# Glycosylation
# in
# *Dictyostelium discoideum*

Characterisation and location of glycans
on the spore coat protein SP96
in wild-type and fucose mutants of the cellular slime mold

A dissertation submitted to the

Fachbereich 9 Chemie der
Bergischen Universität-Gesamthochschule Wuppertal

For the degree of
Doctor of Natural Science (Doctor rerum naturalium)

presented by
Marcus Mreyen, Dipl.Chemiker
Wuppertal, 2000

*Dedicated to my aunt*
*Elke Barentzen*
*who passed away*

# *Acknowledgements*

*First and foremost I thank Prof. Günther Vogel (Bergische Universität Wuppertal, Department of Biochemistry) for the support in realising my own ideas and for his interest and comments during all stages of this work. Equally I thank Prof. Keith Williams (Macquarie University, Department of Biological Sciences, Sydney) for giving me the opportunity to work in his lab on an interesting topic, and for the support and the enthusiasm that flowed from him.*
*Further thanks to him for giving me the chance of using equipment at Proteome Systems Limited and for being part of a fantastic team.*

*A big thanks also to my inofficial supervisors Dr. Nicki Packer and Dr. Martin Slade who supported me with ideas, motivation and discussion whenever it was necessary. I also thank both for proofreading the chapters of this thesis.*

*Thanks to Prof. Dr. Altenbach (Bergische Universität Wuppertal) for taking over the third assessment.*

*I am also indebted to Dr. Brad Walsh for making it possible to use the equipment of APAF (Australian Proteome Analysis Facility) and to all the people who introduced me to the use of these things.*

*Thanks to Matt Harrison and Dr. Niclas Karlsson for help with the ESI–MS, Dr. D. Jardine for GC-MS analysis, Simon Gates and Angella Connolly* for *amino acid analysis and Dr. Peter Karuso for help with NMR analyis.*

*Prof. H. Freeze (The Burnham Institute, La Jolla, USA) and Prof. C. West (University of Florida, College of Medicine) are acknowledged for providing the monoclonal antibody AD7.5 and 83.5.*

*Thanks to my "lab mates", Alamgir, Lucy, Tash, Mark, Jenny, Sharon, Eva and Gerd for innumerable discussions and all the other members of the lab who made my stay at Macquarie University a great time.*
*Thanks to the members of PSL for the easy integration of my person into the team and for giving me an impression of perfect team work.*

*I would like to thank my colleagues in Germany, Christine, Tim, Andreas, Stefan and Guido for many discussion and for keeping in touch with me during my time in Australia.*

*My special thanks to my former girlfriend Ulrike for her big support during my study and the time of my PhD.*

*I would like to thank my parents for financial support, my aunt, who died in a tragic manner in the last month of my writing and my uncle, for their belief in my skills and person.*

*Many thanks to all the people who made my time in Australia an unforgettable experience and by far more than only a research visit.*

*This work was funded by Deutschen Akademischen Austauschdienst DAAD and Graduierten-förderung des Landes NRW*

***Thanks for this!!***

# Table of Content

# List of Figures

# List of Tables

# Zusammenfassung

Der zelluläre Schleimpilz *Dictyostelium discoideum* zeigt in Bezug auf posttranslationale Modifikationen viele Parallelen zu höherentwickelten Organismen. Aufgrund dieser Tatsache und wegen seiner einfachen Kultivierung wurde dieser Eukaryont herangezogen um die funktionelle Bedeutung von Glycanstrukturen an Proteinen zu untersuchen. Mutanten mit einem reduzierten Fucosylierungsgrad zeigen phänotypische Veränderungen in der Lebensfähigkeit ihrer Sporen und präsentieren im Vergleich zum Wildtyp veränderte glycanabhängige Epitope an deren Proteinen. Bislang fehlte eine detaillierte Charakterisierung der vorhandenen Glycanstrukturen in Wildtyp und Mutanten. Diese ist jedoch erforderlich, um den beobachteten Phänotyp auf molekularer Ebene erklären zu können. Anhand des prominenten Sporenmantelproteins SP96 wurde in der vorliegenden Arbeit eine eindeutige strukturelle Aufklärung der in den verschiedenen Mutanten vorliegenden posttranslationalen Modifikationen erreicht.

Am Protein des Wildtyps konnten etwa 20 an Serine gebundene Fucose-Einheiten, sowie etwa 60 Einheiten der neuartigen, phosphodiester gebundenen Disaccharidstruktur (Fuc(α1-3)GlcNAc-α-1-P) an Serine nachgewiesen werden. Diese wurde mittels GC-MS und LC-MS Analysen, sowie durch 1D- und 2D-NMR-Spektroskopie eindeutig identifiziert.

Die Quantifizierung der posttranslationalen Modifikationen am SP96 Protein erfolgte mittels HPAEC-PAD Monosaccharid- und GC-MS Analyse, sowie eines colorimetrischen Phosphattests und zeigte, daß im Protein des Wildtyps über 70% der Serine glykosyliert vorliegen. Die Abnahme der Fucose in den zwei Mutanten auf etwa 20 % bzw. < 5% im Vergleich zum Wildtyp konnte in einer Mutante eindeutig dem Verlust der terminalen Fucose von der Disaccharideinheit, in einer anderen Mutante dem weiteren Verlust der *O*-glycosydisch gebundenen Fucose zugeordnet werden. Die Lokalisierung der vorhandenen Glycanstrukturen erfolgte nach enzymatischer Spaltung des Proteins und anschließender Edman-Sequenzierung der mittels Antikörper markierbaren Fragmente. Darüberhinaus wurde ein bereits in *D. discoideum* etabliertes Expressionssystem benutzt, um Peptidmotive aus SP96 auf ihre potentielle Glykosylierbarkeit hin zu untersuchen.

Es konnte gezeigt werden, daß alle Modifikationen an der Aminosäure Serin vorliegen, wobei der überwiegende Teil am C-terminus des Proteins lokalisiert zu sein scheint. Die identifizierten, quantifizierten und lokalisierten Glycanstrukturen konnten abschließend in direkten Zusammenhang zu den Epitopen monoklonaler Antikörper gebracht werden.

Die phänotypischen Beobachtungen lassen sich so erklären, daß der Verlust der direkt an Serin gebundenen Fucose, im Gegensatz zu der terminal am Disaccharid gebundenen, einen größeren Einfluß auf die korrekte Eingliederung von SP96 in den Komplex von Sporenmantelproteinen hat. Dies führt letztendlich zu einem permeableren Sporenmantel.

# Abstract

In relation to post-translational modifications the cellular slime mold *Dictyostelium discoideum* has many features in common with higher eukaryotic cells. Therefore this easy-to-handle eukaryotic organism was chosen to study the relationship of glycosylation patterns on proteins to their function. Mutants with a lack in fucosylation have a phenotype of less viable spores and present different glycan dependant epitopes on proteins. Until now no detailed study has been done to characterise the different oligosaccharide structures in order to elucidate the relationship of the glycan to function and to relate the glycan structures to the antibody epitopes.

In this thesis the exact differences in the post-translational modifications of wild-type and mutants of *D. discoideum* were analysed on the purified major spore coat protein SP96.

The wild-type protein contains approximately 20 serine linked fucose residues and approximately 60 sites with a newly identified, *O*-linked phosphodiester-linked disaccharide structure (Fuc(α1-3)GlcNAc-α-1-P) attached to serine. This new structure was identified by GC-MS and LC-MS analysis as well as by 1D and 2D-NMR spectroscopy.

It could be shown that the decreased amount of fucose in one mutant to 20% of the wild-type content was the result of a missing terminal fucose on the newly identified disaccharide structure. Another mutant, with a further decreased amount of fucose (less than 5% of the wild-type content) lost furthermore the *O*-linked fucose residues.

The identified glycan structures were correlated with the epitopes of some monoclonal antibodies. This allowed some conclusions as to the mutations in the different mutants to be drawn.

The quantification of post-translational modifications of the SP96 protein was determined by HPAEC-PAD monosaccharide and GC-MS analysis as well as by a colourimetric phosphate assay. This showed that more than 70% of the serines in wild-type SP96 were glycosylated.

The localisation of modifications on SP96 was also partially achieved. To localise some of the glycan structures on SP96 it was enzymatically digested, and peptide fragments detected by monoclonal antibodies were used for Edman sequencing. A second approach used an already established *D. discoideum* expression system to analyse recombinant peptide motifs of SP96 for their *O*-glycosylation. It was shown that serine was the only modified hydroxyamino acid and that the main location of the glycosylation was the serine-rich C-terminus of the protein.

It can be concluded that *O*-linked fucose has a higher influence on the correct integration of SP96 in the spore coat than the terminal fucose of the disaccharide. This, finally, leads to the increased permeability of the spore coat.

# Publication

Marcus Mreyen, Alan Champion, Supriya Srinivasan, Peter Karuso, Keith L. Williams and Nicolle Packer

**Multiple O-Glycoforms on the Spore Coat Protein SP96 in *Dictyostelium discoideum***

**Fuc(α1-3)GlcNAc-α-1-P-Ser is the major modification**

Journal of Biological Chemistry (2000), Vol.275, No. 16, pp. 12164-12174

## Amino acids

| | | |
|---|---|---|
| Alanine | Ala | A |
| Arginine | Arg | R |
| Asparagine | Asn | N |
| Aspartic acid | Asp | D |
| Cysteine | Cys | C |
| Glutamic acid | Glu | E |
| Glutamine | Gln | Q |
| Glycine | Gly | G |
| Histidine | His | H |
| Isoleucine | Ile | I |
| Leucine | Leu | L |
| Lysine | Lys | K |
| Methionine | Met | M |
| Phenylalanine | Phe | F |
| Proline | Pro | P |
| Serine | Ser | S |
| Threonine | Thr | T |
| Tryptophan | Trp | W |
| Tyrosine | Tyr | Y |
| Valine | Val | V |

# Abbreviations

| | |
|---|---|
| AA | amino acid |
| Ara*f* | Arabinofuranose |
| ax | axial |
| COSY | correlated spectroscopy |
| CP | Cysteine proteinase |
| csA | contact site A |
| dNTP | deoxynucleotide mix |
| EGF | epidermal growth factor |
| eq | equatorial |
| ER | endoplasmic reticulum |
| ESI | electrospray ionization |
| Fuc | Fucose |
| Gal | Galactose |
| GalNAc | *N*-acetylgalactosamine |
| GC | gas chromatography |
| GDP-Fuc | Guanosine diphosphofucose |
| Glc | Glucose |
| GlcNAc | *N*-acetylglucosamine |
| GlcNAc-1-P | *N*-acetylglucosamine-1-phosphate |
| GST | Glutathione S-transferase |
| HPAEC-PAD | high performance anion exchange chromatography |
| HRP | horse radish peroxidase |
| HSQC | heteronulear single quantum coherence |
| mAb | monoclonal antibody |
| Man | Mannose |
| MS | mass spectrometry |
| mS | milli Siemens |
| NeuAc | *N*-acetylneuraminic acid |
| NMR | nuclear magnetic resonance |
| NOE | nuclear Overhauser effect |
| PAGE | polyacrylamide gel electrophoresis |

| | |
|---|---|
| PCR | polymerase chain reaction |
| Pi | anorganic phosphate |
| pm | pico meter |
| ppm | parts per million |
| PsA | prespore specific Antigen |
| PSV | prespore vesicle |
| PVDF | polyvinylidene difluoride |
| Rha | Rhamnose |
| ROESY | rotating-frame NOE spectroscopy |
| rpm | rounds per minute |
| rPsA | recombinant PsA |
| SP | spore coat protein |
| TFA | Trifluoroacetic acid |
| TOCSY | total correlation spectroscopy |
| TOF | time of flight detection |
| U | Unit |
| UDP | uridine diphosphate |
| UV | ultraviolet |
| Xyl | Xylose |
| (v/v) | volume per volume |
| (w/v) | weight per volume |

# 1 Introduction

This dissertation deals with the localisation, qualitative and quantitative determination of post-translational modifications to the spore coat glycoprotein SP96 in the eukaryotic organism *Dictyostelium discoideum.* The vegetative amoebae of *D. discoideum* have many features in common with mammalian cells and are therefore suitable for basic research on differentiation, cell motility, membrane trafficking, cell-cell recognition and different aspects of glycobiology.

The spore, the final stage in the organisms developmental process, is created to survive long starvation periods. The coat surrounding the spore is the protective barrier against environment stress, as for example against protease digestion and dehydration. In recent years progress has been made in understanding spore coat assembly (120, 121, 136, 137). The major part of the spore coat is built of glycoproteins. The functional importance of glycosylation of spore coat proteins could be demonstrated in survival of mutants with a lack in fucosylation. The mutants show a higher permeability of the spore coat which results in a lower viability. However, the detailed structures of protective glycans are still unknown.

Earlier experiments on glycosylation of proteins in *D. discoideum* involved the use of monoclonal antibodies created against late stage glycoproteins of wild-type and mutants lacking fucosylation (10, 11, 48). A panel of antibodies recognised more or less undefined carbohydrate structures on spore coat proteins, including SP96 and PsB as well as cysteine proteinases secreted from vegetative amoebae (12).

That glycosylation is developmentally regulated could be demonstrate using the mutants which lack fucose causing the absence of carbohydrate epitope on proteins of only one stage of development compared to wild-type proteins. Spore coat proteins of the mutant lack the carbohydrate depending epitope, but the same strain is still able to construct the carbohydrate structure recognised by the antibody on the cysteine proteinase, expressed by vegetative growing amoebae.

This developmentally regulated glycosylation and the phenotypic characterisation of fucose mutants led to research focussing on the glycosylation structures recognised by the monoclonal antibodies. A major spore coat protein, SP96, of the wild-type strain and fucose mutants has several carbohydrate epitopes. This glycoprotein is important for the correct assembly of the spore coat. Characterisation of its post-translational modifications and their localisation on SP96 provided an opportunity to relate to the antibody epitopes, and shed light on glycan/ function relationships.

## 1.1 Glycosylation of proteins

Glycoproteins are proteins with different carbohydrates covalently bound at distinct amino acids within the protein backbone. The content of carbohydrate can vary from less than 1 % in some collagens to more than 85 % in some blood group substances (82, 115).

Although studies of glycoproteins date back more than a century, the enormous significance of glycosylation as a common and highly diverse co- and post-translational modification on proteins has really started to be appreciated only in the last decade. Increasing the sensitivity of the methods used to detect and analyse accurately glycosylation has revealed the diversity in both structures and functions of the carbohydrate component of proteins. Glycosylation can effect the physicochemical or biological nature of proteins (Table 1-1). Frequently, a relationship can be observed between the amount of glycosylation on a protein and its function (Figure 1-1). However, whilst the function of glycosylation is becoming well understood in some biological processes, the relationship between the glycan structure and their functions still remains speculative.

Glycosylation of proteins mainly occurs by the action of glycosyltransferases, enzymes which catalyse the transfer of carbohydrate residues from activated sugar donors to the protein backbone or to already attached glycans. These enzymes are predominately resident membrane proteins of the endoplasmic reticulum and the Golgi apparatus. The sequential processing of glycan structures goes through these compartments. Glycosyltransferases are mainly highly specific for their donor and acceptor substrates, resulting in specific linkages of the attached carbohydrate (105).

The carbohydrate components of glycoproteins are often heterogeneous. Variation at a single site of glycosylation is termed microheterogeneity and diversity between sites of glycosylation is termed macroheterogeneity. This combined heterogeneity results in many discrete subsets or glycoforms of a glycoprotein (25, 110). Diversity can be observed of carbohydrates attached to a single protein produced by different tissues or organisms. Even in one cell type glycosylation may vary depending on the cell cycle, state of differentiation and development. It is also known that during the progression of some diseases, for example cancer and Alzheimer, a change of glycosylation in some proteins can be observed (52). Glycosylation is also important during embryonic development but in mammals these stages are not easily manipulated experimentally. Therefore, for a better understanding of this structure/ function relationship in development, it is useful to analyse the glycosylation of less complex organisms, working as a model for differentiation and development. In this case the

glycosylation is easier to manipulate using inhibitors or mutations and changes can be related to phenotypical observations.

| Effect | Functional effect |
|---|---|
| Physicochemical | Modifies solubility, electrical charge, mass, size and viscosity in solutions. Controls protein folding, stabilize protein conformation, confers thermal stability and protection against proteolysis |
| Biological | Modifies protein trafficking and turnover, immunological properties. Modulates activity of enzymes and hormones, cell recognition and adhesion, signal transduction. |

**Table 1-1 Function of the glycan component of glycoproteins.**

Adapted from Sharon and Lis **(82, 115)**.

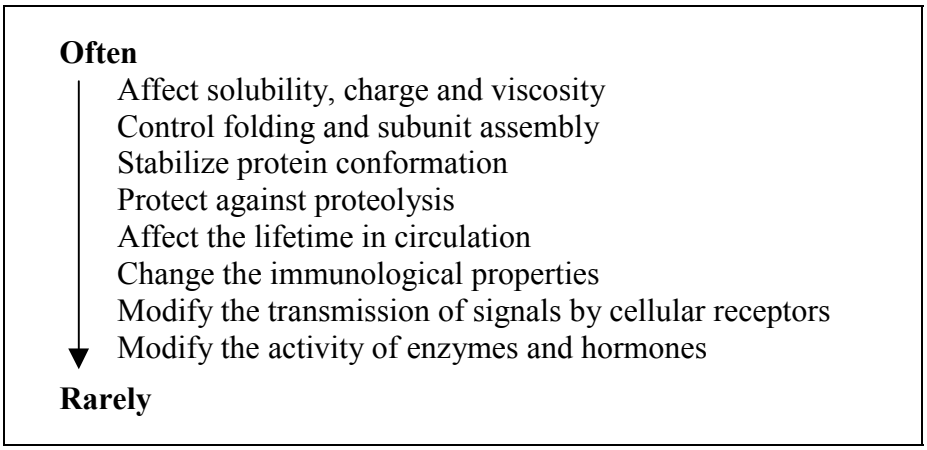



**Figure 1-1 Relationship of glycosylation level/ function**

## 1.2 Types of protein glycosylation

Glycosylation of proteins is classified into different types, depending on the carbohydrate-peptide linking group. A recent review (115) describes three major types of glycosylation (A-C) and furthermore three minor types (D-F).

A) *N*-linked glycosylation: the reducing terminal sugar is covalently bound to the amide group of the amino acid asparagine (77, 110).

B) *O*-linked glycosylation: the reducing terminal sugar is bound to the hydroxyl group of an amino acid, commonly serine (Ser) or threonine (Thr) and less frequently tyrosine (Tyr), hydroxylysine (Hly) or hydroxyproline (Hpr).

C) Glycosylphosphatidylinositol (GPI)-anchor: the carboxy terminus of a protein is attached via ethanolamine to a carbohydrate structure containing inositol, which in turn is connected to a hydrophobic lipid membrane anchor (27).

D) Phosphoglycosylation: the linkage of a carbohydrate through a phosphodiester bond to a hydroxy amino acid, such as Thr or Ser (62).

E) *C*-mannosylation: the indole ring of tryptophan (Trp) has a mannose residue attached (19, 64).

F) Ribosylation: the linkage of the 3-hydroxyl group of ribose to glutamic acid (Glu), asparagine (Asp), arginine (Arg) or cysteine (Cys).

Of special interest in this thesis are *N*-linked, *O*-linked and phosphodiester linkages.

### 1.2.1 *N*-linked glycosylation

*N*-linked glycosylation in eucaryotes is the best analysed form of glycosylation so far. The core structure of $GlcNAc_2Man_5$ is built on the cytoplasmic side of the endoplasmic reticulum (ER), on the lipid carrier molecule dolichol, which consists of 16-20 isoprene units. After translocation of the glycan core structure to the lumen of the ER, it is modified by the addition of four mannose and three glucose residues, resulting in the structure $GlcNAc_2Man_9Glc_3$. Finally this glycan is transferred *en bloc* to the growing polypeptide at the amino acid asparagine residue in the consensus tripeptide sequence Asn-X-Ser/Thr (X≠Pro) (77). This glycosylation occurs co-translational as the growing peptide is

translocated into the ER. After transfer of the glycoprotein to the Golgi apparatus the glycan undergoes reconstruction, being trimmed to a core structure and rebuilt to form three structural types of *N*-linked oligosaccharides. These structures are called A) high mannose-type, B) complex-type and C) hybrid type (Figure 1-2), but may possess a broad range of structural variation.

An important tool in the analysis of *N*-linked glycans are specific *N*-glycosidases, used to release the reducing end of the glycan from the protein. For a complete glycan release it requires denaturation of the protein which is not suitable for functional studies. Functional studies on the effect of *N*-glycans on intact proteins have been done in three ways. Inhibitors like tunicamycin, which inhibit the first glycosylation step, the addition of GlcNAc to dolicholphosphate resulting in proteins lacking *N*-glycans. Secondly mutants defective, or deficient in one or more glycosyl transferases (129). The effect of altered *N*-linked glycosylation goes over a broad range depending on the modified protein (128). For example, most lysosomal enzymes lose the targeting signal for delivery to lysosomes or lose their activity (128). On the other hand, some glycoprotein hormones show complex effects like unusual combinations of subunits, or an agonist being converted into an antagonist by loss of glycosylation or changes in specifity (128). Apart from complex effects, also the simple loss of protection can be observed on proteins resulting from changed conformation (128).

In the last decade, several new carbohydrate structures linked to asparagine have been

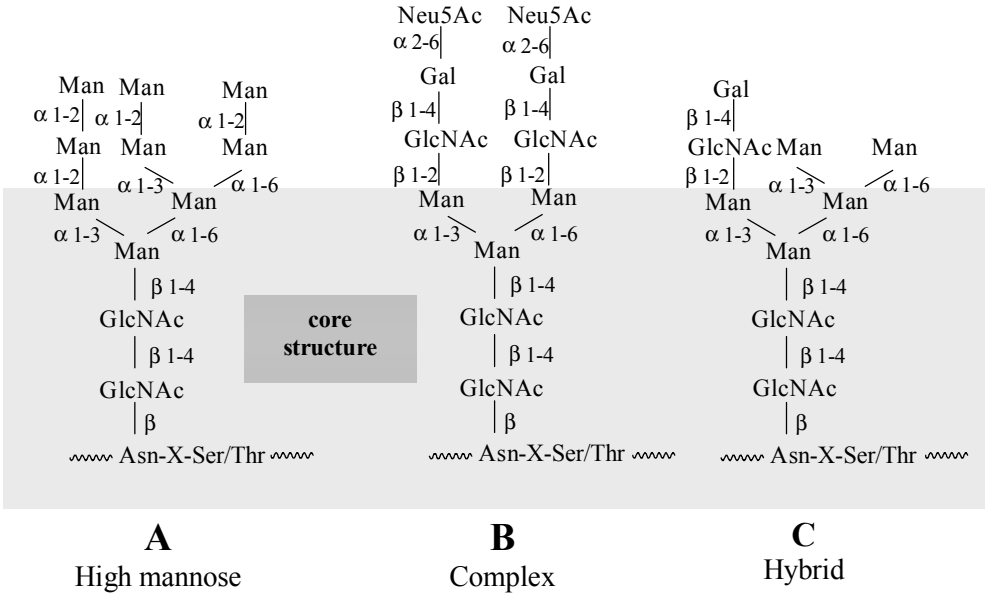



**Figure 1-2 Representative structures of *N*-linked oligosaccharides**

**A)** high mannose–type, **B)** complex-type, **C)** hybrid-type

discovered, mainly in bacterial glycoproteins and all cases fit the tripeptide Asn-X-Ser/Thr consensus sequence mentioned above (115). Table 1-2 shows all known reducing terminal linkages of carbohydrate to asparagine (115).

In a recently published paper it was assumed that the majority of proteins, containing the consensus sequence, will be found to be glycosylated (3) provided that they have a signal sequence i.e. transit to the ER/Golgi.

| Reducing terminal linkage | Amino acid | Occurrence |
|---|---|---|
| βGlcNAc | Asn | Eukaryote |
| βGalNAc | Asn | Archaebacteria |
| βGlc | Asn | Animals (only laminin) |
| | | Archaebacteria |
| Rha | Asn | Eubacteria |

**Table 1-2Carbohydrate-asparagine linkages** (115)

### 1.2.2 *O*-linked glycosylation

*O*-linked glycosylation is less well characterised than *N*-linked glycosylation. The carbohydrate is commonly attached via a hydroxyl group of the amino acid serine or threonine and infrequently to tyrosine, hydroxylysine or hydroxyproline. *O*-glycosylation is mainly a post-translational and post-folding event; therefore, only serine and threonines on the protein surface will be glycosylated (127). *O*-linked glycosylation can be initiated in several subcellular compartments. It may occur in the ER or in the three regions of the Golgi apparatus (proximal-, intermediate- and beyond-compartment). In addition cytosolic *O*-glycosylation, *O*-GlcNAc, occurs in eukaryotes. Furthermore, the presence/ location of glycosyltransferases is cell- or tissue-specific (14).

*O*-linked glycosylation has no defined core glycan structure unlike eukaryotic *N*-linked glycosylation. There is a wide variety of reducing terminal sugar residues (**Table 1-3**). Size may vary from a single unit (for example *O*-GlcNAc) up to elongated

oligosaccharides more than 100 residues in length, as found on some cell surface proteins (115). One problem in analysing *O*-glycans is the lack of a generally applicable enzyme for releasing a broad range of *O*-linked glycans from the protein, like for example, PNGase F for *N*-linked sugars.

Determing the exact localisation of *O*-glycosylation sites is also a very complicated issue, since all chemical methods used to release the *O*-glycan also destroy the protein. For example, peptides cannot be analysed by mass-spectrometry after releasing their *O*-linked modification because commonly the peptide is degraded afterwards. One successful way of identifying the glycosyated sites of native glycoproteins is solid phase Edman sequencing with aqueous phase chromatography, which allows the identification of modified amino acids and the identification of the attached reducing sugar (47). Identification of acceptor sites for *O*-glycosylation has also been carried out using *in vitro* glycosylation assays (6, 94, 95, 122, 133, 152) and more recently using *in vivo* glycosylation systems (71, 91, 92). In the latest *in vivo* glycosylation systems, a carrier protein is expressed with an attached peptide which can be modified by the organisms "glycosylation machinery". The carrier protein is chosen to be easily purified by affinity chromatography and the glycopeptide used for further characterisation (70, 71).

The characterisation of different glycosylation sites should help to predict rules when *O*-glycosylation will occur (46). No peptide specific consensus sequence for *O*-glycosylation can yet be defined. Indeed, with the possible motifs already identified (28, 46, 149), it seems unlikely that a single consensus sequence exists, particularly in view of the larger variety of reducing terminal carbohydrates. So far, two neuronal networks[a] have been developed that predict sites of *O*-glycosylation based on "training sets" compiled from known glycosylation sites (53, 57). One predicts the sites of αGalNAc on secreted mammalian proteins and the other the addition of αGlcNAc to the polypeptide backbone of secreted cell surface proteins of *D. discoideum* (57).

*O*-glycosylation has been observed at single sites or accumulated in clusters in proteins, in contrast to eukaryotic *N*-linked glycosylation with its mostly well separated glycosylation sites. Often *O*-glycosylation can be clustered in serine and/ or threonine rich repeat regions, so called "mucin domains", sometimes interrupted by single proline residues. Mucins are glycoproteins with highly glycosylated domains of oligosaccharides linked via α–GalNAc to Ser/ Thr. The extensive glycosylation on mucins and in the functional domains

[a] http://www.cbs.dtu.dk/services/NetOGlyc/

of many cell-surface proteins is believed to force the protein backbone into a semi-rigid or extended conformation. This "bottlebrush glycosylation" provides spatial separation between functional domains, or projects functional domains away from the cell membrane (68). It is speculated that the high carbohydrate content on mucins influence their physicochemical functions such as hydrodynamic properties, solubility and buffering capabilities and, moreover, has biological function in cell–adhesion, migration and differentiation (130).

| Reducing terminal linkage | Amino acid | Occurrence |
|---|---|---|
| Ara*f* | Hyp | Plants |
| α-Fuc | Ser, Thr | Animals |
| α-Gal | Hyp | Plants, eubacteria |
| α-Gal | Ser | Plants, eubacteria |
| β-Gal | Hyl | Animals (collagen only) |
| β-Gal | Tyr | Eubacteria |
| α-GalNAc | Ser, Thr | Eukaryotes |
| α-Glc | Tyr | Animals (glycogen only) |
| β-Glc | Ser, Thr | Eubacteria, Animals |
| α-GlcNAc | Ser, Thr | Protozoa |
| β-GlcNAc | Ser, Thr | Animals |
| α-Man | Ser, Thr | Yeast, animals |
| β-Xyl | Ser | Animals |

**Table 1-3 Carbohydrate-hydroxy amino acid linkages**

### 1.2.3 Phosphoglycosylation

The first reported example of a protein modified by phosphoglycosylation was an endopeptidase known as Proteinase I, isolated from the cellular slime mold *Dictyostelium discoideum* (54). A released glycan contained GlcNAc-1-$PO_4$, later confirmed as single GlcNAc residues phosphodiester linked to serine (87). A highly specific monoclonal antibody against GlcNAc-1-$PO_4$ was used to identify further phosphoglycosylated proteinases in this

organism (101, 118). Recently, a new phosphoglycosylation (Fuc-β-1-P) was identified on an artificial peptide and further characterised as an antibody epitope in *D. discoideum* (119).

Phosphodiester linked carbohydrates were also identified on secreted proteins of the protozoan flagellate *Leishmania*, where this class of glycosylation could act as a secretion signal (62). In *Trypanosoma cruzi*, a kinetoplastid parasite, phosphoglycosylation was detected on some cell surface proteins (63) (Table 1-4). The variety of glycan structures ranges from single carbohydrates up to long oligosaccharides, which may be formed from phosphodiester-linked disaccharide repeat units. The attached glycans in *Leishmania* are linear or branched, and may be further phosphorylated (62, 67). Phosphoglycosylation is found in repetitive Ser/ Thr rich motifs (145), but at this time no consensus peptide sequence has been reported for phosphoglycosylation (86).

One feature common to all phosphoglycoproteins characterized is a high level immunogenicity, which assist the production of a large number of highly specific monoclonal antibodies (16, 65). The strength of the immune response suggests that phosphoglycosylation is unlikely to occur in mammalian cells.

| Reducing terminal linkage | Amino acid | Occurrence |
|---|---|---|
| Fuc* | Ser | *D. discoideum* |
| GlcNAc* | Ser | *D. discoideum* |
| α-Man | Ser | *Leishmania mexicana* |
| Xyl | Ser/ Thr | *Trypanosoma cruzi* |

**Table 1-4 Phosphoglycosylation adapted from Haynes** (62).
* single carbohydrate unit linked over phosphodiester

## 1.3 *Dictyostelium discoideum*- a model system for developmental glycobiology

Our understanding of glycobiology of multicellular organisms is largely based on studies in a limited number of model systems such as *Drosophila* and the *Muridae* (mouse). *D. discoideum* has a unique life cycle, during which it grows as single cells, but on starvation cells aggregate together to form a multicellular organism and progressing through a series of well defined developmental steps. In this study the cellular slime mold *D. discoideum* has been used as a model to study glycobiology during growth and development. The amoeba exhibits the same types of protein modifications as are found in complex eukaryotes, including sulphation, phosphorylation and *N*-linked and *O*-linked glycosylation (9, 38, 39). There are two reports of sialic acid (NeuNAc) in *D. discoideum* (153, 154), although modern studies have failed to find sialylation in this organism. To date relatively few glycan structures have been solved in *D. discoideum*. Progress so far has relied on immunochemisty. Since many *Dictyostelium* glycans are highly immunogenic in mammals, monoclonal antibodies have been generated against carbohydrate antigens expressed at different stages in development. Although the structures the antibodies recognize are unknown, they are an important tool for studying developmental regulation of the respective epitopes and physiological consequences of their loss in mutant strains (40). The generation of mutants with altered glycosylation patterns is greatly assisted by the cells being haploid and parasexual genetic techniques to compliment mutations (148).

### 1.3.1 *D. discoideum* life cycle

*D. discoideum* is an unicellular free-living amoeba (~10 μm) and was first isolated from temperate forest soil by Raper in 1933 (107). The unicellular amoeba is fed by phagocytosis of bacteria. When all the bacteria are consumed, starvation triggers a complex developmental program (Figure 1-3). Cells migrate into mounds using pulses of cyclic adenosine 3′, 5′-monophosphate (cAMP) as a chemotactic signal. As a result groups of $10^3$-$10^5$ cells become mutually adhesive and undergo a synchronous morphogenesis. The mounds of cells form into a motile, multicellular structure the so called “slug”, that migrates along light and heat gradients. During this time the cells differentiate into two cell types, prespore cells in posterior 2/3 and prestalk cells in the anterior 1/3 of the slug. Approximately 24 h

after aggregation the slug cells form a fruiting body (2-3 mm). The mature fruiting body contains about 70,000 spores, supported by a stalk and basal disc of about 30,000 cells sheathed in cellulose walls to hold up the mass of spores. Each spore encloses a viable, dormant cell. Under proper environmental conditions (17, 18), like exposure to water and a suitable supply of nutrients, the dormant spore germinates, rupturing the spore coat to release a viable, free living amoeba.

*Dictyostelium* is an eukaryotic organism; it is easy to handle and shows a tightly regulated pattern of development. Moreover, strains exist which can also be cultivated in liquid media like bacteria suspension and axenic- and synthetic-media, which provide an easy to handle developmental system.

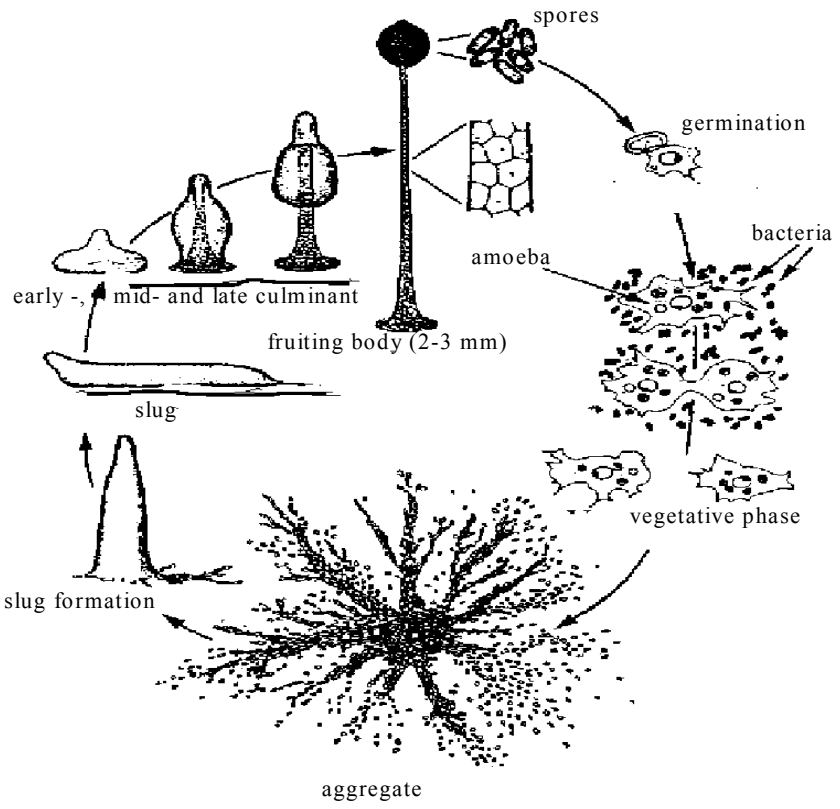



**Figure 1-3 Life cycle of *Dictyostelium discoideum***

### 1.3.2 Spores

The spore is an important developmental stage of the organism, which allows single viable cells to survive long periods of starvation. To protect the enclosed cell from environmental stress, as desiccation, heat and UV light, the individual spores are covered by a multi-layered coat (30). The spore wall outer layer consists of glycoproteins, a middle layer contains cellulose and the inner layer containing further glycoproteins and galactose/ *N*-Acetyl galactosamine polysaccharides and appears to have some physical association with the plasma membrane (140) (Figure 1-4). The cellulose plays an important structural role in the spore coat assembly. Enzymes such as α-Mannosidase and acid phosphatase are located in the inner spore coat and in the soluble matrix surrounding the spore (140). The spore coat is composed of approximately 50 % cellulose and 50 % glycoprotein (140). More than 25 different spore coat proteins have been identified, however some of them have been only described by single research groups.

Immuno-precipitation and immunogold labelling experiments displayed that during the assembly of the spore coat, several structural spore coat proteins (SP) and other essential components are secreted from prespore cells from prespore vesicles (PSVs). The PSVs appear in the posterior prespore cells during the latter half of morphogenesis and at culmination they move to, and synchronously fuse with the plasma membrane. Their contents (proteins and other polysaccharide components) are deposited into the extracellular space and then incorporated into the spore coat (120). Up to know more than 25 different spore coat proteins have been identified, although some of them have been described by only one research group. Four of the major spore coat proteins, SP96[b], SP85 (PsB), SP70 and SP60, are preassembled in a so called PsB protein complex inside the prespore vesicles (120, 136). One important feature for the proper integration of the PsB complex in the spore coat is its cellulose binding activity which is generated by assembly of at least SP96, PsB and SP70 (84, 121). Another spore coat protein SP75 (found within PSVs but not as part of the PsB complex) is intergrated later into the spore coat after secretion and gives further stability to the spore coat. After the fruiting body has been formed some spore coat proteins of the contents of PSVs are more strongly associated with the spore coat and are also disulfide-cross-linked to each other (147). This association is supported by the observation that most of the spore coat proteins have a

[b] the number after the abbreviation SP for spore coat protein is derived from the size of the protein band after SDS-PAGE separation.

relatively high content of cysteine in their amino acid sequence and were only extracted from the spores during SDS or urea treatment and SDS or urea/ reducing conditions, respectively (136, 147). Fluorescence staining with a monoclonal antibody which recognises SP96 (MUD3) demonstrates that SP96 is associated with the outer and inner layer of the coat and therefore seems to traverse the cellulose layer (108, 121, 136). Additional SP96 can be found in the soluble material from between the spores. SP85 is found only in the inner layer of the spore coat, whereas SP75 is in the outer layer. SP70 and SP60 are also found in the outer layer, although a location in the middle or inner spore coat cannot be excluded (40).

The spore coat proteins SP96, SP85 (PsB), SP75, SP70 and SP60 contain frequented amino acids sequences with some similarities to epithelial mucin repeats (143). The similarity to mucin proteins is of particular interest as this indicates a possible protective function in the spore coat. It is almost certain that the spore coat proteins are glycosylated within these mucin-like regions.

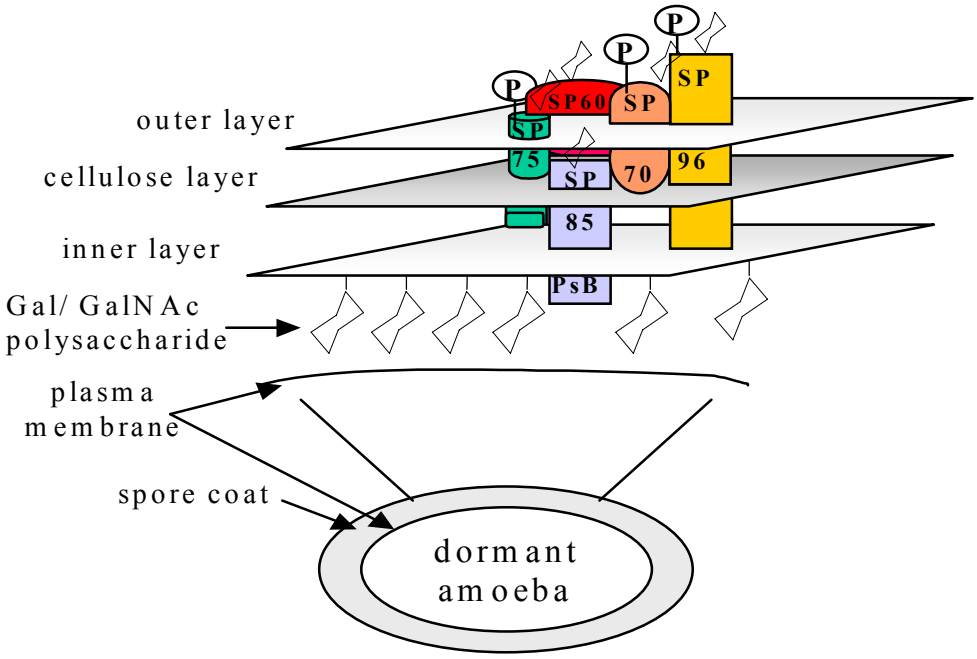



**Figure 1-4 Schematic assemble of the spore coat with integrated PsB spore coat protein complex**

only the best analysed spore coat proteins are shown (References see Table 1-6)

◊ modified by glycosylation, Ⓟ modified by phosphorylation

### 1.3.3 Tools for glycosylation analysis of spore coat proteins

Most studies on the glycobiology of *D. discoideum* have been done using tools such as mutant strains and monoclonal antibodies. Although the exact structures that the antibodies recognize are unknown they have been used to define different types of glycosylation (11) and for identification and purification of fully-, partly-, or non-glycosylated forms of the molecule being studied. An expression system established in *D. discoideum* has been used for the analysis of *O*-linked glycosylation acceptor motifs (24, 70, 71).

#### 1.3.3.1 *D. discoideum* glycosylation mutants

After mutagenesis mutants were isolated which miss a wild-type specific antigen. Up to now various glycosylation mutants have been isolated, whereby at least 5 different genes could be identified. Defect in the *mod*A locus influences *N*-glycosylation and results in incompetence to trim the *N*-glycan. Defect in the *mod*B locus influences *O*-glycosylation and leads to the absence of GlcNAc on proteins. The defects leading to the *mod*C, *mod*D and *mod*E loci mutation all show changes in the fucosylation on *O*-glycans (11, 39).

In glycosylation studies done by Champion *et al*. (10) the strain X22, which was generated from the parent strain NC4 (148), was used as a wild-type. Amoeba of strain X22 were mutagenized, plated out and slugs of individual colonies were screened for the loss of carbohydrate dependent antibody epitopes of MUD50 and/ or MUD62 (see 1.3.3.2.1 and 1.3.3.2.2). Finally two strains, HU2470 and HU2471 with independent mutations in loci *modD* and *modE*, respectively, were found to lack the MUD62 epitope but still retain the MUD50 epitope (10, 11). Biochemical analysis of spores showed in strain HU2470 (*modD*) a decreased fucose content of 17 % and in HU2471 (*modE*) of 5.8 % of the wild-type (X22) level (11). Both strains developed normally and produced spores, although carrying a mutated *mod*E gene had a lower viability after storage (121). The aggregation was somewhat delayed and HU2471 (*modE*) produced slightly smaller colonies than its parent strain X22. Fucosylation is also implicated in strains with a mutant allele in the *modC* which show definite phenotypes (11). The defect in the *modC* locus disables formation of GDP-fucose from GDP-mannose, and therefore the gene may encode the dehydratase or epimerase enzyme activities responsible for this conversion (43). The resulting total lack of fucose is reflected in phenotypes such as slow proliferation, a defective slime sheath and a permeable spore coat. The latter result in a less protected dormant cell with lower viability.

### 1.3.3.2 Carbohydrate specific antibodies

Studies of glycoproteins in *D. discoideum* using mAbs have used two approaches. One approach is commonly termed the "shotgun" approach, where monoclonal antibodies are produced from crude cellular extract containing many different proteins, and the proteins identified by the mAbs are further examined. For example, a preparation of cells or cellular membranes of multicellular-stage cells used as an antigen could be expected to yield mAbs which recognise cell surface molecules, candidates for cell adhesion and cell interaction functions (10, 48, 50). Alternatively, molecules have first been purified by conventional protein chemistry, and then monoclonal antibodies have been produced to these purified molecules. Monoclonal antibodies have also been prepared from mutant strains which lack immunodominant carbohydrate epitopes.

In the following mainly monoclonal antibodies are described which were relevant for the glycan analysis of the spore coat protein SP96 (Table 1-5).

| name of mAb | mAb raised against | Description of epitope and examples of recognised proteins | Glycoantigen | strains (epitope absent)/ mutation |
|---|---|---|---|---|
| MUD50 | Slug sheath proteins of *D. discoideum* | *O*-linked glycan, resistent to mild acid hydrolysis, recognise GlcNAc linkage to Ser/Thr as minimal epitope, contains probably furher phosphate and fucose in PsA. Present on proteins like PsA, csA and PsB. | GA-XX | HU2428, HU2460 *modB* |
| MUD62 | Slug cell membranes of a *modB* mutant | *O*-linked glycosylation, susceptible to mild acid hydrolysis, contains GlcNAc and fucose, linkage unknown,. Present on SP protein SP96, SP75 and on some vegetative cysteine proteinase | GA-X | HU2733, HU2470, HU2471 *modC*/ *modD*/ *modE* |
| MUD166 | partly purified slug membranes of a *modB*/ *modE* double mutant | *O*-linked glycan, it is assumed to recognise a truncated form of MUD62 epitope, sensitive to mild acid hydrolysis, mAb binding can be inhibit by GlcNAc. Present on spore coat proteins of fucose mutants and on some vegetative cysteine proteinase of wild-type and mutants | GA-XIII | X22 wild-type spores No mutation |
| MUD3 | spores | *O*-linked glycan, it is assumed to contain fucose, sensitive to mild acid hydrolysis, not sensitive to periodate. Present on SP96 of X22 and HU2470 | GA-XII | HU2733, HU2471 *modC*/ *modE* |
| 83.5 | slug membrane preparation | *O*-linked glycosylation, susceptible to mild acid hydrolysis, contains GlcNAc and fucose, linkage unknown, minimal epitope Fucβ-1-P, similar to MUD62. Present on SP protein SP96, SP75 and on some vegetative cysteine proteinase | GA-X | HU2733, HU2470, HU2471 *modC*/ *modD*/ *modE* |
| AD7.5 | Purified lysosomal proteinase-1 of *D. discoideum* | recognises GlcNAc-α–1–P phosphodiester linked to Ser. Present on some spore coat proteins and on cysteine proteinases | - | - |

**Table 1-5 mAb as tools for O-linked glycan analysis**

#### 1.3.3.2.1 MUD50

Grant and Williams (48) showed that the monoclonal antibody MUD50, raised against urea extracted extracellular matrix proteins of *D. discoideum*, recognise pronase-insensitive determinants which are on slug cells and sheath proteins. Strains with a mutation in the *modB* gene lose the MUD50 epitope. It has been shown with PsA (**p**respore **s**pecific **A**ntigen) that the reducing terminal sugar fails to be attached to the protein (45). The antigenic determinants are developmentally regulated. Extensive studies on the proteins recognised by this antibody indicate that they are also found preferentially, but not exclusively, in pre–spore cells (2). The antibody defines an *O*-linked glycosylation in *D. discoideum* which is present on PsA, **c**ontact **s**ite **A** (csA), SP85 (PsB) and others (11). The MUD50 epitope is resistant to mild acid hydrolysis. In studies on recombinant PsA (PTVT repeat) and *O*-glycosylation acceptor motifs it was shown that MUD 50 recognises single GlcNAc units bound to serine or threonine (71, 155). In the cell surface molecule PsA it recognises a GlcNAc linkage to threonine (44), further it probably contains fucose and phosphate in PsA (61, 156). MUD 50 reacts with some but not all larger oligosaccharides with the same *modB* dependent reducing sugar. For example, MUD 50 reacts with the natural PsA cell surface molecule produced at the slug stage, despite of the fact that the reducing terminal GlcNAc residues can be further modified by phosphodiester linked GlcNAc, Fuc-GlcNAc or P-Fuc-GlcNAc (156).

#### 1.3.3.2.2 MUD62

Monoclonal antibody MUD62 was produced using extracellular matrix from slugs as the immunogen and it was the major antigen when slug cell membranes of a *modB* mutants (lacking the highly immunogenic MUD50 epitope) were used (10). The antibody defines a *O*-linked glycosylation which is found on wild-type (X22) spore coat proteins SP96 and SP75 (141) and on cysteine proteinases, CP4 and CP5[c] in vegetative cells. The binding on vegetative proteins is much weaker compared to that seen on spore coat proteins (12). So far fucose and GlcNAc have been identified as part of the epitope, but the details of the linkage are still unknown (11, 43). It could be shown that the binding of MUD62 can be inhibited by L-fucose and that the epitope is susceptible to mild acid hydrolysis. Strains with a mutation

[c] A. Champion, G. Harrison, M. Wilkins, M.North, A. Gooly and K. Williams (unpublished work)

in the loci *modD* (HU2470) or *modE* (HU2471) show a decreased level of fucosylation which results in loss of the MUD62 epitope on the spore coat proteins (11). On vegetative cysteine proteinases of strain HU2470 (*modD*) the epitope is still found, despite its absence on spore coat proteins.

#### 1.3.3.2.3 MUD166

Monoclonal antibody MUD166 was prepared against partly purified slug membranes of a *modB*/*modE* double mutant which lacks the immunodominant MUD50 and MUD62 reactive glycoantigens. On western blots it reacts weakly with the SP96 and SP75 of wild-type (X22) slug or spore preparations, but it reacts strongly with SP96 and SP75 of fucose mutants with mutations in *modD* (HU2470) or *modE* (HU2471). Because similar patterns of protein staining are observed with spore preparations of MUD62 in wild-type and MUD166 in fucose mutants, it is assumed that MUD166 recognises a truncated form of the MUD62 epitope. Interestingly MUD166 reacts with the vegetative cysteine proteinases in wild-type strain (X22) and in the fucose mutants HU2470 (*modD*) and HU2471 (*modE*) with a weaker but similar intensity (12). The MUD166 epitope is also similar to the MUD62 epitope in being sensitive to mild acid hydrolysis. Binding of MUD166 can be inhibited by GlcNAc; so the epitope probably consists at least partly of this carbohydrate (11).

#### 1.3.3.2.4 MUD3

Monoclonal antibody MUD3 was raised against spores and is specific for the spore coat protein SP96 (131). It is assumed that the epitope contains at least partly fucose, because, while MUD3 recognises SP96 only in the wild-type (X22) and in the fucose mutant HU2470, the fucose mutant HU2471 with a further decreased level of fucose lacks the epitope. MUD3 is not competed by fucose, so this probably means that the epitope has conformational requirements. The MUD3 epitope is sensitive to mild acid hydrolysis and β-elimination, which is consistent with the hypothesis that the epitope consist at least partly of carbohydrates. However, one strange result was that it is not sensitive to periodate teatment, which attacks cis hydroxyl groups and so would destroy unsubstituted fucose residues (12).

1.3.3.2.5 83.5

Monoclonal antibody 83.5 was raised against glycoconjugates from slug membranes fractionated by wheat germ agglutinin affinity purification (141). The antibody recognises several developmentally regulated proteins including SP96, SP75 on wild-type preparations and is able to label some vegetative cysteine proteinases. Strain HU2470 and HU2471 with a mutation in loci *modD* or *modE* respectively, lack the epitope of 83.5 which is similar but not identical to the MUD62 antibody epitope. Antibody 83.5 recognises a glycoantigen reported to be an *O*-linked fucose and GlcNAc-containing oligosaccharide (43, 109) which is similar, but not identical to the independently isolated MUD62 antibody. Antibody 83.5 is inhibited by L-fucose and the epitope is labile to mild acid (87). Recently published *in vivo* studies on artificial peptides revealed that the essential minimal epitope of antibody 83.5 is a Fucβ-1-P unit phosphodiester linked to serine (119).

1.3.3.2.6 AD7.5

Monoclonal antibody AD7.5 was made against the purified lysosomal proteinase-1, that contains multiple GlcNAc-α-1-P residues in phosphodiester linkage to serine (87). Antibody AD7.5 can be inhibited by UDP-GlcNAc, UDP-Glc, GlcNAc-α-1-P and partially by UDP-GalNAc (87). On a western blot it weakly recognises cysteine proteinases produced by amoebae grown on bacteria. It can be specifically inhibited by UDP-GlcNAc, UDP-Glc, GlcNAc-α-1-P and partially by UDP-GalNAc (87). Spore extract of wild-type (X22) shows labelling with AD7.5 of numerous proteins, but only a weak recognition of SP96. In the fucose mutants HU2470 (*modD*) and HU2471 (*modE*) the protein band pattern is similar to that found in wild-type, although the SP96 band labelled much more stronger.

#### 1.3.3.3 Expression of *O*-linked glycosylation acceptor motifs in *D. discoideum*

A recent review of Jung and Williams highlighted the potential of *D. discoideum* as a host for recombinant (glyco-) protein expression (72). Six different heterologous recombinant proteins have been successfully produced in *D. discoideum.* These include *S. japonicum* Glutathione-S-transferase (GST) (24), rotavirus VP7 (29), human IgE receptor[d], human muscarinic receptor (132), malaria circumsporozoite antigen (31) and human antithrombin III (23).

To analyse motifs for *O*-glycosylation *in vivo* the 3` end of GST gene was so modified to encode a 10-30 amino acid long peptide containing *O*-glycosylation acceptor motif. The expression system was built up of a Ddp2-based extrachromosomal plasmid containing the modified GST gene under control of an *D. discoideum* actin15 promoter (24, 71). The construct includes a *D. discoideum* signal sequence so the fusion protein is secreted. The actin15 promoter is active early in starvation phase causing transformed *D. discoideum* cells to express the GST fusion protein and the peptide acceptor motif modified by the organism´s glycosylation machinery. The secreted fusion protein can be purified using a GST affinity chromatography column and the cleaved acceptor motif used for further glycosylation analysis.

In other studies on the recombinant prespore-specific antigen (rPsA) expressed under the actin 15 promoter it was demonstrated that *D. discoideum* adds single GlcNAc residues to a threonine-rich $(PTVT)_{3-5}$ domain (157). The same glycosylation was observed on GST shorter (PTVT) repeats and proteins fused with repeats where threonine was replaced with serine (71). Interestingly there was a difference in modification between the glycosylation at different stages of development. The recombinant rPsA expressed during vegetative growth contained only single GlcNAc residues, but on the authentic PsA produced by prespore cells during the slug stage of development the reducing terminal GlcNAc is further modified with phosphate and/or GlcNAc and fucose (61, 156). Thus the elaboration of *O*-glycosylation structure is dependent on the developmental stage of the organism.

The GST expression system was further used to identify glycosylated serines and/ or threonines in human mucin-like motifs (70). *D. discoideum* attached GlcNAc to Ser/ Thr residues in mucin repeats, unlike to mammalian glycosylation which has GalNAc on the reducing terminus. Since there has been no clear evidence of GalNAc linked to serine or threonine residues in *D. discoideum* to date, it seems that in this organism GlcNAc might

[d] I.Wilson, M.B. Slade and K.L Williams unpublished work

substitute for GalNAc in this first step of *O*-glycosylation but the peptide motif remains the same.

### 1.3.4 Glycosylation studies on spore coat proteins

Most glycosylation studies on spore coat proteins of *D. discoideum* have been done without knowing the exact glycan structure (40). Detection methods like periodic acid/ Schiff staining, radiolabelling with carbohydrates and detection with monoclonal antibodies, give an impression of a wealth of glycoproteins at this developmental stage (Table 1-6) (43, 102, 137, 140, 147).

Radiolabelling studies have demonstrated that in prespore cells incorporation of Fuc is greatly enhanced (49) and that most fucoproteins are prespore cell specific (79). A problem has been that in earlier studies the same protein has been given different names in different laboratories. However, more recently the major spore coat proteins have been defined by their gene sequences. Some of the already identified spore coat glycoproteins have been targets of further, more detailed, analysis. In particular, PsA and the spore coat proteins of the PsB spore coat complex (SP96, SP85, SP75 and SP70) have received attention (10, 11, 33, 84, 121, 136, 137, 139, 141, 142).

Radioactive labeling experiments revealed that SP96 and SP75 are highly phosphorylated and also glycosylated like most of the other spore coat proteins (1, 2, 20, 22). The post- translational modifications of these proteins are important for protection as has been shown by phenotypical characterisation of mutant strains with a lack in fucosylation. Their spores were less viable as a result of a higher permeability of the spore coat.

One of the most complete studies on the structure of a *D. discoideum* glycoprotein is that done on the Prespore specific Antigen (PsA). The protein size is approximately 30 kDa, it can be labelled with monoclonal antibody MUD50 and has the *mod*B dependent *O*-linked glycosylation. PsA has no *N*-linked glycosylation site as was evident from the published DNA sequence (26). Until quite recently the structure of the *O*-linked oligosaccharides has not been fully defined, but it was concluded that these contain GlcNAc (61, 109), Fuc and phosphate (61). Latest studies have shown that a phosphate group modifies the reducing terminal GlcNAc and that these are further modified by either GlcNAc, Fuc-GlcNAc or P-Fuc-GlcNAc (156). Threonine in a proline rich repeat domain $(PTVT)_{3-5}$ is modified at the reducing terminus with GlcNAc (45). It seems that this repeat domain is a “spacer” domain

which extends the functional domain beyond the cell surface. The glycosylation is necessary to protect the protease sensitive domain from cleavage. This was demonstrated with the proteolytic enzyme papain, where PsA was cleaved from the cell surface of *modB* cells. This did not occur in wild-type cells (45). On a truncated form of PsA which was expressed during the vegetative stage it was shown that a single *N*-acetylglucosamine residue on threonine was the minimal epitope for MUD50 (155). This clearly indicates a developmental regulation of type 2 oligosaccharides. Further it was demonstrated that a substitution in the motif of threonine (PTVT) to serine (PSVS) leads to the same single GlcNAc modification, also recognised by MUD50 (71).

