+), and those that have a disrupted AOX1 gene exhibit slow methanol utilisation (Mut^S). In this way, disruption of the AOX1 gene leaves the yeast with a working copy of the AOX2 gene only, which results in poor growth characteristics on methanol, whilst the growth rate on other carbon sources remains unaffected. The different ways in which the recombinant plasmid can integrate into the host genome are illustrated in Figure 3:2

Figure 3:2 – Following page. The generation of recombinant *P. pastoris* clones using pPIC9. A: Linearisation of the plasmid in the HIS4 region causes insertion at the His4 locus of the host chromosome, and results in a His^+ phenotype for all strains and a Mut phenotype that depends on the host strain (Mut^+ GS115 and Mut^S KM71). B: linearisation of the plasmid in the AOX1 region (either 5'AOX1, 3'AOX1 or AOX1 transcription termination regions), causes insertion in the corresponding AOX1 region of the host chromosome. This results in a similar phenotype to that for insertion at His4 – His^+ all strains, and Mut^+ GS115 and Mut^S KM71. C: linearisation of the plasmid in both 5' and 3' AOX1 regions causes gene replacement of the entire AOX1 coding region. This results in a His^+ Mut^S phenotype and should only be performed in the GS115 strain (taken from Invitrogen, 2002a).


A
5´ AOX1
3´
Gene of Interest
TT
HIS4
his4*
Pichia Genome (his4)
* mutation
HIS4
TT
Gene of Interest
5´ PAOX1
3´ AOX1
his4*
His+
His-
B
HIS4
TT
Gene of Interest
5´ AOX1
3´ AOX1
5´
AOX1 or aox1::ARG4
TT
3´
Pichia Genome (his4)
5´
AOX1 or aox1::ARG4
TT
3´
5´ PAOX1
Gene of Interest
TT
HIS4
3´ AOX1
Expression Cassette
C
5´ PAOX1
Gene of Interest
TT
HIS4
3´ AOX1
Linearized plasmid
5´
AOX1
TT
3´
Pichia genome (his4)
5´ PAOX1
Gene of Interest
TT
HIS4
3´ AOX1
Plasmid integrated into genome

Polymerase chain reaction (PCR) was used to amplify the region of MCP encoding the N-terminal two modules (MCP12) and asymmetric restriction sites were incorporated into the primers (5' EcoR1 and 3' Not1) so that directional cloning was possible. In this way any positive transformants should contain the inserted gene fragment in the correct orientation. The multicloning region of pPIC9 is shown alongside the final construct in Figure 3:3. The PCR product, shown in Figure 3:4, gave a single band of approximately 500 bp which is slightly larger than one would expect for MCP12. This DNA fragment was then ligated into pPIC9 and used to transform *E. coli* Top10 cells. The transformation yielded a large number of transformants, ten of which were worked up by 'miniprep' and tested by restriction digest (EcoR1 and Not1) for the presence of the required DNA fragment. Clones containing inserts of the correct size for MCP12 were then sequenced to confirm their identity. A clone containing an insert of the predicted sequence for MCP12 was subsequently employed to prepare the DNA used to transform the *P. pastoris* strain KM71. This *P. pastoris* strain was transformed by gene insertion in the AOX1 region, and was achieved by linearising the plasmid with Pme1, which cuts the plasmid in the 3' AOX1 region. This strategy should yield *P. pastoris* clones with a His^+ Mut^S phenotype.

A



B EKR↓EA↓EA↓YVE FSDAC....

Figure 3:3 A: The cloning region of pPIC9 showing: the α-factor secretion signal with the KEX2 and STE13 cleavage sites, and the EcoR1 and Not1 restriction sites used for inserting the MCP12 gene fragment. **B: The pPIC9-MCP12 construct in the region of the secretion signal showing:** the KEX2 and STE13 cleavage sites (red); the artefactual cloning residues (purple) and the start of the MCP12 sequence (blue).

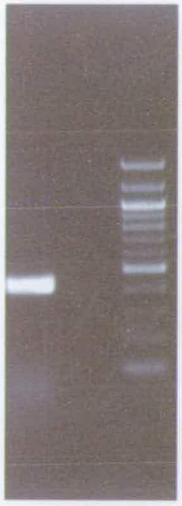


Figure 3:4 PCR amplification of the MCP12 fragment. The gel shows a single band that appeared to be slightly larger than the predicted size for MCP12 containing the EcoR1 and Not1 restriction sites, stop codon, and extra bases added to facilitate efficient restriction digests.

3.2.2. Immuno-screening of recombinant *P. pastoris* clones

The transformation yielded a large number of transformants, and since there is often significant clonal variation amongst recombinant *P. pastoris* clones, expression levels between individual colonies may vary significantly. Moreover, some clones may produce unusually high expression levels, and these are often associated with multi-copy integration events (Clare et al., 1991). These multi copy clones, sometimes referred to as 'jackpot clones', occur spontaneously but are relatively rare, accounting for only 1% of the total number of His^+ transformants. It is therefore desirable to screen a large number of colonies for expression following a *P. pastoris*. One way of doing this is to carry out a large number of 'test inductions', and test each one for protein expression. It is often impractical, however, to perform such an analysis on each of the hundreds of clones generated by a typical *P. pastoris* transformation. This is particularly true for proteins such as MCP12 that cannot easily be assayed. It is therefore useful to find another method of screening. Antibodies to the protein of interest can prove useful in this respect. In this case the recombinant protein should be secreted, therefore if the clones are induced whilst growing on a membrane with a high protein-binding capacity, the protein should become immobilised on the membrane. It should then be a relatively straightforward task to detect the immobilised protein using standard immunostaining techniques.

A colony lift was performed from the initial selective minimal dextrose (MD) plates from the transformation using a membrane with high protein binding capacity. These membranes were transferred to minimal glycerol (MG) plates and incubated for 24 hours at 30 °C. The membranes were subsequently transferred to minimal

methanol (MM) plates, and incubated for a further 48 hours at 30 °C. Following the first 24 hours, the membrane colonies were 'fed' by adding 500 µl of methanol onto the lid of the MM plates. The membranes were subsequently stained as described in section 2.3.1.2. The results of these membrane screens are shown in Figure 3:5 and appear to show, that certain colonies express greater levels of MCP12 than others; the clone indicated in Figure 3:5 was used for further analysis. These results need to be interpreted with caution however, since it is difficult to correct for the size of the colony. They do provide several important pieces of information however, not least of which is that this method of colony screening provides a reasonable assurance that the protein that is subsequently observed is MCP12, since antibodies against the N-terminus of MCP were used for detection.

Figure 3:5 Immuno-screening of the MCP12 pPIC9 KM71 membranes. The membrane shows that most clones express at only a relatively low level and colonies are hard to distinguish from the background. There are a few clones, however that appear to give much stronger signals. The clone marked by the arrow was the one chosen for further analysis.

It had previously been discovered that the N-terminal module of MCP (MCP1) became hyperglycosylated when expressed in *P. pastoris* (O'Leary, 2000), and that these glycans were sensitive to $EndoH_f$. $EndoH_f$ is an endoglycosidase that cleaves

high mannose type N-glycans between the first and second GlcNAc residues (Robbins et al., 1984). That the glycans of MCP1 were sensitive to $EndoH_f$, should come as no surprise given that the glycosylation patterns found in *P. pastoris* are reported to be of the high mannose type (see schematic in Figure 1.12), and this type of glycans are amenable to cleavage using EndoH (Miele et al., 1997). In order to determine whether the MCP12 expressed by *P. pastoris* was also hyperglycosylated; the clone indicated in Figure 3:5, was grown on a test scale (10 ml) and 1 ml samples subjected to SDS-PAGE, both before and after treatment with $EndoH_f$. This SDS-PAGE result, shown in Figure 3:6, shows a smear for the untreated sample that is considerably larger than what would be expected for MCP12 - 28-35 kDa rather than the predicted 16 kDa for non-glycosylated MCP12. Moreover, following enzymatic deglycosylation with $EndoH_f$ the smear was replaced by a single band of the expected molecular weight for MCP12. These data are similar to those obtained with MCP1 (O'Leary, 2000), and provide compelling evidence that the MCP12 species expressed by *P. pastoris* is subject to hyperglycosylation, and that the attached glycans are of the high mannose type. This hyperglycosylation may potentially cause problems with subsequent experiments, by virtue of the extreme heterogeneity of the glycoforms of the protein, and also in the increased size molecular weight of the species.

Figure 3:6 SDS-PAGE showing the effect of EndoH$_f$ treatment on the MCP12 fragment expressed by *P. pastoris*. The EndoH$_f$ treated sample (dg MCP12) shows a discrete band of around 16 kDa, which is close to what would be expected. The untreated sample (gMCP12), on the other hand, shows a smear centred at approximately 30kDa.

3.2.3. Optimisation of Expression of MCP12

There are several conditions to optimise with any new recombinant *P. pastoris* expression attempt, including cell density at induction, aeration, pH, and temperature. The majority of these factors merely contribute to maintaining the culture in its optimal growth conditions. These factors are therefore typically strain specific rather than protein specific. There is, however, one main parameter that is worth optimising for each protein construct, namely length of induction. The aim of adjusting the induction time is to find the optimum balance between the amount of protein produced and modification or degradation. To address this issue, SDS-PAGE was performed on deglycosylated 1 ml samples taken from a culture of MCP12 at

days 2 – 6 post, and this is shown in Figure 3:7. The gel shows that yield increases with increasing induction time, but by days five and six, degradation appears to become significant. It is slightly difficult to interpret such a gel, however, since the method requires that the protein is deglycosylated before running the SDS PAGE. Since the deglycosylation reaction takes place at 37°C over a period of four hours, it is feasible that the observed degradation occurs during the deglycosylation reaction, despite the addition of both PMSF and EDTA as protease inhibitors. Nevertheless, the optimal yield to degradation ratio appears to occur on day five, and therefore for subsequent larger scale growths the culture was induced for five days.

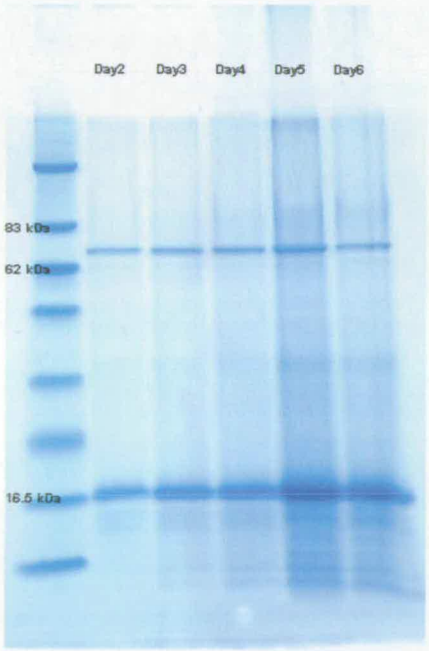


Figure 3:7 Induction time optimisation of MCP12. The SDS PAGE of deglycosylated 1ml samples shows a gradual increase in MCP12 production (band near 16.5 kDa marker) over the six days. By day six however degradation appears to become more of a problem. The band between the 83 kDa and 62 kDa markers is around the predicted size for $EndoH_f$.

Despite the use of antibodies, that had been raised against synthetic peptide corresponding to residues in the N-terminal domain of MCP12, for the membrane screens. Further confirmation was sought that the species observed on the gel was indeed MCP12. This was subsequently achieved through the use of N-terminal protein sequencing. The sequence of the N-terminal ten residues was thus determined to be: YVEFSDACEE, which correspond exactly to the expected sequence for MCP12 preceded by the predicted cloning artefacts. Furthermore, these N-terminal sequencing data indicate both that there was complete processing of the α-factor secretion signal, and that there was no significant proteolytic degradation at the N-terminal end of the protein.

3.3. Purification of MCP12

3.3.1. Initial purification of MCP12

The purification of glycoproteins can often be problematic due to the heterogeneity of the species to be purified. One solution to this problem is to enzymatically cleave the N-linked glycans; however too can prove problematic. The deglycosylation reaction requires prolonged incubation at 37 °C, which as discussed potentially allows any proteases present to degrade the sample. This can be alleviated to some extent by the inclusion of protease inhibitors in the sample; however protease inhibitors are rarely completely effective. In this case is seems logical to try an initial purification step to remove most of the contaminants present in the crude concentrate prior to deglycosylation. For this reason the initial purification step for

all MCP12 preparations was the same, regardless of the glycosylation state required for the final sample. The initial purification was achieved using hydrophobic interaction chromatography (HIC) on a Resource Iso column (Amersham Pharmacia Biotech).

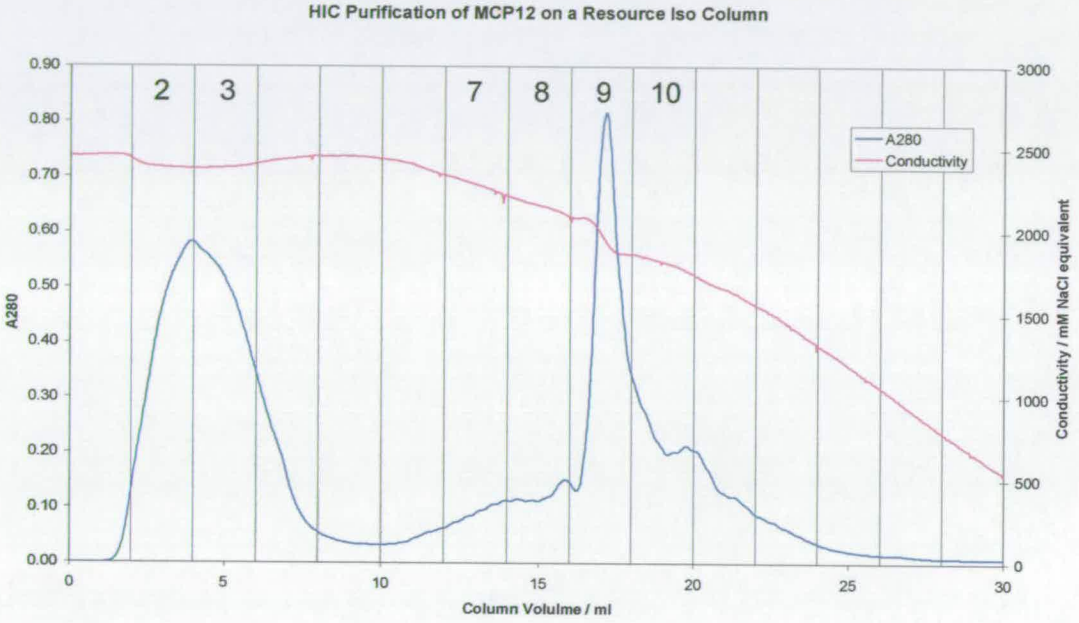


Figure 3:8 Result of HIC purification of crude glycosylated MCP12 on a Resource Iso column. All the fractions that stuck to the column appear to contain MCP12 but with differing glycosylation profiles. These fractions should theoretically all behave equally after deglycosylation. Only the protein from fraction 9 was pooled for further analysis since it is possible that the different behaviour of the neighbouring fractions on HIC may be indicative of unfolded protein.

The chromatogram, shown in Figure 3:8 shows a rather broad main peak that is due to differences in hydrophobicity between the different glycoforms. Examination of SDS-PAGE performed on column fractions both before and after deglycosylation is shown in Figure 3:9 and Figure 3:10 respectively. These SDS-PAGE results, when taken in combination with the A_{280} trace, show that this purification method appears

to separate MCP12 from most of the major contaminants, including the abundant species that causes the crude concentrated culture supernatant to appear deep green in colour. Even though the main peak is very broad, it appears that most of the constituents are MCP12 with different glycoforms, and a band of the correct size for deglycosylated MCP12, is the major species in each of these fractions following treatment with EndoH$_f$. It is possible, that the broadness of the peak arises in part from the presence of unfolded species with different hydrophobicity properties. For this reason only those fractions that made up the central part of the peak were taken for further analysis. Furthermore it can be seen that the flow-through fractions also contain MCP12 albeit in relatively small amounts; this appears to arise predominantly from a small proportion of very heavily glycosylated protein.

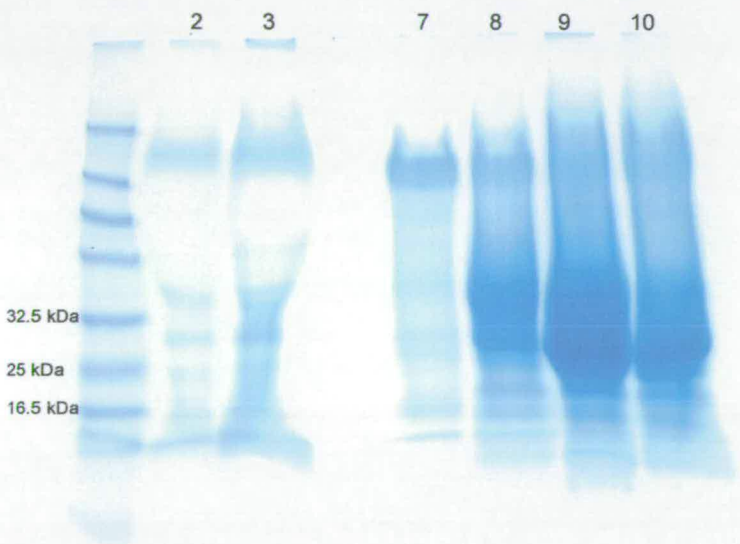


Figure 3:9 SDS-PAGE showing the purification of MCP12 on resource Iso. The glycoprotein samples were concentrated and exchanged into water before being subjected to SDS-PAGE. The rather broad hump apparent in the chromatogram appears to be due to the partial separation of the different glycoforms. This is apparent by observing that the size of the major species in each lane grows progressively smaller with increasing fraction number.

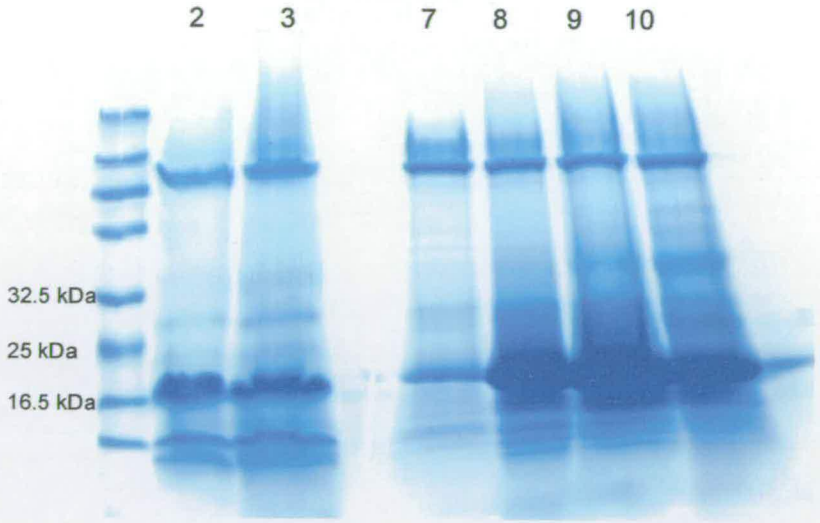


Figure 3:10 SDS PAGE of deglycosylated MCP12 purified on a Resource iso column. A band of around 70kDa is visible in each lane, and corresponds to the expected size for endoH_f. The same amount of protein that was applied to the glycosylated gel (Figure 3:9) was deglycosylated with endoHf and loaded in the same way and despite this there appears to be a much larger amount of protein on this gel compared to the gel in Figure 3:9. This effect is due to the protein being concentrated into a single band following deglycosylation, rather than the diffuse smear that is observed in Figure 3:9. The deglycosylation reaction appears not to have gone to completion, as indicated by the remnants of the higher molecular weight bands that are visible in Figure 3:9. Degradation can also be observed, but it is hard to know whether this degradation existed in the samples prior to deglycosylation or is an artefact of deglycosylation.

3.3.2. Purification of EndoH_f deglycosylated MCP12

The pooled fractions corresponding to the central portion of the main peak from the HIC chromatography were concentrated, and deglycosylated using EndoH_f. EndoH cleavage of N-linked glycans results in a single N-acetylglucosamine (GlcNAc) residue attached to the asparagine residue of the glycosylation sequon, and this enzyme has been used successfully for the removal of high mannose type glycans from recombinant proteins subsequently used for biophysical studies (McAlister et al., 1998). In addition to the deglycosylated protein species, the deglycosylation

reaction should also yield the glycans that were cleaved off the protein, while a fraction of glycoprotein that remains glycosylated. These glycospecies are in addition to the unattached sugars that are naturally produced by *P. pastoris*. Such contaminants can be difficult to remove using conventional chromatography. The majority of these species are likely to contain the mannose residues typical of high mannose type glycans. The lectin Conconavalin A (ConA) from the jack bean *Canavalia ensiformis* binds selectively to the non-reducing terminal mannose; and to a much lesser extent to glucose residues of branched glycans (Hietanen et al., 1989). ConA is readily available pre-immobilised to a sepharose matrix. This may therefore represent an efficient way to remove the cleaved glycans along with the non-deglycosylated protein. The fully deglycosylated protein should not adhere to the column, or should stick only very weakly due to its remaining N-acetylglucosamine residue, since ConA has only a low affinity for GlcNAc. The glycans and non-deglycosylated protein on the other hand should bind tightly to the column. The chromatogram shown in Figure 3:11 shows that whilst most of the protein had been effectively deglycosylated, and passed straight through the column as predicted, there is still a proportion of protein that stuck to the column indicating incomplete deglycosylation. This method therefore proved to be a useful technique for removing the glycans, and non-deglycosylated protein from the deglycosylated MCP12. This purification method was also useful in removing the non-attached sugars that are present in the culture supernatant of *P. pastoris* cultures.

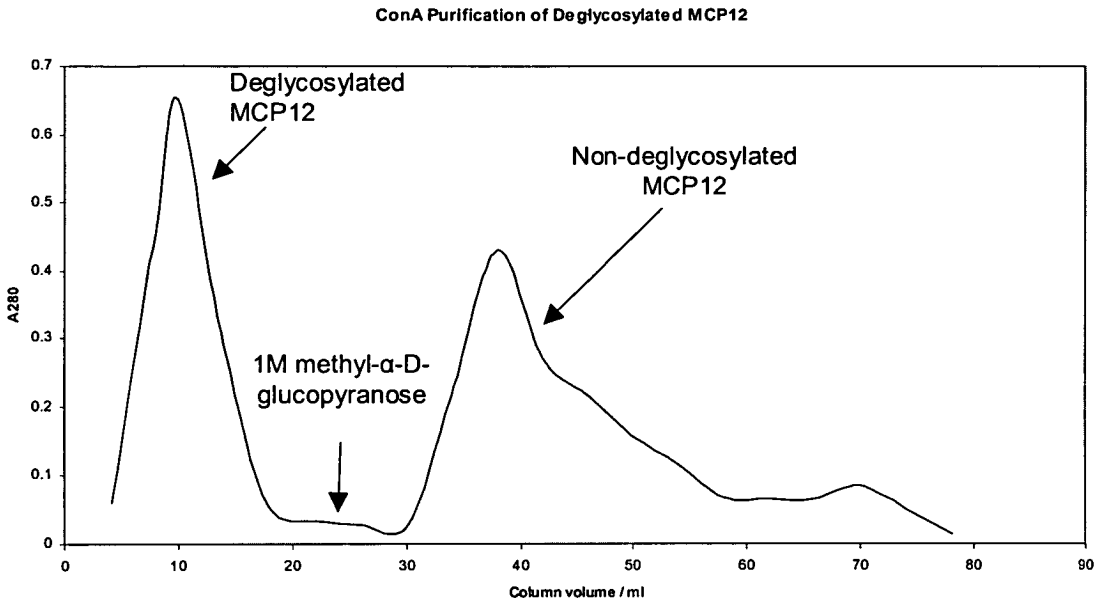


Figure 3:11 ConA purification of deglycosylated MCP12. Fully deglycosylated MCP12 flows straight through the column without retention, the sugars are eluted along with any remaining glycosylated MCP12.

Following EndoH$_f$ deglycosylation and ConA purification, the sample was subjected to anion exchange chromatography, on a MonoQ HR5/5 column (Amersham Pharmacia Biotech) or a Poros20Q column (Applied Biosystems). This step was designed to remove the EndoH$_f$ used for deglycosylation, any degraded MCP12 along with any remaining contaminants. The chromatogram shown in Figure 3:12 shows a single sharp peak, with only minor other impurities. Only the fractions comprising the main sharp peak were pooled and taken for further analysis. The fractions on either side were discarded, despite the presence of trace quantities of MCP12 in these samples (data not shown) to reduce the possibility of carrying modified or improperly folded protein.

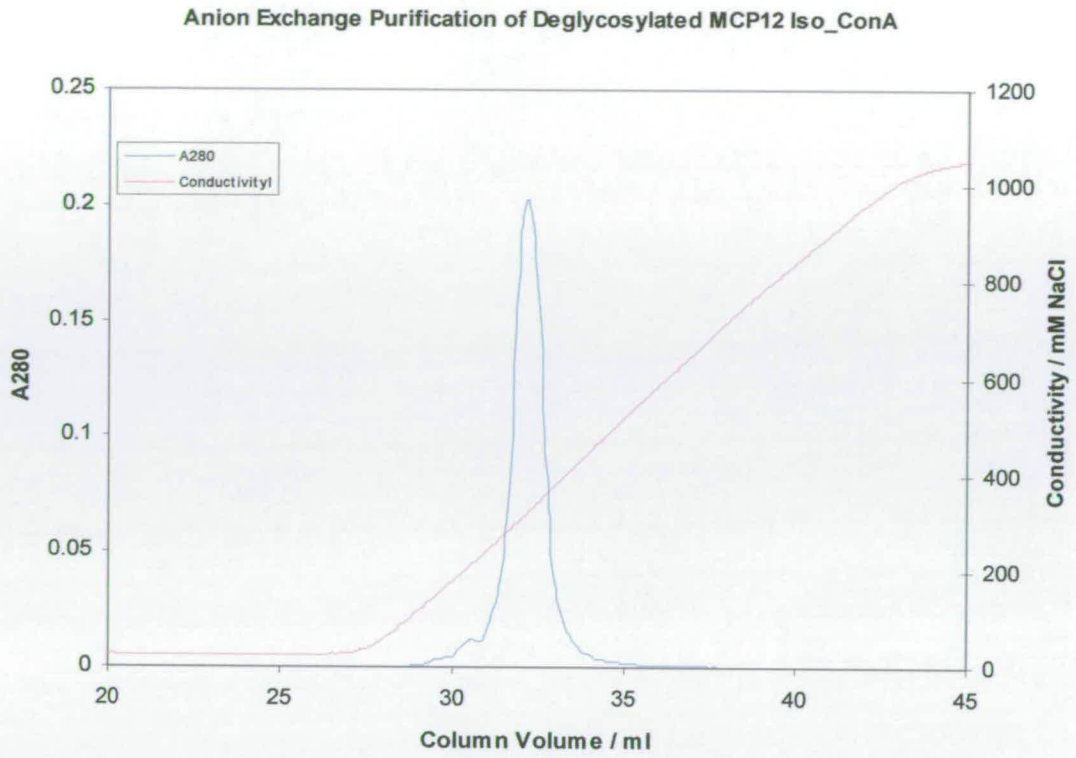


Figure 3:12 Anion exchange purification of deglycosylated MCP12. The chromatogram gives a single peak with a slight tail both leading and following the peak. When the fractions were pooled for further analysis these regions of the chromatogram were excluded.

Following the anion exchange purification step of the $EndoH_f$ deglycosylated MCP12, this protein became purified to homogeneity as judged by SDS PAGE, as illustrated in Figure 3:13. The yield from this recombinant clone was typically in the region of 10 mg of purified deglycosylated MCP12 per litre of induction medium. This expression level was subsequently shown to be sufficient to allow isotopic labelling with ^{15}N and further biophysical characterisation.

Figure 3:13 SDS PAGE of purified deglycosylated MCP12. The gel shows that MCP12 has been purified to homogeneity, as judged by SDS PAGE and runs at approximately the correct molecular weight for MCP12.

3.3.3. Purification of glycosylated MCP12

The heterogeneity of glycoforms present in the glycosylated form of MCP12 causes various difficulties arise. The two most noticeable of these difficulties are the difficulties associated with purifying glycosylated MCP12 away from other contaminating proteins, and those associated with purifying glycosylated MCP12 away from the non-attached sugars that are naturally produced by *P. pastoris*. The presence of large quantities of non-attached sugars that are present in the culture medium of *P. pastoris* means that the lectin affinity chromatography that was effectively used for the removal of contaminating sugars with deglycosylated MCP12 cannot be used for this purpose for glycosylated MCP12. Moreover, the

large quantities of covalently attached sugars, and their excessive heterogeneity, can cause it to behave unpredictably during subsequent chromatography.

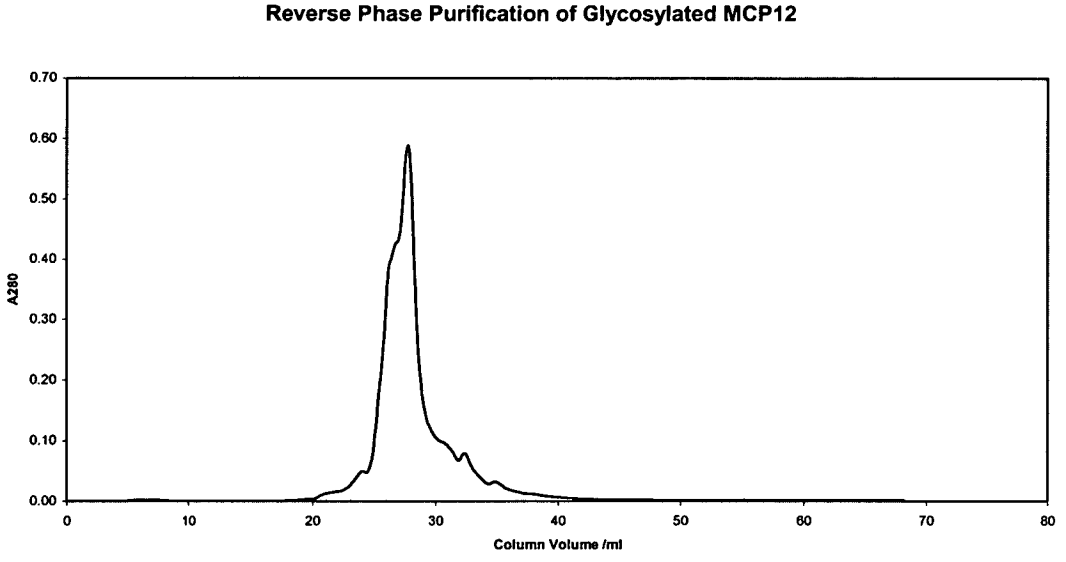


Figure 3:14 Reverse phase purification of glycosylated MCP12. The fractions from the central peak between 25 and 29 ml column volume contained purified glycosylated MCP12 and were pooled for further analysis. Fractions either side contained the remaining impurities and trace quantities of MCP12.

Reverse phase chromatography has been successfully used to remove the unattached *P. pastoris* sugars from recombinant protein samples. It was a logical choice therefore to choose reverse phase chromatography as the final purification step for glycosylated forms of MCP12. The chromatogram shown in Figure 3:14 shows that the majority of the glycosylated MCP12 is eluted form the C4 reverse phase column between 25 ml and 29 ml elution volume. The chromatogram as expected exhibits a relatively broad peak for MCP12, and effectively separates the glycosylated MCP12 from the remaining protein impurities. It remains unclear whether all of the contaminating sugars have been removed from the sample, since there is no practical

method for assessing this. Due to the heterogeneity of this sample it is difficult to determine with any certainty the exact purity of this sample. All of the glycosylated protein is found within the expected region for glycosylated MCP12. Moreover following deglycosylation with $EndoH_f$ only two bands are observed. One of these is the correct size for deglycosylated MCP12, whilst the other is the correct size for $EndoH_f$.

The results in the preceding section demonstrate the successful purification of both glycosylated and enzymatically deglycosylated forms of MCP12.

3.4. Biophysical Characterisation of MCP12

3.4.1. Preliminary remarks

A full understanding of a protein requires biophysical methods of characterisation, which, offer ways to gain insights into everything from the chemical composition to the atomic resolution 3D structure. The results presented in the following section describe the characterisation of both the glycosylated and deglycosylated forms of MCP12 using a number techniques.

3.4.2. Mass spectrometry analysis of MCP12

To confirm that the protein species that had been purified was indeed MCP12, and to establish the integrity of the MCP12 produced, electrospray ionisation mass spectrometry (ES-MS) was performed on a single quadrupole instrument. This type

of MS has a resolution of approximately 0.01 % (Siuzdak, 1996). It therefore has the ability to provide compelling evidence that the protein is the correct species, and to what extent it has been modified or degraded. In this case the ES-MS shown in Figure 3:15 of the intact unlabelled deglycosylated MCP12 gave a major species of 15822 Da, this is 3 Da smaller than the predicted mass of 15825 Da for MCP12 with two GlcNAc residues and four disulphide bonds. This mass difference is difficult to interpret, and could potentially be explained by the hypothesis that only one of the glycosylation sites is recognised, and that the N-terminus has been incompletely processed (losing 203 Da from the loss of one GlcNac, but gaining 200 Da from the addition of EA). This idea would give a species of the correct mass, however it takes no account of the N-terminal sequencing results, which demonstrated complete processing of the secretion signal. It is also possible that three lysine residues have been converted to allysine, loosing 1 Da each. The discrepancy may also arise partly from amide formation at the C-terminus of the protein. The mass difference may also simply be due to a calibration error in the mass spectrometer. Regardless of which is true, it seems to be most likely that the species that was produced was MCP12 with two GlcNAc residues and four disulphide bonds. The species observed at 15859 Da is 37 Da larger than the main peak and probably represents a potassium adduct. The species observed at 15919 Da is 97 Da larger than the main peak and probably represents two potassium adducts and one sodium adduct. The presence of sodium and potassium adducts is an artefact of mass spectrometry, and should in no way alter the properties of the protein.

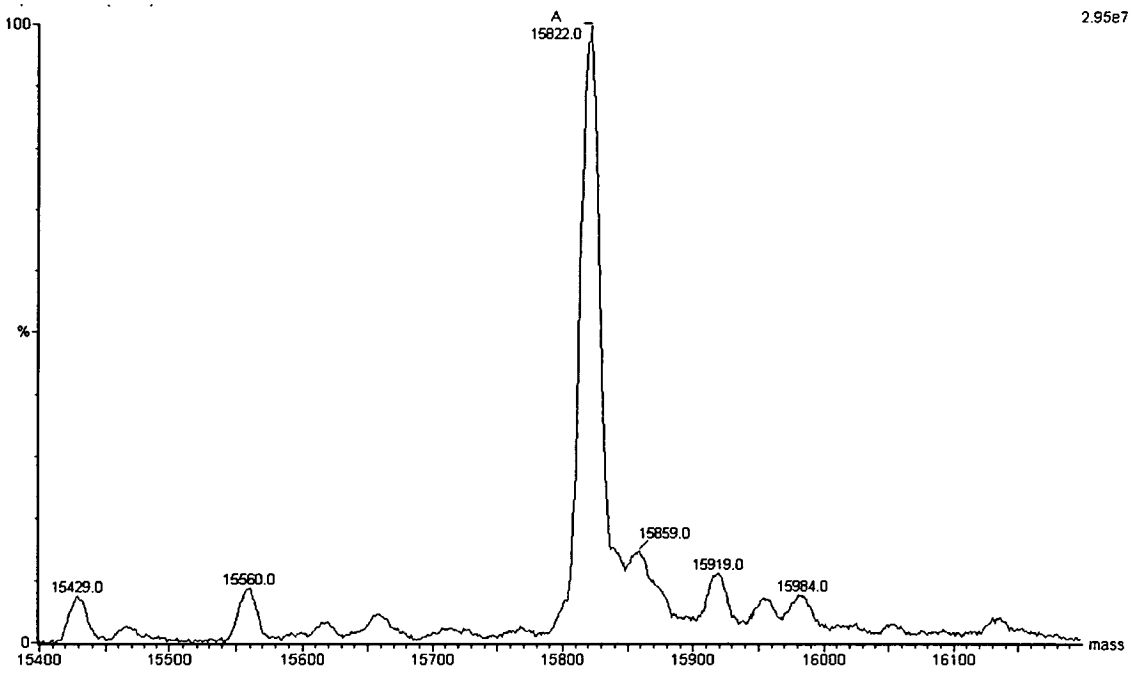


Figure 3:15 Mass spectrum of EndoH$_f$ deglycosylated MCP12. The mass of the main peak is within 3 Da of the expected mass of EndoHf deglycosylated MCP12, with four disulphide bonds, and two GlcNAc residues. Other peaks discussed in the text.

The small peak observed at 15560 Da is 262 Da smaller than the main peak and is consistent with the loss of the N-terminal YV residues. The small peak observed at 15429 Da is 393 Da smaller than the main peak, which is close to the predicted mass for the loss of the N-terminal YVE residues – a predicted mass difference of 391.5 Da. The ES-MS data are consistent with the expected data for EndoHf deglycosylated MCP12 with four disulphide bonds and two GlcNAc residues. These data, when combined with the N-terminal sequencing results indicate that both of the possible N-glycosylation sites in MCP12 have been glycosylated by *P. pastoris*, and that both sets of glycans were successfully removed. These data also indicate that the major portion of the protein is intact, and that only minor quantities of degradation are observed. Moreover, those small amounts of degradation that are observed all appear to involve removal of the cloning artefact amino acids, and

therefore would not be expected to have a detrimental effect on the protein's structure and function.

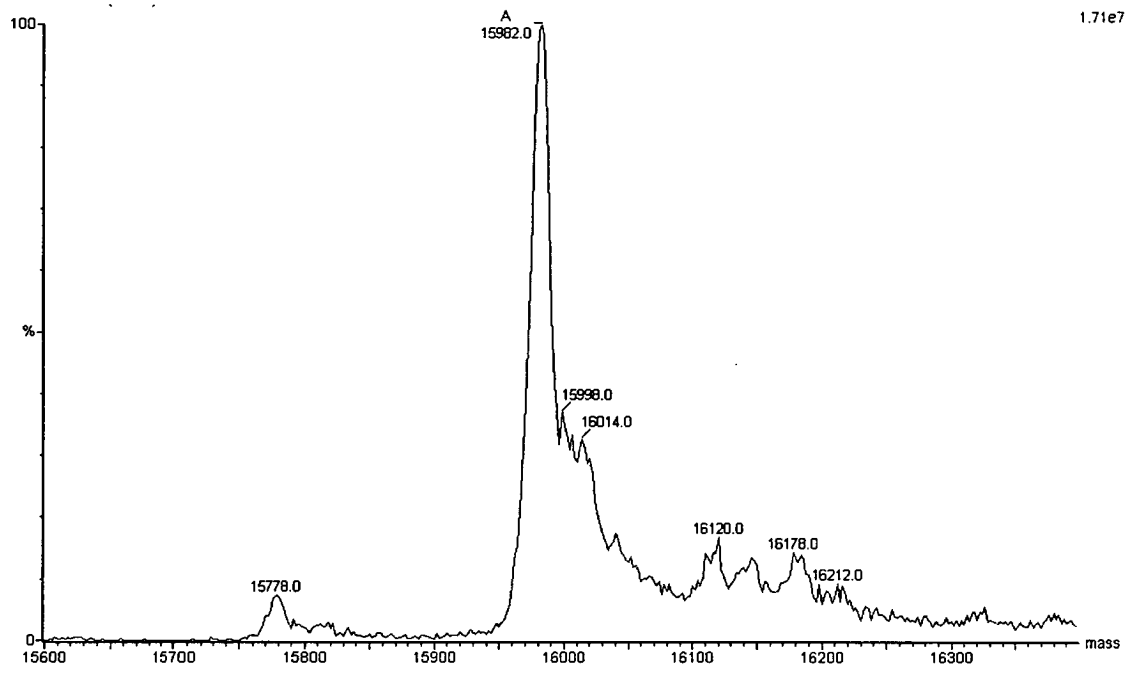


Figure 3:16 Mass spectrum of EndoHf deglycosylated ^{15}N labelled MCP12. The spectrum shows a main peak that is consistent with > 90% incorporation of ^{15}N into MCP12.

In addition to its ability to confirming the identity and integrity of MCP12, ES-MS was also used to confirm the incorporation of ^{15}N into MCP12 in a sample prepared for the purpose of heteronuclear NMR experiments. The mass spectrum shown in Figure 3:16 is of ^{15}N labelled EndoH$_f$ deglycosylated MCP12. The sample of ^{15}N labelled MCP12 had been produced using exactly the same procedure as that for the unlabelled samples with the sole exception that the recombinant *P. pastoris* was cultured in medium that contained ^{15}N ammonium sulphate (2 gl^{-1}) as the sole nitrogen source (Wiles, 1996). The ES-MS spectrum shows a peak with a mass of 15982 Da which is 157 Da larger than would be expected for unlabelled MCP12.

The incorporation of ^{15}N at 100% of possible sites would result in an increase in molecular weight of 170 Da. These ES-MS data are consistent with approximately 92% incorporation of ^{15}N into MCP12. It is impossible use these ES-MS data to assess whether the incorporation of ^{15}N nuclei into MCP12 is uniform; however a site preference for the incorporation of ^{15}N labels has not been reported nor is expected.

3.4.3. NMR spectroscopic analysis of MCP12

3.4.3.1. NMR analysis of deglycosylated MCP12

Of all the biophysical techniques available for the study of proteins, high field NMR spectroscopy is perhaps the most powerful. The chief advantage that NMR spectroscopy has over other techniques is that it is the only biophysical solution technique that is residue specific. With multi-dimensional NMR spectroscopy one can observe the individual nuclei in a protein directly, non-destructively and in an aqueous environment. In favourable cases the optimum NMR conditions may even be near physiological with respect to temperature, salt concentration, and pH. Other solution techniques observe signals that are averaged over the whole protein, or that depend upon the presence of tryptophan or specifically inserted labels to give limited localised information. In addition, the majority of these techniques are either destructive or expose the protein to unnatural conditions. High field multidimensional NMR spectroscopy therefore provides a powerful method of probing a protein in great detail.

The dispersion of resonances over a wide range of -0.4 ppm to 11.2 ppm in the 1D ^{1}H spectrum shown in Figure 3:17 is indicative of a folded protein. The spectrum also shows relatively sharp peaks, indicating that self association was limited. Given the quality of this spectrum of MCP12, it made sense to produce a ^{15}N labelled sample, to facilitate the acquisition of the multidimensional heteronuclear NMR experiments that permit more detailed analysis. A ^{15}N sample was made which had > 90% incorporation of ^{15}N (see Figure 3:16).