The assembly of the spore coat PsB protein complex is now quite well understood (1.3.2, Figure 1-4) and therefore the next step of interest was to elucidate the significance of the extensive protein modifications (136, 137). PsB (SP85) is recognised, like PsA, by MUD50 and also exhibits *mod*B glycosylation. However, little is known about structural details because most studies of the MUD50 epitope have been done on PsA (1.3.3.2.1) (117). Both SP75 and SP70 have potential *N*-glycosylation sites, but probably only SP75 is modified by *N*-linked oligosaccharides since its molecular weight is altered in a strain carrying the *mod*A mutation (140). The *modA* mutation in glycosidase II alters the size and charge of *N*-linked oligosacchrides (36, 37, 41, 42). Both proteins show phosphorylation on serine residues (22). A fucose containing glycan could be released by β-elimination from SP75 which runs as a hexasaccharide in gel chromatography and contains the epitope recognized by antibodies 83.5 and MUD62 (40). Although SP70 is not reactive to the MUD62 antibody, it is fucosylated since its molecular size is reduced in the *mod*C and *mod*E mutants which are deficient in fucose glycosylation (10, 43, 136). There is no direct evidence that SP60 is glycosylated or phosphorylated, but the calculated molecular weight of the unmodified protein without the signal peptide is only 47 kDa compared to the observed 60 kDa suggesting that post-translational modifications may exist (35, 136, 137).

| Protein name | Detection of the post-translational modification | Characteristic | Reference |
|---|---|---|---|
| SP170 | pa/Ss, ca-la | Man, GlcNH2, Fuc | (147) |
| SP103 | pa/Ss, ca-la | Man, GlcNH2, Fuc | (147) |
| SP96/ p112 | ca-la, mAb, [$^{32}$P]$P_i$ | Fuc, GlcNAc, -$PO_4$ | (10, 11, 22, 30, 131) |
| SP90 | pa/Ss, ca-la ? | | (147) |
| SP85/ PsB | mAb | Fuc?, GlcNAc? | (136, 140) |
| SP82 | pa/Ss, ca-la | Man, Glc$NH_2$, Fuc | (147) |
| SP80 | mAb | Fuc | (141) |
| SP76 | pa/Ss, ca-la | Man, Glc$NH_2$, Fuc | (147) |
| SP75 | ca-la, mAb, [$^{32}$P]$P_i$ | Fuc, GlcNAc, -$PO_4$ | (10, 11, 84, 136, 140) |
| SP70/ p78 | ca-la, mAb, [$^{32}$P]$P_i$ | Fuc, -$PO_4$ | (10, 22, 43, 136) |
| SP68 | | | (102) |
| SP60/ p58 | | | (35, 136, 137) |
| SP33 | pa/Ss | | (147) |
| SP29/ PsA | mAb | Fuc, GlcNAc, -$PO_4$ | (61, 142) |

**Table 1-6 Glycoproteins identified as spore coat proteins**
pa/ Ss: periodic acid/ Schiff staining mAb: monoclonal antibody
ca-la: carbohydrate radio-labelling [$^{32}$P]$P_i$: radio-labelling with phosphate

### 1.3.5 Spore coat protein SP96

SP96 is composed of 580 amino acids (**Figure 1-5**). After cleavage of the 20 AA long signal sequence, it is one of the largest spore coat proteins of the PsB protein complex. The calculated mass of the protein component is 57,510 Dalton, but the protein is further post-translationally modified and runs on SDS PAGE at an apparent molecular weight of 96 kDa. Although the amino acid sequence reveals one possible *N*-linked glycosylation site (AA 99; NSS) no *N*-linked glycan could be detected (A. Champion, pers. communication). Possible sites for *O*-linked glycosylation in the protein are 121 serine and 83 threonine residues. More than half (74 Ser and 42 Thr) form a C-terminal serine and threonine rich domain.

```
  1  MRVLLVLVAC  LTYFSGGALA  QSCSSYSGDN  CPSTCFQGSY  DIPCGAQVKY  CTEMKDNCGE
 61  GGDVKCWKDG  SNLPVQTWSS  CVPSELFGPN  GKFKPSEIPN  SSNCPTNCEN  GVEWVNLCGL
121 SCDAKTACCP  DVCQCKGGQT  SGGSTTGSQT  SGGSTSGGST  TGSQTSGGST  TGSQTSGSQT
181 SAGSCSNTQC  PNGFYCQVQG  NNAVCVPQQS  STSGGHQNDP  CDTVQCPYGY  SCESRDGFEA
241 KCTRDEDEPT  HRPTHRPKPP  HDSDKYLCDN  VHCPRGYKCN  AKNGVAKCIA  GYEIPRVCRN
301 IQCPTGYRCE  DHNRNPICVL  EERENPDNCL  TCNDVNCEAS  GLVCVMTRAR  CKVGAAKCCD
361 VQPTCIKPST  IAGSTIASIA  STIASTGSTG  ATSPCSVAQC  PTGYVCVAQN  NVAVSLPRPT
421 TTTGSTSDSS  ALGSTSESSA  SGSSAVSSSA  SGSSAASSSP  SSSAASSSPS  SSAASSSPSS
481 SAASSSPSSS  ASSSSSPSSS  ASSSSAPSSS  ASSSSAPSSS  ASSSSASSSS  ASSAATTAAT
541 TIATTAATTT  ATTTATTATT  TATTTATTTA  ATIATTTAAT  TTATTTATTA  TTTATTTATS
```

**Figure 1-5** **Amino acid sequence of SP96** (34)
underlined sequence: 20 AA long N-terminal signal sequence
**NSS** motif for *N*-linked glycosylation

Radiolabelling experiments demonstrate that SP96 is heavily phosphorylated on serine (2, 21) and that it has at least 50 times more phosphate incorporated than PsA (61, 136, 137). Apart from the possibility of a monoester phosphate it was speculated that phosphate could also occur in a phosphodiester linkage.

Fucose and *N*-acetyl-glucosamine were also identified on the protein by radio-labelling experiments (109). However, the exact *O*-linked glycan structure is still unknown. After the release of glycans by β-elimination, some fucose could be identified as fucitol as a result of being directly attached to the protein backbone (109). It was assumed that SP96 possesses multiple *O*-linked fucosylated glycans (109). Further evidence for the presence of fucose in SP96 is deduced from the creation of mutant strains with a decreased level in fucosylation (10, 11). The SP96 protein of the different strains looses fucose-dependent epitopes of monoclonal antibodies MUD62, mAb 83.5 and MUD3. A new GlcNAc containing epitope recognised by MUD166 and AD7.5, respectively, is revealed in these mutant strains (**Table 1-7**).

| strains | *mod* mutation | Fucosylation level* [%] | $M_r$ of SP96 [kDa] | mAb | | |
|---|---|---|---|---|---|---|
| | | | | MUD62/ 83.5 | MUD3 | MUD166/ AD7.5 |
| X22 | + | 100 | 96 | + | + | - |
| HU2470 | *mod*D | 17 | 93 | - | + | + |
| HU2471 | *mod*E | 5 | 90 | - | - | + |
| HU2733 | *mod*C | 0 | 90 | - | - | + |

**Table 1-7 Antibody definition of fucosylation mutants**

* fucosylation level in spores in % compare to wild-type (X22)

Apparent molecular mass ($M_r$) of SP96 was determined by recognition with the different mAb.

Strength of labelling: +, present at similar level

-, absent or almost absent

## 1.4 Aim of this thesis

The aim of this thesis is the characterisation and the quantification of the different post-translational modifications on the *D. discoideum* spore coat protein SP96. To this end, SP96 of the wild-type (X22) and mutant strains HU2470 and HU2471 with a decreased level of fucosylation were analysed and compared. The results allow elucidation of the differences in glycosylation related to the lost MUD62, mAb 83.5 and MUD3 antibody epitopes. Further work focuses on the localisation of the glycosylation sites in SP96 with its different serine and threonine rich domains. The work required either purified SP96 protein or expression of potential SP96 glycosylation motifs in the *D. discoideum* GST-fusion system. Combining the results found in this thesis with the already published phenotypic characteristics of the different mutant strains allow some speculation about possible glycan structure/ function relationships.

# 2 Material and Methods

## 2.1 Enzymes, antibodies and chemicals

Enzymes:

| | |
|---|---|
| Alkaline phosphatase | Boehringer Mannheim |
| Pretaq™ | Gibco BRL |
| Proteinase K | Boehringer Mannheim |
| Accutaq-LA DNA Polymerase | Sigma |
| BamH1 | Boehringer Mannheim |
| T4 DNA ligase | Promega |

| Antibodies: | Reference |
|---|---|
| MUD3, monoclonal antibody | (131) |
| MUD62, monoclonal antibody | (10) |
| MUD166, monoclonal antibody | (11) |
| 83.5, monoclonal antibody | (141) |
| AD 7.5, monoclonal antibody | (87) |
| Anti *S. japonicum* GST, polyclonal antibody | Sigma |
| I-SPY™ 11, monoclonal antibody, HRP conjugated | Amrad Biotech |
| Sheep anti mouse IgG, HRP conjugated | Silenus Laboratories |
| Sheep anti rabbit IgG, HRP conjugated | Silenus Laboratories |

Chemicals:

All chemicals were of analytical reagent grade and were from Merck unless otherwise indicated. All solutions were prepared with MilliQ-water.

## 2.2 Strains and Vectors

### 2.2.1 *D. discoideum* strains

For the experiments were following *D. disoideum* strains used:

| | |
|---|---|
| X22 | wild-type (148) |
| HU2470 | fucose mutant (10) |
| HU2471 | fucose mutant (10) |
| HU2868 | strain NP2 maintained on *Micrococcus luteus* (146) |
| HU2942 | (this thesis) |
| HU2943 | (this thesis) |
| HU2944 | (this thesis) |
| HU2945 | (this thesis) |

### 2.2.2 *E. coli* strains

*E.coli* cells used for transformation and plasmid preparation

| | |
|---|---|
| *E. coli* DH5α | (56) |

### 2.2.3 Bacteria used as food source for *D. discoideum*

| | |
|---|---|
| *Klebsiella aerogenes* | (148) |
| *Mircococcus luteus* PRF3 | (146) |
| *E. coli* B/r | |

### 2.2.4 Vectors

| | |
|---|---|
| pMUW2442 | (containing GST gene Nsi/KpNI; M. Slade)[e] |
| pMUW2882 | (modified pMUW1632 by W. Dittrich)[1] |
| pMUW110 | (13) |
| pMUW2911 | (this thesis) |
| pMUW2921 | (this thesis) |
| pMUW2931 | (this thesis) |
| pMUW5021 | (this thesis) |

## 2.3 Media, buffers and cell culture

### 2.3.1 Media and buffer

Salt solution

| Compound | Concentration |
|---|---|
| NaCl | 10 mM |
| KCl | 10 mM |
| $CaCl_2$ | 3 mM |

The salts were dissolved in 1 litre $dH_2O$ and autoclaved in aliquots.

HEPES buffer (pH 7.5)

| Compound | Concentration |
|---|---|
| NaCl | 10 mM |
| KCl | 10 mM |
| HEPES (Sigma) | 5 mM |
| $CaCl_2$ | 0.1 mM |

[e] M. Slade (pers. communication)

SM agar (124)

| Compound | Amount/ litre | |
|---|---|---|
| | SM | SM/5 |
| Agar (Calbiochem) | 12.0 g | 12.0 g |
| Special peptone (Oxoid, code L72) | 10.0 g | 2.0 g |
| Yeast extract (Oxoid, code L21) | 10.0 g | 2.0 g |
| Glucose (BDH) | 10.0 g | 2.0 g |
| $MgSO_4 \cdot 7H_20$ | 1.0 g | 1.0 g |
| $K_2HPO_4$ | 1.0 g | 1.0 g |
| $KH_2PO_4$ | 2.2 g | 2.2 g |

Autoclaved agar was cooled to 50°C and poured into (9 cm) sterile petri dishes using 25 ml per plate. Plates were stored in sealed bags at 4°C.

Geneticin (G418) agar

As for SM/5 agar plates, plus 10 µg/ml of filter sterilised G418 (Amresco) added immediately before pouring. Plates stored in sealed bags at 4 °C for up to 10 days.

Water agar

| Compound | Amount/ litre |
|---|---|
| Agar (Calbiochem) | 11.0 g |
| Dihydrostreptomycin sulphate | 0.25 g |

Autoclaved agar was cooled to 50°C and poured into 9 cm sterile petri dishes using 25 ml per plate. Plates were stored in sealed bags at 4°C.

Axenic medium (138)

| Compound | Amount/ litre |
|---|---|
| Glucose (BDH) or maltose | 15.4 g |
| Special peptone (Oxoid, code L72) | 14.3 g |
| Yeast extract (Oxoid, code L21) | 7.15 g |
| Dihydrostreptomycin sulphate | 0.25 g |
| $KH_2PO_4$ | 0.48 g |
| $Na_2HPO_4$ x $2H_20$ | 0.64 g |

Axenic medium was autoclaved and stored in the dark. Before use a sterile stock vitamin solution (5 mg/ml vitamin B12 and 2 mg/ml folic acid) was added to 1:1000.

L-Broth

| Compound | Amount/ litre |
|---|---|
| Tryptone (Oxoid) | 10 g |
| Yeast extract (Oxoid, code L21) | 5 g |
| NaCl | 10 g |

Adjusted to pH 7.0 with NaOH, aliquot and autoclaved.

L-Agar

As for L-Broth, plus 12 g of agar per litre. L-agar was cooled down to 50 °C before pouring into plates (25 ml per plate). Plates were stored in sealed bags at 4 °C.

**Ampicillin agar**

As for L-agar plates, plus 100 µg/ml of filter sterilised ampicillin (Sigma) added immediately before pouring. Plates stored in sealed bags at 4 °C for up to 10 days.

### 2.3.2 Growth of *D. discoideum* cultures on SM plates

Cultures were grown at 21±1°C on lawns of *K. aerogenes* on SM agar plates (9 cm). For weekly maintenance, *D. discoideum* fruiting bodies were collected from single colonies on a sterile loop, suspended in 1.5 ml of sterile salt solution and streaked onto fresh plates of SM agar prespread with bacteria. The spores germinated, and the amoebae fed on the bacteria by phagocytosis creating plaques in the bacterial lawn. Cells behind the edge of the plaque were starved of their food supply and so entered the developmental pathway. Cultures grown on bacterial lawns finish the developmental process and form fruiting bodies after 5-7 days.

To get large quantities of spores, cells were also grown on *K. aerogenes* in 40 x 40 x 2 cm autoclaved steel trails containing 1 litre SM agar.

### 2.3.3 Growth of *D. discoideum* cultures on SM/5 plates

Transformed *D. discoideum* cultures were routinely maintained under G418 selection. Cultures were grown at 21±1 °C on lawns of *M. luteus* on SM/5 agar containing 10 µg G418/ml. For maintenance every 2 weeks fruiting bodies were collected from single colonies, suspended in 1.5 ml of sterile salt solution and restreaked onto fresh plates of SM/5 agar prespread with bacteria.

### 2.3.4 Development of *D. discoideum* slugs on water agar

Spores of *Micrococcus* grown *D. discoideum* transformants were plated out on SM agar plates with *K. aerogenes* as a food source. After approximately 3-4 days the clearance of the bacteria lawn could be observed and the cells were harvested in HEPES buffer; they

were shaken in this buffer until nearly all bacteria were eaten. Afterwards the cells were harvested at 4 °C by centrifugation in a Beckman J2-21 centrifuge (rotor JA10, 3500 rpm, 2150 g, 10 min) and resuspended in a small volume of HEPES buffer before the thick *D. discoideum* suspension was plated out on a water agar plate. The cells started to aggregate within a few hours, so that after 1 day slugs were collected.

### 2.3.5 Harvest and storage of *D. discoideum* spores

For long term storage, spores from five small plates were combined by tapping onto the same petri dish lid. The spores were released and stuck to the inner surface of the lid. These were then suspended in 0.25 ml of sterile horse serum (CSL) using a glass spreader and poured into a prechilled (-20°C) glass vial containing dry, sterile silica crystals. The crystals were evenly coated with the spore suspension by repeatedly hitting the vial sharply in the palm of the hand and turning (the pre-chilling was necessary to prevent overheating the spores at this stage). The vial was then sealed, labelled and placed in a desiccated storage box at 4°C. Spores stored in this way were viable for several years.

### 2.3.6 Growth of *D. discoideum* in liquid culture and harvesting of cells

For axenic culture, spores were inoculated into axenic medium in Erlenmeyer flasks, containing between 10-30 % of their nominal volume with medium, at a density of 2 x $10^5$ spores/ml. Axenic grown cultures had a generation time of 10-12 hours. For transformants selection was maintained with Geneticin (G418; 10µg G418/ml). The cultures were maintained at 21±1 °C on a rotating shaker at 150 rpm and subcultured into fresh axenic medium before cells reached densities of 5x$10^6$ cells/ml.

To grow *D. discoideum* in bacterial suspension, *K. aerogenes* or *E. coli* bacteria grown on SM plates were resuspended in HEPES buffer and density adjusted to ($OD_{460}$ 10). Bacterial suspensions were inoculated with 2 x $10^5$ *D. discoideum* spores per ml. The generation time of bacteria grown cells ranged between 3-4 hours. The *D. discoideum* cells were harvested by centrifugation in a Beckman J2-21 centrifuge (rotor JA10, 3500 rpm, 2150 g, 10 min) at 4 °C.

Spores from small plates and large steel trails were harvested by inverting and tapping the plate firmly on the surface of a bench, taking care not to dislodge the agar before the spores were suspended in buffer used in the first step of extraction (2.4.2.1).

## 2.4 Protein chemistry methods

### 2.4.1 Buffers and solutions

Phosphate buffered saline (PBS) 5x stock (pH 7.2)

| Compound | Concentration |
|---|---|
| $Na_2HPO_4$ | 100 mM |
| $KH_2PO_4$ | 7.5 mM |
| KCl | 130 mM |
| NaCl | 770 mM |

Polyacrylamide gel electrophoresis (PAGE) buffers

| *Lower gel buffer 4 x (pH 8.8; conc. HCl)* | *Upper gel buffer 4 x (pH 6.8; conc. HCl)* |
|---|---|
| Tris 1.5 M | Tris 0.5 M |
| SDS 14.0 mM | SDS 14.0 mM |

Recipe for pouring PAGE gels

| Compound | Amount | | Compound | Amount |
|---|---|---|---|---|
| Resolving gel | 10 % | 12 % | Stacking gel | 4% |
| $dH_2O$ | 4.84 ml | 4.34 ml | $dH_2O$ | 3.2 ml |
| Lower gel buffer | 2.6 ml | 2.6 ml | Upper gel buffer | 1.3 ml |
| Liquigel (*Gradipore) | 2.5 ml | 3.0 ml | Liquigel (*Gradipore) | 0.5 ml |
| 20 % (w/v) APS | 50 µl | 50 µl | 20 % (w/v) APS | 25 µl |
| Temed (BioRad) | 15 µl | 15 µl | Temed (BioRad) | 7.5 µl |

*From a 40% acrylamide stock, acrylamide:bis ratio of 29:1 (Gradipore).

Electrophoresis SDS running buffer, 5 x stock (pH 8.5)

| Compound | Concentration |
|---|---|
| Tris | 125 mM |
| Glycine | 1.0 M |
| SDS | 0.5 % (w/v) |

Reducing sample buffers for SDS-PAGE, 3 x stock

| Compound | Concentration |
|---|---|
| Tris/HCl (pH 6.8) | 65 mM |
| Glycerol | 10 % (v/v) |
| SDS | 2.3 % (w/v) |
| 2-Mercaptoethanol (Sigma) | 5 % (v/v) |
| Bromophenol Blue | 0.25% (w/v) |

## 2.4.2 Purification of spore coat protein SP96

### 2.4.2.1 Extraction of SP96 from the spore coat

All extraction steps were done twice to get as little contamination as possible between the different fractions.

5-10 x $10^{10}$ spores were harvested into 40 ml of extraction solution 1, shaken for 10 min/ room temperature and then centrifuged at 3000 g. The pellet of spores was resuspended in 40 ml extraction solution 2 and the procedure repeated. A third extraction step also used extraction solution 2 (15 ml) with the temperature increased to 100 °C, by incubating in a boiling water bath for 10 min. A fourth extraction step was done at 100 °C under reducing conditions with extraction solution 3 (15 ml). The third extract (Extraction solution 2; 100 °C) contained most of the SP96 protein and was used for further purification.

**Extraction solution 1**

50 mM Tris/HCl pH 8

**Extraction solution 2**

50 mM Tris/HCl pH 8
8 M Urea

**Extraction solution 3**

50 mM Tris/HCl pH 8
8 M Urea
0.7 M 2-Mercaptoethanol

The solution used to extract the spores were stored at room temperature.

### 2.4.2.2 Column chromatography steps

The SP96 positive extract from step 3 was diluted 1:1 with $dH_2O$ and applied to a DEAE cellulose column (DE52, Whatman, 1.4 cm x 10 cm, volume: ~15 ml) with a flow rate of 0.5 ml/ min. After intensive washing (approximately 5 x column volume) with 25 mM Tris/ 4 M urea (pH8) the retained proteins were eluted by steps of 100 mM, 250 mM, 500 mM and 1 M NaCl. SP96 was eluted with 250 mM NaCl and extensively dialysed against coupling buffer (0.1 M $NaHCO_3$/ 0.5 M NaCl (pH 8.3)), incubated for 1 h with Sepharose 4B

to minimize nonspecific binding and the eluate then applied to a MUD3 affinity column (2 ml column size). Monoclonal antibody MUD3 was used to purify SP96 from the wild-type X22 and mutant HU2470 as both strains carry the MUD3-epitope. For the preparation of these columns MUD3 antibody was purified from ascites on an EconoPac Protein A cartridge (Bio-Rad) and then coupled to CNBr activated Sepharose (Pharmacia). Both steps were carried out essentially as described in the protocols of the companies. SP96 fractions eluted from the DEAE column were applied to the MUD3 affinity column with 2 h incubation at 20 ºC and the bound SP96 was eluted with 100 mM diethylamine (pH 11), after which the pH was immediately lowered with 50 µl 1 M Tris (pH 8) solution. The SP96 positive fractions were repurified on the same MUD3 column. It was not necessary to remove the neutralised diethylamine for the repurification step because the SP96 was eluted on the basis of high pH. After the pH was decreased to pH 8 with 1 M Tris the SP96 protein could be bound immediately again on the column. The positive fractions were pooled and concentrated.

Approximately 12-16 nmols of SP96 was extracted from 5 x $10^{10}$ spores of each strain of *D. discoideum*.

### 2.4.3 Purification of recombinant GST fusion proteins

Glutathione-S-transferases (GST) possess selective and reversible binding for reduced glutathione and can be efficiently affinity purified on immobilised glutathione.

The culture supernatant of axenic grown stationary phase *D. discoideum* cells was dialysed extensively against PBS buffer at 4°C to remove free glutathione. For affinity purification of GST-fusion proteins, 500 ml culture supernatant was adjusted to reducing conditions of 2 mM DTT and incubated with 1.5 ml equilibrated glutathione-Agarose (Sigma) overnight at 4 °C. The affinity matrix was slowly packed (flow rate 2-3 ml/ min) into a column (∅ 1 cm) and was washed twice with 50 ml PBSD buffer ( 1 x PBS containing 2 mM DTT (Sigma), pH 7.2). The retained GST binding proteins were eluted in 5 x 1 ml fractions of TE buffer (1 mM EDTA, 10 mM Tris/HCl; pH 8.0) containing 20 mM reduced glutathione (Sigma) and 2 mM DTT.

### 2.4.4 Peptide purification by reversed phase chromatography

The glyco-peptides, which were released from the purified GST-fusion proteins by factor Xa cleavage (2.4.8.3), were separated based on their hydrophobicity on a reverse phase Sephasil™ C8 column using a SMART System (Pharmacia, LKBμ).

The samples (~ 0.5 ml) were applied by multi injections of 100 µl aliquots in intervals of 8 min to a column equilibrated with solution A (0.5 % TFA in $dH_2O$) for 10 min. The used flow rate was 100 µl/ min. 5 min after the last injection a dipolar gradient 0-30 % solvent B (0.45 % TFA in 85 % acetonitrile) was established over a period of 30 min followed by a further linear gradient to 100 % over 30 min. The eluted sample was collected in 200 µl aliquots.

### 2.4.5 SDS polyacrylamide gel electrophoresis (SDS-PAGE)

The discontinuous, one dimensional gel electrophoresis was carried out with the Laemmli buffer system (78) in a Mini-Protean II apparatus (BioRad). 10 % or 12 % (w/v) acrylamide resolving gels (0.5 mm thick) with a 3 % (w/v) stacking gel were mainly used. For protein separation on gradient gels BioRad Ready Gels (4-20 %) were choosen. Samples were suspended in 1 x sample buffer, heated to 95 ºC for 3 min and centrifuged before loading on a gel. Protein bands were visualised with Coomassie Blue or the more sensitive silver nitrate stain methods.

### 2.4.6 Western blotting and staining of proteins on membrane

After SDS-PAGE the proteins were transfered onto nitrocellulose (0.1 µ Schleicher and Schuell) or PVDF membrane (Millipore) using a semi-dry electroblotting system (76) (12V for 1 h) with a discontinuous buffer system.

To visualise proteins on the membrane, a standard staining method with Amido black 10B (Sigma) was used (0.1 g Amido black 10B (Sigma) in 10 ml (methanol:acetic acid :water; 3:1:6 v/v)). The stained bands were then used for amino acid anaylsis, Edman sequencing or glycan analysis.

Glycoproteins were detected on membranes using the Bio-Rad Immun-Blot Kit as described by the manufacturer. This stain is based on immunochemical detection of biotin which is bound to the aldehyde groups of periodate cleaved sugars.

For immunochemical labelling of proteins the membrane was rinsed for 10 min in PBS followed by a 10 min wash with 3 % (w/v) skimmed milk powder (Diploma) in PBS to block non-specific binding. The incubation with the first antibody was done in the same solution used to block the membrane, with the SP96 recognising antibodies for 1 h (hybridoma supernatant) and with the anti GST antibody for 3 h (1:10000 dilution in PBS), on a rocking platform. After rinsing for 3 times for 5 minutes with PBS it was incubated with the sheep anti-mouse antibody or sheep anti-rabbit antibody conjugated to horse radish peroxidase (Silenus) diluted to 1:1500 in 3 % (w/v) skimmed milk powder in PBS for 2 hours or overnight. After washing 3 times for 5 min with PBS to remove the unbound components, the blot was developed using a fresh solution of 4CN substrate.

Developing substrate for horse radish peroxidase

| Compound | volume |
|---|---|
| 4-chloro-1-naphthol (3 mg/ml stock in methanol) | 2 ml |
| PBS | 8 ml |
| $H_2O_2$ (30%, BDH) | 20 µl |

This solution was made fresh for each reaction.

### 2.4.7 Protein quantification

The amount of protein in samples was determined by using the Bio-Rad $D_c$ protein assay in the range of 0.2 mg/ ml up to 1.6 mg/ ml . The calibration standard was bovine serum albumin (BSA). Lower levels of protein in liquid samples or on PVDF membranes were quantified by using amino acid analysis (151).

## 2.4.8 Protein digest

### 2.4.8.1 Pretaq™ (thermostable protease)

SP96 protein (~700 pmol) was dissolved in 120 µl $dH_2O$ and incubated at 80 ºC before adding 90 mU of Pretaq™ enzyme. To determine the optimal digest conditions an aliquot was taken out every 10 min (0-60 min) and frozen immediately to stop the reaction. The samples were separated on SDS-PAGE without boiling which activates the protease. Aliquots of Pretaq™ enzyme (3 Units) were stored at –20 ºC.

### 2.4.8.2 Proteinase K

SP96 protein (~700 pmol) was dissolved in 50 mM Tris/ HCl (pH8) and Proteinase K (90 mU) added and incubated at room-temperature. Samples were taken at time intervals (0-60 min) and the reaction immediately stopped by boiling with SDS-PAGE sample buffer.

### 2.4.8.3 Factor Xa

The factor Xa cleavage was used to release the potential *O*-glycosylation acceptor peptides from the expressed and purified GST fusion protein. The protease recognised specifically the amino acid sequence IEGR and cleaved at the C-terminus of the arginine in this sequence. The digest was done at 30 ºC in elution buffer containing 1 mM $CaCl_2$. The enzyme:substrate ratio was approximately 1:10. The cleavage process was monitored by SDS-PAGE.

## 2.5 Characterisation of post-translational modifications

### 2.5.1 Analysis of phosphorylation

#### 2.5.1.1 Phosphoamino acid analysis

Affinity purified SP96 of strains X22 and HU2470 and HU2471 (20-60 pmol) were gas phase hydrolysed with 6 M HCl for 4h at 110ºC, derivatised with Fmoc as a fluorogenic reagent and analysed for the presence of phosphoamino acids by HPLC (150).

#### 2.5.1.2 Total phosphate analysis after acid hydrolysis

The degree of phosphorylation of SP96 was determined by measuring the amount of inorganic phosphate released by strong acid hydrolysis using the method of Lanzetta *et al.* (80), which was modified for microplate analysis. The method is based on a colour reaction involving formation of phosphomolybdate complexes.

20µl protein samples (20-100 pmol) of SP96 were hydrolysed with 4 µl 10 M $H_2SO_4$ at 170ºC in a sealed tube for 4 h. The samples were transferred into a 96 well microplate, 100 µl of phosphate reagent was added and carefully mixed. After 1 min the reaction was stopped by adding 10 µl of a 25 % (w/v) trisodium citrate solution and after 20 min incubation the absorbance was measured at 630 nm. A potassium dihydrogenphosphate standard curve (0 - 4.0 nmol) was used to calculate the concentration of inorganic phosphate in the samples. To determine the background phosphate in the samples, non hydrolysed SP96 protein samples were also assayed.

#### 2.5.1.3 Enzymatic release of phosphate with alkaline phosphatase

10-20 µl SP96 sample (100-200 pmol) in 30 µl of 100 mM ammonium bicarbonate (pH8.3) and 10 µl (15 mU) alkaline phosphatase were incubated for 16 h at 37ºC. The reaction was stopped with 10 µl 4 M HCl and the released phosphate was detected as in 2.5.1.2.

#### 2.5.1.4 Dephosphorylation with HF

After reductive β-elimination and desalting (2.5.2.3), samples were lyophilized and mixed at 4°C with pre-cooled 40% (v/v) HF (50µl) for 48 h. The HF was neutralised with an equimolar concentration of LiOH, the LiF pelleted and washed. The supernatants were pooled and used for further analysis.

### 2.5.2 Analysis of glycosylation

#### 2.5.2.1 Monosaccharide analysis

To release the monosaccharides, 25 pmol of affinity purified SP96 was hydrolysed with 2 M TFA for 4h at 100ºC. To quantitate the amino sugars, a stronger hydrolysis with 4 M HCl for 4h at 100ºC was carried out. The hydrolysates were analysed on a HPAEC-PAD system (Dionex), by isocratic elution from a CarboPac PA10 column at 1 ml/min with 12 mM NaOH for 20 min. 2-deoxyglucose (1 nmol) was added as internal standard for quantitation of the sugars.

#### 2.5.2.2 Oligosaccharide/ charged glycan analysis

The glycans on SP96 were released by mild acid hydrolysis or non-reductive β-elimination. β-elimination was carried out by dissolving SP96 (1nmol) in 50 µl 100mM NaOH and incubating at room temperature for 4 h. The alkaline sample was directly separated on a CarboPac PA1 column (Dionex) using three sequential linear gradients of 1M sodium acetate in 100 mM NaOH for 1 min, raised to 15% (v/v) over 20 min, to 50% (v/v) over 10 min and to 70 % (v/v) over 10 min. With this method glycans P1 and P2 were isolated from the wild-type X22 and the fucose mutants HU2470 and HU2471, respectively.
Mild acid hydrolysis was achieved by treating the purified SP96 protein sample (0.5 nmol) with 40 mM TFA for 8 min at 100ºC. After drying, the sample was dissolved in water and the released sugars separated as above.

#### 2.5.2.3 GC-MS Analysis

To determine the reducing-terminal linkages of the glycans on SP96, samples were subjected to reductive β-elimination, acetylated and analysed by gas chromatography/mass spectrometry (GC-MS). SP96, from all strains (0.2-0.5 nmol), was dissolved in 400µl 100 mM NaOH, 1 M sodium borohydride and incubated at 45 ºC for 18h. The samples were carefully neutralised with 4 M acetic acid and desalted by anion-exchange chromatography (Bio-Rad AG 50W-X8 200-400 mesh). Dried samples were acetylated with excess anhydrous acetic anhydride (50µl) and anhydrous pyridine (50µl) at 100ºC for 1h. The acetylated alditols were extracted into dichloromethane and analysed on a Fisons MD800 GC-MS using a BPX5 (0.22mm x 25m) column. Samples (1µl) were analysed by injection onto the column at 150ºC, the temperature was held for 5 min, then increased to 300ºC over 30 min and held for 10 min.

#### 2.5.2.4 Electrospray-ionisation mass spectrometry

Glycans released by ß-elimination with 100 mM NaOH at 45 ºC for 4 h (~300 pmol) were desalted through a Hypercarb® (Shandon Scientific, UK) guard column (103). After washing with MilliQ water to remove the salt, the carbohydrates were eluted from the column into the mass spectrometer using a 15 min linear gradient of 0.05% (v/v) TFA in water to 90% (v/v) acetonitrile/0.05% (v/v) TFA at a flow rate of 100 µl/min. Spectra were collected in negative ion mode on an ESI-TOF (LCT; Micromass, UK.). Samples were introduced into the electrospray source through a capillary held at a potential of 3.5 kV. The sample cone voltage was 30V, or 80V for in-source collision induced dissociation (CID). Full scan spectra were acquired over a mass range of 70-1200 Da with 0.7 sec scan duration and 0.7 sec interscan delay. All spectra were externally calibrated with NaI.

#### 2.5.2.5 Nuclear magnetic resonance spectroscopy

The HPLC purified X22-SP96 disaccharide P1 (~100 µg) (2.5.2.2) was dissolved in $D_2O$ (0.5 ml; 99.96 atom%, Aldrich), repeatedly lyophilized from $D_2O$ and filtered into an

NMR tube (PP527, Wilmald). The sample was degassed and equilibrated under an atmosphere of nitrogen.

The NMR data were acquired on a Bruker DRX600 (600 MHz) NMR spectrometer at 27°C and processed using xwinNMR (version 2.6; Bruker). All 2D NMR experiments were run with quadrature detection with an $^{1}H$ spectral width of 6009 Hz and a recycle delay of 2 s. Chemical shifts were referenced to the fucose methyl ($\delta_H$ 1.166 ppm; $\delta_C$ 16.2 ppm). High power $^{1}H$ $\pi/2$ pulses were determined to be 9.5 μs and low power (for DIPSI spin lock) at 25.15 μs. DIPSI sequences were flanked with trim pulses at the same low power of 2 ms duration. $^{13}C$ high power $\pi/2$ pulse was 10.5 μs and a low power pulse of 65 μs was used for GARP decoupling. Gradient pulses were delivered along the z-axis using a 100 step sine program.

Data for 1D experiments were acquired using a WATERGATE 3-9-19 pulse sequence (106) enhanced with gradient selection (116). All pulses were delivered at the same power of 0 decibels and gradients were at the ratio 50:50 of full power. Two dummy scans were used before the acquisition of 128 scans, using a recycle delay of 2 s. 32,000 real points were acquired and zero filled to 64,000 and then Gaussian multiplied for resolution enhancement. Carbon - hydrogen correlation (HSQC) was achieved via a sensitivity enhanced double INEPT transfer using echo/antiecho-TPPI gradient (80:20.1) selection (74, 104, 112). 2,000 data points were collected in t2 (128 scans per increment) with a 1.3 s recycle delay with decoupling during acquisition. In t1, 512 increments were used (10-120 ppm) and the INEPT sequence was optimized for a X-H coupling of 145 Hz. A gradient ratio of 80:20.1 was used to select echo/antiecho-TPPI phase sensitivity.

One dimensional ROSEY spectra were measured using a selective Gaussian pulse on the anomeric protons. A 1000 step Gaussian program (60 ms, 64.6 dB) was used to achieve a $\pi/2$ pulse. A mixing time of 250 ms (13 dB) was used for a continuous wave spin lock. Gradient selection was achieved with a 15% gradient along the z-axis and 10,000 transients were accumulated over 6009 Hz. ROE enhancements were measured as a percentage of the irradiated peak and not compensated for offset from the carrier frequency.

Two dimensional homonuclear Hartman-Hahn transfer spectra (TOCSY) were measured using the MLEV17 (4) pulse sequence flanked with 2 ms low power trim pulses. Water suppression was achieved using a gradient assisted WATERGATE (106, 116) pulse sequence and a gradient ratio of 50:50 was used in the WATERGATE sequence. A mixing time of 30 ms was used for the Hartman-Hahn transfer. Sine bell shifted (90°) apodization

was used in the processing of both dimensions. Homonuclear shift correlation (COSY) with double quantum filter phase sensitivity using States-TPPI method was supplemented with a gradient-assisted WATERGATE sequence (106, 116) to achieve water suppression. A gradient ratio of 30:30 was used in the WATERGATE sequence.

#### 2.5.2.6 Cleavage of the X22-SP96 glycan P1 with periodate

Approximately 3 nmol of purified glycan was treated in the dark with 10 µl 20 mM sodium-periodate for 40 min at room temperature. Desalting was done over a 50 µl carbon column (Carbograph, Alltech) prewashed with 80 % (v/v) acetonitrile/ 0.05 % (v/v) TFA and equilibrated with water before the sample was applied, washed with water and eluted with 100 µl 25 % (v/v) acetonitrile/ 0.05 % (v/v) TFA. After drying the sample was hydrolysed with 2 M TFA/ 100 °C/ 4 h and separated on HPAEC-PAD.

## 2.6 Molecular biological methods

### 2.6.1 Buffers and solutions

5 x TBE buffer (pH 8.3)

| Compound | Concentration |
|---|---|
| Tris | 450 mM |
| Boric acid | 450 mM |
| EDTA | 10 mM |

The buffer was used at a 1x concentration after dilution with 4 vol. of $dH_2O$.

10 x TE stock (pH 8.0)

| Compound | Concentration |
|---|---|
| Tris | 100 mM |
| $Na_2EDTA$ | 10 mM |

Adjusted to pH 8.0 with conc. HCl, autoclaved and used at a 1x concentration.

DNA Gel Loading Buffer

| Compound | Amount |
|---|---|
| glycerol | 30 % (w/v) |
| Orange G (Sigma) | 0.1 % (w/v) |
| Xylene Cyanol | 0.25 % (w/v) |

in TBE buffer (pH 8.3). Aliquots stored at -20°C.

SOC medium

| Compound | Amount (g/l) |
|---|---|
| Tryptone (Oxoid) | 20.0 g |
| Yeast extract (Oxoid, code L21) | 5.0 g |
| NaCl | 0.58 g |
| KCl | 0.18 g |

After autoclaving, glucose (filter sterilised) was added to 20 mM . 1 ml aliquots were stored at –20 °C .

PIPES/ HL5 medium

| Compound | Amount (g/l) |
|---|---|
| Yeast extract (Oxoid, code L21) | 5.0 g |
| Special peptone (Oxoid, code L72) | 10.0 g |
| glucose | 10.0 g |
| PIPES | 6.04 g |

The pH adjusted to 6.9 and autoclaved and stored at room temperature in the dark.

PIPES buffer

| Compound | Concentration |
|---|---|
| PIPES, pH 6.9 | 20 mM |
| NaCl | 140 mM |
| $Na_2HPO_4$ | 0.75 mM |

### 2.6.2 Oligonucleotides

All oligos were dissolved in 10 mM Tris (pH8.3) to a final concentration of 100 pmol/µl. The probability of oligonucleotides forming secondary structures was checked using the Fold RNA prediction in the GCG package.

| Primer Name | Order name | Sequence of oligonucleotide |
|---|---|---|
| "Oligo primer1" | SP96S.r1a | 5´CAGGTGAGTGCTGGATATTGAGAGGAAGATGCAGA TGAAGATGGAGCTGATGATGATGAAGCACGACCTT CGATGGCATCTGAAC-3′ |
| "Oligo primer2" | SP96T.r1a | 5′CAGGTGAGTGCTGGATATTGTGCGGTGGTGGTAG CAGTAGTAGCGGTAGTGGTAGCACGACCTTCGAT GGCATCTGAAC-3′ |
| "Oligo primer 3" | SP96G.r1a | 5′AGGTGAGTGCTGGATATTGAGATCCGGTAGAACCA CCTGAAGTAGATCCTCCAGAGGTTTGTGAACCAGC ACGACCTTCGATGGC-3′ |
| "primer A" | GSTL99.f1 | 5′-TTCAATGCTTGAAGGAGCTGTTT-3′ |
| "primer B" | VBamKpn | 5´-CTAGAGGATCCCCGGGTACCTAAATC-3´ |
| "primer C" | GSTptxt.r4a | 5´-CCGGGGATCCTCAGGTTG-5´ |
| "primer D" | ISPY.r1a | 5′-ACTGGGGATCCTCAGGTGAGTGCTGGATATTG-3′ |
| "primer 1" | SP60.f1a | 5′-CTCTAAGATAACTAATACTCAAACTCA-3´ |
| "primer 2" | SP60.r1a | 5′-TGTTCCAATCACCATTAACTTCACCA-3´ |
| "primer 3" | SP60.f2a | 5′-CCACCTCTAGAATCCCATACTACATTAAAATATTTG-3´ |
| "primer 4" | SP60.r2a | 5′-GCTAATGCATATGTTAAGAGTGATAATAAGGCAATT AATAATGATAAAATCTTCA-3´ |
| "primer 5" | PsAsig.f1 | 5-GCCTTATTATCACTCTTAACATATGCATTAGC-3´ |
| "primer 6" | ISPY.r1a | 5-ACTGGGGATCCTCAGGTGAGTGCTGGATATTG-3′ |
| "primer 7" | Ddp2.ori.xba.r2 | 5´-GTAGATCCTCTAGACGAGCAC-3´ |
| "primer 8" | VBamKpn | 5′-CTAGAGGATCCCCGGGTACCTAAATC-3´ |
| Primer for sequencing | P174GST.f2c | 5´-GTGCCTTGATGCCTTCCCAAA-3´ |
| | gstA89.r2 | 5-CCAAAACAGCTCCTTCAAGCATTGAAATTTCTGC-3´ |
| | GSTfx.R4 | 5´-GCACGACCTTCGATGGCATCTGAACCT-3´ |

**Table 2-1 Oligonucleotides used in PCR reactions**

Primer name: used for the PCR schemes figure 3-30, 3-35 and 3-37 in the result chapter

Factor X cleavage site, ISPY sequence, *Bam*H1 cleavage site, corrected mutation, *Xba*I cleavage site, *Nsi*I cleavage site in PsA sequence

### 2.6.3 PCR (Polymerase chain reaction)

DNA fragments were modified by PCR reactions (111). To guarantee a good annealing of the primers with the template DNA, they were so chosen that their annealing values ($T_{m)}$, calculated using the formula $T_m = 2(A+T) + 4(G+C)$, were similar. This temperature was reduced by further 2 °C for each mismatched base. The annealing temperature used in the PCR was 5 °C lower than the calculated $T_m$ value. The extension time ($E_t$), required to synthesise the PCR product, was calculated as $E_t$=1 min/500 bp. Generally 20-30 cycles were used to amplify the DNA.

The reaction was done in 0.5 ml Eppendorf tubes (Bacto) containing ~8 ng template DNA, 10 pM each primer, 500 µM dNTPs (lithium salts, Boehringer Mannheim) and Accutaq buffer (Sigma) to 20 µl total volume. The liquid was covered by 20 µl mineral oil and the reaction was started during the 94 °C hold period by adding an Accutaq polymerase mixture (5µl Accutaq buffer containing dNTP′s).

PCR program chosen on Hybaid Omnigene PCR machine:

| | |
|---|---|
| 94 °C | 3 min |
| 94 °C | hold |
| [94 °C, 30 sec/ ($T_m$ – 7) °C, 1 min/ 68 °C, $E_t$ min] x | 5 cyles |
| [94 °C, 30 sec/ ($T_m$ – 5) °C, 1 min/ 68 °C, $E_t$ min] x | 15 – 20 cycles |
| 68°C | 8 min |
| 26°C | 1 min |

### 2.6.4 Separation and recovery of DNA using agarose gels

DNA was routinely separated at a constant voltage of 90 V in 1 % agarose gels (w/v) (Progen) for fragments larger than 0.5 kb size and in 2 % (w/v) gels for smaller pieces. Running buffer was 1 x TBE buffer. Ethidium bromide (Sigma, 25 µg/ml) was added before the agarose gel was poured so the DNA bands could be afterwards visualised under UV light. To recover DNA, the bands were cut out and extracted as described in the QIAquick gel extraction kit™ protocol (Qiagen).

### 2.6.5 Precipitation of DNA

To enhance the precipitation of DNA (100 µl sample volume) 1/10 volume of 3 M sodium acetate (pH 7) solution was added to neutralise the negatively charged DNA backbone with sodium ions. 3 volumes ethanol 95 % was added and kept for 20 min at –80 °C. The precipitated DNA was spun for 20 min at 14000 rpm in an Eppendorf centrifuge and the supernatant discarded. The DNA pellet was washed with 500 µl (70 % v/v) ethanol. The pellet was dried for 5 min in the Speedyvac and the DNA dissolved in sterile $dH_2O$ and stored at –20.

### 2.6.6 Restriction digests and ligation

DNA was digested using Boehringer restriction enzyme buffers and 1-10 units (U) enzyme per µg of DNA for 1-2 h at 37 °C. The reaction was stopped by heating for 15 min to 70 °C. DNA fragments were purified by agarose gel electrophoresis.

When two DNA fragments were to be ligated the fragments were gel purified and pooled. When a PCR of the whole plasmid was to be recircularized the ligation was carried out directly on the restriction digest. The digestion was stopped and 1:2 diluted with water before the right condition for the ligation was adjusted with 10 x ligase buffer. To start the reaction 1-2 units T4 DNA ligase (Sigma) was added and incubated at 4 °C overnight. The reaction was stopped by heating for 15 min to 70 °C.

### 2.6.7 DNA sequencing

DNA for sequencing was prepared using Qiagen midiprep filter kit and the sequencing reaction used the Big Dye Terminator RR-Mix (Perkin Elmer) solution. The used primers are listed in Table 2-1. For DNA sequencing gels were run by the sequencing facility at the University of New South Wales or Macquarie University (Sydney, Asutralia).

### 2.6.8 Preparation of electroporation competent *E. coli* cells

10 ml L-broth media was inoculated with *E. coli* DH5α cells stored at –80 °C. The culture was incubated overnight at 37 °C at 200 rpm on an orbital shaker before it was diluted on the next day in 1 L of 37 °C warm L-broth. This flask was shaken vigorous at 200 rpm, until the cells had reached an optical density of 0.5 to 0.6 at 600 nm and therefore still grew logarithmically. The culture was chilled to 4 °C for 30 min and maintained at this temperature for all the next steps. The cells were pelleted at 4000 g for 15 min in a prechilled rotor and as much as possible of the supernatant was removed. To remove the rest of the medium the cells were washed twice with 500 ml chilled sterile water. The cells were resuspended in 20 ml chilled sterile 10 % (v/v) glycerol in $dH_2O$, pelleted again and then finally resuspended in sterile 10 % (v/v) glycerol solution to a density of 1-3 x $10^{10}$ cells/ml. The cells were aliquoted into sterile tubes, snap frozen in liquid nitrogen and stored at –80 °C.

### 2.6.9 Transformation of plasmids into *E. coli*

Electroporation was carried out using a BioRad Gene Pulser set at:
1.8 kV, 25µFD capacitance and 200 ohms resistance
Capacitance Extender 250 µFD

Competent *E. coli* cells were thawed slowly on ice. 40 µl aliquots of *E. coli* were mixed with 2 µl DNA immediately before the solution was pipetted in a 0.1 cm prechilled cuvette (BioRad). After the voltage pulse was applied the cells were immediately resuspended with 200 µl of SOC media. The cells were transferred back in the Eppendorf cup and incubated with shaking at 37 °C for 1 h, before being plated out in different dilutions onto L-agar plates containing 100 µg/ml ampicillin and grown at 37 °C overnight.

### 2.6.10 Plasmid preparation

Isolation of plasmid DNA from *E. coli* was done by using the Qiagen Qiafilter Midi Kits as described in Qiagen Plasmid Purification Handbook (January 1997).

### 2.6.11 Transformation of *D. discoideum*

The DNA was transformed into *D. discoideum* by the calcium phosphate transformation procedure according to Nellen (93). Before transformation, 10 µg pMUW110 vector DNA were mixed with 2 µg expression vector DNA and precipitated to ensure the DNA sterile. The pellet was dissolved in 300 µl $dH_2O$ and used for the calcium phosphate precipitation step.

For the transformation *D. discoideum* strain HU2868 was used. 20 ml HU2868 (1-2 x $10^6$ cells/ml) of ashax growing cells were placed on a plastic petri dish for at least 30 min, allowed the cells to settle down and adhere to the plastic. The medium was carefully replaced by 10 ml of 20 mM PIPES/ HL5 medium containing 100 µg/ml ampicillin to wash the cells. During this time the calcium phosphate precipitation was done. 300 µl DNA was mixed in a glass tube with 38 µl 2 M $CaCl_2$ solution followed by 300 µl 40 mM PIPES buffer and allowed to precipitate for a repeated mixing of the sample for 25 min. As much as possible of the washing media was again carefully removed from the cells and the Ca-DNA precipitate added dropwise onto the cells. The DNA was allowed to coat the cells for 30 min by gentle rocking and then 10 ml of PIPES/ HL5 medium containing 10 µl Ampicillin (10 %) and 10 µl of the fungicide Amphotericin B (500 µg /ml) was added. Three hours later the cells were given an osmotic shock. The medium was removed and the cells covered with 2 ml of 18 % glycerol in 20 mM PIPES buffer for exactly 3 min, before 10 ml medium 20 mM PIPES/ HL5 medium containing 10 µl Ampicillin (10 %) was added. The cells were allowed to recover overnight before the media was replaced with axenic media including 10 µg G418. One day later, the transformed cells were spread on plates of the bacteria *Micrococcus luteus* (32, 146). *M. luteus* is used as the food source because of its resistance against the selective marker G418. The bacteria suspended in salt solution were spread out on SM/5 agar plates containing 5 µg G418 and allowed to dry. The transformant *Dictyostelium* cells were resuspended in the axenic medium and spun down for 3 min at 1000 rpm. The pellet was resuspended in the rest of the bacteria salt solution and spread out on the *M. luteus* spread plates and dried again. After 7-9 days little colonies could be observed. Cells or spores from the colonies were maintained on *Micrococcus* on SM/5 containing 10 µg G418 plates.

# 3 Results

At the beginning of this chapter different monoclonal antibodies are introduced which are an important tool for purification of SP96 and for characterisation of its modifications. The purification of SP96 from different *D. discoideum* strains is described, followed by the qualitative and quantitative characterisation of the glycan structures on the protein. The composition of the glycans was determined by phosphate, monosaccharide and GC-MS analysis. GC-MS analysis was also used to determine the linkage of the sugars to the protein. To elucidate the intact glycan structure, released oligosaccharides were analysed by HPAEC-PAD, LC-MS and NMR spectrometry.

The last part of this chapter deals with the localization of the modified sites in the SP96 protein. The modified hydroxyamino acids are identified by amino acid analysis and the specific sites are identified by two approaches. One approach uses an unspecific digestion enzyme to cleave the protein. The heavily glycosylated part of the protein is resistant to digestion and is partially characterised by Edman sequencing. The second approach is the use of an *in vivo* expression system established in *D. discoideum* itself for the analysis of potential glycosylated motifs. DNA encoding potential motifs for *O*-glycosylation were fused to a gene for a carrier protein, glutathione-S-transferase (GST), and expressed in *D. discoideum*. After purification of the GST fusion protein, the attached motifs were analysed immunochemically for glycosylation.

## 3.1 Overview: Immunochemical detection of SP96 in different strains

An important tool for the work on spore coat proteins is the use of monoclonal antibodies. These antibodies, mostly made against crude protein extracts, recognise highly specific structural elements on the proteins. The binding region, called the epitope, can be an amino acid sequence, a structural element of post-translational modifications, a mixture of both, or the three dimensional structure of the protein. Although only little is known about most of the epitope structures recognised by monoclonal antibodies, their sensitivity and specificity are helpful for the purification and characterisation of the spore coat proteins. For this work different mAbs were used to detect SP96 in the wild-type (X22) and the fucose mutants HU2470 and HU2471. It is known (11, 43, 87, 141) that the epitopes of these antibodies include at least partly different carbohydrate structures.

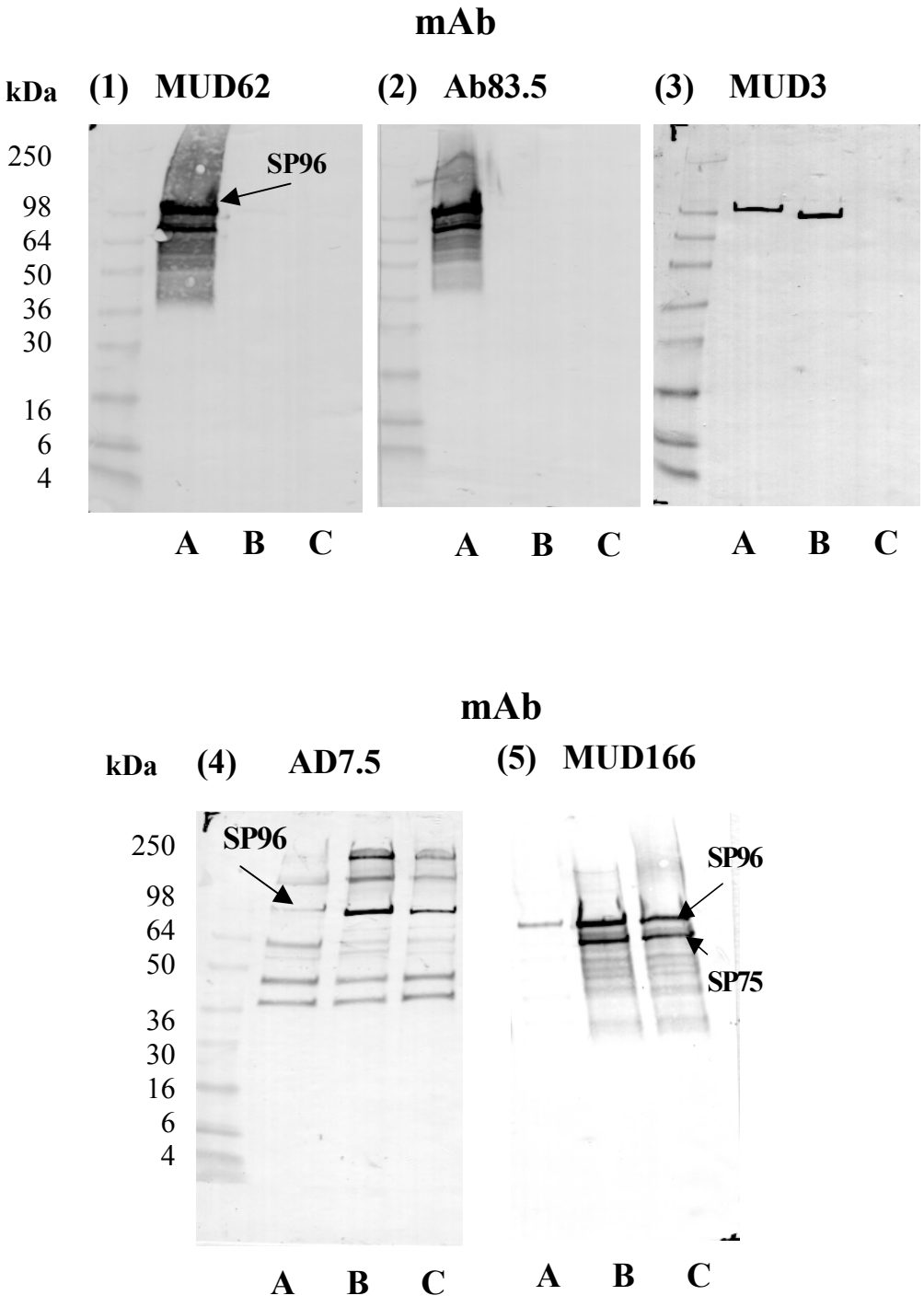



**Figure 3-1 Immunochemical detection of spore coat proteins with mAb MUD62 (1), mAb 83.5 (2), MUD3 (3), AD7.5 (4) and MUD166 (5) in wild-type X22 (A), fucose mutant HU2470 (B) and fucose mutant HU2471 (C).**

The MUD62 (1) and 83.5 (2) monoclonal antibodies display a similar binding pattern. Both show a strong labelling of SP96 and SP75 in the wild-type (Figure 3-1). The decreased level

of fucosylation in the mutants HU2470 (B) and HU2471 (C) results in the loss of the recognised epitope structure and therefore no labelling can be observed with both monoclonal antibodies. MUD3 (3) is the only mAb which recognises specifically SP96 in the wild-type X22 (A) and the fucose mutant HU2470 (B). The MUD3 epitope is not recognised in the fucose mutant HU2471 (C) (Figure 3-1). The difference of about 10kDa between the X22-SP96 protein and HU2470-SP96 is the result of the altered migration due to decreased fucosylation of the mutant protein (11). The monoclonal antibodies MUD166 (4) and AD7.5 (5) (Figure 3-1) show a reciprocal binding intensity for SP96 in mutant versus wild-type in comparison to MUD62 and 83.5. On wild-type SP96 only a very weak labelling can be observed. It seems that here fucose prohibits the binding of MUD166 and AD7.5. In the fucose mutants the SP96 band is strongly visualized by these mAbs which indicates that the epitope is presented after losing fucose. Interesting is the difference of the labelling of SP75 with MUD166 and AD7.5, because in binding studies of cysteine proteinases and of SP96 they showed some similarities.