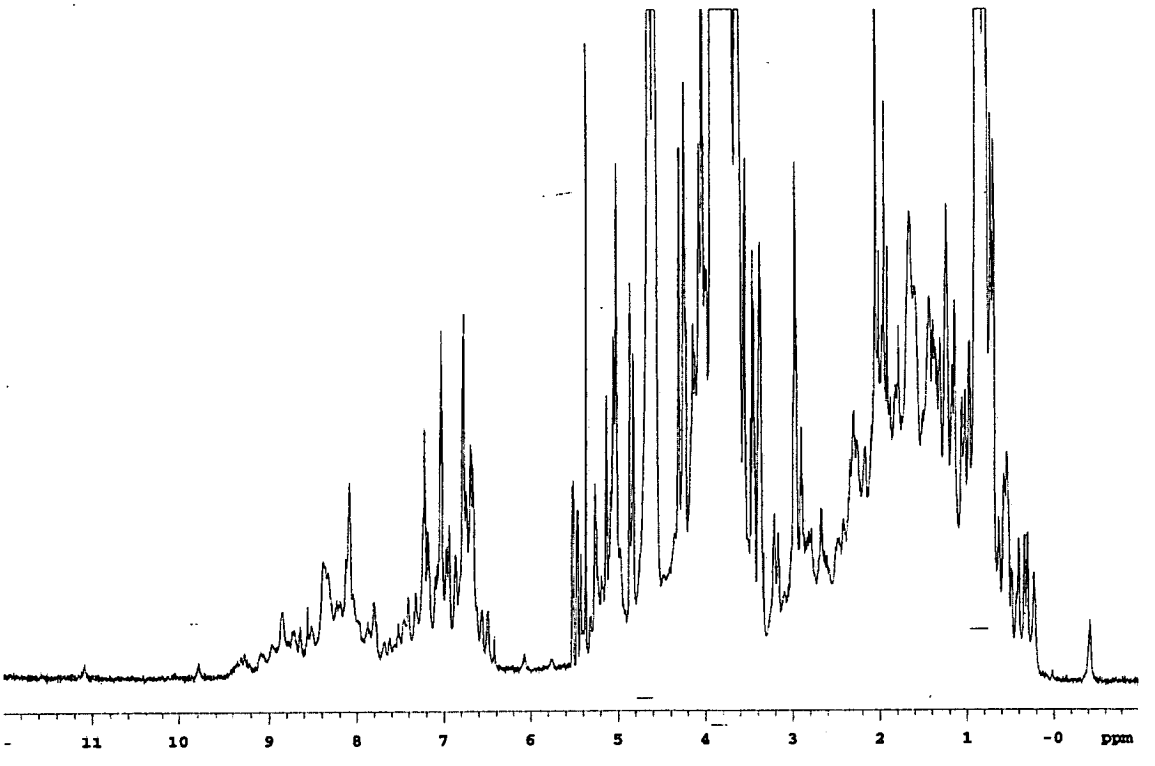


Figure 3:17 1D ^{1}H-NMR spectrum of EndoHf deglycosylated MPC12. The sample was 0.9 mM in 20 mM potassium phosphate at pH 6.0. The spectrum was recorded at 37 °C at 600 MHz proton frequency with presaturation water suppression.

One of the most powerful NMR experiments is the 2D ^{1}H-^{15}N heteronuclear single quantum coherence spectrum (^{1}H-^{15}N HSQC). This experiment produces a single peak for each NH group in a protein. This therefore yields a resonance for each amino acid (with the exception of proline), along with a resonance for the NH group in the side chains of certain amino acids. This experiment therefore provides detailed residue specific information. The ^{1}H-^{15}N HSQC spectrum shown in Figure 3:18 is once again indicative of a folded protein.

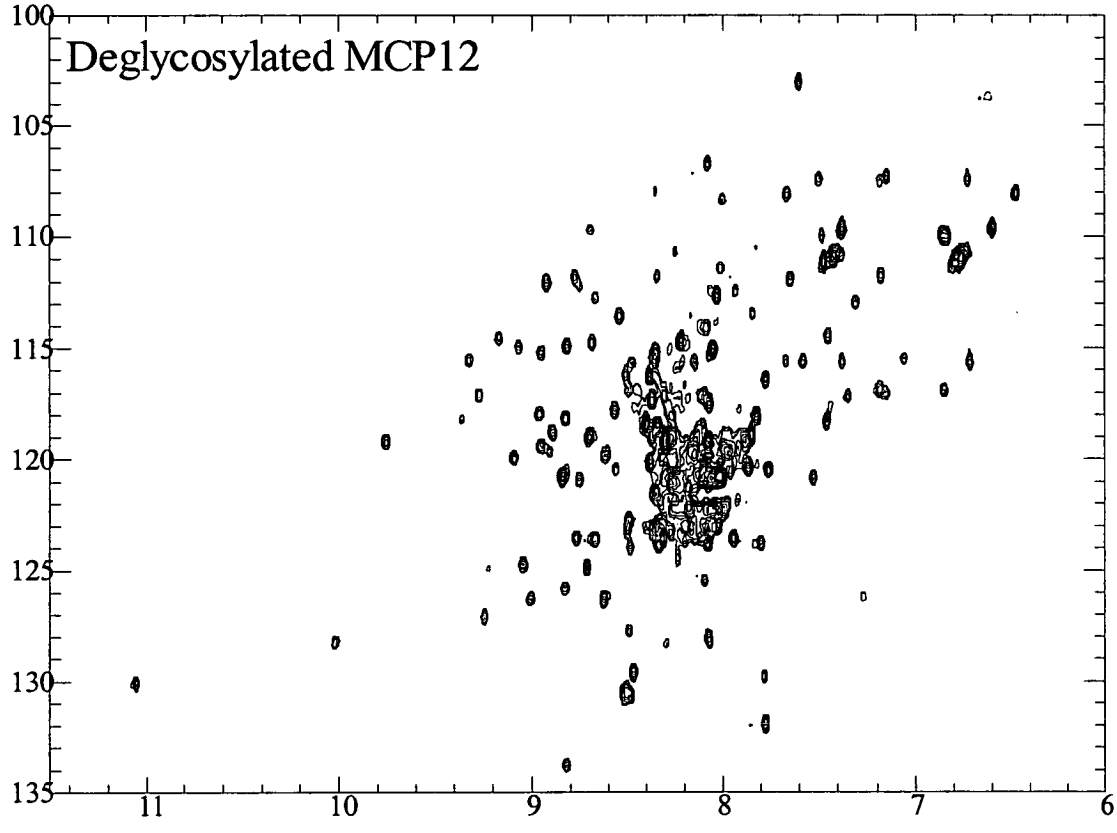


Figure 3:18 ^{1}H ^{15}N HSQC spectrum of deglycosylated MCP12. The sample was 0.9 mM in 20mM potassium phosphate pH 6.0.

Overall, the spectrum indicates that the protein is largely folded with a large number of resonances significantly shifted from random coil region in both the ^{15}N and ^{1}H

dimensions. There is however, a higher than expected degree of overlap in the centre of the spectrum. This region of overlap could be due to a number of factors which include: the presence of unfolded material; aggregation; the presence of ^{15}N labelled impurities, or conformational exchange. Analysis of the number of resonances indicates than 107 resonances are reasonably well dispersed and countable, of these 7 are probably side chain resonances. This results in the number of dispersed backbone NH resonances being approximately 100 and compares unfavourably with the predicted number of resonances. After accounting for the proline residues, the MCP12 species should yield 127 backbone NH resonances indicating too few resonances for both modules in MCP12 to be fully folded. There are however too many resonances to account for only one module being properly folded. Missing resonances can be due to excessive line broadening – a phenomenon that can be caused by a number of factors. Conformational exchange and amide exchange with water on an intermediate (ms) time scale can result in this line broadening, as can aggregation. Aggregation of proteins can be quite a common phenomenon at the concentrations required to give adequate signal to noise in NMR experiments.

To determine whether the presence of the overlapping region was due to a problem with one particular module, or was indicative of a more widespread problem, the ^{1}H-^{15}N HSQC of deglycosylated MCP12 was compared to a previously acquired ^{1}H-^{15}N HSQC spectrum of deglycosylated MCP1. The NMR data of MCP1 had been acquired on a deglycosylated sample of MCP1 previously expressed in *P. pastoris* (O'Leary, 2000). The spectra overlaid in Figure 3:19 indicate that both of the modules are folded - at least to some extent. This can be inferred because a large

number of resonances from the residues of MCP1 can largely be correlated with similar resonances in the spectrum of MCP12. In addition to the resonances that are similar to those from MCP1; the HSQC spectrum of MCP12 contains a good number of extra resonances – presumably from the second module of MCP12.

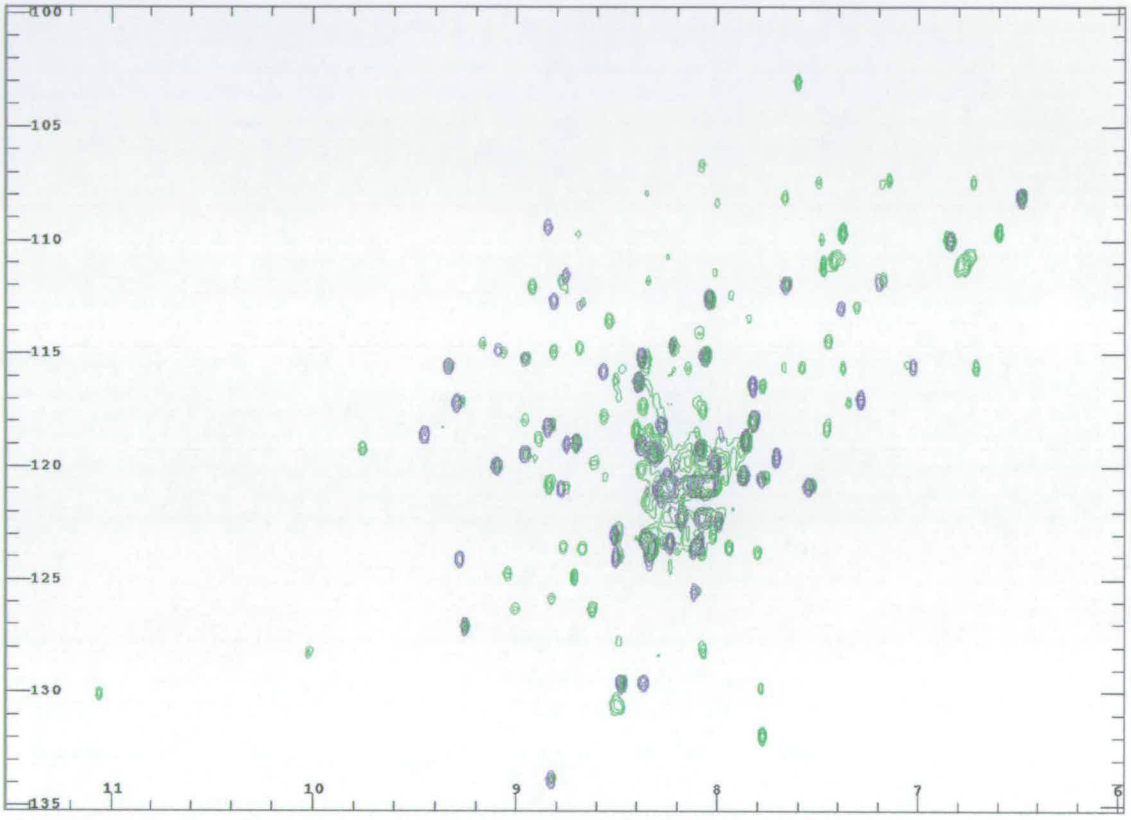


Figure 3:19 Overlay of ^{1}H ^{15}N HSQC spectra of MCP12 (green) and MCP1 (purple). The overlay shows a large number of the resonances from MCP1 have almost identical shifts to resonances in MCP12. In addition there are a large number of extra resonances in the MCP12 spectrum that presumably do not arise from MCP1. Taken together this indicates that both of the modules of MCP12 are to a large extent regularly folded. The central overlap present in the MCP12 spectrum, however appears to be absent from the MCP1 spectrum.

In order to achieve the optimal conditions for NMR experiments on MCP12, ^{1}H-^{15}N HSQC spectra were acquired under a number of different conditions. The conditions optimised included pH, salt concentration, temperature, protein concentration. The

spectrum shown in Figure 3:18 represents HSQC spectrum of deglycosylated MCP12 with the greatest dispersion and least overlap in the random coil region that was recorded. TOCSY spectra were subsequently acquired on deglycosylated MCP12 with a variety of different mixing times, in order to determine whether deglycosylated MCP12 could be suitable for structural studies to be undertaken.

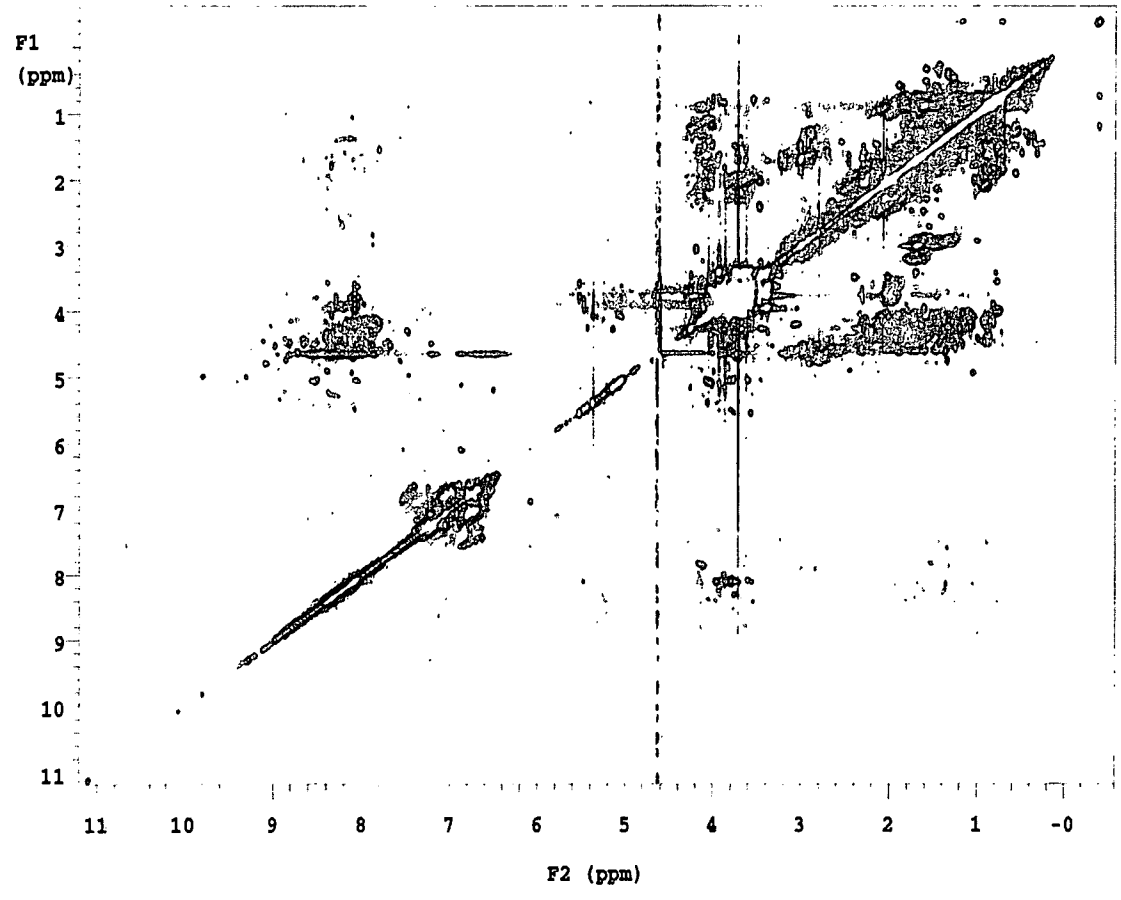


Figure 3:20 TOCSY spectrum of MCP12 with watergate water suppression. The spectrum was acquired at 37 °C; mixing time of 60 ms. The sample was 0.9 mM in 20 mM potassium phosphate pH 6.0.

The 2D ^{1}H TOCSY spectrum shown in Figure 3:20 was acquired with a mixing time of 60 ms and shows a dearth of cross peaks. This is particularly true in the amide region of the spectrum; a phenomenon that is often associated with aggregating

proteins. Aggregating proteins generally have shorter T_2 values than do their monomeric counterparts and this results in a rapid 'fall off' of the signal, with the result that the TOCSY transfer cannot extend far along the amino acid side chain. Interestingly however, the cross peaks that are present in this ^1H TOCSY, also have a rather unusual shape, and appear distorted. It is unlikely that this phenomenon is due entirely to aggregation, and it is possible that conformational exchange may also be responsible. In order to determine whether the unfavourable behaviour of deglycosylated MCP12 was due to aggregation, ^{15}N T_2 values were measured, and these values were typically in the high 80's and low 90s' ms. T_2 values in this range are slightly shorter than the ideal for a protein of this size and may suggest aggregation, however they are not too dissimilar to those obtained from CR1~15-16 (~90 ms). In addition, dilution of the sample had only a negligible effect on the measured ^{15}N T_2 values, indicating that either excessive aggregation was not occurring, or that aggregation was occurring to a similar extent over the range of concentrations tested. Taken together, the results described in this section indicate that deglycosylated MCP12 could be prone to aggregation under NMR conditions, that it is undergoing some form of conformational exchange process; or possibly a combination of both. In order to address these issues, a glycosylated sample of MCP12 was made and analysed, which is the subject of the following section.

3.4.3.2. NMR analysis of glycosylated MCP12

Given the functional and potentially structural importance of the N-glycan found on CCP module 2 of MCP, it is possible that the non-dispersed region of the HSQC

spectrum could be associated with the removal of the N-linked glycans. We therefore decided to produce a sample of MCP12 without removing the N-linked glycans added by *P. pastoris*. It would normally be expected that the presence carbohydrates on a protein would result in considerably broader NMR signals because of the extra mass they would add to the protein. This problem is especially pertinent in this case as the mass of carbohydrate that is attached to the N-glycosylation sites is greater than the mass of the protein.

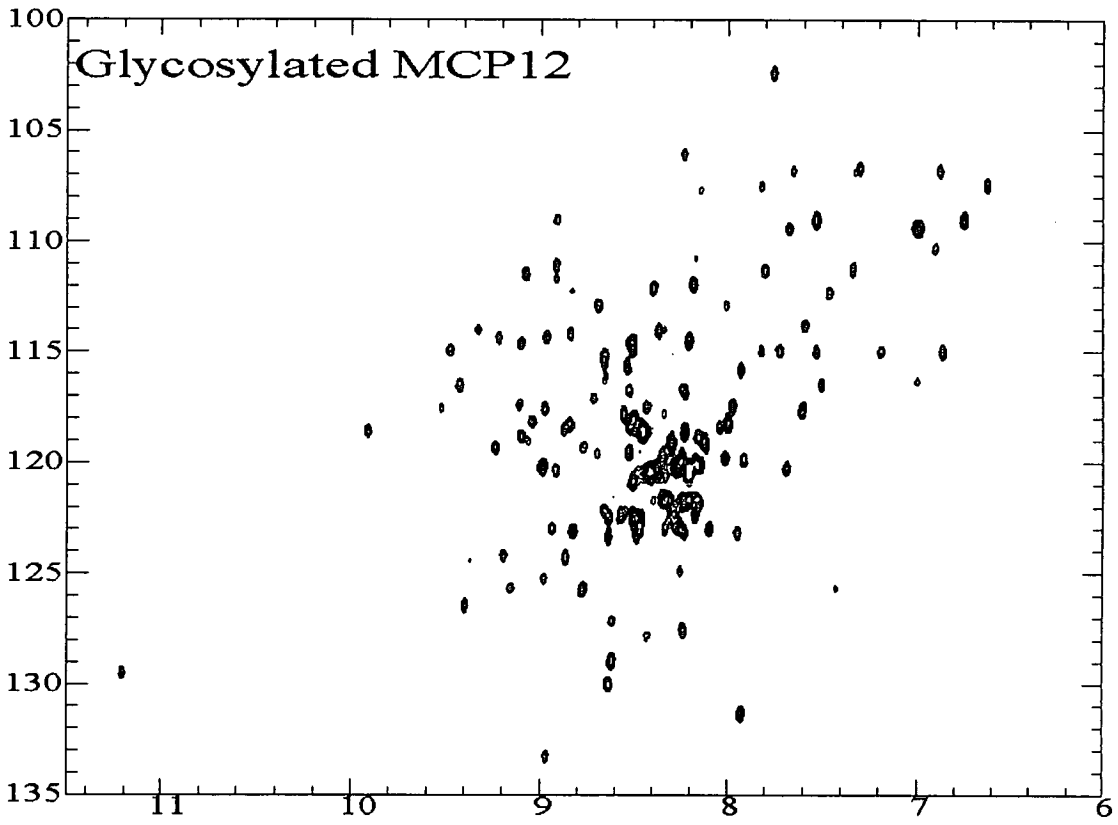


Figure 3:21 ^{1}H - ^{15}N HSQC spectrum of glycosylated MCP12. The sample was 0.9 mM in 20 mM potassium phosphate pH 6.0.

The ^{1}H-^{15}N HSQC spectrum shown in Figure 3:21 once again shows a similar degree of dispersion to that of the deglycosylated sample, however the peaks in this

spectrum appear to be sharper than those of the deglycosylated version. It is also clear that the overlap in the centre of the spectrum is significantly reduced compared to the deglycosylated form. When the ^{1}H-^{15}N HSQC spectra from glycosylated and deglycosylated MCP12 are overlaid (illustrated in Figure 3:22) it becomes clear that there are some quite profound differences between the spectra from two forms of MCP12.

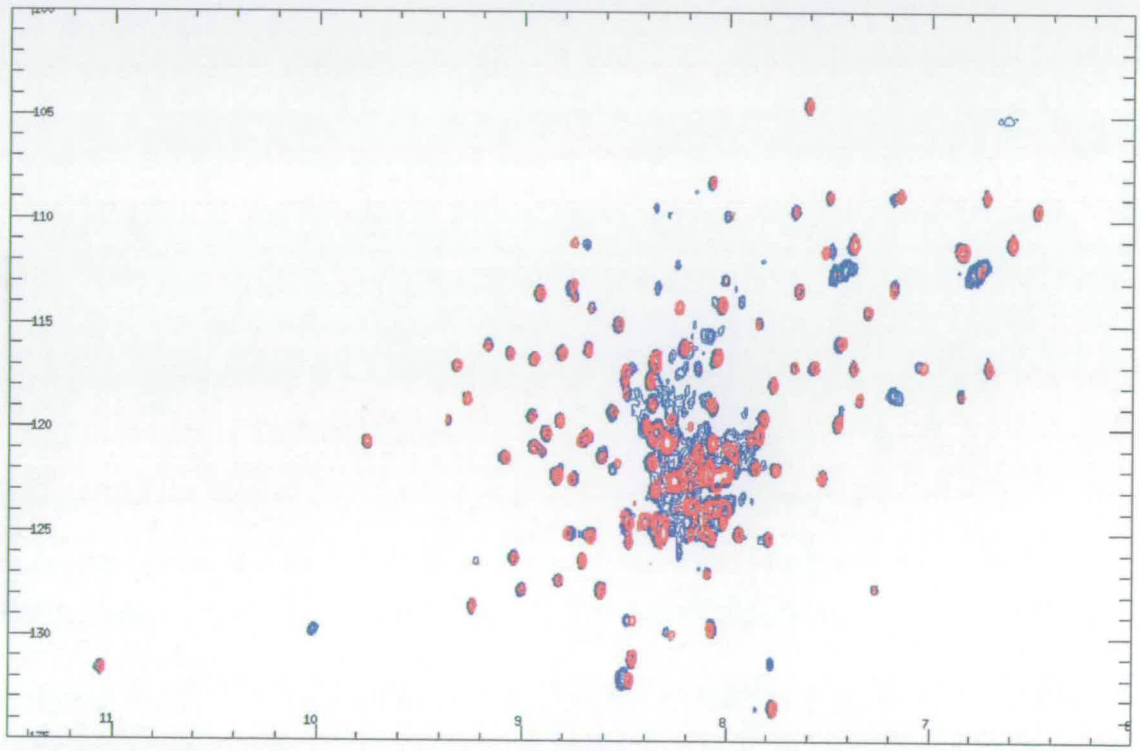


Figure 3:22 Overlay of the ^{1}H - ^{15}N HSQC spectra from glycosylated and deglycosylated MCP12. Both samples were 0.9 mM in 20 mM potassium phosphate pH 6.0.

It can clearly be seen that the deglycosylated MCP12 gives extra resonances that are not observed in the glycosylated sample. The glycosylated sample on the other hand appears not to have any resonances that are not found in both spectra. Given these profound differences between these two spectra, and the known biological

requirement for the glycans it would seem reasonable to think that the glycans may playing some structural role and that this gives rise to the differences between these spectra. It is possible however, that the glycans in the glycosylated sample are simply acting to make the protein more soluble and therefore less prone to aggregation. One would imagine, however that this 'solubility effect' would be compensated for by the increased molecular weight of the glycosylated sample.

3.4.4. Gel filtration of deglycosylated MCP12

In order to determine whether the deglycosylated MCP12 contains dimers or higher order aggregates, the protein was applied to a Superdex 75™ gel filtration column and the chromatogram of which is shown in Figure 3:23. This chromatogram shows a single sharp peak with no evidence of the presence of higher order aggregates. Gel filtration chromatography can also be used to estimate the molecular weight of a protein; however, the technique is shape dependant. It therefore makes sense in this case to calibrate the column using other CCP module constructs which one would expect to have a similar overall topology. Using single, double and triple CCP modules as a calibrant for this column, MCP12 elutes in the volume between the region where double and triple modules elute. This could possibly mean that it is present as a dimer that behaves as though it is smaller than it actually is. It is more likely, however given that MCP12 is of a larger molecular weight than that used to calibrate the column (VCP23 Mr 12.5 kDa, MCP12 Mr 15.5 kDa) that under these conditions deglycosylated MCP12 exists as a monomer. The glycosylated sample

was not applied to the Superdex 75™ column since it was expected that the extreme heterogeneity of the sample would render the results meaningless.

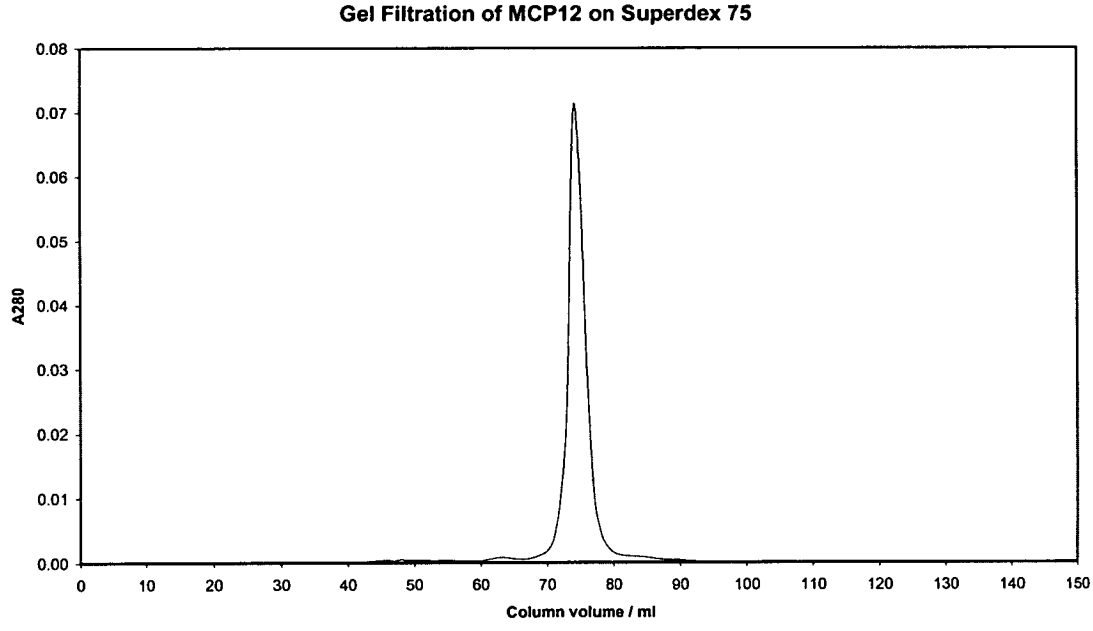


Figure 3:23 Gel filtration of purified MCP12 on a Superdex 75 column. The chromatogram shows a single sharp peak which eluting slightly earlier than would be expected for a protein of this size, when the Pharmacia gel filtration standards are used to calibrate the column. However, this has also been observed with several CCP module constructes, and when the column is calibrated using CCP module constructs, the protein elutes approximately where expected.

Moreover, the observation that the gel filtration of MCP12 yielded a sharp peak indicates that the protein is not in an equilibrium between a monomeric and dimeric state, since in such cases broad peaks are often observed. The results presented thus far indicate that the glycans of MCP12 result in different behaviour between the two forms, however whether this is due to a structural interaction, or intrinsic properties of the glycans themselves remains moot.

3.4.5. Differential Scanning Calorimetry of MCP12

In order to determine whether the glycans of glycosylated MCP12 provide it with increased thermal stability over deglycosylated MCP12 differential scanning calorimetry was performed. Samples of both glycosylated and deglycosylated MCP12 were prepared and subjected to analysis by differential scanning calorimetry (DSC).

DSC is a method by which an insight can be gained into the thermal stability of a protein. The technique is based on the fact that there is an enthalpic difference between the folded and unfolded states of a protein, which arises due to the energy stored in the protein's structure. This difference can be measured upon thermal denaturation of the protein, and can therefore provide data about the thermal stability of the protein. Differential scanning calorimetry measures the thermal energy taken up by a sample upon heating or cooling. This is achieved by comparing the different thermal energy required to heat a sample and reference cell at the same rate. During the unfolding transition, the heat energy uptake (C_p) in measured with respect to the reference cell and allows the transition temperature (T_m) to be calculated. The T_m is defined as the temperature at which 50% of the protein molecules are folded (Cooper et al., 2000). The data is then base line corrected, normalised to take account of the concentration, and fitted to an appropriate model of unfolding. For most proteins this is usually a simple 2-state unfolding model, although a multistate unfolding model may be appropriate for some proteins. DSC yields two independent measures of the enthalpy changed due to the thermal unfolding of a protein, namely the calorimetric enthalpy (ΔH_{cal}) and Van't Hoff enthalpy (ΔH_V). The calorimetric enthalpy is an

absolute, model free measure of the heat energy released upon unfolding, and is proportional to the amount of protein in the calorimeter cell. In effect ΔH_{cal} is a measure of the area under the transition peak. The Van't Hoff enthalpy on the other hand is independent of the protein concentration, and is a measure of the degree of cooperativity of the unfolding process. The value of ΔH_V is dependant on the shape of the transition curve, and assumes a two state reversible equilibrium. The sharper the transition curve, the greater the value for ΔH_V (Plum et al., 1995). For an ideal protein, $\Delta H_{cal} = \Delta H_{VH}$, although there are several reasons why this may be not the case.

The DSC data obtained for deglycosylated MCP12 are shown in Figure 3:24 and were best fitted to a model with two cooperative unfolding transitions, with T_ms of 58.5 °C and 68.5 °C. It should be noted that in this case $\Delta H_{cal} < \Delta H_{VH}$ and this is probably the result of using a multistep unfolding model. The T_m values obtained for MPC12 can be compared to the value of 62 °C that had previously been obtained for $EndoH_f$ deglycosylated MCP1 (O'Leary, 2000). It is tempting to speculate that the value of 58.5 °C corresponds to the melting of MCP1, and that of 68.5 °C corresponds to that for MCP2, although this cannot be proven with these data.

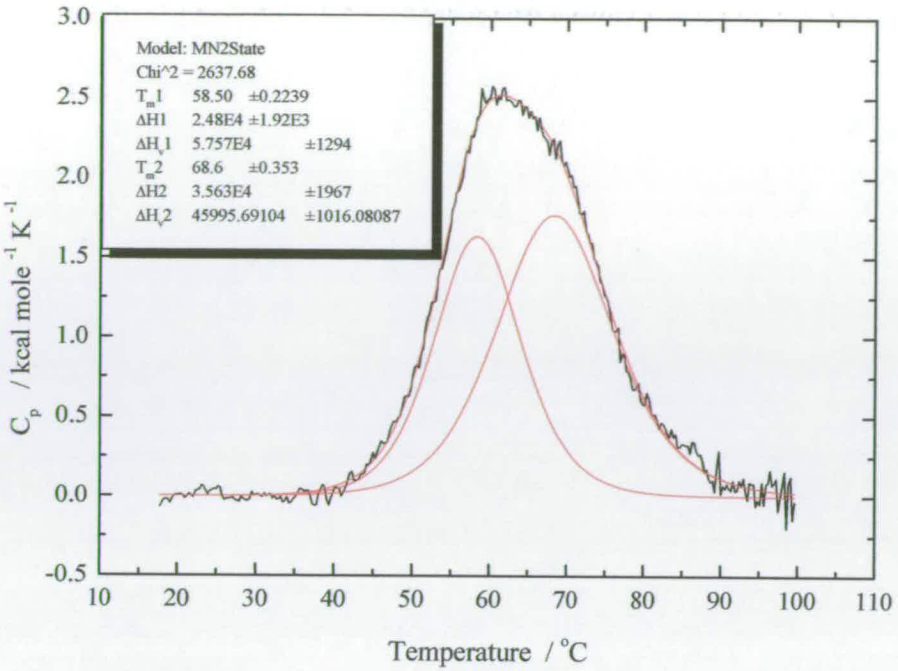


Figure 3:24 Differential scanning calorimetry analysis of EndoHf deglycosylated MCP12. These data fit best to a double melting model. This would indicate that the two modules melt independently, and that one module has a significantly higher melting temperature than the other.

The DSC data for the glycosylated form of MCP12 are shown in Figure 3:25, and were best fitted a simple two step unfolding model with a T_m of 58.5 °C. This would indicate that the presence of the N-linked glycans have destabilised one of the modules and that now both modules appear to share a common melting temperature. It is impossible, however to know from these data which if any module has had its thermal stability reduced by the presence of the N-linked glycans. It is therefore desirable to perform a similar analysis on the individual CCP module fragments.

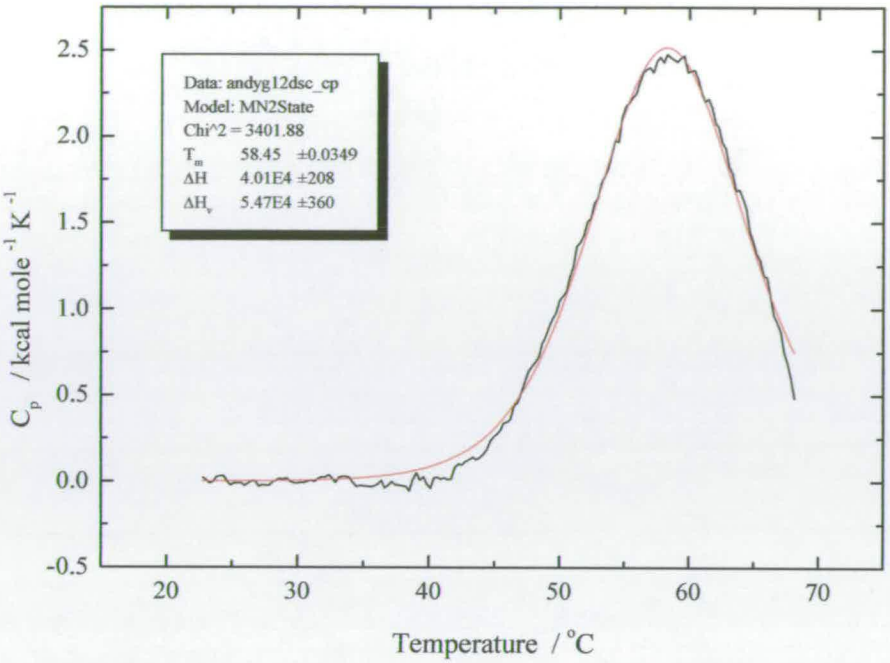


Figure 3:25 Differential scanning calorimetry of glycosylated MCP12. These data fit best to a two step unfolding model with a T_m of 58.5 °C, indicating that both modules share a common melting temperature.

The DSC data for both glycosylated and deglycosylated MCP12, appear to suggest that the glycans have little effect on the thermal stability of one module in MCP12, and that they may have a significant stabilising effect on the other. From the observations in the previous paragraph, it could be inferred that it is the second module that has undergone a lowering of T_m upon deglycosylation, but this is speculative. It has therefore become desirable to perform similar experiments on the second module of MCP on its own, in order to delineate these effects, and these are the subject of subsequent chapters. In addition to recording similar data on the single module construct, it would be useful to determine whether the MCP12 produced is active.

3.5. Functional Characterisation of MCP12

MCP12 appears to be behaving as though it contains some unstructured component. Thus far investigating whether this behaviour is the result of removing the N-linked glycans has proved difficult. It therefore seems useful to investigate the functional characteristics of both the glycosylated and deglycosylated forms of MCP12. This fragment of MCP is thought not to contain any complement activity; however it is thought to be the region that interacts with measles virus. This property, in principal at least, allows one to test MCP12 for activity. The N-linked glycans have been shown to be essential for this function, and the crystal structure of MCP12, expressed in CHO cells, appears to suggest that this requirement for N-linked glycans has some structural basis. Thus far the minimum glycans necessary for function have not been determined, and it is possible that the presence of the GlcNAc following cleavage with $EndoH_f$ may be sufficient. In order to determine whether, the MCP12 fragment expressed by *P. pastoris* is active, and whether EndoH cleavage of the N-linked glycans renders it inactive, a modified version of the virus overlay protein blot assay (VOPBA) was performed (Ohtsuka et al., 1996). The VOPBA result shown in Figure 3:26, shows that hyperglycosylated protein as expressed by *P. pastoris* is able to bind measles virus. Removal of the N-linked glycans, however, results in a reduction of measles virus binding to a level below the detection threshold of this assay. This result also indicates that, as predicted neither deglycosylated MCP1, nor the CR2 construct have the ability to bind detectable levels of measles virus. It would be possible to perform considerably more in-depth analysis on the ability of MCP12 to bind measles virus, and this should yield some interesting information.

Such a study, whilst interesting and informative could be a project in itself, and is not the subject of this thesis. The functional data presented here although preliminary in nature still demonstrate that the hyper-glycosylated protein expressed by *P. pastoris* is active, and therefore presumably is correctly folded, and that the deglycosylated form of MCP12 has a significantly reduced ability to bind to MV. These data add weight to the hypothesis that the deglycosylation of MCP12 disrupts its structure in some manner.

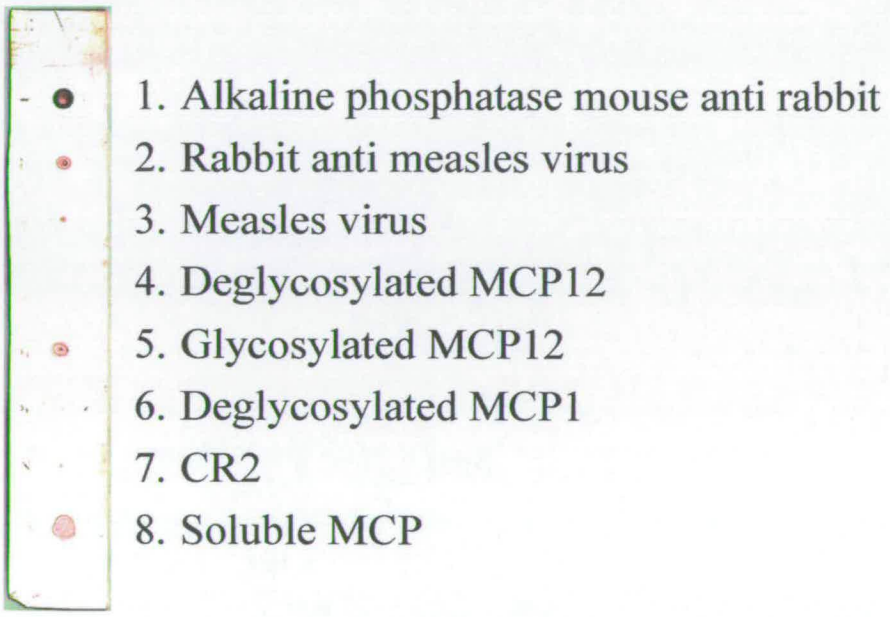


Figure 3:26 Virus overlay protein binding assay showing MCP12 binding to measles virus. Lanes 1-3 represent positive internal controls to demonstrate that all components of the detection system perform as expected. Lane 7 is a negative test control and lane 8 is a positive test control. The three test lanes 4, 5, & 6 indicate that hyperglycosylated MCP12 can bind measles virus, whereas deglycosylated MCP12 and MCP1 cannot bind detectable levels of measles virus.

3.6. Discussion

The results presented in this chapter demonstrate the successful cloning and expression of MCP12 in *P. pastoris* to levels that are sufficient for biophysical characterisation to be performed. Deglycosylated MCP12 was successfully purified to homogeneity as determined by SDS-PAGE, whilst a hyperglycosylated form of MCP12 was purified to a homogeneous protein with heterogeneous glycoforms.

N-terminal sequencing and mass spectrometry data confirmed the identity of the protein as MCP12, and that complete processing of the secretion signal had occurred. Mass spectrometry data also confirmed the integrity of the protein, showing no modifications and only trace amounts of proteolysis, which occurs only in the region of the protein left by cloning artefacts. Preliminary NMR experiments indicate that both the glycosylated and deglycosylated forms of MCP12 show the good dispersion indicative of a structured protein. There is, however, a larger area of overlap in the centre of the ^{1}H-^{15}N HSQC for deglycosylated MCP12 than would normally be expected for a protein of this size. Moreover there are insufficient peaks in the ^{1}H-^{15}N HSQC to account for all the amino acids in the protein. Possible explanations as to why this is the case included: aggregation, regions that are locally unfolded, the presence of contaminants or unfolded material, or the protein being flexible. Comparison of the ^{1}H-^{15}N HSQC spectra of deglycosylated MCP12 and deglycosylated MCP1, shows the presence of MCP1 in a similar state in MCP12 as it is alone. Aggregation would not appear to be the main cause of the overlapped region of the ^{1}H-^{15}N HSQC spectrum, since it remains largely unchanged under a variety of sample conditions.

Functional data suggest that only glycosylated forms of MCP12 are active. It is possible therefore that removal of the N-linked glycans may locally disrupt the structure of MCP12. The NMR data acquired do not contradict this hypothesis and indeed suggest that removal of the N-linked glycans results in an increase in the protein's flexibility. The DSC data obtained from both glycosylated and deglycosylated MCP12 indicate that the glycans have a destabilising effect on one of the modules. The DSC data provide no evidence as to which module is potentially destabilised, since the two modules of MCP12 appear to be behaving independently, and the technique is unable to provide the location specific information required to ascribe the different properties to one module or the other. That the DSC data appear to suggest that the glycans destabilise one of the modules is not necessarily contradictory to the NMR data since DSC records thermal stability, whereas NMR is able to investigate a protein's dynamics. Moreover, this destabilising effect may merely be an artefact of hyperglycosylation as opposed to physiological glycosylation.

The requirement of N-linked glycans for structural integrity appears to be only a rare occurrence, however, the crystal structure of MCP12 also indicates that this might be the case (Casasnovas et al., 1999). Proving this hypothesis is potentially difficult due to the inherent difficulties associated with producing highly purified proteins with defined glycosylation patterns.

It is difficult to delineate the regions of MCP12 that are unstable and therefore, it was decided to express the two modules individually. Since MCP1 had already been

expressed and its solution structure solved (O'Leary, 2000), further efforts were concentrated on MCP. It was anticipated that analysis of the MCP2 single module would provide valuable information about this module, and potentially its behaviour in the MCP12 pair. Expression and characterisation of MCP2 is the subject of the following chapter.