## 3.2 Purification of SP96

### 3.2.1 Strategy

The protein SP96 is part of an external protein complex in the spore coat. A four step sequential extraction of *D. discoideum* spores with (I) 50 mM Tris/HCl (pH8), (II) 50 mM Tris/HCl (pH8)/ 8 M urea at room-temperature, (III) the same solution under boiling conditions and (IV) finally boiling under reduced conditions shows that SP96 is associated with the protein complex in different forms. The first extraction step released only a minor amount of weakly associated SP96 from the protein complex detected by MUD3. The main part of the protein was extracted by boiling with 8 M urea, which breaks down non-covalent bonds. The last step of boiling under reduced conditions released further SP96 from the membrane; therefore, part of the protein must also be integrated by covalent disulfide bonds or be hold in place by proteins that are linked by S-S bonds in the spore coat protein complex.

After extraction of SP96 out of the spore coat the already reported highly negative charge of the protein (22, 43) allowed an ion exchange chromatography purification step to be used to purify the protein. The monoclonal antibody MUD3, which recognised SP96 in the

wild-type X22 and the fucose mutant HU2470 was used as a tool for further affinity purification of SP96 from these strains. Figure 3-2 outlines this process of purification.

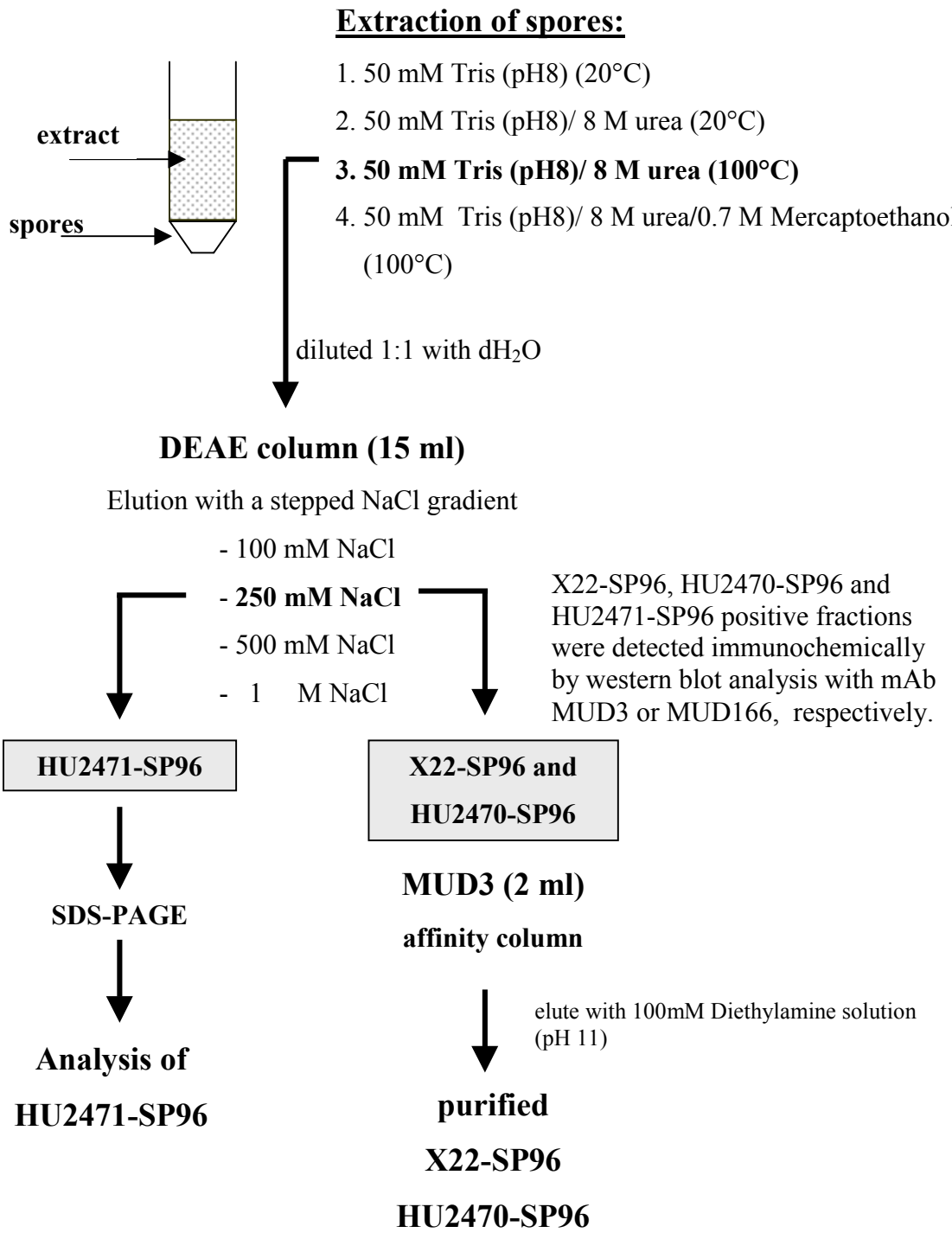



**Figure 3-2 Scheme of SP96 purification**

### 3.2.2 Steps of purification

Spores of the strains X22, HU2470 and HU2471 were harvested and extracted (Material and Methods; 2.4.2). For development of the purification protocol the fraction with the highest content of MUD3 recognising SP96 protein, "hot urea fraction", was chosen. Anion-exchange-chromatography with a stepwise gradient for elution was the first step of purification which was done on a DEAE-anion exchange column. X22-SP96, HU2470-SP96 and HU2471-SP96 positive fractions were identified by western blot analysis with MUD3 or MUD166, respectively (Figure 3-2) and gave an approximately 10 fold purification. A typical elution profile of the anion exchange chromatography is shown in Figure 3-3 for strain HU2471. The corresponding SDS-PAGE of the different eluted protein fractions is presented in Figure 3-4. Both the wild-type strain and also the fucose mutant HU2470 had a similar elution profile and protein pattern on the SDS-PAGE.

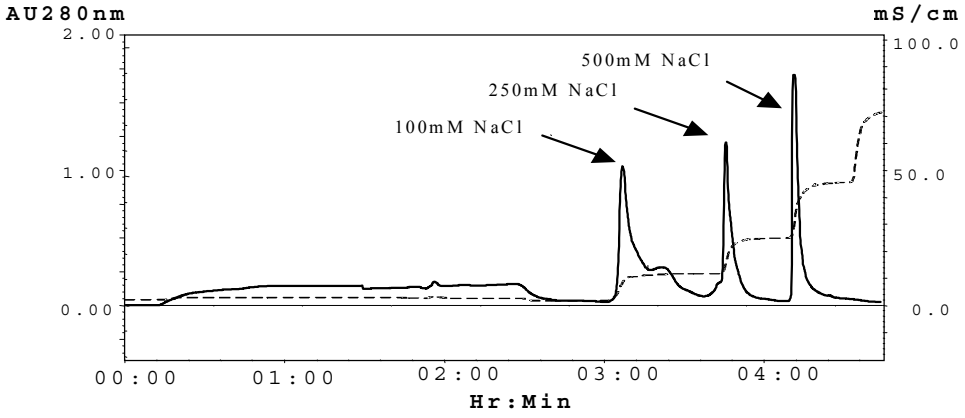



**Figure 3-3 Elution profile DEAE-anion exchange chromatography (strain HU2471)**
DEAE column volume 15 ml.
Protein elution was done with 4 M urea/ 25 mM Tris (pH 8) solution with different NaCl salt concentrations (100 mM, 250 mM, 500 mM and 1 M) up to the point where Absorbance (AU280 nm) reached baseline.
Conductivity of the increased salt concentration was monitored in mS/ cm.

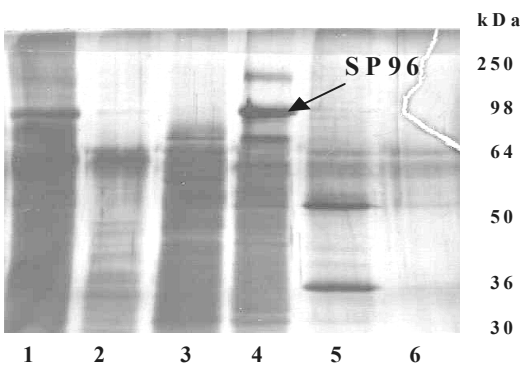



**Figure 3-4 Silver stained SDS-PAGE of eluted protein fractions after DEAE Anion-exchange chromatography of strain HU2471**

1) Extract of spores
2) Flow through
3) 100 mM NaCl fraction
4) 250 mM NaCl fraction
5) 500 mM NaCl fraction
6) 1 M NaCl fraction

Although the level of purification was still quite low, the SP96 protein of each strain was the major protein band in the range of approximately 80-140 kDa for the fractions eluted with 250 mM NaCl. Indeed, the amino acid composition analysis of the SP96 protein band from the different strains, after transfer to a PVDF membrane and hydrolysis, already gave a good match to the ExPASy database[f] amino acid composition of SP96[g].

To be able to quantify the post-translational modifications on SP96 and to be flexible in the choice of methods used to characterise them, it was necessary to get pure SP96 sample in salt-free solution. The wild-type SP96 and the fucose mutant HU2470-SP96 were specifically recognised by the monoclonal antibody MUD3 (Figure 3-1), and so a MUD3 affinity column was used for the next purification step. Eluted fractions were immunochemically screened by MUD3 dot-blot analysis (Figure 3-5). The major two fractions (M) were checked by a silver stained SDS-PAGE which displayed that the SP96 protein in those fractions was still contaminated by four proteins in the range of 55-85 kDa (Figure 3-6). Therefore, the fractions (S) on either side of the main peak were discarded and the main fractions (M) were used for a repeated purification step on the affinity column. Again, only the main fractions were picked, concentrated in a concentrator (Millipore Ultrafree-15 Biomax) and this time desalted in the same concentrator by several water washs. Figure 3-7 shows the purified X22-SP96 and HU2470-SP96 sample after SDS-PAGE. Quite often glycoproteins show a weak stain with the common staining methods, like coomassie or silver stain, but the SP96 band was quite visible and it was the only protein visible in the

[f] http://www.expasy.ch/ (Identification tools; AACompIdent)

[g] SWISS-PROT and TrEMBL Protein Sequence Databases: AC number P14328

silver stained gel (Figure 3-7, **A**). After transfer of the proteins to PVDF membrane a very sensitive glycostain, based on immunochemical detection of biotin which is bound to the aldehyde groups of periodate cleaved sugars, was used to exclude the contamination through glycoproteins. The result showed that the purified SP96 of both strains was homogeneous and that no further glycoprotein contamination was visible (Figure 3-7, **B**). That the visualised band in (Figure 3-7, **A** and **B**) were definitely SP96 could be displayed after immunochemical detection with MUD3 (Figure 3-7, **C**)

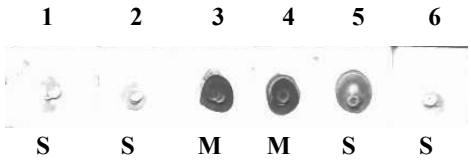
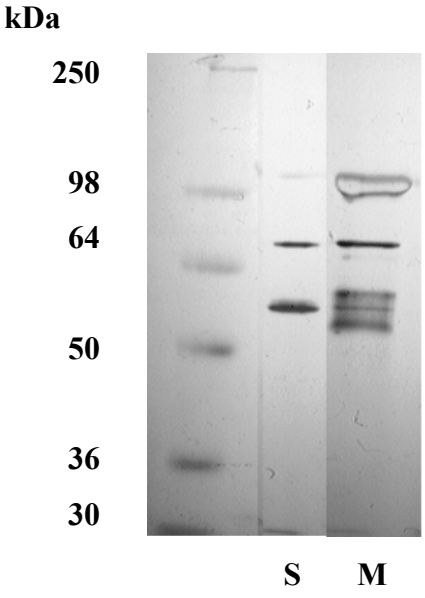



**Figure 3-5 MUD3 dot blot to screen for SP96 positive fractions after stepwise elution (1-6) from the MUD3 affinity column**

**S**) side fraction, **M**) main fraction fraction size 1ml

**Figure 3-6**

**Silver stained SDS-PAGE of a site fraction (S; 2) and a main (M; 3) after MUD3 affinity purification**



Amino acid analysis of the protein preparations from 1 x $10^{10}$ spores showed a yield of 2.4 nmol protein from the X22 wild-type and 3.2 nmol protein from the mutant HU2470. This is equivalent to not less than $10^5$ molecules of SP96 recovered in the hot urea extraction of each mutant and wild-type spore, probably $10^6$ after losses considered. The average amount of total protein after each purification step is shown in **Table 3-1**. The original spore extract (hot urea fraction) from the X22 strain (80 mg) yielded, after approximately 560 times concentration, 144 µg SP96 protein and and a yield of 190 µg SP96 protein from HU2470 (87 mg), after approximately 450 times concentration.

The following characterisation of the post-translational modifications of X22-SP96 and HU2470-SP96 was done from affinity purified salt free sample. Because MUD3 could not be used for purifying HU2471-SP96, all characterisation steps of this protein were done after transfer to a PVDF membrane. SP96 from this mutant was transferred by western blot onto PVDF membrane and visualised with Amido-black after the second purification step (DEAE column).

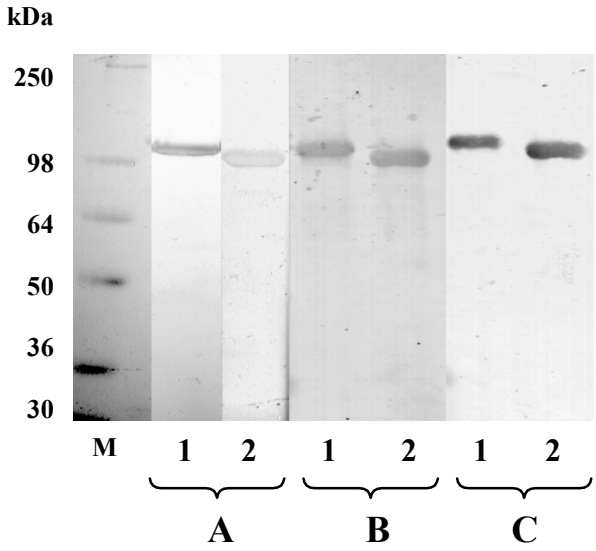



**Figure 3-7 SDS-PAGE of final purified X22-SP96 (1) and HU2470-SP96 (2)**

Separation of affinity purified SP96 from the wild-type (lane 1, about 1µg) and fucose mutant (lane 2, about 1 µg) on a 10 % (w/v acrylamide) SDS-PAGE under reducing conditions. Visualisation in gel with silver nitrate stain **(A)**, after western blot onto nitrocellulose with ImmunoBlot stain for glycoproteins **(B)** and SP96 protein/ carbohydrate specific antibody MUD3 **(C)**. Lane M, molecular mass marker proteins with values in kDa.

| Purification step | Total amount of protein [µg] | |
|---|---|---|
| | X22 | HU2470 |
| 1) Spore-Extract | 80000 | 87000 |
| 2) Anion-exchange | 7300 | 8200 |
| 3) MUD3 affinity (2 x) | 144 | 190 |

**Table 3-1 Summary of the recovery of protein in each purification step**

All values are calculated to the extraction of $10^{10}$ spores.

## 3.3 Identification of post-translational modifications on SP96

The identification of the post-translational modifications was done in two steps (Figure 3-8). The first, a composition analysis, was done to identify the phosphoamino acids and the sugars which are involved in the construction of the modifications. This step removed and took apart the modification of SP96 in detectable units by acid hydrolysis. The second step was a detailed structure analysis of the different modifications. The questions to answer were:

1) How many different modification forms are attached to the protein?
2) How are the modification forms linked to the protein ?
3) What is the size of the modifications ?
4) What is the detailed structure of each modification?

In this step the intact glycans were removed from the SP96 protein by an alkaline β-elimination.

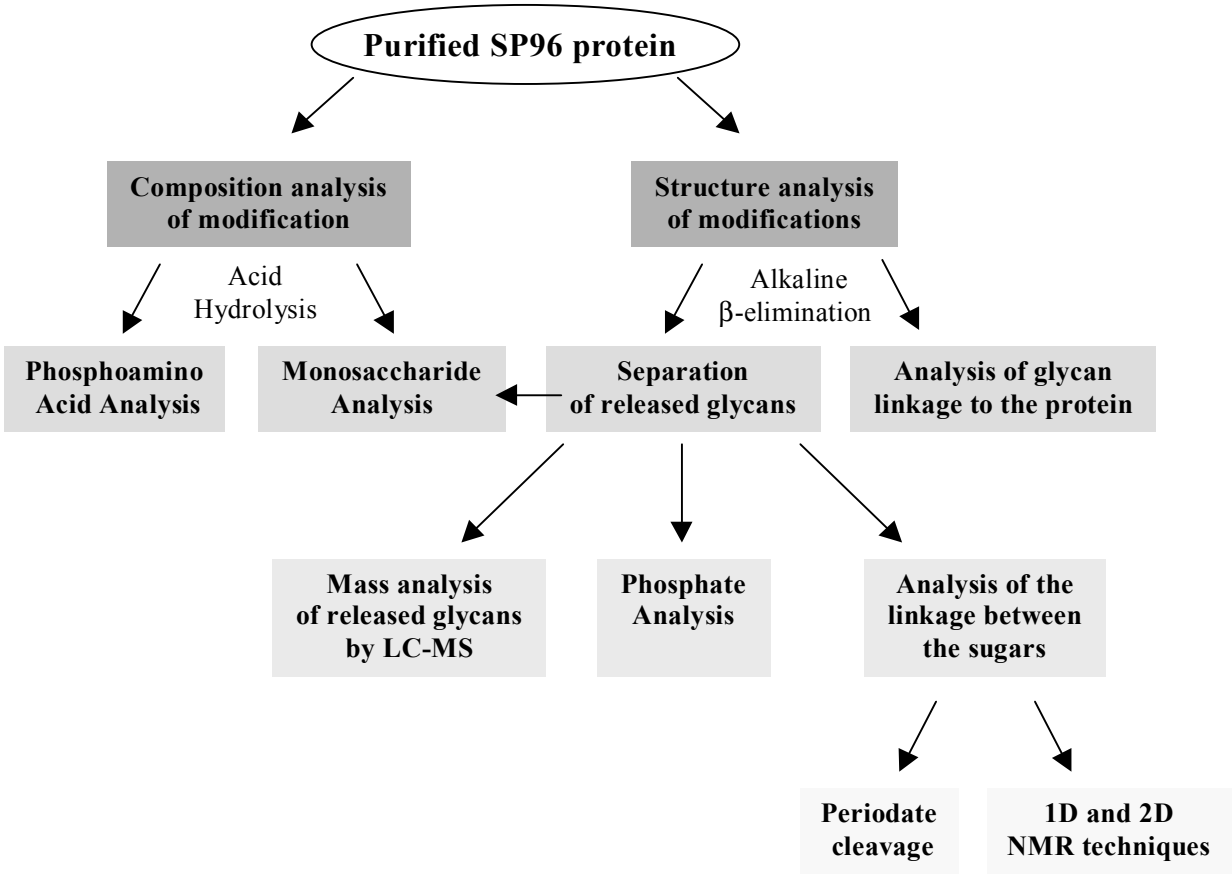



**Figure 3-8 Scheme of identification of post-translational modifications on SP96**

### 3.3.1 Identification of phosphorylated amino acids

The typical acceptors for eukaryotic phosphorylation are the hydroxyamino acids serine (Ser), threonine (Thr) and tyrosine (Tyr). The SP96 amino acid composition includes 121 Ser, 83 Thr and 12 Tyr residues, which are potential sites for phosphorylation and/ or glycosylation.

In the early eighties it was shown by $^{32}P$ radiolabelling experiments that some spore coat proteins of *D. discoideum* were phosphorylated. A major spore coat phosphoprotein was identified by several groups as a 103 kDa (1), 95 kDa (15) and 96 kDa (22) spore coat protein. These proteins correspond to the now named SP96. By comparing the radiolabelled,

hydrolysed protein with standard phosphoamino acids using thin layer chromatography, serine was identified as the only phosphorylated amino acid in those experiments.

In this work it was confirmed that serine was the only phosphorylated amino acid in SP96 from all three strains using the very sensitive Fmoc phosphoamino acid analysis method (150). The acid hydrolysed protein from the strains showed the presence of serine phosphate with no evidence of threonine phosphate or tyrosine phosphate (Figure 3-9). For hydrolysis MUD3-affinity purified X22-SP96 and the HU2470-SP96 samples (approx. 20 pmol) as well as a destained piece of PVDF membrane blotted HU2471-SP96 was used.

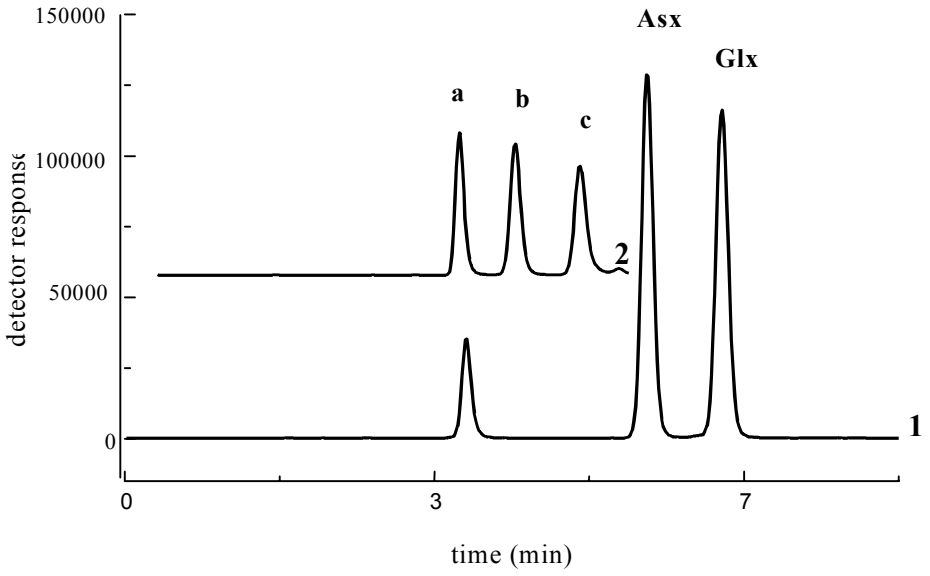


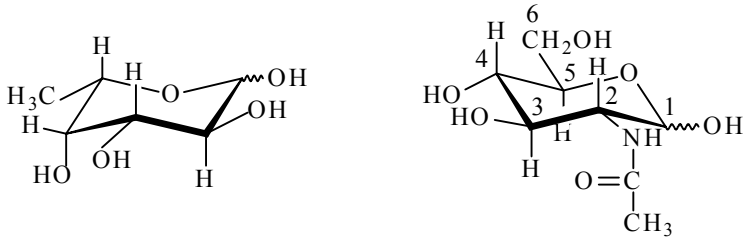

**Figure 3-9 Phosphoamino acid composition**

Hydrolysed and Fmoc derivatised sample of X22-SP96 **(1)** compared to derivatised standard phoshoamino acids **(2)** after reversed phase separation (**a**, phospho-serine; **b**, phospho-threonine; **c**, phospho-tyrosine) Chromatograms of acid hydrolysed SP96 from HU2470 and HU2471 showed the same result.

### 3.3.2 Monosaccharide composition analysis

For monosaccharide analysis, a high performance anion exchange chromatography system with pulsed amperometric detection (HPAEC-PAD) was used. For this all glycan structures on the protein have to be released and hydrolysed by TFA or HCl to monosaccharides which can then be detected in the nmol range. The identification was done by comparing the retention time to the internal standard deoxyglucose.

For monosaccharide analysis, X22-SP96 and HU2470-SP96 protein (25 pmol) was hydrolysed and the monosaccharides separated on a CarboPac PA10 column. Hydrolysis from the HU2471-SP96 protein was done again from PVDF membrane after partial purification by ion-exchange-chromatography without further quantification of the protein. The monosaccharide analysis of the acid hydrolysed proteins showed the presence of fucose (Fuc) (peak 1) and glucosamine (peak 3) in all strains (Figure 3-10, Figure 3-11). Hydrolysis with 4 M HCl, to release all amino sugars, showed no increase in the amount of glucosamine in all strains (data not shown). Glucosamine is the product of hydrolysed *N*-acetylglucosamine (GlcNAc), a common aminosugar in post-translationally modified proteins. The glucose, apparent in the chromatograms (peak 4), is a common contaminant of monosaccharide analysis and varied in all analyses. Using the internal standard (peak 2) the calculated ratio of Fuc:GlcNAc in SP96 was 1.2 : 1 in the wild-type, 0.3 : 1 in HU2470, and 0.04 : 1 in the HU2471 mutant. As the protein amount could not be quantified on the PVDF membrane, it was assumed that the amount of GlcNAc in HU2471-SP96 is comparable to that in the other strains (Chapter 3.4). The fucose content was calculated to drop from 100% in the X22-SP96 to approximately 25% in HU2470-SP96 to 3% in HU2471-SP96. These results coincide with observations done by Champion (11) where the fucose content in spores of these strains was determined. His experiments displayed a decrease of fucose to 15% in HU2470 compared to the wild-type, and to <5% in HU2471.



L-fucose* D-*N*-acetylglucosamine*

**Figure 3-10 Monosaccharides in all analysed strains of SP96**

* shown is the enantiomer of each sugar generally found in biological systems

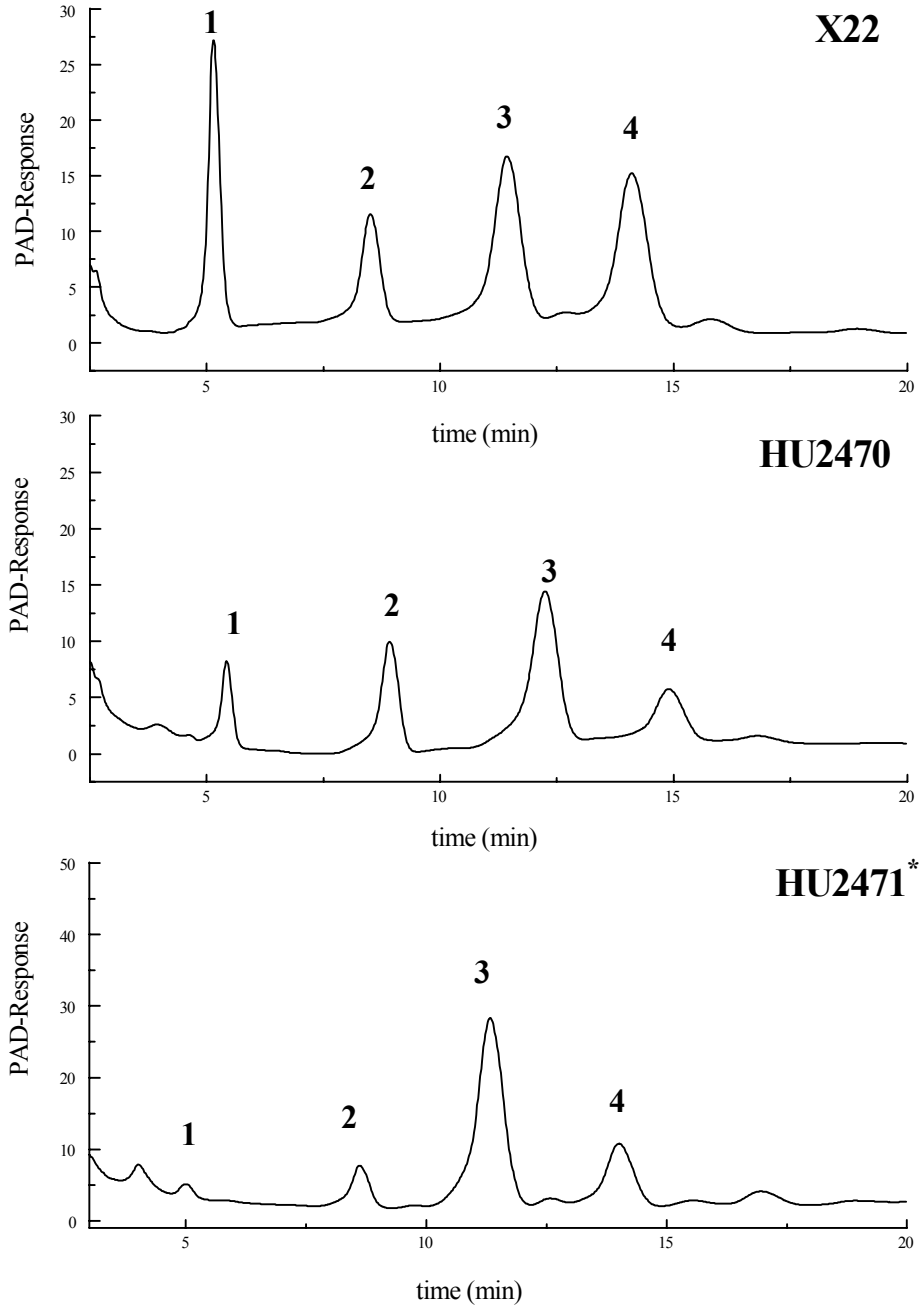



**Figure 3-11 Monosaccharide analysis**

**1)** Fucose; **2)** deoxyglucose (1nmol internal standard); **3)** glucosamine;
**4)** glucose (contamination, as signal fluctuated in different measurements done on the same sample)

* amount of HU2471-SP96 was not quantified before hydrolysis; therefore the ratio of GlcNAc to deoxyglucose is different in comparison to X22-SP96 and HU2470-SP96.

### 3.3.3 Determination of the glycan linkage to the protein by GC-MS analysis

To analyse the linkage of the sugars to the protein the glycans were β-eliminated, reduced and peracetylated for analysis by GC-MS. After β-elimination a single reducing terminal sugar, which is directly attached to the protein, is detected as an acetylated alditol, an open chain sugar form (Figure 3-12). Glycans larger than monosaccharides and phosphorylated sugars cannot be detected by this method. The composition of these larger glycans can then be determined by hydrolysis of the reduced β-eliminated sugars prior to acetylation for GC-MS. Any reducing terminal sugar is detected as an alditol, whereas any sugar attached to a linkage other than the peptide will not be reduced and will maintain an acetylated pyranoside ring structure. The alditols and pyranoses can be readily separated by GC and identified by their different fragmentation spectra obtained by EI-MS. Phosphorylated sugars will not be seen in the conditions of GC-MS so that dephosphorylation of the monosaccharides prior to peracetylation is required to determine the presence of any sugar phosphates.

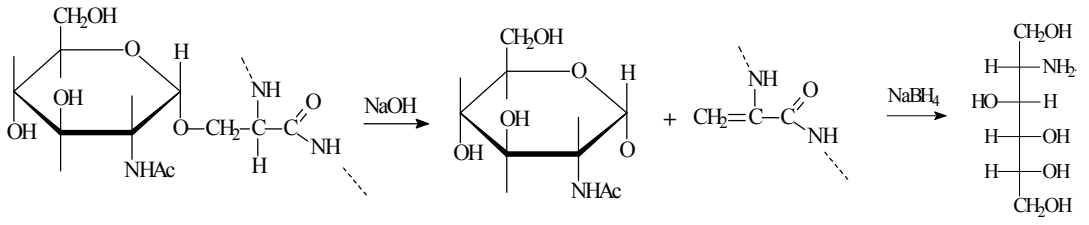



**Figure 3-12 Example of the β-elimination of GlcNAc from serine with subsequent reduction**

The glycans of SP96 of each strain were released by β-elimination using reducing alkaline conditions and aliquoted into four samples, each for a different treatment (Figure 3-13). To see whether a single monosaccharide was directly attached to the protein, the first sample was directly peracetylated (a). The second sample was hydrolysed with TFA and peracetylated to see the whole monosaccharide composition (b). The presence of phosphorylated sugars in the third aliquot was determined after treatment with HF to dephosphorylate the sugars prior to peracetylation (c) and the fourth aliquot was an acid hydrolysate of a similarly dephosphorylated sample (d).

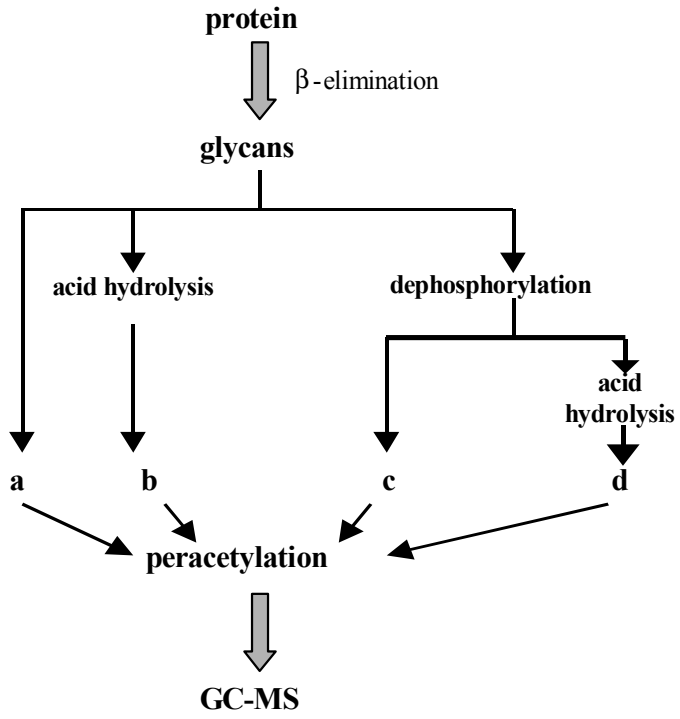



**Figure 3-13 Flow chart of sample preparation for GC-MS**

As an example, the GC-chromatograms of the β-eliminated, acid hydrolysed sample (treatment (b)) of X22-SP96, HU2470-SP96 and HU2471-SP96 are shown in Figure 3-14 . In the X22 and HU2470 sample a large fucitol acetate peak (peak 2) was detected compared to the HU2471 sample where this peak was completely missing. A peak of GlcNAc acetate in the pyranoside form (peak 4) could be observed in the sample of all three strains. The peaks, collectively labelled as 1, are the multiple isomers of deoxyhexosepyranoside (fucose acetate). The peak at around 16 min in the X22 and HU2470 and before at around 15.3 min in the HU2471 strain is of contaminating glucose peracetate which varied in all samples. Peak 3 is the internal standard myo-inositol. The corresponding EI-MS fragmentation spectra (Figure 3-15) were used to identify these sugars. In the HU2471 strain sample, were the glycans were released from the PVDF membrane, further non-sugar signals were observed as byproducts. The results after reductive β-elimination and acetylation (treatment (a)) showed a deoxysugar in the alditol acetate form which, from the monosaccharide analysis, must be fucitol with approximately the same amount in both the wild-type and the HU2470 mutant. The HU2471 mutant completely lacks this sugar modification. Single fucoses are thus attached via a

glycosidic *O*-linkage directly to X22-SP96 and to HU2470-SP96, but not to HU2471-SP96. In both mutants, HU2470-SP96 and HU2471-SP96, a small amount of non-reduced aminosugar in the pyranoside acetate form was detected, which agreed with the monosaccharide analysis where GlcNAc was observed in the mutants (Figure 3-11). Unfortunately the amount of GlcNAc pyranoside in the mutants after β-elimination treatment (a) could not be calculated because of a large contaminating peak. However, GlcNAc pyranoside was not in sufficient quantity to be significant, suggesting that most of the GlcNAc occurs in the mutant in an alkali-labile linkage that could not be reduced.

The GC of the acid hydrolysate of the reductively β-eliminated glycans (treatment (b)) showed approximately the same amount of fucitol as prior to hydrolysis. There was also a large amount of acid-released GlcNAc in the pyranoside form in the wild-type and mutant strains. The presence of the acetylated pyranoside form of the amino sugar reveals that GlcNAc is contained in the glycans of SP96, but is not at the reducing terminal attachment site to the protein.

The presence of a deoxysugar in the pyranoside form was also detected in the wild-type after hydrolysis, which suggests the presence of additional fucose residues which are not directly linked to the protein. This peak was dominant (approximately 30 fold more) in the wild type, hardly detectable in the HU2470 mutant, and not detectable at all in the HU2471 mutant. This absence of fucose in the mutants reflects the *modD352* mutation in HU2470, the *modE353* in HU2471, and therefore the loss of the MUD62 and Ab83.5 epitope (Figure 3-1).

As there was no evidence of N-acetylglucosaminitol after reductive β-elimination, it can be concluded that none of the N-acetylglucosamine is linked directly to the protein. Instead, in the wild-type and the mutants, GlcNAc is located as a distal residue of the sugar modification on the protein.

Dephosphorylation of the reduced β-eliminated glycans (treatment (c)) also showed the fucitol peak but, contrary to the non-dephosphorylated sample, GlcNAc pyranoside was present in all strains. This result strongly suggests that GlcNAc exists as a phosphorylated non-reducing terminal monosaccharide. The phosphate must be attached to the anomeric carbon (C1) position, as its location on any other carbon would result in an alditol acetate rather than an acetylated pyranoside, appearing after reductive β-elimination and dephosphorylation. This implies that GlcNAc is attached to the protein via a phosphodiester bond. Fucose pyranoside was also detected in the wild type at a low level after dephosphorylation but was not detected in mutant protein, even though the same amount of

HU2470-SP96 protein was used for the analysis. The same rationale would indicate the presence of a small amount of fucose in a phosphodiester linkage to the protein in the wild type SP96. Hydrolysis of the dephosphorylated β-eliminated glycans (treatment (d)) yielded the same GC eluted sugar forms (fucitol, GlcNAc pyranoside and Fuc pyranoside) as in treatment (b) (data not shown), since 2M TFA acid hydrolysis removed the phosphate from the sugars to the same extent as the HF treatment.

The results show that the approximately 75 % decreased level of fucose from wild-type to the HU2470 mutant is the consequence of a missing non reducing fucose linkage. This linkage could be explained by a phosphodiester linked fucose to the protein and/or a terminal fucose on the GlcNAc. If it is the former linkage, however, it would be expected to give a much stronger signal of fucose after dephosphorylation (treatment (c) provided it is the main modification, because of the resulting monosacchride.

The further decreased level of fucose between HU2470-SP96 and HU2471-SP96 (Figure 3-11) is thus the result of the missing, directly attached *O*-linked fucose (Figure 3-14).This modification has to be relevant to the MUD3 epitope because of the resulting loss of reactivity to MUD3 in the HU2471 strain compared to the other fucose mutant (HU2470) (Figure 3-1). Although the outcome from the monoaccharide analysis demonstrated that the HU2471 mutant has still approximately 3 % fucose on SP96, the GC-MS analysis did not show a possible linkage for this.

The following possible modifications of the different strains could be deduced from the GC-MS analysis (Figure 3-16)

.

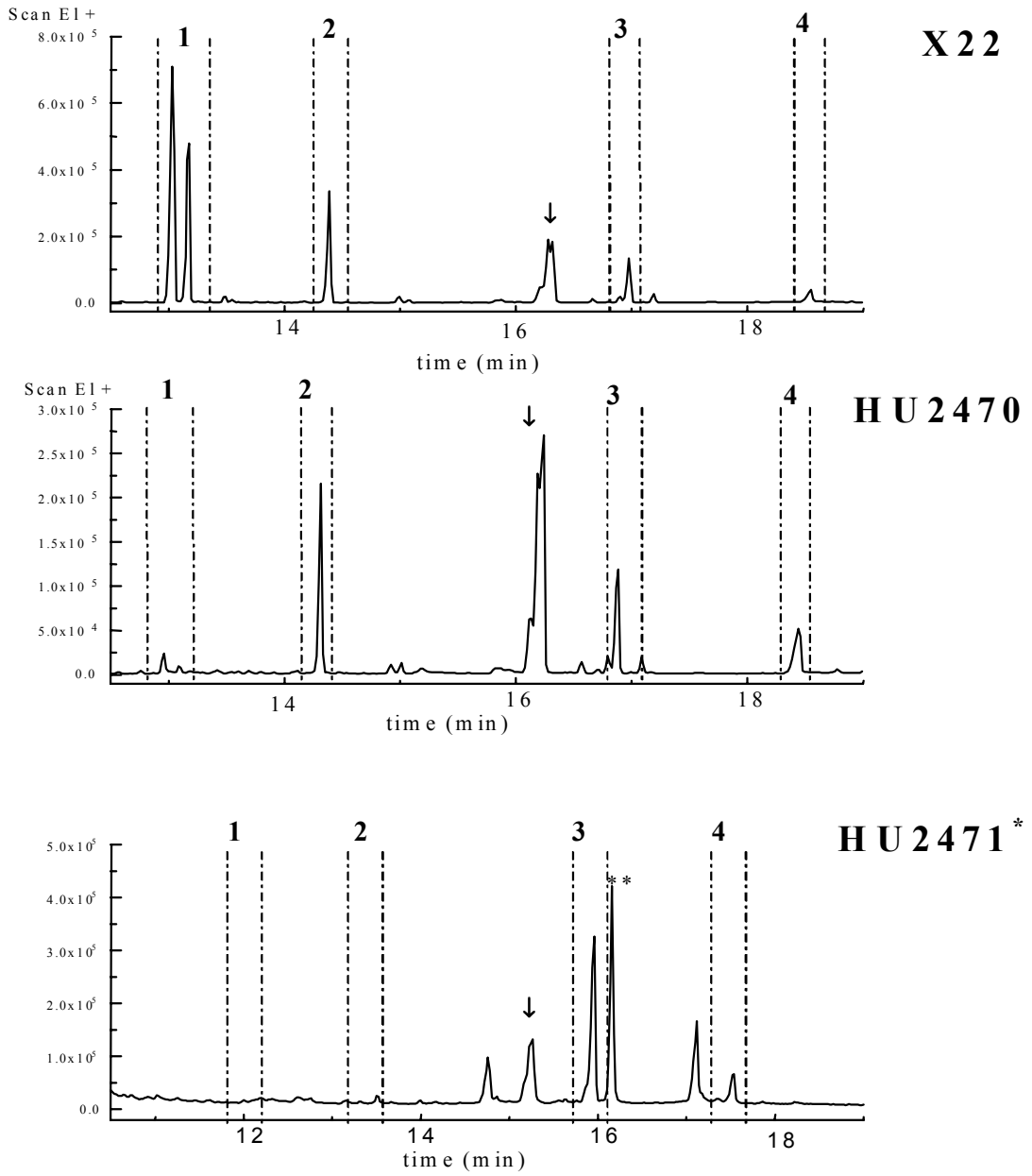



**Figure 3-14 GC spectrum of the released sugars.**

Gas chromatograms of the released sugars of SP96 from X22, HU2470 and HU2471 after β-elimination under reducing conditions, acid hydrolysis and peracetylation. Peak 1 was identified by retention time (dotted lines show window of retention of the isomeric forms of the standard sugars derivatised and run under the same conditions) and EI-MS fragmentation pattern, as acetyl-2,3,4-tri-O-acetyl-6-deoxy-hexopyranoside, which corresponds to fucose peracetate; peak 2 was identified as 1,2,3,4,5-penta-O-acetyl-6-deoxy-hexitol, which is the open chain reduced form of acetylated fucitol; peak 3 was the internal standard myo-inositol (0.5 nmol) and peak 4 was identified as hexopyranose-2-(acetylamino)-2-deoxy-1,3,4,6-tetraacetate, the pyranoside structure of GlcNAc peracetylated.

↓ contamination signal of glucose peracetate which varied in all samples

* sample was separated on a new column; therefore shift of retention time

** observed contamination signal shortly after internal standard peak (at 16.1 min for HU2471), stronger in PVDF membrane prepared sample

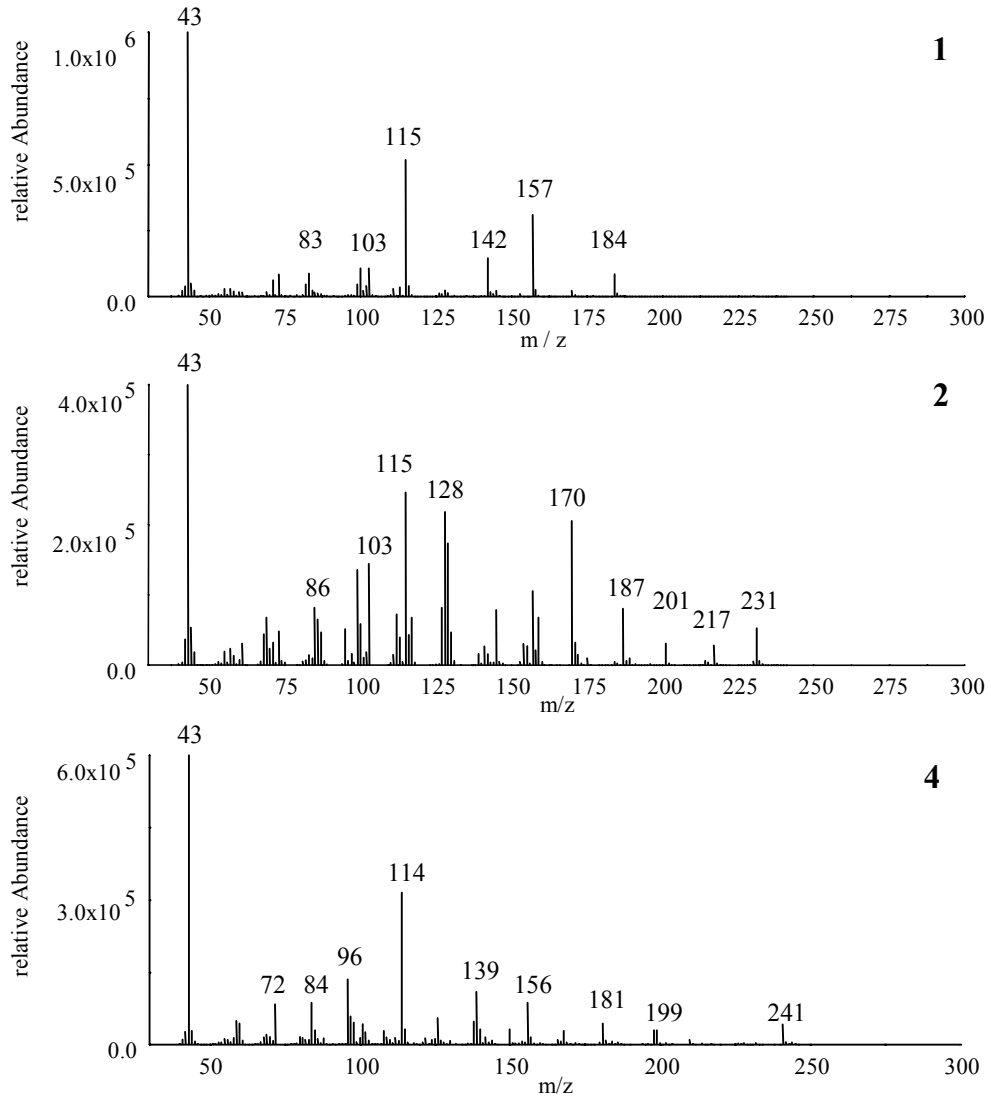



**Figure 3-15 EI-MS spectrum of peaks 1, 2 and 4 of Figure 3-14**

The sugars were identified as
1) acetyl-2,3,4-tri-O-acetyl-6-deoxy-hexopyranoside (fucose peracetate)
2) 1,2,3,4,5-penta-O-acetyl-6-deoxy-hexitol (acetylated fucitol)
4) hexopyranose-2-(acetylamino)-2-deoxy-1,3,4,6-tetraacetate (GlcNAc peracetylated)
by comparison with the standard library (GC-MS Software: Masslab)

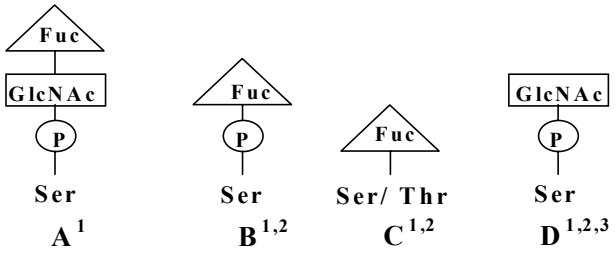



**Figure 3-16 Possible modifications on SP96 deduced from GC-MS analysis**

[1]) X22; [2]) HU2470; [3]) HU2471

### 3.3.4 Release and identification of phosphorylated mono- and disaccharides

For both analyses described previously (HPAEC-PAD and GC-MS) the glycan was hydrolysed to single sugars to determine the composition. To analyse the intact glycan structures, they need to be released without hydrolysis. They can then be analysed by liquid chromatography and mass spectrometry. Common ways to release intact glycans from protein are alkaline β-elimination (0.1 M NaOH) or mild acid hydrolysis (40mM TFA, 8 min, 100°C).

The phospho-amino acid analysis indicated the presence of phosphorylated serines, and the GC-MS results implied that the sugar phosphates were linked to the protein via a phosphodiester bond. Phosphodiester and fucose linkages have both been shown to be susceptible to mild-acid hydrolysis (114). Furthermore, some of the mAbs epitopes on SP96 have been observed to be mild-acid labile[h]. To analyse the mild-acid labile sugar linkages, SP96 protein from the wild-type and mutant strains (~500 pmol) was treated with mild acid and the released sugars separated by HPAEC-PAD. In all cases a large peak at 4 min was observed, which eluted at the same time as monosaccharides on a gradient capable of resolving monosaccharides from monosaccharide phosphates and disaccharides (data not shown). The acid treatment appeared to remove the phosphate from the sugar as well as to hydrolyse any disaccharides present.

Nonreductive β-elimination using 0.1 M NaOH was also used to try to release the glycans from the SP96 protein (1 nmol; X22-SP96, HU2470-SP96 and approximately the same amount HU2471-SP96). The reducing terminus of non substituted sugars undergoes a degradative 'peeling' reaction under alkaline conditions and would not be detectable. The treatment of X22-SP96 with NaOH (Figure 3-17) released one glycan (P1; 14.4 min). The same treatment released a single different glycan (P2 and P3 respectively; 18.1 min) from both HU2470-SP96 and HU2471-SP96 (Figure 3-17). This happened with a shift of retention time of over 4 min and a weaker signal.

Peak P1 was collected, desalted over a graphitised carbon column and an aliquot of the sample hydrolysed. Fucose and GlcNAc were detected by monosaccharide analysis and quantified by comparison to an internal standard (Figure 3-18). Phosphate was assayed after both strong-acid hydrolysis and alkaline-phosphatase treatment of the released P1 glycan. A

[h] A. Champion & L. Browne, unpuplished work

molar ratio of 1:1:1 of Fuc:GlcNAc:phosphate was obtained in the glycan released by β-elimination of the wild-type SP96.

The HPLC peak (Figure 3-17, P2 and P3) at 18.1 min from the alkaline treatment of the mutants HU2470-SP96 and HU2471-SP96 showed the same retention time as a standard of GlcNAc-α-1-P. Chromatography of a mixture of the sample with a GlcNAc-α-1-P standard gave one peak. The small shoulder on the P2 peak was also observed after treatment of the GlcNAc-α-1-P standard with alkali. This might be due to the cyclisation of some of the phosphate on the sugar under alkaline conditions (7).

To check for the possibility of phosphodiester-linked fucose, Fuc-α-1-P was analysed by HPAEC-PAD. Fuc-α-1-P eluted with a retention time of 16.2 min. In the X22-SP96, HU2470-SP96 and the HU2471-SP96 samples, no signal was detected at this retention time. However, the presence of this glycan cannot be discounted, as the response of the detector to phosphorylated fucose was much weaker (1:50) than to the monosaccharides, and therefore less than 5 nmol would be hard to detect. If the glycan were present, then the amount would be >50 x less than the amount of GlcNAc-α-1-P.

Treatment of the samples with mild acid (40mM TFA, 8 min, 100°C) caused the peaks of the P1, P2 and P3 glycans to disappear with a concomitant large increase of monosaccharide eluting at 4 min (data not shown). To investigate this apparent acid lability of GlcNAc-α-1-P, the standard was treated with two different mild-acid hydrolysis conditions which have been reported in the literature (viz: 40 mM TFA/ 100ºC/ 8 min (63) and 10mM HCl/ 100ºC/ 8 min (119)). After both treatments GlcNAc-α-1-P could no longer be detected by CarboPac PA1 chromatography. Also, under these conditions, the glycan P1 apparently lost both the phosphate and the fucose, thus confirming the similar sensitivity of these bonds to mild-acid hydrolysis (114) and explaining why no phosphorylated mono- or disaccharides were detected by the analysis of the mild-acid hydrolysate.

In summary, the wild-type analysis showed the presence of (Fuc, GlcNAc, $P_i$), while both mutants appeared to contain only GlcNAc, $P_i$.

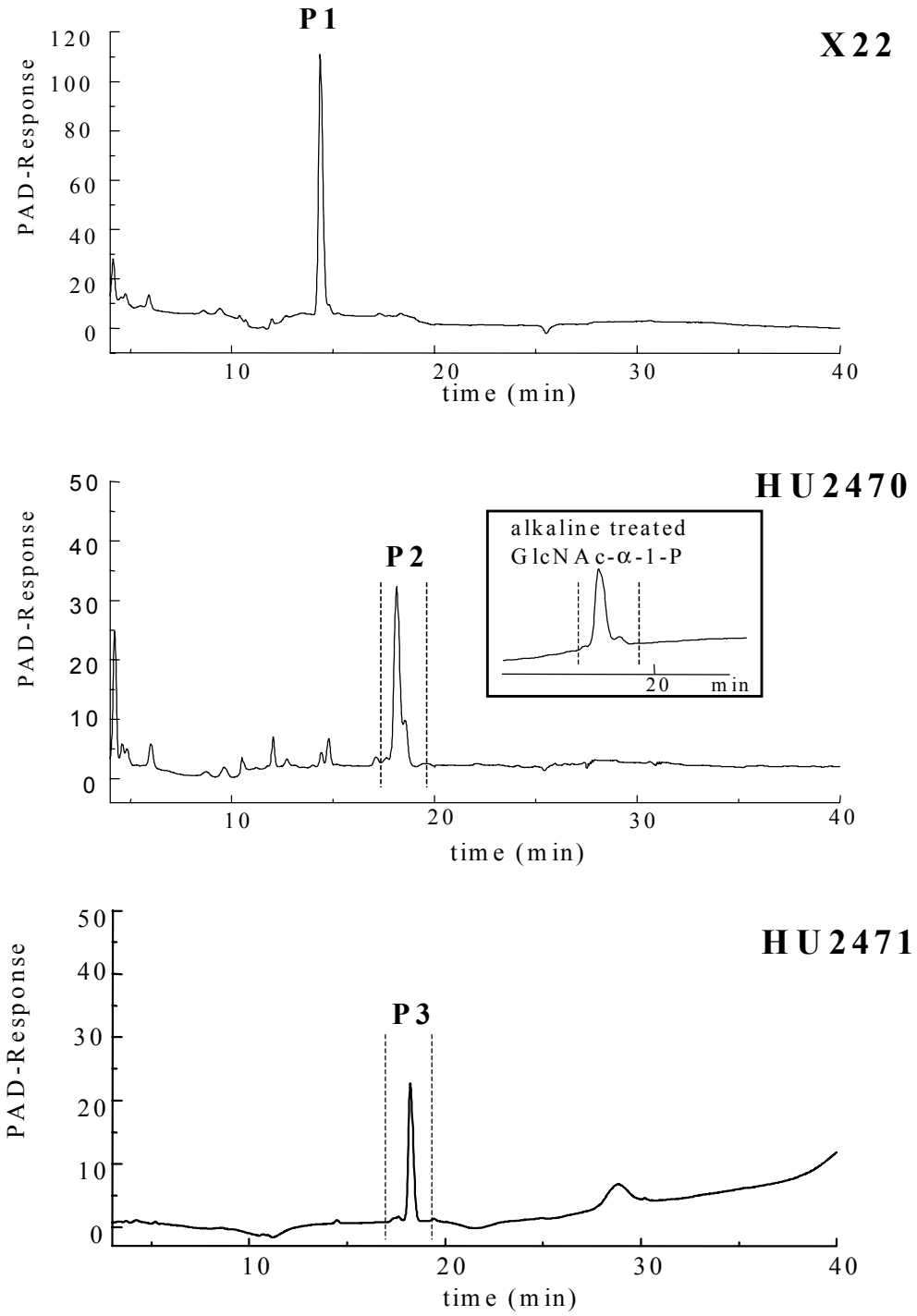



**Figure 3-17 Separation of released phosphorylated sugars by HPAEC-PAD**

After alkaline β-elimination of the SP96 protein (~1nmol), the released phosphorylated sugars were separated on a CarboPac PA1 column: X22 chromatogram of the released sugar (P1: 14.4 min); HU2470 and HU2471 chromatogram of the released sugar (P2 and P3: 18.1 min). This latter peak eluted at the same time as a GlcNAc-α-1-P standard.

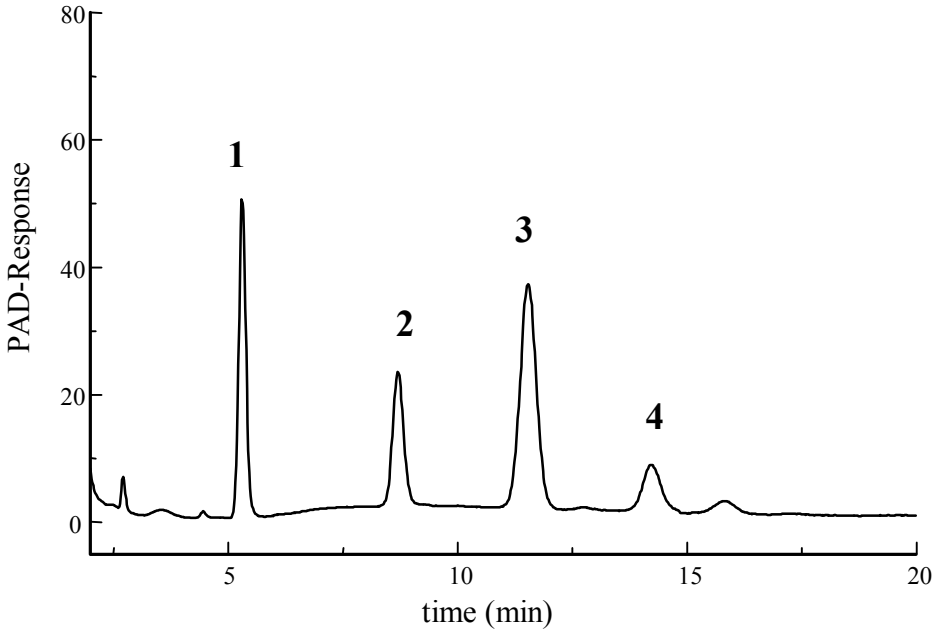



**Figure 3-18 Monosaccharide composition of X22-SP96 glycan P1**

The β-eliminated and HPAEC-PAD separated peak (P1) (Figure 3-17) of wild-type SP96 was desalted on a graphitised carbon column and hydrolysed in 2M TFA. The monosaccharides were identified by their retention times on HPAEC-PAD as fucose (1) and glucosamine (3) in the molar ratio of 1:1. Deoxyglucose (2) was added as an internal-standard for quantitation, and some commonly contaminating glucose (4) was present.