4. Recombinant MCP2 from *P. pastoris*

4.1. Preliminary remarks

The work described in the following section describes the successful expression and purification of the second module of MCP (MCP2) in *P. pastoris*. *P. pastoris* was chosen for expression of MCP2 for the same reasons as it was chosen for expression of MCP12, and because of the success already achieved using this organism in the case of MCP12.

4.2. Cloning and expression of MCP2 in P. pastoris

4.2.1. Cloning of MCP2 in pPICZα

The vector used for the expression of MCP2 in *P. pastoris* was the pPICZα vector, which is distinct from the pPIC9 vector used for the expression of MCP12. This vector, illustrated in Figure 4:1, is also a shuttle vector, but it uses resistance to the antibiotic zeocin as a selective maker in both *E. coli* and *P. pastoris*. One of the primary advantages of using zeocin as a selective marker is that using only a single selective marker for both *E. coli* and *P. pastoris* results in a significantly smaller size of vector (3.6 kb for pPICZα compared to 8.0 kb for pPIC9). The resulting smaller vector size leads to more straightforward downstream DNA manipulations. Moreover, using zeocin as a selective agent in yeast provides the ability to directly select the multicopy integration events that are often associated with higher levels of expression (Vassileva et al., 2001). Taken in concert these attributes makes zeocin an ideal selective marker for *P. pastoris* expression. For these reasons pPICZα was

chosen as the vector for the expression of MCP2 rather than the pPIC9 vector (which was used for the expression of MCP12).

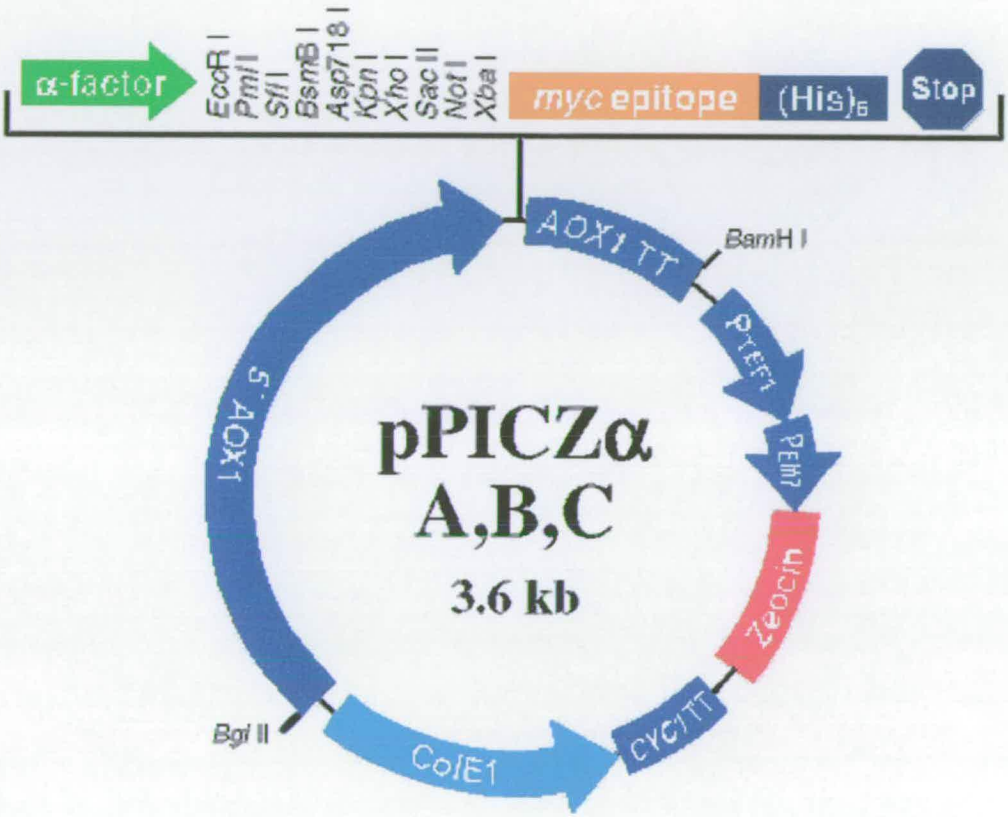


Figure 4:1 The pPICZα expression vector. The features of this vector are similar to pPIC9 but with the use of zeocin selection in both *E. coli* and *P. pastoris* rather than ampicillin and histidine. This results in a smaller vector size, fewer false positives, and the ability to directly select multicopy integration events (taken from Invitrogen, 2002b).

As before, PCR was used to amplify the region coding for MCP2, and asymmetric restriction sites (5' *Eco*R1 & 3' *Not*1) were incorporated into the primers to facillitate directional cloning. The PCR reaction showin in Figure 4:2 shows a single band of approximately 250 bp, which corresponds well with the predicted size of 233 bp for the MCP2 fragment (allowing for all extra sequences included for restriction sites and stop codons). This PCR product was digested with *Eco*R1 and *Not*1 and was

subsequently ligated into pPICZα using a similar strategy to that used for ligating MCP12 into pPIC9.

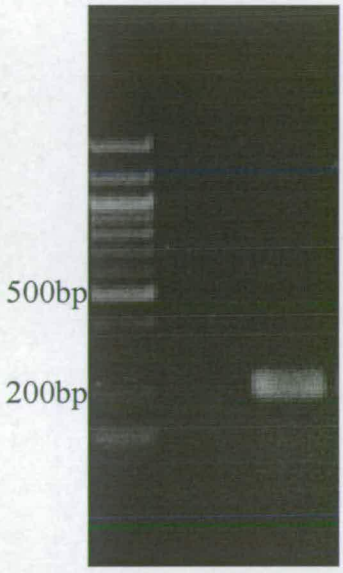


Figure 4:2 PCR amplification of MCP2, showing a single band of the correct size for MCP2

The recombinant plasmid thus genereated was then used to transform chemically competent Top10 *E. coli* cells. The transformation of the *E. coli* cells produced numerous colonies, and plasmids from ten clones that grew on the selective medium (low salt LB agar containing 25 $\mu g.ml^{-1}$ zeocin) were prepared by miniprep and digested with *Xho*1 and *Xba*1. The reason why plasmids were digested with *Xho*1 and *Xba*1, whose restriction sites flank the region containing the insert, rather than the restriction enzymes incorporated into the PCR product (*Eco*R1 and *Not*1), was to allow concatamers to be detected and rejected. A clone that gave an insert of the predicted size was selected, and its DNA sequenced in order to confirm the identity and integrity of the insert.

After confirming that the DNA sequence of the insert was correct this recombinant plasmid was prepared on a larger scale by a midiprep procedure. The plasmid was subsequently linearised with *Sac*1 and used to transfrom the *P. pastoris* KM71-His strain. The KM71-His strain was genetically modified in-house by transforming the original KM71 strain with 'empty' pPIC9 vector, in the AOX1 region, to convert the His^- phenotype of KM71 to a His^+ phenotype (data not shown). This modified strain had been used previously in our laboratory for the expression of several proteins, and had been shown to behave as expected, *i.e.* display the predicted His^+ Mut^S phenotype (data not shown). This strain was chosen for the initial expression attempts because of the success that was achieved using the KM71 strain for the expression of MCP12, and because, it negated the requirement to supplement the growth and induction media with histidine. The latter point resolves the potential future problems associated with subsequent isotopic enrichment. This cloning strategy would be expected to yield recombinant *P. pastoris* with a His^+ Mut^S phenotype.

The transformation of KM71-His with the *Sac*1 linearised MCP2-pPICZα yielded numerous colonies that could grow on the selective medium (YPDS + 100 $\mu g.ml^{-1}$ zeocin). In the absence of an assay for MCP2, or antibodies for this module, screening of the colonies was performed by using small-scale test inductions followed by SDS-PAGE. Five colonies were picked, grown and induced for 4 days, deglycosylated, and screened for expression of MCP2 by SDS-PAGE. The results shown in Figure 4:3 indicate a good level of expression that is better than was achieved with MCP12. The gel also shows that the level of expression of all clones

with the exception of clone 1 is approximately the same to within experimental error, and for this reason clone 4 was chosen at random and used in further analysis. On the basis of the previous observation that MCP12 had been hyperglycosylated by *P. pastoris*, the samples of MCP2 were deglycosylated prior to SDS-PAGE. Subsequent TCA precipitation of the culture supernatant is inefficient, and only a faint smear could be observed on an SDS-PAGE gel.

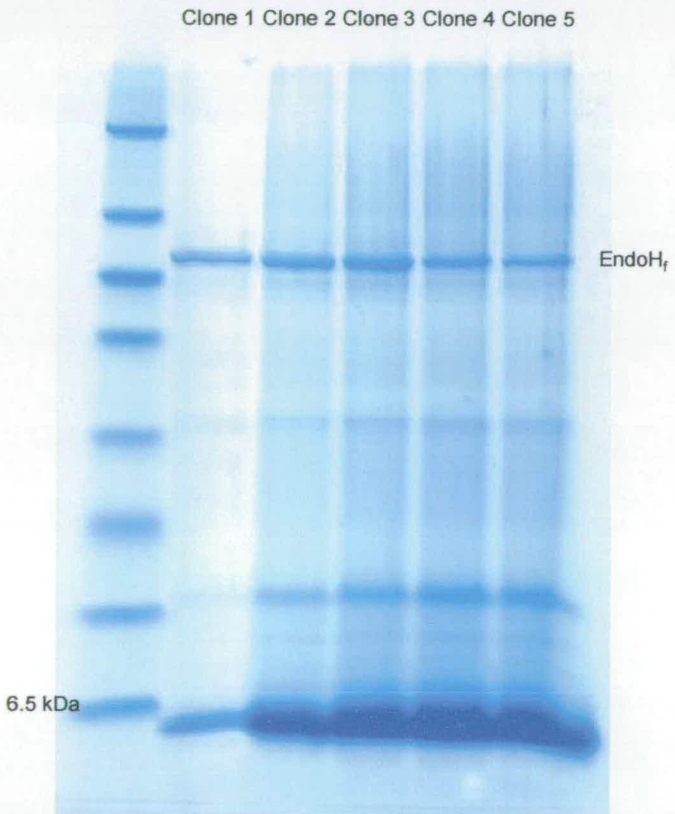


Figure 4:3 SDS-PAGE showing deglycosylated MCP2 test inductions. All of the clones tested express a species of approximately the correct size for MCP2. Moreover, with the exception of clone 1, the clones appear to express this species to approximately the same level.

In order to determine whether MCP2 is in fact been glycosylated by *P. pastoris*, both untreated and deglycosylated samples were subjected to SDS-PAGE. In order to

circumvent the TCA precipitation step, 4 ml of induced culture supernatant was concentrated down to 40 µl using a centrifugal concentrator (5 kDa MWCO). The sample was exchanged into water by three 10 fold dilutions and concentrations. Of this, 10 µl was deglycosylated, and the remainder mock deglycosylated by incubation at 37 °C for the same length of time as the deglycosylated sample. The result of the SDS-PAGE for the glycosylated and deglycosylated samples of MCP2 is shown in Figure 4:4.

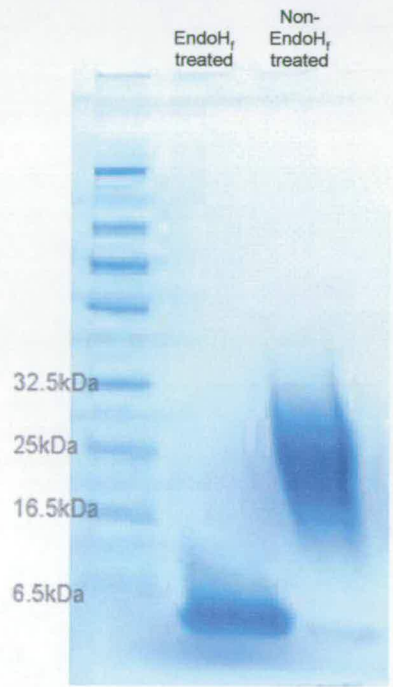


Figure 4:4 SDS PAGE showing the effect of deglycosylating the MCP2 expressed in *P. pastoris* with EndoHf. The glycosylated sample shows a large smear that extends between 16 and 30 kDa. This is in contrast to the deglycosylated sample, which runs as a single band slightly lower than would be expected.

Three times more of the glycosylated MCP2 was loaded onto SDS-PAGE to compensate for its high degree of heterogeneity. This SDS-PAGE result shown in Figure 4:4 shows that the sample produced by *P. pastoris*, is both hyperglycosylated

and highly heterogeneous. This result is not surprising given the experience with MCP12 described in the previous chapter. This result also demonstrates that the N-linked glycans can be cleaved by $EndoH_f$. Furthermore this observation implies that, as is the case with MCP12, the glycosylation of recombinant MCP2 did not comprise complex type glycans, since $EndoH_f$ is unable to cleave complex N-linked glycans.

4.3. Preparative expression and purification of *P. pastoris* MCP2

4.3.1. Large scale growth and expression of MCP2 *P. pastoris*

In order to obtain sufficient protein for subsequent studies, the expression of MCP2 was scaled up in a similar way to that used for the scale up of MCP12. Growth and induction of MCP2 was performed in buffered minimal media. This was done to with the requirement for future isotopic labelling of the recombinant protein module in mind. Typically the initial growth phase was carried out in BMG (2 l) at 29 °C until A_{600} ≈16. The recombinant yeast were then induced by harvesting the cells by centrifugation and resuspending them in BMM (1 l). The cells were incubated in this inducing medium for a total of 5 days at 29 °C and were fed daily with undiluted methanol (1 % v/v). Optimisation of the length of induction was not performed in the case of MCP2, since it seemed likely that it would have similar properties to MCP12 as far as proteolysis was concerned. The temperature for growth and induction was set to 29 °C since it has been demonstrated that expression is greatly inhibited by growth at temperatures significantly above 30 °C and therefore to compensate for possible inaccuracies in the incubator's temperature control.

Moreover it has been demonstrated that whilst growing on methanol at high cell densities, *P. pastoris* produces a significant amount of heat which may result in the temperature of the culture rising above that set by the incubator. During both the growth and induction phases, the culture was agitated vigorously to achieve the high levels of oxygenation required for efficient expression of heterologous proteins. In the induction phase, however, the culture volume in each flask was reduced to 250 ml from 500 ml per 2 l conical flask, to compensate for the culture's increased requirement for oxygen. Following the induction phase, the supernatant was harvested by centrifugation, and PMSF (to 0.5 mM), and EDTA (to 5 mM) added to reduce the likelihood of proteolysis. A sample of the harvested culture supernatant (1 ml) was deglycosylated with $EndoH_f$, TCA precipitated and subjected to SDS-PAGE to determine whether the scale-up had been successful. The resulting band on the SDS-PAGE gel was of a higher intensity than that obtained from the test inductions, presumably because the more favourable growth conditions had resulted in higher levels of expression. Following harvest, all subsequent manipulations prior to purification were carried out either on ice or at 4 °C in a cold room. The culture supernatant was filter sterilised in order to remove any remaining yeast cells, which could subsequently act as a source of proteases. The now clarified culture supernatant was then concentrated to a final volume of 20 ml using ultra filtration (3 kDa MWCO), after which time a further dose of PMSF (0.5 mM) was added. This second dose is required because PMSF is unstable in even slightly acidic aqueous environments. This concentrate was subsequently purified as is described in the following section.

The production of the ^{15}N samples of MCP2 was achieved in a similar way to that described above, with the sole exception that both the growth and expression phases were carried out on media containing 0.2 % (w/v) ^{15}N ammonium sulphate as the sole nitrogen source. These media were prepared by using YNB with neither amino acids nor ammonium sulphate, and then supplementing the media with 2 $g.l^{-1}$ ^{15}N ammonium sulphate. This protocol resulted in the production of MCP2 samples with greater than 90 % incorporation of ^{15}N in the final protein (see section 4.4.1.1 for details).

4.3.2. Initial purification of MCP2

Hydrophobic interaction chromatography on a Resource Iso column was selected as the initial purification step in the purification of MCP2, because this method had proved successful for the purification of MCP12. Once again, the high level of heterogeneity of the MCP2 sample due to the different glycoforms present, meant that the initial purification of MCP2 gave a broad peak covering several fractions – numbers 16-19 in the chromatogram shown in Figure 4:5-A. This chromatogram shows a large unbound fraction that contains only traces of very highly glycosylated MCP2, and a single broad peak of the bound material. SDS-PAGE of the non-deglycosylated fractions from the HIC purification (Figure 4:5-B) shows that the more highly glycosylated forms of MCP2 elute earlier from the column, although each fraction still remains highly heterogeneous. Following deglycosylation with $EndoH_f$, each of the bound fractions gave predominantly a band of the same size on SDS-PAGE, as shown in Figure 4:5-C. It can be seen from these SDS-PAGE gels

that this purifications step effectively removes significant quantities of contaminating proteins. Moreover, it effectively removes almost all the contaminating species that turns the yeast medium a deep green colour. Whilst this cannot be observed easily by looking at the SDS-PAGE gels, it can clearly be seen when one compares the colour of the unbound fractions with the bound fractions.

4.3.3. Purification of deglycosylated MCP2

As with the recombinant MCP12 sample, the pooled fractions corresponding to the central portion of the main peak from the HIC chromatography of MCP2 (fractions 16-18) were concentrated, and deglycosylated by treatment with $EndoH_f$. Based upon the success obtained with MCP12, MCP2 was also purified, following deglycosylation, by affinity chromatography using ConA sepharose. This step removes any non-deglycosylated MCP2 along with the contaminating cleaved glycans. As with MCP12, the deglycosylated MCP2 passed straight through the column in the unbound fractions, whereas the cleaved glycans and any MCP2 still glycosylated, bound to the column, and was subsequently eluted with 1 M methyl-α-D-glucopyranose. The result of this purification step is illustrated in Figure 4:6.

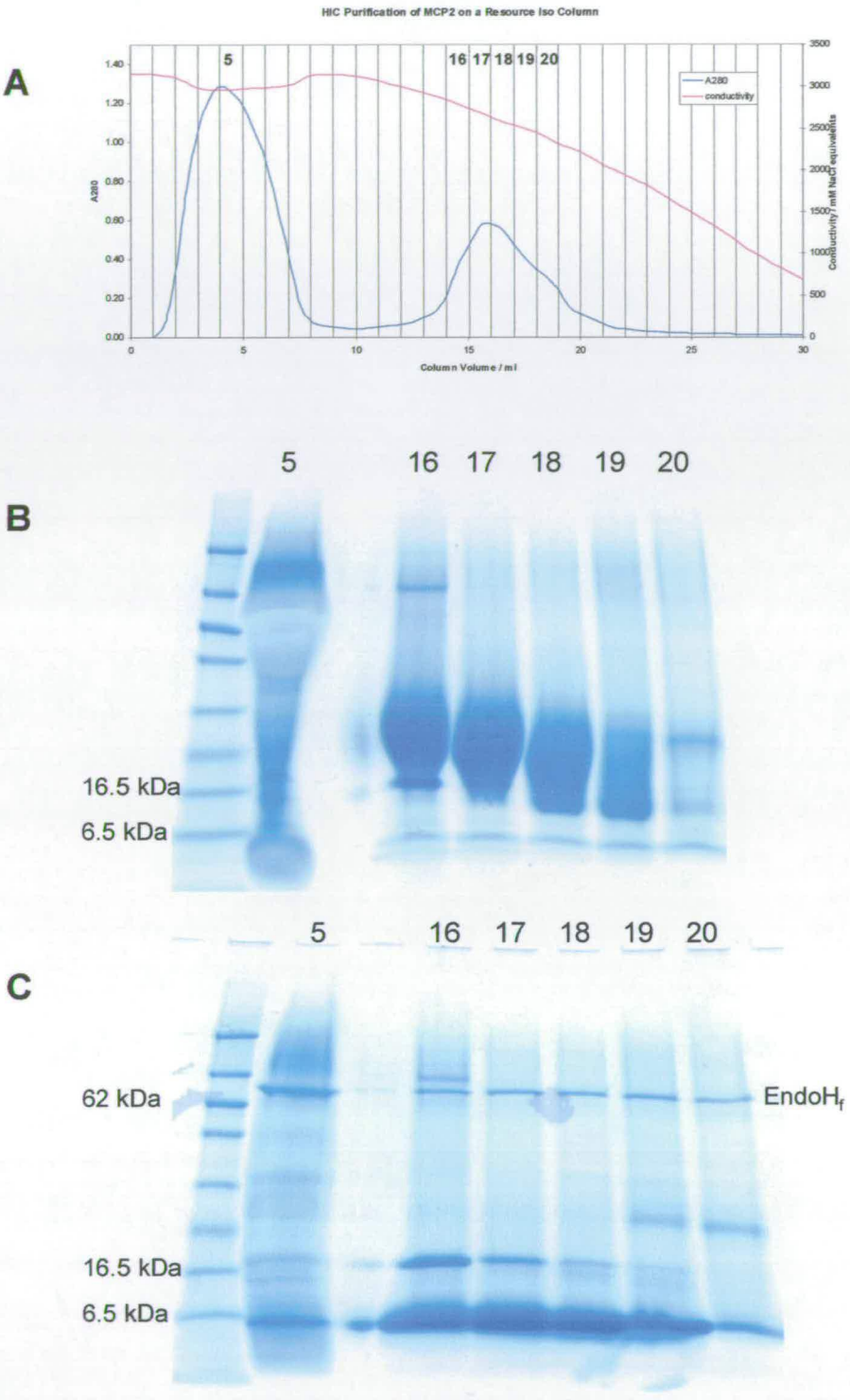


Figure 4:5 HIC purification of crude MCP2 on a Resource Iso column. A: chromatogram showing a large unbound fraction and single broad peak of bound material. B: SDS PAGE of non-deglycosylated fractions. C: SDS PAGE of non-deglycosylated fractions. The gels show than the majority of the bound material is glycosylated MCP2 and that more highly glycosylated species elute from the column earlier.

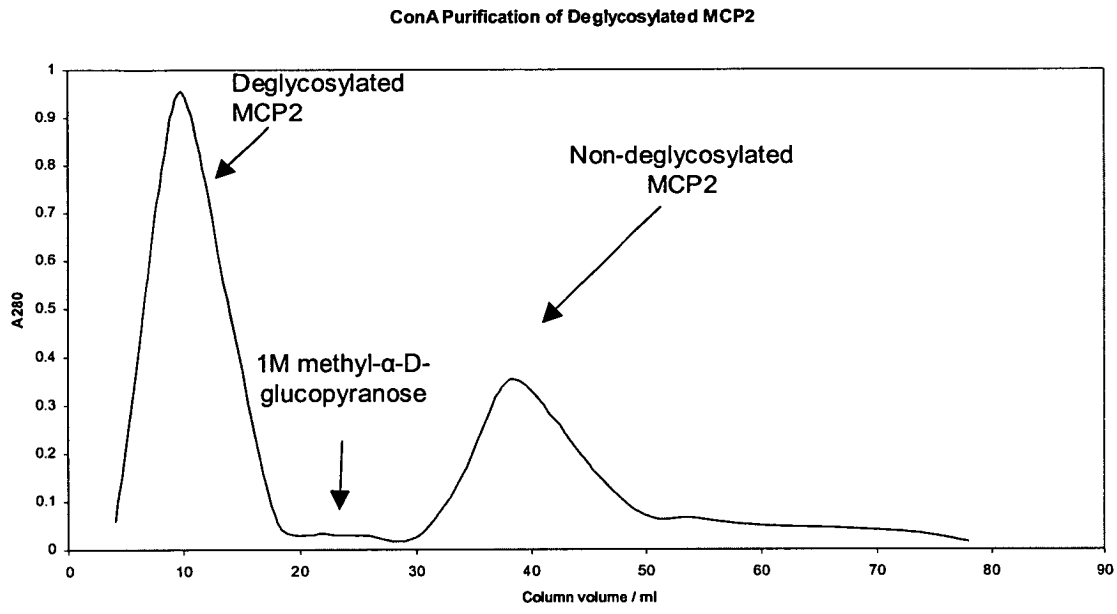


Figure 4:6 ConA purification of deglycosylated MCP2. Deglycosylated MCP2 remains unbound, whereas the cleaved glycans and any non-deglycosylated MCP2 that remained were retained on the column.

Following ConA purification of deglycosylated MCP2 the breakthrough fractions from the ConA column were pooled, concentrated and further purified using anion exchange chromatography on a Poros20Q (Applied Biosystems) column The MCP2 was exchanged into 50 mM Tris, pH 8.0, using a PD10 gravity-fed gel filtration column (Amersham-Pharmacia Biotech), and applied to the Poros20Q column in aliquots of approximately 1 mg. The chromatogram shown in Figure 4:7 shows that the deglycosylated MCP2 elutes from the column as a single peak indicative of a homogeneous protein. Fractions 18 and 19 from this column, that represented the central portion of the single peak, were pooled and used for subsequent analyses.

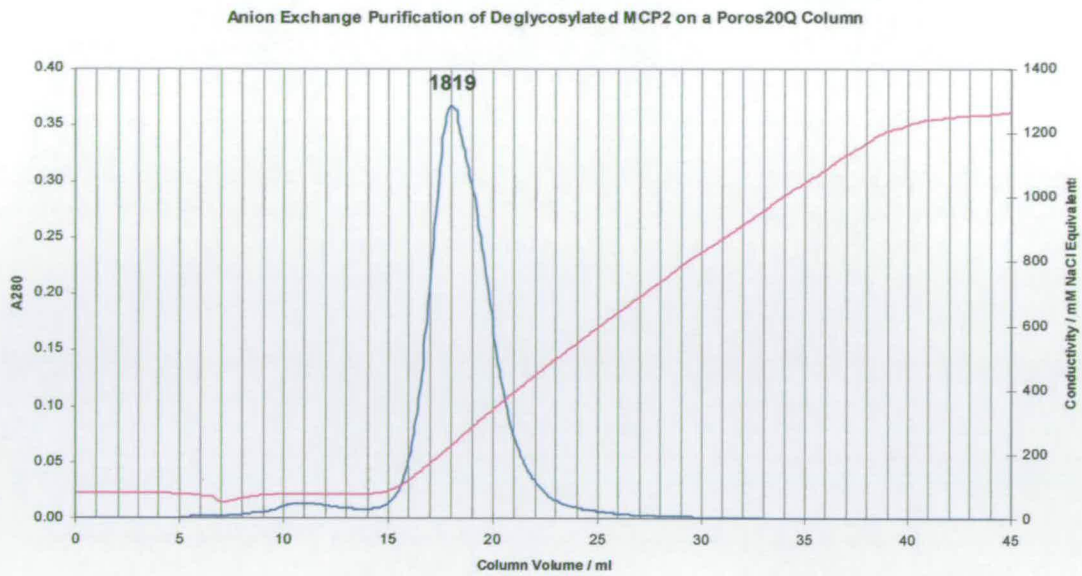


Figure 4:7 Anion exchange purification of deglycosylated MCP2 on a Poros20Q column. The chromatogram shows that deglycosylated MCP2 elutes as a single peak.

Figure 4:8 SDS-PAGE of purified deglycosylated MCP2. The gel shows that deglycosylated MCP2 has been purified to homogeneity as judged by SDS-PAGE.

The SDS-PAGE of the pooled fractions, as shown in Figure 4:8, demonstrates the purification of deglycosylated MCP2 to homogeneity as judged by this method. The identity of the protein sample was further confirmed using N-terminal sequencing. This shows the presence of extra N-terminal EA sequences. These residues are occasionally present at the beginning of proteins secreted from *P. pastoris*, and are the result of inefficient cleavage of the secretion signal. A secondary sequence lacking extra E and A residues was also observed by N-terminal sequencing. These N-terminal sequences were determined to be: **EAEFYRETCP** and **EFYRETCPYI.** With the native MCP2 sequence starting YRET...

4.3.4. Purification of glycosylated MCP2 expressed in *P. pastoris*

In addition to the production of a deglycosylated sample of MCP2, a glycosylated sample was required in order to compare the behaviour of glycosylated and deglycosylated MCP2 with a view to better understanding the effects of the N-linked glycans. Purification of glycosylated MCP2 had many of the problems associated with heterogeneous glycosylation that were observed during the purification of glycosylated MCP12. The purification of the glycosylated sample of MCP2 was achieved in a similar way, namely an initial HIC purification step, followed by reverse phase chromatography. The chromatogram obtained from reverse phase chromatography of glycosylated MCP2 is shown in Figure 4:9, and demonstrates that much of the glycosylated module elutes as a single peak similar to the one for glycosylated MCP12. Demonstrating the purity of the highly heterogeneous glycosylated form of MCP2 is difficult, simply by virtue of its inherent

heterogeneity. In this case it seems reasonable to compare SDS-PAGE of the same sample both before and after treatment with EndoH$_f$. For the untreated sample, this method would be expected to yield the heterogeneous smear previously observed. On the other hand, for the treated sample two bands would be expected, one running at approximately 70 kDa corresponding to the EndoH$_f$ used for deglycosylation, and one at ~ 6 kDa which corresponds to deglycosylated MCP2. This expected profile was indeed observed on the SDS-PAGE shown in Figure 4:10

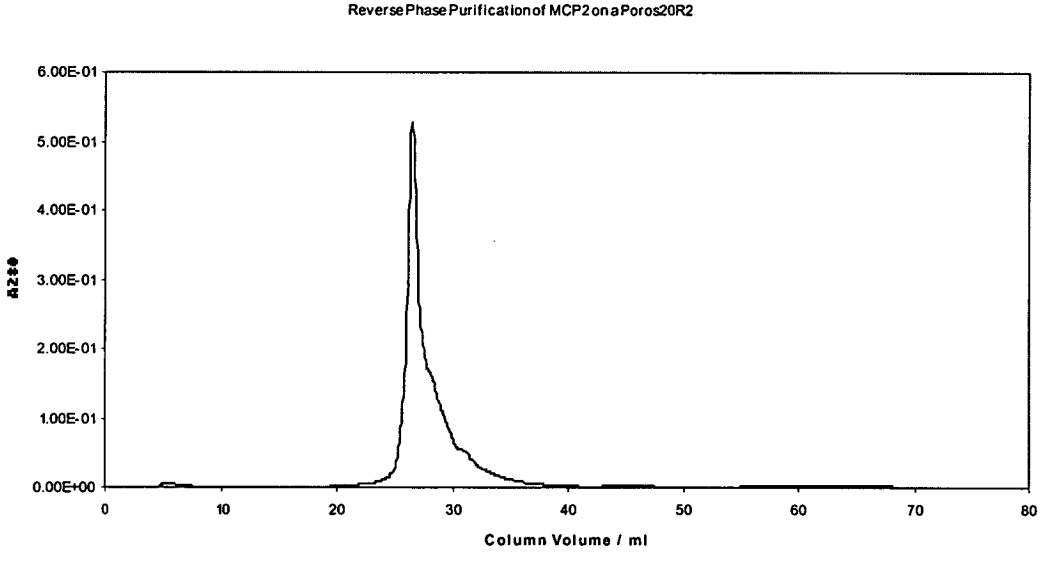


Figure 4:9 Reverse phase purification of glycosylated MCP2. The chromatogram shows that glycosylated MCP2 elutes as a single peak with a slight shoulder, which is unsurprising given the heterogeneity of glycoforms observed.

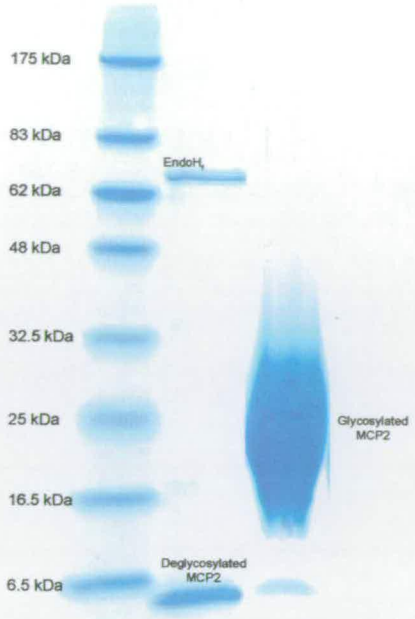


Figure 4:10 SDS-PAGE of purified glycosylated MCP2 after treatment with EndoH$_f$. The gel of the EndoH$_f$ treated sample shows two bands, one runs around the 6.5 kDa marker as would be expected for deglycosylated MCP2 and the other runs at ~70 kDa which corresponds to the expected mass of the EndoH$_f$. The untreated sample shows a highly heterogeneous smear extending primarily between 16 kDa and 32.5 kDa.

The proteins purified as described in this section were then used for subsequent analysis as described in the following section.

4.4. Biophysical Characterisation of MCP2 expressed in P. pastoris

4.4.1. Mass spectrometry analysis of MCP2

In the absence of an assay, evidence was sought to confirm that the species expressed and subsequently purified was indeed chemically intact and correctly folded MCP2. Plasmid DNA sequencing and N-terminal protein sequencing were encouraging but further validation was required. This was achieved using electrospray ionisation

(ES)-MS; the mass spectrum of MCP2 deglycosylated with $EndoH_f$ is shown in Figure 4:11. It reveals two major species that are within 1 Da (lower) of the expected mass of MCP2 with its two disulphide bonds formed, the single remaining GlcNAc residue following $EndoH_f$ treatment, and an N-terminus of either EF... or EAEF.... Each of these two major peaks has two associated minor peaks corresponding to masses +23 Da and +38 Da. It is most likely that these arise from the presence of sodium and potassium adducts respectively. In addition each of the two major peaks is associated with a -18 Da species. Modifications that could explain this difference include succinimide formation from aspartic acid residues, dehydration, and formation of dehydroalanine from serine. Regardless of which of these possibilities is correct, the fraction of the sample that contains this modification is small. It proved impossible to perform ES-MS on the non-deglycosylated form of MCP2 proved due to the high heterogeneity resulting from the presence of different glycoforms.

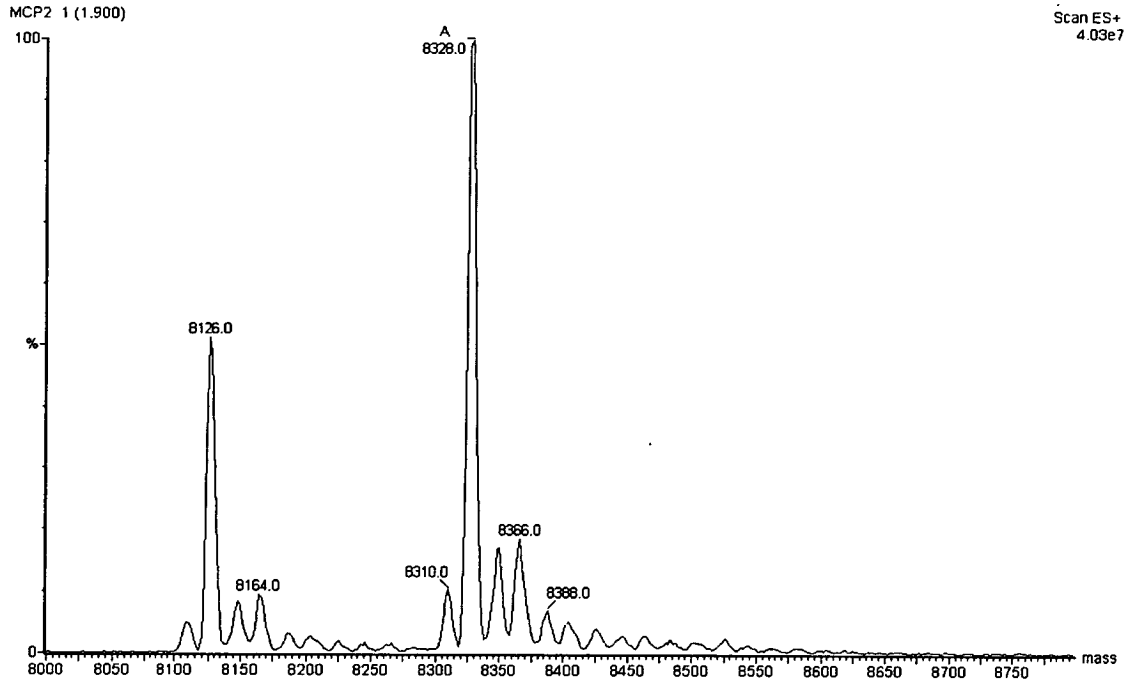


Figure 4:11 Mass spectrum of EndoH$_f$ deglycosylated MCP2. The two species at 8127 kDa and 8327 kDa are within 1 Da of the predicted mass for the protein with two disulphide bonds, a single GlcNAc residue, and N-terminal residues of EF and EAEF respectively. This is the result of incomplete processing of the secretion signal and was also observed with N-terminal sequencing.

4.4.1.1. Verification of ^{15}N incorporation into the ^{15}N deglycosylated MCP2

A sample of each of the glycoforms of MCP2 was uniformly labelled with ^{15}N by growing up the *P. pastoris* cells with ^{15}N ammonium sulphate as the sole nitrogen source, as described elsewhere (Wiles, 1996). ES-MS was used to verify that the MCP2 had been ^{15}N-labelled and to check the level of ^{15}N incorporation. The mass spectrum shown in Figure 4:12 once again shows two main peaks – presumably corresponding to the different N-termini.

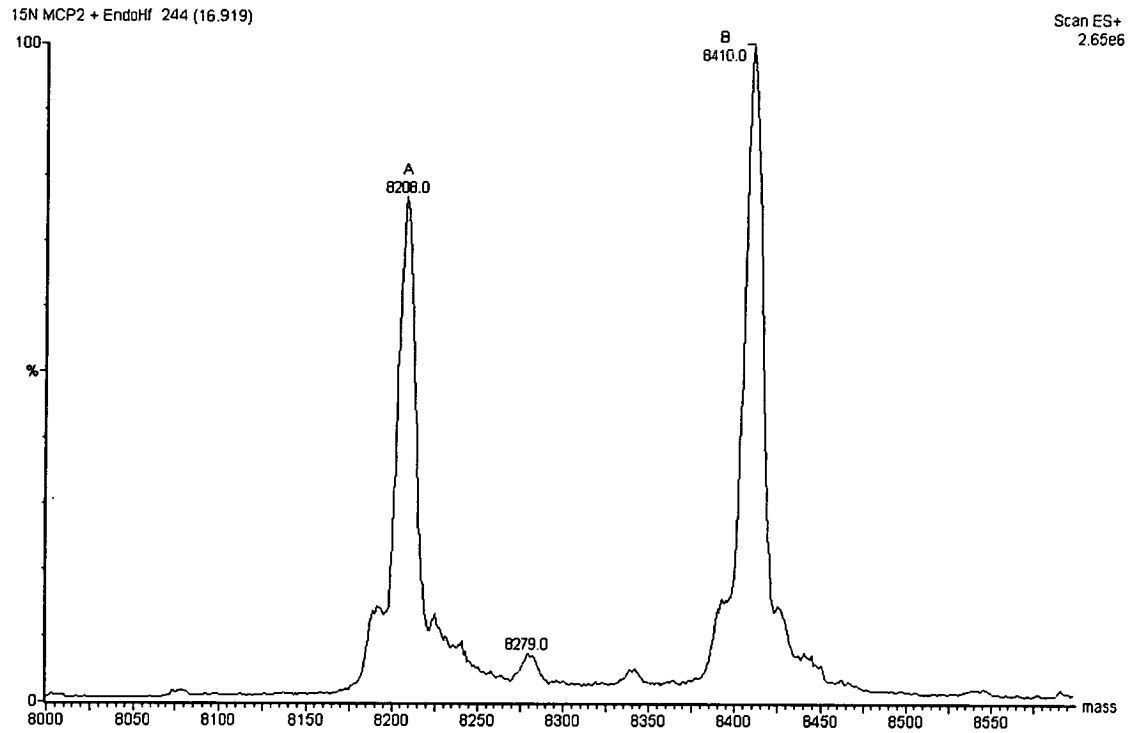


Figure 4:12 Mass spectrum of ^{15}N labelled EndoHf deglycosylated MCP2. The peaks correspond to masses for the major species which are 82 Da greater than the unlabelled equivalents. This indicates that 82 out of the possible 89 nitrogen atoms have ^{15}N nuclei.

The peak corresponding to the major species (N-terminus EAEF...) is 82 Da larger than its counterpart in the unlabelled sample. This mass difference presumably represents the incorporation of 82 ^{15}N atoms out of a possible 89 nitrogen atoms per MCP2 molecule, which would correspond to ~ 92 % incorporation of ^{15}N. The peak corresponding to the minor species (N-terminus EF...) is also 82 Da greater than its unlabelled counterpart. Thus82 out of 87 ^{14}N nuclei have been replaced by ^{15}N nuclei in this case corresponding to ~ 94 % ^{15}N enrichment and consistent within experimental error with the percentage incorporation calculated for the larger species.

4.4.1.2. Verification of the number of disulphide bonds in MCP2

In order to verify that both disulphide bonds had formed the protein was subjected to carboxymethylation using neutralised iodoacetic acid. This reagent would be expected to react with any free sulfhydryls present in the protein. To reduce the ambiguity resulting from the presence of the single GlcNAc residue remaining after $EndoH_f$ treatment, this procedure was performed on PNGaseF-deglycosylated MCP2. Like $EndoH_f$, PNGaseF is an endoglycosidase that cleaves N-linked glycans. Unlike $EndoH_f$, PNGaseF cleaves glycans between the first GlcNAc residue and the asparagine residue of the NXS/T glycosylation sequon. This type of cleavage therefore leaves no sugar residues on the protein, which may be useful when attempting to interpret a mass spectrum. The use of PNGaseF-deglycosylated MCP2 also permits a less ambiguous confirmation that the two major species observed (Figure 4:11) are due to the ragged N-terminus (expected difference 200 Da), rather than the presence or absence of a single GlcNAc residue (predicted mass difference 203 Da).

The PNGaseF-treated sample was subjected to ES-MS (i) prior to any further treatment; (ii) following carboxymethylation; and (iii) following denaturation with urea, reduction with DTT and carboxymethylation. The inclusion of the denatured, reduced, and alkylated sample acts as a positive control to prove that the carboxymethylation reaction progresses as would be expected. The purpose of the denaturing agents was to disrupt any tertiary structure that could bury free –SH groups.

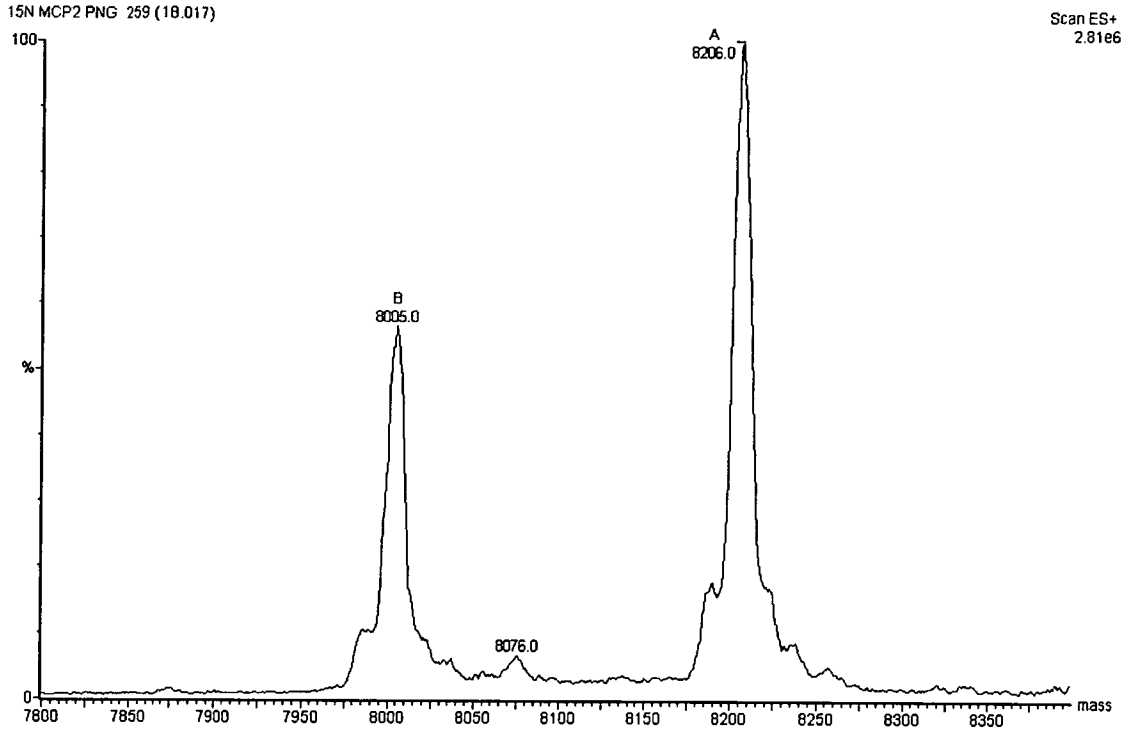


Figure 4:13 Mass spectrum of ^{15}N-labelled PNGaseF-deglycosylated MCP2. The spectrum shows the expected two peaks that correspond to the mass of MCP2 with two disulphide bonds, no GlcNAc residue and the ragged N-terminus (± EA).

The mass spectrum in Figure 4:13 shows two peaks that are within 1 Da of the predicted mass of deglycosylated MCP2 containing two disulphide bonds, the now expected ragged N-terminus, and minus the GlcNAc residue not removed by EndoH$_f$, but cleaved by PNGaseF. These masses can be compared to those obtained with the EndoH$_f$-treated sample. The mass of MCP2 after treatment with PNGaseF is 204 Da lower than that observed for EndoHf-deglycosylated MCP2. This mass difference corresponds to the loss of a ^{15}N-abelled GlcNAc residue (expected mass difference 204 Da). Therefore, these data provide confirmation that the two peaks are indeed due to the presence of the ragged N-terminus, rather than the presence of a proportion of MCP2 that had never been glycosylated.

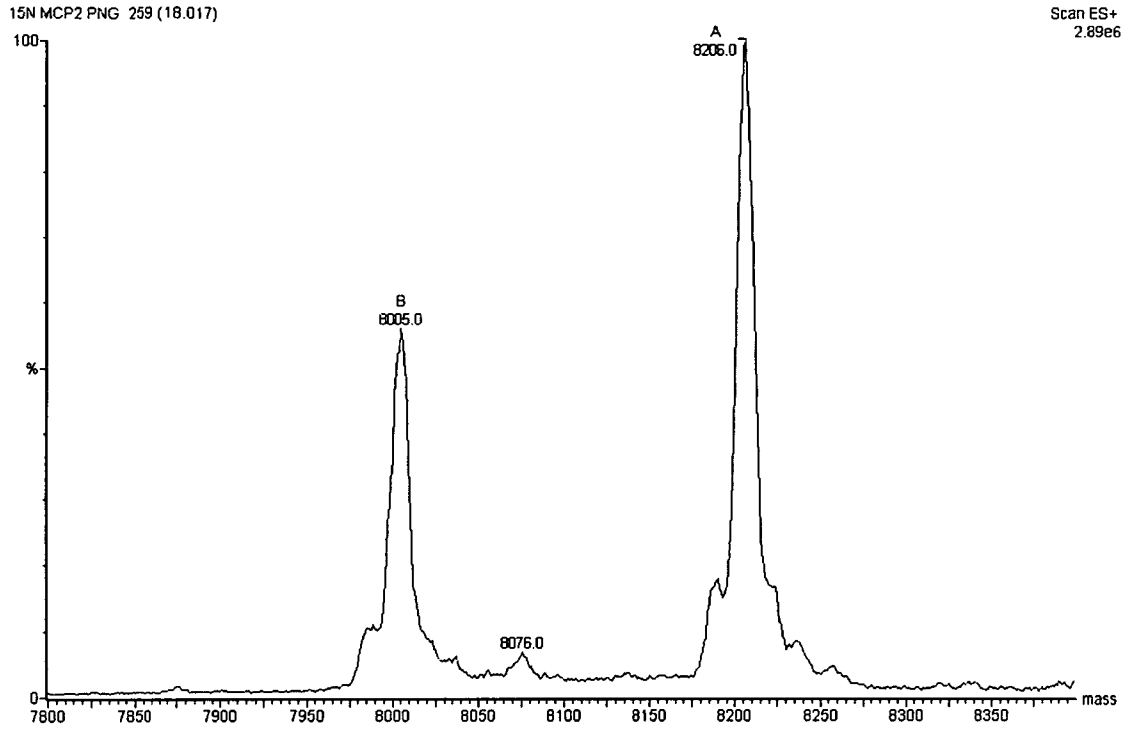


Figure 4:14 Mass spectrum of carboxymethylated ^{15}N labelled MCP2 deglycosylated with PNGaseF.

The mass spectrum of MCP2, treated with neutralised iodoacetic acid, shown in Figure 4:14 reveals no change in mass when compared to the non-carboxymethylated sample (Figure 4:13), indicating that all four cysteines are oxidised, and therefore presumably involved in disulphide bonds. The mass spectrum of denatured, reduced, and alkylated MCP2 shown in Figure 4:15 shows an increase in mass of 235 Da and 236 Da for the smaller and larger peak respectively. This increased mass is in good agreement with the prediction of 232 Da expected for the carboxymethylation of 4 cysteines. Each of the carboxymethylation reactions is predicted to add 58 Da for each free cysteine. It appears, however, that the observed mass is 3 Da larger than would be predicted. This is probably the result of deamidation of the three asparagine residues in the protein. This is not entirely surprising since the reduction reaction is carried out at high pH, and at an elevated

temperature, making deamidation more likely. When taken in concert these data indicate that MCP2 has both its predicted disulphide bonds formed, and that these are reduced following denaturation and treatment with DTT. That two disulphides are present in MCP2, in line with expectations, provides further evidence that the core of the protein may indeed be folded as would be expected.

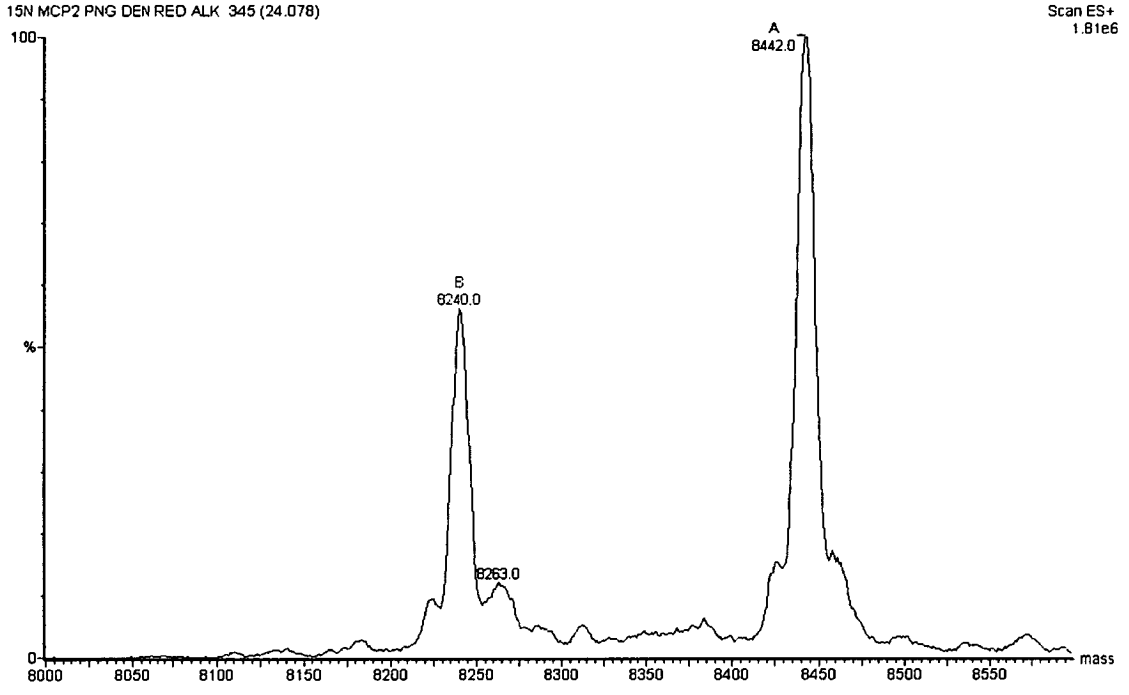


Figure 4:15 Mass spectrum of denatured, reduced, and carboxymethylated PNGaseF-deglycosylated, ^{15}N-labelled MCP2. The spectrum shows an increase in mass of 235 and 236 when compared to non-alkylated MCP2.

The results described in the preceding sections demonstrate the successful expression and purification of MCP2, and subsequent verification that the isolated species contains its predicted two disulphide bonds. These data further indicate that the core of the protein should be folded in a similar way to that found naturally. Following these observations, the protein was subjected to NMR spectroscopy in order to gain further information and therefore a deeper understanding of its structure and

dynamics. The results of the NMR spectroscopy carried out on this protein will be the subject of the following section.

4.4.2. NMR spectroscopic analysis of MCP2

As previously discussed high field NMR can provide a wealth of information about a protein's structure and dynamics. Following the purification of a sample of $EndoH_f$-deglycosylated MCP2, a 1 D ^1H spectrum was acquired at 25 °C under the conditions optimised for MCP12. This is shown in Figure 4:16.

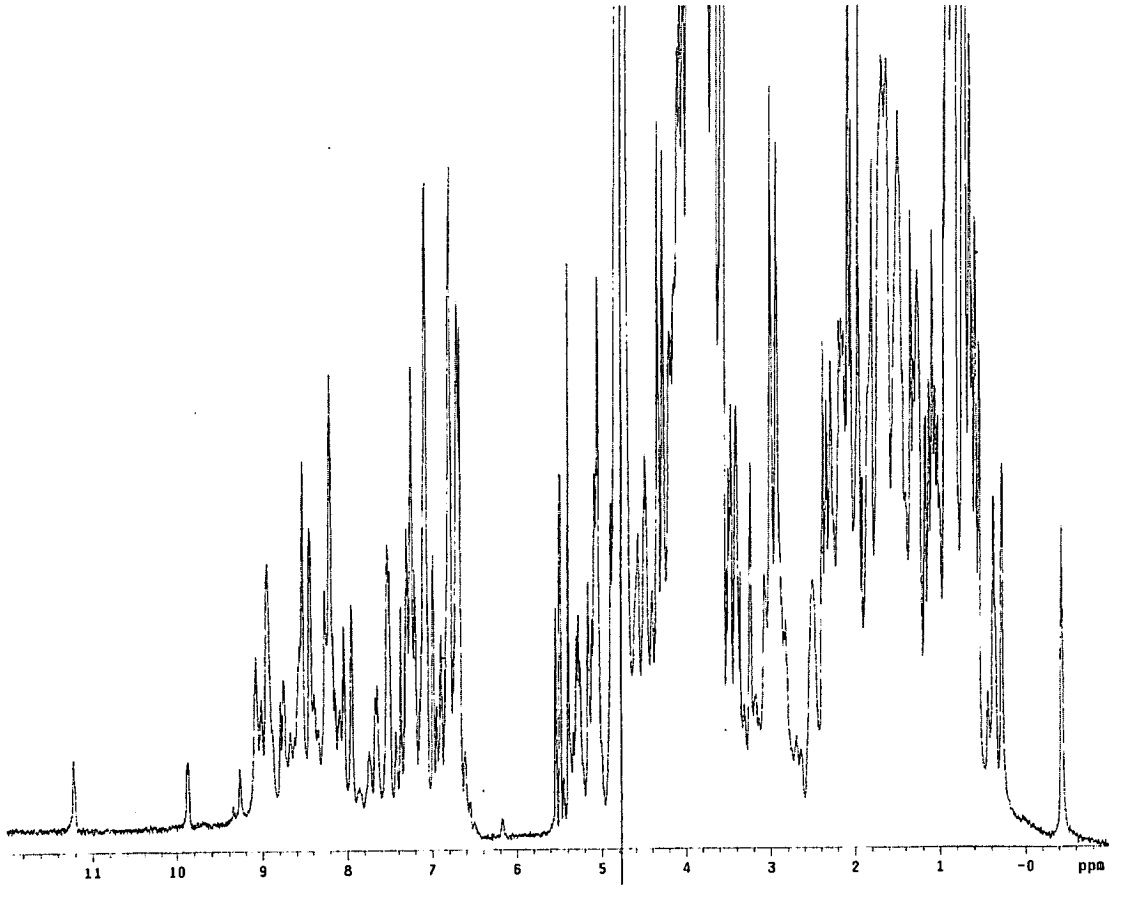


Figure 4:16 1D ^1H spectrum of $EndoH_f$ deglycosylated MCP2. The spectrum was acquired at 25 °C. The sample was 0.9 mM in 20 mM potassium phosphate pH 6.0.

The spectrum shows good dispersion, and narrow line-widths, both of which are associated with folded proteins. A spectrum of the same sample was then acquired at 37 °C as shown in Figure 4:17.

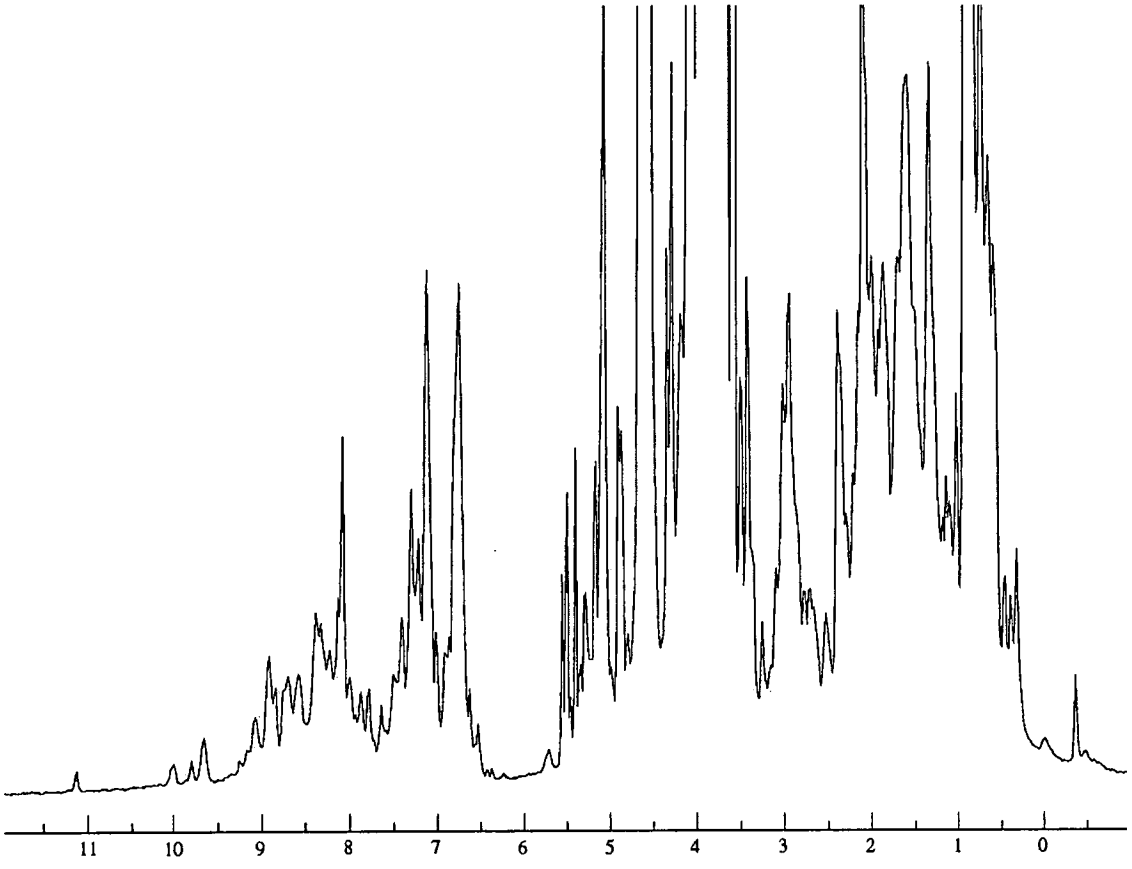


Figure 4:17 1D ^{1}H spectrum of EndoH$_{f}$-deglycosylated MCP2. The spectrum was acquired at 37 °C. The sample was 0.9 mM in 20 mM potassium phosphate pH 6.0.

The spectrum acquired at 37 °C shows several significant differences compared to that acquired at 25 °C. These include the appearance of broad peaks at 10.1 ppm, 9.7 ppm, 5.7 ppm, 0.1 ppm, and -0.5 ppm. Such changes could result from temperature-sensitive differences in the rates of inter-conversion between multiple conformers. For example an intermediate exchange rate that causes extreme broadening (effective

disappearance) of a signal is likely to become more rapid at a higher temperature resulting in a reappearance of the signal in the spectrum. To explore this further, 1 D ^{1}H spectra were also acquired at 42 °C and 10 °C (as shown in Figure 4:18 and Figure 4:19 respectively). Both of these spectra appear to exhibit the same trend, with the broad peaks becoming more prominent as temperature is increased.

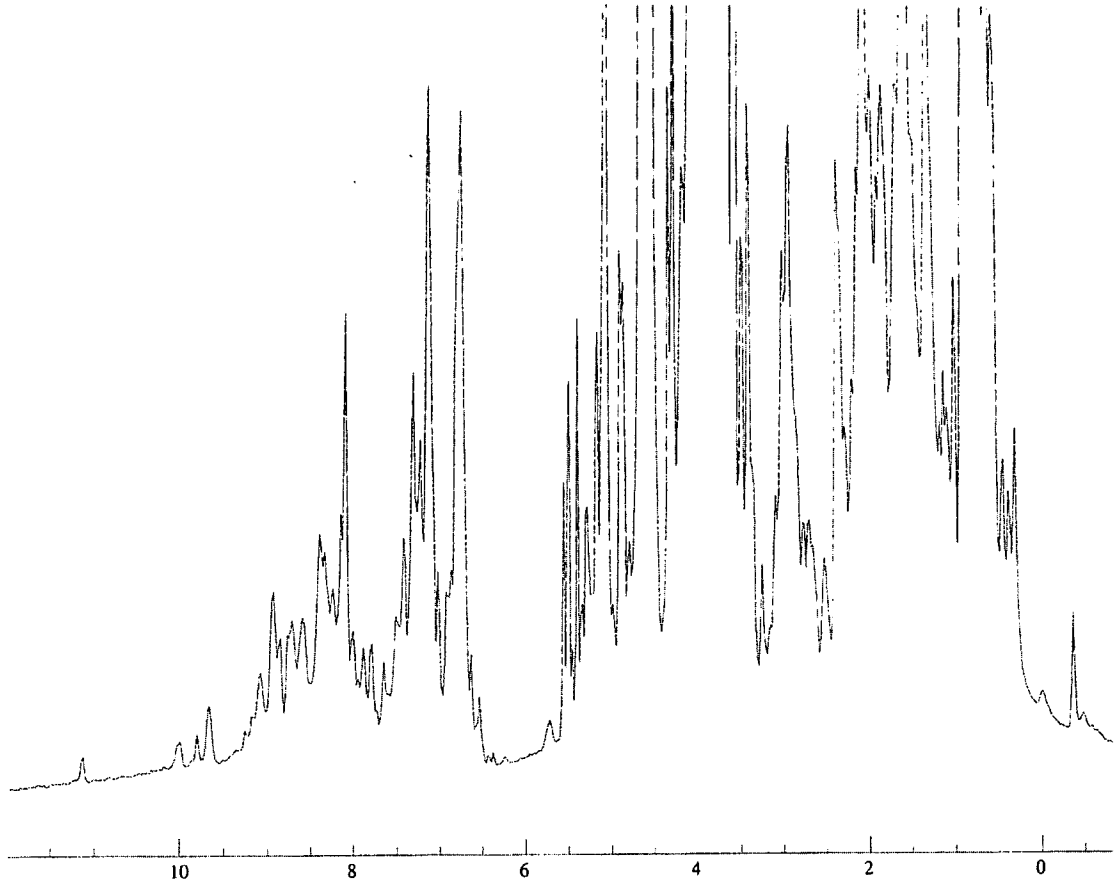


Figure 4:18 1D ^{1}H spectrum of EndoH$_f$-deglycosylated MCP2. The spectrum was acquired at 42 °C. The sample was 0.9 mM in 20 mM potassium phosphate, pH 6.0.

It is interesting to note that the broad peaks eventually disappeared completely following prolonged (>2 weeks) incubation in the fridge. It therefore seems that the fresh sample of deglycosylated MCP2 was not fully folded – but over time (at 4° C)

the folding process reached completion. A possible reason for very slow folding of this protein could be the presence of several X-Pro bonds that might settle into a preferred *cis* or *trans* configuration over an extended time. All subsequent spectra were acquired at 25 °C.

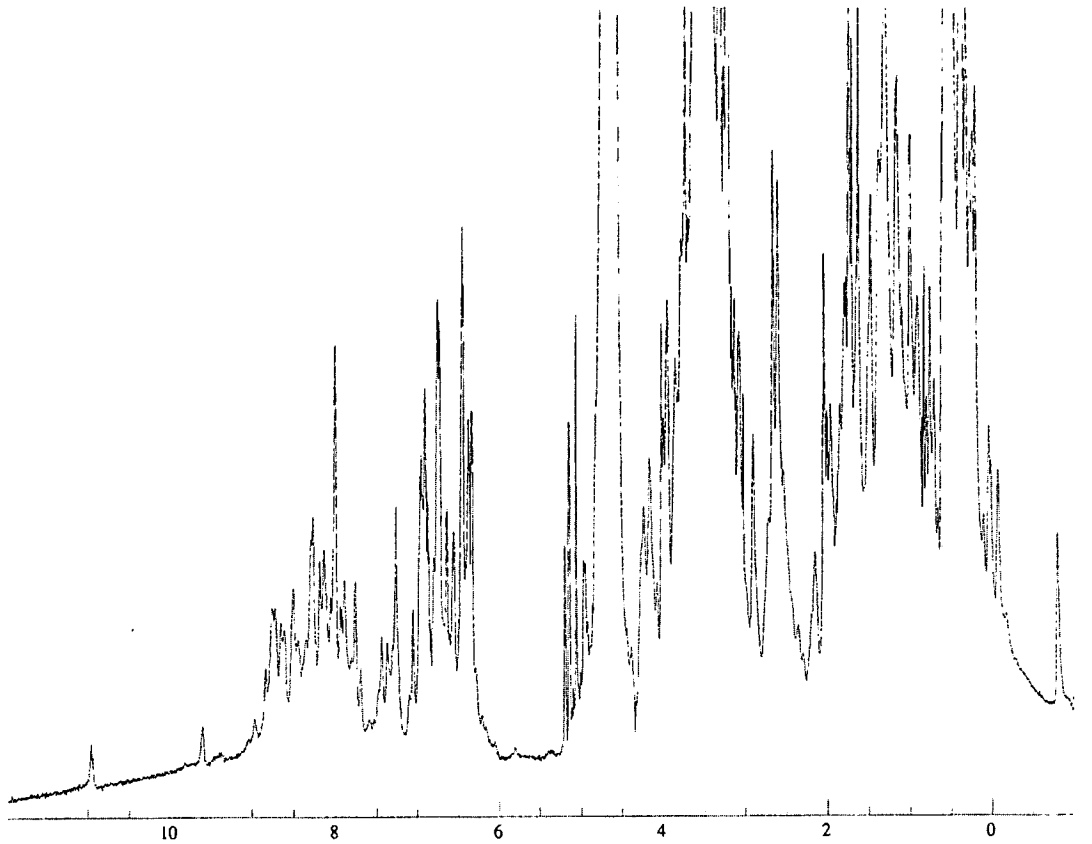


Figure 4:19 1D ^{1}H spectrum of EndoH$_{f}$-deglycosylated MCP2. The spectrum was acquired at 10 °C. The sample was 0.9 mM in 20 mM potassium phosphate pH 6.0.

Following the acquisition of the 1 D ^{1}H spectra, and the selection of 25 °C as a suitable temperature for data acquisition, a 2D ^{1}H, ^{1}H TOCSY spectrum was acquired on deglycosylated MCP2, and the result is shown in Figure 4:20. This spectrum shows the expected extent of dispersion for a protein of this type. For

many side-chains, the TOCSY transfer appears to extend through to the Hγ nuclei. There appears, however, to be significantly fewer peaks in this spectrum than would normally be expected for a protein of this size. This deficit may arise from inefficient magnetisation transfer due to aggregation or conformational exchange on an intermediate time-scale, both of which can result in shorter ^1H-T_2 relaxation times. The excessively large quantities of sugars present in glycosylated MCP2 rendered its 2D ^1H ^1H TOCSY s unusable.

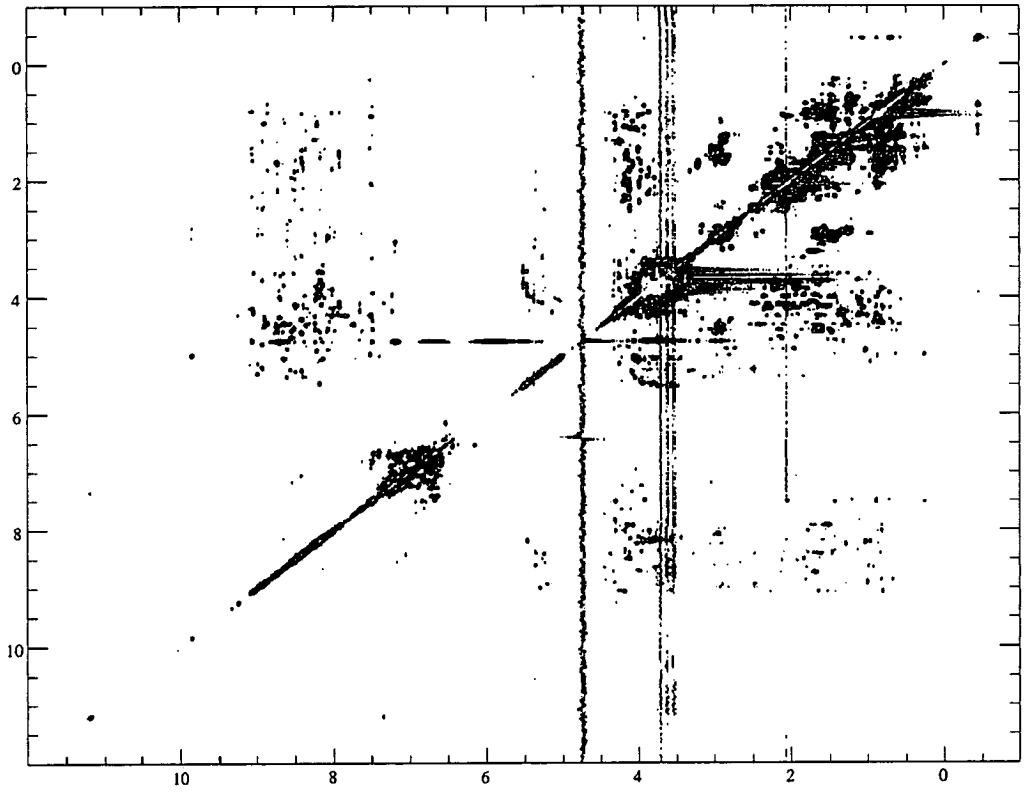


Figure 4:20 2D ^1H ^1H TOCSY spectrum of EndoH$_f$ deglycosylated MCP2 acquired at 25 °C. The sample was 0.9 mM in 20 mM potassium phosphate at pH 6.0.

A homonuclear NOESY was subsequently acquired of deglycosylated MCP2 and is shown in Figure 4:21. The spectrum shows a number of peaks with reasonable dispersion; however the shape of the peaks indicates that the sample may be in conformational exchange. Furthermore, it is noteworthy that there are no NOEs to

the consensus tryptophan residue. Were the protein well structured one would expect, to observe several such NOEs.

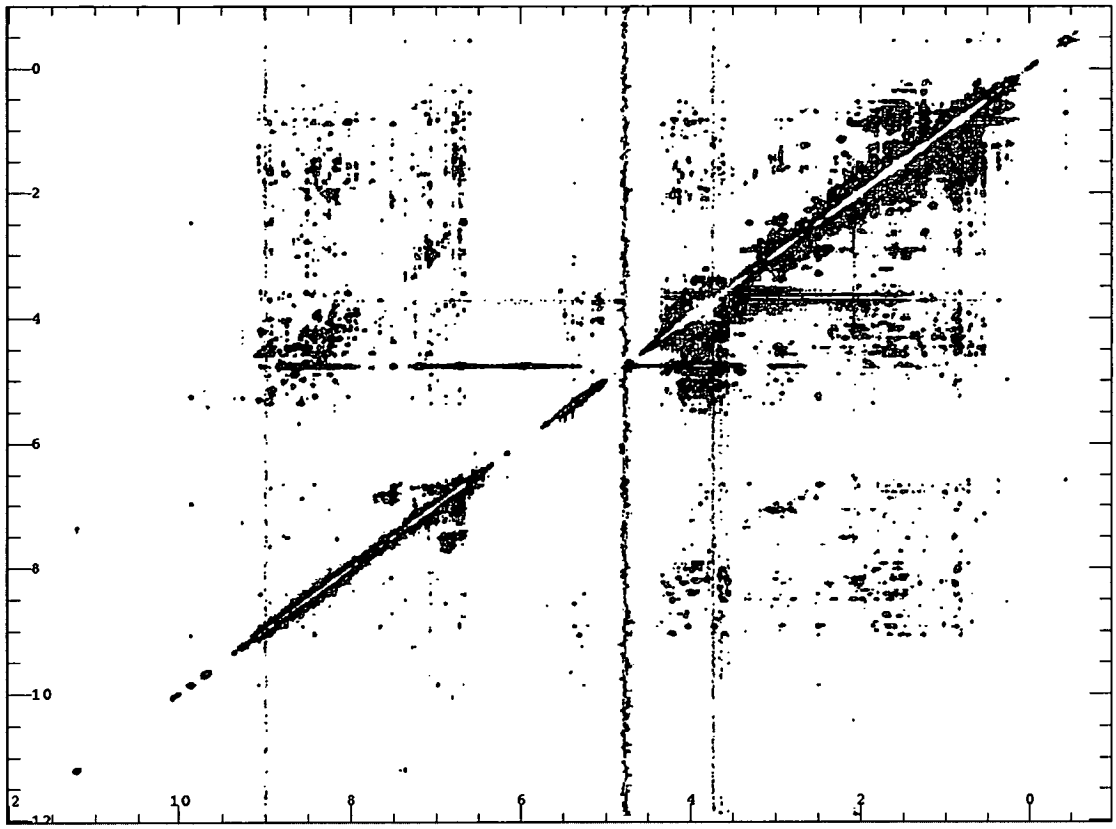


Figure 4:21 NOESY spectrum of deglycosylated MCP2 (100 ms mixing time). The spectrum was acquired at 25 °C on a 600 MHz Varian Innova spectrometer. The sample was 0.9 mM in 20 mM potassium phosphate, pH 6.0

Following the production of a ^{15}N-labelled sample of EndoH$_f$-deglycosylated MCP2 (see Figure 4:12), a more detailed NMR analysis of MCP2 could be performed. Initially ^{1}H ^{15}N HSQC spectra were recorded on a deglycosylated sample of MCP2. The spectrum shown in Figure 4:22 shows good dispersion indicative of a folded protein. There is, however, a considerable amount of line-broadening and overlap in the random coil region of the spectrum. This might be caused by aggregation, the presence of unfolded (or mis-folded) protein, or conformational exchange taking place on an intermediate timescale. It is possible that glycosylation reduces

conformational mobility. This could provide one possible explanation for the functional requirement of the N-linked glycans for both the measles receptor activity and the natural complement activity, which has been recorded both in this study and by others.

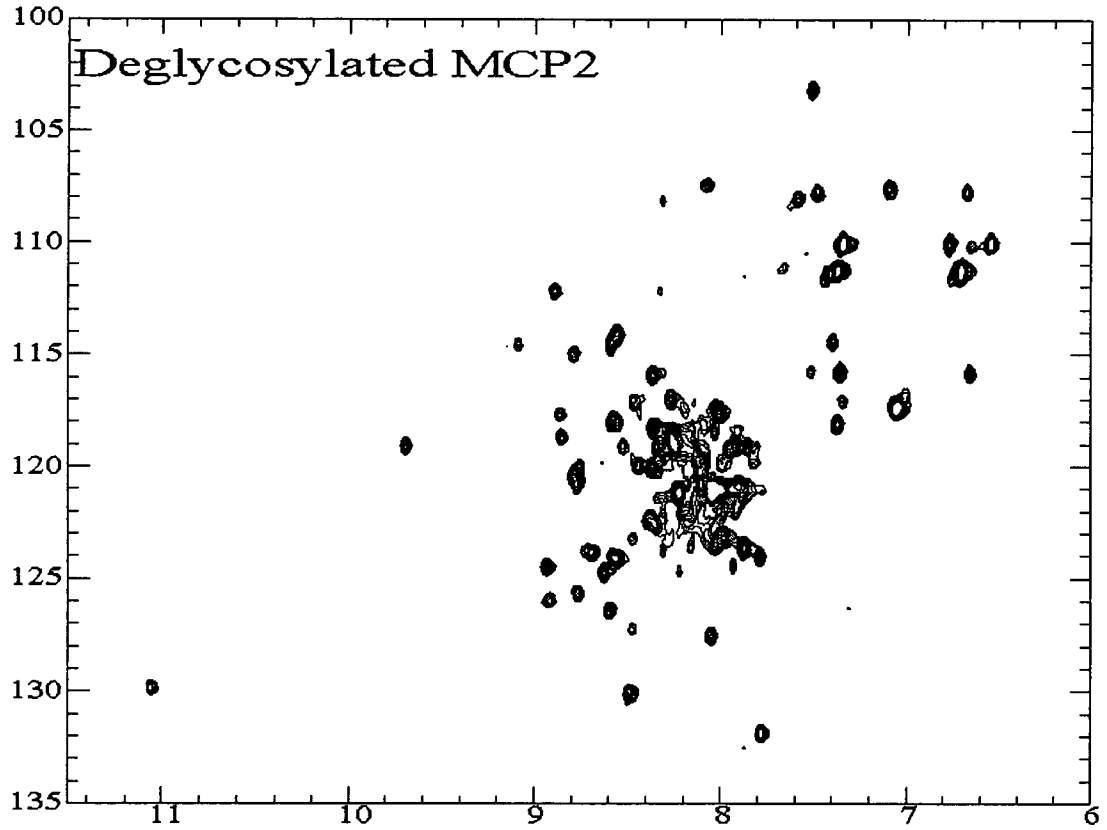


Figure 4:22 ^{1}H, ^{15}N HSQC spectrum of deglycosylated MCP2. The spectrum was acquired on a 0.9 mM sample in 20 mM potassium phosphate, pH 6.0, at 25 °C. The spectrum shows a large degree of dispersion, however there is considerable overlap in the centre of the spectrum.

It was therefore a logical next step to acquire a ^{1}H, ^{15}N HSQC spectrum of glycosylated MCP2. This spectrum, shown in Figure 4:23, is similar to that of deglycosylated MCP2, with a few interesting differences. Notably amongst these, are the considerably reduced amount of overlap in the random coil region, and the

difference in intensity of a small number of resonances, including increased relative intensity of the peaks at 8.6 ppm, 127 ppm, and that at 7.4 ppm, 127 ppm.

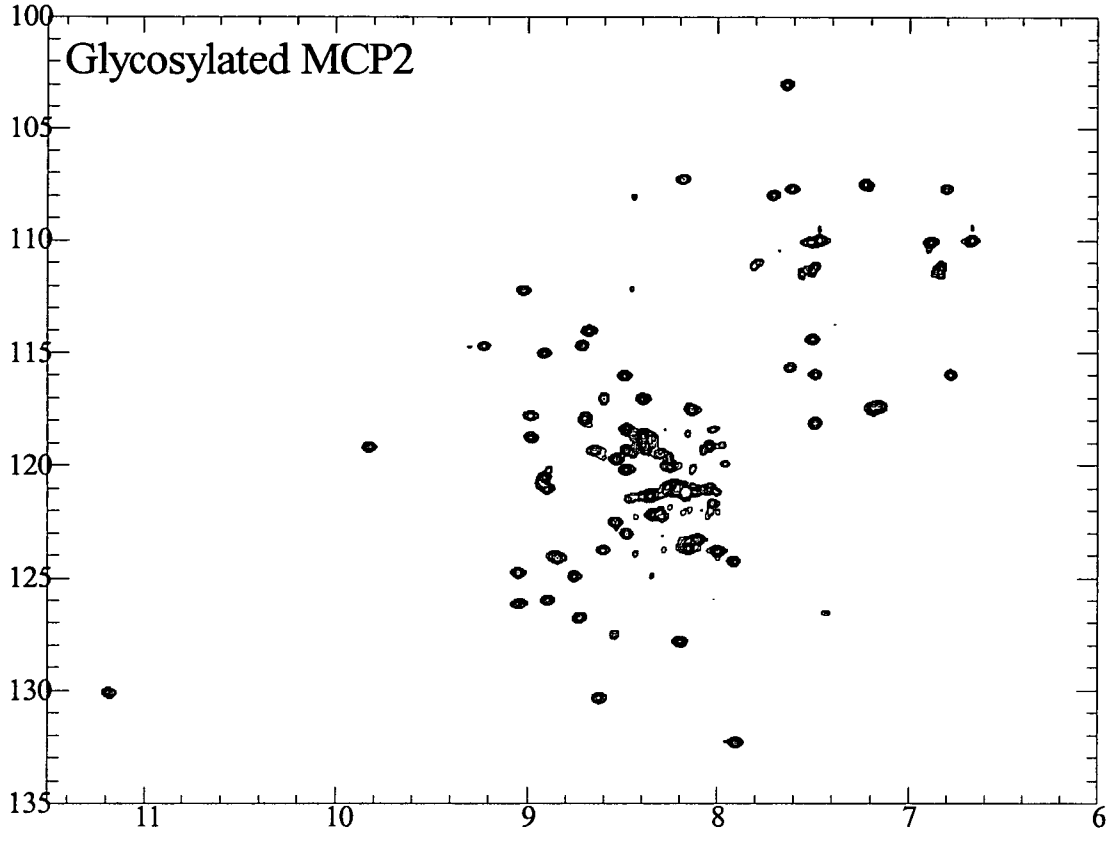


Figure 4:23 HSQC spectrum of glycosylated MCP2. This spectrum shows many features indicative of a tightly folded protein, with well defined peaks and a large degree of dispersion in both the nitrogen and proton dimensions. The sheer size of the glycoprotein and its heterogeneity, however, preclude the use of this version of MCP2 for structural studies.

The differences between these HSQC spectra of MCP2 are subtle, however the spectrum of glycosylated MCP2 shows significantly less overlap in the random coil region. One such possible explanation is that the sugars indeed perform a structural role, as may be inferred from the crystal structure (Casasnovas et al., 1999). A second possibility is that the sugars enhance solubility and inhibit the formation of transient aggregates.

4.4.3. Differential scanning calorimetry analysis of MCP2

To investigate the possibility that the N-linked glycans on module 2 have a stabilising effect, differential scanning calorimetry was performed on both deglycosylated and glycosylated forms.

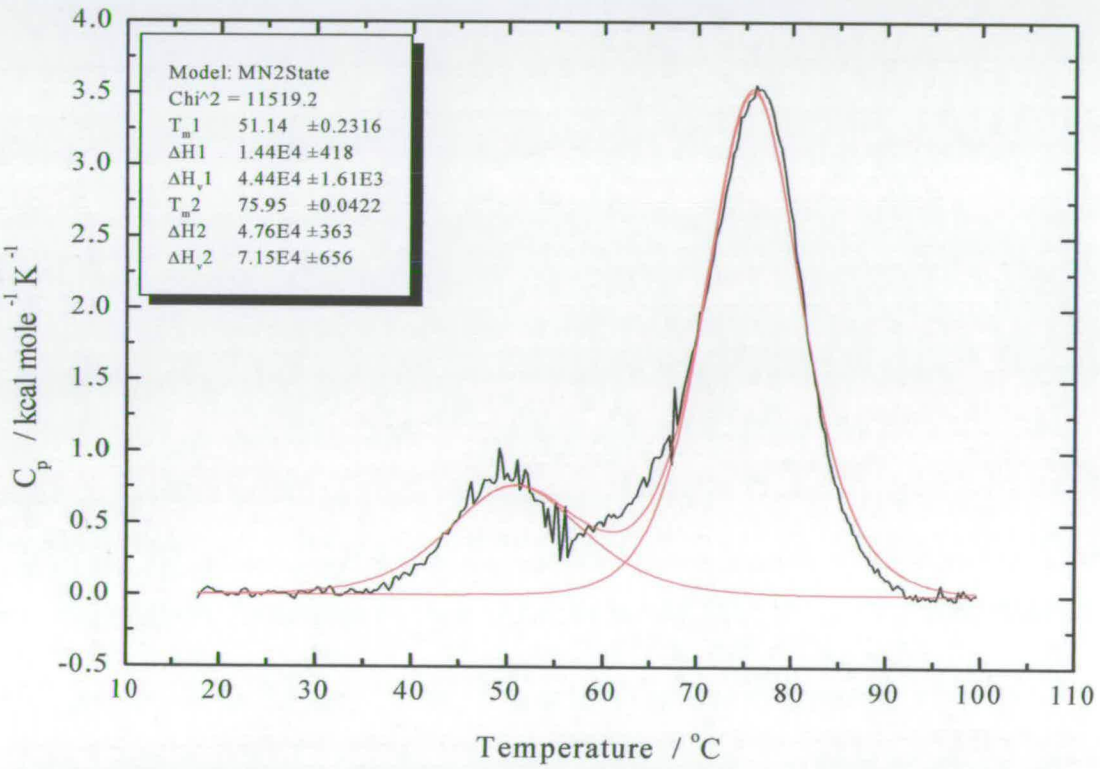


Figure 4:24 Differential scanning calorimetry of EndoH$_f$-deglycosylated MCP2. The data were best fit to a model with two transitions with melting temperatures of 51 °C and 76 °C. The sample was 0.63 mg.ml^{-1}.

The DSC data for EndoH$_f$-deglycosylated MCP2, shown in Figure 4:24, were fitted to a double transition model with T_m values of 51 °C and 76 °C. Possible explanations for the presence of two transitions in the melting profile include the unfolding of two different forms of the protein, or a two-stage co-operative unfolding

process. The presence of two forms of MCP2 in the deglycosylated sample is consistent with two interpretations of the HSQC NMR data (i.e. line-broadening due to inter-conversion between conformers; or the presence of an improperly folded form of the protein). Bearing this in mind, one can attempt to interpret these data. The T_m value of 76 °C is indicative of a protein which is well folded and highly compacted. This transition therefore, probably represents the thermal denaturation of the native conformation of MCP2. The transition with a T_m of 51 °C would represent a protein conformation that is considerably less stable.