### 3.3.5 Mass analysis of the released glycans by LC-MS

Samples of X22-SP96, HU2470-SP96 and HU2471-SP96 protein were treated with alkaline conditions (0.1M NaOH) and desalted over cation exchange resin. The samples were passed through an in-line graphitised carbon cartridge and were analysed by LC-MS-TOF in negative-ion mode (Figure 3-19). Measured with the low cone voltage the spectrum of the X22 sample still showed some glycan TFA adducts (data not shown) which disappeared after increasing the voltage. The major signal in the X22 sample of mass m/z 446.05 correlates with the negative-ion mass of a Fuc-GlcNAc-1-P disaccharide, while the signal at m/z 300.05 is the result of the loss results from this structure. It is not possible to definitively differentiate as to whether this loss has resulted from the mass spectrometry ionisation or reflects the native heterogeneity of the structures. The former interpretation is favoured, however, since only the disaccharide phosphate was seen in the HPAEC–PAD chromatogram of the eliminated wild-type glycans (Figure 3-17). In the spectrum also the masses of dimers of these saccharides were detected at m/z 893.13 (Fuc-GlcNAc-1-P)$_2$ and at m/z 747.10 the defucosylated dimer (Fuc-GlcNAc-1-P+GlcNAc-1-P). Because of the high mass accuracy of an ESI-TOF mass spectrometer, these masses cannot be accounted for by larger oligosaccharide structures. If the glycosylation was composed of repeats of these structures then a mass of m/z 875.12, representing the structure Fuc-GlcNAc-1-P-Fuc-GlcNAc-1-P, would result. This mass was not observed.

The major signal in the HU2470 sample of mass m/z 300.10 correlates with the negative ion mass of a GlcNAc-1-P saccharide, while the signal at m/z 601.13 is the result of dimers of these saccharides (Figure 3-19). Also detected here is its sodium adduct with a mass of m/z 623.16. The spectrum of HU2471 showed the same signals as seen in HU2470 (data not shown).

Although the LC-MS method is much more sensitive than HPAEC-PAD, it could not give clear evidence for the existence of Fuc-1-P (m/z 243.06) in the strains. The tiny signal at m/z 243 in spectrum X22 (arrow in Figure 3-19) may be from Fuc-1-P, but this did not appear in the 30 V spectrum under the conditions used to analyse the Fuc-1-P standard. This result could mean that, if this signal is from the glycan Fuc-1-P, it is only a minor modification on the X22-SP96. There is no evidence that it is a modification in the mutants. Indeed, both the chromatographic and mass spectrometric data support the conclusion that the major modification in the wild–type SP96 is the phosphorylated disaccharide Fuc-GlcNAc-1-

P. The mutants appear to have the terminal fucose missing and only a GlcNAc-1-P is released.

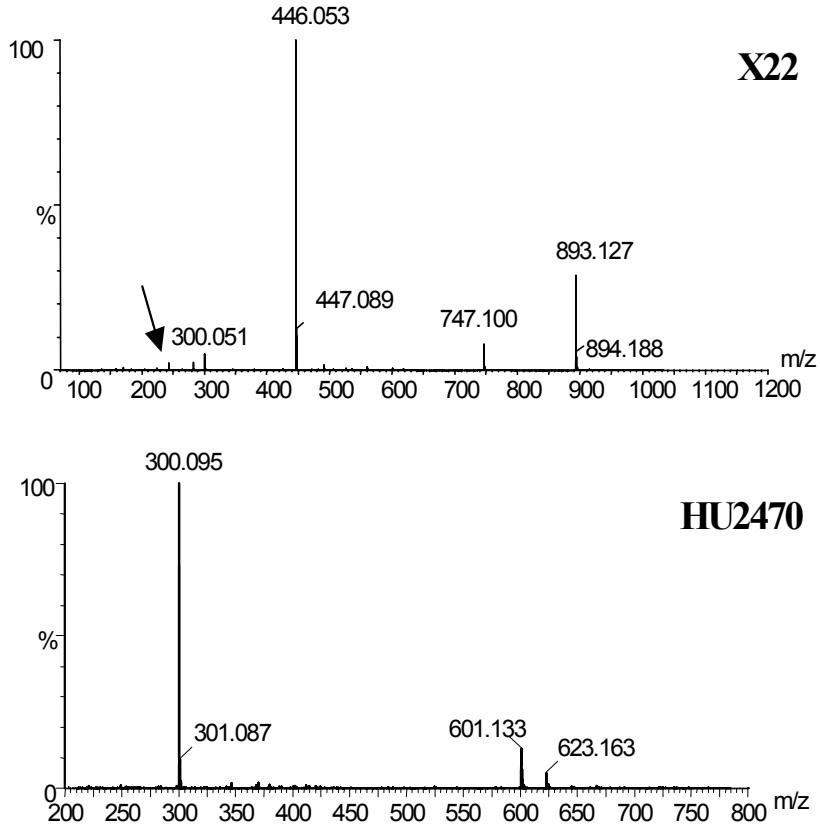



**Figure 3-19 Negative ion LC-ESI-TOF mass analysis of released glycans of X22-SP96 and HU2470-SP96**

The β-eliminated released glycans (100pmol) were desalted on an in-line graphitised carbon-cartridge and eluted directly into an ESI-TOF mass spectrometer and detected in negative-ion mode. The mass spectra of SP96 from X22 and HU2470 were obtained by in-source ionisation with a cone voltage of 80 V(X22) and 30 V (HU2470), respectively.

**X22)** The major ion at m/z 446.05 corresponds to the mass of Fuc-GlcNAc-1-P. A loss of fucose generates the ion m/z 300.05 which is identical to the mass of GlcNAc-1-P. The mass at m/z 893.13 corresponds to a dimer of the Fuc-GlcNAc-1-P and m/z 747.10 to the loss of fucose from that dimer.

The arrow → labelled the possible Fuc-1-P mass (m/z 243.06).

**HU2470)** The major ion at m/z 300.10 corresponds to the mass of GlcNAc-1-P. Also detected were the dimer of this saccharide m/z 601.13 and its sodium adduct m/z 623.16, respectively.

### 3.3.6 Determination of disaccharide phosphate linkages

To determine the complete structure of X22-SP96 glycan P1, the linkage of Fuc to GlcNAc and the anomeric configuration (α or β) had to be elucidated.

To answer these questions different approaches are available. A common wet-chemical method to analyse linkages between sugars is methylation analysis by GC-EI-MS. With the characteristic fragmentation pattern it is possible to infer the linkage of the sugars. For the identification of the anomeric configuration, enzymatic approaches with exo- and endoglycosidases are often used. If available it is also possible to use immediately a highly linkage- and anomeric- specific enzyme. The advantage is here the low amount of glycan needed for analysis. For glycan P1, fucosidases are necessary which are only available in α–1,2; α–1,6 and α–1→(3,4) specificity. A simple periodate cleavage of the glycan P1, used as a pre-experiment, should help to restrict the choice of enzymes.

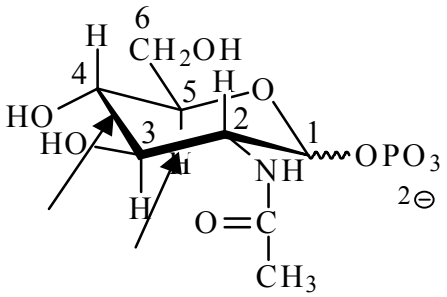
Another approach for the analysis of complete glycan structures is NMR (**n**uclear **m**agnetic **r**esonance) spectroscopy, which is also able to answer all these questions of linkage and anomericity. The development of high resolution NMR (>500 MHz) and two-dimensional (2D) techniques make it possible to resolve complex spectra. The only disadvantage of these techniques is the high amount of purified sample needed. Typically, a 0.5-5 mM solution (1ml) of purified sample is required to conduct 500 MHz $^{1}$H NMR experiments in the Fourier-transform (FT) mode in 2D in a reasonable time.

#### 3.3.6.1 Periodate treatment of glycan P1

Periodate is commonly used for oxidative cleavage of the carbon–carbon bond between vicinal diols. To a smaller extent the cleavage between a hydroxy group and amino alcohols and their derivatives is possible (83). A periodate treatment of a glycan chain would therefore destroy all sugars with vicinal OH groups. Only sugars which generally have or necessitate, because of the linkage to another sugar, no vicinal diols can be detected by monosaccharide analysis. This result excludes some of the sugar linkages. This is helpful to finally restrict the choice of fucosidase.

The glycan was oxidised with periodate, desalted over a carbon column and hydrolysed to release any unoxidised monosaccharide. Terminal fucose would be completely cleaved by this treatment and could not be detected. The fate of the GlcNAc, however,

depended on the position of the attached fucose. Only in case of a fucose linkage to the 6 position of the GlcNAc could it be cleaved between the vicinal diols of positions 3 and 4 (Figure 3-20) and hence, GlcNAc could not be detected anymore by monosacharide analysis. A fucose linkage to the 4 position would result in a slower cleavage of the carbon-carbon bond between the hydroxy group (ring position 3) and the amino alcohol group on position 2. If the glucosamine is still detectable, a linkage to either the 3 or 4 position of GlcNAc is possible. In this case further resolution using an enzymatic approach can be dismissed, because no enzymes are available which differentiate between these linkage positions.




D-*N*-acetylglucosamine-1-phosphate

**Figure 3-20 Periodate cleavage sites on GlcNAc-1-P**

The released glycan P1 was used for periodate cleavage, as well as the same amount of GlcNAc-α-1-P standard used in a control experiment. The result showed complete cleavage of fucose in the periodate treated sample P1 (**B**) compare to the untreated one (**A**) which had a large fucose signal at around 5 min (**1**) (Figure 3-21). The glucosamine signal (**2**) at around 12 min was, however, visible in both chromatograms. The periodate treated standard GlcNAc-α-1-P sample was completely cleaved after this time (data not shown). The lower amount of GlcNAc in sample **B** compared to **A** could be probably explained by to the loss during the carbon-column desalting step, which does not retain monosaccharides quantitatively.

The conclusion drawn from this result was that a 1,6 linkage between Fucose and GlcNAc could be excluded. The last two possible linkages, (1,3 and 1,4), cannot be differentiated by an enzymatic approach, as specific fucosidases for these linkages are not available. Therefore the NMR approach was chosen to determine the linkages.

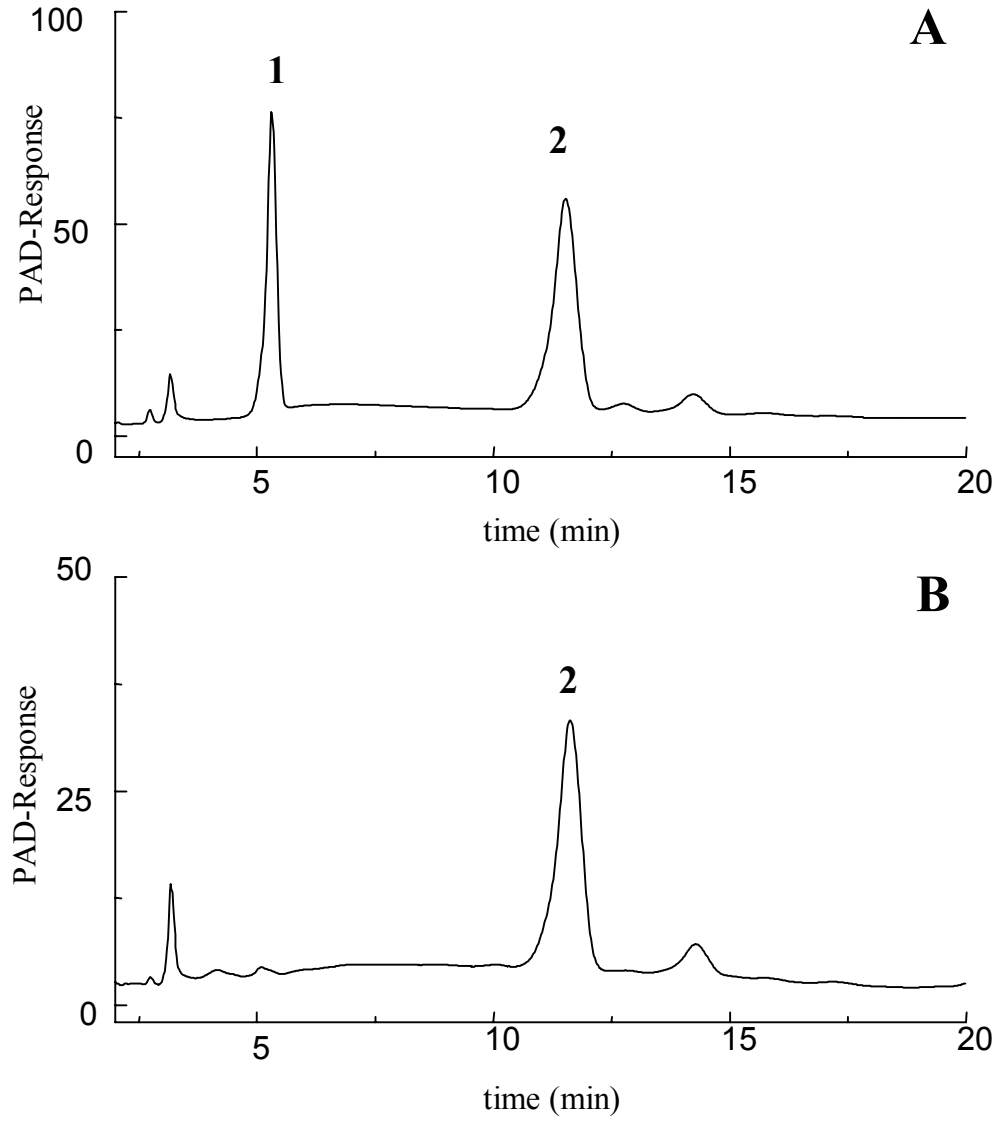



**Figure 3-21 Monosaccharide analysis of periodate treated and untreated glycan P1**
**A)** untreated, hydrolysed P1; **B)** periodate treated, hydrolysed P1

**1)** fucose; **2)** glucosamine

### 3.3.6.2 NMR analysis of glycan P1

The glycan was released from X22-SP96 in liquid sample and from PVDF-membrane blotted bands by β-elimination, before it was further purified twice by HPAEC-PAD chromatography and desalted over a graphitised carbon column. The amount of glycan P1 isolated for the NMR analysis was approximately 120 µg.

#### 3.3.6.2.1 $^{1}$H-NMR

The first NMR experiment performed was a one-dimensional $^{1}$H-NMR, because of its higher sensitivity compared to the two-dimensional experiments. Useful information can already be obtained by this, because of the so called "structural reporter groups", specific types of protons which resonate in characteristic regions of the spectrum. From the Fuc-GlcNAc-1-P glycan the following protons belong to it: anomeric protons from the fucose and from the *N*-acetylglucosamine, the fucose H-5 and $CH_3$ protons and the aminosugar *N*-acetyl $CH_3$ protons. The anomeric protons show generally a chemical shift between 4.5 and 5.5 ppm, whereas the protons of α–anomers are more deep-field shifted than those of the the β type. The observed coupling constant for a $^{3}J_{H1,H2}$ coupling in the α–anomers is ~2-4 Hz, whereas the corresponding β-anomers show coupling values from 7-9 Hz for $^{3}J_{H1,H2}$, because of the larger dihedral angle of ~180 °. Thus, the anomeric configuration of a glycosyl residue can be inferred both from the chemical shift and the coupling constant of its anomeric proton signal in the $^{1}$H-NMR spectrum.

The significant signal overlap in the region of ~3.5-4.2 ppm originate from the so called "sugar skeleton protons". These protons often create non-first-order spectra, which make it very difficult to locate the resonance of individual nuclei and to assign these resonances to specific nuclei in the oligosaccharide structure. To interpret this part of the spectrum two-dimensional (2D) NMR experiments are helpful.

The spectrum of phospho-glycan P1 (Figure 3-22) revealed an anomeric proton with a chemical shift $\delta_H$ 5.37 indicative of α-anomeric sugars. For GlcNAc, H1 was assigned on the basis of the coupling to H2 and phosphorus ($^{3}J_{HP}$=6.9 Hz ). In addition, the small coupling constant to GlcNAcH2 (*J*=3.3Hz**)** is not consistent with an axial proton at H1 where a much larger coupling would be expected. Similarly, a doublet of the fucose anomeric proton with a chemical shift ($\delta_H$ 5.03) was coupled to Fuc H2 (*J* = 4.1Hz) indicating that fucose was also in an α-configuration. Further assignment was possible for the singlet signal of the *N*-acetyl $CH_3$

protons of the aminosugar at ($\delta_H$ 2.05) and the doublet of the fucose $CH_3$ protons at $\delta_H$ 1.17 (out of range of the spectrum shown in Figure 3-22).

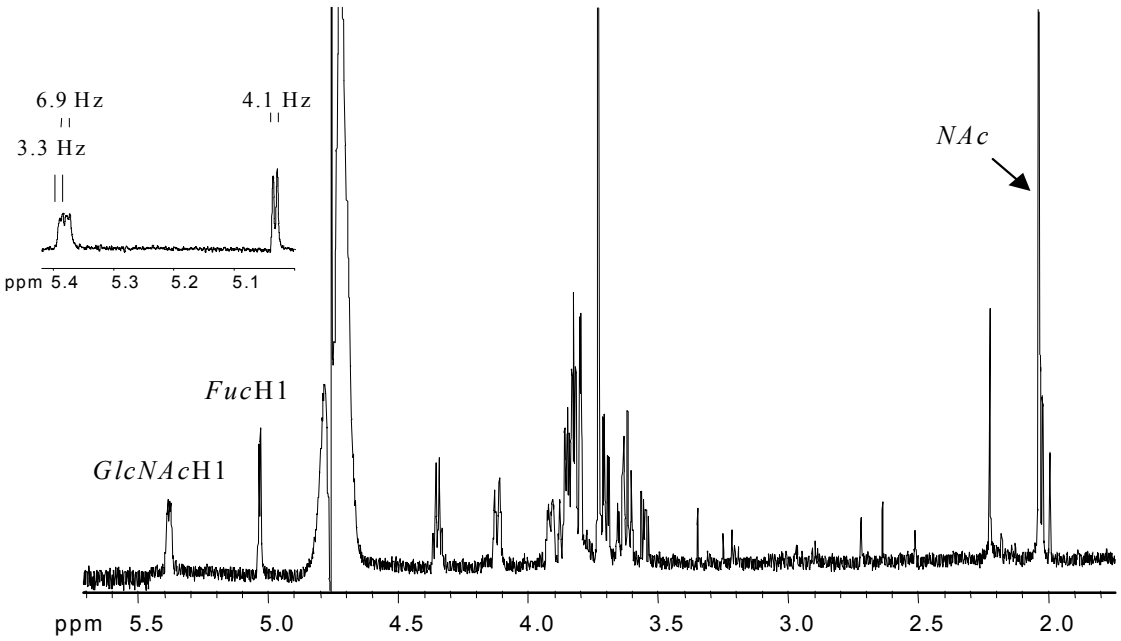



**Figure 3-22 Part of $^1$H-NMR spectrum of purified X22-SP96 glycan P1**
Zoomed location of anomeric protons displayed between 5-5.4 ppm and the not resolved skeleton protons between 3.5 –4-5 ppm.
(120 µg glycan P1 repeatedly lyophilised and redissolved in $D_2O$)

3.3.6.2.2 2D-NMR (COSY, TOCSY and HSQC)

After characterisation of the α-anomeric configuration, the linkage attachment of Fuc to GlcNAc was, however, still unknown. The linkage could only be identified by the assignment of all signals in the spectrum. The strategy for obtaining sequential assignments in $^1$H spectra of carbohydrates basically involves two different types of NMR experiments. The various glycosyl spin systems of a particular carbohydrate are identified by a combination of 2D direct (COSY; *Co*rrelated *S*pectroscop*y*) and relayed (TOCSY; *To*tal *C*orrelation

*S*pectroscop*y*) *J*-correlated spectroscopy. Because of the consecutive allocation of protons around each glycosyl ring, a vicinal *J*-coupled connectivity exists for each residue such that each proton leads to the next proton of the ring. Conveniently, the anomeric protons (H1) of each aldosyl residue resonate in a characteristic region ($4.5<\delta<5.5$ ppm) that contains few other signals. Thus the *J*-connectivity trail for such residues can be initiated in this region. The actual magnitudes of the *J*-coupling reveal the stereochemistry and configuration at each carbon; the complete set of *J*-couplings characterizes the identity of the residue. The TOCSY technique normally permits subspectral editing of the $^{1}$H spectrum for each constituting glycosyl residue and, consequently, the virtually complete assignment of all multiple patterns in the $^{1}$H spectrum.

Another supplementary NMR technique is the heteronuclear 2D NMR spectroscopy. More recently, the sensitivity problem in the detection of $^{1}J$-connectivities between $^{13}$C and $^{1}$H nuclei has been overcome to a large extent by the introduction of the so-called *reversed* i.e. $^{1}$H-detected 2D {$^{13}$C,$^{1}$H} shift-correlation experiment termed HSQC (*H*eteronuclear *S*ingle *Q*uantum *C*oherence). From these experiments a complete {$^{13}$C,$^{1}$H} one-bond shift-correlation map, with a high resolution in the $^{1}$H dimension, could be obtained from the glycan.

To identify the exact linkage between Fuc and GlcNAc in the X22-SP96 phosphoglycan P1, the full $^{13}$C and $^{1}$H assignment of the disaccharide is required. This was achieved using the techniques mentioned above (HSQC, Figure 3-23; TOCSY, Figure 3-24 and COSY Figure 3-25).

Assignment of the Fuc signals was aided by the HSQC (Figure 3-23) and the COSY (Figure 3-25, Table 3-2) spectra. It was found, that in particular, H6 was coupled to H5 ($\delta_H$ 4.36). No correlation was observed in the COSY spectrum between H5 and H4, but the TOCSY spectrum (Figure 3-24) showed a strong correlation between H1 and H2/H3 and a weak correlation to H4 of Fuc. For GlcNAc, H1 was assigned on the basis of the coupling to H2 and phosphorus ($^{3}J_{HP}$=6.9Hz). The HSQC (Figure 3-23**)** was used to help assign the remaining signals. $(H6)_2$ was evident as the only correlation of two protons to the same carbon ($\delta_c$ 61.7). Both C6 protons were coupled to H5 (COSY; $\delta_H$ 3.96) which was in turn coupled to H4 ($\delta_H$ 3.59). The TOCSY spectrum (Figure 3-24) showed correlation between H1 to H2 and H3 and a weak correlation to H4 to complete the assignments. The chemical shifts of Fuc were similar to those previously reported (81), while those of GlcNAc were affected by the α-1-phosphate. In particular, C2 and C5 were shifted by ~2ppm upfield by the

shielding effect of the phosphate. This was also observed in the GlcNAc-α-1-P standard (Table 3-2). Comparisons of the C3, C4 and C5 chemical shifts with those of similar disaccharides (81) indicated that substitution was at C3 (Table 3-2). The chemical shifts of C4 ($\delta_c$ 69.5) and C3 ($\delta_c$ 79.3.) of GlcNAc were similar to those of a α1-3-linkage ($\delta_c$ 69.7, $\delta_c$ 83.4 respectively).

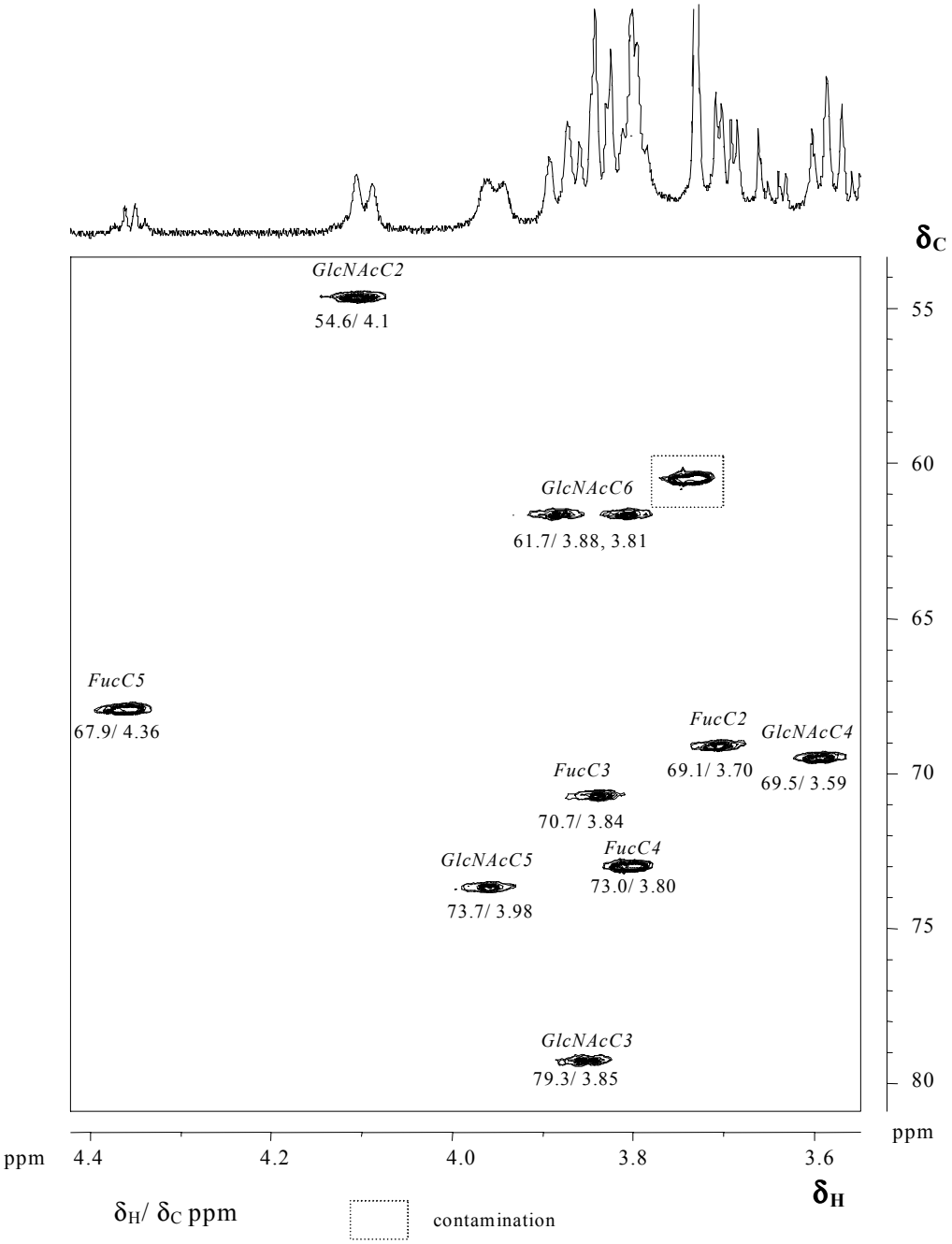



**Figure 3-23 HSQC spectrum of glycan P1 of skeleton protons and carbons**
Sensitivity-enhanced double INEPT (insensitive nuclei enhanced by polarization) transfer was used to correlate carbons and protons in the enonanomeric region, to help assign all the signals of the disaccharide. The *boxed* signal is an impurity in the sample.

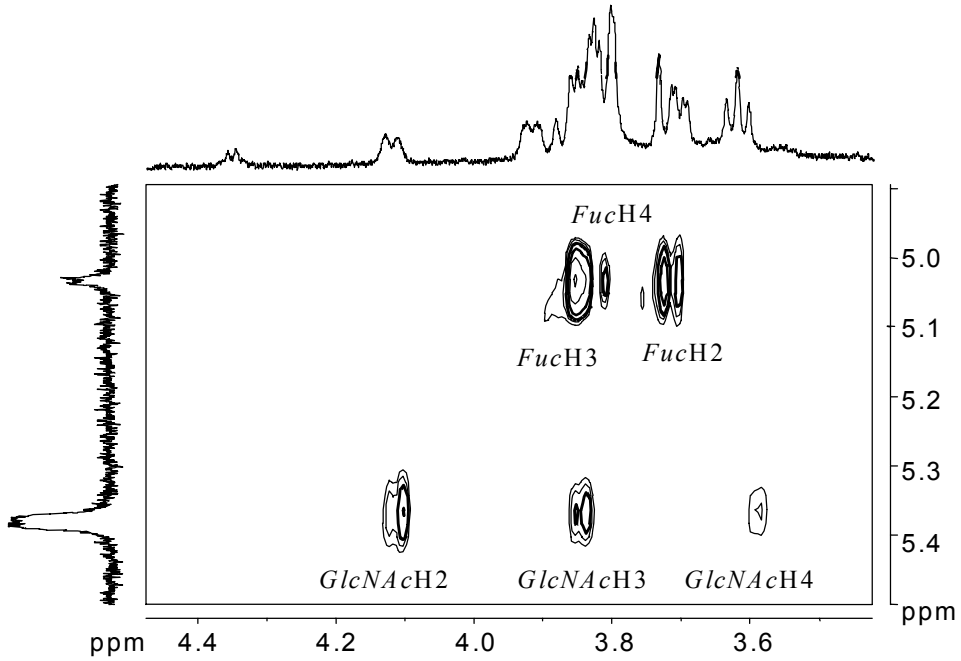



**Figure 3-24 TOCSY spectrum of skeleton protons of glycan P1**
Correlation between the two anomeric protons with the next protons (H-2 and H-3) are evident. Weaker correlations to H-4 can also be observed.

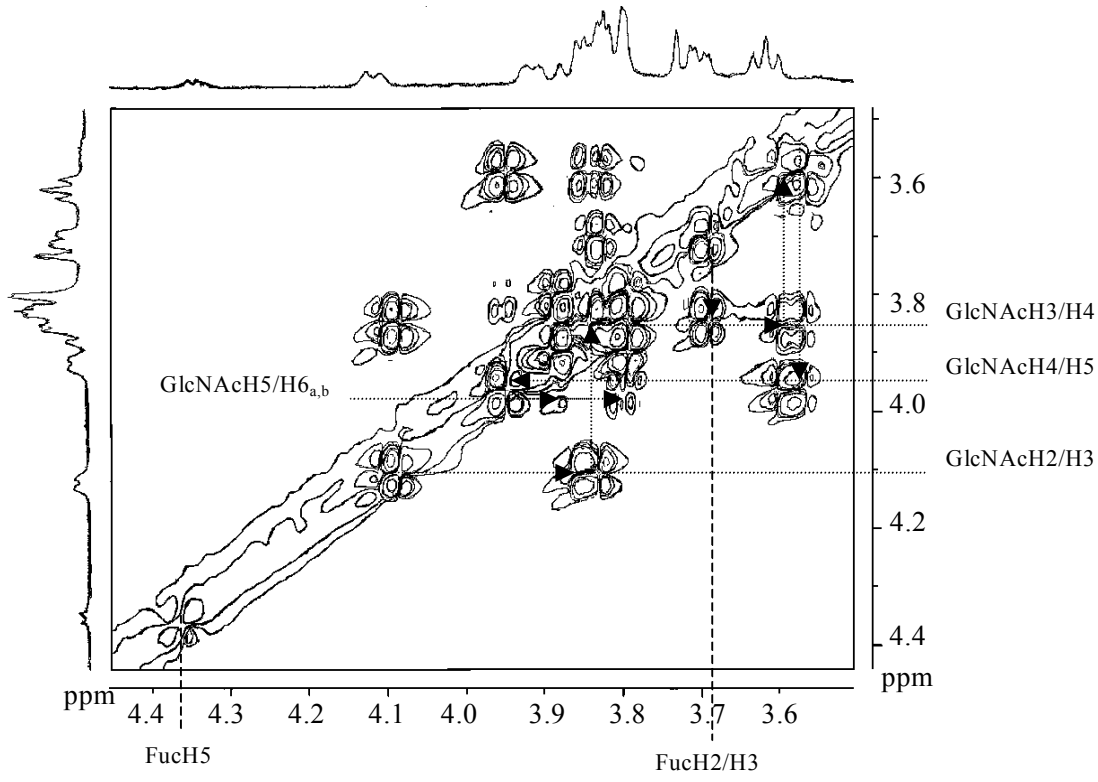



**Figure 3-25 COSY spectrum of skeleton protons of glycan P1**
arrows illustrate the coupled-neighbour protons in GlcNAc (dotted line) and some of Fuc (broken line)

| | (P1) | | GlcNAc[a] | GlcNAc[b] | β-Fuc | Fuc(α-1) | GlcNAc(α-1)P | |
|---|---|---|---|---|---|---|---|---|
| *Assign.* | $^{13}C$ | $^{1}H$ | $^{13}C$ | $^{13}C$ | $^{13}C$ | $^{13}C$ | $^{13}C$ | $^{1}H$ |
| *FucC6* | 16.2 | 1.17 | | | 16.5 | 16.3 | | |
| *Ac* | 23.1 | 2.04 | | | | | 23.1 | 2.04 |
| *GlcNAcC2* | 54.6 | 4.10 | 55.5 | 56.9 | | | 55.2 | 3.92 |
| *GlcNAcC6* | 61.7 | 3.81, 3.88 | 61.7 | 61.0 | | | 61.8 | 3.77, 3.89 |
| *FucC5* | 67.9 | 4.36 | | | 71.9 | 67.9 | | |
| *FucC2* | 69.1 | 3.70 | | | 71.5 | 68.9 | | |
| *GlcNAcC4* | 69.5 | 3.59 | 69.7 | 78.5 | | | 71.1 | 3.49 |
| *FucC3* | 70.7 | 3.84 | | | 74.1 | 70.3 | | |
| *FucC4* | 73.0 | 3.80 | | | 72.4 | 72.7 | | |
| *GlcNAcC5* | 73.7 | 3.96 | 76.2 | 76.1 | | | 73.1 | 3.97 |
| *GlcNAcC3* | 79.3 | 3.85 | 83.4 | 73.5 | | | 72.6 | |
| *GlcNAcC1* | 94.4 | 5.37 | 101.1 | 101.8 | | | 93.7 | 5.36 |
| *FucC1* | 100.8 | 5.03 | | | 104.8 | 100.4 | | |

**Table 3-2 $^{13}C$ and $^{1}H$ chemical shifts of the isolated disaccharide (P1).**

Proton and carbon chemical shifts were referenced to the fucose methyl ($\delta_H$ 1.166, $\delta_C$ 16.2 ppm). For comparison, published $^{13}C$ shifts for residue linked to the [a] 3 and [b] 4 position of GlcNAc (81), α- and β-fucose (5, 81) α-1-GlcNAc-P (Sigma) are given.

3.3.6.2.3 1D-ROESY NMR

A one dimensional ROESY experiment of glycan P1 was chosen to support the interpretation of the 2D-NMR results seen above (3.3.6.2.2).

ROESY (rotating-frame nuclear Overhauser spectroscopy) experiments allow one to probe which protons are close in space to each other (less than 450 pm away). The experiment involves the selective irradiation of an anomeric proton for a given period, followed by signal detection in the absence of irradiation. In general, 1,3-diaxial and 1,2-eq-ax proton pairs in pyranosyl rings will produce intraresidue ROESY peaks. More importantly, however, ROESY peaks may be observed between interresidue proton pairs. Thus, in some cases, ROESY is valuable in determination and confirmation of linkage sites and O-glycosidic conformation in oligosacharides.

The α1-3-linkage was confirmed by the ROESY spectrum (Figure 3-26) which showed a strong interresidue correlation between Fuc H1 and GlcNAcH3 (4.5%). In addition, approximately equal (~1%) ROE's were observed between FucH1, GlcNAcH2 and GlcNAcH4. Also detected were the intraresidue correlations to FucH2 (3.8%).

These data established conclusively the substitution and stereochemistry of the disaccharide as Fuc(α1-3)GlcNAc-α-1-P.

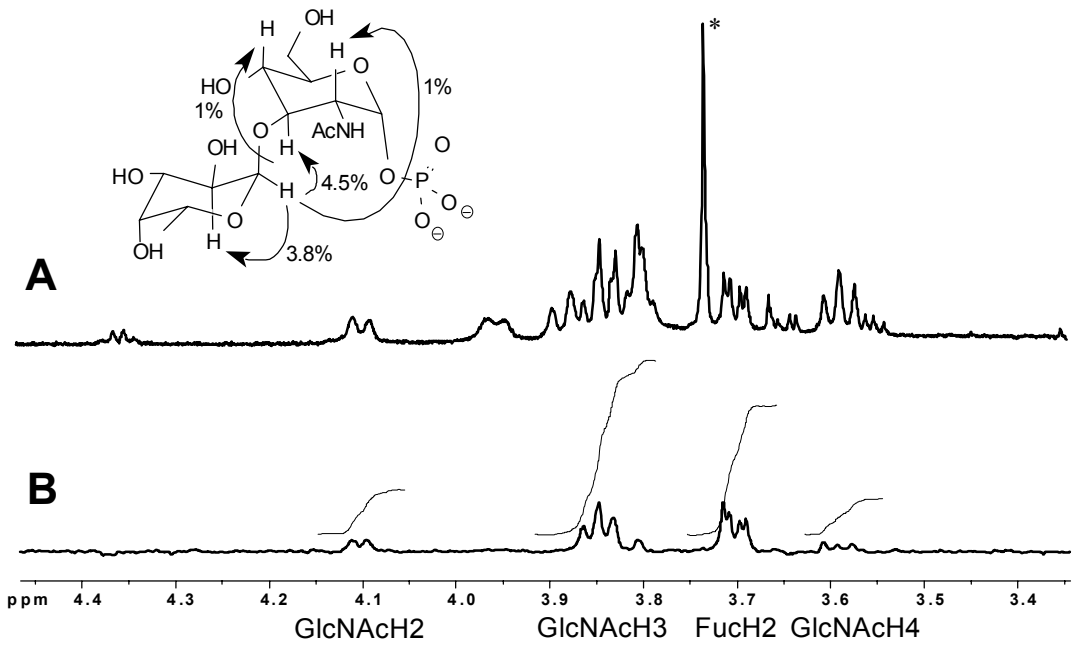



**Figure 3-26 One-dimensional ROESY spectrum of the released disaccharide P1**

Proton NMR spectra of Fuc(α1-3)GlcNAc-α-1-P showing the non-anomeric region. The normal WATERGATE spectrum (A) shows a quartet ($\delta_H$ 4.36) for FucH5. The singlet at (*; $\delta_H$ 3.74) is an unidentified impurity. Spectrum B is a one-dimensional ROESY obtained by selective irradiation of FucH1. Enhancements of GlcNAcH3 (4.5%) and GlcNAcH2/H4 (1%) indicate an α1-3 linkage.

## 3.4 Quantitative analysis of the post-translational modifications on X22-SP96 and HU2470-SP96

Previously, radiolabelling experiments (21, 109) indicated that SP96 may be modified by glycosylation and phosphorylation. Until now the protein has, however, not been purified far enough for quantitative analysis. This kind of information could be helpful to get a better understanding of the possible function of the modifications of this spore coat protein, for example, if a glycan is implicated in protein conformation or in conferring protease resistance to the spore. For the quantitative analysis only the samples of X22-SP96 and HU2470–SP96 were used, because of their high purity after affinity purification.

The quantification of the fucose by monosaccharide analysis showed that fucose was approximately four times higher in wild-type X22-SP96 compared to that in the fucose mutant HU2470-SP96 ( Table 3-3). The quantification of *N*-acetylglucosamine as glucosamine resulting from acid hydrolysis of the protein showed 21% more in the wild type ( Table 3-3). To quantify the degree of phosphorylation of SP96, the purified protein from both strains (20-100 pmol) was hydrolysed with sulfuric acid, and the released inorganic phosphate was measured by adapting a nanomole-sensitive colourimetric phosphate assay (80) to a microtitre plate format. By comparison with a standard curve, the calculated amount of released phosphate was ~17% higher from the X22-SP96 wild-type compared to that in the HU2470-SP96 mutant ( Table 3-3).

To determine whether the phosphate exists as a monoester (on either an amino acid or a sugar), the protein (0.2 nmol) was treated with alkaline phosphatase, which removes singly substituted phosphates, and the released phosphate was detected with the same assay. The limit of detection of the phosphate assay is 200 pmoles and no measurable released phosphate was detected in SP96 protein from both strains. Although the phosphate may be otherwise inaccessible to the enzyme, this suggests that less than 1 % of the total phosphate on the protein was not further modified.

The ratio of GlcNAc to phosphate (1:1) was the same in both the wild-type and the mutant. The slightly lower (approximately 20 %) molar content of each substituent in the mutant may be significant in the number of sites which are substituted or may be a function of the analysis involved as the values lie within the standard error range.

| | nmol/ nmol protein | | |
|---|---|---|---|
| Strain | **Fucose** | **GlcNAc** | **Phosphate** |
| X22 | 78±6 | 66±4 | 63±13 |
| HU2470 | 18±2 | 52± 4 | 54±9 |

**Table 3-3 Compositional analysis of SP96**

GC-MS analysis was also used for quantification of the component sugars. Sugars were released by reductive β-elimination from the purified SP96 protein (0.2-0.5 nmol) of the X22 and HU2470 strain and converted to a volatile form pre- **(a)** and post- **(b)** acid hydrolysis, before they were analysed by gas chromatography mass spectrometry (see 3.3.3). Quantification of the GC peaks was carried out relative to the internal standard myo-inositol. The compositions are shown in Table 3-4. As fucose is susceptible to the acid degradation used in the monosaccharide analysis, these values are necessarily less reliable, and the GC quantitative data (an average of 28 sites) are probably a more accurate estimation of the fucitol which is more stable to acid. A problem in the quantification of GlcNAc in the HU2470 sample was that it eluted close to a contamination peak, as proven by a comparison measured EI-MS spectra with database results. The quantification of the amount of the sugars in the HF treated samples (treatment c and d; 3.3.3) is not shown, as it was apparent that destruction of the monosaccharides was occurring over the long period of incubation in the HF acidic conditions, even at 0°C.

| | nmol/ nmol protein | | | | | |
|---|---|---|---|---|---|---|
| Treatment | **Fucitol acetate** | | **Fucose acetate** | | **GlcNAc acetate** | |
| Strain | X22 | HU2470 | X22 | HU2470 | X22 | HU2470 |
| **(a)** red. β-elimination | 28 | 29 | - | - | - | +* |
| **(b)** red. β-elimination/ acid hydrolysis | 33 | 23 | 56 | 2 | 56 | +* |

**Table 3-4 Quantification of β-eliminated sugars from SP96 by GC-MS**

*signal partly overlapped with a contaminating peak; could be identified as GlcNAc acetate but not quantified.

There are 121 serine and 83 threonine residues in unprocessed SP96. In the 20 amino acid signal sequence, which is cleaved from the mature protein, there is only one serine and one threonine. Therefore, 120 serine and 82 threonine residues have the potential to be O-glycosylated.

The results show that, on average 60 GlcNAc were attached to serine via phosphodiester bonds, which is the only phospho-amino acid in SP96 (see 3.3.1) with no further phosphate on non reducing positions of the sugars. This conclusion was also supported by the 1:1 ratio between GlcNAc and phosphate in both strains ( Table **3-3**). Therefore, about 50 % of the serines are modified by this glycan. As the remaining O-linked fucose is also attached to serine (see 3.5.2), approximately 70 % of this amino acid is modified in total.

As shown by ES-MS analysis (see 3.3.5), the only apparent difference between the wild-type-SP96 and HU2470-SP96 modifications is a missing terminal fucose on the phosphodiester-linked GlcNAc. The total mass of approximately 60 missing fucose residues would be 8.8 kDa. Although migration of glycoproteins on SDS-PAGE is not always proportional to the mass of the glycan, this result is consistent with the predicted mass difference after SDS-PAGE between X22-SP96 and HU2470-SP96 of approximately 10 kDa, seen on the MUD3 western blot (see 3.1).

## 3.5 Location of glycosylation sites in SP96

### 3.5.1 Strategy

After quantification of the post-translational modifications it was now of interest to locate the glycosylation sites in the SP96 protein. A closer look at the 580 amino acid (AA) long sequence (1.3.5) displayed that, as already mentioned earlier, SP96 has many potential *O*-glycosylation sites due to its 120 (20 %) serine and 82 (14 %) threonine residues. Also interesting was that 73 serine and 47 threonine residues were contained in the last 180 amino acids on the C-terminus, resulting in a serine- and a threonine-rich repeat region. A further serine/ threonine mixed region combined with glycine is further upstream in the sequence between amino acids 140-180. Ser/ Thr rich domains are quite often targets for substantial *O*-glycosylation. They are found on mammalian cell surface glycoproteins, like mucins, and immunoglobulins (127). With at least 60 phosphoglycans linked to serine it was therefore assumed that part of the serine-rich tail of SP96 would also be glycosylated. The hydroxy amino acid to which fucose was directly *O*-linked could be either serine or threonine as both these can be β-eliminated to release fucose.

One approach commonly used for characterisation of heavily glycosylated sites in proteins such as mucins, is digestion with an unspecific protease. Highly glycosylated regions resistant to this digestion can be separated in gel and used for further analysis. Glycosylated fragments of SP96 were separated by SDS-PAGE and screened for the MUD3 or MUD62 epitope, before being used for further analysis. Another approach to identify potentially glycosylated sites is to use the established *D. discoideum in vivo* expression system for the identification of O-glycosylation acceptor sites (24, 70, 71). In the case of SP96 three different motifs out of the repeat regions of the protein were examined for their acceptance of *O*-glycosylation.

### 3.5.2 Amino acid analysis pre and post β-elimination

The phosphodiester linked glycans were definitively assigned to serine, yet the hydroxy amino acid to which fucose was O-linked was still not identified.

β-elimination releases O-linked glycans from proteins while, at the same time, converts the modified hydroxy amino acid into its dehydro-form. For example, serine is converted to dehydroalanine (see Figure 3-12) and threonine to dehydroamino-2-butyric acid. The consequence of this conversion is a different retention time of the original amino acid compared to the dehydro amino acid during amino acid analysis. Therefore, the modified hydroxyamino acids can be quantified by the decrease in the amount of detected hydroxyamino acid pre- and post β-elimination.

Purified samples of X22-SP96 and HU2470-SP96 were aliquoted (10-20 pmol each) and at least four determinations were done after β-elimination and hydrolysis (post β-elimination) as well as after hydrolysis only (pre β-elimination) with 6 M HCl. The results pre and post β–elimination were normalised to isoleucine and shown in Figure 3-27.

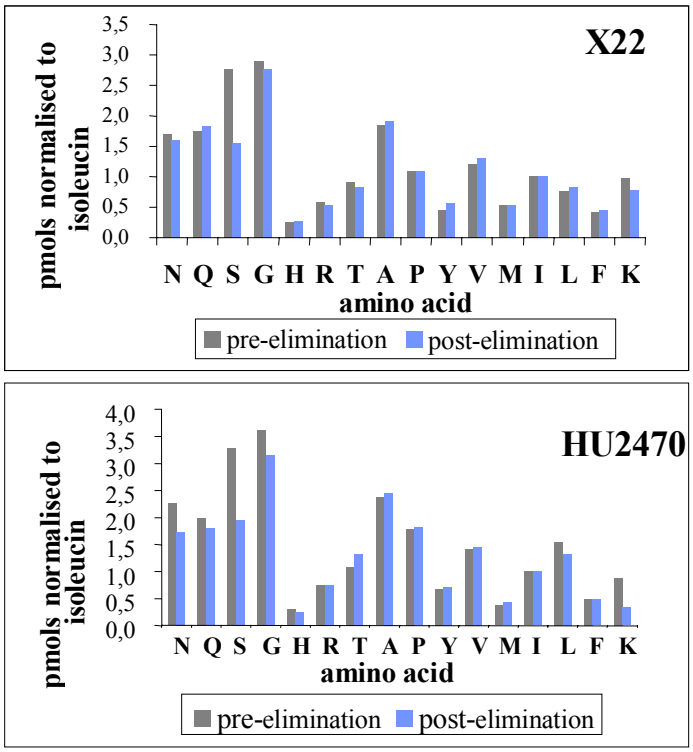



**Figure 3-27**

**Amino acid analysis of X22-SP96 and HU2470-SP96 pre and post β-elimination**

Amino acids are displayed by a one letter code

In both the wild type and the mutant, the amino acid ratio after normalisation to isoleucine showed only a drop in the amount of serine. Although this result would not exclude the possibility of a small amount of modified threonine, any difference is within the error of the analysis. This shows that serine is probably the only modified hydroxyamino acid in SP96 and that, therefore, up to 70 % of the serines are modified by glycosylation, phosphorylation or phosphoglycosylation.

### 3.5.3 Non-specific protease digest of SP96

The non-specific proteases Pretaq™ and Proteinase K were used to digest SP96 so that the glycosylated, more resistant part of the protein could be further analysed. These proteases are common used in glycobiology for analysis of glycopeptides. The Pretaq™ enzyme is an extracellular alkaline serine protease which was isolated out of hot spring bacteria *Thermus* sp. strain Rt41A and was originally used to remove proteins during the DNA isolation procedure. This aggressive protease is also stable in the presence of denaturing reagents and has its optimum activity at pH 8 and 90°C. The other protein used, Proteinase K, is also a serine protease which has a good activity at room temperature and pH 8.

X22-SP96 protein was at first incubated for different times with a fixed amount of Pretaq™ and Proteinase K . After 20 min, in both protease samples no uncleaved SP96 could be detected with MUD3. A longer incubation time resulted in a complete cleavage and the loss of the MUD3 and MUD62 epitope. The sample preparation for SDS-PAGE of the Pretaq™ digest was done without boiling, because of the high activity of the enzyme under reduced conditions and at high temperatures. Although the products of the digestion were not uniformly reproducible, several bands often appeared at the same mass. A digest with a decreased amount of protease (20 mU) and a longer incubation time was also not able to yield a better reproducibility of the results.

In both samples the smallest MUD3 detected fragment was around 55 kDa. A weak labelling of fragments at 62 kDa and 92 kDa was also observed. The difference between the Pretaq™ and Proteinase K digest was that in the latter one the two larger fragments were labelled more strongly by MUD3 compared to the band at 55 kDa. The progress of the digest was done after SDS-PAGE by western blot analysis with MUD3, silverstain and an immuno-glycostain (Figure 3-28). All the MUD3 detected fragments showed also the MUD62

epitope (data not shown). The band at 55 kDa of the SDS-PAGE separated sample was used for a western blot to a PVDF membrane and was chosen for analysis by Edman degradation.

The glycopeptide was sequenced up to 9 cycles from the N-terminus. As expected, the interpretation of the results was difficult, because of the multiple starting sites which resulted from the use of an unspecific protease. Nevertheless, there was a possible sequence homology to the sequence between amino acid 359-368 of SP96 (Table 3-5). This result showed that the more digestion-resistant part of SP96 started around 60 amino acids upstream of the serine / threonine-rich tail of SP96. This part of the protein appeared to be modified by O-linked fucose and the phosphodiester linked Fuc(α1-3)GlcNAc-α-1-P disaccharide, because of the positive MUD3 and the MUD62 epitope reactivity, respectively.

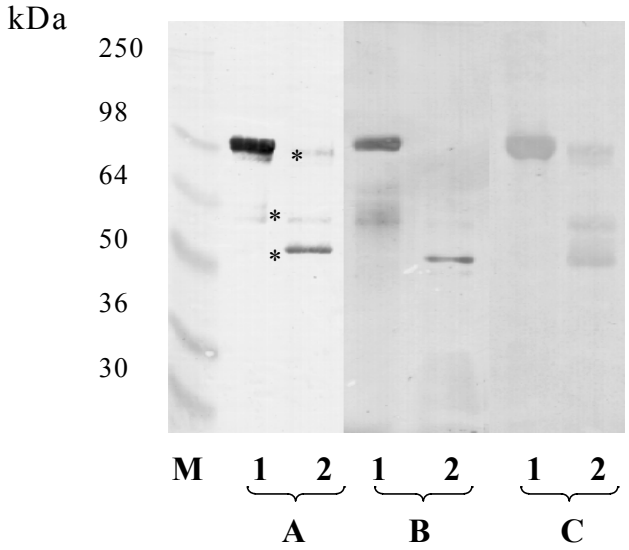



**Figure 3-28**

**X22-SP96 digestion with the unspecific protease Pretaq™**

**1**) undigested ; **2**) digested
**A**) immunoblot MUD3; **B**) silver stain;
**C**) immuno glycostain
*) labeled fragments at approximately 94, 62 and 55 kDa.

| N-terminal sequence of SP96 fragment (55 kDa) |
|---|
| **DVQPTCIKP** |

**Table 3-5**
**N-terminal sequence of the first 9 amino acids of the 55 kDa fragment obtained by Pretaq™ digest of SP96**

### 3.5.4 An *in vivo* system for the identification of *O*-glycosylation acceptor sites in *D. discoideum*

An alternative to the identification of *O*-glycosylation sites in original glycoproteins is the use of short-candidate acceptor peptides for *in vivo* or *in vitro* studies. In *in vivo* studies different potential acceptor peptides are expressed and the glycopeptides characterised. Usually short synthetic peptides (10-20 amino acids) are used in combination with diverse glycosyl-transferases (99, 119, 134, 135).

To analyse regions of SP96 for *O*-glycosylation acceptor sites advantage was taken of *D. discoideum in vivo* expression system (24, 71). For glutathione-S-transferase (GST) fusion proteins, the C-terminus of GST is fused with the candidate acceptor peptides via a factor Xa-specific endonuclease cleavage site. The *D. discoideum* cells secrete the GST-peptide fusion protein in the liquid media from which it can be purified by affinity chromatography using glutathione agarose.

The expression system utilized two different plasmids, one integrating vector pMUW110 (9 kb) and one extrachromosomal expression vector pMUW2882 (3.8 kb)(Figure 3-29) and maintains 100 copies of extrachromosomal expression plasmids per cell (13).

The integrating vector contains a neomycin phosphotransferase (Neo) gene conferring resistance to geneticin (G418) and the Ddp2 Rep gene. On co-transformation of the two plasmids into *D. discoideum* and selection with G418, pMUW110 integrates into the chromosome and the product of the Rep gene can be expressed. The Ddp2 rep gene acts in *trans* on the Ddp2 origin of replication to ensure the autonomous replication of pMUW2882 in the nucleus of *D. discoideum*.

The expression vector contains the *D. discoideum* plasmid Ddp2 origin of replication and an expression cassette, consisting of an actin15 promoter, the secretion signal of the glycoprotein PsA fused in frame to the *Schistosoma japonicum* glutathione-S-transferase (sj 26) gene elongated with the 12 bp stretch of the factor Xa –cleavage site followed by a unique cloning site (*Bam*HI) and a polyadenylation signal.

It should be mentioned that native spore coat protein SP96 is expressed in a later developmental stage compared to the peptide fusion protein which is expressed in this system under the control of the actin 15 promoter. Therefore, differences in glycosylation are possible. For example, SP29 (PsA) is originally also expressed at a later developmental stage like SP96, and glycosylation analysis of acceptor motifs analysed with this expression system show less complex glycosylation than identified on native PsA (61, 142, 155). The target of

the protein is also different as the GST-fusion protein is secreted while authentic SP96 is targeted to prespore vesicles.

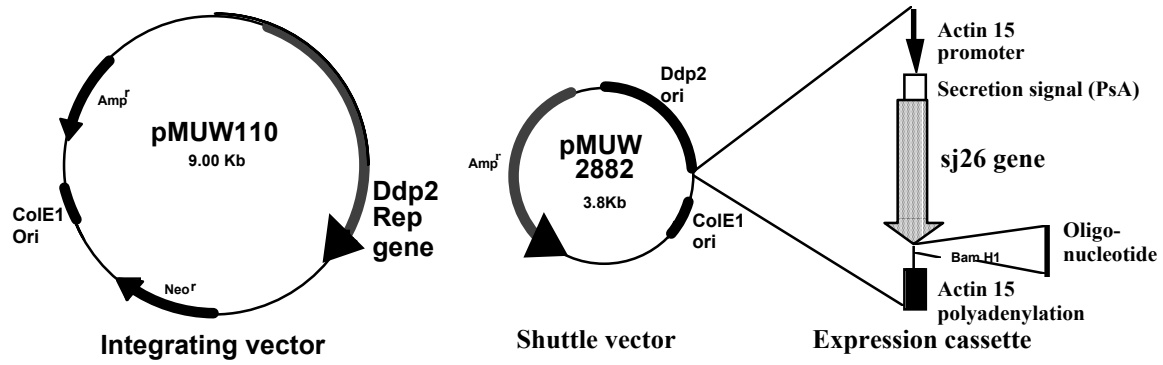



**Figure 3-29 Two vector expression system**

### 3.5.4.1 Three repeat regions in SP96 as possible motifs for *O*-glycosylation

In SP96 three serine and/ or threonine rich repeat regions stand out (3.5.1). These are the serine/ threonine mixed region AA 140-180 (with 13 Ser and 11 Thr ), a serine rich region at AA 425-535 (73 Ser) and a threonine rich sequence at AA 535-600 (42 Thr). Although it was already shown that serine was probably the only modified amino acid in SP96, a similar threonine rich sequence was also chosen for comparison. Three 12-17 amino acid long sequences were chosen as the possible motif for O-glycosylation (Table 3-6). Each peptide also contained a C-terminal I-SPY antibody epitope (QYPALT) for immunochemical detection of the peptides (I-SPY™, AMRAD BIOTECH).

| motif | amino acid area in SP96 | theoretical mass | amino acid sequence |
|---|---|---|---|
| SP96G-motif | 146-163 | 2041 | **AGSQTSGGSTGGSTGS**QYPALT |
| SP96T-motif | 550-562 | 1767 | **ATTTATTATTTA**QYPALT |
| SP96S-motif | 500-514 | 1855 | **ASSSSAPSSSASSS**QYPALT |

**Table 3-6 Chosen amino acid sequences of SP96 for analysis of *O*-glycosylation sites**

Bold letters display the sequence of SP96, normal letters display the sequence-tag (I-SPY™)

### 3.5.4.2 Construction of the expression vectors and their insertion in *D. discoideum*

By exchanging existing motifs against new ones, several generations of expression vectors containing different *O*-glycosylation acceptor motifs were available. In the vectors used by Jung *et al*. (70, 71) a small section of DNA encoding the peptides is inserted between NcoI/ KpnI close to the factor Xa cleavage sites. However, these vectors were subsequently formed by further mutations in the GST gene (M. Slade, personal communication). A later generation of vectors corrected these mutations, improved the codon usage and the cloning efficiency in *E. coli* by replacing parts of the gyrase codon. This second generation of vectors lack convenient cloning sites for replacing the fusion peptide, since it has been observed that vector cloning in AT rich sequences in *D. discoideum* was less stable than that by PCR reactions (M. Slade, personal communication). To have a better control over the exchange of the old against the new motif and to be able to correct one mutation on the pMUW2882 plasmid in the GST gene (mutation methionine to isoleucine), the reconstruction of the plasmid was done by three separate polymerase chain reactions (PCR) . The exact oligonucleotide sequences of the primers are listed in chapter 2.6.2.

The first PCR step was done by using three different oligonucleotides encoding for the SP96G, SP96T and the SP96S-motif ("oligo primer1-3"), and a primer which overlaps with part of the GST gene (Primer A) to correct the mutation (methionine to isoleucine) (Figure 3-30, PCR 1). The "oligo primers1-3" coding the SP96 sequences showed an overlap on the 3´ site with a factor X cleavage site which exists on plasmid pMUW2882. The 5´ site of the SP96 sequences flanks an epitope sequence of a commercially available antibody (I-SPY sequence tag). The first PCR product gave a band at approximately 470 bp. The DNA was extracted and used again in a later stage as a primer (PCR 3). The second PCR was to amplify the whole plasmid and was produced with the "Primer B" and "Primer C" (Figure 3-30, PCR 2). The result showed a 3.8 kb band on the 1% agarose gel which was also extracted out of the gel. The last PCR was done with three primers (PCR 1 product of the different motifs, Primer B and Primer D) and replaced the old *O*-glycosylation acceptor motif against the new, SP96, motif (Figure 3-30, PCR 3). The primer created in PCR 1 and "Primer B" were both able to anneal with the pMUW2882 plasmid, whereas the latter inserted a *Bam*HI restriction site to the 3´end . "Primer D" was only able to anneal to the 5´ end of PCR 1 product primer and inserted the *Bam*HI cutting site which was necessary for ligation to the 5´ end. Lastly, the plasmid DNA was digested with the restriction enzyme *Bam*HI and self-ligated to produce circular plasmids.

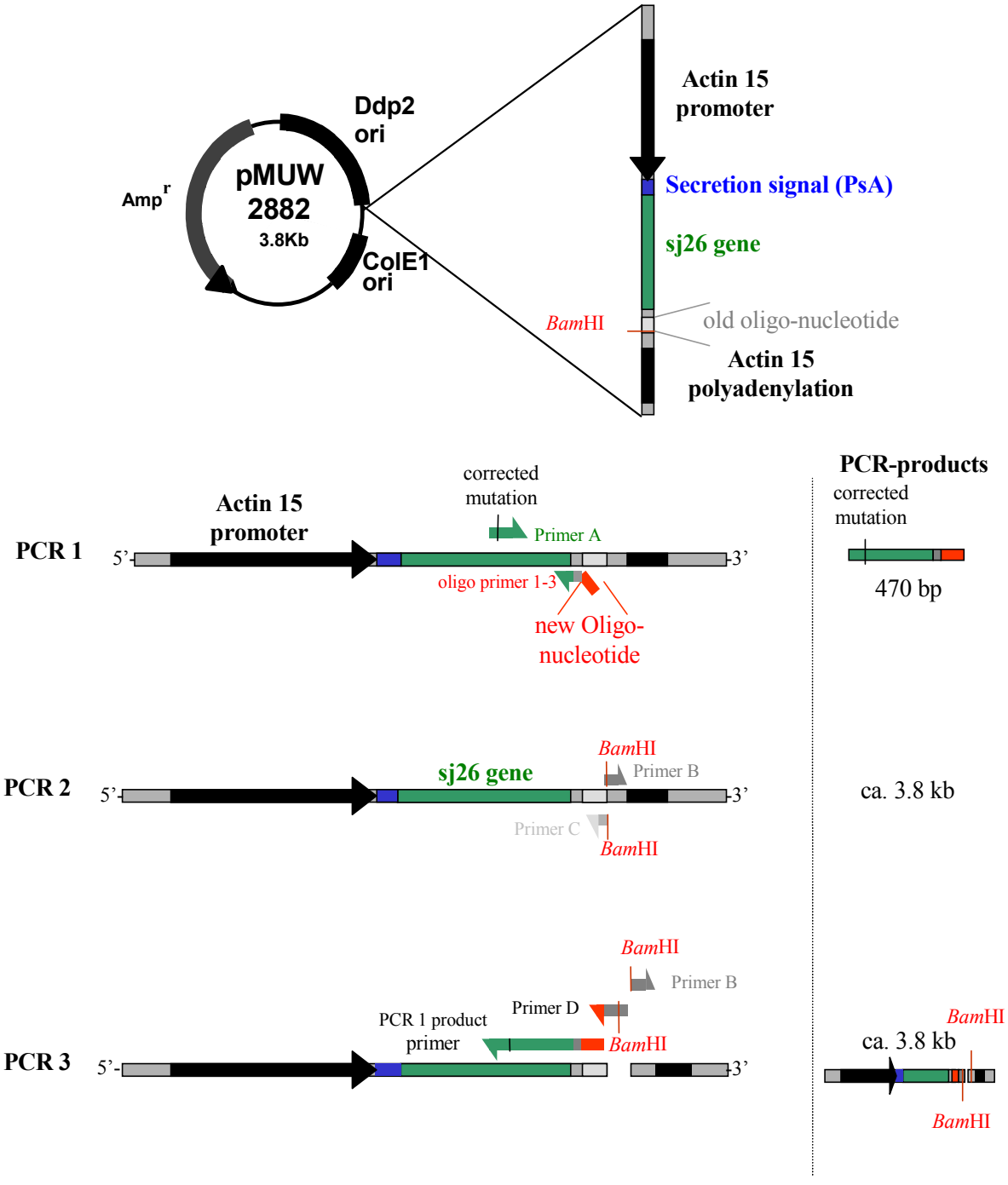



**Figure 3-30 PCR reactions used for the construction of the expression vectors**

This isolated plasmid DNA was electroporated into *E. coli* cells and transformant clones selected by ampicillin resistance. Four transformants were picked out for each

construct and tested by PCR for the presence of the GST-SP96 peptide fusion (primer: P174GST.f2c/ ISPY. r1a; Material and Methods 2.6.2). The expected 256 bp large fragment was visible in three of the SP96S-motif transformants, in three of the SP96T-motif and in all four of SP96G-motif vectors (data not shown). Plasmids were prepared from two transformants of each construct and their GST genes sequenced. The old mutation (bp 1197, isoleucine → methionine) was corrected in all three constructs, but unexpectedly some new mutations were detected (see below). It seemed the used AccuTaq polymerase mix lost its "proofreading" activity and therefore some new misincorporations had taken place. The plasmids chosen for further work were pMUW2911 (SP96G-motif), pMUW2921 (SP96T-motif) and pMUW2931 (SP96S-motif) (Sequences are listed in the appendix).

Vector pMUW2911 (SP96G-motif) had two mutations, one at 994 bp in the PsA secretion signal, where the amino acid lysine was converted to glutamic acid, the second changing isoleucine to valine in the GST gene at base 1190. Vector pMUW2921 contained only one mutation, in the factor Xa cleavage site at 1590 bp (glycine to serine), which could effect a cleavage of the *O*-glycosylation acceptor motif from the GST protein. The pMUW2931 vector showed two mutations: one converted the amino acid at 995 bp from lysine this time to threonine, the second mutation was at 1510 bp and changed isoleucine to threonine.

In the next step, the three expression vectors were co-transformed by means of the integrating vector pMUW110 into *D. discoideum* strain HU2868 by a calcium precipitation method (93). Strain HU2868 is derived from strain NP2 (32) and was maintained on *Micrococcus luteus*. This bacterial food source was chosen because of its resistance against the antibiotic Geniticin (G418), the selective marker for transformed *D. discoideum*. Typically 5-10 *D. discoideum* transformants were obtained with a transformation of $10^7$cells. One transformant of each containing the vector was characterised in detail (Table 3-7).

| Motif of SP96 | Name of the vector | Name of the *D.discoideum* transformant |
|---|---|---|
| G-motif | pMUW2911 | HU2942 |
| T-motif | pMUW2921 | HU2944 |
| S-motif | pMUW2931 | HU2943 |

**Table 3-7 SP96 repeat region and the corresponding vector and transformed strain, respectively**

### 3.5.4.3 Expression and purification of recombinant GST fusion protein

For expression of recombinant GST the transformants were usually grown in glucose containing axenic media under G418 selection. Only during the final scale-up were cultures grown in maltose as a carbon source because of the higher expression rate (M. Slade, personal communications). The use of the PsA signal sequence resulted in a secretion of the fusion proteins after having been processed through the endoplasmic reticulum (ER) and Golgi apparatus. Previous studies have shown that the maximum expression levels under control of the actin 15 promoter were obtained when the cells entered the stationary phase. Growth rate slows at approximately $6*10^6$ cells/ml and generally reached $1*10^7$ cells/ml before the supernatant was collected. A fourfold higher concentration of the culture media of all three transformants was required for a weak detection of the expressed protein in western blot analysis with anti GST antibody (data not shown).

For further purification, the culture supernatant had to be dialysed extensively against PBS to remove free glutathione. The supernatant was reduced with DTT and GST fusion proteins which were purified by affinity–chromatography using glutathione Sepharose. The yield of GST protein was approximately 20-30 µg per 1 L culture supernatant. This is much lower than previous results using this expression system, where between 500 µg and mg´s of GST protein could be purified (70, 71). It was noticed that the weak GST band observed in concentrated media was not related to the amount of GST after affinity purification. It appeared that the GST underwent degradation caused by the reducing conditions necessary for the affinity purification, although all purification steps were done at 4°C. It is known that *D. discoideum* secrets some cysteine protease in the media (11, 96, 98), but the use of the cysteine protease inhibitor E64 did not increase the yield. So far no explanation can be given for this observation.

The calculated mass of the GST peptides was around 30 kDa, that of *S. japonicum* GST was approximately ~28 kDa, and that of the acceptor motif 2 kDa. The apparent size of purified GST (HU2943 strain; S-motif) was ~35 kDa, the GST-T motif (HU2944 strain) and the GST-G motif (HU2942 strain) were approximately ~33 kDa and ~36 kDa size, respectively (Figure 3-31). All GST bands were higher than expected from their theoretical mass and could therefore be post-translationally modified. The GST-T motif was used as a negative control, as threonine was not found to be modified in SP96 (3.5.2), but here it had a larger than expected protein mass. The G-motif (HU2942) gave a second weak band at 33 kDa which could be the breakdown product of the top band or due to a partly

glycosylated motif. A glycostain based on immunochemical detection of Biotin, which was bound over the aldehyde groups of periodate cleaved sugars, reacted strongly with GST-S motif (HU2943), but weakly with GST-T motif (HU2944) This could also be an artefact showing a detectable band with GST-G motif (HU2942) (data not shown). Previously studies have reported that GST expressed without an attached acceptor motif was not modified by glycosylation (69).

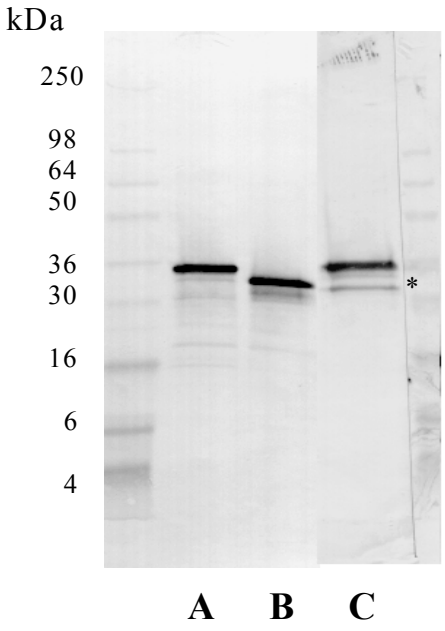



**Figure 3-31 Western blot of purified GST-acceptor motif of different strains separated on a gradient gel and stained with anti GST antibody**

**A**) HU2943 (S-motif); **B**) HU2944 (T-motif); **C**) HU2942 (G-motif)

* breakdown product of the top band or only partly glycosylated motif

#### 3.5.4.4 Immunochemical characterisation of the different SP96 motifs

Monoclonal antibodies recognising SP96 epitopes (3.1) were used to screen against carbohydrate dependent epitopes. The antibody MUD3 which recognises the wild-type and fucose mutant (HU2470) SP96 did not show any labelling on the western-blot of all three GST fusion proteins. This was not surprising as MUD3 is a developmentally regulated epitope, i.e., the necessary glycosyltransferase is probably not present at the stage the GST-fusion proteins are expressed. The immunochemical detection with MUD62 antibody, which was able to recognise the terminal fucose of the phosphodiester bound Fucose(α1,3)GlcNAc-unit on wild-type SP96 (3.3.5), reacted weakly with only the GST-S motif and a ~1 kDa smaller band. The latter could be a fragment or only partly glycosylated (Figure 3-32). The same pattern was displayed by the immunochemical detection with Ab 83.5, which reacts similarly as MUD62 (Figure 3-32). Both antibodies did neither detect the GST-T motif nor the GST-G motif. Also the use of the sensitive mAb MUD166 both displayed a labelling of the S-motif and the G-motif fusion protein. However, the bands were much stronger than seen on the MUD62 or Ab83.5 western-blot, respectively. The high sensitivity of this mAb also lead to the detction of some weak contamination bands. Monoclonal antibody AD7.5 which recognises GlcNAc-1-P (87) also reacted with the GST-S motif and the GST-G motif as expected because of the similarity to the MUD166 epitope (this study). This antibody is less sensitive in comparison to MUD166, and therefore only the stronger fusion protein bands were observed. Neither MUD166 nor AD7.5 reacted with GST-T motif fusion protein. All monoclonal antibodies recognising SP96 post-translational modifications were unable to react with GST proteins after the C-terminal peptide motif was cleaved with factor Xa (data not shown).

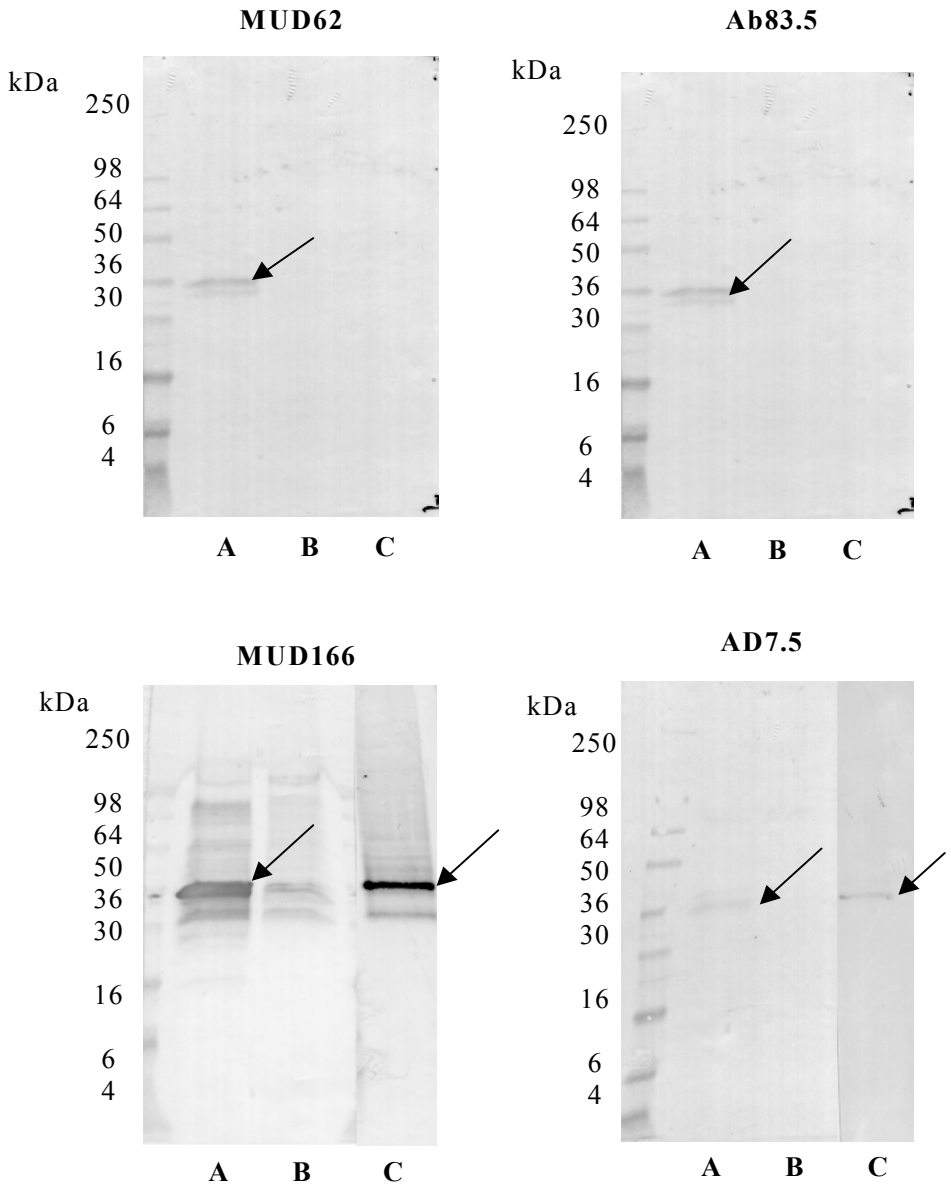



**Figure 3-32 Western blot analysis of GST fusion proteins of different strains with SP96 recognising mAb MUD62, Ab83.5, MUD166 and AD7.5**
**A**) HU2943 (S-motif); **B**) HU2944 (T-motif); **C**) HU2942 (G-motif)
Approximately 0.4 µg purified GST sample of each strain was used

SP96 recognising antibodies were unable to detect a glycan structure on the T-motif, which was consistent with earlier results (3.5.2). However, a very weak band seen in a glycostain suggests there may be a low level of glycosylation. Jung et al. reported *O*-glycosylation with this expression system on a threonine-rich repeat region of a MUC2-like peptide and showed that directly *O*-linked GlcNAc glycosylation was present (70). This

glycosylation was biochemically analysed and also detected with an antibody called MUD50. Because of the similarity of the two threonine-rich repeat sequences (Table 3-8), MUD50 was used to screen against *O*-linked GlcNAc on the T-motif fusion protein. In this way it could be and demonstrated that the MUD50 epitope was present on the HU2944 peptide fusion protein. This result was surprising because the MUD50 epitope is normally expressed at a different time and differently targeted and is further not present on native SP96 (Figure 3-33). Another interesting observation was the spontaneous cleavage of the acceptor motif, which resulted in the loss of the MUD50 epitope as is also shown in Figure 3-33(C). This cleavage could be accelerated by using factor Xa protease. An possible explanation could be the mutation in the factor Xa cleavage site found in the vector pMUW2921 (3.5.4.2).

| | | |
|---|---|---|
| MUC2-like peptide | P **T** T **T** P I T **T** T **T** T **T** V **T** P **T** P **T** P **T** G **T** Q T | (70) |
| SP96-T motif | A T T T A T T A T T T A | this study |

**Table 3-8 Comparison of a MUC2-like peptide sequence with the SP96 T motif used in this study**
bold T: with α-GlcNAc modified threonine (70)

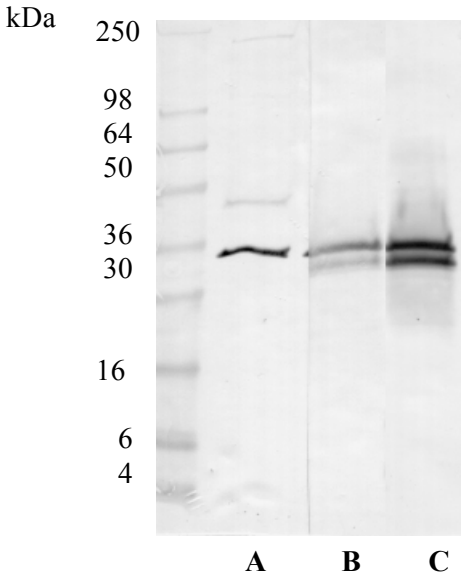



**Figure 3-33 Identification of T-motif glycosylation by using O-linked GlcNAc recognising mAb MUD50**

**A**: GST labeled mAb MUD50
**B**: GST labeled with anti *S. japonicum* GST
**C**: partly cleaved GST labeled with anti *S. japonicum* GST

#### 3.5.4.5 Separation of the released acceptor motif

The amount of purified GST-fusion protein was quite low in all transformants. Nevertheless, an attempt was made to recover the factor Xa cleaved acceptor motifs for mass-spectrometry analysis by MALDI-MS or LC-MS. The glycopeptide mass could be used to calculate the number of glycosylated sites on each motif.

The peptide-fusion proteins from HU2942 and HU2944 were factor Xa cleaved and separated over a reversed phase column (Sephasil C8). The peptides were eluted with an acetonitrile gradient from 0-85 % at a flow rate of 100 µl/min and samples were collected every 2 minutes. Positive fractions for the S-motif, eluted at 52 % v/v acetonitrile, were identified by dot blot with MUD166 and with MUD50 for the T-motif 49 % v/v acetonitrile, respectively (Figure 3-34).

The acetonitrile concentration needed to elute the S- and the T-motif in this study differs from Jung´s results who eluted four different glycosylated motifs with 16-22 % v/v acetonitrile and characterised them using mass-spectrometry (70, 71). These differences raise doubts concerning the interpretation of the spectra in Figure 3-34. As fractions were only tested by dot blot with MUD166 and MUD50, but not with anti GST antibody it one cannot could exclude that the positive peaks results from non-cleaved peptide fusion proteins. It is possible that in an incomplete digest the amount of released peptide was too small to be detected by UV absorption and immunochemical detection. All analysis done on the positive fractions by mass-spectrometry with MALDI-MS and LC-MS did not give any positive results, although both are sensitive analytical methods. Both mass-spectrometers were adjusted to a mass window capable to detect peptides in the range from 500-3500 m/z. Hence non-cleaved protein (~ 35 kDa) would be out of range. A further interpretation could be that the amount of cleaved peptides were too low or were not ionised because of a high content of glycosylation.

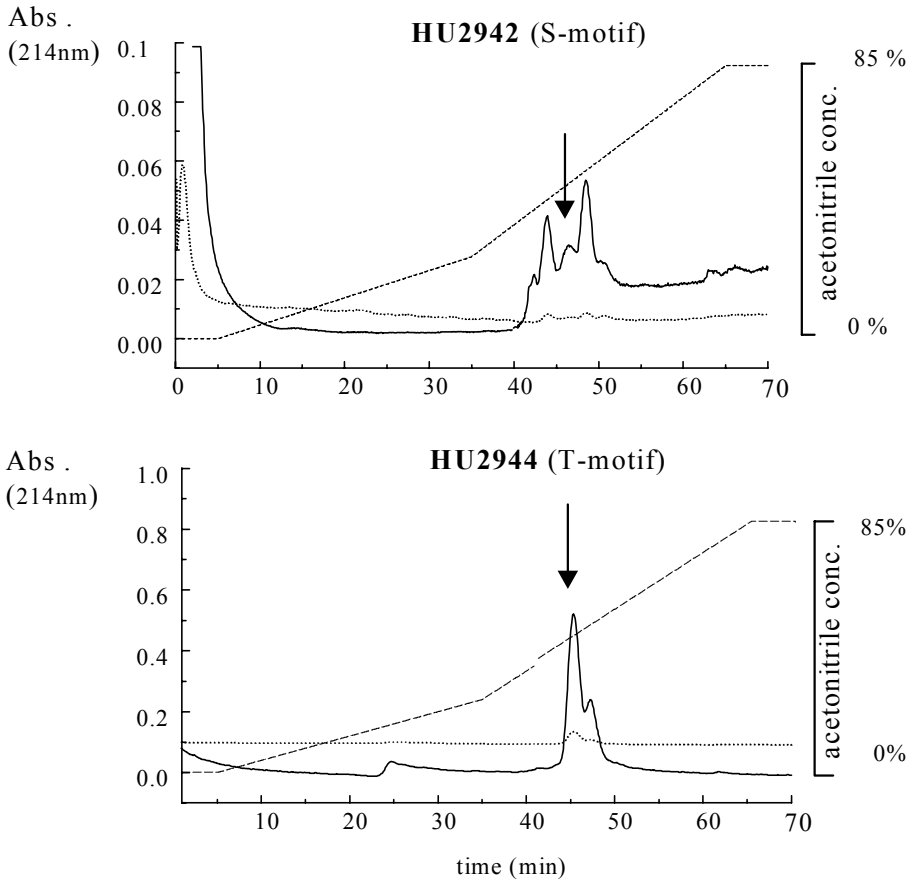



**Figure 3-34 Reversed phase separation of GST S- and T-motif fusion proteins after factor Xa cleavage, respectively**

Positive fractions indicated by arrows were identified in HU2942 by MUD166 and in HU2944 by MUD50 dot blot
Dashed lines: acetonitrile concentration. Dotted lines: absorbance at 280 nm, solid lines: absorbance at 214nm

### 3.5.5 Analysis of T-motif glycosylation in dependence of the expression time

The GST expression results shown in 3.5.4.4 demonstrated that the chosen repeat regions out of SP96 are acceptor motifs for *O*-glycosylation. However, it is possible that the glycosylation could be dependent on the time of expression. For example, the MUD50 epitope detected on the GST-T motif fusion protein produced during the stationary phase was not present on native SP96 expressed in a later developmental stage. To support this result the actin 15 promoter of the expression vector pMUW2921 was exchanged by the promoter of the spore coat protein SP60, the only well described promoter of a later stage (55, 144). SP60, like SP96, is synthesised exclusively in prespore cells at the tipped aggregate stage of development (125). Transformants with expression plasmid containing SP60 promoter were grown on bacteria plates so they were able to go through the developmental program. At the slug stage the SP60 promoter would be active and the peptide fusion protein would be expressed. Slugs were used for immunochemical analysis with MUD50 to compare it with the results seen Figure 3-33.

#### 3.5.5.1 Construction of the expression vector with the SP60 promoter and the insertion in *D. discoideum*

The promoter region in the plasmid was again amplified by different PCR reactions. The schematic PCR steps are displayed in Figure 3-35 and Figure 3-36. The first PCR reaction to isolate the SP60 promoter was done on purified DNA from *D. discoideum* strain HU2868 as a template with two primers ("primer 1" and "primer 2"; exact sequence see Material and Methods 2.6.2), which enclosed the SP60 promoter region and the 5´end of the SP60 gene (**PCR 1**, sequence shown in appendix). The reverse "primer 2" within this SP60 gene was chosen for a relatively high GC rich sequence to ensure a correct annealing. The final PCR product was an approximately 1 kb fragment which was used as a template in a second PCR. The primers in the second PCR step (**PCR 2**) inserted a *Xba*I cleaving site on the 3´ end and on the 5´ end an overlap with the PsA sequence with an *Nsi*I cleavage site. The resulting PCR product was here 750 bp (Figure 3-38, 1). The step is size serving as a control for the right PCR product in the first PCR step. The DNA encoding the PsA secretion signal, the *S. japonicum* GST and the SP96 T-motif were amplified in **PCR 3** using the pMUW2921 plasmid as a template and the "primer 5" and "primer 6" (Figure 3-38, 2). The PCR2 (~750 bp) and PCR3 products (~750 bp), as well as the primers 3 and 6 were used in **PCR 4** to fuse

the SP60 promoter with the PsA- *S. japonicum* GST SP96T-motif sequence to produce a 1,5 kbp large fragment (Figure 3-38, 3). **PCR 5** amplified most of the plasmid vector pMUW2442 to give a product of 2,85 kbp (Figure 3-38, 4). The "primer 7" inserted an *Xba*I site at the start of the promoter (Actin 15) (Figure 3-36). In the final step, the 2,85 kbp fragment from PCR 5 and the 1,5 kbp fragment of PCR 4 were digested with restriction enzymes *Bam*H1 and *Xba*I and ligated (Figure 3-37). The newly constructed plasmid with the SP60 promoter was sequenced and named pMUW5021.

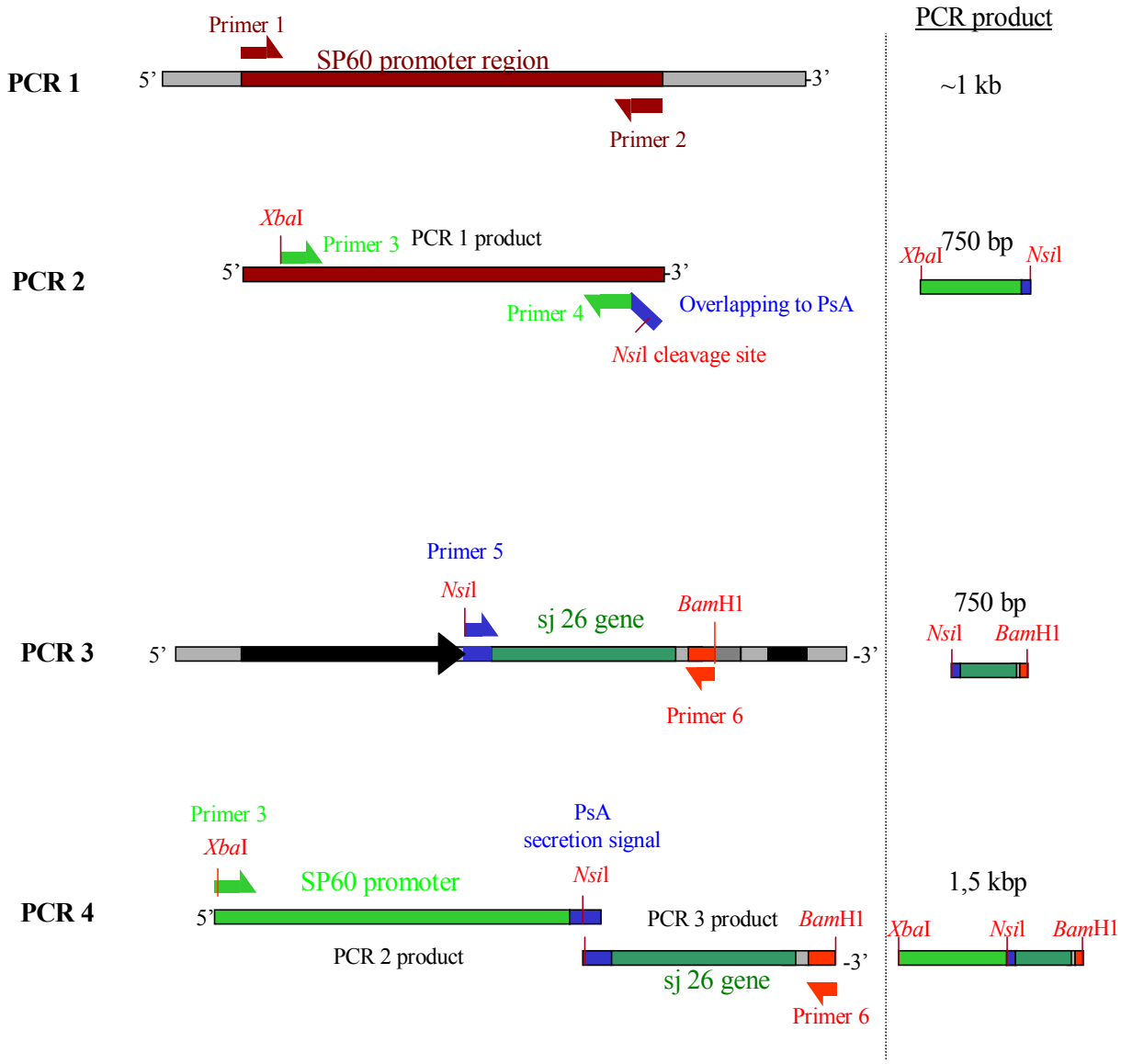



**Figure 3-35 Schema of PCR reactions used to construct plasmid with an exchanged promoter region (Actin15 against SP60 promoter)**

PCR1: Template was *D. discoideum* DNA from strain HU2868 to multiply SP60 promoter sequence
PCR3: Template was plasmid pMUW2921 (T-motif, construction Figure 3-30) to multiply PsA-sj26-T-motif sequence

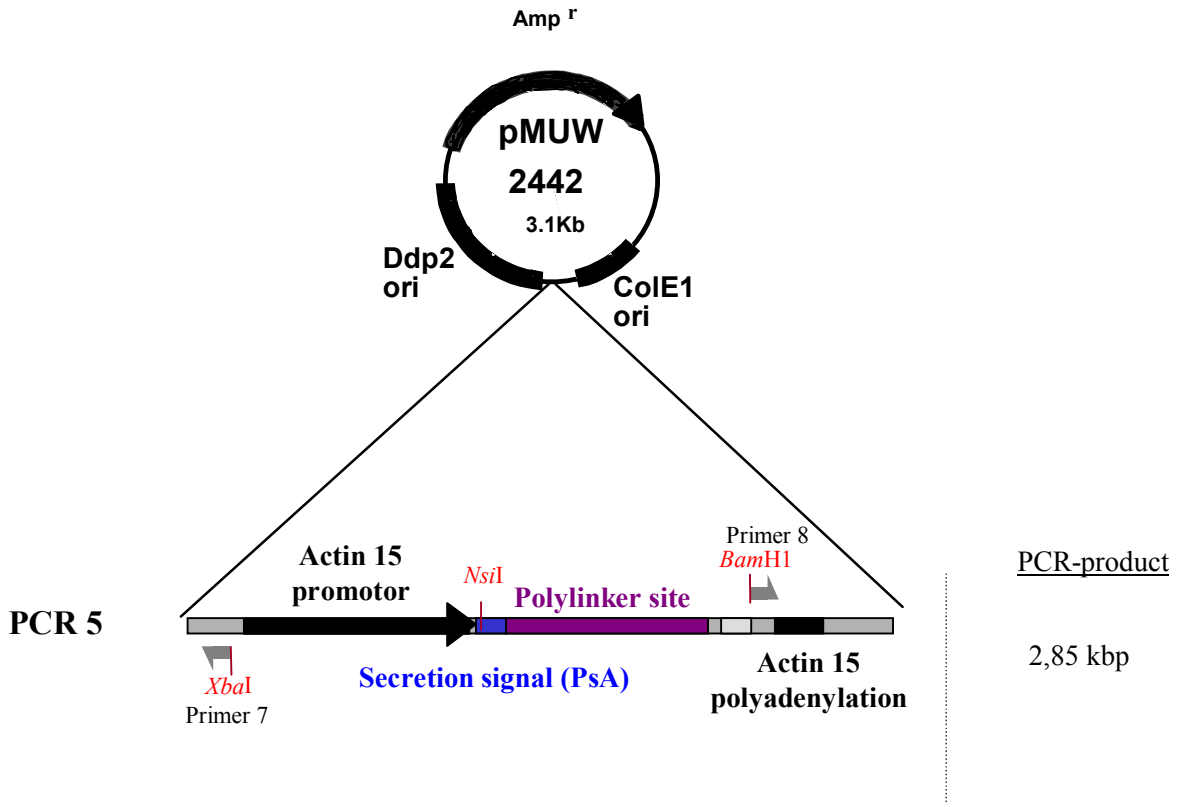



**Figure 3-36 Original plasmid pMUW2442 used as a template in PCR5 step to amplify the part of the plasmid (Amp$^r$, Ddp2 ori and ColE1 ori)**

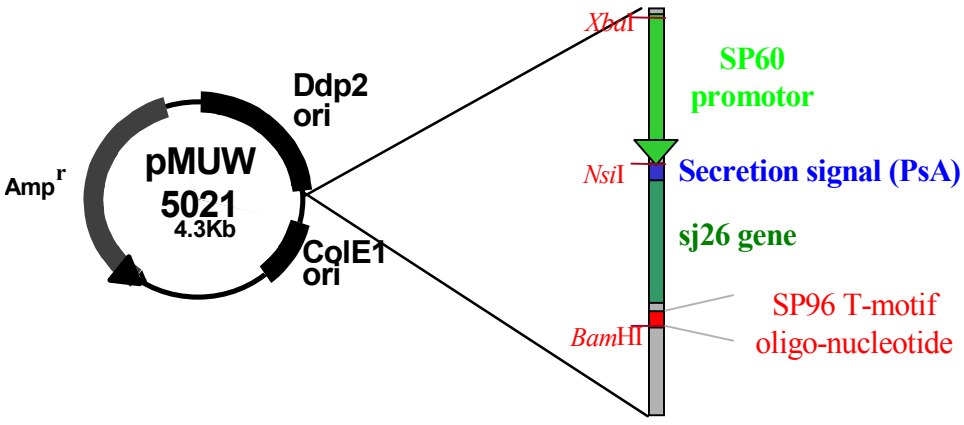



**Figure 3-37 Constructed plasmid pMUW5021 containing the SP60 promoter**

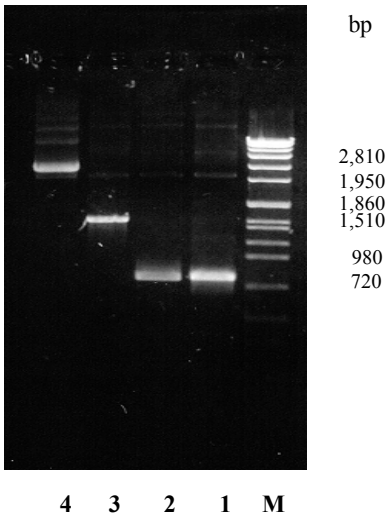



**Figure 3-38 Agarose gel with PCR products of reactions PCR2-PCR5**

**M**: SPP1/EcoR1 marker
**1** : PCR 2 (primer3/ primer4)
**2** : PCR 3 (primer5/ primer6)
**3** : PCR4 (primer3/ primer6)
**4** : PCR5 (primer7/ primer8)

### 3.5.5.2 Immunochemical characterisation of SP96 T-motif in slugs expressing GST-fusion protein

In this experiment it was not absolutely clear how the organism can secrete protein in the slug stage, therefore also samples of the slime sheath of the slugs was taken and used for analysis. After slugs (approximately 20) were collected in reducing sample buffer. The proteins were separated by SDS-PAGE, transferred to nitrocellulose and the GST bands were stained immunochemically with anti GST antibody.

The western blot with anti *S. japonicum* GST antibody did not label any protein band in the slime sheath sample (data not shown). The slug sample showed a low amount of expressed peptide fusion protein which made further analysis difficult. The GST expressed during the slug stage reacted with the GST antibody only as a very weak band at ~70 kDa

(Figure 3-39, 1), compared to the GST fusion protein at 33 kDa isolated during the stationary phase out of the axenic media (Figure 3-39, 2). The slug GST fusion protein band (1) had approximately twice the size compared to the protein band labelled (2) in Figure 3-39. Therefore, it was suspected that the GST might have dimerised. GST is naturally a homodimeric molecule stabilized by salt bridges and hydrophobic contacts (85), but should be monomeric under the chosen conditions. Interestingly, a similar GST fusion protein band pattern was detected during the analysis of supernatant of bacteria grown cells, using Actin 15 promoter (Figure 3-40) instead of the SP60 one. The Actin 15 promoter expressed sample was taken while the cells were in the stationary phase, at a cell density of approximately $1{,}5*10^7$ cells/ml. However, only the lower band (~33 kDa) could be affinity purified (data not shown). The higher band (~65 kDa) was only observed in bacterially grown cells.

The purpose of using the SP60 promoter was to elucidate and compare the MUD50 binding capability on the T-motif during different expression times. Secreted GST-T motif fusion protein from strain HU2944 in axenic growth presented the MUD50 epitope (Figure 3-33). However, the protein band at ~70 kDa of sample 1 (Figure 3-39) expressed by SP60-promoter HU2945 did not show a labelling with MUD50, even after a 5 fold concentration. The result is consistent with observations made on native SP96 protein, but nevertheless should be taken with skepticism, because of the low amount of sample. It is still possible that the observation is the result of the missing carbohydrate epitope of MUD50 or that the amount of protein was under the detection limit.

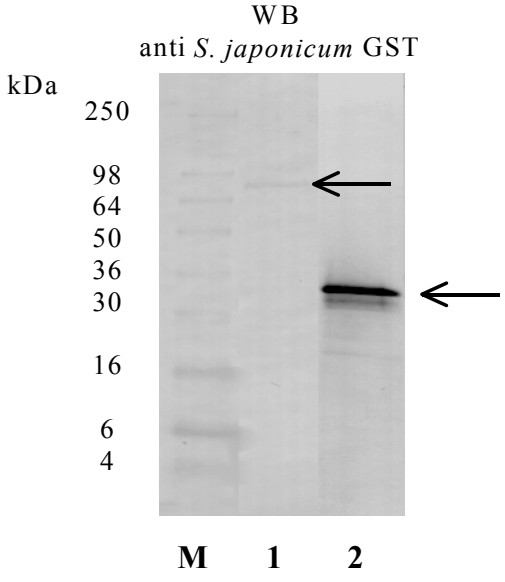



**Figure 3-39 Expression of GST T-motif fusion protein at different times**
**1)** Slugs of strain HU2945 (SP60 promoter) grown on bacteria
**2)** Out of supernatant purified GST T-motif fusion protein of strain HU2944 grown in axenic media

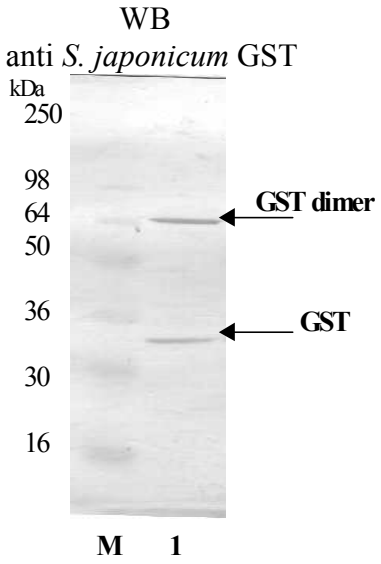



**Figure 3-40 GST fusion protein of strain HU2944 grown in bacteria/ buffer solution and secreted into the buffer**

Separated on a SDS-PAGE and detected with anti *S. japonicum* GST antibody.

## 3.6 Prediction of modified sites in SP96 by different databases

Two databases were consulted for prediction of modified sites in native SP96 and in the SP96 motifs expressed as GST-fusion proteins.

The first database (DictyOGlyc 1.1[i]) predicts *O*-GlcNAc glycosylation sites in *D. discoideum* based on a small data set of known glycosylation sites and sites not glycosylated for this class of *O*-glycosyation in *D. discoideum*. Although the database is limited because of the small data set, it was used to make predictions on the three repeat motifs of SP96 and on the whole SP96 sequence. It was of interest to see if the MUD50 positive SP96-T motif shows potential sites for this glycosylation. AA 96 in SP96 sequence, or AA 17 in the SP96-G motif predicts only for one a small potential over the threshold (see Appendix). SP96 is not recognised by MUD50, nor is the SP96-G motif. Therefore, it is unlikely that the database prediction is correct. For the MUD50 positive SP96-T motif no positive prediction for glycosylation was made.

The second database (NetPhos 2.0[j]) is used for prediction of phosphorylated sites in eucaryotic proteins. The search was done with the sequence of the SP96-G, SP96-S and SP96–T motif and whole SP96. The result (see Appendix) predicts potential for 88 of the 121 serine, 8 of the 83 threonine and 6 of the 12 tyrosine residues in the SP96 sequence. The prediction that serine is the main phosphorylated amino acid is in harmony with experimental work which gives no hint for threonine and tyrosine phosphorylation. The positive prediction for serine in this database covered the whole SP96 tail region (AA457-600). The results of the different motifs showed no potential for threonine phosphorylation in all three motifs. There was a potential for 2 of the 5 serines in the SP96-G motif and for 5 of the 10 serines in the SP96-S motif to be phosphorylated. While it is unknown whether motifs used for phosphorylation are comparable to motifs of phosphoglycosylation, the predicted results are in harmony with the observed intensities of immunochemical labelling with MUD166 of the different GST-SP96 motifs (3.5.4.4).

[i] http://www.cbs.dtu.dk/services/NetOGlyc/

[j] http://www.cbs.dtu.dk/services/NetPhos/

# 4 Discussion

The spore is the dormant stage of the asexual developmental cycle of *D. discoideum.* It is a resistant structure capable of surviving for long periods. Spores are surrounded by three different layers, in which proteins are embedded. These build up a shield and protect the dormant amoeba from environmental stress. Some of the proteins carry post-translational modifications, and previous studies have shown that a lack in fucosylation on modifications increases the permeability of the spore coat. This makes the spores less viable with time (43). There is a picture emerging of the assembly of the spore coat, involving a complex of proteins (137) and cellulose binding (84, 158). In order to better understand the nature of the spore coat, detailed structural information is required for each of the component glycoproteins.

In previous studies changes of post-translational modifications have been analysed in these proteins, in wild-type and mutants with a decreased amount of fucose, using monoclonal antibodies. Inhibition studies, comparison of binding patterns and radiolabelling experiments helped to get some information about the differences in modification, though details about the exact structures could not be obtained by these methods (10, 11, 22, 30, 43, 84, 136, 141, 142, 147). It was further observed that some of the detected carbohydrate epitopes on spore coat proteins were also present on proteins expressed during the earlier, vegetative stage.

To get a better understanding how glycosylation fulfils the protective function in spores, how glycosylation changes during development and which glycosyltransferases are regulated over the time of expression, it is necessary to obtain more information about the carbohydrate structures on these proteins.

This thesis clarifies the post-translational modifications of the major spore coat protein, SP96, from the wild-type X22, the *modD* fucose mutant HU2470 and the *mod*E fucose mutant HU2471 of *D. discoideum*. The detailed structural differences between the glycosylation of the SP96 protein in the different strains has been determined and related to the carbohydrate dependent epitopes recognised by the monoclonal antibodies MUD62, MUD166, MUD3 and mAb83.5. The quantification of these modifications, which is done here for the first time on SP96, in combination with partial localisation of the modifications on the already known sequence of SP96 (34), is an important step toward understanding how the protein is embedded in the spore coat and how it protects the spore.

Some of these studies concerning of the differences in post-translational modifications of wild-type X22 and fucose mutant HU2470, as well as their quantitative analysis have already been published Mreyen *et al*. (90).

## 4.1 Multiple forms of modifications on SP96

Before commencing these studies there was a fragmentary amount of information about SP96. The presence of phosphoserine was an early discovery (20, 22) and this helped explain the acidic isoelectric point (pI) observed by 2D-PAGE (22, 43). The presence of more than one SP96 spot on 2D-gels could indicate differently phoshorylated forms of the protein. In early studies (20) it was speculated that the phosphate group could be part of a phosphodiester linkage.

This study shows that, if there is phosphoserine on SP96, it represents only a minor part of the modified serine residues. Most of the phosphate-modified serine residues are moreover elongated by the addition of *N*-acetylglucosamine and fucose in the wild-type. It was shown that the novel phosphoglycan structure (Fuc(α1-3)GlcNAc-α-1-P-Ser) is a major component of the modification of SP96. More than 50% of serine residues in SP96 are modified with this glycan. A further 15-20% of the serines are modified by *O*-linked fucose. If present at all, phosphoserine accounts for less than 1% of the modifications. The phosphoglycosylation of SP96 may be analogous to the phosphoglycan assembly observed in the protozoan parasite *Leishmania* although the glycans are different. In a *Leishmania* acid phosphatase, more than 60% of serine residues of a serine/threonine rich domain are modified by Man-α-1-P-Ser and elongated by other sugars (66, 145). It may be that the abundant *O*-glycans stabilize and protect the enzyme from proteases (66).

Mutants of *D. discoideum* and monoclonal antibodies have helped to understand the *O*-glycosylation of SP96 (Table 4-1). Comparing the structures of the modification on the mutants confirmed the ideas about the nature of the epitopes of the mAbs MUD62 and MUD166, which had been developed from a previous study of fucose mutants (11). Champion *et al.* (11) observed that the GA-X epitope, recognised by MUD62 on wild-type SP96, was competed by fucose, as was a previous, independently isolated anti-carbohydrate antibody, mAb 83.5 (141). It was reported that this antigen GA-X is an *O*-linked GlcNAc-containing oligosaccharide (43). The fucose mutants *modD* HU2470 and *modE* HU2471 lost this epitope and a new epitope appeared (GA-XIII) which was competed by GlcNAc; this

epitope is recognised by MUD166. All these observations can now be explained (Figure 4-1). The *modD* mutant HU2470-SP96 and the *modE* mutant HU2471-SP96 are missing the terminal fucose from the phosphodiester-linked Fuc(α1-3)GlcNAc-α-1-P-Ser disaccharide. This fucose residue is necessary for recognition by MUD62. Therefore the mutants present a new terminal sugar structure, GlcNAc-α-1-P-Ser, which is recognised by MUD166 and also by AD7.5, the latter of which recognises a GlcNAc-α-1-P-Ser glycan structure on cysteine proteinases of *D. discoideum* (87). Specifically, the *modD* mutation appears to affect the activity of the (α1-3) fucosyltransferase, because of its selective effect on the GA-X (MUD62) epitope in developing cells, compared with the retention of the fucose depending GA-XX (MUD3) epitope (involving Fuc-Ser) in these cells.

Srikrishna *et al.* (119) suggest that the mAb 83.5 epitope involves Fuc-β-1-P-Ser. Based on experiments performed in this thesis the evidence is that there are only minor amounts of this structure on SP96. However, the evidence to date suggests that mAb 83.5 and MUD62 have similar recognition patterns and here we show that MUD62 identifies Fuc(α1-3)GlcNAc-α-1-P-Ser. On this basis it is suggested that mAb 83.5 is also likely to identify this structure. Metha *et al.* (87), however, have dismissed this structure as the epitope for mAb 83.5 in studies on proteinase 1. Cysteine proteinase 1 from *D. discoideum* displays the mAb 83.5 epitope, contains as post-translational modification the phosphoglycan GlcNAc-α-1-P and further fucose in an unknown linkage. Treatment with α-fucosidase did not release any fucose, therefore it was concluded that fucose does not exist in a linkage to GlcNAc-α-1-P. Certainly, most studies with both mAb 83.5 and MUD62 antibodies on later growth stage proteins conclude that both antibodies recognise fucose as part of the epitope. It appears now that, at least for mAb83.5, there may be two different glyco-antigens (90, 119). It is still unclear if MUD62 is also able to recognise Fuc-β-1-P-Ser.

Carbohydrate dependent epitopes are found on spore coat proteins which are expressed during the multicellular stage of the organism, and also on proteins, especially cysteine proteinases, of vegetative growing cells. Wild-type cysteine proteinases of vegetative cells show both epitopes, the MUD62 and the MUD166, compared to spore coat proteins of this strain which show only the MUD62. Interestingly the *modD* mutation only partially inhibits fucosylation in prespore/ spore cells. The spore coat proteins of a strain containing the *modD* mutation show the MUD166 epitope, but lack the mAb83.5 and MUD62 epitope GA-X, whereas vegetative cysteine proteinases are still able to express this epitope (12). It seems that the activation of different glycosyltransferases is developmentally regulated. It is

shown that Fuc(α1-3)GlcNAc-α-1-P-Ser is the dominant glycan structure in the MUD62 epitope of wild-type SP96. This does not mean that the structure has to be identical to the glycan structure in the MUD62 epitope on the cysteine proteinases. In studies done by Srikrishna *et al.* on an artificial peptide the less complex Fuc-β-1-P-Ser phosphoglycan was identified as a minimal epitope for mAb 83.5 (119). Because of the similarity of the binding pattern of mAb 83.5 and MUD62, it cannot be excluded that MUD62 binds in the same way and also recognises the Fuc-β-1-P-Ser phosphoglycan structure. A detailed analysis of the epitopes of these two antibodies on the cysteine proteinase could elucidate this question.

It has further been shown that incorporation of fucose into glycoproteins is increased by more than a factor of 3 in *D. discoideum* during early development (126). Also observed was an increasing amount of secreted fucosidase after 4 hours of starvation. It is possible that the developmentally regulated α-L-fucosidase participates in defucosylation as part of a process of carbohydrate-structure alteration preceding and accompanying late development. The Fuc(α1-6)GlcNAc linkages were most susceptible, followed by Fuc(α1-2)Gal and Fuc(α1-3)Glc, but not Fuc(α1-3)GlcNAc and Fuc(α1-4)GlcNAc linkages under these conditions. This observation is in harmony with the identified Fuc(α1-3)GlcNAc linkage on the SP96 spore coat protein, because a susceptibility of the linkage would interfere with the construction of this kind of glycan linkage. On the contrary fucosidases from human serum show a higher susceptibility for the Fuc(α1-3)GlcNAc, the Fuc(α1-4)GlcNAc and the Fuc(α1-2)Gal linkages and only a poor cleavage rate for Fuc(α1-6)GlcNAc (113).

Late expressed proteins in *Dictyostelium* quite often contain *O*-glycans composed of fucose and GlcNAc, showing variation in their arrangements and types of linkages. The glycans detected on native SP29 (PsA) all have reducing terminal GlcNAc. In some cases the sites are further modified by addition of either GlcNAc, Fuc-GlcNAc or P-Fuc-GlcNAc through a phosphodiester bond (156). The glycans detected on SP96 are also composed of Fuc, GlcNAc and phosphate, but with a different arrangement. It was shown that *N*-acetylglucosamine in the reducing position directly attached to threonine or serine is the minimal epitope of MUD50 (44, 71). This kind of observation that glycans can vary in their linkages but not in their sugar composition has already been made on the Lewis blood group antigens where linkages in α1-4 and α1-3 between fucose and *N*-acetylglucosamine are found (115) .

The *modE* mutant HU2471 shows a reduced fucosylation of < 5% of the wild-type (11). SP96 from this mutant had a fucose content of approximately 3% of wild-type. This

reduction is beyond the missing terminal fucose of the phosphodiester-linked disaccharide the result of the missing *O*-linked fucose monosaccharide residue attached directly to serine. This form of glycosylation has been discovered on a small number of secreted proteins involved in the blood coagulation cascade or clot dissolution in humans. This structure is also found within epidermal growth-factor like (EGF) modules (58, 59), which are typically 30 to 40 amino acids in length and are a common structural motif found on numerous secreted and cell surface proteins in animals (8). They often mediate protein-protein interactions. More recently *O*-linked fucose was detected on a mammalian cell surface receptor, responsible as an essential player in a wide variety of developmental cascades (88).