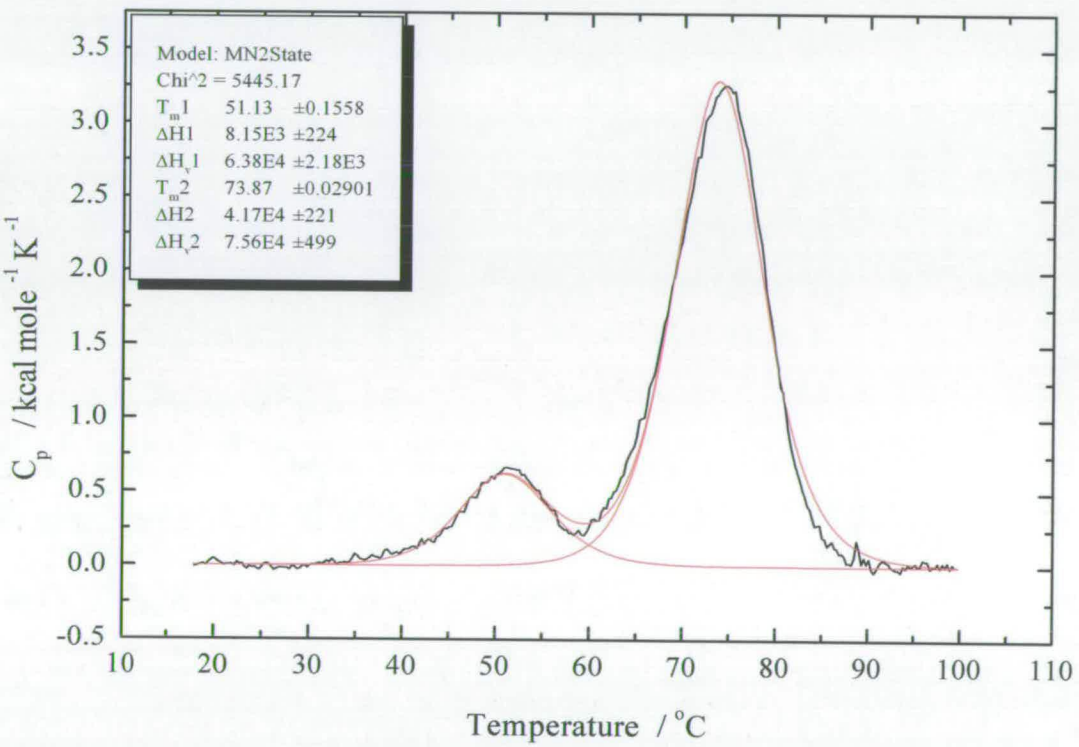


Figure 4:25 Differential scanning calorimetry of glycosylated MCP2. The data were again fitted to a two transition model of unfolding, with melting temperatures of 51 °C and 74 °C. The sample was o.54 mg.ml^{-1}.

The DSC data for glycosylated MCP2 (shown in Figure 4:25) were also fitted to a two transition model of unfolding. In this case, melting temperatures of 51 °C and 74 °C were obtained. That similar two transition DSC traces were obtained for both

glycosylated and deglycosylated MCP2 (Figure 4:26) implies that glycosylation is not having a very major affect either on the thermal stability, or the unfolding pathway of the protein.

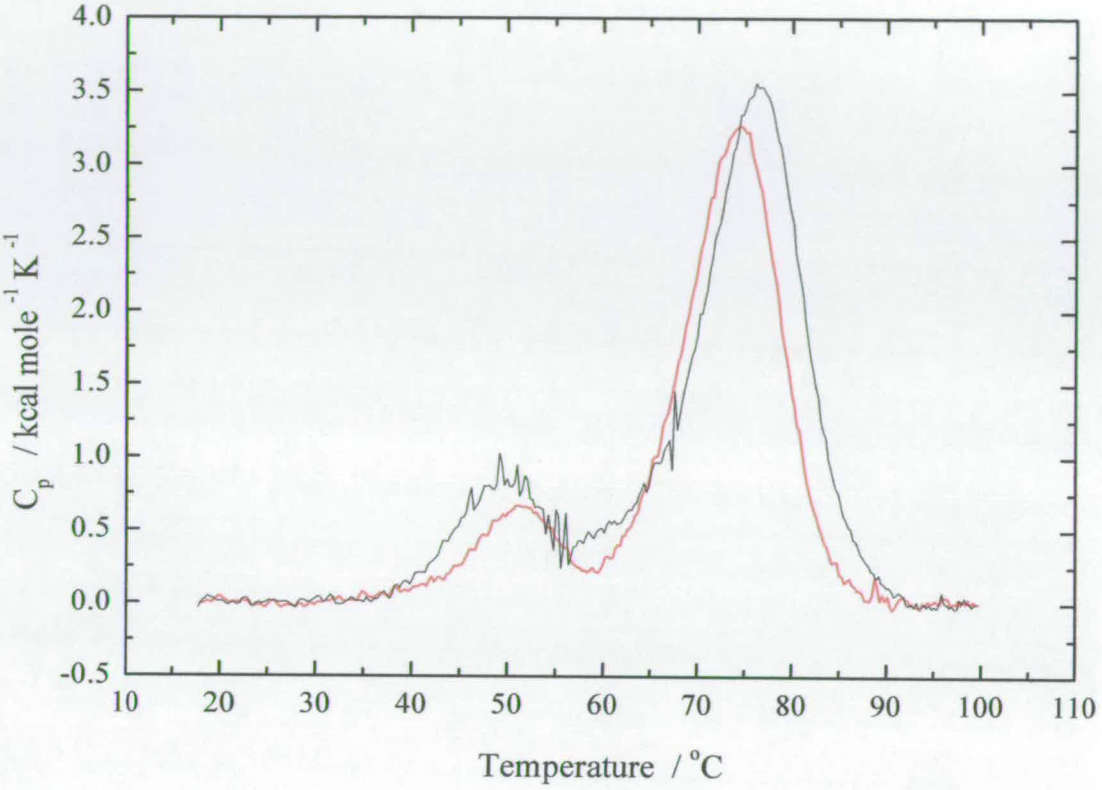


Figure 4:26 Overlay of DSC of glycosylated (red) and deglycosylated (black) MCP2.

4.5. Discussion

The results presented in this chapter demonstrate the successful cloning and expression of MCP2 in *P. pastoris* at levels sufficient for biophysical characterisation. Deglycosylated MCP2 was successfully purified to homogeneity as determined by SDS-PAGE, whilst the hyperglycosylated form of MCP2 was purified

to a homogeneous protein species that is heterogeneous solely with respect to glycosylation.

N-terminal sequencing and mass spectrometry data confirmed the identity of the protein as MCP2, and revealed that complete processing of the secretion signal had not occurred - an extra EA, or EAEA, sequence remained at the N-terminus of the recombinant protein following signal sequence processing. Mass spectrometry experiments also confirmed the lack of chemical modifications to the recombinant protein, and was further used to verify that the two predicted disulphide bonds had formed. Preliminary NMR experiments indicated that both the glycosylated and deglycosylated forms of MCP12 show the good dispersion indicative of a structured protein. There was, however, a larger area of overlap in the centre of the $^{1}H,^{15}$NHSQC from deglycosylated MCP2 than would normally be expected for a protein of this size. In the case of glycosylated MCP2, on the other hand, this is not the case. While a possible explanation for this is transient aggregation of deglycosylated MCP2this seems improbable since the problem persisted over a range of sample conditions. A more likely explanation is that conformational exchange on an intermediate time-scale is responsible. Evidence for inter-conversion between conformers was observed in the ^{1}H NMR spectra of fresh MCP2 samples recorded at various temperatures. While the broad peaks that appeared at higher temperatures failed to appear in older samples, implying that a slow folding process was involved, other dynamic events on this intermediate time-scale cannot be ruled out. CCP modules do not possess a high proportion of regular secondary structure and much of the main chain lies in flexible loops and turns. Further NMR analysis was not

performed on deglycosylated since it was thought that the acquired spectra were not of sufficient quality for a full structural study to be undertaken.

The question as to the affect of glycans could be better addressed through the production of a sample of MCP2 that has a more natural glycosylation pattern. This can best be achieved through the use of an alternative expression system that does not hyperglycosylate recombinant proteins. Moreover, it is possible that the act of deglycosylation is the cause of the apparently less ordered structure of deglycosylated MCP2. To determine whether this is the case, a sample of native sequence that has never been glycosylated is required. To achieve these goals, two other expression systems were employed. *Drosophila melanogaster* S2 cells were used to generate a more naturally glycosylated form of MCP2; and *E. coli* was used to generate a large amount of non-glycosylated protein that could subsequently be refolded. The use of these alternative expression systems will form the subject of the following chapter.

5. Recombinant Mcp2 from Other Organisms

5.1. Preliminary remarks

Given the problems encountered producing a sample well-folded sample of MCP2, two alternative approaches were tried. In order to discount the hypothesis that it was the act of deglycosylation that was responsible for the apparent differences between the spectra of glycosylated and deglycosylated MCP2, a sample with native sequence that had never been glycosylated was required. In this respect expression in *E. coli* appears to be a reasonable option. Furthermore, CCP modules have, in the past, been successfully expressed and refolded from material expressed in *E. coli* (Dodd et al., 1995). The second approach was to attempt to express the protein in a host that is capable of attaching a far smaller and better defined glycosylation profile. The ideal host for this purpose would appear to be an insect expression system, and a *D. melanogaster* expression system in particular seems appropriate since expression levels of 10 – 35 $mg.l^{-1}$ have been achieved (Ivey-Hoyle et al., 1991; Johanson et al., 1995). The work described in the following sections describes the use of these hosts for the expression of MCP2.

5.2. Cloning and expression of MCP2 in E. coli

5.2.1. The pET system

The pET system was chosen as a starting point for expression studies of MCP2 in *E. coli*, since it has a proven record of generating large amounts of recombinant proteins. It also has several diverse vectors with different features that can help when

expressing difficult proteins. The pET system illustrated in Figure 5:1, is centred around the use of the bacteriophage T7 RNA polymerase. In the pET system, the gene of interest is cloned behind the T7 RNA polymerase promoter. Under normal circumstances of course, *E. coli* doesn't produce T7 RNA polymerase, and wild type strains do not contain this gene. This facilitates the cloning of genes encoding toxic proteins, for example, since basal protein expression is negligible. The expression hosts used in the pET system are lysogens for the bacteriophage DE3, and contain the lacI gene and a copy of the T7 RNA polymerase gene under the control of the lacUV5 promoter. The lacUV5 promoter is inducible by the unmetabolisable lactose analogue, isopropyl thiogalactoside (IPTG). Therefore the DE3 lysogenic hosts produce T7 RNA polymerase after the addition of IPTG to the culture medium. The polymerase is then able to undertake transcription of the gene in a pET vector that is under the control of the T7 RNA polymerase promoter. Thus, this system is a simple and effective method of producing inducible recombinant protein in enabled host strains.

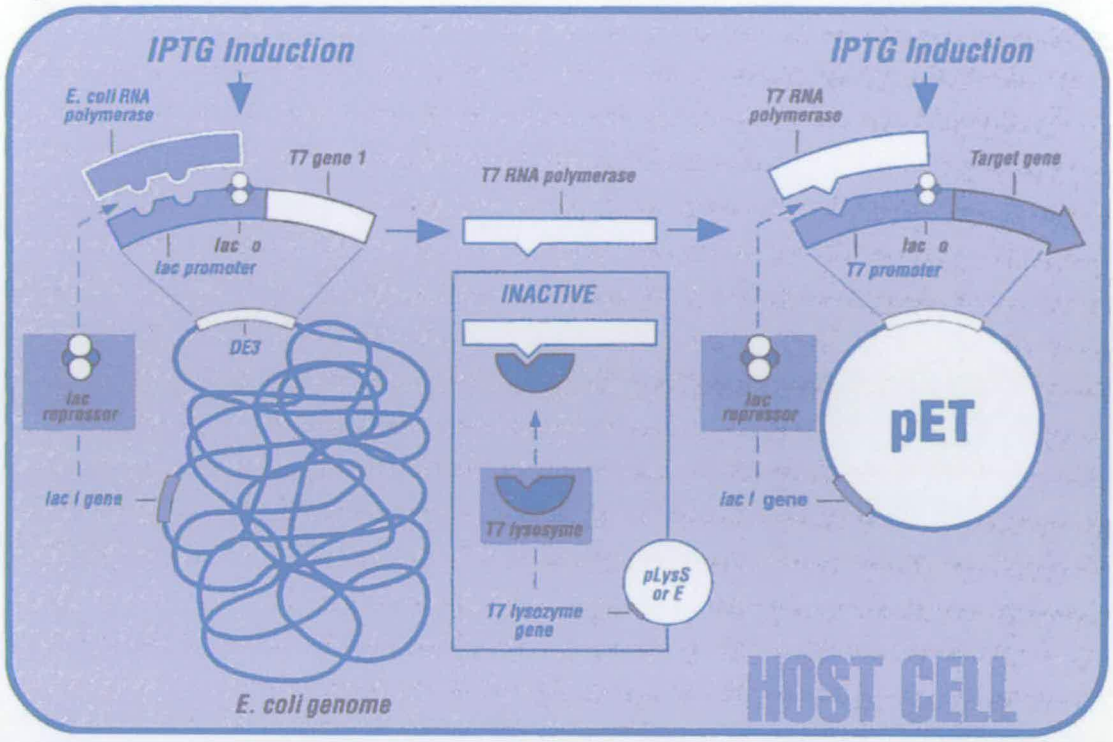


Figure 5:1 Overview of the pET system. The gene of interest is cloned in a pET vector behind the T7 RNA polymerase promoter. Following transformation into a DE3 host strain, expression is induced by adding IPTG to the culture medium. This induces the production of T7 RNA polymerase from a gene on the DE3 lysogen that is under the control of the lacUV5 promoter, (taken from Novagen, 2002)

5.2.2. Cloning and expression of MCP2 in pET15b

The expression vector pET15b shown in Figure 5:2 was chosen for initial work on production of MCP2 in *E. coli*. The resulting construct will henceforth be referred to as MCP2 pET15b. This vector is one of the simplest expression vectors in the pET expression system, with only the inclusion of a thrombin cleavable N-terminal His tag for purification. This vector uses ampicillin resistance as a selectable marker.

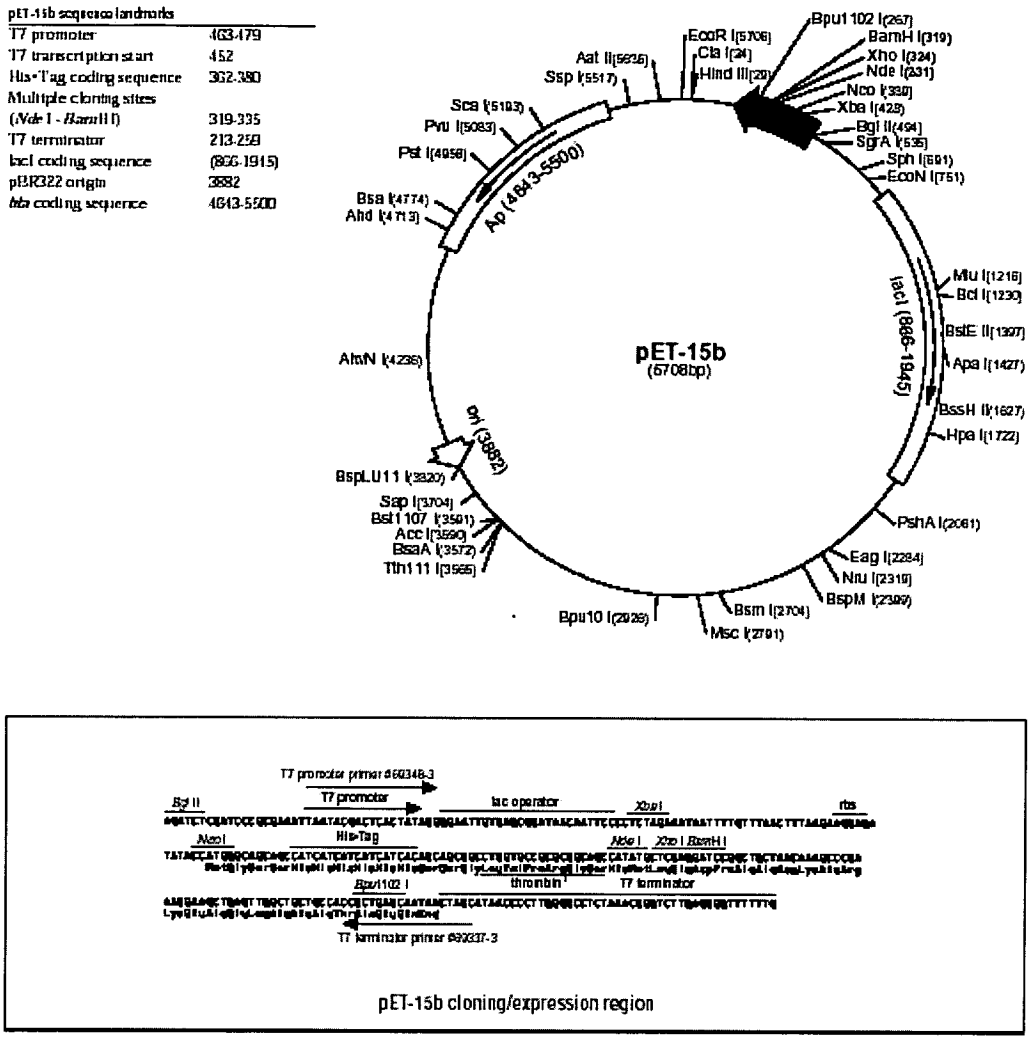


Figure 5:2 Vector map of pET15b showing the key features (taken from Novagen, 2002)

The region coding for MCP2 was amplified by PCR, and the agarose gel shown in Figure 5:3 reveals a single band of approximately the correct size for MCP2. This fragment was then inserted into pET15b between the restriction sites *Nco*1 and *Xho*1. The plasmid was then sequenced to verify the identity and accuracy of the insert, and a plasmid containing an insert of the published sequence for MCP2 was used to transform *E. coli* BL21 DE3 cells. Transformation of *E. coli* was performed

immediately prior to every growth for all the following expression studies. A single colony from a transformation was used to inoculate LB and samples taken of both the soluble and insoluble fractions of induced and uninduced samples.

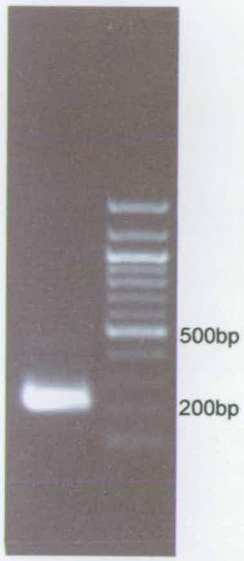


Figure 5:3 PCR of MCP2 for cloning into pET15b. The gel shows a single band of approximately the correct size for MCP2

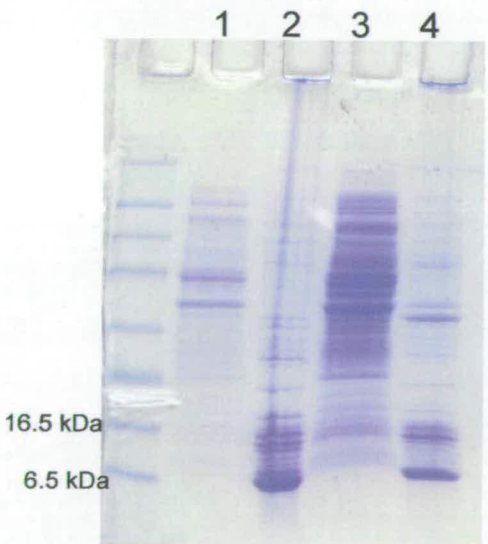


Figure 5:4 SDS PAGE of MCP2 pET15b. The gel shows that MCP2 is expressed in reasonable quantities, but is exclusively in the insoluble fraction (lanes 2 and 4) . The presence of MCP2 in the uninduced fraction (lane 4) indicates that expression of the fragment is not dependant upon induction with IPTG.

The SDS PAGE shown in Figure 5:4 shows the soluble (lane 1) and insoluble (lane 2) fractions from the induced samples of MCP2, alongside the soluble (lane 3) and insoluble (lane 4) fractions from the induced samples of MCP2. The gel shows a band of the correct size for the MCP2 fragment in the insoluble fractions from both the induced and uninduced samples. Since it had previously been observed that BL21 DE3 cells transformed with pET15b without an insert do not produce a band of this size, it seemed probable that this band was MCP2. The fact that this band is also present in the induced samples (albeit at a lower level) indicated that the plasmid was "leaky" i.e. induction was not a prerequisite for expression. This phenomenon has previously been observed with other proteins expressed using the pET system and is usually the result of basal expression of the T7 RNA polymerase gene. If desired this basal expression can be reduced by using host strains containing pLysS or pLysE (Dubendorff et al., 1991; Studier, 1991). There seems little point in attempting to repress this basal expression in this case, however, since MCP2 is apparently not toxic to *E.coli*.

5.2.3. Cloning and expression of MCP2 in pET32a

Since the MCP2 expressed in pET15b was found entirely in the insoluble fraction, it was decided to explore alternative strategies with the aim of achieving expression of soluble material. One method of solubilising proteins expressed in *E. coli* is to express them as fusion proteins. One possible fusion partner is thioredoxin, which is

encoded in the pET32a vector illustrated in Figure 5:5. This fusion tag is particularly appropriate for the expression of disulphide bonded proteins since it has been shown to catalyse the formation of disulphide bonds in the cytoplasm of *E. coli* (Novagen, 2002). In addition to its ability to aid the formation of disulphide bonds, thioredoxin has also been shown to increase the solubility of proteins expressed in *E. coli* (Garcia-Ortega et al., 2000; Sachdev et al., 1998). For these reasons it seemed reasonable to use pET32 to express MCP2 in a soluble and, potentially, disulphide bonded form. Another reason for choosing the pET32 vector was the ease with which the gene for MCP2 could be inserted into the pET32 plasmid. It had previously been observed that genes cloned into pET15b between *Nco*1 and *Xho*1 could be digested with *Nde*1 and *Xho*1 and inserted into pET32 (S. Blein, University of Edinburgh. personal communication). This would allow the direct insertion of the MCP2 construct into pET32 without the need for a further PCR step. The resulting construct will henceforth be referred to as MCP2 pET15_32. This method for inserting a gene into pET32 does, however, produce a fusion protein that contains two His tags and two thrombin cleavage sites. The presence of the two His tags poses no serious problems, and indeed may actually aid the purification. The presence of two thrombin cleavage sequences on the other hand, could result in a less efficient cleavage of the fusion tags. This potential problem could possibly be alleviated by cleaving the fusion tags using a larger amount of thrombin. If this construct proved useful, one could subsequently reclone the insert in a cleaner fashion.

pET-32a(+) sequence landmarks	
T7 promoter	764-780
T7 transcription start	783
Trx•Tag coding sequence	366-692
His•Tag coding sequence	327-344
S•Tag coding sequence	249-293
Multiple cloning sites (*Nco* I - *Xho* I)	158-217
His•Tag coding sequence	140-157
T7 terminator	26-72
lacI coding sequence	1171-2250
pBR322 origin	3681
bla coding sequence	4445-5302
f1 origin	5431-5889

The maps for pET-32b(+) and pET-32c(+) are the same as pET-32a(+) (shown) with the following exceptions: pET-32b(+) is a 5899bp plasmid; subtract 1bp from each site beyond *Bam*H I at 198. pET-32c(+) is a 5901bp plasmid; add 1bp to each site beyond *Bam*H I at 198 except for *Eco*R V, which cuts at 209.

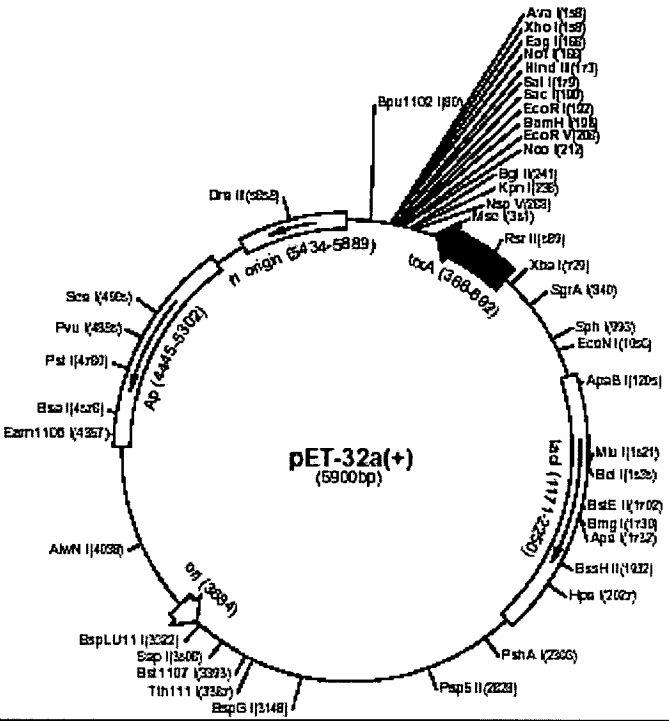


Figure 5:5 Vector map of pET32a showing the key features of the thioredoxin tag, ampicillin resistance, and multicloning region (taken from Novagen, 2002).

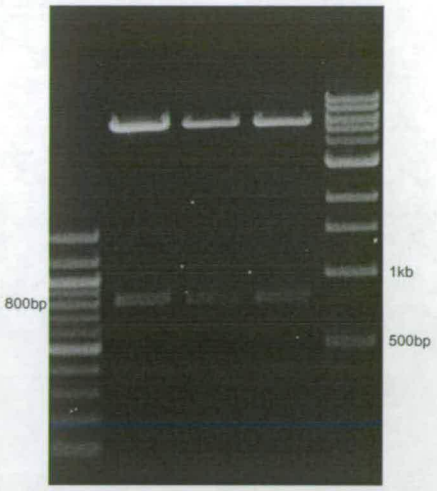


Figure 5:6 1.4% agarose gel showing MCP2 pET15b _pET32 cut *Xho*1 *Xba*1. The gel shows that all the clones tested contained an insert of the correct size for MCP2 with the thioredoxin tag, 2* His tags, and 2* thrombin cleavage sites

The agarose gel shown in Figure 5:14 shows the digestion with *Xho*1 and *Xba*1 of pET32 plasmids thought to contain the MCP2 insert. The gel reveals the presence of the insert in all of the plasmids tested. These plasmids were then sequenced to varify the identity and accuracy of the insert, and a clone containing an insert of the predicted sequence for MCP2 was used to transform *E. coli* BL21 DE3 cells. The Origami strain of *E. coli*, which has also been shown to be effective at forming disulphides in the cytoplasm, was also transformed using this plasmid, but no positive clones were obtained using this strain. Soluble and insoluble fractions of induced and uninduced samples of MCP2 pET15_32 were prepared in a similar way to those for MCP2 pET15b, and subjected to SDS-PAGE. The gel shown in Figure 5:7 reveals that in contrast to MCP2 pET15b all or most of the MCP2 is now found in the soluble fraction as a band of approximately 25 kDa. The use of a thioredoxin tag was therefore successful in producing a soluble form of MCP2 expressed in *E. coli*. It is also interesting to note that once again MCP2 is found in the uninduced samples albeit at a lower level.

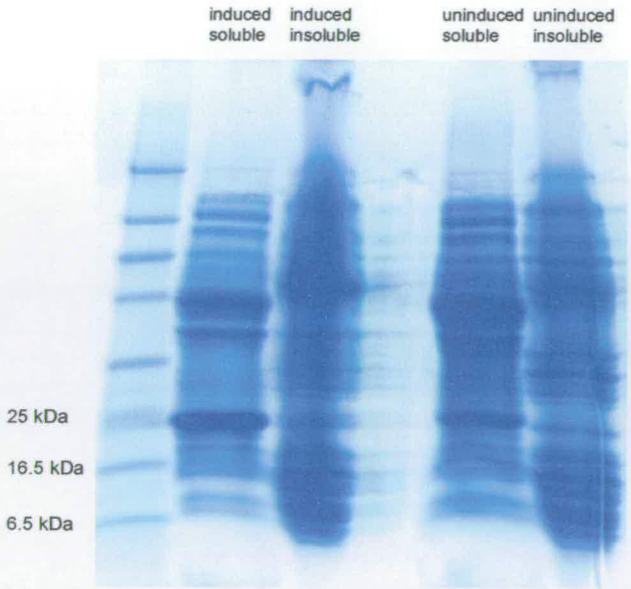


Figure 5:7 Induction tests of MCP expressed in *E. coli* BL21. SDS PAGE showing a band at 25 kDa in the soluble fractions of both the induced and uninduced samples. This band is the correct size for MCP2 with the thioredoxin tag, two His tags and a thrombin cleavage sequences. This band was not present in either the gel for BL21 without the MCP2 pET15b32a plasmid (data not shown) or the BL21 containing MCP2 pET15b (Figure 5:4). It can also be seen that the band is present at only very small levels in either of the insoluble fractions. It therefore seems that the addition of the thioredoxin tag has converted all the expressed MCP2 from the insoluble to the soluble fraction. In addition as with MCP2 pET15b, induction with IPTG is not required for protein expression.

5.2.4. Large scale expression of MCP2 pET15_32

Since it was observed that MCP2 was expressed by both induced and uninduced cells, it seemed prudent to exploit this property to develop a more efficient protocol for expression. It has been noted by others in our laboratory that *E. coli* can be grown satisfactorily by inoculating 250 ml of LB medium containing 100 $\mu g.ml^{-1}$ ampicillin (LB-amp100) directly with a single colony from a freshly transformed plate. It therefore seemed reasonable to modify this procedure to express MCP2.

This was achieved by inoculating 4 × 250 ml LB-amp100 each with a single colony from a freshly transformed plate of MCP2 pET15-32 BL21. This culture was incubated overnight at 37 °C with shaking at 250 rpm, and processed the following morning for purification of MCP2. This protocol offers significant advantages over the standard growth–induction protocols commonly used for expressing proteins in *E. coli*. Not least of these advantages are the time saving (almost two days) and the speed at which the harvested pellet is purified – thus reducing the chances of degradation. The cells expressing the MCP2 fusion protein were subjected to chemical lysis with the "Bugbuster" reagent (Novagen). Following lysis, the soluble fraction was applied to a charged nickel affinity column and bound protein eluted with a step gradient of 1 M imidazole. The chromatogram shown in Figure 5:8 reveals that the MCP2 fusion protein elutes from the column as a single peak. Following the initial purification of MCP2 on nickel affinity resin, the thioredoxin and His tags were cleaved with thrombin, and then the sample was re-applied to the nickel affinity column. This last step removes the cleaved His tags, and some of the non-specifically bound proteins, whilst the cleaved protein passes through the column without binding. N-terminal sequencing yielded a sequence of **GSHMYRETCP** indicating that the thioredoxin and His tags had been completely removed.

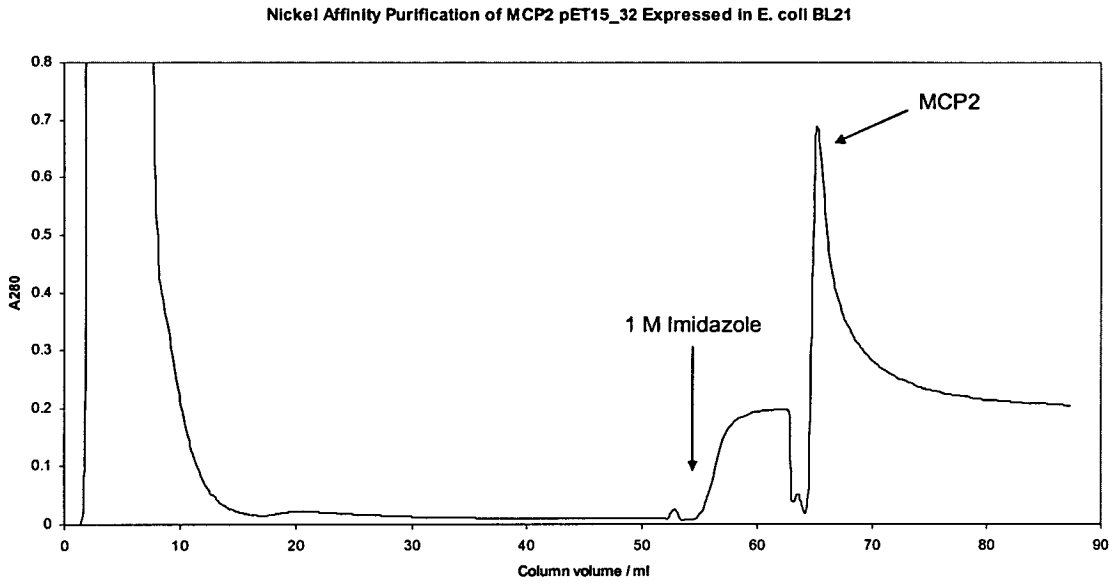


Figure 5:8 Chromatogram showing the purification of MCP2 on a nickel affinity column. The rise and dip in A_{280} between 55 ml and 65 ml is due to the protein purification machine equilibrating to elution buffer with the column off line.

The final purification step was gel filtration on a Superdex 75 column. The chromatogram shown in Figure 5:9 reveals that a small proportion of material elutes at approximately 78 ml. This compares to the gel filtration of MCP12 which elutes from the column at approximately 75 ml, and indicates that a small proportion of the sample exists as a dimer. This is presumably the result of non-specific dimerisation, since the protein had not been refolded at this stage, and this behaviour was not observed with MCP2 expressed by *P. pastoris*. Since the sample was soluble at this stage it was decided to acquire a 1D ^{1}H NMR spectrum, in order to determine whether a refolding step was necessary. This spectrum, shown in Figure 5:10 shows almost none of the dispersion seen in the MCP2 produced by *P. pastoris* that is characteristic of folded proteins. It was therefore concluded that this form of MCP2 was almost completely unfolded, and therefore requires a refolding step.

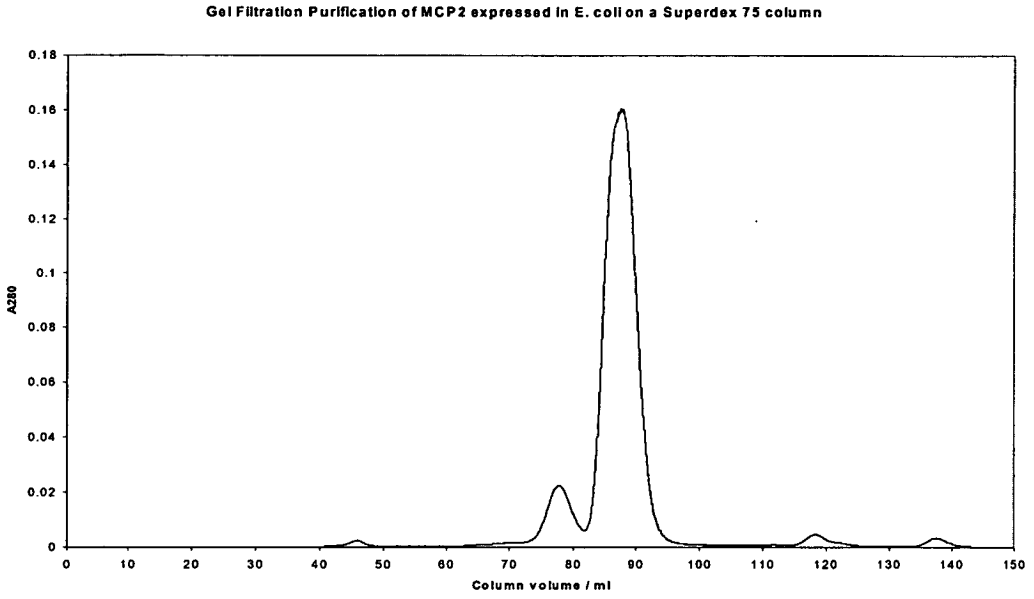


Figure 5:9 Purification of MCP2 expressed in *E. coli* on a Superdex 75 column. The chromatogram shows the existence of a small proportion of material that behaves as a dimer.

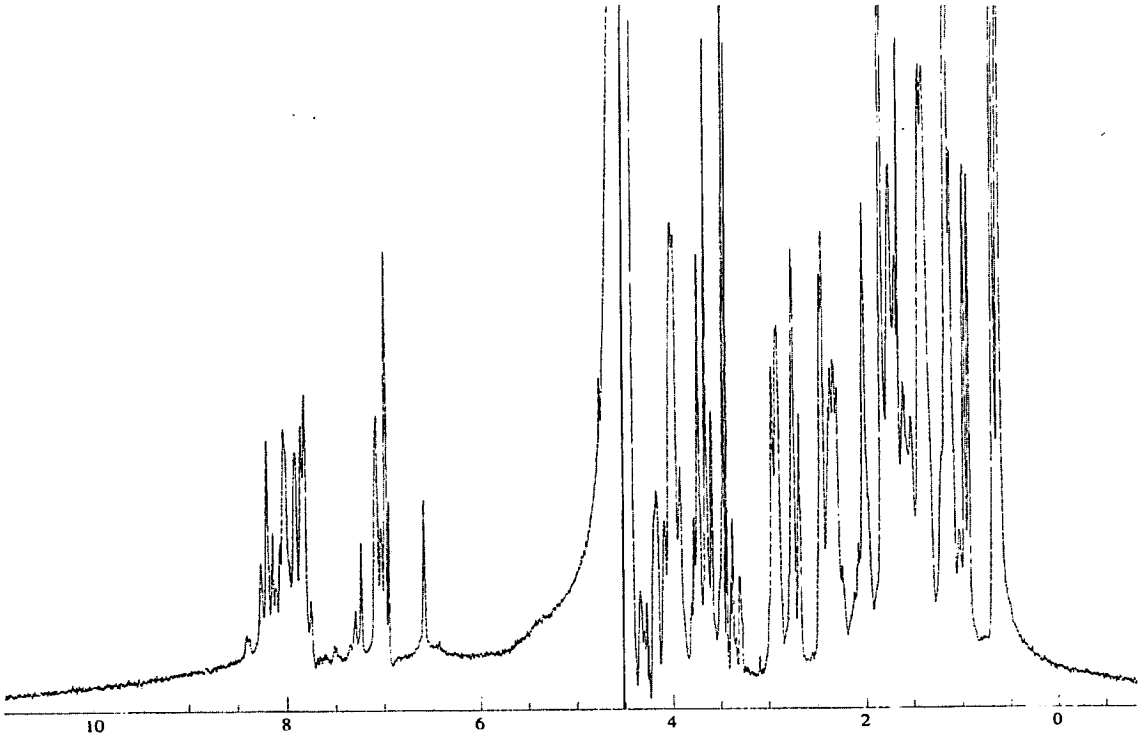


Figure 5:10 1D 1H NMR spectrum at 25 °C of MCP2 expressed in *E. coli* before refolding. The sample was 0.8 mM in 20 mM potassium phosphate, pH 6.0. The spectrum shows almost no dispersion from random coil.

5.2.5. Refolding of MCP2 expressed in *E. coli*

The refolding of proteins that contain disulphide bonds can prove problematic and even in favourable cases the process results in the loss of large amounts of protein. The problem is particularly challenging in proteins for which there is no easy functional assay of the correctly folded protein. In the case of MCP2, the refolding step was performed using a protocol similar to that used to successfully refold MCP1 (O'Leary, 2000). The protein was purified as described above, and then denatured using 8 M urea, and the disulphide bonds disrupted by the inclusion of 2 mM reduced glutathione (GSH) and 0.2 mM oxidised glutathione (GSSG). The protein was then dialysed against 100 mM Tris, pH 8.0, 2 mM GSH, 0.2 mM GSSG containing 5 mM EDTA to help prevent possible metal catalysed oxidation, and proteolysis. The refolded sample was then further purified by anion exchange chromatography under the conditions used for the purification of MCP2 expressed in *P. pastoris*, and this chromatogram shown in Figure 5:11 is similar to that obtained for MCP2 expressed in *P. pastoris*. Following the refolding step a 1D ^{1}H NMR spectrum was acquired as shown in Figure 5:12. This spectrum shows considerably more dispersion than that of the non-refolded sample, indicating that the refolding protocol had caused the protein to become more structured. Comparison of the spectra from the refolded material with deglycosylated MCP2 expressed by *P. pastoris* shows that the spectra are largely similar, however it appears that there is still more dispersion in the deglycosylated MCP2 than in the refolded MCP2. Possible reasons for this include the structure of MCP2 being further destabilised by having no glycans at all or that the initial steps of glycosylation are required in order to attain the correct fold.

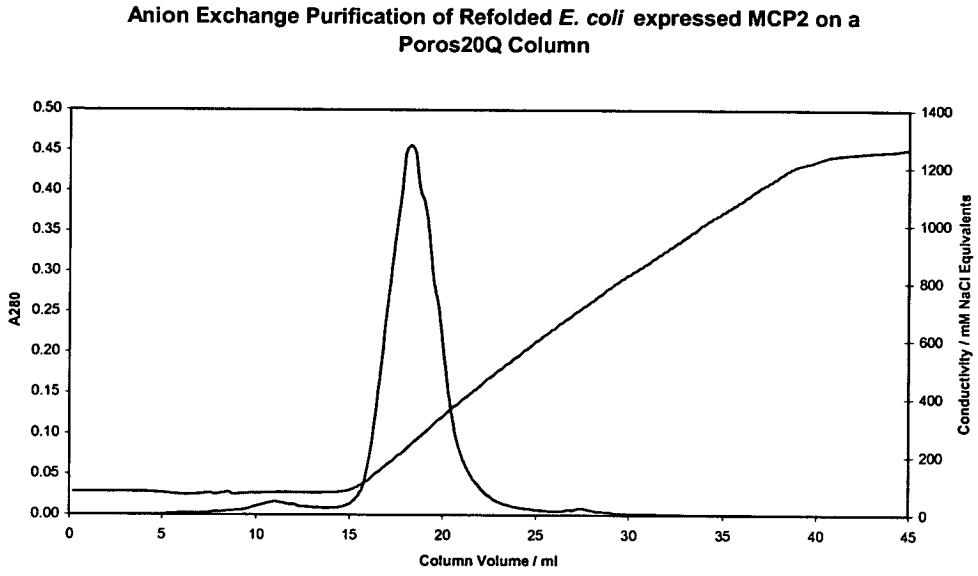


Figure 5:11 Purification of 'refolded' *E. coli*-expressed MCP2. The chromatogram appears to be very similar to that for the *P. pastoris*-expressed MCP2.