Comparison of results from immmunochemical and glycan characterisation between wild-type X22, mutant HU2470 and mutant HU2471 demonstrate that the observed loss of the MUD3 (GA-XII) epitope on HU2471-SP96 is definitely the result of the missing *O*-fucose linked to serine. The only protein which is recognised by MUD3 is SP96, and that only in the X22 and the HU2470 strain. SP96 of these strains is also the only protein in *D. discoideum* where this glycan could be detected so far. There the MUD3 epitope is less common in comparison to epitopes of MUD62, MUD166, MUD50, AD7.5 and 83.5. It was also suspected that the *O*-linked fucose glycan structure is only part of the epitope. The epitope could also be a combination between carbohydrate and amino acid sequence. This would also explain the specificity for SP96. It is more unlikely that the antibody recognises a glycan stabilized conformation, because of its binding after denaturing conditions (SDS-PAGE/ western-blot analysis), although minor renaturation under these conditions cannot be excluded. The *modE* mutation appears to prevent fucosylation globally, because of the loss of two different fucose-dependent epitopes on SP96 (GA-X and GA-XII) and of the GA-X epitope on cysteine proteinases. A possible step affected by this gene is transport of GDP-fucose into the Golgi apparatus, rather than a fucosyltransferase (11). On the other hand, strains carrying the *modC* mutation have a complete lack of fucose and are probably unable to transform GDP-Man to GDP-Fuc. Therefore, they lose all fucose-dependent epitopes.

| Strain | Mutation | Antibody recognition | postulated glycan structure involved in mAb epitopes | Postulated defect |
|---|---|---|---|---|
| X22 | - | MUD62 | Fuc(α1-3)GlcNAc-α-1-P-Ser | - |
| | | mAb 83.5 | Fuc(α1-3)GlcNAc-α-1-P-Ser/ Fuc-β-1-P-Ser | |
| | | MUD3 | Fuc-Ser | |
| HU2470 | *modD* | MUD166 AD7.5 | GlcNAc-α-1-P-Ser | α1-3 fucosyltransferase |
| | | MUD3 | Fuc-Ser | |
| HU2471 | *modE* | MUD166 AD7.5 | GlcNAc-α-1-P-Ser | GDP fucose transport |
| HU2733 | *modC* | MUD166 AD7.5 | GlcNAc-α-1-P-Ser | Formation of GDP-fucose from GDP-mannose |

**Table 4-1 SP96 in *D. discoideum* strains X22, HU2470, HU2471 and HU2733; the epitopes, postulated glycan structures and possible defects in the mutant strains**

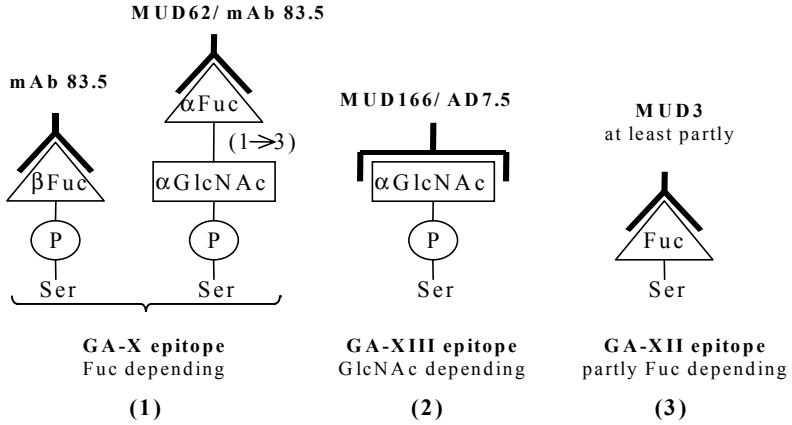



**Figure 4-1 Simplified model of mAb binding sites on SP96**.
Identified post-translational modifications and their postulated antibody epitopes, on the spore coat protein SP96. (1) on spore coat proteins from the wild-type ; (2) on some spore coat proteins from the fucose mutants HU2470, HU2471 and HU2733 with a mutation in *modD*, *modE* and *modC*; (3) on spore coat protein SP96 from the wild-type and the *modD* mutant HU2470. The mutants HU2471 and HU2733, with further decreased amount of fucose, lose this epitope.

## 4.2 Location of modifications on SP96

SP96 has on the C-terminal 180 amino acids a serine/ threonine rich tail where 74 of the 120 serines and 47 of the 82 threonines are located. In this tail region of SP96[k] (protein sequence, see 1.2.8) there are two repeat regions: one serine-rich (ASSSSAPSSSA) and one threonine rich repeat (TTTATTA) located in the last 65 amino acids. N-terminal for these regions is a repeat with both serine and threonine (AGSQTSGGSTSG).

It was shown that all post-translational modifications found on SP96 are localized on serine and that any modification on threonine can almost be excluded. Edman sequence analysis and immunochemical analysis with MUD62 of digested SP96 demonstrated that most of the phosphoglycans are on the serine-rich tail, which also contains *O*-linked fucose (MUD3 epitope). Whether the *O*-linked fucose is only in the tail region or possibly also further upstream on the protein is still unknown. Up to now only one amino acid motif for the addition of fucose as a reducing terminal sugar to threonine/ serine has been described from epidermal growth factor regions in human of some multidomain proteins (Cys-X-X-Gly-Gly-**Thr/Ser**-Cys) (59). This motif is not found in the SP96 sequence.

The analysis with the GST expression system of three serine/ threonine repeat regions of SP96 as potential *O*-glycan acceptor sites confirmed the result of the location of the phosphoglycosylation in the serine rich tail and further identified this modification in the mixed serine/ threonine sequence closer to the N-terminus in the SP96 sequence.

Interestingly the dominant epitope, expressed during this earlier vegetative stage, is that recognised by MUD166 instead of MUD62 which is the main epitope found in the spore cells. In *D. discoideum* it has been shown that lysosomal cysteine proteinases expressed during the vegetative stage have a serine-rich domain, similar to SP96, which is the target for GlcNAc-α-1-P-Ser modification (101). The cysteine proteinase CP6 has 44% serine in 100 amino acids, while CP7 contains 43% serine in its 130 amino-acid domain. The latter contains several short repeat motifs such as (SGSG, SQSQ, SQSA and SGSA) (101).

That the phosphodiester-linked glycan structure is on the serine rich tail is also supported by the database[l] prediction of phosphorylated amino acids which predicts that all serines in the tail region have the potential for becoming phosphorylated. These results, in combination with the calculated ~60 phosphodiester glycan units per protein molecule, suggest a “bottle-brush” like glycan orientation on the SP96 tail. This is also observed on the

[k] Swiss-Prot P14328

serine/ threonine rich domains of animal and human mucins which have carbohydrate chains with up to 20 sugar residues (123). The reducing terminal sugar in mucins is, however, a GalNAc which is further elongated. No phosphodiester-linked glycans have been reported in vertebrates.

In previous studies it has been shown that labelling of vegetatively expressed cysteine proteinases with MUD62 and mAb83.5 antibodies was weaker than that observed on multicellular-stage proteins (12). A weaker immunolabelling was also observed on the expressed GST-S motif and GST-G motif fusion protein with MUD62 antibodies compared to MUD166. This is consistent with the fucose labelling in the development of *D. discoideum,* where the fucose content increases in the late development. A possible explanation could be that there is a lower level of α1-3 fucosyltransferase at this stage of expression which results in less Fuc(α1-3)GlcNAc-α-1-P-Ser modification. Another possibility is that Fuc-β-1-P-Ser is expressed at the vegetative stage instead of Fuc(α1-3)GlcNAc-α-1-P-Ser. At least for the mAb 83.5 it is proven that the Fuc-β-1-P-Ser glycan is a possible epitope, and that the antibodies bind to this modification with a lower affinity. Unfortunately, no accurate mass could be obtained of the modified *O*-glycosylation acceptor peptides of the GST-fusion protein, which would have been helpful to estimate the number of modified sites and glycan composition. Up to now there was only minor evidence for the Fuc-β-1-P-Ser modification on native SP96 protein and facts for MUD62 binding to this structure.

There is also a variation in the intensity of the MUD166 and AD7.5 labelling between SP96 and the GST fusion proteins. Both antibodies show similar band intensities with the same amount of HU2470-SP96 protein. The same amount of expressed GST-S motif or GST-G motif peptides, however, displayed strong differences in intensity in favour of MUD166. Thus it seems that AD7.5 recognises a more conformational-dependent epitope which is better presented on the SP96 protein than on short acceptor peptides. *In vitro* studies on *N*-acetylglucosamine-1-phosphotransferase specificity with different acceptor motifs in *D. discoideum* demonstrated that a serine flanked by alanines is a 13-fold better acceptor for GlcNAc-α-1-P-Ser modification than serine flanked by valines or by basic amino acids (86). This result is confirmed by comparing immunochemical results of MUD166 and AD7.5 western blot analysis for the GST-S motif, where four serines are flanked by alanine, and the GST-G motif, where no serine is flanked by alanine. Here the same tendency can be observed

[1] http://www.cbs.dtu.dk/service/NetPhos/

as that described by Metha *et al*. (86). The prediction from the database for phosphorylation also showed a higher content of potential modified serine residues in the GST-S motif than in the GST-G. Although nothing is known about the comparability between phosphorylation and phosphoglycosylation of acceptor motifs, this result fits with the observations. Comparison of different immuno-blot patterns of MUD166 and AD7.5 of total spore protein extract is also consistent with the suggestion that there are unknown differences in their epitopes. Although both antibodies seem to need GlcNAc-α-1-P-Ser in the epitope, a significant difference in labelling of SP75 can be observed. MUD166 antibody labels the SP75 band in both mutants, while there was no SP75 detection with AD7.5. On the other hand, AD7.5 makes some protein bands visible which are not detected by MUD166.

An interesting feature was the immunochemical labelling of the expressed threonine rich GST-T acceptor motif of SP96 with the monoclonal antibody MUD50. The MUD50 epitope (minimal epitope: GlcNAc-Thr/Ser) is not found on native SP96. In addition to sheath proteins, the MUD50 epitope occurs on glycoproteins found primarily, but not exclusively, in prespore cells (2). This theronine-rich motif was chosen as a negative control for glycosylation, as there was no evidence for threonine glycosylation in SP96. As all motifs were expressed during the vegetative stage under the control of an actin15 promoter it may be assumed that this type of glycosylation is a developmentally regulated process. A similar observation of positive MUD50 labelling was made on a threonine-rich motif of SP29 (PsA), expressed with the same GST expression system, although in this case the native protein has this epitope. Further analysis on this motif showed, however, that on the earlier expressed acceptor motif only a truncated glycan structure is added. The minimal epitope for MUD50 is a single GlcNAc residue directly linked to threonine (71). The *D. discoideum* database[m] prediction for *O*-GlcNAc glycosylation of the SP96-T motif was negative although the MUD50 epitope was detected. An exact identification of the glycosylation sites on this motif could help to improve the reliability of this prediction neural network.

In this expression system the replacement of the actin15 (vegetative stage) promoter by a SP60 promoter, active during the multicellular stage, should demonstrate the developmentally regulated glycosylation on the GST-T motif of SP96. Under control of the SP60 promoter the expressed GST fusion protein of slug cells could not be labelled with MUD50, which strongly supports the conjecture that the different glycosylation is

[m] http://www.cbs.dtu.dk/service/NetOGlyc/

developmentally regulated. Unfortunately the amount of expressed protein was too low for further analysis.

An interesting phenomenon of peripheral importance was the observed GST dimer. It is still not understood why a GST fusion protein dimer is expressed in bacterially grown *D. discoideum* cells. This was observed under control of the actin15 promoter, from bacteria liquid-culture grown *Dictyostelium* cells, and with the SP60 promoter out of slug cells. The GST dimer can, however, be compared with the monomeric form, not be affinity-purified. This phenomenon has already been observed during expression of a soluble form of a human mast cell IgE receptor (Wilson, Slade and Williams, unpublished). Under the chosen conditions, the naturally occurring homodimeric form of GST should not be observed.

## 4.3 Structure/Function relationship

The link between the identified glycan structures on SP96 of the analysed *D. discoideum* strains and their phenotypical observations allows no definite conclusions of structure/ function relationship. However, some observations can be made.

The Fuc-Ser modification is apparently more important to the structural integrity of the spore coat than the fucose of the Fuc(α1-3)GlcNAc-α-1-P-Ser structure, since mutants carrying mutations in the *modC* or *modE* genes (which lack all or nearly all fucose containing structures) exhibit shorter viability on storage than strains carrying a *modD* mutation (which only lacks the phosphodiester-linked disaccharide containing fucose). How far the loss of the retaining GlcNAc-α-1-P-Ser modification influences the structural integrity is still unknown. The amount of Fuc-Ser with ~30 modified sites in SP96 is quite high and therefore could be expected to have a strong conformational influence on SP96.

Extraction studies done on X22 wild-type and HU2470 mutant spores demonstrated that SP96 is released in different forms from the spore coat (12, 51, 89, 121). It can be extracted in a soluble form by washing spores with buffer, a stronger associated form by boiling spores in high urea concentration, and a disulfide-linked form by boiling in high urea concentration under reducing conditions (89). All extracted forms of SP96 are modified with the MUD62 or the MUD166 epitope. SP96 with the MUD3 epitope, however, is mainly extracted in the strongly associated form and the covalently linked form. This clearly demonstrates the relevance of the Fuc-Ser modification on SP96 for the strong assembly of the spore coat.

A *modC* mutant with a complete loss of fucosylation showed after germination a 45 kDa breakdown product of SP96 which remained sedimentable with the coat. Here it was reasoned that SP96 was cleaved either because of the unmasking of a proteolytically sensitive site owing to the defect in glycosylation, or because of a greater accessibility of a protease to SP96 in the spore coat matrix (43). This is also supported by the observed activation of cysteine proteinases after germination of spores (97), in which wild-type SP96 is unaffected by this digestion (43).

Recently the Fuc-Ser modification was identified on two sites on a EGF-like domain from a human blood coagulation factor. The glycosylation was believed to affect the interaction of full-length coagulation factor with tissue factor by influencing its $Ca^{2+}$ binding affinity (73). Also recently identified was *O*-fucose as monosaccharide and elongated species in subsets of EGF modules on a large cell surface receptor, which was speculated to influence receptor-ligand interaction (88). Compared to these examples, SP96 is much more profoundly modified. *O*-linked glycans have also been shown to be important for the stability of glycoproteins. This is illustrated very well on a granulocyte colony stimulation factor (G-CSF) where one *O*-linked sugar protects the molecule against polymerisation at 37 °C and against heat denaturation (100). In this case it is possible that the sugar chain protects a cysteine (in the ionized state) against oxidising radical attack (60).

The high level "bottle brush" like glycosylation with the newly identified Fuc(α1-3)GlcNAc-α-1-P-Ser provides SP96 with a very hydrophilic tail. It is difficult to estimate the compactness of this tail on the spore surface, but it may be assumed that it sticks out of the spore coat and helps to protect the dormant amoeba from protease digestion and other environmental influences. The glycans may also be relevant for building up hydrogen bonds between water molecules to form a hydrated coat around the spore. This would prevent it drying out and further stabilise the pH of the environment surrounding the spore.

In mucins several of these "bottle brush" like glycosylation domains can be observed, where they are very resistant to protease digestion and partly function as spatial separation between functional domains, or project functional domains away from cell membrane (68). SP96 has on the C-terminal site of the heavily glycosylated serine-rich region an unmodified threonine-rich tail, which probably does not work as a functional domain.

The loss of the terminal Fuc of this disaccharide, as observed in the *modD* mutant strain, shows only minor phenotypical changes, and up to now no mutant with a lack of the remaining GlcNAc-α-1-P-Ser structure has been obtained. How far this is an experimental

problem, that means that a knock out of the relevant enzyme for phosphoglycosylation is lethal, or if no one has tried it yet is unknown. Cysteine proteinases of *D. discoideum* which also have phosphoglycans (GlcNAc-1-P-Ser or Man-6-P, respectively) are found in separate vesicles. It is therefore speculated that they are sorted out via their mutually exclusive carbohydrate modifications (40). Although it is possible that the phosphoglycosylation on SP96 functions as protection against proteolysis, it has to be mentioned that phosphodiester linkages are generally less stable than most of the direct glycosidic linkages to proteins. Mucins, for example, display reducing terminal GalNAc-elongated sugar chains having protective function. A phosphodiester-linked glycan is more acid labile than the linkage of a reducing sugar to the protein. However it is unknown how far this is relevant to the organism under real environmental condition. That spores are able to survive acidic treatments is shown on spores, which pass through the very acidic digestive systems of birds and are still able to germinate afterwards.

## 4.4 Further prospects

To get a better understanding of *O*-glycosylation function in *D. discoideum* further work should focus on the analysis of cysteine proteinases and other spore coat proteins. To get a better understanding about the function of phosphoglycosylation during the different stages of development it is necessary to create mutants, which lose this modification on spore coat proteins and cysteine proteinases, respectively. With a phenotypical analysis of these mutants it should be possible to assess functions for these modifications.

In summary, one can say that SP96 is the first spore coat protein of *D. discoideum* wich has been analysed in detail and found to have further newly identified phosphoglycan modification Fuc(α1-3)GlcNAc-α-1-P-Ser. SP96 seems to be modified by at least three different *O*-glycosylation forms (Fuc-Ser and Fuc(α1-3)GlcNAc-α-1-P-Ser and possibly a minor amount Fuc-P-Ser), which were localized at different clusters on the protein backbone. This is the first example of a protein with such diverse *O*-glycosylation. A knowledge of these glycan structures together with the phenotype correlation with mutations and antibody recognition makes it possible to speculate on the relationship between glycosylation and function of the spore coat proteins. Different glycan structures on SP96 have been assigned to antibody epitopes which is an important step forward to the characterisation of these often used tools.

# 5 References


1. Akalehiywot, T., and C. H. Siu. 1983. Phosphorylation of spore coat proteins of *Dictyostelium discoideum*. *Can. J. Biochem. Cell Biol.* 61:996-1001.
2. Alexander, S., E. Smith, L. Davis, A. A. Gooley, S. B. Por, L. Browne, and K. L. Williams. 1988. Characterisation of an antigenically related family of cell-type specific proteins implicated in slug migration in *Dictyostelium discoideum*. *Differentiation*. 38:82-90.
3. Apweiler, R., H. Hermjakob, and N. Sharon. 1999. On the frequency of protein glycosylation, as deduced from analysis of the SWISS-PROT database. *Biochim. Biophys. Acta.* 1473:4-8.
4. Bax, A., and D. G. Davis. 1985. Practical aspects of two-dimensional transverse NOE spectroscopy. *J. Magn . Reson.* 63:207-213.
5. Bock, K., and C. Peterson. 1983. Carbon-13 Nuclear Magnetic Resonance Spectroscopy of Monosaccharides. *Advances in Carbohydrate Chemistry and Biochemistry*. 41:27-66.
6. Brockhausen, I., D. Toki, J. Brockhausen, S. Peters, T. Bielfeldt, A. Kleen, H. Paulsen, M. Meldal, F. Hagen, and L. A. Tabak. 1996. Specificity of O-glycosylation by bovine colostrum UDP-GalNAc: polypeptide alpha-N-acetylgalactosaminyltransferase using synthetic glycopeptide substrates. *Glycoconj. J.* 13:849-856.
7. Brown, D. M. 1963. Phosphorylation. *In* Advances in Organic Chemistry. Vol. 3. R. Raphael, E. Taylor, and H. Wynberg. Interscience Publishers. 75-146.
8. Campbell, I. D., and P. Bork. 1993. *Curr. Opin. Struct. Biol.* 3:385-392.
9. Cardelli, J. A., J. Schatze, J. M. Bush, J. Richardson, D. Ebert, and H. Freeze. 1991. Biochemical and genetic analysis of the biosynthesis, sorting, and secretion of *Dictyostelium* lysosomal enzymes. *Dev. Genetics*. 11: 454-463.
10. Champion, A., A. A. Gooley, M. Callaghan, M. I. Carrin, R. L. Bernstein, E. Smith, and K. L. Williams. 1991. Immunodominant carbohydrate determinants in the multicellular stages of *Dictyostelium discoideum*. *J. Gen. Micro.* 137: 2431-2438.
11. Champion, A., K. Griffiths, A. A. Gooley, B. Y. Gonzalez, M. Gritzali, C. M. West, and K. L. Williams. 1995. Immunochemical, genetic and morphological comparison of fucosylation mutants of *Dictyostelium discoideum*. *Microbiology*. 141:785-797.
12. Champion, A. C. 1995. Aspects of glycosylation in *Dictyostelium discoideum*. *In* School of Biological Sciences. Macquarie-University, Sydney, Australia.
13. Chang, A. C., M. B. Slade, and K. L. Williams. 1990. Identification of the origin of replication of the eukaryote *Dictyostelium discoideum* nuclear plasmid Ddp2. *Plasmid*. 24:208-217.
14. Clausen, H., and E. P. Bennett. 1996. A family of UDP-GalNAc: polypeptide N-acetylgalactosaminyl-transferases control the initiation of mucin-type O-linked glycosylation. *Glycobiology*. 6:635-646.
15. Coffman, D. S., B. H. Leichtling, and H. V. Rickenberg. 1981. Phosphoproteins in *Dictyostelium discoideum*. *J Supramol Struct Cell Biochem*. 15:369-385.
16. Cooper, R., A. R. de Jesus, and G. A. Cross. 1993. Deletion of an immunodominant *Trypanosoma cruzi* surface glycoprotein disrupts flagellum-cell adhesion. *J. Cell Biol.* 122:149-156.

17. Cotter, D. A., and K. B. Raper. 1966. Spore germination in *Dictyostelium discoideum*. *Proc Natl Acad Sci U S A*. 56:880-887.

18. Cotter, D. A., and K. B. Raper. 1968. Spore germination in strains of *Dictyostelium discoideum* and other members of the Dictyosteliaceae. *J Bacteriol*. 96:1690-1695.

19. de Beer, T., J. F. Vliegenthart, A. Loffler, and J. Hofsteenge. 1995. The hexopyranosyl residue that is C-glycosidically linked to the side chain of tryptophan-7 in human RNase Us is alpha-mannopyranose. *Biochemistry*. 34:11785-11789.

20. Delaney, S. J., D. G. Wilkinson, and B. D. Hames. 1983. Phosphorylation of spore coat proteins during development of *Dictyostelium discoideum*. *Biochem J*. 212:699-703.

21. Devine, K. M., J. E. Bergmann, and W. F. Loomis. 1983. Spore coat proteins of *Dictyostelium discoideum* are packaged in prespore vesicles. *Dev. Biol.* 99:437-446.

22. Devine, K. M., J. H. Morrissey, and W. F. Loomis. 1982. Differential synthesis of spore coat proteins in prespore and prestalk cells of *Dictyostelium*. *Proc. Natl. Acad. Sci. U S A*. 79:7361-7365.

23. Dingermann, T., E. M. Troidl, M. Broker, and K. Nerke. 1991. Expression of human antithrombin III in the cellular slime mould *Dictyostelium discoideum*. *Appl. Microbiol. Biotechnol.* 35:496-503.

24. Dittrich, W., K. L. Williams, and M. B. Slade. 1994. Production and secretion of recombinant proteins in *Dictyostelium discoideum*. *Bio/Technology*. 12:614-618.

25. Dwek, R. A. 1996. Glycopinion. *Glycoconj. J.* 9:1-12.

26. Early, A. E., J. G. Williams, H. E. Meyer, S. B. Por, E. Smith, K. L. Williams, and A. A. Gooley. 1988. The structural characterisation of the *Dictyostelium discoideum* prespore specific gene D19 and of its product, the cell surface glycoprotein, PsA. *Molec. Cell Biol.* 8:3458-3466.

27. Eckert, V., P. Gerold, and R. T. Schwarz. 1997. GPI-anchors: structure and function. *In* Glycosciences: Status and Perspectives. H.J.Gabius and S.Gabius. Chapman & Hall GmbH, Weinheim, Germany. 223-238.

28. Elhammer, A. P., R. A. Poorman, E. Brown, L. L. Maggiora, J. G. Hoogerheide, and F. J. Kezdy. 1993. The specificity of UDP-GalNAc:polypeptide N-acetylgalactosaminyl-transferase as inferred from a database of *in vivo* substrates and from the *in vitro* glycosylation of proteins and peptides. *J. Biol. Chem.* 268:10029-10038.

29. Emslie, K. R., J. M. Miller, M. B. Slade, P. R. Dormitzer, H. B. Greenberg, and K. L. Williams. 1995. Expression of the rotavirus SA11 protein VP7 in the simple eukaryote *Dictyostelium discoideum*. *J. Virol.* 69:1747-1754.

30. Erdos, G. W., and C. M. West. 1989. Formation and organisation of the spore coat of *Dictyostelium discoideum*. *Exp. Mycology*. 13:169-182.

31. Fasel, N., C. Begdadi-Rais, M. Bernard, C. Bron, G. Corradin, and C. D. Reymond. 1992. *Dictyostelium discoideum* as an expression host for the circumsporozoite protein of *Plasmodium falciparum*. *Gene*. 111:157-163.

32. Fisher, P. R., and Z. Wilczynska. 1990. Isolation of signal transduction mutants of *D. discoideum* by random insertion mutagenesis with an integrating shuttle vector and a novel method of selecting transformants. *In* EMBO/ NSF Workshop 1990
Developmental and Molecular Biology of *Dictyostelium*.

33. Fosnaugh, K., D. Fuller, and W. F. Loomis. 1994. Structural roles of the spore coat proteins in *Dictyostelium discoideum*. *Dev. Biol.* 166:823-825.

34. Fosnaugh, K. L., and W. F. Loomis. 1989. Sequence of the *Dictyostelium discoideum* spore coat gene SP96. *Nucleic. Acids Res.* 17:9489.

35. Fosnaugh, K. L., and W. F. Loomis. 1989. Spore coat genes SP60 and SP70 of *Dictyostelium discoideum*. *Mol. Cell Biol.* 9:5215-5218.

36. Free, S. J., and R. T. Schimke. 1978. Effects of post-translational modification mutation on different developmentally regulated glycosidases in *Dictyostelium discoideum*. *J. Biol. Chem.* 253:4107-4111.

37. Free, S. J., R. T. Schimke, H. Freeze, and W. F. Loomis. 1978. Characterization and genetic mapping of *modA*, a mutation in the post-translational modification of the glycosidases of *Dictyostelium discoideum*. *J. Biol. Chem.* 253:4102-4106.

38. Freeze, H. 1986. Modifications of lysosomal enzymes in *Dictyostelium discoideum. Mol. Cell. Biochem.* 72: 47-65.

39. Freeze, H. H. 1992. Developmental Glycobiology of *Dictyostelium discoideum*. *In Cell Surface Carbohydrates and Cell Development*. M. Fukuda. CRC Press Inc, Boca Raton, USA. p285-317.

40. Freeze, H. H. 1997. *Dictyostelium discoideum* glycoproteins: using a model system for organismic glycobiology. *In* New Comprehensive Biochemistry. Vol. 29B Glycoproteins II. J. Montreuil, Vliengenthart, J.F.G. and Schachter H. Elsevier, Cambridge. 89-121.

41. Freeze, H. H., and A. L. Miller. 1981. Mod A: a post-translational mutation affecting phosphorylated and sulfated glycopeptides in *Dictyostelium discoideum*. *Mol. Cell. Biochem.* 35:17-27.

42. Freeze, H. H., R. Yeh, A. L. Miller, and S. Kornfeld. 1983. The mod A mutant of *Dictyostelium discoideum* is missing the alpha 1,3-glucosidase involved in asparagine-linked oligosaccharide processing. *J. Biol. Chem.* 258:14880-14884.

43. Gonzalez-Yanes, B., R. B. Mandell, M. Girard, S. Henry, O. Aparicio, M. Gritzali, R. D. Brown, G. W. Erdos, and C. M. West. 1989. The spore coat of a fucosylation mutant in *Dictyostelium discoideum*. *Dev. Biol.* 133:576-587.

44. Gooley, A. A. 1992. Characterisation of the glycoprotein PsA. School of Biological Sciences, Macquarie University, Sydney, Australia, PhD Thesis.

45. Gooley, A. A., R. Marschalek, and K. L. Williams. 1992. Size polymorphisms due to changes in the number of O-glycosylated tandem repeats in the *Dictyostelium discoideum* glycoprotein PsA. *Genetics*. 130:749-756.

46. Gooley, A. A., and K. L. Williams. 1994. Towards characterizing O-glycans: the merits of *in vivo* and *in vitro* approaches in seeking peptide motifs specifying O-glycosylation sites. *Glycobiology*. 4:413-417.

47. Gooley, A. A., and K. L. Williams. 1997. How to find, identify and quantitate the sugars on proteins. *Nature*. 385:557-559.

48. Grant, W. N., and K. L. Williams. 1983. Monoclonal antibody characterisation of slime sheath: the extra-cellular matrix of *Dictyostelium discoideum*. *EMBO J.* **2**:935-940.

49. Gregg, J. H., and G. C. Karp. 1978. Patterns of cell differentiation revealed by L-[3H]fucose incorporation in *Dictyostelium*. *Exp. Cell. Res.* 112:31-46.

50. Gregg, J. M., M. Krefft, A. Haas-Kraus, and K. L. Williams. 1982. Antigenic differences between prespore cells of *Dictyostelium discoideum* and *Dictyostelium mucoroides* using monoclonal antibodies. *Experimental Cell Research*. 142:229-233.

51. Griffiths, K. R. 1996. O-glycosylation and spore maturation in *Dictyostelium discoideum*. *In* School of Biological Sciences. Macquarie University, Sydney, Australia.

52. Guevara, J., B. Espinosa, E. Zenteno, L. Vazguez, J. Luna, G. Perry, and R. Mena. 1998. Altered glycosylation pattern of proteins in Alzheimer disease. *J. Neuropathol. Exp. Neurol.* 57:905-914.

53. Gupta, R., H. Birch, K. Rapacki, S. Brunak, and J. E. Hansen. 1999. O-GLYCBASE version 4.0: a revised database of O-glycosylated proteins. *Nucleic Acids Res.* 27:370-372.

54. Gustafson, G. L., and L. A. Milner. 1980. Occurrence of N-acetylglucosamine-1-phosphate in proteinase 1 from *Dictyostelium discoideum*. *J. Biol. Chem.* 255:7208-7210.

55. Haberstroh, L., and R. A. Firtel. 1990. A spatial gradient of expression of a cAMP-regulated prespore cell-type-specific gene in Dictyostelium. *Genes Dev*. 4:596-612.

56. Hanahan, D. 1983. Studies on transformation of *Escherichia coli* with plasmids. *J Mol Biol*. 166:557-580.

57. Hansen, J. E., O. Lund, N. Tolstrup, A. A. Gooley, K. L. Williams, and S. Brunak. 1998. NetOglyc: prediction of mucin type O-glycosylation sites based on sequence context and surface accessibility. *Glycoconj. J.* 15:115-130.

58. Harris, R. J., V. T. Ling, and M. W. Spellman. 1992. O-linked fucose is present in the first epidermal growth factor domain of factor XII but not protein C. *J. Biol. Chem.* 267:5102-5107.

59. Harris, R. J., and M. W. Spellman. 1993. O-linked fucose and other post-translational modifications unique to EGF modules. *Glycobiology*. 3:219-224.

60. Hasegawa, M. 1993. A thermodynamic model for denaturation of granulocyte colony-stimulating factor: O-linked sugar chain suppresses not the triggering deprotonation but the succeeding denaturation. *Biochim. Biophys. Acta.* 1203:295-297.

61. Haynes, P., M. A. J. Ferguson, A. A. Gooley, J. W. Redmond, and K. L. Williams. 1993. Post-translational modification of PsA, a cell surface glycoprotein from *Dictyostelium discoideum*: structure of the glycosylphosphatidylinositol membrane anchor and partial structure of O-linked oligosaccharides attached to the carboxyl terminal domain. *Eur. J. Biochem.* 216:729-737.

62. Haynes, P. A. 1998. Phosphoglycosylation: a new structural class of glycosylation? *Glycobiology*. 8:1-5.

63. Haynes, P. A., M. A. Ferguson, and G. A. Cross. 1996. Structural characterization of novel oligosaccharides of cell-surface glycoproteins of *Trypanosoma cruzi*. *Glycobiology*. 6:869-878.

64. Hofsteenge, J., D. R. Muller, T. de Beer, A. Loffler, W. J. Richter, and J. F. Vliegenthart. 1994. New type of linkage between a carbohydrate and a protein: C-glycosylation of a specific tryptophan residue in human RNase Us. *Biochemistry*. 33:13524-13530.

65. Ilg, T., D. Harbecke, M. Wiese, and P. Overath. 1993. Monoclonal antibodies directed against *Leishmania* secreted acid phosphatase and lipophosphoglycan. Partial characterization of private and public epitopes. *Eur. J. Biochem.* 217:603-615.

66. Ilg, T., Y. D. Stierhof, D. Craik, R. Simpson, E. Handman, and A. Bacic. 1996. Purification and structural characterization of a filamentous, mucin- like proteophosphoglycan secreted by *Leishmani*a parasites. *J. Biol. Chem.* 271:21583-21596.

67. Ilg, T., Y. D. Stierhof, M. Wiese, M. J. McConville, and P. Overath. 1994. Characterization of phosphoglycan-containing secretory products of *Leishmania*. *Parasitology*. 108:63-71.

68. Jentoft, N. 1990. Why are proteins O-glycosylated? *TIBS*. 15:291-294.

69. Jung, E. 1998. Defining acceptor motifs for O-glycosylation in *Dictyostelium discoideum*. *In* School of Biological Sciences. Macquarie University, Sydney, Australia.

70. Jung, E., A. A. Gooley, N. H. Packer, P. Karuso, and K. L. Williams. 1998. Rules for the addition of O-linked N-acetylglucosamine to secreted proteins in *Dictyostelium discoideum--in vivo* studies on glycosylation of mucin MUC1 and MUC2 repeats. *Eur. J. Biochem.* 253:517-524.

71. Jung, E., A. A. Gooley, N. H. Packer, M. B. Slade, K. L. Williams, and W. Dittrich. 1997. An *in vivo* approach for the identification of acceptor sites for O- glycosyltransferases: motifs for the addition of O-GlcNAc in *Dictyostelium discoideum*. *Biochemistry*. 36:4034-4040.

72. Jung, E., and K. L. Williams. 1997. The production of recombinant glycoproteins with special reference to simple eukaryotes including *Dictyostelium discoideum*. *Biotechnol. Appl. Biochem.* 25:3-8.

73. Kao, Y. H., G. F. Lee, Y. Wang, M. A. Starovasnik, R. F. Kelley, M. W. Spellman, and L. Lerner. 1999. The effect of O-fucosylation on the first EGF-like domain from human blood coagulation factor VII. *Biochemistry*. 38:7097-7110.

74. Kay, L. E., P. Keifer, and T. Saarinen. 1992. Pure absorption gradient enhanced heteronuclear single quantum correlation spectroscopy with improved sensitivity. *J. Am. Chem. Soc.* 114:10663-10665.

75. Kessin, R. H., G. G. Gundersen, V. Zaydfudim, and M. Grimson. 1996. How cellular slime molds evade nematodes. *Proc. Natl. Acad. Sci. U S A*. 93:4857-4861.

76. Khyse-Anderson, J. 1984. Electroblotting of multiple gels: a simple apparatus without buffer tank for rapid transfer of protein from polyacrylamide to nitrocellulose. *J. Biochem. Biophys. Methods.* 10:203-209.

77. Kornfeld, R., and S. Kornfeld. 1985. Assembly of asparagine-linked oligosaccharides. *Annu. Rev. Biochem.* 54:631-634.

78. Laemmli, U. K. 1970. Cleavage of the structural proteins during the assembly of the head of bacteriophage *T4*. *Nature*. 227: 680-685.

79. Lam, T. Y., and C. H. Siu. 1981. Synthesis of stage-specific glycoproteins in *Dictyostelium discoideum* during development. *Dev. Biol.* 83:127-137.

80. Lanzetta, P. A., L. J. Alvarez, P. S. Reinach, and O. A. Candia. 1979. An improved assay for nanomole amounts of inorganic phosphate. *Anal. Biochem.* 100:95-97.

81. Lemieux, R. U., K. Bock, L. T. J. Delbaere, S. Koto, and V. S. Rao. 1980. *Can. J. Chem.* 58:631-653.

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

83. March, J. 1992. Advanced Organic Chemistry: Reactions, Mechanisms and Structure. John Wiley & Sons, New York. 1174.

84. McGuire, V., and S. Alexander. 1996. PsB multiprotein complex of *Dictyostelium discoideum*. Demonstration of cellulose binding activity and order of protein subunit assembly. *J. Biol. Chem.* 271:14596-14603.

85. McTigue, M. A., S. L. Bernstein, D. R. Williams, and J. A. Tainer. 1995. Purification and crystallization of a schistosomal glutathione S-transferase. *Proteins*. 22:55-57.

86. Mehta, D. P., J. R. Etchison, R. Wu, and H. H. Freeze. 1997. UDP-GlcNAc:Ser-protein N-acetylglucosamine-1-phosphotransferase from *Dictyostelium discoideum* recognizes serine-containing peptides and eukaryotic cysteine proteinases. *J Biol Chem*. 272:28638-28645.

87. Mehta, D. P., M. Ichikawa, P. V. Salimath, J. R. Etchison, R. Haak, A. Manzi, and H. H. Freeze. 1996. A lysosomal cysteine proteinase from *Dictyostelium discoideum* contains N-acetylglucosamine-1-phosphate bound to serine but not mannose-6- phosphate on N-linked oligosaccharides. *J. Biol. Chem.* 271:10897-10903.

88. Moloney, D. J., L. H. Shair, F. M. Lu, J. Xia, R. Locke, K. L. Matta, and R. S. Haltiwanger. 2000. Mammalian Notch1 is modified with two unusual forms of O-linked glycosylation found on epidermal growth factor-like modules. *Journal of Biological Chemistry*. 275:9604-9611.

89. Mreyen, M. 1996. Untersuchung posttranslationaler Modifikationen in *D. discoideum*. *In* Biochemie. Bergische Universität Wuppertal, Wuppertal, Deutschland.

90. Mreyen, M., A. Champion, S. Srinivasan, P. Karuso, K. L. Williams, and N. H. Packer. 2000. Multiple O-glycoforms on the spore coat protein SP96 in *Dictyostelium discoideum*. Fuc(alpha1-3)GlcNAc-alpha-1-P-Ser is the major modification. *J. Biol. Chem.* 275:12164-12174.

91. Muller, S., S. Goletz, N. Packer, A. Gooley, A. M. Lawson, and F. G. Hanisch. 1997. Localization of O-glycosylation sites on glycopeptide fragments from lactation-associated MUC1. All putative sites within the tandem repeat are glycosylation targets *in vivo*. *J. Biol. Chem.* 272:24780-24793.

92. Nehrke, K., F. K. Hagen, and L. A. Tabak. 1996. Charge distribution of flanking amino acids influences O-glycan acquisition *in vivo*. *J. Biol. Chem.* 271:7061-7065.

93. Nellen, W., C. Silan, and R. A. Firtel. 1984. DNA-mediated transformation in *Dictyostelium discoideum*: regulated expression of an actin gene fusion. *Mol. Cell. Biol.* 4:2890-2898.

94. Nishimori, I., N. R. Johnson, S. D. Sanderson, F. Perini, K. Mountjoy, R. L. Cerny, M. L. Gross, and M. A. Hollingsworth. 1994. Influence of acceptor substrate primary amino acid sequence on the activity of human UDP-N-acetylgalactosamine:polypeptide N-acetylgalactosaminyltransferase. Studies with the MUC1 tandem repeat. *J. Biol. Chem.* 269:16123-16130.

95. Nishimori, I., F. Perini, K. P. Mountjoy, S. D. Sanderson, N. Johnson, R. L. Cerny, M. L. Gross, J. D. Fontenot, and M. A. Hollingsworth. 1994. N-acetylgalactosamine glycosylation of MUC1 tandem repeat peptides by pancreatic tumor cell extracts. *Cancer Res.* 54:3738-3744.

96. North, M. J. 1982. Proteolytic activities in *Dictyostelium discoideum* detected with chromogenic peptide substrates. *Experimental Mycology*. 6:345-352.

97. North, M. J., D. A. Cotter, and K. J. Franek. 1990. *Dictyostelium discoideum* spore germination: increases in proteinase activity are not directly coupled to the emergence of myxamoebae. *J. Gen. Micro.* 136:835-840.

98. North, M. J., K. J. Franek, and D. A. Cotter. 1990. Differential secretion of *Dictyostelium discoideum* proteinases. *J. Gen. Micro.* 136:827-833.

99. OConnell, F. K. Hagen, and L. A. Tabak. 1992. The influence of flanking sequence on the O-glycosylation of threonine in vitro. *J Biol Chem*. 267:25010-25018.

100. Oh-eda, M., M. Hasegawa, K. Hattori, H. Kuboniwa, T. Kojima, T. Orita, K. Tomonou, T. Yamazaki, and N. Ochi. 1990. O-linked sugar chain of human granulocyte colony-stimulating factor protects it against

polymerization and denaturation allowing it to retain its biological activity. *J Biol Chem*. 265:11432-11435.

101. Ord, T., C. Adessi, L. Wang, and H. H. Freeze. 1997. The cysteine proteinase gene cprG in *Dictyostelium discoideum* has a serine-rich domain that contains GlcNAc-1-P. *Arch. Biochem. Biophys.* 339:64-72.
102. Orlowski, M., and W. F. Loomis. 1979. Plasma membrane proteins of *Dictyostelium*: the spore coat proteins. *Dev. Biol.* 71:297-307.
103. Packer, N. H., M. A. Lawson, D. R. Jardine, and J. W. Redmond. 1998. A general approach to desalting oligosaccharides released from glycoproteins. *Glycoconj. J.* 15:737-747.
104. Palmer, A. G., III, J. Cavanagh, and P. Wright. 1991. Sensitivity improvement in proton-detected two-dimensional heteronuclear correlation NMR spectroscopy. *J. Magn. Reson.* 93:151-170.
105. Paulson, J. C., and K. J. Colley. 1989. Glycosyltransferases. Structure, localization, and control of cell type-specific glycosylation. *J. Biol. Chem.* 264:17615-17618.
106. Piotto, M., V. Saudek, and V. Sklenar. 1992. *J. Biomol. NMR*. 2:661-666.
107. Raper, K. B. 1935. *Dictyostelium discoideum*, a new species of slime mold from decaying forest leaves. *J. Agr. Res.* 50:135-147.
108. Richardson, D. L., and W. F. Loomis. 1992. Disruption of the sporulation-specific gene spiA in *Dictyostelium discoideum* leads to spore instability. *Genes Dev.* 6:1058-1070.
109. Riley, G. R., C. M. West, and E. J. Henderson. 1993. Cell differentiation in *Dictyostelium discoideum* controls assembly of protein-linked glycans. *Glycobiology*. 3:165-177.
110. Rudd, P. M., and R. A. Dwek. 1997. Glycosylation: heterogeneity and the 3D structure of proteins. *Crit. Rev. Biochem. Mol. Biol.* 32:1-100.
111. Saiki, R. K., S. Scharf, F. Faloona, K. B. Mullis, G. T. Horn, H. A. Erlich, and N. Arnheim. 1985. Enzymatic amplification of beta-globin genomic sequences and restriction site analysis for diagnosis of sickle cell anemia. *Science*. 230:1350-1354.
112. Schleucher, J., M. Schwendinger, M. Sattler, and P. Schmidt. 1994. A general enhancement scheme in heteronuclear multidimensional NMR employing pulsed field gradients. *J. Biomol. NMR*. 4:301-306.
113. Schopohl, D., A. Muller-Taubenberger, B. Orthen, H. Hess, and W. Reutter. 1992. Purification and properties of a secreted and developmentally regulated alpha-L-fucosidase from *Dictyostelium discoideum*. *J. Biol. Chem.* 267:2400-2405.
114. Sharon, N. 1975. Complex Carbohydrates: Their Chemistry, Biosynthesis and Function. Addison-Wesley Publishing Company.
115. Sharon, N., and H. Lis. 1997. Glycoproteins: Structure and Function. *In* Glycosciences: Status and Perspectives. H. J. a. G. Gabius, S. Chapman & Hall GmbH, Weinheim. 133-154.
116. Sklenar, V., M. Piotto, and Leppik. 1993. Gradient-tailored water suppression for proton-nitrogen-15 HSQC experiments optimized to retain full sensitivity. *J. Magn. Reson., Ser. A*. 102:241-245.
117. Smith, E., A. A. Gooley, G. C. Hudson, and K. L. Williams. 1989. Glycoproteins that exhibit extensive polymorphisms in *Dictyostelium discoideum*. *Genetics*. 122:59-64.
118. Souza, G. M., J. Hirai, D. P. Mehta, and H. H. Freeze. 1995. Identification of two novel *Dictyostelium discoideum* cysteine proteinases that carry N-acetylglucosamine-1-P-modification. *J. Biol. Chem.* 270:28938-28945.

119. Srikrishna, G., L. Wang, and H. H. Freeze. 1998. Fucosebeta-1-P-Ser is a new type of glycosylation: using antibodies to identify a novel structure in *Dictyostelium discoideum* and study multiple types of fucosylation during growth and development. *Glycobiology*. 8:799-811.

120. Srinivasan, S., H. Alexander, and S. Alexander. 1999. The prespore vesicles of *Dictyostelium discoideum*. Purification, characterization, and developmental regulation. *J. Biol. Chem.* 274:35823-35831.

121. Srinivasan, S., K. R. Griffiths, V. McGuire, A. Champion, K. L. Williams, and S. Alexander. 2000. The cellulose-binding activity of the PsB multiprotein complex is required for proper assembly of the spore coat and spore viability in *Dictyostelium discoideum*. *Microbiology*. 146:1829-1839.

122. Stadie, T. R., W. Chai, A. M. Lawson, P. G. Byfield, and F. G. Hanisch. 1995. Studies on the order and site specificity of GalNAc transfer to MUC1 tandem repeats by UDP-GalNAc: polypeptide N-acetylgalactosaminyltransferase from milk or mammary carcinoma cells. *Eur. J. Biochem.* 229:140-147.

123. Strous, G. J., and J. Dekker. 1992. Mucin-type glycoproteins. *Crit Rev Biochem Mol Biol*. 27:57-92.

124. Sussman, M. 1987. Cultivation and synchronous morphogenesis of *Dictyostelium* under controlled experimental conditions. *Methods Cell Biology*. 28: 9-29.

125. Tasaka, M., M. Hasegawa, T. Ozaki, M. Iwabuchi, and I. Takeuchi. 1990. Isolation and characterization of spore coat protein (sp96) gene of *Dictyostelium discoideum*. *Cell. Differ. Dev.* 31:1-9.

126. Tschursin, E., G. R. Riley, and E. J. Henderson. 1989. Differential regulation of glycoprotein sulfation and fucosylation during growth of *Dictyostelium discoideum*. *Differentiation*. 40:1-9.

127. Van den Steen, P., P. M. Rudd, R. A. Dwek, and G. Opdenakker. 1998. Concepts and principles of O-linked glycosylation. *Crit. Rev. Biochem. Mol. Biol.* 33:151-208.

128. Varki, A. 1993. Biological roles of oligosaccharides: all of the theories are correct. *Glycobiology*. 3:97-130.

129. Varki, A., and M. J.D. 1995. Oligosaccharides in vertebrate development. *Sem. Dev.Biol*. 6:127-138.

130. Verma, M., and E. A. Davidson. 1994. Mucin genes: structure, expression and regulation. *Glycoconj. J.* 11:172-179.

131. Voet, L., M. Krefft, M. Bruderlein, and K. L. Williams. 1985. Flow cytometer study of anterior-like cells in *Dictyostelium discoideum*. *J. Cell Science*. 75:423-435.

132. Voith, G., and T. Dingermann. 1995. Expression of the human muscarinic receptor gene m2 in *Dictyostelium discoideum*. *Biotechnology (N Y)*. 13:1225-1229.

133. Wandall, H. H., H. Hassan, E. Mirgorodskaya, A. K. Kristensen, P. Roepstorff, E. P. Bennett, P. A. Nielsen, M. A. Hollingsworth, J. Burchell, J. Taylor-Papadimitriou, and H. Clausen. 1997. Substrate specificities of three members of the human UDP-N-acetyl-alpha-D-galactosamine:Polypeptide N-acetylgalactosaminyltransferase family, GalNAc-T1, -T2, and -T3. *J. Biol. Chem.* 272:23503-23514.

134. Wang, Y., N. Agrwal, A. E. Eckhardt, R. D. Stevens, and R. L. Hill. 1993. The acceptor substrate specificity of porcine submaxillary UDP-GalNAc:polypeptide N-acetylgalactosaminyltransferase is dependent on the amino acid sequences adjacent to serine and threonine residues. *J. Biol. Chem.* 268:22979-22983.

135. Wang, Y., G. F. Lee, R. F. Kelley, and M. W. Spellman. 1996. Identification of a GDP-L-fucose:polypeptide fucosyltransferase and enzymatic addition of O-linked fucose to EGF domains. *Glycobiology*. 6:837-842.

136. Watson, N., V. McGuire, and S. Alexander. 1994. The PsB glycoprotein complex is secreted as a preassembled precursor of the spore coat in *Dictyostelium discoideum*. *J. Cell. Sci.* 107:2567-2579.

137. Watson, N., K. L. Williams, and S. Alexander. 1993. A developmentally regulated glycoprotein complex from *Dictyostelium discoideum*. *J Biol. Chem.* 268:22634-22641.

138. Watts, D. J., and J. M. Ashworth. 1970. Growth of myxamoebae of the cellular slime mould *Dictyostelium discoideum* in axenic culture. *J. Biochem.* 119.: 171-174.

139. West, C. M., and G. W. Erdos. 1988. The expression of glycoproteins in the extracellular matrix of the cellular slime mould *Dictyostelium discoideum. Cell Differentiation*. 23:1-16.

140. West, C. M., and G. W. Erdos. 1990. Formation of the *Dictyostelium* spore coat. *Develop. Genet.* 11:492-506.

141. West, C. M., G. W. Erdos, and R. Davis. 1986. Glycoantigen expression is regulated both temporally and spatially during development in the cellular slime moulds *Dictyostelium discoideum* and *D. mucoroides. Mol. Cell. Biochem.* 72:121-140.

142. West, C. M., and W. F. Loomis. 1985. Absence of a carbohydrate modification does not affect the level or subcellular localisation of three membrane glycoproteins in *mod*B mutants of *Dictyostelium discoideum*. *J. Biol. Chem.* 260:13803-13809.

143. West, C. M., J. Mao, H. van der Wel, G. W. Erdos, and Y. Zhang. 1996. SP75 is encoded by the DP87 gene and belongs to a family of modular *Dictyostelium discoideum* outer layer spore coat proteins. *Microbiology*. 142:2227-2243.

144. Widdowson, D. C., J. A. Proffitt, P. S. Jagger, A. J. Richards, and B. D. Hames. 1990. Developmental expression and characterization of the gene encoding spore coat protein SP60 in *Dictyostelium discoideum*. *Mol. Microbiol.* 4:951-960.

145. Wiese, M., T. Ilg, F. Lottspeich, and P. Overath. 1995. Ser/Thr-rich repetitive motifs as targets for phosphoglycan modifications in *Leishmania mexicana* secreted acid phosphatase. *Embo J.* 14:1067-1074.

146. Wilczynska, Z., and P. R. Fisher. 1994. Analysis of a complex plasmid insertion in a phototaxis-deficient transformant of *Dictyostelium discoideum* selected on a *Micrococcus luteus* lawn. *Plasmid*. 32:182-194.

147. Wilkinson, D. G., J. Wilson, and B. D. Hames. 1983. Synthesis of spore proteins during development of *Dictyostelium discoideum*. *Biochem. J.* 215:567-574.

148. Williams, K. L., and P. C. Newell. 1976. A genetic study of aggregation in the cellular slime mould *Dictyostelium discoideum* using complementation analysis. *Genetics*. 82:287-307.

149. Wilson, I. B. H., Y. Gavel, and G. von Heijne. 1991. Amino acid distributions around O-linked glycosylation sites. *Biochem. J.* 275:529-534.

150. Yan, J. X., N. H. Packer, L. Tonell, K. Ou, M. R. Wilkins, J.-C. Sanchez, A. A. Gooley, D. F. Hochstrasser, and K. L. Williams. 1997. High sample throughput phosphoamino acid analysis of proteins separated by one and two-dimensional gel electrophoresis. *J. Chromatogr.* 764:201-210.

151. Yan, J. X., M. R. Wilkins, K. Ou, A. A. Gooley, K. L. Williams, J. C. Sanchez, O. Golaz, C. Pasquali, and D. F. Hochstrasser. 1996. Large-scale amino-acid analysis for proteome studies. *J. Chromatogr. A*. 736:291-302.

152. Yoshida, A., M. Suzuki, H. Ikenaga, and M. Takeuchi. 1997. Discovery of the shortest sequence motif for high level mucin-type O-glycosylation. *J. Biol. Chem.* 272:16884-16888.

153. Yoshida, M., G. Fuse, T. Matui, and S. Ouchi. 1992. Identification of sialic acids in cell adhesion molecule, contact site A from *Dictyostelium discoideum*. *Biochem. Biophys. Res. Comm.* 188:794-798.

154. Yoshida, M., S. Yokota, and S. Ouchi. 1997. Characterization and distribution of O-glycosylated carbohydrates in the cell adhesion molecule, contact site A, from *Dictyostelium discoideum*. *Exp. Cell. Res.* 230:393-398.

155. Zachara, N. 1994. Identification of a single O-linked N-acetylglucosamine on recombinant Prespore specific Antigen (PsA), a glycoprotein produced by *Dictyostelium discoideum*. School of Biological Sciences, Macquarie University, Sydney, Australia, Hons Thesis.

156. Zachara, N. 1998. Sequencing Proteins, Sequencing Sugars: A Unified Approach. *In* School of Biological Sciences. Macquarie University, Sydney, NSW, Australia.

157. Zachara, N. E., N. H. Packer, M. D. Temple, M. B. Slade, D. R. Jardine, P. Karuso, C. J. Moss, B. C. Mabbutt, P. M. Curmi, K. L. Williams, and A. A. Gooley. 1996. Recombinant prespore-specific antigen from Dictyostelium discoideum is a beta-sheet glycoprotein with a spacer peptide modified by O-linked N- acetylglucosamine. *Eur. J. Biochem.* 238:511-518.

158. Zhang, Y., R. D. Brown, Jr., and C. M. West. 1998. Two proteins of the *Dictyostelium* spore coat bind to cellulose *in vitro*. *Biochemistry*. 37:10766-10779.

# 6 Appendix

## NetPhos 2.0 Prediction Results

**Name: SP96 Length: 600**

```
MRVLLVLVACLTYFSGGALAQSCSSYSGDNCPSTCFQGSYDIPCGAQVKYCTEMKDNCGEGGDVKCWKDGSNLPVQTWSS
CVPSELFGPNGKFKPSEIPNSSNCPTNCENGVEWVNLCGLSCDAKTACCPDVCQCKGGQTSGGSTTGSQTSGGSTSGGST
TGSQTSGGSTTGSQTSGSQTSAGSCSNTQCPNGFYCQVQGNNAVCVPQQSSTSGGHQNDPCDTVQCPYGYSCESRDGFEA
KCTRDEDEPTHRPTHRPKPPHDSDKYLCDNVHCPRGYKCNAKNGVAKCIAGYEIPRVCRNIQCPTGYRCEDHNRNPICVL
EERENPDNCLTCNDVNCEASGLVCVMTRARCKVGAAKCCDVQPTCIKPSTIAGSTIASIASTIASTGSTGATSPCSVAQC
PTGYVCVAQNNVAVSLPRPTTTTGSTSDSSALGSTSESSASGSSAVSSSASGSSAASSSPSSSAASSSPSSSAASSSPSS
SAASSSPSSSASSSSSPSSSASSSSAPSSSASSSSAPSSSASSSSASSSSASSAATTAATTIATTAATTTATTTATTATT
TATTTATTTAATIATTTAATTTATTTATTATTTATTTATS

.......................SY.............Y.....................................
..............S.........................................TS.........S..S.S..S.
.....S......S..S.S.T...S..........Y..................S.....................S......
..T......T...T........S..Y..........Y.....................................
.....................................................S....S.............S....S.......
...Y.................TT..S.S..S...S.S.SS.S..S..S.S.S.....SSS.SSS..SSS.SSS..SSS.SS
S..SSS.SSS.SSSSS.SSS.SSSS..SSS.SSSS..SSS.SSSS.SSSS.SS.........................
.......T.............................S
```

Ser: 88 Thr: Tyr:

Ser: 8 Thr: 6 Tyr:

Serine predictions

| Name | Pos | Context | Score | Pred |
|---|---|---|---|---|
| | | v | | |
| SP96 | 15 | LTYFSGGAL | 0.019 | . |
| SP96 | 22 | ALAQSCSSY | 0.066 | . |
| SP96 | 24 | AQSCSSYSG | 0.056 | . |
| SP96 | 25 | QSCSSYSGD | 0.584 | *S* |
| SP96 | 27 | CSSYSGDNC | 0.432 | . |
| SP96 | 33 | DNCPSTCFQ | 0.041 | . |
| SP96 | 39 | CFQGSYDIP | 0.092 | . |
| SP96 | 71 | WKDGSNLPV | 0.002 | . |
| SP96 | 79 | VQTWSSCVP | 0.010 | . |
| SP96 | 80 | QTWSSCVPS | 0.030 | . |
| SP96 | 84 | SCVPSELFG | 0.262 | . |
| SP96 | 96 | KFKPSEIPN | 0.861 | *S* |
| SP96 | 101 | EIPNSSNCP | 0.271 | . |
| SP96 | 102 | IPNSSNCPT | 0.026 | . |
| SP96 | 121 | LCGLSCDAK | 0.004 | . |
| SP96 | 141 | GGQTSGGST | 0.925 | *S* |
| SP96 | 144 | TSGGSTTGS | 0.353 | . |
| SP96 | 148 | STTGSQTSG | 0.454 | . |
| SP96 | 151 | GSQTSGGST | 0.988 | *S* |
| SP96 | 154 | TSGGSTSGG | 0.789 | *S* |
| SP96 | 156 | GGSTSGGST | 0.939 | *S* |
| SP96 | 159 | TSGGSTTGS | 0.854 | *S* |

| | | | | |
|---|---|---|---|---|
| SP96 | 163 | STTGSQTSG | 0.454 | . |
| SP96 | 166 | GSQTSGGST | 0.987 | *S* |
| SP96 | 169 | TSGGSTTGS | 0.353 | . |
| SP96 | 173 | STTGSQTSG | 0.614 | *S* |
| SP96 | 176 | GSQTSGSQT | 0.987 | *S* |
| SP96 | 178 | QTSGSQTSA | 0.867 | *S* |
| SP96 | 181 | GSQTSAGSC | 0.291 | . |
| SP96 | 184 | TSAGSCSNT | 0.699 | *S* |
| SP96 | 186 | AGSCSNTQC | 0.008 | . |
| SP96 | 210 | VPQQSSTSG | 0.049 | . |
| SP96 | 211 | PQQSSTSGG | 0.311 | . |
| SP96 | 213 | QSSTSGGHQ | 0.635 | *S* |
| SP96 | 231 | PYGYSCESR | 0.008 | . |
| SP96 | 234 | YSCESRDGF | 0.667 | *S* |
| SP96 | 263 | PPHDSDKYL | 0.802 | *S* |
| SP96 | 340 | NCEASGLVC | 0.004 | . |
| SP96 | 369 | CIKPSTIAG | 0.767 | *S* |
| SP96 | 374 | TIAGSTIAS | 0.629 | *S* |
| SP96 | 378 | STIASIAST | 0.083 | . |
| SP96 | 381 | ASIASTIAS | 0.447 | . |
| SP96 | 385 | STIASTGST | 0.481 | . |
| SP96 | 388 | ASTGSTGAT | 0.867 | *S* |
| SP96 | 393 | TGATSPCSV | 0.958 | *S* |
| SP96 | 396 | TSPCSVAQC | 0.037 | . |
| SP96 | 415 | NVAVSLPRP | 0.497 | . |
| SP96 | 425 | TTTGSTSDS | 0.996 | *S* |
| SP96 | 427 | TGSTSDSSA | 0.985 | *S* |
| SP96 | 429 | STSDSSALG | 0.204 | . |
| SP96 | 430 | TSDSSALGS | 0.838 | *S* |
| SP96 | 434 | SALGSTSES | 0.987 | *S* |
| SP96 | 436 | LGSTSESSA | 0.894 | *S* |
| SP96 | 438 | STSESSASG | 0.665 | *S* |
| SP96 | 439 | TSESSASGS | 0.970 | *S* |
| SP96 | 441 | ESSASGSSA | 0.987 | *S* |
| SP96 | 443 | SASGSSAVS | 0.144 | . |
| SP96 | 444 | ASGSSAVSS | 0.705 | *S* |
| SP96 | 447 | SSAVSSSAS | 0.975 | *S* |
| SP96 | 448 | SAVSSSASG | 0.424 | . |
| SP96 | 449 | AVSSSASGS | 0.897 | *S* |
| SP96 | 451 | SSSASGSSA | 0.936 | *S* |
| SP96 | 453 | SASGSSAAS | 0.355 | . |
| SP96 | 454 | ASGSSAASS | 0.439 | . |
| SP96 | 457 | SSAASSSPS | 0.976 | *S* |
| SP96 | 458 | SAASSSPSS | 0.867 | *S* |
| SP96 | 459 | AASSSPSSS | 0.986 | *S* |
| SP96 | 461 | SSSPSSSAA | 0.966 | *S* |
| SP96 | 462 | SSPSSSAAS | 0.960 | *S* |
| SP96 | 463 | SPSSSAASS | 0.770 | *S* |
| SP96 | 466 | SSAASSSPS | 0.983 | *S* |
| SP96 | 467 | SAASSSPSS | 0.867 | *S* |
| SP96 | 468 | AASSSPSSS | 0.986 | *S* |
| SP96 | 470 | SSSPSSSAA | 0.966 | *S* |
| SP96 | 471 | SSPSSSAAS | 0.960 | *S* |
| SP96 | 472 | SPSSSAASS | 0.770 | *S* |
| SP96 | 475 | SSAASSSPS | 0.983 | *S* |
| SP96 | 476 | SAASSSPSS | 0.867 | *S* |
| SP96 | 477 | AASSSPSSS | 0.986 | *S* |
| SP96 | 479 | SSSPSSSAA | 0.966 | *S* |
| SP96 | 480 | SSPSSSAAS | 0.960 | *S* |
| SP96 | 481 | SPSSSAASS | 0.770 | *S* |
| SP96 | 484 | SSAASSSPS | 0.983 | *S* |

| | | | | |
|---|---|---|---|---|
| SP96 | 485 | SAASSSPSS | 0.867 | *S* |
| SP96 | 486 | AASSSPSSS | 0.986 | *S* |
| SP96 | 488 | SSSPSSSAS | 0.974 | *S* |
| SP96 | 489 | SSPSSSASS | 0.969 | *S* |
| SP96 | 490 | SPSSSASSS | 0.972 | *S* |
| SP96 | 492 | SSSASSSSS | 0.985 | *S* |
| SP96 | 493 | SSASSSSSP | 0.989 | *S* |
| SP96 | 494 | SASSSSSPS | 0.955 | *S* |
| SP96 | 495 | ASSSSSPSS | 0.838 | *S* |
| SP96 | 496 | SSSSSPSSS | 0.995 | *S* |
| SP96 | 498 | SSSPSSSAS | 0.985 | *S* |
| SP96 | 499 | SSPSSSASS | 0.969 | *S* |
| SP96 | 500 | SPSSSASSS | 0.972 | *S* |
| SP96 | 502 | SSSASSSSA | 0.982 | *S* |
| SP96 | 503 | SSASSSSAP | 0.987 | *S* |
| SP96 | 504 | SASSSSAPS | 0.629 | *S* |
| SP96 | 505 | ASSSSAPSS | 0.767 | *S* |
| SP96 | 508 | SSAPSSSAS | 0.987 | *S* |
| SP96 | 509 | SAPSSSASS | 0.849 | *S* |
| SP96 | 510 | APSSSASSS | 0.873 | *S* |
| SP96 | 512 | SSSASSSSA | 0.982 | *S* |
| SP96 | 513 | SSASSSSAP | 0.987 | *S* |
| SP96 | 514 | SASSSSAPS | 0.629 | *S* |
| SP96 | 515 | ASSSSAPSS | 0.767 | *S* |
| SP96 | 518 | SSAPSSSAS | 0.987 | *S* |
| SP96 | 519 | SAPSSSASS | 0.849 | *S* |
| SP96 | 520 | APSSSASSS | 0.873 | *S* |
| SP96 | 522 | SSSASSSSA | 0.972 | *S* |
| SP96 | 523 | SSASSSSAS | 0.989 | *S* |
| SP96 | 524 | SASSSSASS | 0.766 | *S* |
| SP96 | 525 | ASSSSASSS | 0.963 | *S* |
| SP96 | 527 | SSSASSSSA | 0.984 | *S* |
| SP96 | 528 | SSASSSSAS | 0.989 | *S* |
| SP96 | 529 | SASSSSASS | 0.727 | *S* |
| SP96 | 530 | ASSSSASSA | 0.939 | *S* |
| SP96 | 532 | SSSASSAAT | 0.941 | *S* |
| SP96 | 533 | SSASSAATT | 0.890 | *S* |
| SP96 | 600 | TTATS---- | 0.680 | *S* |

Threonine predictions

| Name | Pos | Context | Score | Pred |
|---|---|---|---|---|
| SP96 | 12 | VACLTYFSG | 0.084 | . |
| SP96 | 34 | NCPSTCFQG | 0.053 | . |
| SP96 | 52 | VKYCTEMKD | 0.133 | . |
| SP96 | 77 | LPVQTWSSC | 0.270 | . |
| SP96 | 106 | SNCPTNCEN | 0.085 | . |
| SP96 | 126 | CDAKTACCP | 0.072 | . |
| SP96 | 140 | KGGQTSGGS | 0.794 | *T* |
| SP96 | 145 | SGGSTTGSQ | 0.096 | . |
| SP96 | 146 | GGSTTGSQT | 0.446 | . |
| SP96 | 150 | TGSQTSGGS | 0.441 | . |
| SP96 | 155 | SGGSTSGGS | 0.450 | . |
| SP96 | 160 | SGGSTTGSQ | 0.096 | . |
| SP96 | 161 | GGSTTGSQT | 0.446 | . |

| | | | | |
|---|---|---|---|---|
| SP96 | 165 | TGSQTSGGS | 0.441 | . |
| SP96 | 170 | SGGSTTGSQ | 0.096 | . |
| SP96 | 171 | GGSTTGSQT | 0.446 | . |
| SP96 | 175 | TGSQTSGSQ | 0.152 | . |
| SP96 | 180 | SGSQTSAGS | 0.527 | *T* |
| SP96 | 188 | SCSNTQCPN | 0.315 | . |
| SP96 | 212 | QQSSTSGGH | 0.026 | . |
| SP96 | 223 | DPCDTVQCP | 0.266 | . |
| SP96 | 243 | EAKCTRDED | 0.735 | *T* |
| SP96 | 250 | EDEPTHRPT | 0.954 | *T* |
| SP96 | 254 | THRPTHRPK | 0.923 | *T* |
| SP96 | 305 | IQCPTGYRC | 0.033 | . |
| SP96 | 331 | DNCLTCNDV | 0.117 | . |
| SP96 | 347 | VCVMTRARC | 0.049 | . |
| SP96 | 364 | DVQPTCIKP | 0.065 | . |
| SP96 | 370 | IKPSTIAGS | 0.140 | . |
| SP96 | 375 | IAGSTIASI | 0.043 | . |
| SP96 | 382 | SIASTIAST | 0.270 | . |
| SP96 | 386 | TIASTGSTG | 0.015 | . |
| SP96 | 389 | STGSTGATS | 0.033 | . |
| SP96 | 392 | STGATSPCS | 0.087 | . |
| SP96 | 402 | AQCPTGYVC | 0.024 | . |
| SP96 | 420 | LPRPTTTTG | 0.138 | . |
| SP96 | 421 | PRPTTTTGS | 0.936 | *T* |
| SP96 | 422 | RPTTTTGST | 0.943 | *T* |
| SP96 | 423 | PTTTTGSTS | 0.290 | . |
| SP96 | 426 | TTGSTSDSS | 0.313 | . |
| SP96 | 435 | ALGSTSESS | 0.058 | . |
| SP96 | 536 | SSAATTAAT | 0.301 | . |
| SP96 | 537 | SAATTAATT | 0.366 | . |
| SP96 | 540 | TTAATTIAT | 0.149 | . |
| SP96 | 541 | TAATTIATT | 0.360 | . |
| SP96 | 544 | TTIATTAAT | 0.164 | . |
| SP96 | 545 | TIATTAATT | 0.063 | . |
| SP96 | 548 | TTAATTTAT | 0.224 | . |
| SP96 | 549 | TAATTTATT | 0.354 | . |
| SP96 | 550 | AATTTATTT | 0.199 | . |
| SP96 | 552 | TTTATTTAT | 0.201 | . |
| SP96 | 553 | TTATTTATT | 0.300 | . |
| SP96 | 554 | TATTTATTA | 0.106 | . |
| SP96 | 556 | TTTATTATT | 0.082 | . |
| SP96 | 557 | TTATTATTT | 0.178 | . |
| SP96 | 559 | ATTATTTAT | 0.224 | . |
| SP96 | 560 | TTATTTATT | 0.300 | . |
| SP96 | 561 | TATTTATTT | 0.180 | . |
| SP96 | 563 | TTTATTTAT | 0.201 | . |
| SP96 | 564 | TTATTTATT | 0.300 | . |
| SP96 | 565 | TATTTATTT | 0.180 | . |
| SP96 | 567 | TTTATTTAA | 0.122 | . |
| SP96 | 568 | TTATTTAAT | 0.506 | *T* |
| SP96 | 569 | TATTTAATI | 0.017 | . |
| SP96 | 572 | TTAATIATT | 0.101 | . |
| SP96 | 575 | ATIATTTAA | 0.128 | . |
| SP96 | 576 | TIATTTAAT | 0.242 | . |
| SP96 | 577 | IATTTAATT | 0.350 | . |
| SP96 | 580 | TTAATTTAT | 0.224 | . |
| SP96 | 581 | TAATTTATT | 0.354 | . |
| SP96 | 582 | AATTTATTT | 0.199 | . |
| SP96 | 584 | TTTATTTAT | 0.201 | . |
| SP96 | 585 | TTATTTATT | 0.300 | . |
| SP96 | 586 | TATTTATTA | 0.106 | . |

| | | | | |
|---|---|---|---|---|
| SP96 | 588 | TTTATTATT | 0.082 | . |
| SP96 | 589 | TTATTATTT | 0.178 | . |
| SP96 | 591 | ATTATTTAT | 0.224 | . |
| SP96 | 592 | TTATTTATT | 0.300 | . |
| SP96 | 593 | TATTTATTT | 0.180 | . |
| SP96 | 595 | TTTATTTAT | 0.201 | . |
| SP96 | 596 | TTATTTATS | 0.103 | . |
| SP96 | 597 | TATTTATS- | 0.085 | . |
| SP96 | 599 | TTTATS--- | 0.080 | . |

Tyrosine predictions

| Name | Pos | Context | Score | Pred |
|---|---|---|---|---|
| SP96 | 13 | ACLTYFSGG | 0.428 | . |
| SP96 | 26 | SCSSYSGDN | 0.594 | *Y* |
| SP96 | 40 | FQGSYDIPC | 0.680 | *Y* |
| SP96 | 50 | AQVKYCTEM | 0.075 | . |
| SP96 | 195 | PNGFYCQVQ | 0.621 | *Y* |
| SP96 | 228 | VQCPYGYSC | 0.036 | . |
| SP96 | 230 | CPYGYSCES | 0.086 | . |
| SP96 | 266 | DSDKYLCDN | 0.953 | *Y* |
| SP96 | 277 | CPRGYKCNA | 0.685 | *Y* |
| SP96 | 292 | CIAGYEIPR | 0.316 | . |
| SP96 | 307 | CPTGYRCED | 0.329 | . |
| SP96 | 404 | CPTGYVCVA | 0.977 | *Y* |

## NetPhos 2.0 Prediction Results

**Name: SP96-T Length: 12**

ATTTATTATTTA
.....

Ser: 0 Thr: 0 Tyr: 0

Threonine predictions

| Name | Pos | Context | Score | Pred |
|---|---|---|---|---|
| SP96-T | 2 | ---ATTTAT | 0.178 | . |
| SP96-T | 3 | --ATTTATT | 0.375 | . |
| SP96-T | 4 | -ATTTATTA | 0.259 | . |
| SP96-T | 6 | TTTATTATT | 0.082 | . |
| SP96-T | 7 | TTATTATTT | 0.178 | . |
| SP96-T | 9 | ATTATTTA- | 0.047 | . |
| SP96-T | 10 | TTATTTA-- | 0.111 | . |
| SP96-T | 11 | TATTTA--- | 0.093 | . |

## NetPhos 2.0 Prediction Results

**Name: SP96-S Length: 14**

ASSSSAPSSSASSS
.......SSS.SS.

Ser: 5 Thr: 0 Tyr: 0

Serine predictions

| Name | Pos | Context | Score | Pred |
|---|---|---|---|---|
| | | v | | |
| SP96-S | 2 | ---ASSSSA | 0.008 | . |
| SP96-S | 3 | --ASSSSAP | 0.024 | . |
| SP96-S | 4 | -ASSSSAPS | 0.008 | . |
| SP96-S | 5 | ASSSSAPSS | 0.214 | . |
| SP96-S | 8 | SSAPSSSAS | 0.987 | *S* |
| SP96-S | 9 | SAPSSSASS | 0.849 | *S* |
| SP96-S | 10 | APSSSASSS | 0.788 | *S* |
| SP96-S | 12 | SSSASSS-- | 0.783 | *S* |
| SP96-S | 13 | SSASSS--- | 0.864 | *S* |
| SP96-S | 14 | SASSS---- | 0.314 | . |

## NetPhos 2.0 Prediction Results

**Name: SP96-G Length: 16**

AGSQTSGGSTGGSTGS
.....S..S.......

Ser: 2 Thr: 0 Tyr: 0

Serine predictions

| Name | Pos | Context | Score | Pred |
|---|---|---|---|---|
| | | v | | |
| SP96-G | 3 | --AGSQTSG | 0.012 | . |
| SP96-G | 6 | GSQTSGGST | 0.914 | *S* |
| SP96-G | 9 | TSGGSTGGS | 0.649 | *S* |
| SP96-G | 13 | STGGSTGS- | 0.232 | . |
| SP96-G | 16 | GSTGS---- | 0.013 | . |
| | | ^ | | |

Threonine predictions

| Name | Pos | Context | Score | Pred |
|---|---|---|---|---|
| SP96-G | 5 | AGSQTSGGS | 0.478 | . |
| SP96-G | 10 | SGGSTGGST | 0.366 | . |
| SP96-G | 14 | TGGSTGS-- | 0.032 | . |

# DictyOGlyc 1.1 Prediction Results

**Name: SP96 Length: 600**

```
MRVLLVLVACLTYFSGGALAQSCSSYSGDNCPSTCFQGSYDIPCGAQVKYCTEMKDNCGEGGDVKCWKDGSNLPVQTWSS 80
CVPSELFGPNGKFKPSEIPNSSNCPTNCENGVEWVNLCGLSCDAKTACCPDVCQCKGGQTSGGSTTGSQTSGGSTSGGST160
TGSQTSGGSTTGSQTSGSQTSAGSCSNTQCPNGFYCQVQGNNAVCVPQQSSTSGGHQNDPCDTVQCPYGYSCESRDGFEA240
KCTRDEDEPTHRPTHRPKPPHDSDKYLCDNVHCPRGYKCNAKNGVAKCIAGYEIPRVCRNIQCPTGYRCEDHNRNPICVL320
EERENPDNCLTCNDVNCEASGLVCVMTRARCKVGAAKCCDVQPTCIKPSTIAGSTIASIASTIASTGSTGATSPCSVAQC400
PTGYVCVAQNNVAVSLPRPTTTTGSTSDSSALGSTSESSASGSSAVSSSASGSSAASSSPSSSAASSSPSSSAASSSPSS480
SAASSSPSSSASSSSSPSSSASSSSAPSSSASSSSAPSSSASSSSASSSSASSAATTAATTIATTAATTTATTTATTATT560
TATTTATTTAATIATTTAATTTATTTATTATTTATTTATS

.............................................................................. 80

..............G.............................................................160

.............................................................................240

.............................................................................320

.............................................................................400

.............................................................................480

.............................................................................560
.....................................
```

| Name | Residue | Number | Potential | Threshold | Assignment |
|---|---|---|---|---|---|
| SP96 | Thr | 0012 | 0.0116 | 0.8213 | . |
| SP96 | Ser | 0015 | 0.0097 | 0.7184 | . |
| SP96 | Ser | 0022 | 0.0132 | 0.7124 | . |
| SP96 | Ser | 0024 | 0.0141 | 0.6575 | . |
| SP96 | Ser | 0025 | 0.0209 | 0.6595 | . |
| SP96 | Ser | 0027 | 0.0140 | 0.5715 | . |
| SP96 | Ser | 0033 | 0.0283 | 0.6804 | . |
| SP96 | Thr | 0034 | 0.0130 | 0.6974 | . |
| SP96 | Ser | 0039 | 0.0160 | 0.6874 | . |
| SP96 | Thr | 0052 | 0.0184 | 0.6625 | . |
| SP96 | Ser | 0071 | 0.1903 | 0.5586 | . |
| SP96 | Thr | 0077 | 0.0160 | 0.6944 | . |
| SP96 | Ser | 0079 | 0.2230 | 0.7254 | . |
| SP96 | Ser | 0080 | 0.2159 | 0.7334 | . |
| SP96 | Ser | 0084 | 0.2026 | 0.7294 | . |
| SP96 | Ser | 0096 | 0.6995 | 0.6045 | G |
| SP96 | Ser | 0101 | 0.0187 | 0.6255 | . |
| SP96 | Ser | 0102 | 0.3943 | 0.5915 | . |
| SP96 | Thr | 0106 | 0.0177 | 0.6025 | . |
| SP96 | Ser | 0121 | 0.0402 | 0.7254 | . |
| SP96 | Thr | 0126 | 0.0109 | 0.6365 | . |
| SP96 | Thr | 0140 | 0.0146 | 0.5576 | . |
| SP96 | Ser | 0141 | 0.0093 | 0.5616 | . |
| SP96 | Ser | 0144 | 0.4789 | 0.5965 | . |
| SP96 | Thr | 0145 | 0.0118 | 0.5995 | . |
| SP96 | Thr | 0146 | 0.0135 | 0.5975 | . |
| SP96 | Ser | 0148 | 0.3566 | 0.5845 | . |
| SP96 | Thr | 0150 | 0.0106 | 0.5865 | . |
| SP96 | Ser | 0151 | 0.0099 | 0.5765 | . |
| SP96 | Ser | 0154 | 0.2008 | 0.5735 | . |

| | | | | | |
|---|---|---|---|---|---|
| SP96 | Thr | 0155 | 0.0148 | 0.5865 | . |
| SP96 | Ser | 0156 | 0.0103 | 0.5895 | . |
| SP96 | Ser | 0159 | 0.4549 | 0.5965 | . |
| SP96 | Thr | 0160 | 0.0118 | 0.6015 | . |
| SP96 | Thr | 0161 | 0.0137 | 0.5995 | . |
| SP96 | Ser | 0163 | 0.3636 | 0.5835 | . |
| SP96 | Thr | 0165 | 0.0106 | 0.5875 | . |
| SP96 | Ser | 0166 | 0.0097 | 0.5725 | . |
| SP96 | Ser | 0169 | 0.4455 | 0.5875 | . |
| SP96 | Thr | 0170 | 0.0114 | 0.5945 | . |
| SP96 | Thr | 0171 | 0.0135 | 0.5935 | . |
| SP96 | Ser | 0173 | 0.3501 | 0.5835 | . |
| SP96 | Thr | 0175 | 0.0102 | 0.5895 | . |
| SP96 | Ser | 0176 | 0.0157 | 0.5715 | . |
| SP96 | Ser | 0178 | 0.0722 | 0.5855 | . |
| SP96 | Thr | 0180 | 0.0116 | 0.6035 | . |
| SP96 | Ser | 0181 | 0.0099 | 0.6015 | . |
| SP96 | Ser | 0184 | 0.0372 | 0.6425 | . |
| SP96 | Ser | 0186 | 0.0296 | 0.6115 | . |
| SP96 | Thr | 0188 | 0.0271 | 0.6405 | . |
| SP96 | Ser | 0210 | 0.0318 | 0.5715 | . |
| SP96 | Ser | 0211 | 0.0224 | 0.5576 | . |
| SP96 | Thr | 0212 | 0.0152 | 0.5745 | . |
| SP96 | Ser | 0213 | 0.0164 | 0.5785 | . |
| SP96 | Thr | 0223 | 0.0191 | 0.6635 | . |
| SP96 | Ser | 0231 | 0.0400 | 0.6844 | . |
| SP96 | Ser | 0234 | 0.0137 | 0.5935 | . |
| SP96 | Thr | 0243 | 0.0150 | 0.5556 | . |
| SP96 | Thr | 0250 | 0.0527 | 0.5136 | . |
| SP96 | Thr | 0254 | 0.1092 | 0.5306 | . |
| SP96 | Ser | 0263 | 0.1295 | 0.5266 | . |
| SP96 | Thr | 0305 | 0.0230 | 0.6325 | . |
| SP96 | Thr | 0331 | 0.0116 | 0.6794 | . |
| SP96 | Ser | 0340 | 0.0123 | 0.7374 | . |
| SP96 | Thr | 0347 | 0.0145 | 0.7763 | . |
| SP96 | Thr | 0364 | 0.0184 | 0.6345 | . |
| SP96 | Ser | 0369 | 0.1434 | 0.6445 | . |
| SP96 | Thr | 0370 | 0.0166 | 0.6375 | . |
| SP96 | Ser | 0374 | 0.0271 | 0.7454 | . |
| SP96 | Thr | 0375 | 0.0240 | 0.7384 | . |
| SP96 | Ser | 0378 | 0.0228 | 0.7743 | . |
| SP96 | Ser | 0381 | 0.0767 | 0.7624 | . |
| SP96 | Thr | 0382 | 0.0138 | 0.7494 | . |
| SP96 | Ser | 0385 | 0.0149 | 0.6665 | . |
| SP96 | Thr | 0386 | 0.0120 | 0.6465 | . |
| SP96 | Ser | 0388 | 0.0588 | 0.6045 | . |
| SP96 | Thr | 0389 | 0.0180 | 0.5975 | . |
| SP96 | Thr | 0392 | 0.0253 | 0.6285 | . |
| SP96 | Ser | 0393 | 0.0441 | 0.6425 | . |
| SP96 | Ser | 0396 | 0.0658 | 0.6655 | . |
| SP96 | Thr | 0402 | 0.0283 | 0.6814 | . |
| SP96 | Ser | 0415 | 0.0203 | 0.7074 | . |
| SP96 | Thr | 0420 | 0.0948 | 0.5456 | . |
| SP96 | Thr | 0421 | 0.0194 | 0.5486 | . |
| SP96 | Thr | 0422 | 0.0288 | 0.5616 | . |
| SP96 | Thr | 0423 | 0.0159 | 0.5596 | . |
| SP96 | Ser | 0425 | 0.1348 | 0.5586 | . |
| SP96 | Thr | 0426 | 0.0218 | 0.5755 | . |
| SP96 | Ser | 0427 | 0.0137 | 0.5815 | . |
| SP96 | Ser | 0429 | 0.0173 | 0.6085 | . |
| SP96 | Ser | 0430 | 0.0145 | 0.6195 | . |

| | | | | | |
|---|---|---|---|---|---|
| SP96 | Ser | 0434 | 0.0704 | 0.6155 | . |
| SP96 | Thr | 0435 | 0.0188 | 0.6205 | . |
| SP96 | Ser | 0436 | 0.0121 | 0.6005 | . |
| SP96 | Ser | 0438 | 0.0127 | 0.5775 | . |
| SP96 | Ser | 0439 | 0.0142 | 0.5865 | . |
| SP96 | Ser | 0441 | 0.0133 | 0.5965 | . |
| SP96 | Ser | 0443 | 0.0922 | 0.6245 | . |
| SP96 | Ser | 0444 | 0.0337 | 0.6395 | . |
| SP96 | Ser | 0447 | 0.0143 | 0.6615 | . |
| SP96 | Ser | 0448 | 0.0203 | 0.6545 | . |
| SP96 | Ser | 0449 | 0.0133 | 0.6465 | . |
| SP96 | Ser | 0451 | 0.0149 | 0.6035 | . |
| SP96 | Ser | 0453 | 0.0196 | 0.6085 | . |
| SP96 | Ser | 0454 | 0.0216 | 0.6035 | . |
| SP96 | Ser | 0457 | 0.1355 | 0.5795 | . |
| SP96 | Ser | 0458 | 0.0180 | 0.5785 | . |
| SP96 | Ser | 0459 | 0.1487 | 0.5695 | . |
| SP96 | Ser | 0461 | 0.0569 | 0.5596 | . |
| SP96 | Ser | 0462 | 0.0173 | 0.5765 | . |
| SP96 | Ser | 0463 | 0.0146 | 0.5805 | . |
| SP96 | Ser | 0466 | 0.1630 | 0.5775 | . |
| SP96 | Ser | 0467 | 0.0179 | 0.5745 | . |
| SP96 | Ser | 0468 | 0.1484 | 0.5675 | . |
| SP96 | Ser | 0470 | 0.0555 | 0.5596 | . |
| SP96 | Ser | 0471 | 0.0173 | 0.5765 | . |
| SP96 | Ser | 0472 | 0.0146 | 0.5795 | . |
| SP96 | Ser | 0475 | 0.1630 | 0.5765 | . |
| SP96 | Ser | 0476 | 0.0179 | 0.5735 | . |
| SP96 | Ser | 0477 | 0.1484 | 0.5675 | . |
| SP96 | Ser | 0479 | 0.0555 | 0.5596 | . |
| SP96 | Ser | 0480 | 0.0173 | 0.5765 | . |
| SP96 | Ser | 0481 | 0.0146 | 0.5795 | . |
| SP96 | Ser | 0484 | 0.1622 | 0.5785 | . |
| SP96 | Ser | 0485 | 0.0167 | 0.5765 | . |
| SP96 | Ser | 0486 | 0.1518 | 0.5656 | . |
| SP96 | Ser | 0488 | 0.0586 | 0.5586 | . |
| SP96 | Ser | 0489 | 0.0124 | 0.5656 | . |
| SP96 | Ser | 0490 | 0.0178 | 0.5666 | . |
| SP96 | Ser | 0492 | 0.0134 | 0.5675 | . |
| SP96 | Ser | 0493 | 0.0161 | 0.5715 | . |
| SP96 | Ser | 0494 | 0.1817 | 0.5646 | . |
| SP96 | Ser | 0495 | 0.0139 | 0.5526 | . |
| SP96 | Ser | 0496 | 0.1358 | 0.5476 | . |
| SP96 | Ser | 0498 | 0.0576 | 0.5576 | . |
| SP96 | Ser | 0499 | 0.0129 | 0.5675 | . |
| SP96 | Ser | 0500 | 0.0175 | 0.5705 | . |
| SP96 | Ser | 0502 | 0.0148 | 0.5705 | . |
| SP96 | Ser | 0503 | 0.0174 | 0.5845 | . |
| SP96 | Ser | 0504 | 0.1900 | 0.5795 | . |
| SP96 | Ser | 0505 | 0.0114 | 0.5715 | . |
| SP96 | Ser | 0508 | 0.0463 | 0.5695 | . |
| SP96 | Ser | 0509 | 0.0142 | 0.5805 | . |
| SP96 | Ser | 0510 | 0.0206 | 0.5725 | . |
| SP96 | Ser | 0512 | 0.0152 | 0.5685 | . |
| SP96 | Ser | 0513 | 0.0165 | 0.5805 | . |
| SP96 | Ser | 0514 | 0.1771 | 0.5785 | . |
| SP96 | Ser | 0515 | 0.0112 | 0.5715 | . |
| SP96 | Ser | 0518 | 0.0435 | 0.5695 | . |
| SP96 | Ser | 0519 | 0.0145 | 0.5805 | . |
| SP96 | Ser | 0520 | 0.0153 | 0.5765 | . |
| SP96 | Ser | 0522 | 0.0138 | 0.5815 | . |

| | | | | | |
|---|---|---|---|---|---|
| SP96 | Ser | 0523 | 0.0161 | 0.5925 | . |
| SP96 | Ser | 0524 | 0.0157 | 0.5975 | . |
| SP96 | Ser | 0525 | 0.0108 | 0.5965 | . |
| SP96 | Ser | 0527 | 0.0132 | 0.5835 | . |
| SP96 | Ser | 0528 | 0.0155 | 0.5945 | . |
| SP96 | Ser | 0529 | 0.0143 | 0.6025 | . |
| SP96 | Ser | 0530 | 0.0119 | 0.5965 | . |
| SP96 | Ser | 0532 | 0.0137 | 0.6115 | . |
| SP96 | Ser | 0533 | 0.0124 | 0.6195 | . |
| SP96 | Thr | 0536 | 0.0121 | 0.6485 | . |
| SP96 | Thr | 0537 | 0.0117 | 0.6575 | . |
| SP96 | Thr | 0540 | 0.0118 | 0.6884 | . |
| SP96 | Thr | 0541 | 0.0138 | 0.6914 | . |
| SP96 | Thr | 0544 | 0.0178 | 0.6914 | . |
| SP96 | Thr | 0545 | 0.0113 | 0.6794 | . |
| SP96 | Thr | 0548 | 0.0123 | 0.6385 | . |
| SP96 | Thr | 0549 | 0.0142 | 0.6225 | . |
| SP96 | Thr | 0550 | 0.0185 | 0.6155 | . |
| SP96 | Thr | 0552 | 0.0179 | 0.6105 | . |
| SP96 | Thr | 0553 | 0.0125 | 0.6035 | . |
| SP96 | Thr | 0554 | 0.0181 | 0.6095 | . |
| SP96 | Thr | 0556 | 0.0136 | 0.6135 | . |
| SP96 | Thr | 0557 | 0.0110 | 0.6115 | . |
| SP96 | Thr | 0559 | 0.0185 | 0.6085 | . |
| SP96 | Thr | 0560 | 0.0124 | 0.6055 | . |
| SP96 | Thr | 0561 | 0.0189 | 0.6125 | . |
| SP96 | Thr | 0563 | 0.0183 | 0.6075 | . |
| SP96 | Thr | 0564 | 0.0125 | 0.6005 | . |
| SP96 | Thr | 0565 | 0.0187 | 0.6155 | . |
| SP96 | Thr | 0567 | 0.0213 | 0.6275 | . |
| SP96 | Thr | 0568 | 0.0125 | 0.6395 | . |
| SP96 | Thr | 0569 | 0.0158 | 0.6555 | . |
| SP96 | Thr | 0572 | 0.0110 | 0.6914 | . |
| SP96 | Thr | 0575 | 0.0236 | 0.6834 | . |
| SP96 | Thr | 0576 | 0.0146 | 0.6814 | . |
| SP96 | Thr | 0577 | 0.0140 | 0.6525 | . |
| SP96 | Thr | 0580 | 0.0120 | 0.6245 | . |
| SP96 | Thr | 0581 | 0.0142 | 0.6155 | . |
| SP96 | Thr | 0582 | 0.0183 | 0.6145 | . |
| SP96 | Thr | 0584 | 0.0178 | 0.6095 | . |
| SP96 | Thr | 0585 | 0.0125 | 0.6025 | . |
| SP96 | Thr | 0586 | 0.0181 | 0.6085 | . |
| SP96 | Thr | 0588 | 0.0136 | 0.6125 | . |
| SP96 | Thr | 0589 | 0.0110 | 0.6105 | . |
| SP96 | Thr | 0591 | 0.0188 | 0.6085 | . |
| SP96 | Thr | 0592 | 0.0124 | 0.6055 | . |
| SP96 | Thr | 0593 | 0.0187 | 0.6095 | . |
| SP96 | Thr | 0595 | 0.0195 | 0.5785 | . |
| SP96 | Thr | 0596 | 0.0129 | 0.5496 | . |
| SP96 | Thr | 0597 | 0.0262 | 0.5186 | . |
| SP96 | Thr | 0599 | 0.0414 | 0.5046 | . |
| SP96 | Ser | 0600 | 0.0398 | 0.4876 | . |

## DictyOGlyc 1.1 Prediction Results

**Name: SP96-T motif Length: 12**

```
ATTTATTATTTA
............
```

| Name | Residue | Number | Potential | Threshold | Assignment |
|---|---|---|---|---|---|
| SP96-T | Thr | 0002 | 0.0141 | 0.3658 | . |
| SP96-T | Thr | 0003 | 0.0147 | 0.3837 | . |
| SP96-T | Thr | 0004 | 0.0252 | 0.3857 | . |
| SP96-T | Thr | 0006 | 0.0166 | 0.4117 | . |
| SP96-T | Thr | 0007 | 0.0121 | 0.4067 | . |
| SP96-T | Thr | 0009 | 0.0233 | 0.3548 | . |
| SP96-T | Thr | 0010 | 0.0210 | 0.3358 | . |
| SP96-T | Thr | 0011 | 0.0404 | 0.3418 | . |

## DictyOGlyc 1.1 Prediction Results

**Name: SP96-S motif Length: 14**

```
ASSSSAPSSSASSS
..............
```

| Name | Residue | Number | Potential | Threshold | Assignment |
|---|---|---|---|---|---|
| SP96-S | Ser | 0002 | 0.0174 | 0.2908 | . |
| SP96-S | Ser | 0003 | 0.0199 | 0.3268 | . |
| SP96-S | Ser | 0004 | 0.2404 | 0.3298 | . |
| SP96-S | Ser | 0005 | 0.0128 | 0.3488 | . |
| SP96-S | Ser | 0008 | 0.0471 | 0.3807 | . |
| SP96-S | Ser | 0009 | 0.0161 | 0.3757 | . |
| SP96-S | Ser | 0010 | 0.0336 | 0.3448 | . |
| SP96-S | Ser | 0012 | 0.0471 | 0.3128 | . |
| SP96-S | Ser | 0013 | 0.0511 | 0.3028 | . |
| SP96-S | Ser | 0014 | 0.0488 | 0.3058 | . |

## DictyOGlyc 1.1 Prediction Results

**Name: SP96-G motif Length: 16**

AGSQTSGGSTGGSTGS
.................

| Name | Residue | Number | Potential | Threshold | Assignment |
|---|---|---|---|---|---|
| SP96-G | Ser | 0003 | 0.0433 | 0.3408 | . |
| SP96-G | Thr | 0005 | 0.0120 | 0.3797 | . |
| SP96-G | Ser | 0006 | 0.0105 | 0.3997 | . |
| SP96-G | Ser | 0009 | 0.0811 | 0.4237 | . |
| SP96-G | Thr | 0010 | 0.0130 | 0.4187 | . |
| SP96-G | Ser | 0013 | 0.0471 | 0.3468 | . |
| SP96-G | Thr | 0014 | 0.0485 | 0.3418 | . |
| SP96-G | Ser | 0016 | 0.3274 | 0.3488 | . |

# pMUW 2921 (SP96T-motif)

**Start of Ddp2 ori**

```
CGATAGGCAATTTATTTTTATATCTATCTAAAAAAAAACTAGGAAAAATGAATGTCATCAAATAGTATTTTAACATTTTTTTTTTTTTTT
        10         20         30         40         50         60         70         80         90

TTAAAAAAAAAGTGTCATGACAAAAAAAAAAAAAGTGTCATGACAAAAAAAAAAAAAAAAAAGAGGGGAAAGTAATTATAACTAGGTTAGT
       100        110        120        130        140        150        160        170        180

TTTTTATAATTTTTACATATTTGTTAATAACTTTTAATTTTGAATCATATATGATATTACATCGTCCCGTTGAAAAAAAAAAAAAAAATT
       190        200        210        220        230        240        250        260        270

TTTTTTTTCAAACATTTTCATTTTTTAAAAAATGATATAAAATTTTAAACTAAACTATTTTATTAAATACAAACATATAACTTTATCTTA
       280        290        300        310        320        330        340        350        360

ATCAATTTTTTTGGTTTATACATATTTATGTTCGTACTGAAGTATAGATCGATCTTATTACTAAAGTTTCAAAAGTTTTAAAAAAAATTA
       370        380        390        400        410        420        430        440        450

AAGGGGGTAAATATATAACTTTCTGTTTTTTTCAATTCTGTCATGACAAAAAAAAAAAAAGTGTCATGACAAAAAAAAAAAAAAAAAAAT
       460        470        480        490        500        510        520        530        540
```

**End of Ddp2 ori**

**<Oligo Ddp2ori.xba.r2**

```
TTATTTCTTCAATAGGTATTGAAATGACCTCCGTTTTTAATAAAAAGTATATATTTGTGCTCGTCTAGCGGATCTACAAATTAATTAATC
```

```
       550        560        570        580        590        600        610        620        630
```

**Actin 15 promotor**

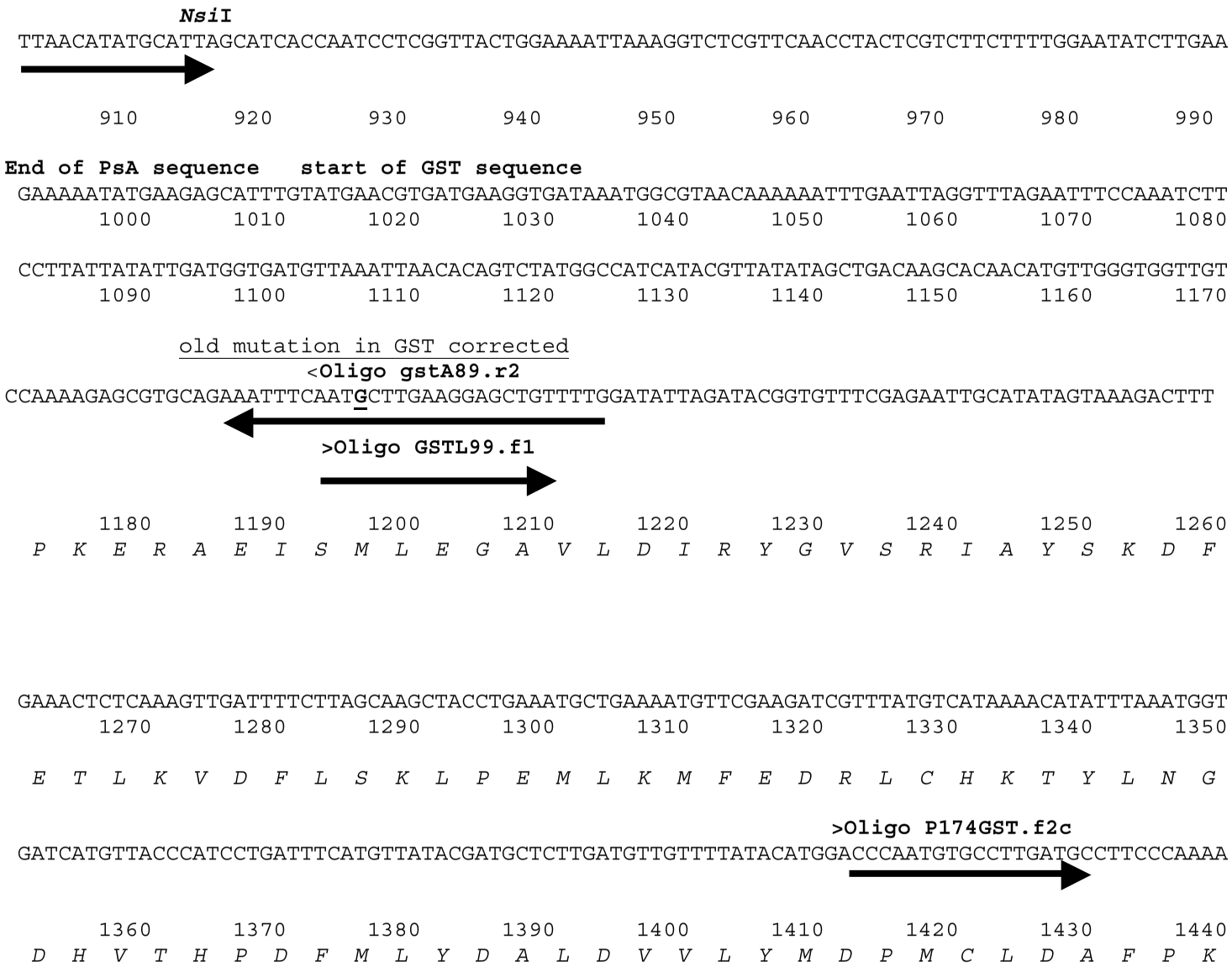
```
CCATCAAATCTTTAAAAAAAAAAATGGTTTAAAAAAACTTGGGTTGGTTAATTATTATTTGAAAATTTTAAAACCCAAATTAAAAAAAAA
       640        650        660        670        680        690        700        710        720

AAATGGGATTCAAAAATTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTCAGATTGCATAAAAAGATTTTTTTTTT
       730        740        750        760        770        780        790        800        810
```

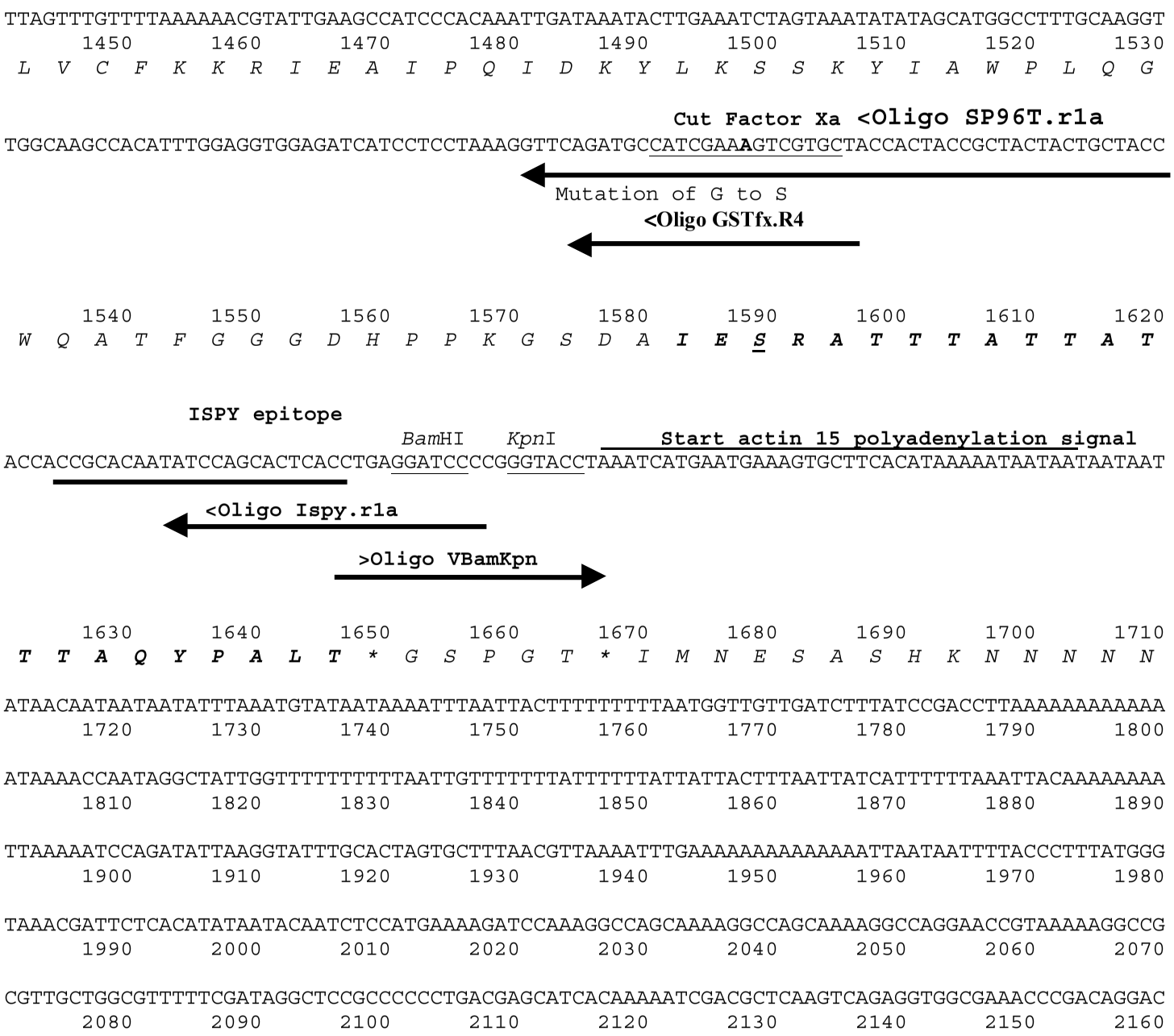
**Start PsA leader sequence >OligoPsAsig.f1**

```
TTTTTTTCTTATTTCTTAAAACAAATAAATTAAATTAAATAAAAAATAAAAATGAAATTCCAACATACATTTATTGCCTTATTATCACTC
```



```
       820        830        840        850        860        870        880        890        900
```


NsiI
TTAACATATGCATTAGCATCACCAATCCTCGGTTACTGGAAAATTAAAGGTCTCGTTCAACCTACTCGTCTTCTTTTGGAATATCTTGAA
910 920 930 940 950 960 970 980 990
End of PsA sequence start of GST sequence
GAAAAATATGAAGAGCATTTGTATGAACGTGATGAAGGTGATAAATGGCGTAACAAAAAATTTGAATTAGGTTTAGAATTTCCAAATCTT
1000 1010 1020 1030 1040 1050 1060 1070 1080
CCTTATTATATTGATGGTGATGTTAAATTAACACAGTCTATGGCCATCATACGTTATATAGCTGACAAGCACAACATGTTGGGTGGTTGT
1090 1100 1110 1120 1130 1140 1150 1160 1170
old mutation in GST corrected
<Oligo gstA89.r2
CCAAAAGAGCGTGCAGAAATTTCAATGCTTGAAGGAGCTGTTTTGGATATTAGATACGGTGTTTCGAGAATTGCATATAGTAAAGACTTT
>Oligo GSTL99.f1
1180 1190 1200 1210 1220 1230 1240 1250 1260
P K E R A E I S M L E G A V L D I R Y G V S R I A Y S K D F
GAAACTCTCAAAGTTGATTTTCTTAGCAAGCTACCTGAAATGCTGAAAATGTTCGAAGATCGTTTATGTCATAAAACATATTTAAATGGT
1270 1280 1290 1300 1310 1320 1330 1340 1350
E T L K V D F L S K L P E M L K M F E D R L C H K T Y L N G
>Oligo P174GST.f2c
GATCATGTTACCCATCCTGATTTCATGTTATACGATGCTCTTGATGTTGTTTTATACATGGACCCAATGTGCCTTGATGCCTTCCCAAAA
1360 1370 1380 1390 1400 1410 1420 1430 1440
D H V T H P D F M L Y D A L D V V L Y M D P M C L D A F P K


TTAGTTTGTTTTAAAAAACGTATTGAAGCCATCCCACAAATTGATAAATACTTGAAATCTAGTAAATATATAGCATGGCCTTTGCAAGGT
1450 1460 1470 1480 1490 1500 1510 1520 1530
L V C F K K R I E A I P Q I D K Y L K S S K Y I A W P L Q G
Cut Factor Xa
<Oligo SP96T.r1a
TGGCAAGCCACATTTGGAGGTGGAGATCATCCTCCTAAAGGTTCAGATGCCATCGAAAGTCGTGCTACCACTACCGCTACTACTGCTACC
Mutation of G to S
<Oligo GSTfx.R4
1540 1550 1560 1570 1580 1590 1600 1610 1620
W Q A T F G G G D H P P K G S D A I E S R A T T T A T T A T
ISPY epitope
BamHI KpnI
Start actin 15 polyadenylation signal
ACCACCGCACAATATCCAGCACTCACCTGAGGATCCCCGGGTACCTAAATCATGAATGAAAGTGCTTCACATAAAAATAATAATAATAAT
<Oligo Ispy.r1a
>Oligo VBamKpn
1630 1640 1650 1660 1670 1680 1690 1700 1710
T T A Q Y P A L T * G S P G T * I M N E S A S H K N N N N N
ATAACAATAATAATATTTAAATGTATAATAAAATTTAATTACTTTTTTTTTAATGGTTGTTGATCTTTATCCGACCTTAAAAAAAAAAAA
1720 1730 1740 1750 1760 1770 1780 1790 1800
ATAAAACCAATAGGCTATTGGTTTTTTTTTTAATTGTTTTTTTATTTTTTATTATTACTTTAATTATCATTTTTTAAATTACAAAAAAAA
1810 1820 1830 1840 1850 1860 1870 1880 1890
TTAAAAATCCAGATATTAAGGTATTTGCACTAGTGCTTTAACGTTAAAATTTGAAAAAAAAAAAAAAATTAATAATTTTACCCTTTATGGG
1900 1910 1920 1930 1940 1950 1960 1970 1980
TAAACGATTCTCACATATAATACAATCTCCATGAAAAGATCCAAAGGCCAGCAAAAGGCCAGCAAAAGGCCAGGAACCGTAAAAAGGCCG
1990 2000 2010 2020 2030 2040 2050 2060 2070
CGTTGCTGGCGTTTTTCGATAGGCTCCGCCCCCCTGACGAGCATCACAAAAATCGACGCTCAAGTCAGAGGTGGCGAAACCCGACAGGAC
2080 2090 2100 2110 2120 2130 2140 2150 2160

```
TATAAAGATACCAGGCGTTTCCCCCTGGAAGCTCCCTCGTGCGCTCTCCTGTTCCGACCCTGCCGCTTACCGGATACCTGTCCGCCTTTC
      2170      2180      2190      2200      2210      2220      2230      2240      2250

TCCCTTCGGGAAGCGTGGCGCTTTCTCATAGCTCACGCTGTAGGTATCTCAGTTCGGTGTAGGTCGTTCGCTCCAAGCTGGGCTGTGTGC
      2260      2270      2280      2290      2300      2310      2320      2330      2340

ACGAACCCCCCGTTCAGCCCGACCGCTGCGCCTTATCCGGTAACTATCGTCTTGAGTCCAACCCGGTAAGACACGACTTATCGCCACTGG
      2350      2360      2370      2380      2390      2400      2410      2420      2430


CAGCAGCCACTGGTAACAGGATTAGCAGAGCGAGGTATGTAGGCGGTGCTACAGAGTTCTTGAAGTGGTGGCCTAACTACGGCTACACTA
      2440      2450      2460      2470      2480      2490      2500      2510      2520

GAAGGACAGTATTTGGTATCTGCGCTCTGCTGAAGCCAGTTACCTTCGGAAAAAGAGTTGGTAGCTCTTGATCCGGCAAACAAACCACCG
      2530      2540      2550      2560      2570      2580      2590      2600      2610

CTGGTAGCGGTGGTTTTTTTGTTTGCAAGCAGCAGATTACGCGCAGAAAAAAAGGATCTCAAGAAGATCCTTTGATCTTTTCTACGGGGT
      2620      2630      2640      2650      2660      2670      2680      2690      2700


CTGACGCTCAGTGGAACGAAAACTCACGTTAAGGGATTTTGGTCATGAGATTATCAAAAAGGATCTTCACCTAGATCCTTTTAAATTAAA
      2710      2720      2730      2740      2750      2760      2770      2780      2790

AATGAAGTTTTAAATCAATCTAAAGTATATATGAGTAAACTTGGTCTGACAGTTACCAATGCTTAATCAGTGAGGCACCTATCTCAGCGA
      2800      2810      2820      2830      2840      2850      2860      2870      2880

TCTGTCTATTTCGTTCATCCATAGTTGCCTGACTCCCCGTCGTGTAGATAACTACGATACGGGAGGGCTTACCATCTGGCCCCAGTGCTG
      2890      2900      2910      2920      2930      2940      2950      2960      2970

CAATGATACCGCGAGACCCACGCTCACCGGCTCCAGATTTATCAGCAATAAACCAGCCAGCCGGAAGGGCCGAGCGCAGAAGTGGTCCTG
      2980      2990      3000      3010      3020      3030      3040      3050      3060

CAACTTTATCCGCCTCCATCCAGTCTATTAATTGTTGCCGGGAAGCTAGAGTAAGTAGTTCGCCAGTTAATAGTTTGCGCAACGTTGTTG
      3070      3080      3090      3100      3110      3120      3130      3140      3150

GCATTGCTACAGGCATCGTGGTGTCACGCTCGTCGTTTGGTATGGCTTCATTCAGCTCCGGTTCCCAACGATCAAGGCGAGTTACATGAT
      3160      3170      3180      3190      3200      3210      3220      3230      3240

CCCCCATGTTGTGCAAAAAAGCGGTTAGCTCCTTCGGTCCTCCGATCGTTGTCAGAAGTAAGTTGGCCGCAGTGTTATCACTCATGGTTA
      3250      3260      3270      3280      3290      3300      3310      3320      3330
```

```
TGGCAGCACTGCATAATTCTCTTACTGTCATGCCATCCGTAAGATGCTTTTCTGTGACTGGTGAGTACTCAACCAAGTCATTCTGAGAAT
      3340       3350       3360       3370       3380       3390       3400       3410       3420

AGTGTATGCGGCGACCGAGTTGCTCTTGCCCGGCGTCAATACGGGATAATACCGCGCCACATAGCAGAACTTTAAAAGTGCTCATCATTG
      3430       3440       3450       3460       3470       3480       3490       3500       3510

GAAAACGTTCTTCGGGGCGAAAACTCTCAAGGATCTTACCGCTGTTGAGATCCAGTTCGATGTAACCCACTCGTGCACCCAACTGATCTT
      3520       3530       3540       3550       3560       3570       3580       3590       3600

CAGCATCTTTTACTTTCACCAGCGTTTCTGGGTGAGCAAAAACAGGAAGGCAAAATGCCGCAAAAAAGGGAATAAGGGCGACACGGAAAT
      3610       3620       3630       3640       3650       3660       3670       3680       3690

GTTGAATACTCATACTCTTCCTTTTTCAATATTATTGAAGCATTTATCAGGGTTATTGTCTCATGAGCGGATACATATTTGAATGTATTT
      3700       3710       3720       3730       3740       3750       3760       3770       3780

AGAAAAATAAACAAATAGGGGTTCCGCGCACATTTCCCCGAAAAGTGCCACCTGACGTCTAAGAAACCATTATTATCAT
  3790       3800       3810       3820       3830       3840       3850
```

# pMUW 2931 (SP96S-motif)

**Start of Ddp2 ori**

```
CGATAGGCAATTTATTTTTATATCTATCTAAAAAAAAACTAGGAAAAATGAATGTCATCAAATAGTATTTTAACATTTTTTTTTTTTTTT
         10         20         30         40         50         60         70         80         90

TTAAAAAAAAAGTGTCATGACAAAAAAAAAAAAAGTGTCATGACAAAAAAAAAAAAAAAAAAGAGGGGAAAGTAATTATAACTAGGTTAGT
       100        110        120        130        140        150        160        170        180

TTTTTATAATTTTTACATATTTGTTAATAACTTTTAATTTTGAATCATATATGATATTACATCGTCCCGTTGAAAAAAAAAAAAAAAATT
       190        200        210        220        230        240        250        260        270

TTTTTTTTCAAACATTTTCATTTTTTAAAAAATGATATAAAATTTTAAACTAAACTATTTTATTAAATACAAACATATAACTTTATCTTA
       280        290        300        310        320        330        340        350        360

ATCAATTTTTTTGGTTTATACATATTTATGTTCGTACTGAAGTATAGATCGATCTTATTACTAAAGTTTCAAAAGTTTTAAAAAAAATTA
       370        380        390        400        410        420        430        440        450

AAGGGGGTAAATATATAACTTTCTGTTTTTTTCAATTCTGTCATGACAAAAAAAAAAAAAGTGTCATGACAAAAAAAAAAAAAAAAAAAT
       460        470        480        490        500        510        520        530        540
```

**End of Ddp2 ori**

**<Oligo Ddp2ori.xba.r2**

```
TTATTTCTTCAATAGGTATTGAAATGACCTCCGTTTTTAATAAAAAGTATATATTTGTGCTCGTCTAGCGGATCTACAAATTAATTAATC
```

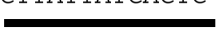
```
       550        560        570        580        590        600        610        620        630
```