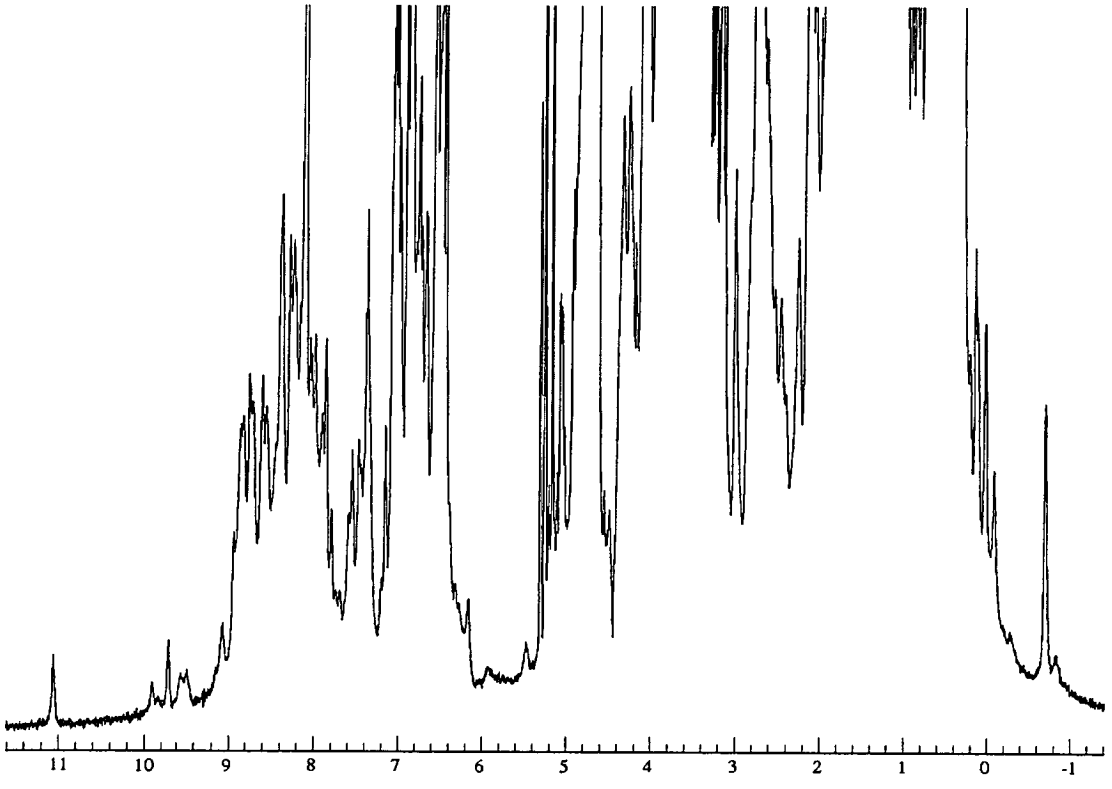


Figure 5:12 1D ^{1}H spectrum of 'refolded' MCP2. The spectrum was acquired at 25 °C. The sample was 0.8 mM in 20 mM potassium phosphate pH 6.0.

5.3. Cloning and expression of MCP2 in drosophila S2 cells

It proved impossible to generate MCP2 with a small and relatively homogeneous glycosylation profile in *P. pastoris*. The choice then was to find a host capable of such a task. It is widely believed that the phenomenon of hyperglycosylation is more common in yeasts such as *S. cerevisiae* than it is in *P. pastoris*. It therefore seemed unnecessary to attempt to express this fragment another yeast host. Instead it was decided to express this fragment in a higher eukaryotic host. The choice of which host to choose rested on several factors, namely: the glycosylation profile produced; the level of expression expected; the media requirements of the host cell line; the equipment requirements of the technique; and the ease of mastering the technique. In each of these respects, one host appeared to be more suitable than the others available, namely S2 cells from the *D. melanogaster* expression system (DES). This is a non-lytic expression system where expression levels of 10 – 35 $mg.l^{-1}$ have been recorded (Ivey-Hoyle et al., 1991; Johanson et al., 1995). Proteins expressed with the DES give a glycosylation profile of the high mannose type (typically Man9GlcNAc2). Whilst complex N-glycans cannot normally be produced in this host, this is unnecessary since it has been demonstrated that either complex, or high mannose, glycans are sufficient for functional activity. The S2 cells are able to grow in serum free medium, which can potentially be isotopically labelled, and are relatively easy to culture. The cells are cultured at room temperature either in loosely adherent or suspension cultures. Moreover transfections are generally relatively efficient. To allow glycosylation and disulphide bond formation, it is necessary that the protein enters the secretory pathway. With this in mind the, the expression vector pMT/BiP/V5-His shown in Figure 5:13 was used. This vector uses

the inducible drosophila metalothianein (MT) promoter, which is very tightly regulated, but can be induced to high levels by the addition of copper sulphate or cadmium chloride to the growth medium (Angelichio et al., 1991). Copper sulphate is the prefered choice as an inducer since it is less toxic to the cells. Additionally, this vector uses the secretion signal from the drosophila 'BiP' protein to target the heterologous protein to the secretory pathway and subsequently into the culture medium via the endoplasmic reticulum (Kirkpatrick et al., 1995). The pMT/BiP/V5-his-A vector has ampicillin resistance as a selectable marker for cloning in *E. coli*, but it has no selectable marker for the drosophila S2 cells. This has the effect that generation of a stable cell line requires the expression vector to be co-transfected with a helper plasmid. In the case of MCP2, pCoHygro served this purpose.

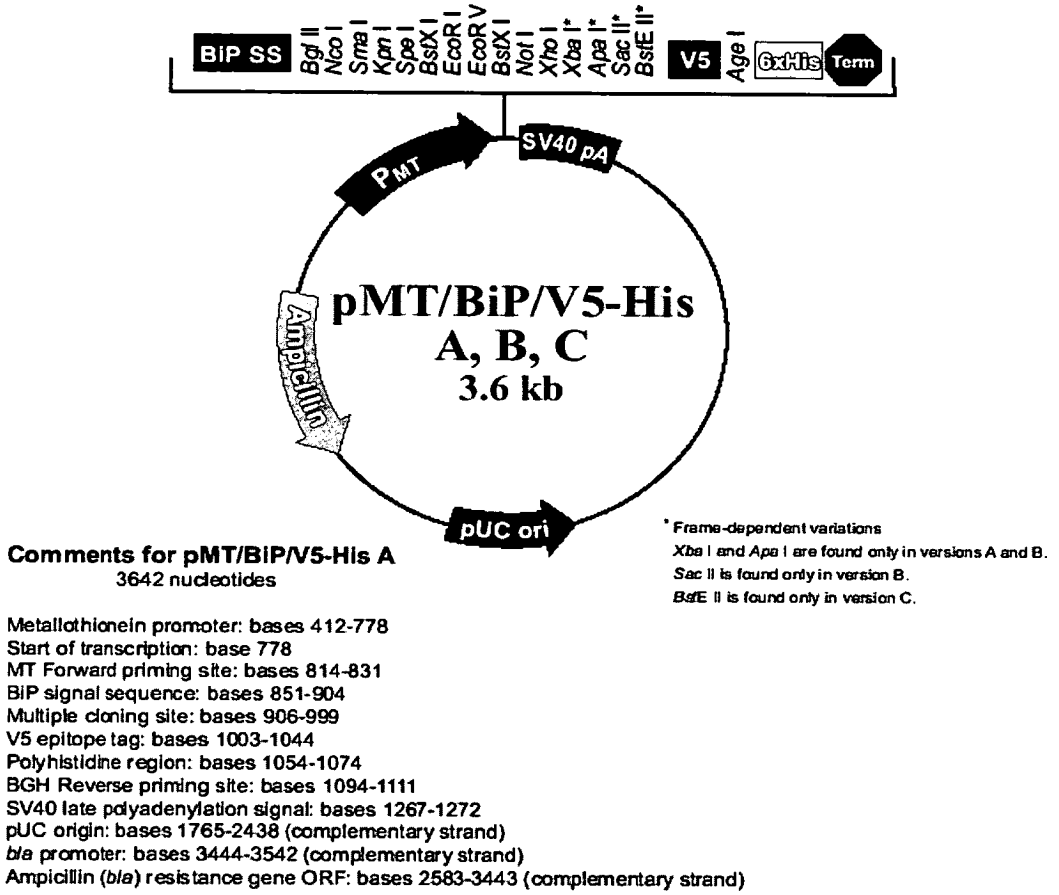


Figure 5:13 Vector map of the drosophila secreted expression vector pMT/BiP/V5-His.

The MCP2 coding sequence was obtained by digestion of MCP2 pET15b with *Nco*1 and *Xho*1, thus leaving the N-terminal His tag and thrombin cleavage sequence from pET15b. This fragment was ligated into pMT/BiP/V5-his-A, and used to transform *E. coli* OneShot Top10 cells, and positive transformants were selected using ampicillin resistance. The transformation was highly efficient due to the 'cut and paste' strategy, and therefore the plasmid from a single positive clone was prepared by "miniprep" and digested with *Nco*1 and *Xho*1. Subsequent electrophoresis on a 1.4 % agarose gel is shown in Figure 5:14 and indicates that the cloning was

successful. The identity and sequence of the clone was confirmed using automated DNA sequencing.

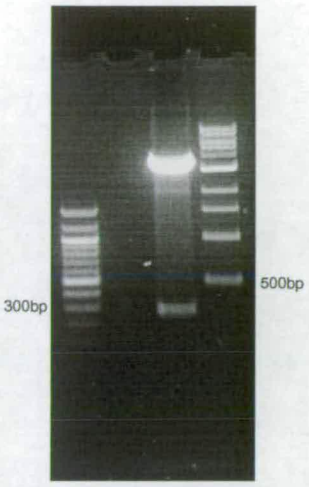


Figure 5:14 1.4% agarose gel showing MCP2 pMT digested with *Nco*1 and *Xho*1. The gel shows an insert of approximately 300bp. This agrees well with the predicted size of 320bp for MCP2 with a His tag and a thrombin cleavage sequence

Plasmid was introduced into S2 cells according to the calcium phosphate transfection protocol from the DES manual. A stable cell line expressing MCP2 was required, therefore the S2 cells (3×10^6) were cotransfected with MCP2 pMT/BiP/V5-his-A and pCoHygro (19 μg MCP2 pMT/BiP/V5-his-A and 1 μg pCoHygro). A stable cell line was achieved by selection with 300 $\mu g.ml^{-1}$ hygromycin-B for three weeks and three days, post transfection. Cells were induced for 24 hours using 500 μM copper sulphate. The supernatant was harvested by centrifugation, TCA-precipitated, and subjected to SDS-PAGE. No MCP2 could be observed on this gel. To determine whether any significant amount of MCP2 had been produced, purification on a nickel affinity column was performed. Due to the presence of the N-terminal His tag carried over from pET15b, it should be possible to purify this protein on a nickel afinity column. Therefore a larger-scale growth was performed, and the cells induced as before once the culture had been expanded to 20ml. This medium was

then concentrated to 1.0 ml using centrifugal concentrators with a MWCO of 3 kDa. The medium was then exchanged into 20 mM sodium phosphate, 100 mM sodium chloride, pH 7, and applied to a charged nickel affinity column. Following washing the bound protein was eluted with 350 mM imidazole, concentrated and subjected to SDS-PAGE. Following SDS-PAGE the proteins were transferred to a nitrocellulose membrane, and immunostained using anti His tag antibodies. This immunoblot blot is shown in Figure 5:15, and shows a faint band of approximately the correct size for the expected MCP2 construct, although it does run slightly smaller than would be expected.

Figure 5:15 Western blot of MCP2 expressed in drosophila S2 cells. The blot shows a faint band of approximately the correct size for this MCP2 construct.

This suggested the level of expression achieved using the DES was very low. It is possible that the expression levels are in fact higher than they first appear to be due to cleavage of the His tag, and a consequent reduction in sensitivity of the immuno-

detection system. The level of expression achieved using this system could only be estimated but was thought to be in the 10s of micrograms per litre range. Such expression levels would be considerably lower than those required for any biophysical analysis to be performed, and for this reason a larger scale-growth was not attempted. It was also presumed that the increases in expression levels required for biophysical characterisation (approximately 100-fold) would be unattainable.

5.3.1. Discussion

The results described in this chapter demonstrate the successful expression of MCP2 in *E. coli*, at levels sufficient for the subsequently required refolding step. NMR spectroscopic analysis of the 'native' MCP2 expressed in *E. coli* yielded spectra that were consistent with an almost completely unstructured protein. Following refolding however, the spectra were considerable improved. Comparison of the spectra generated by deglycosylated MCP2 with refolded MCP2 showed a number of similarities including the appearance of the dispersed peak at 11.1 ppm that is presumed to be the consensus tryptophan residue. Other similarly well dispersed peaks were also observed following refolding, however the spectra from refolded MCP2 suggest that refolded MCP2 is less well ordered than that of deglycosylated MCP2. This observation may imply a trend, whereby having no glycans results in a protein with a less ordered structure than one with a single GlcNAc residue, which in turn has a less well ordered structure than the protein with a native glycosylation profile. One cannot discount however, the possibility that the difference in the quality of these spectra is due to inefficient refolding of the *E. coli* material.

MCP2 was also successfully expressed in drosophila S2 cells, although in this case the quantities of protein obtained were insufficient for any significant analysis to be performed. It may have been possible to increase the levels of protein obtained using this system; however this was not attempted since the necessary increases (~ 100 fold) were thought to be unrealistically high.

6. SLAM1

6.1. Preliminary remarks

The recent discovery that, in addition to MCP, SLAM could serve as a receptor for measles virus (MV) prompted interest in the structural characterisation of the relevant region of SLAM. A direct comparison of the interactions of MV with SLAM and MCP could shed light on the nature of the initial virus-receptor interaction. The initial aim was to produce significant quantities of the domain of SLAM that interacts with MV. This work is described in the following sections.

6.2. Cloning of SLAM into pGEM-T

Since no direct source of the SLAM cDNA was readily available, a cDNA library was used as a template for amplification. Difficulties associated with using a cDNA library, include the complexity of the template, and the low abundance of the target cDNA. This results in increased risk of generating non-specific amplification products, both from mispriming, and due to contamination. In addition, commercial cDNA sources are expensive, and cannot be regenerated in house.

In an attempt to address some of these drawbacks, it was decided to clone the whole cDNA region encoding SLAM into a suitable cloning vector. The cDNA clone generated using this approach, once sequenced, would serve as a template for amplification of the required fragments of SLAM in all subsequent expression experiments. This approach permits the production of a copious supply of SLAM

cDNA from a single aliquot of a commercial cDNA library. In addition, this strategy facilitates improved primer design, since the primer sites are far less restrictive than those required for expression. A vector map of the cloning vector chosen for this purpose, pGEM-T, is shown in Figure 6:1. This vector has the advantage that it requires no restriction sites to be incorporated into the primers. Instead the insert is ligated by means of A-T base pairing. The vector is linear, with a single overhanging T at each 5' end, and pairing is achieved with the presence of a single overhanging A on the 3' ends of the PCR product. PCR products generated using Taq DNA polymerase, naturally have this 3' overhanging A (Clark, 1988), while PCR products generated by most of the popular proof-reading polymerases lack this overhang (Newton et al., 1997). The proof-reading DNA polymerase used in this case produces blunt-ended PCR products and therefore the 3' overhanging A must be added artificially by incubation with Taq in the presence of dATP (Promega, 2002). The pGEM-T vector also has the advantage of using blue-white screening as a result of the cloning region being located in the *lac* Z gene. Colonies not containing an insert therefore appear blue-black upon incubation in the presence of X-gal.

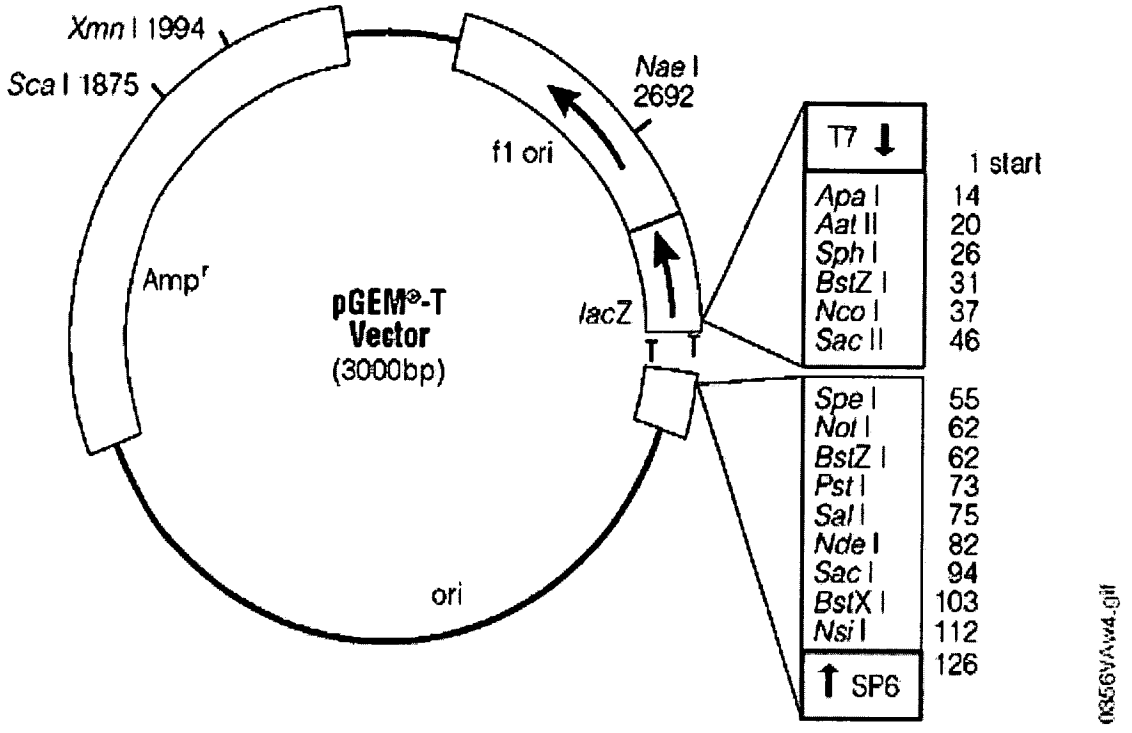


Figure 6:1 Vector map of the cloning vector pGEM-T showing its major features. The vector uses ampicillin resistance for selection in *E. coli*, and PCR products are inserted by means of A-T base pairing (taken from Promega, 2002).

The gene encoding SLAM, was amplified from the QuickClone cDNA library using PCR alongside a negative control as shown in Figure 6:2. The negative control was performed under exactly the same conditions, with the exception that no primers were included in the reaction mix. Surprisingly the negative control yielded a band of 1.5 kb, which was also present in the experimental reaction and was probably the result of contamination in the PCR mixture. In addition to the band found in the negative control, a band of just less than 1 kb was also seen. This band agrees well with the predicted length of the SLAM fragment, which was 936 bp.

Figure 6:2 Agarose gel showing the amplification of SLAM from the QuickClone cDNA library. The gel shows the presence of a band of the predicted size for the gene encoding SLAM, in the test reaction, which is absent from the negative control.

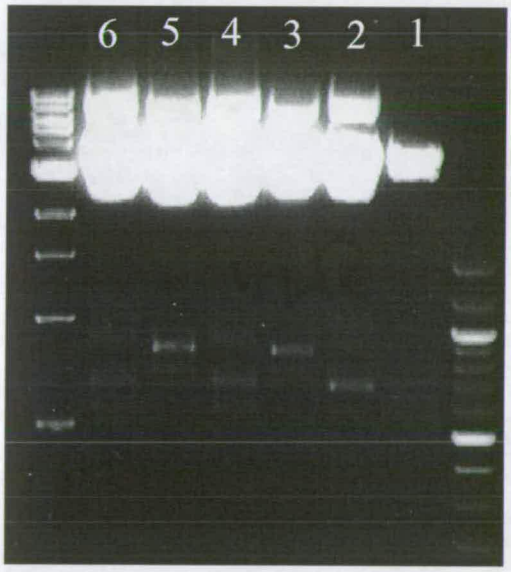


Figure 6:3 Digestion of SLAM pGEM-T with *Nco*1 and *Nde*1. The digested plasmids 3 and 5 give an insert of the predcted size for SLAM.

This band was purified from the gel, A-tailed using *Taq* DNA polymerase in the presence of dATP, ligated into the cloning vector pGEM-T, and used to transform chemically competent *E. coli* Top10 cells. The transformation yielded

approximately 500 colonies of which approximately 40 % were white colonies, signifying that they contain an insert. Six of these colonies were subjected to "miniprep", and the plasmids digested with *Nco*1 and *Nde*1 to reveal the size of the insert, as shown in Figure 6:3. Whilst it is clear that the digests did not go to completion, inserts of several different sizes can be seen. Representatives of each fragment length were sequenced, and clone 3 gave the published sequence for SLAM. This plasmid construct was then used as a template in the subsequent attempts to clone the SLAM1 fragment for expression in *P. pastoris*.

6.3. Expression of SLAM1 in *P. pastoris*

6.3.1. Cloning of SLAM1 in pPICZα

The N-terminal domain of SLAM had previously been shown to be necessary and sufficient for measles virus binding (Ono et al., 2001b). By comparing SLAM~1 both functionally and structurally to MCP12 it should be possible to understand better the initial steps of a measles virus infection.

The N-terminal immunoglobulin (Ig) domain of SLAM contains a putative disulphide bond, and is predicted be N-glycosylated at several sites. Therefore, for the reasons already described in chapter 3, initial attempts to express this domain of SLAM were performed in *P. pastoris*. The region encoding the N-terminal domain was amplified by PCR using the sequenced SLAM pGEM-T clone 3 as a template. Primers that incorporated asymmetric restriction sites (5' *Eco*R1 and 3' *Not*1) were

used to facilitate directional cloning required for the insert to be cloned in-frame with the alpha-factor secretion signal. The PCR product was then ligated into the *P. pastoris* expression vector pPICZα (see Figure 4:1), and used to transform chemically competent *E. coli* Top10 cells. The expression vector pPICZα was chosen over pPIC9 for the reasons detailed in section 4.2.1. The transformation yielded a large number of transformants, eight of which were selected for miniprep, and the plasmids were digested with *Eco*R1 and *Not*1, to determine the presence and size of any insert. The result shown in Figure 6:4 demonstrates that all the plasmids contained an insert of the correct size for SLAM1. Clones 2 and 6) were sequenced and both contained an insert of the published sequence for SLAM1. Clone 6 was taken and used in subsequent experiments.

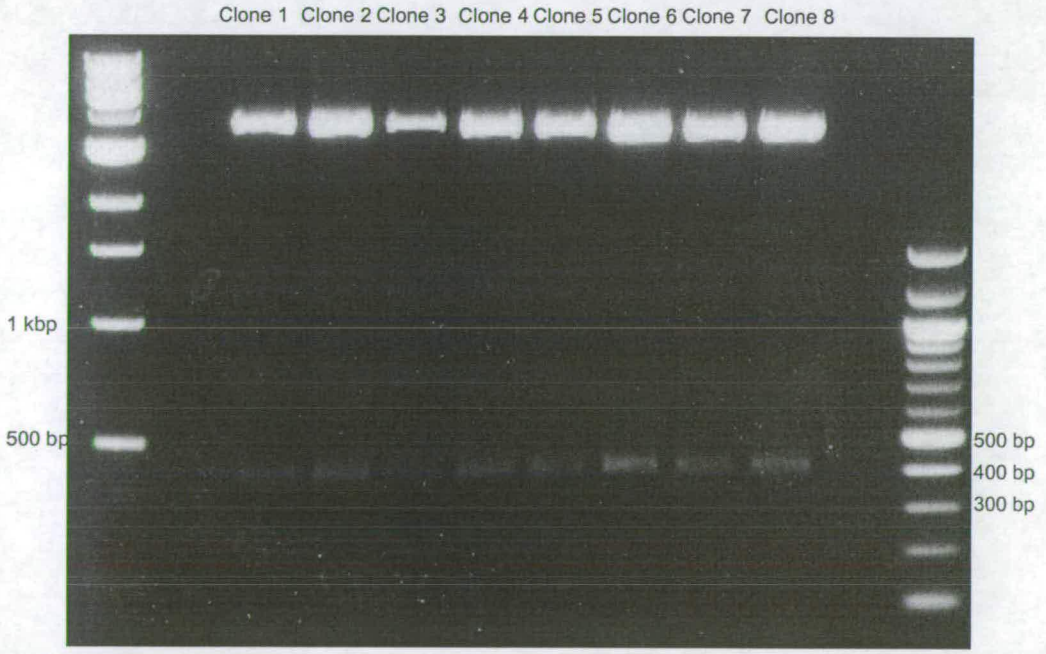


Figure 6:4 Agarose gel showing SLAM1 pPICZa clones digested with *Eco*R1 and *Not*1. The gel shows that each clone contained an insert of approximately 400 bp which corresponds well with the predicted size of 420 bp for the SLAM1 fragment.

The SLAM1 pPICZα plasmid was then linearised using *Sac*1 and 10 μg used to transform *P. pastoris* X33 cells by electroporation (1.5 kV, 200 Ω, 25 μF). The X33 strain of *P. pastoris* is a wild type strain that exhibits Mut^+ growth characteristics on methanol. The *P. pastoris* transformation yielded approximately 100 colonies that were able to grow on the selective medium (YPDS + 100 $\mu g.ml^{-1}$ zeocin). As is always the case for protein expression in *P. pastoris* it is useful to screen a number of colonies in search of a multi insertion event that often results in higher levels of expression (Higgins, 1998). Expression testing was performed using test inductions followed by SDS-PAGE. The induced culture supernatant was treated with $EndoH_f$ prior to SDS-PAGE in light of the experiences of hyperglycosylation gained with MCP. The SDS-PAGE gel is shown in Figure 6:5. It shows a band of approximately the predicted size for SLAM1 (15.6 kDa), however significant degradation is observed with each clone. Clone 12 was chosen for further analysis since it appeared to offer the optimum balance between expression level and degradation. It is possible that the observed degradation takes place during the deglycosylation reaction rather than during the induction phase, since the deglycosylation reaction requires incubation of the crude culture supernatant at 37 °C for extended periods.

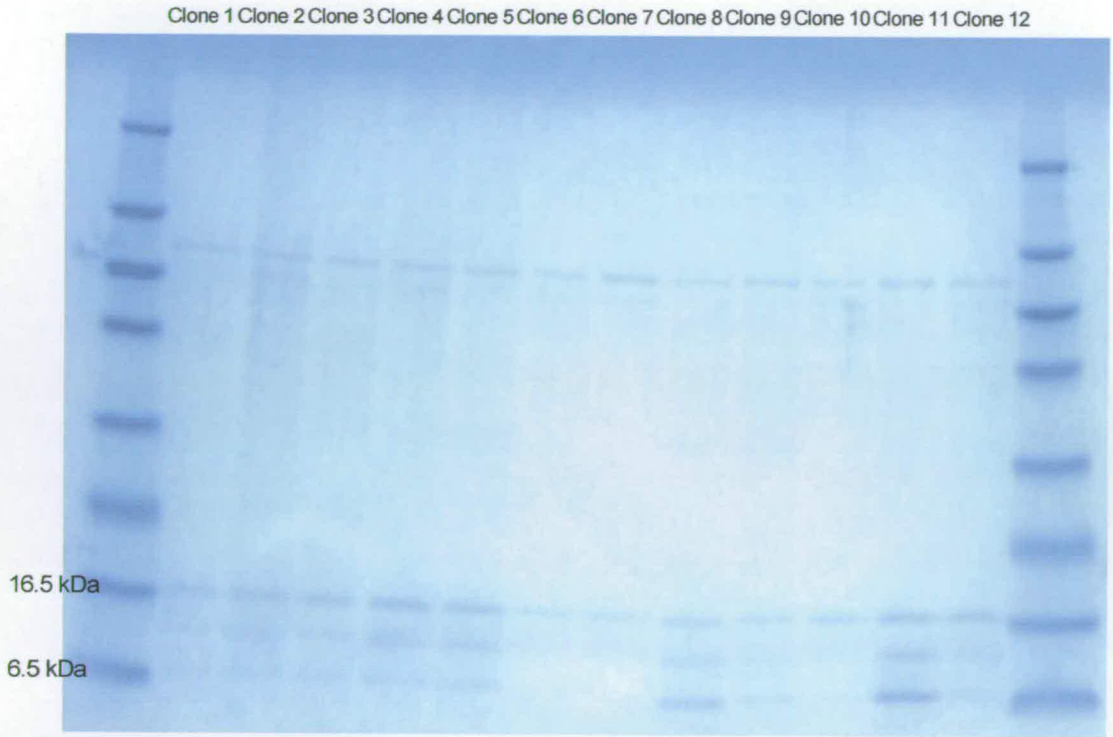


Figure 6:5 Test inductions for SLAM1. The gel shows a significant amount of degradation and apparent differences in expression levels. Clone 12 was chosen since it appeared to have the optimum balance between expression level and degradation.

6.3.2. Optimisation of expression conditions for SLAM1 in X33

As previously alluded to, the growth characteristics for the X33 strain are significantly different from the KM71 strain. These differences occur due to the methanol utilisation phenotype of X33 being Mut^+ compared to the Mut^S phenotype of the KM71 strain. Mut^S strains of *P. pastoris* exhibit considerably faster growth on methanol, compared to Mut^+ strains, and therefore it is worthwhile, in this strain, optimising the growth conditions for the induction phase. This was investigated using four induction flasks, each containing 250 ml of SLAM1 X33 in BMM. Each of these cultures originated from the same resuspended cells obtained during the

growth phase. In each flask the starting methanol concentration was kept at only 0.5% (v/v), to prevent cell death due to methanol poisoning before the cells had time to express the alcohol oxidase gene. The culture was subsequently induced for four days with daily feeds of methanol at concentrations of 0.5 %, 1 %, 2 % and 4 % (v/v). Samples were taken at regular intervals and A_{600}'s measured to follow growth. The culture supernatant was retained in each case to determine the level of expression by SDS-PAGE and the growth curves are shown in Figure 6:6. These indicate that growth rate and extent of growth are highest for the culture grown on 2 % methanol, and that increasing the methanol concentration between 0.5 % and 2 % results in significantly higher growth. Furthermore, 4 % methanol results in less growth than 1 % methanol, indicating that growth is inhibited by the higher levels of methanol. The optimum methanol concentration appears to lie between 2 % and 4 %. The level of SLAM1 expression for each of the methanol concentrations tested was also monitored based on estimates using SDS-PAGE following treatment of the relevant culture supernatants with $EndoH_f$. The gel shown in Figure 6:7,indicates that SLAM1 expression also happened to be optimal at 2 % Further growths were thus performed using a methanol concentration of 2 % during the induction phase.

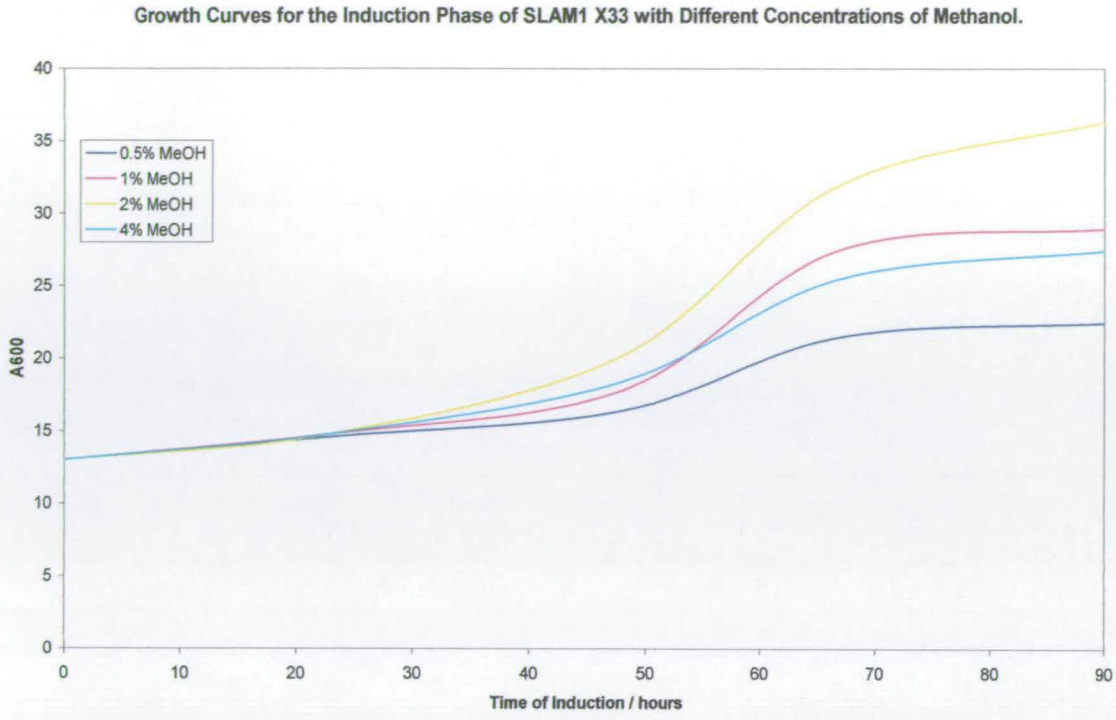


Figure 6:6 Growth curves for the induction phase of SLAM1 in *P. pastoris* X33, with different concentrations of methanol. The graph shows that both the rate and extent of growth of the cells increases with increasing methanol concentration, until at 4 % methanol the growth becomes inhibited. This result suggests that the optimum concentration of methanol for the growth of SLAM1 X33 cells during induction is 2 %.

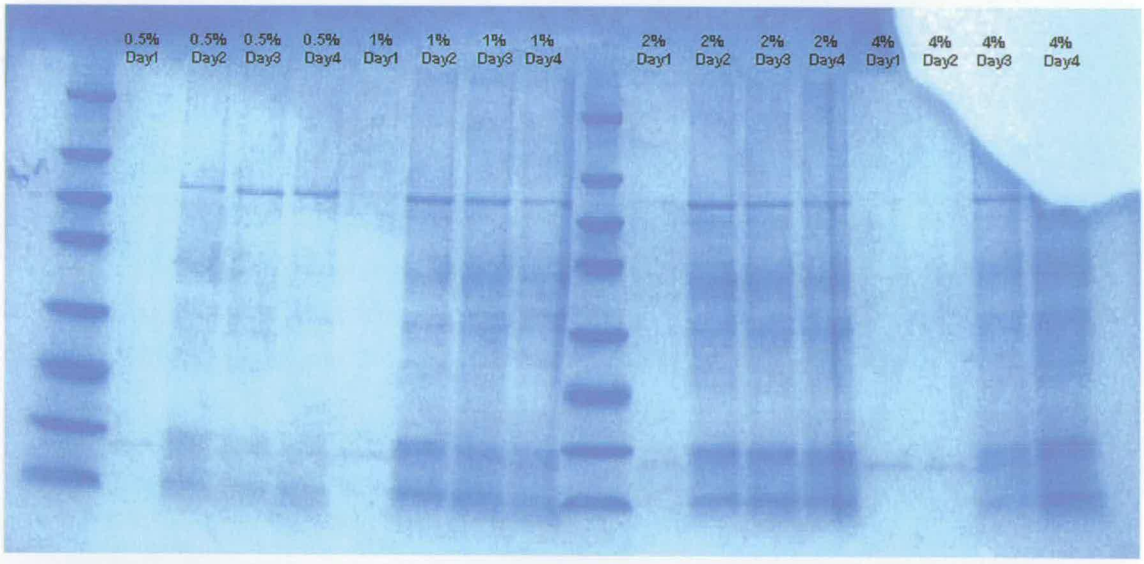


Figure 6:7 SDS-PAGE showing optimisation of SLAM1. The gel shows that the optimum methanol concentration for induction of SLAM1 is 2 % and that the optimum induction time is four days.

Degradation, as had been observed in the test inductions, was also observed during the growth optimisation test. While expression levels increased with increasing induction time, the proportion of SLAM1 that is degraded remains relatively constant throughout the induction period. This suggests that the optimum induction time for SLAM1 is four days, and therefore subsequent growths were performed using this induction time. The level of expression of SLAM1 was estimated on the basis of SDS-PAGE to be in the region of 10 mg per litre of BMM. This estimate suggests that SLAM1 is expressed at a comparable level to that achieved for the MCP fragments.

6.3.3. Verification that the expressed species was indeed SLAM1

To determine that the species observed on SDS-PAGE was indeed SLAM, an immuno-blot was performed. Following SDS-PAGE, the gel was blotted onto a nitrocellulose membrane, and immuno-stained using anti-SLAM antibodies as the primary antibody (Nakamura et al., 2001). Antibodies were raised against was the N-terminal ten residues of human SLAM. The immuno-blot shown in Figure 6:8 clearly demonstrates that deglycosylated SLAM1 is recognised strongly, and that the non-deglycosylated samples give little or no reaction. This supports the notion that the protein expressed is indeed SLAM1. Moreover, the fact that the epitope used to raise primary antibody was mapped to the N-terminus of human SLAM, implies that the expressed protein must have an intact N-terminus.

Figure 6:8 Immunoblot of SLAM1. The blot shows that the expressed species is recognised strongly by anti-SLAM antibodies. The deglycosylated samples yield a single, strongly stained band of the correct size for SLAM1, whereas the untreated sample either stains only poorly or the band is highly diffuse.

In addition the observation that non-deglycosylated SLAM1 does not react with anti-SLAM antibodies, whereas deglycosylated SLAM1 reacts strongly provides reasonable evidence that the recombinant SLAM1 expressed by *P. pastoris* is glycosylated. Moreover, these data further suggest that the recombinant SLAM1 is hyperglycosylated for the following reason. If it were not hyperglycosylated, one would expect to observe a reasonably discrete band, with a lower mobility on SDS-PAGE, which is not observed in this case. The almost complete absence of an observable reaction to anti-SLAM antibodies could result from the glycans masking the epitope for the anti-SLAM antibodies. Since there is no glycosylation site in the epitope used to raise the antibodies, this would seem to be unlikely. The other possible explanation would be that hyperglycosylation of SLAM1 results in a very

diffuse smear on a gel, like that observed with MCP, and that the concomitant reduction in the quantity of protein in each band reduces the signal from the immuno-staining procedure to below the threshold of detection.

6.4. Attempts to purify the recombinant SLAM

6.4.1. Purification of SLAM1 on ConA sepharose

Before describing the purification of SLAM1 it is worth recapping what is already known about this protein. It is highly glycosylated, of approximately 15.5 kDa, and has a predicted pI close to 9 (using the Expasy protparam protein parameter calculator - http://us.expasy.org/tools/protparam.html). One can attempt to exploit any of these properties during purification. Based on the fact that SLAM1 expressed in *P. pastoris* is highly glycosylated, and that the glycosylation pattern found in *P. pastoris* proteins is of the high mannose type, it was decided that the use of ConA sepharose could make a useful first step of purification. It should allow the protein to undergo partial purification prior to deglycosylation, thus reducing the risk of proteolysis and degradation. The culture supernatant was concentrated to 20 ml using ultrafiltration (5 kDa MWCO) as described for the MCP fragments. The protein was then exchanged into the ConA binding buffer (500 mM NaCl; 50 mM Tris, pH 7.5) using a PD10 desalting gel filtration column. This solution (2 ml) was applied to a gravity-fed ConA sepharose column (5 ml) (Amersham Pharmacia Biotech), and 2 ml fractions collected. Unbound protein was washed away with binding buffer, and bound proteins eluted with a step-gradient of elution buffer (500

mM NaCl; 50 mM Tris pH 7.5; 2 M methyl-α-D-gluco-pyranoside). The presence of protein was traced using absorbance at 280 nm.

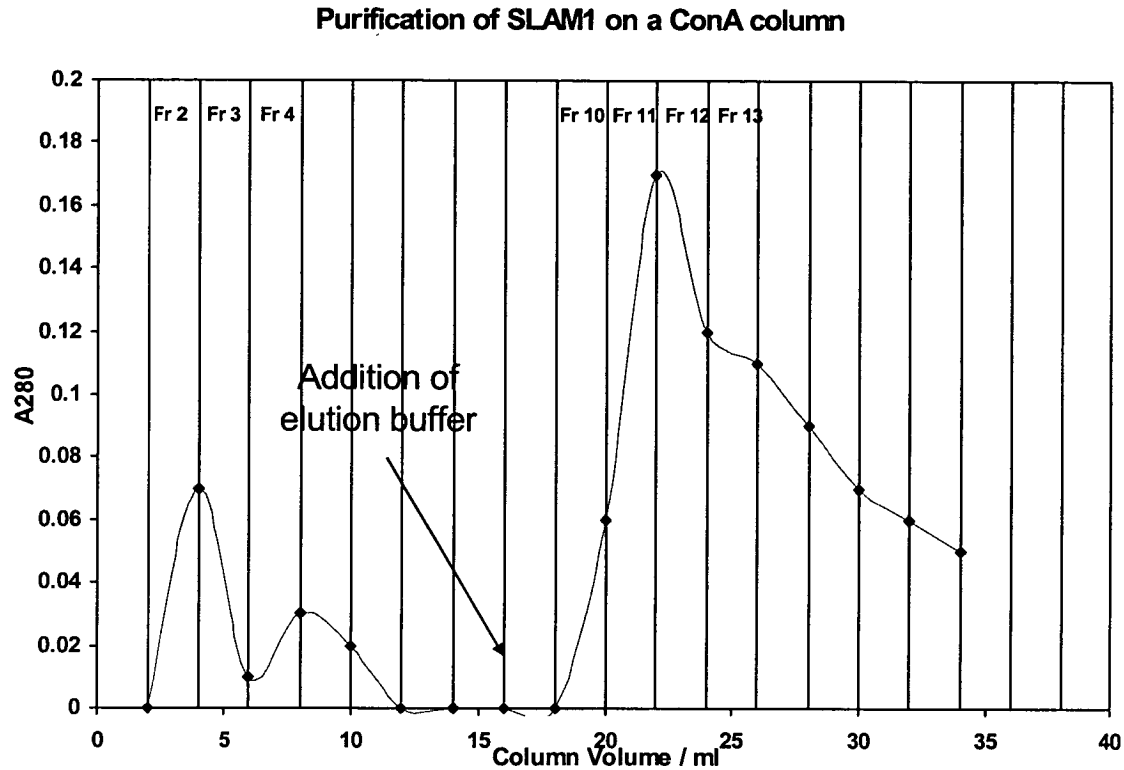


Figure 6:9 ConA purification of SLAM1. The chromatogram shows a significant amount of protein in both the bound and unbound fractions, indicating that a degree of separation had been achieved.

The chromatogram shown in Figure 6:9 initially appears promising, with a significant amount of protein in the unbound fractions, and a proportion bound to the column and eluted subsequently with elution buffer. The presence of SLAM1 was confirmed for the fractions numbered in the chromatogram (Figure 6:9), by using SDS-PAGE after $EndoH_f$-deglycosylation of the fractions The resulting gel is shown in Figure 6:10, and appears to indicate that there is significant SLAM1 present in both the bound and unbound fractions. Moreover, there is a significant quantity of a

larger, contaminating protein (~ 65 kDa), present in the bound fractions. The quantity of crude concentrate applied to this column (estimated to be less than 1 mg of total protein) should be easily within the binding capacity of the column (~ 30 mg of porcine thyroglobulin per ml of gel).