**Actin 15 promotor**

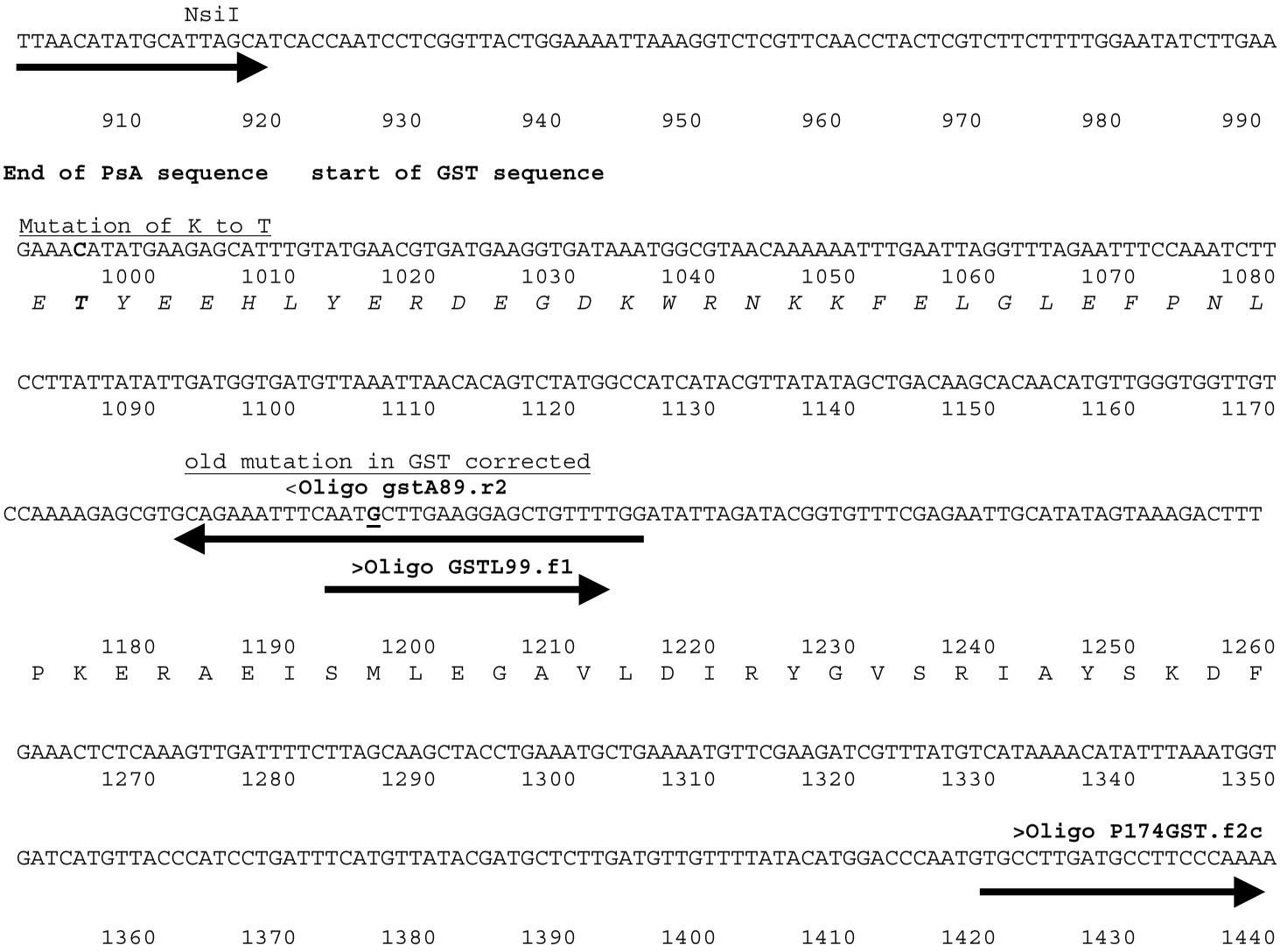
```
CCATCAAATCTTTAAAAAAAAAAATGGTTTAAAAAAACTTGGGTTGGTTAATTATTATTTGAAAATTTTAAAACCCAAATTAAAAAAAAA
       640        650        660        670        680        690        700        710        720

AAATGGGATTCAAAAATTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTCAGATTGCATAAAAAGATTTTTTTTTT
       730        740        750        760        770        780        790        800        810
```

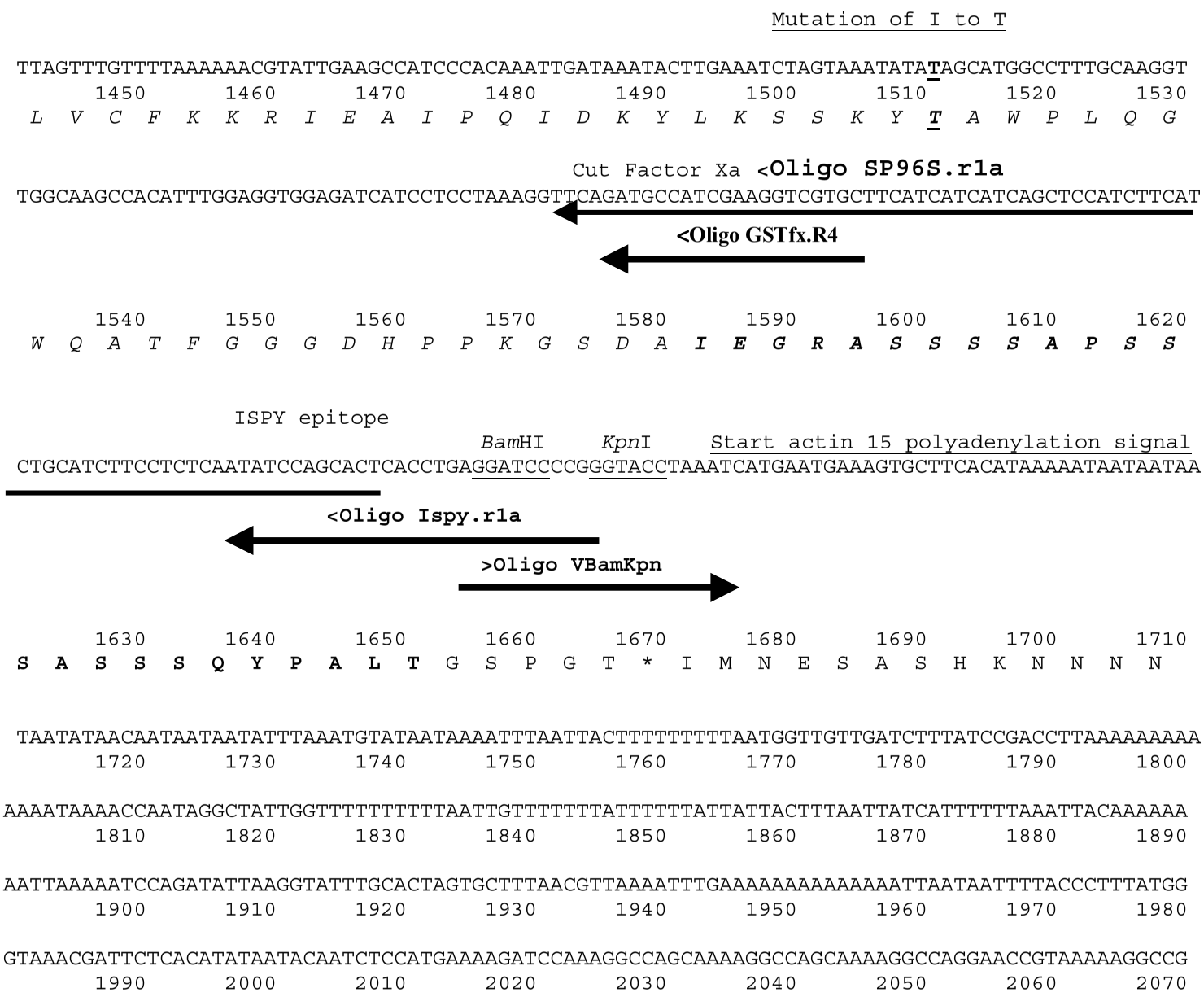
**Start PsA leadersequence  >OligoPsAsig.f1**

```
TTTTTTTCTTATTTCTTAAAACAAATAAATTAAATTAAATAAAAAATAAAAATGAAATTCCAACATACATTTATTGCCTTATTATCACTC
```



```
       820        830        840        850        860        870        880        890        900
```


NsiI
TTAACATATGCATTAGCATCACCAATCCTCGGTTACTGGAAAATTAAAGGTCTCGTTCAACCTACTCGTCTTCTTTTGGAATATCTTGAA
910 920 930 940 950 960 970 980 990
End of PsA sequence start of GST sequence
Mutation of K to T
GAAACATATGAAGAGCATTTGTATGAACGTGATGAAGGTGATAAATGGCGTAACAAAAAATTTGAATTAGGTTTAGAATTTCCAAATCTT
1000 1010 1020 1030 1040 1050 1060 1070 1080
E T Y E E H L Y E R D E G D K W R N K K F E L G L E F P N L
CCTTATTATATTGATGGTGATGTTAAATTAACACAGTCTATGGCCATCATACGTTATATAGCTGACAAGCACAACATGTTGGGTGGTTGT
1090 1100 1110 1120 1130 1140 1150 1160 1170
old mutation in GST corrected
<Oligo gstA89.r2
CCAAAAGAGCGTGCAGAAATTTCAATGCTTGAAGGAGCTGTTTTGGATATTAGATACGGTGTTTCGAGAATTGCATATAGTAAAGACTTT
>Oligo GSTL99.f1
1180 1190 1200 1210 1220 1230 1240 1250 1260
P K E R A E I S M L E G A V L D I R Y G V S R I A Y S K D F
GAAACTCTCAAAGTTGATTTTCTTAGCAAGCTACCTGAAATGCTGAAAATGTTCGAAGATCGTTTATGTCATAAAACATATTTAAATGGT
1270 1280 1290 1300 1310 1320 1330 1340 1350
>Oligo P174GST.f2c
GATCATGTTACCCATCCTGATTTCATGTTATACGATGCTCTTGATGTTGTTTTATACATGGACCCAATGTGCCTTGATGCCTTCCCAAAA
1360 1370 1380 1390 1400 1410 1420 1430 1440


Mutation of I to T
TTAGTTTGTTTTAAAAAACGTATTGAAGCCATCCCACAAATTGATAAATACTTGAAATCTAGTAAATATATAGCATGGCCTTTGCAAGGT
1450 1460 1470 1480 1490 1500 1510 1520 1530
L V C F K K R I E A I P Q I D K Y L K S S K Y T A W P L Q G
Cut Factor Xa <Oligo SP96S.r1a
TGGCAAGCCACATTTGGAGGTGGAGATCATCCTCCTAAAGGTTCAGATGCCATCGAAGGTCGTGCTTCATCATCATCAGCTCCATCTTCAT
<Oligo GSTfx.R4
1540 1550 1560 1570 1580 1590 1600 1610 1620
W Q A T F G G G D H P P K G S D A I E G R A S S S S A P S S
ISPY epitope
BamHI KpnI Start actin 15 polyadenylation signal
CTGCATCTTCCTCTCAATATCCAGCACTCACCTGAGGATCCCCGGGTACCTAAATCATGAATGAAAGTGCTTCACATAAAAATAATAATAA
<Oligo Ispy.r1a
>Oligo VBamKpn
1630 1640 1650 1660 1670 1680 1690 1700 1710
S A S S S Q Y P A L T G S P G T * I M N E S A S H K N N N N
TAATATAACAATAATAATATTTAAATGTATAATAAAATTTAATTACTTTTTTTTTAATGGTTGTTGATCTTTATCCGACCTTAAAAAAAAA
1720 1730 1740 1750 1760 1770 1780 1790 1800
AAAATAAAACCAATAGGCTATTGGTTTTTTTTTTAATTGTTTTTTTATTTTTTATTATTACTTTAATTATCATTTTTTAAATTACAAAAAA
1810 1820 1830 1840 1850 1860 1870 1880 1890
AATTAAAAATCCAGATATTAAGGTATTTGCACTAGTGCTTTAACGTTAAAATTTGAAAAAAAAAAAAAATTAATAATTTTACCCTTTATGG
1900 1910 1920 1930 1940 1950 1960 1970 1980
GTAAACGATTCTCACATATAATACAATCTCCATGAAAAGATCCAAAGGCCAGCAAAAGGCCAGCAAAAGGCCAGGAACCGTAAAAAGGCCG
1990 2000 2010 2020 2030 2040 2050 2060 2070

```
CGTTGCTGGCGTTTTTCGATAGGCTCCGCCCCCCTGACGAGCATCACAAAAATCGACGCTCAAGTCAGAGGTGGCGAAACCCGACAGGAC
      2080      2090      2100      2110      2120      2130      2140      2150      2160

TATAAAGATACCAGGCGTTTCCCCCTGGAAGCTCCCTCGTGCGCTCTCCTGTTCCGACCCTGCCGCTTACCGGATACCTGTCCGCCTTTC
      2170      2180      2190      2200      2210      2220      2230      2240      2250

TCCCTTCGGGAAGCGTGGCGCTTTCTCATAGCTCACGCTGTAGGTATCTCAGTTCGGTGTAGGTCGTTCGCTCCAAGCTGGGCTGTGTGC
      2260      2270      2280      2290      2300      2310      2320      2330      2340

ACGAACCCCCCGTTCAGCCCGACCGCTGCGCCTTATCCGGTAACTATCGTCTTGAGTCCAACCCGGTAAGACACGACTTATCGCCACTGG
      2350      2360      2370      2380      2390      2400      2410      2420      2430

CAGCAGCCACTGGTAACAGGATTAGCAGAGCGAGGTATGTAGGCGGTGCTACAGAGTTCTTGAAGTGGTGGCCTAACTACGGCTACACTA
      2440      2450      2460      2470      2480      2490      2500      2510      2520

GAAGGACAGTATTTGGTATCTGCGCTCTGCTGAAGCCAGTTACCTTCGGAAAAAGAGTTGGTAGCTCTTGATCCGGCAAACAAACCACCG
      2530      2540      2550      2560      2570      2580      2590      2600      2610

CTGGTAGCGGTGGTTTTTTTGTTTGCAAGCAGCAGATTACGCGCAGAAAAAAAGGATCTCAAGAAGATCCTTTGATCTTTTCTACGGGGT
      2620      2630      2640      2650      2660      2670      2680      2690      2700

CTGACGCTCAGTGGAACGAAAACTCACGTTAAGGGATTTTGGTCATGAGATTATCAAAAAGGATCTTCACCTAGATCCTTTTAAATTAAA
      2710      2720      2730      2740      2750      2760      2770      2780      2790

AATGAAGTTTTAAATCAATCTAAAGTATATATGAGTAAACTTGGTCTGACAGTTACCAATGCTTAATCAGTGAGGCACCTATCTCAGCGA
      2800      2810      2820      2830      2840      2850      2860      2870      2880

TCTGTCTATTTCGTTCATCCATAGTTGCCTGACTCCCCGTCGTGTAGATAACTACGATACGGGAGGGCTTACCATCTGGCCCCAGTGCTG
      2890      2900      2910      2920      2930      2940      2950      2960      2970

CAATGATACCGCGAGACCCACGCTCACCGGCTCCAGATTTATCAGCAATAAACCAGCCAGCCGGAAGGGCCGAGCGCAGAAGTGGTCCTG
      2980      2990      3000      3010      3020      3030      3040      3050      3060

CAACTTTATCCGCCTCCATCCAGTCTATTAATTGTTGCCGGGAAGCTAGAGTAAGTAGTTCGCCAGTTAATAGTTTGCGCAACGTTGTTG
      3070      3080      3090      3100      3110      3120      3130      3140      3150

GCATTGCTACAGGCATCGTGGTGTCACGCTCGTCGTTTGGTATGGCTTCATTCAGCTCCGGTTCCCAACGATCAAGGCGAGTTACATGAT
      3160      3170      3180      3190      3200      3210      3220      3230      3240

CCCCCATGTTGTGCAAAAAAGCGGTTAGCTCCTTCGGTCCTCCGATCGTTGTCAGAAGTAAGTTGGCCGCAGTGTTATCACTCATGGTTA
      3250      3260      3270      3280      3290      3300      3310      3320      3330
```

```
TGGCAGCACTGCATAATTCTCTTACTGTCATGCCATCCGTAAGATGCTTTTCTGTGACTGGTGAGTACTCAACCAAGTCATTCTGAGAAT
      3340       3350       3360       3370       3380       3390       3400       3410       3420

AGTGTATGCGGCGACCGAGTTGCTCTTGCCCGGCGTCAATACGGGATAATACCGCGCCACATAGCAGAACTTTAAAAGTGCTCATCATTG
      3430       3440       3450       3460       3470       3480       3490       3500       3510

GAAAACGTTCTTCGGGGCGAAAACTCTCAAGGATCTTACCGCTGTTGAGATCCAGTTCGATGTAACCCACTCGTGCACCCAACTGATCTT
      3520       3530       3540       3550       3560       3570       3580       3590       3600

CAGCATCTTTTACTTTCACCAGCGTTTCTGGGTGAGCAAAAACAGGAAGGCAAAATGCCGCAAAAAAGGGAATAAGGGCGACACGGAAAT
      3610       3620       3630       3640       3650       3660       3670       3680       3690

GTTGAATACTCATACTCTTCCTTTTTCAATATTATTGAAGCATTTATCAGGGTTATTGTCTCATGAGCGGATACATATTTGAATGTATTT
      3700       3710       3720       3730       3740       3750       3760       3770       3780

AGAAAAATAAACAAATAGGGGTTCCGCGCACATTTCCCCGAAAAGTGCCACCTGACGTCTAAGAAACCATTATTATCAT
  3790       3800       3810       3820       3830       3840       3850
```

# pMUW 2911 (SP96G-motif)

**Start of Ddp2 ori**

```
CGATAGGCAATTTATTTTTATATCTATCTAAAAAAAAACTAGGAAAAATGAATGTCATCAAATAGTATTTTAACATTTTTTTTTTTTTTT
         10         20         30         40         50         60         70         80         90

TTAAAAAAAAAGTGTCATGACAAAAAAAAAAAAAGTGTCATGACAAAAAAAAAAAAAAAAAAGAGGGGAAAGTAATTATAACTAGGTTAGT
        100        110        120        130        140        150        160        170        180

TTTTTATAATTTTTACATATTTGTTAATAACTTTTAATTTTGAATCATATATGATATTACATCGTCCCGTTGAAAAAAAAAAAAAAAATT
        190        200        210        220        230        240        250        260        270

TTTTTTTTCAAACATTTTCATTTTTTAAAAAATGATATAAAATTTTAAACTAAACTATTTTATTAAATACAAACATATAACTTTATCTTA
        280        290        300        310        320        330        340        350        360

ATCAATTTTTTTGGTTTATACATATTTATGTTCGTACTGAAGTATAGATCGATCTTATTACTAAAGTTTCAAAAGTTTTAAAAAAAATTA
        370        380        390        400        410        420        430        440        450

AAGGGGGTAAATATATAACTTTCTGTTTTTTTCAATTCTGTCATGACAAAAAAAAAAAAAGTGTCATGACAAAAAAAAAAAAAAAAAAAT
        460        470        480        490        500        510        520        530        540
```

**End of Ddp2 ori**

**<Oligo Ddp2ori.xba.r2**

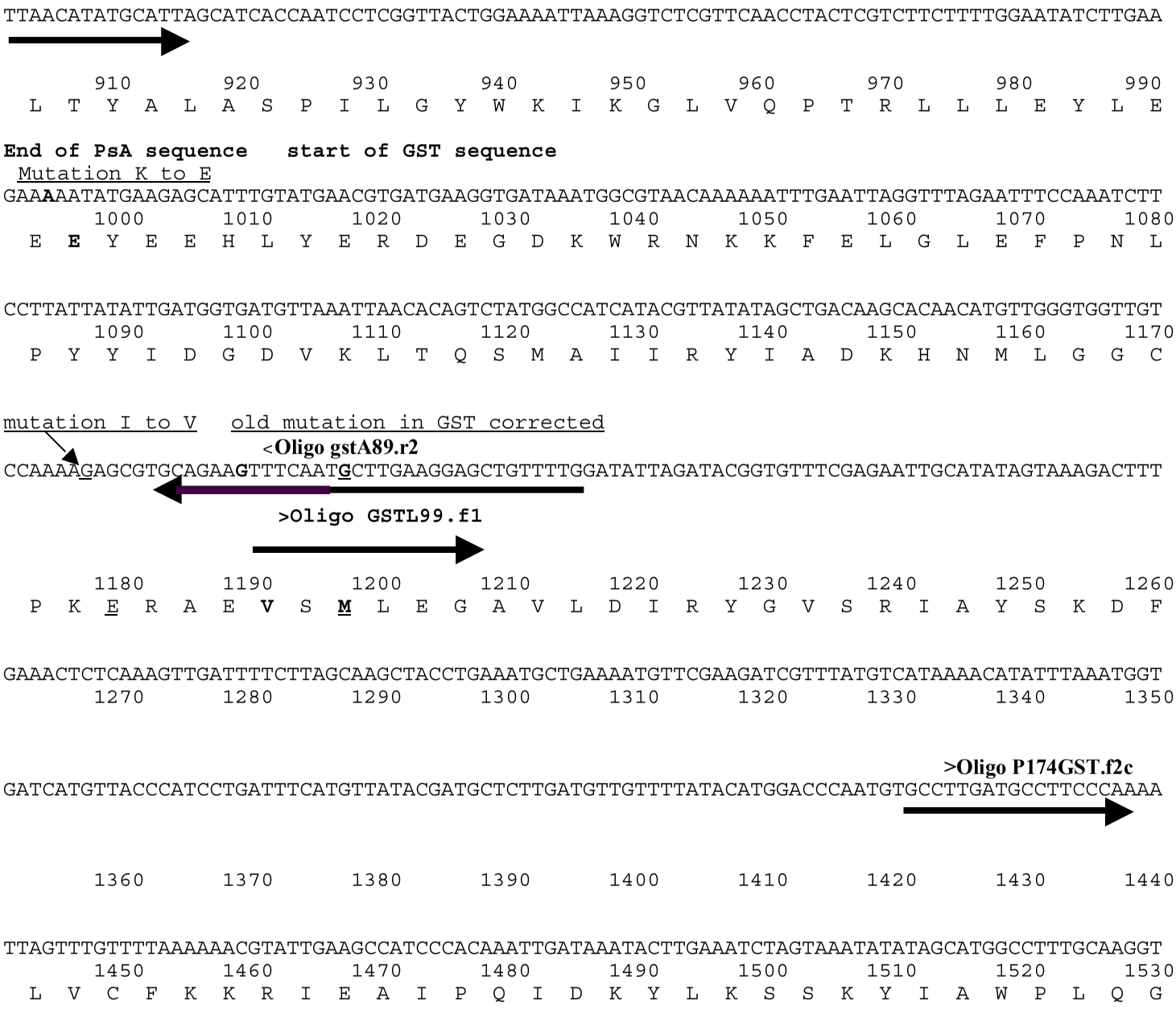
```
TTATTTCTTCAATAGGTATTGAAATGACCTCCGTTTTTAATAAAAAGTATATATTTGTGCTCGTCTAGCGGATCTACAAATTAATTAATC

        550        560        570        580        590        600        610        620        630
```

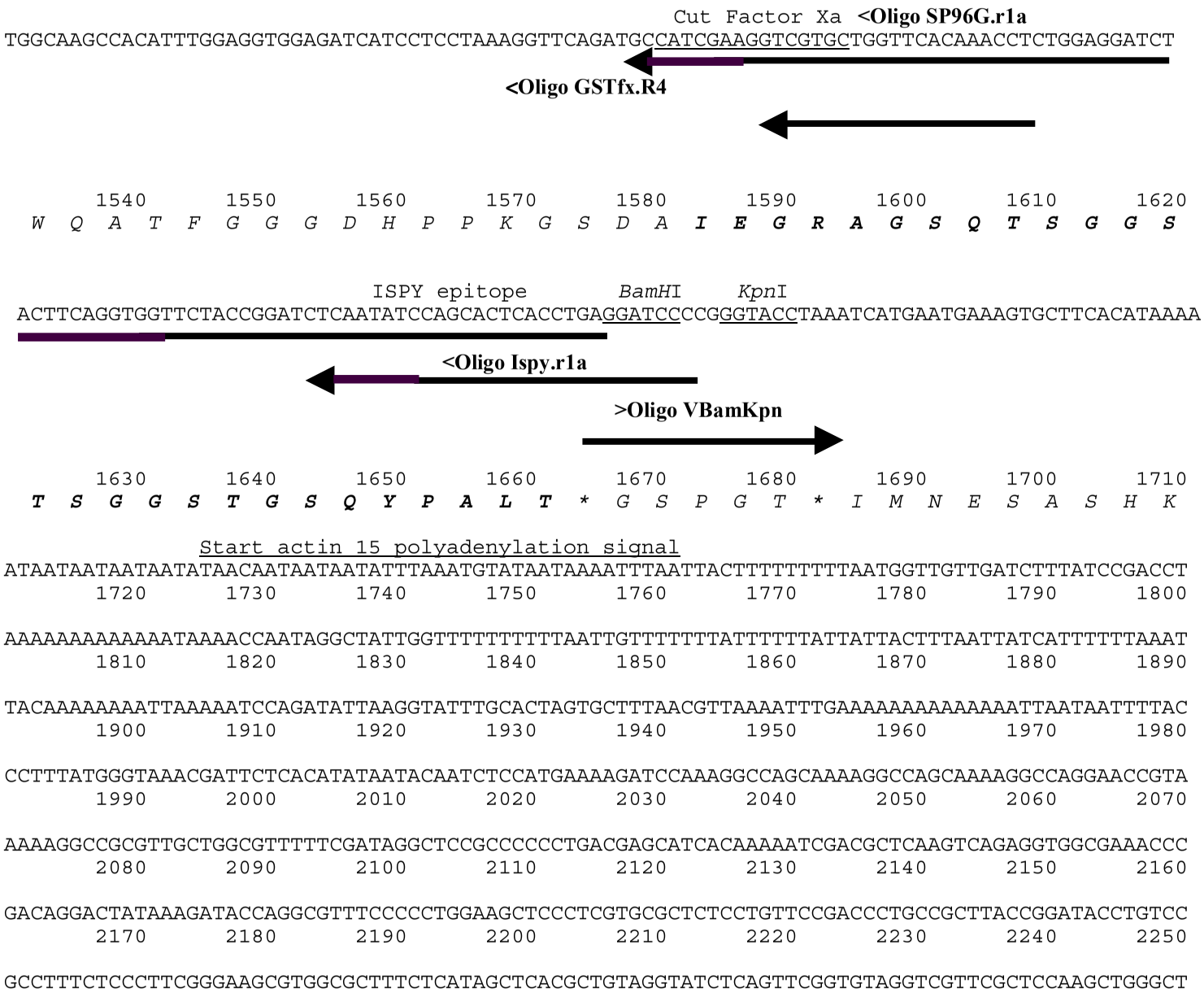
**Actin 15 promotor**

```
CCATCAAATCTTTAAAAAAAAAAAATGGTTTAAAAAAACTTGGGTTGGTTAATTATTATTTGAAAATTTTAAAACCCAAATTAAAAAAAAA
        640        650        660        670        680        690        700        710        720

AAATGGGATTCAAAAATTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTCAGATTGCATAAAAAGATTTTTTTTTTT
        730        740        750        760        770        780        790        800        810
```

**Start PsA leader sequence** **>OligoPsAsig.f1**

```
TTTTTTTCTTATTTCTTAAAACAAATAAATTAAATTAAATAAAAAATAAAAATGAAATTCCAACATACATTTATTGCCTTATTATCACTC

        820        830        840        850        860        870        880        890        900
```


TTAACATATGCATTAGCATCACCAATCCTCGGTTACTGGAAAATTAAAGGTCTCGTTCAACCTACTCGTCTTCTTTTGGAATATCTTGAA
910 920 930 940 950 960 970 980 990
L T Y A L A S P I L G Y W K I K G L V Q P T R L L L E Y L E
End of PsA sequence start of GST sequence
Mutation K to E
GAAAAATATGAAGAGCATTTGTATGAACGTGATGAAGGTGATAAATGGCGTAACAAAAAATTTGAATTAGGTTTAGAATTTCCAAATCTT
1000 1010 1020 1030 1040 1050 1060 1070 1080
E E Y E E H L Y E R D E G D K W R N K K F E L G L E F P N L
CCTTATTATATTGATGGTGATGTTAAATTAACACAGTCTATGGCCATCATACGTTATATAGCTGACAAGCACAACATGTTGGGTGGTTGT
1090 1100 1110 1120 1130 1140 1150 1160 1170
P Y Y I D G D V K L T Q S M A I I R Y I A D K H N M L G G C
mutation I to V
old mutation in GST corrected
<Oligo gstA89.r2
CCAAAAGAGCGTGCAGAAGTTTCAATGCTTGAAGGAGCTGTTTTGGATATTAGATACGGTGTTTCGAGAATTGCATATAGTAAAGACTTT
>Oligo GSTL99.f1
1180 1190 1200 1210 1220 1230 1240 1250 1260
P K E R A E V S M L E G A V L D I R Y G V S R I A Y S K D F
GAAACTCTCAAAGTTGATTTTCTTAGCAAGCTACCTGAAATGCTGAAAATGTTCGAAGATCGTTTATGTCATAAAACATATTTAAATGGT
1270 1280 1290 1300 1310 1320 1330 1340 1350
>Oligo P174GST.f2c
GATCATGTTACCCATCCTGATTTCATGTTATACGATGCTCTTGATGTTGTTTTATACATGGACCCAATGTGCCTTGATGCCTTCCCAAAA
1360 1370 1380 1390 1400 1410 1420 1430 1440
TTAGTTTGTTTTAAAAAACGTATTGAAGCCATCCCACAAATTGATAAATACTTGAAATCTAGTAAATATATAGCATGGCCTTTGCAAGGT
1450 1460 1470 1480 1490 1500 1510 1520 1530
L V C F K K R I E A I P Q I D K Y L K S S K Y I A W P L Q G


Cut Factor Xa <Oligo SP96G.r1a
TGGCAAGCCACATTTGGAGGTGGAGATCATCCTCCTAAAGGTTCAGATGCCATCGAAGGTCGTGCTGGTTCACAAACCTCTGGAGGATCT
<Oligo GSTfx.R4
1540 1550 1560 1570 1580 1590 1600 1610 1620
W Q A T F G G G D H P P K G S D A I E G R A G S Q T S G G S
ISPY epitope BamHI KpnI
ACTTCAGGTGGTTCTACCGGATCTCAATATCCAGCACTCACCTGAGGATCCCCGGGTACCTAAATCATGAATGAAAGTGCTTCACATAAAA
<Oligo Ispy.r1a
>Oligo VBamKpn
1630 1640 1650 1660 1670 1680 1690 1700 1710
T S G G S T G S Q Y P A L T * G S P G T * I M N E S A S H K
Start actin 15 polyadenylation signal
ATAATAATAATAATATAACAATAATAATATTTAAATGTATAATAAAATTTAATTACTTTTTTTTTAATGGTTGTTGATCTTTATCCGACCT
1720 1730 1740 1750 1760 1770 1780 1790 1800
AAAAAAAAAAAAATAAAACCAATAGGCTATTGGTTTTTTTTTTAATTGTTTTTTTATTTTTTATTATTACTTTAATTATCATTTTTTAAAT
1810 1820 1830 1840 1850 1860 1870 1880 1890
TACAAAAAAAATTAAAAATCCAGATATTAAGGTATTTGCACTAGTGCTTTAACGTTAAAATTTGAAAAAAAAAAAAAAATTAATAATTTTAC
1900 1910 1920 1930 1940 1950 1960 1970 1980
CCTTTATGGGTAAACGATTCTCACATATAATACAATCTCCATGAAAAGATCCAAAGGCCAGCAAAAGGCCAGCAAAAGGCCAGGAACCGTA
1990 2000 2010 2020 2030 2040 2050 2060 2070
AAAAGGCCGCGTTGCTGGCGTTTTTCGATAGGCTCCGCCCCCCTGACGAGCATCACAAAAATCGACGCTCAAGTCAGAGGTGGCGAAACCC
2080 2090 2100 2110 2120 2130 2140 2150 2160
GACAGGACTATAAAGATACCAGGCGTTTCCCCCTGGAAGCTCCCTCGTGCGCTCTCCTGTTCCGACCCTGCCGCTTACCGGATACCTGTCC
2170 2180 2190 2200 2210 2220 2230 2240 2250
GCCTTTCTCCCTTCGGGAAGCGTGGCGCTTTCTCATAGCTCACGCTGTAGGTATCTCAGTTCGGTGTAGGTCGTTCGCTCCAAGCTGGGCT

```
       2260       2270       2280       2290       2300       2310       2320       2330       2340

GTGTGCACGAACCCCCCGTTCAGCCCGACCGCTGCGCCTTATCCGGTAACTATCGTCTTGAGTCCAACCCGGTAAGACACGACTTATCGCC
       2350       2360       2370       2380       2390       2400       2410       2420       2430

ACTGGCAGCAGCCACTGGTAACAGGATTAGCAGAGCGAGGTATGTAGGCGGTGCTACAGAGTTCTTGAAGTGGTGGCCTAACTACGGCTAC
       2440       2450       2460       2470       2480       2490       2500       2510       2520

ACTAGAAGGACAGTATTTGGTATCTGCGCTCTGCTGAAGCCAGTTACCTTCGGAAAAAGAGTTGGTAGCTCTTGATCCGGCAAACAAACCA
       2530       2540       2550       2560       2570       2580       2590       2600       2610

CCGCTGGTAGCGGTGGTTTTTTTGTTTGCAAGCAGCAGATTACGCGCAGAAAAAAAGGATCTCAAGAAGATCCTTTGATCTTTTCTACGGG
       2620       2630       2640       2650       2660       2670       2680       2690       2700

GTCTGACGCTCAGTGGAACGAAAACTCACGTTAAGGGATTTTGGTCATGAGATTATCAAAAAGGATCTTCACCTAGATCCTTTTAAATTAA
       2710       2720       2730       2740       2750       2760       2770       2780       2790

AATGAAGTTTTAAATCAATCTAAAGTATATATGAGTAAACTTGGTCTGACAGTTACCAATGCTTAATCAGTGAGGCACCTATCTCAGCGA
       2800       2810       2820       2830       2840       2850       2860       2870       2880
ATCTGTCTATTTCGTTCATCCATAGTTGCCTGACTCCCCGTCGTGTAGATAACTACGATACGGGAGGGCTTACCATCTGGCCCCAGTGCTG
       2890       2900       2910       2920       2930       2940       2950       2960       2970

CAATGATACCGCGAGACCCACGCTCACCGGCTCCAGATTTATCAGCAATAAACCAGCCAGCCGGAAGGGCCGAGCGCAGAAGTGGTCCTG
       2980       2990       3000       3010       3020       3030       3040       3050       3060

CAACTTTATCCGCCTCCATCCAGTCTATTAATTGTTGCCGGGAAGCTAGAGTAAGTAGTTCGCCAGTTAATAGTTTGCGCAACGTTGTTG
       3070       3080       3090       3100       3110       3120       3130       3140       3150

GCATTGCTACAGGCATCGTGGTGTCACGCTCGTCGTTTGGTATGGCTTCATTCAGCTCCGGTTCCCAACGATCAAGGCGAGTTACATGAT
       3160       3170       3180       3190       3200       3210       3220       3230       3240

CCCCCATGTTGTGCAAAAAAGCGGTTAGCTCCTTCGGTCCTCCGATCGTTGTCAGAAGTAAGTTGGCCGCAGTGTTATCACTCATGGTTA
       3250       3260       3270       3280       3290       3300       3310       3320       3330

TGGCAGCACTGCATAATTCTCTTACTGTCATGCCATCCGTAAGATGCTTTTCTGTGACTGGTGAGTACTCAACCAAGTCATTCTGAGAAT
       3340       3350       3360       3370       3380       3390       3400       3410       3420

AGTGTATGCGGCGACCGAGTTGCTCTTGCCCGGCGTCAATACGGGATAATACCGCGCCACATAGCAGAACTTTAAAAGTGCTCATCATTG
       3430       3440       3450       3460       3470       3480       3490       3500       3510

GAAAACGTTCTTCGGGGCGAAAACTCTCAAGGATCTTACCGCTGTTGAGATCCAGTTCGATGTAACCCACTCGTGCACCCAACTGATCTT
       3520       3530       3540       3550       3560       3570       3580       3590       3600

CAGCATCTTTTACTTTCACCAGCGTTTCTGGGTGAGCAAAAACAGGAAGGCAAAATGCCGCAAAAAAGGGAATAAGGGCGACACGGAAAT
```

```
      3610       3620       3630       3640       3650       3660       3670       3680       3690
GTTGAATACTCATACTCTTCCTTTTTTCAATATTATTGAAGCATTTATCAGGGTTATTGTCTCATGAGCGGATACATATTTGAATGTATTT
      3700       3710       3720       3730       3740       3750       3760       3770       3780
AGAAAAATAAACAAATAGGGGTTCCGCGCACATTTCCCCGAAAAGTGCCACCTGACGTCTAAGAAACCATTATTATCAT
      3790       3800       3810       3820       3830       3840       3850
```

# pMUW 5021 (SP60 promoter SP96T-motif)

**Start of Ddp2 ori**

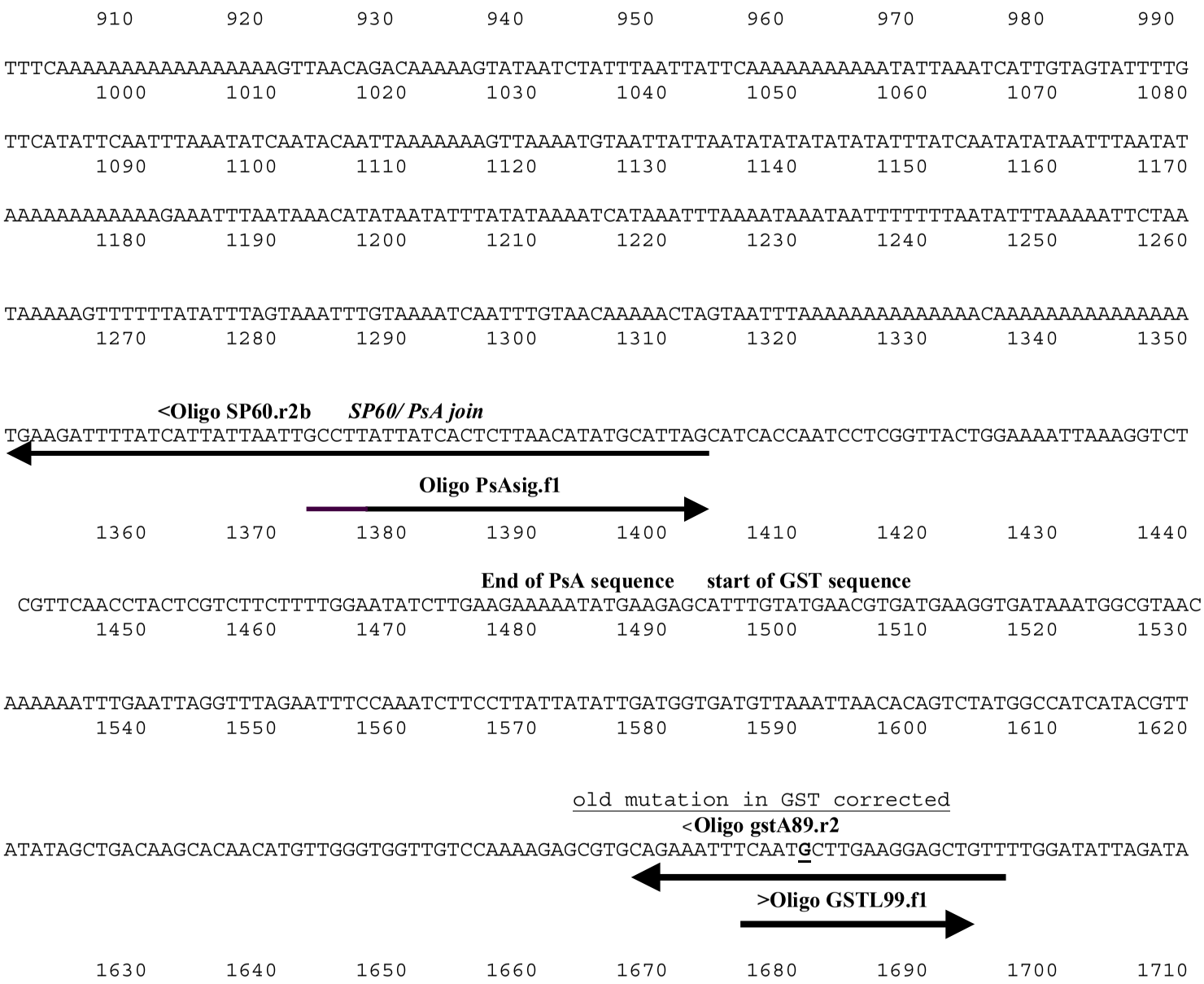
```
CGATAGGCAATTTATTTTTATATCTATCTAAAAAAAAACTAGGAAAAATGAATGTCATCAAATAGTATTTTAACATTTTTTTTTTTTTTT
        10        20        30        40        50        60        70        80        90

TTAAAAAAAAAGTGTCATGACAAAAAAAAAAAAAGTGTCATGACAAAAAAAAAAAAAAAAAAGAGGGGAAAGTAATTATAACTAGGTTAGT
       100       110       120       130       140       150       160       170       180

TTTTTATAATTTTTACATATTTGTTAATAACTTTTAATTTTGAATCATATATGATATTACATCGTCCCGTTGAAAAAAAAAAAAAAAATT
       190       200       210       220       230       240       250       260       270

TTTTTTTTCAAACATTTTCATTTTTTAAAAAATGATATAAAATTTTAAACTAAACTATTTTATTAAATACAAACATATAACTTTATCTTA
       280       290       300       310       320       330       340       350       360

ATCAATTTTTTTGGTTTATACATATTTATGTTCGTACTGAAGTATAGATCGATCTTATTACTAAAGTTTCAAAAGTTTTAAAAAAAATTA
       370       380       390       400       410       420       430       440       450

AAGGGGGTAAATATATAACTTTCTGTTTTTTTTCAATTCTGTCATGACAAAAAAAAAAAAAGTGTCATGACAAAAAAAAAAAAAAAAAAT
       460       470       480       490       500       510       520       530       540
```

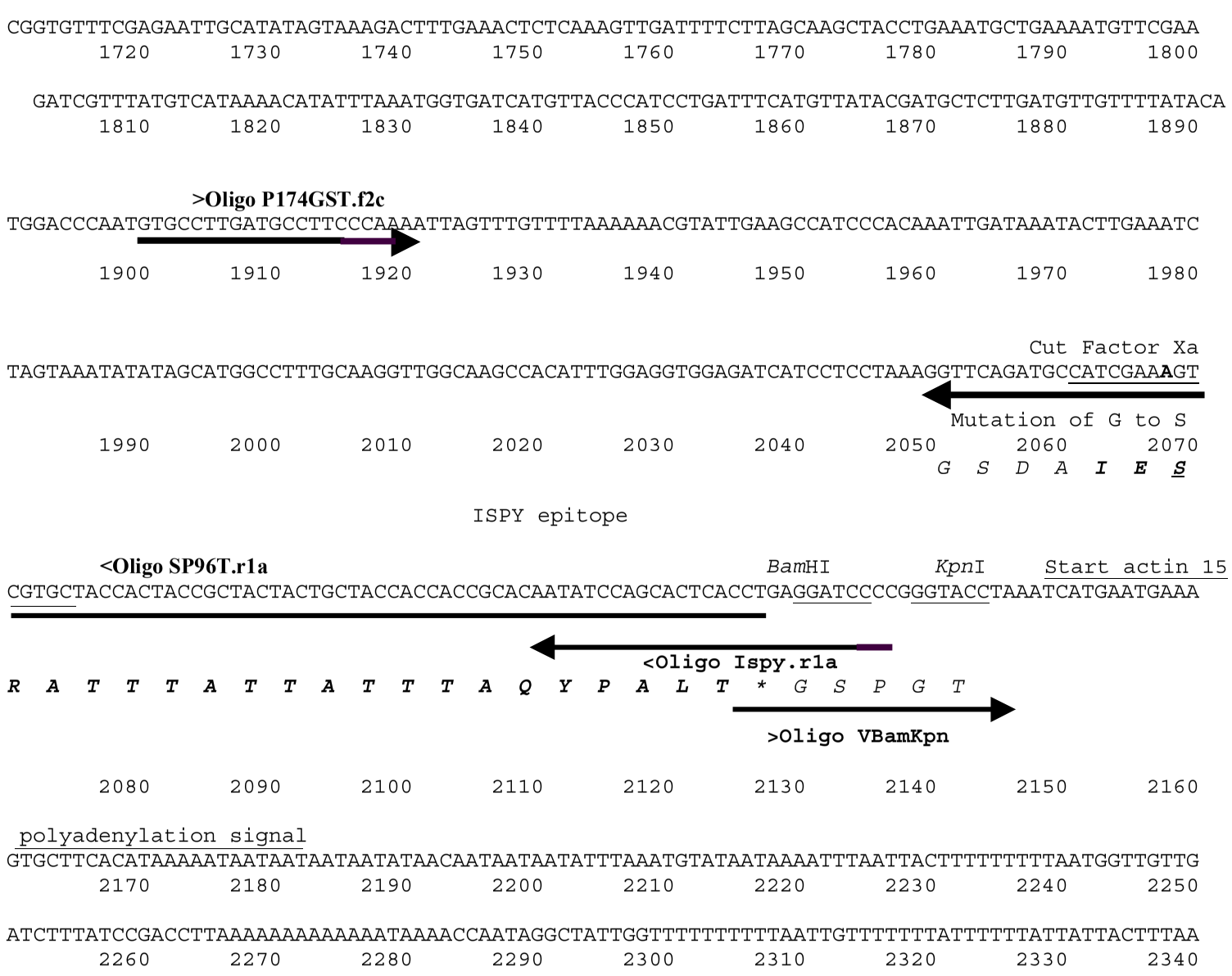
**End of Ddp2 ori**

**>Oligo SP60.f2a**

*XbaI*

```
TTATTTCTTCAATAGGTATTGAAATGACCTCCGTTTTTAATAAAAAGTATATATTTGTGCTCGTCTAGAATCCCATACTACATTAAAATAT
       550       560       570       580       590       600       610       620       630

TTGTATATCAATCAGAAATTTAAATTAAATTGTTTTTTTTTTCACACACCCACACACTAATTTACCCCATTTTTTACATTTCTCATATAAA
▶
       640       650       660       670       680       690       700       710       720

AATTATTTATTGTACATTGTTCAATATTCATTCACACATTAACACACTTTCAACTCAATTATTTTTTTTTTCCAAAAAAACAATAATATATA
       730       740       750       760       770       780       790       800       810

TATACACTGTGAGAATTTTCTATTAATAGCGATAGAAAAAAATTTTATTTTCAAACACACTCCCAACACACAAGCATATGAAAAAACTCAC
       820       830       840       850       860       870       880       890       900

ACCAATTATAATTTGTAAAACATAGAACAAATTTAATAATTATTTTTTATTAAATTGTATTATTTTTTTTTATTATTTATTTATTTTTTAT
```


910 920 930 940 950 960 970 980 990
TTTCAAAAAAAAAAAAAAAAAGTTAACAGACAAAAAGTATAATCTATTTAATTATTCAAAAAAAAAAATATTAAATCATTGTAGTATTTTG
1000 1010 1020 1030 1040 1050 1060 1070 1080
TTCATATTCAATTTAAATATCAATACAATTAAAAAAAGTTAAAATGTAATTATTAATATATATATATTTATCAATATATAATTTAATAT
1090 1100 1110 1120 1130 1140 1150 1160 1170
AAAAAAAAAAAAGAAATTTAATAAACATATAATATTTATATAAAATCATAAATTTAAAATAAATAATTTTTTTAATATTTAAAAATTCTAA
1180 1190 1200 1210 1220 1230 1240 1250 1260
TAAAAAGTTTTTTATATTTAGTAAATTTGTAAAATCAATTTGTAACAAAAACTAGTAATTTAAAAAAAAAAAAAACAAAAAAAAAAAAAA
1270 1280 1290 1300 1310 1320 1330 1340 1350
<Oligo SP60.r2b
SP60/ PsA join
TGAAGATTTTATCATTATTAATTGCCTTATTATCACTCTTAACATATGCATTAGCATCACCAATCCTCGGTTACTGGAAAATTAAAGGTCT
Oligo PsAsig.f1
1360 1370 1380 1390 1400 1410 1420 1430 1440
End of PsA sequence
start of GST sequence
CGTTCAACCTACTCGTCTTCTTTTGGAATATCTTGAAGAAAAATATGAAGAGCATTTGTATGAACGTGATGAAGGTGATAAATGGCGTAAC
1450 1460 1470 1480 1490 1500 1510 1520 1530
AAAAAATTTGAATTAGGTTTAGAATTTCCAAATCTTCCTTATTATATTGATGGTGATGTTAAATTAACACAGTCTATGGCCATCATACGTT
1540 1550 1560 1570 1580 1590 1600 1610 1620
old mutation in GST corrected
<Oligo gstA89.r2
ATATAGCTGACAAGCACAACATGTTGGGTGGTTGTCCAAAAGAGCGTGCAGAAATTTCAATGCTTGAAGGAGCTGTTTTGGATATTAGATA
>Oligo GSTL99.f1
1630 1640 1650 1660 1670 1680 1690 1700 1710


CGGTGTTTCGAGAATTGCATATAGTAAAGACTTTGAAACTCTCAAAGTTGATTTTCTTAGCAAGCTACCTGAAATGCTGAAAATGTTCGAA
1720 1730 1740 1750 1760 1770 1780 1790 1800
GATCGTTTATGTCATAAAACATATTTAAATGGTGATCATGTTACCCATCCTGATTTCATGTTATACGATGCTCTTGATGTTGTTTTATACA
1810 1820 1830 1840 1850 1860 1870 1880 1890
>Oligo P174GST.f2c
TGGACCCAATGTGCCTTGATGCCTTCCCAAAATTAGTTTGTTTTAAAAAACGTATTGAAGCCATCCCACAAATTGATAAATACTTGAAATC
1900 1910 1920 1930 1940 1950 1960 1970 1980
Cut Factor Xa
TAGTAAATATATAGCATGGCCTTTGCAAGGTTGGCAAGCCACATTTGGAGGTGGAGATCATCCTCCTAAAGGTTCAGATGCCATCGAAAGT
Mutation of G to S
1990 2000 2010 2020 2030 2040 2050 2060 2070
G S D A I E S
ISPY epitope
<Oligo SP96T.r1a
BamHI
KpnI
Start actin 15
CGTGCTACCACTACCGCTACTACTGCTACCACCACCGCACAATATCCAGCACTCACCTGAGGATCCCCGGGTACCTAAATCATGAATGAAA
<Oligo Ispy.r1a
R A T T T A T T A T T T A Q Y P A L T * G S P G T
>Oligo VBamKpn
2080 2090 2100 2110 2120 2130 2140 2150 2160
polyadenylation signal
GTGCTTCACATAAAAATAATAATAATATAACAATAATAATATTTAAATGTATAATAAAATTTAATTACTTTTTTTTTAATGGTTGTTG
2170 2180 2190 2200 2210 2220 2230 2240 2250
ATCTTTATCCGACCTTAAAAAAAAAAAAATAAAACCAATAGGCTATTGGTTTTTTTTTTAATTGTTTTTTTATTTTTTATTATTACTTTAA
2260 2270 2280 2290 2300 2310 2320 2330 2340

```
TTATCATTTTTTAAATTACAAAAAAAATTAAAAATCCAGATATTAAGGTATTTGCACTAGTGCTTTAACGTTAAAATTTGAAAAAAAAAAA
      2350      2360      2370      2380      2390      2400      2410      2420      2430

AAATTAATAATTTTACCCTTTATGGGTAAACGATTCTCACATATAATACAATCTCCATGAAAAGATCCAAAGGCCAGCAAAAGGCCAGCAA
      2440      2450      2460      2470      2480      2490      2500      2510      2520

AAGGCCAGGAACCGTAAAAAGGCCGCGTTGCTGGCGTTTTTCGATAGGCTCCGCCCCCCTGACGAGCATCACAAAAATCGACGCTCAAGTC
      2530      2540      2550      2560      2570      2580      2590      2600      2610

AGAGGTGGCGAAACCCGACAGGACTATAAAGATACCAGGCGTTTCCCCCTGGAAGCTCCCTCGTGCGCTCTCCTGTTCCGACCCTGCCGCT
      2620      2630      2640      2650      2660      2670      2680      2690      2700

TACCGGATACCTGTCCGCCTTTCTCCCTTCGGGAAGCGTGGCGCTTTCTCATAGCTCACGCTGTAGGTATCTCAGTTCGGTGTAGGTCGTT
      2710      2720      2730      2740      2750      2760      2770      2780      2790
 GCTCCAAGCTGGGCTGTGTGCACGAACCCCCCGTTCAGCCCGACCGCTGCGCCTTATCCGGTAACTATCGTCTTGAGTCCAACCCGGTAA
      2800      2810      2820      2830      2840      2850      2860      2870      2880

GACACGACTTATCGCCACTGGCAGCAGCCACTGGTAACAGGATTAGCAGAGCGAGGTATGTAGGCGGTGCTACAGAGTTCTTGAAGTGGTG
      2890      2900      2910      2920      2930      2940      2950      2960      2970

GCCTAACTACGGCTACACTAGAAGGACAGTATTTGGTATCTGCGCTCTGCTGAAGCCAGTTACCTTCGGAAAAAGAGTTGGTAGCTCTTGA
      2980      2990      3000      3010      3020      3030      3040      3050      3060

TCCGGCAAACAAACCACCGCTGGTAGCGGTGGTTTTTTTGTTTGCAAGCAGCAGATTACGCGCAGAAAAAAAGGATCTCAAGAAGATCCTT
      3070      3080      3090      3100      3110      3120      3130      3140      3150

TGATCTTTTCTACGGGGTCTGACGCTCAGTGGAACGAAAACTCACGTTAAGGGATTTTGGTCATGAGATTATCAAAAAGGATCTTCACCTA
      3160      3170      3180      3190      3200      3210      3220      3230      3240

GATCCTTTTAAATTAAAAATGAAGTTTTAAATCAATCTAAAGTATATATGAGTAAACTTGGTCTGACAGTTACCAATGCTTAATCAGTGAG
      3250      3260      3270      3280      3290      3300      3310      3320      3330

GCACCTATCTCAGCGATCTGTCTATTTCGTTCATCCATAGTTGCCTGACTCCCCGTCGTGTAGATAACTACGATACGGGAGGGCTTACCAT
      3340      3350      3360      3370      3380      3390      3400      3410      3420

CTGGCCCCAGTGCTGCAATGATACCGCGAGACCCACGCTCACCGGCTCCAGATTTATCAGCAATAAACCAGCCAGCCGGAAGGGCCGAGCG
      3430      3440      3450      3460      3470      3480      3490      3500      3510

TCAGAAGTGGTCCTGCAACTTTATCCGCCTCCATCCAGTCTATTAATGTTGCCGGGAAGCTAGAGTAAGTAGTTCGCCAGTTAATAGTTTG
      3520      3530      3540      3550      3560      3570      3580      3590      3600
```

```
CGCAACGTTGTTGGCATTGCTACAGGCATCGTGGTGTCACGCTCGTCGTTTGGTATGGCTTCATTCAGCTCCGGTTCCCAACGATCAAGGC
       3610      3620      3630      3640      3650      3660      3670      3680      3690

GAGTTACATGATCCCCCATGTTGTGCAAAAAAGCGGTTAGCTCCTTCGGTCCTCCGATCGTTGTCAGAAGTAAGTTGGCCGCAGTGTTATC
       3700      3710      3720      3730      3740      3750      3760      3770      3780

ACTCATGGTTATGGCAGCACTGCATAATTCTCTTACTGTCATGCCATCCGTAAGATGCTTTTCTGTGACTGGTGAGTACTCAACCAAGTCA
       3790      3800      3810      3820      3830      3840      3850      3860      3870

TTCTGAGAATAGTGTATGCGGCGACCGAGTTGCTCTTGCCCGGCGTCAATACGGGATAATACCGCGCCACATAGCAGAACTTTAAAAGTGC
       3880      3890      3900      3910      3920      3930      3940      3950      3960

TCATCATTGGAAAACGTTCTTCGGGGCGAAAACTCTCAAGGATCTTACCGCTGTTGAGATCCAGTTCGATGTAACCCACTCGTGCACCCAA
       3970      3980      3990      4000      4010      4020      4030      4040      4050

CTGATCTTCAGCATCTTTTACTTTCACCAGCGTTTCTGGGTGAGCAAAAACAGGAAGGCAAAATGCCGCAAAAAAGGGAATAAGGGCGACA
       4060      4070      4080      4090      4100      4110      4120      4130      4140

CGGAAATGTTGAATACTCATACTCTTCCTTTTTCAATATTATTGAAGCATTTATCAGGGTTATTGTCTCATGAGCGGATACATATTTGAAT
       4150      4160      4170      4180      4190      4200      4210      4220      4230

GTATTTAGAAAAATAAACAAATAGGGGTTCCGCGCACATTTCCCCGAAAAGTGCCACCTGACGTCTAAGAAACCATTATTATCAT
       4240      4250      4260      4270      4280      4290      4300
```

# Curriculum vitae

**Personal data**

| | |
|---|---|
| Name | : Marcus Mreyen |
| Date of birth | : 27.09.1969 |
| Place of birth | : Essen |
| Parents | : Manfred and Edeltraud Mreyen née Rurainski |

**School Education**

| | |
|---|---|
| Primary school | : 1975 to 1979 |
| Secondary school | : 1979 to 1986 |

**Occupational career**

| | |
|---|---|
| Apprenticeship | : 1986 to 1989 (Chemielaborant an der Universität Dortmund) |
| Matriculation | : 1989 to 1990 (Fachoberschule) |
| Study | : 1990 to 1996 (Studium an der Bergische Universität Wuppertal) |
| degree | : 02.12.1996 Diplom Chemiker<br>(The practical part of the thesis (Diplomarbeit) was done at the Macquarie University and at the Macquarie University Centre for Analytical Biotechnology (MUCAB), Sydney, NSW 2109, Australia) |
| Studying for a doctorate | : 1997 to present (Bergische Universität Wuppertal)<br>(Practical parts of the PhD thesis were mostly done at the Macquarie University, the Australian Proteome Analysis Facility (APAF) and Proteome Systems Limited (PSL), Sydney, NSW, Australia) |