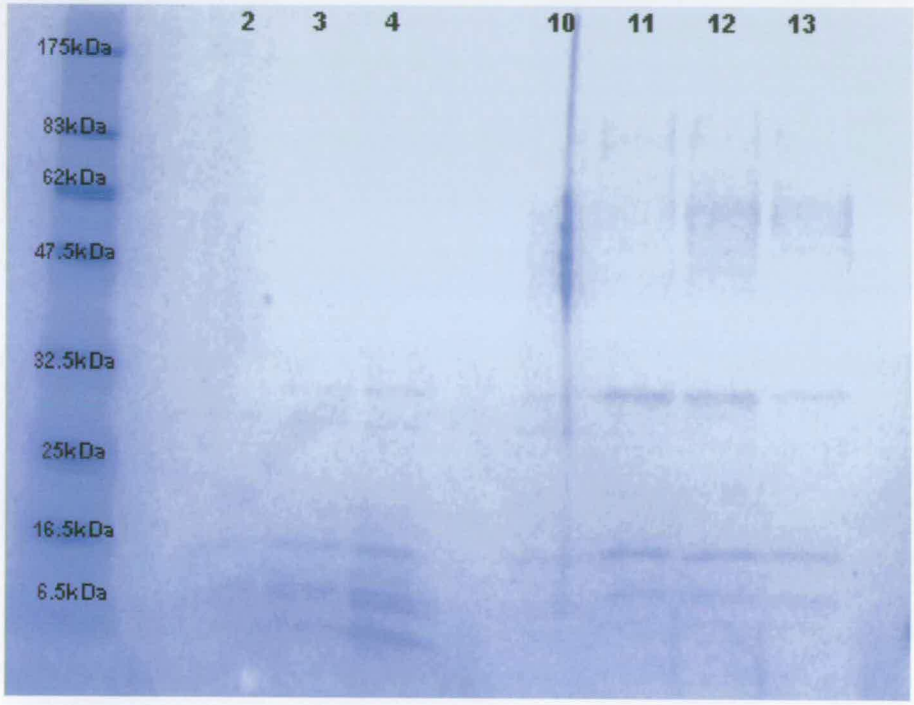


Figure 6:10 SDS-PAGE showing the result of purification of SLAM1 on ConA. The gel shows that significant amounts of SLAM1 are present in both the bound and unbound fractions. Moreover, large amounts of a ~ 70 kDa protein are present as a contaminant in the bound fractions.

It was speculated that the reason for poor binding of SLAM1 to ConA sepharose was the presence of non-covalently attached carbohydrates that have previously been found on proteins expressed by *P. pastoris* (Denton et al., 1998). In order to determine whether this was the case, a protein blot was prepared with glycosylated and deglycosylated SLAM1. This blot was stained using the glycoprotein detection

kit (Sigma) which is an adaptation of the periodic acid-Schiff method (Devine, 1990). This staining method effectively stains vicinal diol groups that are found on many of the glycans found in glycoproteins resulting in a deep magenta colour.

Figure 6:11 Blots of SLAM1 stained using the glycoprotein detection kit from Sigma. The blot shows the presence of significant amounts of glycans, whose mobility on SDS-PAGE remains relatively unchanged following treatment with $EndoH_f$ and / or PNGaseF.

The blot shown in Figure 6:11 indicates that the samples indeed contain large amounts of glycans that have a low mobility on SDS-PAGE. That the mobility of these glycans is largely unaffected by treatment with $EndoH_f$ and /or PNGaseF, may imply that the glycans are non-specifically bound to proteins; that they are unattached to any proteins; or that the glycoprotein to which they are attached is

insensitive to either $EndoH_f$ or PNGaseF. That glycans may co-purify with proteins expressed in *P. pastoris* has previously been observed in the case of bovine beta-lactoglobulin (Denton et al., 1998). It is interesting , however, that the glycans appear to have a similar mobility to the contaminating protein found in the bound fractions from the ConA purification step. Together with the lack of sensitivity to $EndoH_f$ and PNGaseF, this implies a different form of glycosylation (eg O-linked glycosylation). The presence of such large quantities of glycans in the crude SLAM1 samples, which may be binding to the ConA column in preference to SLAM1, could cause a dramatic reduction in the capacity of the column. For this reason it may prove useful to attempt to remove some of these glycans prior to purification with ConA sepharose.

6.4.2. Reverse phase purification of SLAM1

It still remains a goal to use ConA affinity chromatography as a purification step for the production of SLAM1, but it appears that the majority of the non-covalently attached glycans need to be removed first. Reverse phase chromatography is a suitable method for this initial purification step since glycans are generally less hydrophobic than most proteins. Moreover, reverse phase chromatography could in any case provide a useful means of purifying SLAM1 from other contaminating proteins. Crude, concentrated SLAM1 was applied to a C4 reverse phase chromatography column pre-equilibrated with 90 % H_2O (0.05 % v/v TFA), 10 %

acetonitrile (0.05 % TFA). Bound proteins were eluted with a linear gradient to 60 % acetonitrile.

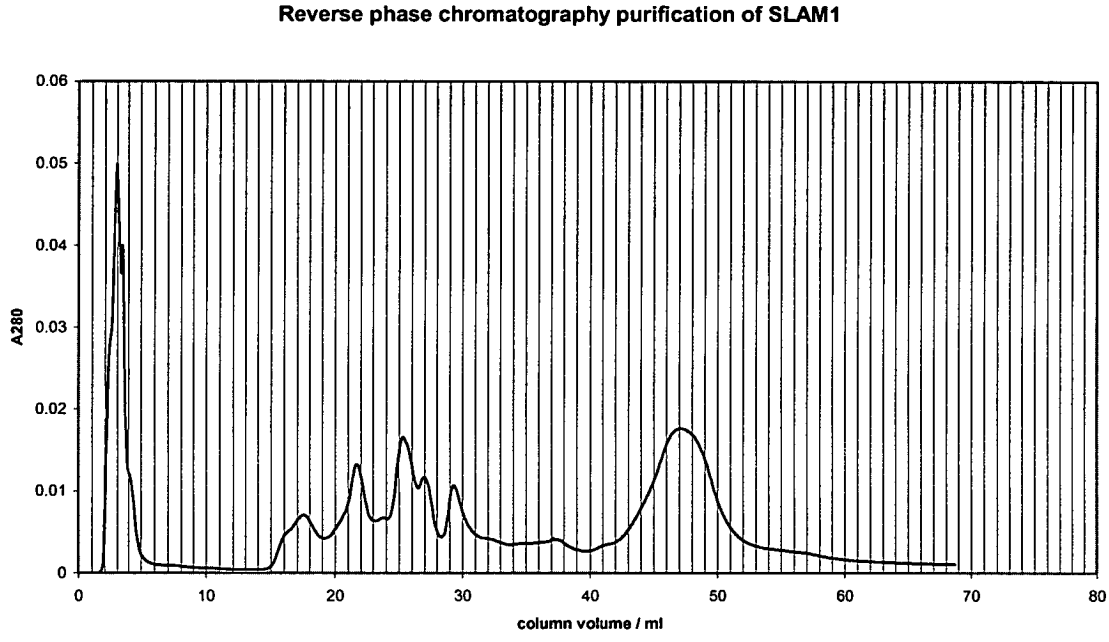


Figure 6:12 Purification of SLAM1 on reverse phase chromatography. Fractions of 1 ml were collected deglycosylated and subjected to SDS-PAGE.

The reverse phase chromatogram shown in Figure 6:12 appears to show that several species have been separated. The fractions were lyophilised, deglycosylated with EndoH_f and subjected to SDS-PAGE. This gel shown in Figure 6:13, indicates that the SLAM1 co-elutes with large amounts of a protein of approximately 80 kDa, which appears to resemble the protein that co-purifies with SLAM1 following ConA purification. Further attempts to purify SLAM1 by reverse phase chromatography used stationery phases with different hydrophobicities, but the fraction containing SLAM1 always co-purified with this protein (data not shown).

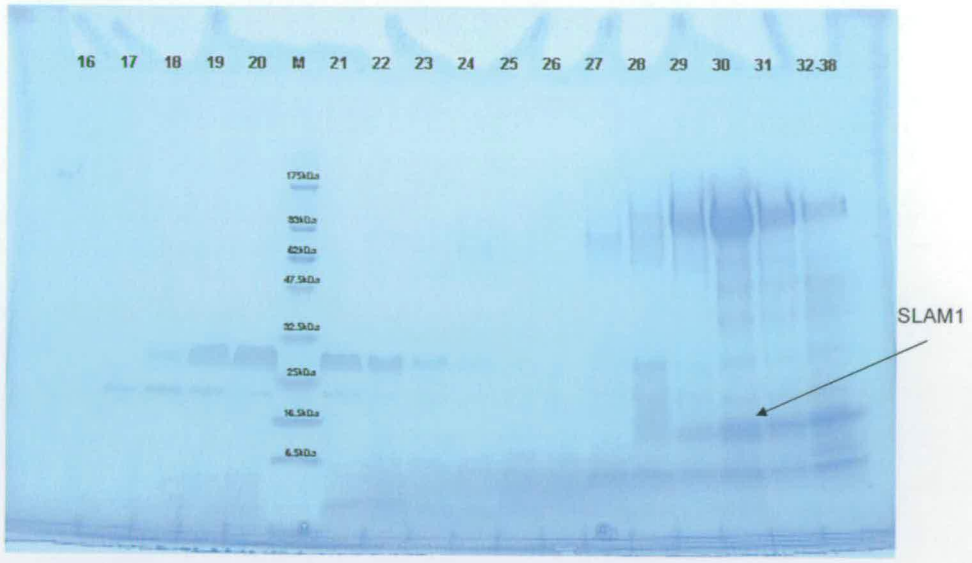


Figure 6:13 SDS-PAGE showing the purification of SLAM1 on a C4 reverse phase column.

The difficulties experienced purifying SLAM1 using reverse phase chromatography necessitated the exploration of other chromatography techniques.

6.4.3. Hydrophobic interaction chromatography purification of SLAM1

Given the success achieved in purifying MCP fragments using HIC, it was thought that this type of chromatography could also prove useful for purifying SLAM1. Although HIC works on similar principles to reverse phase chromatography, the fact that the protein is probably unfolded in the latter but not in the former provides the potential for differential results. To this end a number of different HIC stationery phases were explored, including phenyl, octyl, butyl, and isopropyl substituted media. Mobile phases containing 2 M, 1.5 M, and 1 M $(NH_4)_2SO_4$ were also tested, and of these binding to a phenyl Resource column (Amersham Pharmacia Biotech) in 2 M $(NH_4)_2SO_4$, 50 mM Tris, pH 7.5, yielded the most promising results.

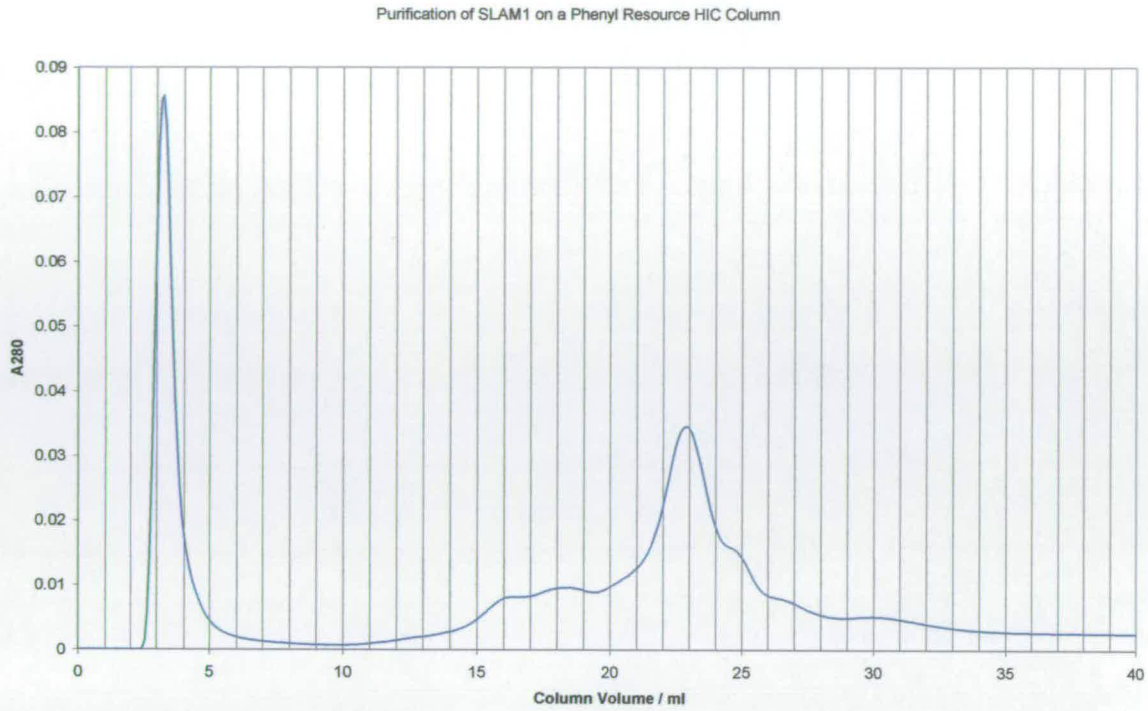


Figure 6:14 Hydrophobic interaction chromatography purification of SLAM1 on a phenyl resource column

Figure 6:15 SDS-PAGE showing hydrophobic interaction chromatography purification of SLAM1 on a phenyl resource column

The crude SLAM1 concentrate was applied to a phenyl Resource column pre-equilibrated with 2 M $(NH_4)_2SO_4$ 50 mM Tris pH 7.5. The chromatogram shown in Figure 6:14 appears promising, as it resembles the chromatogram obtained for the MCP fragments. The bound peak is broad, but this would be expected given the heterogeneity of the glycosylation profile. The SDS-PAGE gel shown in Figure 6:15 reveals, however, that once again the SLAM1 fragment co-purifies with abundant quantities of a protein of ~ 80 kDa. This contaminating protein is presumably the same material that had co-purified with SLAM1 during the previous purification attempts.

6.4.4. Cation exchange purification of SLAM1

Given the unusually high predicted pI of the SLAM1 fragment, cation exchange chromatography appeared to be a potentially useful method of purification. The SLAM1 was deglycosylated prior to cation exchange chromatography to reduce its heterogeneity, and thus to facilitate uniform behaviour during cation exchange. Initial attempts at cation exchange on a MonoS HR 5/5 column (Amersham Pharmacia Biotech) were performed in 50 mM MES pH 6.0, and the chromatogram shown in Figure 6:16 indicates that the SLAM1 fragment did not bind to the column; indeed this was subsequently confirmed by SDS-PAGE of the flow-through fractions (shown in Figure 6:17).

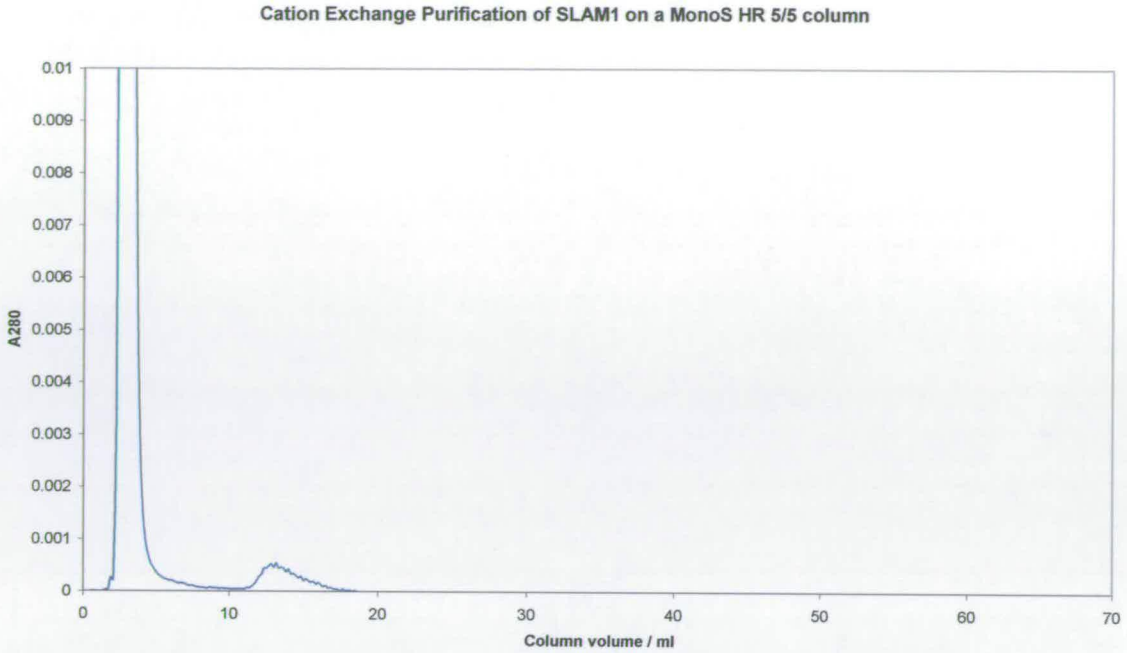


Figure 6:16 Cation exchange purification of SLAM1 on a MonoS Hr 5/5 column. The protein was applied to the column in 50 mM MES pH 6.0. It can be seen from the chromatogram that negligible quantities of protein are present in the bound fractions.

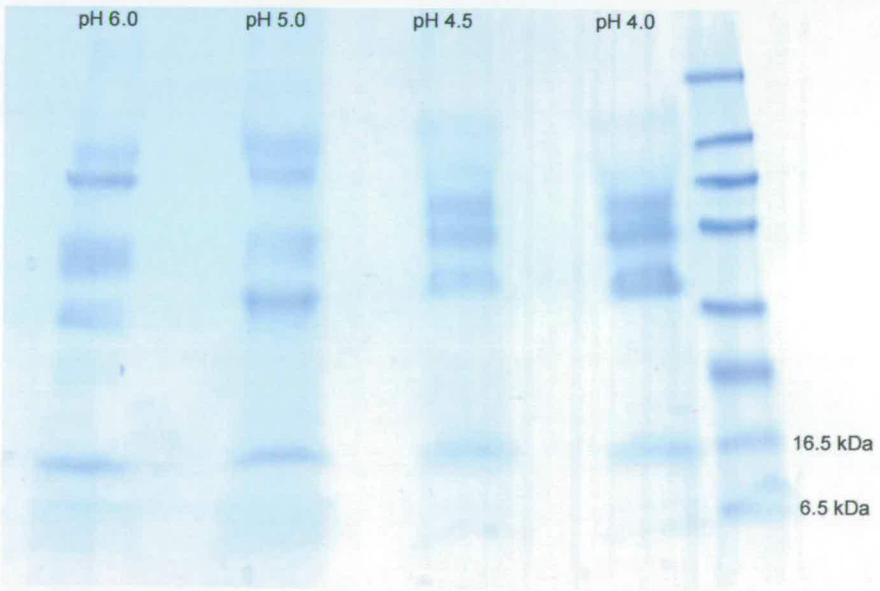


Figure 6:17 SDS-PAGE of the flow-through fractions from the MonoS HR 5/5 run under a number of different pH conditions

Since the SLAM1 fragment did not bind to the MonoS HR 5/5 column under a range of conditions, it was decided to attempt cation exchange. Sodium acetate buffers (20 mM) at pH 5.0; 4.5; and 4.0 were used to equilibrate the column. SDS-PAGE showing the flow through fractions at each of these different pH conditions is shown in Figure 6:17, and demonstrates that the SLAM1 passed through and did not bind to the resin.

6.5. Discussion and future work

The results presented in this chapter describe the cloning and expression of SLAM1 in *P. pastoris*. Following optimisation of the growth and induction conditions, SLAM1 was expressed at levels of approximately 5 - 10 mg per litre of induction medium. Confirmation that this material was indeed SLAM1 was through the use of anti-SLAM antibodies during immuno-blot analysis. As was the case with the MCP fragments, it was determined that SLAM1 was hyperglycosylated - this is perhaps not surprising given the number of glycosylation sites present in the protein.

Despite the success achieved with the expression of SLAM1, a detailed analysis of this protein was negated by the difficulties encountered during the subsequent purification attempts. Under all of the conditions tested SLAM1 either failed to bind to the column, or co-purified with abundant quantities of a contaminating protein of ~ 80 kDa. To address this problem, gel filtration chromatography on a Superdex 75 column (Amersham Pharmacia Biotech) was considered; however the viscosity of crude material from *P. pastoris* would have a detrimental effect on both the resolution and integrity of the column and was therefore not attempted.

Following the failure of the battery of chromatographic techniques tested above, it is probable that a purification tag may be required. A number of factors, however, complicate the use of such tags in this instance. The available tags in the pPICZα vector is a C-terminal $(His)_6$ tag, that follows the C-*myc* antibody epitope. This tag is uncleavable, and would result in the addition of a minimum of 23 extra amino acids, adding approximately an extra 20 % to the size of the protein. In addition, $(His)_6$ tags, or the subsequent immobilised metal ion adsorption chromatographic (IMAC) purification have been shown to affect the quality of a number of proteins expressed for structural biology (for a review see Ramage et al., 2002). Moreover, the successful use of such tags for the purification of proteins expressed in *P. pastoris* is not widely documented. It may also be possible to express SLAM1 in *E. coli*, as it only has one putative disulphide bond and may be amenable to refolding; however time constraints prevented this from being attempted.

7. CONCLUSIONS AND FUTURE WORK

The current debate as to the possible identities of the measles virus receptor is still on-going. A detailed appreciation of the initial processes in a measles virus infection, however, will only be realised with an in depth understanding of the structure and physiochemical behaviour of the viral receptor.

It was the aim of this study to express and characterise the candidate receptors, as a first step in a larger effort to elucidate both their 3D structures, and their precise modes of interaction with the binding partners on the virus. In the process fundamental questions were addressed regarding the functions of the N-linked glycans of the second module of MCP.

Initial efforts were directed towards a study of the region of MCP identified as being necessary and sufficient for measles virus binding – MCP12 (residues F1→ L131). The methylotrophic yeast, *P. pastoris* was selected as the most appropriate host, and was successfully engineered to secret MCP12 at levels sufficient for biophysical characterisation (~10 mg MCP12 per litre of induction medium).

The recombinant protein was hyperglycosylated. The excessive extent and heterogeneity of glycosylation rendered a detailed NMR spectroscopic analysis of this glycosylated form impractical. For such a study to be performed on MCP12, either the glycosylation sites would have to be removed by mutagenesis, or the glycans would require removal post-expression. The former route was not explored since it had already been demonstrated that mutation of the glycosylation site in

module 1 of MCP had a detrimental effect on the protein conformation and resulted in a largely undefined structure (O'Leary, 2000) - this is despite the lack of any known functional effect of the glycan on module 1 of MCP. Preliminary NMR analysis of enzymatically deglycosylated MCP12 indicated that this protein was not fully folded. Comparison with NMR spectra of recombinant MCP1 revealed that while the first module appeared to be properly folded, the second module was not. This observation fits with the known functional requirement of glycosylation in module 2 of MCP (Maisner et al., 1996), and with the observation in the crystal structure of relatively extensive contacts between these glycans and the polypeptide chain (Casasnovas et al., 1999) in module 2. It would have been possible to pursue further NMR studies of this sample of MCP12. Further efforts, however, were considered unlikely to advance significantly understanding of the structure of MCP12. The crystal structure is already known, and a solution structure of MCP1 has been solved – therefore the main benefit of a solution structure of MCP12 would lie in the information it could provide on the intermodular interface and this relies on properly folded protein that yields high quality NMR spectra.

The observation that glycosylation might be critical for structural integrity merited further investigation. To this end, a glycosylated sample was analysed in a similar manner to that used for the deglycosylated protein.. Whilst detailed NMR analysis of glycosylated MCP12 could not be performed for the reasons given above, several experiments were nevertheless possible. The data acquired from these experiments implied that glycosylated MCP12 had a better defined structure than did deglycosylated MCP12 and that therefore the glycans play a structural role. Whilst

uncommon similar findings have been reported with other proteins (De Beer et al., 1996; van Zuylen et al., 1997; Wyss et al., 1995). Further analysis of the hyperglycosylated and deglycosylated forms of MCP12 by DSC indicated that these glycans did not directly stabilise the protein fold. The lack of a stabilising effect could be an artefact of hyperglycosylation as opposed to physiological glycosylation. When taken together these data led to the conclusion that producing the component modules of MCP separately would be useful for further progress. Since MCP1 has been characterised in isolation, further efforts were focussed on MCP2.

The success achieved with the expression of MCP12 in *P. pastoris* led to its use for expression of the single module fragment of MCP2. The fragment was also successfully secreted, although at slightly higher levels than those obtained for MCP12. As predicted the MCP2 fragment was also hyperglycosylated, with approximately a three fold greater mass of sugar than protein. The NMR experiments performed on glycosylated MCP2 yielded spectra indicative of a highly structured protein. Detailed analysis of hyperglycosylated MCP2 by NMR spectroscopy could not be performed for the reasons given above. The NMR experiments performed on deglycosylated MCP2 on the other hand yielded spectra that were indicative of a protein with a more poorly defined structure, and was potentially overly flexible. DSC analysis on both hyperglycosylated and deglycosylated MCP2, however indicated that the glycans have no significant effect on the thermal stability of MCP2. At first analysis, these data appear to be inconsistent with the NMR data, however the lack of a significant stabilising effect of these glycans could also be an artefact of hyperglycosylation. Additionally it is

possible that the flexibility of a protein can be reduced without necessarily increasing its thermal stability.

To discount the hypothesis that it is the deglycosylation reaction itself that is responsible for the observed differences between the spectra of glycosylated and deglycosylated MCP2, a sample with native sequence that had never been glycosylated was required. Refolding a protein is often associated with high protein losses, and therefore it is desirable to express a large quantity of potentially unfolded protein prior to refolding. This end is usually most successfully achieved through expression in *E. coli* which generally gives higher levels of recombinant protein than other organisms. Initial expression attempts for MCP2 in E. coli produced MCP2 in sufficient quantities, however the protein was found entirely in inclusion bodies, which can cause significant downstream problems. Using molecular biology to prevent the protein going into inclusion bodies, can often prove more successful than attempting to solubilising inclusion bodies. Efforts were therefore focused on generating soluble protein through the use of the thioredoxin fusion tag. Expression of the subsequently generated fusion protein resulted in the fragment being found entirely in the soluble fractions. NMR spectroscopic analysis of the purified soluble protein, revealed a spectrum with almost no dispersion from random coil. The NMR experiments performed on refolded E. coli MCP2 however yielded spectra with considerably more dispersion, indicating that the refolded MCP2 has acquired some structure. Comparison of the spectra from refolded MCP2 and deglycosylated MCP2 shows that whilst the spectra are similar indicating that refolding was largely successful.

Expression of MCP2 from *D. melanogaster* S2 cells, proved successful, however the quantities of protein produced were insufficient for any significant characterisation to be performed. Moreover, the requirement to increase expression levels by approximately 100 fold appears unfeasible. It therefore remains a goal to produce MCP with a small and well defined glycosylation profile on module 2. In light of the experience gained through using different expression systems, it seems likely that some non-recombinant method may prove more successful. One such possibility is through the use of N-(beta-saccharide) haloacetamides (Wong et al., 1994) to specifically attach synthetic glycans to free cysteines in a recombinant protein. In this case, of particular interest would be the use of this type of system to attach the chitobiose saccharides, to a mutant protein where the asparagine residue of the glycosylation site has been replaced with a cysteine. This approach could potentially mimic the interactions seen in the crystal structure.

It appears to be an intrinsic property of certain CCP modules to exhibit high flexibility. The failure to produce a CCP module containing proteins that give the very high quality spectra required for structural studies is a problem that has been encountered previously in our and other laboratories. Examples of these include VCP12, CR2~12, and GABAB receptor 1A modules 1-2. Other, CCP module proteins prove readily amenable to structural studies including VCP23 (Henderson et al., 2001), and fH19-20 (A. Herbert, University of Edinburgh, unpublished results).

The work carried out on SLAM was performed in an attempt to gain a fundamental understanding of the mode of action of measles virus binding. This insight would be

the result of considerable characterisation and preferably structural data, on both of these proteins. Whilst this goal was perhaps beyond the scope of this thesis, recombinant *P. pastoris* has been generated that expresses SLAM1 at significant levels. Its purification proved troublesome due to both the continued presence of a large contaminating protein, and the heterogeneity associated with hyperglycosylation. The use of purification tags and alternative expression systems could potentially provide a route to the generation of this fragment with the yields and purity required for biophysical studies.

Appendix A: Buffers and Media Recipies

Towbin buffer:	Tris Base		3.03gl-1
	Glycine		14.4gl-1
	Methanol		100mll-1
TAE (Tris Acetate EDTA) buffer	Tris Base		4.84 gl-1
	Glacial acetic acid		1.142 gl-1
	0.5M EDTA		2 mll-1
TGS (Tris Glycine SDS) buffer	Tris base	25 mM	3.02 gl-1
	Glycine	250 mM	18.8 gl-1
	SDS	0.1% (w/v)	1 gl-1
2* SDS PAGE loading buffer	Tris pH 6.8	100mM	
	Dithiothreitol	200 mM	
	SDS	4% (w/v)	
	Bromophenol blue	0.2% (w/v)	
	Glycerol	20% (v/v)	
LB (Lauria Bertani) medium	Tryptone	1% (w/v)	10 gl-1
	Yeast Extract	0.5% (w/v)	5 gl-1
	NaCl	1% (w/v)	10 gl-1
	pH 7.0		
Low salt LB medium	Tryptone	1% (w/v)	10 gl-1
	Yeast Extract	0.5% (w/v)	5 gl-1
	NaCl	0.5% (w/v)	5 gl-1
	pH 7.0		
YPD (Yeast Extract, Peptone, Dextrose) mediun	Yeast extract	1%	10 gl-1
	Peptone	2%	20 gl-1
	Dextrose	2%	20 gl-1
BMG (Buffered Minimal Glycerol) medium	potassium phosphate p	100mM	
	YNB	1.34%	
	Biotin	4*10-5%	
	Glycerol	1%	
BMM (Buffered Minimal Methanol) medium	potassium phosphate p	100mM	
	YNB	1.34%	
	Biotin	4*10-5%	
	Methanol	1%	
MD (Minimal Dextrose) medium	YNB	1.34%	
	Biotin	4*10-5%	
	Dextrose	2%	

References

ACHSTETTER, T. (1989). Regulation of alpha-factor production in Saccharomyces cerevisiae: a-factor pheromone-induced expression of the MF alpha 1 and STE13 genes. *Mol Cell Biol* **9,** 4507-14.

ADAMS, E. M., BROWN, M. C., NUNGE, M., KRYCH, M. & ATKINSON, J. P. (1991). Contribution of the repeating domains of membrane cofactor protein (CD46) of the complement system to ligand binding and cofactor activity. *J Immunol* **147,** 3005-11.

ALTMANN, F., STAUDACHER, E., WILSON, I. B. & MARZ, L. (1999). Insect cells as hosts for the expression of recombinant glycoproteins. *Glycoconj J* **16,** 109-23.

ANGELICHIO, M. L., BECK, J. A., JOHANSEN, H. & IVEY-HOYLE, M. (1991). Comparison of several promoters and polyadenylation signals for use in heterologous gene expression in cultured Drosophila cells. *Nucleic Acids Res* **19,** 5037-43.

ATABANI, S. F., BYRNES, A. A., JAYE, A., KIDD, I. M., MAGNUSEN, A. F., WHITTLE, H. & KARP, C. L. (2001). Natural measles causes prolonged suppression of interleukin-12 production. *J Infect Dis* **184,** 1-9.

BACZKO, K., BILLETER, M. & TER MEULEN, V. (1983). Purification and molecular weight determination of measles virus genomic RNA. *J Gen Virol* **64,** 1409-13.

BAKER, K. A., DUTCH, R. E., LAMB, R. A. & JARDETZKY, T. S. (1999). Structural basis for paramyxovirus-mediated membrane fusion. *Mol Cell* **3,** 309-19.

BALLOU, L., GOPAL, P., KRUMMEL, B., TAMMI, M. & BALLOU, C. E. (1986). A mutation that prevents glucosylation of the lipid-linked oligosaccharide precursor leads to underglycosylation of secreted yeast invertase. *Proc Natl Acad Sci U S A* **83,** 3081-5.

BANERJEE, A. K. (1987). Transcription and replication of rhabdoviruses. *Microbiol Rev* **51,** 66-87.

BANEYX, F. (1999). Recombinant protein expression in Escherichia coli. *Curr Opin Biotechnol* **10,** 411-21.

BARLOW, P. N., BARON, M., NORMAN, D. G., DAY, A. J., WILLIS, A. C., SIM, R. B. & CAMPBELL, I. D. (1991). Secondary structure of a complement control protein module by two- dimensional 1H NMR. *Biochemistry* **30,** 997-1004.

BARRINGTON, R., ZHANG, M., FISCHER, M. & CARROLL, M. C. (2001). The role of complement in inflammation and adaptive immunity. *Immunol Rev* **180,** 5-15.

BARTZ, R., BRINCKMANN, U., DUNSTER, L. M., RIMA, B., TER MEULEN, V. & SCHNEIDER-SCHAULIES, J. (1996). Mapping amino acids of the measles virus hemagglutinin responsible for receptor (CD46) downregulation. *Virology* **224,** 334-7.

BARTZ, R., FIRSCHING, R., RIMA, B., TER MEULEN, V. & SCHNEIDER-SCHAULIES, J. (1998). Differential receptor usage by measles virus strains. *J Gen Virol* **79,** 1015-25.

BECKFORD, A. P., KASCHULA, R. O. & STEPHEN, C. (1985). Factors associated with fatal cases of measles. A retrospective autopsy study. *S Afr Med J* **68,** 858-63.

BEGUM, N. A., MURAKAMI, Y., MIKATA, S., MATSUMOTO, M., HATANAKA, M., NAGASAWA, S., KINOSHITA, T. & SEYA, T. (2000). Molecular remodelling of human CD46 for xenotransplantation: designing a potent complement regulator without measles virus receptor activity. *Immunology* **100,** 131-9.

BELLINI, W. J., ENGLUND, G., RICHARDSON, C. D., ROZENBLATT, S. & LAZZARINI, R. A. (1986). Matrix genes of measles virus and canine distemper virus: cloning, nucleotide sequences, and deduced amino acid sequences. *J Virol* **58,** 408-16.

BELLINI, W. J., ENGLUND, G., ROZENBLATT, S., ARNHEITER, H. & RICHARDSON, C. D. (1985). Measles virus P gene codes for two proteins. *J Virol* **53,** 908-19.

BERGGARD, K., LINDAHL, G., DAHLBACK, B. & BLOM, A. M. (2001). Bordetella pertussis binds to human C4b-binding protein (C4BP) at a site similar to that used by the natural ligand C4b. *Eur J Immunol* **31,** 2771-80.

BERSCH, B., HERNANDEZ, J. F., MARION, D. & ARLAUD, G. J. (1998). Solution structure of the epidermal growth factor (EGF)-like module of human complement protease C1r, an atypical member of the EGF family. *Biochemistry* **37,** 1204-14.

BLIXENKRONE-MOLLER, M., BERNARD, A., BENCSIK, A., SIXT, N., DIAMOND, L. E., LOGAN, J. S. & WILD, T. F. (1998). Role of CD46 in measles virus infection in CD46 transgenic mice. *Virology* **249,** 238-48.

BLOM, A. M., BERGGARD, K., WEBB, J. H., LINDAHL, G., VILLOUTREIX, B. O. & DAHLBACK, B. (2000). Human C4b-binding protein has overlapping, but not identical, binding sites for C4b and streptococcal M proteins. *J Immunol* **164,** 5328-36.

BOLT, G. & PEDERSEN, I. R. (1998). The role of subtilisin-like proprotein convertases for cleavage of the measles virus fusion glycoprotein in different cell types. *Virology* **252,** 387-98.

BRODBECK, W. G., MOLD, C., ATKINSON, J. P. & MEDOF, M. E. (2000). Cooperation between decay-accelerating factor and membrane cofactor protein in protecting cells from autologous complement attack. *J Immunol* **165,** 3999-4006.

BUCKLAND, R., GERALD, C., BARKER, R. & WILD, T. F. (1987). Fusion glycoprotein of measles virus: nucleotide sequence of the gene and comparison with other paramyxoviruses. *J Gen Virol* **68,** 1695-703.

BUCKLAND, R. & WILD, T. F. (1997). Is CD46 the cellular receptor for measles virus? *Virus Res* **48,** 1-9.

BURDA, P. & AEBI, M. (1999). The dolichol pathway of N-linked glycosylation. *Biochim Biophys Acta* **1426,** 239-57.

CARROLL, M. C. (2000). The role of complement in B cell activation and tolerance. *Adv Immunol* **74,** 61-88.

CASALI, P. & OLDSTONE, M. B. (1983). Immune complexes in viral infection. *Curr Top Microbiol Immunol* **104,** 7-48.

CASASNOVAS, J. M., LARVIE, M. & STEHLE, T. (1999). Crystal structure of two CD46 domains reveals an extended measles virus- binding surface. *Embo J* **18,** 2911-22.

CATTANEO, R., REBMANN, G., BACZKO, K., TER MEULEN, V. & BILLETER, M. A. (1987). Altered ratios of measles virus transcripts in diseased human brains. *Virology* **160,** 523-6.

CEREGHINO, J. L. & CREGG, J. M. (2000). Heterologous protein expression in the methylotrophic yeast Pichia pastoris. *FEMS Microbiol Rev* **24,** 45-66.

CHEN, C. & OKAYAMA, H. (1987). High-efficiency transformation of mammalian cells by plasmid DNA. *Mol Cell Biol* **7,** 2745-52.

CHEN, X., VAN VALKENBURGH, C., FANG, H. & GREEN, N. (1999). Signal peptides having standard and nonstandard cleavage sites can be processed by Imp1p of the mitochondrial inner membrane protease. *J Biol Chem* **274,** 37750-4.

CHUI, D., SELLAKUMAR, G., GREEN, R., SUTTON-SMITH, M., MCQUISTAN, T., MAREK, K., MORRIS, H., DELL, A. & MARTH, J. (2001). Genetic remodeling of protein glycosylation in vivo induces autoimmune disease. *Proc Natl Acad Sci U S A* **98,** 1142-7.

CLARE, J. J., ROMANOS, M. A., RAYMENT, F. B., ROWEDDER, J. E., SMITH, M. A., PAYNE, M. M., SREEKRISHNA, K. & HENWOOD, C. A. (1991). Production of mouse epidermal growth factor in yeast: high-level secretion using Pichia pastoris strains containing multiple gene copies. *Gene* **105,** 205-12.

CLARK, J. M. (1988). Novel non-templated nucleotide addition reactions catalyzed by procaryotic and eucaryotic DNA polymerases. *Nucleic Acids Res* **16,** 9677-86.

COCKS, B. G., CHANG, C. C., CARBALLIDO, J. M., YSSEL, H., DE VRIES, J. E. & AVERSA, G. (1995). A novel receptor involved in T-cell activation. *Nature* **376,** 260-3.

COOPER, A., NUTLEY, M. A. & WADOOD, A. (2000). Differential Scanning Microcalorimetry. In *Protein-Ligand Interactions: hydrodynamics and calorimetry* (ed. B. Z. Chowdhry), pp. 287-318. Oxford University Press, Oxford.

CREGG, J. M., BARRINGER, K. J., HESSLER, A. Y. & MADDEN, K. R. (1985). Pichia pastoris as a host system for transformations. *Mol Cell Biol* **5,** 3376-85.

CREGG, J. M., MADDEN, K. R., BARRINGER, K. J., THILL, G. P. & STILLMAN, C. A. (1989). Functional characterization of the two alcohol oxidase genes from the yeast Pichia pastoris. *Mol Cell Biol* **9,** 1316-23.

CURRAN, M. D. & RIMA, B. K. (1988). Nucleotide sequence of the gene encoding the matrix protein of a recent measles virus isolate. *J Gen Virol* **69,** 2407-11.

DAVIS, S. J., IKEMIZU, S., WILD, M. K. & VAN DER MERWE, P. A. (1998). CD2 and the nature of protein interactions mediating cell-cell recognition. *Immunol Rev* **163,** 217-36.

DAVITZ, M. A., GURNETT, A. M., LOW, M. G., TURNER, M. J. & NUSSENZWEIG, V. (1987). Decay-accelerating factor (DAF) shares a common carbohydrate determinant with the variant surface glycoprotein (VSG) of the African Trypanosoma brucei. *J Immunol* **138,** 520-3.

DE BEER, T., VAN ZUYLEN, C. W., LEEFLANG, B. R., HARD, K., BOELENS, R., KAPTEIN, R., KAMERLING, J. P. & VLIEGENTHART, J. F. (1996). NMR studies of the free alpha subunit of human chorionic gonadotropin. Structural influences of N-glycosylation and the beta subunit on the conformation of the alpha subunit. *Eur J Biochem* **241,** 229-42.

DENTON, H., SMITH, M., HUSI, H., UHRIN, D., BARLOW, P. N., BATT, C. A. & SAWYER, L. (1998). Isotopically labeled bovine beta-lactoglobulin for NMR studies expressed in Pichia pastoris. *Protein Expr Purif* **14,** 97-103.

DEVINE, P. L., WARREN, J.A. (1990). Glycoprotein detection on Immobilon PVDF transfer membrane using the Periodic Acid/Schiff Reagent. *Biotechniques* **8,** 492.

DIAMOND, L. E., QUINN, C. M., MARTIN, M. J., LAWSON, J., PLATT, J. L. & LOGAN, J. S. (2001). A human CD46 transgenic pig model system for the study of discordant xenotransplantation. *Transplantation* **71,** 132-42.

DODD, I., MOSSAKOWSKA, D. E., CAMILLERI, P., HARAN, M., HENSLEY, P., LAWLOR, E. J., MCBAY, D. L., PINDAR, W. & SMITH, R. A. (1995). Overexpression in Escherichia coli, folding, purification, and characterization of the first three short consensus repeat modules of human complement receptor type 1. *Protein Expr Purif* **6,** 727-36.

DORIG, R. E., MARCIL, A., CHOPRA, A. & RICHARDSON, C. D. (1993). The human CD46 molecule is a receptor for measles virus (Edmonston strain). *Cell* **75,** 295-305.

DORIG, R. E., MARCIL, A. & RICHARDSON, C. D. (1994). CD46, a primate-specific receptor for measles virus. *Trends Microbiol* **2,** 312-8.

DOWLING, P. C., BLUMBERG, B. M., MENONNA, J., ADAMUS, J. E., COOK, P., CROWLEY, J. C., KOLAKOFSKY, D. & COOK, S. D. (1986). Transcriptional map of the measles virus genome. *J Gen Virol* **67,** 1987-92.

DUBENDORFF, J. W. & STUDIER, F. W. (1991). Controlling basal expression in an inducible T7 expression system by blocking the target T7 promoter with lac repressor. *J Mol Biol* **219,** 45-59.

DUNLAP, R. C., MILSTIEN, J. B. & LUNDQUIST, M. L. (1983). Identification and characterization of measles virus 50S RNA. *Intervirology* **19,** 169-75.

EL KASMI, K. C. & MULLER, C. P. (2001). New strategies for closing the gap of measles susceptibility in infants: towards vaccines compatible with current vaccination schedules. *Vaccine* **19,** 2238-44.

ERLENHOEFER, C., WURZER, W. J., LOFFLER, S., SCHNEIDER-SCHAULIES, S., TER MEULEN, V. & SCHNEIDER-SCHAULIES, J. (2001). CD150 (SLAM) is a receptor for measles virus but is not involved in viral contact-mediated proliferation inhibition. *J Virol* **75,** 4499-505.

FELGNER, P. L. & RINGOLD, G. M. (1989). Cationic liposome-mediated transfection. *Nature* **337,** 387-8.

GARCIA-ORTEGA, L., LACADENA, J., LACADENA, V., MASIP, M., DE ANTONIO, C., MARTINEZ-RUIZ, A. & MARTINEZ DEL POZO, A. (2000). The solubility of the ribotoxin alpha-sarcin, produced as a recombinant protein in Escherichia coli, is increased in the presence of thioredoxin. *Lett Appl Microbiol* **30,** 298-302.

GEORGIOU, G. & VALAX, P. (1996). Expression of correctly folded proteins in Escherichia coli. *Curr Opin Biotechnol* **7,** 190-7.

GERALD, C., BUCKLAND, R., BARKER, R., FREEMAN, G. & WILD, T. F. (1986). Measles virus haemagglutinin gene: cloning, complete nucleotide sequence analysis and expression in COS cells. *J Gen Virol* **67,** 2695-703.

GERLIER, D., LOVELAND, B., VARIOR-KRISHNAN, G., THORLEY, B., MCKENZIE, I. F. & RABOURDIN-COMBE, C. (1994). Measles virus receptor properties are shared by several CD46 isoforms differing in extracellular regions and cytoplasmic tails. *J Gen Virol* **75,** 2163-71.

GIANNAKIS, E., JOKIRANTA, T. S., ORMSBY, R. J., DUTHY, T. G., MALE, D. A., CHRISTIANSEN, D., FISCHETTI, V. A., BAGLEY, C., LOVELAND, B. E. & GORDON, D. L. (2002). Identification of the streptococcal M protein binding site on membrane cofactor protein (CD46). *J Immunol* **168,** 4585-92.

GODA, S., TAKANO, K., YAMAGATA, Y., KATAKURA, Y. & YUTANI, K. (2000). Effect of extra N-terminal residues on the stability and folding of human lysozyme expressed in Pichia pastoris. *Protein Eng* **13,** 299-307.

GOMBART, A. F., HIRANO, A. & WONG, T. C. (1992). Expression and properties of the V protein in acute measles virus and subacute sclerosing panencephalitis virus strains. *Virus Res* **25,** 63-78.

GRIFFIN, D. E., WARD, B. J. & ESOLEN, L. M. (1994). Pathogenesis of measles virus infection: an hypothesis for altered immune responses. *J Infect Dis* **170 Suppl 1,** S24-31.

GUTHRIDGE, J. M., RAKSTANG, J. K., YOUNG, K. A., HINSHELWOOD, J., ASLAM, M., ROBERTSON, A., GIPSON, M. G., SARRIAS, M. R., MOORE, W. T., MEAGHER, M., KARP, D., LAMBRIS, J. D., PERKINS, S. J. & HOLERS, V. M. (2001). Structural studies in solution of the recombinant N-terminal pair of short consensus/complement repeat domains of complement receptor type 2 (CR2/CD21) and interactions with its ligand C3dg. *Biochemistry* **40,** 5931-41.

HAMMOND, C., BRAAKMAN, I. & HELENIUS, A. (1994). Role of N-linked oligosaccharide recognition, glucose trimming, and calnexin in glycoprotein folding and quality control. *Proc Natl Acad Sci U S A* **91,** 913-7.

HARA, T., KURIYAMA, S., KIYOHARA, H., NAGASE, Y., MATSUMOTO, M. & SEYA, T. (1992). Soluble forms of membrane cofactor protein (CD46, MCP) are present in plasma, tears, and seminal fluid in normal subjects. *Clin Exp Immunol* **89,** 490-4.

HAURI, H. P., KAPPELER, F., ANDERSSON, H. & APPENZELLER, C. (2000). ERGIC-53 and traffic in the secretory pathway. *J Cell Sci* **113 (Pt 4),** 587-96.

HEGEDUS, D. D., PFEIFER, T. A., THEILMANN, D. A., KENNARD, M. L., GABATHULER, R., JEFFERIES, W. A. & GRIGLIATTI, T. A. (1999). Differences in the expression and localization of human melanotransferrin in lepidopteran and dipteran insect cell lines. *Protein Expr Purif* **15,** 296-307.

HELLWAGE, J., MERI, T., HEIKKILA, T., ALITALO, A., PANELIUS, J., LAHDENNE, P., SEPPALA, I. J. & MERI, S. (2001). The complement regulator factor H binds to the surface protein OspE of Borrelia burgdorferi. *J Biol Chem* **276,** 8427-35.

HENDERSON, C. E., BROMEK, K., MULLIN, N. P., SMITH, B. O., UHRIN, D. & BARLOW, P. N. (2001). Solution structure and dynamics of the central CCP module pair of a poxvirus complement control protein. *J Mol Biol* **307,** 323-39.

HIETANEN, J. & TARKKANEN, A. (1989). Glycoconjugates in exfoliation syndrome. A lectin histochemical study of the ciliary body and lens. *Acta Ophthalmol (Copenh)* **67,** 288-94.

HIGGINS, D. R., AND CREGG, J. M. (1998). *Pichia Protocols.* Humana Press, Totowa, NJ., USA.

HOFSTEENGE, J., MULLER, D. R., DE BEER, T., LOFFLER, A., RICHTER, W. J. & VLIEGENTHART, J. F. (1994). New type of linkage between a carbohydrate and a protein: C-glycosylation of a specific tryptophan residue in human RNase Us. *Biochemistry* **33,** 13524-30.

HOLMSKOV, U., MALHOTRA, R., SIM, R. B. & JENSENIUS, J. C. (1994). Collectins: collagenous C-type lectins of the innate immune defense system. *Immunol Today* **15,** 67-74.

HORIKAMI, S. M. & MOYER, S. A. (1995). Structure, transcription, and replication of measles virus. *Curr Top Microbiol Immunol* **191,** 35-50.

HORIKAMI, S. M., SMALLWOOD, S., BANKAMP, B. & MOYER, S. A. (1994). An amino-proximal domain of the L protein binds to the P protein in the measles virus RNA polymerase complex. *Virology* **205,** 540-5.

HORVAT, B., RIVAILLER, P., VARIOR-KRISHNAN, G., CARDOSO, A., GERLIER, D. & RAROURDIN-COMBE, C. (1996). Transgenic mice expressing human measles virus (MV) receptor CD46 provide cells exhibiting different permissivities to MV infections. *J Virol* **70,** 6673-81.

HSU, E. C., IORIO, C., SARANGI, F., KHINE, A. A. & RICHARDSON, C. D. (2001). CDw150(SLAM) is a receptor for a lymphotropic strain of measles virus and may account for the immunosuppressive properties of this virus. *Virology* **279,** 9-21.

HSU, E. C., SABATINOS, S., HOEDEMAEKER, F. J., ROSE, D. R. & RICHARDSON, C. D. (1999). Use of site-specific mutagenesis and monoclonal antibodies to map regions of CD46 that interact with measles virus H protein. *Virology* **258,** 314-26.

INVITROGEN. (2002a). *Pichia Expression Kit - A Manual of Methods for Expression of Recombinant Proteins in Pichia pastoris.* Invitrogen Corporation, Carlsbad, CA, USA.

INVITROGEN. (2002b). *pPICZα A, B, and C - Pichia expression vectors for selection on Zeocin™ and purification of secreted, recombinant proteins.* Invitrogen Corporation, Carlsbad, CA, USA.

ISAACS, S. N., KOTWAL, G. J. & MOSS, B. (1992). Vaccinia virus complement-control protein prevents antibody-dependent complement-enhanced neutralization of infectivity and contributes to virulence. *Proc Natl Acad Sci U S A* **89,** 628-32.

IVEY-HOYLE, M., CULP, J. S., CHAIKIN, M. A., HELLMIG, B. D., MATTHEWS, T. J., SWEET, R. W. & ROSENBERG, M. (1991). Envelope glycoproteins from biologically diverse isolates of immunodeficiency viruses have widely different affinities for CD4. *Proc Natl Acad Sci U S A* **88,** 512-6.

IWATA, K., SEYA, T., YANAGI, Y., PESANDO, J. M., JOHNSON, P. M., OKABE, M., UEDA, S., ARIGA, H. & NAGASAWA, S. (1995). Diversity of sites for measles virus binding and for inactivation of complement C3b and C4b on membrane cofactor protein CD46. *J Biol Chem* **270,** 15148-52.

JARVIS, G. A. (1995). Recognition and control of neisserial infection by antibody and complement. *Trends Microbiol* **3,** 198-201.

JOHANSON, K., APPELBAUM, E., DOYLE, M., HENSLEY, P., ZHAO, B., ABDEL-MEGUID, S. S., YOUNG, P., COOK, R., CARR, S., MATICO, R. & ET AL. (1995). Binding interactions of human interleukin 5 with its receptor alpha subunit. Large scale production, structural, and functional studies of Drosophila-expressed recombinant proteins. *J Biol Chem* **270,** 9459-71.

JOHNSTONE, R. W., LOVELAND, B. E. & MCKENZIE, I. F. (1993a). Identification and quantification of complement regulator CD46 on normal human tissues. *Immunology* **79,** 341-7.

JOHNSTONE, R. W., RUSSELL, S. M., LOVELAND, B. E. & MCKENZIE, I. F. (1993b). Polymorphic expression of CD46 protein isoforms due to tissue-specific RNA splicing. *Mol Immunol* **30,** 1231-41.

JULIUS, D., BLAIR, L., BRAKE, A., SPRAGUE, G. & THORNER, J. (1983). Yeast alpha factor is processed from a larger precursor polypeptide: the essential role of a membrane-bound dipeptidyl aminopeptidase. *Cell* **32,** 839-52.

KADNER, A., CHEN, R. H., FARIVAR, R. S., SANTERRE, D. & ADAMS, D. H. (2001). Transplantation of MCP Cardiac Xenografts into Baboon Recipients. *Transplant Proc* **33,** 770-772.

KALLSTROM, H., BLACKMER GILL, D., ALBIGER, B., LISZEWSKI, M. K., ATKINSON, J. P. & JONSSON, A. B. (2001). Attachment of Neisseria gonorrhoeae to the cellular pilus receptor CD46: identification of domains important for bacterial adherence. *Cell Microbiol* **3,** 133-43.

KALLSTROM, H., ISLAM, M. S., BERGGREN, P. O. & JONSSON, A. B. (1998). Cell signaling by the type IV pili of pathogenic Neisseria. *J Biol Chem* **273,** 21777-82.

KALLSTROM, H., LISZEWSKI, M. K., ATKINSON, J. P. & JONSSON, A. B. (1997). Membrane cofactor protein (MCP or CD46) is a cellular pilus receptor for pathogenic Neisseria. *Mol Microbiol* **25,** 639-47.

KAPLAN, L. J., DAUM, R. S., SMARON, M. & MCCARTHY, C. A. (1992). Severe measles in immunocompromised patients. *Jama* **267,** 1237-41.

KARP, C. L., WYSOCKA, M., WAHL, L. M., AHEARN, J. M., CUOMO, P. J., SHERRY, B., TRINCHIERI, G. & GRIFFIN, D. E. (1996). Mechanism of suppression of cell-mediated immunity by measles virus. *Science* **273,** 228-31.

KATZ, M. (1995). Clinical Spectrum of Measles. In *Measles Virus*, vol. 191. *Current Topics in Microbiology and Immunology* (ed. M. Billeter), pp. 1-12. Springer-Verlag, Berlin Heilelberg.

KIRKITADZE, M. D. & BARLOW, P. N. (2001). Structure and flexibility of the multiple domain proteins that regulate complement activation. *Immunol Rev* **180,** 146-61.

KIRKPATRICK, R. B., GANGULY, S., ANGELICHIO, M., GRIEGO, S., SHATZMAN, A., SILVERMAN, C. & ROSENBERG, M. (1995). Heavy chain dimers as well as complete antibodies are efficiently formed and secreted from Drosophila via a BiP-mediated pathway. *J Biol Chem* **270,** 19800-5.

KITAMURA, M., MATSUMIYA, K., YAMANAKA, M., TAKAHARA, S., HARA, T., MATSUMOTO, M., NAMIKI, M., OKUYAMA, A. & SEYA, T. (1997). Possible association of infertility with sperm-specific abnormality of CD46. *J Reprod Immunol* **33,** 83-8.

KOBUNE, F., SAKATA, H. & SUGIURA, A. (1990). Marmoset lymphoblastoid cells as a sensitive host for isolation of measles virus. *J Virol* **64,** 700-5.

KROSHUS, T. J., SALERNO, C. T., YEH, C. G., HIGGINS, P. J., BOLMAN, R. M., 3RD & DALMASSO, A. P. (2000). A recombinant soluble chimeric complement inhibitor composed of human CD46 and CD55 reduces acute cardiac tissue injury in models of pig-to- human heart transplantation. *Transplantation* **69,** 2282-9.

KRYCH-GOLDBERG, M., MOULDS, J. M. & ATKINSON, J. P. (2002). Human complement receptor type 1 (CR1) binds to a major malarial adhesin. *Trends Mol Med* **8,** 531-7.

LACHMANN, P. J. (1991). The control of homologous lysis. *Immunol Today* **12,** 312-5.

LAEMMLI, U. K. (1970). Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* **227,** 680-5.

LAROCHE, Y., STORME, V., DE MEUTTER, J., MESSENS, J. & LAUWEREYS, M. (1994). High-level secretion and very efficient isotopic labeling of tick anticoagulant peptide (TAP) expressed in the methylotrophic yeast, Pichia pastoris. *Biotechnology (N Y)* **12,** 1119-24.

LAW, S. K. (1983). Non-enzymic activation of the covalent binding reaction of the complement protein C3. *Biochem J* **211,** 381-9.

LAW, S. K. A. & REID, K. B. M. (1995). *Complement*, 2nd edition. IRL Press.

LECOUTURIER, V., FAYOLLE, J., CABALLERO, M., CARABANA, J., CELMA, M. L., FERNANDEZ-MUNOZ, R., WILD, T. F. & BUCKLAND, R. (1996). Identification of two amino acids in the hemagglutinin glycoprotein of measles virus (MV) that govern hemadsorption, HeLa cell fusion, and CD46 downregulation: phenotypic markers that differentiate vaccine and wild-type MV strains. *J Virol* **70,** 4200-4.

LECOUTURIER, V., RIZZITELLI, A., FAYOLLE, J., DAVIET, L., WILD, F. T. & BUCKLAND, R. (1999). Interaction of measles virus (Halle strain) with CD46: evidence that a common binding site on CD46 facilitates both CD46 downregulation and MV infection. *Biochem Biophys Res Commun* **264,** 268-75.

LEHRMAN, M. A. (2001). Oligosaccharide-based information in endoplasmic reticulum quality control and other biological systems. *J Biol Chem* **276,** 8623-6.

LI, L. & QI, Y. (2002). A novel amino acid position in hemagglutinin glycoprotein of measles virus is responsible for hemadsorption and CD46 binding. *Arch Virol* **147,** 775-86.

LINDAHL, G., SJOBRING, U. & JOHNSSON, E. (2000). Human complement regulators: a major target for pathogenic microorganisms. *Curr Opin Immunol* **12,** 44-51.

LISZEWSKI, M. K., FARRIES, T. C., LUBLIN, D. M., ROONEY, I. A. & ATKINSON, J. P. (1996). Control of the complement system. *Adv Immunol* **61,** 201-83.

LISZEWSKI, M. K., LEUNG, M., CUI, W., SUBRAMANIAN, V. B., PARKINSON, J., BARLOW, P. N., MANCHESTER, M. & ATKINSON, J. P. (2000). Dissecting sites important for complement regulatory activity in membrane cofactor protein (MCP; CD46). *J Biol Chem* **275,** 37692-701.

LISZEWSKI, M. K., LEUNG, M. K. & ATKINSON, J. P. (1998). Membrane cofactor protein: importance of N- and O-glycosylation for complement regulatory function. *J Immunol* **161,** 3711-8.

LISZEWSKI, M. K., POST, T. W. & ATKINSON, J. P. (1991). Membrane cofactor protein (MCP or CD46): newest member of the regulators of complement activation gene cluster. *Annu Rev Immunol* **9,** 431-55.

LOVELAND, B. E., JOHNSTONE, R. W., RUSSELL, S. M., THORLEY, B. R. & MCKENZIE, I. F. (1993). CD46 (MCP) confers protection from lysis by xenogeneic antibodies. *Transplant Proc* **25,** 396-7.

LUND, G. A., TYRRELL, D. L., BRADLEY, R. D. & SCRABA, D. G. (1984). The molecular length of measles virus RNA and the structural organization of measles nucleocapsids. *J Gen Virol* **65,** 1535-42.

LUNDBLAD, R. L. (1999). Glycosylation in Pichia pastoris. *Biotechnol Appl Biochem* **30,** 191-2.

MAISNER, A., ALVAREZ, J., LISZEWSKI, M. K., ATKINSON, D. J., ATKINSON, J. P. & HERRLER, G. (1996). The N-glycan of the SCR 2 region is essential for membrane cofactor protein (CD46) to function as a measles virus receptor. *J Virol* **70,** 4973-7.

MAISNER, A. & HERRLER, G. (1995). Membrane cofactor protein with different types of N-glycans can serve as measles virus receptor. *Virology* **210,** 479-81.

MAISNER, A., SCHNEIDER-SCHAULIES, J., LISZEWSKI, M. K., ATKINSON, J. P. & HERRLER, G. (1994). Binding of measles virus to membrane cofactor protein (CD46): importance of disulfide bonds and N-glycans for the receptor function. *J Virol* **68,** 6299-304.

MANCHESTER, M., ETO, D. S., VALSAMAKIS, A., LITON, P. B., FERNANDEZ-MUNOZ, R., ROTA, P. A., BELLINI, W. J., FORTHAL, D. N. & OLDSTONE, M. B. (2000). Clinical isolates of measles virus use CD46 as a cellular receptor. *J Virol* **74,** 3967-74.

MANCHESTER, M., GAIRIN, J. E., PATTERSON, J. B., ALVAREZ, J., LISZEWSKI, M. K., ETO, D. S., ATKINSON, J. P. & OLDSTONE, M. B. (1997). Measles virus recognizes its receptor, CD46, via two distinct binding domains within SCR1-2. *Virology* **233,** 174-84.

MARKOWITZ, L. E., CHANDLER, F. W., ROLDAN, E. O., SALDANA, M. J., ROACH, K. C., HUTCHINS, S. S., PREBLUD, S. R., MITCHELL, C. D. & SCOTT, G. B. (1988). Fatal measles pneumonia without rash in a child with AIDS. *J Infect Dis* **158,** 480-3.

MARTINET, W., SAELENS, X., DEROO, T., NEIRYNCK, S., CONTRERAS, R., MIN JOU, W. & FIERS, W. (1997). Protection of mice against a lethal influenza challenge by immunization with yeast-derived recombinant influenza neuraminidase. *Eur J Biochem* **247,** 332-8.

MAVADDAT, N., MASON, D. W., ATKINSON, P. D., EVANS, E. J., GILBERT, R. J., STUART, D. I., FENNELLY, J. A., BARCLAY, A. N., DAVIS, S. J. & BROWN, M. H. (2000). Signaling lymphocytic activation molecule (CDw150) is homophilic but self-associates with very low affinity. *J Biol Chem* **275,** 28100-9.

MCALISTER, M. S., DAVIS, B., PFUHL, M. & DRISCOLL, P. C. (1998). NMR analysis of the N-terminal SRCR domain of human CD5: engineering of a glycoprotein for superior characteristics in NMR experiments. *Protein Eng* **11,** 847-53.

MERI, S. & PANGBURN, M. K. (1990). Discrimination between activators and nonactivators of the alternative pathway of complement: regulation via a sialic acid/polyanion binding site on factor H. *Proc Natl Acad Sci U S A* **87,** 3982-6.

MERI, S. & PANGBURN, M. K. (1994). Regulation of alternative pathway complement activation by glycosaminoglycans: specificity of the polyanion binding site on factor H. *Biochem Biophys Res Commun* **198,** 52-9.

MERI, T., JOKIRANTA, T. S., HELLWAGE, J., BIALONSKI, A., ZIPFEL, P. F. & MERI, S. (2002). Onchocerca volvulus microfilariae avoid complement attack by direct binding of factor H. *J Infect Dis* **185,** 1786-93.

MIELE, R. G., NILSEN, S. L., BRITO, T., BRETTHAUER, R. K. & CASTELLINO, F. J. (1997). Glycosylation properties of the Pichia pastoris-expressed recombinant kringle 2 domain of tissue-type plasminogen activator. *Biotechnol Appl Biochem* **25,** 151-7.

MINAGAWA, H., TANAKA, K., ONO, N., TATSUO, H. & YANAGI, Y. (2001). Induction of the measles virus receptor SLAM (CD150) on monocytes. *J Gen Virol* **82,** 2913-7.

MOLINA, H., BRENNER, C., JACOBI, S., GORKA, J., CAREL, J. C., KINOSHITA, T. & HOLERS, V. M. (1991). Analysis of Epstein-Barr virus-binding sites on complement receptor 2 (CR2/CD21) using human-mouse chimeras and peptides. At least two distinct sites are necessary for ligand-receptor interaction. *J Biol Chem* **266,** 12173-9.

MORRA, M., LU, J., POY, F., MARTIN, M., SAYOS, J., CALPE, S., GULLO, C., HOWIE, D., RIETDIJK, S., THOMPSON, A., COYLE, A. J., DENNY, C., YAFFE, M. B., ENGEL, P., ECK, M. J. & TERHORST, C. (2001). Structural basis for the interaction of the free SH2 domain EAT-2 with SLAM receptors in hematopoietic cells. *Embo J* **20,** 5840-52.

MORRIS, A. & ALDULAIMI, D. (2002). New evidence for a viral pathogenic mechanism for new variant inflammatory bowel disease and development disorder? *Mol Pathol* **55,** 83.

MUIR, T. W., SONDHI, D. & COLE, P. A. (1998). Expressed protein ligation: a general method for protein engineering. *Proc Natl Acad Sci U S A* **95,** 6705-10.

MULDER, L. C., MORA, M., CICCOPIEDI, E., MELLI, C., NUTI, S., MARINUCCI, G., BRUZZONE, P., LAZZERI, M., LORENZINI, R., ALFANI, D. & ET AL. (1995). Mice transgenic for human CD46 and CD55 are protected from human complement attack. *Transplant Proc* **27,** 333-5.

MULDER, L. C., MORA, M., LAZZARI, M., BOSCHI, E., CICCOPIEDI, E., MARINUCCI, G., MELLI, C., BRUZZONE, P., ALFANI, D., CORTESINI, R. & ROSSINI, M. (1996). Characterisation of Complement-Mediated Liver Damage and Protection in Control and Transgenic Mice for Human Complement Blockers MCP and DAF. *Transplant Proc* **28,** 127-129.

MUNTONI, F., BROCKINGTON, M., BLAKE, D. J., TORELLI, S. & BROWN, S. C. (2002). Defective glycosylation in muscular dystrophy. *Lancet* **360,** 1419-21.

MURTHY, K. H., SMITH, S. A., GANESH, V. K., JUDGE, K. W., MULLIN, N., BARLOW, P. N., OGATA, C. M. & KOTWAL, G. J. (2001). Crystal structure of a complement control protein that regulates both pathways of complement activation and binds heparan sulfate proteoglycans. *Cell* **104,** 301-11.

NAIM, H. Y., EHLER, E. & BILLETER, M. A. (2000). Measles virus matrix protein specifies apical virus release and glycoprotein sorting in epithelial cells. *Embo J* **19,** 3576-85.

NAKAMURA, H., ZARYCKI, J., SULLIVAN, J. L. & JUNG, J. U. (2001). Abnormal T cell receptor signal transduction of CD4 Th cells in X-linked lymphoproliferative syndrome. *J Immunol* **167,** 2657-65.

NANICHE, D., VARIOR-KRISHNAN, G., CERVONI, F., WILD, T. F., ROSSI, B., RABOURDIN-COMBE, C. & GERLIER, D. (1993). Human membrane cofactor protein (CD46) acts as a cellular receptor for measles virus. *J Virol* **67,** 6025-32.

NEWTON, C. R. & GRAHAM, A. (1997). *PCR*, 2nd edition. Bios Scientific Publishers Limited, Oxford.

NIEWIESK, S., SCHNEIDER-SCHAULIES, J., OHNIMUS, H., JASSOY, C., SCHNEIDER-SCHAULIES, S., DIAMOND, L., LOGAN, J. S. & TER MEULEN, V. (1997). CD46 expression does not overcome the intracellular block of measles virus replication in transgenic rats. *J Virol* **71,** 7969-73.

NILSSON, I. & VON HEIJNE, G. (2000). Glycosylation efficiency of Asn-Xaa-Thr sequons depends both on the distance from the C terminus and on the presence of a downstream transmembrane segment. *J Biol Chem* **275,** 17338-43.

NORMAN, D. G., BARLOW, P. N., BARON, M., DAY, A. J., SIM, R. B. & CAMPBELL, I. D. (1991). Three-dimensional structure of a complement control protein module in solution. *J Mol Biol* **219,** 717-25.

NOVAGEN. (2002). *pET System Manual 10th edition.*

NUSSBAUM, O., BRODER, C. C., MOSS, B., STERN, L. B., ROZENBLATT, S. & BERGER, E. A. (1995). Functional and structural interactions between measles virus hemagglutinin and CD46. *J Virol* **69,** 3341-9.

OHGIMOTO, S., OHGIMOTO, K., NIEWIESK, S., KLAGGE, I. M., PFEUFFER, J., JOHNSTON, I. C., SCHNEIDER-SCHAULIES, J., WEIDMANN, A., TER MEULEN, V. & SCHNEIDER-SCHAULIES, S. (2001). The haemagglutinin protein is an important determinant of measles virus tropism for dendritic cells in vitro. *J Gen Virol* **82,** 1835-44.

OHTSUKA, N., YAMADA, Y. K. & TAGUCHI, F. (1996). Difference in virus-binding activity of two distinct receptor proteins for mouse hepatitis virus. *J Gen Virol* **77 (Pt 8),** 1683-92.

OKADA, N., LISZEWSKI, M. K., ATKINSON, J. P. & CAPARON, M. (1995). Membrane cofactor protein (CD46) is a keratinocyte receptor for the M protein of the group A streptococcus. *Proc Natl Acad Sci U S A* **92,** 2489-93.

O'LEARY, J. M. (2000). Chemical Synthesis, Recombinant Expression, and Structural Characterisation of Complement Protein Modules. PhD thesis, University of Edinburgh.

OMER, M. I. (1999). Measles: a disease that has to be eradicated. *Ann Trop Paediatr* **19,** 125-34.

ONO, N., TATSUO, H., HIDAKA, Y., AOKI, T., MINAGAWA, H. & YANAGI, Y. (2001a). Measles viruses on throat swabs from measles patients use signaling lymphocytic activation molecule (CDw150) but not CD46 as a cellular receptor. *J Virol* **75,** 4399-401.

ONO, N., TATSUO, H., TANAKA, K., MINAGAWA, H. & YANAGI, Y. (2001b). V domain of human SLAM (CDw150) is essential for its function as a measles virus receptor. *J Virol* **75,** 1594-600.

OSTERHAUS, A. D., DE VRIES, P. & VAN BINNENDIJK, R. S. (1994). Measles vaccines: novel generations and new strategies. *J Infect Dis* **170 Suppl 1,** S42-55.

OWENS, S. R. (2002). Injection of confidence: The recent controversy in the UK has led to falling MMR vaccination rates. *EMBO Rep* **3,** 406-9.

PAIFER, E., MARGOLLES, E., CREMATA, J., MONTESINO, R., HERRERA, L. & DELGADO, J. M. (1994). Efficient expression and secretion of recombinant alpha amylase in Pichia pastoris using two different signal sequences. *Yeast* **10,** 1415-9.

PANGBURN, M. K. (2000). Host recognition and target differentiation by factor H, a regulator of the alternative pathway of complement. *Immunopharmacology* **49,** 149-57.

PANGBURN, M. K. (2002). Cutting edge: localization of the host recognition functions of complement factor h at the carboxyl-terminal: implications for hemolytic uremic syndrome. *J Immunol* **169,** 4702-6.

PANUM, P. (1939). Observations made during the epidemic of measles on the Faroe Islands in the year 1846. *Med Classics* **3,** 839-886.

PARODI, A. J. (2000). Protein glucosylation and its role in protein folding. *Annu Rev Biochem* **69,** 69-93.

PATTERSON, J. B., THOMAS, D., LEWICKI, H., BILLETER, M. A. & OLDSTONE, M. B. (2000). V and C proteins of measles virus function as virulence factors in vivo. *Virology* **267,** 80-9.

PERKINS, S. J. & GOODSHIP, T. H. (2002). Molecular modelling of the C-terminal domains of factor H of human complement: a correlation between haemolytic uraemic syndrome and a predicted heparin binding site. *J Mol Biol* **316,** 217-24.

PETERSON, J. R., ORA, A., VAN, P. N. & HELENIUS, A. (1995). Transient, lectin-like association of calreticulin with folding intermediates of cellular and viral glycoproteins. *Mol Biol Cell* **6,** 1173-84.

PLUM, G. E. & BRESLAUER, K. J. (1995). Calorimetry of proteins and nucleic acids. *Curr Opin Struct Biol* **5,** 682-90.

PROMEGA. (2002). *pGEM®-T and pGEM®-T Easy Vector Systems - Technical Manual No. 042.* Promega Corporation, Madison, WI, USA.

PUNNONEN, J., COCKS, B. G., CARBALLIDO, J. M., BENNETT, B., PETERSON, D., AVERSA, G. & DE VRIES, J. E. (1997). Soluble and membrane-bound forms of signaling lymphocytic activation molecule (SLAM) induce proliferation and Ig synthesis by activated human B lymphocytes. *J Exp Med* **185,** 993-1004.

PURTILO, D. T., CASSEL, C. K., YANG, J. P. & HARPER, R. (1975). X-linked recessive progressive combined variable immunodeficiency (Duncan's disease). *Lancet* **1,** 935-40.

RAM, S., CULLINANE, M., BLOM, A. M., GULATI, S., MCQUILLEN, D. P., MONKS, B. G., O'CONNELL, C., BODEN, R., ELKINS, C., PANGBURN, M. K., DAHLBACK, B. & RICE, P. A. (2001). Binding of C4b-binding protein to porin: a molecular mechanism of serum resistance of Neisseria gonorrhoeae. *J Exp Med* **193,** 281-95.

RAM, S., SHARMA, A. K., SIMPSON, S. D., GULATI, S., MCQUILLEN, D. P., PANGBURN, M. K. & RICE, P. A. (1998). A novel sialic acid binding site on factor H mediates serum resistance of sialylated Neisseria gonorrhoeae. *J Exp Med* **187,** 743-52.

RAMAGE, P., HEMMIG, R., MATHIS, B., COWAN-JACOB, S. W., RONDEAU, J. M., KALLEN, J., BLOMMERS, M. J. J., ZURINI, M. & RUDISSER, S. (2002). Snags with tags: Some observations made with $(His)_6$-tagged proteins. *Life Science News* **11,** 1-4.

REDDY, P., CARAS, I. & KRIEGER, M. (1989). Effects of O-linked glycosylation on the cell surface expression and stability of decay-accelerating factor, a glycophospholipid-anchored membrane protein. *J Biol Chem* **264,** 17329-36.

REEKE, G. N., JR., BECKER, J. W., CUNNINGHAM, B. A., WANG, J. L., YAHARA, I. & EDELMAN, G. M. (1975). Structure and function of concanavalin A. *Adv Exp Med Biol* **55,** 13-33.

REID, K. B. (1983). Proteins involved in the activation and control of the two pathways of human complement. *Biochem Soc Trans* **11,** 1-12.

REITTER, J. N., MEANS, R. E. & DESROSIERS, R. C. (1998). A role for carbohydrates in immune evasion in AIDS. *Nat Med* **4,** 679-84.

ROBBINS, P. W., TRIMBLE, R. B., WIRTH, D. F., HERING, C., MALEY, F., MALEY, G. F., DAS, R., GIBSON, B. W., ROYAL, N. & BIEMANN, K. (1984). Primary structure of the Streptomyces enzyme endo-beta-N-acetylglucosaminidase H. *J Biol Chem* **259,** 7577-83.

RONNETT, G. V., KNUTSON, V. P., KOHANSKI, R. A., SIMPSON, T. L. & LANE, M. D. (1984). Role of glycosylation in the processing of newly translated insulin proreceptor in 3T3-L1 adipocytes. *J Biol Chem* **259,** 4566-75.

RUDD, P. M., ENDO, T., COLOMINAS, C., GROTH, D., WHEELER, S. F., HARVEY, D. J., WORMALD, M. R., SERBAN, H., PRUSINER, S. B., KOBATA, A. & DWEK, R. A. (1999). Glycosylation differences between the normal and pathogenic prion protein isoforms. *Proc Natl Acad Sci U S A* **96,** 13044-9.

RUDD, P. M., MERRY, A. H., WORMALD, M. R. & DWEK, R. A. (2002). Glycosylation and prion protein. *Curr Opin Struct Biol* **12,** 578-86.

RUDD, P. M., WORMALD, M. R. & DWEK, R. A. (1998). Glycosylation and the immune system. *J Protein Chem* **17,** 519.

SACHDEV, D. & CHIRGWIN, J. M. (1998). Solubility of proteins isolated from inclusion bodies is enhanced by fusion to maltose-binding protein or thioredoxin. *Protein Expr Purif* **12,** 122-32.

SAITO, H., SATO, H., ABE, M., HARATA, S., AMANO, K., SUTO, T. & MORITA, M. (1992). Isolation and characterization of the measles virus strains with low hemagglutination activity. *Intervirology* **33,** 57-60.

SAMBROOK, J., FRITSCH, E. F. & MANIATIS, T. (1989). *Molecular Cloning: a laboratory manual*, 2nd edition. Cold Spring Harbor Laboratory Press, New York.

SAMBROOK, J. & RUSSELL, D. (2000). *Molecular Cloning A Laboratory Manual*, 3rd edition. Cold Spring Harbor Laboratory Press, New York.

SAMUEL, O. & SHAI, Y. (2001). Participation of two fusion peptides in measles virus-induced membrane fusion: emerging similarity with other paramyxoviruses. *Biochemistry* **40,** 1340-9.

SANGER, F., NICKLEN, S. & COULSON, A. R. (1977). DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci U S A* **74,** 5463-7.

SANTORO, F., KENNEDY, P. E., LOCATELLI, G., MALNATI, M. S., BERGER, E. A. & LUSSO, P. (1999). CD46 is a cellular receptor for human herpesvirus 6. *Cell* **99,** 817-27.

SAYOS, J., WU, C., MORRA, M., WANG, N., ZHANG, X., ALLEN, D., VAN SCHAIK, S., NOTARANGELO, L., GEHA, R., RONCAROLO, M. G., OETTGEN, H., DE VRIES, J. E., AVERSA, G. & TERHORST, C. (1998). The X-linked lymphoproliferative-disease gene product SAP regulates signals induced through the co-receptor SLAM. *Nature* **395,** 462-9.

SCHLENKER, T. L., BAIN, C., BAUGHMAN, A. L. & HADLER, S. C. (1992). Measles herd immunity. The association of attack rates with immunization rates in preschool children. *Jama* **267,** 823-6.

SCHNEIDER-SCHAULIES, J., DUNSTER, L. M., KOBUNE, F., RIMA, B. & TER MEULEN, V. (1995a). Differential downregulation of CD46 by measles virus strains. *J Virol* **69,** 7257-9.

SCHNEIDER-SCHAULIES, J., DUNSTER, L. M., SCHWARTZ-ALBIEZ, R., KROHNE, G. & TER MEULEN, V. (1995b). Physical association of moesin and CD46 as a receptor complex for measles virus. *J Virol* **69,** 2248-56.

SCHNEIDER-SCHAULIES, S. & TER MEULEN, V. (2002). Measles virus and immunomodulation: molecular bases and perspectives. In *Expert Reviews in Molecular Medicine*, pp. http://www.expertreviews.org/02004696h.htm.

SEYA, T., HARA, T., MATSUMOTO, M., KIYOHARA, H., NAKANISHI, I., KINOUCHI, T., OKABE, M., SHIMIZU, A. & AKEDO, H. (1993). Membrane cofactor protein (MCP, CD46) in seminal plasma and on spermatozoa in normal and "sterile" subjects. *Eur J Immunol* **23,** 1322-7.

SHIGEKAWA, K. & DOWER, W. J. (1988). Electroporation of eukaryotes and prokaryotes: a general approach to the introduction of macromolecules into cells. *Biotechniques* **6,** 742-51.

SHINKEL, T. A., COWAN, P. J., BARLOW, H., AMINIAN, A., ROMANELLA, M., LUBLIN, D. M., PEARSE, M. J. & D'APICE, A. J. (1998). Expression and functional analysis of glycosyl-phosphatidyl inositol- linked CD46 in transgenic mice. *Transplantation* **66,** 1401-6.

SIUZDAK, G. (1996). *Mass Spectrometry for Biotechnology*. Academic Press, San Diego & London.

SLIEKER, L. J., MARTENSEN, T. M. & LANE, M. D. (1986). Synthesis of epidermal growth factor receptor in human A431 cells. Glycosylation-dependent acquisition of ligand binding activity occurs post-translationally in the endoplasmic reticulum. *J Biol Chem* **261,** 15233-41.

SMITH, B. O., MALLIN, R. L., KRYCH-GOLDBERG, M., WANG, X., HAUHART, R. E., BROMEK, K., UHRIN, D., ATKINSON, J. P. & BARLOW, P. N. (2002). Structure of the C3b Binding Site of CR1 (CD35), the Immune Adherence Receptor. *Cell* **108,** 769-80.

STALKUP, J. R. (2002). A review of measles virus. *Dermatol Clin* **20,** 209-15, v.

STUDIER, F. W. (1991). Use of bacteriophage T7 lysozyme to improve an inducible T7 expression system. *J Mol Biol* **219,** 37-44.

TAKEDA, M., SAKAGUCHI, T., LI, Y., KOBUNE, F., KATO, A. & NAGAI, Y. (1999). The genome nucleotide sequence of a contemporary wild strain of measles virus and its comparison with the classical Edmonston strain genome. *Virology* **256,** 340-50.

TANAKA, K., MINAGAWA, H., XIE, M. F. & YANAGI, Y. (2002). The measles virus hemagglutinin downregulates the cellular receptor SLAM (CD150). *Arch Virol* **147,** 195-203.

TATSUO, H., ONO, N., TANAKA, K. & YANAGI, Y. (2000). SLAM (CDw150) is a cellular receptor for measles virus. *Nature* **406,** 893-7.

THOMAS, J. G., AYLING, A. & BANEYX, F. (1997). Molecular chaperones, folding catalysts, and the recovery of active recombinant proteins from E. coli. To fold or to refold. *Appl Biochem Biotechnol* **66,** 197-238.

THORLEY, B. R., MILLAND, J., CHRISTIANSEN, D., LANTERI, M. B., MCINNES, B., MOELLER, I., RIVAILLER, P., HORVAT, B., RABOURDIN-COMBE, C., GERLIER, D., MCKENZIE, I. F. & LOVELAND, B. E. (1997). Transgenic expression of a CD46 (membrane cofactor protein) minigene: studies of xenotransplantation and measles virus infection. *Eur J Immunol* **27,** 726-34.

UHLMANN, V., MARTIN, C. M., SHEILS, O., PILKINGTON, L., SILVA, I., KILLALEA, A., MURCH, S. B., WALKER-SMITH, J., THOMSON, M., WAKEFIELD, A. J. & O'LEARY, J. J. (2002). Potential viral pathogenic mechanism for new variant inflammatory bowel disease. *Mol Pathol* **55,** 84-90.

UJVARI, A., ARON, R., EISENHAURE, T., CHENG, E., PARAG, H. A., SMICUN, Y., HALABAN, R. & HEBERT, D. N. (2001). Translation rate of human tyrosinase determines its N-linked glycosylation level. *J Biol Chem* **276,** 5924-31.

VAN ZUYLEN, C. W., KAMERLING, J. P. & VLIEGENTHART, J. F. (1997). Glycosylation beyond the Asn78-linked GlcNAc residue has a significant enhancing effect on the stability of the alpha subunit of human chorionic gonadotropin. *Biochem Biophys Res Commun* **232,** 117-20.

VASSILEVA, A., CHUGH, D. A., SWAMINATHAN, S. & KHANNA, N. (2001). Expression of hepatitis B surface antigen in the methylotrophic yeast Pichia pastoris using the GAP promoter. *J Biotechnol* **88,** 21-35.

VOGELSTEIN, B. & GILLESPIE, D. (1979). Preparative and analytical purification of DNA from agarose. *Proc Natl Acad Sci U S A* **76,** 615-9.

VON PIRQUET, C. (1908). Verhalten der Kutanen Tuberkulin-reaktion wahrend der Masern. *Deutsch Med Wochenschr* **34,** 1297-1300.

WANG, G., LISZEWSKI, M. K., CHAN, A. C. & ATKINSON, J. P. (2000). Membrane cofactor protein (MCP; CD46): isoform-specific tyrosine phosphorylation. *J Immunol* **164,** 1839-46.

WARD, B. J., JOHNSON, R. T., VAISBERG, A., JAUREGUI, E. & GRIFFIN, D. E. (1991). Cytokine production in vitro and the lymphoproliferative defect of natural measles virus infection. *Clin Immunol Immunopathol* **61,** 236-48.

WELLS, L., VOSSELLER, K. & HART, G. W. (2001). Glycosylation of nucleocytoplasmic proteins: signal transduction and O-GlcNAc. *Science* **291,** 2376-8.

WHITE, C. E., KEMPI, N. M. & KOMIVES, E. A. (1994). Expression of highly disulfide-bonded proteins in Pichia pastoris. *Structure* **2,** 1003-5.

WILD, T. F. (1999). Measles vaccines, new developments and immunization strategies. *Vaccine* **17,** 1726-9.

WILES, A. P. (1996). D.Phil thesis, Oxford.

WILES, A. P., SHAW, G., BRIGHT, J., PERCZEL, A., CAMPBELL, I. D. & BARLOW, P. N. (1997). NMR studies of a viral protein that mimics the regulators of complement activation. *J Mol Biol* **272,** 253-65.

WOELK, C. H., JIN, L., HOLMES, E. C. & BROWN, D. W. (2001). Immune and artificial selection in the haemagglutinin (H) glycoprotein of measles virus. *J Gen Virol* **82,** 2463-74.

WONG, S. Y., GUILE, G. R., DWEK, R. A. & ARSEQUELL, G. (1994). Synthetic glycosylation of proteins using N-(beta-saccharide) iodoacetamides: applications in site-specific glycosylation and solid-phase enzymic oligosaccharide synthesis. *Biochem J* **300 (Pt 3),** 843-50.

WONG, T. C., YANT, S., HARDER, B. J., KORTE-SARFATY, J. & HIRANO, A. (1997). The cytoplasmic domains of complement regulatory protein CD46 interact with multiple kinases in macrophages. *J Leukoc Biol* **62,** 892-900.

WYSS, D. F., CHOI, J. S., LI, J., KNOPPERS, M. H., WILLIS, K. J., ARULANANDAM, A. R., SMOLYAR, A., REINHERZ, E. L. & WAGNER, G. (1995). Conformation and function of the N-linked glycan in the adhesion domain of human CD2. *Science* **269,** 1273-8.

XIE, M., TANAKA, K., ONO, N., MINAGAWA, H. & YANAGI, Y. (1999). Amino acid substitutions at position 481 differently affect the ability of the measles virus hemagglutinin to induce cell fusion in monkey and marmoset cells co-expressing the fusion protein. *Arch Virol* **144,** 1689-99.

XU, C., MAO, D., HOLERS, V. M., PALANCA, B., CHENG, A. M. & MOLINA, H. (2000). A critical role for murine complement regulator crry in fetomaternal tolerance. *Science* **287,** 498-501.

XU, R., AYERS, B., COWBURN, D. & MUIR, T. W. (1999). Chemical ligation of folded recombinant proteins: segmental isotopic labeling of domains for NMR studies. *Proc Natl Acad Sci U S A* **96,** 388-93.

YANAGI, Y. (2001). The cellular receptor for measles virus--elusive no more. *Rev Med Virol* **11,** 149-56.